

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE
PROVOZNĚ EKONOMICKÁ FAKULTA

MOŽNOSTI NEURONOVÝCH SÍTÍ

disertační práce

Autor: Mgr. Roman Biskup
Školitel: prof. RNDr. Anna Čermáková, CSc.

Praha 2009

Na tomto místě bych rád poděkoval především své školitelce prof. RNDr. Anně Čermákové, CSc. za její připomínky vznesené v průběhu celé práce jakožto i za vedení v průběhu mého doktorského studia.

Dále bych rád poděkoval členům-kolegům z Katedry aplikované matematiky a informatiky Ekonomické fakulty Jihočeské univerzity v Českých Budějovicích a členům Katedry systémového inženýrství Provozně ekonomické fakulty České zemědělské univerzity v Praze za vytvoření pracovního prostředí a trpělivost. Jmenovitě bych rád poděkoval za podporu a rady doc. RNDr. Evě Vaněčkové, CSc., doc. Ing. Tomáši Šubrtovi, Ph.D., Ing. Janě Friebelové, Ph.D., PhDr. Marku Šulistovi, Ing. Michaelu Rostovi, Ph.D., Ing. Martině Houškové Beránkové, Ph.D., Ing. Milanovi Houškovi, Ph.D. a Ing. Petru Veselkovi.

Mé poděkování bych chtěl také směřovat na adresu grantové agentury Zemědělské fakulty Jihočeské univerzity v Českých Budějovicích, díky níž jsem z interního grantu (14/06) „*Možnosti neuronových sítí a data miningu pro hodnocení faktorů harmonizace regionálního rozvoje sociálně ekonomického potenciálu regionů*“ mohl financovat nákup softwaru „*STATISTICA Neuronové sítě 7.1*“ využitého pro realizaci výpočtu v této práci.

Děkuji také všem ze svého okolí, kteří mě při práci morálně podporovali a omlouvám se jim za ústrky i opomíjení, kterých se jim od mě v souvislosti s mým doktorským studiem dostalo.

Obsah

1	Úvod	14
2	Cíl a metodika	18
2.1	Cíle disertační práce	18
2.2	Metodika a prostředky dosažení cílů disertační práce	19
I	Teoretická část	21
3	Úvod do problematiky neuronových sítí	23
3.1	Motivace pro využívání neuronových sítí	23
3.2	Biologická motivace	26
3.3	K dějinám neuronových sítí	32
4	Modely neuronových sítí	37
4.1	Stavební jednotky sítě	37
4.2	Neuronová síť	41
4.3	Kolekce neuronových sítí	46
5	Učení neuronových sítí	49
5.1	Klasifikace učení dle vztahu ke znalosti požadovaného výstupu	50
5.2	Klasifikace učení dle principu modifikace vah	51
5.3	Backpropagation	52
5.4	Genetické učení	57
5.5	Přetrénování a přeúčnění neuronové sítě	61
6	Softwarová prostředí a neuronové sítě	62
6.1	STATISTICA Neuronové sítě	62
6.2	Neurex	67

6.3	NetVisualiser	70
6.4	Další programy	71
7	Konkrétní příklady využití neuronových sítí	73
7.1	Prohledávané databáze	73
7.2	Konkrétní studované využití neuronových sítí	75
8	Expertní systémy	78
8.1	Motivace a historie	78
8.2	Tvorba expertního systému	79
8.3	Klasifikace expertních systémů dle způsobu jejich tvorby	81
II	Materiál a metody	84
9	Materiál a metody	85
9.1	Materiál	85
9.2	Metody	86
III	Aplikační část	88
10	Diskriminace a neuronové sítě	89
10.1	Motivace	89
10.2	Soubor dat	89
10.3	Klasická statistická analýza – diskriminační analýza	91
10.4	Diskriminace prostřednictvím neuronových sítí	92
10.5	Závěr	120
11	Expertní systém a neuronové sítě	123
11.1	Motivace	123
11.2	Soubor dat	125
11.3	Diskriminace prostřednictvím neuronových sítí	129
11.4	Závěr	136
11.5	Expertní systém pro poskytovatele	140
12	Časové řady a neuronové sítě	144
12.1	Motivace	144

12.2	Soubor dat	145
12.3	Analýza časové řady	146
12.4	Modelování a predikce v časových řadách	150
12.5	Exponenciální vyrovnaní	151
12.6	Neuronové sítě	152
12.7	Přetrénování na aktuálních datech	160
12.8	Závěr	164
13	Závěr	168
14	Použité zdroje	171
15	Vlastní publikace	183
A	Soubory dat	185
A.1	Migrace	185
A.2	Splátkový prodej	202
A.3	Nezaměstnanost	221
B	Grafické opory	225
B.1	Nezaměstnanost	225

Seznam tabulek

3.1	Sumarizace rozdílů práce počítačů a nervových sítí	25
3.2	Srovnání mozek – počítač	26
4.1	Klasifikovatelná množina vícevrstevným perceptronem	45
5.1	Terminologie genetických algoritmů převzatá z biologické motivace . .	58
7.1	Výkonnost srovnávaných metod (týdenní predikce)	77
10.1	Výsledky diskriminační funkční analýzy – migrace	91
10.2	Klasifikační matice (diskriminační analýza) – migrace	92
10.3	Rozdělení souboru dat pro trénování neuronové sítě	94
10.4	Výkonnost trénovacích metod použitelných pro trénování dopředné sítě 7:7-10-10-10-1:1	97
10.5	Výkonnost různých architektur dopředné neuronové sítě trénované metodou rychlého šíření chyby	99
10.6	Výkonnost neuronové sítě 7:7-10-10-10-1:1 pro metodu rychlého šíření chyby v závislosti na počtu epoch	100
10.7	Výsledky citlivostní analýzy vícevrstevných dopředných sítí 1–12	102
10.8	Výkonnost neuronové sítě 7:7-30-30-30-1:1 (VVP 7) opakovaně trénované	103
10.9	Výsledky citlivostní analýzy vícevrstevné dopředných sítí 7:7-30-30-30-1:1 (VVP 7) opakovaně trénované	104
10.10	Výkonnost trénování pravděpodobnostní neuronové sítě 7:7-211(621)-2:1 v závislosti na volbě vyhlazovacího faktoru	106
10.11	Výsledky citlivostní analýzy pravděpodobnostních sítí 7:7-211-2:1 v závislosti na volbě vyhlazovacího faktoru	109
10.12	Výkonnost trénování RBF sítě v závislosti na počtu semilokálních jednotek	110

10.13	Výkonnost trénování neuronových sítí vybraných heuristikou – 50 nejlepších	112
10.14	Výkonnost trénování neuronových sítí testovaných heuristikou – 1 000 sítí	112
10.15	Celková výkonnost (Celk.1) neuronových sítí vybraných heuristikou .	116
10.16	Výsledky citlivostní analýzy neuronových sítí vybraných heuristikou .	117
10.17	Výkonnost sítí realizující degenerované řešení	118
10.18	Klasifikační matice – neuronové sítě vytvořené „manuálně“	119
10.19	Klasifikační matice – neuronové sítě vytvořené heuristikou	120
11.1	Průměry a relativní četnosti ve třech skupinách – splátkový prodej . .	127
11.2	Testy významnosti Mahalanobisových vzdáleností	128
11.3	Průměry a relativní četnosti ve dvou skupinách – splátkový prodej . .	129
11.4	Výkonnost trénování neuronových sítí vytipovaných parametrů	135
12.1	Vývoj míry nezaměstnanosti v letech 2004–2009	147
12.2	Sezónnost v modelech konstantní a proporcionální sezónnosti	148
12.3	Vyrovňovací konstanty, hodnoty <i>M.A.E.</i> a Theilův koeficient nesouladu pro nejlepší modely	151
12.4	Hodnoty <i>M.A.E.</i> , Theilova koeficientu nesouladu a trénovací metody pro vybrané neuronové sítě	154
12.5	Výkonnost vybraných neuronových sítí pro modelování míry nezaměstnanosti 2004–2009	156
12.6	Variabilita predikcí v závislosti na počtu neuronů ve skryté vrstvě . .	158
12.7	Výkonnost vybraných neuronových sítí pro modelování trendu míry nezaměstnanosti 2004–2009	159
12.8	Variabilita predikcí v závislosti na počtu neuronů ve skryté vrstvě . .	159
12.9	Vyrovňovací konstanty a hodnoty <i>M.A.E.</i> pro zaktualizované modely exponenciálního vyrovnání	161
12.10	Výkonnost vybraných modelů pro jejichž tvorbu byly použity i hodnoty březnové a dubnové míry nezaměstnanosti pro modelování trendu míry nezaměstnanosti 2004–2009	163
12.11	Variabilita predikcí v závislosti na počtu neuronů ve skryté vrstvě . .	163
12.12	Predikce míry nezaměstnanosti prostřednictvím použitých modelů – exponenciální vyrovnání	165
12.13	Predikce míry nezaměstnanosti prostřednictvím použitých modelů – neuronové sítě	166

A.1	Soubor dat použitý pro diskriminaci dle migrace v Jihočeském regionu	186
A.2	Soubor dat použitý pro diskriminaci dle výsledků splácení spotřebitel- ských úvěrů	202
A.3	Vývoj míry nezaměstnanosti v letech 1993–2009 a hodnoty sezónně očistěné	221

Seznam obrázků

3.1	Schématické znázornění neuronu – izolovaný neuron	27
3.2	Schématické znázornění neuronu – zapojení v síti	29
4.1	Schématické znázornění formálního neuronu	38
4.2	Nejčastěji používané aktivační funkce	39
4.3	Schéma RBF neuronové sítě 7:7-83-1:1	42
4.4	Rozdělení neuronových sítí	43
4.5	Schéma dopředné neuronové sítě 6:8-10-5-1:1	45
8.1	Struktura procesu rozhodování	78
10.1	Schéma třívrstvé dopředné neuronové sítě 7:7-10-10-10-1:1	96
10.2	Průběh chyby učení pro trénovací a validační množinu	99
10.3	ROC křivky pro trénovací a celkovou výkonnost trénovacích algoritmů	101
10.4	Schéma pravděpodobnostní neuronové sítě 7:7-211(621)-2:1	105
10.5	Krabicové grafy – závislost výkonnosti jednotlivých typů sítí (50 nej- lepších sítí vybraných heuristikou)	114
10.6	Schéma pravděpodobnostní (Bayesovské) neuronové sítě 7:7-311-2-2:1	115
11.1	Schéma vstupu neuronové sítě 6:8-...	130
12.1	Vývoj míry nezaměstnanosti v České republice a ve vybraných okre- sech ČR v období leden 1999–duben 2009	146
12.2	Očištěná časová řada míry nezaměstnanosti o sezónní složku leden 2003– duben 2009	148
12.3	Modely vývoje míry nezaměstnanosti v České republice od října 2008 do prosince 2009 (včetně predikce)	152
12.4	Schéma vstupu neuronové sítě 1:12-... (časové řady)	153

12.5	Pedikce vývoje míry nezaměstnnosti v České republice do prosince 2009 – vícevrstvá dopředná síť – VVP 1:12-...-1:1	156
12.6	Predikce vývoje míry nezaměstnnosti v České republice do prosince 2009 – vícevrstvá dopředná síť – VVP 1:12-...-1:1 (T) (trend) + mechanické vyrovnání (sezónnost)	160
12.7	Zaktualizované modely vývoje míry nezaměstnnosti v České republice od října 2008 do prosince 2009 (včetně predikce)	162
12.8	Zaktualizované modely vývoje míry nezaměstnnosti v České republice od října 2008 do prosince 2009 (včetně predikce)	164
B.1	Predikce vývoje míry nezaměstnnosti v České republice prostřednic- tvím 80 vícevrstvých dopředných sítí VVP 1:12-1-1:1 – leden 2004– prosinec 2009	226
B.2	Predikce vývoje míry nezaměstnnosti v České republice prostřednic- tvím kolekce vytvořené z 80 vícevrstvých dopředných sítí VVP 1:12-1- 1:1 – leden 2004–prosinec 2009	226
B.3	Predikce vývoje míry nezaměstnnosti v České republice prostřednic- tvím 80 vícevrstvých dopředných sítí VVP 1:12-2-1:1 – leden 2004– prosinec 2009	227
B.4	Predikce vývoje míry nezaměstnnosti v České republice prostřednic- tvím kolekce vytvořené z 80 vícevrstvých dopředných sítí VVP 1:12-2- 1:1 – leden 2004–prosinec 2009	227
B.5	Predikce vývoje míry nezaměstnnosti v České republice prostřednic- tvím 80 vícevrstvých dopředných sítí VVP 1:12-3-1:1 – leden 2004– prosinec 2009	228
B.6	Predikce vývoje míry nezaměstnnosti v České republice prostřednic- tvím kolekce vytvořené z 80 vícevrstvých dopředných sítí VVP 1:12-3- 1:1 – leden 2004–prosinec 2009	228
B.7	Predikce vývoje míry nezaměstnnosti v České republice prostřednic- tvím 80 vícevrstvých dopředných sítí VVP 1:12-5-1:1 – leden 2004– prosinec 2009	229
B.8	Predikce vývoje míry nezaměstnnosti v České republice prostřednic- tvím kolekce vytvořené z 80 vícevrstvých dopředných sítí VVP 1:12-5- 1:1 – leden 2004–prosinec 2009	229

B.9	Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím 80 vícevrstevných dopředných sítí VVP 1:12-8-1:1 – leden 2004–prosinec 2009	230
B.10	Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím kolekce vytvořené z 80 vícevrstevných dopředných sítí VVP 1:12-8-1:1 – leden 2004–prosinec 2009	230
B.11	Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím 80 vícevrstevných dopředných sítí VVP 1:12-15-1:1 – leden 2004–prosinec 2009	231
B.12	Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím kolekce vytvořené z 80 vícevrstevných dopředných sítí VVP 1:12-15-1:1 – leden 2004–prosinec 2009	231
B.13	Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím 80 vícevrstevných dopředných sítí VVP 1:12-3-2-1:1 – leden 2004–prosinec 2009	232
B.14	Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím kolekce vytvořené z 80 vícevrstevných dopředných sítí VVP 1:12-3-2-1:1 – leden 2004–prosinec 2009	232
B.15	Predikce trendu vývoje míry nezaměstnanosti v České republice prostřednictvím 80 vícevrstevných dopředných sítí VVP 1:12-1-1:1 (T) – leden 2004–prosinec 2009	233
B.16	Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím kolekce vytvořené z 80 vícevrstevných dopředných sítí VVP 1:12-1-1:1 (T) – leden 2004–prosinec 2009	233
B.17	Predikce vývoje trendu míry nezaměstnanosti v České republice prostřednictvím 80 vícevrstevných dopředných sítí VVP 1:12-2-1:1 (T) – leden 2004–prosinec 2009	234
B.18	Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím kolekce vytvořené z 80 vícevrstevných dopředných sítí VVP 1:12-2-1:1 (T) – leden 2004–prosinec 2009	234
B.19	Predikce vývoje trendu míry nezaměstnanosti v České republice prostřednictvím 80 vícevrstevných dopředných sítí VVP 1:12-3-1:1 (T) – leden 2004–prosinec 2009	235
B.20	Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím kolekce vytvořené z 80 vícevrstevných dopředných sítí VVP 1:12-3-1:1 (T) – leden 2004–prosinec 2009	235

B.21	Predikce vývoje trendu míry nezaměstnanosti v České republice prostřednictvím 80 vícevrstvých dopředných sítí VVP 1:12-5-1:1 (T) – leden 2004–prosinec 2009	236
B.22	Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím kolekce vytvořené z 80 vícevrstvých dopředných sítí VVP 1:12-5-1:1 (T) – leden 2004–prosinec 2009	236
B.23	Predikce vývoje trendu míry nezaměstnanosti v České republice prostřednictvím 80 vícevrstvých dopředných sítí VVP 1:12-8-1:1 (T) – leden 2004–prosinec 2009	237
B.24	Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím kolekce vytvořené z 80 vícevrstvých dopředných sítí VVP 1:12-8-1:1 (T) – leden 2004–prosinec 2009	237
B.25	Predikce vývoje trendu míry nezaměstnanosti v České republice prostřednictvím 80 vícevrstvých dopředných sítí VVP 1:12-15-1:1 (T) – leden 2004–prosinec 2009	238
B.26	Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím kolekce vytvořené z 80 vícevrstvých dopředných sítí VVP 1:12-15-1:1 (T) – leden 2004–prosinec 2009	238
B.27	Predikce vývoje trendu míry nezaměstnanosti v České republice prostřednictvím 80 vícevrstvých dopředných sítí VVP 1:12-3-2-1:1 (T) – leden 2004–prosinec 2009	239
B.28	Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím kolekce vytvořené z 80 vícevrstvých dopředných sítí VVP 1:12-3-2-1:1 (T) – leden 2004–prosinec 2009	239

Kapitola 1

Úvod

V současné době existuje nespočet statistických metod zabývajících se vytěžováním dat, ať již za účelem predikce, klasifikace, či jiným. Všechny tyto metody jsou použitelné za splnění určitých předpokladů. Tento fakt je svým způsobem svazující a vymezuje prostor použitelnosti jednotlivých metod. Jako alternativa k těmto metodám se jeví využití umělých¹ neuronových sítí. Ty jsou do jisté míry ušetřeny problémů spjatých s ověřováním předpokladů. Neuronové sítě nepracují na základě modelů apriori založených na matematických modelech platných za již zmíněných podmínek, u nichž je navíc velmi podstatné nastavení parametrů modelu.

Práce neuronových sítí není striktně algoritmická a je tedy v podstatě odlišná od procesů pracujících na základě algoritmů klasických metod. Umělá neuronová síť kopíruje schopnost živého organismu učit se z příkladů a řešit úlohy, jež klasický počítač není schopen řešit pomocí algoritmů, ať už je to z důvodu chybějícího analytického popisu, nebo z důvodu příliš složité analýzy. Vždy není třeba znát algoritmus. Stačí mít jen dostatečnou množinu dat, která se předloží neuronové síti a ona sama nastaví parametry tak, aby co nejlépe odpovídaly zákonitostem v této množině. Při tomto procesu tvorby si neuronová síť může nejen zapamatovat všechny vzorové příklady, ale ještě navíc zobecní jejich zákonitosti (což je jistě výhoda). Toto zobecnění pak umožní řešit jiné, pro neuronovou síť nové a neznámé případy na základě podobnosti, čehož lze výhodně využít k predikci či ke klasifikaci dat.

¹Přestože například v anglické literatuře je běžné používání plného názvu umělé neuronové sítě (Artificial Neural Networks), respektive akronymu ANN, česká odborná literatura uvádí pouze termín neuronové sítě. Z kontextu by však vždy mělo být zřejmé, zda-li se jedná o umělé neuronové sítě nebo jejich vzory – biologické neuronové sítě. Na některých místech práce je adjektivum „umělé“ použito pro zdůraznění toho, že se jedná o matematické modely atp. (viz např. podkapitola 3.1). Nejedná se tedy o nedůslednost.

Jak uvádí například StatSoft [81], technologii neuronových sítí lze použít prakticky v každé situaci, v níž je cílem nalezení neznámé proměnné nebo vlastnosti na základě známých pozorování nebo naměřených hodnot (tzn. v nejrůznějších typech regresí, klasifikací a časových řad). Je-li k dispozici dostatečné množství „historických“ dat a existují-li mezi nimi objektivní vztahy nebo množina vztahů (neuronové sítě jsou relativně tolerantní k šumovým jevům) existuje také reálná šance, že je bude možno prostřednictvím neuronových sítí modelovat.

- **Automatické řízení, kybernetika, umělá inteligence, optimalizace a expertní systémy v této oblasti**

- řízení procesů (např. monitorování strojního zařízení v průmyslových procesech a průběžné nastavování řídicích parametrů),
- samoobslužné mechanismy subsystémů (např. řízení skleníků, budov, dopravní signalizace),
- „samoučící se“ zahradní a zemědělská technika operující na omezeném území,
- rozpoznávání poruch strojů (např. podle zvuku motorů, podle vlivu vibrací nebo akustických zátěží) a následné plánování preventivní údržby,
- provozování strojů (např. odhady spotřeby paliv na základě snímaných měření),
- plánování využití prostoru,
- strategie maximalizace výkrmu hospodářských zvířat,
- rozpoznávání kvality výrobku, ...

- **Zpracování obrazu a řeči**

- optické rozpoznávání textů, písma, obrazů a podpisů (např. určování pravosti podpisu nejen podle celkového vzhledu písma, ale i rychlosti pohybu psacího náčiní při podpisu),
- převod ručně psaných textů do elektronické podoby,
- zpracování obrazu (např. sledování obrazů z londýnských stanic metra, pro rozpoznání toho, kdy je stanice zaplněná, prázdná nebo poloprázdná bez ohledu na světelné podmínky či přítomnost-nepřítomnost vlakové soupravy ve stanici),
- převod mluvené řeči do písemné podoby, odezírání ze rtů,
- syntéza řeči (např. Nettetalk, který se učil vytvářet fonémy na základě psaného textu),
- jazyková analýza (např. identifikace klíčových frází, slov atd. v domorodých jihoamerických jazycích), ...

- **Lékařství (diagnostika)**

- vyhodnocování lékařských snímků (např. odhad velikosti tumoru prostaty),
- detekce a ohodnocení v medicíně (např. detekce epileptických záchvatů),
- chemická diagnostika a analýza vzorků, ...

- **Ekonomika, marketing a expertní systémy v této oblasti**

- předpověď finančních časových řad (i chaotických) – predikce vývoje směnných kurzů, cenných papírů (např. společnost LBS Capital Management tvrdí, že se její obchodní výsledky podstatně zlepšily použitím vícevrstvé sítě perceptronů pro předpovědi cen akcií),
- předpověď reakce spotřebitelů na nový druh zboží,
- úvěrová rizika (např. řešení klasického problému – na základě dotazníku rozhodnout, zda je uchazeč o úvěr vhodnou osobou z hlediska úvěrového rizika),
- rozpoznání nevhodného manažeringu firmy, rozpoznávání finančního zdraví podniku,
- výběr skupiny zákazníků pro hromadnou korespondenci (např. identifikace zákazníků, u kterých je nejvyšší pravděpodobnost příznivé reakce na hromadnou korespondenci na základě informací z existující databáze zákazníků), ...

- **Kompres dat, transformace a analýza signálu**

- komprese obrázků,
- diagnostika výsledků (např. rozpoznávání odběrových diagramů, koronových výbojů), ...

Protože se dá očekávat že neuronové sítě nemusí být samospásným prostředkem pro řešení problémů, při kterých i klasické statistické metody nedosahují vysoce uspokojivých výsledků, je třeba hledat řešení, jak jejich validní výkonnost zvýšit. Takovým řešením jsou například kolekce sítí, které dokáží konstruovat modely, které dávají nestranné odhady s možností odhadu stability/validity těchto modelů i v případě omezené množiny vzorů použitých v procesu učení.

Disertační práce se zaměřuje na problematiku možností využití neuronových sítí. Jejím cílem je prozkoumání možností uplatnění neuronových sítí a jejich kolekcí pro analýzu ekonomických dat a porovnání výkonnosti i způsobu práce neuronových sítí s klasickými statistickými metodami. Konkrétně se jedná o prozkoumání možností uplatnění neuronových sítí pro diskriminaci, tvorbu expertního systému a predikci v časových řadách. Popisované řešené problémy nejsou pouze ilustrativní, naopak

výsledky z nich vzešlé byly využity při řešení grantů a projektů. Předkládaná disertační práce se také zabývá potvrzením respektive zpochybněním obecně přijímaných paradigmat o způsob práce neuronových sítí na reálných datech.

Kapitola 2

Cíl a metodika

2.1 Cíle disertační práce

Oblast neuronových sítí je velice široká a možnosti jejich uplatnění ještě širší. Disertační práce proto nemůže být zaměřena na všechny aspekty, které se k neuronovým sítím vztahují. Hlavní cíl práce, kterým je prozkoumání možností uplatnění neuronových sítí a jejich kolekcí pro analýzu ekonomických dat, bude konkretizován pomocí následujících dílčích cílů:

1. **Navržení expertního systému založeného na neuronovém učení**

Vytvořit expertní systém založený na neuronové síti použitelný jako nástroj pro kvalifikované rozhodování. Takto vytvořený expertní systém založit na možnosti neuronových sítí učit na základě datové množiny.

2. **Vytvoření a použití metodiky pro použití vybraných typů neuronových sítí a jejich kolekcí k řešení konkrétních problémů při analýze ekonomických dat**

Na základě vytvořené metodiky použít neuronové sítě na reálných datech pro diskriminaci, tvorbu expertního systému a predikci v časových řadách.

3. **Porovnání výkonnosti neuronových sítí a jejich kolekcí problémů s výkonností klasických metod při řešení konkrétních problémů**

Prostřednictvím vhodných ukazatelů porovnat výkonnost neuronových sítí a klasických statistických metod v případě diskriminace a predikce. Důraz přitom klást na validitu řešení získaného neuronovými sítěmi.

4. **Analyzování práce (klasifikace a predikce) neuronových sítí a jejich kolekcí s ohledem na jejich kritické použití**

Při porovnávání výkonnosti zohledňovat nejen výkonnost, ale také stabilitu a

způsob řešení, které bude prostřednictvím neuronových sítí získáno. Výsledky analyzovat s ohledem na možné přeúčnění a přetrénování neuronových sítí respektive nesystémové řešení.

5. Navržení hybridního přístupu pro modelování v časových řadách

Pro predikce v časových řadách využít hybridní model kombinující klasické metody s technologií neuronových sítí tak, aby bylo možno dosáhnout vysoké stability řešení i pro relativně krátké časové řady.

6. Potvrzení/zpochybnění obecně přijímaných paradigmat o způsobu práce neuronových sítí

V souvislosti s topologiemi neuronových sítí vhodných pro řešení zvolených problémů potvrdit respektive zpochybnit obecně přijímaná paradigmata o tvorbě sítě, o apriorním nastavení parametrů ovlivňujících trénovací proces, o volbě trénovací metody a o zajištění validity řešení (generalizační schopnosti sítě) svázané s problematikou přeúčnění a přetrénování neuronové sítě.

2.2 Metodika a prostředky dosažení cílů disertační práce

Teoretická část práce bude spočívat v implementaci poznatků získaných studiem dostupné české i zahraniční literatury a bude v sobě kombinovat přehled odborné literatury, přehled o současném stavu řešené problematiky a popisu vybraných metod neuronového učení použitých při zpracování disertační práce. V této části budou uvedeny odkazy na nejdůležitější práce, které vývoj oblasti neuronových sítí ovlivnily.

Následně budou vytipovány vhodné specializované softwary s implementovanou neuronovou sítí a proběhne analýza jejich možností s ohledem na budoucí aplikace. Získané softwary budou z hlediska uživatele zevrubně představeny. Následným krokem bude zajištění financování některého z komerčních softwarů s implementovanou neuronovou sítí a odborné zahraniční literatury (podání a získání grantu na vytvořený projekt).

Bude navržen koncept expertního systému založeného na neuronovém učení a proběhne vytvoření prázdného expertního systému pro zvolenou problematiku.

Výběr a analyzování reálných problémů vhodných pro ověření možností neuronových sítí bude uskutečněn s ohledem na studovaný obor a zaměření pracoviště předkladatele disertační práce. Následně budou zajištěny soubory dat použitelné pro trénování neuronových sítí. Primární volba vhodných vstupních a výstupních ukaza-

telů proběhne dle doporučení odborníků.

Aplikační část práce bude spočívat ve využití neuronových sítí pro analýzu reálných problémů a ve srovnání výsledků získaných prostřednictvím neuronových sítí vybraných architektur (topologií) a klasických statistických metod. Součástí aplikace neuronových sítí bude analýza:

1. procesu trénování s ohledem na faktory, které proces trénování ovlivňují,
2. výsledků získaných prostřednictvím natrénovaných neuronových sítí různých architektur respektive topologií,
3. výkonnosti neuronových sítí prostřednictvím validačních metod,
4. využití kolekcí sítí pro dosažení přesnějších předpovědí.

V závěru (kapitola 13) budou shrnuty dosažené poznatky a zhodnoceno naplnění stanovených cílů.

Použitá symbolika je v souladu s obvyklou matematickou konvencí a v případě možné nejjasnosti je symbolika upřesněna.

Část I

Teoretická část

Teoretická část práce v sobě kombinuje přehled odborné literatury, která je předmětem disertační práce, přehled o současném stavu problematiky, která se zabývá danou problematikou a popisuje vybrané metody použité při zpracování disertační práce.

Nejvíce zdrojů ovlivňujících vývoj neuronových sítí přináší podkapitola 3.3; jedná se jak o zdroje teoretické tak o zdroje popisující implementace, modifikace a aplikace neuronových sítí. Teoretická část přináší i vlastní výsledky, kterými jsou přetransformování vzorců pro metodu zpětného šíření chyby (backpropagation) do sjednocené symboliky, navržení „vhodnostních“ transformací pro výběr jedinců do další generace při genetickém učení neuronových sítí, jakož i popsání využití právě genetických algoritmů pro trénování neuronových sítí (podkapitola 5.4.2) a klasifikaci expertních systémů s ohledem na způsob vkládání znalosti do těchto systémů (podkapitola 8.3). Odborná literatura je citována i při vlastní aplikaci neuronových sítí pro řešení reálných problémů, které přináší i část III.

Konkrétní výsledky uváděné v odborné literatuře, kterých bylo dosaženo s využitím neuronových sítí, přináší zejména kapitola 7. O současném stavu problematiky vypovídá také kapitola 6. Ta přináší informace o softwarových balících implementujících neuronové sítě.

V teoretické části práce jsou citovány i vlastní publikace, prezentující autorův zájem o neuronové sítě jakož i zájem o klasické statistické metody (řazeny podle pořadí vzniku) a to zejména Čermáková – Biskup [115], Biskup – Čermáková [110], Biskup – Čermáková [111], Biskup – Čermáková [109], Biskup – Čermáková [112], Biskup – Čermáková [113], Biskup [108], Čermáková – Biskup [117], Čermáková – Biskup [116], Biskup – Čermáková [114], Šulista et al. [120], Čermáková – Biskup [118] a Čermáková – Biskup [119]. Více lze o zdrojích vytvořených předkladatelem disertační práce nalézt na str. 183.

Kapitola 3

Úvod do problematiky neuronových sítí

Úvod do problematiky neuronových sítí je v rámci třetí kapitoly rozdělen do tří částí. V podkapitole 3.1 jsou popsány výhody práce neuronových sítí ve srovnání s prací klasických počítačů von neumannovské architektury. Podkapitola 3.2 je zaměřena na biologickou motivaci umělých neuronových sítí. Jsou zde zmíněny struktury a principy chování biologických neuronových sítí, které sloužili jako inspirace pro tvůrce umělých neuronových sítí. Na jejich základě jsou následně v kapitole 4 popsány matematické modely neuronu, sítě a učení neuronové sítě. Historický vývoj spjatý s neuronovými sítěmi je naznačen v podkapitole 3.3, v které jsou shrnuty jak světlé okamžiky výzkumu, tak i neplodná, „hluchá“ období, kdy byl výzkum neuronových sítí a jejich učení prakticky zastaven. Podkapitola 3.3 tvoří rámec pro představení tvůrců neuronové koncepce, zásadních prací s tematikou neuronových sítí a institucích zabývajících se problematikou neuronových sítí.

3.1 Motivace pro využívání neuronových sítí

Podkapitola 3.1 navazuje na kapitolu 1 a přináší další motivace pro využívání neuronových sítí prostřednictvím systematického srovnání umělých a biologických neuronových sítí.

3.1.1 Srovnání neuronových sítí a „klasických“ počítačů

Hakl – Holeňa [22] uvádí, že každý počítač lze chápat jako zařízení, které zobrazuje množinu vstupních dat (např. koeficienty lineární soustavy rovnic nebo nerovnic) na množinu dat jiných (např. řešení této soustavy). Nebudou-li na takto pracující počítač kladeny další zužující podmínky, lze konstatovat, že i nervový systém individuálního živočicha je jakýmsi prototypem (každý je jiný) počítače, a mnohou duševní činnost či vegetativní aktivitu živočicha lze považovat za realizaci jakéhosi výpočtu. Mezi těmito dvěma paralelními světy „živých počítačů“ a „klasických počítačů“ existují však zásadní rozdíly. Sumarizaci těch z nich, o nichž se lze domnívat, že mají vztah k principu jejich činnosti, uvádí tabulka 3.1.

Obdobné schéma (tabulka 3.2), zobrazující paralelu mezi umělou neuronovou sítí a mozkem, je zmíněno v Kriesel [41].

Rozdíly uvedené v tabulce 3.2 ospravedlňují názor, že schopnosti živých organismů efektivně zpracovávat informace ze svého okolí (v mnohých ohledech a případech výrazně efektivněji než nejvýkonnější počítače) jsou dány především způsobem komunikace mezi jednotlivými neurony a že tyto neurony lze považovat za jakési elementární, principem shodné stavební kameny nervových center.

Dá se očekávat, jak uvádí Hakl – Holeňa [22], že analýza struktur, které jsou velmi jednoduchým modelem nervových center živočichů (na základě dosud známých poznatků z oblasti anatomie, biochemie a neurologie), může být přínosem jak pro výzkum živých organismů, tak i opačně – pro výzkum a implementaci nových typů vysoce paralelních počítačových architektur, které umožní řešit problémy, pro které na počítačích odvozených od principu Turingova stroje¹ neexistuje efektivní algoritmizace.

Jak uvádí například StatSoft [81] technologii neuronových sítí lze použít prakticky v každé situaci, v níž je cílem nalezení neznámé proměnné nebo vlastnosti na základě známých pozorování nebo naměřených hodnot (tzn. v nejrůznějších typech regresí, klasifikací a časových řad), je-li k dispozici dostatečné množství „historických“ dat a existují-li mezi nimi objektivní vztahy nebo množina vztahů (neuronové sítě jsou relativně tolerantní k šumovým jevům).

¹Turingův stroj je teoretický model počítače, který se skládá z: 1) procesorové jednotky, tvořené konečným automatem, 2) programu ve tvaru pravidel přechodové funkce (tabulka) a 3) potenciálně nekonečné pásky pro zápis mezivýsledků. Využívá se např. pro modelování algoritmů v teorii vyčíslitelnosti. Jedná se o popis idealizovaného mechanismu, který má možnost se podle tabulkou daného předpisu pohybovat po pásce a číst z ní znaky nějaké abecedy nebo je zapisovat.

Tabulka 3.1: Sumarizace rozdílů práce počítačů a nervových sítí

(Zdroj: Hakl – Holeňa [22])

Počítač	Nervová soustava
• Existence centrálního procesoru provádějícího veškerou činnost. • Nízká paralelizace, výjimečně více procesorů (≤ 1000) komunikujících mezi sebou. • Přesná znalost způsobu vyhodnocení a zpracování informace (stroje typu RAM – Random Access Machine, RASP – Random Access Stored Program, Turingův stroj). • Nutnost detailní algoritmizace výpočtů, podložená hlubokým teoretickým pozadím (např. nutnost existence poznatků lineární algebry pro řešení soustav lineárních rovnic). • Velmi vysoká rychlost početních elementárních operací (dnes $\simeq 3$ GHz). • Přesně definovaná architektura procesoru.	• Centrální procesor nebyl v nervových tkáních dosud objeven a nic nepotvrzuje jeho existenci. • Velmi vysoký počet jednodušších (z hlediska přenosu informace) procesorů-neuronů ($\simeq 10^{12}$), velmi vysoká hustota propojení mezi nimi (10^3 pro každý neuron). • Velmi matná představa o činnosti jednotlivých elementů, téměř žádná představa o způsobu komunikace mezi elementy. • Zdá se, že schopnost vyhodnocování různých situací je integrální vlastností nervových systémů bez potřeby pochopení a uvědomění si způsobu zpracování informace. • Pomalý přenos informace (řádu milisekund). • Velký počet lokálních elementů vzájemně propojených, s velkou variabilitou hustoty a způsobu spojení.

3.1.1.1 Odlišnosti v práci neuronových sítí oproti klasickým metodám – přehled

Následující výčet (dle Berka [84] rozšířeno) shrnuje odlišnosti v práci neuronových sítí a klasických nástrojů pro řešení reálných problémů. Bylo by nesprávné neuronové sítě jen vychvalovat, proto je na tomto místě ke každé „dobré“ vlastnosti neuronové

Tabulka 3.2: Kvantitativní a kvalitativní srovnání mezi mozkem a počítačem
(Zdroj: Kriesel [41])

	Mozek	Počítač
Počet výpočetních jednotek	$\approx 10^{11}$	$\approx 10^9$
Typ výpočetní jednotky	neurony	tranzistory
Typ počítání	masivně paralelní	zpravidla sériové
Ukládání dat	asociativní	adresované
Reakční doba	$\approx 10^{-3} \text{ s}$	$\approx 10^{-9} \text{ s}$
Teoretické zařazení případu	$\approx 10^{13} \text{ s}^{-1}$	$\approx 10^{18} \text{ s}^{-1}$
Skutečné zařazení případu	$\approx 10^{12} \text{ s}^{-1}$	$\approx 10^{10} \text{ s}^{-1}$

sítě uvedena i ta „nepříjemná“. Výhody a nevýhody neuronových sítí spolu souvisí a jsou poplatné práci neuronových sítí daných byť vhodných typů.

- **Neuronové sítě vs. symbolické metody**
 - + Vhodnější nástroj pro numerické atributy
 - Nalezené znalosti jsou obtížně interpretovatelné
- **Neuronové sítě vs. expertní systémy**
 - + Relativně snadná tvorba expertního systému
 - + Možnost práce s neurčitostí
 - Báze znalostí není oddělená od inferenčního mechanismu
 - Obtížné vysvětlování práce systému, netransparentnost
- **Neuronové sítě vs. „klasický“ počítač**
 - + Možnost fungovat i při částečném poškození
 - + Schopnost učit se
 - Není oddělená paměť a procesor
 - Uložená informace není lokalizovatelná

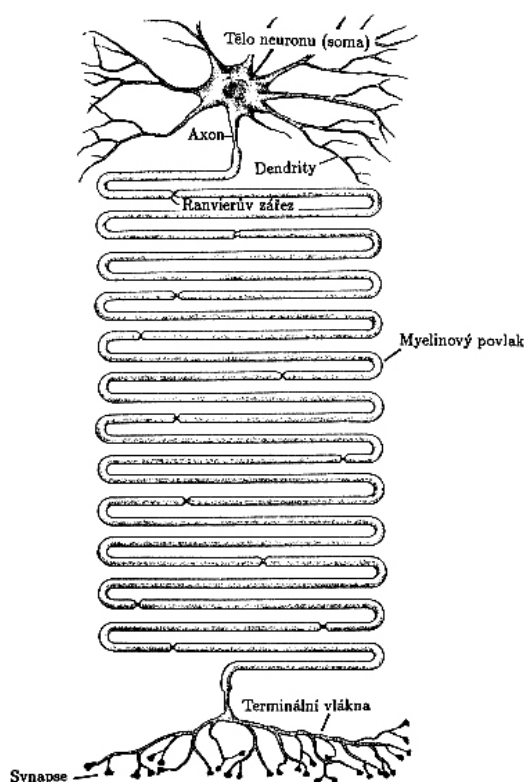
3.2 Biologická motivace

Umělé neuronové sítě jsou jak strukturou tak způsobem práce motivovány biologickými vzory a proto je těmto biologickým vzorům v rámci úvodu do problematiky neuronových sítí věnována podkapitola 3.2.

3.2.1 Neurony jako základní stavební kameny

Neurony jsou živé buňky, které byly donuceny dlouhým vývojem specializovat se na co nejúčelnější *zpracování*, *uchovávání* a *přenos* informací. Jsou základem celé velmi složité struktury neuronové sítě, charakterizované především způsobem a složitostí jejich vzájemného propojení. Kromě toho je však nutno vzít v úvahu, že vnitřní struktura samotných neuronů není o nic jednodušší a že existují oprávněné odhady, že se může blížit složitosti mozku jako takového.

Ne každý neuron je stejný, ale ve všech případech, jak uvádí např. Novák et al. [64], má základní tělo (*soma*), z něhož vychází jediný, i když různým způsobem a v různé vzdálenosti od somatu rozvětvený výstup – *axon*, nebo též *neurit*. Vstup informací z jiných neuronů se může dít buď přímo na soma, nebo též na více či méně rozvětvené vstupní větve – *dendrity*, případně někdy, avšak poměrně zřídka, též přímo na jeho axon (viz obrázek 3.1, na kterém je vidět vlastní tělo neuronové buňky s buněčným jádrem a řadou výběžků rozvětvujících se ve vstupní větve, a obrázek 3.2).



Obrázek 3.1: Schématické znázornění neuronu – izolovaný neuron

(Zdroj: Novák et al. [64, str. 42])

3.2.1.1 Neurony jako součást neuronové sítě

Jednotlivé neurony jsou navzájem propojeny do velmi složitých (*biologických*) *neuronových sítí*. Tedy například mozek je pak možno chápat jako mimořádně komplikovanou soustavu mnoha takových dílčích, byť i tak složitých sítí. Novák et al. [64] uvádí, že na jeden neuron připadá v průměru asi 10 tisíc až 100 tisíc spojení s jinými neurony. To znamená, že celková informační mohutnost neuronových sítí lidského mozku, měřená součinem počtu neuronů a počtu jejich možných vzájemných spojení, dosahuje řádu asi 0,1 až 10 trilionů². Zatím co na jistý neuron může působit větší množství vstupů, je axon u každého neuronu jen jediný, byť též většinou velmi bohatě rozvětvený. Místa spojení jednotlivých neuronů se nazývají *synapse*. I pojem synapse bude mít v dalším výkladu stěžejní význam pro definování tentokrát již celé neuronové sítě. Soudí se (Novák et al. [64]), že informační mohutnost a výkonnost mozku závisí především na složitosti a funkční dokonalosti vzájemného propojení jeho neuronových sítí a na schopnosti jejich adaptivity.

Významnou vlastností neuronů výrazně ovlivňující chování nejen jich samých, ale i celých neuronových sítí, je skutečnost, že reakce somatu každého neuronu na souhrn stimulujících podnětů, přicházejících na vstupy, se projeví až po překročení určité jeho hodnoty – *prahové úrovně* (*biasu*). Soma teprve poté zpracovává vstupní signály do tzv. *vnitřní aktivity neuronu* a ty pak transformuje na výstupní signál.

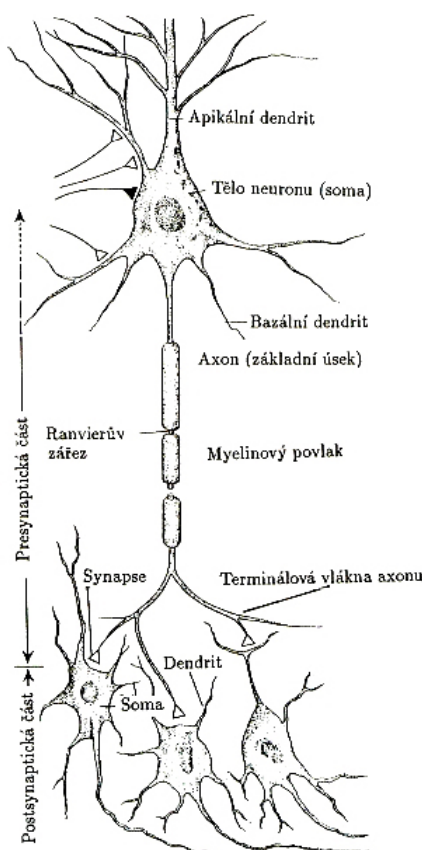
Spojení jednoho konkrétního neuronu s ostatními neurony jsou zobrazeny na obrázku 3.2, kde je zřejmé, že některé spoje jsou realizovány přes dendrity, jiné přímo na tělo neuronu.

Přestože během života organismu dochází obecně k velmi častému nahrazování doživších buněk buňkami novými, u nervových buněk to neplatí. Neurony, které odumrou, nejsou nahrazovány. Novák et al. [64] uvádí, že celkový počet neuronů klesá u člověka denně asi o 10 tisíc, což činí za 75 let života asi 0,2 až 0,5 % z jejich celkového počtu. Naproti tomu synapse vznikají celý život v souvislosti s tím, jak se během života organismus učí novým poznatkům a jak si „zapamatovává“ nové skutečnosti a zážitky.

Počet synapsí vznikajících v průběhu života, jejich velikost a složitost závisí především na průběhu procesu učení organismu novým poznatkům, k němuž nutně v průběhu jeho života dochází. Rozprostření synaptických spojů po celé struktuře mozku je tedy neustále velmi dynamicky se vyvíjející soustavou.

Podkapitola 4.1.1 uvádí, jak lze biologický neuron schématicky modelovat pro

²Pro úplnost, mozek je složen z 20 až 100 miliard neuronů.



Obrázek 3.2: Schématické znázornění neuronu – zapojení v síti

(Zdroj: Novák et al. [64, str. 43])

potřeby implementace neuronové sítě. V této kapitole bude taktéž popsán model funkce neuronu motivovaný skutečnostmi zmíněnými výše.

3.2.2 Neuronové sítě

Za neuronovou síť bude považována taková struktura pro rozprostřené paralelní zpracování informací, která se skládá z jistého (obvykle velmi vysokého) počtu N výkonných prvků (angl. processing elements)³. Každý výkonný prvek neuronové sítě může na svých vstupech přijímat současně libovolný konečný počet různých vstupních informací a na další prvky předávat libovolný konečný počet shodných informací

³Pojem výkonného prvku má dle Novák et al. [64] značný význam. U přirozených neuronových sítí jsou výkonnými prvky jednotlivé biologické neurony, u umělých neuronových sítí pak jejich funkční, matematické, či fyzikální modely.

o stavu svého jediného, avšak velmi rozvětveného výstupu.

Synapse samy působí jako mezineuronová rozhraní uzpůsobená nejen pro funkci spínacího mechanismu, ale též pro vznik paměťových stop. Ve složitých neuronových sítích lidského mozku se uplatňuje několik druhů paměťových mechanismů. Důležitým činitelem při působení synapsí je jejich tzv. průchodnost. Tu je možno vyjádřit v jednoduchém modelu neuronu pomocí váhových činitelů – *synaptických vah* – viz podkapitola 4.1.1. Průchodnost jisté synapse však není konstantní, ale může se velmi značně adaptivně funkčně přizpůsobit. Tato vlastnost synapse se nazývá synaptickou plasticitou. Díky synaptické plasticitě se může při každém šíření signálu průchodnost synapsí měnit, což se děje při procesu zvaném *učení*.

Pro úvahy o chování neuronových sítí jsou zajímavé především takové přístupy, které respektují dynamiku chování neuronů při jejich vzájemném působení a též mechanismus jejich adaptace v reálném čase za stále se měnících vnějších podmínek a okolního prostředí. Jedním z nejpropracovanějších způsobů popisu takových jevů je přístup, který navrhl STEPHEN GROSSEBERG (např. Grossberg [21]). Grossbergův metodický aparát vede, jak uvádí Novák et al. [64], ke dvěma soustavám diferenciálních rovnic popisujících vnitřní aktivitu neuronu a dynamiku synaptických vah.

Biologické neurony, jak bylo zmíněno na začátku této kapitoly, jsou poměrně složité a těžko reprezentovatelné. Za výchozí práci modelování neuronů je všeobecně považován článek McCulloch – Pitts [51], ve kterém autoři MCCULLOCH a PITTS navrhli základní, velmi zjednodušený model neuronu (viz podkapitola 4.1.1 a obrázek 4.1).

3.2.3 Funkce učení

Pro většinu pracovních úseků neuronových sítí je charakteristická schopnost učení, tj. adaptace jejich vlastností podle vnějších okolností, předkládaných vzorů a podle kladených požadavků. Učení se děje podle jistých zákonitostí (někdy též algoritmů – v anglické literatuře je užíván též termín *learning law*). Funkce učení charakterizuje proces nastavení (adaptace) sítě i jejích výkonných prvků podle požadovaných vztahů mezi vstupními a výstupními signály sítě. Změny ve výkonných prvcích jsou obvykle uskutečňovány především prostřednictvím váhových koeficientů jednotlivých vstupních signálů na ně působících. Navíc učení je bezprostředně spjato s pamětí, neboť pro uchování naučených zákonitostí je třeba paměť. Bez ní tedy nelze žádný proces učení uskutečnit, avšak proces učení není možno na ni redukovat.

Jednotlivé postupy a paradigmaty učení lze rozlišovat podle několika hledisek.

1. Prvním, které bere v úvahu Novák et al. [64], je, zda předmětem učení jsou jednotlivá fakta, stimuly, nebo zda jsou to též vztahy mezi nimi a poznatky o nich platící. Podle toho můžeme v zásadě rozlišit dvě základní paradigmaty učení:
 - *Učení neasociativní*, při němž je organismu předkládán jednou či opakovaně jistý jednoduchý stimulus, nebo skupina stimulů bez uvažování jejich vzájemných souvislostí. Úkolem organismu je pak zachovat vhodné z těchto stimulů ve své paměti a podle potřeby si je umět vybavit (většinou opakovaně).
 - *Učení asociativní*, při němž cílem učení je extrakce vztahů existujících mezi daným stimulem a stimulem jiným (nebo jejich celou skupinou), nebo mezi skupinou stimulů navzájem, případně mezi stimuly a chováním organismu.
2. Dalším významným rozlišovacím hlediskem je, zda se učící organismus učí sám, bez cizí pomoci, nebo zda k tomu používá vnější asistenci. Podle toho rozlišujeme:
 - *učení bez učitele*,
 - *učení s učitelem*.

Oba tyto druhy učení jsou široce používány, a to jak při učení živých tvorů, tak při učení umělých neuronových sítí a neuropočítačů na nich založených. Někde mezi oběma uvedenými způsoby učení stojí hodnocené (známkované) učení, u něhož učící se neuronová síť nedostává sice pro své učení k dispozici žádné učební vzory, ale dílčí výsledky procesu učení jsou průběžně hodnoceny (známkovány) tak, aby proces učení vedl k žádoucímu cíli. Hecht-Nielsen [29] charakterizuje tyto způsoby učení jako dosud méně vhodné pro obecné aplikace a vidí jejich užití spíše v oblasti řízení a oblasti optimalizačních úloh. Naopak Kriesel [41] toto učení uvádí jako jeden ze základních typů učení a nazývá jej *posilujícím učením*.

Neuronové sítě se obecně mohou učit změnou vhodného souboru svých parametrů. U umělých neuronových sítí to obvykle bývá změnou přenosových funkcí a synaptických vah, prahových funkcí apod. V úvahu přichází i změna počtu výkonných prvků sítě a její konfigurace.

3.3 K dějinám neuronových sítí

Jako každý přírodovědný obor, má i teorie neuronových sítí za sebou bohatou historii. V Kriesel [41] je historie představena kompaktně jako časový přehled. Stejným způsobem je historie uspořádána v předložené práci. Je však doplněna o další důležité mezníky a slouží jako logický rámec pro představení stěžejních prací, jež ovlivnily vývoj oboru.

Dějiny umělých neuronových sítí mají svůj počátek v časných padesátých letech dvacátého století a to skoro současně s dějinami programovatelných elektronických počítačů⁴.

3.3.1 Počátky

Již 1943 popisují WARREN MCCULLOCH a WALTER PITTS velmi jednoduchý typ neuronové sítě a ukazují, že sama síť tohoto typu dovede vypočítat prakticky každou logickou nebo také aritmetickou funkci. Ve své práci McCulloch – Pitts [51] vytvoří model neuronu, používaný prakticky dodnes.

1947 popisují PITTS a MCCULLOCH praktickou oblast použití neuronové sítě (v jejich práci z roku 1943 nebylo nic takového uvedeno), konkrétně poznávání prostorových vzorů pomocí neuronové sítě.

1949 formuluje DONALD OLDING HEBB v Hebb [27] klasické Hebbovské pravidlo učení⁵, které ve své všeobecné formě představuje podstatu skoro všech způsobů učení neuronových sítí.

1950 zastává neuropsycholog KARL LASHLEY tezi, že paměťové ukládání informací v mozku se realizuje *distribuovaně (rovnoměrně rozděleně)*. Zdůvodňuje svoje teze na pokusech s krysami, u kterých pouze velikost, nikoliv místo zničené nervové tkáně, ovlivňuje jejich výkonnost – nalézt cestu bludištěm.

⁴Vzhledem k tomu, že se jedná o mladý vědní obor, tak mnoho osob spojených dokonce i s počátky tohoto oboru stále ještě žije. Mnozí z „průkopníků“ umělých neuronových sítí stále aktivně vědecky působí.

⁵Tzv. *Hebbova hypotéza* (1949): Změna váhy spoje mezi neurony je úměrná korelaci aktivit těchto neuronů.

3.3.2 Doba rozkvětu

1951 rozvíjejí MARVIN MINSKY ve své disertační práci Minsky [54] neuropočítač *Snark*, který dokáže automaticky nastavit své váhy. Snark ale nikdy nebyl prakticky použitý. Přínos i neblahé důsledky práce Marvina Minského (viz dále) shrnul ve svém příspěvku Kelemen [38] Jozef Kelemen.

1957 vynalézá FRANK ROSENBLATT tzv. *perceptron*, který je zobecněním McCullochova a Pittsova modelu pro reálný číselný obor parametrů. Více se lze o přínosu Franka Rosenblatta dozvědět např. v Šíma – Neruda [86].

1957–1958 rozvíjí FRANK ROSENBLATT, CHARLES WIGHTMAN a jejich spolupracovníci v MITu⁶ první úspěšný neuropočítač, *Mark I Perceptron*, který má 20×20 pixelů velký senzor rozeznávající jednoduché číslice a 512 motorově poháněných potenciometrů – pro proměnné nastavení vah (co variabilní váha, jeden potenciometr). FRANK ROSENBLATT popisuje v Rosenblatt [74] neuronovou síť jako model mozku s ohledem na ukládání informací.

1959 popisuje FRANK ROSENBLATT různé varianty perceptronu, formuluje a dokazuje svůj *Teorém o konvergenci perceptronu*.

1960 představují BERNARD WIDROW a MARCIAN E. HOFF *Adaline*⁷, rychlý a přesný učící se adaptivní systém popsany např. v Widrow – Marcian E. Hoff [96].

Téhož roku zakládají ROGER BARRON a LEWEY GILSTRAP první firmu na aplikace neurovýpočtů.

1961 představuje KARL STEINBUCH techniky asociativní paměti, které jsou považovány za předchůdce dnešních neuronových asociativních pamětí.

1965 dává NILS NILSON ve své knize *Učící se stroje* (Nilson [62]) přehled o pokroku výzkumu neuronových sítí a pokroku práce z přelomu padesátých a šedesátých let dvacátého století. Obecně se přijímá, že základní principy učícího se a tedy, jak se běžně používá „inteligentního“ systému, jsou již objeveny, což bylo z dnešního pohledu bezmezná přecenění dané situace. Na druhou stranu to ale umělým neuronovým sítím zajistilo velkou popularitu a přineslo postačující prostředky na výzkum.

⁶Massachusetts Institute of Technology, viz [49], konkrétně v Laboratoři umělé inteligence

⁷ADaptive LINear Element

1969 provádějí MARVIN MINSKY a SEYMOUR PAPERT ve své knize Minsky – Papert [55] podrobnou matematickou analýzu perceptronu. Ukazují, že perceptron neumí reprezentovat mnoho důležitých problémů⁸ a upozorňují na toto přecenění. Popularita umělých neuronových sítí je rázem ta tam a prostředky na výzkum se náhle nedostávají. Následné zjištění, že robustnější modely vykazují stejné problémy a prognóza, že celý obor je mrtvý, způsobují téměř kompletní ztrátu prostředků na výzkum na příštích 15 let. Tak mylné prognózy to z dnešního pohledu byly. Navíc zanedbávání analytického výzkumu a provádění pompézních experimentů s umělou inteligencí, připomínající alchymii dílen Rudolfa II., jen přispělo ke ztrátě zájmu o umělé neuronové sítě a umělou inteligenci vůbec.

3.3.3 Dlouhé ticho a pomalejší obnova

Přes malou atraktivitu oboru byly v tomto období, které odstartovalo v roce 1969 kniha Minsky – Papert [55] a ukončeno v polovině devadesátých let minulého století, položeny teoretické základy pro renesanci neuronových sítí.

1972 představuje TEUVO KOHONEN model *linárního asociátoru* – model asociativní paměti. Tento neuropsychologický pohled je ve stejném roce prezentován nezávisle na JAMES A. ANDERSONovi, který dospěl ke stejným výsledkům.

1973 používá CHRISTOPH VON DER MALSBURG nelineární a biologicky lépe motivovaný neuronový model.

1974 vyvíjí PAUL WERBOS ve své disertaci Werbos [95] na Harvardu princip učení – *backpropagation*. Ten se stane významným a „znovuobjeveným“ teprve o 10 let později.

1976–1980 a později prezentuje STEPHEN GROSSBERG mnoho prací (např. Grossberg [21]; výčet publikací lze nalézt na Johnson [37], kde je mnoho prací dostupných v elektronické podobě), v nichž je matematicky přesně analyzováno velké množství neuronových modelů. Grossberg se v nich velmi podrobně věnuje problémům neuronových sítí řízených adaptivně, bez předem známých asociací. Díky spolupráci s GAILEM CARPENTEREM následně vznikají modely teorie adaptivní resonance, krátce ART (mimo jiné Carpenter – Grossberg [7]).

⁸Jedním z triviálních problémů, který neumí perceptron řešit, je logická funkce XOR – eXclusive OR, tj. tzv. vylučovací disjunkce.

1982 popisuje TEUVO KOHONEN *samoorganizující se mapy*⁹ (Kohonen [39]), které po svém objeviteli dostaly jméno – *Kohonenova síť*. Obšírněji o této problematice pojednává o několik let později vydaná kniha Kohonen [40], které je TEUVO KOHONEN autorem.

1983 představují FUKUSHIMA, MIYAKU a ITO neuronový model *Neocognitron*¹⁰ – viz např. Fukushima – Miyaku [19], který slouží k rozpoznávání ručně psaných symbolů. *Neocognitron* tak představuje rozšíření již v roce 1975 vyvinutého *Cognitronsu* a popsaneého v Fukushima [18].

Ve stejném roce ukazují KIRKPATRICK, GELATT a VECCHI, jak lze prostřednictvím Boltzmannova stroje¹¹ řešit některé optimalizační úlohy, např. *problém obchodního cestujícího*.

3.3.4 Renesance

Pod vlivem JOHNA HOPFIELDa a jiných vědců začínají další vědci opět věřit v budoucnost oborů spjatých s technologií neuronových sítí. Navíc další publikace Rumelhart et al. [75] o metodě backpropagation od RUMELHARTa, HINTONA a WILLIAMSE pomalu obnovily období rozkvětu oboru neuronových sítí.

1985 uveřejňuje HOPFIELD článek (Hopfield – Tank [31]), ve kterém popsal řešení problému obchodního cestujícího prostřednictvím *Hopfieldovy sítě*.

1986 jsou ve velké míře publikovány články o učícím algoritmu backpropagation. Dle Šíma – Neruda [86] své výsledky publikují zejména badatelé z tzv. *PDP skupiny*¹² ve sborníku, jenž edituje DAVID RUMELHART a JAMES MCCLELLAND (David Rumelhart and James McClelland [12]). V tomto sborníku se mimo jiné objevuje článek RUMELHARTa, HINTONA a WILLIAMSE – Rumelhart et al. [75], kteří v něm popisují učící algoritmus zpětného šíření chyby (backpropagation) pro vícevrstvou neuronovou síť. Tím vyřešili problém, který se Minskému a Papertovi v šedesátých letech minulého století jevil jako nepřekonatelná překážka pro využití a další rozvoj neuronových sítí.

⁹SOM – self organizing maps

¹⁰Neocognitron je vícevrstvá hierarchická neuronová síť. První verze sítě Neocognitron byla založena na principu učení bez učitele. Původní dvě základní verze navržené profesorem Fukushimou se od sebe odlišují především použitým principem učení. Jedná se o Neocognitron využívající: 1) učení bez učitele a 2) učení s učitelem.

¹¹Boltzmannův stroj je symetrická binární zpětnovazební síť se skrytou vrstvou.

¹²Parallel Distributed Processing Group

Backpropagation je dosud nejpoužívanější učící metodou dopředných neuronových sítí. Jak již bylo naznačeno, backpropagation se objevil již dříve. Mezi vědce, kteří ho navrhli již dříve, patří např. ARTHUR BRYSON, YU-CHI HO či již zmíněný PAUL WERBOS.

1987 v práci Hecht-Nielsen [29] a Hornik et al. [32] ukazují jejich autoři cestu k obecnějšímu poznání ve směru reprezentovatelnosti funkcí. Ta vede ke Kolmogorově teorému o řešení tzv. třináctého Hilbertova problému. Tím je univerzálnost neuronových sítí jako aproximátorů matematicky dokázána. Blíže se touto problematikou zabývá kniha Šíma – Neruda [86] ve své třetí části.

V San Diegu se téhož roku koná první větší konference specializovaná na neuronové sítě (*IEEE International Conference on Neural Networks*). Na ní se schází na tu dobu neuvěřitelných 1700 účastníků a je založena mezinárodní společnost pro výzkum neuronových sítí *INNS*¹³ – [36] .

3.3.5 Od toho času . . .

O rok později po svém vzniku začíná *INNS* vydávat svůj časopis *Neural Networks*– [60]. V následujících letech vznikají další specializované časopisy, jako např. *Neural Computation* (1989) – [58], *IEEE Transactions on Neural Networks* (1990) – [35] a mnoho jiných (např. v Praze vychází od roku 1991 mezinárodní časopis *Neural Network World* – [59]).

Až doposud probíhá doslova explozivní vývoj v oblasti výzkumu, který už není rozdrobený. Mnoho renomovaných univerzit založilo nové výzkumné ústavy zabývající se neuronovými sítěmi a vyhlásilo výukové programy zaměřené na neurovýpočty. Tento trend pokračuje až dodnes, kdy se zdá, že široký záběr výzkumu a vynaložené investice neodpovídají kvalitě dosažených výsledků. Teprve budoucnost ověří správnost těchto premis.

¹³International Neural Network Society

Kapitola 4

Modely neuronových sítí

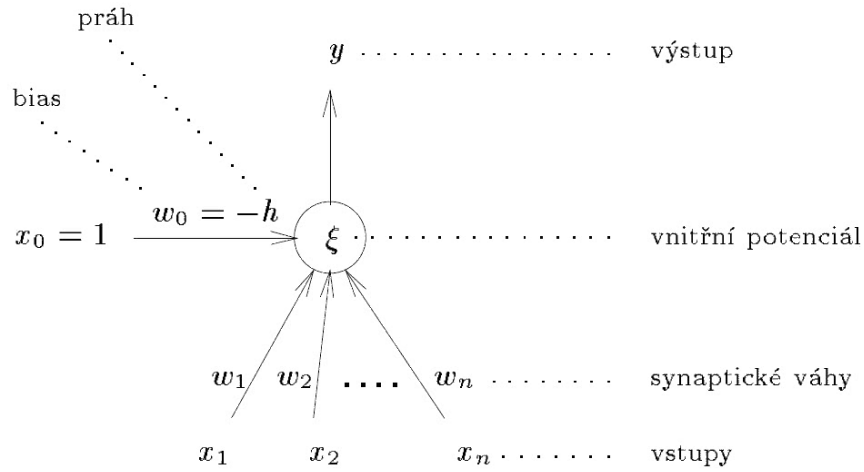
Kapitola 4 pojednává o vybraných zástupcích a obecných koncepcích neuronových sítí. Nejprve je však třeba definovat základní funkční jednotku neuronové sítě a tou je neuron.

4.1 Stavební jednotky sítě

4.1.1 Formální neuron

Základní stavební jednotkou matematického modelu neuronové sítě je tzv. *formální neuron*. Ten je získán ze zjednodušeného biologického neuronu. Práce neuronu je pak také motivována svým biologickým vzorem. Schématická struktura formálního neuronu je znázorněna na obrázku 4.1.

Vzhledem k tomu, že se neuron a následně i neuronová síť v čase vyvíjí (parametry sítě, parametry učení, atp.), je při tvorbě jejich definic nutno používat *čas*. Čas je ukazatelem průběhu práce neuronové sítě. Časová kalkulace se zároveň dělí do diskrétních časových intervalů – práce neuronové sítě je tedy nahlížena epizodicky. Kriesel [41] označuje aktuální čas (přítomnost) symbolem t , následující časový interval symbolem $t + 1$ a předcházející časový interval symbolem $t - 1$. Vztahují-li se různé veličiny (např. $w_{i,j}$ nebo y_i) k nějakému určitému času, je čas značen formou horního indexu, tj. $(w_{i,j})^t$ resp. $(y_i)^t$.



Obrázek 4.1: Schématické znázornění formálního neuronu dle Šíma – Neruda [86]

4.1.1.1 „Klasický“ formální neuron

Formální neuron představený v práci McCulloch – Pitts [51] (dále jen neuron) má n obecně reálných *vstupů* $x_1, \dots, x_n$, reprezentujících dendrity¹. Vstupy jsou ohodnoceny taktéž obecně reálnými *synaptickými váhami* $w_1, \dots, w_n$, jež určují jejich propustnost. Ve shodě s neurofyziologickou motivací uvedenou v podkapitole 3.2.1 jsou obvykle uvažovány též záporné synaptické váhy. Tím je pak vyjádřen inhibiční charakter vah. Vážená suma vstupních hodnot je často snižovaná o práh neuronu (bias)² Θ . Prah však může také vstupovat jako proměnná až do přenosové funkce – viz dále. Model práce celého neuronu s obecněji pojatým vlivem prahové hodnoty uvádí vztah 4.1:

$$y = \sigma \left(\sum_{i=1}^n x_i w_i, \Theta \right), \quad (4.1)$$

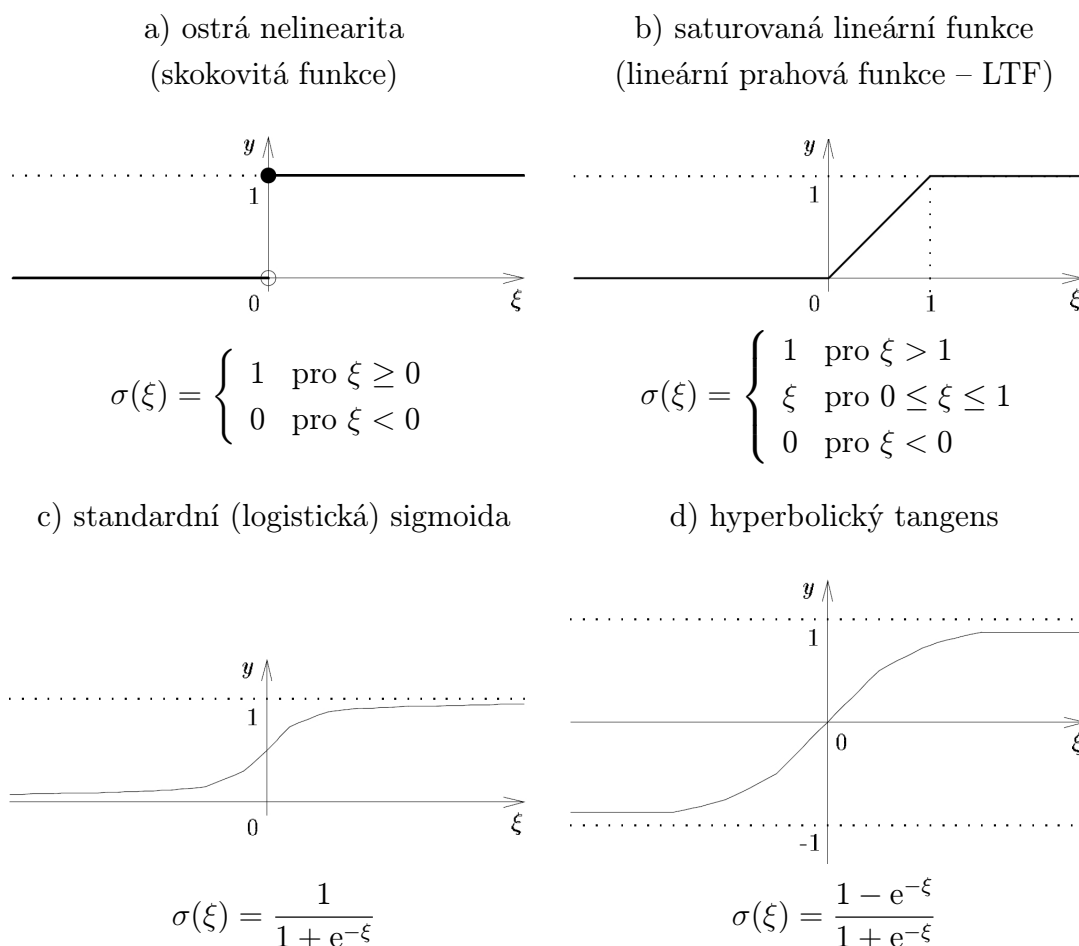
kde $\sum_{i=1}^n x_i w_i$ bývá označováno jako ξ a nazývá se *vnitřním potenciálem neuronu*.

Právě hodnota vnitřního potenciálu ξ po dosažení prahové hodnoty Θ indukují po transformaci *výstup* (*stav*) daného neuronu y , čímž je modelován elektrický impuls axonu. Obecně nelineární transformace vnitřního potenciálu $y = \sigma(\xi)$ při současném

¹Reálně se počet pohybuje v desítkách vstupů na každý umělý neuron.

²Na obrázku 4.1 je bias respektive práh zobrazen různými (opačnými) hodnotami w_0 a $-h$. Označení w_0 odpovídá realizaci biasu, jako jedné z vah. Např. Kriesel [41] zavádí koncepci tzv. *On-neuronu*, který modeluje prahové hodnoty pro celou síť. Více lze nalézt v Kriesel [41]. Z jiného pohledu předvádí bias na jednu z vah v Šíma – Neruda [86].

zohlednění prahové hodnoty Θ je dána *aktivační (přenosovou) funkcí*³ σ . Nejčastěji používané aktivační funkce mají tvar a matematický předpis uvedený na obrázku 4.2. Pro zjednodušení zápisu a lepší představu je obrázek 4.2 vytvořen tak, že $\Theta = 0$.



Obrázek 4.2: Nejčastěji používané aktivační funkce dle Šíma – Neruda [86]

Přenosové funkce jsou, jak uvádí Novák et al. [64], vlastně dvojí:

- funkce σ typu LTF (linear threshold, resp. transfer function)
- funkce σ typu SAF (sigmoidal semilinear activation function)

První typ je ovšem možno chápat jako zvláštní (limitní) případ přenosových funkcí druhého typu.

³Pro tuto funkci je v literatuře používáno různé označení. Část prací hovoří – podobně jako zde – o přenosové funkci (transfer function, či transmission function), některé práce ji označují jako prahovou funkci (threshold function), jinde je používáno označení „obecně nelineární funkce“.

4.1.1.2 Formální neuron dle Kriesel [41]

Německá technická preciznost, reprezentovaná v této rešerši Krieselovou knihou Kriesel [41], popisuje a definuje složení neuronu a to (od vstupu signálu do neuronu po jeho výstup)⁴ jako:

Spojení – synapse: Přes spojení jsou přenášena data mezi neurony, přičemž váhy na synapsích mají buďto zesilující, nebo zeslabující vliv.

Propagační funkce: Propagační funkce f_{prop} pro každý neuron j přijímá výstupy $o_{i_1}, \dots, o_{i_n}$ ⁵ ostatních neuronů $i_1, \dots, i_n$, z nichž existuje spojení k neuronu j , a zpracovává je s ohledem na váhy jednotlivých spojení $w_{i,j}$ k hodnotě síťového vstupu neuronu net_j pro j -tý neuron, který může být použit pro aktivační funkci:

$$f_{prop} : \{o_{i_1}, \dots, o_{i_n}\} \times \{w_{1,j}, \dots, w_{n,j}\} \rightarrow net_j. \quad (4.2)$$

Nejčastěji bývá používána vážená suma vstupů, tj.

$$net_j = \sum_i (o_i \cdot w_{i,j}). \quad (4.3)$$

Vstup neuronu – vnitřní potenciál: Vstup neuronu net_j je tedy výsledkem propagační funkce.

Prahová hodnota / Práh / Bias: Prahová hodnota Θ_j je jednoznačně přiřazená neuronu j a značí místo největšího stoupání aktivační funkce – viz dále.

Konkrétní hodnota aktivace – stav: Konkrétní hodnota aktivace udává stupeň aktivity neuronu a často se označuje termínem *aktivace*. Stupeň aktivace a_j ,⁶ přiřazený jednoznačně k neuronu j , je definován prostřednictvím aktivační funkce f_{act} jako:

$$a_j(t) = f_{act}(net_j(t), a_j(t-1), \Theta_j), \quad (4.4)$$

⁴Úmyslně je v této části ponechána symbolika originálu Kriesel [41]. Proti předchozímu popisu neuronu tak, jak je uváděno v české literatuře, toto značení obsahuje složitější indexování, jež umožňuje rozlišovat i jednotlivé neurony a přímo pak budovat model neuronové sítě.

⁵V předchozím značeno $x_1, \dots, x_n$. Díky způsobu zavedení, s jakým je pracováno v Kriesel [41] – viz předchozí poznámka pod čarou, nelze úplně přesně převést definici formálního neuronu dle Kriesel [41] na definici prezentovanou např. v Šíma – Neruda [86], která přímo vychází z McCullocha Pittsova přístupu.

⁶V předchozím značeno y , stejně jako např. v Šíma – Neruda [86].

jinak zapsáno $f_{act} : net_j(t) \times a_j(t-1) \times \Theta_j \rightarrow a_j(t)$. Aktivační funkce je dle Kriesel [41] zároveň i transformací značenou dříve σ (viz Šíma – Neruda [86]).

Výstupní funkce: Výstupní funkce⁷ f_{out} neuronu j vypočítává hodnotu, jež je předávána dalším neuronům, ke kterým existuje spojení toho z j -tého:

$$f_{out} : a_j \rightarrow o_j. \quad (4.5)$$

Často je touto funkcí volena identita $o_j = a_j$.

4.1.2 „Alternativní“ neurony

Výše popsaný formální neuron je typický pro tzv. dopřednou síť. Existují však i jiné typy sítí, jejichž jednotky se obvykle nenazývají neurony, ale *lokální*, či *semilokální jednotky*. Ty mají úplně jinou přechodovou funkci než formální neuron uvedený výše. Přenosová funkce u lokálních jednotek vypočte v prvním kroku vzdálenost vstupu od středu určeného parametry lokální jednotky a v kroku druhém transformuje nejčastěji Gaussovou funkcí (obecně nelineární funkcí) tuto vzdálenost na hodnotu výstupní. Lokální jednotky jsou základem například pro RBF⁸ neuronové sítě. Schéma RBF architektury je znázorněno na obrázku 4.3. V této síti jsou lokální jednotky „skryty“ v střední části sítě.

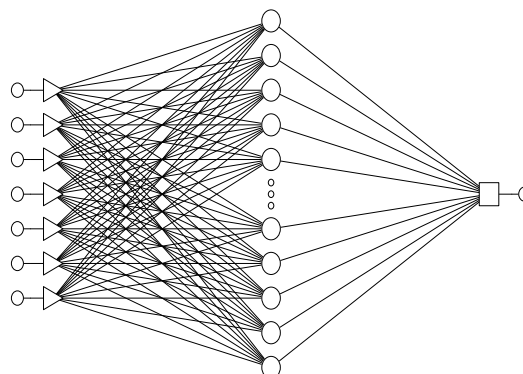
Ze zevrubného popisu lokální jednotky vyplývá, že jednotky se chovají v jistém smyslu duálně k perceptronům. Zatímco perceptron globálně rozděluje vstupní prostor na dva podprostory, lokální jednotky mají relevantní výstup lokalizován do okolí bodu určeného svými parametry. Více např. Šíma – Neruda [86, str. 177–129] nebo Wang – Fu [94].

4.2 Neuronová síť

Umělá neuronová síť se skládá z formálních neuronů vzájemně propojených tak, že výstup neuronu je zároveň vstupem obecně více jiných neuronů. Počet neuronů a jejich vzájemné propojení v síti určuje *architekturu neuronové sítě*. Z hlediska využití rozlišujeme v síti *vstupní*, *pracovní* (*skryté*, *mezilehlé*) a *výstupní neurony*. Stavby všech

⁷Na rozdíl od přenosové funkce tak, jak ji chápali např. McCulloch – Pitts [51], nemá obvykle žádnou další funkci, tj. předává přímo aktivaci.

⁸RBF – **R**adial **B**asis **F**unction



Obrázek 4.3: Schéma RBF neuronové sítě 7:7-83-1:1 – 83 lokálních jednotek je „skryto“ v střední části sítě

(Vytvořeno v programu STATISTICA Neuronové sítě 7.1)

neuronů v síti určují stav neuronové sítě a synaptické váhy všech spojení představují *konfiguraci neuronové sítě*.

Neuronovou síť lze chápat jako *naučitelné zobrazení* – funkci. Kriesel [41] definuje neuronovou síť jako uspořádanou trojici (N, V, w) se dvěma množinami N, V a funkcí w . N označuje množinu *neuronů*, V uspořádanou množinu $\{(i, j) | i \in N, j \in N\}$, jejíž prvky se nazývají *spojení* (synapse) neuronu i s neuronem j , a funkce $w : V \rightarrow \mathbb{R}$ definuje synaptické váhy $w((i, j))$; krátce se označují symbolem $w_{i,j}$. Pro neexistující spojení od neuronu i k neuronu j se váha $w_{i,j}$ buď nedefinuje, nebo je definovaná jako $w_{i,j} = 0$. Váhy souhrnně lze reprezentovat čtvercovou maticí nazývanou maticí sousednosti.

V Holeňa [30] je neuronová síť definovaná jako orientovaný graf, kde neurony jsou reprezentovány vrcholy grafu a spoje mezi neurony realizují hrany grafu. V této práci bude na neuronovou síť nahlíženo výhradně jako na zobrazení.

4.2.1 Klasifikace neuronových sítí

4.2.1.1 Klasifikace dle funkce

Neuronové sítě se většinou dělí do tří skupin: *asociativní paměti*, *klasifikátory* a *aproximátory*.

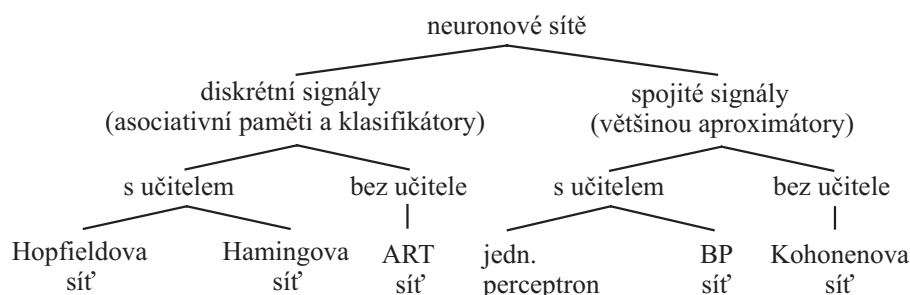
Asociativní paměti jsou takové sítě, jež odpovídají vzorem. Ten je v síti asociován

(logicky spojen) se vstupním vzorem. Podle druhu asociace se dále asociativní paměti rozdělují na *autoasociativní* (vstupní i výstupní vzor je podobný; tyto paměti se používají například k rozpoznávání poškozených vzorů) a *heteroasociativní* (vstupy i výstupy jsou různé). Nejznámější asociativní pamětí je Hopfieldova síť.

Klasifikátory jsou sítě, jež rozpoznají vstupní vzor a zařadí ho do vybrané třídy. Typickým zástupcem tohoto druhu sítí je např. Hamingova síť.

Poslední skupinou z této klasifikace neuronových sítí jsou aproximátory. Jejich funkcí je vytvoření vlastního vnitřního modelu na základě vstupních informací, jenž následně aproximuje skutečný systém. To pak dovoluje síti predikovat události. V současné době se aproximátory stávají silným konkurentem například i tak mocnému nástroji, jakým je regresní analýza. Do této skupiny patří asi nejznámější neuronová síť vůbec a tou je vícevrstvá dopředná síť, někdy též označovaná jako backpropagation síť. Této síti bude věnována další pozornost v souvislosti s jejím učením, viz kapitola 5.

Rozdělení neuronových sítí dle jejich funkce je přehledně uvedeno na obrázku 4.4



Obrázek 4.4: Rozdělení neuronových sítí dle jejich funkce

(Zdroj: Chmelenský [34])

4.2.1.2 Klasifikace dle způsobu učení

Kromě předchozího způsobu rozdělení lze sítě ještě dělit na sítě pracující *s učitelem* a na sítě *bez učitele*. Sítě pracující s učitelem se snaží napodobit předkládané vzory a generalizovat podstatu informací. Neuronová síť se tedy učí z příkladů. Naproti tomu sítě pracující bez učitele si musí najít svou vlastní reprezentaci pouze ze známých vstupních dat a zadaného kritéria, jemuž se snaží vyhovět. Mezi sítě s učitelem patří např. Hopfieldova síť a mezi sítě bez učitele řadíme např. Kohonenovu síť. Blíže se

pojmu učení věnuje kapitola 5.

4.2.1.3 Klasifikace dle zpracovávaného signálu

Dalším možným rozdělením neuronových sítí je rozdělení dle druhu zpracovávaného signálu. Ten může být buď *diskrétní*, nebo *spojitý*. Diskrétní signály zpracovává např. Hamingova nebo ART síť a spojitě signály může zpracovávat např. backpropagation síť nebo Kohonenova síť. Protože většina implementací umělých neuronových sítí je ve formě počítačových programů (MatLab, SAS, Statistica aj.), znamená to, že je pracováno vesměs s diskrétně kódovanými signály. Existují však i hardwarové struktury a optické sítě zpracovávající i signály spojitě.

4.2.1.4 Klasifikace dle architektury sítě

Dělení sítí lze taktéž založit na typu architektury sítě, viz Kriesel [41].

Feedforward – dopředná síť

Neurony jsou v architektuře dopředné sítě řazeny do vrstev: jedné vstupní vrstvy, n výkonných vrstev (skrytých pro vnějšek sítě) a jedné výstupní vrstvy. Spojení mezi každým jedním neuronem smí u dopředné sítě směřovat výhradně k neuronu v následující vrstvě. Na obrázku 4.5 je zobrazena dopředná síť se vstupní vrstvou osmi neuronů (trojúhelníčky), dvěma skrytými vrstvami o deseti a pěti neuronech a s jedním výstupním neuronem⁹. Tato síť byla použita pro diskriminaci „platičů“ a „neplatičů“ v příspěvku Biskup [108].

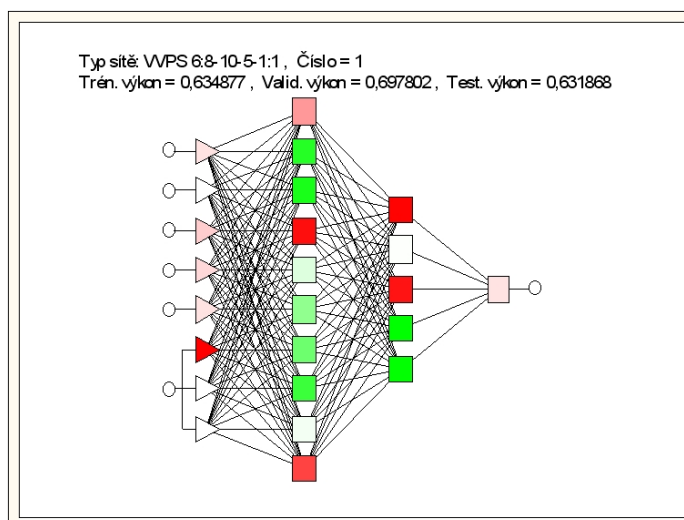
Kriesel [41] uvádí, jak rozsáhlá musí být např. síť typu vícevrstvého perceptronu¹⁰ (typický zástupce dopředné sítě) pro reprezentaci různých množin, viz tabulka 4.1.

Někdy jsou u dopředných sítí uvažována i tzv. *zkracující spojení* (viz Kriesel [41]), která přeskakují jednu nebo více úrovní (vrstev). Také tyto spoje však smí výhradně směřovat směrem k výstupní vrstvě.

O zobecnění vícevrstvého perceptronu – dopředné sítě – pojednává podkapitola 5.3.1.

⁹Barevné zvýraznění neuronu vypovídá aktivitě jednotlivých neuronů, jak je popsáno na str. 130 v kapitole 11.

¹⁰Vícevrstvý perceptron je vícevrstvá dopředná síť, jejíž neurony ve výstupní vrstvě mají pouze binární výstupy.



Obrázek 4.5: Schéma dopředné neuronové neuronové sítě s dvěma skrytými vrstvami
(Vytvořeno v programu STATISTICA Neuronové sítě 7.1)

Tabulka 4.1: Klasifikovatelná množina vícevrstevným perceptronem

(Zdroj: Kriesel [41])

Počet skrytých vrstev	Klasifikovatelná množina
1	nadrovina
2	konvexní polygon
3	každá libovolná množina
4	také každá libovolná množina, tedy žádný další užitek

Zpětnovazebná síť

V tomto typu sítě se jako zpětně spojený neuron označuje takový neuron, který může ovlivňovat jakýmkoliv způsobem jakékoliv spojení. Zpětnovazební síť umožňují jak spojení běžné v dopředné síti, tak spojení neuronu se sebou samým. Tento typ spojení je označován jako *přímá zpětná vazba*. Neurony se tak samy omezují a posilují, aby dospěly ke své aktivační hranici. Vazby běžně nepřípustné v dopředné síti jsou tzv. *nepřímé zpětné vazby* a *příčné zpětné vazby*. První z nich jsou spojení směrem ke vstupní vrstvě, kde neuron může sám sebe ovlivňovat skrze okliku (cyklus). Druhý typ vazby je spojení neuronů v jedné rovině (vrstvě). Takové spojení pak často způsobuje, že každý neuron brání dalším neuronům ve vrstvě v aktivaci a posiluje tak sebe sama a pouze nejsilnější neuron je aktivní (Systém – „Vítěz bere vše“).

Úplně propojená síť

Úplně (kompletně) propojená síť dovoluje spojení mezi všemi neurony (tj. všechny typy vazeb) kromě přímých zpětných vazeb. Spojení musí být symetrická. Populárním příkladem je *samoorganizující se mapa* od Kohonena. V takovém typu sítě může každý neuron principiálně udržovat spojení s každým neuronem – tedy každý neuron, jak vysvětluje Kriesel [41], se může stát vstupním neuronem. Proto zde zpravidla není používáno přímé zpětné spojení a neexistují ani žádné jasně definované vrstvy.

4.3 Kolekce neuronových sítí

Komplet nebo také kolekce (Ensemble) je soubor neuronových sítí kooperujících za účelem tvorby kvalitnějších předpovědí. V principu lze uvažovat o kolekcích, které vypočítávají výstup kolekce nebo spolehlivost¹¹.

4.3.1 Kolekce pro výpočet výstupu sítě

Do kolekce pro výpočet výstupu sítě může být zařazena jakákoli skupina neuronových sítí s obecně různými vstupy i výstupy. Pokud však mají sítě různé výstupy, pak má výstup kolekce vícenásobný výstup¹². Kolekce sítí proto může být použita například pro tvorbu modelu s vícenásobným výstupem, kde každý výstup je vytvářen samostatně jinou sítí.

Pokud mají sítě uspořádané do kolekce stejný výstup, pak kolekce sítí odhaduje výstupní hodnoty prostřednictvím kombinací výstupů jednotlivých neuronových sítí. Výstupy kolekce jsou v závislosti na typu výstupu tvořeny následujícím způsobem:

1. KATEGORIÁLNÍ výstupy sítí, tj. výstupy sítí použité např. pro klasifikaci, jsou v rámci kolekce sítí kombinovány prostřednictvím módu výstupu sítí. V případě vícenásobného módu kolekce označí daný vzor za neznámý.
2. NUMERICKÉ výstupy sítí, tj. výstupy sítí použitých pro předpověď reálné hodnoty, jsou v rámci kolekce sítí průměrovány.

Výstupy jednotlivých sítí jsou v rámci kolekce váženy prostřednictvím volitelných vah nezávisle na tom, zda se jedná o klasifikaci nebo o predikci reálné hodnoty.

¹¹Spolehlivost není chápána tak, jak s ní pracuje statistika (tj. konfidence), ale jako aktivace výstupu sítě – viz podkapitola 4.3.2.

¹²Dokonce i za předpokladu, že sítě do kolekce zařazené měly každá pouze jediný výstup avšak predikující hodnoty jiného ukazatele, má kolekce vícenásobný výstup.

Software STATISTICA 7.1 Neuronové sítě implicitně volí tyto váhy rovny jedné, ale umožňuje jejich modifikaci.

4.3.2 Kolekce pro výpočet spolehlivosti

Kolekce pro výpočet spolehlivosti jsou primárně určeny pro úlohy klasifikační a jejich použití má více omezení než kolekce pro výpočet výstupu sítě. Kolekce kombinuje (průměruje) přímo hodnoty aktivizačních funkcí výstupních neuronů sítí zařazených do kolekce. Pro predikci reálné hodnoty dávají kolekce pro výpočet spolehlivosti stejné výsledky jako kolekce pro výpočet výstupu sítě, neboť spolehlivost je zde rozuměna bezprostředně hodnota aktivizační funkce výstupního neuronu/neuronů. I pro klasifikaci jsou kombinovány hodnoty aktivací výstupních neuronů a nikoli výsledky klasifikace neuronovou sítí po aplikaci klasifikačního prahu. Aktivizační práh je volen až pro hodnotu aktivace kolekce, a proto kolekce pro výpočet spolehlivosti na rozdíl od kolekcí pro výpočet výstupu sítě vždy dospějí k rozhodnutí.

4.3.3 Důvody pro použití kolekcí

Existuje několik důvodů pro použití kolekcí sítí. Některé z nich jsou uvedeny například v StatSoft [81], jiné byly navrženy a ověřeny v rámci aplikační části (kapitola 11 a 12) předkladatelem disertační práce.

1. Kolekce umožňují pohodlné uspořádání neuronových sítí, predikujících proměnné/vstupy aniž by všechny úlohy musely být bezprostředně realizovány jedinou sítí. Vícenásobné výstupy (jedné) složité sítě bývají totiž často negativně ovlivněny nežádoucími vazbami (šíření nežádoucích signálů mezi neurony) ve skrytých vrstvách. Výsledkem mohou být i nekvalitní predikce. Kolekce umožňují vypočítávat každý výstup samostatně a negativnímu vlivu nežádoucích vazeb zabránit.
2. Kolekci z různě natrénovaných sítí lze seznámit se všemi trénovacími vzory a přitom je možno zabránit přetrénování neuronových sítí do kolekce zařazených, tj. zabránit přetrénování celé kolekce. Jsou tedy důležitým nástrojem v boji proti přetrénování a zlepšují tak generalizační schopnosti při použití neuronových sítí. „Průměrování“ predikcí získaných prostřednictvím neuronových sítí různých architektur a topologií¹³ na různých trénovacích a validačních množi-

¹³Topologie sítě je konkrétnějším popisem neuronové sítě než architektura sítě. Topologii sítě tvoří jednak její architektura, tj. vzájemné propojení neuronů v síti, a dále pak zvolené typy přenosových

nách snižuje různorodost modelů a přitom zachovává věrohodnost těchto modelů. Sítě trénované na různých trénovacích a validačních množinách prostřednictvím cross-validace (křížového ověření), opakovaného náhodného výběru a bootstrapu¹⁴ je možno následně uspořádat do jednoho celku – kolekce. V StatSoft [81] je uvedeno, že mnoho teoretických prací ukazuje, že očekávaná výkonnost kolekcí je vyšší, než průměrná výkonnost sítí do kolekce zařazených. Tento předpoklad byl ověřen při trénování neuronových sítí v souvislosti s modelováním časových řad¹⁵.

3. Kolekce sítí umožňuje díky využití cross-validace, opakovaného náhodného výběru a bootstrapu při tvorbě trénovací, validační a testovací množiny vytvořit *nestranné* odhady výkonnosti jednotlivých sítí.
4. Pro kolekci je možný výpočet variability predikcí a tedy i spolehlivosti modelů.

funkcí a způsob modifikace volných parametrů.

¹⁴Základy metod pro ověřování kvalit modelů prostřednictvím opakovaných výběrů byly položeny v článcích Stone [82] (cross-validace) a Efron [14] (bootstrapu). Rozličné varianty bootstrapu jsou stále aktuální a používají se i mimo problematiku neuronových sítí – viz např. diplomová práce Petrásek [71], v níž je proveden teoretický popis variant několika bootstrapových metod. Popis využití zmíněných metod pro evaluaci trénování neuronových sítí přináší i publikace věnované přímo problematice neuronových sítí: Šíma – Neruda [86], StatSoft [81] či Haykin [25].

¹⁵Výsledky uvedené v tabulce 12.5 na str. 156 a tabulce 12.7 na str. 159 ukazují, že chyby ($M.A.E$) pro kolekci sítí jsou skutečně nižší, než průměrná chyba sítí, do kolekce zařazených.

Kapitola 5

Učení neuronových sítí

Proces vývoje sítě zmíněný v podkapitole 4.2 se nazývá *učením (trénováním)* neuronové sítě. Učením se mění propojení a stav neuronů a adaptují se váhy. V souvislosti se změnou těchto charakteristik v čase je vhodné celkovou dynamiku neuronové sítě rozdělit do tří oddělených částí – dynamik. S ohledem na dynamiky pak uvažujeme tři režimy práce sítě: *organizační* (změna topologie), *aktivní* (změna stavu) a *adaptivní* (změna konfigurace). Toto dělení neodpovídá plně neurofyzilogické skutečnosti, neboť v nervové soustavě probíhají příslušné změny současně, což mimo jiné činí biologickou neuronovou síť stále ještě výkonově nedosažitelnou (u lidského mozku také nečekají jedny neurony na druhé), ale implementace je tímto docela výrazně zjednodušená. Uvedené dynamiky neuronové sítě jsou obvykle určeny počátečním stavem a pravidlem. Toto pravidlo určuje vývoj dané charakteristiky sítě v čase (tj. změny topologie, stavu a konfigurace).

Konkrétním dynamikám odpovídají různé modely neuronových sítí, jež se zdají být vhodné pro řešení určitých tříd úloh. To znamená, že pro konkrétní specifikaci modelu neuronové sítě stačí definovat jeho organizační, aktivní a adaptivní dynamiku. Jednotlivé typy neuronových sítí již byly zmíněny v podkapitole 4.2.1.

Neuronová síť by se mohla učit mnohé, otázkou pak zůstává způsob implementace. Teoreticky by se mohla neuronová síť, dle Kriesel [41], učit tím, že:

1. rozvíjí nová spojení,
2. ruší existující spojení,
3. mění váhy spojení,
4. mění prahové hodnoty neuronů,
5. obměňuje jednu nebo více ze tří funkcí neuronu (propagační, aktivační a výstupní funkce),

- 6. vznikají nové neurony, nebo
- 7. zanikají existující neurony.

V dalším bude pozornost věnována třetí a čtvrté možnosti, jež jsou nejběžnější. Druhou možnost lze realizovat prostřednictvím třetí – nastavením příslušné váhy na nulu a dalším netrénováním této váhy. První možnost – vytvoření nových spojení – lze realizovat podobně jako druhou.

Možnost 5 je těžko implementovatelná, navíc není intuitivně ani biologicky motivovaná. Možnosti 6 a 7 poskytují během trénování neuronové sítě nejen dobře nastavené váhy, ale i optimalizace topologie sítě.

5.1 Klasifikace učení dle vztahu ke znalosti požadovaného výstupu

Nechť učení neuronové sítě probíhá tak, že jsou modifikovány váhy na spojeních mezi neurony. Přitom se předpokládá, že lze definovat co je *žádoucí* odezvou neuronové sítě a že existuje *trénovací množina* vzorů k učení. Pak lze dle Kriesel [41] rozlišit následující tři typy učení. Obvykle jsou uváděny, tak jako v podkapitole 3.2.3, způsoby dva.

5.1.1 Učení bez učitele

Při tomto typu učení (angl. *unsupervised learning*, *self-organized learning*) jsou síti předkládány pouze vzorové vstupy; síť se snaží identifikovat podobný vzor a zařadit jej do podobné kategorie. Jako nejznámější je uváděna *Kohonenova samoorganizující se mapa*. I takovéto učení je z pohledu biologické motivace přijatelné, neboť také živé systémy při svém učení vytváří skupiny podobných objektů, ke kterým pak přistupují stejně nebo podobně.

5.1.2 Posilující učení

Při tomto typu učení (angl. *reinforcement learning*) je síti dodávána po proběhnutí průchodu sítě jen pravdivostní hodnota, která definuje, zda výsledek je správný nebo není. Oproti následujícímu způsobu učení je zřejmé, že i zde síť hledá vlastní realizaci a přesný požadovaný trénovací výstup nemusí existovat. Právě sem by bylo možno zahrnout učení prostřednictvím „genetických“ algoritmů. Pravdivostní

hodnota, kterou obdrží síť jako odezvu na výsledek své aktivní dynamiky, by byla kritériální funkcí – viz podkapitola 5.4.2.

5.1.3 Učení s učitelem

Při učení neuronové sítě s učitelem (*supervised learning*) musí existovat trénovací množina vzorových vstupů a požadovaných odpovědí neuronové sítě. Pro každý do sítě vložený trénovací vzor je výsledek „vypočtený“ neuronovou sítí porovnán se správným řešením a dle difference jsou změněny váhy sítě. Cílem učení je změna vah tak, aby síť po natrénování samostatně nejen přiřadila vzorovému vstupu i požadovaný výstup, ale aby i do té doby neznámému vstupu dodala přijatelný výsledek, tedy aby uměla *generalizovat*. Tento způsob učení sice není plně biologicky přijatelný, ale je účelnější než ostatní, a proto velmi vhodný. Typickým zástupcem učení s učitelem je backpropagation.

5.2 Klasifikace učení dle principu modifikace vah

5.2.1 Účelové učení

Účelové učení má za cíl nalézt takový soubor vah, který odpovídá vhodné extrémní hodnotě účelové funkce. Autorem jedné z nejvýznamnějších zákonitostí tohoto druhu je prof. BERNARD WIDROW (viz např. Widrow – Marcian E. Hoff [96]). Jeho postup učení je jedním z nejvýkonnějších, neboť konverguje k optimu z jakéhokoli výchozího bodu.

5.2.2 Soutěžní učení

Soutěžní učení je dalším významným druhem učení. Je založeno na tom, že před každým cyklem učení je proveden výběr mezi uvažovanými výkonnými prvky. Ten z nich, který v soutěžení s ostatními zvítězí, získá také možnost provést modifikaci svých vah (nebo naopak vah všech ostatních nezvítězivších výkonných prvků). Například soutěžního učení založené na myšlence, že určitá uvažovaná vrstva výkonných prvků může uspořádat svoje váhy tak, aby příslušné váhové vektory byly rozděleny po celém prostoru $\mathbb{R}^n$ s hustotou pravděpodobnosti co nejlépe aproximující hustotu, se kterou jsou pro učení vybírány vstupní vektory je známo jako tzv. Kohonenův způsob učení vícevrstevných neuronových sítí (viz Kohonen [40]).

5.2.3 Časoprostorové učení

Časoprostorové učení je dokonce založeno na modifikaci vah nejen podle velikosti jistých vstupních signálů, ale též podle jejich změn v čase. Jsou zde tedy brána v úvahu též hlediska časové dynamiky sítě.

Problematika zákonitosti učení neuronových sítí je velmi široká a stále se prudce vyvíjí. Kromě poznávání nových, účinnějších a rychlejších procesů se však v poslední době začíná obracet pozornost též k otázkám citlivosti učebních procesů na změny jejich parametrů a jejich přípustné odchylky.

5.3 Backpropagation

Pro popis učení neuronové sítě byl vybrán přelomový princip učení – backpropagation, viz historie neuronových sítí (podkapitola 3.3). Jak již bylo naznačeno dříve, je neuronová síť určena architekturou a topologií. Aby bylo možno konkrétní princip učení představit, je nejprve třeba definovat topologii konkrétněji, než bylo provedeno v podkapitole 4.2. Architektura dopředné sítě již byla popsána.

5.3.1 Topologie dopředné neuronové sítě

V následujícím bude popsán dosti obecný model vícevrstvé dopředné neuronové sítě. Neuronová síť bude mít vstupní vrstvu s n_0 neurony (nultá vrstva), p skrytých vrstev, každou s n_i neurony, $i = 1, \dots, p$, a výstupní vrstvu ($p+1$ vrstva) s n_{p+1} neurony. Neurony vstupní a výstupní vrstvy budou mít lineární přenosovou funkci $\sigma(\xi)$:

$$\sigma(\xi) = \xi, \quad (5.1)$$

pro $i = 0, p+1$ a $j = 1, \dots, n_i$ a „ ξ “ je příslušná vstupní hodnota (dále již nebude připomínáno).

Všem skrytým neuronům bude přiřazena jako přenosová funkce sigmoida:

$$\sigma(\xi) = \frac{1}{1 + e^{-\xi}}, \quad (5.2)$$

pro $i = 1, \dots, p$ a $j = 1, \dots, n_i$, viz graf 4.2 c) na str. 39.

Vstupní vrstva bude sloužit, jak je u dopředné sítě zvykem, jen k distribuci vektoru vstupních hodnot $\mathbf{x} = (x_1, \dots, x_{n_0})$ k první skryté vrstvě, tj. $y_{0,j} = x_j$, pro $j = 0, \dots, n_0$. Váhy zde proto nemá smysl uvažovat. Stejně tak do dalších úvah nejsou

neurony vstupní vrstvy zařazovány a index i se tedy nepohybuje od 0 do $p+1$, ale obvykle

$$i = 1, \dots, p+1.$$

Hodnota vnitřního potenciálu j -tého neuronu v i -té nevstupní vrstvě bude značena $\xi_{i,j}$, $i = 1, \dots, p+1$ a $j = 1, \dots, n_i$, kde n_i je počet neuronů v i -té vrstvě. Množina vstupních neuronů j -tého neuronu v i -té vrstvě se obvykle značí $i_{\leftarrow}$.¹ Podobně množina výstupních neuronů j -tého neuronu v i -té vrstvě bude značena symbolem $i^{\rightarrow}$. Každý nevstupní neuron je dále charakterizován hodnotou biasu $\Theta_{i,j}$, $i = 1, \dots, p+1$ a $j = 1, \dots, n_i$ a vektorem synaptických vah $(w_{i,j,1}, \dots, w_{i,j,n_{i-1}})$. Jak již bylo naznačeno v podkapitole 4.1.1, je vhodné zařadit bias do vektoru synaptických vah.

Zobecněný vektor synaptických vah $\mathbf{w}_{i,j} = (w_{i,j,0}, w_{i,j,1}, \dots, w_{i,j,n_{i-1}})$ bude mít jako svůj první člen bias, tedy $w_{i,j,0} = -\Theta_{i,j}$. V tomto pojetí odpovídá bias synaptické váze jednotkového vstupu uměle přidaného nevstupního neuronu ($y_0 = 1$), který se stává vstupem pro všechny neurony všech vrstev sítě kromě vrstvy vstupní. Vektorem $\mathbf{y}_i = (y_{i,1}, \dots, y_{i,n_i})$, pro $i = 1, \dots, p$, budou značeny hodnoty $y_{i,j} = \sigma(\xi_{i,j})$, pro $j = 1, \dots, n_i$ – viz vzorec 5.2, reprezentující celkový výstup i -té skryté vrstvy, tj. výstup první skryté vrstvy $\mathbf{y}_1$ a tak podobně. Výstup poslední (výstupní) vrstvy je značen $\mathbf{y}_{p+1} = (\xi_{p+1,1}, \dots, \xi_{p+1,n_{p+1}})$, výstup první (vstupní) vrstvy $\mathbf{y}_0 = (x_1, \dots, x_{n_0})$. Rovnice pro výpočet vnitřního potenciálu lze vyjádřit s využitím maticového počtu pro všechny nevstupní neurony:²

$$\xi_{i,j} = \mathbf{w}_{i,j} \cdot (1, \mathbf{y}_{i-1})^T, \quad (5.3)$$

kde $i = 1, \dots, p+1$ a $j = 1, \dots, n_i$.

5.3.2 Učení zpětným šířením chyby

Neuronovou síť lze chápat jako naučitelné zobrazení³ F , jehož výstup $\mathbf{y}_{p+1}$ je závislý na vahách příslušných každému neuronu sítě a vstupních hodnotách, tedy $\mathbf{y}_{p+1} = F(\mathbf{w}, \mathbf{x})$, kde $\mathbf{w}$ je vektor všech zobecněných synaptických vah $\mathbf{w}_{i,j}$, $i = 1, \dots, p$ a $j = 1, \dots, n_i$, celé neuronové sítě a $\mathbf{x}$ je vstup sítě. Z hlediska práce neuronové sítě, jak bylo zmíněno na začátku této kapitoly, existuje několik dynamik

¹Index j může být vynechán, neboť je zřejmé, že množina vstupních neuronů je společná pro celou vrstvu. Hodnota $|i_{\leftarrow}|$ značí počet vstupních neuronů a je rovna n_{i-1} .

²Alespoň pro chvíli se tak dá vyhnout třem dolním indexům ve vzorcích. Pro srovnání: $\xi_{i,j} = \sum_{c=0}^{n_{i-1}} w_{i,j,c} \cdot y_{i-1,c}$.

³Viz podkapitola 4.2 a přístup popsany v Kriesel [41].

neuronové síti. Pro uživatele využívajícího již adaptovanou neuronovou síť (aktivní dynamika sítě) je samozřejmě podstatný pouze celkový výstup sítě $\mathbf{y}_{p+1}$. V průběhu učení se však musí počítat s hodnotami všech neuronů. V souvislosti s popisovaným modelem a adaptivní dynamikou sítě se jedná o adaptace synaptických vah a biasů skrytých a výstupních neuronů. Konkrétně pro dopřednou síť byl vyvinut učící mechanismus ve světové literatuře známý pod názvem „*backpropagation*“. V češtině se pro tento způsob učení vžilo pojmenování *učení se zpětným šířením chyby*.

Tato metoda spadá mezi takzvané způsoby *učení s učitelem*, viz klasifikace neuronových sítí na str. 42. Existuje tedy množina trénovacích vzorů (nechť počet trénovacích vzorů je s) ve formě vektoru $\mathbf{v}_{(k)} = (\mathbf{x}_{(k)}, \mathbf{d}_{(k)})$, kde $\mathbf{x}_{(k)}$ je (k) -tý trénovací vstup a $\mathbf{d}_{(k)}$ je (k) -tý k němu požadovaný výstup. Tyto hodnoty předkládáme neuronové síti, aby si je svým způsobem spojila (asociovala si je, naučila se). Při této metodě jsou v prvním kroku⁴ $t = 0$ náhodně nastaveny složky vektoru zobecněných synaptických vah $(\mathbf{w}_{i,j})^0$ na hodnoty z intervalu $(0;1)$ a v průběhu učení ($t > 0$) jsou pak vhodně adaptovány⁵. Neuronová síť je trénována na tzv. trénovací množině (trénování = adaptace hodnot vah a biasů) a v tomto procesu se pravidelně střídá aktivní a adaptivní dynamika práce sítě. Neuronová síť vypočte v t -tém kroku výstup $\mathbf{y}_{p+1,(k)}^t$ příslušný aktuální váhové konfiguraci $\mathbf{w}^t$ a vstupu $\mathbf{x}_{(k)}$. Podle míry neshody výstupu sítě $\mathbf{y}_{p+1,(k)}^t$ s požadovaným výstupem $\mathbf{d}_{(k)}$ se pak váhy adaptují. K posouzení neshody neuronové sítě pro (k) -tý trénovací vzor slouží *chybová funkce*, která popisuje vývoj *parciální chyby sítě* v čase. Parciální chyba sítě $E_{(k)}^t$ v čase t , jež je závislá na k -tém vzoru a na vektoru aktuální váhové konfigurace $\mathbf{w}^t$, může mít tvar⁶:

$$E_{(k)}^t = \frac{1}{2} (\mathbf{y}_{p+1,(k)}^t - \mathbf{d}_{(k)}) (\mathbf{y}_{p+1,(k)}^t - \mathbf{d}_{(k)})^T, \quad (5.4)$$

$k = 1, \dots, s$.

Celková chyba sítě E^t v čase t , s níž síť na trénovací množině pracuje, je závislá na celé trénovací množině a vektoru aktuální váhové konfigurace, tedy:

$$E^t = \sum_{k=1}^s E_{(k)}^t. \quad (5.5)$$

⁴Práce sítě obvykle probíhá v diskretních krocích, viz poznámka v podkapitole 4.2.

⁵Existují však i jiné algoritmy nastavení vah, které mohou vést k rychlejšímu a efektivnějšímu způsobu učení – viz např. Škutová [87].

⁶ $E_{(k)}$, uvedená ve vzorci (5.4), je tzv. Euklidovskou vzdáleností mezi požadovaným a reálným výstupem sítě. Vedle Euklidovy vzdálenosti se někdy pro trénování sítě používá tzv. *entropie*. Entropie je definovaná jako suma násobků požadované hodnoty a logaritmu chyby na každé jednotce, což odpovídá maximálně-věrohodnému odhadu nastavení parametrů sítě.

Cílem adaptace sítě je minimalizace celkové chyby sítě (5.5). Tento problém není vzhledem ke komplikovaným vztahům v prostoru sítě triviální. K jeho řešení se používá nejjednodušší gradientní metoda, pro niž jsou díky volbě diferencovatelných přenosových funkcí, kterými logistická sigmoida i lineární funkce jsou, předpoklady splněny. Adaptace probíhá v diskrétních krocích jednotlivých učících cyklů a změna vah je rovna zápornému gradientu chybové funkce pro $\mathbf{w}_{i,j}^{t-1}$ a k -tý trénovací vzor. Nová hodnota vektoru synaptických vah $\mathbf{w}_{i,j}^t$, kde t značí cyklus učení, se vypočítá jako:

$$\mathbf{w}_{i,j}^t = \mathbf{w}_{i,j}^{t-1} - \alpha \cdot \frac{\partial E^{(k)}}{\partial \mathbf{w}_{i,j}} (\mathbf{w}_{i,j}^{t-1}, \mathbf{v}^{(k)}), \quad (5.6)$$

pro všechny váhové vektory, kde použité značení je ve shodě s předcházejícím. Adaptace vah definovaná vztahem 5.6 je založena na Hebbově hypotéze (viz Hebb [27]) zmíněné v podkapitole 3.3.

Vzorec (5.6) lze rozepsáním pro jednotlivé váhy $w_{i,j,c}$, $i = 1, \dots, p$, $j = 1, \dots, n_i$, $c = 0, \dots, n_{i-1}$ a s využitím věty o derivaci složené funkce zapsat jako (5.7)⁷ a (5.8). Vzorec (5.7) se používá pro adaptaci synaptických vah neuronů výstupní vrstvy, vzorec (5.8) pro úpravu synaptických vah skrytých neuronů. Použité značení je opět ve shodě s předcházejícím.

$$\begin{aligned} w_{p+1,j,c}^t &= w_{p+1,j,c}^{t-1} - \alpha \cdot \frac{\partial E^{(k)}}{\partial w_{p+1,j,c}} (w_{p+1,j,c}^{t-1}, \mathbf{v}^{(k)}) = \\ &= w_{p+1,j,c}^{t-1} - \alpha \cdot \frac{\partial E^{(k)}}{\partial y_{p+1,j}} \cdot \frac{\partial y_{p+1,j}}{\partial \xi_{p+1,j}} \cdot \frac{\partial \xi_{p+1,j}}{\partial w_{p+1,j,c}} (w_{p+1,j,c}^{t-1}, \mathbf{v}^{(k)}) = \\ &= w_{p+1,j,c}^{t-1} - \alpha \cdot (y_{p+1,j,(k)} - d_{j,(k)}) \cdot y_{p,j,(k)} \end{aligned} \quad (5.7)$$

$$\begin{aligned} w_{i,j,c}^t &= w_{i,j,c}^{t-1} - \alpha \cdot \frac{\partial E^{(k)}}{\partial w_{i,j,c}} (w_{i,j,c}^{t-1}, \mathbf{v}^{(k)}) = \\ &= w_{i,j,c}^{t-1} - \alpha \cdot \frac{\partial E^{(k)}}{\partial y_{i,j}} \cdot \frac{\partial y_{i,j}}{\partial \xi_{i,j}} \cdot \frac{\partial \xi_{i,j}}{\partial w_{i,j,c}} (w_{i,j,c}^{t-1}, \mathbf{v}^{(k)}) = \\ &= w_{i,j,c}^{t-1} - \alpha \left(\sum_{r=1}^{n_{i+1}} \frac{\partial E_k}{\partial y_{i+1,r}} (w_{i,j,c}^{t-1}, \mathbf{v}^{(k)}) \cdot y_{i+1,r} (1 - y_{i+1,r}) w_{i+1,r,c} \right) \cdot \\ &\quad \cdot (y_{i-1,j,(k)} (1 - y_{i-1,j,(k)})) \cdot y_{i-1,j,(k)} \end{aligned} \quad (5.8)$$

⁷Formálně je derivace u vzorce 5.7 zapisována pro $\mathbf{v}^{(k)}$. Je však zřejmé, že pro změnu konkrétní váhy $w_{p,j,c}$ je třeba sice celý vstup $\mathbf{x}^{(k)}$, ale pouze požadovaný výstup j -tého výstupního neuronu, tj. $d_{j,(k)}$. To již neplatí pro vzorec 5.8!

Vzorce (5.7) a (5.8) vypadají velice komplikovaně, avšak jejich složitost pramení především z nutnosti korektního indexování jednotlivých výstupů neuronů a jim příslušných synaptických vah. „Přehlednější“ způsob zápisu výpočtů definuje ve své práci Hristev [33]. Takové zavedení, jak ho definuje Hristev [33], je pro tuto práci neúčelné, protože používá velice speciální matematický aparát pro práci s maticemi a vektory. V práci Šíma – Neruda [86] jsou vzorce pro zpětné šíření chyby odvozeny o něco obecněji. Jednak je zavedeno jiné indexování a jako přenosová funkce je volena sigmoida se strmostí charakterizovanou parametrem λ (viz vzorec 5.9) obecně různým pro každý neuron j .⁸

$$\sigma_j(\xi_j) = \frac{1}{1 + e^{-\lambda_j \cdot \xi_j}} \quad (5.9)$$

Ve vzorcích (5.7) a (5.8) se oproti již definovaným symbolům vyskytuje parametr α . Parametr α se nazývá *parametr učení*. Parametr učení určuje rychlost učení a je volen z intervalu $(0; 1)$. Pokud je zvolen parametr učení příliš malý, konverguje metoda sice jistě, ale za to velmi pomalu a ve většině případů jen k některému lokálnímu minimu chybové funkce. Zvolíme-li parametr příliš vysoký ($\alpha \rightarrow 1$), chybová funkce a tím i nastavení zobecněných synaptických vah obvykle konverguje, ale může také divergovat. Ve vzorci (5.8) bylo pro výpočet použito parciálních derivací „chyb“ neuronů, do nichž vede spoj od j -tého neuronu i -té vrstvy. To znamená, že při výpočtu zobecněných synaptických vah postupujeme v každém kroku od výstupní vrstvy postupně až k první vstupní vrstvě. Odtud název učení se zpětným šířením chyby. Proces adaptace vah probíhá přes všechny trénovací vzory, a to většinou v několika opakujících se cyklech trochu nadneseně dle staré antické zásady „repetitio est mater studiorum“⁹.

Adaptace vah končí v okamžiku, kdy je buď dosaženo absolutní shody výstupů sítě se vzory, nebo se celková chyba sítě snížila pod danou mez. Další ukončující podmínkou procesu může být zadaný počet opakování, nebo pokles změny celkové chyby sítě pod danou mez. Jak již bylo řečeno, metoda zpětného šíření chyby založená na gradientní metodě často vede k nalezení lokálního minima chybové funkce. Aby se tento nedostatek odstranil, používají se často modifikované metody zpětného šíření chyby, například metody zahrnující *moment setrvačnosti*, či metody se zvětšující se hodnotou parametru učení. Více lze nalézt v Šíma – Neruda [86], Holeňa [30] nebo Haykin [25].

⁸Sigmoida, uvažovaná v předvedeném odvození, by měla ve vyjádření vzorce 5.9 všechny parametry λ_j rovna jedné.

⁹Opakování jest matkou studia (učení, moudrosti).

5.3.3 Shrnutí

Trénování neuronových sítí pomocí algoritmu backpropagation je v současné době nejpopulárnější a je využíváno například při trénování neuronových sítí pro predikci v časových řadách. Při modifikaci základní varianty je dosahováno slušných výsledků bez nebezpečí uvíznutí v lokálním minimu celkové chyby sítě, či divergenci v nastavení vah. Nevýhodou oproti klasickým (statistickým) metodám je nemožnost odhadnutí chyby výsledku dosaženého prostřednictvím neuronové sítě.

5.4 Genetické učení

Přestože backpropagation je jistě stěžejním učícím algoritmem, bude v rámci této rešerše zmíněn ještě jeden přístup, který je také založen na biologické motivaci stejně jako samotné neuronové sítě. Navíc genetický algoritmus slouží v používaném softwaru STATISTICA Neuronové sítě 7.1 vedle citlivostní analýzy k výběru vstupních proměnných modelu.

5.4.1 Motivace ke genetickému učení

Myšlenka genetických algoritmů je založena na základě biologické genetiky a Darwinovy evoluční teorie. Postupně jsou tvořeny generace jedinců, jež se předem daným způsobem vyvíjejí z generace předchozí. Při matematickém popisu genetických algoritmů bývá používána terminologie známá z biologických oborů¹⁰, viz tabulka 5.1 (levý sloupec biologický význam, pravý pak matematická paralela).

V souvislosti s Darwinovou evoluční teorií je princip genetických algoritmů následující: *Cílem je vybrat optimálního jedince (algoritmus), který se nejlépe hodí pro dané prostředí (daný úkol)*. Na počátku je proto z parametrického prostoru, jenž je určen alelami, náhodně vygenerována množina jedinců, určená svými genotypy (1. generace). Pro jednotlivé jedince se spočte tzv. *kritériální funkce*, což je funkce charakterizující vhodnost jedince (v případě neuronových sítí se jedná o celkovou chybu sítě). Následující generace je pak předem daným způsobem (viz dále) tvořena z předchozí generace. Tento postup vycházející ze skutečných principů přírody je jedním z nejkompexnějších přístupů. Druhá generace vzniká z první generace, obecně

¹⁰Jakkoliv se jeví oblasti informatiky a genetiky od sebe vzdálené, oblast genetických algoritmů (podobně jako neuronových sítí) dokazuje, že motivace přírodou pro vývoj nových technologií je velice užitečná.

Tabulka 5.1: Terminologie genetických algoritmů převzatá z biologické motivace – levý sloupec biologický význam, pravý matematická paralela

Populace	
sled generací vývojově	uspořádaná množina generací
Generace	
skupina jedinců, na níž je aplikována teorie o vývoji druhů (selekce)	množina jedinců, na níž je prováděn výběr pro vytvoření nové generace
Jedinec	
nositel genetické informace	reprezentant vektoru genotypu
Genotyp	
genetický popis organismu	celkový vektor, jenž kóduje řešený problém
Fenotyp	
fyzický projev genotypu; např. je-li u člověka příslušný chromozóm typu XY (= genotyp), je daný člověk mužem (= fenotyp)	výsledek transformace zakódování; např. binární vyjádření 101 (= genotyp) odpovídá číslu 5 v dekadickém, tedy pro člověka běžném, zápisu (= fenotyp)
Chromozóm	
část molekuly DNA	vektor obsahující jednu či více složek celkového vektoru, jenž kóduje řešený problém, např. sekvence symbolů z nějaké abecedy (mohou to být čísla, znaky nebo jejich kombinace)
Gen	
elementární jednotka genetické informace	jedna určitá složka vektoru (chromozomu), konkrétní symbol
Alela	
určité místo (pozice) v chromozómu, jedna z konkrétních forem genu	množina hodnot, kterých mohou geny nabývat
Fitness	
síla jedince v generaci, od ní se odvozuje šance k přežití	síla jedince v generaci, odvozuje se z kritériální funkce

$(n + 1)$ -ní generace vzniká z generace n -té. Obvykle jsou respektována 4 pravidla (principy), která říkají, jak by měla následující generace z té předcházející vzniknout.

1. Do další generace na základě nejvyšší vhodnosti postupuje část jedinců přímo, princip tzv. *elitářství* (dále princip (1)).
2. Druhá část nově vznikající generace je z předchozí generace vybrána náhodně, což představuje princip *nepohlavního rozmnožování* (dále princip (2)).
3. Třetí část vzniká *zkombinováním genotypů dvou rodičů* s výsledkem dvou potomků, u nichž je genotyp vytvořen náhodnou záměnou některých sobě odpovídajících genů (dále princip (3)). Záměna genů může být přesně dána, nebo je volena náhodně.
4. Poslední skupina nové generace je odvozena od těch předchozích, nově vzniknuvších skupin, pomocí *mutace* – náhodné změny genu (dále princip (4)). V modelech genetických algoritmů bývá mutace řešena pomocí generátorů náhodných čísel a určený gen je buďto zaměněn náhodným číslem, nebo je gen původní změněn v závislosti na vygenerovaném čísle.

5.4.2 Genetické učení neuronových sítí

Pro popis trénování neuronové sítě prostřednictvím genetických algoritmů bude použita architektura a topologie předložená při výkladu zpětného šíření chyby. Navíc bude použita celková chyba sítě (vzorec (5.5)) představovat míru vhodnosti jedince.

V prvním kroku je náhodně vygenerováno dostatečné množství (m) jedinců (=neuronových sítí s různým nastavením volných parametrů), jejichž genotypem jsou hodnoty volných parametrů trénované sítě. Pro každého jedince je vypočtena hodnota kriteriální funkce $E^{l,1}$, $l = 1, \dots, m$, jež slouží k ukončení genetického učení, respektive k ohodnocení každého z jedinců. Z hodnoty kriteriální funkce $E^{l,1}$ je dále vypočítána vhodnost $v^{l,1}$. Vhodnost je transformovaná hodnota kriteriální funkce do intervalu $\langle 0, 1 \rangle$, která umožňuje vybrat do další generace zástupce s možností zvýhodňovat lepší jedince a přitom zachovat princip náhody. Za „vhodnostní“ transformaci se nabízí zvolit (Biskup – Čermáková [109]):

$$v^{l,1} = 1 - \frac{E^{l,1}}{\max_{1 \leq l \leq m} \{E^{l,1}\}}, \quad (5.10)$$

nebo

$$v^{l,1} = 1 - \frac{Q^{l,1}}{\sum_{l=1}^m Q^{l,1}}, \quad (5.11)$$

kde $Q^{l,1} = C - E^{l,1}$, pro $l = 1, \dots, m$ a $C \in \mathbb{R}$ je maximální chyba sítě či její odhad.

Vzorec (5.10) může být s výhodou použit při *turnajovém výběru jedinců*, kdy proti sobě soupeří dva jedinci. Výsledek jejich soupeření závisí pouze na náhodě a vhodnosti. Tedy každému jedinci je vygenerováno náhodně číslo, jež je váženo vhodností, jedinec s vyšším číslem vítězí. Tímto způsobem jsou přímo vybráni elitní jedinci – princip (1) (viz principy zmíněné od str. 59), tj. ti, kteří měli nejvíce vítězství, a dále pak kandidáti ke křížení – princip (2). Náhodně, bez ohledu na vhodnost, jsou do nové generace zařazeni další jedinci – princip (3). Nakonec je na náhodně vybraných jedincích provedena mutace, a to buď na jedincích nově vzniknuvších, nebo jsou vybraní jedinci zdvojeni a mutace je provedena na kopii.

Vzorec (5.11) bývá volen při postupu nazývaném *ruletová selekce*. Nejprve jsou vybráni elitní jedinci podle absolutní velikosti vhodnosti. Výběr pro křížení pak probíhá tak, že každému jedinci je přiřazena výše kruhu (rulety) dle hodnoty vhodnosti a tím šance postoupit do další generace (nechat si „vytočit“ postup na ruletě). Principy náhodnosti a mutace jsou obdobné postupu popsanému výše.

Již popsanými principy je vytvořena nová generace, jež je zdrojem pro generaci následující. Výběr a modifikace jedinců probíhá opět podle pravidel (1)–(4), uvedené vzorce zůstávají v platnosti, jen index „1“ se změní na příslušný index generace. Proces generování nových generací je ukončen v okamžiku, kdy je dosaženo hodnoty předepsané celkové chyby sítě, nebo proveden předepsaný počet kroků.

5.4.3 Shrnutí

Postup při genetickém učení je formálně méně složitý než backpropagation. Při výběru přenosové funkce není nutné omezení na její dvojnásobnou diferencovatelnost nutnou pro gradientní metodu. Genetický algoritmus prohledávající parametrický prostor s m jedinci v populaci pracuje zhruba tak rychle, jako m^3 izolovaných hledačů (viz Šíma – Neruda [86]). Mechanismus genetického algoritmu v němž jedinci procházejí parametrický prostor a navzájem se ovlivňují pomocí předem daných pravidel – genetických operátorů, bývá označován jako *implicitní paralelismus*. Tím je dosahováno synergetického efektu, díky němuž populace jedinců dosáhne kýženého řešení rychleji. Křížení a mutace zabraňují uvíznutí algoritmu v lokálních minimech celkové chyby sítě.

5.5 Přetrénování a přeurčení neuronové sítě

Na závěr kapitoly věnované učení neuronových sítí je nutné uvést, že dokonalé trénování nemusí být pro výsledné vlastnosti neuronové sítě vždy to nejlepší, neboť při něm může síť ztratit velmi ceněnou schopnost *generalizaci*. Proto se množina dat, jež reprezentují reálný problém, rozděluje na *trénovací* a *testovací* množinu. Pomocí trénovací množiny se adaptují synaptické váhy a testovací množina slouží k ověření trénovanosti sítě, a to včetně její generalizační schopnosti (viz Šíma – Neruda [86]). V ideálním případě konverguje celková chyba sítě na trénovací množině k nule (= stav stoprocentní „natrénovanosti“ sítě na trénovací množinu), zatímco chyba sítě na testovací množině nejprve klesá a následně roste. Stav minima chyby na testovací množině je optimálním okamžikem pro ukončení učení, neboť od tohoto okamžiku dál síť ztrácí svou schopnost zobecňovat a je příliš fixována na trénovací množinu – stává se přetrénovanou (overlearning).

Vzhledem k tomu, že i data z testovací množiny se zapojují do trénování, bývá doporučováno rozdělení dat dokonce na tři množiny. Vedle trénovací a testovací množiny se doporučuje vytvoření tzv. *validační* množiny, na níž jsou ověřovány vlastnosti neuronové sítě. Někteří autoři např. Haykin [25] význam validační a testovací množiny zaměňují. Validační množina se pak podílí na trénování, zatímco testovací množina slouží pro odhad vlastností sítě na neznámých datech. Takto je na množiny nahlíženo i v předkládané disertační práci.

Značná fixace na trénovací množinu se projevuje také při tzv. *přeurčení* (overfitting), tj. při zbytečně bohaté architektuře sítě. Množství volných parametrů vstupujících do procesu trénování, jež je přímo úměrné počtu výkonných jednotek, umožňuje z pohledu generalizace nevhodné přizpůsobení sítě tréninkové množině. Zvláště v případě pozdě ukončeného procesu trénování.

I přes mnohá doporučení pro volbu architektury sítě (např. Šíma – Neruda [86], Haykin [25], Vondrák [93], aj.) je konstrukce obvykle heuristická. Buďto je začínáno s malou architekturou a ta je následně obohacována, nebo – naopak – výchozí bohatý model je postupně zjednodušován – *prořezáván* (pruning). Zjednodušování modelu nabízí také program STATISTICA Neuronové sítě 7.1, který disponuje podporou malých vah (Šíma – Neruda [86], StatSoft [81]), nebo vypouštěním vstupních neuronů na základě citlivosti (StatSoft [81]).

Kapitola 6

Softwarová prostředí a neuronové sítě

Kapitola 6 obsahuje jednak představení softwarového prostředí využívaného při zpracovávání této disertační práce (podkapitola 6.1) a dále pak programů, které byly využívány předtím, než byl autorovi program STATISTICA Neuronové sítě 7.1 k dispozici (podkapitola 6.2 a 6.3). Podkapitola 6.4 přináší výčet několika dalších softwarových prostředí, ve kterých je technologie neuronových sítí implementována.

6.1 STATISTICA Neuronové sítě

Program STATISTICA Neuronové sítě 7.1 je produktem softwarové firmy StatSoft¹ a je lokalizován do českého jazyka. Jedná se o balík, který může fungovat samostatně nebo v návaznosti na standardní i rozšiřující produkty STATISTICA firmy StatSoft (samozřejmě jsou kompatibilní jen stejné verze; v současnosti je na trhu verze 8). Bližší představení produktu tak, jak ho na svých webových stránkách prezentuje StatSoft CR, nabízejí následující kapitoly popisující již verzi 8 (odlišnosti oproti předchozí využívané verzi budou pro zajímavost vyznačeny).

Konkrétní používané funkce softwaru budou ještě zmíněny a komentovány v aplikační práci při řešení konkrétních úkolů.

¹Více o produktech i společnosti jako takové lze nalézt na: StatSoft CR [80].

6.1.1 Motivace pro použití programu STATISTICA

Popularita metodologie neuronových sítí rychle stoupá v nejrůznějších oblastech od základního výzkumu po vytěžování dat (data mining), v oblastech predikcí v komerční sféře, řízení rizik, marketingu a dalších. Systém *STATISTICA Automatizované neuronové sítě*² (dále jen *SANS*) je jednou z nejpokročilejších a nejvýkonnějších aplikací technologie neuronových sítí dostupných na současném trhu. Systém nabízí množství jedinečných prostředků a je určen nejen pro specialisty v oboru neuronových sítí (kterým nabízí mimořádný rozsah typů sítí a trénovacích algoritmů), ale i pro nové uživatele této technologie (pro které je k dispozici *Automatický vyhledávač sítě*, nástroj, který uživatele provede všemi kroky potřebnými pro vytváření neuronových sítí).

6.1.2 Způsob implementace neuronových sítí

SANS podporuje mnoho důležitých tříd z neuronových sítí pro řešení problémů reálného světa včetně vícevrstevných perceptronů, RBF sítí a samoorganizujících map.

Výjimečná snadnost použití kombinovaná s překvapivým analytickým výkonem, tak charakterizuje StatSoft [81] *SANS*. Dále uvádí, že představovaný modul obsahuje automatický vyhledávač sítě, který provede uživatele krok po kroku postupem pro vytvoření skupiny rozdílných sítí a výběrem nejvhodnější sítě s nejlepším výkonem (což je úloha, která by jinak vyžadovala zdoluhavý postup metodou „pokusů a omylů“ při solidních znalostech základní teorie).

SANS v sobě integruje pre- a post-processing obsahující výběr dat, kódování jmenovitých hodnot, škálování, normalizaci a nahrazování chybějících hodnot s interpretací pro klasifikaci, regresi a problematiku časových řad, vyspělé, vysoce optimalizované trénovací algoritmy včetně metody konjugovaných gradientů a metody Broyden-Fletcher-Goldfarb-Shanno (BFGS)³. Při tom všem nabízí *SANS* plnou kontrolu nad všemi aspekty, které ovlivňují chování neuronové sítě, jako je aktivací a chybová funkce nebo komplexnost sítě a navíc nabízí podporu pro kombinování sítí a síťových architektur následně organizovaných v sadách neuronových sítí pro vytváření kolekcí.

Plná integrace v rámci systému STATISTICA je samozřejmostí. Všechny výsledky, grafy, reporty atd. mohou být dále upravovány pomocí grafických a analytic-

²Ve verzi 7 se tento modul jmenoval jen *Neuronové sítě*.

³Tento algoritmus není ve verzi 7.1 k dispozici.

kých nástrojů systému STATISTICA. Lze tak např. provádět další analýzy reziduí, vytvářet komentované souhrnné reporty apod. Jedná se tedy o komplexní grafickou a statistickou zpětnou vazbu poskytující interaktivní průzkumné analýzy, která může pokračovat využitím kompletního makra pro všechny analýzy, programováním vlastní analýzy za pomoci neuronové sítě v prostředí STATISTICA Visual Basic nebo volat systém *SANS* z libovolné aplikace podporující COM. To znamená automaticky provádět analýzy s neuronovými sítěmi v tabulkách MS Excel, či zahrnout procedury neuronových sítí do vlastních aplikací vyvinutých v C, C++, C#, Java apod⁴.

SANS obsahuje funkcionalitu, která napomáhá v kritických okamžicích návrhu neuronových sítí. Zahrnuje nejen vyspělé architektury neuronových sítí a trénovací algoritmy, ale též inovativní přístup k návrhu síťové architektury s použitím specifických a smysluplných chybových funkcí umožňujících interpretaci dosažených výsledků.

6.1.3 Práce s daty

Systém *SANS* je plně kompatibilní se systémem STATISTICA, tudíž je zde k dispozici velký výběr nástrojů pro editaci (přípravu) dat pro analýzy (transformace, výběr podmínek případů, nástroje pro ověření dat atd.), protože používání neuronových sítí představuje více než jen „krmení“ neuronové sítě daty. Data musí být obvykle pro vstup do neuronové sítě připravena. Důležité je to též pro korektní interpretaci výstupu z neuronové sítě. *SANS* obsahuje automatické škálování dat pro vstup i výstup, je zde i automatický záznam nominálních proměnných (např. pohlaví = muž, žena), obsahuje one-of-N kódování. *SANS* též obsahuje prostředky pro zacházení s chybějícími daty. Má k dispozici prostředky pro přípravu a interpretaci dat z časových řad. Navíc jako všechny analýzy STATISTICA, program může být „připojen“ k databázi prostřednictvím nástroje pro in-place-database processing, nebo může být navázán na aktivní data tak, že modely jsou „přetrénovány“ nebo použity (např. pro vypočítání predikovaných hodnot nebo klasifikaci) automaticky pokaždé, jakmile se změní data.

⁴Vše lze samozřejmě provádět jen v případě současného vlastnictví další balíků obsahujících zmíněné procedury.

6.1.4 Automatické vyhledání sítě

Automatické vyhledávání sítě zahrnuté v systému *SANS* je nástroj poskytující výpočet velkého počtu různých architektur neuronových sítí různé komplexnosti a výběr toho nejlepšího souboru specifických architektur pro řešený problém.

SANS si automaticky ponechává kopie nejlepších nalezených sítí. Ty mohou být kdykoliv znovu využity. Užitečnost a prediktivní validita sítě může být automaticky odhadnuta se zahrnutím testovacího a validačního vzorku a vyhodnocením velikosti a efektivity sítě jakož i ceny nesprávné klasifikace.

Rozšíření výkonu *SANS* podporuje několik voleb pro úpravu sítě. Je možno zvolit lineární výstupní vrstvu pro sítě použité v regresní úloze nebo softmax aktivační funkci pro odhad pravděpodobnosti v klasifikační úloze. Chybová funkce křížové entropie, založená na modelech teorie informací, je taktéž zahrnuta a je zde i řada specializovaných aktivačních funkcí včetně exponenciální, hyperbolického tangentu, logistické sigmoidy a funkce sinus pro skryté a výstupní neurony.

6.1.5 Trénování neuronové sítě

Pokud je třeba experimentovat s architekturami a typy neuronových sítí, je možno se spolehnout na kvalitu a rychlost trénovacího algoritmu sítě. *SANS* podporuje nejlepší známé vyspělé trénovací algoritmy, jako metodu konjugovaného gradientu a BFGS⁵, obsahuje řadu trénovacích algoritmů pro shlukovou analýzu, které jsou založeny na dobře známém Kohonenovu algoritmu pro samoorganizující se mapy.

Iterativní trénovací procedura implementovaná v *SANS* je doplněna automatickým sledováním obou chyb, trénovací a nezávislé testovací chyby. Trénování může být kdykoliv ukončeno manuálně, nebo lze specifikovat podmínku pro zastavení v závislosti na dosažení cílové chyby, nebo zhoršující se výběrové chybě. I když dojde k přeučení, *SANS* si automaticky ponechá kopii nejlepší objevené sítě, která je automaticky použita jako nejlepší řešení. Když je trénování ukončeno, lze na závěr vyzkoušet výkonnost na trénovacích, testovacích a validačních vzorcích.

SANS umožňuje vlastnosti natrénované sítě zobrazit např. prostřednictvím bodových grafů či 3D grafů povrchů odezvy, které napomáhají uživateli porozumět chování sítě.

Uživatel si může vybrat více modelů (a kolekcí). *SANS* pak bude zobrazovat, po-

⁵Tzv. rychlé trénovací algoritmy druhého řádu, které mohou konvergovat mnohem rychleji než algoritmy prvního řádu (gradientní metoda).

kud je to možné, výsledky komparativním způsobem (např. zobrazením odezвовých křivek pro několik modelů v jednom grafu, nebo prezentací předpovědí více modelů v jedné tabulce). Tato vlastnost je zvláště užitečná pro srovnání více modelů trénovaných na stejné datové sadě, navíc všechny statistiky jsou generovány nezávisle pro trénovací, testovací a validační vzorek.

Vypočítané celkové statistiky obsahují střední chybu sítě, tzv. matici záměn pro klasifikační problém (sumarizuje správné a nesprávné klasifikace přes všechny třídy) a korelaci pro regresní problém – vše je vypočítáno automaticky. Kohonenovy sítě obsahují okno topologické mapy, které umožňuje vizuálně prozkoumat aktivace jednotek během analýzy dat.

6.1.6 Příklady reálných aplikací

Technologii neuronových sítí lze použít, jak uvádí StatSoft [81] prakticky v každé situaci, v níž je cílem nalezení neznámé proměnné nebo vlastnosti na základě známých pozorování nebo naměřených hodnot (tzn. v nejrůznějších formách regresí, klasifikací a časových řad), je-li k dispozici dostatečné množství „historických“ dat a existují-li mezi nimi objektivní vztahy nebo množina vztahů (neuronové sítě jsou relativně tolerantní k šumovým jevům). Dále lze neuronové sítě výhodně použít pro vyšetřovací analýzu pro případ hledání shluků dat (Kohonenovy sítě).

V Elektronické učebnici statistiky StatSoft [81], v kapitole neuronové sítě, lze nalézt obsažnou diskusi o teoretických úvahách týkajících se oblastí vhodných pro úspěšné využití technologie neuronových sítí. Následující seznam uvádí výběr reprezentativních příkladů použití. V žádném případě však seznam nepokrývá veškeré oblasti užití této technologie.

6.1.7 Volitelný generátor zdrojového kódu sítě

Generátor zdrojového kódu je volitelný přídatný modul, který dává uživateli možnost stavět jeho vlastní aplikace na základě řešení nalezeného systémem *SANS*. Modul vytváří zdrojový kód pro sestavení neuronové sítě (v C nebo PMML), který lze následně zkompileovat do vlastního programu. Modul je určen především pro podnikové programátory, kteří potřebují převést nalezená vysoce optimalizovaná řešení do svých stávajících aplikací určených k řešení opakujících se analytických problémů⁶.

⁶Generátor je přídatná **placená** aplikace.

6.1.8 Zkušenosti s využitím softwaru STATISTICA 7.1 Neuronové sítě

Program STATISTICA 7.1 Neuronové sítě byl zakoupen díky „štědrosti“ Zemědělské fakulty Jihočeské univerzity v Českých Budějovicích z interního grantu (14/06): *Možnosti neuronových sítí a data miningu pro hodnocení faktorů harmonizace regionálního rozvoje sociálně ekonomického potenciálu regionů*. Jedná se o komplexní komerční software, který umožňuje rozsáhlé využití neuronových sítí (viz výše) v relativně uživatelsky příjemném prostředí. Výsledky získané díky tomuto softwaru byly prezentovány v příspěvcích Biskup – Čermáková [114], Čermáková – Biskup [117], Biskup [108], Čermáková – Biskup [116], Čermáková – Biskup [118] a Čermáková – Biskup [119]. Právě na výpočtech provedených v tomto programu je založena aplikační část disertační práce, která mimo jiné hodnotí i práci tohoto programu.

6.2 Neurex

Program Neurex 4.0 je výtvozem Ivo Vondráka a jedná se o expertní systém založený na bázi neuronových sítí. Díky bohatší dokumentaci (Vondrák [93]) a jednoduchosti programu bude věnována pozornost právě této verzi, přestože existuje i verze 5.1 vytvořená v roce 2005. Následující podkapitola se opírá právě o jím vytvořenou publikaci Vondrák [93] a tvoří paralelu k obecněji popsanému principu učení.

6.2.1 Koncepce systému

Pomocí programu Neurex 4.0 lze v podstatě řešit úlohy diagnostického typu, kde je dána na jedné straně množina jakýchsi symptomů (příznaků), ke kterým hledána odezva (diagnóza). Tento typ úloh je běžný v celé řadě výrobních technologií, kdy je třeba na základě projevů výrobního systému usuzovat na jeho stav nebo případnou poruchu. Také problematika hodnocení kvality výrobků je v podstatě diagnostikou, kde příznaky jsou tvořeny charakteristikami, např. u odlitků je to doba, teplota či způsob lití, chemické složení apod. Cílem je pak určit pravděpodobnost té či oné vady.

Otázkou je, co vše budeme potřebovat, abychom mohli náš problém implementovat prostřednictvím neuronové sítě systému Neurex. Pokusme se tyto požadavky definovat v následujících několika bodech:

Co vše je nutné, aby bylo možno konkrétní problém implementovat prostřednictvím neuronové sítě systému Neurex, shrnují následující body:

1. Je nutné stanovit vstupní a výstupní faktory popisující řešený problém. Za předpokladu, že každý vstupní faktor bude reprezentován jedním neuronem vstupní vrstvy a stejně tak každý výstupní faktor bude reprezentován jedním neuronem výstupní vrstvy, pak je dán počet těchto neuronů. Úkolem zůstává pouze stanovení, který neuron v pořadí odpovídá kterému faktoru, tedy jest nutno provést přiřazení faktorů jednotlivým neuronům. Z hlediska úplné definice architektury sítě zbývá ještě stanovit, kolika vnitřními vrstvami a s jakým počtem neuronů bude tvořena naše neuronová síť. Zde neexistuje žádné exaktní pravidlo, nicméně platí, že je nutná alespoň jedna vnitřní vrstva, což je také ve většině případů i postačující počet. Počet neuronů této vnitřní vrstvy je určen následujícími podmínkami:

- (a) počet neuronů vnitřní vrstvy by měl být menší než je počet neuronů vstupní či výstupní vrstvy podle toho, která z nich obsahuje větší počet neuronů. Nedodržení této podmínky obvykle vede k snížení schopnosti generalizovat vytvořené znalosti a výrazně se prodlužuje doba adaptace.
- (b) Počet neuronů vnitřní vrstvy by měl být větší než je počet neuronů vstupní či výstupní vrstvy podle toho, která z nich je menší. Nedodržení této podmínky vede k tzv. kompresi dat a existuje určité nebezpečí ztráty schopnosti rozpoznat všechny prvky našeho problémového prostoru. Neuronová síť se pak stává neadaptovatelnou.

Například architektury 30–15–5 a 10–9–7 respektují podmínky na rozdíl od 20–30–5 a 20–5–8, které nerespektují první resp. druhou podmínku.

2. V následujícím je nutné vytvořit trénovací množinu obsahující vzory, které popisují řešenou problematiku. Každý takový vzor je tvořen jednak vektorem vyjadřujícím hodnoty vstupních faktorů a dále vektorem vyjadřujícím hodnoty výstupních faktorů. Tyto hodnoty jsou obecně reprezentovány reálnými čísly (pravdivostní hodnoty lze vyjádřit procenty) z přípustného, předem definovaného intervalu. Odtud pak vyplývá, že horní mez intervalu vyjadřuje maximální a dolní mez intervalu vyjadřuje minimální stav excitace vstupního neuronu. Zde je nutné si uvědomit nebezpečí, které může vést k problémům s adaptací neuronové sítě. Vzniká v případě, že tytéž vstupní vektory u několika vzorů povedou k rozdílným výstupům. Proto je Neurex, jak uvádí Vondrák [93], vybaven kontrolou a poměrným způsobem přiřazuje jistotu rozhodnutím, jež nejsou v trénovací množině jednoznačná.

Trénovací množina může být tvořena empirickými hodnotami, nebo modelovými hodnotami definovanými experty. V prvním případě se doporučuje z celého

populačního souboru vybrat pouze některé, pokud možno typické vzory, které se použijí v trénovací množině. Po následné adaptaci sítě se pak využijí zbylé vzorky k testování schopnosti generalizovat. Jinými slovy, pokud je neuronová síť schopna na základě několika vzorů trénovací množiny odvodit všechny zbylé, pak je schopna generalizovat a má předpoklady pro spolehlivé použití. Může však nastat i situace, že některé vzory se učí velmi špatně (tyto jsou v systému Neurex průběžně identifikovány), tj. něčím se odlišují od ostatních. Může se jednat např. o špatně změřené nebo zadané vzory nebo vzory tvořící spíš výjimku než podstatu problému („out of logic“). V obou případech je vhodné tyto vzory z trénovací množiny vyřadit.

3. Adaptace neuronové sítě je klíčovým bodem vytvoření znalostí pro síť s definovanou topologií na trénovací množině vzorů. Před zahájením vlastního procesu učení je ještě nutno specifikovat několik charakteristik: inicializační hodnotu strmosti sigmoid neuronů, hodnotu koeficientu učení a hodnotu koeficientu vlivu předchozího kroku. Na počátku je k dispozici homogenní neuronová síť se sigmoidami o stejné strmosti a s náhodně generovanými synaptickými vahami. V případě, že v procesu adaptace použita metoda parametrické backpropagation, pak dojde k postupné změně strmostí při současné změně prahů a synaptických vah.

Dále je také třeba explicitně stanovit počet iterací, které proběhnou nad jednotlivými vzory při průchodu celou trénovací množinou. Poslední nastavovaná charakteristika je dána tzv. mezní podmínkou procesu adaptace reprezentující maximální chybu odezvy neuronů výstupní vrstvy. Je-li zvolena např. hodnotu 10 %, pak odezva sítě pro každý vzor musí být reprezentována excitací každého jednotlivého výstupního neuronu s maximální chybou 10 %. Proces adaptace probíhá až do okamžiku dosažení této podmínky.

4. Vlastní proces konzultace s adaptovanou neuronovou sítí již nevyžaduje další zvláštní požadavky. Je nutné pouze zadat hodnoty vstupních faktorů a následuje takřka okamžitá odezva sítě reprezentovaná předložením hodnot výstupních faktorů. Díky rychlosti odezvy a jednoduchosti vlastního zadávání vstupů se proto nabízí i možnost experimentování na bázi „co se stane, když změním ...“, tedy provádění jakési analýzy senzitivnosti. Kromě toho systém umožňuje i konzultaci s neúplnou informací na vstupu, tedy v případě, že některá z hodnot vstupních faktorů není známa (v systému je takový atribut označen symbolem otazníku namísto skutečné hodnoty). Hodnoty výstupních faktorů vypočtené neuronovou sítí pak nebudou jednoznačné a budou se pohybovat v intervalech

odrážejících neurčitost na vstupu.

Čtyři body definující koncepci systému Neurex lze vyjádřit jako:

1. definice topologie sítě, faktorů a základních charakteristik,
2. definice trénovací množiny včetně její kontroly,
3. učení – adaptace neuronové sítě,
4. konzultace s adaptovanou neuronovou sítí.

Poněkud zdlouhavá definice topologie, trénovací množiny a následné adaptace v programu Neurex je jednorázovou akcí, a proto zatěžuje jen jednou, zato konzultování s naučenou neuronovou sítí je časově nenáročné.

Součástí programu Neurex 4.0 je také modul určený pro shlukování a přípravu dat k trénování. Vše je podrobně popsáno v Vondrák [93] včetně obecného popisu práce dopředné neuronové sítě.

6.2.2 Zkušenosti s využitím softwaru Neurex 4.0

Výsledky vypočtené prostřednictvím programu Neurex 4.0 byly prezentovány prostřednictvím příspěvků Biskup – Čermáková [111], Biskup – Čermáková [112] a Biskup – Čermáková [113]. O aplikaci expertního systému se jednalo pouze v Biskup – Čermáková [112]. To vypovídá o univerzálnosti softwaru Neurex 4.0. Nevýhodou tohoto programu je poměrně malá stabilita a náročnost na hardwarové vybavení počítače.

6.3 NetVisualiser

Program NetVisualiser je vizualizační program vytvořený Bc. Petrem Marťánem. Demonstruje možnosti některých neuronových sítí, konkrétně sítí známých pod jménem:

- lineární perceptron,
- backpropagation,
- RBF síť,
- Kohonenova síť (Kohonenova mapa).

Uvedené modely neuronových sítí zvolil autor proto, že reprezentují základní typy neuronových sítí. Program byl vytvořen jako součást diplomové práce Marťán [48]. V ní je možno nalézt jak popis implementace neuronových sítí, tak obecný přehled o použitých neuronových sítích. Kompletní výsledky (diplomovou práci i program samotný) je dostupný on-line (Marťán [47]).

6.3.1 O programu

Program NetVisualiser umožňuje podle zadaných parametrů vytvářet modely neuronových sítí a následně je učit. Pro učení si uživatel může vybrat z několika různých rychlostí učení a různých režimů vizualizace procesu učení. V průběhu tohoto učení je dokonce možno měnit parametry sítě a tím ovlivňovat proces adaptace sítě. Efektivnost učení lze současně vizualizovat grafem chyby učení či statistikou sítě. Vizualizaci učení neuronové sítě lze využít v režimu klasifikace vzorů pro dvojrozměrná data a v režimu aproximace funkcí pro data jednorozměrná.

Program pracuje s modely neuronových sítí ve dvou základních režimech – klasifikace vzorů a aproximace funkcí. Natrénovanou neuronovou síť je možné kdykoli uložit a později ji opět načíst pro další trénování nebo demonstraci jejích vlastností.

Tento program také umožňuje uživateli práci s trénovacími či testovacími daty. V případě, že se jedná o dvojrozměrná data, lze tato data pomocí grafického uživatelského rozhraní vytvářet přímo v programu. Program navíc dokáže trénovací data automaticky generovat. Vždy existuje při práci s programem možnost uložení aktuálních dat a načtení uložených či externích dat. Program také podporuje automatickou normalizaci dat do intervalu $(0, 1)$ nebo zašumění vzorů.

6.3.2 Zkušenosti s využitím softwaru NetVisualiser

Díky programu NetVisualiser bylo možno realizovat trénování neuronových sítí a vytvořit tak příspěvky Biskup – Čermáková [110] a Biskup – Čermáková [113]. Jak je uvedeno výše, program má především funkci vizualizační, nicméně vypočítává váhy neuronové sítě (realizuje trénování – adaptivní dynamiku neuronové sítě), se kterými lze pracovat ve vlastní neuronové síti (aktivní dynamika neuronové sítě).

6.4 Další programy

Mezi další programy, které využívají technologii neuronových sítí, patří jak komerční, tak nekomerční programy, v následujícím přehledu. Přehled je jen ilustrativní, neboť neuronové sítě jsou v současnosti předmětem zájmu jak vědců, tak komerčního sektoru. Vzhledem k zaměření autora a jeho školitelky seznam obsahuje zejména zavedené statistické softwary.

- **NeuralDimension** je souhrnné označení pro programy, které nabízejí řešení prostřednictvím neuronových sítí implementovaných do různých programů jako

je MS Excel, MATLAB, apod.

[61]

- Jak již bylo zmíněno dříve, existuje rozšíření matematického softwaru *MATLAB*, které rozšiřuje jeho možnosti o tvorbu, implementaci a simulaci neuronových sítí. Toto rozšíření je prodáváno pod jménem Neural Network ToolboxTM.

The MathWorks [89]

- V balíku **Enterprise Miner**TM společnosti *SAS* je na výběr z několika architektur neuronových sítí, jež mají sloužit pro data mining.

SAS Institute [76]

- Nejen softwarovou podporou implementace neuronových sítí, ale i vývojem na tomto poli se zabývá *Institute of Parallel and Distributed High-Performance Systems (IPVR)*, který vyvinul program **SNNS**⁷. Program je určen pro operační systém Linux.

The SNNS Team [90]

- Stejně jako firma *SAS* i firma *SPSS* implementovala do svých data miningových balíků systém využívající možnosti neuronových sítí – **SPSS Neural Networks**TM.

SPSS [79]

- Mnoho dalších softwarů je k dispozici pouze k výukovým účelům, např. **Neuronek**⁸. Takové programy mají obvykle implementovaný jen jeden druh sítě a je u nich možno jen výjimečně měnit topologii sítě. Na trhu jsou samozřejmě i další komerční balíky (**EasyNN-plus software** – blíže [13], **BrainMaker** – blíže [4]) poskytující uživateli určitou volnost. Existence obou typů softwarů svědčí o tom, že neuronové sítě stále ještě prožívají období znovuvzkříšení.

⁷Stuttgart Neural Network Simulator

⁸Počítačový program, který realizuje fáze učení, testování a vybavování třívrstvé neuronové sítě – více v Heger – David [28].

Kapitola 7

Konkrétní příklady využití neuronových sítí

V kapitole 7 je konkretizováno využití neuronových sítí. Citované práce nejsou natolik zásadní, aby rozvíjely teorii neuronových sítí (takové práce jsou uvedeny v podkapitole 3.3), ukazují však, že neuronové sítě mají reálné aplikace napříč všemi oblastmi lidské činnosti.

7.1 Prohledávané databáze

7.1.1 Agris

Základní databáze, v níž byly vyhledávány práce s tematikou neuronových sítí byla databáze Agris. Bylo nalezeno 546 záznamů, z toho 194 článků obsahovalo alespoň abstrakt. Ve 40 případech byly neuronové sítě použity v ekonomii, či ekonomice, ve 2 případech v čistě finanční oblasti, v 7 případech v souvislosti s obchodem a v 5 případech byly prostřednictvím neuronových sítí provedeny predikce v časových řadách

Podle očekávání se všechny práce se zabývaly zemědělskou, nebo biologickou tematikou. Jednalo se hlavně o:

- predikce na základě získaných informací – predikce následného stavu, modelování budoucích stavů (mikrobiologie) – např. Hamm – Brorsen [23], Al-Alawi – Zaibet [1],
- plánování optimalizovaného využití – výstavba drenáží, využití prostoru, do-

pravní problémy – např. Fujii et al. [17],

- samostatné rozhodování – ovládání skleníků, ovládání traktorů a strojů – např. Noguchi et al. [63],
- rozpoznávání (klasifikace) – rozpoznávání poruch na základě zvukových či barevných projevů, třídění na základě ne přesně klasifikovaných kritérií – např. Sato et al. [77],
- srovnání – srovnání klasických technik (regresní analýza, shlukování) s neuronovými sítěmi – např. Mainland [45], Markiewicz [46],
- teoretické práce – např. Veselý [92].

7.1.2 Další databáze

S problematikou neuronových sítí se lze setkat prakticky v jakémkoli typu literatury, počínaje seminárními, bakalářskými a diplomovými pracemi a konče výzkumnými záměry či grantovými projekty. Prohledávány byly tedy i jiné databáze např. databáze *ABI/Inform (R) Global Edition*, *Dissertation Abstracts*, *Periodical Abstracts Research II Edition*, *Wilson Business Abstracts*, *Social Sciences Abstracts Full Text* a další databáze dostupné studentům ČVUT (např. IEEE). V podkapitole 7.1.2 jsou uvedeny články, které byly získány právě z těchto databází.

Článek Worzala et al. [97] obsahuje srovnání neuronových sítí a klasické regresní analýzy pro předpověď ceny realit. Autor uvádí, že nelze prokázat nadřazenost neuronových sítí jakožto analytického nástroje, neboť použití neuronových sítí není vždy úspěšné. V článku zabývajícím se určením finančního zdraví podniku Lacher et al. [42] se autoři opírají o výhody neuronových sítí v porovnání s diskriminační analýzou. Ta je totiž použitelná jen při splnění předem daných předpokladů (nezávislost, normalita, atd.), zatímco neuronové sítě se apriory o žádné takové předpoklady neopírají. Článek Swanson – White [85] se zabývá predikcemi. Teoretická práce Li et al. [44] představuje zdokonalení Hopfieldovy neuronové sítě. V článku Fieldsend – Singh [16] se mimo jiné autoři zabývají změněnou metodikou výpočtu chybové funkce¹, která není jako obvykle založená na Euklidově vzdálenosti mezi požadovaným a vypočteným výstupem sítě. Autoři tohoto článku zdokonalují výkonnost neuronové sítě a neuronovou síť s takovou topologií nazývají „Paretovou vývojovou neuronovou sítí“.

¹Chybová funkce jde definována v podkapitole 5.3.

7.2 Konkrétní studované využití neuronových sítí

Přestože již v předchozích podkapitolách je citována řada publikací dokazujících široké možnosti využití neuronových sítí, podkapitola 7.2 cituje ty, které předkladatele disertační práce ke studiu neuronových sítí „přivedly“ a poskytly ucelené informace o této problematice. V současné době je k dispozici mnoho zdrojů, přičemž většina technicky zaměřených článků se opírá o zatím nepřekonanou knihu Šíma – Neruda [86]. Reference českých publikací o neuronových sítích často uvádějí i další jméno spojené s umělou inteligencí, a to jméno doc. Ing. Miroslava Šnorka, CSc.² Osob spjatých s výzkumem na poli neuronových sítí je relativně mnoho. Některé z nich, stejně tak i skupiny kolem nich vytvořené, hrají velkou roli i v mezinárodním měřítku. Za všechny budiž uvedena výzkumná skupina *NCG*³.

Ne všechny dostupné informace jsou však spolehlivé. Navíc vzhledem k tomu, že mnoho prací vzniká souběžně, není úplně zřejmé, kdo byl kým ovlivněn, respektive zda autoři nečerpali ze stejného zdroje. Rok vydání není často tím nejspolehlivějším vodítkem.

7.2.1 Expertní systémy

Jedním z výhodných způsobů využití neuronových sítí jsou expertní systémy. Obecný přehled funkce expertních systémů lze získat v Berka et al. [2] nebo v kapitole 8.

Použití neuronové sítě jako expertního systému lze nalézt v diplomové práci Prskavec [72] vytvořené na ČVUT pro rozpoznávání obrazců částečných výbojů. Vlastní implementace sítě proběhla v programu MATLAB a Neurex 5.1⁴ Cílem diplomové práce bylo vytvořit neuronovou síť pro rozpoznávání základních obrazců částečných výbojů s cílem určit druh výbojové činnosti, resp. lokalizovat zdroje částečných výbojů, což se autorovi podařilo.

Stejným tématem se o několik let později zabýval ve své diplomové práci Jiří Chmelenský. V abstraktu diplomové práce Chmelenský [34] uvádí: „V programu Matlab[®]6 a jeho Neural Network Toolboxu byla vytvořena neuronová síť pro rozpoznávání obrazců částečných výbojů. Vhodnou volbou topologie, trénovací množiny

²Výčet nesčetných aktivit doc. Ing. Miroslava Šnorka, CSc. spjatých s problematikou neuronových sítí lze nalézt např. na Šnorek [83]. Za všechny jeho publikace na jsou na tomto místě uvedeny publikace Šnorek [88] a Mráz et al. [56].

³Computational Intelligence Group – [9]

⁴Představení programu Neurex, byť verze 4, je v příloze 6.2.

a trénovacího algoritmu se podařilo naučit neuronovou síť nejen rozpoznávat základní obrazce částečných výbojů, ale také eliminovat vliv okolních rušení (tyristorové, náhodné a šum na pozadí). Důvodem práce bylo pomocí návrhu neuronové sítě minimalizovat okolní rušení v průběhu měření částečných výbojů, které má negativní vliv na naměřené hodnoty použitím vhodného zapojení (např. můstkového).“

7.2.2 Predikce v časových řadách

Analýzou vysokofrekvenčních časových řad se zabývá Mastný [50]. Autor se ve své práci pokusil predikovat prostřednictvím dopředné sítě s jednou skrytou vrstvou s 21 neurony směnný kurz EUR/USD. Síť měla 9 vstupů (devizové kurzy EUR/USD, GBP/USD, JPY/USD za tři po sobě jdoucí předchozí období) a jeden výstup – predikci devizového kurzu EUR/USD. Autor použil tzv. intervenční časové řady. V závěru pak uvádí, že neuronové sítě mají dobré vlastnosti jak pro predikci kurzu, tak pro predikci kolísání kurzu.

Analýzou časových řad pro predikci potřeby vody s pomocí neuronových sítí se hodlá ve své disertační práci zabývat Ing. Martin Němec v tezí k doktorské práci Němec [57] vymezuje cíl následujícím způsobem: „V disertační práci se budu zabývat výzkumem predikce potřeby vody pomocí neuronových sítí, a to jak z krátkodobého, tak z dlouhodobého hlediska na časových řadách s hodinovou potřebou vody. Takto malý krok byl zvolen záměrně, neboť tyto řady vykazují poměrně velký rozptyl a nahodilé výchyly, a proto je nelze s úspěchem modelovat pomocí klasických matematických metod.“ Za stěžejní úkoly považuje tento doktorand výběr vhodné neuronové sítě, vhodný popis časové řady, volbu výpočetního prostředku, aplikaci natrénované neuronové sítě na podobné časové řady (tj. validaci) i porovnání získaných výsledků s výsledky získanými prostřednictvím klasických statistických metod. Více lze nalézt právě v Němec [57].

Článek Garcia – Kirschen [20], zaměřený na predikce na konkurenčním trhu s elektřinou, používá neuronové sítě jako součást moderních data-miningových metod sloužících k predikci. Výkonnost neuronových sítí je porovnána s výkonností klasických statistických metod, např. s exponenciálním vyrovnaním či ARIMA modely. Při trénování byly jako nejvhodnější architektury k měsíční predikci shledány vícevrstvé perceptrony a RBF neuronové sítě. Výkonnost použitých metod byla porovnávána prostřednictvím střední absolutní relativní chyby (MARE)⁵ a odmocniny ze střední

⁵Terminologie a značení bylo ponecháno tak, jak je použito v práci Garcia – Kirschen [20]. V případě střední absolutní relativní chyby se by se dala použít terminologie opsat následovně:

kvadratické chyby (RMSE⁶). Výsledky článku ukazují, že neuronové sítě pracují s lepšími výsledky než ostatní metody, viz tabulka 7.1. Autoři článku (Garcia, Kirschen) navíc připouštějí, že kvalita predikcí je stále nižší, než by bylo potřeba. Pokud je však neuronová síť „přetrénována“ na co nejaktuálnějších datech, přesnost predikce prostřednictvím neuronových sítí se zlepšuje.

Tabulka 7.1: Výkonnost srovnávaných metod (týdenní predikce)

(Zdroj: Garcia – Kirschen [20] – upraveno)

ARIMA		Exp. vyrovnání		Neuronová síť	
RMSE (MW)	MARE	RMSE (MW)	MARE	RMSE (MW)	MARE
558	0,7	657	0,8	380	0,3

Využitím neuronových sítí pro predikci v časových řadách se obšírněji zabývá také Ing. Marek Obitko, Ph.D. Ten například na webových stránkách Obitko [65], představuje možnosti neuronových sítí a genetických algoritmů.

MARE je průměrná absolutní odchylka měřená relativně.

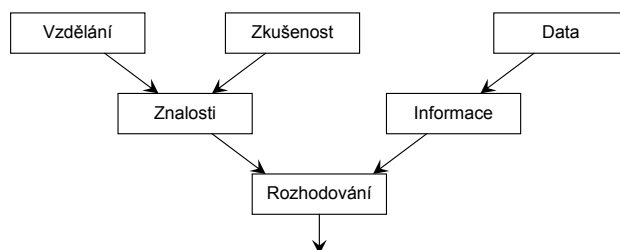
⁶RMSE = $\sqrt{(NIV_{actual} - NIV_{forecasted})^2}$; obecně RMSE (root-mean-square error) = odmocnina z průměru čtvercových odchylek u každého prvku trénovací množiny.

Kapitola 8

Expertní systémy

8.1 Motivace a historie

Každá instituce, firma, podnik se při svém rozhodování opírá o *znalosti*, které jsou důležité pro řešení rozličných problémů (jak investovat, komu půjčit, jak odstranit závadu, atp). Znalosti jsou dle Turban – Frenzel [91] definovány jako informace, jež byly zorganizovány a analyzovány tak, aby byly srozumitelné a použitelné pro řešení problémů, rozhodování či učení. Znalosti jsou těm, kteří o ně stojí (dále uživatelům), obvykle zprostředkovávány odborníky v daném oboru (dále experty). Faktory vstupující do obecného procesu rozhodování symbolicky zobrazuje schéma na obrázku 8.1, přičemž expert představuje znalost, se kterou je na základě dostupných informací rozhodováno. Vysoce odbornou práci expertů musí uživatel obvykle velice velkoryse ocenit, přičemž horentní sumy za tyto služby jsou často vynucovány pouhým nedostatkem expertů.



Obrázek 8.1: Struktura procesu rozhodování

(Zdroj: Berka et al. [2, str. 44] – upraveno)

Právě honoráře za služby kvalitních expertů byly jedním z důvodů, které vedly ke vzniku *expertních systémů*. Expertní systémy mohou stejně jako experti zprostředkovávat znalost nutnou ke kvalifikovanému rozhodování.

První expertní systém vznikl od roku 1965 a jeho vývoj pokračoval až do roku 1983. Tento systém byl založen na poznatcích z výzkumu systému General Problem Solver a jmenoval se *DENDRAL*. Jeho tvůrci byli EDWARD FEIGENBAUM, JOSHUA LEDERBERG, BRUCE BUCHANAN a CARL DJERASSI (publikováno např. v Buchanan – Feigenbaum [6]). Projekt vznikl na Stanfordské univerzitě a byl součástí projektů týkajících se využití umělé inteligence. DENDRAL byl určen pro identifikaci a popis sloučenin v analytické chemii a pracoval na základě spektrální analýzy. Jakkoli lze diskutovat o prvenství právě DENDRALu, je všeobecně uznáváno, že to byl skutečně první úspěšný systém, který byl po léta propracováván a zdokonalován.

8.2 Tvorba expertního systému

Geneze expertního systému je velice složitá a podílí se na ní vedle skupiny expertů, jejichž znalosti jsou podkladem pro pravidla k vyhodnocování informací, skupina programátorů realizující tvorbu počítačového programu a skupina znalostní inženýrů¹ starajících se o implementaci znalostí do počítačového programu.

Každá z uvedených skupin má při tvorbě expertního systému nezastupitelnou roli a je nedílnou součástí týmu. Zdar celého procesu tvorby systému je postaven na následujících premisách:

- Ochotě expertů vyzradit – prodat – své znalosti.
- Schopnosti expertů popsat své rozhodovací postupy.
- Kvalitní komunikaci mezi experty a částí týmu starající se o implementaci.
- Existencí vhodné reprezentace znalosti v expertním systému.

Paradoxem první premisy je skutečnost, že expert vyrazující své znalosti se jako takový zbavuje své výjimečnosti a právem se obává, že bude nahrazen expertním systémem, který pomáhá vytvářet. V souvislosti tím vzniká otázka, kdo je, či kdo by měl být odpovědný za rozhodování expertního systému. Expert poskytující znalosti, či programátor systému? V současné době expertní systémy nevystupují v rolích expertů, ale jsou chápány pouze jako podpůrné prostředky pro vlastní práci uživatele

¹ *Znalostní inženýr*: obvykle se jedná o pracovníka s hlubokými znalostmi provozních systémů, aplikačního softwaru a nástrojů pro vývoj softwaru, který nese odpovědnost za design a vývoj systémů, včetně komunikačních systémů a sítí.

v roli experta – manažera. Výsledné rozhodnutí tedy opět leží na bedrech člověka, který by měl být zodpovědný za svá rozhodnutí (správná i chybná).

Do druhé premisy se promítá rozpor mezi vysokou erudovaností experta a jeho neschopností své znalosti předat. Přitom předání znalostí musí být vysoce strukturované a srozumitelné. Z toho je zřejmé, jak kompaktní tým musí tvořit programátor a expert (třetí premisa). Obě profese spojuje znalostní inženýr. Ten řídí programátory a experty tak, aby byly naplněny představy zadavatele, respektive budoucího uživatele. Znalostní inženýr například transformuje požadavky experta tak, aby programátor mohl tyto požadavky zapracovat do systému, nebo naopak tlumočí požadavky programátora na formu podání vědomostí expertem².

Čtvrtou premisu splňují například některé topologie neuronových sítí – viz podkapitola 8.3.3.

Expertní systémy jsou dle typu řešených úloh jsou obvykle rozdělovány na:

- *diagnostické* – tyto provádějí diagnostiku a tvoří většinu všech současných expertních systémů; nacházejí uplatnění např. v lékařství (stanovení choroby), v oblasti PC (řešení on-line vzniklé závady), v ekonomii a marketingu (odhadování finančního zdraví podniku, bonity klienta atd.);
- *plánovací* – tvoří necelých 10 % všech expertních systémů; úkolem je plánování postupů při zadaných činnostech a plánování vůbec;
- *dedikované* – doplňují do 10 % plánovací expertní systémy; jedná se o kombinaci předchozích dvou typů.

V souvislosti s aplikační částí disertační práce bude pozornost věnována výhradně diagnostickým expertním systémům³.

²Znalostní inženýr je vlastně „tlumočníkem“ mezi programátorem a expertem. Nejenže každý z nich mluví jinou řečí, ale každý úplně jinak uvažuje a mnohdy své požadavky na fungování systému nedokáže svému protějšku vhodně formulovat.

³Například v Berka et al. [2] je uvedeno třídění do 11 typů. Expertní systémy pro: *design*, *diagnózu*, *instrukce*, *interpretace*, *řízení*, *monitorování*, *plánování*, *predikci*, *preskripci*, *simulace* a *selekcii*.

8.3 Klasifikace expertních systémů dle způsobu jejich tvorby

8.3.1 Klasický expertní systém

Většina expertních (diagnostických) systémů pracuje tak, že uživatel zodpovídá posloupnost otázek, po jejichž vyhodnocení by měl expertní systém stanovit (podobně jako expert) konečný verdikt (diagnózu). V závislosti na již zodpovězených otázkách je systémem volena další, konkrétní otázka, která zpřesňuje informaci nutnou pro rozhodnutí. Tento logický postup práce expertního systému založeného na kladení a zodpovídání otázek označovaný jako *If-then-else*⁴ zpravidla systému sice umožňuje zprostředkovat znalost, tvorba systému je však poměrně časově náročná a klade vysoké nároky na jeho tvůrce. Celý expertní systém naznačený výše má tvar rozhodovacího stromu, kde terminály (listy stromu) obsahují rozhodnutí.

Vytvoření takového rozhodovacího stromu je velice náročné. Je třeba si uvědomit, kolik kombinací při dané množině otázek a dané množině odpovědí může nastat. Expert buď musí při tvorbě systému dát odpověď na všechny možné kombinace otázek, nebo musí být v jádru expertního systému využit inteligentní řídicí mechanismus, který si umí odpovědi na kombinace otázek expertem nezodpovězené sám vydedukovat – viz podkapitola 8.3.2 a 8.3.3. Při tvorbě expertního systému popsaného v podkapitole 8.3.1 spolupracuje celý tým na tvorbě explicitních pravidel expertního systému.

8.3.2 Expertní systém s expertem „napůl“

V některých případech nedokáže expert přesně formulovat pravidla, podle kterých se rozhoduje. Přesněji, expert, ani celý tým nemusí být schopen algoritmizovat pravidla pro tvorbu expertního systému. Řešení této situace nabízí následující postup:

1. Expert pouze identifikuje podstatné vstupní informace (dále faktory), které zvažuje při svém rozhodování a
2. pro různé reálné kombinace těchto faktorů formuluje svá rozhodnutí.

Expert tedy pouze sděluje, jak by se v dané situaci rozhodl, nikoli proč by se tak rozhodl. Tuto znalost je třeba expertní systém „naučit“.

Způsob tvorby expertního systému popsaný v podkapitole 8.3.2 je vhodný i v případě, kdy nelze skutečnou situaci zmapovat (o tom pojednává podkapitola 8.3.3),

⁴{Premisa}-{důsledek při platnosti premisy}-{důsledek při neplatnosti premisy}.

tj. vytvořit dostatečně širokou množinu dvojic faktorů a výstupních parametrů jednoznačně určujících vhodné rozhodnutí.

Možnost, jak se vyhnout situaci, že pravidla formuluje přímo expert, a přitom vytvořit vhodný expertní systém, nabízí vedle algoritmů klasifikačních a regresních stromů právě *neuronová síť*. Neuronová síť tedy nahrazuje mechanismus formulace pravidel pro rozhodování na základě vybraných faktorů. Tím svým způsobem pouze kopíruje znalost experta, které se naučila.

Oproti přístupu v podkapitole 8.3.1 nejsou jednotlivé skupiny týmu tak těsně propojeny. Pracují samostatněji. Experti konzultují danou problematiku, programátoři zajišťují učení expertního systému a znalostní inženýři zprostředkovávají komunikaci mezi experty a programátory.

8.3.3 Expertní systém bez experta

Pokud není problém získat reprezentativní množinu dvojic faktorů a výstupních parametrů, tj. jsou k dispozici faktory a určitý výsledek získaný pozorováním, na základě nichž je jednoduché určit správné rozhodnutí (např. je známo, jaká investice přinesla největší zisk, kdo bez problémů splatil půjčku, kde vznikla závada, atp.), není experta zapotřebí. Stačí pouze trénovací množina faktorů a požadovaných odpovědí expertního systému a neuronová síť. Např. síť, jejíž adaptivní dynamika je založena na učení s učitelem. Vhodná je síť z řady klasifikátorů (asociativní paměti), tedy například Hopfieldova či Kohenova síť, nebo neuronová síť typu backpropagation⁵, RBF síť⁶ nebo pravděpodobnostní neuronová síť. Pomocí „neuronového“ učení⁷ se pak nastaví parametry sítě tak, aby pro dané faktory odpovídala tak, jak má.

Problémem hodným experta však i zde zůstává výběr faktorů pro rozhodování, tj. výběr podstatných informací pro vhodné kvalifikované rozhodnutí a způsob odpovědi expertního systému. Odpověď může být buď jednoznačná (ano/ne, zařazení do určité skupiny), nebo ji může tvořit skupina aposteriorních parametrů, na jejichž základě bude učiněno rozhodnutí. V případě úplného vyřazení experta nahrazuje neuronová síť jak mechanismus rozhodování na základě informací, tak proces vytváření znalostí na základě reálné situace.

Při tvorbě expertního systému bez experta má nejvýraznější roli v rámci týmu programátor, který volí/programuje metodu učení expertního systému. Expertní systém

⁵Backpropagation = dopředná síť – viz podkapitola 5.3.1.

⁶popis odlišnosti výkonných jednotek RBF sítě je uveden v podkapitole ??

⁷Viz kapitola 5.

je učen na základě souvislostí, které „skrývá“ množina dvojic faktorů a výstupních parametrů.

Poznámka: Na kapitulu 8 navazuje kapitola 11 popisující tvorbu expertního systému pro určování bonity žadatele o spotřebitelský úvěr.

Část II

Materiál a metody

Kapitola 9

Materiál a metody

9.1 Materiál

S ohledem na architekturu disertační práce je materiál (datové soubory) součástí aplikační části a je specifikován v podkapitolách se shodným názvem „Soubor dat“. Na str. 89 je specifikován soubor pro diskriminaci obcí dle salda migrace (podkapitola 10.2), na str. 125 je specifikován soubor využitý pro tvorbu expertního systému určujícího bonitu žadatele o spotřebitelský úvěr na koupi mobilního telefonu (podkapitola 11.2) na str. 145 je specifikován soubor dat pro analýzu a predikci míry nezaměstnanosti v České republice (podkapitola 12.2).

Uvedené soubory dat jsou zařazeny do příloh A.1, A.2 a A.3. Tato podkapitola uvádí jen jejich přehled a několik poznámek ke vzniku souborů dat.

9.1.1 Migrace v Jihočeském regionu

Zdrojem primárních dat – ekonomických ukazatelů – ovlivňujících migraci byly databáze Ministerstva financí (Ministerstvo financí [53]), Ministerstva práce a sociálních věcí (OKsystem [66]) a Českého statistického úřadu (Český statistický úřad [11]), pobočka v Českých Budějovicích. Některé ukazatele byly přepočteny na stejný základ (jednoho ekonomicky aktivního obyvatele¹). Více je o ekonomických ukazatelích a celém souboru dat uvedeno v podkapitole 10.2.

¹Mezi ekonomicky aktivní jsou zahrnuty všechny osoby, které patří mezi zaměstnané osoby, zaměstnavatele, samostatně výdělečně činné, pracující důchodce, pracující studenty a učně, ženy na mateřské dovolené, osoby v základní nebo civilní vojenské službě, ve vazbě a výkonu trestu nebo osoby nezaměstnané.

9.1.2 Bonita klientů spotřebitelských úvěrů

Data pro vytvoření expertního systému určujícího bonitu žadatele o spotřebitelský úvěr na koupi mobilního telefonu byly porovnány z jednotlivých žádostí vyplněných žadatelem. Po uplynutí doby potřebné k zjištění platební morálky klientů byly údaje doplněny o informace o bezproblémovém splacení úvěru eventuálně o výši dlužné částky. Více o faktorech použitých pro určování bonity klienta je uvedeno v podkapitole 11.2.

9.1.3 Vývoj nezaměstnanosti v České republice

Registrovaná míra nezaměstnanosti tak, jak ji vypočítává Ministerstvo práce a sociálních věcí, byla získána právě z portálu tohoto ministerstva v sekci *Zaměstnanost – Statistiky nezaměstnanosti* (OKsystem [67]). Podrobněji je používána časová řada popsána v podkapitole 12.2.

9.2 Metody

Aplikační část práce byla zpracována tak, aby bylo možno realisticky vyhodnotit možnosti neuronových sítí. Diskriminace a modelování v časových řadách bylo provedeno prostřednictvím neuronových sítí a poté porovnáno s výsledky klasických nástrojů statistiky. K řešení problémů byly vybrány vícevrstvé dopředné sítě, pravděpodobnostní neuronové sítě a RBF sítě, tedy sítě, jejichž trénování je založeno na tzv. učení s učitelem². Motivace pro řešení právě těchto problémů jsou uvedeny v podkapitolách se shodným názvem „Motivace“ – podkapitoly 10.1, 11.1 a 12.1.

Některé z využívaných metod byly popsány v části I, jiné jsou prostřednictvím odkazů na literaturu zmíněny přímo v aplikační části. Souhrnný výčet všech metod by byl neúčelný a nepřehledný, neboť oblast neuronových sítí a souvisejících oblastí je velmi široká.

9.2.1 Metody pro měření výkonnosti

Výkonnost diskriminačních nástrojů byla hodnocena standardně prostřednictvím procenta správně klasifikovaných objektů a byla provedena taktéž validace těchto výkonností. Vhodnost modelů časových řad byla hodnocena jednak mírami neshody

²Jak již bylo vysvětleno v úvodu, popis architektury neuronových sítí, stejně tak jako způsoby jejich trénování, jsou součástí především I části.

(zejména $M.A.E.$, $M.S.E.$, $M.A.P.E.$ a $S.D.$ poměr) a Theilovým U koeficientem nesouladu. Přesné definice ukazatelů výkonnosti/vhodnosti neuronových sítí jsou s ohledem na architekturu disertační práce uvedeny v jednotlivých kapitolách – zejména v podkapitolách 10.4.1 a 12.4.

9.2.2 Implementace metod – softwarová prostředí

Výpočty prezentované v předkládané disertační práci byly provedeny prostřednictvím statistického softwaru STATISTICA Neuronové sítě 7.1, STATISTICA komplet 6.1 Cz a pro sumarizaci či dílčí výpočty byl použit program MS Excel 2003. STATISTICA je renomovaný modulární statistický software, který umožňuje využívat specializované moduly jak jednotlivě, tak v kombinaci. Popis možností modulu STATISTICA Neuronové sítě 7.1 lze najít v podkapitole 6.1, nebo na StatSoft [81], či StatSoft CR [80]. Konkrétní využití softwaru pro řešení vybraných problémů přináší zejména část III.

Samotná práce je vysázena prostřednictvím profesionálního typografického systému L^AT_EX.

Část III

Aplikační část

Kapitola 10

Diskriminace a neuronové sítě

Tato kapitola se zabývá srovnáním výsledků diskriminace získaných prostřednictvím metod klasické diskriminační analýzy (publikováno v Čermáková – Biskup [117]) a diskriminací provedenou prostřednictvím neuronových sítí. Přitom důraz je kladen na problematiku neuronových sítí.

10.1 Motivace

Jihočeský region byl zaměřen na zemědělskou výrobu realizovanou převážně v příhraniční horské a podhorské výrobní oblasti. V souvislosti s útlumem zemědělské výroby bylo možno předpokládat zvýšenou migraci obyvatel do vnitrozemí. Aby se minimalizovalo nebezpečí vyliďňování či rychlého „stárnutí“ obcí, bylo třeba hledat – identifikovat – faktory, které migraci ovlivňují a prostřednictvím nichž by bylo možno trend pomalého vyliďňování obcí zvrátit. Identifikací faktorů rozumíme:

1. zjištění, které z uvažovaných ekonomických ukazatelů významně ovlivňují znaménko salda migrace,
2. zjištění, zda zvolené ekonomické ukazatele stačí ke kvalitní predikci znaménka salda migrace.

Některé výsledky, které byly prezentovány v rámci této kapitoly, byly použity při řešení MSM 6007665806.

10.2 Soubor dat

Explorační analýze byly podrobeny ukazatele zjištěné v 621 obcích Jihočeského regionu. Data byla získána z databáze Ministerstva financí, Ministerstva práce a

sociálních věcí a Českého statistického úřadu, pobočky v Českých Budějovicích.

Aby byly obce z hlediska vybraných ekonomických ukazatelů srovnatelné, byly některé z ukazatelů přepočteny na stejný základ – na jednoho ekonomicky aktivního obyvatele (dale jen EAO). Spojité vstupní proměnné jsou výhodnější jak pro přístup strojového učení reprezentovaný neuronovými sítěmi, tak při použití metod diskriminační analýzy.

Bylo zvoleno 12 ekonomických ukazatelů, z nichž bylo na základě výsledků korelační analýzy vybráno 7 ekonomických ukazatelů s nízkým stupněm těsnosti vzájemné korelační vazby. Jednalo se o následující ukazatele:

1. **nadmořská výška obce** (výš) – nadmořská výška v m n. m. (pro zvýšení homogenity dat byla při klasické diskriminační analýze použita logaritmická transformace¹),
2. **hustota obyvatel** (hustob) – počet obyvatel na 1 ha katastrálního území,
3. **míra nezaměstnanosti** (míranez),
4. **stavební aktivita v obci** (bytnaek) – počet dokončených bytů na 1 EAO,
5. **zaměstnanecká „absorpce“ obce** (zamnaek) – počet obyvatel zaměstnaných v obci připadajících na 1 EAO,
6. **podnikatelská aktivita v obci** (fyznaek) – počet osob samostatně výdělečně činných připadajících na 1 EAO,
7. **význam soukromého zemědělství** (rolnaek) – počet soukromě hospodařících rolníků připadajících na 1 EAO.

Věcná významnost ekonomických ukazatelů byla konzultována s odborníky a je uvedena v příspěvku Čermáková – Biskup [116].

Diskriminováno bylo do dvou skupin podle znaménka salda migrace (obce se záporným – *Z-migrace*, nezáporným – *N-migrace* saldem migrace).

Použitý soubor dat složený z vybraných, výše uvedených, přepočtených ekonomických ukazatelů a přiřazení do skupiny dle salda migrace je uveden v tabulce A.1 v příloze A.1. Tabulka A.1 navíc obsahuje ještě informace i rozdělení dat na trénovací, testovací a validační.

¹V tabulkách značeno **logvýš**.

10.3 Klasická statistická analýza – diskriminační analýza

V podkapitole 10.3 jsou uvedeny pouze závěry publikace Čermáková – Biskup [117]. Pro účely předložené disertační práce není podrobná citace nutná. Z metod výběru diskriminátorů byla použita kroková metoda vpřed. Výběr byl po pěti krocích ukončen – výsledky viz tabulka 10.1.

10.3.1 Výsledky klasifikace prostřednictvím diskriminační analýzy

Ve výsledném diskriminačním modelu existují pouze 3 statisticky významné, a tím i důležité, diskriminátory: poloha obce z hlediska nadmořské výšky, stavební aktivita v obci a zaměstnanecká absorpce obce. Statistická významnost diskriminace byla testována pomocí Pressovy q -statistiky a zjistilo se, že klasifikační funkce se statisticky vysoce významnou spolehlivostí diskriminují objekty do dvou tříd ($q = 55,11$, $\chi^2_{0,01}(1) = 6,63$). Také vzdálenost mezi centroidy obou tříd byla signifikantní.

Tabulka 10.1: Výsledky diskriminační funkční analýzy

(Zdroj: Čermáková – Biskup [117])

N=621	Výsledky diskriminační funkční analýzy					
	krok 5, poč. prom. v modelu: 5; grupovací: diskř (2 skup)					
	Wilks. lambda: 0,96968 přibliž $F(5, 615) = 3,847$ $p < 0,002$					
	Wilks. Lambda	Parc. Lambda	F na vyj (-1,615)	Úroveň p	Toler.	1-toler. R^2
logvýš	0,980	0,989	6,728	0,01	0,947	0,052
bytnaek	0,976	0,993	4,032	0,045	0,987	0,013
zamnaek	0,976	0,994	3,92	0,048	0,968	0,031
míranez	0,973	0,996	2,411	0,121	0,941	0,06
hustob	0,973	0,997	1,880	0,171	0,95	0,05

Na druhou stranu hodnoty jak Wilksovy lambdy, tak parciální lambdy jsou blízké jedné. Diskriminační síla ukazatelů tedy není příliš vysoká, ačkoli tyto ukazatele signifikantně diskriminují do dvou tříd. Signifikantní diskriminaci lze proto spíše přičíst rozsahu souboru než účinnosti diskriminátorů. Interpretovat statistickou významnost jako důkaz kvality modelu by nebylo správné.

Z výsledků provedených analýz lze vyvodit závěr, že pokud se občané Jihočeského regionu stěhují, nejsou důvodem ani nepříznivé přírodní podmínky, ani nezaměstnanost, či malá šance získat byt či zaměstnání v místě. Tento závěr podporuje i analýza klasifikační matice (viz tabulka 10.2).

Tabulka 10.2: Klasifikační matice – diskriminační analýza

(Zdroj: Čermáková – Biskup [117] – upraveno)

diskr	Klasifikační matice		
	Řádky: pozorované klasifikace		
	Sloupce: předpovězené klasifikace		
	% správných	Z-migrace p=0,359	N-migrace p=0,641
Z-migrace	10,314	23	200
N-migrace	95,477	18	380
Celkem	64,895	41	580

Z údajů uvedených v tabulce 10.2 je zřejmé, že se prostřednictvím klasifikačních funkcí sice podařilo správně klasifikovat celkem téměř 64,89 % objektů, avšak velmi nízkou úspěšnost klasifikace vykazovala právě záporná migrace.

Jako stěžejní literatura pro zpracování statistické analýzy prostřednictvím diskriminační analýzy byly použity knihy: Hebák – Hustopecký [26] a McLachlan [52].

10.4 Diskriminace prostřednictvím neuronových sítí

V podkapitole 10.4 jsou shrnuty a komentovány výsledky diskriminace prostřednictvím neuronových sítí různých topologií, tj. neuronových sítí architektur vhodných pro diskriminaci, trénovaných prostřednictvím dostupných metod. Nejedná se jen o prezentování výsledků, ale také o genezi diskriminace, která proběhla prostřednictvím programu STATISTICA 7.1 Neuronové sítě. Rozdělením souboru dat tak, jak je běžné pro trénování neuronových sítí se zabývá podkapitola 10.4.1. Následující podkapitoly (10.4.2–10.4.4) popisují použitý přístup, který byl autorem předložené disertační práce zvolen pro řešení problému a obsahují výsledky, které umožňují porovnání diskriminační výkonnosti testovaných sítí a trénovacích algoritmů. Podkapi-

toly 10.4.5 a 10.4.6 pak využívají pro řešení problému heuristiku programu STATISTICA 7.1 Neuronové sítě.

Odborná literatura uvádí mnohá doporučení, která by měla uživateli pomoci při volbě optimální topologie neuronové sítě a případně při výběru trénovacího algoritmu. Teoretické práce prokazují značnou diskriminační sílu neuronových sítí. Ve většině případů se tyto studie opírají o generovaná data vhodná pro daný problém a danou topologii neuronové sítě. Hledání optimálního řešení prostřednictvím neuronových sítí na reálných datech je stále otevřený problém.

10.4.1 Soubor dat pro neuronové sítě

Vzhledem k tomu, že strojové učení neklade žádné požadavky na homogenitu dat, nebylo nutno soubor dat homogenizovat.

Požadovaný výstup sítě byl znám – diskriminace ke skupině obcí se záporným respektive nezáporným saldem migrace. Trénování neuronové sítě proto probíhalo jako tzv. učení s učitelem (vice podkapitola 5.1.3). Aby mohla být výkonnost diskriminace prostřednictvím neuronových sítí (dále jen výkonnost) srovnatelná s výsledky klasické diskriminační analýzy, byla výkonnost měřena prostřednictvím procenta správně klasifikovaných objektů – viz například tabulka 10.2 na str. 92 a tabulka 10.19, respektive 10.18 na str. 120, respektive 119.

Soubor dat sestávající se z 621 objektů byl pro učení neuronových sítí využit dvojím způsobem²:

a) K porovnání výkonnosti jednotlivých typů neuronových sítí a jednotlivých topologií sítí. Pro tento účel byl soubor dat rozdělen do tří podmnožin přibližně stejného rozsahu – trénovací, testovací a validační. Důvody pro toto rozdělení jsou uvedeny v podkapitole 5.5, přičemž význam jednotlivých množin byl takový, jak ho uvádí Haykin [25], tj:

- **trénovací množina** sloužila pro trénování neuronové sítě,
- **validační množina** sloužila pro odhad validační chyby (případně včasné zastavení trénování, atp.),
- **trénovací množina** sloužila pro porovnání modelů a odhad chyby, se kterou neuronová síť pracovala.

Program STATISTICA 7.1 Neuronové sítě navíc umožňuje zařadit část případů do množiny, jež bude při výpočtech ignorována.

²Výkonnosti, příslušné jednomu ze způsobu trénování, jsou uvedeny v tabulkách 10.4, 10.5, 10.6, 10.8, 10.10, 10.12, 10.13 a 10.14.

Ve výsledných tabulkách prezentujících výkonnost jednotlivých sítí je pak uvedena výkonnost natrénované sítě vzhledem k trénovací, testovací a validační množině (značeno „*Trén.*“, „*Test.*“ a „*Valid.*“) a celková výkonnost (značeno „*Celk_1*“) natrénované sítě na trénovací množině (211 obcí) vzhledem k celému souboru dat. Přestože program STATISTICA 7.1 Neuronové sítě dokáže náhodně rozdělit soubor dat do zmíněných tří podmnožin, bylo pro zajištění srovnatelnosti výsledků využito rozdělení pevné (viz tabulka A.1). Do trénovací množiny bylo zahrnuto 211 obcí (33,98 %), do testovací množiny 201 obcí (32,37 %) a do validační množiny 209 obcí (33,36 %). Blíže kontingenční tabulka 10.3.

- b) Pro porovnání výsledků diskriminace prostřednictvím diskriminační analýzy a prostřednictvím neuronových sítí. Pro tento účel byl celý soubor – všechny obce – trénovací množinou a výkon sítě měřen pouze prostřednictvím trénovacího výkonu. Trénovací výkon je při absenci testovací a validační množiny zároveň celkovým výkonem sítě. Ten je v tabulkách označen „*Celk_2*“ a uvádí celkovou výkonnost sítě natrénované na všech 621 obcích vzhledem k celému souboru dat.

Tabulka 10.3: Rozdělení souboru dat pro trénování neuronové sítě

Saldo	Množina (počty obcí)			Celkem
	trénovací	testovací	validační	
Z-migrace	91	70	62	223
N-migrace	120	139	139	398
Celkem	211	209	201	621

Každá z natrénovaných sítí prezentovaných níže je pro snazší orientaci označena kódem a pořadovým číslem (např. „VVP 3“, „RBF 1“, „PNS 5“, ...). Kód v sobě primárně nese informaci o architektuře sítě a pořadové číslo pak odkazuje na konkrétní topologii respektive odlišný způsob trénování sítě. Označení je obvykle prvotně použito v některé z tabulek s výsledky.

10.4.2 Vícevrstvá dopředná síť

První z architektur neuronových sítí vhodných pro diskriminaci je vícevrstvá dopředná síť, která byla popsána v podkapitole 5.3.1. Program STATISTICA 7.1 Neuronové sítě disponuje dopřednou sítí s maximálně třemi skrytými vrstvami. Tento počet skrytých vrstev je pro diskriminaci dostačující – viz tabulka 4.1, podkapitola 4.2.1.

Vytvořená vícevrstvá dopředná síť musela být s ohledem na soubor dat složena ze vstupní vrstvy obsahující 7 vstupních neuronů a výstupní vrstvy s jediným neuronem. Následně bylo třeba zvolit počet skrytých vrstev a počet neuronů v každé z nich. Na str. 99 je představeno a prostřednictvím jejich výkonnosti (v tabulce 10.5) porovnáno několik dopředných sítí s různým počtem skrytých vrstev a různým počtem neuronů. Volbou počtu skrytých vrstev a počtu neuronů sice skončila volba architektury neuronové sítě, bylo však třeba ještě specifikovat zbytek topologie sítě. Výsledky pro různá nastavení trénování vícevrstevných dopředných sítí (viz výčet níže) jsou uvedeny na str. 96–105 této práce. Konkrétně byl volen³:

- způsob posuzování chyby sítě – součet čtverců chyby nebo entropii (koncept viz podkapitola 5.3.2),
- počet fází – jedna nebo dvě – a počet epoch v každé fázi⁴,
- způsob počátečního nastavení vah – náhodné rovnoměrné v předem zvoleném intervalu nebo náhodné gaussovské, definované střední hodnotou a směrodatnou odchylkou,
- podmínka ukončení trénovacího procesu (ukončení je založeno na cílové trénovací/testovací chybě, respektive zlepšení trénovací/testovací chyby),
- způsob nastavení klasifikačních prahů (automatické nebo manuální),
- metoda učení – nabízeny jsou: metoda backpropagation, metoda sdružených gradientů, Quasi-Newtonova metoda, Levenbergerova-Marquardtova metoda, metoda rychlého šíření chyby a metoda Delta-bar-delta – včetně zadání parametrů zvolené metody (například u metody backpropagation se jednalo o parametr učení a setrvačnosti, respektive jejich změny v průběhu učení, případně přidání gaussovského šumu).

Dále bylo nutno rozhodnout, zda má být zvolená neuronová síť zjednodušována, tj. mají-li být v průběhu učení vyřazovány neurony (včetně vstupních⁵) s nízkými vahami. Jak se nastavení parametrů projevilo na výkonnosti vícevrstevných dopředných sítí ukazují tabulky 10.4–10.8 uvedené v následujících odstavcích.

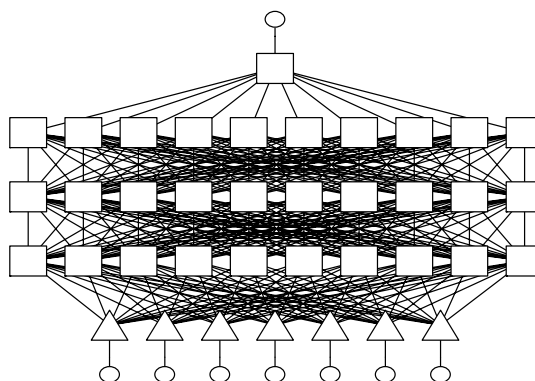
³Dostupné a relevantní volby v programu STATISTICA 7.1 Neuronové sítě. Pro jiné typy úloh než je diskriminace (např. aproximace, tj. tam, kde je výstup sítě spojitý) lze měnit aktivační funkci výstupních neuronů – lineární nebo sigmoidní s pevným parametrem, atp.

⁴Tato volba nabízí možnost trénovat neuronovou síť dvěma různými metodami po sobě a tím vylepšit výsledky trénování.

⁵Vstupní neurony mohou být také vyřazeny až po proběhnutí učení na základě nízké citlivosti vstupních proměnných.

10.4.2.1 Trénovací metody a výkonnost sítě

Stěžejním faktorem adaptivní dynamiky práce neuronové sítě je jistě trénovací metoda, jež modifikuje váhové nastavení sítě pro dosažení minimální trénovací chyby. Trénování byla experimentálně podrobena dopředná neuronová síť se 7 vstupními neurony, třemi 3 skrytými vrstvami, v každé po 10 neuronech a jedním výstupním neuronem (viz obrázek 10.1)⁶.



Obrázek 10.1: Schéma třívrstvé dopředné neuronové sítě 7:7-10-10-10-1:1

(Vytvořeno v programu *STATISTICA Neuronové sítě 7.1*)

V tabulce 10.4 jsou uvedeny výkonnosti (trénovací, validační, testovací a celkové – viz definice výkonů sítě na str. 93) různých trénovacích metod vícevrstvé dopředné sítě zobrazené na obrázku 10.1. Trénování probíhalo shodně pro všechny metody v 500 epochách⁷. Počáteční inicializace vah proběhla náhodně. Každá z vah byla programem *STATISTICA 7.1 Neuronové sítě* náhodně vybrána z intervalu $(0; 1)$, tak, jak zmiňuje podkapitola 5.3.2. Chyba sítě, používaná při trénování, byla vypočítávána prostřednictvím součtu čtverců celkové chyby – viz vzorec (5.5), str. 54. Výkonnost sítí⁸ byla proto převážně ovlivněna použitou trénovací metodou⁹.

⁶Výraz 7:7-10-10-10-1:1 znamená {7 vstupů}:-{7 vstupních neuronů}-{10 neuronů v 1. skryté vrstvě}-{10 neuronů v 2. skryté vrstvě}-{10 neuronů v 3. skryté vrstvě}-{1 výstupní neuron}:{1 výstup}.

⁷Jedna trénovací epocha v programu *STATISTICA 7.1 Neuronové sítě* odpovídá adaptaci sítě pro každý trénovací vzor (právě jednou).

⁸Ačkoli je architektura všech sítí stejná – 7:7-10-10-10-1:1, z pohledu topologie se jedná o různé neuronové sítě. Proto je používán termín „výkonnost sítí“ a ne „výkonnost sítě“.

⁹Výsledky trénování ovlivňovalo také počáteční nastavení vah. O tom bude ještě zmínka na str. 103. Srovnání výkonností to však významně neovlivnilo.

Tabulka 10.4: Výkonnost trénovacích metod použitelných pro trénování dopředné sítě 7:7-10-10-10-1:1

Metoda	Výkon				
	Trén.	Valid.	Test.	Celk_1	Celk_2
VVP 1 Backpropagation	0,754	0,478	0,542	0,593	0,614
VVP 2 Sdružených gradientů	0,654	0,493	0,577	0,575	0,533
VVP 3 Quasi-Newtonova	0,806	0,512	0,567	0,630	0,583
VVP 4 Levenbergerova-Marquardtova	0,9	0,498	0,512	0,639	0,597
VVP 5 Rychlého šíření chyby	0,929	0,483	0,537	0,652	0,717
VVP 6 Delta-bar-delta	0,682	0,526	0,522	0,578	0,641

Z výsledků (tabulka 10.4) je zřejmé, že trénovací metody byly různě „úspěšné“. Celková výkonnost sítě trénované na trénovací množině (Celk_1) kolísala v rozmezí 57,5 %–65,2 %. Přestože variabilita trénovacího výkonu byla poměrně značná (27,5 %), není pro porovnávání výkonnosti neuronových sítí trénovaných jednotlivými metodami směrodatná. Seriózní možnost porovnání výkonnosti nabízí výkon validační, testovací a celkový (Celk_1). Pro ně je variabilita výkonností neuronových sítí trénovaných jednotlivými metodami poměrně nízká. Například rozdíl mezi validační výkonností nejúspěšnější a nejméně úspěšné metody je menší než 5 % – viz tabulka 10.4. Validační i testovací výkon má stejnou vypovídací váhu, protože ani jedna z chyb – testovací a validační – nevstupovala do procesu učení. Celkové výkonnosti neuronových sítí natrénované různými metodami (viz tabulka 10.4 sloupce Celk_1 a Celk_2) nebyly příliš vysoké (do 72 %). Proč je někdy hodnota Celk_1 vyšší než hodnota Celk_2 je diskutováno na str. 103.

Výkonnost dopředné neuronové sítě není ovlivňována jen zvolenou trénovací metodou. Přestože byly testovány různé kombinace nastavení zmíněné na str. 95, byla pro další srovnávání výkonnosti (prezentované v textu disertační práce) vybrána nejvhodnější trénovací metoda vzhledem k použitému souboru dat – metoda rychlého šíření chyby – viz tabulka 10.4, sloupce Celk_1 a Celk_2. Motivem pro testování dalších topologií dopředných neuronových sítí bylo pochopitelně nalezení sítě s vyšší výkonností.

10.4.2.2 Architektura a výkonnost sítě

Některé trénovací metody dosáhly trénovací výkonnosti přes 80 % a až trénování na bohatší trénovací množině – všech 621 obcích – způsobilo, že se nepodařilo zna-

lost o saldu migrace v závislosti na zvolených ukazatelích a mohutnosti trénovací množiny do struktury sítě rozložit stejně dobře jako pro menší množinu trénovacích vzorů (když trénování probíhalo na 211 obcích – viz tabulka 10.4). Na základě těchto výsledků se jevílo jako rozumné architekturu sítě obohatit¹⁰.

V okamžiku, kdy byl počet neuronů ve skrytých vrstvách zvýšen na 30, – tj. byla uvažována třívrstvá dopředná neuronová síť 7:7-30-30-30-1:1 – končila část adaptací této sítě (bez ohledu na trénovací algoritmus) divergencí vah sítě¹¹.

Pro srovnání výkonností neuronových sítí s různě bohatou architekturou byla vybrána nejúspěšnější trénovací metoda pro dopřednou síť 7:7-10-10-10-1:1 a soubor dat – tj. metoda rychlého šíření chyby. Výkonnost trénovací metody rychlého šíření chyby, jejíž nastavení vah i přes bohatost architektury konvergovalo, shrnuje tabulka 10.5. Ta také uvádí srovnání výkonnosti s dalšími dopřednými sítěmi s naopak chudší architekturou. Trénování vždy probíhalo v 500 epochách.

Zatímco někteří autoři doporučují větší počet neuronů v první skryté vrstvě, než je počet neuronů ve vrstvě vstupní, někteří naopak tento přístup nedoporučují. Byli-li počet neuronů ve skrytých vrstvách volen tak, aby jejich počet klesal od sedmi k jednomu¹², klesala i celková výkonnost sítě (viz tabulka 10.5). Výkonnost trénování vícevrstvé dopředné sítě 7:7-6-4-2-1:1 a vícevrstvé dopředné sítě 7:7-4-1:1 (s pouze jednou skrytou vrstvou!) v 500 epochách metodou rychlého šíření chyby je uvedena také v tabulce 10.5. I zde je variabilita validačního výkonu nízká. Rozdíl mezi výkonností nejlepší a nejhorší architektury je nižší než 5 %.

Bohatší síť (více skrytých vrstev, více neuronů ve skrytých vrstvách) by měla zajistit dostatečnou „kapacitu“ sítě pro možné rozproštění závislosti výstupních faktorů na vstupních faktorech do prostoru vah. To bylo v případě postižení závislosti salda migrace na ekonomických ukazatelích prokázáno. Celková výkonnost bohatších sítí

¹⁰Důvodem pro další nezlepšování výkonnosti sítě v průběhu trénování (vedle nebohaté architektury vícevrstvé neuronové sítě) se mohlo stát uvíznutí trénovací metody v pasti lokálního minima chybové funkce tak, jak o tom pojednáno v posledním odstavci podkapitoly 5.3.2. Ani zařazení momentu setrvačnosti, ani proměnná volba parametru učení, výkonnost sítě nezlepšily. Lze se tedy domnívat, že metody opravdu dosáhly globálního nikoliv lokálního minima chybové funkce, a bohatší architektura sítě by mohla výkonnost sítě zlepšit. Experimentálně bylo navíc ověřeno, že i drobné, avšak necitlivé, operování s parametrem učení a momentem setrvačnosti vede k divergenci vah neuronové sítě.

¹¹Trénování neuronové sítě bylo ovlivňováno volbou topologie a mnohdy pro stejné počáteční nastavení končilo zcela jinak – buďto diametrálně (konvergence/divergence), nebo jen nastavením vah a tím i výsledkem diskriminace (výkonností, diskriminační funkcí sítě).

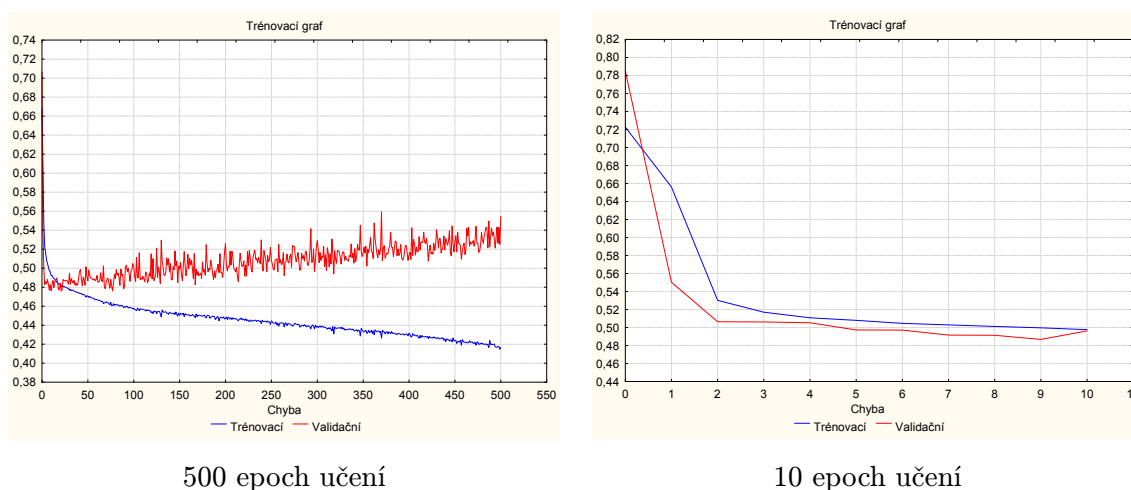
¹²Důvody pro volbu jednoduché architektury uvádí podkapitola 6.2.1 a to v souvislosti s využíváním softwaru Neurex.

Tabulka 10.5: Výkonnost různých architektur dopředné neuronové sítě trénované metodou rychlého šíření chyby

Architektura	Výkon				
	Trén.	Valid.	Test.	Celk.1	Celk.2
VVP 5 7:7-10-10-10-1:1	0,929	0,483	0,537	0,652	0,717
VVP 7 7:7-30-30-30-1:1	0,995	0,507	0,542	0,684	0,792
VVP 8 7:7-6-4-2-1:1	0,621	0,459	0,517	0,516	0,631
VVP 9 7:7-4-1:1	0,663	0,474	0,512	0,551	0,631

(VVP 5 a VVP 7) v porovnání s výkonností chudších sítí (VVP 8 a VVP 9) byla více jak o 10 % vyšší. Bohatší architektura sítě však nezajistila zásadní vylepšení celkového výkonu sítě zvolené na počátku, tj. vícevrstvé dopředné sítě s architekturou 7:7-10-10-10-1:1. Vondrák [93] navíc namítá, že nadměrná „košatost“ sítě vede výhradně ke ztrátě schopnosti generalizovat a síť se stává náchylnou k přetrénování, ale o tom již byla zmínka v podkapitole 5.5. Navíc tato námitka se s ohledem na bohatost architektury a použitý soubor dat ukázala neopodstatněnou – viz validační a testovací výkonnost v tabulce 10.5, sloupce Valid. a Test.

10.4.2.3 Generalizace jako podstatný faktor výkonnosti sítě



Obrázek 10.2: Průběh chyby učení pro trénovací a validační množinu – vícevrstvá dopředná síť 7:7-10-10-10-1:1, metoda rychlého šíření chyby

(Vytvořeno v programu STATISTICA Neuronové sítě 7.1)

Mezi další způsoby, jak zlepšit vlastnosti natrénované neuronové sítě, patří včasné ukončení trénovacího procesu. Ztráta generalizační schopnosti sítě je patrná na obrázku 10.2. Na levé straně je zobrazen průběh trénování, při němž docházelo ke spojitému poklesu trénovací chyby, zatímco chyba na validační množině měla tendenci růst. Na pravé straně jsou zobrazeny první epochy učení, při nichž klesá shodně jak trénovací, tak validační chyba. S ohledem na tyto výsledky je zřejmé, že trénování v 500 epochách bylo zbytečným luxusem, neboť validační chyba začala růst cca po 10 epochách. Po 10 epochách tedy začíná být neuronová síť „přeučená“ a fixovaná na obce, zařazené v trénovací množině. Přestože se testovací a validační chyba zvyšovala, pokles chyby trénovací byl natolik podstatný, že i přes snížení výkonnosti na testovací a validační množině, celková výkonnost s počtem trénovacích epoch rostla. Porovnání výkonnosti dopředné neuronové sítě 7:7-10-10-10-1:1 trénované metodou rychlého šíření chyby v závislosti na počtu epoch, je uveden v tabulce 10.6.

Tabulka 10.6: Výkonnost neuronové sítě 7:7-10-10-10-1:1 pro metodu rychlého šíření chyby v závislosti na počtu epoch

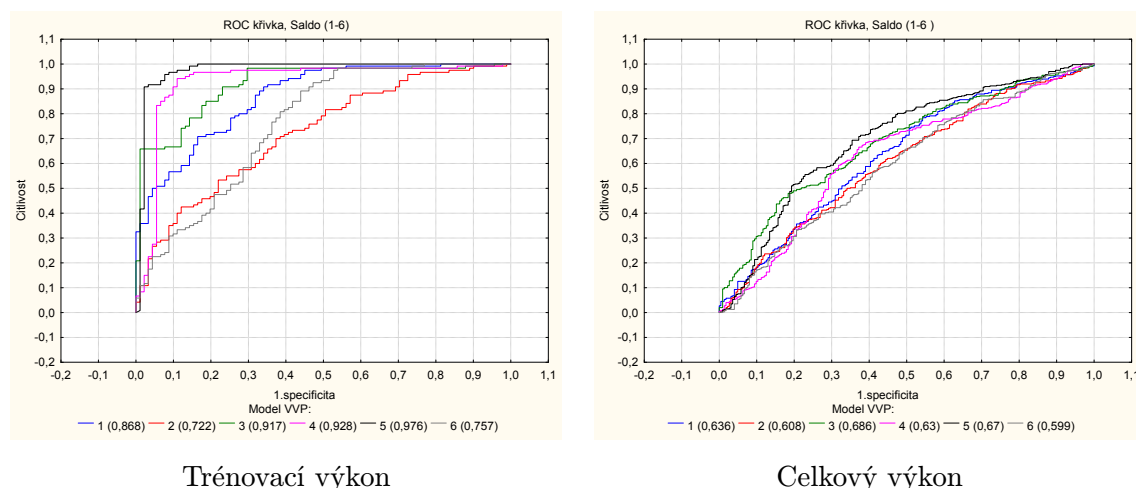
Počet epoch	Výkon				
	Trén.	Valid.	Test.	Celk.1	Celk.2
VVP 10 1 epocha	0,412	0,502	0,417	0,444	0,471
VVP 11 10 epoch	0,54	0,512	0,512	0,522	0,564
VVP 12 100 epoch	0,782	0,498	0,478	0,588	0,631
VVP 5 500 epoch	0,929	0,483	0,537	0,652	0,717

10.4.2.4 Diagnostika klasifikačních možností sítě a jednotlivých diskriminátorů

U diskriminačních úloh probíhá po natrénování neuronové sítě optimalizace klasifikačního prahu pro dosažení maximální výkonnosti. Obrázek 10.3 zobrazuje ROC křivky¹³ pro trénovací data (vlevo) a celková data (vpravo) a sítě trénované algoritmy zmíněnými v tabulce 10.4 na str. 97. Na obrázku 10.3 je zřejmé, že zatímco pro trénovací množinu je nalezení optimálního klasifikačního prahu možné, nalézt optimální

¹³Betinec [3]: „ROC křivky slouží k hodnocení a grafickému znázornění chování klasifikátorů při klasifikaci do dvou tříd. Narozdíl od tradičně užívaných měr kvality, jako jsou např. celková správnost (accuracy) nebo celková chyba (misclassification error), netrpí zkreslením, k němuž dochází při nevyváženém zastoupení jednotlivých tříd v populaci, proto se s nimi lze setkat v širokém spektru aplikací od medicíny, přes bankovníctví, až po humanitní vědy.“

klasifikační práh napříč skupinami je prakticky nemožné. Problematikou ROC křivek se zabývá například Repíková [73] nebo Betinec [3].



Obrázek 10.3: ROC křivky pro trénovací a celkovou výkonnost trénovacích metod – viz tabulka 10.4

(Vytvořeno v programu STATISTICA Neuronové sítě 7.1)

Až doposud byly pro data 621 obcí vytvářeny sítě, které svou sílu opíraly o všech sedm diskriminátorů. Výsledky (podkapitola 10.3) diskriminace provedené prostřednictvím klasické diskriminační analýzy naznačily, že některé diskriminátory je možno vynechat. Celý postup tvorby diskriminační funkce popisuje článek Čermáková – Biskup [117].

Při tvorbě neuronových sítí lze ke zjednodušení modelu použít funkci automatického *prořezávání* neuronů. Software STATISTICA 7.1 Neuronové sítě umožňuje automaticky vypouštět neurony na základě:

- nízkého zapojení libovolného neuronu do adaptivní dynamiky sítě (prakticky neprobíhá významná změna nízkých vah/parametrů při procesu učení sítě),
- nízkého zapojení libovolného neuronu do aktivní dynamiky sítě (daný neuron není aktivní při výpočtu výstupu sítě, tj. má prakticky stejnou – obvykle nízkou – hodnotu výstupu),
- citlivostní analýzy pro vstupní neurony (daný vstupní neuron „nezvyšuje“ výkonnost neuronové sítě).

Automatické prořezávání bylo využito až při automatickém generování sítí. Při porovnávání neuronových sítí VVP 1–VVP 12 by automatické prořezávání neuronů mohlo vést ke vzniku zcela jiných sítí (jednodušších) než bylo zamýšleno a požadováno.

Tabulka 10.7: Výsledky citlivostní analýzy vícevrstevných dopředných sítí 1–12

Sít		Citlivost						
		výšnm	míranez	hustob	bytnaek	fyznaek	rolnaek	zamnaek
VVP 1	Podíl	1,017	1,048	1	1,009	1,013	1,021	1,012
	Pořadí	3	1	7	6	4	2	5
VVP 2	Podíl	1	1,012	1,001	1,007	1,002	1,012	0,998
	Pořadí	6	2	5	3	4	1	7
VVP 3	Podíl	1,005	1,073	1,043	1,108	1,058	1,116	1,045
	Pořadí	3	4	7	2	5	1	6
VVP 4	Podíl	1,061	1,109	1,035	1,051	1,041	1,041	1,054
	Pořadí	2	1	7	4	6	5	3
VVP 5	Podíl	1,096	1,088	1,014	1,072	1,11	1,078	1,023
	Pořadí	2	3	7	5	1	4	6
VVP 6	Podíl	1,007	1,014	0,996	1,008	1,013	0,984	0,995
	Pořadí	4	1	5	3	2	7	6
VVP 7	Podíl	1,007	1,014	0,996	1,008	1,013	0,984	0,995
	Pořadí	4	1	5	3	2	7	6
VVP 8	Podíl	1,019	1,044	1,031	1,019	1,009	1,06	1,023
	Pořadí	5	2	3	6	7	1	4
VVP 9	Podíl	0,997	0,996	1,003	1,01	0,99	1,011	0,999
	Pořadí	5	6	3	2	7	1	4
VVP 10	Podíl	0,986	1,003	1	0,987	0,998	0,999	1,004
	Pořadí	7	2	3	6	5	4	1
VVP 11	Podíl	1,001	0,99	1,002	0,996	1,002	1,004	0,988
	Pořadí	4	6	3	5	2	1	7
VVP 12	Podíl	1,015	1,069	1,003	1,04	1,066	1,012	1,005
	Pořadí	4	1	7	3	2	5	6

V tabulce 10.7 jsou uvedeny podíly a pořadí citlivosti jednotlivých diskriminátorů pro používané vícevrstvé dopředné sítě, trénované na trénovací množině. Pokud je poměr citlivosti nižší než 1, lze předpokládat (viz StatSoft [81]), že vyřazení daného diskriminátoru nesníží nebo dokonce zvýší výkonnost sítě. Přestože výkonnost sítí nebyla neuspokojivá (viz tabulky 10.4–10.6), byl vliv jednotlivých diskriminátorů nízký. Navíc pořadí vlivu diskriminátorů bylo pro každou z uvedených sítí různé (viz tabulka 10.7). To potvrzuje domněnku, že žádný z diskriminátorů není lepší než jiný. Na základě průměrného pořadí citlivosti ($\bar{T}$) jednotlivých diskriminátorů pro

různé sítě se můžeme domnívat, že mezi vhodné diskriminátory lze zahrnout míru nezaměstnanosti ($\bar{T} = 2,636$) a význam soukromého zemědělství ($\bar{T} = 2,909$), zatímco mezi špatné diskriminátory patří hustota obyvatel ($\bar{T} = 5,182$) a zaměstnanecká „absorpce“ obce ($\bar{T} = 5$). Tyto výsledky se diametrálně liší od výsledků diskriminační analýzy (viz tabulka 10.1 na str. 91).

10.4.2.5 Očekávané i neočekávané výsledky trénování dopředné sítě

Na závěr podkapitoly zabývající se výsledky trénování vícevrstvé dopředné sítě je třeba připomenout, že trénování některých topologií neuronových sítí je proces striktně náhodný. Výsledek trénování neovlivňuje jen uživatelem definovaná topologie, ale i průběh učení a nastavení vah. Každá z natrénovaných sítí je unikátním výsledkem, jednoznačně definovaný pouze topologií sítě a vektorem synaptických vah.

Tabulka 10.8 shrnuje výkonnosti jedné konkrétní vícevrstvé dopředné sítě¹⁴ trénované za „stejných“ podmínek. Rozdílná výkonnost byla způsobena pouze počátečním náhodným nastavením vah a následně průběhem trénování, který probíhal v 500 epochách¹⁵. Daleko odlišnějších výsledků bylo dosahováno u sítí s jednodušší architekturou, nebo u sítí trénovaných v nižším počtu epoch¹⁶.

Tabulka 10.8: Výkonnost neuronové sítě 7:7-30-30-30-1:1 (VVP 7) opakovaně trénované

Trénování	Výkon			
	Trén.	Valid.	Test.	Celk.1
1	0,995	0,483	0,488	0,689
2	0,957	0,507	0,502	0,659
3	0,995	0,488	0,522	0,671
4	1	0,512	0,522	0,681

Odlišnost výsledků trénování se pochopitelně projevila i v citlivostní analýze jednotlivých diskriminátorů. Výsledky zobrazené v tabulce 10.9 sice shodně přiřazuje

¹⁴Úmyslně byla zvolena síť, u níž by se dalo očekávat, že díky své bohatosti a délce učení dospěje ke stejné výkonnosti. Nestalo se.

¹⁵V případě 4. trénování zobrazeném v tabulce 10.8 je vidět, že bylo dosaženo maximálního trénovacího výkonu, tj. někdy pouhá trpělivost může přinést dílčí výsledky. O výkonnosti sítě ale vypovídá validační, testovací a celková výkonnost, která není ani v tomto příliš silná.

¹⁶Tento výsledek se pochopitelně dá očekávat, neboť čím kratší je trénovací proces, tím více je výsledek závislý na počátečním nastavení. Trénovací výkon pro jednu trénovací epochu osciloval v rozsahu cca 20 %.

nízkou diskriminační schopnost hustotě obyvatel a zaměstnanecké „absorpci“ obce, ale jednoznačně nejvýznamnější diskriminátor neurčují.

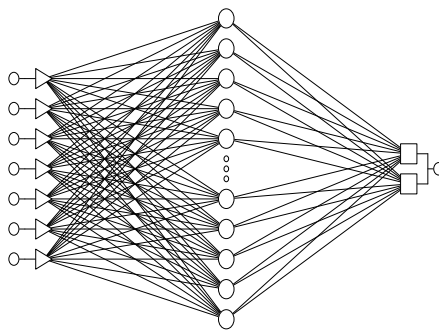
Tabulka 10.9: Výsledky citlivostní analýzy vícevrstvé dopředných sítě 7:7-30-30-30-1:1 (VVP 7) opakovaně trénované

Trénink		Citlivost						
		výšnm	míranez	hustob	bytnaek	fyznaek	rolnaek	zamnaek
1	Podíl	1,073	1,129	1,013	1,116	1,067	1,063	1,04
	Pořadí	3	1	7	2	4	5	6
2	Podíl	1,104	1,104	1,015	1,136	1,101	1,178	1,062
	Pořadí	4	3	7	2	5	1	6
3	Podíl	1,13	1,145	1,045	1,128	1,095	1,126	1,066
	Pořadí	2	1	7	3	5	4	6
4	Podíl	1,094	1,136	1,022	1,137	1,086	1,187	1,044
	Pořadí	4	3	7	2	5	1	6

Trénování vícevrstvé dopředné sítě přineslo i některé „paradoxní“ výsledky. Při tréninku sítě na celém souboru dat (všechny obce) bylo dosaženo nižšího celkového výkonu sítě (sloupec Celk.2 v tabulce 10.4), než v případě, kdy byla k trénování použita jen její část (sloupec Celk.1 v tabulce 10.4). Uvedený paradox lze vysvětlit tím, že celý soubor dat mohl obsahovat více problémových, netypických případů (obcí), které „kazily“ proces učení. Paralelu s takovými výsledky „učení“ je možno pozorovat v reálném životě. Pokud totiž učení organismu probíhá na základě všech, tj. i nevhodných, vzorů z okolí, je možné, že se buď učí špatné „věci“, nebo proces učení není úspěšný. Organismus je ve své snaze naučit se „zmatený“ a proces učení ve svém výsledku selže, popřípadě selže již při samotném procesu učení. Ale i klasické statistické metody často selžou, pokud je pohled na soubor dat příliš nekritický. Příkladem jsou např. nevhodné interpretace výsledků regresní (vhodnost modelu, parametry modelu) a korelační analýzy (síla závislosti měřená koeficientem/indexem korelace) při výskytu vlivných hodnot atp.

10.4.3 Pravděpodobnostní neuronová síť

Další vhodnou sítí pro řešení klasifikačních úloh je pravděpodobnostní neuronová síť¹⁷, někdy též nazývaná Bayesovská síť. Trénování pravděpodobnostní neuronové sítě (adaptivní dynamika sítě) probíhá – na rozdíl od vícevrstvé dopředné sítě – velice rychle, avšak výpočty natrénovanou neuronovou sítí (aktivní dynamika sítě) jsou náročné jak na čas, tak na kapacitu počítače. Podstatnou výhodou pravděpodobnostní sítě jsou její dobré vlastnosti v případě klasifikace souboru dat s odlehlými hodnotami. Z uvedených důvodů je v předložené práci trénována pro diskriminaci obcí i pravděpodobnostní neuronová síť.



Obrázek 10.4: Schéma pravděpodobnostní neuronové sítě 7:7-211(621)-2:1

(Vytvořeno v programu *STATISTICA Neuronové sítě 7.1*)

10.4.3.1 Vyhlazovací faktor a výkonnost sítě

Výsledky trénování Bayesovských sítí ovlivňuje tzv. *vyhlazovací faktor*, kterým je v případě použití vícerozměrného normálního rozdělení pro každou třídu směrodatná odchylka. Optimální volbou vyhlazovacího faktoru se zabývá například Paprzycki

¹⁷Pravděpodobnostní neuronovou sítí je rozuměna třívrstvá dopředná síť, která odhaduje (např. jádrové odhady) na základě trénovací množiny hustotu pravděpodobnosti pro každou třídu. Jako hustota pravděpodobnosti bývá volena obvykle Parzenova hustota pravděpodobnosti, nebo hustota vícerozměrného normálního rozdělení. Vstupní vrstva je úplně propojena se všemi neurony skryté vrstvy, kterých je právě tolik, jako je trénovacích vzorů. Výstupní vrstva obsahuje neuron pro každou třídu. Ten, který vypočte vyšší hodnotu „vyhrává“ – tj. do té třídy bude zařazen vstupní objekt. Poměrně shrnujícím a poučným článkem, v němž je popsán Bayesovský přístup k problematice neurčitosti a jejího řešení prostřednictvím neuronových, je článek Hanzelka [24].

et al. [69], kde jeho autoři doporučují jeho nastavení v intervalu $\langle 1,25;1,55 \rangle$.¹⁸ Výsledky, zjištěné na zkoumaném souboru dat, však ukázaly jiné optimální hodnoty¹⁹ – viz tabulka 10.10.

Sloupce „Celk_1“ a „Celk_2“ v tabulce 10.10 uvádějí celkovou výkonnost natrénované sítě na trénovací množině a na celé množině tak, jak bylo zavedeno již dříve. Pro určení Celk_1 je trénována pravděpodobnostní neuronová síť 7:7-211-2:1, kdežto pro určení Celk_2 pravděpodobnostní neuronová síť 7:7-621-2:1. 211 respektive 621 skrytých neuronů odpovídá počtu trénovacích vzorů – viz poznámka pod čarou číslo 17 na str. 105. Prakticky se tedy jedná o dvě různé sítě, přestože jsou značeny stejně – tj. PNS i , $i = 1, \dots, 8$. (viz obrázek 10.4).

Výkonnost pravděpodobnostní neuronové sítě je bez ohledu na volbu vyhlazovacího faktoru vyšší než 64 %. Oproti vícevrstvé dopředné síti nedává pro opakované trénování za „stejných“ podmínek jiné výsledky, což je důsledkem jak topologie sítě, tak způsobem trénování. Tj. každé datové množině přísluší pro zvolenou topologii (včetně vyhlazovacího parametru) právě jedna natrénovaná síť²⁰.

Tabulka 10.10: Výkonnost trénování pravděpodobnostní neuronové sítě 7:7-211(621)-2:1 v závislosti na volbě vyhlazovacího faktoru

Vyhlazovací faktor	Výkon				
	Trén.	Valid.	Test.	Celk_1	Celk_2
PNS 1 $\sigma = 0,01$	1	0,483	0,443	0,646	1
PNS 2 $\sigma = 0,1$	0,834	0,579	0,632	0,683	0,7
PNS 3 $\sigma = 0,5$	0,569	0,675	0,682	0,641	0,641
PNS 4 $\sigma = 1$	0,569	0,665	0,692	0,641	0,641
PNS 5 $\sigma = 1,3$	0,569	0,665	0,692	0,641	0,641
PNS 6 $\sigma = 1,5$	0,569	0,665	0,692	0,641	0,641
PNS 7 $\sigma = 2$	0,569	0,665	0,692	0,641	0,641
PNS 8 $\sigma = 5$	0,569	0,665	0,692	0,641	0,641

Výsledky uvedené v tabulce 10.10 ukazují, že maximální výkonnosti bylo dosa-

¹⁸Pravděpodobnostní neuronová síť obvykle pracuje s normalizovanými numerickými proměnnými, takže takováto volba parametru jasně svědčí o tendenci „vyhladit“ ostré klasifikování a vyhnout se přetrénování neuronové sítě.

¹⁹Z pohledu pouhé maximalizace klasifikačního výkonu a získání nedegenerovaného řešení (více na str. 107)!

²⁰Samozřejmě pokud by bylo pracováno s daty symbolicky. Rozdílné výsledky může způsobit použití odlišného softwaru, různá přesnost při (numerických) výpočtech, atp.

ženo pro malé hodnoty vyhlazovacího faktoru. Při volbě velmi nízkého vyhlazovacího faktoru ($\sigma = 0,01$) byla síť schopná na základě zvolených diskriminátorů přesně diskriminovat všechny obce do dvou skupin a to jak pro trénovací množinu – 211 obcí, tak pro celkový datový soubor – 621 obcí. Validací a testovací výkonnost (viz první řádek tabulky – PNS 1) však byla pro tuto hodnotu vyhlazovacího faktoru neuspokojivá. Nejvyšší celkový výkon pro trénování na trénovací množině byl dosažen při hodnotě vyhlazovacího faktoru $\sigma = 0,1$. Faktor $\sigma = 0,01$, který sice maximalizoval výkonnost sítě pro trénovací množinu, dával celkový výkon sítě srovnatelný s celkovými výkony pro vyhlazovací faktory s hodnotami vyššími nebo rovnými než 0,5.²¹

10.4.3.2 Generalizace jako podstatný faktor výkonnosti sítě

Z výsledků uvedených v tabulce 10.10 je také patrné, že neuronové síti PNS 1 se podařilo ve svém parametrickém prostoru zakódovat úplnou informaci o migraci v Jihočeském kraji²². Na druhou stranu tato síť vykazovala přesprílišnou fixovanost na trénovací množinu a tedy nízkou generalizační schopnost. Obce Jihočeského kraje byly rozděleny znovu a to tak, aby trénovací množina obsahovala „většinu“ obcí (náhodně bylo do trénovací množiny vybráno 90 %–95 % všech obcí) a validace probíhala na zbytku (tj. malý zlomek případů). Vyhlazovací faktor byl nastaven na 0,01. Po zkušenostech z minulého trénování nebylo překvapující, že trénovací výkonnost dosahovala opět 100 %.

Generalizační schopnosti natrénovaných sítí posuzované validačními výkony byly nízké²³. Validací výkonnosti se pohybovaly od 40 % do 50 %, výjimečně do 60 %. Přitom poměr obcí, klasifikovaných neuronovou sítí jako obce se zápornou migrací, byl vyšší než v původním souboru. Chování natrénované pravděpodobnostní neuronové sítě svědčí – vedle nízkého validačního výkonu oproti výkonu trénovacímu – o fixovanosti sítě na trénovací data. Trénovací množina obsahovala větší množství

²¹Sítě s vyhlazovacím faktorem vyšším než 0,5 zařadily všechny obce mezi obce s nezápornou migrací, tj. maximalizovaly svou výkonnost tak, že všechny obce zařadily do početnější skupiny dle salda migrace v trénovací množině. Takto „degenerované“ řešení klasifikačního problému bylo získáno již u vícevrstvé dopředné sítě a to při trénování metodou Delta-bar-delta.

²²100% natrénovanosti bylo možné dosáhnout díky tomu, že se v souboru dat nevyskytovaly dvě obce se stejnými ukazateli a přitom jiným saldem migrace.

²³Mimo jiné probíhalo trénování na pravděpodobnostních neuronových sítích: 7:7-570-2:1, 7:7-564-2:1, 7:7-562-2:1, 7:7-556-2:1, 7:7-550-2:1, ... Různý počet skrytých neuronů a tedy trénování formálně jiných sítí (architekturou) odpovídá počtu obcí v trénovací množině. O tom byla již několikrát zmínka.

obcí s nezápornou migrací²⁴, a proto je dokázala neuronová síť na nich natrénovaná diskriminovat. Pokud byl neuronové síti tohoto typu předložen neznámý vzor (z validační množiny), měla tendenci jej spíše zařadit mezi obce se zápornou migrací²⁵. Toto chování pravděpodobnostní neuronové sítě se liší od trénování pravděpodobnostních neuronových sítí s vyhlazovacím parametrem vyšším nebo rovným než 0,5.

I přes fixovanost pravděpodobnostní neuronové sítě na trénovací data zapříčiněnou nízkým vyhlazovacím faktorem se podařilo ověřit tvrzení vyslovené na začátku podkapitoly 10.4.3 – tvrzení, že testovací a validační výkony pravděpodobnostních sítí jsou minimálně srovnatelné s testovacími a validačními výkony vícevrstvé dopředné sítě²⁶.

10.4.3.3 Diagnostika klasifikačních možností sítě a jednotlivých diskriminátorů

Analýza citlivosti pro pravděpodobnostní neuronové sítě (viz tabulka 10.11) přinesla obdobné výsledky, jako analýza citlivosti u vícevrstvých dopředných sítí. Za diskriminátory s nejnižší citlivostí byly označeny: hustota obyvatel a zaměstnanecká „absorpce“ obce. Podíly spolehlivosti rovny jedné²⁷ pro PNS 4–8 respektive i PNS 3 naznačují necitlivost k různému (možná jakémukoliv) vlivu jednotlivých diskriminátorů pro diskriminaci provedenou pravděpodobnostní neuronovou sítí s vyhlazovacím parametrem větším než 0,5.

10.4.4 RBF síť

Vhodným nástrojem pro diskriminaci je i RBF síť. V podkapitole 4.1.2 byla popsána základní stavební jednotka RBF sítě, tzv. semilokální jednotka. Z těchto jednotek je složena skrytá vrstva sítě (viz obrázek 4.3 na str. 42). V procesu trénování

²⁴Taková je však reálná situace – viz tabulka 10.3 na str. 94.

²⁵Síť se chovala podle pravidla: „To, co neznám/nepoznám, přiřadím k méně početné skupině. Tu znám méně, a proto by tam neznámý vzor mohl patřit.“

²⁶Validační výkony vícevrstvých dopředných sítí v podkapitole 10.4.2 sice nebyly optimalizovány a byly nižší, než kdyby byly sítě trénovány jen do okamžiku, kdy se validační výkon začal zvyšovat, takže toto srovnání je poněkud zavádějící. Ale i při srovnání validačních a testovacích výkonů pravděpodobnostních sítí a dopředných sítí vytvořených v podkapitolách 10.4.5 a 10.4.6, u nichž optimalizace validačního výkonu proběhla, byly tyto výkony u pravděpodobnostních sítí vyšší.

²⁷Odchytky hodnot podílu citlivosti jednotlivých faktorů se projeví až na 5. desetinném místě, a přestože bylo možno určit pořadí citlivosti jednotlivých diskriminátorů, nelze ho považovat za signifikantní.

Tabulka 10.11: Výsledky citlivostní analýzy pravděpodobnostních sítí 7:7-211-2:1 v závislosti na volbě vyhlazovacího faktoru

Trénink		Citlivost						
		výšnm	míranez	hustob	bytnaek	fyznaek	rolnaek	zamnaek
PNS 1	Podíl	1,102	1,116	1,009	1,036	1,071	1,128	1,01
	Pořadí	3	2	7	5	4	1	6
PNS 2	Podíl	1,024	1,022	0,998	1,02	1,018	1,027	1,004
	Pořadí	2	3	7	4	5	1	6
PNS 3	Podíl	1,001	1,001	1	1,001	1	1,001	1
	Pořadí	2	1	7	3	5	4	6
PNS 4–8	Podíl	1	1	1	1	1	1	1
	Pořadí	1	3	7	2	6	4	5

RBF sítě jsou nejprve lokalizovány semilokální jednotky, následně jsou určeny jejich parametry a nakonec dochází k optimalizaci lineárního výstupu celé sítě. Program STATISTICA 7.1 Neuronové sítě nabízí k lokalizaci – „rozprostření“ – skrytých jednotek dvě metody: náhodný výběr z trénovacích vzorů a metodu „k-průměrů“. Nastavení parametrů přenosové funkce může proběhnout:

- manuálně (shodný rozptyl pro všechny jednotky),
- heuristicky/automaticky (shodný rozptyl pro všechny jednotky) a
- prostřednictvím metody „k-nejbližších sousedů“.

Optimalizace lineárního výstupu sítě probíhá v programu STATISTICA 7.1 Neuronové sítě pomocí pseudoinverze (singulární dekompozice, SVD – Singular Value Decomposition).

Trénovací výkonnosti RBF sítí s různými počty semilokálních jednotek byly podle očekávání různé – viz tabulka 10.12. Zajímavé bylo zjištění, že zvyšující se trénovací výkonnost (korespondující se zvyšujícím se počtem semilokálních jednotek) neměla za následek zhoršení validační a testovací výkonnosti, tj. nedocházelo k přeučení sítě. Byl-li do skryté vrstvy RBF sítě zvolen počet semilokálních jednotek rovný počtu vzorů, dosáhla RBF síť absolutní trénovací výkonnosti²⁸. Pro posuzování chyb sítě byl zvolen součet čtverců chyb²⁹, centroidy pro lokalizaci semilokálních jednotek byly

²⁸Asymetričnost (neúplnost) tabulky je dána maximálním počtem skrytých neuronů (= semilokálních jednotek), které lze v architektuře sítě zvolit. Proto bylo možno sítě RBF 8–10 trénovat pouze na celém souboru dat a část tabulky je prázdná.

²⁹Pro tento soubor dat dával součet čtverců chyb použitý pro posuzování chyby lepší výsledky (vyšší výkonnost po trénování) než v případě, kdy bylo použito ke stejnému účelu entropie.

vybrány z trénovacích vzorů a rozptyl byl odhadnut heuristicky/automaticky shodně pro všechny jednotky. Trénování probíhalo na normovaných vstupních proměnných, tj. hodnoty jednotlivých numerických diskriminátorů byly standardizovány.

Tabulka 10.12: Výkonnost trénování RBF sítě v závislosti na počtu semilokálních jednotek

Počet jednotek	Výkon				
	Trén.	Valid.	Test.	Celk.1	Celk.2
RBF 1 7	0,626	0,522	0,557	0,568	0,538
RBF 2 20	0,649	0,502	0,562	0,572	0,573
RBF 3 50	0,706	0,512	0,498	0,573	0,622
RBF 4 100	0,829	0,474	0,532	0,614	0,641
RBF 5 150	0,877	0,512	0,567	0,654	0,663
RBF 6 200	0,986	0,440	0,478	0,638	0,742
RBF 7 211	1	0,531	0,537	0,692	0,734
RBF 8 350					0,842
RBF 9 500					0,915
RBF 10 621					1

Také trénování RBF sítě (hodnoceno prostřednictvím trénovacího výkonu a výsledky klasifikace) je procesem značně závislým na počáteční volbě parametrů – zejména rozložením semilokálních jednotek. Na druhou stranu se neprokázalo, že by celkovou výkonnost RBF sítě (Celk.1 a Celk.2)³⁰ ovlivňovala metoda zvolená pro lokalizaci semilokálních jednotek, tj. zda byla jejich centra volena náhodně na základě trénovacích vzorů, nebo byla určována prostřednictvím metody k-průměrů. Výsledky příliš neovlivňoval ani výběr metod pro vhodnou volbu rozptylu RBF funkcí (heuristicky/automaticky nebo prostřednictvím metody k-nejbližších sousedů³¹).

Výsledky citlivostní analýzy i analýza ROC křivek byly obdobné výsledkům získaným u sítí předchozích typů a proto nejsou na tomto místě znovu prezentovány.

10.4.5 Automatické vyhledávání sítí 1

Program STATISTICA 7.1 Neuronové sítě disponuje heuristikou, která nabízí pomoc při tvorbě a trénování neuronových sítí. Na počátku řešení problému stačí spe-

³⁰Například pro RBF síť se 100 skrytými neurony se celková výkonnost Celk.1 pohybovala kolem 60 %.

³¹Manuální nastavení rozptylu RBF funkcí nebylo využíváno.

cifikovat jakou úlohu³² je třeba prostřednictvím neuronových sítí řešit. Míra pomoci při řešení problému má dvě úrovně:

1. **Vlastní návrh sítě** – vytvoří, dle kontextové nápovědy softwaru, neuronovou síť daného typu a architektury se zvolenými vstupními/výstupními proměnnými.
2. **Průvodce řešením** – dle kontextové nápovědy softwaru vytvoří a otestuje neuronové sítě pro analýzu dat a predikce. Automaticky navrhne sítě pro řešení daného problému.

Mód Vlastní návrh sítě umožňuje poměrně nezávislé nastavení konkrétního typu neuronové sítě a „automatika“ řeší jen dílčí problémy adaptivní dynamiky, jakými jsou např. nastavení hodnoty klasifikačního prahu, rozložení semilokálních jednotek, prořezávání neuronů, včasné ukončení trénovacího procesu s ohledem na validační chybu a diverzitu, atp. Módu vlastní návrh sítě bylo omezeně použito při trénování sítí popsaných v podkapitolách 10.4.2–10.4.4.

Průvodce řešením programu STATISTICA 7.1 Neuronové sítě dle nastavených parametrů vytváří a trénuje neuronové sítě, které mezi sebou následně porovnává a vybírá z nich ty nejvýkonnější. Aby bylo možno porovnat úspěšnost heuristiky s výsledky získanými „manuálně“, bylo i zde zvoleno stejným způsobem rozdělení obcí do tří množin. Díky tomu lze srovnat nejen celkový výkon – Celk_1, ale také výkonnost trénovací, validační a testovací. Automatické trénování na celém souboru dat nemělo smysl, neboť heuristika by neměla dle čeho volit/zlepšovat/optimalizovat parametry procesu učení a vybírat nejlepší sítě – chyběla by validační a testovací výkonnost, zejména validační chyba. Na uživateli nezávislé výsledky budou shrnuty v podkapitole 10.4.6.

Prostřednictvím heuristiky v módu Průvodce řešením bylo testováno 1000 sítí a z nich software vybral 50 nejlepších. Mezi testované sítě patřila:

- vícevrstvá dopředná síť – maximálně čtyřvrstvá (dvě skryté vrstvy) s nejvýše 30-ti skrytými neurony v každé z vrstev,
- pravděpodobnostní síť,
- RBF síť – s maximálně 211 semilokálními jednotkami.

S ohledem na srovnatelnost sítí vzniklých prostřednictvím heuristiky se sítěmi vytvořenými v podkapitolách 10.4.2–10.4.4 byla zvolena právě omezení zmíněná výše. Pokud by nebyly parametry nastaveny tímto způsobem, program STATISTICA 7.1

³²STATISTICA 7.1 Neuronové sítě umožňuje řešit regresní a klasifikační úlohy, úlohy shlukové analýzy a úlohy z oblasti časových řad.

Neuronové sítě by navrhl daleko jednodušší/chudší architektury sítí³³. Minimální a maximální hodnoty výkonnosti na jednotlivých množinách shrnuje tabulka 10.13 a tabulka 10.14. Matice kategorizovaných krabicových grafů vizualizuje na obrázku 10.5 výkonnost 50-ti nejlepších sítí.

Tabulka 10.13: Výkonnost trénování neuronových sítí vybraných heuristikou – 50 nejlepších

Ukazatel	Výkon			
	Trén.	Valid.	Test.	Celk.1
Minimum	0,479	0,488	0,483	0,517
Maximum	0,692	0,675	0,692	0,652

Tabulka 10.14: Výkonnost trénování neuronových sítí testovaných heuristikou – 1 000 sítí

Ukazatel	Výkon			
	Trén.	Valid.	Test.	Celk.1
Minimum	0,412	0,426	0,453	0,440
Maximum	0,697	0,675	0,692	0,652

10.4.5.1 Induktivní zhodnocení výsledků

Následující induktivní úvahy byly vytvářeny pro 50 nejlepších sítí vytvořených prostřednictvím heuristiky programu STATISTICA 7.1 Neuronové sítě. Prostřednictvím Friedmanova testu se nepodařilo prokázat ($Q = 4,68$, $p\text{-value} = 0,096$), že by na 5% hladině významnosti byla jiná distribuce alespoň jedné z výkonností – trénovací, validační a testovací. To jest testována byla hypotéza pro tři závislé soubory:

H₀: $F_{\text{Trén.}}(x) = F_{\text{Valid.}}(x) = F_{\text{Test.}}(x)$ (distribuce hodnot trénovacího, validačního a testovacího výkonu jsou shodné), proti

H_A: $\neg H_0$ (alespoň jedna distribuce hodnot trénovacího, validačního a testovacího výkonu je odlišná od ostatních)

³³Pro vícevrstvou dopřednou síť pouze třívrstvou síť s 11-ti skrytými neurony v pouze jedné skryté vrstvě a pro RBF síť pouze 105 semilokálních jednotek.

a na 5% hladině významnosti nebylo možno nulovou hypotézu zamítnout. Heuristika tedy zabránila přetrénování neuronových sítí³⁴, které bylo možno sledovat při manuálním trénování (viz např. obrázek 10.2 na str. 99). Při přetrénování by byl jistě evidentní rozdíl mezi trénovací výkonností a výkonností validační a testovací. Heuristika například u vícevrstvých sítí ukončovala trénovací proces po maximálně 3 trénovacích epochách. Ani pro všech 1000 heuristikou vybraných sítí neprobíhal trénovací proces příliš dlouho – maximální doba trvání trénovacího procesu (tj. do okamžiku, než se validační chyba začala zvyšovat) byla 69 trénovacích epoch. RBF sítě vytvořené heuristikou obsahovaly maximálně 42 semilokálních jednotek. Pravděpodobnostní neuronová síť velice brzy našla takové nastavení parametrů, jež maximalizovalo její validační výkon (0,665). Bohužel toto nastavení vedlo k degenerované klasifikaci zmíněné již na str. 107.

Kruskalův-Wallisův test³⁵, který testoval

H₀: $F_{VVP}(x) = F_{PNS}(x) = F_{RBF}(x)$ (distribuce hodnot výkonnosti vícevrstvých dopředných neuronových sítí, pravděpodobnostních neuronových sítí a RBF sítí je shodná), proti

H_A: $\neg H_0$ (alespoň jedna distribuce výkonnosti vícevrstvých dopředných neuronových sítí, pravděpodobnostních neuronových sítí a RBF sítí je odlišná od zbylých typů sítí),

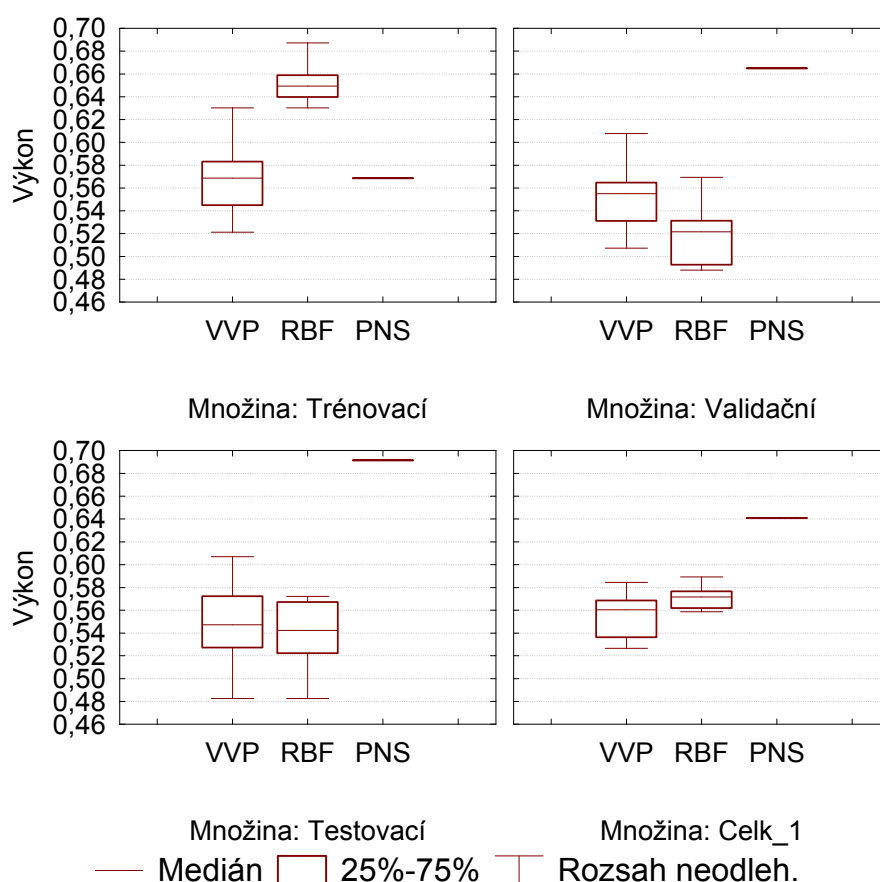
podle očekávání prokázal (p -value = 0), odlišnou výkonnost – trénovací, validační, testovací i celkovou – jednotlivých typů testovaných sítí, viz krabicové grafy na obrázku 10.5.

Naprosto nesrovnatelné výsledky mezi manuálním nastavením a heuristikou byly získány u vícevrstvých dopředných sítí. Více než s počty neuronů ve skrytých vrstvách operovala heuristika s počátečním nastavením vah³⁶. Heuristika – maximalizující validační výkonnost – velice omezila výsledný počet trénovacích epoch. To mělo za následek poměrně nízký trénovací výkon v porovnání s výsledky, jež jsou popsány

³⁴Průměrná pořadí naopak ukazují, že heuristikou vybrané neuronové sítě mají nižší výkon validační než trénovací – $\bar{T}_{\text{Trén.}} = 2,18$, $\bar{T}_{\text{Valid.}} = 1,76$ a pro úplnost $\bar{T}_{\text{Test.}} = 2,06$.

³⁵Parametrický test ANOVA nebylo možno použít pro silné porušení normality dat, navíc byla prokázána nestejná variabilita (Levenův test, $F = 12,608$, p -value = 0) výkonnosti jednotlivých typů sítí – krabicové grafy na obrázku 10.5 tuto situaci jasně zobrazují (viz variabilita u pravděpodobnostní neuronové sítě).

³⁶Program STATISTICA 7.1 Neuronové sítě sice plně zpřístupňuje parametry natrénované sítě, z nichž mnohé plyne (počet neuronu, počet vrstev, trénovací metoda, výsledné nastavení vah, atp.), ale jakých počátečních parametrů (vyhlazovací parametry, míry učení, atp.) bylo použito se dopátrat nedá.



Obrázek 10.5: Krabicové grafy – závislost výkonnosti jednotlivých typů sítí (50 nejlepších sítí vybraných heuristikou)

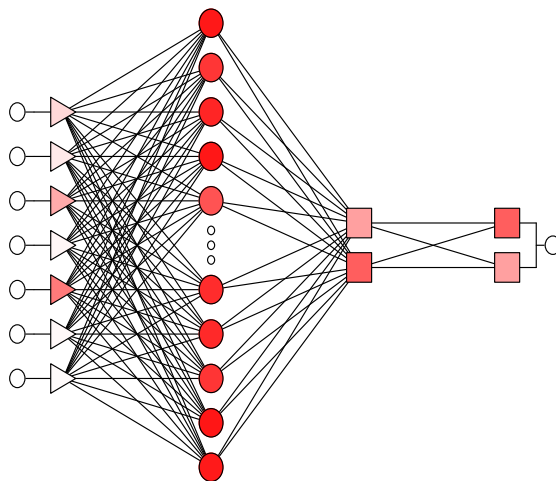
(Vytvořeno v programu STATISTICA komplet 6.1 Cz)

v podkapitole 10.4.2. Validační a testovací výkony byly oproti sítím trénovaným převážně v 500 epochách podle očekávání vyšší.

10.4.6 Automatické vyhledávání sítí 2

Výsledky prezentované v podkapitole 10.4.6 byly získány tak, že programu STATISTICA 7.1 Neuronové sítě byla nechána maximální volnost. Neproběhlo tedy žádné rozdělování obcí do trénovací, validační a testovací množiny, žádné navyšování počtu neuronů a semilokálních jednotek. Program rozdělil data do dvou přibližně stejně početných množin. 311 obcí využil pro trénování (trénovací množina) a zbytek (310 obcí) použil k porovnání a výběru vhodných sítí – 155 obcí pro výběr z trénovaných sítí

(validační množina) a 155 obcí pro porovnání výkonnosti sítí na obcích nezařazených do tréninku (testovací množina). Natrénované sítě byly vybírány na základě nejvyšší validační výkonnosti³⁷ a diversity sítí. Z toho důvodu nedosahovaly výsledné sítě vysokých trénovacích výkonností (do 70 %). Nejlepší sítí, kterou heuristika pro diskriminaci vytvořila, byla označena RBF neuronová síť. Schéma její architektury je uvedeno na obrázku 4.3, str. 42.



Obrázek 10.6: Schéma pravděpodobnostní (Bayesovské) neuronové sítě 7:7-311-2-2:1
(Vytvořeno v programu *STATISTICA Neuronové sítě 7.1*)

Mezi další trénované sítě heuristika zvolila a natrénovala vícevrstvou dopřednou síť s jednou skrytou vrstvou a pravděpodobnostní neuronovou síť (obrázek 10.6). Oproti obrázku 10.4 je na obrázku 10.6 zobrazena čtveřice, nikoli dvojice, výstupních neuronů. Dva „nadpočetné“ neurony (viz definice pravděpodobnostní neuronové sítě na str. 105) realizují zařazení ztrátové matice³⁸ do trénování pravděpodobnostní neuronové sítě. Váhy vzešlé z trénování však vliv ztrátové matice³⁹ eliminovaly.

Neuronovým sítím, které heuristika samostatně vybrala, bylo předloženo všech

³⁷Validační výkonnost, respektive validační chyba sloužila taktéž k ukončování iteračního procesu učení u vícevrstvé dopředné sítě.

³⁸Ztráta způsobená nesprávnou klasifikací nemusí mít pro řešení problému stejně důsledky. Ztrátová matice specifikuje, která z nesprávných klasifikací má vážnější důsledky – přináší vyšší ztrátu. Na hlavní diagonále ztrátové matice se nachází nuly a mimo ni ztráta způsobená nesprávnou klasifikací. Takto specifikovaná ztráta ovlivňuje proces trénování, který se snaží zabránit nesprávným klasifikacím s vyššími ztrátami.

³⁹Program *STATISTICA 7.1 Neuronové sítě* zvolil symetrickou ztrátovou matici, jež samozřejmě

7 vybraných diskriminátorů. Dle typu neuronové sítě se podařilo dosáhnout celkové výkonnosti Celk_1 i 68 % (pro vhodné topologie a trénovací algoritmy). Podstatné výsledky pro nejúspěšnější neuronové sítě jednotlivých architektur uvádí tabulka 10.15. Mnoho neuronových sítí však dosahovalo po natrénování daleko horších výsledků. Část neuronových sítí vybraných heuristikou programu diskriminovalo degenerovaně. Bez bližšího hodnocení funkčnosti sítí by pak mohlo překvapit, že mnoho sítí (různých architektur) diskriminovalo s celkovou výkonností cca 64,1 %. Přitom takto vysoké diskriminace dosahovaly i sítě, které měly pouze jeden vstupní neuron! Tedy pouze jeden diskriminátor byl schopen diskriminovat s 64,1% spolehlivostí. Bohužel tyto sítě podobně jako pravděpodobnostní sítě diskriminovaly degenerovaně. Drobné odchylky v celkové výkonnosti byly způsobeny několika málo obcemi, které byly klasifikovány jako obce se zápornou migrací – některé správně některé chybně.

Tabulka 10.15: Celková výkonnost (Celk_1) neuronových sítí vybraných heuristikou

Typ sítě	VVP H (7:7-5-1:1)	RBF H (7:7-83-1:1)	PNS H (7:7-311-2-2:1)
Celk_1	0,591	0,676	0,643

Trénink vícevrstevných dopředných sítí obvykle končil po prvních 100 epochách trénování prostřednictvím metody backpropagation. Pokud heuristika nechala probíhat trénování i v druhé etapě, pak bylo toto trénování (metodou sdružených gradientů) ukončeno nejpozději po 5-ti trénovacích epoch. Pro RBF sítě byly v procesu trénování lokalizovány semilokální jednotky prostřednictvím metody k-průměrů (pouze jednou metodou náhodné volby z trénovacích vzorů) a rozptyl byl nastavován metodou k-nejbližších sousedů.

nemohla výsledky ovlivnit. Při experimentálním manuálním nastavení ztrátové matice za účelem kompenzování „degenerovanosti“ klasifikace tak, jak navrhuje StatSoft [81], nebylo dosaženo vhodných výsledků. Volba ztrátové matice, která by penalizovala zařazení obce se zápornou migrací mezi obce s nezápornou migrací měla následující efekt:

- nižší relativní penalizace – nižší než 3:1 (oproti zařazení obce s nezápornou migrací mezi obce se zápornou migrací) – nezpůsobila změnu chování natrénované sítě, všechny obce byly sítě stále označeny za obce s nezápornou migrací;
- vyšší relativní penalizace – vyšší než 5:1 – způsobila úplnou změnu chování natrénované sítě, všechny obce byly označeny naopak za obce se zápornou migrací;
- při relativní penalizaci mezi 3:1–5:1 docházelo k postupnému posunu od jedné k druhé degenerované klasifikaci.

Využití ztrátové matice žádným způsobem nezvýšilo výkonnost sítě.

Hodnocení významnosti diskriminátorů bylo provedeno prostřednictvím citlivostní analýzy. Podíly a pořadí citlivosti pro vybrané neuronové sítě natrénované heuristikou (VVP H, RBF H a PNS H) zobrazuje tabulka 10.16. Jednotlivé podíly se od sebe liší jen nepatrně a jsou velice citlivé i na nastavení samotné sítě. Například citlivostní analýza provedená u RBF sítě s parametry 7:7-68-1:1 má pořadí důležitosti diskriminátorů odlišné od sítě RBF 7:7-83-1:1 (tj. RBF H). Pořadí diskriminátorů u RBF 7:7-68-1:1 určené na základě citlivostní analýzy je (vzestupně): rolnaek, fyznaek, míranek, hustob, výšnm, zamnaek a bytnaek, zatímco u RBF H (viz tabulka 10.16): rolnaek, výšnm, míranek, fyznaek, zamnaek, hustob a bytnaek.

Tabulka 10.16: Výsledky citlivostní analýzy neuronových sítí vybraných heuristikou

Sít		Citlivost						
		výšnm	míranek	hustob	bytnaek	fyznaek	rolnaek	zamnaek
VVP H	Podíl	0,993	0,998	1,005	1,002	0,999	0,997	1,002
	Pořadí	7	5	1	2	4	6	3
RBF H	Podíl	1,065	1,043	1,011	1,01	1,041	1,075	1,025
	Pořadí	2	3	6	7	4	1	5
PNS H	Podíl	0,998	0,999	1	1	0,998	0,997	1
	Pořadí	6	4	1	3	5	7	2

10.4.7 Výsledky klasifikace prostřednictvím neuronových sítí

V podkapitolách 10.4.2–10.4.6 byla jediným měřítkem úspěšnosti neuronových sítí (v závislosti na volbě topologie sítě a metodě trénování sítě) jejich výkonnost. Tyto kapitoly taktéž zmiňují mnoho výsledků, které již na tomto místě nejsou opakovány ani ve větší míře shrnovány. Podkapitola 10.4.7 přináší výsledky samotné diskriminace – diskriminační matice, které vypovídají o způsobu diskriminace⁴⁰ Z hodnot klasifikačních matic uvedených v tabulkách 10.18 a 10.19 na str. 119 a 120 je zřejmé, že neuronové sítě, stejně jako metody diskriminační analýzy (viz klasifikační matice, tabulka 10.2, str. 92) lépe diskriminovaly obce s nezápornou migrací. Rozložení špatně a dobře zařazených obcí s ohledem na saldo migrace se však lišilo.

⁴⁰O degenerované klasifikaci sice již byla řeč, ale bližší hodnocení způsobu klasifikace neuronovými sítěmi zatím nebylo zmíněno ani prezentováno. Jedná se tedy o první komplexní hodnocení způsobu klasifikace, k němuž bylo prostřednictvím neuronových sítí dosaženo.

Tabulky v podkapitolách 10.4.2–10.4.7 jsou doplněny o dvojí typ zvýraznění. Fialové hodnoty zobrazují výkonnosti sítí, které pracují tak, že označí všechny obce za obce s nezápornou migrací. Tato výkonnost (trénovací, validační, testovací a celkovou) dosahuje hodnot zobrazených v tabulce 10.17. Tučné hodnoty v tabulkách zvýrazňují skutečnost, že bylo dosaženo výkonnosti vyšší, než v případě degenerovaného řešení.

Tabulka 10.17: Výkonnost sítí realizující degenerované řešení

Výkon				
Trén.	Valid.	Test.	Celk_1	Celk_2
0,569	0,665	0,692	0,641	0,641

10.4.7.1 Neuronové sítě s manuálním nastavením trénovacího procesu

V tabulce 10.18 jsou uvedeny jako příklad neuronové sítě typů prezentovaných v podkapitolách 10.4.2–10.4.4, trénované na základě manuálního nastavení (tj. zástupce vícevrstevných dopředných sítí, pravděpodobnostních neuronových sítí a RBF sítí).

Jako nejlépe natrénované sítě byly autorem předkládané práce vybrány sítě, jež dosahovaly uspokojivé⁴¹ nikoli maximální celkové! Maximální celkové výkonnosti totiž mnohdy dosahovaly sítě:

- přetrénované – vysoká trénovací výkonnost zvyšovala i celkovou výkonnost Celk_1, přestože validační a testovací výkonnost byla nízká (např. VVP 5, VVP 7 a PNS 1);
- diskriminujících necitlivě – zařazují všechny obce mezi obce s nezápornou migrací (na str. 107 byl pro tento způsob klasifikace zaveden pojem degenerované řešení – viz poznámka pod čarou 21 na str. 107); nebo
- náhodně a nestabilně pro vhodnou volbu náhodně volených parametrů v průběh trénovacího procesu (opět např. VVP 7 – viz tabulka 10.8 na str. 103).

Opírat se taktéž pouze o trénovací (třeba tak, že celý datový soubor bude použit pro trénování) nebo testovací výkonnost, považuje autor předkládané práce o zavádějící. Trénovací výkonnost vypovídá o schopnosti sítě rozložit klasifikační znalost do

⁴¹Termín *uspokojivý* reprezentuje celkovou výkonnost blížící se výkonnosti sítí, jež diskriminovaly degenerovaně.

Tabulka 10.18: Klasifikační matice – neuronové sítě vytvořené „manuálně“

diskr	Klasifikační matice		
	Řádky: pozorované klasifikace Sloupce: předpovězené klasifikace		
	% správných	Z-migrace	N-migrace
– VVP 5 –			
Z-migrace	51,203	149	142
N-migrace	77,576	74	256
Celkem	65,217	223	398
– PNS 2 –			
Z-migrace	57,571	100	74
N-migrace	72,483	123	324
Celkem	68,277	223	398
– RBF 6 –			
Z-migrace	49,658	145	147
N-migrace	76,292	78	251
Celkem	63,768	223	398

parametrického prostoru sítě – tj. o dostatečné bohatosti architektury sítě a vhodnosti topologie –, testovací výkonnost pak o generalizačních schopnostech sítě – tj. je vhodná pro porovnávání výkonnosti topologií jednotlivých testovaných sítí. Samostatně hodnoceny však nevypovídají tyto výkonnosti o popisované a analyzované reálné situaci.

Zobrazení všech výsledků by bylo neúčelné a nezajímavé. Proto byl od každé architektury vybrán pouze jeden svým způsobem charakteristický zástupce – viz tabulka 10.18.

10.4.7.2 Neuronové sítě s automatickým-heuristickým nastavením trénovacího procesu

U automaticky trénovaných neuronových sítí je, dle soudu předkladatele této práce, celková výkonnost vhodným ukazatelem výkonnosti neuronových sítí. Heuristika prostřednictvím validačního výkonu minimalizuje validační výkon, takže nedochází k přeučení neuronových sítí. Trénovací výkonnost, jejíž maximální hodnota není heuristikou upřednostňována, se v závislosti na počtu vzorů (viz str. 114) zařazených

Tabulka 10.19: Klasifikační matice – neuronové sítě vytvořené heuristikou

diskr	Klasifikační matice		
	Řádky: pozorované klasifikace Sloupce: předpovězené klasifikace		
	% správných	Z-migrace	N-migrace
– VVP H –			
Z-migrace	62,332	136	87
N-migrace	57,286	166	232
Celkem	59,259	223	398
– RBF H –			
Z-migrace	53,929	151	129
N-migrace	78,886	72	269
Celkem	67,663	223	398
– PNS H –			
Z-migrace	0,448	1	222
N-migrace	100,000	0	398
Celkem	64,251	1	620

do trénovací množiny významně projevuje právě na celkové výkonnosti sítě. Celková výkonnost u sítí vytvořených heuristikou programu STATISTICA 7.1 Neuronové sítě vypovídá, na rozdíl od neuronových sítí s manuálním nastavením trénovacího procesu, o popisované a analyzované reálné situaci.

Na základě klasifikačních matic je vidět, že prostřednictvím např. nejlépe automaticky natrénované neuronové sítě RBF H – viz tabulka 10.19 – se podařilo správně klasifikovat téměř 68 % objektů.

10.5 Závěr

Ukázalo se, že neuronové sítě, stejně jako diskriminační analýza, byly povětšinou úspěšnější při klasifikaci obcí s nezáporným saldem migrace (viz tabulka 10.2 na str. 92, tabulka 10.18 na str. 119 a tabulka 10.19 na str. 120). Zatímco klasifikace vícevrstevnými dopřednými sítěmi a RBF sítěmi se pohybovaly kolem podobného procenta úspěšné diskriminace do obou skupin obcí, pravděpodobnostní neuronové sítě zejména ty s vyšším ($\sigma > 0,5$) vyhlazovacím faktorem nebyly schopny správně diskriminovat obce se záporným saldem migrace (podobně jako diskriminační analýza).

Například z tabulky 10.19 je zřejmé, že pravděpodobnostní síť PNS H označila jen jedinou obci, která by měla mít, dle vnitřního nastavení natrénované sítě, záporné saldo migrace. To, ale není nic překvapivého, neboť jak diskriminační analýza, tak pravděpodobnostní neuronové sítě jsou založeny na Bayesovské statistice.

Porovnáme-li dále výsledky diskriminačních výkonů, pak diskriminační analýza používala 5 diskriminátorů, kdežto neuronové sítě 7 diskriminátorů. Celková výkonnost neuronových sítí byla vyšší než výkonnost diskriminační analýzy. Přestože bylo na celé množině dat možno neuronovou síť úplně natrénovat, nelze 100% výkonnost sítě považovat za optimální. Neuronové sítě s takto vysokou výkonností byly příliš vázány na trénovací množinu – byly přeučené (viz výsledky ověřující validační výkonnost pravděpodobnostních neuronových sítí s vyhlazovacím parametrem $\sigma = 0,01$ na str. 107). Výsledky, které byly dosaženy při automatickém trénování, naznačují že i při maximalizaci validačního výkonu bylo možno dosáhnout celkové diskriminační výkonnosti blízké 70% (viz Celk_1 pro RBF 7 v tabulce 10.12 na str. 110)

Pokud byla použita takováto architektura neuronové sítě použita pro trénování na celé množině, dosáhla výkonosti přesahující 73 % (viz Celk_1 opět pro RBF 7). To je cca o 8 % vyšší diskriminační schopnost, než měla použitá diskriminační analýza. S ohledem na výkonnost degenerovaných řešení však nelze s uspokojením tvrdit, že by výsledky byly výborné. Podíl na celkovém relativním neúspěchu diskriminačního procesu lze spatřovat ve výběru, dostupnosti a relevantnosti sledovaných ukazatelů.

Výsledky statistických analýz prokázaly (viz tabulka 10.1 na str. 91), že statisticky významnými diskriminátory do dvou tříd dle znaménka salda migrace jsou poloha obce z hlediska nadmořské výšky, možnost bydlení a zaměstnanecká „absorpce“ obce. Neuronové sítě za nejlepší diskriminátory označovaly míru nezaměstnanosti a význam soukromého zemědělství (viz tabulka 10.7 na str. 102, tabulka 10.9 na str. 104, tabulka 10.11 na str. 109 a tabulka 10.16 na str. 117). Aby však bylo možno z hodnot diskriminátorů spolehlivě predikovat, zda se obce budou či nebudou vylidňovat, je třeba k uvedeným ekonomickým ukazatelům hledat ještě ukazatele další.

Klasifikační pravidla vytvořená v procesu trénování neuronové sítě byla velice nestabilní a měnila se v závislosti na architektuře a dokonce i na topologii sítě. Další ukazatele by jistě pomohly lépe rozložit klasifikační znalost do prostředí neuronové sítě. Analyzované faktory neumožnily dosažení celkově uspokojivějších výsledků.

Příspěvek Čermáková – Biskup [117] nepovažuje citlivostní analýzou pro neuronové sítě za tak podstatný nástroj jakými jsou testy významnosti diskriminátorů u diskriminační analýzy. Ukázalo, že nestabilita pořadí významnosti diskriminátorů získaná pro neuronové sítě souvisí nejen se statistickou významností diskriminačních

schopností diskriminátorů. Souvisí i s motivem trénovacího procesu – dosáhnout maximální výkonu. Nestabilita řešení ovlivněná všemi faktory vstupujícími do procesu trénování neuronových, však svědčí o obecně nedostatečné diskriminační síle zvolených ekonomických ukazatelů.

Jakkoli se dalo předpokládat, že zvolené ukazatele budou pro diskriminaci obcí z pohledu salda migrace (jejich relevantnost byla potvrzena odborníky) dostačující, ukázalo se, že by bylo vhodné zařadit ještě další ukazatele. Rozhodování obyvatel o setrvání, nebo odchodu z obce zřejmě není ovlivňováno pouze ekonomickými ukazateli. Pro obyvatele Jihočeského kraje zřejmě zůstávají důležité i tradiční hodnoty a proto by bylo vhodné zaměřit pozornost také na sociální situaci v obci, kterou ovlivňují mezilidské vztahy, rodinné zázemí a zvyklosti, atp. Nezanedbatelný vliv na rozhodnutí odejít z obce za prací má i systém sociální podpory České republiky. Někteří občané, mající v obci zajištěné zázemí se mohou spokojit s příjmem, jenž ji přináší podpora nezaměstnanosti. Podobně by se daly vysvětlit nedostatky i ostatních ukazatelů, které byly až doposud považovány za stěžejní pro ochotu občanů změnit místo, kde žijí.

Kapitola 11

Expertní systém a neuronové sítě

V rámci aplikačních možností neuronových sítí vytvořil předkladatel disertační práce expertní systém, využitelný v oblasti splátkového prodeje mobilních telefonů. S ohledem na zdroj dat byl expertní systém (dále je označován jako *systém A*) navrhován tak, aby pomohl prodejci minimalizovat možné riziko, které mu hrozí, jestliže spotřebitel úvěr nesplácí. Cílem realizace expertního systému je zdokonalení pravidel tak, aby se minimalizovala ztráta prodejce. Minimalizací rizika se rozumí zdokonalení pravidel pro posuzování bonity¹ zákazníků, pořizujících zboží na splátky (dále jen klienti). V podkapitole 11.5 je uveden poněkud komplexnější přístup, který vytváří expertní systém využitelný pro určování bonity klienta² nikoliv prodejcem, ale samotnou společností poskytující spotřebitelské úvěry (dále jen poskytovatelem).

11.1 Motivace

Obava poskytovatele, že klient nebude splácet svůj závazek, je pochopitelná. V případě nesplácení úvěru přichází o zapůjčenou hotovost a musí podstoupit řadu akcí k zajištění navrácení prostředků. Snaha o vymezení skupiny perspektivních klientů (tj. vyřazení problémových klientů ze současných akceptovatelných klientů) by mohla vést k neúnosnému omezení potenciálních klientů. Jakým způsobem by mohl být ex-

¹Bonita (z pohledu finančnictví) charakterizuje schopnost dlužníka dostát svým finančním závazkům, obecně odráží míru schopností, možností nebo kvalit.

²Vzhledem k tomu, že se jedná o stejné osoby jako při tvorbě systému A, budou označováni stejně – klienti. Tedy ať je na klienta nahlíženo jako na zákazníka firmy prodávající mobilní telefony, nebo na potenciálního zákazníka společnosti poskytující spotřebitelské úvěry, stále se jedná o tu samou osobu. Klientem je navíc označován jak žadatel, tak následně i spotřebitel, kterému již byl mobilní telefon na úvěr prodán.

pertní systém využít tak, aby k zužování skupiny potenciálních klientů nedocházelo, naznačuje podkapitola 11.5.2.

Je v zájmu prodejce „předvídat“ chování klientů a z toho důvodu odhadovat/určovat jejich bonitu? V jeho zájmu to pochopitelně je. Vzhledem k tomu, že prodejce může doporučit/nedoporučit každého jednotlivého potenciálního klienta na základě „subjektivního“ dojmu, ovlivňuje i on procento identifikovaných neplatičů³. Prodejce je zákazníkem-partnerem poskytovatele a ztrácí kredit, pokud poskytovatele připravuje o „bezproblémový“ zisk, měl by se proto pokusit bonitu klienta „odhadnout“. Expertní systém by mohl svým způsobem nahradit odhad/dojem prodávajícího exaktní znalostí, která by v případě neuronové sítě mohla být „extrahována“ přímo z dat.

Bonita klienta je odhadována předem, v závislosti na typu závazku. U klienta se hodnotí zejména:

- jeho finanční situace (finanční stabilita, likvidita, rentabilita, míra zadluženosti atd.),
- jeho majetková situace,
- jeho očekávaná důchodová situace, kvalita podnikatelského záměru, reálnost splatit poskytnuté úvěry,
- technická a obchodní zdatnost, postavení na trhu, kvalita managementu,
- zkušenosti poskytovatele z dřívějšího obchodního spojení s klientem.

Pro tvorbu expertního systému bez experta je nutná trénovací množina (viz podkapitola 8.3.3). Při zjišťování bonity klienta však nelze provést věrohodné šetření, na jehož základě by taková trénovací množina vznikla. Je těžké si představit, že

- klient vyplní pravdivě do dotazníku potřebné údaje o tom, zda je ochoten respektive neochoten dodržet podmínky plynoucí ze závazku, který s poskytovatelem uzavírá,
- poskytovatel uzavře smlouvy s každým klientem, a poté na základě výsledků splácení či nesplácení úvěru vytvoří trénovací množinu nutnou pro trénování expertního systému.

Trénování neuronových sítí proto probíhá na základě informací o klientech, kterým již byl spotřebitelský úvěr přidělen.

Soubor dat tvoří vektory obsahující skupinu vstupních faktorů a informaci o „platební morálce“ každého klienta, který získal spotřebitelský úvěr – více podkapitola 11.2.

³V tomto okamžiku pochopitelně potencionálních neplatičů.

11.2 Soubor dat

Soubor dat analyzovaný v kapitole 11 pochází z prostředí splátkového prodeje mobilních telefonů. Data byla poskytnuta nejmenovanou firmou (provozovny: Klatovy, Prachatice a Strakonice) a jedná se o informace o klientech, kteří získali spotřebitelský úvěr od poskytovatele – jedné z předních firem spotřebitelského financování v České republice.

Po konzultaci s prodejci byly jako validní apriorní faktory ovlivňující „vhodnost“ klienta z pohledu bezproblémového dodržení závazku splatit úvěr navrženy:

- celkový měsíční příjem klienta v českých korunách (spojitá proměnná);
- cena výrobku v českých korunách (spojitá proměnná);
- počet splátek (diskrétní proměnná);
- informace o splacení předchozích závazků – tj. měl-li klient při splacení předchozích závazků problémy, či nikoli (alternativní případně ordinální proměnná);
- počet a výše souběžně běžících spotřebitelských úvěrů – míra finanční zátěže klienta vzniklá splácením úvěrů.

Bohužel ne všechny informace bylo možno získat. Soubor dat byl sestaven z údajů, jež o klientovi zjišťuje poskytovatel⁴. Zjišťování dalších údajů nad rámec údajů uvedených v žádosti o spotřebitelský úvěr (jakkoliv by byl jejich přínos pro určení bonity relevantní) nebylo možno provést z následujících důvodů:

- vyplňování delších dotazníků by nadměrně zpomalovalo prodej a v případě dobrovolnosti by snižovalo rozsah souboru⁵;
- byla obava ze sankcí prodejce firmou spotřebitelského financování, která se na výzkumu nebyla ochotna podílet⁶.

⁴Zákony České republiky poměrně striktně omezují informace, jež mohou poskytovatelé vyžadovat a jak s nimi mohou nakládat. Právní předpisy zabývající se nejen ochranou osobních údajů, ale i samotným poskytováním spotřebitelského úvěru uvádí následující výčet:

- Zákon č. 513/1991 Sb. obchodní zákoník [107] – § 497–507 a jeho pozdější změny – zejména [104] a [99] (celkově bylo změn 20)
- Zákon č. 321/2001 Sb. o některých podmínkách sjednávání spotřebitelského úvěru [103]
- Zákon č. 21/1992 Sb. o bankách [98]
- Zákon č. 101/2000 Sb. o ochraně osobních údajů a o změně některých zákonů [100] a jeho pozdější změny [101], [106], [102] a [105]
- a další

⁵Navíc se dá předpokládat, že by tento vzorek nebyl reprezentativním vzorkem klientů.

⁶Z téhož samého důvodu si poskytovatel nedovolil poskytnout ani informace o klientech, jimž nebyl spotřebitelský úvěr přidělen a jejichž vyplněné žádosti ani nearchivoval. Využití těchto dat

V průběhu šetření došlo i ke změně metodiky zaznamenávání údajů, a proto data tvořila nehomogenní skupinu záznamů, kterou bylo nutno revidovat. Informace o měsíci a provozovně, ve které byla smlouva uzavřena, *věku*, *pohlaví*, rodinném stavu, vzdělání, *zaměstnání*, úrovni *měsíčního čistého příjmu* a výši spotřebitelského úvěru společně s typem pojištění tohoto úvěru a počtem dohodnutých splátek byly zjišťovány pouze u části klientů. U zbytku klientů byly zjištěny pouze informace o měsíci a provozovně, ve které byla smlouva uzavřena, *věku* a *pohlaví* spotřebitele, *zařazení do jedné z předem definovaných kategorií* – podnikatel, zaměstnanec s pracovní smlouvou na dobu neurčitou a důchodce, úrovni *měsíčního čistého příjmu*, o výši spotřebitelského úvěru společně s typem pojištění tohoto úvěru a počtem dohodnutých splátek.

Spotřebitelské úvěry (731 schválených úvěrů s úplnými daty) byly sjednávány v druhé polovině roku 2004. Výčet faktorů použitých pro určování bonity klienta je uveden níže:

1. **Věk spotřebitele (roky)** (vek). Lze předpokládat, že věk ovlivňuje bonitu klienta. Klienti z různých věkových kategorií mají odlišný přístup k plnění závazků.
2. **Pohlaví spotřebitele** (pohlaví). I když předjímat odlišnost v chování mužů a žen se na první pohled nejeví jako správné, ze zkušenosti je známo, že se muži a ženy svým přístupem k plnění závazků liší.
3. **Čistý měsíční příjem (Kč)** (prijem). Výše příjmu sama o sobě sice není determinujícím faktorem ovlivňujícím ochotu klienta splácet úvěr. Lze však uvažovat o tom, že klienti s nižším příjmem se mohou snáze dostat do finančních obtíží.
4. **Výše spotřebitelského úvěru (Kč)** (uver). Výše spotřebitelského úvěru společně s počtem dohodnutých splátek přímo vypovídá o zátěži, kterou klientovi splácení úvěru měsíčně přináší.
5. **Počet dohodnutých splátek** (splátek). Počet dohodnutých splátek zejména ve vzájemné interakci s výší spotřebitelského úvěru vypovídá o výši disponibilních prostředků klienta⁷.
6. **Typ, respektive absence, pojištění** (pojisteni). Pojištění je obvykle poskytovatelem vyžadováno. Je vyjádřením jeho nedůvěry/nejistoty. Pojištění navíc „dobrovolně“ uzavírají klienti, kteří si uvědomují možné problémy při splácení úvěru. Pojištění samozřejmě spotřebitelský úvěr prodražuje.

S ohledem na počet splátek (10, 12, 20, 22, 30, 33 a 50) bylo samozřejmě nutno

předpokládá systém B popsáný v podkapitole 11.5.

⁷Vzhledem k tomu, že poskytovatel často navyšuje počet splátek, zařazení tohoto faktoru „odhaluje“ také splátkovou politiku poskytovatele.

počkat do doby, než bylo možno platební morálku klientů vyhodnotit. Jak známo, politika poskytovatelů je poměrně nekompromisní a nesplacení byt' jediné koruny, činí z klienta „černou ovci“. Přitom mnoho nedopatření generující označení „neplatič“ by se dalo připsat spíše na vrub poskytovateli, který například špatně vypočetl splátku při předčasném splacení úvěru apod. Informace o platební morálce získané od poskytovatele bylo nutno revidovat. Klienti byli na základě výsledků splácení rozděleni do tří skupin:

- **Neplatič** – klient, který buďto vůbec úvěr splácet nezačal, nebo velice záhy se splácením přestal – jeho dluh vůči poskytovateli zůstal jak relativně, tak absolutně, velký (92 klientů).
- **Problémový platič** – klient, u kterého se sice vyskytly při splácení problémy, avšak jeho platební morálka by se dala považovat za dobrou (15 klientů).
- **Platič** – klient, který bez větších problémů (viz poznámky výše) splatil celý spotřebitelský úvěr (624 klientů).

Situaci napříč souborem dat popisuje tabulka 11.1. V tabulce jsou uvedeny podmíněné průměry respektive podmíněné relativní četnosti sledovaných faktorů v závislosti na příslušnosti k jedné ze tří skupin dle výsledků splácení. Tabulka 11.1 i předchozí výčet ukazují, že příslušnost klientů do jednotlivých skupin je značně nerovnoměrná – počet platičů značně převyšuje počty v ostatních skupinách. Tato situace se ovšem dá očekávat, neboť klienti byli označeni za bonitní expertním systémem poskytovatele a jeví se pro něj býti bezproblémovými klienty.

Tabulka 11.1: Průměry a relativní četnosti ve skupinách rozdělení na platiče, problémové platiče a neplatiče

diskr	Průměry				Relativní četnosti					
	věk	příjem	úvěr	splátek	pohlaví		pojištění			N
					muž	žena	A	B	NE	
Platiči	35,5	11 371	7 320	19,5	0,566	0,434	0,737	0,013	0,25	624
Problémoví	32,7	10 320	9 586	22,7	0,667	0,333	0,733	0,067	0,2	15
Neplatiči	30,9	9 885	7 890	20,5	0,576	0,424	0,826	0,033	0,141	92
Všichni	34,8	11 163	7 439	19,7	0,569	0,431	0,748	0,016	0,235	731

Na základě věcné analýzy a analýzy vzdálenosti mezi centroidy jednotlivých skupin dle výsledků splácení bylo provedeno zjednodušení situace. Na základě Mahalanobiosovy vzdálenosti a ní založeném testu nelze tvrdit, že centroid skupiny problémových platičů je signifikantně odlišný od centroidu neplatičů (viz tabulka 11.2), byla

skupina problémových platičů spojena se skupinou neplatičů. S ohledem na cíl tvorby expertního systému A, tj. pomoci prodejci pro odhalení klientů, které poskytovatel považuje za neplatiče, bylo toto spojení logické. Nadále bylo tedy segmentováno na „Platiče“ a „Neplatiče“ – novou situaci napříč souborem dat popisuje tabulka 11.3. Spojením se alespoň částečně snížila disproporce četností jednotlivých skupin, přestože i nadále tato disproporce působí problémy – viz klasifikace prostřednictvím pravděpodobnostních neuronových sítí.

Pouhá „naivní“ klasifikace využívající značného nepoměru mezi zastoupením platičů a neplatičů dosahuje 85,4% klasifikační výkonnosti⁸. Výkonnosti sítí byly tedy značně kriticky porovnávány s touto vysokou klasifikační výkonností⁹ a pozornost byla věnována také způsobu klasifikace. Pokud sítě klasifikovaly „naivně“, byl jejich způsob práce, stejně jako v kapitole 10, označen za degenerovaný.

Tabulka 11.2: Testy významnosti Mahalanobisových vzdáleností

(Zdroj: Čermáková – Biskup [117])

Třída	Testy významnosti pro Mahalanobisovy vzdálenosti 2					
	F testy s a 722, stup. volnosti					
	Sigma-omezená parametrizace					
	Platiči F	Platiči p	Problémoví F	Problémoví p	Neplatiči F	Neplatiči p
Platiči			2,025	0,0497	4,055	0,0002
Problémoví	2,025	0,0497			1,112	0,3532
Neplatiči	4,055	0,0002	1,112	0,3532		

V souboru dat je každý klient identifikován 7 faktory, z nichž 6 lze považovat za faktory ovlivňující bonitu. Sedmý faktor indikuje do jedné z uvažovaných bonitních skupin – Platič/Neplatič. Soubor dat je uveden v tabulce A.2 v příloze A.2.¹⁰

⁸Pokud neuronová síť klasifikovala klienty tímto způsobem, byl její způsob práce označen jako degenerovaný.

⁹Pro lepší orientaci je dosažení výkonnosti sítě diskriminující degenerovaně zvýrazněno **fialově**, dosažení vyšší výkonnosti **tučně**.

¹⁰Z dat byly odstraněny všechny identifikační prvky, které by mohly klienty, prodejce a poskytovatele nějakým způsobem identifikovat. S daty tedy bylo pracováno čistě jako s hodnotami již bez zpětné vazby k jednotlivým klientům a konkrétním spotřebitelským smlouvám.

Tabulka 11.3: Průměry a relativní četnosti ve skupinách rozdělení na platiče a neplatiče

diskr	Průměry				Relativní četnosti					
	věk	příjem	úvěr	splátek	pohlaví		pojištění			N
					muž	žena	A	B	NE	
Platiči	35,5	11 371	7 320	19,5	0,565	0,435	0,731	0,012	0,257	624
Neplatiči	31,1	9 946	7 128	20,8	0,589	0,411	0,813	0,037	0,15	107
Všichni	34,8	11 163	7 439	19,7	0,569	0,431	0,748	0,016	0,235	731

11.2.1 Ekonomická situace

Následující hodnoty představují ekonomickou situaci z pohledu poskytovatele a prezentují finanční zainteresovanost poskytovatele i prodejce na vhodně pracujícím expertním systému. Za sledovaných šest měsíců měli klienti, jimž byl schválen spotřebitelský úvěr na nákup (spolufinacování nákupu) mobilního telefonu, celkově zaplatit 5 437 553 Kč.¹¹ Přitom poskytovatel bezprostředně po ukončení¹² smluvního vztahu nevybral 783 259 Kč. Avšak i klienti, kteří měli při splácení problémy (vznikla jim jakkoliv malá pohledávka vůči poskytovateli), zaplatili při splátkách 459 604 Kč. Poskytovatel „bez problémů“ vybral 85,6 % celkové částky, což odpovídá 85,4 % bezproblémových klientů¹³.

11.3 Diskriminace prostřednictvím neuronových sítí

Stejně tak jako v kapitole 10, je i zde neuronová síť použita jako nástroj diskriminace.

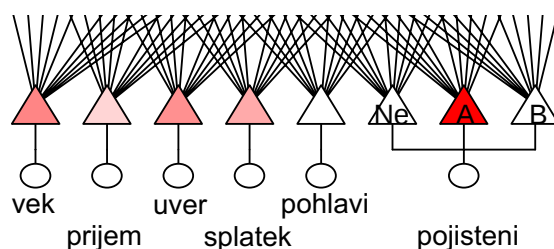
Vstupní vrstva neuronové sítě použité v podkapitole 10.4 (str. 94) nemá stejný počet neuronů jako je počet faktorů. Důvodem jsou kategoriální vstupy (pohlaví a pojištění). Úrovně proměnné „pohlaví“ může přenést jeden neuron prostřednictvím

¹¹Celkovou sumu za prodej telefonu (včetně mobilních služeb), kterou díky splátkovému prodeji získal prodejce, nebylo možno zpětně zjistit. Navíc prodejce spolupracoval (minoritně) i s dalšími poskytovateli spotřebitelských úvěrů). Dle navrhované koncepce expertního systému využitelného pro prodejce (systém A) byl tento systém vytvářen k prosté minimalizaci počtu neplatičů. Celková výše ztráty by tedy byla zajímavější právě pro poskytovatele.

¹²Ještě před zahájením opravných opatření – upomínkami počínaje a návrhem na exekuci konče.

¹³Klienti považovaní v dalším za neplatiče nezaplatili 764 301 Kč a splatili pouze 266 853 Kč.

binárního vstupu. Vstup úrovně proměnné „pojištění“ do sítě musí řešit trojice neuronů¹⁴. Přitom každý z neuronů odpovídá jedné z možných hodnotúrovni kategoriální vstupní proměnné. Každý ze vstupních neuronů, odpovídající úrovni kategoriální proměnné, dostává na svůj vstup jednotkový signál, zatímco ostatní neurony mají vstup nulový. Situace je dobře patrná na obrázku 11.1, jenž symbolicky zobrazuje vstupy (kroužky) a neurony vstupní vrstvy (trojúhelníčky).



Obrázek 11.1: Schéma vstupu neuronové sítě 6:8-...
(Vytvořeno v programu STATISTICA Neuronové sítě 7.1 – upraveno)

První 4 vstupy zleva odpovídají numerickým vstupům (věk, příjem, úvěr a splátek). Hodnoty těchto vstupů jsou přímo přenášeny na vstupní neurony. Poslední dva vstupy (zleva) odpovídají kategoriálním vstupům (pohlaví a pojištění). Na vstup určující pohlaví navazuje pouze jeden neuron, zatímco na vstup přinášející informace o pojištění navazují tři neurony. Obrázek je doplněn o označení toho, který neuron je aktivován při konkrétní úrovni faktoru „pojištění“. Barevné zvýraznění informuje o míře aktivace daného neuronu. Bílá barva reprezentuje nulovou aktivaci, sytě červená barva pak maximální (kladnou) aktivaci. Intenzita odstínu je úměrná míře aktivace¹⁵. Obrázek 11.1 zobrazuje klienta-muže (0 = muž, 1 = žena) – nulová aktivace pátého vstupního neuronu zleva, s pojištěním typu A – nulová aktivace šestého a osmého vstupního neuronu a maximální – jednotková – aktivace sedmého vstupního neuronu.

Navázání ostatních částí sítě (s ohledem na architekturu) je totožné jako v případě, kdy jsou vstupy striktně numerické, jak tomu bylo například v podkapitole 10.4.

¹⁴Obecně n neuronů pro n různých úrovní, jichž může kategoriální – „nenumernický“ – vstup nabývat.

¹⁵Sytě zelená reprezentuje maximální zápornou aktivaci. V případě vstupů této úlohy samozřejmě záporné hodnoty nepřicházejí v úvahu. To ale neplatí o šíření signálu v síti, jak je vidět např. již na obrázku 4.5 na str. 45.

Označení výkonností sítí bude stejné jak bylo použito a zavedeno v kapitole 10 na str. 93. Protože v kapitole 11 je důraz je kladen na vytvoření co nejlepšího expertního systému klasifikujícího platiče a neplatiče, nebude proces trénování popsán tak podrobně, jako v kapitole 10.

11.3.1 Trénování sítí na celém souboru dat

Výsledky uvedené v podkapitole 11.3.1 byly získány prostřednictvím průvodce řešením programu STATISTICA 7.1 Neuronové sítě a následnou optimalizací manuálního nastavení parametrů některých sítí. Trénování probíhalo na souboru dat všech klientech. Průvodce řešením testoval 1000 neuronových sítí a vybral 20 nejlepších z nich. Nejlepších výkonností bylo dosaženo prostřednictvím pravděpodobnostních ($\text{Celk_2} < 0,854$) a RBF sítí ($\text{Celk_2} \leq 1$) – výsledky jsou komentovány níže.

11.3.1.1 Pravděpodobnostní neuronová síť

Prakticky všechny pravděpodobnostní sítě vybrané heuristikou programu STATISTICA 7.1 Neuronové sítě diskriminovaly degenerovaně. Vzhledem k tomu, že proces trénování pravděpodobnostní neuronové sítě je ovlivněn pouze jedním parametrem, lze poměrně jednoznačně sledovat jeho vliv na výslednou výkonnost sítě a způsob klasifikace.

Podobně jako v kapitole 10 bylo možno natrénovat pravděpodobnostní neuronovou síť na datech všech 731 klientů tak, aby dosáhla 100% výkonnosti. Té bylo dosahováno pro vyhlazovací parametr $\sigma \leq 0,01$. Zvýšení vyhlazovacího faktoru na hodnotu 0,02 vedlo ke snížení tréninkové a tedy celkové výkonnosti sítě (Celk_2) na hodnotu **0,992**, kdy 5 neplatičů bylo chybně zařazeno mezi platiče a 1 platič chybně zařazen mezi neplatiče. Stabilní výkonnosti $\text{Celk_2} = 0,854$ bylo dosaženo pro vyhlazovací parametr $\sigma \geq 0,29$. Jedná se o výborný výsledek? Nikoli. Výkonnosti bylo dosaženo tím, že všichni klienti byly označeny jako platiči.

Hodnoty vyhlazovacích parametrů zmíněné v předchozím odstavci vymezují jeho meze, tj. $\sigma \in \langle 0,01; 0,29 \rangle$. Mimo tyto meze nemá smysl vyhlazovací parametr volit. Pro hodnoty $\sigma \leq 0,01$ bude pravděpodobnostní neuronová síť přetrénovaná a pro hodnoty $\sigma \geq 0,29$ bude diskriminovat degenerovaně. Například pro vyhlazovací parametr $\sigma = 0,045$ zařadila síť cca polovinu neplatičů mezi platiče (53 : 54) a v případě platičů neudělala ani jednu chybu. Pokud by takto dokázala pracovat neuronová síť i pro neznámé klienty, dokázala by snížit ztrátu na polovinu. Jakou finanční úsporu by to mohlo přinést shrnuje podkapitola 11.2.1.

Výsledky citlivostí analýzy ukázaly, že bez ohledu na hodnotu vyhlazovacího faktoru voleného z intervalu $\langle 0,01; 0,29 \rangle$, byly jako nejvlivnější faktory označeny pojistní a věk. Faktory příjem a úvěr byly označeny za méně významné. Čím více byla neuronová síť fixována na tréninkovou množinu, tím více prostřednictvím podílu citlivosti trvala na zařazení všech faktorů. Pro vyhlazovací parametr $\sigma = 0,01$, tj. nejvyšší hodnotu vyhlazovacího parametru, pro kterou bylo možno dosáhnout 100% výkonnost, měl nejnižší podíl citlivosti faktor příjem (6,742). Vyřazení tohoto faktoru vedlo ke snížení celkové výkonnosti sítě – Celk.2 na hodnotu **0,99** (6 chybně zařazených neplatičů a 1 chybně zařazený platič). Vyřazení faktoru věk (dle citlivostní analýzy velmi vlivný) vedlo ještě k vyššímu snížení celkové výkonnosti sítě – Celk.2 na hodnotu **0,964** (26 chybně zařazených neplatičů).

Vyřazení obou kategoriálních vstupů (pojištění a pohlaví), které byly označeny za vlivné, výkonnost sítě pro vyhlazovací parametr $\sigma = 0,01$ paradoxně nesnížilo a byla tedy zachována 100% úspěšnost diskriminace i s pouhými čtyřmi diskriminátory. Vzhledem k tomu, že síť trénovaná s takto nízkou hodnotou vyhlazovacího faktoru byla nutně přetrénovaná, bylo vhodné nahlížet na ni spíše jako na databázi údajů než jako na znalostní pravidla. Proto právě faktory s vyšším počtem různých hodnot/úrovní umožňovaly lepší klasifikaci, byť byly prostřednictvím analýzy citlivosti označeny za méně významné než faktory kategoriální – tj. faktory s malým počtem různých úrovní. Postupným odstraňováním jednotlivých faktorů až po jediný faktor – věk – se celková výkonnost sítě snižovala až na 85,4 %. Bohužel práce této sítě byla degenerovaná. Ani snižování vyhlazovacího faktoru σ až na hodnotu 0,000001 nepřineslo významné zlepšení výkonnosti. Pravděpodobnostní neuronová síť s jedním vstupem – věkem – ($\sigma < 0,01$) správně označila pouze jednoho neplatiče a všechny ostatní neplatiče zařadila mezi platiče.

Byl-li zařazen faktor s vyšším počtem různých hodnot než měl faktor věk, tj. byli ignorován výsledek automatického odstraňování vstupních proměnných a výsledek citlivostní analýzy, pracovala neuronová síť s vyšší výkonností. Tuto podmínku splňoval faktor příjem (původně nejméně preferovaný vstupní faktor). Pravděpodobnostní neuronová síť s jedním vstupem – příjem – a hodnotou vyhlazovacího faktoru $\sigma = 0,000001$ dokázala oproti degenerovanému řešení správně klasifikovat 16 neplatičů a 622 platičů, přičemž chybně zařadila 2 platiče, a jak bylo možno očekávat, nejvíce chybovala při zařazení neplatičů (91 neplatičů bylo chybně zařazeno mezi platiče). Výkonnost sítě tak vzrostla na 87,3 % – tj. o necelá 2% oproti řešení degenerovanému. Nelze však předpokládat, že zlepšení výkonnosti sítě bude významné i mimo trénovací množinu.

Na základě řady provedených manuálně konstruovaných simulací je možno se domnívat, že přetrénovaná síť pracuje tak, že nevytváří znalostní pravidla ale databázi údajů (viz výsledky v předchozí odstavce). Citlivostní analýza není založena na motivu dosažení maximální výkonnosti sítě a její doporučení nevedou nutně ke zvýšení výkonnosti sítě, zvláště pokud se jedná o síť přetrénovanou.

11.3.1.2 RBF síť

RBF síť vybrané heuristikou programu STATISTICA 7.1 Neuronové sítě použily při svém trénování pro lokalizaci semilokálních jednotek náhodný výběr z trénovacích vzorů, odhad rozptylu jako parametru semilokálních jednotek byl proveden prostřednictvím metody „k-nejbližších sousedů“ a optimalizace lineárního výstupu sítě prostřednictvím pseudoinverze. Maximální výkon byl dosažen při zařazení maximálního počtu semilokálních jednotek, tj. 731 jednotek. 100% natrénovanosti bylo dosaženo i pro síť se 730 jednotkami, nebo pro síť, která nezahrnula kategoriální vstupní faktory. Při snížení počtu semilokálních jednotek na polovinu, klesl celkový výkon RBF sítě na cca 85 %. Bylo dosaženo klasifikační výkonnosti degenerovaného řešení. RBF síť však neklasifikovala způsobem, který byl za degenerovaný označen. Oproti pravděpodobnostním sítím RBF síť chybovala shodně pro platiče i neplatiče¹⁶.

Analýza citlivosti přiřadila všem faktorům přibližně stejný podíl citlivosti, jenž klesal s celkovou výkonností sítě (Celk.2) podobně jako u pravděpodobnostní sítě. Daleko významnější zhoršení výkonnosti než u pravděpodobnostních neuronových sítí způsobilo vyřazení spojitých faktorů. I při maximálním počtu semilokálních jednotek klesla výkonnost sítě na 78,3 % při vyloučení faktoru příjem a na 58 % při vyloučení faktoru věk. I takto konstruované RBF sítě chybovaly stejně při klasifikaci platičů a neplatičů.

11.3.1.3 Vícevrstvá dopředná síť

Sítě navržené heuristikou (vícevrstvé dopředné sítě s maximálně dvěmi skrytými vrstvami po 30-ti neuronech) nebyly schopny klasifikovat s výkonností vyšší než 72 %.

¹⁶Stejně procento chybně klasifikovaných bylo zjištěno pro libovolné nastavení počtu semilokálních jednotek. Minimální celkové výkonnosti Celk.2 (stále vyšší než 50%) bylo dosahováno i pro pouhé dvě semilokální jednotky. Prováděné experimenty probíhaly s hodnotou rozptylu vypočtenou prostřednictvím metody k-nejbližších sousedů s parametrem 1, tj. rozptyl každé semilokální jednotky byl nastaven jako vzdálenost k nejbližšímu sousedovi, tak jak navrhuje Šíma – Neruda [86]. Lokalizace jednotek probíhala metodou k-průměrů.

Zlepšení výkonnosti přineslo až dodatečné zařazení třetí skryté vrstvy. Tu heuristika programu STATISTICA 7.1 Neuronové sítě zařadit nedokáže. Celkové výkonnosti sítě (Celk_2) přesahující 90 % bylo dosaženo po dlouhotrvajícím tréninku metodou sdružených gradientů pro dostatečně bohatou síť (30 neuronů v každé ze skrytých vrstev). Lze tedy předpokládat, že i zde bylo dosaženo vysoké trénovací výkonnosti na úkor generalizace. Bohatá dopředná neuronová síť (bohatá na počet skrytých vrstev a počet neuronů) do svého parametrického prostoru kódovala spíše klasifikační pravidla než datovou množinu. Tak, jak zavádí pojmy informace a znalost kapitola 8, reprezentovala vícevrstvá dopředná síť spíše pouhou informaci o platících a neplatících na poli splátkového prodeje mobilních telefonů ve vztahu k vybraným faktorům, než znalost o chování klientů na základě zvolených faktorů¹⁷.

11.3.2 Trénování s evaluací generalizace

Zhodnocení výkonnosti jednotlivých typů sítí bylo provedeno prostřednictvím cross-validace, opakovaného náhodného výběru a bootstrapu. Bez ohledu na způsob ověření bylo dosaženo podobných výsledků. Tréninkové, validační, testovací a celkové (Celk_1) výkonnosti však nejsou pouhými průměry jednotlivých výkonností dosažených pro trénování na jednotlivých výběrech.

Tabulka 11.4 uvádí výkonnosti kolekce sítí¹⁸ vytvořené z topologicky identických neuronových sítí trénovaných na dvaceti různých podmnožinách celého souboru dat. Stejně tak, pokud není uvedeno jinak, jsou všechny ostatní výsledky (např. procenta špatně klasifikovaných klientů) odhadnuty na základě opakovaného trénování. Rozdělení vzorů-klientů do jednotlivých podmnožin¹⁹ (trénovací, validační a testovací) bylo provedeno programem STATISTICA 7.1 Neuronové sítě. Přednastavené poměry pro zařazení do jednotlivých podmnožin se shodovaly s doporučeními, která lze získat při studiu literatury zmíněné v poznámce pod čarou číslo 14.

Evaluace výkonnosti pravděpodobnostních neuronových sítí ukázala, že mnohdy vysoký trénovací výkon nezůstane zachován pro výkon validační a testovací. Způsob klasifikace pro vytípané vyhlazovací faktory zmíněné v podkapitole 11.3.1 shrnuje

¹⁷K podobným závěrům vedou i výsledky celé podkapitoly 11.3.1.

¹⁸Kolekce sítí umožňují diskriminaci, která je založena na znalosti rozložené ve všech sítích v kolekci obsažených, je tedy funkčním diskriminačním nástrojem. Kolekce sítí byly popsány v podkapitole 4.3 a s ohledem na práci kolekcí sítí v případě klasifikace, byly někteří klienti označeni za nezařaditelné – tj. za platíce i neplatíce je považoval stejný počet sítí z vytvořené kolekce.

¹⁹Matematicky by bylo korektnější použití slova podskupina namísto podmnožina, neboť bootstrap je založen na výběru s opakováním.

Tabulka 11.4: Výkonnost trénování neuronových sítí vytipovaných parametrů

Síť	Výkon			
	Trén.	Valid.	Test.	Celk.1
PNS ($\sigma = 0,01$)	1	0,777	0,805	0,953
PNS ($\sigma = 0,045$)	0,974	0,801	0,817	0,93
PNS ($\sigma = 0,29$)	0,854	0,852	0,862	0,863
RBF ₃₆₇	0,934	0,522	0,505	0,891
RBF ₁₈₀	0,846	0,649	0,667	0,84

následují výčet.

- Pravděpodobnostní neuronová síť, která dosáhla pro $\sigma = 0,01$ na trénovací množině 100% výkonnost, chybovala maximálně u 15 % platičů a u 80 % neplatičů nezařazených do procesu trénování (tj. klientů zařazených do validační a testovací množiny).
- Dle očekávání byla výkonnost pravděpodobnostní neuronové sítě se $\sigma = 0,045$, při tréninku na datech všech klientů (viz podkapitola 11.3.1, str. 131) nadhodnoceným odhadem celkové výkonnosti. Optimistických výsledků na trénovací množině, kde se síť mýlila maximálně u 0,32 % platičů (maximálně jeden špatně zařazený platič) a u 24,53 % neplatičů, nebylo podle očekávání dosaženo na validační a testovací množině. Pro neznámé klienty sice síť chybovala u 13,84 % platičů, ale vždy se mýlila ve více než 65 % neplatičů. Dokonce ani celková výkonnost nedosáhla výkonnosti sítě klasifikující degenerovaně (viz tabulka 11.4).
- Pravděpodobnostní neuronové síť se $\sigma > 0,29$ klasifikovaly, až na drobné výjimky, na všech 3 množinách degenerovaně.

Vypouštění vstupních faktorů vedlo k potvrzení domněnky vyslovené v podkapitole 11.3.1, tj. domněnky, že neuronové sítě spíše než znalost, zprostředkovávají pouhou informaci o situaci:

- Pro $\sigma = 0,01$ nezměnilo vyřazení kategoriálních faktorů (pohlaví a pojisteni) příliš ani výkonnost sítě ani způsob klasifikace sítí. Avšak zvýšení vyhlazovacího faktoru na hodnotu $\sigma = 0,045$ vedlo k degenerované klasifikaci prostřednictvím pravděpodobnostní neuronové sítě opírající se pouze o spojité faktory věk, příjem, úvěr a splátek; 91,59 % neplatičů bylo zařazeno chybně.
- Vyřazení nejvlivnějšího kvantitativního faktoru – věku – sice snížilo trénovací výkonnost sítě pro zvolené hodnoty σ , avšak validační a testovací výkonnosti zů-

staly přibližně stejné. Procento chybně klasifikovaných klientů celkově vzrostlo. Procenta nesprávných klasifikací klientů nezařazených do trénování dosahovala v případě neplatičů podobných hodnot jako při vyřazení kategoriálních znaků.

Evaluace RBF sítí proběhla pro sítě, které obsahovaly buď maximální možný počet semilokálních jednotek nebo poloviční počet z maximálního počtu jednotek. Vzhledem k tomu, že RBF sítě neměly tendenci klasifikovat degenerovaně, byla jejich validační a trénovací výkonnost daleko nižší než v případě pravděpodobnostních neuronových sítí. Výsledky jsou uvedeny v tabulce 11.4, na str. 135.

- Při tréninku RBF sítě, jejíž počet semilokálních jednotek odpovídal počtu trénovacích vzorů – RBF 6:8-367-1:1 – se procento správně klasifikovaných klientů zařazených do procesu trénování se nelišilo v závislosti na tom, zda se jednalo o platiče, či neplatiče. Na testovací množině již procento správně zařazených platičů a neplatičů shodné nebylo. Rozdíl v procentu správně klasifikovaných klientů nezařazených do procesu trénování byl maximálně 20%, v závislosti na tom, zda se jednalo o platiče, či neplatiče. Jak však uvádí tabulka 11.4, RBF sítě neklasifikovaly neznámé klienty degenerovaně.
- Výsledky jednodušší neuronové sítě RBF 6:8-180-1:1, ukazují, že zařazení maximálního počtu semilokálních jednotek způsobilo přeúčtení sítě, neboť validační a testovací výkonnosti této jednodušší sítě jsou cca o 10 % vyšší než u sítě RBF 6:8-367-1:1. RBF síť se 180 semilokálními jednotkami klasifikovala spolehlivěji platiče než neplatiče.

Hodnotit kvalitu RBF sítí s ohledem na možné vypouštění vstupních faktorů se vzhledem k jejich slabé validační a testovací výkonnosti nejevilo účelné. Právě nízká diskriminační schopnost používaných faktorů způsobila, že klasifikace prostřednictvím RBF sítí, jež jsou pro řešení tohoto problému obvykle nejspolehlivější, byla horší než klasifikace prostřednictvím pravděpodobnostních neuronových sítí.

11.4 Závěr

Je nutno přiznat, že na základě vybraných faktorů nedokázaly neuronové sítě zlepšit pravidla poskytovatele pro spolehlivější stanovení bonity klienta. Pro vhodné nastavení parametrů sítí (vyhlazovací faktor, počet semilokálních jednotek a jejich parametrů) sice dokázaly reprezentovat data o klientech, avšak hodnocení validačního a testovacího výkonu vedlo k závěru, že takto natrénované sítě nereprezentují znalost, která by na základě zvolených faktorů umožňovala obecné hodnocení bonity klienta,

ale pouze a jen data o klientech.

Jako nejpoužitelnější (nikoli z pohledu nejvyšší výkonnosti!) se jevila pravděpodobnostní neuronová síť, jejíž tréninkový proces byl specifikován volbou vyhlazovacího faktoru $\sigma = 0,045$ (dále jen $PNS_{0,045}$). Tato síť byla schopna dosáhnout vysokého trénovacího výkonu a zároveň zachovala i vysoký validační a trénovací výkon. Neuronová síť $PNS_{0,045}$ nediskriminovala degenerovaně. Nebyla sice schopna dosáhnout 85,4% správné klasifikace pro neznámé klienty jako síť pracující degenrovaně, ale byla schopna správně zařadit 86,16 % jí neznámých platičů a mýlila ve více než 65 % případech jí neznámých neplatičů.

Expertní systém, který by se měl rozhodovat jen na základě faktorů věk, příjem, úvěr, splátek, pohlaví a pojištění – viz podkapitola 11.2, nemá smysl vytvářet, neboť by nedokázal zlepšit klasifikační pravidla používaná poskytovatelem. Faktory, jež by mohly zlepšit výkonnost neuronových sítí, jsou uvedeny v podkapitole 11.4.1. Nedostatečnou schopnost sítí správně klasifikovat mohlo vedle omezené volby vstupních faktorů způsobit i značně nestabilní prostředí splátkového prodeje, které právě ve sledované době zažívalo boom. Právě nepřehledná situace té doby však byla motivací pro otestování moderního nástroje pro diskriminaci, kterým neuronové sítě jsou.

Kritické hodnocení způsobu klasifikace sítě $PNS_{0,045}$ znamená, že na každého správně klasifikovaného neplatiče připadají dva špatně klasifikování platiči. Síť $PNS_{0,045}$ dokázala odhalit 35 % neplatičů. Tedy odhalení neplatících klientů na jedné straně znamená dvojnásobnou ztrátou platících klientů na straně druhé. Zisk a ztráta počítána klienty není z pohledu poskytovatele relevantní. Zatímco schválení úvěru klientovi, který se stane neplatičem, může způsobit ztrátu ve výši celé investované částky, schválení úvěru klientovi, který úvěr bez problémů splatí, přinese pouze cca zisk 10 %. Pro tyto úvahy by však bylo třeba získat úplné informace přímo od poskytovatele a případný zisk respektive ztrátu přesně vyčíslit, neboť úvahy o zařazení ztrátové matice do procesu trénování pravděpodobnostní neuronové sítě nejsou podloženy reálnými daty a byly vytvářeny jen na základě obecných ekonomických úvah.

Pokud by byl-lo pracováno s touto, daty nepodloženou domněnkou, a k její realizaci by byla použita ztrátová matice (popsaná na str. 115), bylo by nutno poněkud změnit celý přístup nahlížení na problém. Ztrátová matice by musela být experimentálně nastavena tak, aby ztráta špatně zařazeného neplatiče byla přiměřeně vyšší, než ztráta v případě chybně zařazeného platiče²⁰. Celková výkonnost pravděpodob-

²⁰Pokud by platila premisa o cca 10% zisku z platícího klienta – platiče, pak by ztráta špatně zařazeného neplatiče musela být destkrát vyšší, než ztráta v případě chybně zařazeného platiče.

nostní neuronové sítě měřená procentem správně zařazených klientů by poklesla, což bylo ověřeno. Hodnocení výkonnosti prostřednictvím procenta správně klasifikovaných vzorů by nemohlo být považováno za relevantní a muselo by se změnit, protože chyba při trénování je založena na jiném kritériu, než výsledná výkonnost sítě. I výkonnost by musela být posuzována z pohledu finančního (zisk, ztráty) a nikoliv z pohledu zisku respektive ztráty bonitního klienta. Hodnocení na základě finanční ztráty/zisku by bylo důležité spíše z pohledu poskytovatele nikoli prodejce, z jehož pohledu byl problém řešen. Prodejce se snaží minimalizovat především počet klientů a přitom poskytovateli přinášet maximální počet bezproblémových zákazníků-klientů a sobě tím maximální zisk spojený s přízní poskytovatele (výhodné splátkové kalendáře pro zákazníky, benevolentnější pravidla pro poskytnutí úvěru, atp.).

11.4.1 Návrh faktorů umožňujících zlepšení správné klasifikace

Výsledky snahy o vytvoření a natrénování neuronové sítě jako expertního nástroje pro určování bonity klienta ukázaly, že použité diskriminátory k úspěšnější klasifikaci nedostačovaly. B. Relevantní odhad bonity klienta by mohly zlepšit informace o:

- a) jeho zaměstnání,
- b) problémech při dřívějším splácení úvěru,
- c) počtu a výši souběžně běžících úvěrů.

Add a). Informace o zaměstnání sice byla jednou z informací, kterou klient-žadatel musel v žádostech o spotřebitelský úvěr vyplnit. Bohužel změna metodiky neumožnila faktor zaměstnání mezi diskriminátory zařadit, neboť u části klientů byl zjišťován přímo typ práce – profese²¹ a u části klientů jen informace o tom, zda-li se

²¹Výpis různých úrovní tohoto kategoriálního znaku (převáděno do mužského rodu) nabízí následující výčet: administrativní pracovník, asistent, automechanik, bagrista, barman, brusič, CNC programátor, číšník, dělník, detektiv, dispečer, ekonom, elektrikář, elektromechanik, elektromontér, elektrikář, energetik, inspektor, instalatér, jiný, kadeřník, kontrolor jakosti, kovoobráběč, krejčí, kuchař, kvalitář, lékař, malíř, natěrač, manažer kvality, manažer prodeje, mateřská dovolená, mechanik, mistr, montér, nástrojař, obchodní poradce, obchodní zástupce, obsluha čerpací stanice, obchodní asistent, ochranka, operátor tkalcovny, opravář, ostraha, OSVČ, ošetřovatel, pečovatel, pedikúra, pekař, pletař, pokladník, pokojský, pokrývač, policejní inspektor, poštovní doručovatel, prodávač, programátor CNC, radiologický asistent, realitní činnost, referent, řezník, řidič, sanitář, seřizovač, skladník, směnárník, soustružník, starosta, státní zaměstnanec, strážník, strážný, svářeč, šič, tajemník, technik, technolog, tesař, topenář, topič, traktorista, truhlář, účetní, učitel, uklízeč, úředník, vedoucí potravin, vedoucí skladu, vedoucí směnový, vedoucí výjezdu, voják, zámečník, zástupce

jedná o podnikatele (OSVČ), zaměstnance s pracovní smlouvou na dobu neurčitou či důchodce – blíže viz podkapitola 11.2. Právě druhá metodika zaznamenávání formy příjmu klienta se jeví jako rozumnější.

Add b). Informace o problémech lze přímo od klienta získat těžko. Není v jeho zájmu se k dřívějšímu nesplácení přiznávat. Poskytovatelům jsou sice k dispozici databáze, jež poskytují informace o občanech, kteří v minulosti pochybili, nebo – bohužel – byli za neplatiče nezaslouženě označeni.

Jak uvedl Osvaag [68], registrů, kam se klient může při nesplácení úvěru dostat, je celá řada. Např.:

CRÚ Centrální registr úvěrů provozovaný Českou národní bankou eviduje aktuální údaje o právnických a fyzických osobách. Jedná se o souhrn informací o finančních závazcích a jejich historii. V databázi CRÚ však nejsou evidovány spotřebitelské úvěry fyzických osob, hypoteční úvěry fyzických osob, ručitelské závazky klientů, ani údaje o depozitních účtech (běžné účty bez povoleného debetu, spořicí, termínové účty). Nicméně pochybení při splácení jakéhokoli závazku má vypovídací schopnost. Více například na CzechTrade [10].

BRKI Bankovní registr klientských informací eviduje fyzické osoby (občany s RČ a podnikatele s IČ). Provozovatelem je Czech Banking Credit Bureau ([5]). Údaje o půjčkách, tedy o dluzích klientů, zadávají do bankovního registru všechny banky. Shromažďuje pozitivní i negativní informace k prověřovanému subjektu za účelem získání uceleného obrazu hospodaření klienta a jeho finančních toků ovlivňující jeho bonitu. Informace se vymazávají po čtyřech letech.

NRKI Nebankovní registr klientských informací sdružující společnosti v oblasti leasingu a splátkového prodeje provozuje Leasing And Loan Credit Bureau (Leasing And Loan Credit Bureau [43]). Informace o závazcích, včasnosti jejich plnění a zajištění jsou v databázi ještě čtyři roky od ukončení smluvního vztahu s členem registru.

SOLUS Eviduje podnikatelské subjekty s IČ a fyzické osoby s RČ. Jde o negativní rejstřík, tedy databázi těch, jejichž platební morálka byla či je špatná. Více přímo na SOLUS [78].

CERD Centrální registr dlužníků. Informační systém, který umožňuje vyhledávat nesplacené závazky ekonomických subjektů registrovaných v České republice.

vedoucího, zdravotní laborant, zdravotní sestra, zdravotník a zedník. Je zřejmé, že by s ohledem na charakter znaku muselo dojít k překódování tohoto znaku – 54 profesí bylo zařazeno pouze jednou. Na to, jak profese sloučit (optimálně s ohledem na bonitu klienta) však nebylo k dispozici dostatek dat. Navíc vyplněné informace svědčí o značné „lidové tvořivosti“.

Zprostředkovává zobrazení údajů vedených v jednotlivých zdrojových registrech, ze kterých čerpá data. Provozuje ho CERD ČR. Více přímo na CERD System [8].

V době sběru a zpracovávání dat, nebyly tyto registry tak běžné, přestože například registr SOLUS je v provozu již do roku 1999. Navíc registry neobsahovaly relevantní informace²².

Add c). Počet a výše souběžně běžících úvěrů informuje o míře finanční zátěže vzniklé klientovi splácením jiných úvěrů. Tato zátěž by mohla působit problémy při splácení nového závazku a fakticky snižuje příjem klienta. Některé registry zmíněné informace obsahují, ale úplný výčet souběžně běžících úvěrů by musel přiznat klient sám. Jak k tomu klienta motivovat, bez možnosti finanční penalizace za zatajení těchto informací, však není zřejmé.

11.5 Expertní systém pro poskytovatele

V podkapitole 11.5 je navržen prázdný expertní systém (dále jen *systém B*) pro určování bonity žadatelů (klientů) o spotřebitelský úvěr tak, aby jej mohl využívat poskytovatel. Publikace Biskup – Čermáková [112], z níž konstrukce systému B vychází, je zaměřena na využití neuronových sítí pro vytvoření expertního systému bez experta (viz podkapitola 8.3.3) a popisuje možnosti tohoto expertního systému.

S ohledem na reálné využití systému B nezbyvá, než se smířit s tím, že žádný poskytovatel neuvolní všechny potřebné informace pro trénink neuronové sítě. Odhalil by tak svou bonitní politiku a zveřejnil svá citlivá ekonomická data. Systém A a systém B jsou odlišné expertní systémy. Systém B popsáný v podkapitole 11.5 je v porovnání se systémem A sofistikovanější a jeho aplikace stejně jako ověření funkčnosti by byly daleko náročnější. Vytvoření systému A, popsáného v podkapitole 11.3, umožnilo představení možností využití neuronových sítí pro tvorbu expertního systému určujícího bonitu klienta alespoň z pohledu prodejce.

²²V registrech nebylo možno najít prodejci „známé“ neplatiče, jejichž pohledávky byly například již vymáhány soudně, či neplatiče, kteří byli celostátně hledáni pro podvody se spotřebitelskými úvěry. Naopak v registru neplatičů bylo možno nalézt majitele společnosti, která poskytla data, pro jednu pozdě zaplacenou fakturu mobilnímu operátorovi za vyúčtování hovorů na soukromém čísle.

11.5.1 Popis reálné situace z pohledu poskytovatele

Určování bonity klienta a následné přidělení či zamítnutí úvěru bylo je a v současné době automatizované. Ve většině případů probíhá on-line prostřednictvím formulářů vyhodnocovaných přímo PHP serverem, přičemž rozhodovací pravidla (tj. expertní systém) jsou součástí PHP kódu. Spotřebiteli je úvěr buď přidělen, zamítnut, nebo zamítnut s možností osobní komunikace s operátorem. Často se stává, že je úvěr schválen za jiných podmínek (změna splátkového kalendáře, apod.), neboť v původně zvolené konfiguraci je pro poskytovatele riskantní. Na klientovi pak záleží, zda na poskytovatelem navržené podmínky přistoupí.

Poskytovatel by se měl pokoušet zlepšovat/aktualizovat pravidla pro poskytování úvěru v závislosti na informacích, které poskytuje aktuální situace na trhu. Při tom musí vycházet ze své vyzkoušené bonitní politiky²³. Na jedné straně by poskytovatel měl ze stávajících klientů, považovaných dle pravidel za bonitní, vyřadit ty, kteří obvykle neplní své závazky a na straně druhé by mohl zařadit mezi bonitní klienty některé z klientů, kteří až doposud úvěr nemohli získat, ale kombinace hodnot jejich ekonomických faktorů naznačuje vysokou (byť praxí neověřenou) bonitu. Jak tuto situaci řešit prostřednictvím neuronových sítí je uvedeno v podkapitole 11.5.2.

11.5.2 Návrh expertního systému pro poskytovatele

Navržený expertní systém založený na neuronových sítích musí, dle požadavku uvedeného v podkapitole 11.5.1, dokázat rozlišit klienty na bonitní a nebonitní dle stávajících pravidel a zároveň bonitní a nebonitní dle reálné situace. Tomuto požadavku byl přizpůsoben i návrh výstupních kritérií. Pro trénování sítí byla zvažována výstupní kritéria:

- 1) bonita klienta,
- 2) příslušnost ke skupině klientů, kterým byl zamítnut respektive nezamítnut spotřebitelský úvěr.

Znalost by tedy reprezentovaly dvě sítě. Vstupem obou sítí by byly podstatné faktory, popisující klienta zjišťované v rámci žádosti o spotřebitelský úvěr podobně jako u systému A.

²³Poskytovatel má zmapovanou oblast podstatných faktorů ovlivňujících bonitu klienta a vytvořená pravidla pro rozhodování. Nutno dodat, že určování bonity klienta je založeno především na výši jeho příjmu, na případných minulých zkušenostech firmy s klientem a libovůli prodejce. Prodejce má totiž možnost přímo doporučit či nedoporučit každého z klientů.

Jedna neuronová síť by byla vytvořena tak, aby dokázala reprezentovat znalosti o bonitě klientů, jimž byl spotřebitelský úvěr poskytnut. Po natrénování by tedy určovala bonitu klientů²⁴.

Druhá neuronová síť by byla vytvořena tak, aby reprezentovala aktuální pravidla používaná pro přidělování úvěru. Tato neuronová síť by byla natrénovaná na celé množině žádostí o splátkový prodej a umožnila by rozhodovat o přidělení úvěru za stávajících pravidel bez toho, aby musela být tato pravidla pro poskytnutí úvěru explicitně vyjádřena²⁵.

Kdyby byl jako bonitní klient určován pouze ten, který splnil kritérium vypočtené druhou neuronovou sítí, a jehož bonita byla na základě kritéria vypočteného první neuronovou sítí dostatečná, zužovala by se skupina klientů, kteří by se svou žádostí o splátkový prodej úspěšili. Zužovat klientelu – byť za cenu vytipování bezproblémového klienta –, není cílem poskytovatele. Schválení úvěru klientovi by se tedy mohlo odůvodnit i v případě, že přísluší do skupiny klientů, kteří nesplní kritérium vypočtené druhou neuronovou sítí, za předpokladu, že na základě prvního kritéria je označen jako bonitní.

Jak by se na tuto situaci dalo nahlížet? Klient by sice za běžných okolností úvěr nedostal (druhé kritérium), ale na základě hodnocení jeho bonity (první kritérium) tak, jak ho klasifikuje znalost vytvořená v neuronové sítí, se jeví jako bezproblémový. V tomto okamžiku se jedná o jistou extrapolaci²⁶ pravidel. Extrapolaci, neboť nebyl-li zatím povolen spotřebitelský úvěr pro takového, ani podobného klienta, nelze zcela spoléhat na hodnocení bonity dle prvního kritéria. Neuronová síť vrátí hodnotu prvního kritéria na základě jejího vnitřního parametrického nastavení a vstupních dat, ale vzhledem k absenci vzoru v trénovací množině je tato hodnota jen jakýmsi odhadem možné splátkové bezproblémovosti klienta.

Nyní tedy záleží jen na marketingové strategii a směru, jakým se chce poskytovatel pokusit, s určitým rizikem, posunout meze pro přidělování spotřebitelských úvěrů. Své rozhodnutí opírá o expertní systém, který by měl reprezentovat znalost chování klientů, a nejedná se tedy jen o rozhodnutí na základě „dojmu“. V případě ať již

²⁴Pouze u těchto klientů lze zjistit jejich chování, viz poznámka o nemožnosti přidělení úvěru každému klientovi (podkapitola 11.2).

²⁵Z matematického pohledu jde o nepřesně použité vyjádření, neboť explicitní vyjádření výpočtu neuronovou sítí jistě existuje. Nicméně znalost v síti uložená nelze efektivně analyzovat jinak, než pomocí například odezvvé plochy.

²⁶U neuronových sítí se namísto termínu extrapolace používá termín generalizace. Ten je však obsahově o něco širší.

úspěšného (předpokládá se) nebo neúspěšného posunutí mezí pro přidělení spotřebitelského úvěru, získává poskytovatel minimálně nová a aktuální data o chování klientů a může provést aktualizaci celého expertního systému. Tedy jak neuronové sítě vypovídající o aktuálním stavu pro popis bonity klienta, tak neuronové sítě vypovídající o aktuálním stavu pravidel pro poskytnutí úvěru.

Takto opakovaně znovunatrénované neuronové sítě na nově vzniknuvší, respektive rozšířené, trénovací množině reprezentují nová pravidla (váhové nastavení sítě), jež v konečném důsledku vymezují pravidla pro přiznání spotřebitelského úvěru. Vzhledem k tomu, že situace je dynamická, jsou neuronové sítě jako základ expertního systému vhodným nástrojem, neboť „přetrénování“²⁷, respektive „dotrénování“²⁸ neuronové sítě není problémem, stejně jako modifikace její architektury, pokud ta stávající již nevyhovuje.

Kdyby situace byla neměnná dalo by se očekávat, že proces znovunatrénování bude konvergovat k optimálním rozhodovacím pravidlům, tj. obě sítě by dávaly stejně předpovědi bonity. Reálná situace taková není. Neuronové sítě umožňují sledovat stále se měnící situaci a flexibilně měnit znalost chování klientů v nich uloženou prostřednictvím opakovaného trénování vhodných sítí.

Situace by nemusela být řešena dvěma oddělenými sítěmi. Pro vytvoření kompaktního expertního systému by bylo výhodné použít kolekci sítí (viz podkapitola 4.3), která by měla dva výstupy tak, jak byly definovány výše.

²⁷Tj. natrénování neuronové sítě na jiném souboru dat – např. aktuálnějším, nebo rozšířeném.

²⁸Provedení adaptace parametrů neuronové sítě pouze na souboru dat nových klientů za předpokladu, že trénink bude proveden s nižší učící konstantou. To zajistí zachování původní znalosti rozprostřené v síti, jež je vylepšována/měněna na základě aktuálních dat. Taková adaptace je možná například u vícevrstvé dopředné sítě, ale u pravděpodobnostní neuronové sítě není možná.

Kapitola 12

Časové řady a neuronové sítě

V této kapitole jsou zkoumány možnosti neuronových sítí pro predikci v časové řadě. Speciálně je pracováno se sezónními časovými řadami a na základě zobecnění zkušeností tréninku neuronových sítí je vytvořena metodika výběru a kombinace sítí v kolekce pro predikci vývoje ekonomických ukazatelů. Obdobně jako v předchozích kapitolách je nejdříve odkazováno na publikované výsledky získané prostřednictvím klasických statistických metod a tyto výsledky jsou konforntovány s výsledky získanými prostřednictvím neuronových sítí.

V podkapitole 12.6.4 je navržena hybridní metoda, která kombinuje při modelování klasické statistické metody s použitím neuronových sítí.

12.1 Motivace

Vývoj ekonomiky spotřební společnosti má periodický charakter, který je charakterizován jak sezónními výkyvy, tak dlouhodobým vývojem – trendem. V současné době prochází nejen země vyspělého světa ekonomickou krizí, která je označována za celosvětovou. Přestože není jasné, jak se současnou recesí bojovat, administrativy všech států se snaží snížit dopad této krize na obyvatelstvo. Monitorování vývoje ukazatelů, charakterizujících krizovou situaci, stejně tak jako monitorování vlivu přijatých opatření pro zmírnění dopadů, jsou důležitým aspektem, který může pomoci současnou situaci analyzovat. Právě statistické metody jsou vhodným nástrojem pro analýzu dosavadního vývoje a predikci vývoje budoucího.

Jedním z relevantních ukazatelů, které vypovídají o negativních tendencích v ekonomice, je míra nezaměstnanosti. Výsledky za první čtvrtletí roku 2009 vydané Českým statistickým úřadem a Ministerstvem práce a sociálních věcí potvrzují negativní vliv

ekonomické krize na míru nezaměstnanosti v České republice. Průměrné hodnoty počtu nezaměstnaných za první čtvrtletí, vypočítávané dle metodiky Mezinárodní organizace práce očištěné o sezónní výkyvy, vzrostly v porovnání se čtvrtým čtvrtletím roku 2008 o 65,9 tisíc osob¹ (zdroj: Český statistický úřad [11]). V historii sledování vývoje míry nezaměstnanosti (od roku 1993) nebyl zaznamenán takto rekordní nárůst míry nezaměstnanosti. K 30. 4. 2009 bylo v České republice evidováno 456,7 tisíc nezaměstnaných (zdroj: OKsystem [66]). Lze prostřednictvím analýzy míry nezaměstnanosti zachytit zvrát v trendu a predikovat budoucí vývoj s uspokojivou přesností? Právě na tuto otázku se snaží odpovědět kapitola 12, která své výsledky opírá o metodologii neuronových sítí i klasických statistických metod.

Některé výsledky, které byly prezentovány v rámci této kapitoly, byly použity při řešení MSM 6007665806.

12.2 Soubor dat

Vzhledem k tomu, že metodika výpočtu míry nezaměstnanosti je nejednotná, bylo nutno určit zdroj, ze kterého bude sestavena časová řada vývoje nezaměstnanosti. Data byla čerpána z databáze Ministerstva práce a sociálních věcí (zdroj: OKsystem [67])².

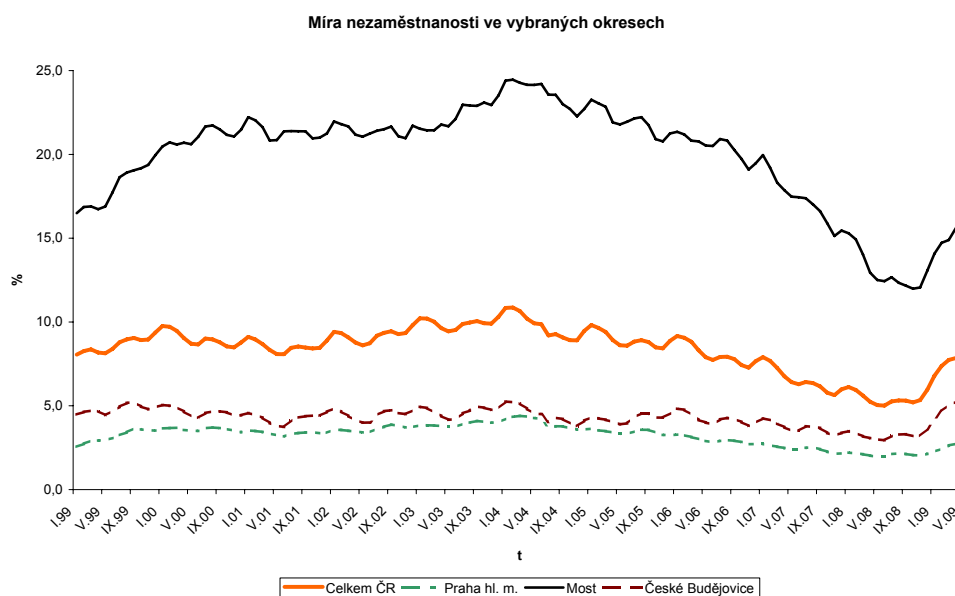
Dle OKsystem [66] je nezaměstnaným každý evidovaný uchazeč o zaměstnání, který může bezprostředně nastoupit do zaměstnání při nabídce vhodného pracovního místa, tj. evidovaný uchazeč, který nemá žádnou objektivní překážku pro přijetí zaměstnání. Takový uchazeč bývá obvykle označován za dosažitelného uchazeče o zaměstnání. Za dosažitelného se nepovažuje uchazeč o zaměstnání ve vazbě, ve výkonu trestu, uchazeč v pracovní neschopnosti, uchazeč, který je zařazen na rekvalifikační kurzy, nebo uchazeč, který vykonává krátkodobé zaměstnání, a dále uchazeč, který pobírá peněžitou pomoc v mateřství, nebo kterému je poskytována podpora v nezaměstnanosti po dobu mateřské dovolené.

Hodnoty míry nezaměstnanosti v České republice od roku 1993 jsou uvedeny v přílohách – příloha A.3 na str. 221. V této příloze jsou rovněž uvedeny i sezónně očištěné řady. Očištění od sezónnosti (klasickými metodami) provedl předkladatel práce. Graf vývoje míry nezaměstnanosti je uveden na obrázku 12.1 na str. 146. Na grafu je

¹Prostý mezičtvrtletní nárůst nezaměstnaných činil 97,2 tisíc osob.

²Míra nezaměstnanosti, jak ji vypočítává Ministerstvo práce a sociálních věcí, se nazývá „registrovaná míra nezaměstnanosti“. Používání plného názvu by zbytečně komplikovalo popisky a text disertační práce a proto nebylo používáno. Další mírou nezaměstnanosti je například tzv. „Harmonizovaná míra nezaměstnanosti“, kterou zpracovává EUROSTAT a slouží pro mezinárodní srovnání.

vedle vývoje míry nezaměstnanosti v celé České republice zobrazen rovněž vývoj míry nezaměstnanosti v Praze (Praha hl. m. – okres s jednou z nejnižších měr nezaměstnanosti v ČR), v okrese Most (okres s tradičně nejvyšší mírou nezaměstnanosti) a vývoj míry nezaměstnanosti v okrese České Budějovice.



Obrázek 12.1: Vývoj míry nezaměstnanosti v České republice a ve vybraných okresech ČR v období leden 1999–duben 2009

Do roku 2004 byla při výpočtu míry nezaměstnanosti používána jiná metodika³. Z toho důvodu nelze pro analýzu vývoje míry nezaměstnanosti použít celou časovou řadu a pro tvorbu modelů byla tedy použita časová řada až od ledna roku 2004 (tabulka 12.1).

12.3 Analýza časové řady

Z charakteru sledovaného ukazatele lze usuzovat na skutečnost, že se jedná o periodickou časovou řadu. Při tvorbě modelu pro vývoj míry nezaměstnanosti bylo uvažováno jak o modelu konstantní sezónnosti s měsíční periodicitou, tak o modelu

³Míra nezaměstnanosti do roku 2004 byla vypočítávána ze všech uchazečů o zaměstnání – nikoliv z dosažitelných uchazečů o zaměstnání – a proto byla v porovnání s aktuální mírou nezaměstnanosti vyšší.

Tabulka 12.1: Vývoj míry nezaměstnanosti v letech 2004–2009

	2004	2005	2006	2007	2008	2009
leden	9,69	9,82	9,16	7,90	6,12	6,80
únor	9,68	9,65	9,06	7,68	5,94	7,38
březen	9,47	9,41	8,82	7,26	5,61	7,74
duben	9,05	8,92	8,31	6,76	5,24	7,85
květen	8,84	8,61	7,90	6,41	5,04	
červen	8,87	8,59	7,74	6,29	5,01	
červenec	9,20	8,83	7,92	6,42	5,26	
srpen	9,28	8,91	7,93	6,35	5,34	
září	9,08	8,80	7,79	6,16	5,31	
říjen	8,92	8,49	7,43	5,78	5,20	
listopad	8,90	8,43	7,28	5,64	5,34	
prosinec	9,47	8,88	7,67	5,98	5,96	

Data převzata z databáze Ministerstva práce a sociálních věcí (OKsystem [67]) – rok 2004–2009.

proporcionální sezónnosti. Předpoklad o modelu konstantní sezónnosti byl podpořen testem. Na 5% hladině významnosti se podařilo zamítnout nulovou hypotézu ($H_0: \beta_1 = \beta_2 = \dots = \beta_{12} = 0$) ve prospěch alternativní ($H_A: \beta_i \neq 0$ alespoň pro dvě sezóny, $i = 1, \dots, n$), neboť hodnota testového kritéria $F = 3,915$ ležela v kritickém oboru $K = \{F : F \geq F_{0,95}(11, 44)\}$, kde $F_{0,95}(11, 44) = 2,014$.⁴

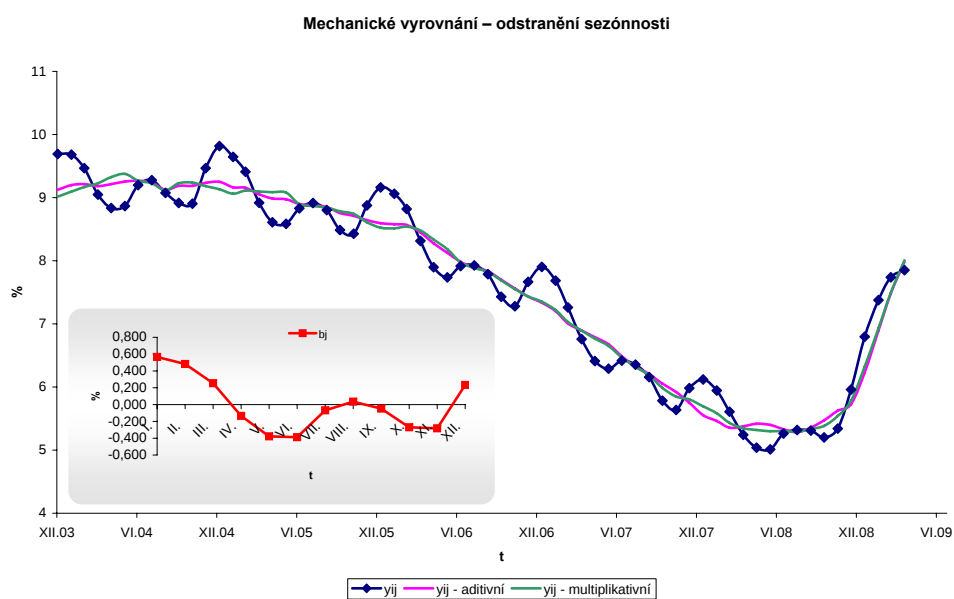
Metodou mechanického vyrovnání byly pro oba modely odhadnuty sezónní složky – viz tabulka 12.2. Hodnoty uvedené v tabulce 12.2 potvrzují a kvantifikují skutečnost, že v zimních a letních měsících se nezaměstnanost zvyšuje (v souvislosti s ukončením sezónních prací a nástupem absolventů na trh práce).

Graf (obrázek 12.2) ukazuje, že na první pohled optimistické zpomalení negativního vývoje v měsíci dubnu je ovlivněno spíše sezónními výkyvy, než výrazným zpomalením vzrůstající míry nezaměstnanosti. Po očištění o sezónnost – ať už konstantní, či proporcionální – lze stále hovořit o cca 0,5% stoupající míře nezaměstnanosti. Stejně tak – na základě očištěné časové řady – stagnace poklesu míry nezaměstnanosti v České republice započala již v dubnu loňského roku. Od října již docházelo ke zhoršování situace, jak již bylo uvedeno v podkapitole 12.1.

⁴p-value = 0,000.

Tabulka 12.2: Sezónnost v modelech konstantní a proporcionální sezónnosti

	b_j	$(1 + c_j)$
leden	0,567	1,075
únor	0,483	1,064
březen	0,254	1,033
duben	-0,134	0,980
květen	-0,377	0,947
červen	-0,388	0,946
červenec	-0,068	0,993
srpen	0,035	1,005
září	-0,047	0,995
říjen	-0,270	0,966
listopad	-0,282	0,964
prosinec	0,234	1,031



Obrázek 12.2: Očištěná časová řada míry nezaměstnanosti o sezónní složku leden 2003–duben 2009

12.3.1 Výběr metod pro modelování míry nezaměstnanosti

Výběr vhodných modelů a topologií neuronových sítí pro modelování vývoje míry nezaměstnanosti v České republice musel zohledňovat následující skutečnosti:

1. míra nezaměstnanosti je ovlivňována sezónními výkyvy v rámci kalendářního roku (bylo prokázáno výše),
2. časová řada míry nezaměstnanosti je jistě autokorelovaná a to tak, že pro vysvětlení aktuální míry nezaměstnanosti má nejvyšší váhu hodnota v předchozím měsíci a vliv hodnot míry nezaměstnanosti směrem do minulosti klesá,
3. použitý statistický software (STATISTICA komplet 6.1 Cz) vyžaduje časovou řadu dlouhou alespoň pět kompletních period a pro neuronovou síť je potřebná co nejdelší časová řada⁵,
4. nezaměstnanost je ovlivňována celou řadou faktorů, které dokáží vysvětlit změny v chování vývoje míry nezaměstnanosti, respektive existuje souvislost mezi těmito faktory a mírou nezaměstnanosti.

Pro modelování vývoje míry nezaměstnanosti byly s ohledem na skutečnosti 1 a 2 vybrány vícevrstvé dopředné sítě⁶ a z klasických statistických metod metoda exponenciálního vyrovnaní. Skutečnost zmíněná v bodu 3 osvětluje výběr délky časové řady. Přestože zařazení intervenčních respektive korelovaných časových řad (skutečnost č. 4) obvykle dokáže vylepšit predikční schopnosti modelů, jejich použití bylo zavrhnuto z následujících důvodů:

- Faktory, které by ovlivnily vývoj míry nezaměstnanosti tvoří časové řady v pravém slova smyslu, ale právě tak k nim přistupuje neuronová síť. Jaký vliv by na míru nezaměstnanosti mělo zavedení „šrotovného“⁷, zvýšení financování programů na rekvalifikaci, atp. nelze předjímat, neboť v historii nemají obdoby.
- Časové řady, jakými jsou růst HDP, obchodní bilance atp. jsou stejně jako míra nezaměstnanosti endogenními proměnnými systému, které jsou vývojem

⁵Němec [57] ve své práci *Predikce potřeby vody pomocí neuronových sítí* uvádí, že je vhodné, aby trénovací množina obsahovala alespoň tolik prvků, kolik má neuronová síť volných parametrů, jinak dochází ke ztrátě robustnosti a obecnosti neuronové sítě! Tedy neuronová síť by neměla mít více vah, než-li je počet hodnot k učení neuronové sítě.

⁶RBF sítě nejsou z logiky věci vhodné pro predikce v časových řadách. Při výraznější změně trendu by totiž na vstup sítě přicházely hodnoty značně vzdálené od lokalizovaných jednotek a predikce vytvořené RBF sítí by nebyly podložené.

⁷Šrotovným je označován státní příspěvek na nákup nového auta majiteli, který nechá ekologicky zlikvidovat své staré auto. Příspěvek má kromě obnovy automobilového parku pomoci zabránit stagnaci v automobilovém průmyslu a tedy propouštění pracovníků v tomto průmyslovém odvětví a v odvětvích na něj navazujících. Tato pobídka má v různých zemích různé podmínky.

ekonomiky ovlivňovány stejně jako samotná míra nezaměstnanosti.

- Intervenční časové řady zvyšují počet parametrů odhadovaných jak při konstrukci modelu, tak při trénování neuronové sítě. V případě vývoje míry nezaměstnanosti však neexistuje tak dlouhá časová řada, aby bylo možno tyto parametry efektivně odhadnout.

Predikční schopnost vytvořených modelů byla hodnocena prostřednictvím Theilova koeficientu nesouladu. Proto byly hodnoty míry nezaměstnanosti v březnu a dubnu roku 2009 (tj. nejčerstvější hodnoty sledovaného ukazatele v době odevzdávání disertační práce) ponechány mimo proces tvorby modelu a trénování neuronových sítí.

12.4 Modelování a predikce v časových řadách

Pro posouzení vhodnosti modelů byly v první fázi použity charakteristiky $M.A.E.$ ⁸, $M.S.E.$ ⁹, $M.A.P.E.$ ¹⁰, $S.D.$ poměr¹¹. Porovnání mezi exponenciálním vyrovnáním a neuronovými sítěmi bylo provedeno prostřednictvím charakteristiky $M.A.E.$ Základním ukazatelem vhodnosti neuronové sítě pro modelování časové řady je však $S.D.$ poměr.

Predikční schopnost modelů byla evaluována pomocí Theilova koeficientu nesouladu¹², vypočítávaného podle vzorce (12.1):

$$U = \frac{\sqrt{\sum_{h=1}^{n-N} (P_N(N+h) - y_{N+h})^2}}{\sqrt{\sum_{h=1}^{n-N} (P_N(N+h))^2} + \sqrt{\sum_{h=1}^{n-N} (y_{N+h})^2}}, \quad (12.1)$$

kde

⁸střední absolutní chyba odhadu

⁹střední kvadratická chyba odhadu

¹⁰střední absolutní procentická chyba odhadu

¹¹poměr směrodatné odchylky chyby odhadu ku směrodatné odchylce empirických hodnot

¹²Theilovo U koeficient

N	je délka časové řady použité pro tvorbu modelu/trénování neuronové sítě
n	je délka celé časové řady
y_t	empirické hodnoty sledovaného ukazatele
$P_N(N + h)$	v čase N predikované hodnoty do časového horizontu h (predikce začínají v $t = (N + 1)$ -ní hodnotě), $h = 1, \dots, n - N$

Testová statistika U nabývá hodnot z intervalu $\langle 0; 1 \rangle$. Pokud je $U = 0$, jsou predikční schopnosti modelu excelentní (predikce je shodná s empirickými hodnotami). Zatímco $M.A.E.$, $M.S.E.$, $M.A.P.E.$ a $S.D.$ poměr jsou míry shody modelu s pozorovanými hodnotami použitými pro konstrukci modelu/trénování neuronových sítí, Theilův koeficient nesouladu je mírou shody hodnot predikovaných modelem v porovnání s hodnotami naměřenými ex post. Jedná se tedy o validaci modelu.

12.5 Exponenciální vyrovnání

Vhodné modely a vyrovnávací konstanty pro modely založené na exponenciálním vyrovnání byly vybrány na základě analýzy charakteristik $M.A.E.$, $M.S.E.$ a $M.A.P.E.$. Kompletní výsledky byly prezentovány v článku Čermáková – Biskup [118]. Byly zvažovány jak modely konstantní tak modely proporcionální sezónnosti s exponenciálním a tlumeným trendem. Výsledky jsou uvedeny v tabulce 12.3.

Tabulka 12.3: Vyrovnávací konstanty, hodnoty $M.A.E.$ a Theilův koeficient nesouladu pro nejlepší modely

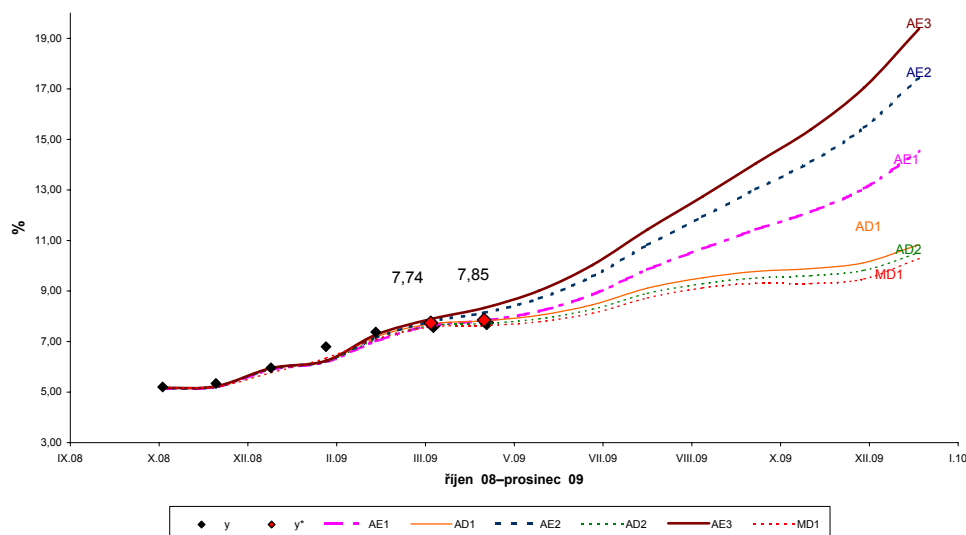
Ozn.	Sezónnost	Trend	Kritérium	$M.A.E.$	α	γ	δ	φ	U
AE1	konst.	exp.	$M.A.E.$	0,098	0,9	0,1	0,4		0,005
AD1	konst.	tlum.	$M.A.E.$	0,099	0,8	0,1		0,9	0,002
AE2	konst.	exp.	$M.S.E.$	0,099	0,9	0,1	0,6		0,013
AD2	konst.	tlum.	$M.S.E.$	0,101	0,7	0,1		0,9	0,008
AE3	konst.	exp.	$M.A.P.E.$	0,098	0,9	0,1	0,8		0,022
MD1	prop.	tlum.	$M.A.P.E.$	0,103	0,7	0,1		0,9	0,013

konst. = konstantní, prop. = proporcionální, exp. = exponenciální, tlum. = tlumený,

U = Theilův koeficient nesouladu

Grafické zobrazení empirických hodnot i predikcí je na obrázku 12.3. Hodnoty použité pro ověření predikčních schopností modelů (březnová a dubnová míra neza-

městnanosti) jsou zvýrazněny červeně (červeně vyplněný kosočtverec). Predikovaná míra nezaměstnanosti prostřednictvím optimálních modelů se pohybovala v rozmezí 7,58 %–7,89 % (březen) a 7,63 %–8,33 % (duben) – více výsledků je uvedeno v souhrnné tabulce 12.12 na str. 165, hodnoty Theilova koeficientu nesouladu v rozmezí 0,002–0,022 (viz tabulka 12.3).



Obrázek 12.3: Modely vývoje míry nezaměstnanosti v České republice od října 2008 do prosince 2009 (včetně predikce)

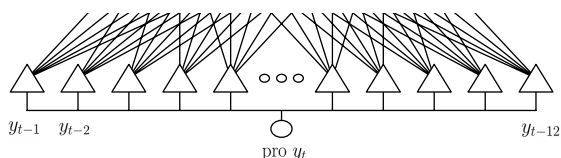
Podrobná tabulka výsledků spolu s predikcemi až do konce roku 2009 je uvedena v článku Čermáková – Biskup [118] a byla podkladem pro vytvoření grafu na obrázku 12.3.

12.6 Neuronové sítě

12.6.1 Modelování vývoje prostřednictvím neuronových sítí

Trénování neuronových sítí pro modelování vývoje v časových řadách má svá specifika. Tréninkové vzory jsou složeny z hodnot *časového okna* a predikované hodnoty. S ohledem na předpoklad měsíční sezónnosti časové řady bylo zvoleno časové okno dlouhé dvanáct hodnot.

Program STATISTICA Neuronové sítě 7.1 umožňuje trénování neuronové sítě bez složité tvorby trénovací množiny¹³. Stačí určit, že neuronové sítě budou použity pro modelování časových řad, identifikovat modelovanou časovou řadu, případně řady intervenční, vymezit časové okno a hodnotu, která má být predikována (y_t , y_{t+1} , atp.). Samozřejmě je nutné zadat další náležitosti, běžné při tvorbě a trénování neuronových sítí. V případě časových řad je počet neuronů vstupní vrstvy roven délce časového okna – viz obrázek 12.4, výstupní vrstva má jeden neuron.



Obrázek 12.4: Schéma vstupu neuronové sítě 1:12-... (časové řady)

(Vytvořeno v programu STATISTICA Neuronové sítě 7.1 – upraveno)

Na rozdíl od klasických statistických metod přistupuje neuronová síť k časové řadě jako k celku. Hodnotu budoucí predikuje pouze na základě hodnot minulých. Neprovádí dekompozici časové řady. Obitko [65] uvádí, že neuronová síť je natrénována na datech z minulosti s nadějí, že objeví skryté závislosti a bude jich schopna využít pro predikci. Neuronová síť tedy nepředstavuje explicitně daný model časové řady; slouží jako černá skříňka, která predikuje budoucí vývoj.

Pro trénování neuronových sítí jsou trénovací, validační a testovací vzory vybírány naprosto libovolně z celé časové řady¹⁴, což není při systematické práci klasických statistických metod běžné.

12.6.2 Využití heuristiky pro trénování neuronových sítí

Pokud by v případě vývoje nezaměstnanosti (62 hodnot) měla být volena neuronová síť tak, aby odpovídala podmínce uvedené v poznámce 5 na str. 149, mohla by dopředná neuronová síť s časovým oknem dlouhým 12 hodnot obsahovat maximálně

¹³Tvorba trénovací množiny pro software s implementovanou neuronovou sítí bez podpory modelování časových řad je popsána například v příspěvku Biskup – Čermáková [111], ve kterém byl pro predikci kurzu použit program Neurex 4.0.

¹⁴Každému vzoru – specifikovanému hodnotou y_t – přísluší také hodnoty časového okna $y_{t-1}, y_{t-2}, \dots, y_{t-k}$, z toho vyplývá, že pro trénování mohou být použity pouze vzory y_t , pro které platí $t > k$.

jednu skrytou vrstvu se čtyřmi skrytými neurony. To ještě za předpokladu, že by v procesu trénování byla použita celá časová řada.

I přesto bylo možno natrénovat daleko složitější neuronové sítě, které umožňovaly reálné predikce. Padesát nejlépe natrénovaných vícevrstvých dopředných sítí prostřednictvím heuristiky softwaru STATISTICA Neuronové sítě 7.1 predikovalo hodnoty březnové a dubnové míry nezaměstnanosti v rozmezí 7,3 %–7,45 % a 6,62 %–8,22 %. Heuristika často zkrátila časové okno a obvykle volila velmi jednoduché sítě. Predikované hodnoty do delšího časového horizontu byly obvykle nereálné. Charakteristiky dvou neuronových sítí, které byly do tohoto výběru heuristikou zařazeny a přitom měly bohatou architekturu, jsou uvedeny v tabulce 12.4 a zobrazeny na obrázku 12.5 na str. 156. Hodnoty *M.A.E.* a Theilova koeficientu nesouladu pro tyto dvě sítě umožňují bezprostřední srovnání výkonnosti sítí (tabulce 12.4) s výkonností exponenciálního vyrovnání (tabulka 12.3 na str. 151).

Tabulka 12.4: Hodnoty *M.A.E.*, Theilova koeficientu nesouladu a trénovací metody pro vybrané neuronové sítě

Ozn.	Architektura sítě	S.D.	<i>M.A.E.</i>	Průběh tréninku	<i>U</i>
VVP1	s12 VVP 1:12-20-8-1:1	0,06	0,069	100 epoch metodou back-propagation, 162 epoch metodou sdružených gradientů – b	0,016
VVP2	s12 VVP 1:12-20-9-1:1	0,06	0,056	100 epoch metodou back-propagation, 200 epoch metodou sdružených gradientů – b	0,041

S.D. = S.D. poměr, s = délka časového okna, b = obnovena nejlepší síť,

U = Theilův koeficient nesouladu

Bohaté sítě, vytvořené pro predikci vývoje míry nezaměstnanosti, sice dosahovaly srovnatelných a lepších hodnot charakteristik *M.A.E.* než modely založené na exponenciálním vyrovnání, ale jejich generalizační schopnost byla velmi slabá, což ukázaly právě predikce do delšího časového horizontu. Ukázalo se tedy, že nelze spoléhat na vypočítané trénovací, validační a testovací výkony neuronových sítí, neboť nepřináší relevantní měřítko výkonnosti. Jejich predikční výkonnost ve srovnání s exponenciálním vyrovnáním nadhodnocují¹⁵.

¹⁵Není třeba opakovat, že dostatečně bohaté sítě trénované na celé délce časové řady dosahovaly

Následující podkapitola simuluje zhoršování stability řešení pro predikci v časové řadě s měsíční sezónností a řešením tohoto problému v závislosti na zvyšujícím se počtu neuronů ve skryté vrstvě.

12.6.3 Využití kolekcí jednoduchých sítí pro predikci v časových řadách

Výsledků srovnatelných s výsledky exponenciálního vyrovnání – měřeno prostřednictvím charakteristiky $M.A.E.$ – dosahovaly i velmi jednoduché vícevrstvé dopředné sítě. Zvyšující se variabilita v predikcích však ukazovala nestabilitu řešení.

V tabulce 12.5 jsou uvedeny výsledky vícevrstvé dopředné síť s jednou skrytou vrstvou s 1, 2, 3, 5, 8 a 15 neurony a dopředné sítě s dvěma skrytými vrstvami s třemi a dvěma skrytými neurony. Pro každou architekturu bylo natrénováno 80 neuronových sítí. Trénování probíhalo na 38 vzorech, validace pak na 12 vzorech. Trénování probíhalo prostřednictvím kombinace algoritmů zpětného šíření chyby (backpropagation) a sdružených gradientů. Trénování bylo ukončováno tak, aby validační chyba sítě zůstala co nejnižší. Vhodnost jednotlivých sítí byla posuzována prostřednictvím hodnot $M.A.E.$ – průměrné, minimální a maximální hodnoty.

Následně byla ze všech sítí stejné architektury (vždy 80 sítí) vytvořena kolekce, která byla použita pro nestrannou predikci a výpočet Theilova koeficientu nesouladu z této predikce. Trénovací a validační schopnost kolekce sítí byla hodnocena prostřednictvím $S.D.$ poměru. Výsledky jsou uvedeny v tabulce 12.5 a graficky vyjádřeny v přílohách (příloha B.1.1 na str. 225–232).

Pro ověření zhoršující se stability predikce bylo vytvořeno 580 sítí a 7 kolekcí. Výhody použití kolekce sítí místo jedné konkrétní sítě, byť s lepšími charakteristikami, byly uvedeny v podkapitole 4.3.3. Na tomto místě jsou konkretizovány pouze důvody bezprostředně související s řešeným problémem – tj. predikcí vývoje míry nezaměstnanosti.

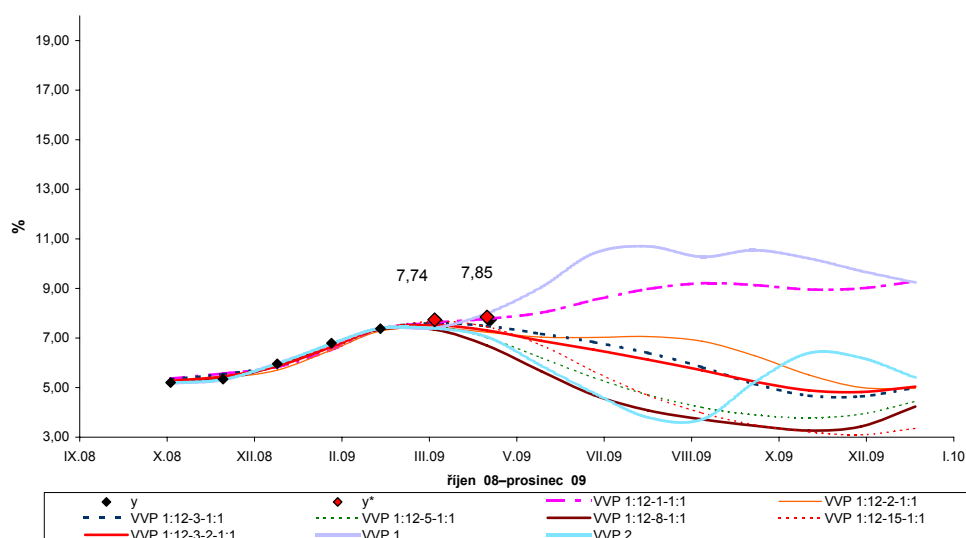
1. Do trénování neuronových sítí byly zapojeny všechny použitelné vzory. Každá ze sítí je trénována na **jiné** trénovací množině (38 vzorů) a zbytek vzorů (12 vzorů) je použit pro optimalizaci (validaci) neuronové sítě (viz podkapitola 5.5). Proto, i přes zařazení všech vzorů do trénovacího procesu, bylo možno zabránit přetrénování neuronové sítě, tedy i celé kolekce.

absolutní přesnosti v predikcích – hodnota $M.A.E$ byla pro takové sítě rovna nule, ale jejich vhodnost natrénování pro neznámé hodnoty je neověřitelná.

Tabulka 12.5: Výkonnost vybraných neuronových sítí pro modelování míry nezaměstnanosti 2004–2009

Síť	<i>M.A.E.</i>				<i>S.D.</i> poměr		<i>U</i>
	prům.	min.	max.	kol.	Trén.	Valid.	
VVP 1:12-1-1:1	0,206	0,105	1,281	0,14	0,171	0,170	0,006
VVP 1:12-2-1:1	0,17	0,064	0,343	0,121	0,131	0,196	0,033
VVP 1:12-3-1:1	0,112	0,07	0,213	0,088	0,087	0,119	0,018
VVP 1:12-5-1:1	0,094	0,06	0,179	0,066	0,071	0,117	0,041
VVP 1:12-8-1:1	0,091	0,046	0,208	0,052	0,066	0,131	0,058
VVP 1:12-15-1:1	0,081	0,041	0,279	0,045	0,057	0,119	0,018
VVP 1:12-3-2-1:1	0,12	0,06	1,256	0,074	0,098	0,12	0,027

2. Pro kolekce byla vypočtena variabilita predikcí a tedy i spolehlivost modelů a to prostřednictvím variačního koeficientu predikcí jednotlivých sítí kolekce (viz tabulka 12.6 na str. 158).



Obrázek 12.5: Predikce vývoje míry nezaměstnanosti v České republice do prosince 2009 – vícevrstvá dopředná síť – VVP 1:12-...-1:1

Hodnoty *M.A.E.*, uvedené v tabulce 12.5, ukazují pro vícevrstvou dopřednou síť s alespoň 3 skrytými neurony srovnatelné výsledky, jakých bylo dosaženo prostřednictvím exponenciálního vyrovnaní. Nutno dodat, že trénovací proces neuronových

sítí byl ukončován v okamžiku před začátkem zvyšování validační chyby. Bylo tedy zabráňováno přeučení, ale i dosažení nižších (lepších) hodnot $M.A.E.$

Hodnoty $S.D.$ poměru pro kolekci sítí (viz dva předposlední sloupce v tabulce 12.5) ukazují, že více neuronů ve skryté vrstvě umožňuje lepší predikci pro známé hodnoty, ale že zajištění generalizační schopnosti sítě zastavovalo trénovací proces na přibližně stejných validačních chybách. Grafy na obrázcích B.1–B.14 na str. 226–232 však ukazují, že i přes použití validační množiny v trénovacím docházelo k přeučení – viz značně nehomogenní predikce pro jednotlivé sítě do delšího časového horizontu. Přeučení bylo možné zejména díky volbě bohatý sítí¹⁶

Zařazení druhé skryté vrstvy (VVP 1:12-3-2-1:1) nepřineslo výrazně odlišných výsledků – viz poslední řádek v tabulce 12.5, přestože počet parametrů sítě VVP 1:12-3-2-1:1 je nižší než sítě VVP 1:12-5-1:1, která má také pět skrytých neuronů, ale zařazených do jedné vrstvy.

Grafické zobrazení výsledků predikcí prostřednictvím kolekcí analyzovaných architektur neuronových sítí je zobrazeno na obrázku 12.5. Obrázek 12.5 ukazuje, že sítě s větším počtem neuronů (více než tři neurony ve skryté vrstvě/vrstvách) mají tendenci v delším časovém horizontu predikovat velmi nízké míry nezaměstnanosti – viz šedé hodnoty v tabulce 12.13 na str. 166.

Reálnost predikcí lze posoudit prostřednictvím Theilova koeficientu nesouladu uvedeného v posledním sloupci tabulky 12.5. Růst variačního koeficientu predikce shrnuje tabulka 12.6 a naznačuje spolehlivost predikcí. Zatímco se variabilita predikce pro první období výrazně zvýší až pro vícevrstvou dopřednou síť s 8 skrytými neurony, pro druhé období stačí 5 skrytých neuronů, ve třetím období je vysoká variabilita i pro nejjednodušší síť.

Výsledky popsané v podkapitole 12.6.3, vedly předkladatele disertační práce k vytvoření hybridního systému kombinujícího klasický (statistický) přístup s technologií neuronových sítí. Tento přístup je popsán v následující podkapitole.

12.6.4 Využití kolekcí jednoduchých sítí pro tvorbu hybridního modelu

Vzhledem ke způsobu práce neuronových sítí použitých pro modelování časových řad, které se při procesu učení snaží odhalit chování všech složek časové řady současně, by měl být parametrický prostor (počet skrytých vrstev a v nich obsažených

¹⁶Přeučení z příčiny volby příliš bohaté neuronové sítě pochopitelně zařazení validační množiny do trénovacího procesu nezabrání.

Tabulka 12.6: Variabilita predikcí v závislosti na počtu neuronů ve skryté vrstvě

Síť	Variační koeficient (%)		
	březen	duben	květen
VVP 1:12-1-1:1	6,15	8,33	12,32
VVP 1:12-2-1:1	8,12	12,33	16,98
VVP 1:12-3-1:1	5,08	12,14	21,45
VVP 1:12-5-1:1	6,70	20,28	36,25
VVP 1:12-8-1:1	16,36	37,42	73,34
VVP 1:12-15-1:1	12,79	31,15	60,66
VVP 1:12-3-2-1:1	8,48	17,34	25,95

neuronů) neuronových sítí dostatečně široký. Tento požadavek je ale v přímém rozporu s omezeným počtem trénovacích vzorů, který je běžný pro časové řady s měsíční periodou sledování.

Cílem podkapitoly 12.6.4 byla analýza možností vytvořeného hybridního modelu kombinujícího klasický statistický přístup s technologií neuronových sítí. Neuronové sítě byly použity pro modelování trendu, zatímco sezónní složky byly odhadnuty prostřednictvím metod mechanického vyrovnání. Časová řada míry nezaměstnanosti byla nejprve očištěna o sezónní složku (byla uvažována konstantní sezónnost) a trénování neuronové sítě proběhlo na této očištěné řadě – viz tabulka A.3 na str. 221, sloupec y_3 . Pro možnost porovnání výsledků s neuronovými sítěmi použitými pro modelování celé časové řady, byla provedena obdobná simulace jako v podkapitole 12.6.3.

Pro odlišení budou neuronové sítě/kolekce použité pouze pro odhad trendu označeny vedle vyjádření architektury také symbolem (T) – např. VVP 1:12-1-1:1 (T). Číselné ukazatele stejně tak jako grafy prezentují opět složenou časovou řadu – tj. odhad trendu prostřednictvím neuronové sítě a vliv sezónních složek odhadnutých prostřednictvím metod mechanického vyrovnání. Výjimku tvoří jen některé grafy zařazené do příloh, kde je skutečnost, že se jedná o vizualizaci pouze trendové složky jasně vyznačena.

Charakteristiky uvedené v tabulce 12.7 ukazují, že odhad trendu prostřednictvím neuronových sítí a jejich kolekcí je lepší než odhad chování všech složek časové řady simultánně. Hlavním přínosem této dekompozice je zajištění mnohem vyšší spolehlivosti/vyrovnanosti predikcí. Tato skutečnost je zřejmá z grafů na obrázcích B.15–B.28 na str. 233–239 a numericky je potvrzena v tabulce 12.8.

Grafické zobrazení výsledků predikcí prostřednictvím kolekcí analyzovaných ar-

Tabulka 12.7: Výkonnost vybraných neuronových sítí pro modelování trendu míry nezaměstnanosti 2004–2009

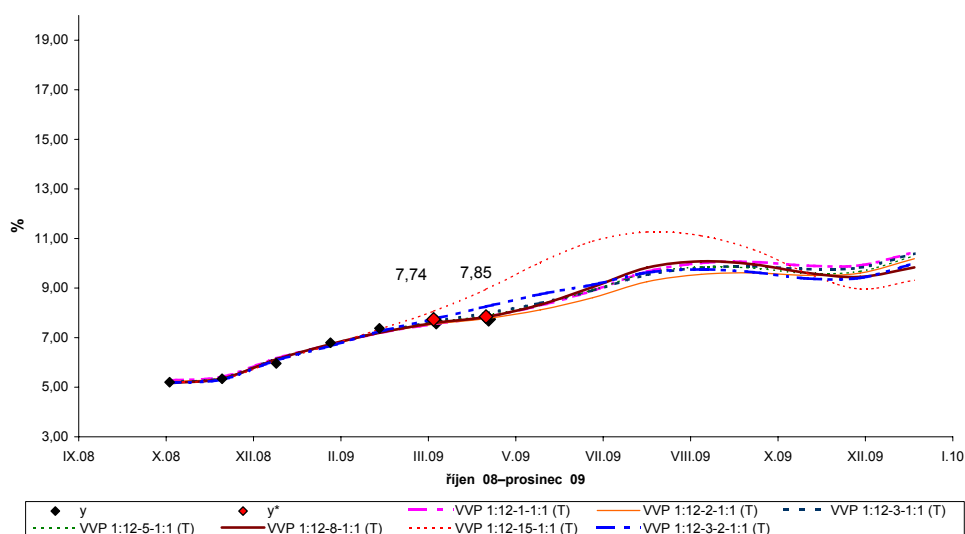
Síť	<i>M.A.E.</i>				<i>S.D.</i> poměr		<i>U</i>
	prům.	min.	max.	kol.	Trén.	Valid.	
VVP 1:12-1-1:1 (T)	0,116	0,051	1,24	0,072	0,1	0,108	0,009
VVP 1:12-2-1:1 (T)	0,074	0,05	0,135	0,057	0,069	0,072	0,01
VVP 1:12-3-1:1 (T)	0,074	0,051	0,162	0,051	0,067	0,075	0,007
VVP 1:12-5-1:1 (T)	0,07	0,041	0,158	0,049	0,061	0,08	0,006
VVP 1:12-8-1:1 (T)	0,068	0,036	0,127	0,046	0,057	0,09	0,007
VVP 1:12-15-1:1 (T)	0,061	0,031	0,257	0,037	0,05	0,086	0,05
VVP 1:12-3-2-1:1 (T)	0,068	0,045	0,156	0,05	0,061	0,073	0,018

Tabulka 12.8: Variabilita predikcí v závislosti na počtu neuronů ve skryté vrstvě

Síť	Variační koeficient (%)		
	březen	duben	květen
VVP 1:12-1-1:1 (T)	4,49	7,07	8,18
VVP 1:12-2-1:1 (T)	5,34	8,54	10,5
VVP 1:12-3-1:1 (T)	6,19	11,97	16,95
VVP 1:12-5-1:1 (T)	7,79	18,97	26,69
VVP 1:12-8-1:1 (T)	9,22	29,61	38,89
VVP 1:12-15-1:1 (T)	13,85	30,42	43,41
VVP 1:12-3-2-1:1 (T)	8,84	15,11	19,33

chitektur neuronových sítí pro trendovou složku sezónnosti odhadnuté metodou mechanického vyrovnaní je zobrazeno na obrázku 12.6 a predikované hodnoty shrnuty v tabulce 12.13 na str. 166. Na obrázku 12.6 je vidět, že tyto hybridní modely, u nichž je trend predikován sítěmi s větším počtem neuronů, nemají tendenci dávat do delšího časového horizontu predikce velmi nízké míry nezaměstnanosti tak, jak tomu bylo v případě, kdy vývoj byl modelován pouze prostřednictvím neuronových sítí. Rovněž hodnoty Theilova koeficientu nesouladu (poslední sloupec v tabulce 12.7) jasně vypovídají o lepší predikční schopnosti hybridních modelů. Všechny odhady trendů pro dopředné vícevrstvé sítě s jednou vrstvou obsahující 8 a méně skrytých neuronů predikovaly na dvě období s přesností $U < 0,01$. Výrazně stoupla také spolehlivost modelů zařazených do kolekce měřená prostřednictvím variačního koeficientu predikcí. Variační koeficient predikcí se pohyboval u nejlepšího modelu založeného na kolekci sítí

typu VVP 1:12-3-1:1 (T) od 6,19 % do 16,95 % pro predikce na 1–3 období.



Obrázek 12.6: Predikce vývoje míry nezaměstnanosti v České republice do prosince 2009 – vícevrstvá dopředná síť – VVP 1:12-...-1:1 (T) (trend) + mechanické vyrovnání (sezónnost)

12.7 Přetrénování na aktuálních datech

Jsou-li jednou vytvořeny použitelné modely, které se osvědčily pro predikci v časových řadách lze tyto modely použít i pro predikci na základě aktuálních dat. Zaktualizovat je možno jak modely založené na exponenciálním vyrovnání, tak modely založené na neuronových sítích a jejich kolekcí. Hodnoty ponechané pro ověření predikčních schopností modelů (březnová a dubnová míra nezaměstnanosti) byly použity pro odhad vyrovnávacích konstant a trénování neuronových sítí a v rámci podkapitoly 12.7 vznikly nové zaktualizované modely.

Pro odlišení budou zaktualizované modely obsahovat ve svém značení symbol (A) – např. AD2 (A), VVP 1:12-5-1:1 (A), VVP 1:12-1-1:1 (AT), atp. V souhrnných tabulkách 12.12 a 12.13 na str. 165 a 166 jsou barevně odlišeny hodnoty teoretické (černá barva), hodnoty predikující míru nezaměstnanosti na dvě období dopředu (červená barva) a hodnoty predikující do delšího časového horizontu (šedá barva). Je zřejmé, že pro zaktualizované modely jsou tato období jiná, než pro modely popisované v podkapitolách 12.5 a 12.6.2–12.6.4. Predikce jsou uspořádány tak, aby

blízko sebe byly podobné modely, tj. modely původní s modely zaktualizovanými a pro modely založené na stejných architekturách vícevrstvých dopředných sítí.

12.7.1 Exponenciální vyrovnnání

Kompletní výsledky pro modely založené na exponenciálním vyrovnnání byly publikovány v článku Čermáková – Biskup [119]. Tabulka 12.9 obsahuje výsledky, které ukazují, že vyrovnávací konstanty se příliš nezměnily (viz srovnání s tabulkou 12.3 na str. 151). Zato predikce na základě zaktualizovaných modelů se liší; viz graf na obrázku 12.7 v porovnání s grafem na obrázku 12.3 na str. 152, nebo predikce v tabulce 12.12 na str. 165. Rozdíl v predikci je způsoben volbou trendu. Dle vypočítaných hodnot charakteristiky $M.A.E.$ však není zřejmé, je-li vhodnější tlumený, nebo exponenciální trend. Na základě zkušeností opřených o hodnoty Theilova koeficientu nesouladu lze předpokládat spíše trend tlumený. Tuto domněnku podporuje také předpoklad, že opatření vlády budou směřovat ke zbrzdění/zastavení nárůstu míry nezaměstnanosti.

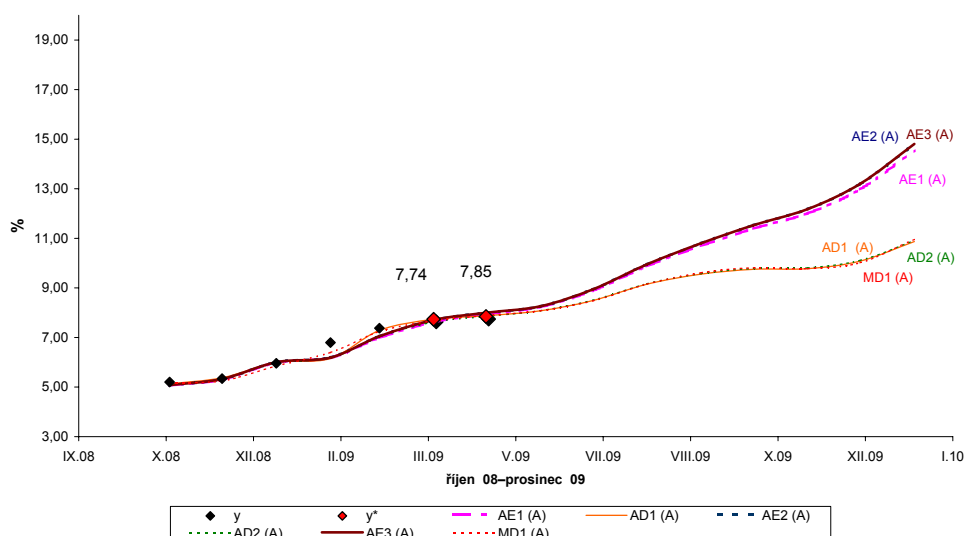
Tabulka 12.9: Vyrovnávací konstanty a hodnoty $M.A.E.$ pro zaktualizované modely exponenciálního vyrovnnání

Ozn.	Sezónnost	Trend	Kritérium	$M.A.E.$	α	γ	δ	φ
AE1 (A)	konst.	exp.	$M.A.E.$	1,003	0,9	0,1	0,4	
AD1 (A)	konst.	tlum.	$M.A.E.$	1,002	0,9	0,1		0,9
AE2 (A)	konst.	exp.	$M.S.E.$	1,004	0,9	0,2	0,5	
AD2 (A)	konst.	tlum.	$M.S.E.$	0,105	0,7	0,1		0,9
AE3 (A)	konst.	exp.	$M.A.P.E.$	1,003	0,9	0,1	0,5	
MD1 (A)	prop.	tlum.	$M.A.P.E.$	0,094	0,9	0,1		0,9

konst. = konstantní, prop. = proporcionální, exp. = exponenciální, tlum. = tlumený

12.7.2 Neuronové sítě, jejich kolekce a hybridní modely

Na základě aktuálních hodnot byly přetrénovány sítě a kolekce vícevrstvých dopředných sítí s jednou skrytou vrstvou s 1, 2, 3 a 5 neurony pro modelování celé časové řady simultánně a kolekce vícevrstvých dopředných sítí s jednou skrytou vrstvou s 1, 2, 3, 5 a 8 neurony pro hybridní modely. Přetrénovány byly tedy pouze ty kolekce, jejichž variační koeficienty predikcí nepřevýšily pro první období 10 %. Pro každou architekturu bylo opět natrénováno 80 neuronových sítí. Trénování nyní probíhalo na



Obrázek 12.7: Zaktualizované modely vývoje míry nezaměstnanosti v České republice od října 2008 do prosince 2009 (včetně predikce)

40 vzorech, validace pak na 12 vzorech. Trénování probíhalo prostřednictvím kombinace algoritmů zpětného šíření chyby (backpropagation) a sdružených gradientů. Trénování bylo ukončováno tak, aby validační chyba sítě byla co nejnižší. Vhodnost jednotlivých sítí byla posuzována prostřednictvím hodnot $M.A.E.$ – průměrné, minimální a maximální hodnoty.

Následně byla ze všech sítí stejné architektury (vždy 80 sítí) vytvořena kolekce, která byla použita pro nestrannou predikci. Trénovací a validační schopnost kolekce sítí byla hodnocena prostřednictvím $S.D.$ poměru. V tabulce 12.10 jsou uvedeny výsledky jak pro modely založené čistě na neuronových sítích, tak výsledky pro hybridní modely. Hodnoty uvedené v této tabulce ukazují, že vhodnost modelů měřená prostřednictvím charakteristiky $M.A.E.$ se pro dané architektury neuronových sítí prakticky nezměnila (viz tabulka 12.5 a 12.7 na str. 156 a 159).

Tabulka 12.11 oproti tabulce 12.6 (str. 158) a tabulce 12.8 (str. 159) ukazuje, že variabilita predikcí pro srovnatelné období je nižší. Pravděpodobně může být tento pokles způsoben buď tím, že trénovací množina je bohatší o dva vzory, nebo samotným vývojem míry nezaměstnanosti. Tabulka 12.11 oproti tabulce 12.6 (str. 158) a tabulce 12.8 (str. 159) obsahuje v sloupcích březen a duben vypočtené variační koeficienty predikcí pro období zařazené do trénování neuronových sítí.

Teoretické hodnoty a predikce míry nezaměstnanosti v České republice od října 2008

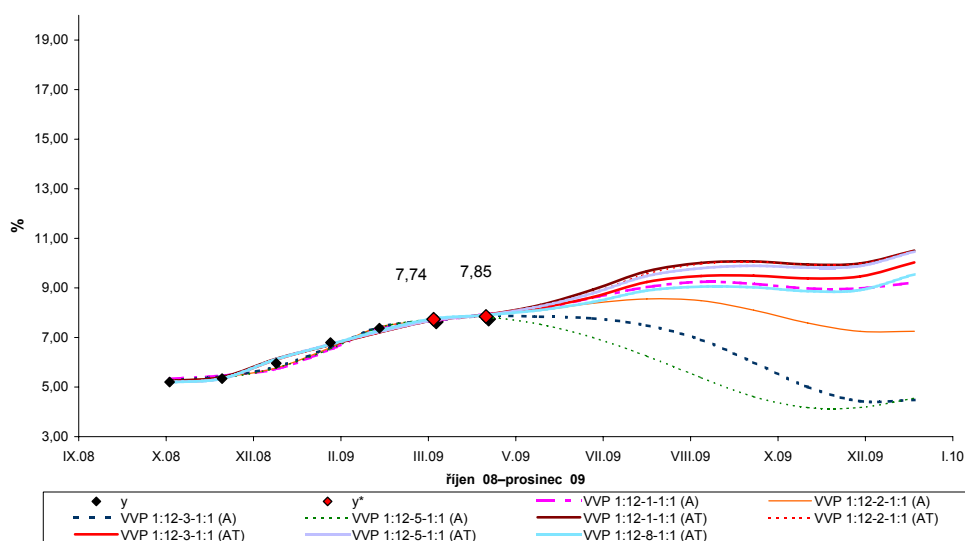
Tabulka 12.10: Výkonnost vybraných modelů pro jejichž tvorbu byly použity i hodnoty březnové a dubnové míry nezaměstnanosti pro modelování trendu míry nezaměstnanosti 2004–2009

Síť	M.A.E.				S.D. poměr	
	prům.	min.	max.	kol.	Trén.	Valid.
VVP 1:12-1-1:1 (A)	0,171	0,107	1,24	0,12	0,114	0,169
VVP 1:12-2-1:1 (A)	0,119	0,07	0,217	0,099	0,101	0,121
VVP 1:12-3-1:1 (A)	0,074	0,051	0,162	0,088	0,09	0,109
VVP 1:12-5-1:1 (A)	0,102	0,046	0,207	0,068	0,081	0,134
VVP 1:12-1-1:1 (AT)	0,099	0,05	1,21	0,066	0,086	0,096
VVP 1:12-2-1:1 (AT)	0,079	0,056	0,152	0,059	0,073	0,078
VVP 1:12-3-1:1 (AT)	0,073	0,044	0,136	0,054	0,067	0,073
VVP 1:12-5-1:1 (AT)	0,07	0,048	0,132	0,049	0,064	0,072
VVP 1:12-8-1:1 (AT)	0,07	0,04	0,161	0,045	0,06	0,09

Tabulka 12.11: Variabilita predikcí v závislosti na počtu neuronů ve skryté vrstvě

Síť	Variační koeficient (%)				
	březen	duben	květen	červen	červenec
VVP 1:12-1-1:1 (A)	3,21	3,6	4,75	6,07	6,97
VVP 1:12-2-1:1 (A)	1,5	2,37	6,16	11,62	16,54
VVP 1:12-3-1:1 (A)	1,32	1,73	5,68	13,16	24,94
VVP 1:12-5-1:1 (A)	1,53	1,77	6,07	15,93	32,75
VVP 1:12-1-1:1 (AT)	2,27	2,48	3,63	4,25	4,71
VVP 1:12-2-1:1 (AT)	1,69	1,57	3,86	6,61	9,43
VVP 1:12-3-1:1 (AT)	1,54	1,67	5,6	11,79	18,03
VVP 1:12-5-1:1 (AT)	1,37	1,37	4,54	10,11	14,44
VVP 1:12-8-1:1 (T)	2,97	4,43	9,08	18,17	26,46

do prosince 2009 jsou zobrazeny na grafu na obrázku 12.8. Konkrétní teoretické a predikované hodnoty pro období ledna 2009–srpna 2009 jsou tabelovány v tabulce 12.13 na str. 166.



Obrázek 12.8: Zaktualizované modely vývoje míry nezaměstnanosti v České republice od října 2008 do prosince 2009 (včetně predikce)

12.8 Závěr

V kapitole 12 byly analyzovány možnosti neuronových sítí pro predikce v časových řadách. Bylo prokázáno, že i pro poměrně krátkou časovou řadu, lze natrénovat neuronové sítě tak, aby v porovnání s klasickými statistickými postupy dávaly srovnatelné nebo dokonce lepší výsledky. Pro nejlepší model založený na exponenciálním vyrovnání – model konstantní sezónnosti s tlumeným trendem – je $M.A.E. = 0,099$ a $U = 0,002$, pro nejlepší model založený pouze na kolekci neuronových sítí – VVP 1:12-1-1:1 – je $M.A.E. = 0,14$ a $U = 0,0062$. Konečně nejlepší hybridní model založený na odhadu trendu prostřednictvím kolekce neuronových sítí – VVP 1:12-3-1:1 – a sezónních složek metodou mechanického vyrovnání lze charakterizovat hodnotami $M.A.E. = 0,051$ a $U = 0,007$.¹⁷

Ukazuje se, že pokud byly trénovány neuronové sítě na primárních datech, vykazovaly modely na nich založené vyšší hodnoty charakteristik $M.A.E$ i U , než modely založené na neuronových sítích trénovaných na očištěných datech (např. mechanickým vyrovnáním), k nimž byla sezónnost ex post přidána. Lze proto pro poměrně krátkou časovou řadu doporučit spíše druhý přístup k řešení problému a nespolehat se

¹⁷Kolekce založené na sítích typu VVP 1:12-5-1:1 a VVP 1:12-8-1:1 dosáhly lepších predikčních výkonů, ale s ohledem na spolehlivost predikce byl zvolena kolekce ze sítě VVP 1:12-3-1:1.

na to, že neuronová síť bude schopna modelovat všechny složky simultánně. Nejlepší výsledky jak z hlediska hodnoty $M.A.E$ ($M.A.E. < 0,08$),¹⁸, tak hodnoty Theilova koeficientu nesouladu ($U < 0,01$) přineslo použití neuronových sítí pro modelování trendu sezónně očištěné řady. Pro ilustraci jsou predikované hodnoty vytvořené prostřednictvím nejkvalitnějších modelů na dvě období uvedeny v souhrnné tabulce 12.12 a 12.13 na následujících stranách.

Tabulka 12.12: Predikce míry nezaměstnanosti prostřednictvím použitých modelů – exponenciální vyrovnání

	I.09	II.09	III.09	IV.09	V.09	VI.09	VII.09
Míra nezaměstnanosti	6,796	7,377	7,738	7,850			
AE1	6,194	7,014	7,642	7,824	8,203	8,860	9,849
AE1 (A)	6,183	6,988	7,625	7,936	8,233	8,889	9,876
AD1	6,217	7,183	7,701	7,819	8,042	8,456	9,114
AD1 (A)	6,153	7,287	7,735	7,877	8,075	8,490	9,149
AE2	6,224	7,150	7,795	8,139	8,707	9,582	10,823
AE2 (A)	6,196	7,048	7,706	7,998	8,256	8,938	9,949
AD2	6,206	7,068	7,634	7,712	7,899	8,278	8,906
AD2 (A)	6,172	7,040	7,626	7,850	8,079	8,504	9,168
AE3	6,239	7,268	7,888	8,331	9,014	10,027	11,430
AE3 (A)	6,188	7,051	7,708	7,999	8,258	8,938	9,949
MD1	6,347	7,086	7,580	7,626	7,784	8,118	8,725
MD1 (A)	6,396	7,237	7,680	7,881	8,076	8,473	9,152

hodnoty teoretické (černá barva),

hodnoty predikující míru nezaměstnanosti na dvě období dopředu (červená barva)

hodnoty predikující do delšího časového horizontu (šedá barva)

Dále bylo experimentálně ověřeno, že dostatečně bohatá struktura sítě vedoucí k lepším hodnotám charakteristiky $M.A.E$. než v případě optimálně použitého exponenciálního vyrovnání způsobuje nestabilní výsledky. Nestranné odhady lze i v tomto případě získat prostřednictvím kolekce sítí. Jednoduché síťové architektury (do tří skrytých neuronů) použité pro modelování celé časové řady a z nich vytvořené kolekce sice nejsou tak kvalitní (dle hodnot charakteristik $M.A.E. \in \langle 0,088; 0,14 \rangle$), ale dávají kvalitní predikce ($U < 0,02$). Do delšího časového horizontu však podobně

¹⁸VVP 1:12-1-1:1 (T), 1:12-15-1:1 (T) a 1:12-3-2-1:1 (T) nebyly do srovnání vybrány.

Tabulka 12.13: Predikce míry nezaměstnanosti prostřednictvím použitých modelů – neuronové sítě

	I.09	II.09	III.09	IV.09	V.09	VI.09	VII.09
Míra nezaměstnanosti	6,796	7,377	7,738	7,850			
VVP 1	6,744	7,404	7,407	7,997	9,048	10,416	10,700
VVP 2	6,774	7,390	7,390	7,047	5,888	4,785	3,799
VVP 1:12-1-1:1	6,527	7,363	7,619	7,776	8,021	8,533	8,980
VVP 1:12-1-1:1 (A)	6,517	7,372	7,669	7,933	8,161	8,607	9,030
VVP 1:12-1-1:1 (T)	6,710	7,192	7,548	7,840	8,301	8,898	9,661
VVP 1:12-1-1:1 (AT)	6,701	7,209	7,697	7,918	8,311	8,929	9,672
VVP 1:12-2-1:1	6,519	7,285	7,412	7,234	7,038	7,024	7,060
VVP 1:12-2-1:1 (A)	6,593	7,422	7,691	7,934	8,095	8,381	8,555
VVP 1:12-2-1:1 (T)	6,701	7,192	7,544	7,769	8,121	8,612	9,257
VVP 1:12-2-1:1 (AT)	6,713	7,186	7,676	7,949	8,295	8,813	9,568
VVP 1:12-3-1:1	6,648	7,378	7,596	7,481	7,169	6,839	6,399
VVP 1:12-3-1:1 (A)	6,605	7,417	7,692	7,869	7,838	7,769	7,484
VVP 1:12-3-1:1 (T)	6,713	7,220	7,647	7,983	8,392	8,906	9,532
VVP 1:12-3-1:1 (AT)	6,717	7,219	7,707	7,913	8,198	8,628	9,238
VVP 1:12-5-1:1	6,717	7,390	7,483	7,010	6,207	5,408	4,710
VVP 1:12-5-1:1 (A)	6,539	7,404	7,747	7,815	7,531	7,002	6,246
VVP 1:12-5-1:1 (T)	6,708	7,218	7,612	7,884	8,350	8,923	9,611
VVP 1:12-5-1:1 (AT)	6,730	7,223	7,699	7,907	8,259	8,783	9,479
VVP 1:12-8-1:1	6,760	7,355	7,340	6,695	5,661	4,700	4,083
VVP 1:12-8-1:1 (T)	6,736	7,198	7,588	7,836	8,327	9,042	9,819
VVP 1:12-8-1:1 (AT)	6,719	7,280	7,763	7,900	8,106	8,430	8,893
VVP 1:12-15-1:1	6,621	7,371	7,669	7,456	6,712	5,662	4,692
VVP 1:12-15-1:1 (T)	6,706	7,336	8,065	8,953	10,032	10,887	11,263
VVP 1:12-3-2-1:1	6,644	7,383	7,518	7,306	6,889	6,531	6,127
VVP 1:12-3-2-1:1 (T)	6,667	7,247	7,753	8,256	8,751	9,135	9,623

hodnoty teoretické (černá barva),

hodnoty predikující míru nezaměstnanosti na dvě období dopředu (červená barva)

hodnoty predikující do delšího časového horizontu (šedá barva)

jako modely založené na exponenciálním vyrovnaní¹⁹ predikují příliš nekonzistentně.

¹⁹Odlíšnost predikcí pro exponenciální a tlumený trend.

Vzhledem k tomu, že modelování v časových řadách je dynamický proces, bylo na závěr nutné aktualizovat nejlepší modely prezentované v kapitole 12. Jejich vhodnost se projeví až s odstupem času, nicméně, protože jsou založeny na modelech vysoce funkčních, lze očekávat, že i tyto modely budou kvalitní. Aktualizované včetně predikcí jsou analyzovány v podkapitole 12.7 a výsledky zařazeny do tabulky 12.13.

Kapitola 13

Závěr

Předkládaná disertační práce se zaměřuje na prozkoumání možností uplatnění neuronových sítí a jejich kolekcí při analýze ekonomických dat. Má spíše aplikační charakter. Za nejdůležitější vlastní přínosy disertační práce lze považovat tyto:

1. Přehled literatury zabývající se problematikou neuronových sítí

Byl vybudován ucelený přehled dostupné literatury a zdrojů z teorie a aplikací neuronových sítí. V práci je uvedeno více než 75 titulů. V rámci popisu vývoje oboru (podkapitola 3.3) a konkretizování využití neuronových sítí (kapitola 7) jsou uvedeny odkazy jak na základní literaturu, tak na vědecké články, které dokumentují současný stav řešené problematiky. Literární rešerše dokazuje, že neuronové sítě jsou stále moderním nástrojem se širokým spektrem aplikačních možností. Vývoj neuronových sítí rozhodně ještě nebyl ukončen. V posledních letech mnoho prací přináší vylepšení a úpravy stávajících architektur a topologií neuronových sítí, či trénovacích algoritmů. Díky možnostem současné výpočetní techniky vznikají dokonce i nové přístupy k tvorbě struktur neuronových sítí. Přesto neuronové sítě představují dosud ne plně využitý nástroj k řešení reálných problémů.

2. Přehled softwarů s implementovanou neuronovou sítí

Z nabídky softwarů s implementovanou neuronovou sítí byla vybrán a uveden jak přehled programů, které jsou dostupné bezplatně, tak přehled programů, které jsou čistě komerční. Přehled je uveden v kapitole 6. Tři programové produkty – Neurex, NetVisualiser a zejména STATISTICA Neuronové sítě – jsou popsány podrobněji, neboť sloužily jako nástroje pro výpočty v předkládané disertační práci a v publikacích s ní spojených. Peníze na nákup profesionálního

softwaru STATISTICA Neuronové sítě 7.1 získal předkladatel disertační práce díky grantu č. 14/06 *Možnosti neuronových sítí a data miningu pro hodnocení faktorů harmonizace regionálního rozvoje sociálně ekonomického potenciálu regionů*, který podal a obhájil v roce 2006. Z tohoto grantu byly také zakoupeny knihy Haykin [25] a Wang – Fu [94], které se bezprostředně věnují problematice neuronových sítí.

3. Navržení expertního systému založeného na neuronovém učení

V kapitole 8 byl navržen a vytvořen prázdný expertní systém, který byl konkrétním způsobem rozvinut v podkapitole 11.5 a v omezeném rozsahu zrealizován při určování bonity žadatele o spotřebitelský úvěr.

4. Prozkoumání možností uplatnění neuronových sítí pro analýzu ekonomických dat

Plnění cílů 2–4 nebylo s ohledem na logické vazby oddělováno – probíhalo simultánně – a proto i závěry jsou na tomto místě prezentovány společně. Kapitola 4 a 5 vytváří teoretický základ pro analýzu ekonomických dat prostřednictvím neuronových sítí. S ohledem na šíři problematiky byla pozornost věnována především dopředným neuronovým sítím vhodným k tzv. učení s učitelem nebo posilujícímu učení. V této souvislosti byly představeny relevantní algoritmy – backpropagation a genetické učení neuronových sítí. Předkladatel disertační práce sjednotil symboliku, do této symboliky přetransformoval vzorce pro metodu zpětného šíření chyby (backpropagation), navrhl vhodnostní transformace pro genetické učení a spojení genetických algoritmů pro trénování neuronové sítě.

Jádro disertační práce tvoří kapitoly 10–12, v nichž jsou ukázány, komentovány a analyzovány možnosti neuronových sítí pro analýzu ekonomických dat. Jedná se zejména o diskriminační schopnosti neuronových sítí typu: *vícevrstvá dopředná síť*, *pravděpodobnostní neuronová síť* a *RBF síť* – tj. sítě vhodné pro učení s učitelem a predikční schopnosti *vícevrstvé dopředné sítě*, použité pro modelování vývoje v časových řadách. Aplikační část práce se opírala o část teoretickou a zabývala se 1) prozkoumáním možností uplatnění neuronových sítí a jejich kolekcí, 2) porovnáním jejich výkonnosti s výkonností klasických statistických metod, 3) kritickým komentováním práce neuronových sítí a jejich kolekcí a 4) vytvořením závěrů k samotným analyzovaným problémům (viz podkapitoly 10.5, 11.4 a 12.8).

5. Navržení hybridního přístupu pro modelování v časových řadách

Přestože neuronové sítě byly vždy používány po předchozí analýze založené na výsledcích statistických metod a v dílčí krocích trénování samy implementovaly klasické statistické postupy, byl pro modelování v časových řadách navržen hybridní přístup. Ten v sobě kombinuje použití mechanického vyrovnání pro odhad sezónní složky a kolekcí sítí pro nestranný odhad a predikci trendu. Díky tomu bylo možno stabilně modelovat relativně krátkou časovou řadu se sezónní složkou.

6. Potvrzení/zpochybnění obecně přijímaných paradigmat o způsobu práce neuronových sítí

Experimentálně – avšak na reálných nikoli simulovaných datech – byly ověřeno, jaké problémy způsobuje přeúčtení a přetrénování neuronových sítí. Také bylo ukázáno, že ne všechny metody určené k zabránění přeúčtení vedou ke kýženým cílům a že přestože neuronové sítě nejsou samospásným řešením všech problémů dokáží být významným konkurentem klasickým statistickým metodám v úlohách klasifikace a predikce v časových řadách. V rámci řešení byly ověřeny možnosti použití například ztrátové matice a kolekcí sítí.

S některými výsledky prezentovanými v předkládané disertační práci byla odborná veřejnost seznámena prostřednictvím těchto prací (řazeny podle pořadí vzniku): Čermáková – Biskup [115], Biskup – Čermáková [110], Biskup – Čermáková [111], Biskup – Čermáková [109], Biskup – Čermáková [112], Biskup – Čermáková [113], Biskup [108], Čermáková – Biskup [117], Čermáková – Biskup [116], Biskup – Čermáková [114], Šulista et al. [120], Čermáková – Biskup [118] a Čermáková – Biskup [119].

Hlavním cílem disertační práce bylo prozkoumání možností uplatnění neuronových sítí a jejich kolekcí při analýze ekonomických dat. Na základě výše uvedených přínosů je možno považovat tento cíl za splněný.

Použité zdroje

- [1] AL-ALAWI, S. – ZAIBET, L. Forecasting Fish Exports in the Sultanate of Oman Using Artificial Neural Networks. *International Refereed Research Journal Dirasat, Agricultural Sciences*, 27(2):s. 190–198, 2000.
- [2] BERKA, P. – JIRKŮ, P. – VEJNAROVÁ, J. *Expertní systémy*. Praha: VŠE Praha, 1998. ISBN 80-7079-873-4.
- [3] BETINEC, M. Použití ROC křivek pro hodnocení klasifikátorů. In ANTOCH, J. – DOHNAL, G. (Ed.) *ROBUST 2006*, s. 25–34, Praha, 2006 [cit. 2008-12-14]. Jednota českých matematiků a fyziků. Dostupné z URL: <<http://www.karlin.mff.cuni.cz/~antoch/robust06/robust2006.pdf>>. ISBN 80-7015-073-4.
- [4] *BrainMaker – BrainMaker Neural Network Software*. California Scientific [online], 2009 [cit. 2008-12-14]. Dostupné z URL: <<http://www.calsci.com/BrainIndex.html>>.
- [5] *BRKI – bankovní registr klientských informací*. [online], [cit. 2008-12-14]. Dostupné z URL: <<http://www.brki.cz/>>.
- [6] BUCHANAN, B. G. – FEIGENBAUM, E. A. Dendral and meta-dendral: Their applications dimension. *Artificial Intelligence*, 11(1–2): s. 5–24, 1978.
- [7] CARPENTER, G. A. – GROSSBERG, S. ART 2: Stable Self-organization of Pattern Recognition Codes for Analog Input Patterns. *Applied Optics*, 26: s. 4919–4930, 1987.
- [8] CERD System. *CERDTM – Centrální registr dlužníků*. [online], [cit. 2008-12-14]. Dostupné z URL: <<http://www.centralniregistrdluzniku.cz/>>.

- [9] *Computational Intelligence Group*. [online], 2006 [cit. 2008-12-14]. Dostupné z URL: <<http://ncg.felk.cvut.cz/>>.
- [10] CzechTrade. *Centrální registr úvěrů*. BusinessInfo.cz – Oficiální portál pro podnikání a export. [online], 2003–2009 [cit. 2008-12-14]. Dostupné z URL: <<http://www.businessinfo.cz/cz/clanek/bankovnictvi/centralni-registr-uveru/1000464/4014/>>.
- [11] Český statistický úřad. *ČSÚ – Český statistický úřad*. [online], 2009 [cit. 2008-12-14]. Dostupné z URL: <<http://www.czso.cz/>>.
- [12] David Rumelhart and James McClelland (Ed.). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition I, II*, Cambridge, MA, 1986. MIT Press.
- [13] *EasyNN-plus*. Neural Planner Software Ltd. [online], 2002 [cit. 2008-12-14]. Dostupné z URL: <<http://www.easynn.com/>>.
- [14] EFRON, B. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1): s. 1–26, 1979. Dostupné z URL: <<http://www.math.tau.ac.il/~yekutieli/MA%20seminar/Efron%201979.pdf>>.
- [15] FANTA, P. *Neuronové sítě ve společenských vědách*. Praha: Karolinum, 2000. ISBN 80-246-0175-3.
- [16] FIELDSEND, J. E. – SINGH, S. Pareto Evolutionary Neural Networks. *IEEE Transactions On Neural Networks*, 16(2): s. 338–354, March 2005.
- [17] FUJII, Y. et al. A Study on the Development of a System for Optimal Distribution of Vegetables and Fruits by Applying a Recurrent Neural Network. *Bulletin of the Faculty of Agriculture Tottori University*, 49: s. 79–85, 1996.
- [18] FUKUSHIMA, K. Cognitron: A Self-Organizing Multilayered Neural Network. *Biological Cybernetics*, 20(3–4): s. 121–136, September 1975.
- [19] FUKUSHIMA, K. – MIYAKU, S. Neocognitron: A New Algorithm for Pattern Recognition Tolerant of Deformations and Shifts in Position. *Pattern Recognition*, 15(6): s. 455–469, 1982.

- [20] GARCIA, M. P. – KIRSCHEN, D. S. Forecasting System Imbalance Volumes in Competitive Electricity Markets. *IEEE Transactions on Power Systems*, 21 (1): s. 240–248, February 2006.
- [21] GROSSBERG, S. Behavioral Contrast in Short-term Memory: Serial Binary Memory Models or Parallel Continuous Memory Models? *Journal of Mathematical Psychology*, 3: s. 199–219, 1978.
- [22] HAKL, F. – HOLEŇA, M. *Úvod do teorie neuronových sítí*. Praha: Ediční středisko ČVUT Praha, 1997. Dostupné z URL: <http://neuron.tuke.sk/math.chtf.stuba.sk/pub/vlado/NN_books_texts/Hakl_Holena_Skripta_FJFI_CVUT.pdf>. ISBN 80-01-01716-8.
- [23] HAMM, L. – BRORSEN, B. W. Forecasting Hog Prices with a Neural Network. *Journal of Agribusiness*, 15(1): s. 37–54, 1997.
- [24] HANZELKA, D. Bayesovská umělá inteligence. In *Kognícia a umelý život V – Slovensko-český seminár o kognícii a umelom živote*, s. 1–15, 2005 [cit. 2008-12-14]. Dostupné z URL: <<http://hilbert.chtf.stuba.sk/KUZV/download/kuzv-hanzelka.pdf>>.
- [25] HAYKIN, S. *Neural Networks: A Comprehensive Foundation*. Second Edition. Practise Hall: Tom Robbins, 1999. ISBN 0-13-273350-1.
- [26] HEBÁK, P. – HUSTOPECKÝ, J. *Vícerozměrné statistické metody s aplikacemi*. : Praha: SNTL – Nakladatelství technické literatury, 1 edition, 1987. ISBN 99-00-02139-X.
- [27] HEBB, D. O. *The Organization of Behavior: A Neuropsychological Theory*. NewYork: Wiley-Interscience, 1949. ISBN 0805843000.
- [28] HEGER, M. – DAVID, J. NEURONEK – program pro výuku neuronových sítí. In *XXVI. Seminár ASR 2001 – Instrumets and Control*, Ostrava, 2001. Dostupné z URL: <<http://www.fs.vsb.cz/akce/2001/asr2001/Proceedings/papers/25.pdf>>. ISBN 80-7078-890-9.
- [29] HECHT-NIELSEN, R. Counterpropagation Networks. In Caudill & Butler (Ed.) *Proceedings of the IEEE First International Conference on Neural Networking*. IEEE Press, 1987.

- [30] HOLEŇA, M. *Matematické základy umělých neuronových sítí*. Studijní text pro posluchače 5. ročníku oboru MI, prosinec 1996.
- [31] HOPFIELD, J. J. – TANK, D. W. 'Neural' Computation of Decisions in Optimization Problems. *Biological Cybernetics*, 52: s. 141–152, 1985.
- [32] HORNIK, K. – STINCHCOMBE, M. – WHITE, H. Multilayer Feedforward Networks Are Universal Approximators. *Neural Networks*, 2: s. 359–366, 1989.
- [33] HRISTEV, R. M. Matrix Techniques in Artificial Neural Networks. Thesis for the degree of master of science in mathematics in the university of canterbury in 2000, University of Canterbury, 1.26a 2001 [cit. 2008-12-14]. Dostupné z URL: <www.scilab.org/publications/ANN/Matrix_ANN_1.26a.ps.gz>.
- [34] CHMELENSKÝ, J. Neuronová síť pro rozpoznávání obrazců částečných výbojů. Diplomová práce, ČVUT, Fakulta elektrotechnická, Katedra elektroenergetiky, 2003.
- [35] *IEEE Transactions on Neural Networks*. IEEE [online], 2008 [cit. 2008-12-14]. Dostupné z URL: <<http://ieee-cis.org/pubs/tnn/>>.
- [36] *International Neural Network Society*. Posytron Engineering srl. [online], 2003 [cit. 2008-12-14]. Dostupné z URL: <<http://www.inns.org/>>.
- [37] JOHNSON, M. *Stephen Grossberg – webpages*. [online], 2008 [cit. 2008-12-14]. Dostupné z URL: <<http://cns.bu.edu/Profiles/Grossberg.html/>>.
- [38] KELEMEN, J. *Od umělých neuronů ke počítačové emotivitě (idea decentralizace v díle Marvina Minskyho)*, 2007.
- [39] KOHONEN, T. Self-organized Formation of Topologically Correct Feature Maps. *Biological Cybernetics*, 43(1): s. 59–69, 1982.
- [40] KOHONEN, T. *Self-Organizing Maps*. 30 / *Springer Series in Information Sciences*. Berlin, Heidelberg: Springer, 1995, 1997, 2001.
- [41] KRIESEL, D. *Ein kleiner Überblick über Neuronale Netze*. Bonn: Rheinische Friedrich-Wilhelms-Universität Bonn, 2007 [cit. 2008-12-14]. Dostupné z URL: <<http://filestorage.dkriesel.com/neuronalenetze-de-delta-dkrieselcom.pdf>>.

- [42] LACHER, R. C. et al. A Neural Network for Classifying the Financial Health of a Firm. *European Journal of Operational Research*, s. 53–65, 1995.
- [43] Leasing And Loan Credit Bureau. *NRKI – Nebankovní registr clientských informací*. [online], [cit. 2008-12-14]. Dostupné z URL: <<http://www.llcb.cz/>>.
- [44] LI, K. – CUI, L.-J. – ZHANG, Y.-F. A Modified Fuzzy Clustering Based on Multisynapse Neural Network. In *Proceedings of the Fourth International Conference on Machine Learning and Cybernetics*, s. 1665–1669, Guangzhou, 2005.
- [45] MAINLAND, D. D. Econometrics or Neural networks: A Case Study of Marketing Margins in the Meat Industry in the UK. In H. Kure & I. Thysen & A. R. Kristensen (Ed.) *First European conference for information technology in agriculture*, Copenhagen (Denmark), 1997.
- [46] MARKIEWICZ, A. O wykorzystaniu sieci neuronowych w badaniu kondycji finansowej przedsiębiorstw (Neural Networks Application in Research of Financial Condition Research). *Folia Universitatis Agriculturae Stetinensis. Oeconomica*, s. 225–232, 2002.
- [47] MARTĚÁN, P. *Stránky Petra Marťána pro demonstraci schopností a principů neuronových sítí a genetických algoritmů*. [online], 2001 [cit. 2008-12-14]. Dostupné z URL: <<http://neuralnets.czweb.org/index.htm>>.
- [48] MARTĚÁN, P. Vizualizace vybraných metod strojového učení. Diplomová práce, Masarykova univerzita, Fakulta informatiky, Katedra elektroenergetiky, 2002 [cit. 2008-12-14]. Dostupné z URL: <<http://neuralnets.czweb.org/dpPetrMartan2002.pdf>>.
- [49] *Massachusetts Institute of Technology*. [online], 2008 [cit. 2008-12-14]. Dostupné z URL: <<http://www.mit.edu/>>.
- [50] MASTNÝ, V. Modelování intradenního vývoje devizových kurzů pomocí neuronových sítí. In TOMŠÍK, P. (Ed.) *Firma a konkurenční prostředí 2003*, s. 173–181, Brno, 2003. KONVOJ s.r.o. ISBN 80-7157-695-6.
- [51] MCCULLOCH, W. S. – PITTS, W. A Logical Calculus of the Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics*, 5: s. 115–133, 1943.

- [52] MCLACHLAN, G. J. *Discriminant analysis and statistical pattern recognition*. : John Wiley & Sons (14 May 1992), 1992. ISBN 0471615315.
- [53] Ministerstvo financí. *Webové stránky Ministerstva financí České republiky*. [online], 2005–2009 [cit. 2008-12-14]. Dostupné z URL: <<http://www.mfcr.cz/cps/rde/xchg/>>.
- [54] MINSKY, M. L. *Theory of Neural-Analog Reinforcement System and its Applications to the Brain-Model Problem*. Disertační práce, University of Princeton, Princeton, NJ, 1954.
- [55] MINSKY, M. L. – PAPERT, S. A. *Perceptrons*. 292 Main Street, Cambridge, MA 02142: The MIT Press, 1 edition, 1969.
- [56] MRÁZ, M. – HÁNA, K. – ŠNOREK, M. Rozpoznávání bdělostního stavu (úrovně pozornosti) člověka ze signálu EEG pomocí neuronové sítě RBF. In *Intelligentní systémy pro praxi*, s. 23–24. Ostrava: ADaM, 2004. ISBN 80-239-3680-8.
- [57] NĚMEC, M. *Predikce potřeby vody pomocí neuronových sítí*. Teze disertace k získání akademického titulu „doktor“, ve zkratce ”Ph.D.”, 2005 [cit. 2008-12-14]. Dostupné z URL: <http://kzei.fsv.cvut.cz/pdf/dis_nemec_2006_cz.pdf>.
- [58] *Neural Computation*. Atypion Systems, Inc. [online], 2008 [cit. 2008-12-14]. Dostupné z URL: <<http://www.mitpressjournals.org/loi/neco>>.
- [59] *Neural Network World*. Acad Sciences [online], 2008 [cit. 2008-12-14]. Dostupné z URL: <<http://www.nnw.cz/>>. ISSN 1210-0552.
- [60] *Neural Networks*. Posytron Engineering srl. [online], 2003 [cit. 2008-12-14]. Dostupné z URL: <<http://www.inns.org/nnjournal.asp>>.
- [61] *Neuro Dimension*. [online], 2009 [cit. 2008-12-14]. Dostupné z URL: <<http://www.nd.com/>>.
- [62] NILSON, N. J. *Learning Machines*. New York: McGraw-Hill, 1965.
- [63] NOGUCHI, N. – ISHII, K. – TERAOKA, H. Optimal Control of Agricultural Vehicles by Neural Networks, 2: Optimal Schedule of Working Route by Use of

- Genetic Algorithms. *Journal of the Japanese Society of Agricultural Machinery*, 56(2):83–92, 1994.
- [64] NOVÁK, M. – FABER, J. – KUFUDAKI, O. *Neuronové sítě a informační systémy živých organismů*. Dlouhá 39, Praha 1 – 110 00: Grada a.s., 1 edition, 1993. ISBN 80-85424-95-9.
- [65] OBITKO, M. *Interactive tutorials, mainly from the area of artificial intelligence*. [online], 1999 [cit. 2008-12-14]. Dostupné z URL: <<http://www.obitko.com/tutorials/>>.
- [66] OKsystem. *MPSV – Integrovaný portál Ministerstva práce a sociálních věcí*. [online], 2002–2009 [cit. 2008-12-14]. Dostupné z URL: <<http://portal.mpsv.cz/>>.
- [67] OKsystem. *Statistiky nezaměstnanosti – Integrovaný portál Ministerstva práce a sociálních věcí – Integrovaný portál Ministerstva práce a sociálních věcí*. [online], 2002–2009 [cit. 2008-12-14]. Dostupné z URL: <<http://portal.mpsv.cz/sz/stat/nz/mes/>>.
- [68] OSVAAG, M. K. Registr dlužníků: černý puntík na několik let. [online], 2007 [cit. 2008-12-14]. Dostupné z URL: <<http://www.penize.cz/18855-registr-dluzniku-cerny-puntik-na-nekolik-let>>.
- [69] PAPRZYCKI, M. – BOWERS, S. – COSTEINES, A. Applying Probabilistic Neural Networks to the Multifont Recognition Problem with Large Training Set. In KUMAR, A. N. – RUSSELL, I. (Ed.) *Proceedings of the Twelfth International Florida Artificial Intelligence Research Society Conference*, s. 336–339. AAAI Press, 1999 [cit. 2008-12-14]. Dostupné z URL: <<http://www.aaai.org/Papers/FLAIRS/1999/FLAIRS99-060.pdf>>. ISBN 1-57735-080-4.
- [70] PATTERSON, D. *Artificial Neural networks – Theory and Applications*. Prentice Hall, 1996.
- [71] PETRÁSEK, J. Metody bootstrap pro závislá pozorování. Diplomová práce, Univerzita Karlova v Praze, Matematicko-fyzikální fakulta, 2008 [cit. 2008-12-14]. Dostupné z URL: <http://artax.karlin.mff.cuni.cz/~petrj3am/texts/Diploma_thesis.pdf>.

- [72] PRSKAVEC, L. Neuronová síť pro rozpoznávání obrazců částečných výbojů. Diplomová práce, ČVUT, Fakulta elektrotechnická, Katedra elektroenergetiky, 2001.
- [73] REPÍKOVÁ, K. Použití ROC křivek při klasifikaci ekonomických objektů. Diplomová práce, Masarykova univerzita, Přírodovědecká fakulta, 2006 [cit. 2008-12-14]. Dostupné z URL: <http://is.muni.cz/th/152278/prif_m/Diplomova_prace.pdf?lang=en>.
- [74] ROSENBLATT, F. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review*, 65(6): s. 386–408, 1958.
- [75] RUMELHART, D. – HINTON, G. – WILLIAMS, R. Learning Internal Representations by Error Propagation. In David Rumelhart & James McClelland (Ed.) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, I, s. 318–362, Cambridge, MA, 1986. MIT Press.
- [76] SAS Institute *SAS® Enterprise Miner™ – Neural Networks*. [online], 2009 [cit. 2008-12-14]. Dostupné z URL: <<http://www.sas.com/technologies/analytics/datamining/miner/neuralnet/index.html>>.
- [77] SATO, K. – HOKI, M. – SALOKHE, V. M. Voice Recognition by Neural Network under Tractor Noise. *Transactions of the ASAE*, 36(4): s. 1223–1227, 1993.
- [78] SOLUS. *SOLUS – Zájmové sdružení právnických osob*. [online], 2008 [cit. 2008-12-14]. Dostupné z URL: <<http://www.solus.cz/>>.
- [79] SPSS. *SPSS Neural Networks™*. [online], 2008 [cit. 2008-12-14]. Dostupné z URL: <http://www.spss.com/statistics/neural_networks/>.
- [80] StatSoft CR. *StatSoft CR*. [online], 2004 [cit. 2008-12-14]. Dostupné z URL: <<http://www.statsoft.cz/page/index.php>>.
- [81] StatSoft. *Electronic Statistics Textbook*. [On-line příručka], 1999 [cit. 2008-12-14]. Dostupné z URL: <<http://www.statsoft.cz/page/index2.php?pg=navigace&nav=31>>.
- [82] STONE, M. Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society*, B36(1): s. 111–133, 1974.

- [83] *Stránky doc. Ing. Miroslava Šnorka, CSc.* [online], 2008 [cit. 2008-12-14]. Dostupné z URL: <http://cs.felk.cvut.cz/webis/people_/snorek.html>.
- [84] *Stránky doc. Ing. Petra Berky, CSc.* [online], [cit. 2008-12-14]. Dostupné z URL: <<http://sorry.vse.cz/~berka/>>.
- [85] SWANSON, N. R. – WHITE, H. A Model Selection Approach to Real-time Macroeconomic Forecasting using Linear Models and Artificial Neural Networks. *Review of Economics and Statistics*, s. 540–550, 1997.
- [86] ŠÍMA, J. – NERUDA, R. *Teoretické otázky neuronových sítí*. Praha: Matfyzpress, 1996. Dostupné z URL: <<http://www.cs.cas.cz/~sima/kniha.pdf>>. ISBN 80-85863-18-9.
- [87] ŠKUTOVÁ, J. Inicializační metody dvouvrstevných dopředných neuronových sítí. In Smutný & Škutová (Ed.) *XXXIII. Seminář ASR '2008 „Instruments and Control“*, s. 307–313. VŠB-TUO, Ostrava, 2008 [cit. 2008-12-14]. Dostupné z URL: <<http://www.fs.vsb.cz/akce/2008/asr2008/proceedings/307.pdf>>. ISBN 978-80-248-1727-9.
- [88] ŠNOREK, M. *Neuronové sítě a neuropočítače*. Praha: Vydavatelství ČVUT, 2002. ISBN 80-01-02549-7.
- [89] The MathWorks. *The MathWorks – Neural Network Toolbox*. [online], 2005 [cit. 2008-12-14]. Dostupné z URL: <<http://www.mathworks.com/products/neuralnet/>>.
- [90] The SNNS Team. *Stuttgart Neural Network Simulator*. [online], 2001 [cit. 2008-12-14]. Dostupné z URL: <<http://www.ra.cs.uni-tuebingen.de/SNNS/>>.
- [91] TURBAN, E. – FRENZEL, L. E. *Expert Systems and Applied Artificial Intelligence*. Prentice Hall Professional Technical Reference, 1992. ISBN 0024216658.
- [92] VESELÝ, A. Logický a neuronový přístup k umělé inteligenci (Logical and Connectionist Approach to Artificial Intelligence). *Zemědělská Ekonomika UZPI*, 46(12): s. 568–574, 2000.
- [93] VONDRÁK, I. *Neurex 4.0 – Expertní systém na bázi neuronových sítí*. Uživatelská příručka, 1993 [cit. 2008-12-14]. Dostupné z URL: <http://skolniftp.sh.cvut.cz/ftp/neroztridene/fel_zlababa.sh>.

- cvut.cz/DES_-_Datab%E1zov%E9_a_expertn%ED_syst%E9my/neurex_manual_kompletni/NEUREX40.DOC>.
- [94] WANG, L. – FU, X. *Data Mining with Computational Intelligence*. Berlin, Heidelberg: Springer-Verlag, 2005. ISBN 3-540-24522-7.
- [95] WERBOS, P. J. *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. Disertační práce, aplikovaná matematika, Harvard University, Cambridge, MA, 1974.
- [96] WIDROW, B. – MARCIAN E. HOFF, J. Generalization and Information Storage in Networks of ADALINE 'Neurons'. *Self Organizing Systems*, s. 435–461, 1962.
- [97] WORZALA, E. – LENK, M. – SILVA, A. An Exploration of Neural Networks and its Application to Real Estate Valuation. *Journal of Real Estate Research*, s. 185–201, 1995.
- [98] *Zákon č. 21/1992 Sb. o bankách*. Ministerstvo vnitra. [online], 2003–2009 [cit. 2008-12-14]. Dostupné z URL: <http://portal.gov.cz/wps/portal/_s.155/701/.cmd/ad/.c/313/.ce/10821/.p/8411/_s.155/701?PC_8411_number1=21/1992&PC_8411_l=21/1992&PC_8411_ps=10#10821>.
- [99] *Zákon č. 36/2008 Sb. změna zákona o ochraně spotřebitele a změna dalších zákonů*. Ministerstvo vnitra. [online], 2003–2009 [cit. 2008-12-14]. Dostupné z URL: <http://portal.gov.cz/wps/portal/_s.155/701/.cmd/ad/.c/313/.ce/10821/.p/8411/_s.155/701?PC_8411_name=obchodn%C3%AD%20z%C3%A1kon%C3%ADk&PC_8411_l=36/2008&PC_8411_ps=25#10821>.
- [100] *Zákon č. 101/2000 Sb. o ochraně osobních údajů a o změně některých zákonů*. Ministerstvo vnitra. [online], 2003–2009 [cit. 2008-12-14]. Dostupné z URL: <http://portal.gov.cz/wps/portal/_s.155/701/.cmd/ad/.c/313/.ce/10821/.p/8411/_s.155/701?PC_8411_name=o%20ochran%C3%AC%20osobn%C3%ADch%20%C3%BAtaj%C3%B9&PC_8411_l=101/2000&PC_8411_ps=10#10821>.
- [101] *Zákon č. 177/2001 Sb. změna zákona o ochraně osobních údajů a zákoníku práce*. Ministerstvo vnitra. [online], 2003–2009 [cit. 2008-12-14]. Dostupné z URL: <http://portal.gov.cz/wps/portal/_s.155/701/.cmd/ad/.c/

313/.ce/10821/.p/8411/_s.155/701?PC_8411_name=o%20ochran%C3%AC%20osobn%C3%ADch%20%C3%BAaj%C3%B9&PC_8411_l=177/2001&PC_8411_ps=10#10821>.

- [102] *Zákon č. 310/2002 Sb. změna zák. o ochraně utajovaných skutečností a někt. dalších zák.* Ministerstvo vnitra. [online], 2003–2009 [cit. 2008-12-14]. Dostupné z URL: <http://portal.gov.cz/wps/portal/_s.155/701/.cmd/ad/.c/313/.ce/10821/.p/8411/_s.155/701?PC_8411_name=o%20ochran%C3%AC%20osobn%C3%ADch%20%C3%BAaj%C3%B9&PC_8411_l=310/2002&PC_8411_ps=10#10821>.
- [103] *Zákon č. 321/2001 Sb. o některých podmínkách sjednávání spotřebitelského úvěru.* Ministerstvo vnitra. [online], 2003–2009 [cit. 2008-12-14]. Dostupné z URL: <http://portal.gov.cz/wps/portal/_s.155/701/.cmd/ad/.c/313/.ce/10821/.p/8411/_s.155/701?PC_8411_number1=321/2001%20Sb.&PC_8411_l=321/2001&PC_8411_ps=10#10821>.
- [104] *Zákon č. 370/2000 Sb. změna obchodního zákoníku.* Ministerstvo vnitra. [online], 2003–2009 [cit. 2008-12-14]. Dostupné z URL: <http://portal.gov.cz/wps/portal/_s.155/701/.cmd/ad/.c/313/.ce/10821/.p/8411/_s.155/701?PC_8411_name=obchodn%C3%AD%20z%C3%A1kon%C3%ADk&PC_8411_l=370/2000&PC_8411_ps=25#10821>.
- [105] *Zákon č. 439/2004 Sb. změna zákona o ochraně osobních údajů.* Ministerstvo vnitra. [online], 2003–2009 [cit. 2008-12-14]. Dostupné z URL: <http://portal.gov.cz/wps/portal/_s.155/701/.cmd/ad/.c/313/.ce/10821/.p/8411/_s.155/701?PC_8411_name=o%20ochran%C3%AC%20osobn%C3%ADch%20%C3%BAaj%C3%B9&PC_8411_l=439/2004&PC_8411_ps=10#10821>.
- [106] *Zákon č. 450/2001 Sb. změny některých zákonů týkajících se veřejné správy.* Ministerstvo vnitra. [online], 2003–2009 [cit. 2008-12-14]. Dostupné z URL: <http://portal.gov.cz/wps/portal/_s.155/701/.cmd/ad/.c/313/.ce/10821/.p/8411/_s.155/701?PC_8411_name=o%20ochran%C3%AC%20osobn%C3%ADch%20%C3%BAaj%C3%B9&PC_8411_l=450/2001&PC_8411_ps=10#10821>.
- [107] *Zákon č. 513/1991 Sb. obchodní zákoník.* Ministerstvo vnitra. [online], 2003–2009 [cit. 2008-12-14]. Dostupné z URL: <http://portal.gov.cz/wps/portal/_s.155/701/.cmd/ad/.c/313/.ce/10821/.p/8411/_s.155/

701?PC_8411_name=0bchodn%C3%AD%20z%C3%A1kon%C3%ADk&PC_8411_l=513/
1991&PC_8411_ps=10#10821>.

Vlastní publikace

- [108] BISKUP, R. Předpovídání bonity klienta pomocí neuronové sítě a diskriminační analýzy. In *Sborník příspěvků z doktorské vědecké konference THINK TOGETHER 2006*, 1, s. 202–206, Praha, únor 2006. Česká zemědělská univerzita v Praze, Provozně ekonomická fakulta. ISBN 80-213-1474-5.
- [109] BISKUP, R. – ČERMÁKOVÁ, A. Genetické učení neuronových sítí. In *Sborník prací z mezinárodní vědecké konference Agrární perspektivy XIII (Trvale udržitelný rozvoj agrárního sektoru – výzvy a rizika)*, II, s. 586–590, Praha, 2004. Česká zemědělská univerzita v Praze, Provozně ekonomická fakulta. ISBN 80-213-1190-8.
- [110] BISKUP, R. – ČERMÁKOVÁ, A. Vícevrstevná síť – učení zpětným šířením chyby. In *Sborník příspěvků studentů DSP z konference s mezinárodní účastí*, II, s. 9–14, České Budějovice, 2005. Jihočeská univerzita v Českých Budějovicích, Zemědělská fakulta. ISBN 80-7040-677-1.
- [111] BISKUP, R. – ČERMÁKOVÁ, A. Neuronové sítě a predikce časových řad. In BISKUP, R. – REMEŠ, R. (Ed.) *Kvantitativní metody v ekonomii 2004*, s. 19–26, České Budějovice, září 2004. Jihočeská univerzita v Českých Budějovicích, Zemědělská fakulta, Katedra aplikované matematiky a informatiky. ISBN 80-7040-691-7.
- [112] BISKUP, R. – ČERMÁKOVÁ, A. Návrh expertního systému založeného na neuronových sítích. In *Kvantitatívne metódy v ekonómii*, s. 12–18, Nitra, prosinec 2004. Slovenská štatistická a demografická spoločnosť. ISBN 80-88946-40-9.
- [113] BISKUP, R. – ČERMÁKOVÁ, A. Regresní analýza a neuronové sítě. *Forum Statisticum Slovacum*. December 2005, 1, 3, s. 106–111. ISSN 1336-7420.

- [114] BISKUP, R. – ČERMÁKOVÁ, A. Využití neuronových sítí při kvantifikaci socioekonomických regionálních dat. In *Sborník prací z mezinárodní vědecké konference Agrární perspektivy XV (Zahraniční obchod a globalizační procesy)*, II, s. 693–697, Praha, 2006. Česká zemědělská univerzita v Praze, Provozně ekonomická fakulta. Dostupné z URL: <<http://www.agris.cz/etc/textforwarder.php?iType=2&iId=152794&PHPSESSID=3e>>. ISBN 80-213-1531-8, MSM 6007665806.
- [115] ČERMÁKOVÁ, A. – BISKUP, R. Základní představa o neuronových sítích. In BISKUP, R. – REMEŠ, R. (Ed.) *Kvantitativní metody v ekonomii 2004*, s. 40–45, České Budějovice, 2004. Jihočeská univerzita v Českých Budějovicích, Zemědělská fakulta, Katedra aplikované matematiky a informatiky. ISBN 80-7040-691-7.
- [116] ČERMÁKOVÁ, A. – BISKUP, R. Porovnání účinnosti dvou statistických metod při analýze vybraných regionálních dat. In *Agroregion 2006 – Zvyšování konkurenceschopnosti v zemědělství*, Sekce 3 / 3, s. 32–36, České Budějovice, 2006. Jihočeská univerzita v Českých Budějovicích, Zemědělská fakulta. ISBN 80-7040-871-5, MSM 6007665806.
- [117] ČERMÁKOVÁ, A. – BISKUP, R. Analýza diskriminačních schopností vybraných ekonomických ukazatelů. *Forum Statisticum Slovacum*. 2006, 1, 3, s. 11–17. ISSN 1336-7420. MSM 6007665806.
- [118] ČERMÁKOVÁ, A. – BISKUP, R. Comparison of Inertia f Statistical Methods for Forecasting of Unemployment Under Conditions of Economic Depression. *Lucrări științifice*. 2009, XI, 1, s. 165–172. ISSN 1453-1410. MSM 6007665806.
- [119] ČERMÁKOVÁ, A. – BISKUP, R. Performance Evaluation of Statistical Methods for Forecasting of Unemployment Rate. *Acta Universitatis Bohemiae Meridionales*. 2009, XII, 2, s. XX–XX. ISSN 1212-3285. MSM 6007665806. V tisku.
- [120] ŠULISTA, M. – BISKUP, R. – REMEŠ, R. Implementace natrénované neuronové sítě. In *Tvorba softwaru 2007*, s. 188–193, Ostrava, červen 2007. VŠB – Technická univerzita Ostrava. ISBN 978-80-248-1427-8, MSM 6007665806.

Příloha A

Soubory dat

A.1 Migrace

Následující tabulka uvádí hodnoty 7 ekonomických ukazatelů použitých pro diskriminaci v kapitole 10:

výš – nadmořská výška obce: nadmořská výška v m n. m.,

hustob – hustota obyvatel: počet obyvatel na 1 ha katastrálního území,

míranez – míra nezaměstnanosti ,

bytnaek – stavební aktivita v obci: počet dokončených bytů na 1 EAO,

zamnaek – zaměstnanecká „absorpce“ obce: počet obyvatel zaměstnaných v obci připadajících na 1 EAO,

fyznaek – podnikatelská aktivita v obci: počet osob samostatně výdělečně činných připadajících na 1 EAO,

rolnaek – význam soukromého zemědělství: počet soukromě hospodařících rolníků připadajících na 1 EAO.

klasifikační proměnnou:

slado – saldo migrace: obce se záporným – *Z-migrace* nezáporným – *N-migrace* saldem migrace,

a přiřazení do jedné z datových množin použitých pro trénování a měření výkonnosti natrénování neuronových sítí:

množina: *testovací*, *trénovací* a *validační* – vysvětlení viz podkapitola 5.5,

Tabulka A.1: Soubor dat použitý pro diskriminaci dle migrace v Jihočeském regionu

výš	míranez	hustob	bytnaek	fyznaek	rolnaek	zamnaek	saldo	mnozina
478	3,990	4,937	0,007	0,359	0,011	0,069	Z-migrace	testovací
445	1,900	0,232	0,000	0,298	0,021	2,404	N-migrace	trénovací
413	2,930	5,975	0,019	0,331	0,010	0,218	N-migrace	testovací
522	3,940	0,883	0,008	0,289	0,014	0,522	N-migrace	trénovací
455	14,890	0,344	0,021	0,319	0,064	0,106	N-migrace	trénovací
413	3,350	1,074	0,109	0,385	0,023	0,856	N-migrace	validační
419	1,830	0,214	0,000	0,101	0,101	0,092	N-migrace	testovací
408	3,950	0,365	0,013	0,303	0,040	0,105	N-migrace	validační
403	4,000	0,108	0,000	0,260	0,160	0,100	N-migrace	validační
450	4,650	0,237	0,012	0,291	0,116	0,256	N-migrace	trénovací
389	1,280	0,299	0,000	0,301	0,038	0,294	N-migrace	validační
381	4,790	17,143	0,004	0,349	0,005	1,257	Z-migrace	validační
550	12,370	0,290	0,000	0,309	0,103	0,753	N-migrace	validační
385	3,540	0,781	0,027	0,398	0,062	0,221	N-migrace	validační
494	4,380	0,277	0,004	0,287	0,008	0,251	N-migrace	testovací
450	4,030	15,025	0,005	0,420	0,008	0,344	N-migrace	trénovací
480	5,450	0,258	0,000	0,218	0,018	0,309	Z-migrace	validační
446	5,670	0,416	0,000	0,299	0,062	0,680	N-migrace	validační
490	1,020	1,119	0,020	0,347	0,010	0,133	Z-migrace	testovací
425	3,800	0,606	0,005	0,277	0,022	0,233	Z-migrace	validační
521	6,320	0,292	0,000	0,379	0,084	0,021	N-migrace	validační
432	5,670	0,302	0,018	0,234	0,066	0,408	N-migrace	validační
476	3,660	0,941	0,031	0,378	0,055	0,549	N-migrace	testovací
410	2,960	0,680	0,021	0,343	0,019	0,359	N-migrace	trénovací
424	5,490	0,242	0,000	0,244	0,067	1,312	Z-migrace	testovací
463	2,130	0,174	0,000	0,298	0,064	0,298	N-migrace	validační
474	4,880	0,191	0,000	0,207	0,134	0,598	Z-migrace	trénovací
472	3,080	0,740	0,015	0,462	0,000	0,493	N-migrace	testovací
403	1,450	0,270	0,000	0,189	0,131	0,218	N-migrace	testovací
550	1,800	0,820	0,027	0,279	0,036	0,045	N-migrace	testovací
394	3,640	0,514	0,013	0,343	0,023	0,623	N-migrace	validační
420	3,580	0,814	0,009	0,340	0,025	0,559	N-migrace	trénovací
493	4,550	0,141	0,000	0,159	0,091	0,569	Z-migrace	validační
543	8,710	0,195	0,012	0,268	0,021	0,381	N-migrace	validační
486	2,650	0,221	0,006	0,268	0,021	0,533	N-migrace	validační
442	7,890	0,187	0,000	0,250	0,066	0,092	N-migrace	validační
494	10,000	0,368	0,000	0,300	0,050	0,000	N-migrace	testovací
480	5,410	0,353	0,009	0,271	0,117	0,162	N-migrace	testovací

Pokračování na další stránce

Tabulka A.1 – Pokračování z předchozí stránky

výš	míranez	hustob	bytnaek	fyznaek	rolnaek	zamnaek	saldo	mnozina
385	2,690	1,768	0,003	0,392	0,014	0,872	N-migrace	validační
467	3,390	0,840	0,017	0,364	0,051	0,178	N-migrace	testovací
490	6,060	0,315	0,030	0,333	0,000	0,152	N-migrace	trénovací
523	6,060	0,422	0,000	0,162	0,040	0,182	N-migrace	validační
451	5,260	0,310	0,000	0,219	0,041	0,453	N-migrace	trénovací
488	2,510	0,320	0,000	0,276	0,005	0,241	N-migrace	validační
494	8,590	0,196	0,003	0,184	0,066	0,134	N-migrace	validační
554	9,880	0,317	0,000	0,296	0,161	0,259	N-migrace	trénovací
560	15,870	0,214	0,000	0,190	0,032	0,325	N-migrace	trénovací
493	2,510	0,691	0,013	0,339	0,030	0,435	N-migrace	validační
460	5,410	0,255	0,009	0,343	0,027	0,162	N-migrace	validační
448	0,000	0,216	0,000	0,306	0,139	0,333	N-migrace	validační
515	3,420	0,635	0,004	0,325	0,031	0,312	N-migrace	testovací
457	3,920	0,145	0,007	0,261	0,111	0,274	N-migrace	testovací
468	4,290	0,557	0,021	0,358	0,050	0,772	N-migrace	trénovací
440	3,940	0,711	0,007	0,254	0,054	0,061	N-migrace	validační
505	4,220	0,426	0,005	0,318	0,037	0,431	N-migrace	trénovací
389	3,850	2,528	0,045	0,413	0,012	0,608	N-migrace	trénovací
570	4,720	0,369	0,003	0,258	0,031	0,072	N-migrace	testovací
438	7,920	0,242	0,000	0,248	0,149	0,287	Z-migrace	testovací
476	8,150	0,150	0,007	0,341	0,074	0,230	Z-migrace	testovací
513	6,670	0,192	0,000	0,200	0,100	0,400	N-migrace	testovací
515	3,130	0,220	0,000	0,344	0,031	0,188	N-migrace	validační
405	3,790	0,670	0,000	0,349	0,023	0,644	N-migrace	testovací
407	11,700	0,311	0,000	0,138	0,053	0,181	Z-migrace	trénovací
470	4,030	1,077	0,073	0,395	0,016	0,395	N-migrace	trénovací
433	1,510	0,324	0,005	0,327	0,101	0,151	Z-migrace	validační
512	1,690	1,036	0,017	0,372	0,003	0,254	N-migrace	testovací
541	4,700	0,325	0,001	0,254	0,022	0,711	N-migrace	validační
502	4,890	0,312	0,003	0,245	0,058	0,300	N-migrace	trénovací
419	3,330	0,309	0,000	0,297	0,017	0,375	N-migrace	trénovací
471	0,000	0,287	0,000	0,277	0,015	0,508	Z-migrace	validační
460	10,570	0,142	0,000	0,301	0,073	0,089	N-migrace	trénovací
398	2,410	0,349	0,012	0,265	0,028	0,406	N-migrace	validační
398	1,550	0,600	0,000	0,310	0,016	2,232	Z-migrace	testovací
406	2,720	0,639	0,020	0,408	0,027	0,428	N-migrace	testovací
433	7,140	0,176	0,000	0,298	0,036	0,202	N-migrace	testovací
393	4,500	1,531	0,013	0,341	0,010	0,434	N-migrace	validační
480	2,670	7,452	0,005	0,411	0,011	0,312	N-migrace	testovací
471	4,750	0,472	0,009	0,324	0,027	0,327	N-migrace	testovací

Pokračování na další stránce

Tabulka A.1 – Pokračování z předchozí stránky

výš	míranez	hustob	bytnaek	fyznaek	rolnaek	zamnaek	saldo	mnozina
397	4,530	0,222	0,004	0,214	0,045	1,256	N-migrace	testovací
568	8,680	0,320	0,000	0,275	0,030	0,423	Z-migrace	trénovací
515	3,890	1,844	0,041	0,379	0,037	0,158	N-migrace	validační
474	3,870	1,774	0,010	0,382	0,019	0,135	N-migrace	trénovací
535	4,060	0,340	0,010	0,228	0,030	0,208	Z-migrace	trénovací
494	5,130	0,371	0,013	0,475	0,090	0,218	Z-migrace	testovací
614	7,940	0,342	0,026	0,344	0,016	0,365	N-migrace	testovací
484	3,490	1,595	0,006	0,278	0,031	0,489	Z-migrace	trénovací
463	6,530	0,436	0,015	0,243	0,039	0,481	N-migrace	trénovací
443	6,590	0,152	0,000	0,304	0,066	5,696	N-migrace	trénovací
458	3,810	0,891	0,004	0,285	0,021	0,853	N-migrace	validační
362	6,280	1,928	0,000	0,312	0,008	0,756	N-migrace	validační
399	2,310	1,093	0,017	0,352	0,029	0,254	N-migrace	validační
438	5,240	4,058	0,010	0,402	0,005	0,233	N-migrace	testovací
396	4,820	2,475	0,006	0,313	0,024	0,247	Z-migrace	trénovací
503	8,930	0,347	0,018	0,384	0,027	0,232	N-migrace	trénovací
500	0,000	0,036	0,000	0,214	0,286	0,357	N-migrace	testovací
505	3,300	0,316	0,033	0,372	0,052	0,198	N-migrace	trénovací
405	5,830	1,482	0,019	0,350	0,058	1,632	N-migrace	testovací
427	6,410	0,339	0,000	0,308	0,076	0,337	N-migrace	validační
530	4,200	0,176	0,000	0,245	0,091	0,154	Z-migrace	trénovací
407	3,700	0,817	0,005	0,275	0,000	0,116	Z-migrace	trénovací
435	0,000	0,159	0,000	0,143	0,048	0,048	N-migrace	trénovací
390	3,930	2,574	0,003	0,299	0,007	0,415	N-migrace	testovací
485	4,920	0,281	0,008	0,320	0,057	0,377	N-migrace	testovací
394	4,520	0,342	0,010	0,291	0,010	0,608	N-migrace	validační
518	6,960	0,222	0,000	0,234	0,006	0,633	N-migrace	testovací
443	2,550	0,182	0,000	0,233	0,073	0,222	N-migrace	testovací
661	11,760	0,221	0,001	0,203	0,037	0,359	N-migrace	validační
575	8,800	0,520	0,007	0,308	0,025	0,153	N-migrace	trénovací
533	14,390	0,090	0,000	0,182	0,061	0,318	N-migrace	validační
568	9,790	0,225	0,004	0,279	0,045	0,358	N-migrace	testovací
670	17,220	0,320	0,000	0,225	0,070	0,149	Z-migrace	validační
728	18,880	0,166	0,002	0,527	0,044	0,326	Z-migrace	trénovací
492	9,370	6,384	0,001	0,334	0,009	0,903	Z-migrace	testovací
618	15,240	0,137	0,020	0,214	0,025	0,456	Z-migrace	testovací
424	11,110	0,405	0,225	0,336	0,034	0,404	N-migrace	validační
708	14,710	0,245	0,003	0,451	0,012	0,342	Z-migrace	validační
512	4,310	0,613	0,009	0,321	0,022	0,614	N-migrace	testovací
651	17,500	0,372	0,000	0,146	0,021	0,693	Z-migrace	trénovací

Pokračování na další stránce

Tabulka A.1 – Pokračování z předchozí stránky

výš	míranez	hustob	bytnaek	fyznaek	rolnaek	zamnaek	saldo	mnozina
776	12,470	0,230	0,008	0,353	0,024	0,713	N-migrace	testovací
674	17,870	0,246	0,002	0,309	0,029	0,295	Z-migrace	validační
525	0,000	0,126	0,000	0,333	0,143	0,095	Z-migrace	testovací
671	12,230	0,414	0,000	0,314	0,014	0,435	N-migrace	trénovací
540	12,930	0,278	0,000	0,278	0,013	0,594	N-migrace	testovací
537	8,530	1,760	0,002	0,276	0,005	0,869	Z-migrace	trénovací
521	6,380	0,683	0,007	0,324	0,030	0,303	Z-migrace	trénovací
776	8,100	0,265	0,000	0,313	0,000	0,873	Z-migrace	testovací
663	13,400	0,465	0,002	0,223	0,002	0,535	Z-migrace	testovací
681	13,580	0,185	0,002	0,199	0,044	0,397	N-migrace	testovací
794	23,210	0,044	0,000	0,250	0,125	0,179	N-migrace	validační
540	14,350	0,272	0,000	0,306	0,048	0,277	N-migrace	validační
550	10,380	0,223	0,047	0,245	0,028	0,462	Z-migrace	trénovací
635	6,580	0,341	0,004	0,341	0,021	0,230	N-migrace	trénovací
558	2,590	0,394	0,010	0,202	0,016	0,109	N-migrace	validační
675	13,950	0,323	0,023	0,219	0,088	0,093	N-migrace	testovací
760	45,210	0,040	0,000	0,295	0,089	0,082	Z-migrace	validační
755	17,330	0,033	0,000	0,327	0,033	0,293	Z-migrace	trénovací
670	14,650	0,143	0,000	0,289	0,048	0,407	N-migrace	testovací
535	13,460	0,334	0,000	0,327	0,096	0,510	N-migrace	trénovací
528	13,860	0,134	0,000	0,362	0,066	0,675	Z-migrace	trénovací
623	15,980	0,097	0,000	0,129	0,052	0,222	N-migrace	validační
632	6,450	0,237	0,026	0,194	0,039	0,123	Z-migrace	trénovací
540	5,000	0,335	0,067	0,325	0,058	1,608	Z-migrace	validační
675	8,740	0,437	0,016	0,213	0,027	1,398	N-migrace	testovací
776	21,770	0,097	0,000	0,379	0,016	0,137	Z-migrace	trénovací
548	8,220	3,041	0,001	0,276	0,011	0,675	Z-migrace	trénovací
592	15,780	1,399	0,002	0,259	0,008	0,453	N-migrace	testovací
695	5,170	0,244	0,000	0,362	0,121	0,155	N-migrace	validační
571	10,350	0,377	0,000	0,354	0,018	0,650	N-migrace	validační
473	6,270	0,772	0,003	0,317	0,025	0,194	N-migrace	validační
620	6,630	0,386	0,000	0,316	0,036	0,342	Z-migrace	trénovací
578	13,040	0,158	0,000	0,261	0,304	0,043	N-migrace	validační
516	0,000	0,232	0,000	0,160	0,200	0,200	N-migrace	validační
515	4,350	0,126	0,000	0,305	0,065	0,196	Z-migrace	testovací
549	4,050	0,287	0,000	0,405	0,068	0,041	N-migrace	trénovací
506	9,800	0,152	0,013	0,353	0,026	0,033	Z-migrace	testovací
580	0,000	0,147	0,000	0,182	0,061	0,333	N-migrace	validační
560	6,060	0,161	0,000	0,121	0,167	0,091	N-migrace	trénovací
488	17,890	0,465	0,000	0,263	0,147	0,495	N-migrace	trénovací

Pokračování na další stránce

Tabulka A.1 – Pokračování z předchozí stránky

výš	míranez	hustob	bytnaek	fyznaek	rolnaek	zamnaek	saldo	mnozina
512	12,350	0,341	0,000	0,162	0,074	0,529	N-migrace	validační
461	15,480	0,053	0,000	0,250	0,155	0,060	N-migrace	testovací
477	12,780	0,225	0,000	0,207	0,049	0,429	Z-migrace	testovací
525	7,480	0,312	0,000	0,206	0,009	0,421	Z-migrace	trénovací
489	5,760	2,878	0,004	0,241	0,006	0,827	N-migrace	trénovací
511	14,760	0,204	0,000	0,258	0,020	1,488	N-migrace	trénovací
519	8,410	0,152	0,003	0,270	0,038	0,136	N-migrace	trénovací
544	8,970	0,191	0,000	0,256	0,054	0,206	N-migrace	trénovací
577	8,660	1,186	0,003	0,274	0,016	0,963	Z-migrace	trénovací
466	28,490	0,174	0,000	0,193	0,053	0,323	Z-migrace	trénovací
420	7,290	0,548	0,000	0,307	0,043	0,671	N-migrace	validační
565	3,030	0,216	0,000	0,152	0,030	0,000	N-migrace	testovací
515	0,000	0,139	0,000	0,118	0,353	0,059	N-migrace	validační
473	7,840	0,216	0,020	0,510	0,098	1,921	N-migrace	validační
455	7,840	0,168	0,020	0,255	0,000	0,176	N-migrace	testovací
452	5,110	0,248	0,007	0,336	0,058	0,518	N-migrace	validační
441	4,000	0,179	0,000	0,180	0,160	0,320	N-migrace	testovací
566	13,040	0,136	0,000	0,304	0,087	0,000	Z-migrace	testovací
458	4,260	0,232	0,000	0,277	0,085	0,341	N-migrace	validační
461	5,260	0,203	0,013	0,276	0,092	0,099	N-migrace	validační
427	4,170	0,504	0,000	0,209	0,028	0,487	Z-migrace	testovací
537	6,670	0,226	0,067	0,300	0,033	0,200	N-migrace	testovací
462	7,320	0,223	0,000	0,305	0,000	2,647	Z-migrace	validační
450	5,560	0,300	0,010	0,293	0,020	1,152	Z-migrace	trénovací
484	11,110	0,068	0,000	0,232	0,071	0,162	N-migrace	validační
609	18,920	0,179	0,000	0,216	0,162	0,027	N-migrace	trénovací
555	16,000	0,244	0,000	0,160	0,260	0,220	N-migrace	testovací
593	14,580	0,219	0,000	0,146	0,208	0,000	Z-migrace	validační
496	4,940	0,321	0,019	0,220	0,049	0,201	Z-migrace	testovací
558	5,790	0,167	0,000	0,265	0,025	0,422	N-migrace	trénovací
484	2,380	0,221	0,024	0,381	0,071	1,642	Z-migrace	testovací
507	21,540	0,244	0,000	0,185	0,046	0,046	N-migrace	testovací
525	10,000	0,436	0,000	0,195	0,011	0,479	Z-migrace	trénovací
458	9,680	0,163	0,000	0,387	0,129	0,323	N-migrace	trénovací
524	7,140	0,274	0,000	0,228	0,021	0,635	N-migrace	testovací
492	6,820	0,461	0,003	0,239	0,016	0,703	Z-migrace	trénovací
476	4,840	0,384	0,007	0,285	0,054	0,527	Z-migrace	trénovací
614	4,760	0,242	0,000	0,159	0,143	0,063	Z-migrace	trénovací
475	5,160	3,071	0,004	0,290	0,015	0,937	N-migrace	testovací
539	3,850	0,072	0,000	0,231	0,116	0,039	N-migrace	validační

Pokračování na další stránce

Tabulka A.1 – Pokračování z předchozí stránky

výš	míranez	hustob	bytnaek	fyznaek	rolnaek	zamnaek	saldo	mnozina
526	5,880	0,129	0,000	0,323	0,088	0,176	N-migrace	testovací
439	4,720	0,452	0,033	0,270	0,025	0,642	N-migrace	trénovací
414	3,410	0,329	0,000	0,296	0,068	0,045	Z-migrace	trénovací
513	6,980	0,443	0,016	0,248	0,054	0,202	N-migrace	validační
474	8,970	0,234	0,000	0,167	0,038	0,730	Z-migrace	validační
575	7,030	0,300	0,000	0,293	0,034	0,575	N-migrace	validační
450	8,000	0,483	0,000	0,244	0,040	0,527	Z-migrace	validační
535	8,900	0,253	0,021	0,228	0,061	0,342	N-migrace	trénovací
424	6,460	0,878	0,006	0,250	0,029	0,388	N-migrace	testovací
426	6,060	0,313	0,000	0,327	0,055	0,194	N-migrace	trénovací
438	7,690	0,383	0,007	0,205	0,033	0,513	N-migrace	validační
588	7,730	0,412	0,003	0,254	0,015	0,603	Z-migrace	validační
543	10,710	0,131	0,000	0,286	0,054	0,089	Z-migrace	trénovací
507	7,490	2,424	0,002	0,261	0,010	0,450	Z-migrace	validační
474	10,460	0,143	0,013	0,333	0,007	1,053	N-migrace	validační
460	9,840	0,143	0,003	0,279	0,114	0,416	Z-migrace	trénovací
625	5,750	0,188	0,012	0,219	0,058	0,920	Z-migrace	trénovací
481	13,330	0,215	0,010	0,185	0,031	0,010	N-migrace	testovací
443	30,200	0,168	0,000	0,200	0,063	0,235	N-migrace	testovací
458	7,690	0,183	0,000	0,154	0,096	1,019	N-migrace	testovací
464	8,290	0,399	0,016	0,337	0,021	0,316	N-migrace	validační
487	9,410	0,183	0,000	0,318	0,129	0,659	Z-migrace	validační
478	6,900	0,189	0,000	0,257	0,088	0,489	Z-migrace	validační
511	12,200	0,223	0,000	0,366	0,098	0,415	N-migrace	validační
425	9,620	0,207	0,000	0,058	0,019	1,597	N-migrace	testovací
423	0,000	0,137	0,000	0,250	0,025	0,025	N-migrace	trénovací
575	6,790	0,330	0,000	0,294	0,054	0,380	N-migrace	testovací
468	0,000	0,314	0,011	0,232	0,021	0,611	N-migrace	validační
472	6,380	0,191	0,000	0,185	0,034	0,272	Z-migrace	trénovací
512	3,160	0,136	0,000	0,274	0,084	0,390	Z-migrace	validační
475	5,450	0,341	0,015	0,322	0,124	1,164	N-migrace	trénovací
502	3,740	0,372	0,000	0,206	0,103	0,112	N-migrace	validační
427	8,330	0,167	0,000	0,208	0,292	0,042	N-migrace	validační
512	12,910	0,592	0,002	0,254	0,015	0,523	Z-migrace	testovací
434	2,630	0,083	0,026	0,395	0,026	0,053	Z-migrace	validační
470	5,980	0,125	0,017	0,342	0,017	0,777	N-migrace	trénovací
514	13,120	0,228	0,004	0,227	0,050	0,245	N-migrace	trénovací
545	16,960	0,078	0,000	0,232	0,021	0,405	Z-migrace	validační
450	7,860	0,225	0,002	0,292	0,042	0,634	N-migrace	validační
545	8,270	0,457	0,000	0,311	0,032	0,464	Z-migrace	validační

Pokračování na další stránce

Tabulka A.1 – Pokračování z předchozí stránky

výš	míranez	hustob	bytnaek	fyznaek	rolnaek	zamnaek	saldo	mnozina
442	6,990	0,234	0,005	0,296	0,113	0,274	N-migrace	trénovací
532	9,240	0,487	0,000	0,210	0,019	0,962	Z-migrace	validační
618	6,900	0,548	0,000	0,204	0,026	0,858	Z-migrace	trénovací
454	6,570	0,571	0,006	0,250	0,031	0,872	N-migrace	validační
440	12,070	0,182	0,000	0,190	0,103	0,017	N-migrace	validační
487	10,490	0,446	0,000	0,105	0,091	0,210	N-migrace	validační
434	4,640	0,901	0,003	0,326	0,015	0,940	Z-migrace	trénovací
448	12,500	0,152	0,000	0,219	0,156	0,063	N-migrace	testovací
513	3,480	0,151	0,000	0,252	0,061	0,800	Z-migrace	testovací
540	13,330	0,238	0,000	0,333	0,044	0,000	N-migrace	trénovací
478	0,000	0,128	0,027	0,243	0,135	0,027	N-migrace	testovací
545	18,180	0,101	0,045	0,227	0,091	0,000	Z-migrace	trénovací
498	9,810	0,200	0,003	0,215	0,051	0,256	Z-migrace	testovací
483	6,000	0,193	0,000	0,240	0,060	0,080	N-migrace	validační
423	2,860	0,082	0,029	0,343	0,172	0,486	N-migrace	testovací
465	0,000	0,332	0,000	0,212	0,045	0,212	Z-migrace	validační
594	2,220	0,302	0,000	0,155	0,126	0,777	Z-migrace	trénovací
505	3,330	0,101	0,033	0,433	0,100	0,333	N-migrace	trénovací
512	47,620	0,103	0,000	0,143	0,095	0,000	N-migrace	testovací
428	6,370	0,224	0,000	0,328	0,078	0,257	N-migrace	testovací
463	11,330	0,346	0,005	0,269	0,028	0,453	N-migrace	validační
441	5,210	0,363	0,000	0,417	0,010	0,615	N-migrace	trénovací
475	8,080	0,175	0,000	0,384	0,091	0,253	N-migrace	testovací
442	7,940	0,294	0,000	0,291	0,090	0,344	N-migrace	trénovací
493	8,500	0,600	0,000	0,275	0,052	0,301	N-migrace	validační
450	8,620	0,141	0,017	0,207	0,086	0,534	N-migrace	validační
400	7,000	1,016	0,002	0,367	0,026	0,599	N-migrace	validační
448	8,990	0,251	0,000	0,346	0,041	0,560	N-migrace	testovací
385	8,210	0,324	0,000	0,296	0,085	0,534	N-migrace	trénovací
486	12,090	0,222	0,000	0,440	0,286	0,121	N-migrace	testovací
447	9,350	0,201	0,000	0,224	0,140	0,084	Z-migrace	trénovací
456	3,570	0,220	0,000	0,241	0,036	0,585	N-migrace	validační
382	6,000	0,311	0,000	0,280	0,110	0,090	Z-migrace	testovací
427	5,660	0,197	0,000	0,245	0,283	0,208	N-migrace	testovací
535	9,920	0,329	0,008	0,347	0,074	0,306	N-migrace	testovací
471	7,760	0,329	0,000	0,320	0,068	0,557	N-migrace	testovací
666	12,300	0,357	0,002	0,289	0,105	0,359	Z-migrace	validační
447	5,000	0,210	0,007	0,286	0,029	0,486	Z-migrace	testovací
414	12,500	0,092	0,000	0,396	0,146	0,208	N-migrace	trénovací
373	8,980	0,310	0,000	0,233	0,069	0,311	Z-migrace	validační

Pokračování na další stránce

Tabulka A.1 – Pokračování z předchozí stránky

výš	míranez	hustob	bytnaek	fyznaek	rolnaek	zamnaek	saldo	mnozina
460	8,890	0,378	0,004	0,316	0,058	0,262	N-migrace	testovací
457	15,430	0,128	0,000	0,282	0,069	0,388	Z-migrace	testovací
525	10,010	0,285	0,001	0,390	0,043	0,392	N-migrace	testovací
398	11,540	0,115	0,000	0,269	0,154	1,116	Z-migrace	validační
444	14,910	0,190	0,000	0,210	0,184	0,140	Z-migrace	trénovací
478	6,350	0,236	0,000	0,286	0,048	0,191	N-migrace	trénovací
482	0,000	0,272	0,000	0,154	0,154	0,077	N-migrace	testovací
432	7,370	0,149	0,000	0,358	0,126	0,274	Z-migrace	trénovací
441	8,700	0,072	0,000	0,348	0,109	0,239	N-migrace	validační
450	11,940	0,207	0,000	0,276	0,194	0,567	N-migrace	trénovací
461	8,800	2,209	0,010	0,333	0,016	0,883	Z-migrace	testovací
478	0,000	0,100	0,000	0,250	0,050	0,000	Z-migrace	trénovací
412	10,040	0,435	0,005	0,345	0,057	0,345	N-migrace	testovací
433	9,770	0,752	0,000	0,304	0,070	0,579	Z-migrace	testovací
461	10,340	0,165	0,000	0,267	0,146	0,224	Z-migrace	validační
439	14,290	0,214	0,000	0,286	0,191	0,238	Z-migrace	testovací
435	23,680	0,197	0,000	0,421	0,210	0,263	N-migrace	validační
439	14,290	0,149	0,014	0,343	0,100	0,029	Z-migrace	testovací
499	8,000	0,158	0,040	0,200	0,040	0,040	N-migrace	testovací
426	9,300	0,164	0,023	0,349	0,070	1,139	N-migrace	testovací
390	15,430	0,379	0,000	0,383	0,031	0,994	Z-migrace	trénovací
451	6,450	0,225	0,000	0,387	0,000	0,226	Z-migrace	trénovací
421	7,140	0,172	0,000	0,270	0,119	0,865	N-migrace	trénovací
383	13,870	0,181	0,000	0,214	0,081	0,584	N-migrace	testovací
465	11,430	0,056	0,000	0,300	0,043	0,071	Z-migrace	testovací
398	7,170	4,708	0,001	0,326	0,010	0,844	Z-migrace	validační
454	14,040	0,392	0,017	0,281	0,112	0,202	N-migrace	testovací
462	13,640	0,150	0,000	0,273	0,273	0,136	Z-migrace	trénovací
383	8,710	0,818	0,002	0,302	0,024	0,668	N-migrace	validační
480	10,290	0,318	0,000	0,250	0,103	0,279	N-migrace	trénovací
478	11,330	0,196	0,000	0,256	0,044	0,020	N-migrace	validační
568	6,210	0,303	0,000	0,338	0,138	0,421	Z-migrace	validační
386	6,220	0,425	0,000	0,284	0,040	0,098	N-migrace	trénovací
438	11,460	0,199	0,000	0,271	0,125	0,521	Z-migrace	validační
386	10,470	0,386	0,000	0,298	0,089	0,675	Z-migrace	validační
500	9,230	0,471	0,000	0,354	0,067	0,375	Z-migrace	trénovací
387	7,750	0,147	0,000	0,217	0,047	0,395	Z-migrace	trénovací
439	7,100	0,177	0,006	0,266	0,112	0,272	Z-migrace	trénovací
413	13,290	0,202	0,000	0,294	0,084	0,392	N-migrace	trénovací
491	2,380	0,216	0,000	0,119	0,095	0,071	Z-migrace	trénovací

Pokračování na další stránce

Tabulka A.1 – Pokračování z předchozí stránky

výš	míranez	hustob	bytnaek	fyznaek	rolnaek	zamnaek	saldo	mnozina
393	8,330	0,377	0,000	0,278	0,125	0,597	N-migrace	trénovací
414	10,260	0,177	0,000	0,436	0,077	0,026	Z-migrace	validační
426	14,080	0,069	0,000	0,366	0,014	0,141	N-migrace	validační
465	17,350	0,430	0,000	0,296	0,020	1,276	Z-migrace	trénovací
437	14,580	0,243	0,000	0,198	0,219	0,156	N-migrace	trénovací
475	7,890	0,166	0,000	0,132	0,039	0,145	Z-migrace	testovací
425	11,110	0,056	0,000	0,472	0,361	0,278	Z-migrace	trénovací
435	10,340	0,389	0,000	0,303	0,021	1,461	Z-migrace	validační
418	7,020	0,212	0,000	0,228	0,088	0,000	N-migrace	trénovací
492	6,150	0,501	0,000	0,296	0,036	0,677	N-migrace	validační
488	4,230	1,091	0,005	0,254	0,032	0,307	Z-migrace	trénovací
463	8,110	0,204	0,000	0,162	0,027	1,244	N-migrace	validační
563	12,700	0,231	0,000	0,262	0,191	0,397	Z-migrace	trénovací
402	5,490	0,438	0,000	0,373	0,033	0,791	Z-migrace	trénovací
374	8,930	0,146	0,009	0,196	0,116	0,161	N-migrace	testovací
447	12,900	0,122	0,032	0,194	0,065	0,065	Z-migrace	trénovací
656	7,030	0,918	0,055	0,320	0,047	0,578	Z-migrace	testovací
478	0,000	0,171	0,000	0,063	0,016	0,000	N-migrace	testovací
895	14,190	0,040	0,000	0,471	0,052	0,284	Z-migrace	trénovací
588	8,160	0,349	0,027	0,313	0,082	0,102	N-migrace	testovací
490	6,980	0,171	0,000	0,419	0,023	0,000	N-migrace	trénovací
826	5,430	0,110	0,016	0,287	0,140	0,807	Z-migrace	trénovací
540	11,650	0,268	0,000	0,262	0,165	0,097	N-migrace	validační
523	4,240	0,709	0,003	0,332	0,047	1,104	N-migrace	testovací
665	11,760	0,162	0,029	0,147	0,235	0,000	N-migrace	testovací
474	7,300	0,269	0,000	0,208	0,079	0,326	Z-migrace	testovací
621	0,000	0,189	0,061	0,364	0,273	0,485	Z-migrace	validační
805	11,170	0,069	0,000	0,330	0,071	0,244	N-migrace	trénovací
496	8,410	0,230	0,000	0,274	0,235	0,288	N-migrace	trénovací
504	6,210	1,294	0,037	0,374	0,035	0,960	N-migrace	testovací
528	4,760	0,525	0,000	0,286	0,082	0,360	N-migrace	testovací
758	6,780	0,131	0,008	0,237	0,042	0,212	N-migrace	trénovací
465	6,900	0,446	0,000	0,288	0,035	0,046	Z-migrace	trénovací
613	12,000	0,136	0,000	0,520	0,280	0,080	N-migrace	testovací
925	22,540	0,074	0,000	0,465	0,070	0,564	N-migrace	testovací
762	14,340	0,136	0,000	0,327	0,032	0,586	Z-migrace	trénovací
960	5,800	0,829	0,000	0,464	0,000	0,392	N-migrace	validační
1062	9,000	0,037	0,000	0,620	0,080	0,710	Z-migrace	validační
663	10,060	0,718	0,000	0,278	0,041	0,231	Z-migrace	testovací
566	1,610	0,200	0,016	0,451	0,064	0,032	Z-migrace	trénovací

Pokračování na další stránce

Tabulka A.1 – Pokračování z předchozí stránky

výš	míranez	hustob	bytnaek	fyznaek	rolnaek	zamnaek	saldo	mnozina
765	9,920	0,455	0,011	0,391	0,019	0,225	Z-migrace	validační
559	5,430	0,453	0,001	0,333	0,136	0,332	N-migrace	trénovací
568	7,140	0,384	0,000	0,190	0,024	0,417	N-migrace	validační
460	10,530	0,147	0,000	0,369	0,000	0,105	Z-migrace	testovací
425	4,490	0,258	0,000	0,213	0,112	0,213	N-migrace	testovací
411	8,740	0,255	0,003	0,262	0,117	0,411	N-migrace	testovací
575	0,650	0,133	0,000	0,267	0,182	0,397	Z-migrace	validační
745	7,250	0,261	0,005	0,280	0,073	0,083	N-migrace	trénovací
431	6,900	0,439	0,000	0,426	0,081	1,185	Z-migrace	testovací
432	6,410	1,033	0,017	0,297	0,025	0,628	N-migrace	validační
896	12,200	0,060	0,000	0,366	0,073	0,268	Z-migrace	trénovací
737	19,500	0,093	0,000	0,399	0,022	0,418	Z-migrace	testovací
1025	4,170	0,040	0,000	0,626	0,021	0,917	N-migrace	testovací
438	13,330	0,098	0,000	0,267	0,200	0,067	N-migrace	testovací
600	0,000	0,257	0,000	0,214	0,000	0,071	N-migrace	validační
561	5,200	3,034	0,001	0,327	0,012	0,880	Z-migrace	trénovací
770	19,230	0,166	0,000	0,269	0,282	0,141	N-migrace	testovací
738	6,180	0,434	0,004	0,446	0,110	0,749	N-migrace	testovací
780	5,650	0,021	0,008	0,726	0,081	0,791	Z-migrace	trénovací
834	9,430	0,070	0,000	0,522	0,050	0,974	N-migrace	trénovací
458	7,860	0,498	0,002	0,255	0,073	0,439	N-migrace	testovací
765	6,250	0,425	0,031	0,353	0,089	0,143	N-migrace	trénovací
765	3,550	0,474	0,005	0,340	0,107	0,264	Z-migrace	trénovací
501	4,900	0,305	0,007	0,231	0,035	1,764	Z-migrace	validační
509	6,350	0,348	0,000	0,222	0,016	0,397	Z-migrace	trénovací
532	4,350	0,176	0,000	0,479	0,087	0,044	N-migrace	validační
732	5,070	0,390	0,014	0,301	0,093	0,792	N-migrace	validační
694	7,010	1,011	0,000	0,365	0,017	0,764	Z-migrace	validační
509	7,270	0,369	0,009	0,354	0,227	0,259	N-migrace	testovací
525	5,980	0,837	0,000	0,336	0,034	0,430	N-migrace	trénovací
760	11,270	0,377	0,005	0,323	0,017	0,488	Z-migrace	validační
747	9,090	0,311	0,000	0,364	0,091	0,667	N-migrace	validační
593	11,490	0,147	0,040	0,322	0,023	0,103	N-migrace	trénovací
642	4,550	0,116	0,046	0,273	0,319	0,000	N-migrace	testovací
565	3,450	0,327	0,000	0,255	0,138	0,035	N-migrace	validační
786	7,450	0,081	0,000	0,273	0,050	0,422	N-migrace	testovací
732	7,560	0,506	0,006	0,353	0,036	0,404	Z-migrace	testovací
725	2,170	0,456	0,000	0,412	0,065	0,043	N-migrace	validační
775	27,540	0,130	0,000	0,348	0,029	0,522	N-migrace	validační
651	5,360	0,436	0,009	0,223	0,089	0,241	Z-migrace	testovací

Pokračování na další stránce

Tabulka A.1 – Pokračování z předchozí stránky

výš	míranez	hustob	bytnaek	fyznaek	rolnaek	zamnaek	saldo	mnozina
446	11,460	0,396	0,001	0,285	0,037	0,335	Z-migrace	validační
526	11,800	0,289	0,002	0,341	0,021	0,388	N-migrace	testovací
448	8,000	0,527	0,000	0,327	0,040	0,213	N-migrace	trénovací
457	12,630	0,183	0,011	0,147	0,063	0,042	N-migrace	validační
440	6,630	1,545	0,001	0,288	0,015	1,163	N-migrace	testovací
545	0,000	0,113	0,000	0,276	0,207	0,172	N-migrace	trénovací
518	11,110	0,127	0,000	0,278	0,000	0,444	Z-migrace	testovací
431	5,560	0,187	0,000	0,222	0,334	0,000	N-migrace	trénovací
432	8,000	0,161	0,000	0,293	0,000	0,480	N-migrace	validační
432	13,330	0,304	0,000	0,338	0,071	0,252	N-migrace	testovací
554	11,270	0,243	0,000	0,366	0,070	0,239	N-migrace	validační
377	8,710	0,404	0,000	0,233	0,081	0,185	N-migrace	validační
529	10,990	0,191	0,000	0,286	0,088	0,429	N-migrace	trénovací
558	5,430	0,407	0,002	0,281	0,100	0,378	Z-migrace	validační
484	11,160	0,371	0,000	0,236	0,039	0,713	N-migrace	trénovací
531	4,240	0,337	0,000	0,331	0,017	0,271	N-migrace	testovací
468	8,940	0,280	0,000	0,224	0,117	0,369	Z-migrace	validační
450	5,130	0,448	0,000	0,333	0,038	0,603	N-migrace	testovací
660	8,260	0,219	0,000	0,240	0,091	0,091	Z-migrace	validační
440	18,000	0,213	0,000	0,340	0,040	0,000	N-migrace	testovací
575	9,770	0,293	0,000	0,308	0,090	0,338	N-migrace	validační
455	19,440	0,236	0,000	0,444	0,167	0,555	N-migrace	validační
460	42,860	0,143	0,000	0,500	0,214	0,000	Z-migrace	validační
458	5,410	0,192	0,000	0,135	0,081	0,243	Z-migrace	validační
412	5,520	0,438	0,000	0,248	0,124	0,041	N-migrace	validační
510	0,000	0,211	0,000	0,378	0,000	0,027	N-migrace	trénovací
618	9,210	0,168	0,000	0,329	0,184	0,316	Z-migrace	testovací
515	15,850	0,449	0,000	0,244	0,134	0,037	Z-migrace	validační
457	4,900	0,746	0,000	0,304	0,059	1,039	N-migrace	trénovací
473	5,810	0,282	0,000	0,407	0,198	0,349	N-migrace	validační
453	23,080	0,275	0,000	0,192	0,038	0,269	N-migrace	testovací
525	8,400	0,234	0,000	0,260	0,076	0,176	Z-migrace	validační
545	10,200	0,375	0,000	0,224	0,020	0,398	Z-migrace	trénovací
510	13,810	0,192	0,000	0,204	0,050	0,392	Z-migrace	trénovací
513	21,430	0,252	0,000	0,214	0,024	0,048	N-migrace	trénovací
412	6,360	1,335	0,000	0,333	0,054	0,499	N-migrace	trénovací
460	17,140	0,318	0,000	0,143	0,057	0,443	N-migrace	testovací
474	5,750	0,197	0,000	0,207	0,081	0,230	N-migrace	trénovací
524	6,520	0,121	0,000	0,348	0,087	0,348	N-migrace	trénovací
595	11,710	0,286	0,000	0,180	0,045	0,117	Z-migrace	testovací

Pokračování na další stránce

Tabulka A.1 – Pokračování z předchozí stránky

výš	míranez	hustob	bytnaek	fyznaek	rolnaek	zamnaek	saldo	mnozina
398	11,480	0,271	0,016	0,213	0,180	0,640	N-migrace	testovací
524	16,130	0,188	0,000	0,387	0,194	0,032	Z-migrace	validační
431	7,140	0,243	0,018	0,268	0,250	0,196	N-migrace	trénovací
476	0,000	0,080	0,000	0,154	0,308	0,000	N-migrace	trénovací
483	10,000	0,311	0,000	0,240	0,040	0,020	N-migrace	testovací
529	3,330	0,275	0,000	0,200	0,000	1,765	Z-migrace	testovací
506	6,670	0,208	0,000	0,334	0,167	0,033	N-migrace	trénovací
434	6,820	0,345	0,000	0,236	0,132	0,300	N-migrace	testovací
503	3,130	0,177	0,000	0,094	0,094	0,000	N-migrace	validační
458	10,660	0,236	0,000	0,205	0,049	0,180	N-migrace	trénovací
565	9,200	0,593	0,000	0,316	0,035	0,589	Z-migrace	validační
474	5,360	0,221	0,018	0,339	0,107	0,036	Z-migrace	testovací
454	6,190	0,559	0,031	0,279	0,062	1,238	Z-migrace	trénovací
483	6,900	0,640	0,000	0,355	0,059	0,179	N-migrace	testovací
494	7,270	0,272	0,009	0,273	0,018	0,327	N-migrace	validační
519	15,380	0,077	0,000	0,077	0,154	0,538	N-migrace	validační
530	0,000	0,080	0,000	0,500	0,350	0,000	Z-migrace	trénovací
480	7,320	0,449	0,000	0,220	0,049	0,293	N-migrace	trénovací
454	5,940	0,289	0,000	0,168	0,218	0,059	N-migrace	testovací
406	5,750	0,841	0,000	0,368	0,069	0,184	Z-migrace	testovací
459	15,570	0,169	0,008	0,270	0,057	0,246	Z-migrace	testovací
450	11,110	0,256	0,000	0,222	0,056	0,481	Z-migrace	trénovací
541	22,000	0,399	0,000	0,260	0,120	0,080	N-migrace	validační
438	9,620	0,309	0,000	0,327	0,038	0,500	N-migrace	validační
450	10,450	0,320	0,000	0,269	0,097	0,642	N-migrace	validační
463	12,500	0,333	0,000	0,275	0,083	1,850	N-migrace	validační
599	17,650	0,092	0,000	0,412	0,412	0,000	Z-migrace	validační
418	7,360	0,403	0,000	0,209	0,043	0,515	N-migrace	validační
450	10,800	0,439	0,000	0,202	0,099	0,610	N-migrace	trénovací
482	9,760	0,185	0,000	0,293	0,098	0,024	N-migrace	testovací
513	0,000	0,234	0,000	0,419	0,161	1,194	N-migrace	testovací
418	27,500	0,141	0,000	0,250	0,175	0,075	Z-migrace	testovací
400	8,780	0,376	0,000	0,344	0,061	0,480	N-migrace	testovací
491	10,910	0,290	0,000	0,200	0,042	0,521	N-migrace	validační
413	12,900	0,194	0,000	0,290	0,290	0,000	Z-migrace	trénovací
528	6,860	0,203	0,000	0,225	0,069	0,833	Z-migrace	validační
445	2,170	0,268	0,000	0,391	0,260	0,217	Z-migrace	testovací
391	6,760	0,394	0,000	0,266	0,077	0,468	Z-migrace	trénovací
521	6,670	0,173	0,000	0,200	0,400	0,934	N-migrace	trénovací
455	9,500	0,456	0,000	0,271	0,064	0,910	N-migrace	validační

Pokračování na další stránce

Tabulka A.1 – Pokračování z předchozí stránky

výš	míranez	hustob	bytnaek	fyznaek	rolnaek	zamnaek	saldo	mnozina
413	7,060	0,545	0,000	0,275	0,130	1,334	N-migrace	trénovací
423	6,060	0,560	0,000	0,303	0,038	1,621	Z-migrace	validační
423	10,070	0,714	0,013	0,262	0,020	0,544	N-migrace	testovací
512	8,330	0,399	0,005	0,271	0,010	0,598	N-migrace	trénovací
498	20,690	0,164	0,000	0,414	0,241	0,034	N-migrace	trénovací
462	11,460	0,193	0,000	0,198	0,073	0,365	N-migrace	testovací
395	13,790	0,421	0,000	0,241	0,000	0,034	N-migrace	testovací
438	5,930	0,514	0,000	0,271	0,034	1,178	Z-migrace	trénovací
441	3,730	0,366	0,000	0,328	0,142	0,209	N-migrace	trénovací
393	7,210	6,789	0,002	0,300	0,012	1,056	N-migrace	validační
585	11,360	0,250	0,000	0,250	0,045	0,318	Z-migrace	trénovací
442	7,270	0,267	0,000	0,200	0,182	0,055	N-migrace	validační
415	6,500	0,420	0,000	0,208	0,038	0,432	N-migrace	validační
458	16,340	0,325	0,000	0,196	0,098	0,386	N-migrace	trénovací
435	6,980	0,304	0,000	0,171	0,016	0,132	Z-migrace	testovací
388	8,960	0,578	0,002	0,329	0,031	0,257	N-migrace	trénovací
454	10,430	0,218	0,000	0,304	0,000	0,104	N-migrace	validační
498	13,100	0,291	0,000	0,286	0,214	0,179	N-migrace	testovací
484	4,640	0,311	0,007	0,225	0,040	0,563	Z-migrace	validační
459	3,030	0,175	0,000	0,333	0,061	0,212	N-migrace	validační
513	3,700	0,172	0,000	0,111	0,056	0,907	Z-migrace	trénovací
468	0,000	0,073	0,000	0,087	0,087	0,348	N-migrace	validační
488	10,530	0,227	0,000	0,176	0,053	0,105	Z-migrace	testovací
509	7,140	0,204	0,000	0,286	0,054	0,107	Z-migrace	validační
706	10,000	0,270	0,000	0,167	0,200	0,000	N-migrace	trénovací
441	9,090	0,189	0,000	0,303	0,182	0,091	N-migrace	trénovací
398	7,490	1,839	0,004	0,344	0,024	0,869	N-migrace	validační
459	8,960	0,344	0,004	0,258	0,168	0,287	Z-migrace	trénovací
461	7,130	1,550	0,003	0,342	0,019	0,620	N-migrace	validační
555	6,180	0,473	0,000	0,365	0,045	0,365	Z-migrace	validační
567	8,820	0,196	0,000	0,353	0,176	0,265	N-migrace	validační
513	17,860	0,176	0,000	0,286	0,071	0,000	N-migrace	validační
493	6,560	0,345	0,000	0,344	0,213	0,590	Z-migrace	validační
432	13,790	0,231	0,000	0,827	0,000	0,000	N-migrace	validační
406	7,460	2,715	0,001	0,377	0,011	0,605	Z-migrace	validační
457	0,000	0,150	0,000	0,198	0,105	0,070	N-migrace	trénovací
416	11,110	0,150	0,009	0,282	0,248	0,350	N-migrace	validační
520	11,940	0,198	0,060	0,363	0,100	0,274	N-migrace	validační
664	10,000	0,161	0,000	0,200	0,200	0,300	N-migrace	trénovací
424	9,780	0,288	0,000	0,380	0,065	0,380	Z-migrace	trénovací

Pokračování na další stránce

Tabulka A.1 – Pokračování z předchozí stránky

výš	míranez	hustob	bytnaek	fyznaek	rolnaek	zamnaek	saldo	mnozina
455	8,560	0,521	0,000	0,358	0,048	0,460	N-migrace	testovací
420	4,170	0,435	0,000	0,209	0,209	0,250	Z-migrace	validační
431	8,820	0,082	0,000	0,323	0,118	0,088	Z-migrace	testovací
469	8,380	0,211	0,000	0,365	0,150	0,407	Z-migrace	testovací
455	2,350	0,392	0,000	0,447	0,094	0,094	N-migrace	trénovací
440	9,300	0,123	0,000	0,372	0,000	0,163	Z-migrace	validační
554	5,170	0,203	0,002	0,256	0,044	0,320	N-migrace	validační
552	2,700	0,302	0,000	0,149	0,027	0,392	Z-migrace	trénovací
435	3,570	0,153	0,000	0,345	0,107	0,214	Z-migrace	testovací
410	8,490	0,239	0,000	0,349	0,066	0,038	Z-migrace	trénovací
465	5,730	0,536	0,009	0,381	0,023	0,473	N-migrace	trénovací
454	3,450	0,148	0,000	0,621	0,104	0,449	N-migrace	trénovací
476	15,840	0,435	0,000	0,218	0,020	0,188	N-migrace	testovací
417	12,200	0,270	0,024	0,342	0,098	0,244	N-migrace	trénovací
565	5,560	0,151	0,014	0,278	0,222	0,111	N-migrace	trénovací
478	10,730	0,193	0,000	0,311	0,186	0,260	Z-migrace	testovací
455	6,380	0,050	0,000	0,362	0,021	0,340	N-migrace	trénovací
420	7,410	0,133	0,000	0,371	0,056	0,130	N-migrace	validační
521	7,140	0,258	0,000	0,196	0,143	0,054	N-migrace	testovací
555	6,540	0,503	0,004	0,361	0,064	0,392	N-migrace	trénovací
548	2,390	0,413	0,000	0,331	0,056	1,510	N-migrace	trénovací
588	5,260	0,199	0,000	0,158	0,026	0,118	N-migrace	testovací
488	5,500	0,695	0,045	0,329	0,034	0,651	N-migrace	testovací
486	0,000	0,145	0,000	0,313	0,188	0,000	N-migrace	validační
578	7,070	0,346	0,002	0,319	0,049	0,345	N-migrace	testovací
475	8,820	0,194	0,000	0,353	0,059	0,118	Z-migrace	trénovací
418	4,740	0,691	0,005	0,400	0,042	0,390	N-migrace	trénovací
429	12,700	0,141	0,000	0,270	0,032	0,111	N-migrace	validační
450	4,090	0,493	0,005	0,387	0,065	0,464	N-migrace	testovací
460	9,090	0,317	0,000	0,333	0,000	0,121	Z-migrace	trénovací
423	4,760	0,431	0,000	0,254	0,048	0,095	N-migrace	testovací
558	0,000	0,314	0,028	0,254	0,183	0,085	N-migrace	trénovací
490	3,920	0,345	0,000	0,255	0,098	0,157	N-migrace	testovací
460	4,410	0,359	0,000	0,279	0,059	1,235	Z-migrace	testovací
502	4,450	0,441	0,001	0,445	0,052	0,474	N-migrace	validační
419	12,500	0,239	0,000	0,313	0,125	0,422	N-migrace	validační
456	14,490	0,226	0,000	0,246	0,159	0,029	Z-migrace	testovací
443	24,000	0,209	0,000	0,320	0,280	0,040	N-migrace	testovací
453	4,570	0,878	0,004	0,279	0,022	0,808	N-migrace	testovací
515	5,710	0,188	0,000	0,371	0,143	0,086	N-migrace	trénovací

Pokračování na další stránce

Tabulka A.1 – Pokračování z předchozí stránky

výš	míranez	hustob	bytnaek	fyznaek	rolnaek	zamnaek	saldo	mnozina
450	7,450	0,718	0,000	0,323	0,056	0,043	N-migrace	trénovací
535	8,720	0,313	0,035	0,319	0,147	0,275	N-migrace	testovací
468	12,500	0,253	0,000	0,250	0,063	0,000	N-migrace	testovací
505	6,060	0,194	0,008	0,265	0,114	0,401	N-migrace	testovací
425	3,540	0,364	0,035	0,319	0,009	0,266	N-migrace	testovací
465	7,690	0,151	0,000	0,269	0,385	0,282	Z-migrace	validační
503	8,260	0,223	0,000	0,281	0,074	0,694	N-migrace	trénovací
464	5,430	0,434	0,003	0,306	0,029	0,700	Z-migrace	testovací
395	5,390	1,460	0,037	0,340	0,025	2,022	N-migrace	validační
578	0,000	0,248	0,000	0,133	0,067	0,433	N-migrace	validační
627	4,760	0,249	0,000	0,262	0,071	0,119	N-migrace	validační
625	10,940	0,159	0,000	0,422	0,063	1,844	Z-migrace	trénovací
415	22,060	0,348	0,000	0,404	0,059	0,044	N-migrace	testovací
443	4,350	0,139	0,000	0,102	0,116	0,131	Z-migrace	testovací
503	4,310	0,166	0,024	0,287	0,034	0,881	N-migrace	trénovací
421	4,720	0,156	0,000	0,244	0,016	0,307	N-migrace	validační
487	3,330	0,236	0,000	0,533	0,033	0,000	N-migrace	testovací
482	3,550	0,788	0,006	0,432	0,041	0,444	N-migrace	testovací
519	5,000	0,253	0,013	0,288	0,125	0,500	N-migrace	testovací
405	7,220	0,195	0,000	0,423	0,041	0,248	N-migrace	validační
418	5,990	0,351	0,010	0,322	0,045	0,227	Z-migrace	trénovací
655	6,120	0,120	0,020	0,245	0,000	0,000	Z-migrace	validační
400	3,970	1,393	0,000	0,298	0,048	0,258	Z-migrace	trénovací
500	8,820	0,182	0,000	0,206	0,088	0,059	Z-migrace	trénovací
475	3,390	0,204	0,000	0,229	0,042	0,085	Z-migrace	validační
443	2,260	0,388	0,000	0,331	0,075	2,350	Z-migrace	validační
423	10,870	0,374	0,000	0,326	0,283	0,109	N-migrace	testovací
399	4,930	8,773	0,002	0,362	0,008	0,708	Z-migrace	testovací
430	4,830	0,310	0,000	0,348	0,063	0,145	N-migrace	testovací
498	7,250	0,221	0,000	0,305	0,145	0,754	Z-migrace	testovací
475	5,450	0,173	0,000	0,327	0,091	0,036	N-migrace	trénovací
465	5,770	0,134	0,019	0,289	0,385	0,519	N-migrace	testovací
447	3,930	0,682	0,004	0,384	0,026	1,031	N-migrace	validační
643	7,110	0,212	0,005	0,171	0,095	0,436	N-migrace	validační
405	6,350	3,655	0,002	0,357	0,017	0,906	N-migrace	trénovací
449	3,740	0,341	0,003	0,371	0,037	0,823	N-migrace	testovací
439	7,860	0,262	0,003	0,321	0,044	0,299	N-migrace	trénovací
525	4,290	0,789	0,014	0,393	0,050	0,665	Z-migrace	trénovací
419	6,450	0,285	0,013	0,323	0,161	0,181	N-migrace	trénovací
479	3,700	0,142	0,000	0,222	0,148	0,259	N-migrace	validační

Pokračování na další stránce

Tabulka A.1 – Pokračování z předchozí stránky

výš	míranez	hustob	bytnaek	fyznaek	rolnaek	zamnaek	saldo	mnozina
410	9,780	0,169	0,000	0,266	0,060	0,473	Z-migrace	validační
437	6,570	5,829	0,002	0,381	0,011	0,872	Z-migrace	trénovací
455	10,530	0,132	0,000	0,132	0,132	0,053	Z-migrace	testovací
455	7,300	0,448	0,000	0,316	0,041	0,349	N-migrace	testovací
418	1,650	0,177	0,000	0,248	0,066	0,248	Z-migrace	trénovací
405	6,310	0,277	0,000	0,433	0,081	0,667	Z-migrace	trénovací
423	7,830	0,142	0,017	0,374	0,052	0,200	Z-migrace	trénovací
465	10,000	0,324	0,000	0,327	0,145	0,227	Z-migrace	trénovací
407	7,970	2,210	0,001	0,321	0,010	0,696	Z-migrace	trénovací
519	4,400	0,245	0,011	0,297	0,022	0,121	N-migrace	validační
432	6,750	0,177	0,000	0,276	0,068	0,252	N-migrace	testovací
555	20,450	0,138	0,000	0,182	0,045	0,023	Z-migrace	testovací
414	9,090	0,223	0,000	0,364	0,036	0,036	Z-migrace	validační
601	2,300	0,124	0,012	0,242	0,081	0,023	Z-migrace	trénovací
606	12,500	0,084	0,000	0,188	0,125	0,000	Z-migrace	validační
438	14,290	0,157	0,000	0,357	0,036	0,929	N-migrace	validační
422	11,930	0,291	0,000	0,340	0,165	0,184	N-migrace	trénovací
466	8,700	0,287	0,000	0,196	0,087	0,000	N-migrace	validační
445	5,130	0,326	0,013	0,269	0,192	0,141	Z-migrace	validační
439	15,450	0,381	0,000	0,264	0,091	0,100	N-migrace	trénovací
449	8,160	0,564	0,005	0,296	0,026	0,316	N-migrace	testovací
471	3,700	0,503	0,000	0,432	0,074	0,427	N-migrace	testovací
424	3,850	0,336	0,000	0,424	0,064	0,026	N-migrace	trénovací

Data převzata z databáze Ministerstva financí (Ministerstvo financí [53]), Ministerstva práce a sociálních věcí (OKsystem [66]) a Českého statistického úřadu, pobočky v Českých Budějovicích (Český statistický úřad [11]) – rok 2006.

A.2 Splátkový prodej

Následující tabulka uvádí hodnoty 6 ekonomických ukazatelů použitých pro diskriminaci v kapitole 11:

vek – věk spotřebitele ,

pohlavi – pohlaví spotřebitele ,

prijem – čistý měsíční příjem ,

uver – výše spotřebitelského úvěru ,

splatek – počet dohodnutých splátek ,

pojisteni – typ respektive absence pojištění .

klasifikační proměnné:

bonita3: *Neplatič, Problémový, Platič*,

bonita2: *Neplatič, Platič*.

Tabulka A.2: Soubor dat použitý pro diskriminaci dle výsledků splácení spotřebitelských úvěrů

vek	pohlavi	prijem	uver	splatek	pojisteni	bonita3	bonita2
24	muž	15000	5841	10	Ne	Platič	Platič
48	žena	8500	4671	20	A	Platič	Platič
32	žena	6500	3098	30	A	Neplatič	Neplatič
59	muž	4500	3490	20	A	Platič	Platič
20	muž	10000	5332	10	Ne	Platič	Platič
23	žena	7000	11232	20	A	Platič	Platič
40	žena	5500	4564	30	A	Platič	Platič
32	muž	15000	9162	20	A	Platič	Platič
25	muž	11000	4895	30	A	Platič	Platič
21	žena	7200	6280	10	Ne	Platič	Platič
44	muž	15000	7371	20	A	Platič	Platič
48	žena	11000	7369	30	A	Platič	Platič
35	muž	6400	4499	30	A	Neplatič	Neplatič
54	žena	13500	8616	30	A	Platič	Platič
51	muž	8600	4564	30	A	Platič	Platič
34	muž	12500	12591	10	A	Platič	Platič
59	muž	7500	6480	10	A	Neplatič	Neplatič
22	žena	10000	6704	20	A	Platič	Platič
40	žena	7000	7624	30	A	Platič	Platič
69	muž	7500	7002	20	Ne	Platič	Platič
41	žena	8000	4491	20	A	Platič	Platič
23	žena	6000	5202	20	A	Platič	Platič

Pokračování na další stránce

Tabulka A.2 – Pokračování z předchozí stránky

vek	pohlavi	prijem	uver	splatek	pojisteni	bonita3	bonita2
41	muž	12000	7624	30	A	Platič	Platič
42	žena	9000	7542	20	A	Platič	Platič
21	žena	9000	7002	20	A	Platič	Platič
38	žena	7000	7002	20	A	Platič	Platič
28	muž	10000	3290	10	Ne	Platič	Platič
31	muž	18000	9919	30	A	Problémový	Neplatič
40	muž	18000	8134	30	A	Platič	Platič
45	muž	12000	11862	20	A	Platič	Platič
40	muž	13000	3499	10	Ne	Platič	Platič
32	žena	6500	3706	30	A	Neplatič	Neplatič
25	žena	10500	4041	20	A	Platič	Platič
31	muž	13000	8991	20	A	Platič	Platič
25	žena	12000	9724	10	A	Neplatič	Neplatič
53	muž	14500	13974	30	A	Platič	Platič
26	muž	12000	6936	30	A	Platič	Platič
25	žena	9500	4819	30	A	Platič	Platič
35	muž	6400	9664	30	A	Neplatič	Neplatič
30	muž	20000	6696	10	A	Platič	Platič
35	muž	15000	6579	10	A	Platič	Platič
24	muž	13000	8134	30	A	Neplatič	Neplatič
42	žena	13000	7369	30	A	Platič	Platič
29	žena	10000	7369	30	A	Neplatič	Neplatič
30	žena	8000	9827	20	A	Platič	Platič
22	žena	7500	3524	30	A	Platič	Platič
33	muž	16000	7490	30	A	Platič	Platič
22	žena	8000	6490	20	A	Platič	Platič
47	žena	8500	4990	30	A	Platič	Platič
25	žena	6500	9531	10	Ne	Platič	Platič
33	žena	10000	4490	10	Ne	Platič	Platič
47	žena	6500	7371	20	A	Platič	Platič
29	muž	12000	5841	10	Ne	Platič	Platič
19	žena	8500	5598	10	A	Neplatič	Neplatič
22	žena	7500	6525	10	A	Platič	Platič
22	žena	9000	4995	20	A	Platič	Platič
22	muž	7400	13990	20	A	Platič	Platič
39	žena	8300	4761	10	Ne	Platič	Platič
27	žena	9000	4436	20	A	Platič	Platič
27	žena	20000	6561	10	Ne	Platič	Platič
36	muž	9000	6498	20	A	Platič	Platič
24	muž	9500	6525	20	A	Platič	Platič

Pokračování na další stránce

Tabulka A.2 – Pokračování z předchozí stránky

vek	pohlavi	prijem	uver	splatek	pojisteni	bonita3	bonita2
27	muž	14500	12141	10	Ne	Platič	Platič
25	muž	7500	4041	10	Ne	Platič	Platič
44	žena	8000	3501	10	Ne	Platič	Platič
41	muž	15000	14858	30	A	Neplatič	Neplatič
26	žena	7000	7290	10	A	Neplatič	Neplatič
22	muž	12000	5940	30	A	Platič	Platič
67	muž	5100	3321	20	Ne	Neplatič	Neplatič
41	žena	14000	7911	10	Ne	Platič	Platič
62	muž	8000	3141	10	Ne	Platič	Platič
34	žena	12000	5391	20	A	Platič	Platič
22	žena	7000	7984	20	A	Neplatič	Neplatič
31	žena	11000	4491	20	A	Platič	Platič
20	muž	10000	6480	30	A	Platič	Platič
40	muž	9000	4005	10	Ne	Neplatič	Neplatič
25	žena	24000	4995	10	Ne	Platič	Platič
36	žena	10000	16286	30	A	Problémový	Neplatič
25	žena	6800	13515	30	A	Platič	Platič
24	muž	12000	5994	10	Ne	Platič	Platič
28	muž	13000	9945	30	A	Platič	Platič
42	žena	13000	7395	30	A	Platič	Platič
56	muž	10000	16320	30	A	Platič	Platič
26	muž	12000	8091	20	A	Neplatič	Neplatič
29	muž	15000	7290	10	A	Platič	Platič
37	žena	2570	3492	22	A	Neplatič	Neplatič
33	muž	10000	15984	20	A	Neplatič	Neplatič
22	muž	10000	13515	30	A	Platič	Platič
41	muž	7200	9435	30	A	Platič	Platič
20	žena	12000	12141	10	Ne	Platič	Platič
22	muž	10000	10200	30	A	Platič	Platič
72	muž	7300	6375	30	Ne	Platič	Platič
21	muž	8000	9990	10	A	Platič	Platič
54	žena	4500	9817	40	A	Platič	Platič
72	žena	8000	4203	10	Ne	Platič	Platič
39	muž	5100	3996	10	A	Neplatič	Neplatič
36	muž	7500	7004	30	A	Platič	Platič
26	muž	17000	13230	20	A	Neplatič	Neplatič
26	žena	8000	11592	20	A	Platič	Platič
47	muž	11000	8811	20	A	Platič	Platič
22	žena	7500	6561	20	A	Platič	Platič
22	muž	12000	4491	10	A	Platič	Platič

Pokračování na další stránce

Tabulka A.2 – Pokračování z předchozí stránky

vek	pohlavi	prijem	uver	splatek	pojisteni	bonita3	bonita2
31	muž	12000	8491	30	A	Platič	Platič
29	muž	13000	7395	30	A	Platič	Platič
47	muž	12000	14620	30	A	Platič	Platič
35	muž	15000	4500	20	A	Platič	Platič
26	žena	5200	4203	20	A	Platič	Platič
25	žena	7000	5100	30	A	Neplatič	Neplatič
24	muž	12000	8415	30	A	Platič	Platič
57	žena	7300	3492	20	A	Platič	Platič
43	muž	10000	5100	30	A	Platič	Platič
27	muž	7500	3996	10	A	Platič	Platič
18	žena	8000	3438	10	A	Neplatič	Neplatič
61	žena	6000	7395	30	A	Platič	Platič
35	žena	5000	5100	30	A	Platič	Platič
54	žena	4000	7395	30	A	Neplatič	Neplatič
28	žena	7000	5610	30	A	Neplatič	Neplatič
33	žena	12000	4995	20	A	Platič	Platič
22	žena	8200	13515	30	A	Neplatič	Neplatič
20	muž	9000	14040	20	A	Platič	Platič
30	muž	18000	6715	30	A	Platič	Platič
32	muž	13000	6561	10	A	Platič	Platič
35	muž	16000	4995	20	A	Platič	Platič
32	muž	18000	4491	10	A	Platič	Platič
51	žena	20000	6282	20	A	Platič	Platič
31	žena	8000	4491	10	Ne	Platič	Platič
20	muž	11000	6320	30	A	Neplatič	Neplatič
32	žena	10500	7191	20	A	Platič	Platič
25	muž	9500	11600	10	A	Platič	Platič
33	muž	15000	5391	10	Ne	Platič	Platič
27	muž	10500	9000	20	A	Platič	Platič
64	žena	7500	13515	30	A	Platič	Platič
35	muž	30000	5999	20	A	Platič	Platič
28	muž	10500	13855	30	A	Platič	Platič
32	žena	7000	7191	20	A	Platič	Platič
20	žena	4000	5264	30	A	Problémový	Neplatič
66	žena	7000	14500	30	Ne	Platič	Platič
59	žena	9500	9860	30	A	Platič	Platič
39	muž	11000	3501	10	A	Platič	Platič
19	muž	8000	6000	20	A	Problémový	Neplatič
35	muž	11000	5040	10	A	Platič	Platič
22	muž	14000	12320	20	A	Neplatič	Neplatič

Pokračování na další stránce

Tabulka A.2 – Pokračování z předchozí stránky

vek	pohlavi	prijem	uver	splatek	pojisteni	bonita3	bonita2
35	muž	30000	13515	30	A	Platič	Platič
38	muž	16000	14620	30	A	Problémový	Neplatič
30	muž	14500	6715	30	A	Platič	Platič
29	muž	15000	6561	20	A	Platič	Platič
22	muž	10000	13515	30	A	Platič	Platič
43	žena	10300	14450	30	A	Neplatič	Neplatič
22	žena	7200	8480	20	A	Neplatič	Neplatič
28	žena	8000	9280	30	A	Platič	Platič
20	muž	9000	6300	20	A	Platič	Platič
21	muž	7500	5399	10	Ne	Neplatič	Neplatič
24	žena	6500	7650	20	A	Platič	Platič
28	žena	8000	4491	20	A	Platič	Platič
24	muž	12000	6885	30	A	Platič	Platič
21	muž	9000	6799	20	A	Platič	Platič
29	muž	10000	18360	20	A	Problémový	Neplatič
20	muž	7500	5832	10	Ne	Platič	Platič
29	muž	9000	3510	10	Ne	Neplatič	Neplatič
43	muž	9000	7369	30	A	Platič	Platič
40	muž	15000	3141	10	Ne	Problémový	Neplatič
54	muž	20000	6822	20	A	Platič	Platič
41	muž	13000	10061	20	A	Neplatič	Neplatič
25	žena	10500	4761	20	A	Platič	Platič
30	žena	8000	12402	20	A	Neplatič	Neplatič
34	žena	20000	4761	10	A	Platič	Platič
49	muž	7000	5841	20	A	Platič	Platič
29	žena	8500	7501	30	A	Platič	Platič
23	žena	12000	4041	10	Ne	Platič	Platič
36	žena	10000	3141	20	A	Platič	Platič
48	muž	10000	3492	20	A	Platič	Platič
47	žena	6000	7879	30	A	Platič	Platič
26	žena	11000	6741	10	Ne	Platič	Platič
26	muž	20000	6291	10	Ne	Platič	Platič
35	muž	32000	9891	20	A	Platič	Platič
40	muž	13000	6247	12	A	Neplatič	Neplatič
20	muž	10000	3245	20	A	Platič	Platič
23	žena	13000	4041	10	Ne	Platič	Platič
38	žena	13000	6741	10	Ne	Platič	Platič
21	žena	6000	9882	20	A	Neplatič	Neplatič
39	žena	16000	4311	10	Ne	Platič	Platič
43	muž	10000	6741	10	Ne	Platič	Platič

Pokračování na další stránce

Tabulka A.2 – Pokračování z předchozí stránky

vek	pohlavi	prijem	uver	splatek	pojisteni	bonita3	bonita2
36	žena	13000	4662	20	A	Platič	Platič
54	žena	4000	8082	20	A	Platič	Platič
29	muž	14000	7002	20	A	Neplatič	Neplatič
47	žena	25000	6224	20	A	Platič	Platič
34	žena	15000	5391	20	A	Platič	Platič
28	muž	13000	4392	10	Ne	Platič	Platič
32	muž	15000	3771	11	Ne	Platič	Platič
26	žena	9000	5400	10	Ne	Platič	Platič
42	muž	10000	5373	30	A	Platič	Platič
21	žena	10000	4662	20	A	Neplatič	Neplatič
35	žena	7600	9504	20	A	Platič	Platič
29	muž	16500	7002	20	A	Platič	Platič
50	žena	6400	3402	20	A	Platič	Platič
24	žena	14000	10422	20	A	Platič	Platič
62	žena	8000	3141	11	Ne	Platič	Platič
37	muž	15000	6561	10	A	Neplatič	Neplatič
24	muž	18000	6741	10	Ne	Platič	Platič
23	muž	12200	5391	20	A	Platič	Platič
56	muž	13000	3911	20	A	Platič	Platič
30	muž	20000	12591	10	Ne	Platič	Platič
28	muž	11000	4941	10	Ne	Platič	Platič
47	muž	10000	6224	10	A	Platič	Platič
32	muž	13000	9882	20	A	Platič	Platič
57	žena	7680	12291	30	A	Platič	Platič
48	žena	11400	5382	20	A	Platič	Platič
42	muž	11000	7114	30	A	Platič	Platič
35	žena	6500	12582	20	A	Platič	Platič
53	žena	9100	5265	20	A	Platič	Platič
48	muž	12000	7114	30	A	Platič	Platič
29	muž	20000	8091	20	A	Platič	Platič
20	muž	8000	13944	10	A	Platič	Platič
39	žena	10000	11691	20	A	Platič	Platič
30	žena	9000	9990	20	A	Neplatič	Neplatič
21	muž	6500	5841	10	Ne	Problémový	Neplatič
30	žena	8000	10659	30	A	Platič	Platič
33	žena	7000	11449	30	B	Problémový	Neplatič
26	muž	7500	7290	10	A	Platič	Platič
33	muž	17000	3329	10	A	Platič	Platič
50	žena	12000	4784	20	A	Platič	Platič
29	muž	11000	7592	20	A	Platič	Platič

Pokračování na další stránce

Tabulka A.2 – Pokračování z předchozí stránky

vek	pohlavi	prijem	uver	splatek	pojisteni	bonita3	bonita2
33	žena	17200	6462	20	A	Platič	Platič
54	muž	9000	3771	10	Ne	Platič	Platič
74	muž	7000	3501	20	Ne	Platič	Platič
38	muž	8000	3141	10	A	Platič	Platič
43	žena	12000	6291	20	A	Platič	Platič
46	muž	12000	8899	30	A	Platič	Platič
23	muž	10000	9104	20	A	Platič	Platič
30	muž	14000	9087	10	A	Platič	Platič
36	žena	10100	3990	20	A	Platič	Platič
29	žena	6500	7002	20	A	Neplatič	Neplatič
46	muž	13000	7624	30	A	Platič	Platič
36	žena	8500	5841	10	Ne	Platič	Platič
46	muž	10000	5990	20	A	Platič	Platič
32	žena	6500	3299	30	A	Neplatič	Neplatič
50	žena	7000	4005	20	A	Neplatič	Neplatič
27	muž	9000	5841	10	Ne	Platič	Platič
40	muž	6000	12696	50	A	Platič	Platič
27	žena	13000	3745	20	A	Platič	Platič
22	muž	7000	5028	20	A	Platič	Platič
49	muž	11000	4941	20	A	Platič	Platič
72	žena	8400	3639	10	Ne	Platič	Platič
21	muž	12000	21984	10	Ne	Neplatič	Neplatič
40	muž	17000	4790	5	Ne	Platič	Platič
66	muž	7300	4482	10	Ne	Platič	Platič
37	žena	12000	5391	10	Ne	Platič	Platič
22	muž	9000	4374	20	A	Platič	Platič
27	žena	11000	5202	20	A	Platič	Platič
55	muž	9000	7624	50	A	Platič	Platič
34	žena	10000	3599	5	Ne	Platič	Platič
24	žena	14000	8549	10	Ne	Platič	Platič
39	žena	11000	4499	20	A	Platič	Platič
20	žena	8000	3245	10	Ne	Platič	Platič
58	žena	5129	3762	20	A	Platič	Platič
30	žena	6000	3762	20	A	Platič	Platič
44	muž	6500	3762	10	A	Platič	Platič
60	žena	6000	3992	20	A	Platič	Platič
29	žena	13000	3411	10	Ne	Platič	Platič
22	muž	20000	7542	20	A	Platič	Platič
22	muž	9800	9154	30	A	Neplatič	Neplatič
22	žena	10500	8082	20	A	Platič	Platič

Pokračování na další stránce

Tabulka A.2 – Pokračování z předchozí stránky

vek	pohlavi	prijem	uver	splatek	pojisteni	bonita3	bonita2
53	muž	11702	5202	20	A	Platič	Platič
40	muž	25000	8293	10	Ne	Platič	Platič
53	žena	4500	4662	20	A	Platič	Platič
58	muž	9000	5922	20	A	Platič	Platič
46	muž	12000	4491	10	Ne	Platič	Platič
34	muž	13000	8541	10	Ne	Platič	Platič
22	muž	10000	10782	20	A	Platič	Platič
32	žena	8500	4662	20	A	Platič	Platič
47	muž	12000	8082	20	A	Platič	Platič
28	muž	18000	3402	20	A	Platič	Platič
31	žena	7500	4482	20	A	Neplatič	Neplatič
54	žena	8754	11883	30	A	Platič	Platič
32	žena	5000	10174	30	A	Platič	Platič
56	muž	10000	3492	20	A	Platič	Platič
24	žena	15000	6291	20	A	Platič	Platič
40	muž	25000	7515	10	Ne	Platič	Platič
34	muž	12000	9264	20	A	Platič	Platič
30	žena	7500	6604	30	A	Platič	Platič
29	muž	14000	7542	20	A	Platič	Platič
45	žena	18000	6291	10	Ne	Platič	Platič
35	muž	12000	7002	20	A	Platič	Platič
37	muž	12500	10914	30	A	Platič	Platič
32	muž	15000	5490	10	Ne	Platič	Platič
22	žena	8000	10242	20	A	Platič	Platič
39	žena	10000	8190	20	A	Platič	Platič
63	žena	8050	3492	10	Ne	Platič	Platič
22	žena	11000	14004	20	A	Platič	Platič
40	muž	12000	4005	10	Ne	Platič	Platič
27	muž	16000	10341	10	Ne	Platič	Platič
45	žena	18000	7731	10	Ne	Platič	Platič
36	žena	9500	5391	10	Ne	Platič	Platič
56	muž	11000	4941	10	Ne	Platič	Platič
40	žena	12800	14022	20	A	Platič	Platič
22	muž	13500	3492	10	Ne	Platič	Platič
46	žena	22000	5022	20	A	Platič	Platič
36	žena	7100	5880	10	Ne	Platič	Platič
39	žena	7500	13829	30	A	Platič	Platič
32	žena	10000	6993	10	Ne	Platič	Platič
42	muž	12000	7191	10	Ne	Platič	Platič
29	muž	10000	3492	10	Ne	Platič	Platič

Pokračování na další stránce

Tabulka A.2 – Pokračování z předchozí stránky

vek	pohlavi	prijem	uver	splatek	pojisteni	bonita3	bonita2
40	žena	4500	3402	20	A	Platič	Platič
26	muž	14500	9537	30	A	Platič	Platič
27	muž	15000	11959	30	A	Platič	Platič
27	muž	15000	8376	30	A	Problémový	Neplatič
33	žena	11000	5310	10	Ne	Platič	Platič
47	žena	6700	7182	20	A	Platič	Platič
41	žena	16000	4819	30	A	Platič	Platič
27	muž	10500	11449	30	A	Neplatič	Neplatič
70	žena	7299	9891	11	Ne	Platič	Platič
41	žena	7000	11142	20	A	Platič	Platič
20	žena	7500	11449	30	A	Neplatič	Neplatič
50	muž	15000	7461	20	A	Platič	Platič
45	žena	7500	7369	30	A	Platič	Platič
25	muž	12000	10341	20	A	Platič	Platič
37	žena	11000	5839	30	A	Platič	Platič
38	žena	10000	3402	20	A	Platič	Platič
26	žena	12000	4842	20	A	Platič	Platič
58	muž	9800	6822	20	A	Problémový	Neplatič
29	muž	10000	14022	20	A	Platič	Platič
33	muž	18000	7182	20	A	Platič	Platič
24	žena	9500	5562	20	A	Platič	Platič
21	žena	8000	11959	30	A	Platič	Platič
20	muž	8500	4194	20	A	Platič	Platič
23	muž	10500	8541	10	Ne	Platič	Platič
21	muž	12000	7182	20	A	Platič	Platič
27	žena	7000	11142	20	A	Platič	Platič
24	žena	9000	15804	20	A	Platič	Platič
54	muž	8500	7002	20	A	Platič	Platič
52	žena	14000	5022	20	A	Platič	Platič
20	muž	14000	7640	10	Ne	Platič	Platič
46	žena	9500	5562	20	A	Platič	Platič
25	žena	10000	7542	20	A	Platič	Platič
39	muž	14000	10174	30	A	Platič	Platič
20	žena	10000	11449	30	A	Platič	Platič
32	žena	6000	5382	20	A	Platič	Platič
56	žena	7500	6804	20	A	Platič	Platič
29	muž	11000	6094	30	A	Platič	Platič
20	žena	7500	14254	30	A	Platič	Platič
31	muž	35000	11025	10	Ne	Platič	Platič
25	muž	15000	7731	20	A	Platič	Platič

Pokračování na další stránce

Tabulka A.2 – Pokračování z předchozí stránky

vek	pohlavi	prijem	uver	splatek	pojisteni	bonita3	bonita2
28	žena	8000	5695	30	A	Platič	Platič
48	žena	16100	3699	20	A	Platič	Platič
58	žena	6740	9520	30	A	Platič	Platič
20	žena	8500	4941	11	Ne	Platič	Platič
68	muž	6652	5715	20	Ne	Platič	Platič
34	muž	15000	13572	20	A	Platič	Platič
51	žena	10000	7461	11	Ne	Platič	Platič
36	muž	17000	3141	11	Ne	Platič	Platič
27	muž	14000	6102	20	A	Neplatič	Neplatič
42	muž	12682	3492	20	A	Platič	Platič
38	žena	10500	6349	30	A	Platič	Platič
25	muž	13000	10782	20	A	Platič	Platič
54	žena	4500	4203	20	A	Platič	Platič
22	muž	12000	6299	10	Ne	Platič	Platič
59	žena	7200	5390	20	A	Platič	Platič
27	žena	7000	7542	20	A	Platič	Platič
50	muž	61000	6216	10	Ne	Neplatič	Neplatič
29	muž	10000	7002	20	A	Platič	Platič
31	muž	15000	8982	10	Ne	Platič	Platič
19	žena	5500	6696	10	A	Neplatič	Neplatič
22	muž	10000	4621	20	A	Platič	Platič
34	muž	15159	7624	30	A	Platič	Platič
45	žena	10000	7515	20	A	Platič	Platič
24	muž	12000	6680	20	A	Platič	Platič
30	muž	12500	3228	20	A	Platič	Platič
25	muž	14000	14229	30	A	Platič	Platič
27	muž	12500	10791	10	Ne	Platič	Platič
55	muž	8000	8540	50	A	Platič	Platič
37	muž	19000	4203	10	Ne	Platič	Platič
36	žena	11000	5841	10	Ne	Platič	Platič
41	žena	12000	7624	30	A	Platič	Platič
40	žena	8700	14254	30	A	Neplatič	Neplatič
48	muž	8000	6642	20	A	Platič	Platič
38	žena	7500	5789	20	A	Platič	Platič
33	žena	11500	6299	10	Ne	Platič	Platič
18	muž	9000	4624	10	A	Neplatič	Neplatič
38	žena	7500	6822	20	A	Platič	Platič
34	muž	13000	5999	20	A	Platič	Platič
21	muž	9500	5211	10	Ne	Platič	Platič
37	žena	10500	6741	10	Ne	Platič	Platič

Pokračování na další stránce

Tabulka A.2 – Pokračování z předchozí stránky

vek	pohlavi	prijem	uver	splatek	pojisteni	bonita3	bonita2
22	žena	9000	5391	10	Ne	Platič	Platič
38	muž	16000	5751	10	Ne	Platič	Platič
35	žena	7918	6390	20	A	Platič	Platič
29	muž	8000	3510	20	A	Neplatič	Neplatič
49	muž	15000	3490	10	Ne	Platič	Platič
23	muž	10000	13050	20	A	Platič	Platič
22	muž	14000	8091	10	Ne	Neplatič	Neplatič
32	muž	26000	6291	10	Ne	Platič	Platič
18	muž	8000	5680	10	A	Platič	Platič
30	muž	10000	3510	20	A	Platič	Platič
49	žena	8000	3141	10	Ne	Platič	Platič
21	žena	7000	14040	20	A	Neplatič	Neplatič
34	muž	6670	9810	20	A	Platič	Platič
39	muž	10000	3900	20	A	Platič	Platič
26	muž	8500	6643	30	A	Platič	Platič
53	žena	12000	4050	20	B	Platič	Platič
34	žena	10000	6502	30	A	Platič	Platič
52	muž	11000	11645	30	A	Platič	Platič
24	žena	5900	7020	20	A	Platič	Platič
27	muž	13000	6525	20	A	Platič	Platič
35	žena	7300	4995	10	Ne	Platič	Platič
42	žena	10000	12510	20	A	Platič	Platič
20	žena	6500	6987	30	A	Platič	Platič
22	žena	12000	9000	20	A	Platič	Platič
48	muž	15000	14577	30	A	Platič	Platič
24	muž	10000	4491	10	Ne	Platič	Platič
37	muž	10000	6885	30	A	Neplatič	Neplatič
49	muž	13000	7012	30	A	Platič	Platič
57	žena	5400	5391	10	Ne	Platič	Platič
24	muž	9000	6291	20	A	Platič	Platič
23	muž	9000	3149	10	Ne	Platič	Platič
20	žena	5600	6741	10	Ne	Platič	Platič
31	muž	9500	7395	30	A	Neplatič	Neplatič
33	žena	10000	9435	30	A	Neplatič	Neplatič
38	muž	25000	16320	30	A	Platič	Platič
26	muž	7500	7290	20	A	Platič	Platič
47	muž	3707	3149	33	Ne	Platič	Platič
52	žena	11000	8500	30	A	Platič	Platič
44	žena	13000	6885	30	A	Platič	Platič
33	muž	12000	8982	10	Ne	Platič	Platič

Pokračování na další stránce

Tabulka A.2 – Pokračování z předchozí stránky

vek	pohlavi	prijem	uver	splatek	pojisteni	bonita3	bonita2
18	muž	10000	5032	10	A	Neplatič	Neplatič
36	žena	9000	12591	10	A	Platič	Platič
45	žena	15000	13005	30	A	Platič	Platič
38	žena	10000	6570	10	A	Platič	Platič
31	muž	18500	3150	10	Ne	Platič	Platič
24	muž	11000	5931	10	Ne	Platič	Platič
27	muž	18000	6885	30	A	Platič	Platič
40	muž	15000	7770	30	A	Platič	Platič
21	muž	7500	9775	30	A	Neplatič	Neplatič
59	žena	5400	13993	50	A	Platič	Platič
49	žena	9300	4005	10	Ne	Platič	Platič
24	žena	9500	6375	30	A	Platič	Platič
22	muž	8500	13728	30	A	Platič	Platič
22	žena	12000	9945	30	A	Platič	Platič
27	žena	8400	8993	30	A	Neplatič	Neplatič
44	žena	9500	6375	30	A	Platič	Platič
46	žena	9600	4050	10	Ne	Platič	Platič
30	žena	9200	9945	30	A	Neplatič	Neplatič
35	žena	10000	6375	30	A	Platič	Platič
44	žena	6600	5100	30	A	Platič	Platič
33	žena	6500	4049	22	Ne	Platič	Platič
62	žena	8500	3321	10	Ne	Platič	Platič
33	žena	6000	7293	30	A	Platič	Platič
45	muž	8000	5100	30	A	Platič	Platič
22	muž	7500	5100	30	A	Platič	Platič
40	žena	11000	4950	10	Ne	Platič	Platič
25	žena	4700	9435	30	A	Platič	Platič
37	žena	14000	12282	30	A	Platič	Platič
51	muž	15000	5202	20	A	Platič	Platič
20	muž	8000	7002	20	A	Platič	Platič
31	muž	10500	4482	20	A	Platič	Platič
27	muž	15000	4004	10	Ne	Platič	Platič
26	muž	9200	4662	20	A	Platič	Platič
39	muž	20000	3329	10	Ne	Platič	Platič
46	žena	11000	5382	20	A	Platič	Platič
57	muž	11980	10782	20	A	Platič	Platič
33	muž	9900	5500	20	A	Platič	Platič
19	muž	6500	11322	20	A	Platič	Platič
41	žena	20000	12402	10	A	Platič	Platič
32	muž	9400	10422	20	A	Platič	Platič

Pokračování na další stránce

Tabulka A.2 – Pokračování z předchozí stránky

vek	pohlavi	prijem	uver	splatek	pojisteni	bonita3	bonita2
29	muž	10500	8073	30	A	Platič	Platič
28	muž	10000	8082	20	A	Platič	Platič
21	muž	11000	5382	20	A	Platič	Platič
44	žena	8000	5202	20	A	Platič	Platič
53	žena	10000	3771	10	Ne	Platič	Platič
37	muž	10000	4049	10	Ne	Platič	Platič
66	žena	7000	15282	20	Ne	Platič	Platič
34	žena	20000	7992	20	A	Platič	Platič
57	žena	5000	3402	20	A	Platič	Platič
48	muž	16000	5202	20	B	Platič	Platič
26	muž	11000	7542	20	B	Platič	Platič
55	muž	20000	4482	20	A	Platič	Platič
39	žena	12000	7803	30	A	Platič	Platič
48	muž	10000	5697	10	Ne	Problémový	Neplatič
45	muž	11500	11322	20	A	Platič	Platič
37	muž	11000	8082	20	A	Platič	Platič
22	muž	13500	5382	20	A	Platič	Platič
26	muž	10000	4500	10	Ne	Platič	Platič
27	žena	10500	5040	20	A	Neplatič	Neplatič
42	muž	14000	8010	20	A	Platič	Platič
34	žena	8000	8100	20	A	Neplatič	Neplatič
36	muž	10400	5022	20	A	Platič	Platič
37	muž	12000	5841	10	Ne	Platič	Platič
27	žena	10500	4842	20	A	Neplatič	Neplatič
59	žena	5285	8082	20	A	Platič	Platič
35	muž	15000	3141	20	A	Platič	Platič
28	muž	14310	5841	10	Ne	Platič	Platič
41	muž	14200	9639	30	A	Platič	Platič
24	muž	15000	8082	20	A	Platič	Platič
22	muž	11500	8082	20	A	Platič	Platič
18	muž	12720	11502	20	A	Platič	Platič
25	žena	8500	13995	20	A	Platič	Platič
25	muž	8500	6498	20	A	Platič	Platič
27	muž	10500	4041	10	Ne	Platič	Platič
33	muž	12000	4499	20	A	Platič	Platič
41	žena	9000	5292	10	Ne	Platič	Platič
21	muž	14000	8442	20	A	Platič	Platič
51	muž	30000	9882	20	A	Platič	Platič
33	muž	14000	4662	20	A	Platič	Platič
28	muž	12000	7497	20	A	Platič	Platič

Pokračování na další stránce

Tabulka A.2 – Pokračování z předchozí stránky

vek	pohlavi	prijem	uver	splatek	pojisteni	bonita3	bonita2
39	žena	12618	7182	20	A	Platič	Platič
56	žena	7300	6561	10	Ne	Platič	Platič
51	muž	6000	8899	30	A	Platič	Platič
35	muž	15000	10962	20	A	Platič	Platič
29	žena	7500	3141	10	Ne	Platič	Platič
33	muž	15000	5391	10	Ne	Platič	Platič
35	muž	14000	8991	10	A	Platič	Platič
26	žena	12000	11142	20	A	Platič	Platič
23	muž	20000	25164	20	A	Platič	Platič
20	muž	8500	3599	10	Ne	Platič	Platič
24	muž	20000	9576	30	A	Platič	Platič
49	muž	14000	8442	20	A	Platič	Platič
29	žena	17000	15093	30	A	Platič	Platič
31	žena	8500	5022	20	A	Platič	Platič
53	muž	6700	15084	20	A	Platič	Platič
33	muž	11000	6462	20	A	Platič	Platič
23	žena	8000	11502	20	A	Platič	Platič
32	muž	10000	5562	20	A	Platič	Platič
40	muž	31000	7182	20	A	Platič	Platič
48	muž	13500	7002	20	A	Platič	Platič
29	muž	19000	8442	20	A	Platič	Platič
26	žena	7500	5580	20	A	Platič	Platič
21	žena	7800	7879	30	A	Platič	Platič
50	žena	6954	4819	30	A	Platič	Platič
61	muž	7500	7002	20	A	Platič	Platič
47	muž	9000	4842	20	A	Platič	Platič
19	muž	13000	4662	20	A	Platič	Platič
25	žena	8500	7182	20	A	Platič	Platič
43	žena	7000	7542	20	A	Platič	Platič
42	muž	12000	11504	20	A	Platič	Platič
50	žena	10000	3402	20	A	Platič	Platič
36	muž	16000	10242	20	A	Platič	Platič
40	žena	8000	9666	30	A	Platič	Platič
20	žena	12380	11142	20	A	Platič	Platič
21	muž	6500	5562	20	A	Platič	Platič
31	muž	10500	7542	20	A	Platič	Platič
55	muž	12500	11142	20	A	Platič	Platič
37	žena	8000	5382	20	A	Platič	Platič
29	muž	10000	4049	10	Ne	Platič	Platič
55	muž	14000	8442	20	A	Platič	Platič

Pokračování na další stránce

Tabulka A.2 – Pokračování z předchozí stránky

vek	pohlavi	prijem	uver	splatek	pojisteni	bonita3	bonita2
21	muž	10000	14022	20	A	Neplatič	Neplatič
52	žena	7500	5742	20	A	Platič	Platič
23	žena	7000	7049	10	Ne	Platič	Platič
26	žena	8000	12933	30	A	Problémový	Neplatič
19	žena	6500	10242	20	A	Platič	Platič
49	muž	10000	8802	20	A	Platič	Platič
34	žena	11000	11959	30	A	Platič	Platič
57	muž	11500	14904	20	A	Platič	Platič
26	muž	10000	3942	20	A	Platič	Platič
18	muž	10000	8442	20	A	Platič	Platič
44	žena	6500	7182	20	A	Platič	Platič
44	muž	13500	6299	30	A	Platič	Platič
27	muž	12000	12123	30	A	Platič	Platič
19	muž	15000	11780	20	A	Platič	Platič
27	muž	8000	12123	30	A	Platič	Platič
52	muž	12000	3492	20	A	Platič	Platič
45	žena	6700	5301	10	Ne	Platič	Platič
23	muž	11000	14364	20	A	Platič	Platič
52	žena	9000	15093	30	A	Platič	Platič
23	muž	11000	6462	20	A	Platič	Platič
64	muž	8500	3492	20	A	Platič	Platič
19	žena	8000	11502	20	A	Platič	Platič
25	muž	17000	6384	20	A	Platič	Platič
29	muž	5580	5022	20	A	Platič	Platič
30	muž	23000	7533	30	A	Platič	Platič
41	žena	9000	4662	20	A	Platič	Platič
54	muž	10000	5103	30	A	Platič	Platič
29	žena	5800	5562	20	A	Platič	Platič
31	muž	9000	4320	20	A	Neplatič	Neplatič
30	žena	9000	8442	20	A	Neplatič	Neplatič
57	muž	10000	8442	20	A	Platič	Platič
23	žena	6000	5382	20	A	Neplatič	Neplatič
40	muž	13000	5022	10	A	Neplatič	Neplatič
32	muž	15000	9684	20	A	Platič	Platič
32	muž	11000	3402	20	A	Platič	Platič
42	muž	14000	12886	30	A	Platič	Platič
19	muž	9000	9904	20	A	Problémový	Neplatič
36	žena	8000	8991	20	A	Platič	Platič
34	muž	29780	8091	20	A	Platič	Platič
46	muž	12000	4491	10	Ne	Platič	Platič

Pokračování na další stránce

Tabulka A.2 – Pokračování z předchozí stránky

vek	pohlavi	prijem	uver	splatek	pojisteni	bonita3	bonita2
24	žena	10000	7497	20	B	Platič	Platič
21	muž	9000	6498	20	A	Platič	Platič
27	muž	8100	7290	10	Ne	Platič	Platič
26	muž	11500	4080	20	B	Platič	Platič
29	muž	12500	5490	20	A	Platič	Platič
24	muž	12705	3141	10	Ne	Platič	Platič
19	žena	6000	4896	20	A	Platič	Platič
29	žena	7500	5490	10	A	Platič	Platič
44	žena	8000	5994	20	B	Platič	Platič
20	žena	8000	4491	20	A	Platič	Platič
47	žena	8800	4590	20	A	Platič	Platič
38	muž	6608	10782	30	B	Neplatič	Neplatič
20	muž	8611	8540	20	A	Neplatič	Neplatič
40	muž	13100	12996	20	A	Platič	Platič
44	žena	10000	7542	20	A	Platič	Platič
28	muž	10000	4499	10	Ne	Platič	Platič
22	muž	6000	7497	20	A	Platič	Platič
31	muž	10991	7497	20	A	Platič	Platič
39	žena	15000	5841	20	A	Platič	Platič
37	žena	11000	4392	10	Ne	Platič	Platič
46	muž	20000	6111	10	Ne	Platič	Platič
49	žena	9000	13489	30	A	Platič	Platič
41	muž	10000	9963	30	A	Platič	Platič
46	žena	8500	9180	20	A	Problémový	Neplatič
68	muž	7950	7506	20	Ne	Platič	Platič
23	žena	8000	12492	20	A	Platič	Platič
34	muž	10000	7624	30	B	Platič	Platič
29	žena	6500	9792	10	A	Platič	Platič
39	žena	11500	1584	30	A	Platič	Platič
42	muž	11500	7290	30	B	Neplatič	Neplatič
58	muž	12000	3031	10	Ne	Platič	Platič
40	muž	13000	9531	10	Ne	Neplatič	Neplatič
51	muž	9000	4401	10	A	Platič	Platič
25	muž	11000	4401	20	A	Platič	Platič
23	žena	12000	5994	20	A	Platič	Platič
38	muž	13770	4842	20	A	Platič	Platič
47	muž	14500	3501	20	A	Neplatič	Neplatič
24	muž	9500	18819	30	A	Platič	Platič
51	žena	11500	4990	10	Ne	Neplatič	Neplatič
40	žena	8982	8982	20	A	Platič	Platič

Pokračování na další stránce

Tabulka A.2 – Pokračování z předchozí stránky

vek	pohlavi	prijem	uver	splatek	pojisteni	bonita3	bonita2
21	muž	10000	4999	20	A	Neplatič	Neplatič
21	žena	7000	8134	30	A	Platič	Platič
40	muž	13000	9504	20	A	Platič	Platič
27	muž	11000	6462	20	A	Platič	Platič
21	žena	9000	8091	10	Ne	Platič	Platič
18	žena	6000	8091	10	Ne	Platič	Platič
19	žena	6300	5841	20	A	Platič	Platič
28	žena	15000	5391	20	A	Platič	Platič
43	muž	11500	7938	30	A	Platič	Platič
22	žena	13000	7290	20	A	Platič	Platič
31	muž	19500	15984	20	A	Platič	Platič
24	muž	11000	7290	20	A	Platič	Platič
29	muž	21000	12599	30	A	Platič	Platič
30	muž	14000	6876	30	A	Platič	Platič
24	muž	16000	13499	20	A	Platič	Platič
25	muž	9000	5994	20	A	Platič	Platič
33	muž	22000	14484	30	A	Platič	Platič
53	žena	8000	10579	30	A	Platič	Platič
19	muž	10000	8982	20	A	Neplatič	Neplatič
19	žena	8500	4392	10	Ne	Platič	Platič
18	muž	8000	8061	30	A	Platič	Platič
34	muž	12000	6462	20	A	Platič	Platič
50	muž	10000	15093	30	A	Platič	Platič
23	muž	9500	9409	30	A	Platič	Platič
61	muž	8000	13107	30	A	Neplatič	Neplatič
29	muž	25000	9154	30	A	Platič	Platič
31	žena	7000	4500	20	A	Platič	Platič
36	muž	25000	5399	10	Ne	Platič	Platič
27	muž	13000	7461	10	Ne	Platič	Platič
21	muž	7000	5999	10	Ne	Neplatič	Neplatič
25	žena	10000	5391	10	Ne	Platič	Platič
26	muž	11500	9792	10	Ne	Platič	Platič
32	muž	10500	9775	30	A	Platič	Platič
34	muž	12842	12141	10	A	Neplatič	Neplatič
49	muž	6000	5309	10	Ne	Platič	Platič
25	muž	15000	10701	10	A	Platič	Platič
25	žena	7000	5841	10	Ne	Neplatič	Neplatič
21	žena	5000	8784	20	A	Neplatič	Neplatič
41	muž	15000	4491	10	Ne	Platič	Platič
25	žena	9560	6561	10	Ne	Platič	Platič

Pokračování na další stránce

Tabulka A.2 – Pokračování z předchozí stránky

vek	pohlavi	prijem	uver	splatek	pojisteni	bonita3	bonita2
25	muž	6000	3490	10	Ne	Neplatič	Neplatič
30	žena	9000	5399	10	Ne	Platič	Platič
58	žena	7000	4491	20	A	Platič	Platič
28	muž	15000	8540	10	Ne	Platič	Platič
19	muž	5000	9512	20	A	Platič	Platič
29	žena	8000	15580	30	A	Platič	Platič
48	muž	13000	6561	10	Ne	Platič	Platič
40	muž	25000	15291	20	A	Platič	Platič
38	muž	20000	3140	10	Ne	Platič	Platič
30	muž	13000	5391	10	Ne	Platič	Platič
32	muž	14000	7830	20	A	Platič	Platič
35	žena	7000	13489	20	A	Platič	Platič
40	muž	12000	5841	20	A	Platič	Platič
43	žena	9000	3499	13	Ne	Platič	Platič
29	muž	12000	8091	20	A	Platič	Platič
26	muž	12000	5324	20	A	Platič	Platič
32	muž	12000	3950	10	Ne	Platič	Platič
25	muž	10000	8990	20	A	Platič	Platič
19	žena	10000	9792	20	A	Platič	Platič
40	muž	16000	3690	10	B	Neplatič	Neplatič
28	muž	16000	8990	20	A	Platič	Platič
64	žena	6481	3510	10	Ne	Platič	Platič
42	žena	6300	7290	10	Ne	Platič	Platič
27	muž	15000	8427	20	A	Platič	Platič
37	žena	8000	4491	20	A	Platič	Platič
21	muž	10000	6111	10	A	Platič	Platič
43	muž	17000	13490	20	A	Platič	Platič
28	muž	15000	8702	20	A	Platič	Platič
21	žena	9500	6790	5	Ne	Platič	Platič
29	žena	9000	3869	10	Ne	Platič	Platič
36	muž	20000	4401	10	Ne	Platič	Platič
63	žena	8000	14492	30	A	Platič	Platič
34	muž	15000	3663	10	A	Platič	Platič
28	muž	14000	6498	20	A	Platič	Platič
52	muž	13000	13491	10	Ne	Platič	Platič
22	žena	12000	8991	10	B	Platič	Platič
24	muž	12000	3112	10	Ne	Neplatič	Neplatič
57	muž	15000	13990	20	A	Platič	Platič
32	muž	14000	7290	30	A	Platič	Platič
41	žena	8000	7490	20	A	Platič	Platič

Pokračování na další stránce

Tabulka A.2 – Pokračování z předchozí stránky

vek	pohlavi	prijem	uver	splatek	pojisteni	bonita3	bonita2
24	muž	16000	4995	20	A	Neplatič	Neplatič
22	muž	13000	13991	30	A	Neplatič	Neplatič
40	muž	8000	13168	30	A	Neplatič	Neplatič
29	muž	15000	4990	10	Ne	Platič	Platič
73	žena	7300	6669	10	Ne	Platič	Platič
32	muž	11000	5994	20	A	Platič	Platič
22	žena	10000	6561	20	A	Platič	Platič
23	muž	18000	3490	10	A	Platič	Platič
31	muž	9000	9086	30	A	Neplatič	Neplatič
21	muž	10000	12291	30	A	Platič	Platič
55	žena	9000	3897	20	A	Platič	Platič
27	žena	8000	11988	20	A	Platič	Platič
20	žena	9000	3168	10	A	Neplatič	Neplatič
30	žena	9000	8541	10	Ne	Platič	Platič
22	muž	30000	8991	20	A	Platič	Platič
40	žena	10000	4990	20	A	Platič	Platič
30	muž	11000	6795	20	A	Platič	Platič
32	muž	14000	6669	20	A	Platič	Platič
41	muž	17000	3912	20	A	Platič	Platič
33	žena	9000	3492	20	A	Platič	Platič
23	muž	15000	3992	10	A	Platič	Platič
33	muž	16000	5491	30	A	Platič	Platič
44	muž	7000	17864	30	A	Neplatič	Neplatič
43	muž	10000	7590	30	A	Platič	Platič
27	muž	15000	7290	20	A	Platič	Platič
28	žena	7000	9792	20	A	Platič	Platič
41	muž	5000	6993	20	A	Neplatič	Neplatič
55	muž	13000	8982	10	A	Platič	Platič
29	žena	20000	5112	10	Ne	Platič	Platič

Data byla poskytnuta nejmenovanou firmou, rok 2004–2007.

A.3 Nezaměstnanost

Následující tabulka uvádí hodnoty vývoje míry nezaměstnanosti od roku 1993, které byly použity pro analýzu a predikci v časových řadách v kapitole 12. Míra nezaměstnanosti¹ byla vypočítávána dle metodiky metodiky Ministerstva práce a sociálních věcí (OKsystem [66]). Tabulka taktéž obsahuje hodnoty sezónně očištěných řad:

datum – měsíc.rok míra nezaměstnanosti je zjišťována k poslednímu dni v měsíci

i rok $i = 1, \dots, 11$,

j sezóna $j = 1, \dots, 12$,

t_{ij} $t_{ij} = 12 \cdot (i - 1) + j$, $t_{ij} = 1, \dots, 124$

y míra nezaměstnanosti,

metodika metodiky výpočtu míry nezaměstnanosti, respektive identifikátor určující hodnoty pro tvorbu modelu a pro ověření predikční síly modelu,

y_1 míra nezaměstnanosti očištěná od konstantní sezónnosti, sezónní indexy b_j vypočítané na základě hodnot z let 1993–2008,

y_2 míra nezaměstnanosti očištěná od proporcionální sezónnosti, sezónní indexy $(1 + c_j)$ vypočítané na základě hodnot z let 1993–2008,

y_3 míra nezaměstnanosti očištěná od konstantní sezónnosti (od roku 2004), sezónní indexy b_j vypočítané na základě hodnot z let 2004–2008,

y_4 míra nezaměstnanosti očištěná od proporcionální sezónnosti (od roku 2004), sezónní indexy $(1 + c_j)$ vypočítané na základě hodnot z let 2004–2008.

Tabulka A.3: Vývoj míry nezaměstnanosti v letech 1993–2009 a hodnoty sezónně očištěné

datum	i	j	t_{ij}	y	metodika	y_1	y_2	y_3	y_4
I.99	1	1	1	8,053	stará	7,544	7,585		
II.99	1	2	2	8,267	stará	7,831	7,851		
III.99	1	3	3	8,370	stará	8,169	8,172		
IV.99	1	4	4	8,176	stará	8,359	8,359		
V.99	1	5	5	8,131	stará	8,549	8,564		
VI.99	1	6	6	8,390	stará	8,793	8,825		
VII.99	1	7	7	8,794	stará	8,834	8,831		
VIII.99	1	8	8	8,962	stará	8,920	8,915		

Pokračování na další stránce

¹Míra nezaměstnanosti, jak ji vypočítává Ministerstvo práce a sociálních věcí se správně nazývá „registrovaná míra nezaměstnanosti“. Používání plného názvu by zbytečně komplikovalo popisky a text disertační práce a proto nebylo používáno.

datum	i	j	t_{ij}	y	metodika	y_1	y_2	y_3	y_4
IX.99	1	9	9	9,046	stará	9,050	9,049		
X.99	1	10	10	8,918	stará	9,110	9,129		
XI.99	1	11	11	8,955	stará	9,154	9,178		
XII.99	1	12	12	9,371	stará	9,122	9,103		
I.00	2	1	13	9,762	stará	9,252	9,194		
II.00	2	2	14	9,717	stará	9,281	9,228		
III.00	2	3	15	9,473	stará	9,273	9,250		
IV.00	2	4	16	9,039	stará	9,222	9,242		
V.00	2	5	17	8,706	stará	9,124	9,170		
VI.00	2	6	18	8,660	stará	9,062	9,108		
VII.00	2	7	19	9,011	stará	9,051	9,049		
VIII.00	2	8	20	8,964	stará	8,921	8,916		
IX.00	2	9	21	8,791	stará	8,794	8,793		
X.00	2	10	22	8,543	stará	8,735	8,745		
XI.00	2	11	23	8,487	stará	8,686	8,698		
XII.00	2	12	24	8,777	stará	8,528	8,526		
I.01	3	1	25	9,113	stará	8,604	8,583		
II.01	3	2	26	8,960	stará	8,525	8,510		
III.01	3	3	27	8,680	stará	8,479	8,475		
IV.01	3	4	28	8,332	stará	8,514	8,519		
V.01	3	5	29	8,086	stará	8,504	8,517		
VI.01	3	6	30	8,081	stará	8,483	8,499		
VII.01	3	7	31	8,462	stará	8,502	8,498		
VIII.01	3	8	32	8,536	stará	8,494	8,491		
IX.01	3	9	33	8,476	stará	8,479	8,478		
X.01	3	10	34	8,423	stará	8,614	8,622		
XI.01	3	11	35	8,458	stará	8,658	8,669		
XII.01	3	12	36	8,897	stará	8,648	8,643		
I.02	4	1	37	9,414	stará	8,905	8,866		
II.02	4	2	38	9,342	stará	8,906	8,872		
III.02	4	3	39	9,081	stará	8,881	8,867		
IV.02	4	4	40	8,771	stará	8,954	8,968		
V.02	4	5	41	8,607	stará	9,025	9,065		
VI.02	4	6	42	8,730	stará	9,133	9,183		
VII.02	4	7	43	9,183	stará	9,223	9,222		
VIII.02	4	8	44	9,357	stará	9,314	9,307		
IX.02	4	9	45	9,445	stará	9,448	9,447		
X.02	4	10	46	9,278	stará	9,470	9,497		
XI.02	4	11	47	9,337	stará	9,536	9,570		
XII.02	4	12	48	9,807	stará	9,558	9,527		
I.03	5	1	49	10,221	stará	9,712	9,627		

Pokračování na další stránce

datum	i	j	t_{ij}	y	metodika	y_1	y_2	y_3	y_4
II.03	5	2	50	10,204	stará	9,768	9,691		
III.03	5	3	51	10,016	stará	9,816	9,780		
IV.03	5	4	52	9,633	stará	9,816	9,849		
V.03	5	5	53	9,443	stará	9,861	9,945		
VI.03	5	6	54	9,523	stará	9,926	10,016		
VII.03	5	7	55	9,881	stará	9,921	9,923		
VIII.03	5	8	56	9,968	stará	9,926	9,916		
IX.03	5	9	57	10,052	stará	10,056	10,055		
X.03	5	10	58	9,925	stará	10,117	10,159		
XI.03	5	11	59	9,900	stará	10,099	10,146		
XII.03	5	12	60	10,306	stará	10,057	10,011		
I.04	6	1	61	9,691	nová	9,182	9,128	9,125	9,016
II.04	6	2	62	9,681	nová	9,246	9,195	9,199	9,096
III.04	6	3	63	9,467	nová	9,267	9,244	9,214	9,167
IV.04	6	4	64	9,048	nová	9,231	9,251	9,183	9,228
V.04	6	5	65	8,836	nová	9,253	9,306	9,214	9,326
VI.04	6	6	66	8,866	nová	9,268	9,325	9,254	9,376
VII.04	6	7	67	9,199	nová	9,239	9,238	9,267	9,268
VIII.04	6	8	68	9,276	nová	9,233	9,227	9,241	9,227
IX.04	6	9	69	9,076	nová	9,079	9,078	9,123	9,117
X.04	6	10	70	8,916	nová	9,108	9,127	9,187	9,228
XI.04	6	11	71	8,904	nová	9,103	9,125	9,187	9,237
XII.04	6	12	72	9,467	nová	9,218	9,197	9,234	9,182
I.05	7	1	73	9,817	nová	9,307	9,246	9,250	9,133
II.05	7	2	74	9,646	nová	9,211	9,161	9,164	9,063
III.05	7	3	75	9,410	nová	9,209	9,187	9,156	9,111
IV.05	7	4	76	8,918	nová	9,101	9,118	9,053	9,096
V.05	7	5	77	8,609	nová	9,027	9,067	8,987	9,087
VI.05	7	6	78	8,585	nová	8,988	9,030	8,974	9,080
VII.05	7	7	79	8,830	nová	8,870	8,868	8,899	8,897
VIII.05	7	8	80	8,914	nová	8,872	8,867	8,879	8,867
IX.05	7	9	81	8,803	nová	8,806	8,805	8,851	8,843
X.05	7	10	82	8,486	nová	8,678	8,687	8,757	8,783
XI.05	7	11	83	8,427	nová	8,627	8,637	8,710	8,743
XII.05	7	12	84	8,877	nová	8,628	8,623	8,644	8,610
I.06	8	1	85	9,163	nová	8,653	8,630	8,596	8,524
II.06	8	2	86	9,061	nová	8,626	8,606	8,579	8,513
III.06	8	3	87	8,820	nová	8,619	8,612	8,566	8,540
IV.06	8	4	88	8,314	nová	8,497	8,501	8,449	8,480
V.06	8	5	89	7,898	nová	8,316	8,318	8,276	8,336
VI.06	8	6	90	7,737	nová	8,139	8,138	8,125	8,183

Pokračování na další stránce

datum	i	j	t_{ij}	y	metodika	y_1	y_2	y_3	y_4
VII.06	8	7	91	7,915	nová	7,955	7,949	7,983	7,975
VIII.06	8	8	92	7,926	nová	7,884	7,884	7,891	7,884
IX.06	8	9	93	7,788	nová	7,791	7,790	7,835	7,823
X.06	8	10	94	7,429	nová	7,621	7,604	7,699	7,689
XI.06	8	11	95	7,279	nová	7,478	7,460	7,562	7,552
XII.06	8	12	96	7,666	nová	7,416	7,447	7,433	7,435
I.07	9	1	97	7,902	nová	7,393	7,443	7,336	7,351
II.07	9	2	98	7,684	nová	7,249	7,298	7,202	7,220
III.07	9	3	99	7,261	nová	7,060	7,089	7,007	7,030
IV.07	9	4	100	6,759	nová	6,941	6,910	6,894	6,893
V.07	9	5	101	6,411	nová	6,829	6,752	6,789	6,767
VI.07	9	6	102	6,289	nová	6,691	6,614	6,677	6,651
VII.07	9	7	103	6,418	nová	6,458	6,446	6,487	6,467
VIII.07	9	8	104	6,354	nová	6,312	6,320	6,319	6,321
IX.07	9	9	105	6,157	nová	6,161	6,159	6,205	6,185
X.07	9	10	106	5,783	nová	5,975	5,920	6,053	5,985
XI.07	9	11	107	5,637	nová	5,836	5,777	5,920	5,848
XII.07	9	12	108	5,982	nová	5,733	5,811	5,749	5,802
I.08	10	1	109	6,120	nová	5,610	5,764	5,553	5,693
II.08	10	2	110	5,944	nová	5,508	5,645	5,461	5,584
III.08	10	3	111	5,609	nová	5,408	5,476	5,355	5,431
IV.08	10	4	112	5,242	nová	5,425	5,360	5,377	5,347
V.08	10	5	113	5,039	nová	5,457	5,307	5,417	5,318
VI.08	10	6	114	5,011	nová	5,414	5,271	5,399	5,300
VII.08	10	7	115	5,263	nová	5,303	5,285	5,331	5,302
VIII.08	10	8	116	5,319	nová	5,277	5,291	5,284	5,291
IX.08	10	9	117	5,310	nová	5,314	5,312	5,358	5,334
X.08	10	10	118	5,202	nová	5,393	5,325	5,472	5,384
XI.08	10	11	119	5,341	nová	5,540	5,474	5,624	5,541
XII.08	10	12	120	5,960	nová	5,711	5,790	5,727	5,781
I.09	11	1	121	6,796	nová	6,287	6,401	6,230	6,323
II.09	11	2	122	7,377	nová	6,941	7,006	6,894	6,930
III.09	11	3	123	7,738	aktuální	7,537	7,555	7,484	7,492
IV.09	11	4	124	7,850	aktuální	8,033	8,026	7,985	8,006

Data převzata z databáze Ministerstva práce a sociálních věcí (OKsystem [67]) – rok 1993–2009.

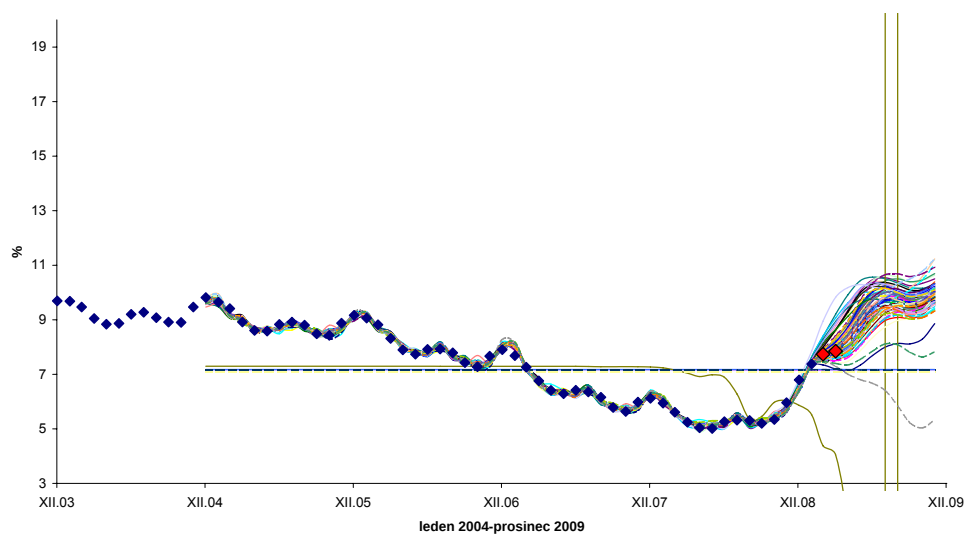
Příloha B

Grafické opory

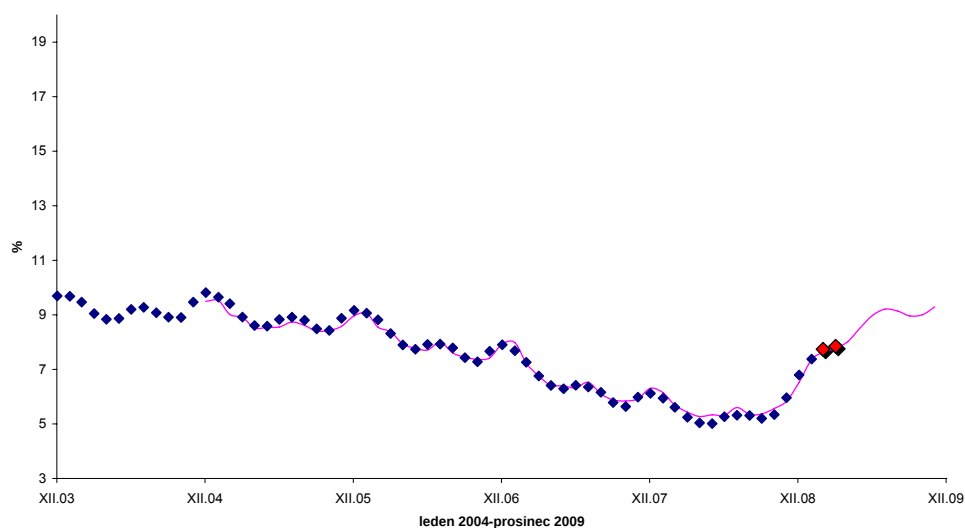
B.1 Nezaměstnanost

B.1.1 Grafické znázornění predikce míry nezaměstnanosti v České republice prostřednictvím dopředných sítí

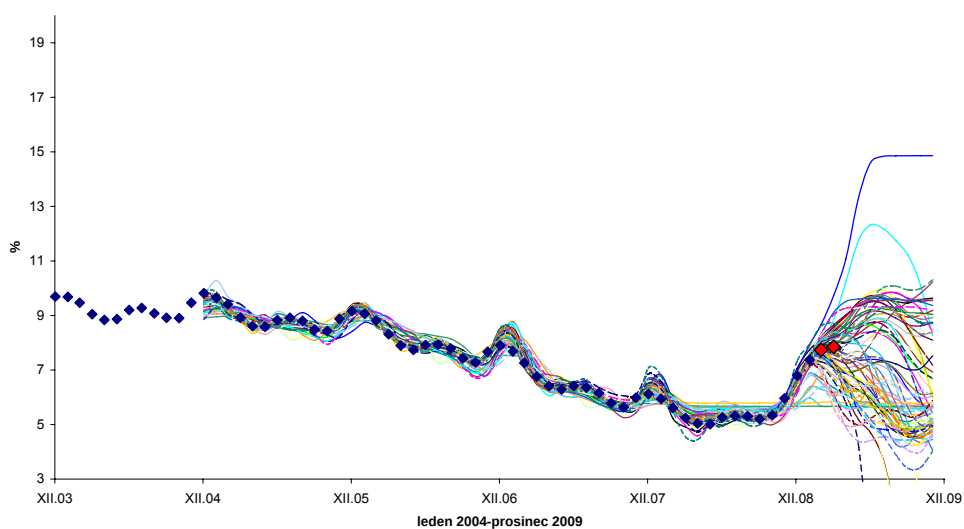
B.1.1.1 Modely tvořeny na základě dat do února 2009



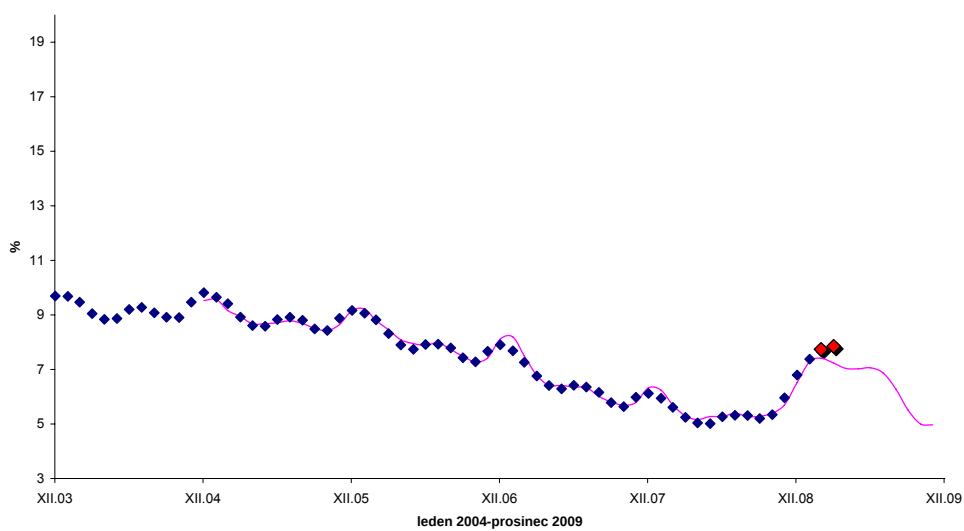
Obrázek B.1: Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím 80 vícevrstvých dopředných sítí VVP 1:12-1-1:1 – leden 2004–prosinec 2009



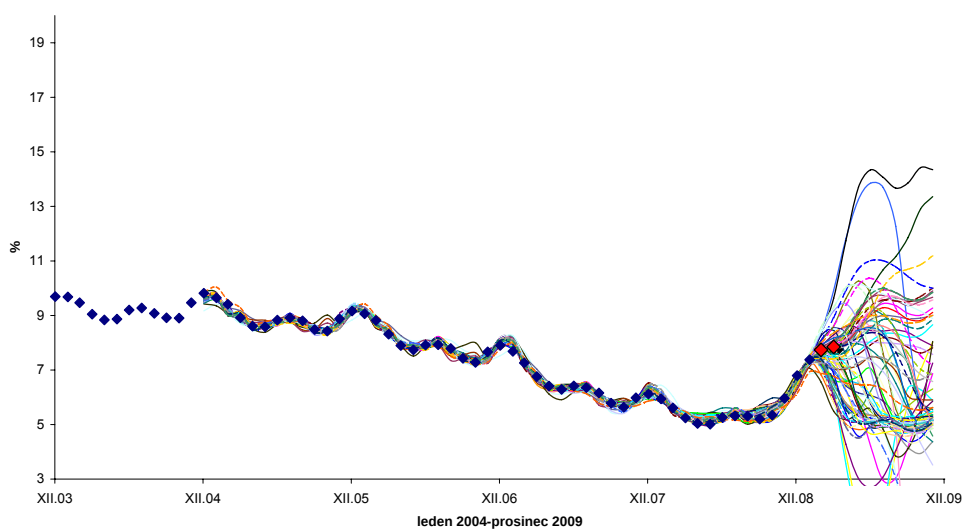
Obrázek B.2: Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím kolekce vytvořené z 80 vícevrstvých dopředných sítí VVP 1:12-1-1:1 – leden 2004–prosinec 2009



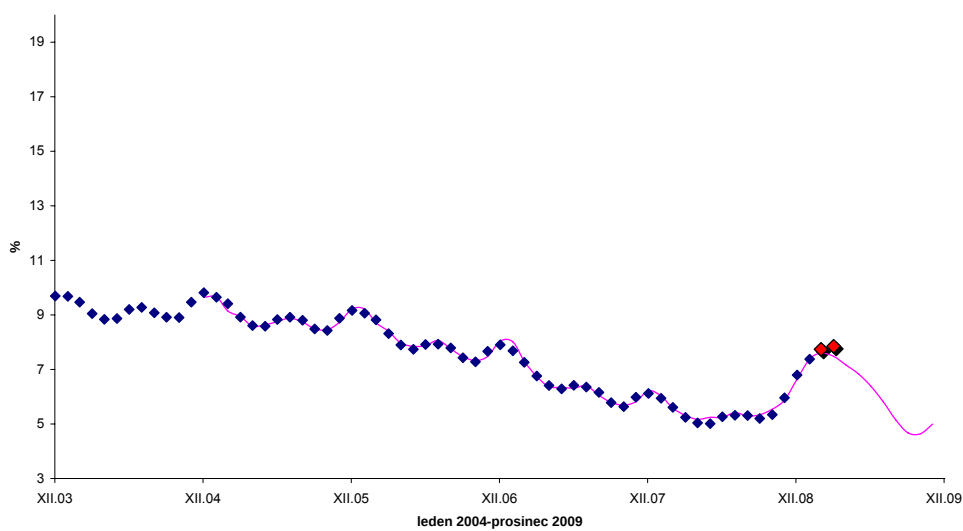
Obrázek B.3: Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím 80 vícevrstvých dopředných sítí VVP 1:12-2-1:1 – leden 2004–prosinec 2009



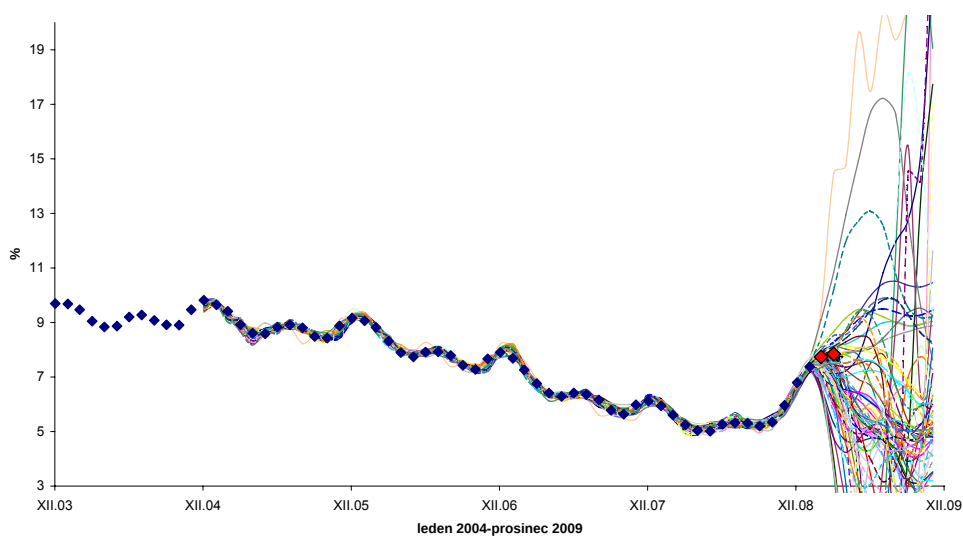
Obrázek B.4: Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím kolekce vytvořené z 80 vícevrstvých dopředných sítí VVP 1:12-2-1:1 – leden 2004–prosinec 2009



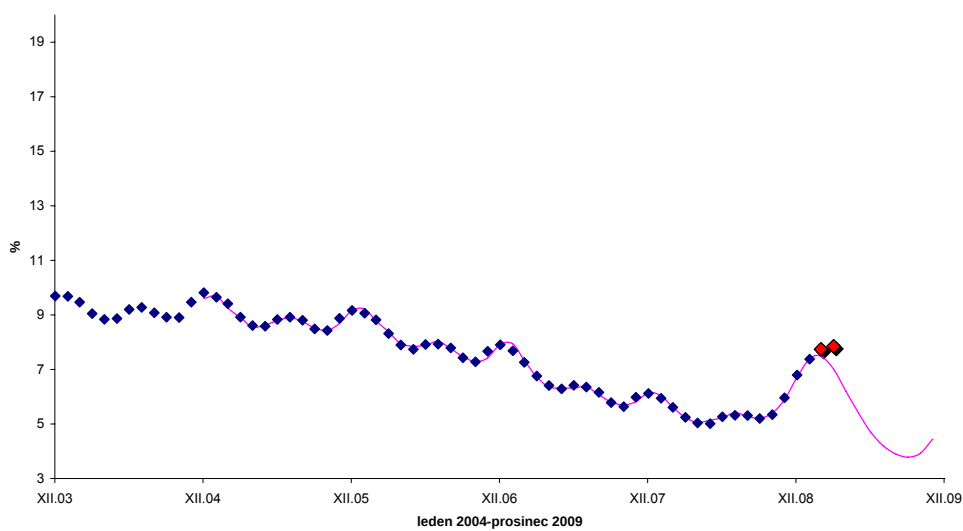
Obrázek B.5: Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím 80 vícevrstvých dopředných sítí VVP 1:12-3-1:1 – leden 2004–prosinec 2009



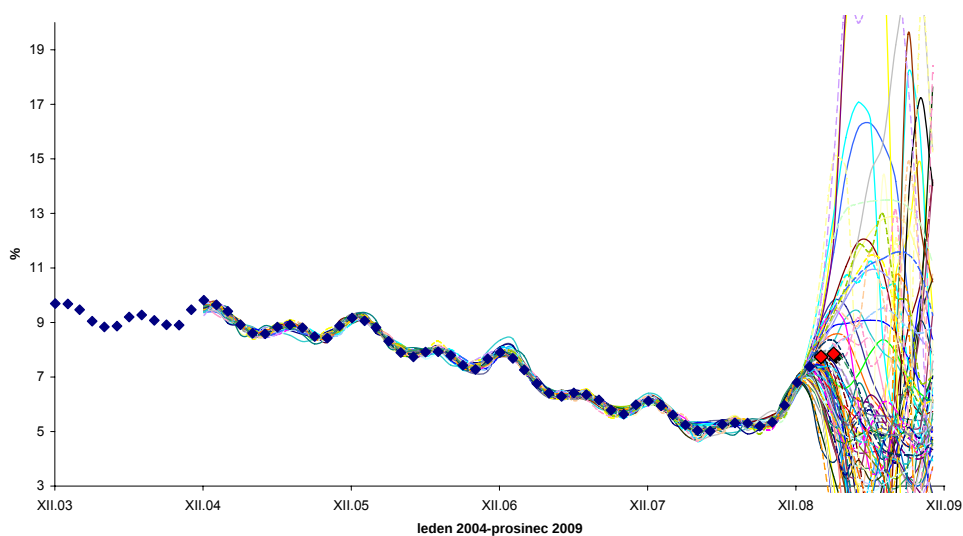
Obrázek B.6: Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím kolekce vytvořené z 80 vícevrstvých dopředných sítí VVP 1:12-3-1:1 – leden 2004–prosinec 2009



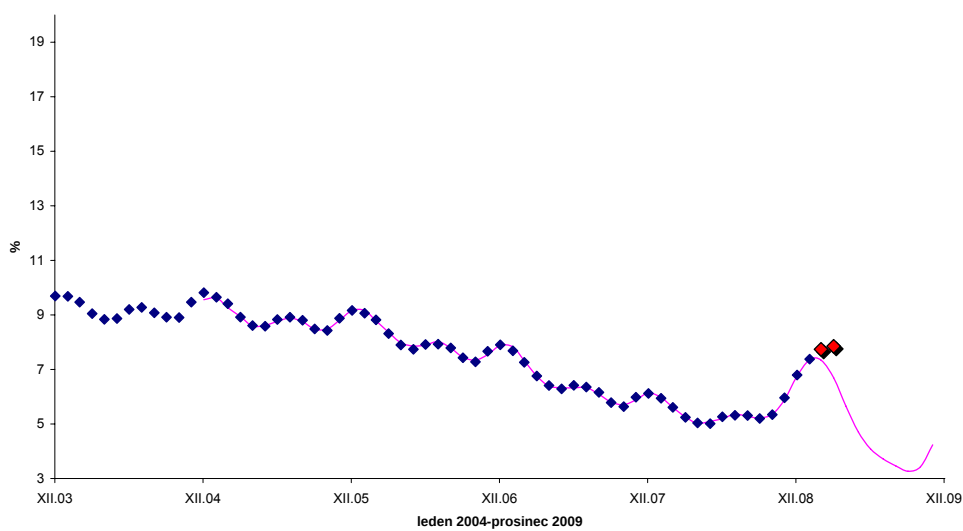
Obrázek B.7: Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím 80 vícevrstvých dopředných sítí VVP 1:12-5-1:1 – leden 2004–prosinec 2009



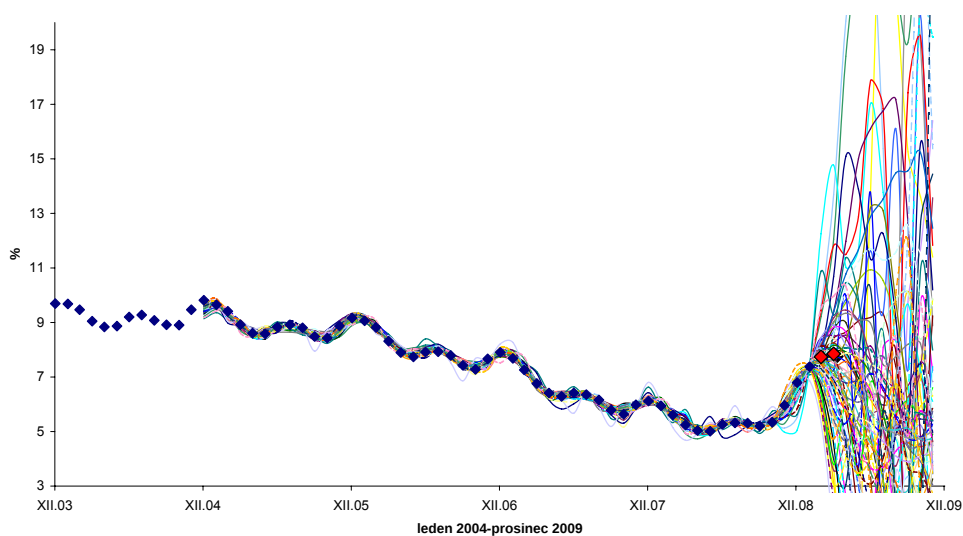
Obrázek B.8: Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím kolekce vytvořené z 80 vícevrstvých dopředných sítí VVP 1:12-5-1:1 – leden 2004–prosinec 2009



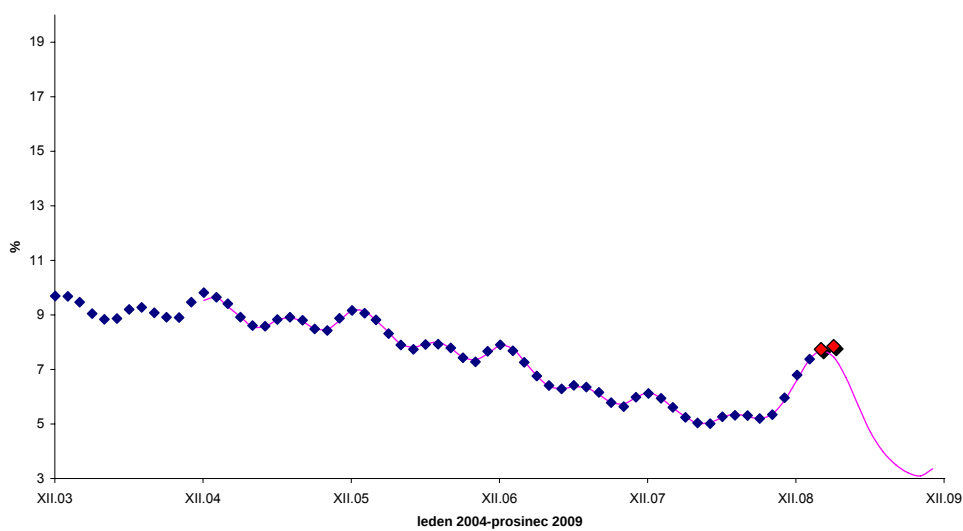
Obrázek B.9: Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím 80 vícevrstvých dopředných sítí VVP 1:12-8-1:1 – leden 2004–prosinec 2009



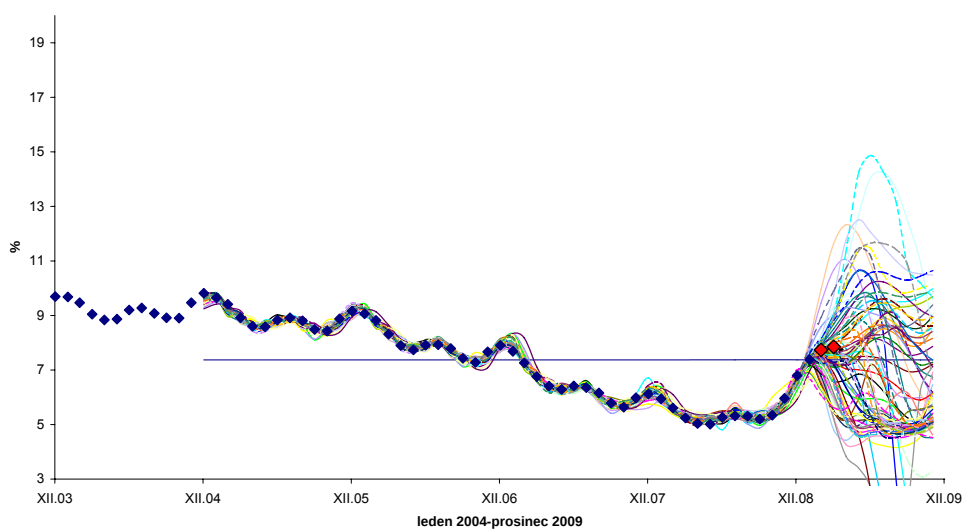
Obrázek B.10: Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím kolekce vytvořené z 80 vícevrstvých dopředných sítí VVP 1:12-8-1:1 – leden 2004–prosinec 2009



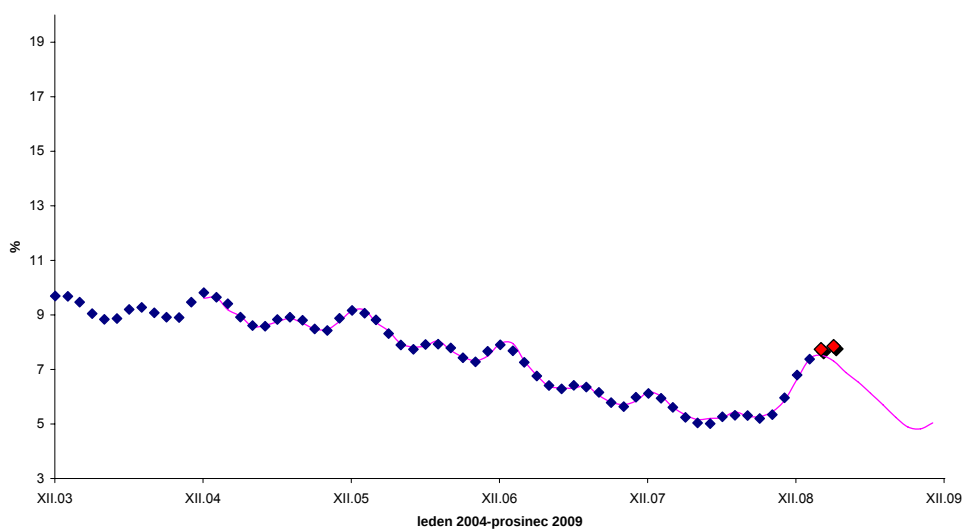
Obrázek B.11: Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím 80 vícevrstvých dopředných sítí VVP 1:12-15-1:1 – leden 2004–prosinec 2009



Obrázek B.12: Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím kolekce vytvořené z 80 vícevrstvých dopředných sítí VVP 1:12-15-1:1 – leden 2004–prosinec 2009

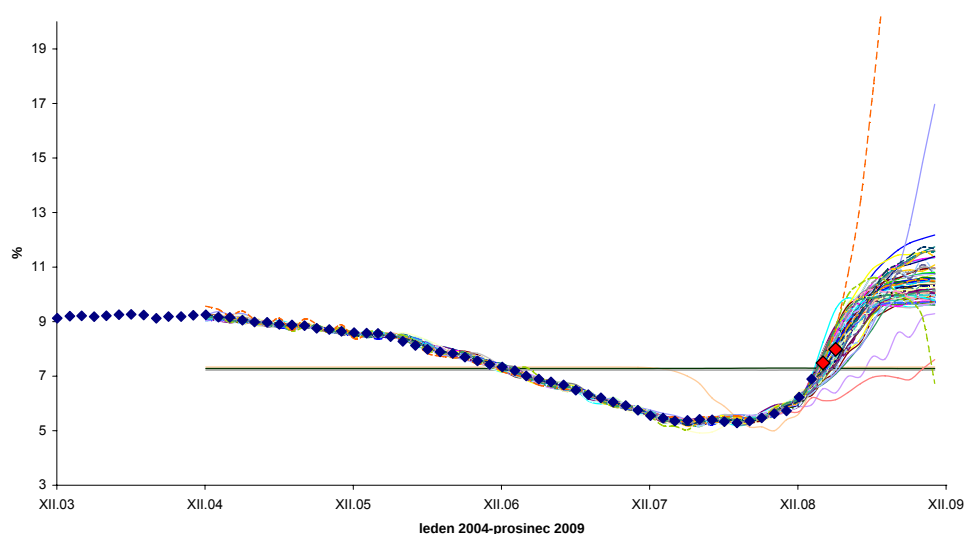


Obrázek B.13: Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím 80 vícevrstvých dopředných sítí VVP 1:12-3-2-1:1 – leden 2004–prosinec 2009

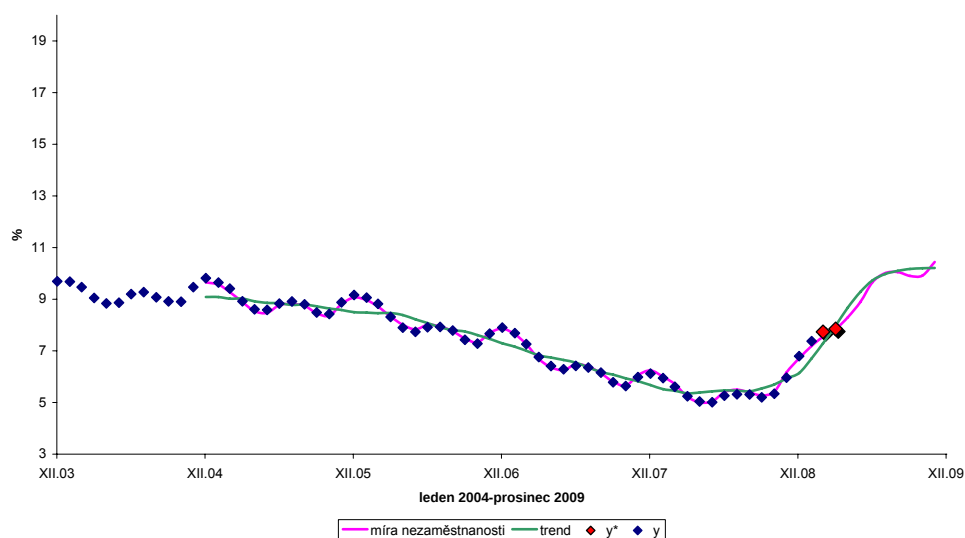


Obrázek B.14: Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím kolekce vytvořené z 80 vícevrstvých dopředných sítí VVP 1:12-3-2-1:1 – leden 2004–prosinec 2009

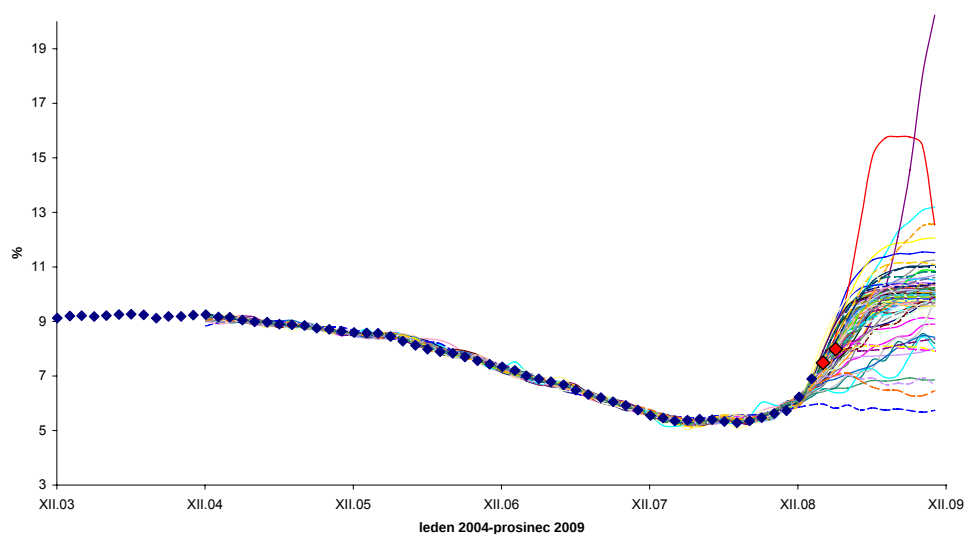
B.1.1.2 Modely tvořeny na základě dat do dubna 2009



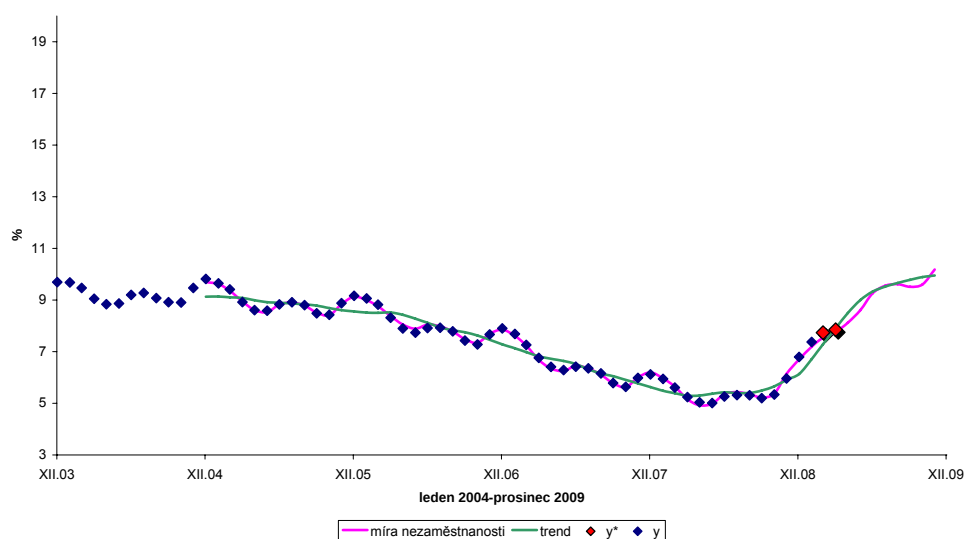
Obrázek B.15: Predikce trendu vývoje míry nezaměstnanosti v České republice prostřednictvím 80 vícevrstvých dopředných sítí VVP 1:12-1-1:1 (T) – leden 2004–prosinec 2009



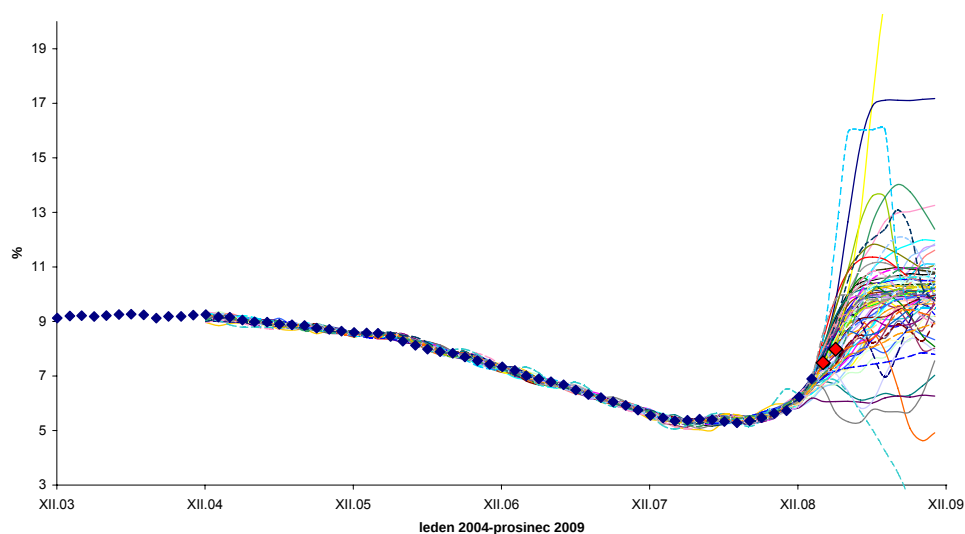
Obrázek B.16: Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím kolekce vytvořené z 80 vícevrstvých dopředných sítí VVP 1:12-1-1:1 (T) – leden 2004–prosinec 2009



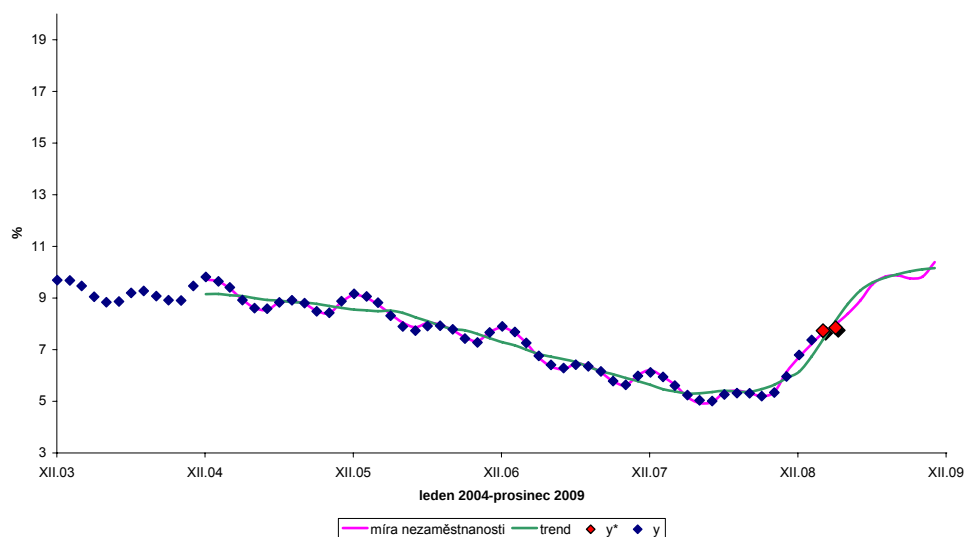
Obrázek B.17: Predikce vývoje trendu míry nezaměstnanosti v České republice prostřednictvím 80 vícevrstevných dopředných sítí VVP 1:12-2-1:1 (T) – leden 2004–prosinec 2009



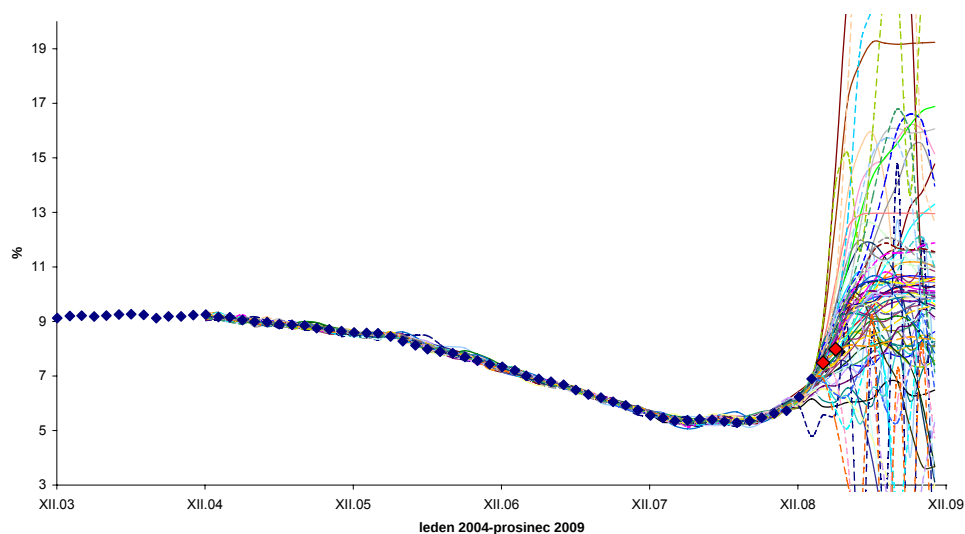
Obrázek B.18: Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím kolekce vytvořené z 80 vícevrstevných dopředných sítí VVP 1:12-2-1:1 (T) – leden 2004–prosinec 2009



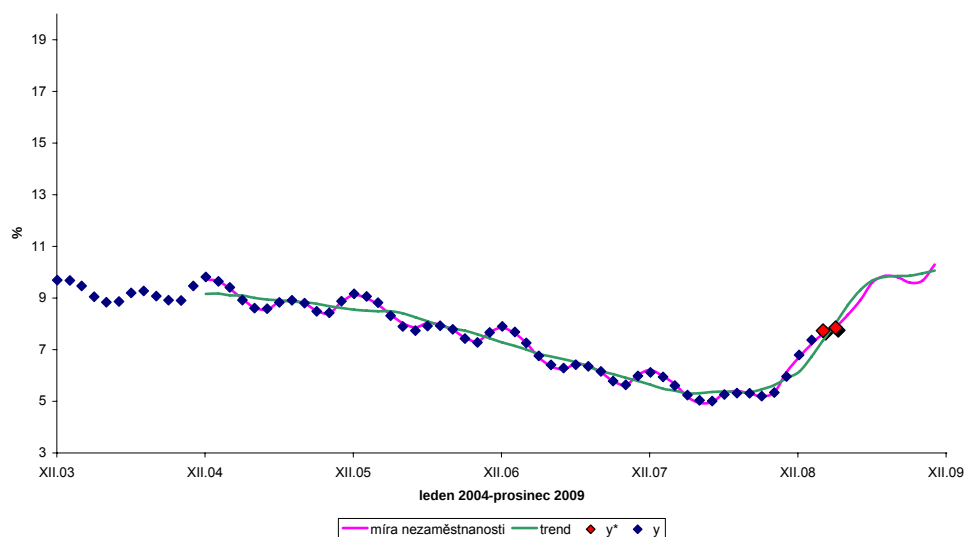
Obrázek B.19: Predikce vývoje trendu míry nezaměstnanosti v České republice prostřednictvím 80 vícevrstvých dopředných sítí VVP 1:12-3-1:1 (T) – leden 2004–prosinec 2009



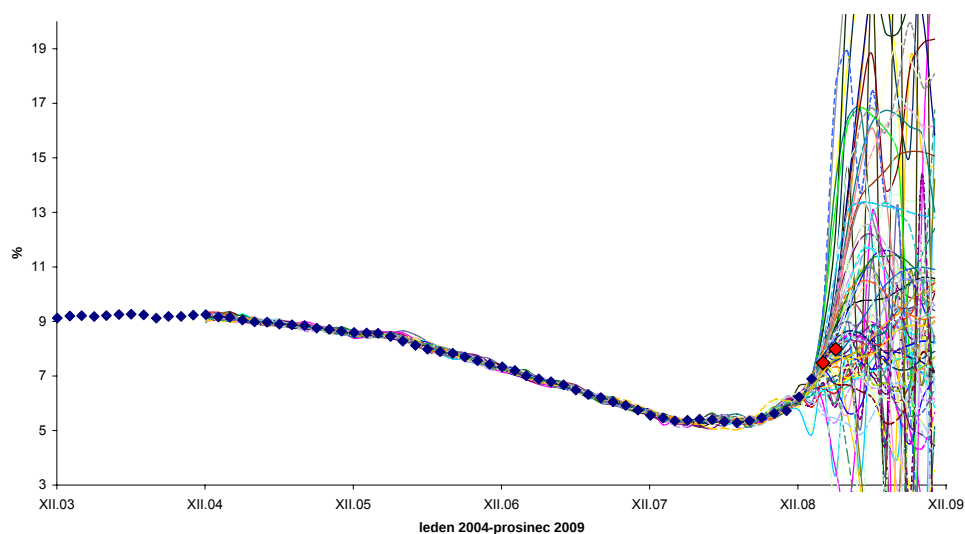
Obrázek B.20: Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím kolekce vytvořené z 80 vícevrstvých dopředných sítí VVP 1:12-3-1:1 (T) – leden 2004–prosinec 2009



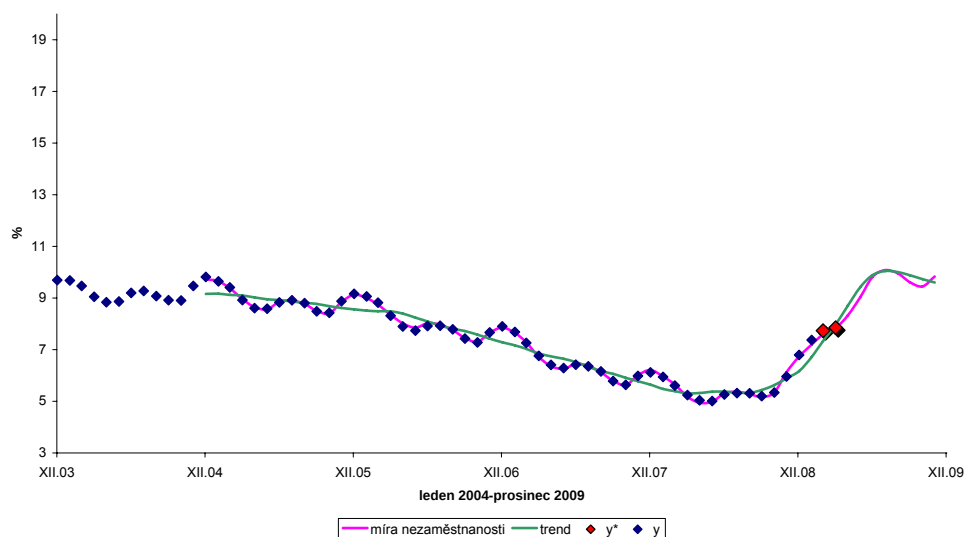
Obrázek B.21: Predikce vývoje trendu míry nezaměstnanosti v České republice prostřednictvím 80 vícevrstevných dopředných sítí VVP 1:12-5-1:1 (T) – leden 2004–prosinec 2009



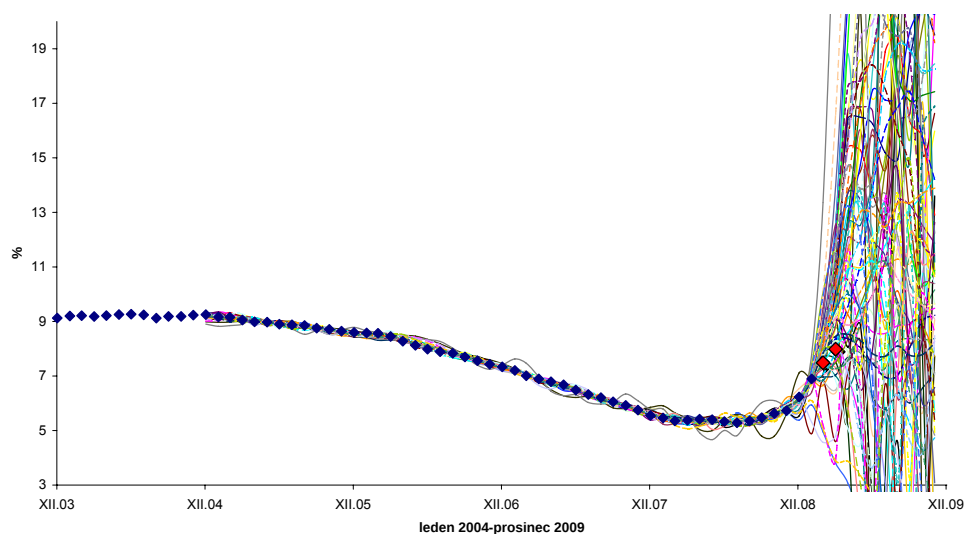
Obrázek B.22: Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím kolekce vytvořené z 80 vícevrstevných dopředných sítí VVP 1:12-5-1:1 (T) – leden 2004–prosinec 2009



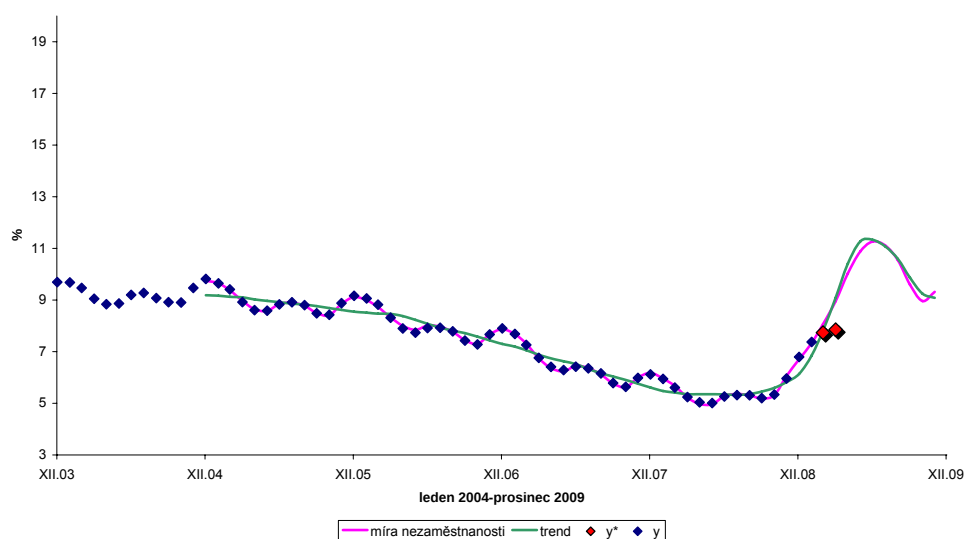
Obrázek B.23: Predikce vývoje trendu míry nezaměstnanosti v České republice prostřednictvím 80 vícevrstvých dopředných sítí VVP 1:12-8-1:1 (T) – leden 2004–prosinec 2009



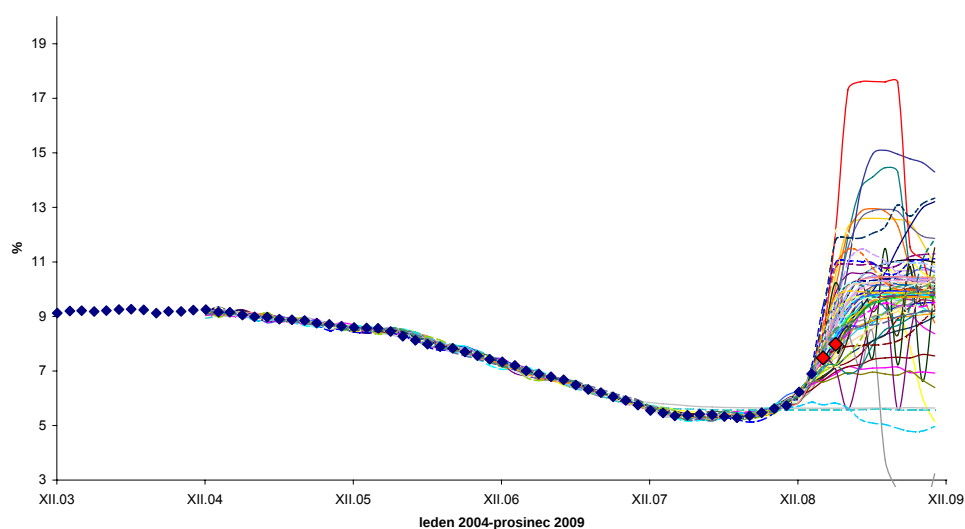
Obrázek B.24: Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím kolekce vytvořené z 80 vícevrstvých dopředných sítí VVP 1:12-8-1:1 (T) – leden 2004–prosinec 2009



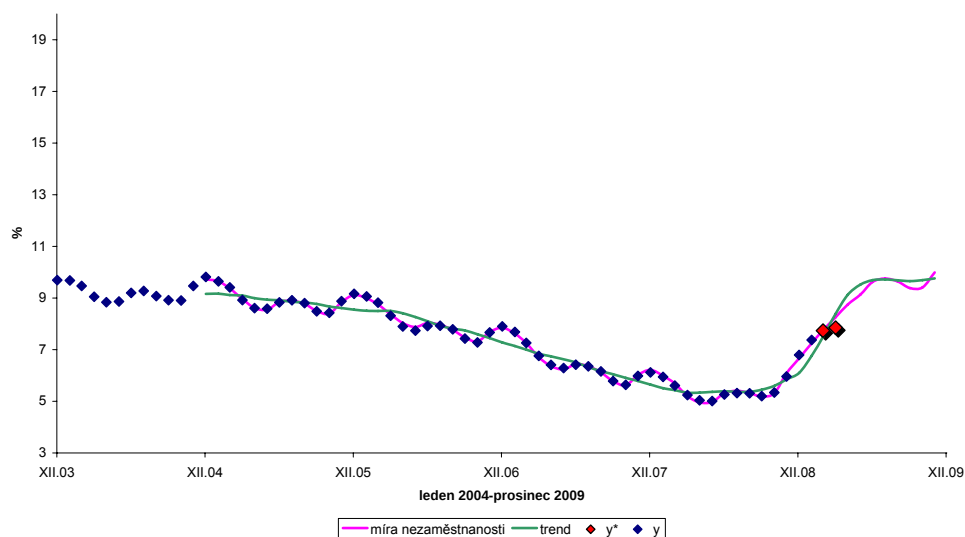
Obrázek B.25: Predikce vývoje trendu míry nezaměstnanosti v České republice prostřednictvím 80 vícevrstevných dopředných sítí VVP 1:12-15-1:1 (T) – leden 2004–prosinec 2009



Obrázek B.26: Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím kolekce vytvořené z 80 vícevrstevných dopředných sítí VVP 1:12-15-1:1 (T) – leden 2004–prosinec 2009



Obrázek B.27: Predikce vývoje trendu míry nezaměstnanosti v České republice prostřednictvím 80 vícevrstevných dopředných sítí VVP 1:12-3-2-1:1 (T) – leden 2004–prosinec 2009



Obrázek B.28: Predikce vývoje míry nezaměstnanosti v České republice prostřednictvím kolekce vytvořené z 80 vícevrstevných dopředných sítí VVP 1:12-3-2-1:1 (T) – leden 2004–prosinec 2009