5. Viacvrstvové neurónové siete

Jednoduchá neurónová sieť (tzv. lineárny model neurónovej siete, ako bol uvedený v kapitole 4), či už jednovrstvová alebo viacvrstvová, je schopná korektne riešiť len obmedzenú triedu problémov — tzv. lineárne separovateľné problémy (viď ďalej podkapitola 5.3.1). Lineárne neurónové siete môžu byť jednoducho zovšeobecnené tak, že aktivity výstupných neurónov sú určené pomocou nelineárnej prechodovej funkcie (ktorá v najjednoduchšom prípade odpovedá tzv. tvrdej nelinearite, napr. skokovej alebo znamienkovej funkcii). Avšak takéto zovšeobecnenie lineárnej neurónovej siete je tiež schopné klasifikovať len lineárne separovateľné problémy. Toto ohraničenie bolo považované za vážny nedostatok neurónových sietí [1]. Teoreticky sa uvažovalo o možnosti zavedenia ďalších (skrytých) vrstiev nelineárnych neurónov do sietí. Žiaľ, nebolo jasné ako adaptovať váhové koeficienty, ktoré sú priradené neurónom zo skrytej vrstvy. Až Rumelhart so spolupracovníkmi [2] navrhli jednoduchý gradientový algoritmus (nazývaný metóda spätného šírenia — angl. *back propagation*) adaptácie viacvrstvových neurónových sietí s dopredným šírením. Týmto sa viacvrstvové neurónové siete stali veľmi populárne a patria medzi univerzálne prístupy teórie neurónových sietí so širokou paletou aplikácií v rôznych oblastiach informatiky a prírodných vied. Naviac bolo dokázané, že neurónové siete tohto typu sú univerzálnym aproximátorom, t.j. sú schopné aproximovať s požadovanou presnosťou ľubovoľnú spojitú funkciu, čiže môžu byť chápané ako univerzálny prostriedok pre regresnú analýzu, kde tvar modelovej funkcie je určený architektúrou neurónovej siete. Pod architektúrou máme na mysli nielen "topológiu" prepojenia neurónov, ale tiež aj nastavenie váhových a prahových koeficientov na určité hodnoty.

5.1 Všeobecný klasifikačný problém

Zavedieme všeobecnú formuláciu klasifikačného problému pomocou pojmu zobrazenia — funkcie definovanej nad dvomi množinami A a B. Tento prístup bude užitočný pre interpretáciu neurónových sietí ako *klasifikátora* alebo *prediktora*. Nech $F(\mathbf{x})$ je funkcia definovaná nad množinou A, ktorá priradí každému elementu $\boldsymbol{x} \in A$ obraz — funkčnú hodnotu z množiny B, $\hat{\boldsymbol{x}} = F(\boldsymbol{x}) \in B$,

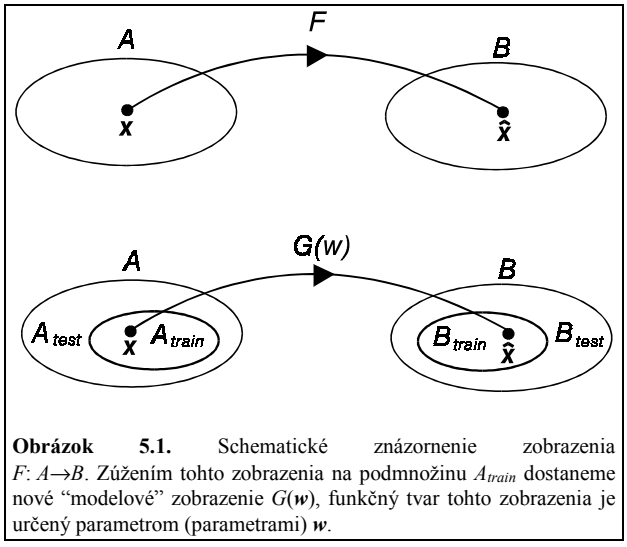
$$F: A \rightarrow B \tag{5.1}$$

Nech $G(\boldsymbol{x},\boldsymbol{w})$ je funkcia, ktorej argumenty sú z konečnej podmnožiny $A_{train} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_r\} \subset A$ (nazývanej *tréningová množina*) a $\boldsymbol{w}$ je parameter (alebo parametre) zobrazenia G, potom $\hat{\boldsymbol{x}} = G(\boldsymbol{x},\boldsymbol{w}) \in B_{train} \subset B$ (pozri obr. 5.1)

$$G(\boldsymbol{w}): A_{train} \rightarrow B_{train} \tag{5.2}$$

Formálne môžeme povedať, že zobrazenie $G(\boldsymbol{w})$ je reštrikcia zobrazenia $F(\boldsymbol{x})$ nad množinou $A_{train} \subset A$. Komplement A_{train} vzhľadom k množine A je označený A_{test} (nazývaný *testovacia množina*), $A_{test}=A \setminus A_{train}$. Predpokladajme, že pre každé $\boldsymbol{x}_i \in A_{train}$ poznáme požadovaný obraz — funkčnú hodnotu $\hat{\boldsymbol{x}}_i$,

$$\boldsymbol{x}_1 \,/\, \hat{\boldsymbol{x}}_1, \boldsymbol{x}_2 \,/\, \hat{\boldsymbol{x}}_2, \ldots, \boldsymbol{x}_r \,/\, \hat{\boldsymbol{x}}_r \tag{5.3}$$



Obrázok 5.1. Schematické znázornenie zobrazenia F: $A \rightarrow B$. Zúžením tohto zobrazenia na podmnožinu A_{train} dostaneme nové "modelové" zobrazenie $G(\boldsymbol{w})$, funkčný tvar tohto zobrazenia je určený parametrom (parametrami) $\boldsymbol{w}$.

Požadované funkčné hodnoty $\hat{\boldsymbol{x}}_i$ sú interpretované ako obrazy funkcie F

$$\hat{\boldsymbol{x}}_i = F(\boldsymbol{x}_i) \qquad (i = 1,2,\ldots,r) \tag{5.4}$$

Cieľom našich úvah je nájsť taký parameter (alebo parametre) $\boldsymbol{w}$ funkcie $G(\boldsymbol{x},\boldsymbol{w})$, aby funkčné hodnoty argumentov z tréningovej množiny A_{train} boli čo najbližšie obrazom funkcie $F(\boldsymbol{x})$ (t.j. požadovaným hodnotám). Definujme *účelovú funkciu*

$$E(\boldsymbol{w}) = \frac{1}{2}\sum_{i=1}^{r}\left(G(\boldsymbol{x}_i,\boldsymbol{w}) - \hat{\boldsymbol{x}}_i\right)^2 \tag{5.5}$$

Táto funkcia vyjadruje sumu kvadrátov odchýlok funkcie $G(\boldsymbol{x},\boldsymbol{w})$ od požadovaných hodnôt $\hat{\boldsymbol{x}}$ braných z tréningovej množiny. Požiadavka, aby vypočítané hodnoty $G(\boldsymbol{x},\boldsymbol{w})$ boli "čo najbližšie" požadovaným hodnotám $\hat{\boldsymbol{x}}$ je realizovaná pomocou požiadavky minimálnosti účelovej funkcie $E(\boldsymbol{w})$ vzhľadom k parametru $\boldsymbol{w}$. Hovoríme, že funkcia $G(\boldsymbol{x},\boldsymbol{w})$ je *adaptovaná*, ak jej parameter $\boldsymbol{w}$ je vybraný tak, aby sa rovnal svojej optimálnej hodnote (t.j. v ktorom má účelová funkcia globálne minimum). Nech $\overline{\boldsymbol{w}}$ je optimálna hodnota parametru $\boldsymbol{w}$ určená nasledujúcim minimalizačným problémom

$$\overline{\boldsymbol{w}} = \arg\min_{\boldsymbol{w} \in W} E(\boldsymbol{w}) \tag{5.6}$$

kde W je množina (priestor) prípustných hodnôt parametra $\boldsymbol{w}$. Adaptovaná funkcia $G(\boldsymbol{x}, \overline{\boldsymbol{w}})$ simuluje pôvodnú funkciu $F(\boldsymbol{x})$ pre hodnoty argumentov z tréningovej množiny A_{train} na základe minimalizačného kritéria (5.6). Naviac, adaptovaná funkcia $G(\boldsymbol{x}, \overline{\boldsymbol{w}})$ sa používa aj pre predpoveď funkčných hodnôt odpovedajúcich argumentom z testovacej množiny A_{test}, t.j. predpokladá sa, že adaptovaná funkcia dobre aproximuje pôvodnú funkciu $F(\boldsymbol{x})$ tiež mimo tréningovej množiny. Naše úvahy môžu byť jednoducho chápané ako klasický regresný problém, v ktorom parametre modelovej funkcie G sú optimalizované (adaptované) tak, aby vypočítané funkčné hodnoty boli blízke požadovaným (experimentálnym) funkčným hodnotám.

Jeden zo základných problémov v každej oblasti prírodných vied je *hľadanie vzťahu — funkcie medzi štruktúrou jej objektov a ich vlastnosťami*. Ideálom je konštrukcia tejto funkcie v analytickom tvare, ktorá vzťahuje vlastnosti objektov k ich štruktúre. V mnohých prípadoch je tento cieľ buď vôbec nerealizovateľný alebo len s veľkými obtiažami. Tento meta-teoretický prístup ku konštrukcii analytických vzťahov pre koreláciu štruktúra verzus vlastnosť v mnohých prípadoch naráža na principiálne problémy tak teoretického, ako aj numerického charakteru. Preto sa pomerne často používa prístup "regresnej analýzy". Modelová funkcia G sa zostrojí na základe určitých úvah a jej voľne adjustovateľné parametre sa určia pomocou minimalizácie účelovej funkcie $E(\boldsymbol{w})$ (vzťah (5.6)). Takto adaptovaná modelová funkcia G sa potom berie ako analytický vzťah medzi štruktúrou prírodovedných objektov a ich vlastnosťami.

Nech O je množina *objektov*, $O=\{o_1,o_2,...\}$. Každý objekt $o \in O$ je popísaný deskriptorom $\boldsymbol{x}$, ktorý charakterizuje jeho *štruktúru*, a klasifikátorom $\hat{\boldsymbol{x}}$, ktorý popisuje jeho *vlastnosti*. Vzťah medzi deskriptorom a klasifikátorom je formálne vyjadrený pomocou hypotetickej funkcie $\hat{\boldsymbol{x}}=F(\boldsymbol{x})$. Ako už bolo spomenuté vyššie, konštrukcia tejto funkcie patrí medzi základné problémy každej vednej oblasti. Alternatívne riešenie tohto problému môže byť uskutočnené pomocou ad-hoc postulovania modelovej funkcie $G(\boldsymbol{x},\boldsymbol{w})$ v analytickom tvare. Parameter (alebo parametre) $\boldsymbol{w}$ je určený podmienkou, aby funkčné hodnoty pre deskriptory z tréningovej množiny boli blízke požadovaným hodnotám klasifikátora. Hlavným cieľom tohto postupu je, že adaptovaná modelová funkcia $G(\boldsymbol{x}, \overline{\boldsymbol{w}})$ je použitá pre klasifikáciu objektov mimo tréningovej množiny. To znamená, že adaptovaná modelová funkcia $G(\boldsymbol{x}, \overline{\boldsymbol{w}})$ je extrapolovaná mimo tréningovej množiny "v dobrej viere", že aj v tomto prípade bude dobre aproximovať funkciu $F(\boldsymbol{x})$, ktorá presne klasifikuje objekty z celej množiny O.

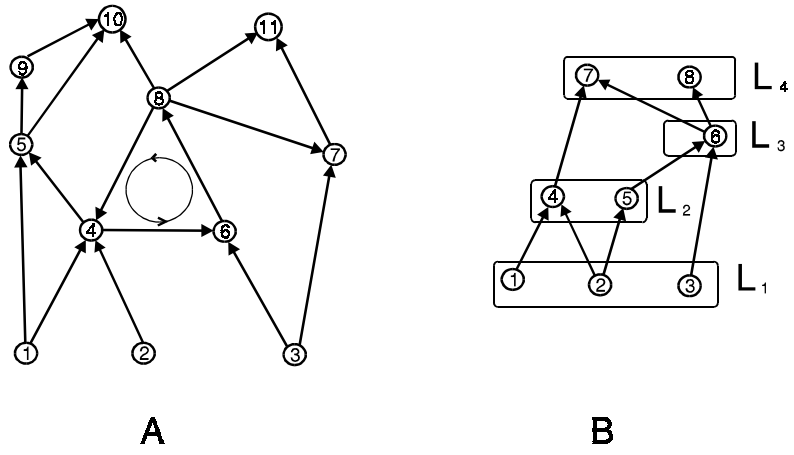
5.2 Definícia neurónovej siete

Paradigma neurónovej siete bude formulovaná pomocou grafovo-teoretického prístupu [3]. Pritom sa vychádza sa z analógie s ľudským mozgom (pozri kapitolu 1) a koncept neurónovej siete bude použitý na konštrukciu modelovej funkcie $G(\boldsymbol{x},\boldsymbol{w})$. Formálne je neurónová sieť určená ako orientovaný graf G=(V,E), pozri obr. 5.2. Výrazy $V=\{v_1,v_2,...,v_N\}$ a $E=\{e_1,e_2,...e_M\}$ označujú neprázdnu vrcholovú množinu resp. hranovú množinu grafu G obsahujúceho N vrcholov (neurónov) a M hrán (spojov). Každý spoj $e \in E$ sa interpretuje ako usporiadaná dvojica dvoch neurónov z množiny V, $e=(v,v')$. Hovoríme, že spoj e začína v neuróne v a končí v neuróne v'. Množina neurónov V je rozložená na disjunktné podmnožiny (pozri obr. 5. 2)

$$V = V_I \cup V_H \cup V_O \tag{5.7}$$

kde V_I obsahuje N_I vstupných neurónov, ktoré sú susedné len s vychádzajúcimi hranami, V_H obsahuje N_H skrytých (angl. *hidden*) neurónov, ktoré sú susedné súčasne s vychádzajúcimi ako aj s vchádzajúcimi hranami, a konečne V_O obsahuje N_O výstupných neurónov, ktoré sú susedné len s vchádzajúcimi hranami. V našich nasledujúcich úvahách budeme vždy predpokladať, že množiny V_I a V_O sú neprázdne, t.j. neurónová sieť obsahuje vždy aspoň jeden vstupný a jeden výstupný neurón.

Pre acyklické neurónové siete (ktoré neobsahujú orientované cykly (pozri graf A na obr. 5.2)) neuróny môžu byť usporiadané do vrstiev (pozri graf B na obr. 5.2)



Obrázok 5.2. Neurónová sieť je definovaná ako orientovaný súvislý graf. Diagram A obsahuje orientovaný graf s jedným cyklom a teda nemôže byť použitý pre definíciu neurónovej siete s dopredným šírením. Diagram B ilustruje možnosť rozkladu vrcholov (neurónov) acyklického orientovaného grafu na vrstvy $L_1,..,L_4$.

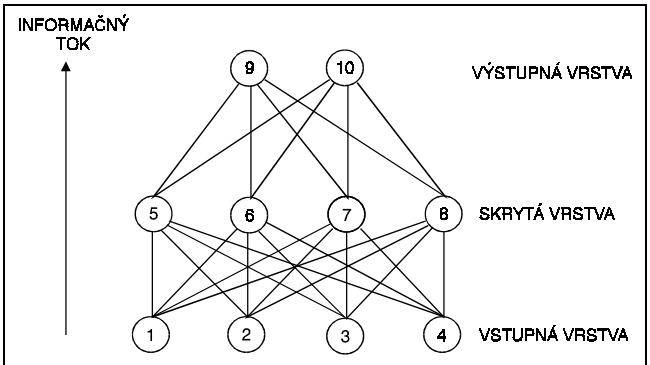
$$V = L_1 \cup L_2 \cup L_3 \cup ... \cup L_t \tag{5.8}$$

kde $L_1=V_I$ je *vstupná vrstva* (obsahuje len vstupné neuróny), L_2, L_3,..., L_{t-1} sú *skryté vrstvy* a L_t je *výstupná vrstva*. Vrstva L_i (pre $1 \le i \le t$) je určená nasledujúcim jednoduchým spôsobom

$$L_i = \{v \in V; d(v) = i + 1\} \tag{5.9}$$

kde vzdialenosť $d(v)$ sa rovná dĺžke maximálnej cesty, ktorá spája daný neurón so vstupným neurónom, potom musí platiť $d(v)=0$, pre $v \in V_I$. Neurónová sieť určená acyklickým grafom je obvykle volená tak, že neuróny z dvoch susedných vrstiev sú poprepájané všetkými možnými spojmi (pozri obr. 5.3). Žiaľ, takýto rozklad množiny neurónov na vrstvy je

možný len pre neurónové siete reprezentované acyklickými grafmi, pre cyklické grafy vzdialenosť $d(v)$ môže nadobúdať ľubovolnú kladnú celočíselnú hodnotu.



Obrázok 5.3. Znázornenie trojvrstvovej neurónovej siete s dopredným šírením.

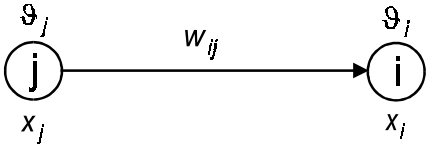
Iný, alternatívny spôsob [3], ako definovať orientovaný graf G je použitie zobrazenia Γ, ktoré priradí každému vrcholu $v \in V$ podmnožinu $\Gamma(v) \subset V$ obsahujúcu tie neuróny, ktoré sú koncovými na spojoch vychádzajúcich z vrcholu v. Neuróny z podmnožiny $\Gamma(v)$ sa nazývajú *nasledovníci* vrcholu v v grafe G. "Inverzné" zobrazenie Γ^{-1} priradí každému vrcholu $v \in V$ podmnožinu $\Gamma^{-1}(v) \subset V$ zloženú z "predchodcov" vrcholu v v grafe G.

Neuróny a spoje sú ohodnotené reálnymi číslami, pozri obr. 5.4. Každý neurón v_i je ohodnotený *prahovým koeficientom* ϑ_i a *aktivitou* x_i. Podobne, každý spoj (v_j,v_i) je ohodnotený váhovým koeficientom (alebo jednoducho, váhou) w_{ij}. Postulujeme, že aktivity skrytých a výstupných neurónov sú určené vzťahom

$$x_i = t(\xi_i) \tag{5.10a}$$

$$\xi_i = \sum_{j \in \Gamma_i^{-1}} w_{ij} x_j + \vartheta_i \tag{5.10b}$$

kde sumácia beží cez neuróny, ktoré sú predchodcami neurónu v_i.



Obrázok 5.4. Neuróny a spoje neurónovej siete sú ohodnotené reálnymi číslami: neuróny sú ohodnotené dvoma rôznymi parametrami, a to prahovými faktormi ϑ_i a aktivitami x_i. Spoje neurónovej siete sú ohodnotené váhovými koeficientmi w_{ij}.

Veličina ξ_i sa nazýva *potenciál* neurónu v_i (analógia tzv. postsynaptického potenciálu, pozri kapitolu 1). *Prechodová (aktivačná) funkcia* $t(\xi)$ z pravej strany (5.10a) je monotónne rastúca funkcia, ktorá vyhovuje nasledujúcim dvom asymptotickým podmienkam: $t(\xi)\to A$, pre $\xi\to-\infty$ a $t(\xi)\to B$, pre $\xi\to\infty$, kde $-\infty<A<B<\infty$. V teórii neurónových sietí sa často využíva nasledujúca "sigmoidálna" funkcia

$$t(\xi)=\frac{B+Ae^{-\xi}}{1+e^{-\xi}} \tag{5.11a}$$

s prvou deriváciou určenou

$$t'(\xi)=\frac{\left[-A+t(\xi)\right]\left[B-t(\xi)\right]}{A+B} \tag{5.11b}$$

Táto prechodová funkcia zobrazuje celú množinu reálnych čísel R na otvorený interval (A,B), formálne t: $R\to(A,B)$. Najčastejšie sa prechodová funkcia (5.11a) využíva pre hodnoty parametrov $A=0$, $B=1$ alebo $A=-1$, $B=1$ (pozri obr. 5.5). Prvý graf odpovedá klasickej sigmoidálnej prechodovej funkcii, zatiaľ čo druhý graf je analógiou hyperbolického tangentu.

Aktivity neurónov tvoria vektor $\boldsymbol{x}=(x_1,x_2,...x_N)$. Tento vektor možno formálne rozložiť na tri podvektory obsahujúce vstupné, skryté a výstupné aktivity

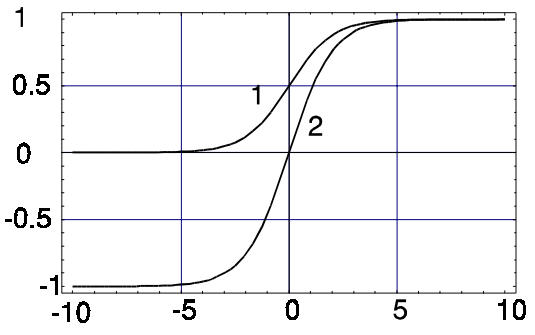
$$\boldsymbol{x}=\boldsymbol{x}_I\oplus\boldsymbol{x}_H\oplus\boldsymbol{x}_O \tag{5.12}$$

Neurónovú sieť s fixovanými váhami a prahovými koeficientmi možno formálne chápať ako funkciu

$$G:R^{N_I}\to(A,B)^{N_O} \tag{5.13}$$

Táto funkcia G priradí vstupnej aktivite $\boldsymbol{x}_I$ (deskriptor) výstupný vektor $\boldsymbol{x}_O$ (klasifikátor) s hodnotami svojich zložiek z otvoreného intervalu (A,B)

$$G(\boldsymbol{x}_I)=\boldsymbol{x}_O \tag{5.14}$$



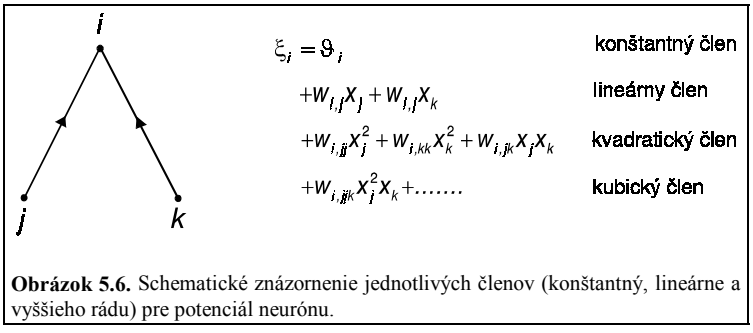
Obrázok 5.5. Priebeh aktivačnej funkcie definovanej (5.11a). Graf 1 odpovedá štandardnej sigmoide ($A=0$, $B=1$), graf 2 je podobný funkcii hyperbolický tangent ($A=-1$,$B=1$).

Skryté aktivity nie sú explicitne uvedené, hrajú len úlohu medzivýsledkov. Ešte niekoľko poznámok k výpočtu aktivít podľa (5.10a-b). Vstupné aktivity sú určené deskriptorom, preto ich pokladáme za fixované. Aktivity skrytých neurónov z druhej vrstvy L_2 môžeme teraz spočítať len použitím vstupných aktivít z vrstvy L_1. Vo všeobecnosti pre výpočet aktivít z vrstvy L_i (kde i>1) musíme poznať len aktivity z nižších vrstiev L_1, L_2,..., L_{i-1} Týmto rekurentným spôsobom môžeme postupne spočítať aktivity všetkých neurónov, ako posledné sa počítajú aktivity výstupných neurónov. Vďaka tomu sa pre neurónové siete reprezentované acyklickým grafom zaužíval názov *neurónové siete* s *dopredným šírením* (angl. *feed-forward neural networks*). Žiaľ, spomínaný jednoduchý postup výpočtu aktivít neurónov je aplikovateľný len na neurónové siete reprezentované acyklickým orientovaným grafom. V prípade, že graf obsahuje orientované cykly, tento postup nie je použiteľný. Rovnice (5.10a-b) sú v tomto prípade spriahnuté a nelineárne. Preto ich riešenie (t.j. skryté a výstupné aktivity) môžeme dosiahnuť len použitím iteračného postupu, a to tak, že štartujeme z počiatočných aktivít, pomocou týchto spočítame nové aktivity a tieto sa v nasledujúcom iteračnom kroku použijú ako vstup pre výpočet nových aktivít. Tento iteračný postup sa opakuje tak dlho, až rozdiel medzi starými a novými aktivitami je menší ako predpísaná presnosť.

V literatúre [4,5] je študovaných ešte mnoho ďalších modifikácií neurónových sietí s dopredným šírením. V ďalšej časti tejto kapitoly uvedieme dva typy, a to neurónovú sieť vyššieho rádu a adaptívnu kombináciu lokálnych neurónových sietí.

5.2.1 Neurónová sieť vyššieho rádu

Jednoduchou možnosťou, ako zovšeobecniť pojem neurónovej siete s dopredným šírením je zovšeobecnenie formuly (5.10b) pre potenciál ξ_i tak, aby obsahovala nielen konštantný člen (prahový faktor) a lineárne členy (vážené aktivity predchádzajúcich neurónov), ale aj členy vyššieho rádu [6] (kvadratické, kubické, ...), pozri obr. 5.6.



Obrázok 5.6. Schematické znázornenie jednotlivých členov (konštantný, lineárne a vyššieho rádu) pre potenciál neurónu.

Potom

$$x_i = t(\xi_i) \qquad (\text{pre } i = 1,2,...,N) \tag{5.15a}$$

$$\xi_i = \vartheta_i + \sum_j w_{i,j} x_j + \sum_{j \le k} w_{i,jk} x_j x_k + ... \tag{5.15b}$$

kde prvá sumácia beží cez všetky $j \in \Gamma_i^{-1}$, druhá sumácia beží cez všetky $j,k \in \Gamma_i^{-1}$, ktoré sú ohraničené podmienkou $j \le k$. Ak formula (5.15b) obsahuje nanajvýš kvadratické členy, potom neurónová sieť sa nazýva sieť druhého rádu. Vo všeobecnosti, členy najvyššieho rádu v (5.15b) určujú rád neurónovej siete. Neurónové siete vyššieho rádu dosahujú podobné výsledky ako neurónové siete prvého rádu, avšak s podstatne menším počtom skrytých neurónov. Je potrebné zdôrazniť, že táto vlastnosť neurónových sietí vyššieho rádu je získaná za "cenu" podstatne horších konvergentných vlastností adaptačného procesu (obrazne povedané, neurónové siete vyššieho rádu sú "viac nelineárne" ako neurónové siete prvého rádu).

5.2.2 Adaptívna kombinácia lokálnych neurónových sietí

Nech $\boldsymbol{N}$=(G,$\boldsymbol{w}$,$\boldsymbol{\vartheta}$) je neurónová sieť s dopredným šírením určená acyklickým orientovaným grafom G, pričom spoje a neuróny sú ohodnotené váhovými koeficientmi $\boldsymbol{w}$ resp. prahovými koeficientmi $\boldsymbol{\vartheta}$. Uvažujme *t lokálnych* neurónových sietí [7,8]

$$\boldsymbol{N}_i = \left(\mathrm{G}^{(i)}, \boldsymbol{w}^{(i)}, \boldsymbol{\vartheta}^{(i)}\right) \qquad (i = 1,2,...,t) \tag{5.16a}$$

a jednu tzv. bránovú (angl. *gating*) neurónovú sieť

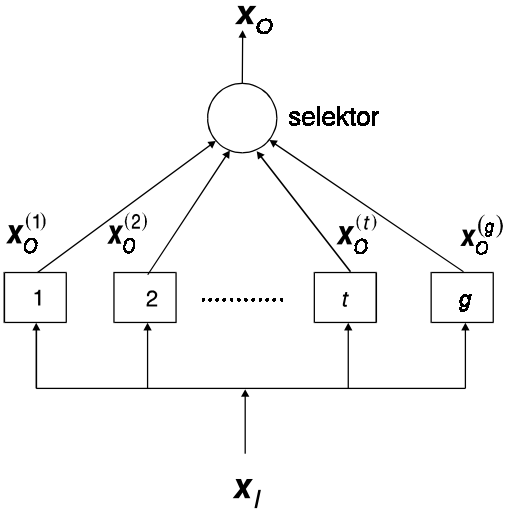
$$\boldsymbol{N}_g = \left(\mathrm{G}^{(g)}, \boldsymbol{w}^{(g)}, \boldsymbol{\vartheta}^{(g)}\right) \tag{5.16b}$$

Lokálne neurónové siete sú určené grafmi $G^{(1)}$, $G^{(2)}$, ..., $G^{(t)}$, ktoré sú ohraničené tak, že všetky obsahujú rovnaký počet vstupných a výstupných neurónov. Orientovaný graf $G^{(g)}$, priradený bránovej neurónovej sieti, má tiež rovnaký počet vstupných neurónov ako lokálne siete, ale počet jeho výstupných neurónov sa rovná počtu t lokálnych sietí (pozri obr. 5.7). Naviac sa predpokladá, že výstupné aktivity bránovej neurónovej siete sú z otvoreného intervalu (0,1), t.j. prechodová funkcia (5.11a) je špecifikovaná parametrami A=0, B=1. Označme ako $\boldsymbol{x}_O^{(i)} = \left(x_1^{(i)}, x_2^{(i)}, \ldots, x_t^{(i)}\right)$ resp. $\boldsymbol{x}_O^{(g)} = \left(x_1^{(g)}, x_2^{(g)}, \ldots, x_t^{(g)}\right)$ výstupné aktivity jednotlivých lokálnych sietí resp. bránovej siete ako odozvu na rovnaký vektor vstupných aktivít $\boldsymbol{x}_I$. Poznamenajme, že všetky zložky týchto vektorov výstupných aktivít sú z otvoreného intervalu (0,1). *Koeficienty proporcionality*, produkované bránovou sieťou, sa určia ako odozva na vstupný vektor $\boldsymbol{x}_I$

$$p_i = \frac{x_i^{(g)}}{\sum_{(j=1)}^{(t)} x_j^{(g)}} \qquad (i = 1,2,\ldots,t) \tag{5.17}$$

Tieto koeficienty nadobúdajú hodnoty z otvoreného intervalu (0,1) a ich suma sa rovná jednej

$$p_1 + p_2 + \ldots + p_t = 1 \tag{5.18}$$



Obrázok 5.7. Znázornenie adaptívnej kombinácie lokálnych neurónových sietí. Selektor realizuje výber výstupnej aktivity $\boldsymbol{x}_O$ na základe výstupných aktivít lokálnych neurónových sietí.

V našich ďalších úvahách tieto koeficienty proporcionality budeme interpretovať ako "pravdepodobnosti" toho, že príslušná lokálna neurónová sieť je použitá na klasifikáciu objektu popísaného deskriptorom — vstupnou aktivitou $\boldsymbol{x}_I$.

Ako sa určí výstupná aktivita $\boldsymbol{x}_O$ adaptívnej kombinácie lokálnych neurónových sietí? Pre vstupnú aktivitu spočítame výstupné aktivity všetkých lokálnych sietí a bránovej siete, $\boldsymbol{x}_O^{(1)}, \boldsymbol{x}_O^{(2)}, \ldots, \boldsymbol{x}_O^{(t)}, \boldsymbol{x}_O^{(g)}$. Výstupná aktivita celej siete môže byť určená ako konvexná[1] kombinácia výstupných aktivít jednotlivých lokálnych sietí, pričom koeficienty konvexnej kombinácie sú určené pomocou koeficientov proporcionality

$$\boldsymbol{x}_O = p_1 \boldsymbol{x}_O^{(1)} + p_2 \boldsymbol{x}_O^{(2)} + \ldots + p_t \boldsymbol{x}_O^{(t)} \tag{5.19}$$

Výstupný vektor $\boldsymbol{x}_O$ považujeme za odozvu adaptívnej kombinácie lokálnych sietí na vektor vstupných aktivít $\boldsymbol{x}_I$. V limitnom prípade — ak len jeden koeficient proporcionality p_i je blízky jednotke a ostatné sú skoro nulové — hovoríme, že i-ta lokálna sieť interpretuje objekt so vstupnou aktivitou $\boldsymbol{x}_I$, a táto sieť je teda "expert" na klasifikáciu daného objektu. Spôsob konštrukcie (výberu) môže byť realizovaný aj tak, že sa vyberie lokálna sieť s maximálnou hodnotou koeficientu proporcionality

$$j = \arg \max_{1 \le i \le t} p_i \tag{5.20}$$

potom $\boldsymbol{x}_O = \boldsymbol{x}_O^{(j)}$. To znamená, že adaptívna kombinácia lokálnych sietí poskytuje ako odozvu na vstupnú aktivitu $\boldsymbol{x}_I$ výstupnú aktivitu tej lokálnej siete, ktorá má maximálny koeficient proporcionality. V našich ďalších úvahách budeme používať formulu (5.19) pre určenie výstupnej aktivity adaptívnej kombinácie lokálnych sietí. Jej hlavnou výhodou pred ostatnými prístupmi (napr. pred prístupom založenom na maximálnej hodnote koeficienta proporcionality (5.20)) je jej "spojitosť" a "diferencovateľnosť".

5.3 Adaptácia neurónovej siete

Adaptácia neurónovej siete spočíva v hľadaní takých prahových a váhových koeficientov, ktoré pre danú dvojicu vstupného a požadovaného výstupného vektora $\boldsymbol{x}_I / \hat{\boldsymbol{x}}_O$ a vypočítaného výstupného vektora $\boldsymbol{x}_O$, určeného vzťahom (5.14), minimalizujú rozdiel medzi výstupnými aktivitami $\boldsymbol{x}_O$ a $\hat{\boldsymbol{x}}_O$. Zostrojme účelovú funkciu

$$E = \frac{1}{2}\left(\boldsymbol{x}_O - \hat{\boldsymbol{x}}_O\right)^2 = \frac{1}{2}\sum_k g_k^2 \tag{5.21a}$$

[1] Konvexná kombinácia vektorov $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ je taká lineárna kombinácia $\alpha_1\boldsymbol{x}_1+\alpha_2\boldsymbol{x}_2+\ldots+\alpha_n\boldsymbol{x}_n$, kde lineárne koeficienty α_i sú nezáporné a ich suma sa rovná jednej.

$$g_k = \begin{cases} (x_k - \hat{x}_k) & (\text{pre } k \in V_O) \\ 0 & (\text{pre } k \notin V_O) \end{cases} \tag{5.21b}$$

kde x_k a $\hat{x}_k$ sú komponenty vektorov $\boldsymbol{x}_O$ resp. $\hat{\boldsymbol{x}}_O$, a $\boldsymbol{a}^2$ je skalárny súčin $\boldsymbol{a}\cdot\boldsymbol{a}=\sum a_i^2$. Cieľom adaptačného procesu je nájdenie takých prahových a váhových koeficientov, ktoré minimalizujú účelovú funkciu E. Pre viac párov vstupných a výstupných vektorov

$$\boldsymbol{x}_I^{(1)} / \hat{\boldsymbol{x}}_O^{(1)}, \boldsymbol{x}_I^{(2)} / \hat{\boldsymbol{x}}_O^{(2)}, \ldots, \boldsymbol{x}_I^{(r)} / \hat{\boldsymbol{x}}_O^{(r)} \tag{5.22}$$

(ktoré tvoria tréningovú množinu), má účelová funkcia (5.21) tvar

$$E = \sum_{i=1}^{r} E^{(i)} \tag{5.23a}$$

$$E^{(i)} = \frac{1}{2}\left(\boldsymbol{x}_O^{(i)} - \hat{\boldsymbol{x}}_O^{(i)}\right)^2 \tag{5.23b}$$

kde $\boldsymbol{x}_O^{(i)}$ je výstupný vektor neurónovej siete, určený vzťahom (5.14) ako odozva na vstupný vektor $\boldsymbol{x}_I^{(i)}$ a $\hat{\boldsymbol{x}}_O^{(i)}$ je požadovaný výstupný vektor priradený vstupu $\boldsymbol{x}_I^{(i)}$.

5.3.1 Adaptačný proces perceptrónu

Perceptrón je najjednoduchšia forma neurónovej siete, ktorá obsahuje len dve vrstvy [1]. Spodná vrstva obsahuje p vstupných neurónov a horná vrstva obsahuje len jeden výstupný neurón (to znamená, že perceptrón neobsahuje skryté neuróny), pozri obr. 5.8. Orientované spoje sú ohodnotené váhovými koeficientmi w_i (kde index i vyjadruje index vstupného neurónu, z ktorého daná hrana vychádza) a výstupný neurón je ohodnotený prahovým koeficientom ϑ. Výstupná aktivita y je určená vzťahom (pozri (5.10a-b))

$$y = t\left(\vartheta + w_1 x_1 + w_2 x_2 + \ldots w_p x_p\right) \tag{5.24}$$

Predpokladajme, že tréningová množina obsahuje r párov $\mathbf{x}_1 / y_1$, $\mathbf{x}_2 / y_2, \ldots, \mathbf{x}_r / y_r$ kde $\boldsymbol{x}_i = \left(x_1^{(i)}, x_2^{(i)}, \ldots, x_p^{(i)}\right)$, pre i=1, 2,..., r, sú vektory vstupných aktivít. Budeme predpokladať, že tréningová množina je *neprotirečivá*, t.j. ak pre rôzne dva indexy i a j platí $\boldsymbol{x}_i=\boldsymbol{x}_j$, potom požadované príslušné aktivity tiež musia byť rovnaké, $y_i=y_j$. Rovnica (5.24) pre i-ty pár z tréningovej množiny má tvar

$$y_i = t\left(\vartheta + w_1 x_1^{(i)} + w_2 x_2^{(i)} + \ldots + w_p x_p^{(i)}\right) \tag{5.25}$$

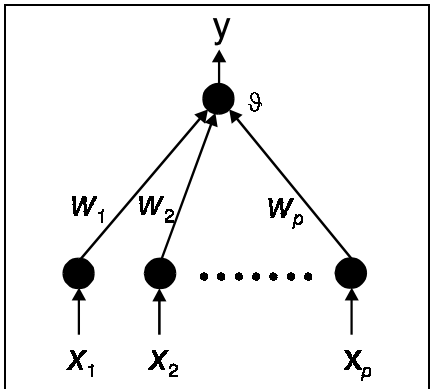
Pretože prechodová funkcia t je monotónne rastúca, musí k nej existovať inverzná funkcia t^{-1}. Potom

$$\vartheta + w_1 x_1^{(i)} + w_2 x_2^{(i)} + \ldots + w_p x_p^{(i)} = t^{-1}(y_i) = \chi_i \tag{5.26}$$

Tieto rovnice tvoria systém lineárnych rovníc pre neznáme $\vartheta, w_1, \ldots, w_p$. Ich maticový tvar je

$$\boldsymbol{A}\boldsymbol{w} = \boldsymbol{\chi} \tag{5.27a}$$

kde $\boldsymbol{A}$ je obdĺžniková matica typu $r \times (p+1)$



Obrázok 5.8. Perceptrón, neurónová sieť s dopredným šírením, obsahujúca len dve vrstvy neurónov (vstupnú a výstupnú) .

$$\boldsymbol{A} = \begin{pmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \ldots & x_p^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & \ldots & x_p^{(2)} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_1^{(r)} & x_2^{(r)} & \ldots & x_p^{(r)} \end{pmatrix} \tag{5.27b}$$

a $\boldsymbol{w}$ resp. $\boldsymbol{\chi}$ sú stĺpcové vektory určené ako $\boldsymbol{w}=(\vartheta, w_1, \ldots, w_p)^{\mathrm{T}}$ resp. $\boldsymbol{\chi}=(\chi_1, \chi_2, \ldots, \chi_r)^{\mathrm{T}}$. To znamená, že adaptačný proces perceptrónu možno pretransformovať na riešenie systému lineárnych rovníc (5.26). Ak tento systém má riešenie, potom povieme, že tréningová množina (obsahujúca r párov tréningových vzorov $\boldsymbol{x}_i/y_i$) je *lineárne separovateľná*. V opačnom prípade, ak systém (5.26) nemá riešenie, perceptrón nemôže byť korektne adaptovaný (hovoríme, že tréningová množina *je lineárne neseparovateľná*).

V prípade, že systém (5.26) nemá riešenie (t.j. podľa Frobeniovej vety [9] vtedy a len vtedy, ak platí hodnosť($\boldsymbol{A}$)≠hodnosť($\boldsymbol{A}|\boldsymbol{\chi}$)), môžeme zostrojiť približné riešenie (v zmysle metódy najmenších štvorcov) použitím prístupu pseudoinverznej matice [9]. Násobme zľava rovnicu (5.27a) transponovanou maticou $\boldsymbol{A}^{\mathrm{T}}$, a dostaneme $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{w}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\chi}$, kde $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ je

pozitívne semidefinitná matica typu $r \times r$. Potom, formálnym aplikovaním matice $(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}$ zľava, dostaneme konečné riešenie (5.27a) v tvare

$$\boldsymbol{w} = \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}\chi \tag{5.28}$$

kde matica $(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}$ sa nazýva pseudoinverzná matica. Riešenie (5.28) sa nazýva *zovšeobecnené riešenie* rovnice (5.27a) a minimalizuje euklidovskú normu $|\boldsymbol{A}\boldsymbol{w} - \boldsymbol{\chi}|$. Ak $|\boldsymbol{A}\boldsymbol{w} - \boldsymbol{\chi}|=0$, potom $\boldsymbol{w}$ je presné riešenie (5.27a). Poznamenajme, že aj keď sme v (5.28) použili explicitný výraz pre pseudoinverznú maticu, tento je platný len za predpokladu, že matica $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ je regulárna (t.j. existuje k nej inverzná matica). V opačnom prípade, ak matica $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ je singulárna (t.j. neexistuje k nej inverzná matica), existujú metódy jej konštrukcie, ktoré nepožadujú znalosť inverznej matice $(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}$. Konštrukcia pseudoinverznej matice $(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}$ patrí medzi štandardné numerické problémy, program pre jej implementáciu je uvedený napr. v monografii [10].

Vyššie uvedený adaptačný proces je ľahko zovšeobecniteľný aj pre perceptróny obsahujúce viac ako jeden výstupný neurón. V tomto prípade pre každý výstupný neurón zostrojíme nezávislé zovšeobecnené riešenie (5.28), ktoré je najlepšie pre daný výstupný neurón. Žiaľ, tento jednoduchý algebraický prístup k adaptácii perceptrónu je neaplikovateľný pre neurónovú sieť obsahujúcu skryté neuróny, pretože nie je možné linearizovať analógiu rovnice (5.25) pomocou inverznej prechodovej funkcie do tvaru systému lineárnych rovníc.

Ilustračný príklad (logická funkcia XOR)

Logická funkcia XOR zohrala v histórii neurónových sietí dôležitú úlohu. Koncom 60-tych rokov Minsky a Papert v známej knihe [1] kritizovali perceptrón ako prístup, ktorý nie je schopný realizovať ľubovoľnú výpočtovú úlohu (napr. XOR funkciu). Na základe tohto pozorovania dospeli k záveru, že neurónové siete (perceptróny patria medzi ne) nie sú univerzálnym výpočtovým prostriedkom.

Tabuľka 5.1. Funkcia XOR

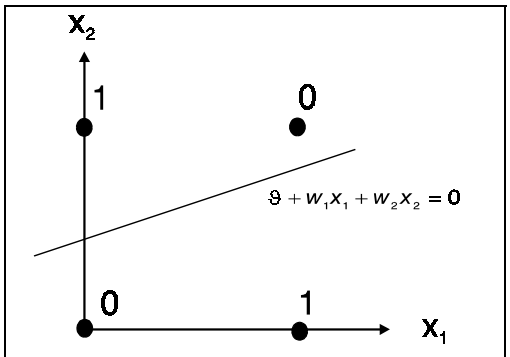
x_1	x_2	y(XOR)
0	0	0
0	1	1
1	0	1
1	1	0

Logická funkcia XOR (vylučujúce alebo) je určená pomocou tab. 5.1. Ide o boolovskú funkciu $y=F(x_1,x_2)$, kde x_1 a x_2 sú nezávislé dvojhodnotové premenné a y je závislá dvojhodnotová premenná. Pokúsime sa realizovať túto funkciu pomocou perceptrónu, ktorý obsahuje dva vstupné neuróny (s aktivitami x_1 a x_2) a jeden výstupný neurón s aktivitou y. Tréningová množina obsahuje 4 objekty z tab. 5.1. Systém (5.27a) má potom tvar

$$\begin{aligned} \vartheta \qquad\qquad &= -Q \\ \vartheta \qquad + w_2 &= Q \\ \vartheta + w_1 \qquad &= Q \\ \vartheta + w_1 + w_2 &= -Q \end{aligned} \tag{5.29}$$

kde Q je dostatočne veľké kladné číslo (určené tak, aby pre funkčné hodnoty prechodovej funkcie platilo $t(Q)=1-\varepsilon$ a $t(-Q)=\varepsilon$, pre malé kladné číslo ε). Systém (5.29) nemá riešenie (o čom sa môžeme jednoducho presvedčiť tak, že od štvrtej rovnice odpočítame druhú a tretiu rovnicu, dostaneme $\vartheta=3Q$, čo je v protirečení s prvou rovnicou).

Ako interpretovať tento výsledok? Argument v prechodovej funkcii vo vzťahu (5.24), ktorý špecifikuje aktivitu výstupného neurónu, možno formálne chápať ako rovnicu roviny $\vartheta+w_1x_1+w_2x_2+...+w_px_p=0$ v p-rozmernom priestore vstupných aktivít. Táto rovina rozdeľuje priestor na dva polpriestory, ktoré sú buď nad rovinou ($\vartheta+w_1x_1+w_2x_2+...+w_px_p>0$) alebo pod rovinou ($\vartheta+w_1x_1+w_2x_2+...+w_px_p<0$). Zhruba povedané, bod ležiaci nad (pod) rovinou má jednotkovú (nulovú) výstupnú aktivitu. Pretože systém (5.29) nemá riešenie, neexistuje rovina, ktorá korektne rozdeľuje 2-rozmerný priestor so súradnicami x_1 a x_2 na dva podpriestory tak, aby body (0,0) a (1,1) ležali pod rovinou a body (0,1) a (1,0) nad rovinou (v tomto prípade ide o priamkou určenú rovnicou $\vartheta+w_1x_1+w_2x_2=0$), pozri obr. 5.9. Môžeme teda povedať, že dve triedy objektov XOR problému sú lineárne neseparovateľné.

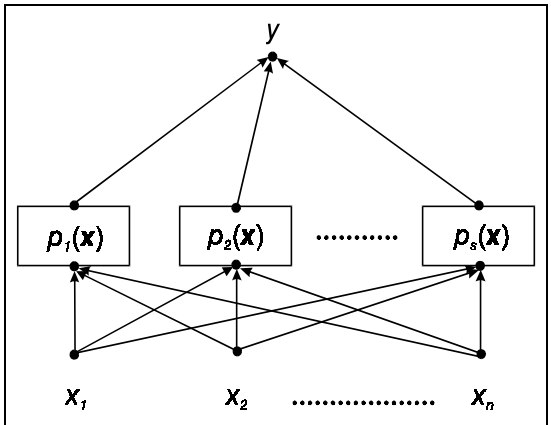


Obrázok 5.9. Znázornenie "nadroviny" (určenej rovnicou $\vartheta+w_1x_1+w_2x_2=0$), ktorá rozdeľuje priestor vstupných aktivít na dva podpriestory.

5.3.2 Adaptačný proces perceptrónu vyššieho rádu

Perceptrón vyššieho rádu je zovšeobecnením obyčajného perceptrónu, a to takým spôsobom, že jeho výstupná aktivita je určená nielen lineárnymi členmi, ale aj kvadratickými, kubickými, atď., členmi (pozri definíciu neurónovej siete vyššieho rádu v podkapitole 5.2.1). Aktivita výstupného neurónu je určená vzťahom

$$y = t\left(\vartheta + w_1 p_1(\boldsymbol{x}) + w_2 p_2(\boldsymbol{x}) + \ldots + w_s p_s(\boldsymbol{x})\right) \tag{5.30}$$



Obrázok 5.10. Schematické znázornenie perceptrónu vyššieho rádu. Obdĺžnikové bloky sú priradené multilineárnym funkciám $p_i(\boldsymbol{x})$.

kde $p_i(\boldsymbol{x}) = x^{\alpha_1} x^{\alpha_2} \ldots x^{\alpha_n}$ sú rôzne multilineárne členy určené exponentmi α_1, α_2, ..., α_p. Diagramatická interpretácia perceptrónu vyššieho rádu je znázornená na obr. 5.10, kde obdĺžnikové bloky reprezentujú multilineárne funkcie $p_i(\boldsymbol{x})$. Ak študujeme perceptrón druhého rádu, potom exponenty vyhovujú buď podmienke $\alpha_1+\alpha_2+...+\alpha_p=1$ (lineárny člen) alebo $\alpha_1+\alpha_2+...+\alpha_p=2$ (kvadratický člen).

Výstupná aktivita je určená vzťahom

$$y = t\left(\vartheta + w_1 x_1 + w_2 x_2 + w_{12} x_1 x_2 + w_{11} x_1^2 + w_{22} x_2^2 + \ldots\right) \tag{5.31}$$

kde sme pre jednoduchosť uviedli len niekoľko prvých členov. Použitím rovnakej linearizačnej procedúry ako v podkapitole 5.3.1 dostaneme rovnaký systém lineárnych rovníc ako (5.27a), matica $\boldsymbol{A}$ má tvar

$$\boldsymbol{A} = \begin{pmatrix} 1 & p_1\left(\boldsymbol{x}^{(1)}\right) & p_2\left(\boldsymbol{x}^{(1)}\right) & \ldots & p_s\left(\boldsymbol{x}^{(1)}\right) \\ 1 & p_1\left(\boldsymbol{x}^{(2)}\right) & p_2\left(\boldsymbol{x}^{(2)}\right) & \ldots & p_s\left(\boldsymbol{x}^{(2)}\right) \\ & & \ldots\ldots\ldots\ldots & & \\ 1 & p_1\left(\boldsymbol{x}^{(r)}\right) & p_2\left(\boldsymbol{x}^{(r)}\right) & \ldots & p_s\left(\boldsymbol{x}^{(r)}\right) \end{pmatrix} \tag{5.32}$$

Prahové a váhové koeficienty sú určené systémom lineárnych rovníc (5.27a), alebo explicitne vzťahom (5.28).

Predpokladajme, že rovnica (5.27a) nemá riešenie, t.j. platí hodnosť($\boldsymbol{A}$)≠hodnosť($\boldsymbol{A}|\boldsymbol{\chi}$). Zavedením nového multilineárneho člena v (5.30) dostaneme novú maticu koeficientov $\boldsymbol{A}'$

(matica $\boldsymbol{A}$ je rozšírená sprava o nový stĺpec), a potom platí jeden z nasledujúcich dvoch prípadov:

(1) Nová matica $\boldsymbol{A}'$ už vyhovuje podmienke hodnosť($\boldsymbol{A}'$)=hodnosť($\boldsymbol{A}'|\boldsymbol{\chi}$), a potom systém (5.28) má riešenie. To znamená, že prahové a váhové koeficienty perceptrónu sú určené riešením systému lineárnych rovníc (5.27a).

(2) Nová matica $\boldsymbol{A}'$ stále nevyhovuje podmienke hodnosť($\boldsymbol{A}'$)= =hodnosť($\boldsymbol{A}'|\boldsymbol{\chi}$), a potom riešenie $\boldsymbol{w}'$ zostrojené pomocou pseudoinverznej matice (5.28) vyhovuje podmienke $|\boldsymbol{A}'\boldsymbol{w}' - \boldsymbol{\chi}| < |\boldsymbol{A}\boldsymbol{w} - \boldsymbol{\chi}|$, t.j. nové riešenie $\boldsymbol{w}'$ je zostrojené s menšou chybou (v zmysle euklidovskej vzdialenosti) ako pôvodné riešenie $\boldsymbol{w}$. To znamená, že ak zavedieme nový multilineárny člen v (5.30), potom prahové a váhové koeficienty určené pomocou pseudoinverznej matice (5.28) poskytujú lepšiu klasifikáciu objektov z tréningovej množiny. V limitnom prípade, ak sme zaviedli dostatočný počet multilineárnych členov môže nastať situácia, že buď podmienka riešiteľnosti hodnosť($\boldsymbol{A}$)=hodnosť($\boldsymbol{A}|\boldsymbol{\chi}$) začne platiť, alebo stále hodnosť($\boldsymbol{A}$)≠hodnosť($\boldsymbol{A}|\boldsymbol{\chi}$), avšak "nepresnosť" $|\boldsymbol{A}\boldsymbol{w} - \boldsymbol{\chi}|$ je už menšia ako predpísané malé kladné číslo ε (riešenie $\boldsymbol{w}$ je "správne" s presnosťou ε). Táto jednoduchá procedúra rozširovania perceptrónu novými multilineárnymi členmi vyššieho rádu je umožnená tým, že použité multilineárne členy $p_1(\boldsymbol{x})$, $p_2(\boldsymbol{x})$, ..., $p_s(\boldsymbol{x})$ sú lineárne nezávislé.

Možno teda povedať, *že perceptrón dostatočne vysokého rádu je schopný klasifikovať korektne s predpísanou presnosťou ε ľubovoľnú neprotirečivú tréningovú množinu*. Toto je veľmi dôležitá vlastnosť perceptrónov vyššieho rádu, ináč povedané, tieto perceptróny sú *univerzálne aproximátory funkcií*, t.j. sú schopné ich aproximovať s ľubovoľnou predpísanou presnosťou.

Ilustračný príklad (logická funkcia XOR)

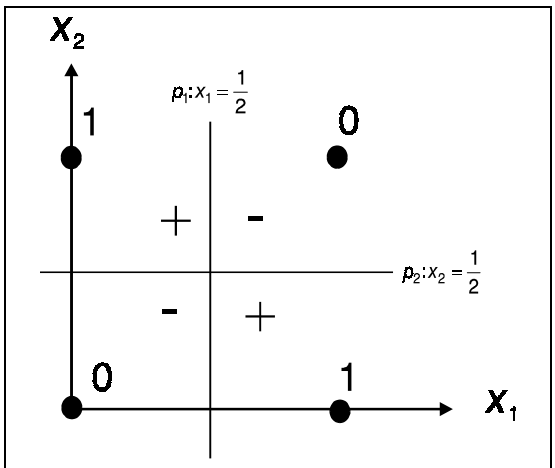
V predchádzajúcom ilustračnom príklade sme ukázali, že obyčajný perceptrón (t.j. perceptrón prvého rádu) nie je schopný klasifikovať logickú funkciu XOR. Toto obmedzenie obyčajného perceptrónu len na lineárne separovateľné vstupné aktivity objektov z tréningovej množiny sa jednoducho odstráni použitím perceptrónu vyššieho rádu. Na klasifikáciu XOR funkcie použijeme perceptrón 2. rádu. Nech potenciál výstupného neurónu obsahuje člen $w_{12}x_1x_2$ (ďalšie členy 2. rádu $w_{11}x_1^2$ a $w_{22}x_2^2$ nie sú uvažované, pretože v dôsledku binárneho charakteru vstupných aktivít x_1 a x_2 sú totožné s členmi 1. rádu). Analógia systému lineárnych rovníc (5.29) má tvar

$$\begin{aligned} \vartheta &= -Q \\ \vartheta + w_2 &= Q \\ \vartheta + w_1 &= Q \\ \vartheta + w_1 + w_2 + w_{12} &= -Q \end{aligned} \qquad (5.33)$$

Tento systém rovníc (pre neznáme ϑ, w_1, w_2 a w_{12}) má jednoznačné riešenie (hodnosť($\boldsymbol{A}$)=hodnosť($\boldsymbol{A}|\boldsymbol{\chi}$)=4), a platí $\vartheta=-Q$, $w_1=w_2=2Q$, $w_{12}=-4Q$. Uvažuj-me funkciu 2. rádu (potenciál výstupného neurónu) určenú implicitne vzťahom $-Q+2Qx_1+2Qx_2-4Qx_1x_2=0$, alebo v zjednodušenom tvare

$$x_1 + x_2 - 2x_1x_2 = \frac{1}{2} \tag{5.34}$$

Táto rovnica má dve nezávislé riešenia, ktoré určujú dve priamky kolmé na osy x_1 alebo x_2



Obrázok 5.11. Funkcia 2. rádu (5.34) rozdeľuje priestor vstupných aktivít (pomocou priamok p_1 a p_2 určených vzťahmi (5.35)) na 4 podpriestory, pričom je dosiahnutá správna interpretácia XOR problému.

$$p_1: x_1 = \frac{1}{2}, \qquad p_2: x_2 = \frac{1}{2} \tag{5.35}$$

Tieto dve priamky určujú hraničné oblasti v rovine x_1-x_2, kde argument (potenciál) prechodovej funkcie v (5.31) mení znamienko (pozri obr. 5.11). Perceptrón druhého rádu je schopný korektne klasifikovať XOR funkciu, priestor vstupných aktivít je rozdelený na 4 podpriestory, kde potenciál nadobúda správne znamienko. Tento jednoduchý príklad môže byť chápaný ako ilustrácia toho, že perceptrón vyššieho rádu je schopný klasifikovať objekty neprotirečivej tréningovej množiny.

5.3.3 Adaptácia neurónovej siete s dopredným šírením

Adaptačný proces neurónovej siete s dopredným šírením, ktorá obsahuje skryté neuróny, nemôže byť uskutočnený takým jednoduchým spôsobom ako pre perceptrón, kde sa adaptačný proces redukuje na riešenie systému lineárnych rovníc (použitím prístupu pseudoinverznej matice). Pre neurónové siete, ktoré obsahujú skryté neuróny, nie je možné linearizovať systém rovníc, ktoré popisujú aktivity skrytých a výstupných neurónov. Preto musíme obrátiť našu pozornosť na takú adaptáciu neurónovej siete, ktorá minimalizuje účelovú funkciu (5.21) alebo (5.23). Túto minimalizáciu nelineárnej účelovej funkcie možno uskutočniť mnohými optimalizačnými metódami, ktoré sú známe v numerickej

matematike [11]. Medzi najefektívnejšie patria tzv. gradientové metódy, založené na použití gradientu účelovej funkcie pre iteratívnu konštrukciu optimálneho riešenia. Pri ich použití musíme poznať gradient účelovej funkcie (t.j. parciálne derivácie $\partial E/\partial \vartheta_i$ a $\partial E/\partial w_{ij}$) a jeho výpočet bude náplňou tejto podkapitoly.

Parciálne derivácie účelovej funkcie E (5.21) vzhľadom k prahovým a váhovým faktorom sú určené vzťahmi (jednoduchá aplikácia formuly pre parciálnu deriváciu zloženej funkcie [12])

$$\begin{aligned} \frac{\partial E}{\partial w_{ij}} &= \frac{\partial E}{\partial x_i}\frac{\partial x_i}{\partial w_{ij}} = \frac{\partial E}{\partial x_i} t'(\xi_i) x_j \\ \frac{\partial E}{\partial \vartheta_i} &= \frac{\partial E}{\partial x_i}\frac{\partial x_i}{\partial \vartheta_i} = \frac{\partial E}{\partial x_i} t'(\xi_i) \end{aligned} \tag{5.36}$$

kde parciálna derivácia $\partial x_i/\partial w_{ij}$ je jednoducho spočítaná pomocou $x_i=t(\xi_i)$ (pozri (5.10)). Potom dostaneme $\partial x_i/\partial w_{ij} = t'(\xi_i)\cdot\partial\xi_i/\partial w_{ij} = t'(\xi_i)x_j$. Podobné úvahy sú aplikovateľné aj pre výpočet parciálnej derivácie $\partial x_i/\partial \vartheta_i$. Porovnaním rovníc (5.36) dostaneme jednoduchý vzťah medzi parciálnymi deriváciami $\partial x_i/\partial w_{ij}$ a $\partial x_i/\partial \vartheta_i$

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial \vartheta_i} x_j \tag{5.37}$$

Výpočet gradientu sa redukuje na výpočet parciálnych derivácií účelovej funkcie vzhľadom k prahovým faktorom, podľa (5.37) parciálne derivácie vzhľadom k váhovým koeficientom sú určené pomocou jednoduchších parciálnych derivácií vzhľadom k prahovým koeficientom. Upriamime našu pozornosť na výpočet parciálnych derivácií $\partial E/\partial x_i$. Ich výpočet závisí od toho, či index i popisuje výstupný neurón alebo skrytý neurón,

$$\begin{aligned} \frac{\partial E}{\partial x_i} &= g_i \qquad (\text{pre } i \in V_O) \\ \frac{\partial E}{\partial x_i} &= \sum_k \frac{\partial E}{\partial x_k}\frac{\partial x_k}{\partial x_i} \quad (pre\ i \in V_H) \end{aligned} \tag{5.38}$$

kde v druhom výraze sumácia beží cez všetky neuróny, ktoré nasledujú za i-tym neurónom. Výraz g_i je určený vzťahom (5.21b), ktorý je nulový pre iné neuróny ako výstupné. Formuly z (5.38) môžeme zjednotiť do jedného vzťahu

$$\frac{\partial E}{\partial x_i} = g_i + \sum_k \frac{\partial E}{\partial x_k}\frac{\partial x_k}{\partial x_i} \tag{5.39}$$

kde je potrebné si uvedomiť, že člen na pravej strane obsahujúci sumáciu, ako už bolo uvedené vyššie, je nulový pre index i popisujúci výstupný neurón. Podobne, ako v (5.36), pre parciálne derivácie z pravej strany (5.39) platí $\partial x_k/\partial x_i = t'(\xi_k)w_{ki}$, dosadením (5.39)

do (5.36) dostaneme konečnú formulu pre výpočet parciálnych derivácií účelovej funkcie E vzhľadom k prahovým faktorom

$$\frac{\partial E}{\partial \vartheta_i} = t'(\xi_i)\left(g_i + \sum_k \frac{\partial E}{\partial \vartheta_k} w_{ki} \right) \qquad \left(\text{pre } i \in \mathsf{V}_H \cup \mathsf{V}_O\right) \tag{5.40}$$

kde derivácia prechodovej funkcie $t'(\xi_i)$ je určená vzťahom (5.11b) a sumácia beží cez všetky neuróny, ktoré sú nasledovníkmi i-teho neurónu.

Vo všeobecnosti možno charakterizovať vzťah (5.40) ako systém lineárnych rovníc, ktorého riešenie určujú parciálne derivácie $\partial E/\partial \vartheta_i$. Pre neurónovú sieť typu dopredného šírenia je možné zostrojiť riešenie tohto systému jednoduchým rekurentným postupom. V prvom kroku vypočítame parciálne derivácie $\partial E/\partial \vartheta_i$ pre výstupné neuróny (výstupná vrstva L_t), pre ktoré platí $\partial E/\partial \vartheta_i = t'(\xi_i)\, g_i$. Pomocou (5.40) potom môžeme spočítať parciálne derivácie $\partial E/\partial \vartheta_i$ pre neuróny z predposlednej vrstvy L_{t-1}. Pri výpočte parciálnych derivácií $\partial E/\partial \vartheta_i$ pre neuróny z vrstvy L_j musíme poznať tieto derivácie z nasledujúcich vrstiev L_{j+1}, L_{j+2}, ..., L_t. Výpočet končí, keď zostrojíme parciálne derivácie pre neuróny z druhej vrstvy L_2. Poznajúc všetky parciálne derivácie $\partial E/\partial \vartheta_i$ pre celú neurónovú sieť, určíme parciálne derivácie $\partial E/\partial w_{ij}$ jednoducho pomocou vzťahu (5.37). Spôsob výpočtu parciálnych derivácií pre neurónovú sieť s dopredným šírením popísaný vyššie, prebiehajúci rekurentne od najvyššej k najnižšej vrstve (t.j. proti smeru šírenia informácie v neurónovej sieti, ktorá prebieha od najnižšej k najvyššej vrstve — tzv. dopredné šírenie) je aj hlavným dôvodom toho, prečo sa tento postup v literatúre často nazýva *spätné šírenie* (angl. *back propagation*). Vzťah podobný formule (5.40) pre výpočet parciálnych derivácií účelovej funkcie vzhľadom k prahovým a váhovým koeficientom bol odvodený v r. 1986 Rumelhartom so spolupracovníkmi [2]. Táto práca je pokladaná za jeden z historických medzníkov rozvoja teórie neurónových sietí, pretože v nej bolo ukázané na mnohých príkladoch (ktoré boli evidentne lineárne neseparovateľné), že zovšeobecnenie perceptrónu tak, aby obsahoval skryté neuróny, spolu s metódou spätného šírenia pre výpočet gradientu účelovej funkcie, je schopné prekonať limity stanovené Minskym a Papertom [1] pre jednoduchý perceptrón neobsahujúci skryté neuróny.

Pri odvodení formuly (5.40) nebol použitý predpoklad, že graf reprezentujúci neurónovú sieť je acyklický (t.j. neurónová sieť je typu dopredného šírenia). Preto táto formula platí pre ľubovoľnú neurónovú sieť, ktorá môže obsahovať aj orientované cykly (tzv. rekurentné siete, pozri kapitolu 6). Avšak v tomto prípade už nie je použiteľný postup spätného šírenia pre výpočet parciálnych derivácií, tieto sú teraz určené ako riešenie systému lineárnych rovníc (5.40) pre parciálne derivácie $\partial E/\partial \vartheta_i$. Totiž sumácia v (5.40) môže obsahovať vo všeobecnosti aj parciálne derivácie $\partial E/\partial \vartheta_i$, ktoré ešte neboli spočítané v predchádzajúcich krokoch rekurentného výpočtu.

Vyššie uvedený postup pre výpočet gradientu účelovej funkcie je ľahko zovšeobecniteľný aj pre účelovú funkciu, ktorá obsahuje viac ako jeden pár vstupno-výstupných vektorov $\boldsymbol{x}_l / \hat{\boldsymbol{x}}_O$ (pozri (5.23a-b)). Potom celkový gradient účelovej funkcie sa jednoducho určí ako suma gradientov spočítaných pomocou (5.40) pre všetky dvojice $\boldsymbol{x}_l / \hat{\boldsymbol{x}}_O$ z tréningovej množiny (5.22)

$$\text{grad } E = \sum_{i=1}^{r} \text{grad } E^{(i)} \tag{5.41}$$

kde účelová funkcia $E^{(i)}$ je definovaná (5.23b) pre *i*-tu dvojicu $\boldsymbol{x}_I / \hat{\boldsymbol{x}}_O$ z tréningovej množiny.

Ak poznáme gradient účelovej funkcie *E*, potom adaptačný proces neurónovej siete je realizovaný minimalizáciou účelovej funkcie *E* vzhľadom k prahovým a váhovým koeficientom. Formálne, adaptovaná neurónová sieť je popísaná koeficientmi, ktoré sú určené ako

$$\left(\overline{\boldsymbol{w}}, \overline{\boldsymbol{\vartheta}}\right) = \arg \min_{(\boldsymbol{w}, \boldsymbol{\vartheta})} E(\boldsymbol{w}, \boldsymbol{\vartheta}) \tag{5.42}$$

Jedným z najjednoduchších spôsobov (súčasne aj najviac používaným), ako realizovať túto minimalizáciu v rámci gradientových optimalizačných metód je metóda *najprudšieho spádu* (angl. *steepest descent*) [11], v ktorej váhové a prahové koeficienty sú rekurentne obnovované pomocou vzťahov

$$\begin{aligned} w_{ij}^{(k+1)} &= w_{ij}^{(k)} - \lambda \frac{\partial E}{\partial w_{ij}} + \mu \Delta w_{ij}^{(k)} \\ \vartheta_j^{(k+1)} &= \vartheta_j^{(k)} - \lambda \frac{\partial E}{\partial \vartheta_j} + \mu \Delta \vartheta_j^{(k)} \end{aligned} \tag{5.43}$$

kde parameter $\lambda>0$ musí byť dostatočne malý (obvykle $\lambda=0{,}01$-$0{,}1$), aby bola zabezpečená monotónna konvergentnosť optimalizačného algoritmu a súčasne dostatočne veľký pre zabezpečenie dostatočne vysokej rýchlosti konvergentnosti. Počiatočné hodnoty prahových a váhových koeficientov $\vartheta_j^{(0)}$ a $w_{ij}^{(0)}$ sú náhodne generované z malého intervalu so stredom v nule, napr. z otvoreného intervalu $(-1,1)$. Posledný člen v (5.43) odpovedá tzv. *momentovému členu*, ktorý je určený pomocou rozdielu koeficientov z posledných dvoch iterácií, $\Delta w_{ij}^{(k)} = w_{ij}^{(k)} - w_{ij}^{(k-1)}$ a $\Delta \vartheta_j^{(k)} = \vartheta_j^{(k)} - \vartheta_j^{(k-1)}$. Momen-tový člen (momentum) je dôležitý pre "obskočenie" lokálnych miním v počia-točnej fáze optimalizácie, hodnota parametru μ sa obvykle volí z intervalu $0{,}5 \le \mu \le 0{,}7$.

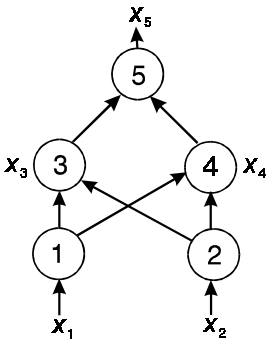
Ilustračný príklad (logická funkcia XOR)

Efektívnosť neurónovej siete obsahujúcej skryté neuróny ilustrujeme na príklade logickej funkcie XOR, ktorá pre obyčajný perceptrón nie je korektne klasifikovateľná. Použitá neurónová sieť bude obsahovať tri vrstvy, prvá vrstva obsahuje dva vstupné neuróny, druhá vrstva dva skryté neuróny a posledná tretia vrstva obsahuje jeden výstupný neurón (pozri obr. 5.12).

Tabuľka 5.2 Aktivity neurónovej siete z obr. 5.12
pre interpretáciu XOR problému

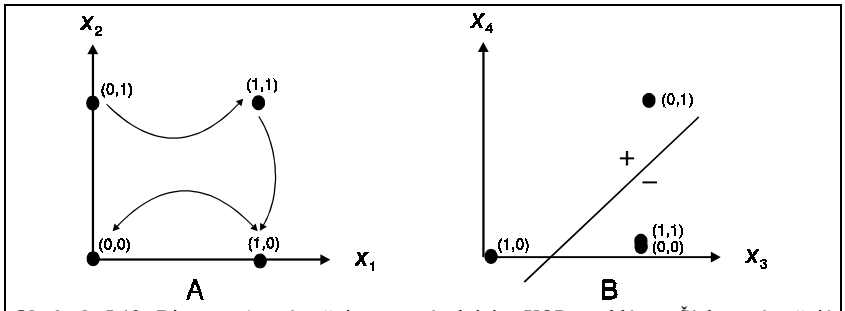
Čís.	x_1	x_2	x_3	x_4	x_5	$\hat{x}_5$
1	0,00	0,00	0,96	0,08	0,06	0,00
2	0,00	1,00	1,00	0,89	0,95	1,00
3	1,00	0,00	0,06	0,00	0,94	1,00
4	1,00	1,00	0,96	0,07	0,05	0,00

Skryté neuróny vytvárajú tzv. vnútornú reprezentáciu funkcie XOR, ktorá už je lineárne separovateľná. Táto skutočnosť, že skryté neuróny sú schopné zaviesť reprezentáciu, v ktorej sú už objekty správne interpretované, je hlavným dôvodom širokého používania neurónových sietí. Parametre adaptačného procesu boli tieto: λ=0,1, μ=0,5; po 400 iteráciach účelová funkcia mala hodnotu E=0,031. Výsledné aktivity skrytých a výstupných neurónov sú uvedené v tab. 5.2. Ak nakreslíme výsledné aktivity skrytých neurónov do roviny x_3-x_4 , vidíme, že tieto poskytujú vnútornú reprezentáciu, ktorá je lineárne separovateľná (pozri obr. 5.13).



Obrázok 5.12. Znázornenie jednoduchej neurónovej siete s dvoma skrytými neurónmi, ktorá je schopná správne interpretovať XOR problém.

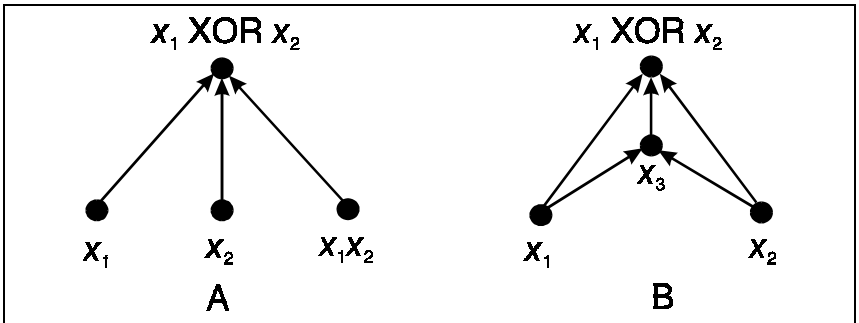
Perceptrón 2. rádu pre XOR funkciu (pozri podkapitolu 5.3.2, ilustračný príklad) je jednoducho realizovateľný pomocou neurónovej siete s jedným skrytým neurónom tak, že zavedieme novú (*funkcionálnu*) vstupnú aktivitu $x_3=x_1 \wedge x_2$ (t.j. nová vstupná aktivita je vytvorením funkcie AND aplikovanej na pôvodné vstupné aktivity x_1 a x_2), pozri obr. 5.14. Perceptrón A môže byť jednoducho pretransformovaný na neurónovú sieť B, ktorá obsahuje jeden skrytý neurón, pričom tento neurón vykonáva logickú funkciu AND.



Obrázok 5.13. Diagram A znázorňuje vstupné aktivity XOR problému. Šípky znázorňujú presun objektov v dôsledku adaptácie siete. Diagram B znázorňuje skryté aktivity neurónovej siete, ktoré sú priradené jednotlivým "obrazcom" XOR problému. Čo je dôležité, skryté aktivity už sú lineárne separovateľné "nadrovinou" určenou $\vartheta_5+w_{45}x_4+w_{35}x_3=0$.

5.3.4 Adaptácia neurónovej siete vyššieho rádu

Výpočet parciálnych derivácií $\partial E/\partial \vartheta_i$ a $\partial E/\partial w_{ji}$ pre neurónové siete vyššieho rádu je analógiou postupu pre neurónové siete s dopredným šírením prvého rádu (pozri podkapitolu 5.3.3). Pre jednoduchosť predpokladajme, že neurónová sieť je druhého rádu, t.j. aktivity (5.15a-b) sú určené lineárnymi a kvadratickými členmi. Zovšeobecnenie pre neurónové siete tretieho alebo vyššieho rádu je jednoduché. Parciálne derivácie účelovej funkcie vzhľadom k váhovým koeficientom prvého a druhého rádu sú určené takto (pozri (5.37))



Obrázok 5.14. Diagram A znázorňuje jednoduchú realizáciu perceptrónu 2. rádu zavedením nového vstupného neurónu, ktorému je priradené vstupná aktivita vyjadrená ako súčin (logická funkcia AND) aktivít prvého a druhého vstupného neurónu. Diagram B znázorňuje najjednoduchšiu neurónovú sieť s jedným skrytým neurónom, ktorého aktivita je realizovaná ako AND funkcia. Pripomeňme, že problém AND je lineárne separovateľný, čiže je možné zostrojiť pre tento problém jednoduchý perceptrón (ktorého výstupný neurón odpovedá skrytému neurónu 3).

$$\frac{\partial E}{\partial w_{i,j}} = \frac{\partial E}{\partial \vartheta_i} x_j$$
$$\frac{\partial E}{\partial w_{i,jk}} = \frac{\partial E}{\partial \vartheta_i} x_j x_k \tag{5.44}$$

Parciálne derivácie $\partial E/\partial \vartheta_i$ sú určené systémom lineárnych rovníc (pozri (5.40))

$$\frac{\partial E}{\partial \vartheta_i} = t'(\xi_i)\left[g_i + \sum_k \frac{\partial E}{\partial \vartheta_k}\left(w_{k,i} + 2w_{k,ii}\, x_i + \sum_{\substack{j\neq k \\ i<j}} w_{k,ij} x_j \right)\right] \tag{5.45}$$

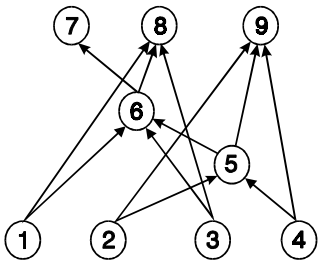
Druhý člen na pravej strane $2w_{k,ii}\, x_i$ odpovedá "čistému" kvadratickému členu $w_{k,ii}x_i^2$ v (5.15b), zatiaľ čo tretí člen $\sum_{(j\neq k, i<j)} w_{k,ij} x_j$ odpovedá "krížovým" členom v (5.15b). Rekurentná formula pre obnovu prahových a váhových koeficientov je analogická formulám (5.43) pre obyčajnú neurónovú sieť prvého rádu.

Ilustračný príklad (sumácia dvoch 2-bitových čísel)

Študujme sumáciu dvoch 2-bitových čísel

$$\begin{array}{r} \alpha_1\ \alpha_2 \\ +\ \alpha_3\ \alpha_4 \\ \hline \alpha_5\ \alpha_6\ \alpha_7 \end{array} \tag{5.46}$$

kde α_i sú binárne čísla. Tab. 5.3 obsahuje všetkých 16 možných realizácií súčtu (5.46), v poslednom stĺpci je uvedená dekadická interpretácia daného súčtu. Rumelhart so spolupracovníkmi [2] navrhli neurónovú sieť špeciálneho tvaru, ktorá obsahuje dva skryté neuróny a je schopná korektne interpretovať súčet (5.46). Skryté neuróny majú význam dvoch medzisúčtov, ktoré sú potrebné pre realizáciu celkového súčtu (medzisúčty pre druhý a prvý stĺpec v (5.46)), pozri obr. 5.15. Avšak s nimi navrhnutou neurónovou sieťou mali vážne problémy pri jej adaptačnom procese.
Z týchto dôvodov Rumelhart so spolupracovníkmi použili na klasifikáciu súčtu (5.46) neurónovú sieť s 3 alebo 4 skrytými neurónmi, ktoré už nemali konvergentné problémy. Tieto problémy adaptačného procesu neurónovej siete znázornenej na obr. 5.15 sa odstránia, ak sa tá bude interpretovať ako neurónová sieť druhého rádu. Po 3000 iteráciách bola adaptácia úspešná, hodnota účelovej funkcie je E=0,03, pre parametre λ=0,1 a μ=0,5. Výsledné aktivity skrytých a výstupných neurónov sú uvedené v tab. 5.4. Posledné tri stĺpce v tejto tabuľke odpovedajú výstupným aktivitám, ktoré sa priamo vzťahujú k požadovaným aktivitám (bitovým premenným) v sumácii (5.46). V prípade neurónovej siete 2. rádu aktivity skrytých neurónov už nemajú význam medzisumácií v (5.46).



Obrázok 5.15. Štruktúra neuróno-vej siete, ktorá je schopná riešiť sumáciu dvoch 2-bitových čísel.

Tabuľka 5.3. Hodnoty binárnych premenných sumácie dvoch 2-bitových čísel z (5.46)

Čís.	α_1	α_2	α_3	α_4	α_5	α_6	α_7	Význam
1	0	0	0	0	0	0	0	0+0=0
2	0	0	0	1	0	0	1	0+1=1
3	0	0	1	0	0	1	0	0+2=2
4	0	0	1	1	0	1	1	0+3=3
5	0	1	0	0	0	0	1	1+0=1
6	0	1	0	1	0	1	0	1+1=2
7	0	1	1	0	0	1	1	1+2=3
8	0	1	1	1	1	0	0	1+3=4
9	1	0	0	0	0	1	0	2+0=2
10	1	0	0	1	0	1	1	2+1=3
11	1	0	1	0	1	0	0	2+2=4
12	1	0	1	1	1	0	1	2+3=5
13	1	1	0	0	0	1	1	3+0=3
14	1	1	0	1	1	0	0	3+1=4
15	1	1	1	0	1	0	1	3+2=5
16	1	1	1	1	1	1	0	3+3=6

Tabuľka 5.4. Hodnoty aktivít skrytých a výstupných neurónov siete určenej obr. 5.15

Čís.	x_5	x_6	x_7	x_8	x_9
1	1,00	1,00	0,00	0,00	0,00
2	0,29	1,00	0,00	0,11	0,95
3	1,00	1,00	0,00	0,93	0,04
4	0,29	0,81	0,01	1,00	0,95
5	0,29	1,00	0,00	0,11	0,95
6	0,00	1,00	0,00	0,88	0,05
7	0,29	0,81	0,01	1,00	0,95
8	0,00	0,99	0,99	0,00	0,05
9	1,00	1,00	0,00	0,93	0,04
10	0,29	0,81	0,00	1,00	0,95
11	1,00	0,13	0,99	0,00	0,04
12	0,29	0,00	1,00	0,10	0,95
13	0,29	0,81	0,01	1,00	0,95
14	0,00	0,16	0,99	0,00	0,05
15	0,29	0,00	1,00	0,10	0,95
16	0,00	0,00	1,00	1,00	0,05

5.3.5 Adaptácia kombinácie lokálnych neurónových sietí

Účelová funkcia pre kombináciu *t* lokálnych sietí má tvar

$$E = \frac{1}{2}\sum_{i=1}^{t} p_i\left(x_O^{(i)} - \hat{x}_O\right)^2 = \frac{1}{2}\sum_{i=1}^{t}\sum_{k} g_k^{(i)2} \tag{5.47a}$$

$$g_k^{(i)} = \begin{cases} p_i\left(x_k^{(i)} - \hat{x}_k\right) & \left(\text{pre } k \in V_O\right) \\ 0 & \left(\text{pre } k \notin V_O\right) \end{cases} \tag{5.47b}$$

kde $x_k^{(i)}$ a $\hat{x}_k$ je vypočítaná resp. požadovaná výstupná aktivita k-teho neurónu i-tej lokálnej neurónovej siete. Tento vzťah pre účelovú funkciu je zovšeobecnením vzťahu (5.21), obsahuje príspevky od každej lokálnej neurónovej siete, pričom tieto sú vážené koeficientmi proporcionality p_i definovanými pomocou výstupných aktivít bránovej siete (pozri podkapitolu 5.2.2). Adaptácia kombinácie lokálnych neurónových sietí spočíva v hľadaní takých prahových a váhových koeficientov lokálnych sietí a bránovej siete, ktoré minimalizujú účelovú funkciu (5.47). Adaptačný proces pre lokálne siete je úplne analogický s adaptačným procesom obyčajnej neurónovej siete popísaným v podkapitole 5.3.3. Výrazy pre výpočet gradientu sú platné aj pre lokálnu sieť s malou modifikáciou, že výrazy g_i v (5.40) sú rozšírené o koeficienty proporcionality. Pre každú lokálnu neurónovú sieť spočítame zvlášť gradient účelovej funkcie a jej prahové a váhové koeficienty sú obnovené pomocou formuly (5.43).

Adaptačný proces kombinácie lokálnych sietí vzhľadom k prahovým a váhovým koeficientom bránovej siete je uskutočniteľný pomocou malej modifikácie adaptačného procesu lokálnych sietí. Parciálne derivácie účelovej funkcie vzhľadom k váhovým koeficientom bránovej siete sú určené vzťahom (pozri (5.37))

$$\frac{\partial E}{\partial w_{ij}^{(g)}} = \frac{\partial E}{\partial \vartheta_i^{(g)}} x_j^{(g)} \tag{5.48a}$$

Parciálne derivácie $\partial E / \partial \vartheta_i^{(g)}$ sú rekurentne určené vzťahom, ktorý má rovnakú formálnu štruktúru ako (5.40)

$$\frac{\partial E}{\partial \vartheta_i^{(g)}} = t'\left(\xi_i^{(g)}\right)\left(g_i^{(g)} + \sum_k \frac{\partial E}{\partial \vartheta_k^{(g)}} w_{ki}^{(g)}\right) \tag{5.48b}$$

kde veličiny $g_k^{(g)}$ určené vzťahom

$$g_k^{(g)} = \begin{cases} \left(\frac{1}{2}\left(\boldsymbol{x}_O^{(k)} - \hat{\boldsymbol{x}}_O\right)^2 - E\right)\left(\sum_j x_j^{(g)}\right)^{-1} & \left(\text{pre } k \in \mathsf{V}_O\right) \\ 0 & \left(\text{pre } k \notin \mathsf{V}_O\right) \end{cases} \tag{5.48c}$$

Jednoduchá diskusia vzťahov (5.48a-c) vedie k nasledujúcim dôležitým záverom: Ak lokálne siete a bránová sieť sú súčasne adaptované pre daný pár $\boldsymbol{x}_I / \hat{\boldsymbol{x}}_O$, potom kombinácia lokálnych sietí smeruje k použitiu len jednej lokálnej siete ku klasifikácii objektu popísaného vstupnými aktivitami $\boldsymbol{x}_I$, pričom výstupná aktivita danej lokálnej siete je blízka požadovanej klasifikácii určenej výstupným vektorom $\hat{\boldsymbol{x}}_O$. To znamená, že koeficienty proporcionality v priebehu adaptačného procesu sa upravia na "binárnu" hodnotu

$$p_j = \begin{cases} 1 & \left(\text{pre } j = i\right) \\ 0 & \left(\text{pre } j \neq i\right) \end{cases} \tag{5.49}$$

pre j=1,2,...,t, kde i je index lokálnej siete poskytujúcej výstupný vektor blízky požadovanému $\hat{\boldsymbol{x}}_O$. Bránová sieť teda rozhoduje tak, že si vyberie jednu lokálnu sieť, ktorá bude použitá pre klasifikáciu daného objektu. Ostatné lokálne siete v dôsledku malosti ich koeficientu proporcionality sa zúčastňujú na tejto klasifikácii zanedbateľnou mierou.

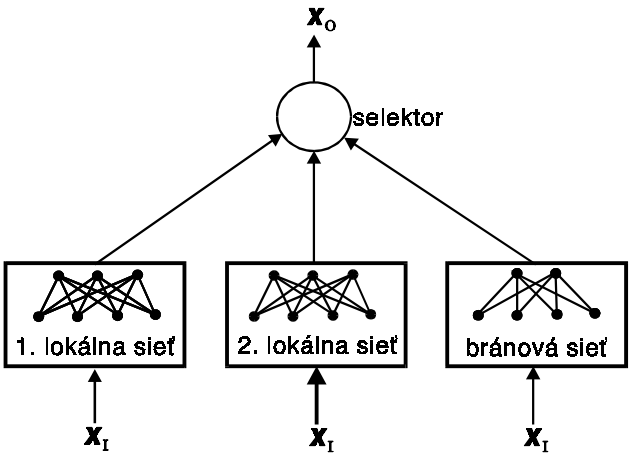
Účelová funkcia (5.47) pre aktuálny adaptačný proces je zovšeobecnená tak, že je sumovaná cez všetky objekty tréningovej množiny (pozri (5.23)). Gradient tejto účelovej funkcie sa potom rovná sume gradientov spočítaných pre jednotlivé objekty (pozri (5.41)).

Ilustračný príklad (sumácia dvoch 2-bitových čísel)

Efektívnosť teórie kombinácie lokálnych neurónových sietí ilustrujeme pomocou klasifikácie sumácie dvoch 2-bitových čísel; tento príklad už bol použitý v predchádzajúcej podkapitole 5.3.4. Ako už bolo zdôraznené, tento príklad obsahujúci 16 objektov je správne interpretovaný neurónovou sieťou, ktorá obsahuje aspoň dva skryté neuróny. Ako ilustračný príklad použijeme kombináciu dvoch lokálnych perceptrónov (neurónových sietí bez skrytých neurónov), ktoré obsahujú štyri vstupné neuróny a tri výstupné neuróny, a bránová sieť tiež predstavuje jednoduchý perceptrón obsahujúci štyri vstupné neuróny a dva (rovnaký počet ako lokálnych sietí) výstupné neuróny (pozri obr. 5.16).

Tréningová množina obsahuje všetkých 16 objektov uvedených v tab. 5.3. Adaptačný proces bol uskutočnený pomocou formúl (5.43) aplikovaných na prahové a váhové koeficienty jednotlivých lokálnych sietí a prahovej siete s parametrami λ=0,1 a μ=0,7. Adaptačný proces bol ukončený po 500 cykloch s hodnotou účelovej funkcie E=0,005. Všetkých 16 objektov bolo korektne klasifikovaných buď prvou alebo druhou lokálnou sieťou. To znamená, že bránová sieť rozdelila tréningovú množinu na dve podmnožiny,

ktoré už sú lineárne separovateľné, čiže korektne interpretovateľné lokálnymi sieťami — perceptrónmi. Dva ilustračné výsledky sú uvedené v tab. 5.5a-b. Tak napríklad, objekt č. 3 (odpovedajúci súčtu 0+2=2) je korektne klasifikovaný prvou lokálnou sieťou (s koeficientom proporcionality p_1=1). Druhá lokálna sieť poskytuje nesprávnu klasifikáciu, ale jej koeficient proporcionality je nulový, p_2=0.



Obrázok 5.16. Schematické znázornenie adaptívnej kombinácie dvoch lokálnych neurónových sietí a jednej bránovej neurónovej siete. Lokálne a bránové siete sú realizované ako dvojvrstvové neurónové siete bez skrytých neurónov (perceptróny).

Tabuľka 5.5a. Výstupné aktivity jednotlivých lokálnych sietí pre interpretáciu objektu č. 3 (pozri tab.5.3)

Objekt č. 3 (0+2=2)				
vstupné aktivity	0	0	1	0
požadované aktivity		0	1	0
vypočítané aktivity				
1. lokálna sieť		0,02	0,99	0,02
2. lokálna sieť		0,91	0,00	0,83
koef. proporcionality	1,00	0,00		

Tabuľka 5.5b. Výstupné aktivity jednotlivých lokálnych sietí pre interpretáciu objektu č. 14 (pozri tab. 5.3)

Objekt č. 14 (3+1=4)				
vstupné aktivity	1	1	0	1
požadované aktivity		1	0	0
vypočítané aktivity				
1. lokálna sieť		0,99	0,01	0,95
2. lokálna sieť		0,99	0,01	0,01
koef. proporcionality	0,00	1,00		

5.4 Neurónová sieť ako univerzálny aproximátor

V počiatkoch histórie neurónových sietí s dopredným šírením [2] bolo venované veľké úsilie tomu, aby sa ukázalo, že tieto neurónové siete s dostatočným počtom skrytých neurónov sú vždy schopné simulovať (aproximovať) zložité binárne alebo spojité funkcie s požadovanou presnosťou. Z pohľadu súčasnosti tieto snahy sú ľahko vysvetliteľné, jednalo sa o prekonanie šoku vyvolaného názorom Minského a Paperta [1], že perceptróny nemajú univerzálny výpočtový charakter. V predchádzajúcej časti tejto kapitoly sme ukázali na rôznych ilustračných príkladoch, že zovšeobecnenie perceptrónu zavedením skrytých neurónov alebo "interakcií" vyššieho rádu medzi neurónmi poskytuje dostatočne flexibilný výpočtový aparát, ktorý je schopný korektne simulovať rôzne zložité binárne funkcie. Hecht-Nielsen v r. 1987 prvý ukázal [13], že trojvrstvové neurónové siete s dopredným šírením a s dostatočným počtom skrytých neurónov sú schopné aproximovať s požadovanou presnosťou každé spojité zobrazenie. V súčasnosti k tomuto problému existuje už pomerne rozsiahla literatúra. Žiaľ, nejedná sa o jednoducho formulovateľný problém. Používajú sa pomerne zložité prostriedky funkcionálnej analýzy a preto sa obmedzíme len na formuláciu základných myšlienok tejto teórie [14].

Študujme spojitú funkciu F, ktorá zobrazuje n-rozmerný priestor R^n na otvorený interval $(0,1)$

$$F: R^n \rightarrow (0,1) \tag{5.50}$$

kde $y=F(\boldsymbol{x})=f(x_1,x_2,...,x_n)$. Tréningová množina A_{train} obsahuje r bodov (aktivít) z n-rozmerného priestoru R^n, $A_{train}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,...,\boldsymbol{x}_r\}$. Nech funkcia

$$t: R \rightarrow (0,1) \tag{5.51}$$

je tzv. *prechodová funkcia*, ktorá v tejto súvislosti je len veľmi všeobecne špecifikovaná ako spojitá a monotónne rastúca, vyhovujúca asymptotickým podmienkam $t(-\infty)=0$ a $t(\infty)=1$. Jednoduchá realizácia týchto všeobecných podmienok sa dá dosiahnuť použitím sigmoidy $t(x)=1/(1+e^{-x})$ (porovnaj s (5.11a) pre $A=0$ a $B=1$). Pretože sme predpokladali, že prechodová funkcia je monotónne rastúca, musí k nej existovať inverzná funkcia $t^{-1}:(0,1)\rightarrow R$. Pre sigmoidu je táto inverzná funkcia určená vzťahom $x=\ln(y/(1-y))$. Podľa Hornika [14] platí nasledujúca veta.

Veta. Pre každé $\varepsilon > 0$ existuje taká funkcia

$$G(\boldsymbol{x}) = \sum_{i=1}^{q} \alpha_i t(\vartheta_i + \boldsymbol{w}_i \cdot \boldsymbol{x}) \tag{5.52a}$$

kde α_i a ϑ_i sú reálne koeficienty, $\boldsymbol{w}_i = \left(w_1^{(i)}, w_2^{(i)}, ..., w_n^{(i)}\right)$ sú vektory obsahujúce n reálnych komponent a $\boldsymbol{x} \cdot \boldsymbol{w}_i = x_1 w_1^{(i)} + x_2 w_2^{(i)} + ... + x_n w_n^{(n)}$ je skalárny súčin vektorov $\boldsymbol{x}$ a $\boldsymbol{w}_i$, že

$$\sum_{k=1}^{r} \left| F(\boldsymbol{x}_k) - G(\boldsymbol{x}_k) \right| < \varepsilon \tag{5.52b}$$

Na základe tejto vety môžeme hovoriť, že funkcia $F(\boldsymbol{x})$ je aproximovaná s presnosťou ε nad tréningovou množinou A_{train} pomocou funkcie $G(\boldsymbol{x})$ určenej (5.52a) pomocou všeobecnej prechodovej funkcie $t(x)$ (realizovanej napr. sigmoidou). Žiaľ, táto veta je len existenčného charakteru, nešpecifikuje parametre funkcie G (napr. koeficienty α_i a ϑ_i a vektory $\boldsymbol{w}_i$), tvrdí len, že táto funkcia existuje a aproximuje funkciu $F(\boldsymbol{x})$ nad tréningovou množinou A_{train} .

Funkcia $G(\boldsymbol{x})$ je jednoducho interpretovateľná neurónovou sieťou s dopredným šírením, ktorá obsahuje jednu vrstvu q skrytých neurónov (pozri obr. 5.17). Koeficienty α_i sú váhy spojov medzi skrytými neurónmi a výstupným neurónom, ϑ_i sú prahové koeficienty skrytých neurónov a vektor $\boldsymbol{w}_i$ obsahuje zložky, ktoré tvoria váhové koeficienty hrán medzi i-tym skrytým neurónom a vstupnými neurónmi. Aktivita výstupného neurónu je v tomto prípade rovná potenciálu výstupného neurónu, zatiaľ čo v štandardnej neurónovej sieti aktivita výstupného neurónu je funkčná hodnota prechodovej funkcie pre jeho potenciál (pozri (5.10a)). Táto reštrikcia je ľahko odstrániteľná použitím predpokladu, že k prechodovej funkcii $y=t(x)$ existuje spojitá inverzná funkcia $x=t^{-1}(y)$. Potom podmienka (5.52b) má tvar

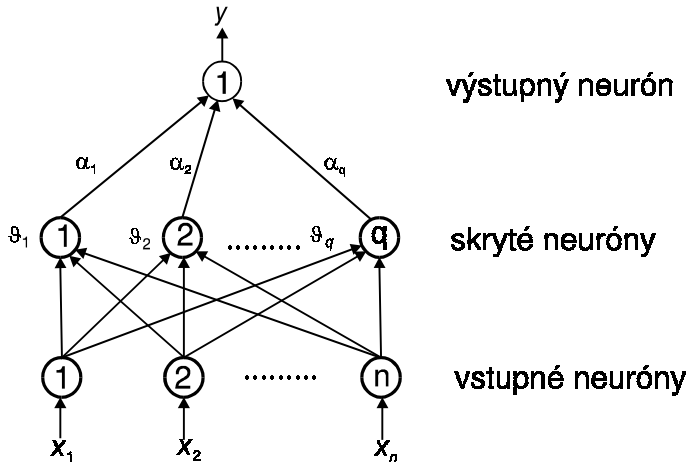
$$\sum_{k=1}^{r} \left| F(\boldsymbol{x}_k) - \tilde{G}(\boldsymbol{x}_k) \right| < \varepsilon' \tag{5.53}$$

kde nová funkcia $\tilde{G}(\boldsymbol{x})$ má tvar

$$\tilde{G}(\boldsymbol{x}) = t\left(\sum_{i=1}^{q} \alpha_i t(\vartheta_i + \boldsymbol{w}_i \cdot \boldsymbol{x})\right) \tag{5.54}$$

Touto jednoduchou modifikáciou vyššie uvedenej vety sme ukázali, že ľubovoľná spojitá funkcia $F(\boldsymbol{x})$, definovaná nad tréningovou množinou A_{train} a s funkčnými hodnotami z otvoreného intervalu (0,1), je aproximovateľná funkciou $\tilde{G}(\boldsymbol{x})$ s požadovanou presnosťou ε'. Čo je dôležité, funkcia $\tilde{G}(\boldsymbol{x})$ je už interpretovateľná neurónovou sieťou s dopredným šírením, ktorá obsahuje jednu vrstvu q skrytých neurónov.

Uvedená veta má principiálnu dôležitosť pre neurónové siete. Zabezpečuje nám, že trojvrstvová neurónová sieť (obsahujúca jednu vrstvu skrytých neurónov) je schopná simulovať s požadovanou presnosťou ľubovolné zobrazenie typu (5.50) definované nad konečnou tréningovou množinou. Týmto máme teda k dispozícii všeobecný prostriedok pre regresnú analýzu funkcií definovaných pomocou "regresnej tabuľky", kde pre nezávislé argumenty sú predpísané funkčné hodnoty (t.j. tréningová množina v zmysle úvodnej podkapitoly 5.1, pozri (5.3)). Teória neurónových sietí poskytuje univerzálny prostriedok pre návrh "modelovej funkcie" v tvare (5.54), kde počet skrytých neurónov a prahové a váhové koeficienty sú regresné parametre. Avšak musíme poznamenať, že hlavný cieľ teórie neurónových sietí s dopredným šírením nie je regresná analýza funkcií definovaných tréningovou množinou (aj keď tento moment v mnohých prípadoch je veľmi dôležitý), ale extrapolácia funkčných hodnôt mimo tréningovej množiny, čiže problém zovšeobecnenia (predikcia a klasifikácia).



Obrázok 5.17. Znázornenie štandardnej neurónovej siete s dopredným šírením, ktorá obsahuje jednu vrstvu skrytých neurónov. Aktivita výstupného neurónu y "simuluje" zobrazenie $F(x)$ pre požadované body z tréningovej množiny.

5.5 Praktické skúsenosti s aplikáciami neurónových sietí na klasifikáciu a predikciu

Viacvrstvové neurónové siete s dopredným šírením patria medzi tie neurónové siete, ktoré sú najčastejšie používané ako univerzálny prostriedok pre klasifikáciu a predikciu. Uvedieme niekoľko praktických skúseností, ako realizovať tieto aplikácie. Najprv budeme študovať problém, ako rozložiť množinu klasifikovaných objektov A na tréningovú a testovaciu množinu, $A=A_{train}\cup A_{test}$. Realizácia tohto rozkladu patrí medzi prvé základné problémy pri aplikáciách neurónových sietí, tréningová množina by mala obsahovať tie objekty z A, ktoré dobre "reprezentujú" ostatné podobné objekty (zahrnuté v testovacej množine A_{test}). Problém rovnakej dôležitosti ako rozklad množiny objektov na tréningovú a testovaciu množinu je aj problém výberu deskriptorov, ktoré sú podstatné pre klasifikáciu objektov. Obvykle sa deskriptory objektov navrhujú "ad-hoc" spôsobom — vyberajú sa také deskriptory, ktoré sú (alebo môžu byť) dôležité pre popis objektov. Z tohto pohľadu vystupuje do popredia problém výberu len tých deskriptorov, ktoré poskytujú rovnakú (alebo o málo horšiu) klasifikáciu objektov ako pôvodná sada deskriptorov. Tento problém budeme riešiť pomocou jednoduchej metódy využívajúcej najbližších susedov v okolí klasifikovaného objektu.

Ďalším dôležitým problémom pri aplikáciách neurónových sietí je navrhnúť vhodnú architektúru neurónovej siete. V našich úvahách sa pre jednoduchosť obmedzíme len na neurónové siete s jednou vrstvou skrytých neurónov (pozri obr. 5.17), pričom pod architektúrou budeme rozumieť počet skrytých neurónov. Na základe vety (uvedenej v podkapitole 5.4), ktorá charakterizuje 3-vrstvovú neurónovú sieť ako univerzálny aproximátor, môžeme očakávať, že s rastom počtu skrytých neurónov bude adaptačný proces neurónovej siete poskytovať lepšie a lepšie výsledky (t.j. hodnota účelovej funkcie (5.23) sa bude asymptoticky blížiť k nule). Tento záver je správny, avšak ak budeme porovnávať predikčné (alebo klasifikačné) schopnosti týchto neurónových sietí, spozorujeme, že od určitého počtu skrytých neurónov sa predikcia neurónovej siete pre objekty z testovacej množiny začne zhoršovať. To znamená, že z hľadiska správnej klasifikácie objektov z testovacej množiny je ďalšie zvyšovanie počtu skrytých neurónov už zbytočné (alebo až nežiaduce, z pohľadu adaptačného procesu neurónovej siete).

Podobný problém je aj s počtom iteračných krokov pri adaptácii neurónových sietí (pre daný počet skrytých neurónov). Ak súčasne sledujeme znižovanie účelovej funkcie (5.23) v priebehu adaptácie, od určitého počtu iteračných krokov spozorujeme, že predikčná schopnosť neurónovej siete sa začne zhoršovať. Podobne ako v predchádzajúcom prípade (zvyšovanie počtu skrytých neurónov), ďalšia adaptácia neurónovej siete je zbytočná, už len zhoršuje jej predikčnú schopnosť.

Problém optimálneho počtu adaptačných krokov úzko súvisí tiež s výberom adaptačnej (minimalizačnej) metódy. V podkapitole 5.3.3 bola diskutovaná jednoduchá modifikácia gradientovej metódy najprudšieho spádu (pozri (5.43)). Aj keď tento prístup je najčastejšie používaný pre adaptáciu viacvrstvových neurónových sietí s dopredným šírením, obvykle je kritizovaný ako veľmi pomalý, vyžadujúci mnoho tisíc iteračných krokov. Z týchto dôvodov sa venuje pozornosť aj iným, efektívnejším optimalizačným metódam numerickej matematiky (napr. Newtonova metóda, metóda združených gradientov alebo metóda premennej metriky, pozri [11]). Ich použitím v teórii neurónových sietí sa dosiahne podstatne rýchlejší proces adaptácie, avšak obvykle za cenu zhoršenia predikčných

schopností neurónovej siete. Obrazne povedané, neurónová sieť je "vynikajúco" adaptovaná na objekty tréningovej množiny (váhové a prahové koeficienty neurónovej siete majú hodnoty odpovedajúce presným hodnotám daného lokálneho minima účelovej funkcie (5.23)), avšak za cenu "preučenia" neurónovej siete s následnou slabou predikčnou schopnosťou.

Na záver tejto podkapitoly uvedieme jednoduchú algoritmizáciu v pascalovskom pseudokóde neurónovej siete s dopredným šírením, ktorá obsahuje jednu vrstvu skrytých neurónov.

5.5.1 Rozklad množiny objektov na tréningovú a testovaciu množinu

Popíšeme jednoduchý spôsob rozkladu množiny objektov A na tréningovú a testovaciu množinu, $A=A_{train}\cup A_{test}$. Predpokladajme, že poznáme nejakú klastrovaciu metódu [15], ktorá nám rozloží množinu A na disjunktné podmnožiny — klastre, ktoré obsahujú "podobné" objekty (z hľadiska metriky použitej v klastrovacej metóde)

$$A = C_1 \cup C_2 \cup ... \cup C_p \tag{5.55}$$

kde i-ty klaster C_i obsahuje n_i objektov z A

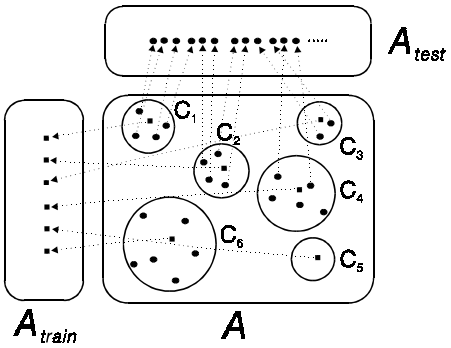
$$C_i = \left\{ o_1^{(i)}, o_2^{(i)}, ..., o_{n_i}^{(i)} \right\} \subset A \tag{5.56}$$

pričom predpokladáme, že objekt $o_1^{(i)} \in C_i$ je ten objekt z i-teho klastra C_i , ktorý leží "najbližšie" k jeho centru. Tento objekt nám v nasledujúcich úvahách bude slúžiť ako "reprezentant" objektov z klastra C_i. Potom tréningová a testovacia množina je určená objektmi

$$\begin{aligned} A_{train} &= \left\{ o_1^{(1)}, o_1^{(2)}, ..., o_1^{(p)} \right\} \\ A_{test} &= \left(C_1 - \left\{ o_1^{(1)} \right\} \right) \cup \left(C_2 - \left\{ o_1^{(2)} \right\} \right) \cup ... \cup \left(C_p - \left\{ o_1^{(p)} \right\} \right) \end{aligned} \tag{5.57}$$

To znamená, že tréningová množina je zložená zo všetkých reprezentantov klastrov a testovacia množina obsahuje zostávajúce objekty (pozri obr. 5.18)). Počet objektov v tréningovej množine je totožný s počtom klastrov, $|A_{train}|=p$ a $|A_{test}|= |A|-p$.

Teória neurónových sietí poskytuje výborný klastrovací prostriedok pomocou *Kohonenovej neurónovej siete* (pozri kapitolu 7), ktorý je ľahko použiteľný aj pre diskutovanú problematiku rozkladu množiny objektov na tréningovú a testovaciu množinu [16]. Obvykle sú výstupné neuróny tejto siete priestorovo uložené na ortogonálnej mriežke typu $N \times N$ (t.j. sieť obsahuje N^2 výstupných neurónov). Adaptačný proces tejto siete spočíva v tom, že objekty množiny A aktivujú len jeden výstupný neurón. Objekty, ktoré aktivujú rovnaký výstupný neurón, môžeme považovať za "podobné". Z teórie Kohonenových neurónových sietí tiež vyplýva, že tieto objekty môžeme tiež ešte podrobnejšie klasifikovať z hľadiska ich "blízkosti" k určitému centru daného výstupného neurónu (napr. minimálnosťou normy rozdielu deskriptorov objektu a váhových koeficientov výstupného neurónu). Tieto "centrálne" objekty nám slúžia ako reprezentanti objektov, ktoré aktivujú dané výstupné neuróny (klastre) a teda tvoria tréningovú množinu. V tejto súvislosti je potrebné bližšie špecifikovať tvar deskriptorov objektov, ktoré sa použijú pre "klastrovanie" množiny A na tréningovú a testovaciu množinu. Podobnosť objektov je v tomto prípade určená nielen ich deskriptormi ale aj ich vlastnosťami. Z týchto dôvodov, pre potreby klastrovania objektov pomocou Kohonenovej siete deskriptory objektov sú rozšírené ešte o vlastnosť (alebo vlastnosti) objektov. Jednoduchá realizácia tohto prístupu je znázornená na obr. 5.19.



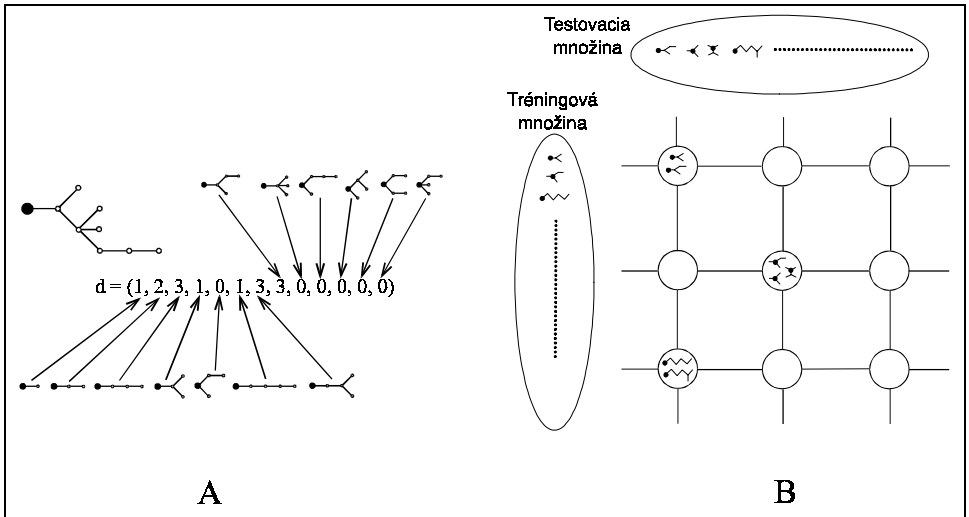
Obrázok 5.18. Schematické znázornenie rozkladu množiny objektov A na tréningovú a testovaciu množinu pomocou rozkladu A na klastre C_1, C_2, ..., C_p . Objekty reprezentované zaplnenými štvorcami odpovedajú tým objektom, ktoré ležia najbližšie k stredom príslušných klastrov, ostatné objekty sú reprezentované zaplnenými krúžkami. Poznamenajme, že klaster C_5 obsahuje len jeden objekt, ktorý patrí do tréningovej množiny.

5.5.2 Optimálny výber deskriptorov

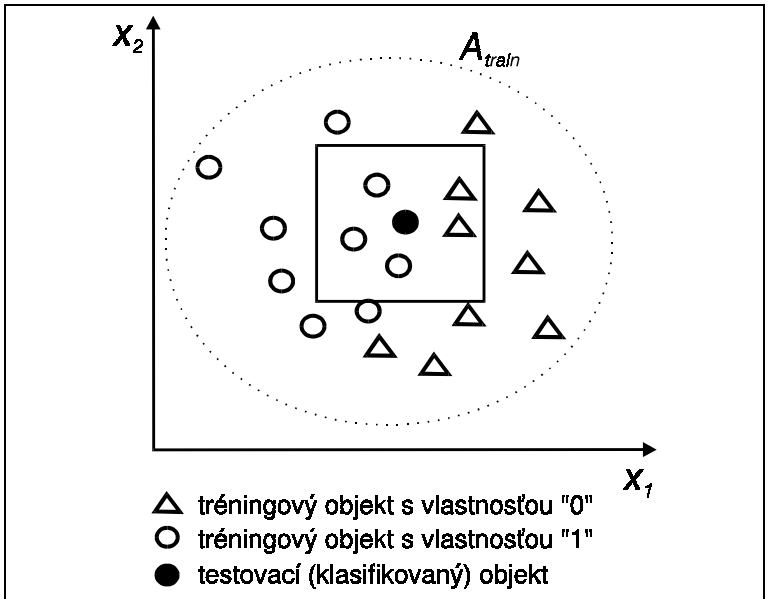
Výber deskriptorov — príznakov, ktoré popisujú vhodným spôsobom objekty, patrí medzi základné úlohy predspracovania dát pre potreby neurónových sietí. V mnohých prípadoch sú deskriptory navrhnuté "ad-hoc", bez podrobnejšieho štúdia ich vzájomnej závislosti a významnosti pre popis objektov. Z týchto dôvodov je dôležité analyzovať použité

deskriptory z pohľadu ich významnosti pre klasifikáciu objektov, z množiny navrhnutých deskriptorov určiť tie, ktoré sú významné pre klasifikáciu testovacích objektov. Pre aplikácie neurónových sietí ako klasifikátorov a prediktorov je určenie optimálneho výberu deskriptorov významné nielen z pohľadu efektívnosti adaptačného procesu, ale tiež aj ako významný medzikrok pre uľahčenie interpretácie výsledkov poskytovaných adaptovanou neurónovou sieťou.

V tejto podkapitole popíšeme klasifikačnú metódu KNN (angl. *K Nearest Neighbor*) [17], ktorej jednoduchá modifikácia je vhodná pre optimálny výber deskriptorov. Základný princíp metódy KNN je znázornený na obr. 5.20. V tejto metóde hrá základnú úlohu vzdialenosť medzi tréningovým objektom a testovacím objektom, táto vzdialenosť môže byť definovaná ako L_1 metrika



Obrázok 5.19. Ilustračný príklad použitia Kohonenovej siete na rozklad množiny objektov *A* na tréningovú a testovaciu množinu prevzatý z publikácie [16], v ktorej sa autori zaoberali použitím neurónových sietí k predikcii ^{13}C NMR (jadrová magnetická rezonancia) chemických posunov nasýtených acyklických uhlovodíkov — alkánov (napr. náplň v zapaľovačoch je tvorená zmesou dvoch alkánov, a to propánu a butánu). Chemický posun je lokálna vlastnosť uhlíkových atómov, preto daný uhlíkový atóm v alkáne je popísaný jeho najbližším okolím (diagram A), t.j. koľkokrát sa daný fragment vyskytuje v alkáne so špecifikovaným uhlíkom. Deskriptor ***d*** obsahuje 13 nezáporných celých čísel, ktoré odpovedajú "frekvenciám výskytu" fragmentov. Pre potreby klastrovania množiny alkánov z hľadiska ^{13}C NMR chemických posunov, tento deskriptor je rozšírený o experimentálne nameraný chemický posun. Takto modifikované deskriptory sú použité pre adaptáciu Kohonenovej neurónovej siete. Po skončení tohto procesu daný deskriptor (alkán) aktivuje práve jeden výstupný neurón. Rôzne alkány, ktoré aktivujú rovnaký výstupný neurón, sú považované za "klaster podobných" alkánov (B).



Obrázok 5.20. Schematické znázornenie metódy KNN. V rovine deskriptorov x_1 a x_2 sú umiestnené objekty tréningovej množiny. Okolo klasifikovaného objektu z testovacej množiny vytvoríme takú sféru (reprezentovanú v dôsledku použitia L_1 metriky štvorcom), aby obsahovala K objektov (v našom prípade K=5). Testovací objekt je klasifikovaný podľa tých tréningových objektov, ktoré sa najviac vyskytujú vo sfére, v tomto prípade má testovací objekt vlastnosť 1.

$$D(\boldsymbol{d}_{train}, \boldsymbol{d}_{test}) = \sum_{i=1}^{n} \left| d_{train}^{(i)} - d_{test}^{(i)} \right| \tag{5.58}$$

kde $\boldsymbol{d}_{train} = \left(d_{train}^{(1)}, d_{train}^{(2)}, \ldots, d_{train}^{(n)}\right)$ a $\boldsymbol{d}_{train} = \left(d_{test}^{(1)}, d_{test}^{(2)}, \ldots, d_{test}^{(n)}\right)$ sú vektory deskriptorov priradené objektu z tréningovej resp. testovacej množiny.
Predpokladajme, že tréningové objekty sú usporiadané tak, že pre testovací objekt s vektorom deskriptorov $\boldsymbol{d}_{train}$ platí

$$D\left(\boldsymbol{d}_{train,1}, \boldsymbol{d}_{test}\right) \le D\left(\boldsymbol{d}_{train,2}, \boldsymbol{d}_{test}\right) \le \ldots \le D\left(\boldsymbol{d}_{train,K}, \boldsymbol{d}_{test}\right) \le D\left(\boldsymbol{d}_{train,K+1}, \boldsymbol{d}_{test}\right) \le \ldots \tag{5.59}$$

K prvých tréningových objektov z tejto postupnosti tvorí okolie (K-rozmernú sféru) testovacieho objektu $\boldsymbol{d}_{test}$. Testovací objekt je klasifikovaný podľa tých objektov z K sféry, ktoré sa v nej najviac vyskytujú. Týmto spôsobom sme schopní klasifikovať každý objekt z testovacej množiny. Formálne, $y_{test} = KNN(\boldsymbol{d}_{test})$, kde y_{test} je vlastnosť priradená testovaciemu objektu s deskriptorom $\boldsymbol{d}_{test}$.

Výraz (5.58) pre vzdialenosť dvoch objektov možno zovšeobecniť tak, že sa zavedú binárne váhy $w_i \in \{0,1\}$, ktoré popisujú, či sa i-ty deskriptor uvažuje (w_i=1) alebo neuvažuje (w_i=0)

$$D(\boldsymbol{d}_{train}, \boldsymbol{d}_{test}) = \sum_{i=1}^{n} w_i \left| d_{train}^{(i)} - d_{test}^{(i)} \right| \tag{5.60}$$

Modifikovaný KNN klasifikátor s binárnymi váhami označíme $KNN_{\boldsymbol{w}}$, alebo $y_{test}^{(\boldsymbol{w})} = KNN_{\boldsymbol{w}}(\boldsymbol{d}_{test})$. Pre binárny váhový vektor $\boldsymbol{w}$ sú výsledky poskytované metódou KNN_w totožné s výsledkami poskytovanými KNN, ak všetky komponenty $\boldsymbol{w}$ sú jednotkové.

Úspešnosť klasifikátora $KNN_{\boldsymbol{w}}$ pri interpretácii objektov z testovacej množiny môže byť popísaná účelovou funkciou

$$f(\boldsymbol{w}) = \frac{1}{|A_{test}|} \sum_{\boldsymbol{d}_{test}} \delta\left(y_{test}^{(req)}, KNN_{\boldsymbol{w}}(\boldsymbol{d}_{test})\right) \tag{5.61}$$

kde $\delta(i,j)$=1 pre $i=j$, $\delta(i,j)$=0 pre $i \neq j$ a $y_{test}^{(req)}$ vyjadruje požadovanú vlastnosť. V prípade, že klasifikátor KNN_w interpretuje správne všetky objekty testovacej množiny, hodnota účelovej funkcie $f(\boldsymbol{w})$ je maximálna (jednotková), jej menšie hodnoty ($0 \leq f(\boldsymbol{w}) < 1$) indikujú, že klasifikátor KNN_w poskytuje nesprávnu interpretáciu. Výraz $1 - f(\boldsymbol{w})$ určuje frakciu objektov testovacej množiny, ktoré sú nesprávne interpretované. Optimálny výber deskriptorov je určený riešením diskrétneho optimalizačného problému

$$\boldsymbol{w}_{opt} = \arg \max_{\boldsymbol{w} \in \{0,1\}^n} f(\boldsymbol{w}) \tag{5.62}$$

kde hľadáme globálne minimum v priestore všetkých binárnych vektorov dĺžky n. Pre malé hodnoty n je problém (5.62) riešiteľný systematickým prehľadávaním celého priestoru riešení (dimenzia priestoru riešení je 2^n), napríklad metódou spätného prehľadávania [18], ktorá môže byť podstatne urýchlená metódou vetiev a hrán (angl. *branch and bound*). Pre väčšie hodnoty n (n>15) už nemožno riešiť optimalizačný problém (5.62) systematickými prehľadávacími algoritmami v dôsledku exponenciálneho rastu CPU času potrebného na riešenie problému. Z týchto dôvodov musíme obrátiť našu pozornosť na také metódy riešenia problému (5.62) ktoré, aj keď sú približné, poskytujú obvykle suboptimálne riešenia blízke optimálnym. V súčasnej informatike sú veľmi populárne tzv. evolučné algoritmy, založené na heuristikách prevzatých z biológie alebo z fyziky, ktoré poskytujú pomerne rýchlo suboptimálne riešenie zložitých optimalizačných problémov spojitého alebo diskrétneho charakteru (pozri kapitolu 9).

5.5.3 Architektúra neurónovej siete a počet adaptačných krokov

Návrh vhodnej architektúry (t.j. topológie grafu určujúceho neurónovú sieť) je zložitý a hlavne numericky náročný problém. Preto sa obmedzíme len na neurónové siete s jednou vrstvou skrytých neurónov (pozri obr. 5.17). Hlavným kritériom pre optimálny návrh

neurónovej siete bude optimálnosť jej klasifikačnej schopnosti, a realizácia tohto návrhu sa bude vykonávať súbežne s určením optimálneho počtu adaptačných krokov. Definujme si dve nasledujúce účelové funkcie (pozri (5.23))

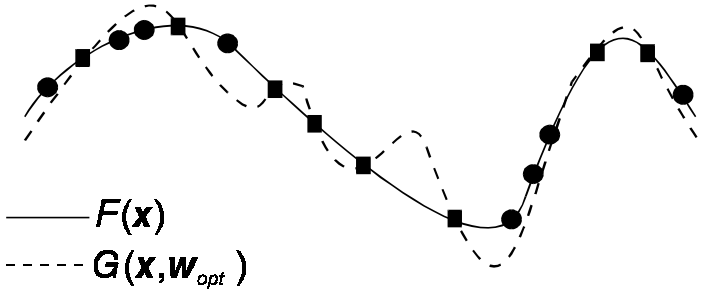
$$E_{train} = \frac{1}{2}\sum_{i}^{A_{train}} \left(G(\boldsymbol{x}_i, \boldsymbol{w}) - \hat{\boldsymbol{x}}_i\right)^2$$
$$E_{test} = \frac{1}{2}\sum_{i}^{A_{test}} \left(G(\boldsymbol{x}_i, \boldsymbol{w}) - \hat{\boldsymbol{x}}_i\right)^2 \tag{5.63}$$

kde E_{train} (E_{test}) je účelová funkcia definovaná pre objekty z tréningovej (testovacej) množiny pre dané hodnoty váhových a prahových koeficientov $\boldsymbol{w}$. Na základe vety o neurónovej sieti ako univerzálnom aproximátore (pozri podkapitolu 5.4) vieme, že pre rastúci počet skrytých neurónov účelová funkcia E_{train} (s adaptovanými váhovými a prahovými koeficientmi) konverguje k nule

$$\lim_{q\to\infty} E_{train} = 0 \tag{5.64}$$

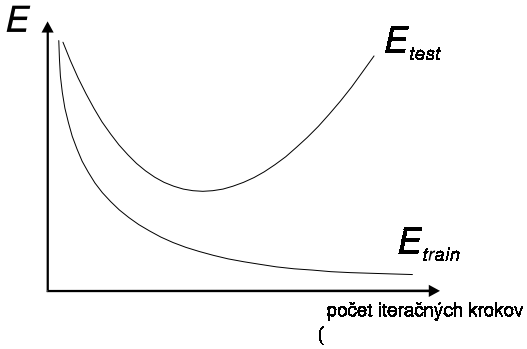
kde q je počet skrytých neurónov. Ako môžeme interpretovať tento dôležitý výsledok teórie neurónových sietí s dopredným šírením a s jednou vrstvou skrytých neurónov? Interpretácia je veľmi podobná analogickej situácii v regresnej analýze s polynomiálnou modelovou funkciou. Zvyšovaním rádu polynómu dostávame stále menšiu a menšiu hodnotu minimalizovanej účelovej funkcie. Flexibilita polynómu rastie s jeho stupňom (pozri obr. 5.21).

Podobný obrázok by sme dostali aj pri štúdiu závislosti schopnosti korektne klasifikovať objekty z testovacej množiny od počtu iteračných krokov pre neurónovú sieť s daným počtom skrytých neurónov. Hodnota účelovej funkcie E_{train} bude klesať s rastom počtu iteračných krokov. Žiaľ, hodnota účelovej funkcie E_{test} bude od určitého počtu iteračných krokov rásť, t.j. zhoršuje sa predikčná schopnosť neurónovej siete s pokračovaním adaptácie neurónovej siete (hovoríme, že neurónová sieť je preučená, pozri obr. 5.22).

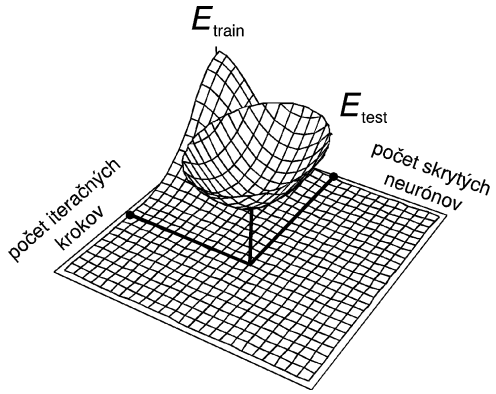


Obrázok 5.21. Schematické znázornenie simulácie "presnej" funkcie *F*, ktorá interpretuje objekty tak z tréningovej (plné štvorce), ako aj z testovacej množiny (plné krúžky) (pozri podkapitolu 5.1). Funkcia *G* reprezentuje adaptovanú neurónovú sieť s dostatočným počtom skrytých neurónov, ktorej adaptácia bola vykonaná vzhľadom k objektom z tréningovej množiny. Funkcia *G* nemusí už presne postihovať objekty z testovacej množiny. Počet "inflexných bodov" funkcie *G* rastie s počtom skrytých neurónov (analógia s polynomiálnou regresiou). Môžeme teda očakávať, že interpretácia objektov z testovacej množiny sa bude zhoršovať s rastom počtu skrytých neurónov.

Z vyššie uvedených úvah vyplýva, že stanovenie optimálneho počtu skrytých neurónov a počtu iteračných krokov vzhľadom pre dané rozdelenie objektov na tréningovú a testovaciu množinu môže byť realizované súčasne. Pre daný počet skrytých neurónov nájdeme optimálny počet iteračných krokov adaptačného procesu. Tento prístup je založený na poznatku, že zatiaľ čo tréningová účelová funkcia E_{train} klesá s rastúcim počtom skrytých neurónov a/alebo rastúcim počtom iteračných krokov, testovacia účelová funkcia E_{test} vykazuje minimum pre určitý počet skrytých neurónov a počet iteračných krokov (pozri obr. 5.23). Tieto hodnoty, v ktorých má E_{test} minimum, sú optimálne pre použitie 3-vrstvovej neurónovej siete pre klasifikáciu objektov z testovacej množiny A_{test}.



Obrázok 5.22. Priebeh účelových funkcií pre tréningovú a testovaciu množinu v závislosti na počte iteračných krokov adaptácie (resp. na počte skrytých neurónov) neurónovej siete. Účelová funkcia E_{train} klesá s rastom počtu iteračných krokov (resp. skrytých neurónov), zatiaľ čo účelová funkcia E_{test} vykazuje minimum.



Obrázok 5.23. Schematické znázornenie priebehu účelových funkcií E_{train} a E_{test} vzhľadom k počtu skrytých neurónov a počtu iteračných krokov. Na rozdiel od účelovej funkcie E_{train}, ktorá monotónne klesá s rastúcim počtom skrytých neurónov a s rastúcim počtom iteračných krokov, účelová funkcia E_{test} má minimum pre optimálny počet skrytých neurónov a iteračných krokov (pozri hrubé čiary na grafe).

5.5.4 Algoritmizácia neurónovej siete s dopredným šírením

Účelom tejto podkapitoly je naznačiť základné princípy algoritmizácie neurónových sietí s dopredným šírením, ktoré obsahujú skryté neuróny. Pre jednoduchosť budeme uvažovať 3-vrstvovú neurónovú sieť, ktorá obsahuje jednu vrstvu skrytých neurónov, pričom neuróny zo susedných vrstiev sú prepojené všetkými možnými spôsobmi, pozri obr. 5.24 .

Aktivity skrytých a výstupných neurónov sú určené vzťahmi (pozri vzťahy (5.10a-b))

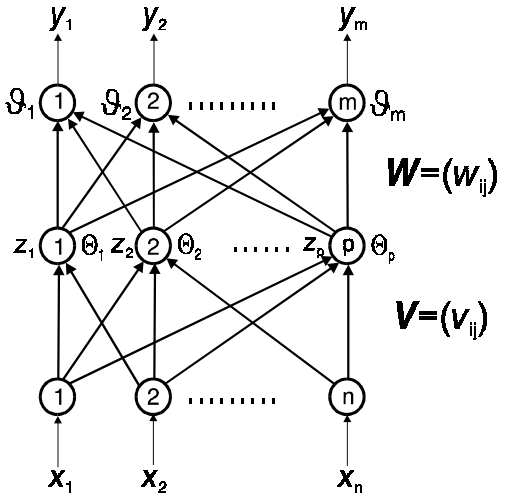
$$z_i = t\left(\sum_{j=1}^{n} v_{ij} x_j + \Theta_i\right) \qquad (\text{pre } i = 1,2,\ldots,p) \tag{5.65a}$$

$$y_i = t\left(\sum_{j=1}^{p} w_{ij} z_j + \vartheta_i\right) \qquad (\text{pre } i = 1,2,\ldots,m) \tag{5.65a}$$

kde $t(\xi)$ je sigmoida určená pomocou (5.11a-b) (s parametrami A=0 a B=1)

$$t(\xi) = \frac{1}{1+e^{-\xi}} \tag{5.66}$$

Grafický priebeh tejto funkcie je znázornený na obr. 5.5, graf 1.



Obrázok 5.24. Trojvrstvová neurónová sieť s dopredným šírením. Vstupná vrstva obsahuje n vstupných neurónov s aktivitami x_1, x_2, ..., x_n. Skrytá vrstva obsahuje p skrytých neurónov, ich aktivity resp. prahové koeficienty sú označené z_1, z_2, ..., z_p resp. Θ_1, Θ_2, ..., Θ_p . Výstupná vrstva obsahuje m výstupných neurónov, ich aktivity resp. prahové koeficienty sú označené y_1, y_2, ..., y_m resp. ϑ_1, ϑ_2, ..., ϑ_m. Váhové koeficienty medzi vstupnou a skrytou vrstvou resp. skrytou a výstupnou vrstvou tvoria maticu $\boldsymbol{V}=(v_{ij})$ resp. maticu $\boldsymbol{W}=(w_{ij})$.

Výpočet aktivít neurónov pre dané váhové a prahové koeficienty sa nazýva aktívna fáza neurónovej siete. Tieto aktivity pre danú neurónovú sieť sa vypočítajú jednoduchým rekurentným postupom: Predpokladajme, že vstupné aktivity x_1, x_2, ..., x_n (deskriptory klasifikovaného objektu) sú známe, potom pomocou (5.65a) zostrojíme aktivity skrytých neurónov z_1, z_2, ..., z_p. Následne, pomocou (5.65b) zostrojíme aktivity výstupných neurónov y_1, y_2, ..., y_m. Uvedený rekurentný spôsob výpočtu aktivít postupuje zdola nahor neurónovou sieťou. Táto skutočnosť sa odráža v názve týchto sietí, ako neurónových sietí s dopredným šírením signálu. Algoritmizácia tohto postupu je uvedená formou pascalovského pseudokódu na obr. 5.25.

```
procedure activities(input :Θ,v,ϑ,w,x;
                         output:z,y);
begin for i:=1 to p do
      begin ξ:=Θ[i];
            for j:=1 to n do ξ:=ξ+v[i,j]*x[j];
            z[i]:=t(ξ);
      end;
      for i:=1 to m do
      begin ξ:=ϑ[i];
            for j:=1 to p do ξ:=ξ+w[i,j]*z[j];
            y[i]:=t(ξ);
      end;
end;
```

Obrázok 5.25. Algoritmizácia v pascalovskom pseudokóde aktívnej fáze neurónovej siete s dopredným šírením, ktorá obsahuje jednu vrstvu skrytých neurónov. Vstupnými parametrami procedúry activities sú vstupné aktivity a váhové a prahové koeficienty, výstupnými parametrami sú skryté a výstupné aktivity. Reálna funkcia t(ξ) je prechodová funkcia definovaná (5.66).

Teraz upriamime našu pozornosť na tzv. adaptačnú fázu neurónovej siete, v ktorej sú upravované (pozri (5.43)) váhové a prahové koeficienty pomocou gradientu účelovej funkcie *E* definovanej (5.41). V prvom kroku budeme študovať konštrukciu gradientu účelovej funkcie (5.21), ktorá je priradená len jednému objektu z tréningovej množiny. Komponenty gradientu sú určené formulami (5.37) a (5.40), ktoré sa jednoducho prepíšu zvlášť pre výstupné neuróny

$$\frac{\partial E}{\partial \vartheta_i} = y_i(1-y_i)\left(y_i - y_{i,req}\right) \qquad \left(\text{pre } i = 1,2,\ldots,m\right) \tag{5.67a}$$

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial \vartheta_i} z_j \qquad \begin{pmatrix} \text{pre } i = 1,2,\ldots,m \\ j = 1,2,\ldots p \end{pmatrix} \tag{5.67b}$$

a zvlášť pre skryté neuróny

$$\frac{\partial E}{\partial \Theta_i} = z_i(1-z_i)\sum_{j=1}^{m}\frac{\partial E}{\partial \vartheta_j} w_{ji} \qquad \left(\text{pre } i = 1,2,\ldots,p\right) \tag{5.68a}$$

$$\frac{\partial E}{\partial v_{ij}} = \frac{\partial E}{\partial \Theta_i} x_j \qquad \begin{pmatrix} \text{pre } i = 1,2,\ldots,p \\ j = 1,2,\ldots n \end{pmatrix} \tag{5.68b}$$

V týchto formulách derivácia prechodovej funkcie $t'(\xi)$ je určená jednoduchým vzťahom $t'(\xi) = t(\xi)(1-t(\xi))$ (pozri vzťah (5.11b)). Formuly (5.67-68) pre výpočet parciálnych derivácií možno realizovať rekurentne tak, že sa postupuje cez neurónovú sieť zhora nadol (t.j. v opačnom smere ako pri výpočte aktivít neurónovej siete). V prvom kroku sa vypočítajú parciálne derivácie účelovej funkcie vzhľadom k prahovým koeficientom výstupných neurónov pomocou (5.67a). Potom sa jednoducho vypočítajú aj parciálne derivácie účelovej funkcie vzhľadom k váhovým koeficientom spojov medzi skrytými a výstupnými neurónmi pomocou (5.67b). Poznajúc túto časť gradientu účelovej funkcie, môžeme pristúpiť k výpočtu tej jeho časti, ktorá odpovedá skrytým neurónom. Podobne ako v predchádzajúcom prípade, pomocou (5.68a) vypočítame parciálne derivácie účelovej funkcie vzhľadom k prahovým koeficientom skrytých neurónov, a potom pomocou (5.68b) vypočítame parciálne derivácie účelovej funkcie vzhľadom k váhovým koeficientom medzi vstupnými a skrytými neurónmi. Algoritmizácia tohto postupu (metódy spätného šírenia) je znázornená v pascalovskom pseudokóde na obr. 5.26.

```
procedure gradient(input :Θ,V,ϑ,W,x,y_req;
                   output:grad_ϑ,grad_w,grad_Θ,grad_v);
begin activities(Θ,V,ϑ,W,x,z,y);
      for i:=1 to m do
      grad_ϑ[i]:=y[i]*(1-y[i])*(y[i]-y_req[i]);
      for i:=1 to m do
      for j:=1 to p do
            grad_w[i,j]:=grad_ϑ[i]*z[j];
      for i:=1 to p do
      begin aux:=0;
            for j:=1 to m do
            aux:=aux+grad_ϑ[j]*w[j,i]
            grad_Θ[i]:=z[i]*(1-z[i])*aux;
      end;
      for i:=1 to p do
      for j:=1 to n do
            grad_v[i,j]:=grad_Θ[i]*x[j];
end;
```

Obrázok 5.26. Výpočet gradientu účelovej funkcie pre dané vektory vstupných aktivít $\boldsymbol{x}$ požadovaných výstupných aktivít $\boldsymbol{y}_{req}$. Proces je inicializovaný výpočtom skrytých a výstupných aktivít pomocou procedúry activities. Prahové a váhové koeficienty sú vstupnými parametrami procedúry. Vypočítaný gradient je výstupným parametrom procedúry.

Parciálne derivácie účelovej funkcie (5.21), ktorá je definovaná nad celou tréningovou množinou, sú určené vzťahom (5.41), t.j. celkový gradient sa rovná sume gradientov pre jednotlivé elementy tréningovej množiny. Pascalovský pseudokód výpočtu celkového gradientu účelovej funkcie je znázornený na obr. 5.27.

```
procedure gradient_total(input :Θ,V,ϑ,W,A_train;
                         output:
                          grad_total_ϑ,, grad_total_w,
                          grad_total_Θ,grad_total_v);
begin for i:=1 to m do grad_total_ϑ[i]:=0;
      for i:=1 to m do
      for j:=1 to p do grad_total_w[i,j]:=0;
      for i:=1 to p do grad_total_Θ[i]:=0;
      for i:=1 to p do
      for j:=1 to n do grad_total_v[i,j]:=0;

      for each pair x/y_req of A_train do
      begin gradient(Θ,V,ϑ,W,x,y_req;
                     grad_ϑ,grad_w,grad_Θ,grad_v);
            for i:=1 to m do
            grad_total_ϑ[i]:=grad_total_ϑ[i]+grad_ϑ[i];
            for i:=1 to m do
            for j:=1 to p do
            grad_total_w[i,j]:=grad_total_w[i,j]
                               +grad_w[i,j];
            for i:=1 to p do
            grad_total_Θ[i]:=grad_total_Θ[i]
                             +grad_Θ[i];
            for i:=1 to p do
            for j:=1 to n do
            grad_total_v[i,j]:=grad_total_v[i,j]
                               +grad_v[i,j];
      end;
end;
```

Obrázok 5.27. Výpočet celkového gradientu účelovej funkcie pre celú tréningovú množinu. Algoritmus je inicializovaný vynulovaním jednotlivých zložiek celkového gradientu. Vlastný výpočet je vnorený do vonkajšieho for-cyklu, ktorý sa opakuje pre všetky páry $\boldsymbol{x}/\boldsymbol{y}_{\text{req}}$ tréningovej množiny A_{train}.

Na záver našich úvah o algoritmizácii neurónových sietí s dopredným šírením, pristúpime k ich adaptačnej fáze, ktorá spočíva v iteračnej úprave prahových a váhových

koeficientov tak, aby účelová funkcia (5.21) bola minimálna (prahové a váhové koeficienty sú určené ako riešenie minimalizačného problému (5.42)). Gradientová minimalizačná metóda najprudšieho spádu je vyjadrená vzťahmi (5.43), ich jednoduchou modifikáciou pre neurónovú sieť s troma vrstvami dostaneme tieto vzťahy

$$\begin{aligned} w_{ij}^{(k+1)} &= w_{ij}^{(k)} - \lambda \frac{\partial E}{\partial w_{ij}} + \mu \Delta w_{ij}^{(k)} \\ \vartheta_i^{(k+1)} &= \vartheta_i^{(k)} - \lambda \frac{\partial E}{\partial \vartheta_i} + \mu \Delta \vartheta_i^{(k)} \end{aligned} \tag{5.69}$$

pre i=1,2,...,m a j=1,2,...,p

$$\begin{aligned} v_{ij}^{(k+1)} &= v_{ij}^{(k)} - \lambda \frac{\partial E}{\partial v_{ij}} + \mu \Delta v_{ij}^{(k)} \\ \Theta_i^{(k+1)} &= \Theta_i^{(k)} - \lambda \frac{\partial E}{\partial \Theta_i} + \mu \Delta \Theta_i^{(k)} \end{aligned} \tag{5.70}$$

pre i=1,2,...,p a j=1,2,...,n, index k popisuje iteračný krok. Symboly Δ sú určené ako rozdiel koeficientov z predchádzajúcich dvoch krokov, tak napr. $\Delta w_{ij}^{(k)} = w_{ij}^{(k)} - w_{ij}^{(k-1)}$. Adaptačný proces je inicializovaný náhodne generovanými prahovými a váhovými koeficientmi, napr. z intervalu (–1,1). Ako je obvyklé v gradientových optimalizačných metódach, adaptačný proces je ukončený, keď hodnota celkového gradientu je menšia ako predpísané malé kladné číslo ε, $|\text{grad } E_{\text{tot}}| < \varepsilon$. Iná alternatíva ukončenia adaptačného procesu je, keď počet iterácií k dosiahne predpísaný počet $k_{\max}$. Pacalovský pseudokód adaptačného procesu je znázornený na obr. 5.28.

Principiálnu dôležitosť v adaptačnom procese neurónovej siete hrá rýchlosť učenia (parameter λ). Tento parameter sa obvykle položí rovný malému kladnému číslu, napr. λ=0,1). V mnohých prípadoch je vhodné tento parameter dynamicky meniť v závislosti na rýchlosti adaptácie neurónovej siete. V prípade, že hodnota účelovej funkcie sa zväčší, potom je potrebné parameter zmenšiť, napr. $\lambda \leftarrow \lambda/10$. V opačnom prípade, ak sa účelová funkcia monotónne zmenšuje, je vhodné zväčšiť parameter λ, napr. $\lambda \leftarrow 2\lambda$. Týmto jednoduchým spôsobom máme zabezpečenú približne optimálnu hodnotu rýchlosti učenia λ.

```
procedure adaptation(input :A_train,k_max,ε,λ,μ;
                     output:Θ,V,ϑ,W);
begin for i:=1 to m do
      begin ϑ[i]:=2*random-1; Δϑ[i]:=0 end;
      for i:=1 to m do
      for j:=1 to p do
      begin w[i,j]:=2*random-1; Δw[i,j]:=0 end;
      for i:=1 to p do
      begin Θ[i]:=2*random-1; ΔΘ[i]:=0 end;
      for i:=1 to p do
      for j:=1 to n do
      begin v[i,j]:=2*random-1; Δv[i,j]:=0 end;
      k:=0; length_grad_E=∞;
      while (k<k_max) and (length_grad_E>ε) do
      begin gradient_total(Θ,V,ϑ,W,A_train;
                           grad_total_ϑ,grad_total_w,
                           grad_total_Θ,grad_total_v);
            for i:=1 to m do
            begin Δ:=-λ*grad_total_ϑ[i]+μ*Δϑ[i];
                  ϑ[i]:=ϑ[i]+Δ; Δϑ[i]:=Δ
            end;
            for i:=1 to m do
            for j:=1 to p do
            begin Δ:=-λ*grad_total_w[i,j]+μ*Δw[i,j];
                  w[i,j]:=w[i,j]+Δ; Δw[i,j]:=Δ
            end;
            for i:=1 to p do
            begin Δ:=-λ*grad_total_Θ[i]+μ*ΔΘ[i];
                  Θ[i]:=Θ[i]+Δ; ΔΘ[i]:=Δ
            end;
            for i:=1 to p do
            for j:=1 to n do
            begin Δ:=-λ*grad_total_v[i,j]+μ*Δv[i,j];
                  v[i,j]:=v[i,j]+Δ; Δv[i,j]:=Δ
            end;
            length_grad_E:=|grad_total_ϑ|+|grad_total_w|+
                           |grad_total_Θ|+|grad_total_v|;
      end;
end;
```

Obrázok 5.28. Algoritmizácia adaptačného procesu neurónovej siete. Procedúra je inicializovaná vynulovaním momentových Δ-členov a náhodnou generáciou prahových a váhových koeficientov z intervalu $(-1,1)$ (premenná random je generátor náhodných čísel s rovnomernou distribúciou z intervalu $\langle 0,1)$). Vonkajší while-cyklus sa opakuje tak dlho, až buď počet iterácií k je väčší ako predpísaný počet k_{max} alebo norma (dĺžka) gradientu length_grad_E je menšia ako požadovaná presnosť ε. Premenné λ resp. μ označujú rýchlosť učenia (malé kladné číslo, napr. λ=0,1) resp. momentový člen (obvykle μ=0,5-0,7). Výstupné parametre procedúry sú prahové a váhové koeficienty adaptovanej neurónovej siete.

Literatúra

[1] M. Minsky and S. Papert. *Perceptrons. An Introduction to Computational Geometry*. The MIT Press, Cambridge, MA, 1969.

[2] D.E. Rumelhart, G.E. Hinton, and R.J. Wiliams. Learning internal representation by error propagation. In: D.E. Rumelhart, J.L. McClelland, and PDP Research Group. *Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Vol 1: Foundation.* The MIT Press, Cambridge, MA, 1987, pp. 318-362.

[3] J. Sedláček. *Úvod do teorie grafů*. Academia, Praha, 1981.

[4] M. Novák. *Neuronové sítě a neuropočítače*. Edice Výber, SENZO a.s., Praha, 1992.

[5] M. Novák, J. Faber a O. Kufudaki. *Neuronové sítě a informační systémy živých organizmů*. Grada, Praha, 1992.

[6] C.L. Giles and T. Maxwell. Learning, invariance, and generalization in high-order neural networks. *Applied Optics* 26:4972-4978, 1987.

[7] R.A. Jacobs, M.I. Jordan, S.J. Nowlan, and G.E. Hinton. Adaptive mixture of local experts. *Neural Computation* 3:79-87,1991.

[8] V. Kvasnička. Adaptive mixture of local neural networks. *Neural Network World* 3:161-174, 1993.

[9] M. Demlová a J. Nagy. *Algebra*. Edice Matematika pro vysoké školy technické, sešit III. SNTL, Praha, 1982.

[10] S. Mika. *Numerické metody algebry*. Edice Matematika pro vysoké školy technické, sešit IV. SNTL, Praha, 1982.

[11] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Wetterling. *Numerical Recipes in Pascal. The Art of Scientific Computing.* Cambridge University Press, Cambridge, UK, 1992.

[12] I. Kluvánek, L. Mišík a M. Švec. *Matematika I*, Alfa, Bratislava, 1971.

[13] R. Hecht-Nielsen. Kolmogorov's mapping neural network existence theorem. *1st IEEE International Conference on Neural Networks*, San Diego, CA, Vol. 3, pp. 11-14, 1987.

[14] K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural Networks* 2:259-366, 1989.

[15] A. Lukasová a J. Šormanová. *Metody shlukové analýzy*. SNTL, Praha, 1985.

[16] D. Svozil, V. Kvasnička, and J. Pospíchal. Neural network prediction of carbon-13 NMR chemical shifts of alkanes. *Journal of Chemical Information and Computer Sciences* 35:924-928, 1995.

[17] P.A. Devijver and J. Kittler. *Pattern Recognition: A Statistical Approach*. Prentice Hall, London, 1982.

[18] L. Kučera. *Kombinatorické algoritmy*. SNTL, Praha, 1983.