

4. Lineárne modely neurónových sietí

V tejto kapitole sa budeme zaoberať najjednoduchšími modelmi neurónových sietí — modelmi ktorých základnou stavebnou jednotkou je formálny neurón s lineárnou aktivačnou funkciou f . Označme výstup formálneho neurónu ako y . Nech aktivity prichádzajúce cez n vstupných kanálov tvoria vektor $\mathbf{x}=(x_1, x_2, \dots, x_n)^T$, kde horný index T označuje transponovaný vektor. Nech sú vstupné kanály váhované pomocou vektora váh $\mathbf{w}=(w_1, w_2, \dots, w_n)^T$. Potom bude výstupná aktivita lineárneho neurónu rovná

$$y = f(\text{net}) = f(\mathbf{w}^T \mathbf{x} + \vartheta), \quad (4.1)$$

kde ϑ je prah aktivácie (excitácie). Najčastejšie je aktivačná funkcia identické zobrazenie a formálny neurón realizuje skalárny súčin vstupného vektora $\mathbf{x}=(x_1, x_2, \dots, x_n)^T$ s váhovým vektorom $\mathbf{w}=(w_1, w_2, \dots, w_n)^T$:

$$y = \mathbf{w}^T \mathbf{x}. \quad (4.2)$$

Samozrejme, takýto model nezohľadňuje saturačné vlastnosti reálnych neurónov (aktivita neurónu sa nemôže zvyšovať priamo úmerne s nárastom postsynaptického potenciálu $\mathbf{w}^T \mathbf{x}$, ale pre veľké hodnoty sa saturuje). Na druhej strane, lineárna povaha modelu nám umožní nájsť analytické riešenia problémov, ktoré v nelineárnych modeloch zohľadňujúcich saturácie neurónov nie je možné získať. Interpretáciou výsledkov aparátom lineárnej algebry možno presne pochopiť, čo sa skrýva za pojmi ako je zovšeobecnenie na základe tréningovej množiny, mechanizmus asociácie vstupov so “zapamätanými” prototypovými vzormi, atď.

Označme váhu j -teho vstupného kanála do i -teho neurónu w_{ij} . Ak neurónová sieť pozostáva z m neurónov reagujúcich na n vstupných kanálov (obr. 4.1), potom váhy všetkých prepojení možno prehľadne usporiadať do váhovej matice

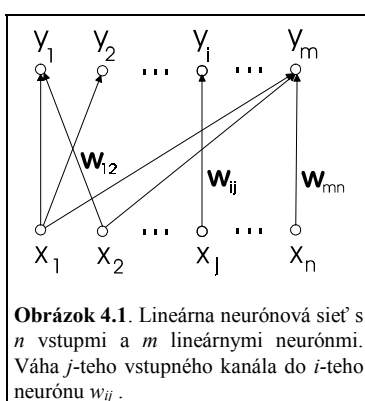
$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2n} \\ \dots & \dots & \dots & \dots \\ w_{m1} & w_{m2} & \dots & w_{mn} \end{bmatrix}. \quad (4.3)$$

Po priložení vektora \mathbf{x} na vstup siete dostávame ako výstup vektor $\mathbf{y}=(y_1, y_2, \dots, y_m)^T$, kde y_1, y_2, \dots, y_m sú aktivácie výstupných neurónov, pričom $\mathbf{y}=\mathbf{W}\mathbf{x}$. Jednovrstvová lineárna neurónová sieť teda realizuje lineárne zobrazenie $\varphi: R^n \rightarrow R^m$. Ak by sme pridali ďalšiu vrstvu, ktorej vstupy by boli aktivácie \mathbf{y} , výstup siete \mathbf{z} by bol lineárny obraz vektora \mathbf{y} , $\mathbf{z}=\mathbf{V}\mathbf{y}$. Zrejme $\mathbf{z}=\mathbf{A}\mathbf{x}$, kde matica $\mathbf{A}=\mathbf{V}\mathbf{W}$. Keďže kompozícia lineárnych zobrazení je opäť lineárne zobrazenie, pridávanie ďalších vrstiev v lineárnych sieťach neprináša žiadny nový kvalitatívny efekt, a preto sa v ďalšom budeme zaoberať iba jednovrstvovými sieťami.

Majme trénovaciu množinu $T = \{(\mathbf{x}^{(1)}, \mathbf{y}^{(1)}), (\mathbf{x}^{(2)}, \mathbf{y}^{(2)}), \dots, (\mathbf{x}^{(N)}, \mathbf{y}^{(N)})\}$, kde $\mathbf{x}^{(i)} \in R^n$ a $\mathbf{y}^{(i)} \in R^m$, $\forall i \in \{1, 2, \dots, N\}$. Sieť bude vykonávať činnosť predpísanú trénovacou množinou T , ak nájdeme takú maticu váh \mathbf{W} , že

$$\mathbf{y}^{(i)} = \mathbf{W}\mathbf{x}^{(i)}, \forall i \in \{1, 2, \dots, N\}. \quad (4.4)$$

Nech \mathbf{X} je matica, ktorej stĺpce tvoria trénovalie vstupy $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}$, a \mathbf{Y} matica,



ktorej stĺpce sú zodpovedajúce “požadované” výstupy $\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \dots, \mathbf{y}^{(N)}$. Potom podmienka (4.4) sa dá prepísať v maticovom tvare

$$\mathbf{Y} = \mathbf{W}\mathbf{X}. \quad (4.5)$$

Pokiaľ je matica \mathbf{X} regulárna¹, požadované váhy prepojení dostávame priamo z (4.5)

$$\mathbf{W} = \mathbf{Y}\mathbf{X}^{-1}. \quad (4.6)$$

Vo všeobecnosti však nemožno predpokladať ani len to, že matica \mathbf{X} je štvorcová (počet vzoriek v trénovej množine je rovný dimenzii vstupných vzorov), nieto ešte lineárnu nezávislosť vstupov $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}$. V takom prípade je (4.6) modifikované na [1, 2]

$$\mathbf{W} = \mathbf{Y}\mathbf{X}^+ \quad (4.7)$$

kde $\mathbf{X}^+ = \mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}$, ak počet riadkov matice \mathbf{X} je menší ako počet jej stĺpcov (počet trénovalích vzoriek N je väčší ako dimenzia vstupov n) a hodnosť matice \mathbf{X} je rovná n . V opačnom prípade $\mathbf{X}^+ = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$, ak pravda, hodnosť matice \mathbf{X} je N . Predpokladáme, že trénovalie príklady korešpondujúce s lineárne závislými vstupmi sme z trénovej množiny vylúčili. Matica \mathbf{X}^+ sa nazýva *pseudoinverzná matica* matice \mathbf{X} a platí

$$\mathbf{X}\mathbf{X}^+\mathbf{X} = \mathbf{X}. \quad (4.8)$$

V ďalšom sa budeme zaoberať prípadom, keď počet trénovalích vzoriek N je menší ako dimenzia vstupov n . Predstavme si, že máme vyriešiť úlohu autoasociácie, teda zapamätať

¹ Štvorcová matica \mathbf{X} dimenzie n sa nazýva regulárna, ak všetky jej riadky (resp. stĺpce) sú lineárne nezávislé. Jej determinant je rôzny od nuly.

si určité “prototypové” vzory $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}$, ktoré tvoria stĺpce matice \mathbf{X} . Trénovacia množina má tvar

$$T = \left\{ (\mathbf{x}^{(1)}, \mathbf{x}^{(1)}) (\mathbf{x}^{(2)}, \mathbf{x}^{(2)}), \dots, (\mathbf{x}^{(N)}, \mathbf{x}^{(N)}) \right\} \quad (4.9)$$

a matica požadovaných výstupov $\mathbf{Y}=\mathbf{X}$. Z (4.7) dostávame pre váhovú maticu

$$\mathbf{W} = \mathbf{X}\mathbf{X}^+ \quad (4.10)$$

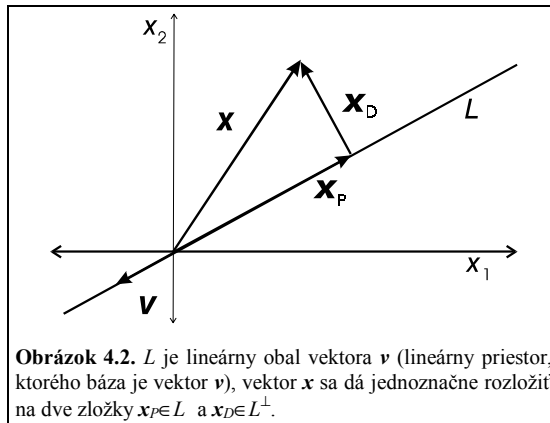
Samozrejme, jedným z možných riešení nášho problému by mohlo byť $\mathbf{W}=\mathbf{I}$, kde \mathbf{I} je jednotková matica dimenzie n . V takomto prípade by sieť len slepo kopírovala na výstup to, čo vidí na vstupe. Takéto riešenie by sme dostali pomocou (4.9), ak by sme mali $N=n$ lineárne nezávislých vstupných vzorov. Ak je počet vstupných vzorov menší ako ich dimenzia, lineárny obal vzorov (množina vektorov, ktoré možno dostať lineárnou kombináciou vzorov) je lineárny podpriestor L lineárneho priestoru R^n . Označme ortogonálny doplnok priestoru L v R^n ako L^\perp . L^\perp je lineárny podpriestor priestoru R^n obsahujúci všetky vektory kolmé na priestor L . Zopakujme si, že vektor je kolmý na priestor L , ak je kolmý na všetky vektory obsiahnuté v L . Zrejme platí, že ak je vektor kolmý na každý vektor z bázy priestoru L , potom je kolmý na priestor L . Každý vektor $\mathbf{x} \in R^n$ sa dá jednoznačne rozložiť $\mathbf{x}=\mathbf{x}_p+\mathbf{x}_D$ na zložky $\mathbf{x}_p \in L$ a $\mathbf{x}_D \in L^\perp$ (obr. 4.2).

Tento fakt sa ľahko dokáže sporom. Nech existujú dva rôzne rozklady

$$\mathbf{x}=\mathbf{x}_{p1}+\mathbf{x}_{D1} \text{ a } \mathbf{x}=\mathbf{x}_{p2}+\mathbf{x}_{D2},$$

kde $\mathbf{x}_{p1}, \mathbf{x}_{p2} \in L$ a $\mathbf{x}_{D1}, \mathbf{x}_{D2} \in L^\perp$. Zrejme $\mathbf{x}_{p1}-\mathbf{x}_{p2}=\mathbf{x}_{D2}-\mathbf{x}_{D1}$. Ľavá strana poslednej rovnosti je lineárna kombinácia vektorov z priestoru L , a teda je to vektor z priestoru L , kým pravá strana z analogických dôvodov je vektor z priestoru L^\perp . Toto je možné, len ak pravá aj ľavá strana sú nulové vektory a teda $\mathbf{x}_{p1}=\mathbf{x}_{p2}$ a $\mathbf{x}_{D1}=\mathbf{x}_{D2}$, čo je v spore s našim predpokladom dvoch rôznych rozkladov.

Dopredu prezradíme, že stratégia, ktorou sa lineárna sieť s váhami $\mathbf{W}=\mathbf{X}\mathbf{X}^+$ zhostí autoasociačnej úlohy je nasledujúca [2, 3]: Vstupné “prototypové” vzorky definujú podpriestor L v celom vstupnom priestore R^n . L je lineárny obal prototypových vstupných vektorov. Sieť považuje každú odchýlku od priestoru L za “pridaný šum”, ktorý vyfiltruje. Pri príchode neznámej vstupnej vzorky \mathbf{x} sieť “predpokladá”, že ide o málo deformovanú podobu niektorého zo vstupných prototypov (uložených ako stĺpce matice \mathbf{X}) a urobí ortogonálny priemet \mathbf{x}_p vektora \mathbf{x} do podpriestoru L , pričom \mathbf{x}_p prehlási za “vyčistenú” verziu vstupu \mathbf{x} . Vektor $\mathbf{x}_D=(\mathbf{x}-\mathbf{x}_p) \in L^\perp$ predstavuje časť, ktorú sieť považuje za deformáciu vstupu vzhľadom na uložené prototypy. Je možný aj iný pohľad, interpretujúci vektor \mathbf{x}_D ako “novátorský aspekt” vstupu \mathbf{x} , vo svetle vzorov predstavených v trénovacej množine. Aby sme dokázali, že hore uvedená stratégia zodpovedá skutočnosti, musíme ukázať, že $\mathbf{W}\mathbf{x}=\mathbf{X}\mathbf{X}^+\mathbf{x}=\mathbf{x}_p$. Inými slovami, je treba dokázať, že matica $\mathbf{X}\mathbf{X}^+$ je operátor vykonávajúci ortogonálnu projekciu na lineárny obal stĺpcov matice \mathbf{X} . Nájdime rozklad $\mathbf{x}=\mathbf{x}_1+\mathbf{x}_2$ vektora \mathbf{x} , že $\mathbf{x}_1 \in L$ a $\mathbf{x}_2 \in L^\perp$. Vieme, že $\mathbf{x}_2 \in L^\perp$ práve vtedy, keď $\mathbf{x}_2 \perp \mathbf{x}^{(i)}$, $i=1, 2, \dots, N$, čo v maticovej formulácii znamená



$$X^T x_2 = o, \quad (4.11)$$

kde o je nulový vektor dimenzie n .

V ďalšom budeme využívať fakt, že riešením sústavy lineárnych rovníc $Au = b$, kde A je matica sústavy s väčším počtom stĺpcov ako riadkov, je

$$u = A^+ b + (I - A^+ A)v. \quad (4.12)$$

v je ľubovoľný vektor rovnakej dimenzie ako vektor x . To sa dá ľahko vidieť, pretože

$$Au = AA^+ b + (A - AA^+ A)v = AA^+ b = b, \quad (4.13)$$

keďže $AA^+ = AA^T (AA^T)^{-1} = I$. Matica X^T je požadovaného typu a teda z (4.11) a (4.12) dostávame

$$x_2 = (X^T)^+ o + (I - (X^T)^+ X^T)v \quad (4.14)$$

Ľahko sa presvedčíme, že

$$(X^T)^+ X^T = X(X^T X)^{-1} X^T = XX^+ \quad (4.15)$$

takže

$$x_2 = (I - XX^+)v. \quad (4.16)$$

Zostáva nájsť "ten pravý" vektor v definujúci ortogonálny rozklad vektora x . Presvedčíme sa, že taký rozklad dostávame práve vtedy, keď $v = x$. Skutočne,

$$x_2^T x_1 = (x - XX^+ x)^T XX^+ x = (x^T - x^T XX^+) XX^+ x = 0, \quad (4.17)$$

lebo XX^+ je symetrická² matica a $XX^+ XX^+ = XX^+$.

² Matica X sa nazýva symetrická, ak $X = X^T$, kde X^T je transponovaná matica k matici X , ktorá vznikne z X zámennou riadkov za stĺpce.

Ukázali sme, že $\mathbf{x}_1 = \mathbf{X}\mathbf{X}^+ \mathbf{x} = \mathbf{x}_p$, čo znamená, že $\mathbf{X}\mathbf{X}^+$ je matica vykonávajúca ortogonálnu projekciu na lineárny obal stĺpcov matice \mathbf{X} (obr. 4.3).

Podotýkame, že ak by sme namiesto $\mathbf{X}\mathbf{X}^+$ použili váhovú maticu $\mathbf{W} = \mathbf{I} - \mathbf{X}\mathbf{X}^+$, dostali by sme ako odozvu na vstup \mathbf{x} vektor $\mathbf{x}_2 = \mathbf{W}\mathbf{x} = (\mathbf{I} - \mathbf{X}\mathbf{X}^+) \mathbf{x} = \mathbf{x}_D$, ktorý sa dá voľne interpretovať ako reprezentant nového a originálneho vo vstupe \mathbf{x} , vzhľadom na tréningové prototypy uložené ako stĺpce matice \mathbf{X} . Takáto neurónová sieť dostala meno detektor novosti (angl. *novelty detector*) [2, 3].

Zatiaľ sme si ukázali, ako pre danú tréningovú množinu T vytvoriť váhovú maticu, ktorá realizuje zobrazenie určené asociačnými párami z T . Okrem iného, je potrebné vypočítať pseudoinverznú maticu k matici \mathbf{X} , čo pri úlohách reálnejšieho rozsahu môže byť výpočtovo veľmi náročné. Navyše, ak by sme po natrénovaní siete na určitej tréningovej množine T chceli pribrať ešte jeden nový tréningový príklad, museli by sme odznova prerátať celú váhovú maticu. Ukážeme si, že v prípade autoasociačnej úlohy sa dá robiť výpočet váhovej matice iteratívne a príchod nového tréningového príkladu predstavuje len pomerne jednoduchý výpočtový úkon príslušne modifikujúci maticu váh \mathbf{W} .

Predstavme si, že stojíme pred nasledujúcou úlohou. K danej báze $\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \dots, \mathbf{u}^{(k)}$ vektorového priestoru L máme vytvoriť ortogonálnu bázu $\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \dots, \mathbf{v}^{(k)}$ definujúcu ten istý priestor L . Algoritmus nazvaný Gramov a Schmidtov (G-S) ortogonalizačný proces realizuje presne takúto úlohu a jeho popis je nasledovný [2]:

1. Položme $\mathbf{v}^{(1)} = \mathbf{u}^{(1)}$
2. V priestore určenom bázou $\mathbf{v}^{(1)}, \mathbf{u}^{(2)}$ nájdime vektor $\mathbf{v}^{(2)}$, tak aby $\mathbf{v}^{(1)} \perp \mathbf{v}^{(2)}$. To sa dá dosiahnuť napríklad takto

$$\mathbf{v}^{(2)} = \mathbf{u}^{(2)} - \frac{\mathbf{v}^{(1)T} \mathbf{u}^{(2)}}{\|\mathbf{v}^{(1)}\|^2} \mathbf{v}^{(1)}. \quad (4.18)$$

Čitateľ sa môže ľahko presvedčiť, že vektor $\mathbf{v}^{(2)}$ je lineárnou kombináciou vektorov $\mathbf{v}^{(1)}, \mathbf{u}^{(2)}$ a skalárny súčin vektorov $\mathbf{v}^{(1)}, \mathbf{v}^{(2)}$ je rovný nule.

3. Bod 2. sa dá rekurentne rozšíriť nasledovne: Nech \tilde{k} vektorov $\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \dots, \mathbf{v}^{(\tilde{k})}$ novej ortogonálnej bázy už bolo určených. Potom k -ty vektor $\mathbf{v}^{(k)}$, od ktorého požadujeme aby ležal v priestore určenom bázou $\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \dots, \mathbf{v}^{(\tilde{k})}, \mathbf{u}^{(k)}$ a bol ortogonálny na všetky vektory $\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \dots, \mathbf{v}^{(\tilde{k})}$, je určený takto:

$$\mathbf{v}^{(k)} = \mathbf{u}^{(k)} - \sum_{i=1}^{\tilde{k}} \frac{\mathbf{v}^{(i)T} \mathbf{u}^{(k)}}{\|\mathbf{v}^{(i)}\|^2} \mathbf{v}^{(i)}. \quad (4.19)$$

Opäť sa dá ľahko overiť, že $\mathbf{v}^{(k)}$ vyhovuje našim požiadavkám.

Predpokladajme, že tréningová množina má N vzoriek $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}$ a \mathbf{x} je vektor neležiaci v ich lineárnom obale. Nech $\tilde{\mathbf{x}}^{(1)}, \tilde{\mathbf{x}}^{(2)}, \dots, \tilde{\mathbf{x}}^{(N)}$ sú vektory určené zo vstupov $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}$ pomocou G-S ortogonalizačného procesu. Zrejme vektor

$$\tilde{\mathbf{x}} = \mathbf{x} - \sum_{i=1}^N \frac{\tilde{\mathbf{x}}^{(i)T} \mathbf{x}}{\|\tilde{\mathbf{x}}^{(i)}\|^2} \tilde{\mathbf{x}}^{(i)} \quad (4.20)$$

je kolmý na lineárny obal vstupov $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}$. Navyše ukážeme, že vektory $\tilde{\mathbf{x}}$ a $\mathbf{x} - \tilde{\mathbf{x}}$ sú na seba kolmé, čo znamená, že $\mathbf{x} - \tilde{\mathbf{x}} = \mathbf{x}_P = \mathbf{X}\mathbf{X}^+ \mathbf{x}$ a $\tilde{\mathbf{x}} = \mathbf{x}_D = (\mathbf{I} - \mathbf{X}\mathbf{X}^+) \mathbf{x}$, kde \mathbf{X} je matica, ktorej stĺpce tvoria vektory $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}$. Vypočítajme skalárny súčin $\tilde{\mathbf{x}}^T (\mathbf{x} - \tilde{\mathbf{x}})$. Keďže súčin matic je asociatívny, dostaneme

$$(\tilde{\mathbf{x}}^{(i)T} \mathbf{x}) \tilde{\mathbf{x}}^{(i)} = \tilde{\mathbf{x}}^{(i)} (\tilde{\mathbf{x}}^{(i)T} \mathbf{x}) = (\tilde{\mathbf{x}}^{(i)} \tilde{\mathbf{x}}^{(i)T}) \mathbf{x}, \quad (4.21)$$

a preto

$$\tilde{\mathbf{x}}^T (\mathbf{x} - \tilde{\mathbf{x}}) = \left(\left[\mathbf{I} - \sum_{i=1}^N \frac{\tilde{\mathbf{x}}^{(i)} \tilde{\mathbf{x}}^{(i)T}}{\|\tilde{\mathbf{x}}^{(i)}\|^2} \right] \mathbf{x} \right)^T \left(\sum_{i=1}^N \frac{\tilde{\mathbf{x}}^{(i)} \tilde{\mathbf{x}}^{(i)T}}{\|\tilde{\mathbf{x}}^{(i)}\|^2} \right) \mathbf{x}. \quad (4.22)$$

Keď označíme maticu $\sum_{i=1}^N \frac{\tilde{\mathbf{x}}^{(i)} \tilde{\mathbf{x}}^{(i)T}}{\|\tilde{\mathbf{x}}^{(i)}\|^2}$ symbolom \mathbf{A} , potom z (4.22) $\mathbf{x}^T (\mathbf{I} - \mathbf{A}^T) \mathbf{A} \mathbf{x} = \mathbf{x}^T (\mathbf{A} - \mathbf{A}^T \mathbf{A}) \mathbf{x} = 0$, lebo $\mathbf{A}^T \mathbf{A} = \mathbf{A}$. To sa dá ľahko dokázať. Stačí si uvedomiť, že

matica \mathbf{A} je symetrická (a teda $\mathbf{A}^T \mathbf{A} = \mathbf{A} \mathbf{A}$) a $\frac{\tilde{\mathbf{x}}^{(i)} \tilde{\mathbf{x}}^{(i)T}}{\|\tilde{\mathbf{x}}^{(i)}\|^2} \frac{\tilde{\mathbf{x}}^{(j)} \tilde{\mathbf{x}}^{(j)T}}{\|\tilde{\mathbf{x}}^{(j)}\|^2} = 0$, pre $i \neq j$, zatiaľ čo

$$\frac{\tilde{\mathbf{x}}^{(i)} \tilde{\mathbf{x}}^{(i)T}}{\|\tilde{\mathbf{x}}^{(i)}\|^2} \frac{\tilde{\mathbf{x}}^{(i)} \tilde{\mathbf{x}}^{(i)T}}{\|\tilde{\mathbf{x}}^{(i)}\|^2} = \frac{\tilde{\mathbf{x}}^{(i)} \tilde{\mathbf{x}}^{(i)T}}{\|\tilde{\mathbf{x}}^{(i)}\|^4} \|\tilde{\mathbf{x}}^{(i)}\|^2 = \frac{\tilde{\mathbf{x}}^{(i)} \tilde{\mathbf{x}}^{(i)T}}{\|\tilde{\mathbf{x}}^{(i)}\|^2}, \text{ pre } i=j. \text{ Ukázali sme, že priemet vektora}$$

$\mathbf{x} = \mathbf{x}^{(N+1)}$ do lineárneho obalu vektorov $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}$ možno určiť ako

$$\tilde{\mathbf{x}}^{(N+1)} = \sum_{i=1}^N \frac{\tilde{\mathbf{x}}^{(i)T} \mathbf{x}^{(N+1)}}{\|\tilde{\mathbf{x}}^{(i)}\|^2} \tilde{\mathbf{x}}^{(i)} = \sum_{i=1}^N \frac{\tilde{\mathbf{x}}^{(i)} \tilde{\mathbf{x}}^{(i)T}}{\|\tilde{\mathbf{x}}^{(i)}\|^2} \mathbf{x}^{(N+1)} = \mathbf{A}_N \mathbf{x}^{(N+1)}, \quad (4.23)$$

kde $\mathbf{A}_N = \sum_{i=1}^N \frac{\tilde{\mathbf{x}}^{(i)} \tilde{\mathbf{x}}^{(i)T}}{\|\tilde{\mathbf{x}}^{(i)}\|^2}$. Ak by sme teraz uvažovali ďalší vstupný vektor $\mathbf{x}^{(N+2)}$, neležiaci v

lineárnom obale vektorov $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}, \mathbf{x}^{(N+1)}$, jeho priemet do tohoto lineárneho obalu dostaneme pomocou $\tilde{\mathbf{x}}^{(N+2)} = \mathbf{A}_{N+1} \mathbf{x}^{(N+2)}$, kde $\mathbf{A}_{N+1} = \sum_{i=1}^{N+1} \frac{\tilde{\mathbf{x}}^{(i)} \tilde{\mathbf{x}}^{(i)T}}{\|\tilde{\mathbf{x}}^{(i)}\|^2}$ a $\tilde{\mathbf{x}}^{(N+1)}$ dostaneme z predošlého kroku ortogonalizačného procesu $\tilde{\mathbf{x}}^{(N+1)} = \mathbf{x}^{(N+1)} - \mathbf{A}_N \mathbf{x}^{(N+1)}$.

Ukázali sme si, že maticu váh prepojení $\mathbf{W} = \mathbf{X}\mathbf{X}^+$ lineárneho autoasociátora môžeme rátať iteračne, vždy pri príchode novej prototypovej vstupnej vzorky, na základe rekurentnej schémy:

1. Inicializuj \mathbf{W}_0 ako nulovú maticu, $\mathbf{W}_0 = \mathbf{0}$.
2. Pre $N > 0$,

$$\mathbf{W}_{N+1} = \mathbf{W}_N + \frac{\tilde{\mathbf{x}}^{(N+1)} \tilde{\mathbf{x}}^{(N+1)T}}{\|\tilde{\mathbf{x}}^{(N+1)}\|^2}, \quad (4.24)$$

kde $\tilde{\mathbf{x}}^{(N+1)} = \mathbf{x}^{(N+1)} - \mathbf{W}_N \mathbf{x}^{(N+1)}$. Model, ktorým sme sa doteraz zaoberali, dostal názov GI (angl. *General Inverse*) [2, 3] podľa pseudoinverznej matice \mathbf{X}^+ použitej pri výpočte matice synaptických váh. V ďalšom si ukážeme iný model lineárnych sietí, ktorý sa líši od predošlého maticou \mathbf{W} synaptických váh.

4.1 Realizácia pamäti pomocou korelačnej matice

V roku 1949 sformuloval kanadský psychológ Hebb hypotézu o plasticite zmeny synaptických váh [4]. Priepustnosť synaptického spojenia vstupného kanála s neurónovou bunkou je priamo úmerná pred- a postsynaptickej aktivite neurónu. Inými slovami, ak je veľká korelácia medzi aktivitou na kanáli z j -teho neurónu do neurónu i a výstupnou aktivitou neurónu i , potom nárast synaptickej váhy kanála je priamo úmerný tejto korelácii. Formálne, ak je trénovacia množina

$$\mathcal{T} = \{(\mathbf{x}^{(1)}, \mathbf{y}^{(1)}), (\mathbf{x}^{(2)}, \mathbf{y}^{(2)}), \dots, (\mathbf{x}^{(N)}, \mathbf{y}^{(N)})\}, \quad (4.25)$$

potom váha w_{ij} bude priamo úmerná korelácii

$$w_{ij} \approx \sum_{k=1}^N x_j^{(k)} y_i^{(k)}, \quad (4.26)$$

čomu v maticovom zápise zodpovedá

$$\mathbf{W} \approx \sum_{k=1}^N \mathbf{y}^{(k)} \mathbf{x}^{(k)T} = \mathbf{Y}\mathbf{X}^T. \quad (4.27)$$

V našich ďalších úvahách budeme uvažovať $\mathbf{W} = \mathbf{Y}\mathbf{X}^T$. Okamžite vidíme, že ak sú vstupné trénovacie vektory $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}$ ortonormálne, dostávame $\mathbf{X}^T = \mathbf{X}^{-1} = \mathbf{X}^+$ a náš “biologicky motivovaný” model je totožný s modelom GI, ktorým sme sa zaoberali v predošlom odseku. Požiadavka ortonormality vstupov však v praxi ťažko obstoí. Preto je namieste otázka, aká je funkcia siete, ak vstupné vektory sú len lineárne nezávislé. Opäť sa pokúsime ilustrovať funkciu siete na autoasociačnej úlohe.

Máme si zapamätať N lineárne nezávislých vstupov $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}$ uložených ako stĺpce matice \mathbf{X} . Zrejme $\mathbf{W} = \sum_{k=1}^N \mathbf{x}^{(k)} \mathbf{x}^{(k)T} = \mathbf{X}\mathbf{X}^T$. Po priložení p -teho vektora $\mathbf{x}^{(p)}$, $p \in \{1, 2, \dots, N\}$ na vstup takejto siete dostávame ako odozvu vektor

$$\begin{aligned} \mathbf{X}\mathbf{X}^T \mathbf{x}^{(p)} &= \left(\sum_{k=1}^N \mathbf{x}^{(k)} \mathbf{x}^{(k)T} \right) \mathbf{x}^{(p)} \\ &= \sum_{k=1}^N \mathbf{x}^{(k)} \mathbf{x}^{(k)T} \mathbf{x}^{(p)} = \mathbf{x}^{(p)} \mathbf{x}^{(p)T} \mathbf{x}^{(p)} + \sum_{k=1, k \neq p}^N \mathbf{x}^{(k)} \mathbf{x}^{(k)T} \mathbf{x}^{(p)} \end{aligned} \quad (4.28)$$

a teda

$$\mathbf{X}\mathbf{X}^T \mathbf{x}^{(p)} = \mathbf{x}^{(p)} \|\mathbf{x}^{(p)}\|^2 + \mathbf{C}(p), \quad (4.29)$$

kde $\mathbf{C}(p) = \sum_{k=1, k \neq p}^N \mathbf{x}^{(k)} \mathbf{x}^{(k)T} \mathbf{x}^{(p)}$ je tzv. "presluch" (angl. *crosstalk*) od iných

vstupných vzoriek. Ak je presluch nenulový, vstupy $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}$ nie sú ortogonálne. Vo všeobecnosti teda na výstupe siete dostávame súčet vektora kolineárneho s daným uloženým vstupom a "presluchového" vektora.

V prípade modelu GI sme si vysvetlili činnosť autoasociátora aparátom lineárnej algebry. Videli sme, že takáto sieť zovšeobecni vstupov trérovacej množiny tak, že vytvorí ich lineárny obal. Odozva siete na nový, dosiaľ nepredložený vstup (ktorý môže predstavovať poškodenú, alebo zašumenú verziu niektorého z trérovacích vzorových vstupov) je ortogonálny priemet vstupu do lineárneho obalu trérovacích vstupov. Činnosť autoasociátora založeného na modeli uvedenom v tejto sekcii si vysvetlíme štatistickými úvahami.

Predstavme si, že trérovacie vstupy $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}$ sú realizácie vektorovej náhodnej premennej $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$, ktorej každá zložka má strednú hodnotu nula ($E(X_i) = 0$, $i = 1, 2, \dots, n$).

Prvok w_{ij} váhovej matice $\mathbf{W} = \mathbf{X}\mathbf{X}^T$ je rovný $w_{ij} = \sum_{k=1}^N x_i^{(k)} x_j^{(k)}$, čo je priamo úmerné

neodchýlenému odhadu $\langle \text{cov}(X_i, X_j) \rangle$ kovariancie

$\text{cov}(X_i, X_j) = E[(X_i - E(X_i))(X_j - E(X_j))] = E[X_i X_j]$ i -tej a j -tej zložky náhodného vektora \mathbf{X} . Váha w_{ij} nesie informáciu o sile vzájomnej (lineárnej) previazanosti náhodných

premenných X_i, X_j . Priložme na vstup siete vektor $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$. i -ta zložka výstupu je daná súčtom príspevkov od všetkých súradníc ováňovaným príslušnými

koreláciami a je priamo úmerná $\sum_{j=1}^n \langle \text{cov}(X_i, X_j) \rangle x_j$. Ak by bola i -ta zložka vstupu x

vplyvom šumu silne zredukovaná a na základe tréningových vstupov je možné sa domnievať, že napríklad existuje silná korelácia medzi i -tou a r -tou, ako aj medzi i -tou a p -tou zložkou vstupov (váhy w_{ir} a w_{ip} sú dominantné), potom ak zložky x_r , x_p vstupu \mathbf{x} sú silne aktivované, spôsobia aj silnú aktiváciu i -tej zložky výstupu. Autokorekcia i -tej zložky vstupu sa udiala na základe pozorovania, že súčasne s r -tou a p -tou zložkou je obvyčajne aktivovaná aj zložka i a zložky r a p mali vysokú aktivitu. Podobná úvaha platí pre štatisticky slabo prepojené zložky vstupov. Tu zasa možno stiahnuť vysokú aktivitu nejakej zložky vstupného vektora, ak na tréningových vstupoch nebola pozorovaná výrazná previazanosť tejto zložky na momentálne aktívne komponenty vstupov.

Model autoasociátora, ktorý sme popísali, sa nazýva pamäť korelačnej matice (angl. *Correlation Matrix Memory*, CMM) [2, 3]. Treba poznamenať, že tento model sa opiera o štatistiky budované nad dvojicami zložiek vstupných vektorov. To môže byť pre zachytenie podstatných korelácií málo, pretože výrazným spôsobom môžu byť prepojené napr. trojice, alebo štvorice vstupných zložiek a na zachytenie tejto skutočnosti by sme potrebovali odhadovať štatistiky vyšších rádov.

Predpoklad, že stredné hodnoty zložiek náhodného vektora X (ktorého realizácie sú prototypové vstupy) sú nulové nie je až taký nerealistický, ako by sa mohlo zdať na prvý pohľad. Vždy je totiž možné urobiť odhad strednej hodnoty vstupov

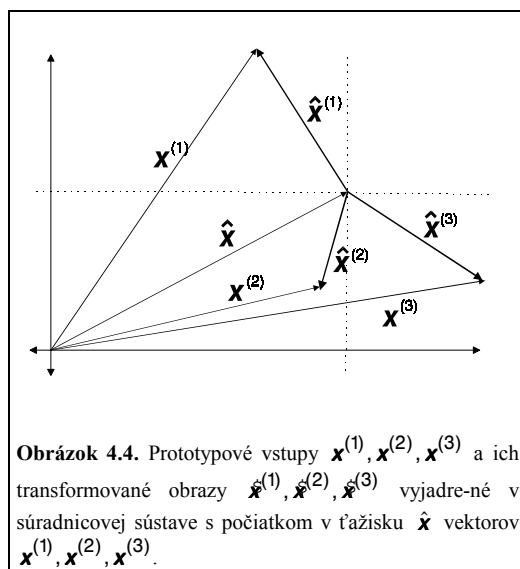
$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{k=1}^N \mathbf{x}^{(k)}, \quad (4.30)$$

a posunúť do nej počiatok súradnicovej sústavy

$$\hat{\mathbf{x}} = \mathbf{x} - \bar{\mathbf{x}}. \quad (4.31)$$

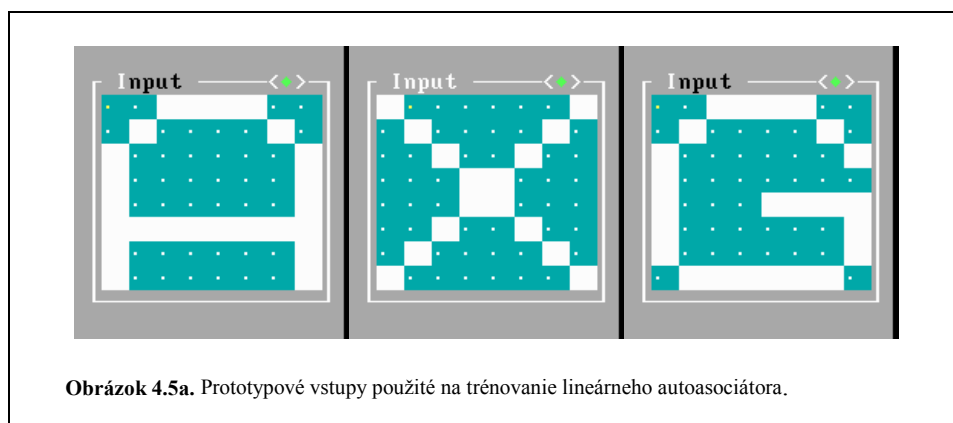
Maticu X potom skonštruujeme z transformovaných prototypových vstupov $\hat{\mathbf{x}}^{(i)}$, $i=1,2,\dots,N$. Odozvu siete na transformovaný vstup opäť transformujeme do pôvodnej súradnicovej sústavy spätnou transformáciou $\mathbf{x} = \hat{\mathbf{x}} + \bar{\mathbf{x}}$.

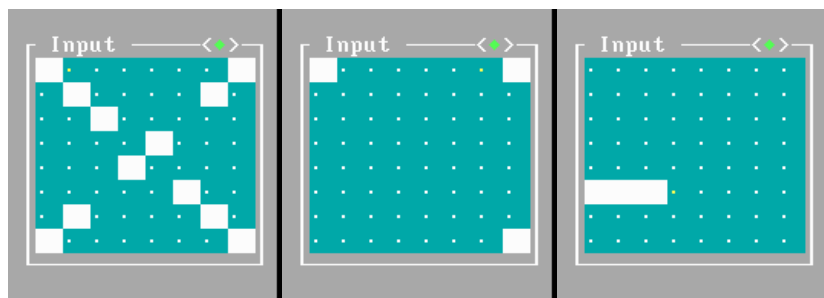
Transformácia (4.31) pomáha zmenšiť presluch medzi prototypovými vstupmi (ako naznačuje obr. 4.4), pretože ak sú vstupné vektory "primknuté k sebe", ich transformované verzie budú navzájom zvierat' väčšie uhly.



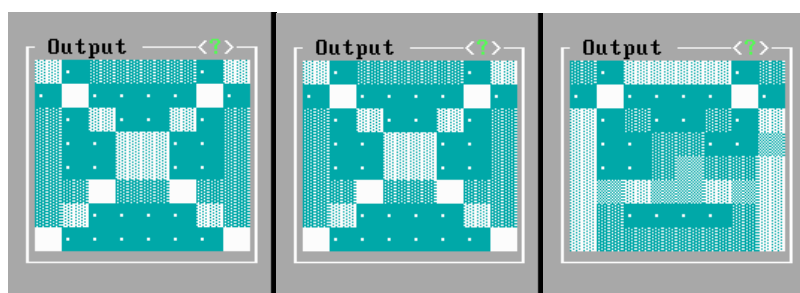
4.2 Príklady lineárnej autoasociácie

Na obr. 4.5a vidíme 3 prototypové vzory priložené na vstup lineárnej autoasociačnej siete so 64 neurónmi. Sieť bola testovaná na 3 porušených vstupoch zobrazených na obrázku 4.5b. Ide o dve “porušené” písmená X a jedno narušené písmeno A. Odozvy siete trénovanej stratégiou CMM sú prezentované na obrázku 4.5c. Obrázok 4.5d zobrazuje odozvy siete trénovanej stratégiou GI. Keďže prototypové vzory nie sú ortogonálne, odozvy siete GI sú kvalitnejšie než odozvy siete CMM.

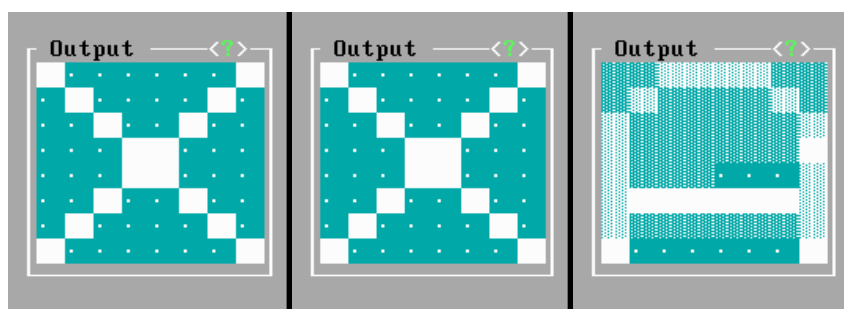




Obrázok 4.5b. Testovacie vstupy pre lineárny asociátor trénovaný na vstupoch zobrazených na obrázku 4.5a.



Obrázok 4.5c. Odozvy siete typu CMM na testovacie vstupy (obrázok 4.5b).



Obrázok 4.5d. Odozvy siete typu GI na testovacie vstupy (obrázok 4.5b).

Literatúra

- [1] F. Štulajter. *Odhady v náhodných procesoch*. Alfa, Bratislava, 1990.
- [2] G.E. Hinton and J.A. Anderson, editors. *Parallel Models of Associative Memory*. Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1989.
- [3] T. Kohonen. *Self-Organization and Associative Memory*. Springer-Verlag, Berlin, 1984.
- [4] D.O. Hebb. *The Organization of Behavior*. Wiley, New York, 1949.