Jednovrstvé sítě II

Hopfieldova síť

John Hopfield, 1982

Základní Hopf. myšlenka:

Naučené vzory si síť zapamatuje jako (pokud možno) hluboká údolí v krajině energetické funkce.

Charakteristika sítě

- Zpětnovazební, neboli rekurentní síť,
  - nelze jednoznačně definovat směr informačního toku, neboli
  - síť obsahuje zpětné vazby mezi buňkami nebo jejich skupinami,
- vybavuje nejbližšího souseda,
  - autoasociativní chování,
- učí se Hebbovsky.

Připomenutí I:

- Co to je to Hebbovské učení?

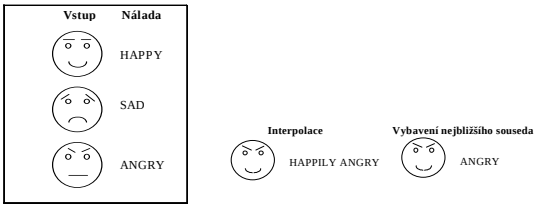
"When an axon of cell *A* is near enough to excite a cell *B* and repeatedly or persistantly takes part in firing it, some growth process or metabolic change takes place in one or both cells, such that *A* 's efficiency, as one of the cells firing *B*, is increased." [Hebb49]

Přeloženo do terminologie umělých neuronových sítí:

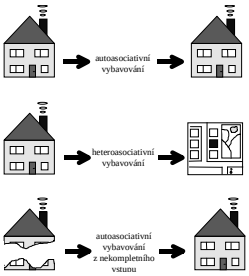
Připomenutí I - pokračování:

- Jsou-li neurony na obou koncích některé synapse aktivovány současně, pak tuto synapsi selektivně posílíme.
- V opačném případě (aktivují-li se nesoučasně), synapsi oslabíme, či dokonce zrušíme.

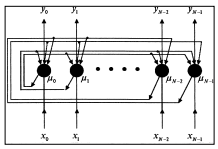
Připomenutí II: vybavení nejbližšího souseda

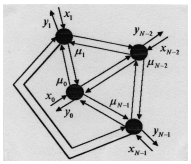


P **III: Autoasociativní x heteroasociativní**



Architektura Hopfieldovy sítě:





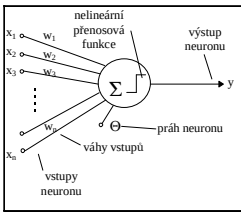
nebo

x_i jsou vstupy, y_i výstupy sítě a μ_i je „okamžitý stav“ neuronu.

Všimněte si:

- Výstup neuronu je spojen se vstupy všech ostatních neuronů, avšak ne se sebou samým!
- Mezi každými dvěma neurony je vzájemné (a uvidíme že i stejně mocné) propojení.
- Neuronů v síti je tolik, kolik má síť vstupů.
- Důležitá poznámka: výstupy *y* sítě jsou po většinu doby jejího fungování neaktivní.

Neuron v Hopfieldově síti



Tedy klasický McCulloch-Pittsův s nulovým prahem

Implementaci zjednodušující konkrétnosti:

- Binární varianta (vstupy i výstupy),
- proměnné nabývají hodnot ± 1,
- v aplikaci „obrázky“ má síť tolik neuronů, kolik je v obrázku pixelů .

Matice vah

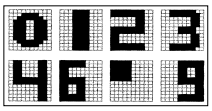
- V dalším budeme potřebovat tzv. matici vah.
- Je to čtvercová matice s N řádky a sloupci.
- N je přitom počet neuronů v síti.
- Jejími prvky budou váhy w_{ij}.

Učení Hopfieldovy sítě

- Nastavení matice vah *W* podle vstupních vzorů takto:

$$w_{ij} = \begin{cases} \sum_{s=0}^{M-1} x_i^s x_j^s, & pro\ i \neq j \\ 0, & pro\ i = j,\ 0 \leq i, j \leq N-1 \end{cases}$$

M vzorů



Kódování prvků vstupního vektoru *X:*

+1 znamená např. bílý bod, -1černý.

Všimněte si tedy:

- Mají-li oba body obrázku stejnou barvu, je hodnota synaptické váhy +1,
- mají-li barvu různou, je hodnota váhy -1.
- Matice *W* je symetrická, protože $w_{ij} = w_{ji}$.
- Prvky na hlavní diagonále jsou nulové.
- Učení jednoho vzoru je jednorázové, vzor stačí předložit jednou! Rozdíl od jiných sítí.

Nedělejte si však velké iluze ...

- Vzorů se síť nemůže naučit příliš mnoho (budeme později kvantifikovat),
- vzory si nesmí být „moc podobné“, musí mít co možná největší Hammingovu vzdálenost,
- síť zadarmo umí i inverzní vzory.

Vybavování Hopfieldovy sítě

- Pozor! Velká zvláštnost - iterace.

Krok 1. Inicializace

Nastavení počátečních stavů podle přeloženého vzoru

$$\mu_i(0) = x_i,\ 0 \leq i \leq N-1$$

x_i je element vzoru, který může nabývat jen hodnot +1 nebo -1.

Vybavování Hopfieldovy sítě

Krok 2. Iterace až do nalezení odpovědi

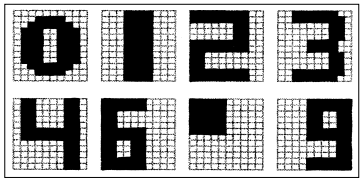
$$\mu_i(t+1) = f\left[\sum_{j=0}^{N-1} w_{ij}\mu_j(t)\right], 0 \leq i \leq N-1$$

Krok 2. provádíme tak dlouho, dokud se mění stavy, tj. dokud nejsou rozdíly mezi těmito a předchozími stavy nulové.

$y_i = \mu_i(t_{posledn'})$ jsou pak přímo jednotlivými body výstupního obrazce.

Tak například

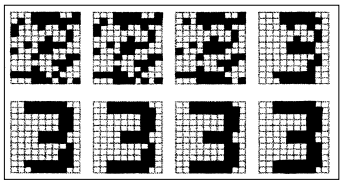
Naučíme Hopfieldovu síť se 120 neurony na tuto sadu znaků:



Matice 10 x 12

Výsledek experimentu vybavování

Síti jsme předložili obraz trojky z 25% poškozený



Iterace při vybavování

Kontrolní otázka: Kdy s iteracemi skončíme?

Vlastnosti Hopfieldovy sítě - souhrn

- Počet vzorů, které lze naučit, je malý. Lze odvodit, že $M < 0{,}15 \times N$, kde N je počet bodů vzoru.
  - Chceme-li naučit 0,…,9, musí mít vzory alespoň 67 bodů.
- Vzory si nesmí být moc podobné,
- váhová matice roste se čtvercem počtu bodů vzoru.
- Velký problém: síť není vůbec imunní vůči posunutí nebo natočení vzoru.

Pozor, častý omyl:

- Hopfieldova síť není klasifikátor! Odpovídá nejbližším sousedem.
- Jak by jí bylo možno upravit, aby se klasifikátorem stala?

FAQ: jak jsme přišli na právě takovou metodu učení?

$$w_{ij} = \begin{cases} \sum_{s=0}^{M-1} x_i^s x_j^s, & pro\ i \neq j \\ 0, & pro\ i = j,\ 0 \leq i, j \leq N-1 \end{cases}$$

Jinými slovy: proč počítáme váhy podle tohoto vzorečku?

Odpověď: plyne to z už dříve uvedené představy hlubokých údolí v energetické krajině.

Jak ale v tomto případě energetickou funkci definovat?

Proč nelze použít energetickou funkci zavedenou v souvislosti s perceptronem?

Odpověď: neumíme určit Eukleidovskou vzdálenost!

Energetickou funkci proto musíme sestavit tak, aby její hodnota byla velká pro velké chyby (kopce) a malá pro malé chyby (údolí). Jedním z možných vyjádření tohoto požadavku je funkce ve tvaru:

$$E = -\tfrac{1}{2} \sum_i \sum_j w_{ij} x_i x_j + \sum_i x_i \theta_i$$

Poznámka 1:

Z následujícího odvození mimo jiné vyplyne, proč je výhodné, aby byly všechny prahy nulové.

Poznámka 2:

Netrapte se, že vás nenapadl právě takovýhle výraz. Proto se tahle síť jmenuje Hopfieldova, a ne třeba ...

Úkol:

- Najděte takovou matici vah *W*, aby energetická funkce měla v místě naučeného vzoru **minimum**.
- Nepoškoďte přitom už naučené vzory (vyhloubená údolí)!

Vyjdeme z naší energetické funkce:

$$E = -\tfrac{1}{2}\sum_i \sum_j w_{ij} x_i x_j + \sum_i x_i \theta_i$$

Chceme, aby byla vždy co nejmenší, tedy co nejzápornější. Aby nebylo třeba měnit prahy v závislosti na vzoru, učiníme všechny **prahy nulové**.

Dále bude výhodné, rozdělíme-li matici vah na dvě části: Jedna bude představovat vliv všech vzorů vyjma právě toho, řekněme, *s*-tého (označme jeho váhu w^*) a druhá část bude právě příspěvkem *s*-tého vzoru. Potom můžeme naši energetickou funkci vyjádřit následovně:

$$E = -\tfrac{1}{2}\sum_i \sum_j w^*_{ij} x_i x_j - \tfrac{1}{2}\sum_i \sum_j w^s_{ij} x^s_i x^s_j = E_s + E_{vse\ bez\ s}$$

šum **signál**

Signál je energie daná *s*-tým vzorem,

zatímco **šum** je zapříčiněn příspěvky od všech ostatních vzorů.

Se šumem nelze nic moc dělat.

Hodnotu signálu však můžeme snížit. Jinými slovy můžeme minimalizovat příspěvek *s*-tého vzoru v energetické funkci.

Učení je přeci nalézání minimální hodnoty energetické funkce.

To odpovídá maximalizaci hodnoty výrazu

$$\sum_i \sum_j w^s_{ij} x^s_i x^s_j$$

.

Hodnota x^s_i je buď +1 nebo -1, ale hodnota $(x^s_i)^2$ je *vždy* kladná. Budeme se tedy snažit dosáhnout závislosti energie na $(x^s_i x^s_j)^2$ a tím dosáhneme největšího součtu. Toho dosáhneme, bude-li

$$w^s_{ij} = x^s_i x^s_j$$

Právě toto nastavení vah minimalizuje energetickou funkci!

Úkol:

- Princip iterací při vybavování si prostudujte ve skriptech!

Problémy:

Energetická funkce

$$E = -\tfrac{1}{2}\sum_i\sum_j w_{ij}^* x_i x_j - \tfrac{1}{2}\sum_i\sum_j w_{ij}^s x_i^s x_j^s = E_s + E_{vse\ bez\ s}$$

Respektuje i rozmístění ostatních minim. Ty mohou reliéf krajiny ovlivnit lokálními minimy, ve kterých může koulející se kulička uvíznout.

Vznikají tzv.**fantomy**, nepravé vzory, které neodpovídají žádným učicím vzorům.

Poznámka

Zbavování sítě fantomů lze přirovnat k léčení neuróz.

Prý: Zlepšování paměti odučováním fantomů probíhá u člověka ve snu. (?)

Odstranění uváznutí

- Natřásáním.
  - Dočasné zvýšení energie sítě umožní kuličce opustit lokální minimum. Snížíme-li natřásání, budou hluboká minima nepřekonatelná.
- Simulovaným žíháním.
  - Tento proces je podobný technologickému procesu žíhání, kdy se materiál po zahřátí postupně ochlazuje. Přechodem jeho struktury do energetického minima se odstraňuje vnitřní pnutí.