

## Organizační dynamika:

- ▶ Cyklická síť se symetrickými spoji (tj. libovolný neorientovaný graf)
- ▶ Množinu všech neuronů značíme  $N$
- ▶ označme  $\xi_j$  vnitřní potenciál a  $y_j$  výstup (stav) neuronu  $j$
- ▶ stav stroje:  $\vec{y} \in \{-1, 1\}^{|N|}$ .
- ▶ označme  $w_{ji}$  *reálnou* váhu spoje od neuronu  $i$  k neuronu  $j$ .
- ▶ žádný neuron nemá bias a předpokládáme  $w_{jj} = 0$  pro  $j \in N$ .

# Boltzmannův stroj - opakování

**Aktivní dynamika:** Stavby neuronů jsou iniciálně nastaveny na hodnoty z množiny  $\{-1, 1\}$ , tj.  $y_j^{(0)} \in \{-1, 1\}$  pro  $j \in N$ .

V  $t$ -tém kroku aktualizujeme náhodně vybraný neuron  $j \in N$  takto: nejprve vypočteme vnitřní potenciál

$$\xi_j^{(t-1)} = \sum_{i \in j_{\leftarrow}}^n w_{ji} y_i^{(t-1)}$$

a poté náhodně zvolíme hodnotu  $y_j^{(t)} \in \{-1, 1\}$  tak, že

$\mathbf{P}[y_j^{(t)} = 1] = \sigma(\xi_j^{(t-1)})$  kde

$$\sigma(\xi) = \frac{1}{1 + e^{-2\xi/T(t)}}$$

Parametr  $T(t)$  se nazývá **teplota** v čase  $t$ .

## Boltzmannův stroj - opakování

- ▶ Velmi vysoká teplota  $T(t)$  znamená, že  $\mathbf{P}[y_j^{(t)} = 1] \approx \frac{1}{2}$  a stroj se chová téměř náhodně.
- ▶ Velmi nízká teplota  $T(t)$  znamená, že buď  $\mathbf{P}[y_j^{(t)} = 1] \approx 1$  nebo  $\mathbf{P}[y_j^{(t)} = 1] \approx 0$  v závislosti na tom, jestli  $\xi_j^{(t)} > 0$  nebo  $\xi_j^{(t)} < 0$ . Potom se stroj chová téměř deterministicky (tj. jako Hopfieldova síť).

# Boltzmannův stroj - reprezentace rozložení

**Cíl:** Chceme sestrojít síť, která bude reprezentovat dané pravděpodobnostní rozložení na množině vektorů  $\{-1, 1\}^{|N|}$ .

**Velmi hrubá a nepřesná idea:** Boltzmannův stroj mění náhodně svůj stav z množiny  $\{-1, 1\}^{|N|}$ .

Když necháme B. stroj běžet dost dlouho s fixní teplotou, potom budou frekvence návštěv jednotlivých stavů nezávislé na iniciálním stavu.

Tyto frekvence budeme považovat za pravděpodobnostní rozložení na  $\{-1, 1\}^{|N|}$  reprezentované B. strojem.

V adaptivním režimu bude zadáno nějaké rozložení na stavech a cílem bude nalézt konfiguraci takovou, že rozložení reprezentované strojem bude odpovídat zadanému rozložení.

# Rovnovážný stav

Fixujeme teplotu  $T$  (tj.  $T(t) = T$  pro  $t = 1, 2, \dots$ ).

Boltzmannův stroj se po jisté době dostane do tzv. *termální rovnováhy*. Tj. existuje čas  $t^*$  takový, že pro libovolný stav stroje  $\gamma^* \in \{-1, 1\}^{|N|}$  platí, že

$$p_N(\gamma^*) := \mathbf{P}[\vec{y}^{(t^*)} = \gamma^*]$$

splňuje  $p_N(\gamma^*) \approx \frac{1}{Z} e^{-E(\gamma^*)/T}$  kde

$$Z = \sum_{\gamma \in \{-1, 1\}^{|N|}} e^{-E(\gamma)/T} \quad E(\gamma) = -\frac{1}{2} \sum_{i,j} w_{ij} y_i^\gamma y_j^\gamma$$

tj. Boltzmannovo rozložení

**Pozn.:** Teorie Markovových řetězců říká, že  $\mathbf{P}[\vec{y}^{(t^*)} = \gamma^*]$  je také dlouhodobá frekvence návštěv stavu  $\gamma^*$ .

Toto platí *bez ohledu na iniciální nastavení neuronů!* Síť tedy reprezentuje rozložení  $p_N$ .

**Problém:** Tak jak jsme si jej definovali má Boltzmannův stroj omezenou schopnost reprezentovat daná rozložení.

Proto množinu neuronů  $N$  disjunktně rozdělíme na

- ▶ množinu **viditelných** neuronů  $V$
- ▶ množinu **skrytých** neuronů  $S$ .

Pro daný stav viditelných neuronů  $\alpha \in \{-1, 1\}^{|V|}$  označme

$$p_V(\alpha) = \sum_{\beta \in \{-1, 1\}^{|S|}} p_N(\alpha, \beta)$$

pravděpodobnost stavu viditelných neuronů  $\alpha$  v termálním ekvilibriu bez ohledu na stav skrytých neuronů.

Cílem bude adaptovat síť podle daného rozložení na  $\{-1, 1\}^{|V|}$ .

## Adaptivní dynamika:

Nechť  $p_d$  je pravděpodobnostní rozložení na množině stavů viditelných neuronů, tj. na  $\{-1, 1\}^{|V|}$ .

Cílem je nalézt konfiguraci sítě  $W$  takovou, že  $p_V$  odpovídá  $p_d$ .

Vhodnou mírou rozdílu mezi rozděleními  $p_V$  a  $p_d$  je relativní entropie zvážená pravděpodobnostmi vzorů (tzv. Kullback Leibler distance)

$$\mathcal{E}(\vec{w}) = \sum_{\alpha \in \{-1, 1\}^{|V|}} p_d(\alpha) \ln \frac{p_d(\alpha)}{p_V(\alpha)}$$

# Boltzmannův stroj - učení

$\mathcal{E}(\vec{w})$  budeme minimalizovat pomocí gradientního sestupu, tj. budeme počítat poslounost vektorů vah  $\vec{w}^{(0)}, \vec{w}^{(1)}, \dots$

- ▶ váhy v  $\vec{w}^{(0)}$  jsou inicializovány náhodně blízko 0
- ▶ v  $\ell$ -tém kroku (zde  $\ell = 1, 2, \dots$ ) je  $\vec{w}^{(\ell)}$  vypočteno takto:

$$w_{ji}^{(\ell)} = w_{ji}^{(\ell-1)} + \Delta w_{ji}^{(\ell)}$$

kde

$$\Delta w_{ji}^{(\ell)} = -\varepsilon(\ell) \cdot \frac{\partial \mathcal{E}}{\partial w_{ji}}(\vec{w}^{(\ell-1)})$$

je změna váhy  $w_{ji}$  v  $\ell$ -tém kroku a  $0 < \varepsilon(\ell) \leq 1$  je rychlost učení v  $\ell$ -tém kroku.

Zbývá spočítat (odhadnout)  $\frac{\partial \mathcal{E}}{\partial w_{ji}}(\vec{w})$ .



# Boltzmannův stroj - učení

Formálním derivováním funkce  $\mathcal{E}$  lze ukázat, že

$$\frac{\partial \mathcal{E}}{\partial w_{ji}} = -\frac{1}{T} \left( \langle y_j^{(t^*)} y_i^{(t^*)} \rangle_{fixed} - \langle y_j^{(t^*)} y_i^{(t^*)} \rangle_{free} \right)$$

- ▶  $\langle y_j^{(t^*)} y_i^{(t^*)} \rangle_{fixed}$  je průměrná hodnota  $y_j^{(t^*)} y_i^{(t^*)}$  v termální rovnováze za předpokladu, že hodnoty viditelných neuronů jsou fixovány na počátku výpočtu dle rozložení  $p_d$ .
- ▶  $\langle y_j^{(t^*)} y_i^{(t^*)} \rangle_{free}$  je průměrná hodnota  $y_j^{(t^*)} y_i^{(t^*)}$  v termální rovnováze bez fixace viditelných neuronů.

Celkově

$$\begin{aligned} \Delta w_{ji}^{(\ell)} &= -\varepsilon(\ell) \cdot \frac{\partial \mathcal{E}}{\partial w_{ji}} (\vec{w}^{(\ell-1)}) \\ &= \frac{\varepsilon(\ell)}{T} \left( \langle y_j^{(t^*)} y_i^{(t^*)} \rangle_{fixed} - \langle y_j^{(t^*)} y_i^{(t^*)} \rangle_{free} \right) \end{aligned}$$

# Boltzmannův stroj - učení

Pro výpočet  $\langle y_j^{(t^*)} y_i^{(t^*)} \rangle_{fixed}$  proved' následující:

▶ Polož  $\mathcal{Y} := 0$  a proved' následující akce  $q$  krát:

1. fixuj náhodně hodnoty viditelných neuronů dle rozložení  $p_d$  (tj. v průběhu následujících kroků je neaktualizuj)
2. simuluj stroj po  $t^*$  kroků

[ Lepší způsob: simuluj stroj tak dlouho, dokud nedosáhne termální rovnováhy při nízké teplotě (zde je možné použít simulované žíhání, tedy pomalé snižování teploty). ]

3. přičti aktuální hodnotu  $y_j y_i$  k proměnné  $\mathcal{Y}$ .

▶  $\mathcal{Y}/q$  bude dobrým odhadem  $\langle y_j^{(t^*)} y_i^{(t^*)} \rangle_{fixed}$  za předpokladu, že  $q$  je dostatečně velké číslo.

$\langle y_j^{(t^*)} y_i^{(t^*)} \rangle_{free}$  se odhadne podobně, pouze se nefixují viditelné neurony (tj. v kroku 1. se zvolí libovolný startovní stav a v následném výpočtu se mohou aktualizovat všechny neurony).

Pro upřesnění analytická verze:

$$\begin{aligned} \langle y_i^{(t^*)} y_j^{(t^*)} \rangle_{fixed} &= \\ &= \sum_{\alpha \in \{-1,1\}^{|V|}} p_d(\alpha) \sum_{\beta \in \{-1,1\}^{|S|}} \frac{p_N(\alpha, \beta)}{p_V(\alpha)} y_j^{\alpha\beta} y_i^{\alpha\beta} \end{aligned}$$

kde  $y_j^{\alpha\beta}$  je výstup neuronu  $j$  ve stavu  $(\alpha, \beta)$ .

$$\langle y_i^{(t^*)} y_j^{(t^*)} \rangle_{free} = \sum_{\gamma \in \{-1,1\}^{|M|}} p_N(\gamma) y_j^\gamma y_i^\gamma$$