Neuronové sítě

Spojité neuronové sítě

Příprava dat

Data $\mathbf{X} \in \mathbb{R}^{m \times n}$ je vhodné před spuštěním perceptronu

- standardizovat rozpětím

$$\mathbf{X}_{ij}^{\mathrm{new}} = \frac{\mathbf{X}_{ij} - m_j}{M_j - m_j} \in \langle 0, 1 \rangle$$

$m_j = \min(\mathbf{X}_{kj})$, $M_j = \max(\mathbf{X}_{kj})$

- standardizovat směrodatnou odchylkou

$$\mathbf{X}_{ij}^{\mathrm{new}} = \frac{\mathbf{X}_{ij} - \mathrm{E}(\mathbf{X}_{kj})}{\mathrm{S}(\mathbf{X}_{kj})} \in N(0, 1)$$

$$\mathrm{E}(\mathbf{X}_{kj}) = \tfrac{1}{m} \sum_{k=1}^{m} \mathbf{X}_{kj}$$

$$\mathrm{S}^2(\mathbf{X}_{kj}) = \tfrac{1}{m-1} \sum_{k=1}^{m} (\mathbf{X}_{kj} - \mathrm{E}(\mathbf{X}_{kj}))^2$$

Lineárně neseparabilní data

Řešení:

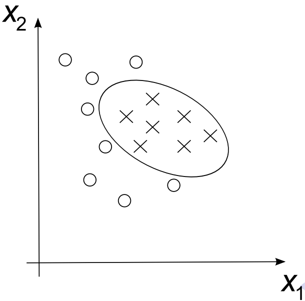
- Síť s jednou či dvěma skrytými vrstvami.
- Rozšíření (vyrovnání) příznakového prostoru

kvadratické

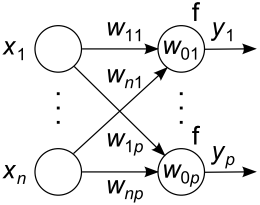
$$x_1, \ldots, x_n, x_1^2, x_1\ x_2, x_1\ x_3, \ldots, x_{n-1}\ x_n, x_n^2$$

kubické

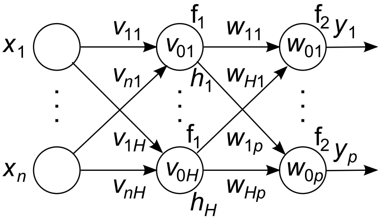
$$x_1, \ldots, x_n, x_1^2, x_1\ x_2, \ldots, x_n^2, x_1^3, x_1^2\ x_2, \ldots, x_n^3$$



Spojité neuronové sítě

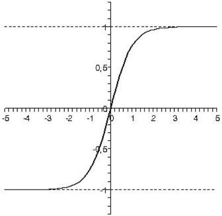


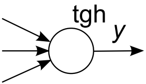
dvouvrstvá síť



třívrstvá síť

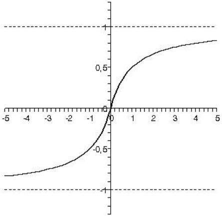
Aktivační funkce

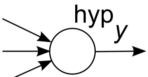




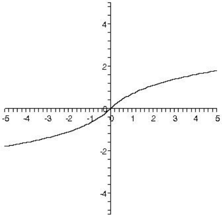
$$f(s) = \mathrm{thg}(s)$$

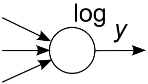
$$s = \sum_{k=0}^{n} w_k x_k$$



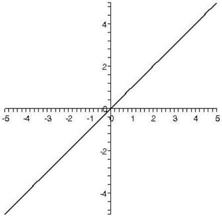


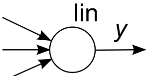
$$f(s) = \frac{s}{1+|s|}$$





$$f(s) = \mathrm{sign}(s) \cdot \ln(1+|s|)$$





$$f(s) = s$$

Klasické umělé neuronové sítě

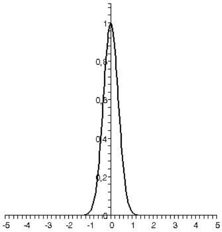
Dvouvrstvé

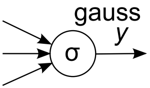
vstup $\rightarrow \mathrm{sign}$	bipolární perceptron
vstup $\rightarrow \mathrm{lin}$	lineární síť
vstup $\rightarrow \mathrm{tgh}$	hladký perceptron

Třívrstvé

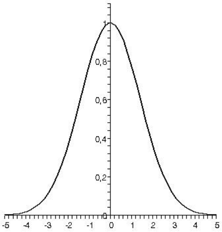
vstup $\rightarrow \mathrm{sign} \rightarrow \mathrm{sign}$	síť bipolárních perceptronů
vstup $\rightarrow \mathrm{tgh} \rightarrow \mathrm{tgh}$	MLP – vícevrstvý perceptron (Multi Layer Perceptron)
vstup $\rightarrow \mathrm{tgh} \rightarrow \mathrm{lin}$	MLL – síť s lineárním výstupem

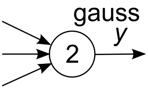
Radial Basis Function network





$$f(s) = e^{-s^2}$$
$$s = \frac{\|\overline{x} - \overline{w}\|}{\sigma}$$
$$\left(\sigma = \frac{1}{2}\right)$$



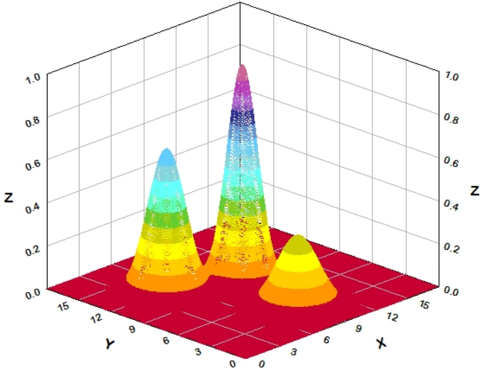


$$f(s) = e^{-s^2}$$
$$s = \frac{\|\overline{x} - \overline{w}\|}{2}$$

Třívrstvá síť

vstup $\rightarrow$ gauss $\rightarrow$ lin RBF – Radial Basis Function network

Radial Basis Function network



Gradientní metoda

počáteční bod $\overline{x} = (x_1, x_2)$

$$x_1^{\text{new}} = x_1 - \lambda \frac{\partial f}{\partial x_1} \qquad \lambda > 0$$
$$x_2^{\text{new}} = x_2 - \lambda \frac{\partial f}{\partial x_2}$$

nevýhody

- pro malé λ je pomalá
- pro velké λ kmitá (přeskakuje řešení)
- nemusíme najít globální minimum

obecný zápis

$$\overline{x}_0 \in \mathbb{R}^n, \lambda > 0$$
$$\overline{x}_{j+1} = \overline{x}_j - \lambda \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right)$$

Stochastická gradientní metoda

$\mathrm{f}(x) = \mathrm{f}_1(x) + \mathrm{f}_2(x)$

- minimalizace f_1 a f_2 se stejnou pravděpodobností $\frac{1}{2}$
- minimalizace průměrné hodnoty $\bar{\mathrm{f}}(x)$

$$\mathrm{f}(x) = 2\,\frac{\mathrm{f}_1(x) + \mathrm{f}_2(x)}{2} = 2\,\bar{\mathrm{f}}(x) = \min$$

obecně

$\mathrm{f}(\overline{x}) = \mathrm{f}_1(\overline{x}) + \cdots + \mathrm{f}_m(\overline{x})$

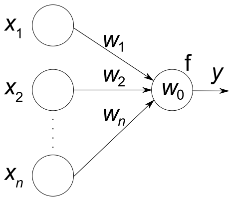
$$\mathrm{f}(\overline{x}) = \sum_{k=1}^{m} \mathrm{f}_k(\overline{x}) = m\,\bar{\mathrm{f}}(\overline{x}) = \min$$

Stochastická gradientní metoda

Algoritmus

- $\overline{x}_0 \in \mathbb{R}^n$, $\lambda_j > 0$, $\mathrm{f}(\overline{x}) = \sum_{k=1}^{m} \mathrm{f}_k(\overline{x})$
- $k \in \hat{m}$ náhodné
  $\overline{x}_{j+1} = \overline{x}_j - \lambda_j \left(\frac{\partial \mathrm{f}_k}{\partial x_1}, \ldots, \frac{\partial \mathrm{f}_k}{\partial x_n} \right)$
- $\lambda_j > 0$, $\sum_{j=0}^{\infty} \lambda_j = \infty$, $\sum_{j=0}^{\infty} \lambda_j^2 < \infty$ (Robins-Moore, 1951)
  $\Rightarrow \lambda_j = \frac{\lambda_0}{1+j}$

Delta učení



- $y = \mathrm{f}(\overline{x}, \overline{w})$
- $y_k = \mathrm{f}(\overline{x}_k, \overline{w})$
- $\triangle_k = y_k^* - y_k$
- $\mathrm{p}_k(\overline{w}) = \frac{1}{2}\triangle_k^2$
- $\mathrm{P}(\overline{w}) = \sum_{k=1}^{m} \mathrm{p}_k(\overline{w}) = \min$

$$\begin{aligned}\mathrm{P}(\overline{w}) &= \sum_{k=1}^{m} \mathrm{p}_k(\overline{w}) = \sum_{k=1}^{m} \frac{1}{2}\triangle_k^2 = \\ &= \sum_{k=1}^{m} \frac{1}{2}(y_k^* - y_k)^2 = \sum_{k=1}^{m} \frac{1}{2}(y_k^* - \mathrm{f}(\overline{x}_k, \overline{w}))^2 = \min\end{aligned}$$

Delta učení

$$\mathrm{P}(\overline{w}) = \sum_{k=1}^{m} \mathrm{p}_k(\overline{w}) = \min$$

- $\overline{w}_0 \in \mathbb{R}^{n+1}$, $\lambda_j > 0$
- $k \in \hat{m}$ náhodné

  $\overline{w}_{j+1} = \overline{w}_j - \lambda_j \left(\frac{\partial \mathrm{p}_k}{\partial w_0}, \ldots, \frac{\partial \mathrm{p}_k}{\partial w_n} \right)$

$$\frac{\partial \mathrm{p}_k}{\partial w_i} = \frac{1}{2} \cdot 2\,(y_k^* - \mathrm{f}(\overline{x}_k, \overline{w})) \cdot (-1) \frac{\partial \mathrm{f}(\overline{x}_k, \overline{w})}{\partial w_i} = -\triangle_k \frac{\partial \mathrm{f}}{\partial w_i}, \quad i = 0, 1, \ldots, n$$

- $\overline{w}_{j+1} = \overline{w}_j + \lambda_j \triangle_k \left(\frac{\partial \mathrm{f}_k}{\partial w_0}, \ldots, \frac{\partial \mathrm{f}_k}{\partial w_n} \right)$

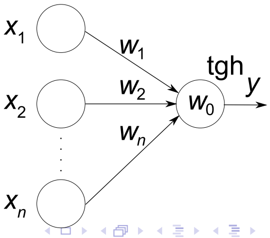
Delta učení

Algoritmus

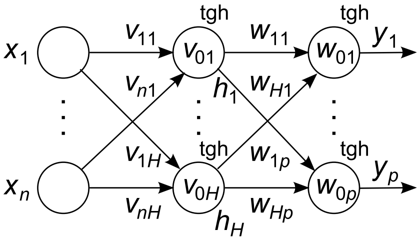
- $\overline{w}_0 \in \mathbb{R}^{n+1}$ (malé), $\lambda_0 > 0$
- náhodná volba k. vzoru, $\lambda_j = \frac{\lambda_0}{1+j}$
- výpočet y_k a $\triangle_k$
- výpočet citlivosti sítě na váhy
- adaptace vah

Učení hladkého perceptronu

- $y = \mathrm{f}(\overline{x}, \overline{w}) = \operatorname{tgh}\left(\sum_{i=0}^{n} w_i x_i\right)$
- $\frac{\partial \operatorname{tgh}\left(\sum_{i=0}^{n} w_i x_i\right)}{\partial w_i} = (1 - y^2) x_i$
- $\overline{w}_{j+1} = \overline{w}_j + \lambda_j \triangle_k (1 - y_k^2)\, \overline{x}_k$



Backpropagation



- vezmeme náhodně vzor $\overline{p} = (\overline{x}, \overline{y}^*)$
- pro výstupy y_s, $s = 1, \ldots, p$:

$$\overline{w}^{\text{new}} = \overline{w} + \lambda \underbrace{\triangle_s(1 - y_s^2)}_{d_s} \overline{h}, \quad \triangle_s = y_s^* - y_s, \quad (h_0 = 1)$$

- pro skryté neurony h_r, $r = 1, \ldots, H$:

$$\overline{v}^{\text{new}} = \overline{v} + \lambda \, \delta_r (1 - h_r^2) \, \overline{x}, \quad \delta_r = \sum_{i=1}^{p} d_i \, w_{ri}$$