MLP architektura

Srpen 2025

V tomto textu je rozebrána architektura MLP (Multi-Layer Perceptron) neuronové sítě.

1 Neuron a architektura

Na obrázku níže je uveden příklad architektury neuronové sítě. Síť má vstupní vrstvu se 2 neurony, dvě skryté vrstvy se 3 a 2 neurony a výstupní vrstvu se 2 neurony.

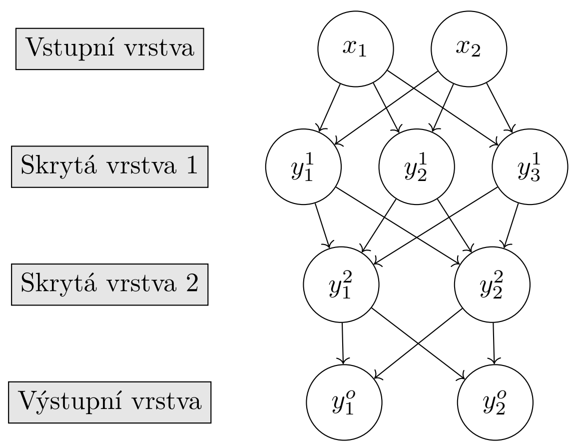


Figure 1: Příklad MLP architektury neuronové sítě.

Vstupní vrstva obsahuje vektor vytvořený z tréninkových dat:

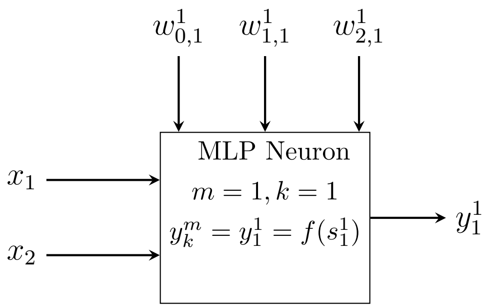
$$\mathbf{X} = (x_0, x_1, x_2, \ldots, x_n)^T \tag{1}$$

kde hodnoty x_i jsou normalizovány na interval $< 0, 1 >$.

Hodnoty neuronů ve vstupní vrstvě jsou:

$$\mathbf{Y^0} = \mathbf{X} \tag{2}$$

Ve skrytých vrstvách mají neurony hodnoty označené jako $\mathbf{Y^k}$. Každý neuron má přiřazené váhy w_i a bias w_0. Následuje ilustrace neuronu.



$w_{j,k}^m$ značí váhu spojení od neuronu j z vrstvy $m-1$ k neuronu k z vrstvy m. Neurony jsou indexovány od 1 a $w_{0,k}^m$ značí bias neuronu k ve vrstvě m. s_k^m je vnitřní potenciál neuronu k ve vrstvě m.

2 Forward propagace

Vnitřní potenciál neuronu se určí pomocí aktivační funkce $f(s)$, kde platí

$$s_k^m = \sum_j w_{jk}^m y_j^{m-1} \tag{3}$$

kde j je index vstupů z předchozí vrstvy $m-1$ a k je index neuronu ve vrstvě m. Hodnotu y_0^{m-1} od neexistujícího neuronu 0 lze definovat jako 1 a tím rovnice platí i pro biasy.

Aktivační funkce $f(s)$ může být definována různým způsobem, například lze použít RELU funkci

$$y_k^m = f(s_k^m) = \max(0, s_k^m) \tag{4}$$

RELU funkce umožňuje vznik hodnot vyšších než 1. Pokud je vyžadováno, aby výstupy y_j nabývaly pouze hodnot od 0 do 1, lze použít jinou aktivační funkci pro výstupní vrstvu.

3 Handling výstupních hodnot

Při aplikaci na rozpoznávání čislic od 0 do 9 je možné použít převod číslice na 10 skutečných výstupů y_k^{skut} následujícím způsobem:

$\mathbf{Y}^{\text{skut}}$(číslice)	y_0^{skut}	y_1^{skut}	y_2^{skut}	y_3^{skut}	y_4^{skut}	y_5^{skut}	y_6^{skut}	y_7^{skut}	y_8^{skut}	y_9^{skut}
$\mathbf{Y}^{\text{skut}}(0)$	1	0	0	0	0	0	0	0	0	0
$\mathbf{Y}^{\text{skut}}(1)$	0	1	0	0	0	0	0	0	0	0
$\mathbf{Y}^{\text{skut}}(2)$	0	0	1	0	0	0	0	0	0	0
$\mathbf{Y}^{\text{skut}}(3)$	0	0	0	1	0	0	0	0	0	0
$\mathbf{Y}^{\text{skut}}(4)$	0	0	0	0	1	0	0	0	0	0
$\mathbf{Y}^{\text{skut}}(5)$	0	0	0	0	0	1	0	0	0	0
$\mathbf{Y}^{\text{skut}}(6)$	0	0	0	0	0	0	1	0	0	0
$\mathbf{Y}^{\text{skut}}(7)$	0	0	0	0	0	0	0	1	0	0
$\mathbf{Y}^{\text{skut}}(8)$	0	0	0	0	0	0	0	0	1	0
$\mathbf{Y}^{\text{skut}}(9)$	0	0	0	0	0	0	0	0	0	1

Například číslici 3 tedy odpovídá vektor skutečných výstupů:

$$\mathbf{Y}^{\text{skut}}(3) = (0, 0, 0, 1, 0, 0, 0, 0, 0, 0)^T \tag{5}$$

Opačný převod 10 výstupů y_i na číslici lze uskutečnit hledáním maximálního výstupu:

$$\text{číslice} = \text{argmax}(\mathbf{Y}) \tag{6}$$

4 Zpětná propagace

Při zpětné propagaci dochází k opačnému průchodu přes síť a k úpravě vah na základě chyby, které se síť při učení dopustila. Chyba se šíří zpět pomocí funkce energie, označené jako E. Pro ukázku použijeme funkci střední kvadratické chyby (MSE):

$$E = \sum_k \frac{1}{2}(y_k - d_k)^2 \tag{7}$$

kde $d_k = y_k^{skut}$ (desired).

Derivace energie vzhledem k vahám v libovolné vrstvě je (řetězové pravidlo):

$$\frac{\partial E}{\partial w_{jk}} = \frac{\partial E}{\partial s_k}\frac{\partial s_k}{\partial w_{jk}} \tag{8}$$

Ve výstupní vrstvě lze dále rozepsat:

$$\frac{\partial E}{\partial w_{jk}} = \frac{\partial E}{\partial y_k^o}\frac{\partial y_k^o}{\partial s_k^o}\frac{\partial s_k^o}{\partial w_{jk}} = \frac{\partial E}{\partial w_{jk}} = -(d_k^o - y_k^o)f'(s_k^o)y_j^{o-1} = \delta^0 y_j^{o-1} \tag{9}$$

Úprava nastavení vah pro výstupní vrstvu směřuje proti gradientu energie, tedy pro výstupní vrstvu platí

$$\Delta w_{jk}^o = \eta(d_k^o - y_k^o)f'(s_k^o)y_j = \eta\delta_k^o y_j^{o-1} \tag{10}$$

kde η je učící rychlost (learning rate). Bez odvození uvedeme, že pro úpravu vah v dalších vrstvách platí obecný vztah

$$\Delta w_{jk}^m = \eta\delta_k^m y_j^{m-1} \tag{11}$$

kde rekurzivně

$$\delta_k^m = \left(\sum_{l=0}^{N_{m+1}} \delta_l^{m+1} w_{kl}^{m+1}\right) f'(s_k^m) \tag{12}$$

a

$$\delta_k^o = (d_k^o - y_k^o)f'(s_k^o) \tag{13}$$