

Date

Rețele neuronale: Tema 1

$$X = [1, 3, 0]$$

$$b = [0, 1, 0, 1]$$

$$W = \begin{bmatrix} 0,3 & 0,1 & -2 \\ -0,6 & -0,5 & 2 \\ -1 & -0,5 & 0,1 \end{bmatrix}$$

$$Y = [0, 1, 0]$$

$$Z = W^T X + b$$

$$z_1 = (1 \cdot 0,3) + (3 \cdot 0,1) + (0 \cdot -2) + 0,1 = 0,7$$

$$z_2 = (-0,6 \cdot 1) + (-0,5 \cdot 3) + (2 \cdot 0) + 0,1 = -2$$

$$z_3 = (-1 \cdot 1) + (-0,5 \cdot 3) + (0,1 \cdot 0) + 0,1 = -2,4$$

$$Z = [0,7, -2, -2,4]$$

$$\hat{Y} = \text{softmax}(Z)$$

$$\begin{aligned} P_1 &= \exp(0,7) : (\exp(0,7) + \exp(-2) + \exp(-2,4)) \\ &= 2,013 : (2,013 + 0,135 + 0,090) \\ &= 2,013 : 2,238 = \underline{0,899} \end{aligned}$$

$$\hat{Y} = [0,899, 0,06, 0,04]$$

$$\begin{aligned} P_2 &= \exp(-2) : (\exp(0,7) + \exp(-2) + \exp(-2,4)) \\ &= 0,135 : 2,238 = \underline{0,06} \end{aligned}$$

$$\begin{aligned} P_3 &= \exp(-2,4) : (\exp(0,7) + \exp(-2) + \exp(-2,4)) \\ &= 0,09 : 2,238 = \underline{0,04} \end{aligned}$$

①

Date

$$\Delta_z L = \hat{Y} - Y = [0,899, -0,94, 0,04]$$

$$\Delta_w L = \nabla_z L x^T$$

$$\nabla_w L_1 = (0,899 - 0) \cdot [0,3 \quad 0,1 \quad -2] \\ = [0,269 \quad 0,089 \quad -1,798]$$

$$\nabla_w L_2 = -0,94 \cdot [-0,6 \quad -0,5 \quad 2] \\ = [0,564 \quad 0,47 \quad 1,88]$$

$$\nabla_w L_3 = 0,04 [-1 \quad -0,5 \quad 0,1] \\ = [-0,04 \quad -0,02 \quad 0,004]$$

$$\nabla_w L = \begin{bmatrix} 0,3 & 0,1 & -2 \\ 0,564 & 0,47 & 1,88 \\ -0,04 & -0,02 & 0,004 \end{bmatrix}$$

$$\nabla_z L = [0,899, -0,94, 0,04]$$

$$W \leftarrow W - \eta \nabla_w L \Rightarrow W \leftarrow \begin{bmatrix} 0,3 & 0,1 & -2 \\ -0,6 & -0,5 & 2 \\ -1 & -0,5 & 0,1 \end{bmatrix} - 0,3 \cdot \begin{bmatrix} 0,3 & 0,1 & -2 \\ 0,564 & 0,47 & 1,88 \\ -0,04 & -0,02 & 0,004 \end{bmatrix}$$

$$W \leftarrow \begin{bmatrix} 0,3 & 0,1 & -2 \\ -0,6 & -0,5 & 2 \\ -1 & -0,5 & 0,1 \end{bmatrix} - \begin{bmatrix} 0,09 & 0,03 & -0,6 \\ 0,169 & 0,141 & 0,564 \\ -0,012 & -0,006 & 0,0012 \end{bmatrix}$$

$$W \leftarrow \begin{bmatrix} -0,6 & 0,07 & -1,4 \\ 0,431 & -0,641 & 1,436 \\ -0,988 & -0,494 & 0,073 \end{bmatrix}$$

②

Date

$$h \leftarrow h - \eta \nabla_h L$$

$$h \leftarrow [0.1 \ 0.1 \ 0.1] - 0.3 \cdot [0.899 \ -0.194 \ 0.04]$$

$$h \leftarrow [0.1 \ 0.1 \ 0.1] - [0.269 \ 0.282 \ 0.012]$$

$$h \leftarrow [-0.169 \ -0.182 \ 0.088]$$

Observation: a lower learning rate (η) means a more accurate result but a longer training time when it comes to iterations.

③