

Homework 1 RN

Arserne Andreea

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

$$W = \begin{bmatrix} 0.3 & 0.1 & -2 \\ -0.6 & -0.5 & 2 \\ -1 & -0.5 & 0.1 \end{bmatrix}$$

$$b = [0.1, 0.1, 0.1]$$

$$y = [0, 1, 0]$$

linear combinations z_i

$$z = W^T x + b = \begin{bmatrix} 0.3 & -0.6 & -1 \\ 0.1 & -0.5 & -0.5 \\ -2 & 2 & 0.1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix} + \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix}$$

$$z_0 = [0.3 \ -0.6 \ -1][1 \ 3 \ 0] + 0.1 = 0.3 - 1.8 + 0.1 = -1.4$$

$$z_1 = [0.1 \ -0.5 \ -0.5][1 \ 3 \ 0] + 0.1 = 0.1 - 1.5 + 0.1 = -1.3$$

$$z_2 = [-2 \ 2 \ 0.1][1 \ 3 \ 0] + 0.1 = -2 + 6 + 0.1 = 4.1$$

$$\hat{y} = \text{softmax}(z) \quad \hat{y}_i = \frac{e^{z_i}}{\sum_j e^{z_j}} \quad i=0 \dots m \quad \sum_j e^{z_j} = e^{-1.4} + e^{-1.3} + e^{4.1} \\ = 0.24 + 0.27 + 60.34 \\ = 60.85$$

$$\hat{y}_0 = \frac{0.24}{60.85} = 0.003$$

$$\hat{y}_1 = \frac{0.27}{60.85} = 0.004$$

$$\hat{y}_2 = \frac{60.34}{60.85} = 0.991$$

gradient cross-entropy loss $\nabla_z L = \hat{y} - y = [0.003 - 0 \ 0.004 - 1 \ 0.991 - 0]$
 $= [0.003 \ -0.996 \ 0.991]$

$$\nabla_W L = \nabla_z L^T \cdot x = \begin{bmatrix} 0.003 \\ -0.996 \\ 0.991 \end{bmatrix} \begin{bmatrix} 1 & 3 & 0 \end{bmatrix} = \begin{bmatrix} 0.003 \cdot 1 & 0.003 \cdot 3 & 0 \\ -0.996 \cdot 1 & -0.996 \cdot 3 & 0 \\ 0.991 \cdot 1 & 0.991 \cdot 3 & 0 \end{bmatrix} \\ = \begin{bmatrix} 0.003 & 0.009 & 0 \\ -0.996 & -2.988 & 0 \\ 0.991 & 2.973 & 0 \end{bmatrix}$$

$$\nabla_b L = \nabla_z L = [0.003 \ -0.996 \ 0.991]$$

step $\eta = 0.1$

$$W \leftarrow W - \eta \nabla_W L = W - 0.1 \nabla_W L = \begin{bmatrix} 0.3 & 0.1 & -2 \\ -0.6 & -0.5 & 2 \\ -1 & -0.5 & 0.1 \end{bmatrix} - \begin{bmatrix} 0.0003 & 0.0009 & 0 \\ -0.099 & -0.298 & 0 \\ 0.0991 & 0.297 & 0 \end{bmatrix} \\ = \begin{bmatrix} 0.299 & 0.099 & -2 \\ -0.501 & -0.202 & 2 \\ -1.099 & -0.797 & 0.1 \end{bmatrix}$$

$$b \leftarrow b - \eta \nabla_b L = [0.1 \ 0.1 \ 0.1] - 0.1 [0.003 \ -0.996 \ 0.991] = b - [0.0003 \ -0.0996 \ 0.0991] \\ = [0.0997 \ 0.1996 \ 0.0009]$$