2. Exercise

$X = [1, 3, 0]$

$$W = \begin{bmatrix} 0.3 & 0.1 & -2 \\ -0.6 & -0.5 & 2 \\ -1 & -0.5 & 0.1 \end{bmatrix} \begin{matrix} -w_1 \\ -w_2 \\ -w_3 \end{matrix}$$

$b = [0.1, 0.1, 0.1]$

$y = [0, 1, 0]$

$\underline{z} = [w_1 \cdot X + b_1,\ w_2 \cdot X + b_2,\ w_3 \cdot X + b_3]$

$w_1 \cdot X + b_1 = 1 \cdot 0.3 + 3 \cdot 0.1 + 0 \cdot (-2) + 0.1 = 0.3 + 0.3 + 0.1 = 0.7$

$w_2 \cdot X + b_2 = 1 \cdot (-0.6) + 3 \cdot (-0.5) + 0 \cdot 2 + 0.1 = -2$

$w_3 \cdot X + b_3 = 1 \cdot (-1) + 3 \cdot (-0.5) + 0 \cdot 0.1 + 0.1 = -2.4$

$\underline{z} = [0.7, -2, -2.4]$

$$\hat{y}_i = \frac{\exp(z_i)}{\sum_j \exp(z_j)}$$

$$\hat{y}_1 = \frac{e^{0.7}}{e^{0.7} + e^{-2} + e^{-2.4}} \simeq \frac{2.0137}{2.2398} \simeq 0.899$$

$$\hat{y}_2 = \frac{e^{-2}}{e^{0.7} + e^{-2} + e^{-2.4}} \simeq \frac{0.1353}{2.2398} \simeq 0.06$$

$$\hat{y}_3 = \frac{e^{-2.4}}{2.2398} \simeq \frac{0.09}{2.2398} \simeq 0.04$$

Gradients

$$L = \frac{1}{2}(\hat{y} - y)^2$$

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial z}{\partial w}$$

$$\frac{\partial L}{\partial \hat{y}} = \hat{y} - y$$

$$\frac{\partial \hat{y}}{\partial z} = \hat{y}(1 - \hat{y})$$

$$\frac{\partial z}{\partial w} = X$$

$$\frac{\partial L}{\partial w} = (\hat{y} - y) \cdot \hat{y}(1 - \hat{y}) \cdot X$$

$$\frac{\partial L}{\partial w_1} = (\hat{y}_1 - y) \cdot \hat{y}_1(1 - \hat{y}_1) \cdot X = (0.899 - 0) \cdot (0.899 - 0.899^2) \cdot [1, 3, 0] = 0.899 \cdot 0.09 \cdot [1, 3, 0] = 0.08 \cdot [1, 3, 0]$$

$$= \underline{[0.08, 0.24, 0]}$$

$$\frac{\partial L}{\partial w_2} = (\hat{y}_2 - y_2) \cdot (\hat{y}_2 - \hat{y}_2^2) \cdot X = -0.94 \cdot 0.05 \cdot [1, 3, 0] = -0,04 \cdot [1, 3, 0] = \underline{[-0.04, -0.12, 0]}$$

$$\frac{\partial L}{\partial w_3} = (\hat{y}_3 - y_3) \cdot (\hat{y}_3 - \hat{y}_3^2) \cdot X = 0.04 \cdot 0.03 \cdot [1, 3, 0] =$$

$= 0.001 \cdot [1,3,0] \simeq [0.001, 0.003, 0]$

$\frac{\partial L}{\partial b} \simeq \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial z}{\partial b} = (\hat{y} - y) \cdot \hat{y}(1-\hat{y}) \cdot 1$

$\frac{\partial L}{\partial b} \simeq [0.08, -0.04, 0.001]$

updates $\eta = 0.1$

$W' \leftarrow W - \eta \nabla_W L$

$b' \leftarrow b - \eta \nabla_b L$

$W_1' \leftarrow W_1 - \eta \nabla_{W_1} L \simeq [0.3, 0.1, -2] - 0.1 \cdot [0.08, 0.24, 0] \simeq [0.22, 0.076, -2]$

$W_2' \leftarrow W_2 - \eta \nabla_{W_2} L \simeq [-0.6, -0.5, 2] - 0.1 \cdot [-0.04, -0.12, 0] \simeq [-0.56, -0.38, 2]$

$W_3' \leftarrow W_3 - \eta \nabla_{W_3} L \simeq [-1, -0.5, 0.1] - 0.1 \cdot [0.001, 0.003, 0] \simeq [-1.0001, -0.5003, 0.1]$

$b' = [0.1, 0.1, 0.1] - 0.1 \cdot [0.08, -0.04, 0.001]$

$= [0.092, 0.104, 0.0999]$