Lungu Andrei - Sebastian

Begin by computing the linear combinations z for each class

$z_1 = W_{11} \cdot x_1 + W_{12} \cdot x_2 + W_{13} \cdot x_3 + b_1$

$z_1 = 0.3 \cdot 1 + 0.1 \cdot 3 + (-2) \cdot 0 + 0.1$

$z_1 = 0.3 + 0.3 + 0.1 = \underline{\underline{0.7}}$

$z_2 = W_{21} \cdot x_1 + W_{22} \cdot x_2 + W_{23} \cdot x_3 + b_2$

$z_2 = -0.6 \cdot 1 + (-0.5) \cdot 3 + 2 \cdot 0 + 0.1$

$z_2 = -0.6 - 1,5 + 0.1 = \underline{\underline{-2.0}}$

$z_3 = W_{31} \cdot x_1 + W_{32} \cdot x_2 + W_{33} \cdot x_3 + b_3$

$z_3 = -1 \cdot 1 + (-0.5) \cdot 3 + 0.1 \cdot 0 + 0.1$

$z_3 = -1 - 1.5 + 0.1 = \underline{\underline{-2.4}}$

$z = [0.7, -2.0, -2.4]$

Apply the softmax function to get the predicted probabilities $\hat{y}$:

$\hat{y} = \text{softmax}(z)$

~~$\hat{y} =$ $\hat{y}_1 = \frac{e^{0.7}}{}$~~

$\text{softmax}(0.7) = \frac{e^{0.7}}{e^{0.7} + e^{-2.0} + e^{-2.4}} = \frac{2.013}{2.013 + 0.135 + 0.090}$

$= \frac{2.013}{2.238} = \underline{\underline{0.899}}$

$\text{softmax}(-2.0) = \frac{e^{-2}}{e^{0.7} + e^{-2.0} + e^{-2.4}} = \frac{0.135}{2.238}$

$= \underline{\underline{0.060}}$

$$\text{softmax}(-2.4) = \frac{e^{-2.4}}{e^{0.7} + e^{-2} + e^{-2.4}} = \frac{0.090}{2.238} = \underline{\underline{0.040}}$$

$\cancel{\hat{y} = [0.899]}$

$\hat{y} = [0.899, 0.060, 0.040]$

Compute the gradient of the loss with respect to z using cross-entropy loss and the true labels y:

$\nabla_z L = \hat{y} - y$ $\qquad \frac{dL}{dz}$

$\cancel{L(\hat{y}, y)}$

$\cancel{L(0.899, 0) = -(0 \log(0.899) + (1-0)\log(1-0.899))}$

$\cancel{\nabla_z L = [0.899, 0.060, 0.040] - [0, 1, 0]}$

$\cancel{= [0.899,}$

Cross-entropy loss.

$L(0.899, 0) = -(0 \log(0.899) + (1-0)\log(1-0.899))$

$= -\log(0.101) = + \underline{\underline{2.29}}$

$L(0.060, 1) = -(1 \log(0.060) + (1-1)\log(1-0.060))$

$= -\log(0.060) = \underline{\underline{2.81}}$

$\cancel{L(0.040, 0) = -(0 \log(0.040) + (1-0)\log(1-0.))}$

$\cancel{= -\log(1)}$

$=$

$L(0.040, 0) = -(0 \log(0.040) + (1-0)\log(1-0.040))$

$= -\log(0.96) = \underline{\underline{0.040}}$

$$\nabla_z L = \hat{y} - y = [0.899, 0.060, 0.050] - [0, 1, 0]$$

$$\nabla_z L = [0.899, -0.94, 0.050]$$

Now, compute the gradients with respect to the weights w and biases b:

$$\nabla_{w_1} L = \nabla_z L \, x$$

$$\nabla_{w_1} L = [0.899, -0.94, 0.050] \cdot [1, 3, 0]$$

$$= [0.899, \cancel{-2.28} -2{,}82, 0]$$

$$\nabla b L = \nabla_z L = [0.899, -0.94, 0.050]$$

Finally, update the weights and biases using a learning rate

$\cancel{b \leftarrow b - \eta \nabla_b L}$

$\cancel{\eta \nabla_b L = 0.1 \cdot [}$

Finally, update the weights and biases using a learning rate η:

$$\eta = 0.1$$

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

$$\eta \cdot \nabla_W L = 0.1 \cdot [0.899, -2.82, 0]$$

$$\eta \cdot \nabla_W L = [0.0899, -0.282, 0]$$

$$W \leftarrow \begin{bmatrix} 0.3 & 0.1 & -2 \\ -0.6 & -0.5 & 2 \\ -1 & -0.5 & 0.1 \end{bmatrix} - [0.0899, -0.282, 0]$$

$$W \leftarrow \begin{bmatrix} 0.2101 & 0.382 & -2 \\ -0.6899 & -0.218 & 2 \\ -1.0899 & -0.218 & 0.1 \end{bmatrix}$$

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

$$\eta \nabla_b L \leftarrow 0.1 \cdot [0.899, -0.94, 0.040]$$

$$\eta \nabla_b L \leftarrow [0.0899, -0.094, 0.004]$$

$$b \leftarrow [0.1, 0.1, 0.1] - [0.0899, -0.094, 0.004]$$

$$b \leftarrow [0.0101, 0.194, 0.096]$$