$x = [1, 3, 0]$ (Even if this writing mode is used, the input vector x should be treated as a column vector)

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

$b = [0.1, 0.1, 0.1]$
(Same mention as for x)
$y = [0, 1, 0]$

1) Begin by computing the linear combinations z for each class:

$$z = W^T x + b$$

$$z = \begin{bmatrix} 0.3 & -0.6 & -1 \\ 0.1 & -0.5 & -0.5 \\ -2 & 2 & 0.1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix} + \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} = \begin{bmatrix} (0.3\times1)+(-0.6\times3)+(-1\times0) \\ (0.1\times1)+(-0.5\times3)+(-0.5\times0) \\ (-2\times1)+(2\times3)+(0.1\times0) \end{bmatrix} + \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix}$$

$$\begin{bmatrix} 0.3-1.8 \\ 0.1-1.5 \\ -2+6 \end{bmatrix} + \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} = \begin{bmatrix} -1.5 \\ -1.4 \\ +4 \end{bmatrix} + \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} = \begin{bmatrix} -1.4 \\ -1.3 \\ 4.1 \end{bmatrix}$$

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

$$\hat{y}_1 = \text{softmax}(z_1) = \frac{\exp(z_1)}{\sum_{j=1}^{3} \exp(z_j)} = \frac{\exp(-1.4)}{\exp(-1.4)+\exp(-1.3)+\exp(4.1)} = \frac{0.25}{0.25+0.27+60.34}$$

$$= \frac{0.25}{60.86} \cong 0.0041$$

$$\hat{y}_2 = \text{softmax}(z_2) = \frac{\exp(z_2)}{\sum_{j=1}^{3} \exp(z_j)} = \frac{\exp(-1.3)}{\exp(-1.4)+\exp(-1.3)+\exp(4.1)} = \frac{0.27}{0.25+0.27+60.34}$$

$$= \frac{0.27}{60.86} \cong 0.0044$$

$$\hat{y}_3 = \text{softmax}(z_3) = \frac{\exp(z_3)}{\sum_{j=1}^{3} \exp(z_j)} = \frac{\exp(4.1)}{\exp(-1.4)+\exp(-1.3)+\exp(4.1)} = \frac{60.34}{0.25+0.27+60.34}$$

$$= \frac{60.34}{60.86} \cong 0.9915$$

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

$$\nabla_z L = \hat{y} - y$$

$$\nabla_z L = \begin{bmatrix} 0.0041 \\ 0.0044 \\ 0.9915 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.0041 \\ -0.9956 \\ 0.9915 \end{bmatrix}$$

4) Now, compute the gradients of the loss with respect to weights W and biases b:

$$\nabla_W L = \nabla_z L \, x^T$$

$$\nabla_b L = \nabla_z L$$

$$\nabla_W L = \begin{bmatrix} 0.0041 \\ -0.9956 \\ 0.9915 \end{bmatrix} \begin{bmatrix} 1 & 3 & 0 \end{bmatrix}$$, but each weight should be considered independently

$$\nabla_{W_1} L = 0.0041 \times \begin{bmatrix} 1 & 3 & 0 \end{bmatrix} = \begin{bmatrix} 0.0041 & 0.0123 & 0 \end{bmatrix}$$

$$\nabla_{W_2} L = -0.9956 \times \begin{bmatrix} 1 & 3 & 0 \end{bmatrix} = \begin{bmatrix} -0.9956 & -2.9868 & 0 \end{bmatrix}$$

$$\nabla_{W_3} L = 0.9915 \times \begin{bmatrix} 1 & 3 & 0 \end{bmatrix} = \begin{bmatrix} 0.9915 & 2.9745 & 0 \end{bmatrix}$$

$$\Rightarrow \nabla_W L = \begin{bmatrix} 0.0041 & 0.0123 & 0 \\ -0.9956 & -2.9868 & 0 \\ 0.9915 & 2.9745 & 0 \end{bmatrix}$$

$$\nabla_b L = \nabla_z L = \begin{bmatrix} 0.0041 \\ -0.9956 \\ 0.9915 \end{bmatrix}$$

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

Let's consider $\eta = 0.1$

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

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

$$W = \begin{bmatrix} 0.3 & 0.1 & -2 \\ -0.6 & -0.5 & 2 \\ -1 & -0.5 & 0.1 \end{bmatrix} - 0.1 \begin{bmatrix} 0.0041 & 0.0123 & 0 \\ -0.9956 & -2.9868 & 0 \\ 0.9915 & 2.9745 & 0 \end{bmatrix}$$

$$W = \begin{bmatrix} 0.3 & 0.1 & -2 \\ -0.6 & -0.5 & 2 \\ -1 & -0.5 & 0.1 \end{bmatrix} - \begin{bmatrix} 0.00041 & 0.00123 & 0 \\ -0.09956 & -0.29868 & 0 \\ 0.09915 & 0.29745 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0.29956 & 0.09877 & -2 \\ -0.50044 & -0.20132 & 2 \\ -1.09915 & -0.79745 & 0.1 \end{bmatrix} \cong \begin{bmatrix} 0.3 & 0.1 & -2 \\ -0.5 & -0.2 & 2 \\ -1.1 & -0.8 & 0.1 \end{bmatrix}$$

$$b = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} - 0.1 \begin{bmatrix} 0.0041 \\ -0.9956 \\ 0.9915 \end{bmatrix} = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} - \begin{bmatrix} 0.00041 \\ -0.09956 \\ 0.09915 \end{bmatrix} = \begin{bmatrix} 0.09959 \\ 0.19956 \\ 0.00085 \end{bmatrix}$$