$$\mathbf{x} = [1, 3, 0],$$

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

$$\mathbf{b} = [0.1, 0.1, 0.1],$$

$$\mathbf{y} = [0, 1, 0].$$

$\eta = 0.02$

Compute the gradient of the loss with respect to $\mathbf{z}$ using the cross-entropy loss and the true labels

$$\nabla_{\mathbf{z}} L = \hat{\mathbf{y}} - \mathbf{y}.$$

Now, compute the gradients with respect to the weights $\mathbf{W}$ and biases $\mathbf{b}$:

$$\nabla_{\mathbf{W}} L = \nabla_{\mathbf{z}} L \mathbf{x}^{\top},$$

$$\nabla_{\mathbf{b}} L = \nabla_{\mathbf{z}} L.$$

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

$$\mathbf{W} \leftarrow \mathbf{W} - \eta \nabla_{\mathbf{W}} L,$$

$$\mathbf{b} \leftarrow \mathbf{b} - \eta \nabla_{\mathbf{b}} L.$$

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

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

2) $\hat{y} = \text{softmax}(z) = \left[\frac{e^{-1.4}}{e^{-1.4} + e^{-1.3} + e^{4.1}} ; \frac{e^{-1.3}}{''} ; \frac{e^{4.1}}{''} \right]$

$= [0.004 ; 0.0044 ; 0.9914]$

3) $\nabla_z L = \begin{bmatrix} 0.004 \\ 0.0044 \\ 0.9914 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.004 \\ -0.9956 \\ 0.9914 \end{bmatrix}$

4) $\nabla_W L = \nabla_z L x^T = \begin{bmatrix} 0.004 \\ -0.9956 \\ 0.9914 \end{bmatrix} \cdot [1, 3, 0] = \begin{bmatrix} 0.004 & 0.012 & 0 \\ -0.9956 & -2.986 & 0 \\ 0.9914 & 2.974 & 0 \end{bmatrix}$

5) $\nabla_b L = \nabla_z L = \begin{bmatrix} 0.004 \\ -0.9956 \\ 0.9911 \end{bmatrix}$

6) $W' = W - \eta \nabla_W L = \begin{bmatrix} 0.3 & 0.1 & -2 \\ -0.6 & -0.5 & 2 \\ -1 & -0.5 & 0.1 \end{bmatrix} - 0.2 \cdot \begin{bmatrix} 0.004 & 0.012 & 0 \\ -0.9956 & -2.986 & 0 \\ 0.9914 & 2.974 & 0 \end{bmatrix}$

$= \begin{bmatrix} 0.2992 & 0.0976 & -2 \\ 0.3956 & 0.0972 & 2 \\ -1.1982 & -1.0548 & 0.1 \end{bmatrix} = W'$

7) $b' = b - \eta \nabla_b L = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} - 0.2 \cdot \begin{bmatrix} 0.004 \\ -0.9956 \\ 0.9911 \end{bmatrix} = \begin{bmatrix} 0.0992 \\ 0.2991 \\ -0.0982 \end{bmatrix}$