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

$$b = \begin{bmatrix} 0.1 & 0.1 & 0.1 \end{bmatrix}$$

$$y = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \quad x = \begin{bmatrix} 1, 3, 0 \end{bmatrix}$$

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

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

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

$$y_1 = \frac{e^{z_1}}{e^{z_1} + e^{z_2} + e^{z_3}} = \frac{e^{-1.4}}{e^{-1.4} + e^{-1.3} + e^{4.1}} = \frac{0.2465}{0.2465 + 0.2725 + 60.3402}$$

$$y_1 = \frac{0.2465}{60.8594} \simeq 0.0040$$

$$y_2 = \frac{0.2725}{60.8594} \simeq 0.0045 \Rightarrow \hat{y} = \begin{bmatrix} 0.0040 \\ 0.0045 \\ 0.9915 \end{bmatrix}$$

$$y_3 = \frac{60.3402}{60.8594} \simeq 0.9915$$

$$\nabla_z L = \begin{bmatrix} 0.0040 \\ 0.0045 \\ 0.9915 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.0040 \\ -0.9955 \\ 0.9915 \end{bmatrix}$$

$$\nabla_W L = \nabla_z L \cdot x^T = \begin{bmatrix} 0.0040 \\ -0.9955 \\ 0.9915 \end{bmatrix} \cdot \begin{bmatrix} 1 & 3 & 0 \end{bmatrix} =$$

$$= \begin{bmatrix} 0.0040 & 0.0040 \cdot 3 & 0 \\ -0.9955 & -0.9955 \cdot 3 & 0 \\ 0.9915 & 0.9915 \cdot 3 & 0 \end{bmatrix} = \begin{bmatrix} 0.0040 & 0.0120 & 0 \\ -0.9955 & -2.9865 & 0 \\ 0.9915 & 2.9745 & 0 \end{bmatrix}$$

$$\nabla_b L = \nabla_z L = \begin{bmatrix} 0.0040 \\ -0.9955 \\ 0.9915 \end{bmatrix}$$

$$W_a = W_i - \eta \nabla_W L \quad \text{considerăm } \eta = 0.1$$

$$W_a = \begin{bmatrix} 0.3 & 0.1 & -2 \\ -0.6 & -0.5 & 2 \\ -1 & -0.5 & 0.1 \end{bmatrix} - 0.1 \begin{bmatrix} 0.0040 & 0.0120 & 0 \\ -0.9955 & -2.9865 & 0 \\ 0.9915 & 2.9745 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0.3 - 0.0004 & 0.1 - 0.0012 & -2 \\ -0.6 + 0.0995 & -0.5 + 0.2986 & 2 \\ -1 - 0.0991 & -0.5 - 0.2974 & 0.1 \end{bmatrix} =$$

$$= \begin{bmatrix} 0.2996 & 0.0988 & -2 \\ -0.5005 & -0.2014 & 2 \\ -1.0991 & -0.7974 & 0.1 \end{bmatrix}$$

$$b_a = b_i - \eta \nabla_b L = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} - 0.1 \begin{bmatrix} 0.0040 \\ -0.9955 \\ 0.9915 \end{bmatrix} =$$

$$= \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} - \begin{bmatrix} 0.0004 \\ -0.0995 \\ 0.0991 \end{bmatrix} = \begin{bmatrix} 0.0996 \\ 0.1995 \\ 0.0009 \end{bmatrix}$$

În funcție de $\eta$, distingem următoarele cazuri:

1) $\eta$ prea mare, are drept consecință overshootingul, loss-ul poate diverge, iar modelul nu învață sau învață încet
2) optimal: convergență eficientă spre soluție, acuratețe bună
3) $\eta$ prea mic: învățare lentă, blocare în minim local sau platou