Loss function can be written as

$$L_i = -f_{y_i} + \log \sum_j e^{f_j}$$

Let's simplify the function f

$$L_i = -w_{y_i} \cdot x_i + \log \sum_j e^{w_j x_i}$$

$$\boxed{L_i = -w_{y_i} x_i + \log\left(\sum_{j \neq y_i} e^{w_j x_i} + e^{w_{y_i} \cdot x_i} \right)}$$

<u>for correct classes.</u>

$$\nabla w_{y_i} \cdot L_i = -x_i + \frac{1}{\left(\sum_{j \neq y_i} e^{w_j x_i} + e^{w_{y_i} x_i} \right)} \times \left(0 + e^{w_{y_i} \cdot x_i} \cdot x_i \right)$$

$$\nabla w_{y_i} \cdot L_i = -x_i + \frac{x_i \cdot e^{w_{y_i} \cdot x_i}}{\sum_j e^{w_j \cdot x_i}}$$

$$\nabla w_{y_i} \cdot L_i = -x_i \left(1 - \frac{e^{w_{y_i} \cdot x_i}}{\sum_j e^{w_j x_j}} \right) = -x_i \left(1 - \frac{\text{correct\_op}}{\text{exp\_sum}} \right)$$

Now for incorrect classes:

$$\boxed{L_i = -w_{y_i}\cdot x_i + \log\left(\sum_{j\neq y_i} e^{w_j x_i} + e^{w_{y_i}\cdot x_i}\right)}$$

↳ we have this relation from the loss function

$$\nabla_{w_j} L_i = 0 + \frac{1 \times \left(e^{w_j x_i}\cdot x_i + 0\right)}{\left(\sum_{j\neq y_i} e^{w_j\cdot x_i} + e^{w_{y_i}\cdot x_i}\right)}$$

$$\boxed{\nabla_{w_j} L_i = \frac{e^{w_j\cdot x_i}\cdot x_i}{\sum_{j} e^{w_j x_i}}} \rightarrow \text{for incorrect classes}$$

NOTE: We see the sigma sign disappears when taking derivative. It happens because we are finding contribution to gradient because of incorrect class-j, not all the incorrect classes. For every incorrect class this is going to be the contribution to the gradient.