Gradient of Multi-class SVM Loss Function

Given:

$$
\begin{aligned}
f(x_i, W)_j &= W_j^T x_i \\
L_i &= \sum_{j \neq y_i} \max(0, W_j^T x_i - W_{y_i}^T x_i + \Delta)
\end{aligned}
$$

The gradient of the loss function with respect to the weight matrix W is:

$$
\begin{aligned}
\delta_{ij} &= \begin{cases} 1 & \text{if } W_j^T x_i - W_{y_i}^T x_i + \Delta > 0 \\ 0 & \text{otherwise} \end{cases} \\
\frac{\partial L_i}{\partial W_{y_i}} &= -\sum_{j \neq y_i} \delta_{ij} x_i \\
\frac{\partial L_i}{\partial W_j} &= \delta_{ij} x_i
\end{aligned}
$$

For all samples:

$$
\begin{aligned}
\frac{\partial L}{\partial W_{y_i}} &= -\sum_{i=1}^{N} \sum_{j \neq y_i} \delta_{ij} x_i \\
\frac{\partial L}{\partial W_j} &= \sum_{i=1}^{N} \delta_{ij} x_i
\end{aligned}
$$

The final gradient of the loss function with respect to W is:

$$
\begin{aligned}
\frac{\partial L}{\partial W} &= \left(\frac{\partial L}{\partial W_1}, \frac{\partial L}{\partial W_2}, \ldots, \frac{\partial L}{\partial W_C} \right) \\
\frac{\partial L}{\partial W_j} &= \sum_{i=1}^{N} \left(\delta_{ij} - \sum_{k \neq y_i} \delta_{ik} \right) x_i
\end{aligned}
$$