Computing the derivative of SVM loss

SVM loss on a single data point (i) :

$$\mathcal{L}_i = \sum_{j \neq y_i} \left[\max(0, \underbrace{w_j^T x_i}_{\text{score for the wrong class}} - \underbrace{w_{y_i}^T x_i}_{\text{score for the correct class (want this to be big enough!)}} + \Delta) \right]$$

example in the problem set)

$X \in \mathbb{R}^{n \times 3073}$

$W \in \mathbb{R}^{3073 \times 10}$

$(1 \times 1) \times (3073 \times 1)$

$\rightarrow 1 \times 3073$

$$dW = \frac{\partial \mathcal{L}}{\partial W} \in \mathbb{R}^{3073 \times 10}$$

$w_j^T x_i$

$w_j \in \mathbb{R}^{3073 \times 1}$ $\quad w_j^T \in \mathbb{R}^{1 \times 3073}$ $\quad x_i \in \mathbb{R}^{1 \times 3073}$ $\quad w_j^T x_i \in \mathbb{R}$

$$dW_i = \underbrace{\frac{\partial \mathcal{L}_i}{\partial W}}_{\text{derivative on a single data}} \begin{cases} j \neq y_i & \dfrac{\partial \mathcal{L}_i}{\partial w_j} = \mathbb{1}(w_j^T x_i - w_{y_i}^T x_i + \Delta > 0) x_i \\ y_i & \dfrac{\partial \mathcal{L}_i}{\partial w_{y_i}} = -\left(\sum_j \mathbb{1}(w_j^T x_i - w_{y_i}^T x_i + \Delta > 0)\right) x_i \end{cases}$$

$\in \mathbb{R}^{1 \times 10}$