Quiz

1. The softmax function takes an C-dimensional vector of scores $\mathbf{s}$ and produces another C-dimensional vector with real values in the range $(0, 1)$ by

$$P_i = \frac{e^{s_i}}{\sum_j e^{s_j}}, \quad \text{for} \quad i = 1, \ldots, C.$$

Sow that the derivative of softmax is given by

$$\frac{\partial P_i}{\partial s_j} = \begin{cases} P_i(1 - P_j), & i = j, \\ -P_i P_j, & i \neq j \end{cases} \tag{1}$$

2. The loss of the softmax function is given by

$$L_i = -\log P_{y_i}.$$

Show that the gradient of the softmax loss w.r.t. $\mathbf{w}_j$ and $\mathbf{w}_{y_i}$ are

$$\nabla_{\mathbf{w}_j} L_i = P_j \mathbf{x}_i, \tag{2}$$
$$\nabla_{\mathbf{w}_{y_i}} L_i = (P_j - 1)\mathbf{x}_i. \tag{3}$$

3. Please compare the similarities between (2)&(3) for Softmax Classifier and (4)&(5) for SVM Classifier in terms of their gradient structure.

$$\nabla_{\mathbf{w}_j} L_i = 1_{(\mathbf{w}_j^T \mathbf{x}_i - \mathbf{w}_{y_i}^T \mathbf{x}_i + 1 > 0)} \mathbf{x}_i, \tag{4}$$

$$\nabla_{\mathbf{w}_{y_i}} L_i = -\left(\sum_{j \neq y_i} 1_{(\mathbf{w}_j^T \mathbf{x}_i - \mathbf{w}_{y_i}^T \mathbf{x}_i + 1 > 0)} \right) \mathbf{x}_i. \tag{5}$$