# Multi Class SVM loss

$n$ = # training example

$m$ = length of each training example

$C$ = # classes

⇒ Let $x^{(i)}$ be $i^{th}$ training Example.

$$x^{(i)} \in R^{(m+1)\times 1}$$

$$x^{(i)} = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$$

and $y^{(i)}$ be it corresponding lable

$$y^{(i)} \in \{1, 2, \cdots C\}$$

⇒ Let $W \in R^{C\times(m+1)}$ be the weights for the linear classifier.

{Score} $\boxed{S^{(i)} = WX^{(i)}} \in R^{C\times 1}$

{Score given to each class for input $x^{(i)}$ given the weights are $W$.}

⇒ Let $W_j$ be $j^{th}$ row of $W$, and $S_j^{(i)}$ be the score given to $j^{th}$ class for input $X^{(i)}$

$$\boxed{S_j^{(i)} = W_j X^{(i)}} \in R$$

$\Rightarrow$ Let $L_i$ be loss for the $i^{th}$ example:

$$\boxed{L_i = \sum_{j \neq y_i} \max(0, S_j^{(i)} - S_{y^{(i)}}^{(i)} + 1)}$$

$\Rightarrow$ The total loss for all training example:

$$\boxed{L = \frac{1}{n} \sum_{i=1}^{n} L_i}$$

★ <u>Calculating Gradient</u>

$$\nabla_W L = \frac{1}{n} \sum_{i=1}^{n} \boxed{\nabla_W L_i}$$

$$\nabla_W L_i = \sum_{j \neq y^{(i)}} \boxed{\nabla_W \max(0, S_j^{(i)} - S_{y^{(i)}}^{(i)} + 1)}$$

$$\nabla_w \max\left(0,\ s_j^{(i)} - s_{y(i)}^{(i)} + 1\right) = \nabla_w L_i^{(j)}$$

\# if $s_j^{(i)} - s_{y(i)}^{(i)} + 1 \leq 0$

$$L_i^{(j)} = 0$$

$$\left(\nabla_w L_i^{(j)}\right)_{a,b} = \frac{\delta L_i^{(j)}}{\delta w_{ab}} = 0$$

$$\nabla_w L_i^{(j)} = 0 \in \mathbb{R}^{c \times (m+1)}$$

\# if $s_j^{(i)} - s_{y(i)}^{(i)} + 1 > 0$

$$L_i^{(j)} = s_j^{(i)} - s_{y(i)}^{(i)} + 1$$

$$= w_j x^{(i)} - w_{y(i)} x^{(i)} + 1$$

$$= \left(w_j - w_{y(i)}\right) x^{(i)} + 1$$

$$\nabla_w L_i^{(j)} = \begin{bmatrix} \nabla_{w_1} L_i^{(j)} \\ \nabla_{w_2} L_i^{(j)} \\ \vdots \\ \nabla_{w_c} L_i^{(j)} \end{bmatrix} \xrightarrow{k\text{th row}} \nabla_{w_k} L_i^{(j)}$$

$$\nabla_{w_k} L_i^{(j)} = \begin{cases} x^{(i)T} & \text{if } k = j \\ -x^{(i)T} & \text{if } k = y^{(i)} \\ 0 & \text{else} \end{cases}$$

$0 \rightarrow \in \mathbb{R}^{1 \times (m+1)}$

$$\nabla_W L_i^{(j)} = \begin{bmatrix} 0 \\ \vdots \\ x^{(i)T} \\ 0 \\ \vdots \\ -x^{(i)T} \\ \vdots \\ 0 \end{bmatrix} \begin{matrix} \\ \\ \rightarrow j^{th} \text{ row} \\ \\ \\ \rightarrow y^{(i)th} \text{ row} \\ \\ \\ \end{matrix}$$

# General Case

$$\nabla_W L_i^{(j)} = 1\{S_j^{(i)} - S_{y^{(i)}}^{(i)} + 1 > 0\} \begin{bmatrix} 0 \\ \vdots \\ x^{(i)T} \\ \vdots \\ -x^{(i)T} \\ \vdots \\ 0 \end{bmatrix} \begin{matrix} \\ \\ \rightarrow j^{th} \text{ row} \\ \\ \rightarrow y^{(i)th} \text{ row} \\ \\ \\ \end{matrix}$$

$$\nabla_W L_i = \sum_{j \neq y^{(i)}} \nabla_W L_i^{(j)}$$

$$\nabla_\omega L_i = \begin{bmatrix} 1\{S_1^{(i)} - S_{y^{(i)}}^{(i)} + 1 > 0\}\, X^{(i)T} \\ 1\{S_2^{(i)} - S_{y^{(i)}}^{(i)} + 1 > 0\}\, X^{(i)T} \\ \vdots \\ -\left(\sum_{j \neq y^{(i)}} 1\{S_j^{(i)} - S_{y^{(i)}}^{(i)} + 1 > 0\}\right) X^{(i)T} \\ \vdots \\ 1\{S_c^{(i)} - S_{y^{(i)}}^{(i)} + 1 > 0\}\, X^{(i)T} \end{bmatrix} \rightarrow y^{(i)\text{th}} \text{ row}$$

$$\nabla_\omega L_i = \begin{bmatrix} 1\{S_1^{(i)} - S_{y^{(i)}}^{(i)} + 1 > 0\}\, 1 \\ 1\{S_2^{(i)} - S_{y^{(i)}}^{(i)} + 1 > 0\}\, 1 \\ \vdots \\ -\left(\sum_{j \neq y^{(i)}} 1\{S_j^{(i)} - S_{y^{(i)}}^{(i)} + 1 > 0\}\right) 1 \\ \vdots \\ 1\{S_c^{(i)} - S_{y^{(i)}}^{(i)} + 1 > 0\}\, 1 \end{bmatrix} \tilde{X}^{(i)}$$

where, $1 \in \mathbb{R}^{1\times(m+1)}$

$1 = [1, 1, \cdots 1]$

$$\tilde{X}^{(i)} = \begin{bmatrix} X_1^{(i)} & & 0 \\ & X_2^{(i)} & \\ 0 & & \ddots \; X_{m+1}^{(i)} \end{bmatrix}$$

$$\boxed{\nabla_\omega L_i = M_i \tilde{X}^{(i)}}$$

$$\boxed{\nabla_\omega L = \frac{1}{n}\sum_{i=1}^{n} M_i \tilde{X}^{(i)}}$$