Hyun-Joon Yang
HW2

1.1) update step: $\vec{w}_{t+1} = \vec{w}_t - \eta \nabla F(\vec{w}_t)$

$$F(\vec{w}_t) = \|\underline{\underline{X}}\vec{w}_t - \vec{y}\|_2^2 + \lambda\|\vec{w}_t\|_2^2 = (\underline{\underline{X}}\vec{w}_t - \vec{y})^T(\underline{\underline{X}}\vec{w}_t - \vec{y}) + \lambda \vec{w}_t^T \vec{w}_t$$

$$\nabla_w F = 2\underline{\underline{X}}^T(\underline{\underline{X}}\vec{w}_t - \vec{y}) + 2\lambda\vec{w}_t = 0$$

$$\Rightarrow \underline{\underline{X}}^T(\underline{\underline{X}}\vec{w}_t - \vec{y}) + \lambda\vec{w}_t \Rightarrow \frac{1}{N}\left[\underline{\underline{X}}^T(\underline{\underline{X}}\vec{w}_t - \vec{y}) + \lambda\vec{w}_t\right]$$

starting at some $\vec{w}_t$ at $t=0$, for $t = 0, 1, 2, \ldots$

$$\vec{w}_{t+1} = \vec{w}_t - \frac{\eta}{N}\left[\underline{\underline{X}}^T(\underline{\underline{X}}\vec{w}_t - \vec{y}) + \lambda\vec{w}_t\right].$$

1.2) $\tilde{\nabla} F(\vec{w}_t) = \vec{x}_n(\vec{x}_n^T\vec{w}_t - y_n) + \lambda\vec{w}_t$ where n is one example s.t. $n \in [N]$ picked uniformly at random and $\vec{x}_n \in \mathbb{R}^{(D+1)}$

starting at some $\vec{w}_t$ at $t=0$, for $t = 0, 1, 2, \ldots$

$$\vec{w}_{t+1} = \vec{w}_t - \eta\left[\vec{x}_n(\vec{x}_n^T\vec{w}_t - y_n) + \lambda\vec{w}_t\right]$$

2.1.1) when $k = y$

$$\frac{\partial l}{\partial a_k} = \frac{-\sum_{k \neq y} e^{a_{k'} - a_y}}{1 + \sum_{k \neq y} e^{a_{k'} - a_y}}$$

when $k \neq y$

$$\frac{\partial l}{\partial a_k} = \frac{e^{a_k - a_y}}{1 + \sum_{k' \neq y} e^{a_{k'} - a_y}} = \frac{e^{a_k - a_y}}{\sum_{k'} e^{a_{k'} - a_y}}$$

$$= \frac{\cancel{e^{-a_y}}(e^{a_k})}{\cancel{e^{-a_y}} \sum_{k'} e^{a_{k'}}} = \frac{e^{a_k}}{\sum_{k'} e^{a_{k'}}} = z_k$$

2.1.2) $$\frac{\partial l}{\partial \underline{W}^{(2)}} = \frac{\partial l}{\partial \vec{a}} \frac{\partial \vec{a}}{\partial \underline{W}^{(2)}} = \frac{\partial l}{\partial \vec{a}} \vec{h}^T$$

$$\frac{\partial l}{\partial \vec{b}^{(2)}} = \frac{\partial l}{\partial \vec{a}} \frac{\partial \vec{a}}{\partial \vec{b}^{(2)}} = \frac{\partial l}{\partial \vec{a}}$$

2.1.3) $$\frac{\partial l}{\partial \vec{u}} = \frac{\partial l}{\partial \vec{a}} \frac{\partial \vec{a}}{\partial \vec{h}} \frac{\partial \vec{h}}{\partial \vec{u}} \Rightarrow \frac{\partial \vec{a}}{\partial \vec{h}} = \underline{W}^{(2)T} \qquad \frac{\partial \vec{h}}{\partial \vec{u}} \Rightarrow \frac{\partial H(u)}{\partial \vec{u}} = \underline{\underline{H}}(\vec{u})$$

$$= \underline{\underline{H}}(\vec{u}) \underline{W}^{(2)T} \frac{\partial l}{\partial \vec{a}}$$

2.1.4) $$\frac{\partial l}{\partial \underline{W}^{(1)}} = \frac{\partial l}{\partial \vec{u}} \frac{\partial \vec{u}}{\partial \underline{W}^{(1)}} = \frac{\partial l}{\partial \vec{u}} \vec{x}^T$$

$$\frac{\partial l}{\partial \vec{b}^{(1)}} = \frac{\partial l}{\partial \vec{u}} \frac{\partial \vec{u}}{\partial \vec{b}^{(1)}} = \frac{\partial l}{\partial \vec{u}}$$

2.2) $\frac{\partial l}{\partial \underline{\underline{W}}^{(1)}} = \frac{\partial l}{\partial \vec{u}} \vec{x}^T = \underline{\underline{H}}(\vec{u}) \underline{\underline{W}}^{(2)T} \frac{\partial l}{\partial \vec{a}} \vec{x}^T$ since $\underline{\underline{W}}^{(2)} = \underline{\underline{0}}$

$= \underline{\underline{0}}$

$\frac{\partial l}{\partial \underline{\underline{W}}^{(2)}} = \frac{\partial l}{\partial \vec{a}} \vec{h}^T = \frac{\partial l}{\partial \vec{a}} \left(\max\{0, \underline{\underline{W}}^{(1)} \vec{x} + b^{(1)}\} \right)^T$ since $\underline{\underline{W}}^{(1)}$ and $b^{(1)} = \underline{\underline{0}}$ and $\vec{0}$

$= \frac{\partial l}{\partial \vec{a}} \left(\max\{0, \vec{0}\} \right)^T = \vec{0}$

$\frac{\partial l}{\partial \vec{b}^{(1)}} = \frac{\partial l}{\partial \vec{u}} = \underline{\underline{H}}(\vec{u}) \underline{\underline{W}}^{(2)T} \frac{\partial l}{\partial \vec{a}}$ since $\underline{\underline{W}}^{(2)} = \underline{\underline{0}}$

$= 0$

since $\frac{\partial l}{\partial \underline{\underline{W}}^{(i)}}$ is always $0 \Rightarrow \eta \frac{\partial l}{\partial \underline{\underline{W}}^{(i)}}$ is always 0 meaning $\underline{\underline{W}}_{t+1}^{(i)} = \underline{\underline{W}}_t^{(i)} - 0 = \underline{\underline{W}}_t^{(i)}$

2.3) $\vec{a} = \underline{\underline{W}}^{(2)} \vec{u} + \vec{b}^{(2)}$

$= \underline{\underline{W}}^{(2)} \left(\underline{\underline{W}}^{(1)} \vec{x} + \vec{b}^{(1)} \right) + \vec{b}^{(2)} = \underline{\underline{W}}^{(2)} \underline{\underline{W}}^{(1)} \vec{x} + \underline{\underline{W}}^{(2)} \vec{b}^{(1)} + \vec{b}^{(2)}$

$\Rightarrow \underline{\underline{U}} = \underline{\underline{W}}^{(2)} \underline{\underline{W}}^{(1)} \qquad \vec{V} = \underline{\underline{W}}^{(2)} \vec{b}^{(1)} + \vec{b}^{(2)}$

$\underline{\underline{U}} \in \mathbb{R}^{K \times D} \qquad \vec{V} \in \mathbb{R}^{K}$

3.1) $F(\vec{w}) = \sum_{i=1}^{N} y_i \vec{w}^T \vec{x}_i - \lambda(\vec{w}^T\vec{w} - 1)$

$$\nabla_w F = \sum_{i=1}^{N} y_i \vec{x}_i - 2\lambda \vec{w} = 0$$

if $x_i \in C_1 \Rightarrow y = 1$, if $x_i \in C_{-1} \Rightarrow y = -1$

$$\hookrightarrow \left(\sum_{i: x_i \in C_1} \vec{x}_i - \sum_{j: x_j \in C_{-1}} \vec{x}_j \right) = 2\lambda \vec{w}$$

$$\frac{1}{2\lambda} \left(\sum_{i: x_i \in C_1} \vec{x}_i - \sum_{j: x_j \in C_{-1}} \vec{x}_j \right) = \vec{w}$$

3.2) $\|\vec{w}\|^2 = 1^2 = \left\| \frac{1}{2\lambda} \left(\sum_i \vec{x}_i - \sum_j \vec{x}_j \right) \right\|^2$

$$1 = \frac{1}{2\lambda} \left\| \sum_i \vec{x}_i - \sum_j \vec{x}_j \right\|^2$$

$$\lambda = \frac{1}{2} \left\| \sum_i \vec{x}_i - \sum_j \vec{x}_j \right\|^2$$

3.3)