ZIYAO YANG 9490544195

Problem set #2

1-1 $\vec{W}_{k+1} = \vec{W}_k + y_k \vec{x}_k$

$\Rightarrow \vec{W}_{k+1}^T = \vec{W}_k^T + y_k \vec{x}_k^T$

$\Rightarrow \vec{W}_{k+1}^T \vec{W}_{opt} = \vec{W}_k \vec{W}_{opt} + y_k \vec{x}_k^T \vec{W}_{opt}$

since two classes are linearly seperable, $\vec{W}_{opt}$ could make no mistakes to predict. Then $y_k \vec{X}_k^T \vec{W}_{opt} > 0$

And $\gamma(\vec{W}_{opt}) = \frac{|\vec{W}_{opt} \vec{x}|}{\|\vec{W}_{opt}\|}$

$\Rightarrow \vec{W}_{k+1}^T \vec{W}_{opt} \geq \vec{W}_k \vec{W}_{opt} + \gamma \|\vec{W}_{opt}\|$

1.2 $\|\vec{W}_{k+1}\|^2 = \vec{W}_{k+1}^T \vec{W}_{k+1}$

$= (\vec{W}_k + y_k \vec{x}_k)^T (\vec{W}_k + y_k \vec{x}_k)$

$= \|\vec{W}_k\|^2 + 2 y_k W_k^T x_k + y_k^2 \|\vec{x}_k\|^2$

since $y_k W_k^T x_k < 0$ and $\|\vec{x}_k\|^2 = 1$, $y_k \in \{-1, 1\}$

then $\|\vec{W}_{k+1}\|^2 = \|\vec{W}_k\|^2 + 2 y_k \vec{W}_k^T \vec{x}_k + y_k^2 \|\vec{x}_k\|^2$

$\leq \|\vec{W}_k\|^2 + y_k^2 \|\vec{x}_k\|^2 \leq \|\vec{W}_k\|^2 + 1$

Therefore, $\|\vec{W}_{k+1}\|^2 \leq \|\vec{W}_k\|^2 + 1$

1.3 $\vec{w}_{k+1}^T \vec{w}_{opt} \geq \vec{w}_M^T \vec{w}_{opt} + \gamma \|\vec{w}_{opt}\|$

$$\geq \vec{w}_{M-1}^T \vec{w}_{opt} + 2\gamma \|\vec{w}_{opt}\|$$

$$\cdots$$

$$\geq \vec{w}_0^T \vec{w}_{opt} + M\gamma \|\vec{w}_{opt}\|$$

since $\vec{w}_0 = 0$

$\vec{w}_{k+1}^T \vec{w}_{opt} \geq M\gamma \|\vec{w}_{opt}\|$

then Due to cachy-schwartz inequality

$$M\gamma \|\vec{w}_{opt}\| \leq \vec{w}_{k+1}^T \vec{w}_{opt} \leq \|w_{k+1}\| \|\vec{w}_{opt}\|$$

so $M\gamma \leq \|w_{k+1}\|$.

similarly $\|w_{k+1}\|^2 \leq \|w_M\|^2 + 1$

$$\leq \|w_{M-1}\|^2 + 2$$

$$\cdots$$

$$\leq \|w_0\|^2 + M \quad , w_0 = 0$$

so $\|w_{k+1}\|^2 \leq M \Rightarrow \|w_{k+1}\| \leq \sqrt{M}$.

since $(M-k)$ iteration make no mistakes and it doesn't matter

Therefore, $M\gamma \leq \|\vec{w}_{k+1}\| \leq \sqrt{M}$.

1.4 $\gamma M \leq \sqrt{M} \Rightarrow M \leq \gamma^{-2}$

2-1 $L(\vec{w},b) = -\sum_n \{ y_n \ln[\sigma(\vec{w}^T\vec{x_n}+b)] + (1-y_n)\ln[1-\sigma(\vec{w}^T\vec{x_n}+b)] \}$

$$\nabla_{\vec{w}} L = -\sum_n \{ y_n \frac{\sigma(z)(1-\sigma(z))\big|_{z=\vec{w}^T\vec{x_n}+b}}{\sigma(\vec{w}^T\vec{x_n}+b)} \cdot \vec{x_n} +$$

$$(1-y_n)\cdot \frac{-\sigma(z)(1-\sigma(z))\big|_{z=\vec{w}^T\vec{x_n}+b}}{1-\sigma(\vec{w}^T\vec{x_n}+b)} \cdot \vec{x_n} \}$$

$$= -\sum_n \{ y_n(1-\sigma(z))\vec{x_n} + (y_n-1)\sigma(z)\cdot\vec{x_n} \}\Big|_{z=\vec{w}^T\vec{x_n}+b}$$

$$= -\sum_n \{ y_n\vec{x_n} - \sigma(\vec{w}^T\vec{x_n}+b)\vec{x_n} \}$$

$$= -\sum_n [y_n - \sigma(\vec{w}^T\vec{x_n}+b)]\cdot\vec{x_n}.$$

so, $w \leftarrow w + \lambda \sum_n^N [y_n - \sigma(\vec{w}^T\vec{x_n}+b)]\cdot\vec{x_n}.$

2-2

$$\sum_{n=1}^{4} [y_n - \sigma(wx_n)]\cdot x_n$$

$$= (0-\sigma(0))\cdot 1 + (1-\sigma(0))\cdot 1 +$$
$$(1-\sigma(0))\cdot 1 + (1-\sigma(0))\cdot 1$$

$$= 1 \qquad \left(\sigma(0) = \frac{1}{1+e^0} = 0.5\right)$$

$$w = 0.001 \longleftarrow w_0 + \lambda \sum_{n=1}^{4} [y_n - \sigma(wx)]\cdot x_n$$

$$= 0 + 0.001 \times 1$$

input	$\sigma(wx)$	predict	label
1	$\sigma(0.001) = 0.50025$	1	0
1	$\sigma(0.001)$	1	1
1	$\sigma(0.001)$	1	1
1	$\sigma(0.001)$	1	1

training accuracy $= \frac{3}{4} = 0.75$

2.3 from 2.2, $w = 0.001$

we could get

input	$\sigma(wx)$	predict	label
-1	$\sigma(-0.001) = 0.49975$	0	0
1	$\sigma(0.001) = 0.50025$	1	1
1	$\sigma(0.001)$	1	0

test accuracy $= \frac{2}{3} = 0.67$

3、 $\frac{\partial L}{\partial v_{jk}} = \frac{\partial L}{\partial \hat{y}_j} \cdot \frac{\partial \hat{y}_j}{\partial o_j} \cdot \frac{\partial o_j}{\partial v_{jk}}$

$$\frac{\partial L}{\partial \hat{y}_j} = -\frac{y_j}{\hat{y}_j}$$

$$\frac{\partial \hat{y}_j}{\partial o_m} = \begin{cases} -\frac{e^{o_j} \cdot e^{o_m}}{[\sum_{i=1}^{3} \exp(o_i)]^2} \\ = -\hat{y}_j \cdot \hat{y}_m & , \quad m \neq j \\ \frac{e^{o_j}}{\sum_{i=1}^{3} \exp(o_i)} - \frac{o_j^2}{\sum_{i=1}^{3} \exp(o_i)} \\ = \hat{y}_j - \hat{y}_j^2 = \hat{y}_j(1-\hat{y}_j) , & m = j \end{cases}$$

$$\frac{\partial L}{\partial o_j} = \sum_{m=1}^{3} \frac{\partial L}{\partial \hat{y}_m} \cdot \frac{\partial \hat{y}_m}{\partial o_j}$$

$$= \frac{\partial L}{\partial \hat{y}_j} \cdot \frac{\partial \hat{y}_j}{\partial o_j} + \sum_{\substack{m=1 \\ m \neq j}}^{3} \frac{\partial L}{\partial \hat{y}_m} \cdot \frac{\partial \hat{y}_m}{\partial o_j}$$

$$= -\frac{y_j}{\hat{y}_j} \cdot \hat{y}_j(1-\hat{y}_j) + \sum_{\substack{m=1 \\ m \neq j}}^{3} \left(-\frac{y_m}{\hat{y}_m}\right)(-\hat{y}_j \cdot \hat{y}_m)$$

$$= -y_j(1-\hat{y}_j) + \hat{y}_j \sum_{\substack{m=1 \\ m \neq j}}^{3} y_m$$

$$= -y_j + \hat{y}_j \sum_{m=1}^{3} y_m$$

So $\frac{\partial L}{\partial v_{jk}} = \frac{\partial L}{\partial o_j} \cdot \frac{\partial o_j}{\partial v_{jk}} = \left(\hat{y}_j \sum_{m=1}^{3} y_m - y_j\right) \cdot z_k$

$$\frac{\partial L}{\partial W_{ki}} = \frac{\partial L}{\partial z_k} \cdot \frac{\partial z_k}{\partial W_{ki}}$$

$$\frac{\partial L}{\partial z_k} = \sum_{j=1}^{3} \frac{\partial L}{\partial \hat{y}_j} \cdot \frac{\partial \hat{y}_j}{\partial \theta_j} \cdot \frac{\partial \theta_j}{\partial z_k}$$

$$= \sum_{j=1}^{3} \left(-y_j + \hat{y}_j \sum_{m=1}^{3} y_m\right) v_{jk}$$

$$\frac{\partial z_k}{\partial W_{ki}} = (1 - z_k^2) \cdot x_i$$

so. $$\frac{\partial L}{\partial W_{ki}} = \sum_{j=1}^{3} \left(-y_j + \hat{y}_j \sum_{m=1}^{3} y_m\right) \cdot v_{jk} \cdot (1 - z_k^2) \cdot x_i$$