Swapnoneel Kayal
200100154

IE643 – Assignment 1

(1)

(a) $H = \{x \in \mathbb{R}^d : \langle w, x \rangle = b\}$

If H is a separating hyperplane: $y^j(w^T x^j + b) > 0$

9 We consider a $\tilde{w}$ whose L_2-norm is β i.e. $\|\tilde{w}\|_2 = \beta$

10 Mathematically, $\|\tilde{w}\|_2 = \sqrt{\langle \tilde{w}, \tilde{w} \rangle} = \sqrt{\sum_{j=1}^{d} (\tilde{w}_j)^2} = \beta$

11 Consider $\|w\|_2 = \gamma$ with $w \neq \vec{0}$ and $\gamma \neq 0$

then $\sqrt{\sum_{j=1}^{d} (w_j)^2} = \gamma$

12

We can consider $\beta / \gamma = c$ $[\because \gamma \neq 0]$ with $c \to$ Scalar

1

$\|cw\|_2 = \sqrt{\langle cw, cw \rangle}$ [Note: $c \in \mathbb{R}$, $w, \tilde{w} \in \mathbb{R}^d$]

2 $= \sqrt{c^2 \langle w, w \rangle}$

$= c\sqrt{\langle w, w \rangle}$

3 $= c\|w\|_2$

$= c\gamma$

4 $= \beta$

$= \sqrt{\sum_{j=1}^{d} \tilde{w}_j^2}$

5

$= \|\tilde{w}\|_2$

6

This implies $\langle cw, cw \rangle = \langle \tilde{w}, \tilde{w} \rangle$

thus $\tilde{w} = cw$

However, $\tilde{w}$ is a separating hyperplane, hence

$y^j(\tilde{w}^T x^j + \tilde{b}) > 0 \;\forall\; j \in [1, \dots, m]$ where $m \to$ # training samples

Substituting, $y^j(cw^Tx^j + \tilde{b}) > 0$ $[\because \tilde{w}^T = (cw)^T = cw^T]$

$y^j(cw^Tx^j + cb + \tilde{b} - cb) > 0 \quad \forall j$
$cy^j(w^Tx^j + b) + y^j(\tilde{b} - cb) > 0 \quad \forall j$

$\Rightarrow \underbrace{cy^j(w^Tx^j + b)}_{>0\ \forall j} + y^j(\tilde{b} - cb) > 0 \quad \forall j$

$>0\ \forall j$ [As $c \in \mathbb{R}$ and $c > 0$ because $c = \beta/\gamma$ and $\beta, \gamma > 0$]

Thus, if $y^j(\tilde{b} - cb) \geq 0\ \forall j$ ~~only~~ $\Rightarrow$ ~~when~~ $\tilde{b} = cb$

This is because when $\tilde{b} > cb$ then $y^j(\tilde{b} - cb) < 0$ for $y^j = -1$ and similarly if $\tilde{b} < cb$ then $y^j(\tilde{b} - cb) < 0$ for $y^j = 1$

$\therefore \tilde{w} = cw$
$\Rightarrow \tilde{b} = cb$

$\tilde{H} = \{x \in \mathbb{R}^d : \langle \tilde{w}, x \rangle = \tilde{b}\}$ with $c = \dfrac{\|\tilde{w}\|_2}{\|w\|_2} = \dfrac{\beta}{\|w\|_2}$

(b) $D = \{(x^1, y^1), \ldots (x^m, y^m)\}$
with $x^j \in \mathbb{R}^d$, $y^j \in \{+1, -1\}$ $\forall j \in \{1, \ldots m\}$

Consider $\max_j \|x^j\|_2 \leq R$

D is linearly separable dataset and there exists a separating hyperplane with w^*, γ such that

$y^j \langle w^*, x^j \rangle \geq \gamma \quad \forall j \in \{1, \ldots m\}$ and $\gamma > 0$

We had proven in class that

$$M \le \frac{R^2 \|w^*\|_2^2}{\gamma^2}$$

mistakes till convergence

We are asked to show $M \le \frac{R^2}{\eta^2}$, $\eta > 0$

In the previous part (a), we showed that if (w, b) is a separating hyperplane then (cw, cb) is also a separating hyperplane when $c > 0$.

$y^j \langle w^*, x^j \rangle > \gamma \quad \forall j \in \{1, \ldots, m\}$

separating hyperplane (w^*), margin (γ)

Consider $\|\tilde{w}\|_2 = \beta$, $\tilde{w} = \mathbb{R}^d \Rightarrow \beta = \sqrt{\sum_{j=1}^d \tilde{w}_j^2}$

Consider $\gamma = \sqrt{\sum_{j=1}^d w_j^{*2}}$

Let $c = \beta / \gamma$ then $\|cw^*\|_2 = \sqrt{\langle cw^*, cw^* \rangle} = c\sqrt{\langle w^*, w^* \rangle}$

$$= c\|w^*\|_2$$
$$= c\gamma$$
$$= \beta$$
$$= \|\tilde{w}\|_2$$

$\therefore \langle cw^*, cw^* \rangle = \langle \tilde{w}, \tilde{w} \rangle$

$\Rightarrow cw^* = \tilde{w}$ with $c > 0$

However, $\tilde{w}$ is a separating hyperplane with $\tilde{\gamma}$ margin ie. $y^j \langle \tilde{w}, x^j \rangle > \tilde{\gamma} \quad \forall j \in \{1, \ldots, m\}$

$$y^j \langle cw^*, x^j \rangle > \tilde{\gamma}$$
$$cy^j \langle w^*, x^j \rangle > \tilde{\gamma}$$
$$y^j \langle w^*, x^j \rangle > \frac{\tilde{\gamma}}{c} \quad \text{as } c > 0$$

But, it is also given to us that

$y^j \langle w^*, x^j \rangle \geq \gamma$

Thus, for both the inequalities to be true, $\frac{\tilde{\gamma}}{c} = \gamma$

i.e. $\tilde{\gamma} = c\gamma$ ~~where~~ where $c = \frac{\beta}{\|w^*\|_2}$ and $\beta \stackrel{c}{=} \|\tilde{w}\|_2$

$$\frac{\|w^*\|_2^2}{\gamma^2} = \frac{(\|w^*\|_2)^2}{(\tilde{\gamma}/c)^2} = \frac{(\beta/c)^2}{(\tilde{\gamma}/c)^2} = \frac{\beta^2}{\tilde{\gamma}^2} = \frac{\|\tilde{w}\|_2^2}{\tilde{\gamma}^2}$$

i.e. if $\tilde{w}$ is a separating hyperplane with margin $\tilde{\gamma}$

then $\frac{\|\tilde{w}\|_2^2}{\tilde{\gamma}^2} = \text{constant} = \frac{\|w^*\|_2^2}{\gamma^2} = \frac{1}{\eta^2}$

Clearly, $\eta \rightarrow$ property of the dataset and is the same for each separating hyperplane.

Previously, we had $M \leq \frac{R^2\|w^*\|_2^2}{\gamma^2}$

$\leq \frac{R^2}{\eta^2}$ with $\eta^2 = \frac{\gamma^2}{\|w^*\|_2^2}$, $\eta > 0$

(2) $w^0 = \{\theta, \theta, \ldots \theta\}^T$, $\theta \in [0,1]$

$D = \{(x^1, y^1), \ldots (x^m, y^m)\}$ ~~is~~ linearly separable dataset

when a mistake is committed $\rightarrow w^{t+1} = w^t + y^t x^t$

when no mistake is committed $\rightarrow w^{t+1} = w^t$

Let γ be margin of separating hyperplane then

$y^j \langle w^*, x^j \rangle \geq \gamma \quad \forall j \in \{1, \ldots, m\}$ where $m = \#$ training samples

Considering $\langle w^*, w^{t+1}\rangle - \langle w^*, w^t\rangle$

$= \langle w^*, w^t + y^t x^t\rangle - \langle w^*, w^t\rangle$

$= \cancel{\langle w^t, w^t\rangle} + y^t\langle w^*, x^t\rangle - \cancel{\langle w^*, w^t\rangle}$

$= y^t\langle w^*, x^t\rangle \geq \gamma$

Note → t is the round when a mistake was committed

If t is the round when no mistake is committed,

$\langle w^*, w^{t+1}\rangle - \langle w^*, w^t\rangle = 0$ $[\because w^{t+1} = w^t]$

So, $\sum_{t=0}^{T}\left(\langle w^*, w^{t+1}\rangle - \langle w^*, w^t\rangle\right)$

$= \sum_{t \in \text{no mistake}}\left(\langle w^*, w^{t+1}\rangle - \langle w^*, w^t\rangle\right)$

$+ \sum_{t \in \text{mistake}}\left(\langle w^*, w^{t+1}\rangle - \langle w^*, w^t\rangle\right)$

$= 0 + \sum_{t \in \text{mistake}} y^t\langle w^*, x^t\rangle$

$\geq \sum_{t \in \text{mistake}} \gamma$

$> M\gamma$

↳ # mistakes till time T

Also, $\langle w^*, w^{T+1}\rangle - \langle w^*, w^0\rangle = \sum_{t=0}^{T}\left(\langle w^*, w^{t+1}\rangle - \langle w^*, w^t\rangle\right)$

$> M\gamma$

Mathematically

$$\sqrt{\sum_{j=1}^{d} \theta w_j^*} = \langle w^*, w^0 \rangle \quad \text{as} \quad w^0 = [\theta, \dots, \theta]^T$$

$$\langle w^*, w^0 \rangle = \sqrt{\theta}\, \|w^*\|_2$$

$$\Rightarrow \langle w^*, w^{T+1} \rangle > Mr + \sqrt{\theta}\, \|w^*\|_2$$

Applying Cauchy-Schwarz Inequality →

$$\langle w^*, w^{T+1} \rangle \leq \|w^*\|_2 \, \|w^{T+1}\|_2$$

But, $\|w^{T+1}\|_2^2 = \|w^T + y^T x^T\|_2^2$ [Assuming a mistake occurred at time T]

$$= \|w^T\|_2^2 + \|y^T x^T\|_2^2 + 2\langle w^T, y^T x^T \rangle$$

$$\|w^{T+1}\|_2^2 = \|w^T\|_2^2 + (y^T)^2 \|x^T\|_2^2 + 2y^T \langle w^T, x^T \rangle$$

∵ A mistake has occurred at time T,

$y^T \langle w^T, x^T \rangle < 0$ as $y^T \neq \langle w^T, x^T \rangle$,
$y^T = \{+1, -1\}$,
$\langle w^T, x^T \rangle = \{+1, -1\}$

Thus, $\|w^{T+1}\|_2^2 = \|w^T\|_2^2 + \|x^T\|_2^2 - 2\langle w^T, x^T \rangle$

$$\leq \|w^T\|_2^2 + \|x^T\|$$

$$\|w^{T+1}\|_2^2 - \|w^T\|_2^2 \leq \|x^T\|_2^2$$

Summing from $t=0$ to $t=T$ →

$$\|w^{T+1}\|_2^2 - \|w^0\|_2^2 \leq \sum_{t \in \text{mistake}} \|x^T\|_2^2 + \sum_{t \in \text{no mistake}} (0) \quad [\because w^{t+1} = w^t]$$

We consider $\|x^t\|_2^2 \leq R^2 \ \forall \ t \in \{1, \dots, T\}$

$$\|w^{T+1}\|_2^2 - \|w^0\|_2^2 \le MR^2$$
$$\|w^{T+1}\|_2^2 \le MR^2 + \|w^0\|_2^2$$

$$\therefore \langle w^*, w^{T+1}\rangle \le \|w^*\|_2 \|w^{T+1}\|_2 \le \|w^*\|_2 \sqrt{MR^2 + \|w^0\|_2^2}$$
$$Mr + \sqrt{\theta}\|w^*\|_2 < \langle w^*, w^{T+1}\rangle$$

~~Thus,~~ $\cancel{(Mr + \sqrt{\theta}\|w^*\|_2)^2 \le \|w^*\|_2^2 \cdot (MR^2 + \|w^0\|_2^2)}$

$$\cancel{M^2r^2 + \theta\|w^*\|_2^2 + 2Mr\sqrt{\theta}\|w^*\|_2 \le MR^2\|w^*\|_2^2 + \|w^*\|_2^2\theta^2 d}$$
$$\cancel{[\because \|w^0\|_2^2 = \theta^2 d]}$$

$$\cancel{M^2r^2 + M(2r\sqrt{\theta}\|w^*\|_2 - R^2\|w^*\|_2^2) + \theta\|w^*\|_2^2 - \|w^*\|_2^2\theta^2 d \le 0}$$

$$Mr + \sqrt{\theta}\|w^*\|_2 < \langle w^*, w^{T+1}\rangle < \|w^*\|_2\sqrt{MR^2 + \|w^0\|_2^2}$$
where $\theta \in [0,1]$

$$Mr < Mr + \sqrt{\theta}\|w^*\|_2 < \langle w^*, w^{T+1}\rangle$$

Thus,
$$Mr < \langle w^*, w^{T+1}\rangle \le \|w^*\|_2 \cdot \sqrt{MR^2 + \|w^0\|_2^2}$$

$$\|w^0\|_2^2 = \sum_{j=1}^{d}\theta^2 = \theta^2 d \quad \text{and} \quad \theta \in [0,1]$$

$$\therefore \|w^0\|_2^2 \le d$$

$$Mr < \langle w^*, w^{T+1}\rangle \le \|w^*\|_2 \cdot \sqrt{MR^2 + \|w^0\|_2^2}$$

$$Mr < \|w^*\|_2\sqrt{MR^2 + d}$$

Hence, $M^2r^2 < \|w^*\|_2^2 \cdot (MR^2 + d)$

$$M^2r^2 - MR^2\|w^*\|_2^2 - d\|w^*\|_2^2 < 0$$

↳ quadratic in M

$(M-\alpha)(M-\beta) \le 0$

$$\alpha, \beta = \frac{R^2 \|w^*\|_2^2 \pm \sqrt{R^4 \|w^*\|_2^4 + 4r^2 d \|w^*\|_2^2}}{2r^2}$$

Thus,

$$\underbrace{\frac{R^2 \|w^*\|_2^2}{2r^2} - \frac{\|w^*\|_2 \sqrt{R^4 \|w^*\|_2^2 + 4r^2 d}}{2r^2}}_{\text{This will be less than 0}} < M < \frac{R^2 \|w^*\|_2^2}{2r^2} + \frac{\|w^*\|_2 \sqrt{R^4 \|w^*\|_2^2 + 4r^2 d}}{2r^2}$$

As $M<0$ is ~~not~~ possible $\rightarrow M < \frac{R^2 \|w^*\|_2^2}{2r^2} + \frac{\|w^*\|_2}{2r^2} \sqrt{R^4 \|w^*\|_2^2 + 4r^2 d}$

(b) In the bounds we have obtained, if $R \gg r$ i.e. $\max_j \|x^j\|_2 \gg r$, we can ignore r in discriminant and get $\Rightarrow$

$$\frac{R^2 \|w^*\|_2^2}{2r^2} - \frac{\|w^*\|_2 \, R^2 \|w^*\|_2}{2r^2} < M < \frac{R^2 \|w^*\|_2^2}{2r^2} + \frac{R^2 \|w^*\|_2^2}{2r^2}$$

i.e. $0 < M < \frac{R^2 \|w^*\|_2^2}{r^2}$

which are the bounds in case of $w^0 = \vec{0}$

In new bounds, the upper bound is higher (less tighter bound) than the upper bound obtained for case when $w^0 = \vec{0}$ i.e. this initialization of w^0 would take more number of mistakes to converge to reach optimal w^*

For general case: $M < \dfrac{R^2\|w^*\|_2^2}{2\gamma^2} + \dfrac{\|w^*\|_2}{2\gamma^2} * \left(\sqrt{R^4\|w^*\|_2^2 + 4\gamma^2\|w^0\|_2^2}\right)$

Now, we ~~a~~ replace d with $\|w^0\|_2^2$

Hence, larger ~~the~~ $\|w^0\|_2 \Rightarrow$ higher is the upper bound on mistakes.

10

(c) The given bound is not tight.

11 Upon solving, we got the following relation $\rightarrow$

12 $M\gamma < M\gamma + \sqrt{\theta}\,\|w^*\|_2$ (as $\theta \in [0,1]$)

$< \langle w^*, w^{T+1}\rangle \leq \|w^*\|_2\sqrt{MR^2 + \|w^0\|_2^2}$

1

Thus, $M\gamma < \|w^*\|_2\sqrt{MR^2 + \|w^0\|_2^2}$

2 $\Rightarrow$ bounds on M are not tight.