Q2/(a) Let $w^1 = [\theta \ \theta \ \dots \ \theta]^T \quad \theta \in [0,1]$

Considering Dataset $D = \{(x^1, y^1) \dots, (x^n, y^n)\}$

where $x^j \in R^d$, $y^j \in \{+1, -1\}$ $\forall j = 1, 2 \dots, n$.

Let $\max \|x^j\|_2 \leq R$

As D is linearly separable $\exists \cdot w^* \in R^d$ & $\gamma > 0$

such that $y^j \langle w^*, x^j \rangle \geq \gamma \quad \forall j = 1, 2 \dots n$

Using Perceptron update $w^{t+1} = w^t + y^t x^t$

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

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

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

$\rightarrow$ 1

For a total progress of T rounds, where M mistakes are made we obtain:

$$\sum_{t=1}^{T} \langle w^*, w^{t+1} \rangle - \langle w^*, w^t \rangle$$

(Since the starting is w^1 we take $t=1$)

$$= \sum_{\substack{t \in \{1, \dots, T\} \\ t: \text{mistake is made in round } t}} \langle w^*, w^{t+1} \rangle - \langle w^*, w^t \rangle$$

$$+ \sum_{\substack{t \in \{1, \dots T\} \\ t: \text{mistake is not made in round } t}} \langle w^*, w^{t+1} \rangle - \langle w^*, w^t \rangle$$

$$= \sum_{t \in \{1, \cdots T\}} \langle w^*, w^{t+1} \rangle - \langle w^*, w^t \rangle$$

t : mistake is made in round t

$$> M\gamma$$

[As total M mistakes have been made and (1)]

$$\sum_{t=1} \langle w^*, w^{t+1} \rangle - \langle w^*, w^t \rangle > M\gamma \quad \text{——} \quad (2)$$

This can be justified as when no mistake is made

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

Now,

$$\sum_{t=1}^{T} \langle w^*, w^{t+1} \rangle - \langle w^*, w^t \rangle$$

$$= \langle w^*, w^{T+1} - w^1 \rangle$$

Using Equation (2) we get

$$M\gamma < \langle w^*, w^{T+1} - w^1 \rangle$$

Using Cauchy-Schwarz Inequality :

$$M\gamma \leq \|w^*\| \, \|w^{T+1} - w^1\|$$

By using Triangle Inequality :

$$M\gamma < \|w^*\| \left[\|w^{T+1}\| + \|w^1\| \right]$$

Considering w' is a d-dimensional space

$$w' = [\theta \cdots \theta]^T_{d\times 1}$$

$$\|w^{T+1}\| + \|w^0\| = \|w^{T+1}\| + \sqrt{d\theta^2}$$

$$M^2\gamma^2 < \|w^*\|^2 [\|w^{T+1}\| + \sqrt{d\theta^2}]^2$$

$$< \|w^*\|^2 [\|w^{T+1}\|^2 + d\theta^2 + 2\sqrt{d\theta^2}\,\|w^{T+1}\|]$$

Therefore from the above Equation:

$$\frac{M^2\gamma^2}{\|w^*\|^2} < [\|w^{T+1}\|^2 + 2\sqrt{d}\,\theta\,\|w^{T+1}\| + d\theta^2]$$

4

Using Perceptron update rule, we get

$$w^{t+1} = w^t + y^t x^t$$

* Since mistake has occured off signs

$$\Rightarrow \|w^{t+1}\|^2 = \|w^t + y^t x^t\|^2$$

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

$$= K$$

5 So

Now from inequality

$$K \leq \|w^t\|^2 + \|x^t\|^2$$

So, $\Rightarrow \|w^{t+1}\|^2 \leq \|w^t\|^2 + \|x^t\|^2$

As $y^t \in [1, -1]$ & whenever update occurs (ie. mistake has occured) $y^t \hat{y}^t \leq 0$ (as $y^t \neq \hat{y}^t$)

$$\Rightarrow y^t \operatorname{sign}(\langle w^t, x^t \rangle) \leq 0$$

$$\Rightarrow y^t (\langle w^t, x^t \rangle \leq 0$$

$$\therefore \|w^{t+1}\|^2 - \|w^t\|^2 \leq \|x^t\|^2$$

$$\leq R \quad \text{——} \quad (6)$$

Using Eqn 6 :

$$\sum_{t=1}^{T} (\|w^{t+1}\|^2 - \|w^t\|^2) = \sum_{\substack{t=\{1,\dots T\} \\ t = \text{mistake} \\ \text{at round } t}} (\|w^{t+1}\|^2 - \|w^t\|^2) + \sum_{\substack{t \in \{1 \dots T\} \\ t : \text{no mistake} \\ \text{at round } T}} (\|w^{t+1}\|^2 - \|w^t\|^2)$$

$$= \sum_{\substack{t \in \{1 \dots T\} \\ \text{mistake at} \\ \text{round } t}} (\|w^{t+1}\|^2 - \|w^t\|^2) \leq MR^2 \quad \hookrightarrow (7)$$

$$\therefore \|w^{T+1}\|^2 - \|w^1\|^2 \leq MR^2$$

$$\Rightarrow \|w^{T+1}\|^2 \leq MR^2 + d\theta^2$$

Using ③

$$\frac{M^2\gamma^2}{\|w^*\|^2} < \left[MR^2 + d\theta^2 + 2\sqrt{d}\,\theta \times \sqrt{MR^2 + d\theta^2} + d\theta^2\right]$$

$$= \left[MR^2 + 2d\theta^2 + 2\sqrt{d}\,\theta \times \sqrt{MR^2 + d\theta^2}\right]$$

Considering $M > 1$,

$$\boxed{\begin{aligned} &MR^2 + d\theta^2 \leq M^2R^2 + M^2 d\theta^2 \\ &\text{and} \quad 2d\theta^2 \leq 2M d\theta^2 \end{aligned}}$$

Combining the two Eqn:

$$\Rightarrow M < \frac{\|w^*\|^2}{\gamma^2}\left[R^2 + 2d\theta^2 + 2\theta\sqrt{d(R^2 + d\theta^2)}\right]$$

End of Part A

P.T.O

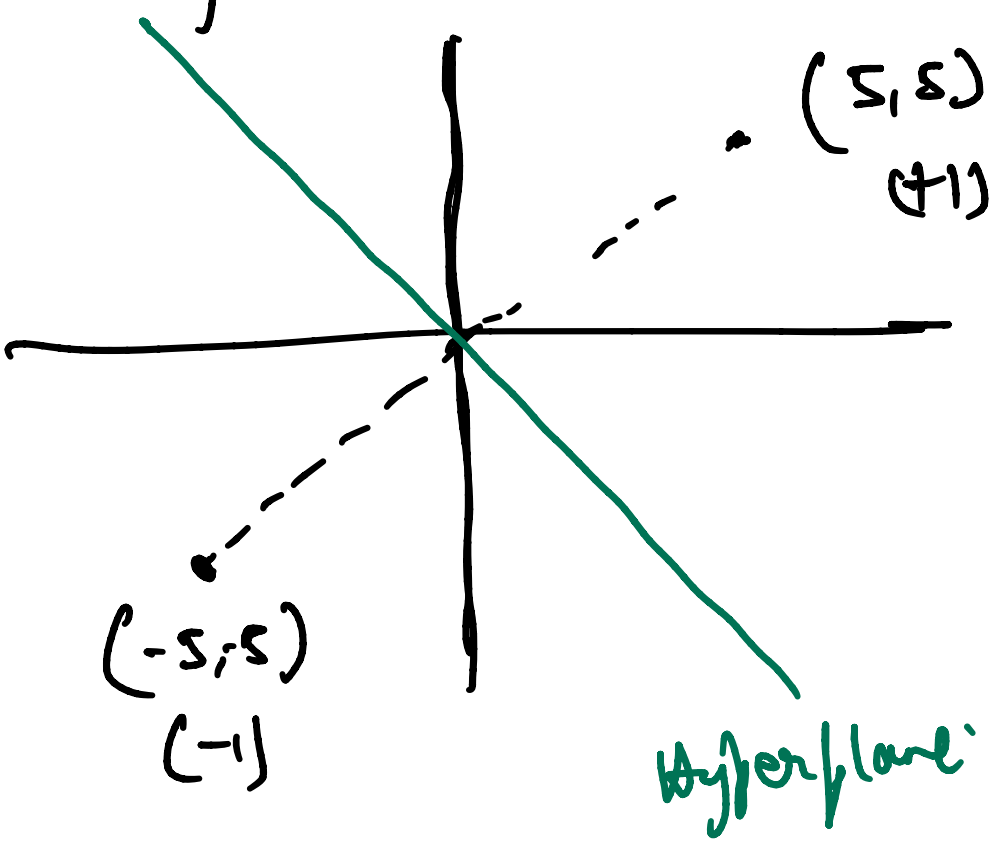
(b) The derivation done in class was for zero weight initialization.

In the current case the bound increases with increase in $d, R, \& \ \theta$. For $\theta = 0$, we obtain the same bound $\frac{\|w^*\|^2}{\gamma^2} R^2$ (Minimum Value)

(Max Value for $\theta = 1$) $\frac{\|w^*\|^2}{\gamma^2}\left[R^2 + 2d + 2\sqrt{d(R^2+d)}\right]$ → (8)

Thus, the bound obtained is greater.

(c) The bound is not tight as we can see from the Example below:



So $w^* = (1,1)$ separates the points. Considering $\gamma = 1$

So putting the above values in the formula (8)

$$= \frac{2}{1}\left[25 + 2\times 2\times 1 + 2\times 1\times \sqrt{(25+2\times 1)\times 2}\right]$$

$R^2 = 25 \quad \gamma = 1 \quad d = 2$

$\theta = 1$

$= 87.394.$

This bound is much greater than required since $w^* = (\frac{1}{2}, \frac{1}{2})$ also satisfies the condition. Thus the bounds are not tight.