1. (a) **[5 marks]** Recall that a hyperplane $H = (w, b)$ for some $w \neq \mathbf{0} \in \mathbb{R}^d$ and $b \in \mathbb{R}$ is defined as $H = \{x \in \mathbb{R}^d : \langle w, x\rangle = b\}$. Show that for every hyperplane H and for every $\beta > 0$, there exists another hyperplane $\tilde{H} = (\tilde{w}, \tilde{b})$ such that $\|\tilde{w}\|_2 = \beta$. Illustrate the relationship between (w, b) and $(\tilde{w}, \tilde{b})$. (Recall that for $u \in \mathbb{R}^d$, $\|u\|_2 = \sqrt{\langle u, u\rangle} = \sqrt{\sum_{i=1}^d |u_i|^2}$ is the ℓ_2 norm of u).

(b) **[5 marks]** Consider a data set $D = \{(x^1, y^1), \ldots, (x^n, y^n)\}$, where $x^j \in \mathbb{R}^d$, $y^j \in \{+1, -1\}, \forall j = 1, 2, \ldots, n$. Let $\max_j \|x^j\|_2 \leq R$. Recall that D is linearly separable if there exist $w^* \in \mathbb{R}^d$ and $\gamma > 0$ such that $y^j \langle w^*, x^j\rangle \geq \gamma$, $\forall j = 1, \ldots, n$. Show that if D is linearly separable, the mistake bound proved in class

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

can be rewritten simply as $M \leq \frac{R^2}{\eta^2}$, where $\eta > 0$ (which might be same as γ or different from γ). **(Hint: Use the previous result about hyperplane.)**

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

now we define an another Hyperplane

$$H' = \{x \in \mathbb{R}^d , \langle \frac{\beta}{\|w\|} w , x\rangle = \frac{\beta}{\|w\|} b\}$$

let $\tilde{\omega} = \frac{\beta}{\|w\|} w \Rightarrow \|\tilde{\omega}\| = \beta$

and assume that $\tilde{b} = \frac{\beta}{\|w\|} b$

then $\tilde{H} = \{x \in \mathbb{R}^d ; \langle \tilde{\omega}, x\rangle = \tilde{b}\}$

$\Rightarrow$ Hence we can always find a Hyperplane

and relation between (w, b) and $(\tilde{\omega}, \tilde{b})$ is-

$$\boxed{\begin{aligned} \tilde{\omega} &= \frac{\beta \omega}{\|\omega\|} \\ \tilde{b} &= \frac{\beta}{\|w\|} b \end{aligned}}$$

(b) Given a data set $D = \{(x^1, y^1), \ldots, (x^n, y^n)\}$
~~ab~~ where $x^j \in \mathbb{R}^d$; $y^j \in \{+1, -1\}$ $\forall j = 1, \ldots, n$

let $\|x^j\|_2 \leq R$

Given D is linearly separable.

$\Rightarrow \exists$ $w^* \in \mathbb{R}^d$ and $\gamma > 0$

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

$\forall j = 1, 2, \ldots, n$

we know that by mistake bound $\Rightarrow$

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

now for $y^j = 1$

$$\langle w^*, x^j \rangle > \gamma$$

$$\Rightarrow \left\langle \frac{\beta}{\|w\|_2} ; x^j \right\rangle > \frac{\gamma\beta}{\|w\|_2}$$

(since $\beta > 0$
$\|w\|_2 > 0$)

and for $y^i = -1$

$$\langle w^*, x^i \rangle < \gamma$$

$$\left\langle \frac{\beta}{\|w^*\|_2} w^*, x^i \right\rangle < \frac{\gamma \beta}{\|w^*\|_2}$$

Hence $(\tilde{\omega}, \tilde{b})$ is also linearly separable.

where $\tilde{\omega} = \frac{\beta w^*}{\|w^*\|_2}$

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

now $\|\tilde{\omega}\|_2^2 = \left\| \frac{\beta (w^*)}{\|w^*\|_2} \right\|_2^2 = \beta^2$

by mistake bound inequality $\Rightarrow$

$$M < \frac{R^2 \|\tilde{\omega}\|_2^2}{\gamma^2}$$

$$M < \frac{R^2 \beta^2}{\gamma^2}$$

let $\eta = \frac{\gamma}{\beta}$

then $\boxed{M \leq \frac{R^2}{\eta^2}}$

2. Consider the perceptron learning algorithm with a starting point $w^0 = \left[\theta\ \theta\ \ldots\ \theta\right]^\top$ where $\theta \in [0,1]$, used to train on a linearly separable data set.

   (a) **[7 marks]** Find a suitable upper bound on the number of mistakes for the choice of starting point w^0 given above.

   (b) **[3 marks]** Compare and contrast the bound you obtained in part (a) with the bound discussed in class. Explain the changes observed in the bound, and explain the dependence of the bound you obtained in part (a) on the choice of w^0.

   (c) **[2 marks]** Justify if your bound is tight.

Q2 (a) Let $w^0 = [\theta\ \theta \ldots \theta]^T$ where $\theta \in [0,1]$.

Consider the data set $D = \{(x^1, y^1), \ldots, (x^n, y^n)\}$, where $x^j \in \mathbb{R}^d$, $y^j \in \{+1, -1\}$, $\forall j = 1, 2, \ldots, n$. Let $\max_j \|x^j\|_2 \leq R$

As D is linearly separable $\exists\, w^* \in \mathbb{R}^d$ and $\gamma > 0$ s.t.

$$y^j \langle w^*, x^j \rangle \geq \gamma, \ \forall j = 1, 2, \ldots, n.$$

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

$$\langle w^*, w^{t+1} \rangle - \langle w^*, w^t \rangle = \langle w^*, w^{t+1} - w^t \rangle$$
$$= \langle w^*, y^t x^t \rangle$$
$$= y^t \langle w^*, x^t \rangle \geq \gamma \quad - (*)$$

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

$$\sum_{t=0}^{T} \langle w^*, w^{t+1} \rangle - \langle w^*, w^t \rangle = \sum_{\substack{t \in \{0, \ldots, T\} \\ t:\text{mistake is} \\ \text{made in round } t}} \langle w^*, w^{t+1} \rangle - \langle w^*, w^t \rangle$$

$$+ \sum_{\substack{t \in \{0, \ldots, T\} \\ t:\text{mistake is} \\ \text{not made in} \\ \text{round } t}} \langle w^*, w^{t+1} \rangle - \langle w^*, w^t \rangle$$

$$= \sum_{\substack{t \in \{0, \ldots, T\} \\ t:\text{mistake is} \\ \text{made in round } t}} \langle w^*, w^{t+1} \rangle - \langle w^*, w^t \rangle$$

$\geq M\gamma$ [As total M mistakes have been made and $(*)$]

$$\therefore \sum_{t=0}^{T} \langle w^*, w^{t+1} \rangle - \langle w^*, w^t \rangle \geq M\gamma \quad - (**)$$

This can be justified as when no mistake is made

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

Now,

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

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

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

$$+ \vdots$$

$$+ \langle w^*, w^1 \rangle - \langle w^*, w^0 \rangle$$

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

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

Using (**) we get

$$M\gamma < \langle w^*, w^{T+1} - w^0 \rangle$$

$$\leq \|w^*\| \, \|w^{T+1} - w^0\|$$ [Using Cauchy-Schwarz Inequality]

$$\therefore M\gamma < \|w^*\| \left[\|w^{T+1}\| + \|w^0\| \right]$$

[By Triangle Inequality]

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

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

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

$$\therefore \frac{M^2\gamma^2}{\|w^*\|^2} < \left[\|w^{T+1}\|^2 + 2\sqrt{d}\,\theta \|w^{T+1}\| + d\theta^2 \right] \quad (***)$$

Using Perceptron update rule, we get

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

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

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

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

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

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

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

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

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

$$\leq R \quad (***) \qquad [\|x^t\|^2 \leq R, \; \forall t = 1,2,\dots,n]$$

$$\therefore \sum_{t=0}^{T} \left(\|w^{t+1}\|^2 - \|w^t\|^2\right)$$

$$= \sum_{\substack{t\in\{0,\dots,T\} \\ t:\text{mistake at round } t}} \left(\|w^{t+1}\|^2 - \|w^t\|^2\right) + \sum_{\substack{t\in\{0,\dots,T\} \\ t:\text{no mistake at round } t}} \left(\|w^{t+1}\|^2 - \|w^t\|^2\right)$$

$$= \sum_{\substack{t\in\{0,\dots,T\} \\ t:\text{mistake at round } t}} \left(\|w^{t+1}\|^2 - \|w^t\|^2\right)$$

$$\leq M R^2 \qquad [\text{Using } (***)]$$

$$\therefore \|\omega^{T+1}\|^2 - \|\omega^0\|^2 \leq MR^2$$

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

Using $(***)$

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

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

Considering $M > 1$,

$$MR^2 + d\theta^2 \leq M^2R^2 + M^2 d\theta^2$$

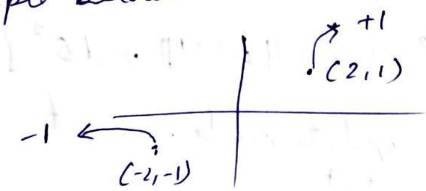
and $2d\theta^2 \leq 2Md\theta^2$

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

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

(b) The bound obtained is greater than the one obtained in class for zero weight initialization. The Bound increases with increase in d, R and θ. For, $\theta = 0$, we obtain the same bound $\frac{\|w^*\|^2}{\gamma^2} R^2$. On the other hand when $\theta = 1$, the bound becomes $\frac{\|w^*\|^2}{\gamma^2} \left[R^2 + 2d + 2\sqrt{d(R^2 + d)} \right]$.

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



The hyper plane with $w^* = (1,1)$ and $\gamma = 1/2$ separates the data

For perceptron with $w = (\frac{1}{2}, \frac{1}{2})$ no mistakes are made.

But the bound obtained is given by

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

$$= \frac{2}{1/4} \left[5 + 2 \times 2 \times \frac{1}{4} + 2 \times \frac{1}{2} \sqrt{2(5 + 2 \times \frac{1}{4})} \right]$$

$$= 74.533$$