ECE 2372 - Homework 3 Avery Peiffer

1.

PLA: $\theta^{j+1} = \begin{cases} \theta^j + y_i \tilde{x}_i & y_i \neq \text{sign}((\theta^j)^T \tilde{x}) \\ \theta^j & \text{otherwise} \end{cases}$

Meaning that $(\theta^{j+1})^T \tilde{x}_i = (\theta^j + y_i \tilde{x}_i)^T \tilde{x}_i$

$= (\theta^j)^T \tilde{x}_i + y_i \tilde{x}_i^T \tilde{x}_i$

$\rightarrow y_i \tilde{x}_i^T \tilde{x}_i$ pushes you in right direction

$y_i = 1$ and $\text{sign}((\theta^j)^T \tilde{x}_i) = -1 \longrightarrow$ gets bigger

$y_i = -1$ and $\text{sign}((\theta^j)^T \tilde{x}_i) = +1 \longrightarrow$ gets smaller

So PLA should converge if data is linearly separable.

(a) $\rho = \min_i |\langle \theta^*, \tilde{x}_i \rangle|$ is distance between closest $\tilde{x}_i$ in training data to hyperplane

$\min_i y_i \langle \theta^*, \tilde{x}_i \rangle = \rho > 0$

$y_i = \text{sign}(\theta^* \tilde{x})$ (z_0)

Closest $\tilde{x}_i$ has a component that lies on the hyperplane and a component orthogonal to the hyperplane $\left(\delta \frac{\omega}{\|\omega\|}\right)$

$\delta \frac{\omega}{\|\omega\|}$ must be greater than zero since the weights are nonzero, which means that the overall ρ must be greater than 0.

(b) $\langle \theta^j, \theta^* \rangle \geq \langle \theta^{j-1}, \theta^* \rangle + \rho$ $\qquad \langle \theta^j, \theta^* \rangle \overset{?}{\geq} j\rho$

Basis

$\theta' = \theta^0 + y_i \tilde{x}_i$

Since $\rho > 0$, adding it to θ^0 will increase it

$\theta' \geq \theta^0$

Now generalize

$$j < \frac{(R^2+1)\|\theta^*\|^2}{\rho^2}$$

$$j\rho^2 < (R^2+1)\|\theta^*\|^2$$

$$j\rho^2 < [\underset{i}{\text{argmax}} \|x_i\|^2 + 1] \|\theta^*\|^2$$

Therefore, we know that $\langle \theta^j, \theta^* \rangle \geq j\rho$ for $j \geq 1$ because j must be smaller than the value in the previous equation, completing the inductive process

(c) $\tilde{x}_{ij}$ was misclassified $\longrightarrow$

$y_{ij} \neq \text{sign}(\theta^{jT} \tilde{x})$

If $y_{ij-1} = -1$ and sign $\theta^{j-1T}\tilde{x} = 1$, update makes $\theta^{jT}\tilde{x}_i$ smaller.

Which makes the norm of $\theta^j \rightarrow \|\theta^j\|^2$, smaller

Meaning it will not be bigger than $\|\theta^{j-1}\|^2 + \|\tilde{x}_{ij}\|^2$

$\therefore \|\theta^j\|^2 \leq \|\theta^{j-1}\|^2 + \|\tilde{x}_{ij}\|^2$

(d) $\|\theta^j\|^2 \leq j(1+R^2)$ where $R = \max_i \|x_i\|$

$\theta^1 = \theta^0 + y_{i1}\tilde{x}_{i1} = y_{i1}\tilde{x}_{i1}$

$\tilde{x}_{i1} = [1, x_{i1}] \longrightarrow \|\tilde{x}_{i1}\|^2 = 1 + \|x_i\|^2$

$\theta^1 = y_{i1}(1 + \|x_i\|^2)$

Since R is the max distance of $\|x_i\|^2$, $\|\theta^j\|$ will never be greater than it.

$\|\theta^j\|^2 \leq j(1+R^2)$

(e) $j \leq \dfrac{(1+R)^2 \|\theta^*\|^2}{\rho^2}$

We know that $\langle \theta^j, \theta^* \rangle \geq j\rho$

Square both sides $\|\theta^j\|^2 \|\theta^*\|^2 \geq j^2\rho^2$

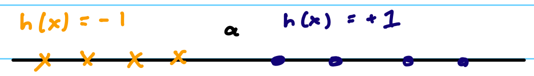
We also know that $\|\theta^j\|^2 \leq j(1+R^2)$

$$\frac{\|\theta^j\|^2 \leq j(1+R^2)}{\|\theta^j\|^2 \|\theta^*\|^2 \geq j^2\rho^2} = \frac{1 \leq 1+R^2}{\|\theta^*\|^2 \geq j\rho^2}$$

$j\rho^2 \leq (1+R^2)\|\theta^*\|$

$$j \leq \frac{(1+R^2)\|\theta^*\|}{\rho^2}$$

3. (a) $h(x) = \text{sign}(x-a)$ for some $a \in \mathbb{R}$



$m_H(n) = n+1$

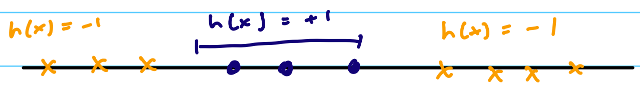
Similarly, for $h(x) = -\text{sign}(x-a)$ for some $a \in \mathbb{R}$, the growth function will also be $n+1$, since this is just flipping the result of the original version.

(b) $h(x) = \begin{cases} +1 & \text{for } x \in [a,b] \\ -1 & \text{otherwise} \end{cases}$

for some $a, b \in \mathbb{R}$

OR

$h(x) = \begin{cases} -1 & \text{for } x \in [a,b] \\ +1 & \text{otherwise} \end{cases}$



$m_H(n) = \binom{n+1}{2} + 1$

$= \frac{1}{2}n^2 + \frac{1}{2}n + 1$

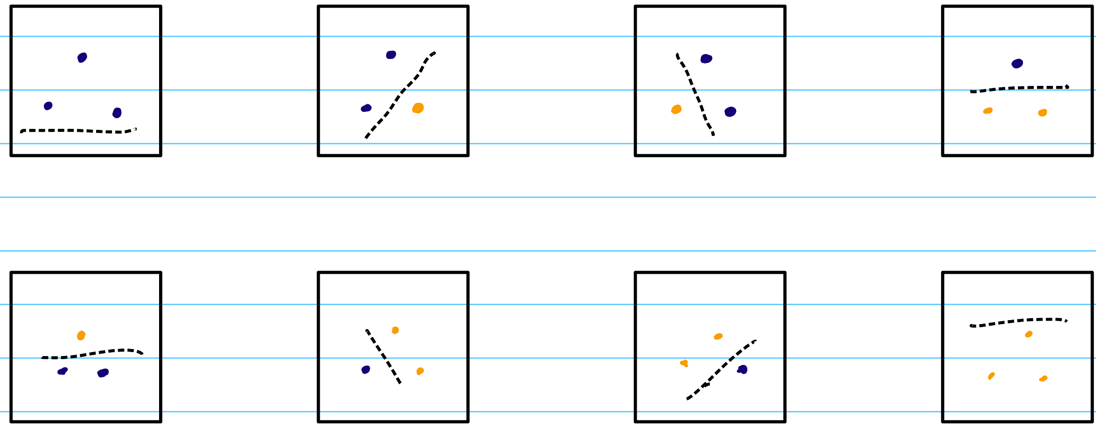
This is the same formula for the opposite case, as well.

4. $h(x) = \begin{cases} +1 & \text{if } \|x-c\| \leq r \\ -1 & \text{otherwise} \end{cases}$ for some $c \in \mathbb{R}^2$, $r \in \mathbb{R}$

(a) $m_H(3) = 8$

Notably, this is a linear classifier

→ Able to draw out configurations for all possible labelings



(b) $m_H(4) < 16$

Can't reach 16 dichotomies because of when a point is inside a triangle $\rightarrow$ a classifier can't be drawn for these dichotomies and we only get 14.