Meng Tianao

20810553

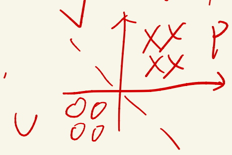
Problem 1.

First, Let's describe the first part of the theorem

Theorem: Let $w_1, \cdots, w_n$ be a set of vectors in a Euclidean space of fixed finite dimention, satisfying the hypothesis that there exists a vector y such that $(w_i, y) > \theta > 0 \quad i = 1, \cdots, N.$ $\boxed{(1)}$

For this part of the theorem, $w_1, \cdots, w_n$, we can consider them as training set, and they can be linearly separable.

For example,



, $W = P \cup (-U)$.

And if the training set can not be linearly separable, then there will not exsist such y vector, which is the perfect assignment of the perceptron.

Then, Let's state another part of the theorem.

Consider the infinite sequence $w_{i_1}, w_{i_2}, \cdots, 1 \le i_k \le N$ for every k, such that each vector $w_1, \cdots, w_N$ ocurs infinitely often. Recursively construct a sequence of vectors $v_0, v_1, \cdots, v_n$, as follows:

v_0 is arbitrary.

$$v_n = \begin{cases} v_{n-1} & \text{if } (w_{i_n}, v_{n-1}) > \theta \\ v_{n-1} + w_{i_n} & \text{if } (w_{i_n}, v_{n-1}) \le \theta \end{cases} \quad \boxed{(2)}$$

The sequence $\{v_n\}$ is convergent - i.e. for some index m,

$v_m = v_{m+1} = v_{m+2} = \cdots = \tilde{v}$

$v_n = v_{n-1} + w_{i_n}$ and $(w_{i_n}, v_{n-1}) \le \theta$ for each n $\boxed{(3)}$

For this part, we can see that at last $(w_i, \tilde{v}) > \theta$ for $i = 1 \cdots n$

The sequence $\{w_{i_n}\}$ represents the training sequence, and $\tilde{v} = y$

For the proof, we only consider those missclassfied sample, and to do correction.

The paper use contradiction to prove the theorem. That is when n is larger enough, which means infinate loop, then we can get (1) and (3) can not be compatible, which means when n is larger enough, then there will not exist such vector y. thus, contradicting our assumption in the theorem, there exist such vector y. $\Rightarrow$ Proved

In the next page. I will show some detail of the proof.

From the training sequence we can know $V_n = V_0 + W_{i_1} + \cdots + W_{i_n}$

Suppose at step n we find one point W_{i_n} and $W_{i_n} \cdot V_{n-1} \leq 0$

then we make correction $V_n = V_{n-1} + W_{i_n}$

then $y \cdot V_n = y \cdot (V_{n-1} + W_{i_n}) = y \cdot V_{n-1} + y \cdot W_{i_n}$

From the inequality (1), we can know $(W_i, y) > \theta$

$$
\begin{aligned}
\therefore \quad y \cdot V_{n-1} + y \cdot W_{i_n} &> y \cdot V_{n-1} + \theta \\
&> y\,(V_{n-2} + W_{i_{n-1}}) + \theta \\
&> y \cdot V_{n-2} + y \cdot W_{i_{n-1}} + \theta \\
&> y \cdot V_{n-2} + 2\theta \\
&\cdots \\
&> y \cdot V_0 + n\theta
\end{aligned}
$$

$\therefore$ From inequality (1). we get $y \cdot V_n > y \cdot V_0 + n\theta$

The using the Cauchy-Schwarts inequality

$$\|y\|^2 \cdot \|V_n\|^2 \geq (y, V_n)^2$$

$$\therefore \quad \|V_n\|^2 \geq \frac{(y, V_n)^2}{\|y\|^2} > \frac{(y \cdot V_0 + n\theta)^2}{\|y\|^2} = \frac{\theta^2}{\|y\|^2}\left\{n + \frac{(V_0, y)}{\theta}\right\}^2$$

Therefore, we can see when we choose $C = \frac{\theta^2}{\|y\|^2}$, $\|y_n\|^2 > Cn^2$ if $(V_0, y) \geq 0$. If $(V_0, y) < 0$, we choose $C = \frac{1}{4}\frac{\theta^2}{\|y\|^2}$, then $\|y_n\|^2 > Cn^2$ if $(V_0, y) < 0$. Thus, from inequality (1), we can get $\|V_n\|^2 > Cn^2$ $\boxed{(4)}$

From inequality (3), we can see $W_{i_n} \cdot V_{n-1} \leq \theta$

$$\therefore \|V_n\|^2 = (V_{n-1} + W_{i_n})(V_{n-1} + W_{i_n})$$

$$= \|V_{n-1}\|^2 + 2W_{i_n} \cdot V_{n-1} + \|W_{i_n}\|^2$$

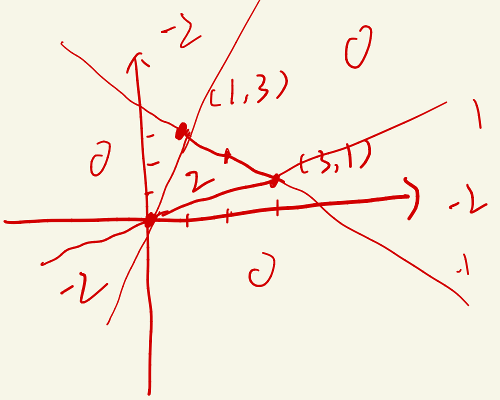
$$\leq \|V_{n-2}\|^2 + 2\underbrace{W_{i_{n-1}} \cdot V_{n-2}}_{\leq \theta} + \|W_{i_{n-1}}\|^2 + 2\underbrace{W_{i_n} \cdot V_{n-1}}_{\leq \theta} + \|W_{i_n}\|^2$$

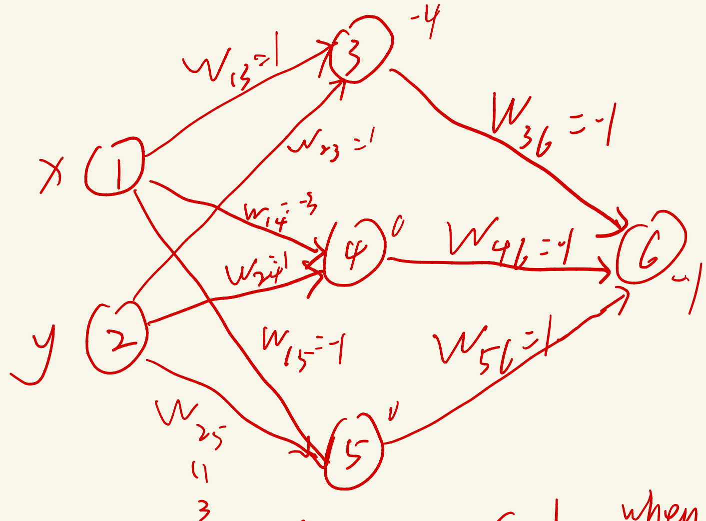
$$\vdots$$

$$\leq \|V_0\|^2 + (2\theta + M)n, \text{ where } M = \max_{i=1,\cdots n} \|W_i\|^2 \quad (5)$$

At last, when we consider n is large enough, then from (4) $\|V_n\| > Cn^2$, from (5) $\|V_n\|^2 \leq \|V_0\|^2 + (2\theta + M)n$, we know $\|V_0\|$, θ and M are constants; therefore, (4) and (5) can not be compatible when n is large enough, which means, when n is infinite, then there will not exist a perfect assignment vector y, which contradicts our theorem first part, thus proved.

Problem 2





$-1+2-1$

$-1-1+1-1$

$1-1+1-1=0$

$1+1+1-1=2$

$1-1+1-1=0$

$-1+1-1-1=-2$

The activation function for vertex 3, 4, 5 $f(x) = \begin{cases} 1 & \text{when } x > 0 \\ -1 & \text{when } x < 0 \\ 0 & \text{when } x = 0 \end{cases}$

The bias we choose for vertex 3, 4, 5, 6 are b_3, b_4, b_5, b_6

we have known that one perception is a linear classfier. and a triangle is composed by three line. Therefore, we can consider each line is represented by one perceptron.

the triangle is $\begin{cases} y = -x+4 \\ y = 3x \\ y = \frac{1}{3}x \end{cases}$

Firstly, we consider 3 is $x+y-4=0$ then $w_{13}=1$ $w_{23}=1$ $b_3=-4$

Then, we consider 4 is $y-3x=0$ then $W_{14}=-3$ $W_{24}=1$ $b_4=0$

At last, we consider 5 is $3y-x=0$ then $W_{15}=-1$ $W_{25}=3$ $b_5=0$

∵ (1,3) is on the decision boundary

∴ Then we design $w_{56}=1$ $b_6=-1$

∵ (3,1) is on the decision boundary ∴ $w_{46}=1$

∵ (0,0) is on the decision boundary ∴ $w_{36}=-1$

Then we can see that the area in the triangle output is 2

∴ we design the activation function of vertex is

$$f_6(x) = \begin{cases} 1 & \text{when } x = 2 \\ 0 & \text{when } x < 2 \end{cases}$$

Problem 3.

$$w^{(k+1)} = w^{(k)} + \Delta w^{(k)}$$

$$\Delta w^{(k)} = \eta \left(t^{(k)} - w^{(k)} x^{(k)} \right) \frac{x^{(k)}}{\|x^{(k)}\|^2}$$

$$\Delta w^{(k+1)} = \eta \left(t^{(k+1)} - w^{(k+1)} x^{(k+1)} \right) \frac{x^{(k+1)}}{\|x^{(k+1)}\|^2}$$

$$\because x^{(k)} = x^{(k+1)} \quad \therefore t^{(k+1)} = t^{(k)}$$

$$\therefore \Delta w^{(k+1)} = \eta \left\{ t^{(k)} - \left[w^{(k)} + \Delta w^{(k)} \right] x^{(k)} \right\} \frac{x^{(k)}}{\|x^{(k)}\|^2}$$

$$= \eta \left[t^{(k)} - w^{(k)} x^{(k)} - \Delta w^{(k)} x^{(k)} \right] \frac{x^{(k)}}{\|x^{(k)}\|^2}$$

$$= \eta \left[t^{(k)} - w^{(k)} x^{(k)} \right] \frac{x^{(k)}}{\|x^{(k)}\|^2} - \eta \Delta w^{(k)}$$

$$= \Delta w^{(k)} - \eta \Delta w^{(k)} = (1 - \eta) \Delta w^{(k)}$$

Problem 4

See the 94.pdf