EE559 - MPR

Homework - 4 - Written Report

RAKSHITHA PANDURANGA
rpandura@usc.edu
7890-1614-34

1)

a) w.k.t $g(x) = w_0 + \underline{w}^T \underline{x}$

The decision surface is a hyperplane, if $g(x)$ is linear.
If x_1 and x_2 are on H, then,

Taking, $g(x_1) - g(x_2) = 0$

$$w^t x_1 + w_0 - (w^t x_2 + w_0) = 0$$

$$w^t x_1 - w^t x_2 = 0$$

$$w^t (x_1 - x_2) = 0$$

∴ The dot product is 0

∴ w is normal to any vector lying in the hyperplane

b) $$d_H = \frac{w^T x + w_0}{\|w\|}$$

Let $\underline{x} = \underline{x}_1 + a\underline{w}$

$$d_H = \frac{\underline{w}^T (x_1 + aw) + w_0}{\|w\|}$$

$$= \frac{\underline{w}^T x_1 + w^T a\underline{w} + w_0}{\|w\|}$$

$$= \frac{(\underline{w}^T x_1 + w_0) + \underline{w}^T a\underline{w}}{\|w\|}$$

$$= \frac{\underline{w}^T \underline{w} a}{\|\underline{w}\|}$$

$$= \frac{\|\underline{w}\|^2 a}{\|\underline{w}\|}$$

$$= \|\underline{w}\| a$$

↓

constant

$\therefore$ $\underline{w}$ points to the +ve side of H if $a > 0$

c) w.k.t $d_H = \frac{g(x)}{\|\underline{w}\|}$ → distance b/w a point $x^{(t)}$ and a hyperplane

$$= \frac{\underline{w}^t \underline{x} + w_0}{\|\underline{w}\|}$$

In augmented space $= \frac{\underline{w}^{(t)} x^{(+)}}{\|\underline{w}\|}$ ~~.~~

$$\boxed{d_H = \frac{g(x^{(+)})}{\|w\|}}$$

w_0 is included in w,

$$\underline{w}^{(+)} = \begin{bmatrix} \underline{w_0} \\ \underline{w}_1 \end{bmatrix} \qquad \underline{x}^{(+)} = \begin{bmatrix} 1 \\ \underline{x} \end{bmatrix}$$

d) weight space at arbitrary point $w^{(t)}$ in a hyperplane $g(x^{(+)})$

$$g(x^{(+)}) = w^{(+)T} x^{(+)} = 0$$

For, $x^{(+)} \perp H$

$\underline{w}_1^{(+)}$ and $w_2^{(+)}$ lie on the hyperplane $g(x^{(+)}) = 0$ — ②

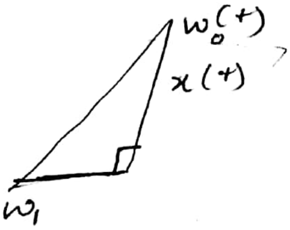
$\therefore$ $\underline{w}_1^{(+)T} x^{(+)} = 0$

$\underline{w}_2^{(+)T} x^{(+)} = 0$

$$\underline{w}_1^{(+)T} x^{(+)} - \underline{w}_2^{(+)T} x^{(+)} = 0$$

$$(\underline{w}_1^{(+)} - (\underline{w}_2^{(+)}) x^{(+)} = 0$$

$$\Rightarrow x^{(+)} \perp H$$



r is the projection of $\underline{w}^{(+)} - w_1^{(+)}$ onto $x^{(+)}$

$$r = \frac{(\underline{w}^{(+)} - \underline{w}_1^{(+)}) \cdot \underline{x}^{(+)}}{\|(\underline{x}^{(+)}\|}$$

$$= \frac{(\underline{w}^{(+)} - \underline{w}_1^{(+)})^T \underline{x}^{(+)}}{\|\underline{x}^{(+)}\|}$$

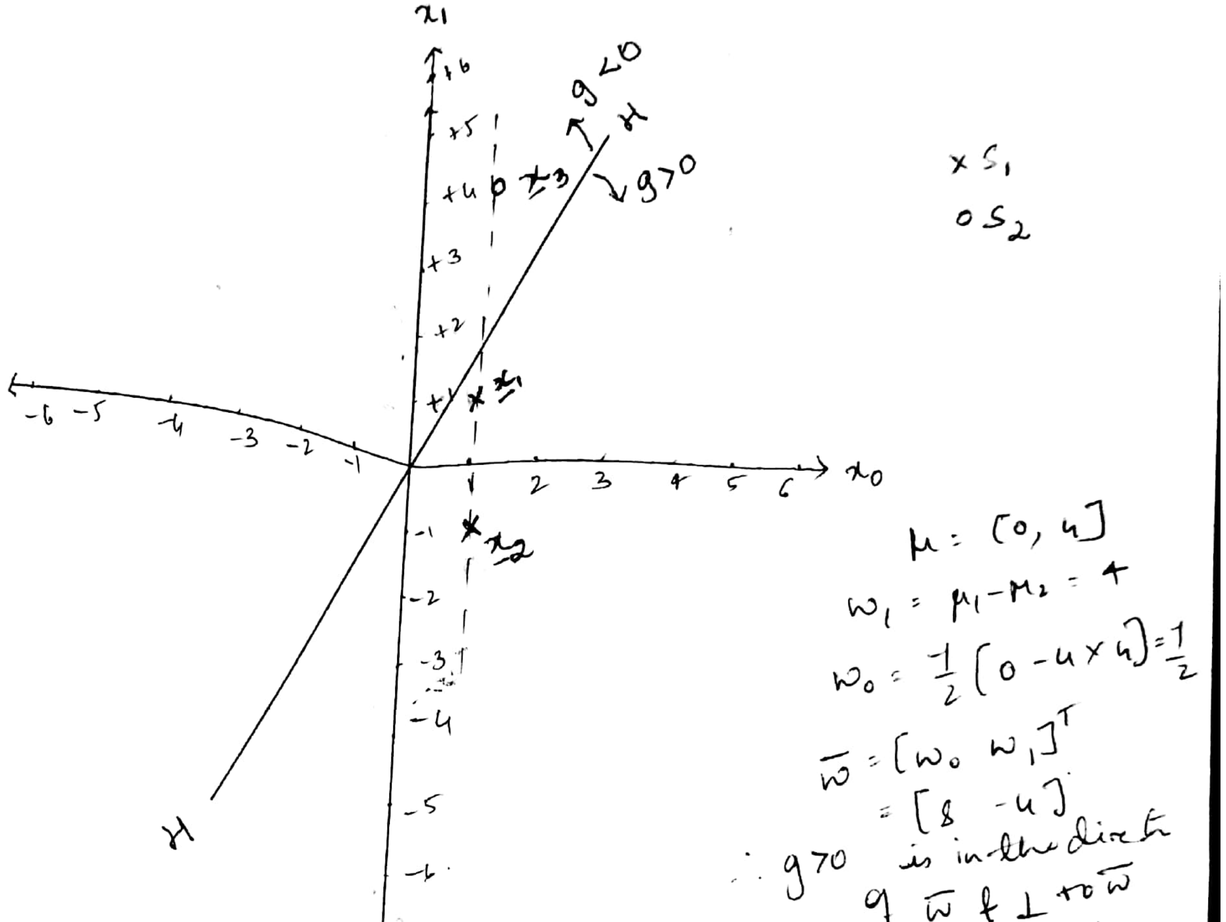
$$= \frac{\underline{w}^{(+)T} x^{(+)} - \underline{w}_1^{(+)T} x^{(+)}}{\|\underline{x}^{(+)}\|}$$

$$= \frac{w^{(+)T} x^{(+)}}{\|w^{(+)}\|}$$

when $r > 0$, it lies on the positive side of hyperplane also depends on which class $x^{(+)}$ is in.

2)

a)



$\mu = [0, 4]$

$w_1 = \mu_1 - \mu_2 = 4$

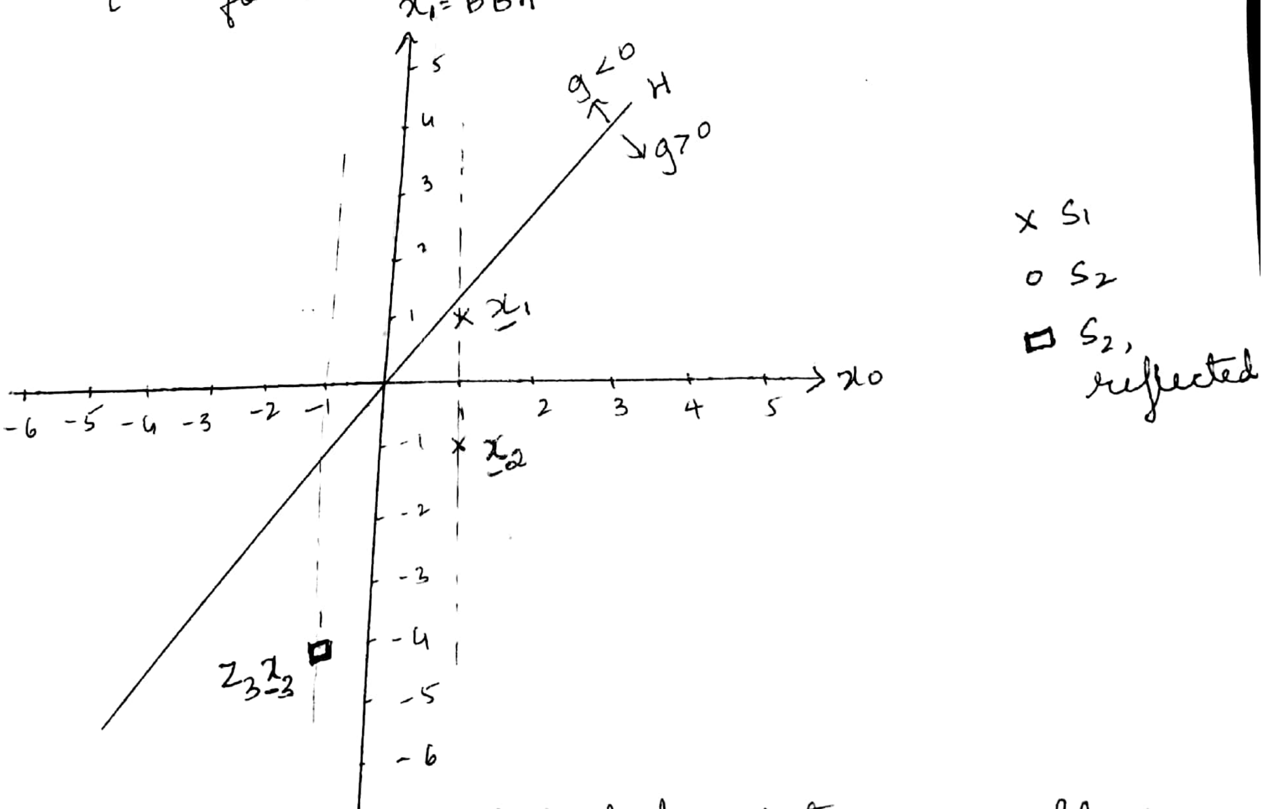
$w_0 = \frac{1}{2}[0 - 4 \times 4] = \frac{1}{2}$

$\bar{w} = [w_0\ w_1]^T$

$= [8\ \ -4]$

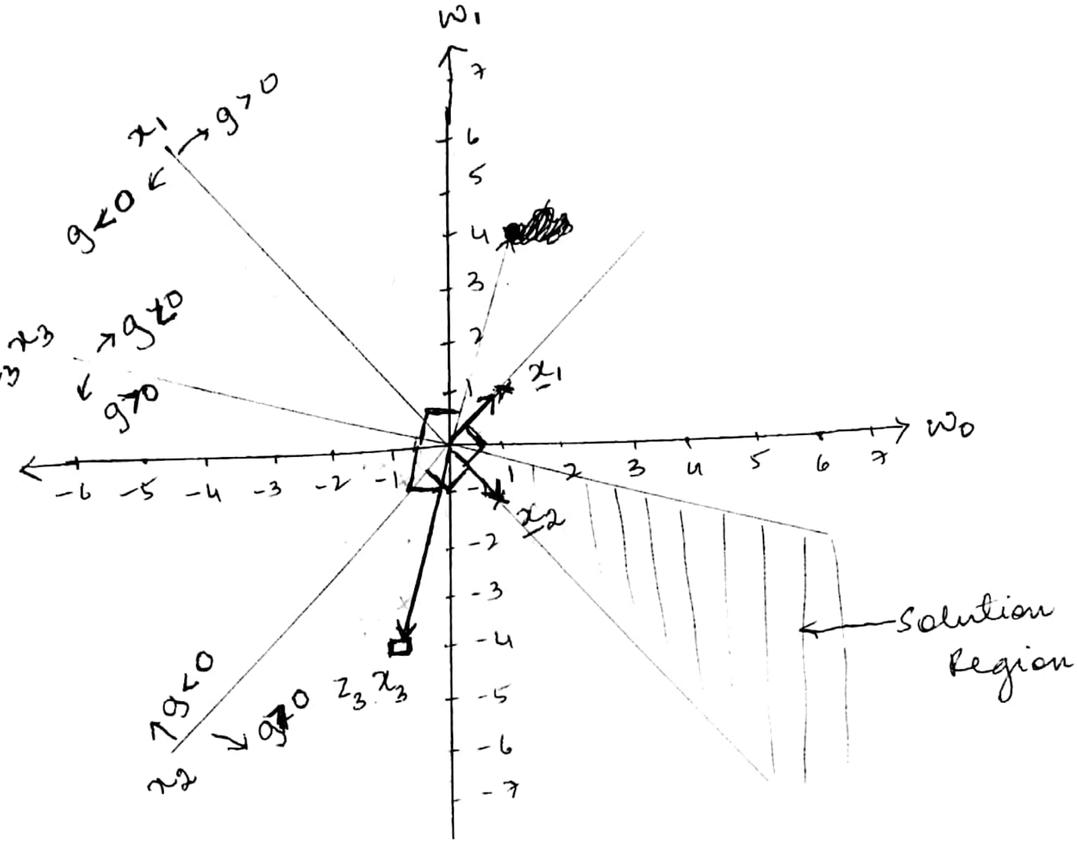
$\therefore g>0$ is in the directn of $\bar{w}$ & $\perp$ to $\bar{w}$

b) Since $Z_n^{(K)} = \begin{cases} +1 & \text{for } S_1 \\ -1 & \text{for } S_2 \end{cases}$, only $x_3^{(2)} = (1, 4)$ gets reflected



H ~~correct~~ classifies the reflected datapoints correctly since it lies on $g>0$

c)



d) $\mu_1 = 0 \quad \mu_2 = 4 \quad w_1 = \mu_1 - \mu_2 = 0 - 4 = -4 \quad w_0 = -\frac{1}{2}[0 - 4 \times 4]$

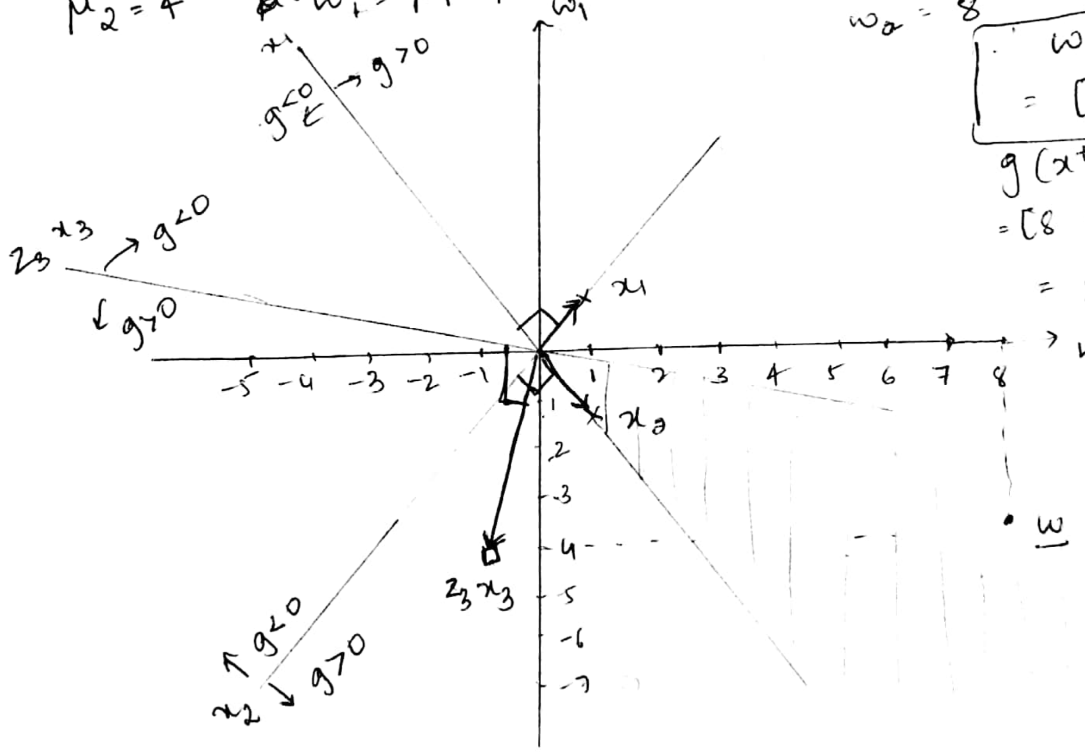
$w_0 = 8$

$$\boxed{w = [w_0 \ w_1]^T = [8, -4]^T}$$

$g(x^+) = w^{+T} x^+$

$= [8 \ -4] \begin{bmatrix} x_0 \\ x_1 \end{bmatrix}$

$= 8x_0 - 4x_1$



Yes, the weight vector $\underline{w}$ lies in the solution region of the weight space

③ (a)

Given → $P(\underline{x})$ is a scalar function of $\underline{x}$
$f(p)$ is a scalar function of p

$\therefore$

$$\nabla f[p(\underline{x})] = \begin{bmatrix} \frac{\partial f[p(\underline{x})]}{\partial p} \\ \vdots \\ \frac{\partial}{\partial p} f[p_d(\underline{x})] \end{bmatrix}$$

Taking for the i^{th} component,

$$\Rightarrow \frac{\partial f}{\partial p}[p(\underline{x})]$$

Applying chain rule,

$$\frac{\partial}{\partial p} f(p) \cdot D[p(\underline{x})]$$

$$\left[\frac{\partial}{\partial p} f(p)\right] \cdot \nabla p(\underline{x})$$

$$\left[\frac{\partial}{\partial p} f(p)\right] D[p(\underline{x})]$$

$$\frac{\partial}{\partial p} f(p) \cdot \nabla p(\underline{x})$$

$$\therefore \boxed{\nabla f(p(\underline{x})) = \frac{\partial}{\partial p} f(p) \cdot \nabla p(\underline{x})}$$ Proved

b)

Relation given,

$$\frac{\partial}{\partial x}\left[x^T M x\right] = (M + M^T)\underline{x}$$

$\cancel{\frac{\partial}{\partial x}\,x}$

Using this

$$\underline{x}^T M \underline{x} = \underline{x}^T \underline{x}$$

when,

$$M = \begin{bmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{bmatrix}_{d \times d}$$

$$\frac{\partial}{\partial x}\left[\underline{x}^T M \underline{x}\right] = \frac{\partial}{\partial t}\left[\underline{x}^T \underline{x}\right] = \left[\begin{bmatrix} 1 & & & 0 \\ & 1 & & \\ 0 & & 1 & \\ & & & 1 \end{bmatrix}_{d \times d} + \begin{bmatrix} 1 & & & 0 \\ & 1 & & \\ 0 & & 1 & \\ & & & 1 \end{bmatrix}_{d \times d}\right]\underline{x}$$

$$= \left[2 I_{d \times d}\right]\underline{x}$$

$$= \begin{pmatrix} 2 & & & 0 \\ & 2 & & \\ & & \ddots & \\ 0 & & & 2 \end{pmatrix}\underline{x}$$

$$\boxed{\nabla_{\underline{x}}\left(\underline{x}^T \underline{x}\right) = 2\underline{x}}$$

c) $\nabla_{\underline{x}}(\underline{x}^T \underline{x})$ $\underline{x} = (\underline{x}_1 \cdots \underline{x}_d)$

$$= \nabla_{\underline{x}} \sum_{i=1}^{d} \underline{x}_i^2$$

$$= \frac{\partial}{\partial \underline{x}} \sum_{i=1}^{d} \underline{x}_i^2$$

$$= \sum_{i=1}^{d} \frac{\partial}{\partial \underline{x}} \underline{x}_i^2$$

$$= \sum_{i=1}^{d} \partial x_i^2 = 2(x_1, x_2 \cdots x_n) = 2\underline{x}$$ //

d)

w.k.t

a) $\nabla_{\underline{x}} f[p(\underline{x})] = \left[\frac{\partial}{\partial p} f(p)\right] \nabla_{\underline{x}} p(\underline{x})$

b) $\nabla_{\underline{x}}(\underline{x}^T \underline{x}) = 2\underline{x}$

$\therefore \nabla_{\underline{x}} \left[(\underline{x}^T \underline{x})^3 \right] \Rightarrow$ $p = \underline{x}^T \underline{x}$

$f(p) = (\underline{x}^T \underline{x})^3$

$$\therefore \frac{\partial}{\partial x} (\underline{x}^T \underline{x})^3 \cdot \nabla_{\underline{x}} (\underline{x}^T \underline{x})$$

$$= 3(\underline{x}^T \underline{x})^2 \underbrace{\frac{\partial}{\partial x} \underline{x}^T \underline{x}}_{2\underline{x}}$$

$$= 3(\underline{x}^T \underline{x})^2 \times 2\underline{x}$$

$$= 6\underline{x}(\underline{x}^T \underline{x})^2$$

$$= 6\underline{x} \underbrace{\left[\sum_{i=0}^{n} x_i^2\right]}_{c}$$

$$= 6\underline{x}\,c$$ //

4)

a) $\nabla_{\underline{w}} \|\underline{w}\|_2$

$$\|\underline{w}\|_2 = \sqrt{\underline{w}^T \underline{w}}$$

$$p(w) = \underline{w}^T \underline{w}$$

$$f(p) = \sqrt{\underline{w}^T \underline{w}}$$

$$\nabla_{\underline{w}} \|\underline{w}\|_2 = \nabla_{\underline{w}} f[p(\underline{w})] = \left[\frac{d}{dp} f(p)\right] \nabla_{\underline{w}} p(\underline{w})$$

$$= \frac{d}{dp}\left[\sqrt{\underline{w}^T \underline{w}}\right] \cdot \underbrace{\nabla_w [\underline{w}^T \underline{w}]}_{2\underline{w}}$$

$$= \frac{1}{\cancel{2}} (\underline{w}^T w)^{-1/2} \cdot \cancel{2}\underline{w}$$

$$= \underline{w} (\underline{w}^T \underline{w})^{-1/2}$$

$$\boxed{\nabla_{\underline{w}} \|\underline{w}\|_2 = \frac{\underline{w}}{\|\underline{w}\|_2}}$$

b) $\nabla_{\underline{w}} \|M\underline{w} - \underline{b}\|_2$

$$\|M\underline{w} - b\|_2 = \sqrt{(M\underline{w} - b)^T (M\underline{w} - \underline{b})}$$

$$= \sqrt{\underline{w}^T M^T M \underline{w} - b^T M \underline{w} - \underline{w}^T M^T b + b^T b}$$

$$= \sqrt{b^T b - 2(M^T b)^T w + w^T M^T M b}$$

b) $\nabla_{\underline{w}} \| M\underline{w} - b \|_2$

$$\| M\underline{w} - b\|_2 = \sqrt{(M\underline{w}-b)^T (M\underline{w}-b)}$$

Let $p = \underbrace{(\underline{\underline{M}}\underline{w}-b)}_{x}{}^T (\underline{\underline{M}}\underline{w}-b) = x^T x$

$$f(p(w)) = p^{1/2}$$

$$\nabla_w \| \underline{\underline{M}}\underline{w} - b\|_2 = \left[\frac{d}{dp} f(p)\right] \cdot \nabla_w p(w)$$

$$= \frac{d}{dp}\left[\sqrt{x^T x}\right] \cdot \nabla_w \underbrace{[x^T x]}_{2x} \frac{d}{dw}(\underline{\underline{M}}\underline{w}-b)$$

$$= \frac{1}{\cancel{2}}(x^T x)^{-1/2} \cdot \cancel{2} x \; (\underline{\underline{M}})$$

$$= \underline{x} \; (\underline{x}^T \cdot \underline{x})^{-1/2} \; (\underline{\underline{M}})$$

$$= \frac{(\underline{\underline{M}}\underline{w}-b)}{\sqrt{(M w-b)^T (M w-b)}} \cdot \underline{\underline{M}}$$

$$\boxed{\nabla_w \| \underline{\underline{M}}\underline{w} - b\|_2 = \frac{(\underline{\underline{M}}\,\underline{w}-\underline{b})}{\| \underline{\underline{M}}\,\underline{w} - b\|_2} \underline{\underline{M}}}$$

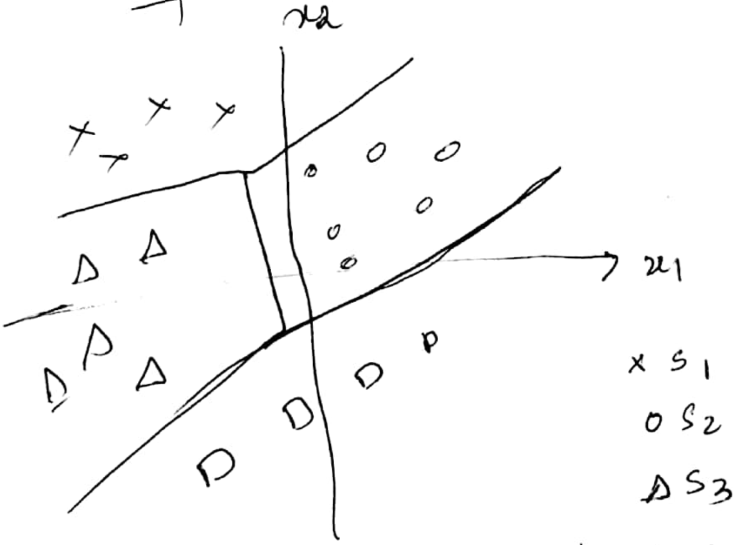
5) Extra Credit problem.

For $c > 2$

Totaly linearly separable implies that all datapoints $x_m^{(i)}$ of class S_i can be separated from all datapoints of all other classes by a linear boundary and this holds $\forall i, i = 1, 2 \ldots c$. Then the data are totaly linearly separable.

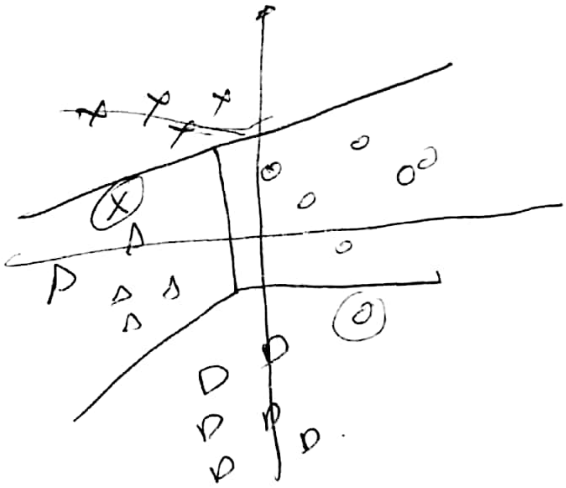
But linear separability implies that if all $g(x)$'s are linear functions, then the data points are linearly separable

Totaly linear separability



Requires that all $g(x)$'s are linear and all points are classified right

linear separability



linear separability requires just linear functions to classify.