SOURABH TIRODKAR.

EE559 HW#3

1) $g(x) = w_0 + w^T x = 0$

$\downarrow$ discriminant func^n

a) P.T. → weight factor $\underline{w}$ is normal to H.

$g(x_1) = w_0 + w^T x_1 = w_0 + w^T x_2 = 0$

$w^T(\underline{x_1 - x_2}) = 0$

Any vector 'v' bet^n x_1 & x_2

$v = x_1 - x_2$

$\downarrow$ lies in same plane as of x_1 & x_2.

$\therefore$ $w^T v = 0$

⇒ weight vector w is normal to vector space v.

⇒ $g(v) = 0$

⇒ v lies on the hyperplane decision boundary.

$\therefore$ $\underline{w}$ is normal to the hyperplane H.

b) S.T. vector $\underline{w}$ points to the positive side of H

Hyperplane divides the feature space:
2 regions for 2 class problem.
c regions for c-class problem.

If a pt x_1 on the hyperplane.
$\Rightarrow g(x_1) = 0$

$x = \underline{x}_1 + a\underline{w} \qquad a > 0$

$g(x) = g(x_1 + a\underline{w}) = g(x_1) + a\,g(\underline{w})$
$= a(w_0 + w^T w)$
$= a w_0 + w^2 > 0.$
$\therefore \underline{g(x) > 0}$

$\Rightarrow$ x lies on the +ve side of hyperplane

$g(x) > 0 \Rightarrow w^T x + w_0 > 0 \Rightarrow$
$w^T x > 0$
$w^T \underline{x} = w \cdot x > 0$

Dot product $\Rightarrow$ +ve
implies both vectors point in the same direction.

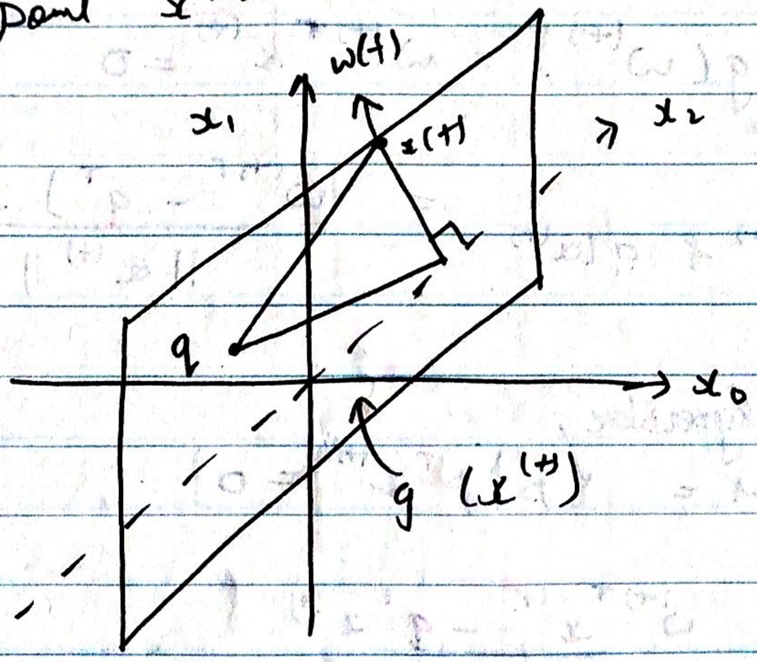
So, a point to Γ_1 (+ve side)

$\rightarrow \underline{w}$ ~~also~~ lies in Γ_1

(+ve side of hyperplane)

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

point $x^{(+)}$



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

Augmented weight vector $w^{(+)} = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix}$

Augmented sample vector $x^{(+)} = \begin{pmatrix} 1 \\ x_1 \end{pmatrix}$

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

$$r = \frac{w^{(+)T} x^{(+)}}{\| w^{(+)} \|} \qquad w^{(+)T} q = 0$$

$g(x^{(+)}) > 0 \quad \forall \; x^{(+)}$ in the +ve side of hyperplane.

$\Rightarrow$ ℓ points in the direction of side of hyperplane.

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

$$\text{Dist}_{w^{(+)} \,\&\, g(x^{(+)})} = \frac{(w^{(+)T} - q^T)\, x^{(+)}}{\|x^{(+)}\|}$$

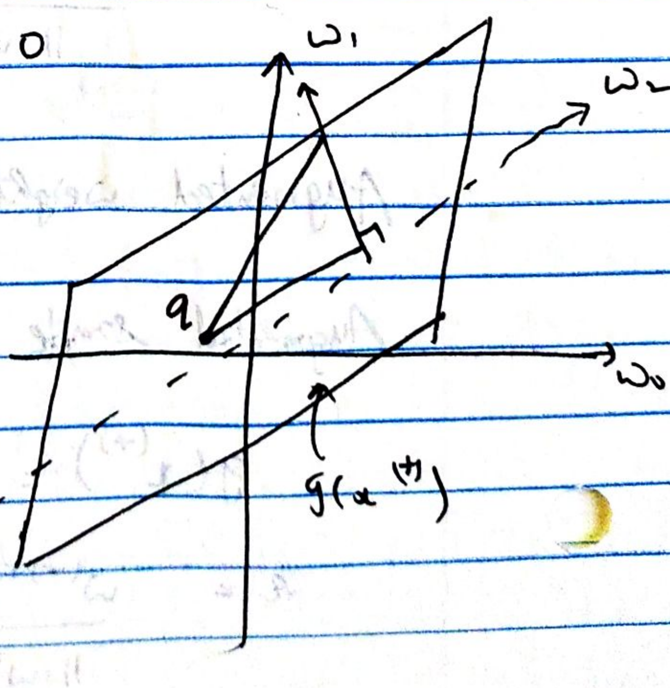
In hyperplane,

$$H = \{ p \mid p^T x^{(+)} = 0 \}$$

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

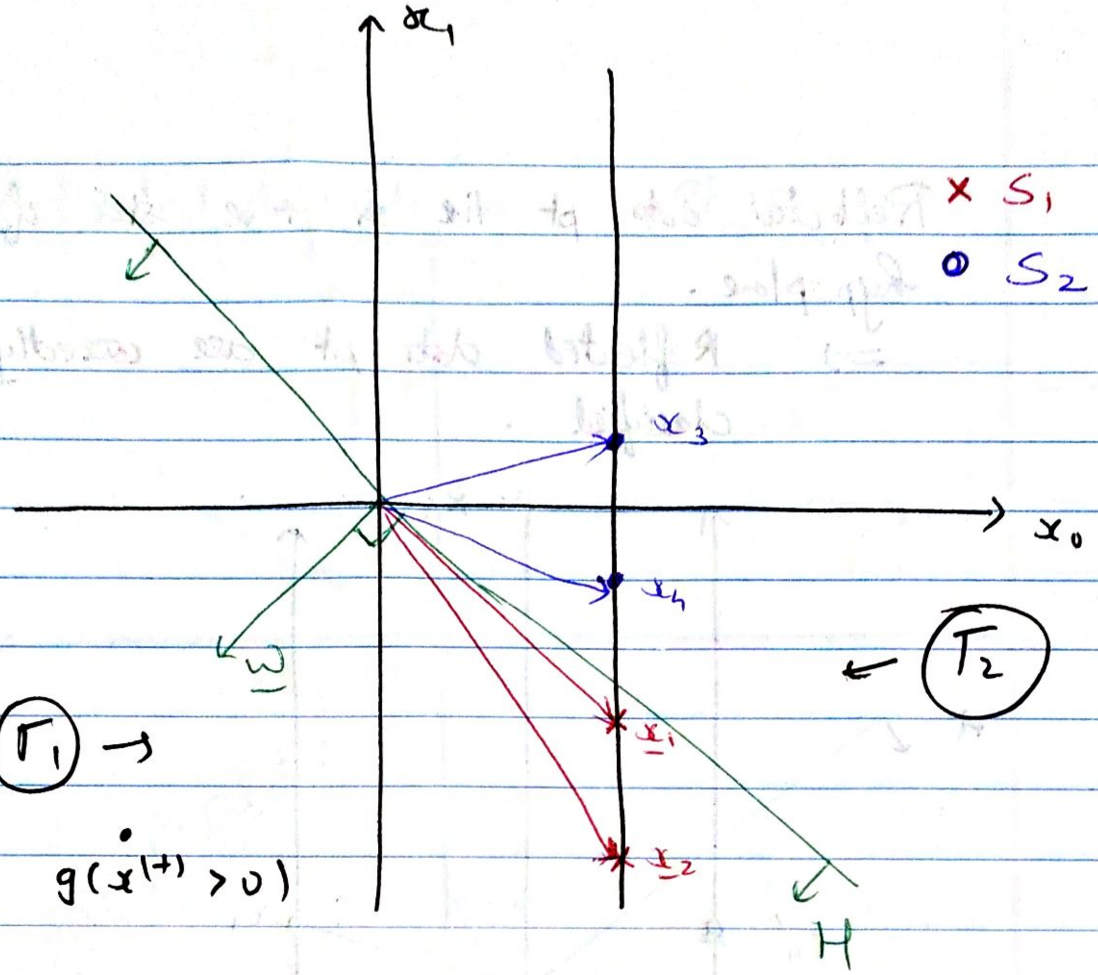
$q \in g(x^{(+)}) \;\therefore\; q^T x^{(+)} = 0$

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



Q.2)

a)



b) Reflected data points :

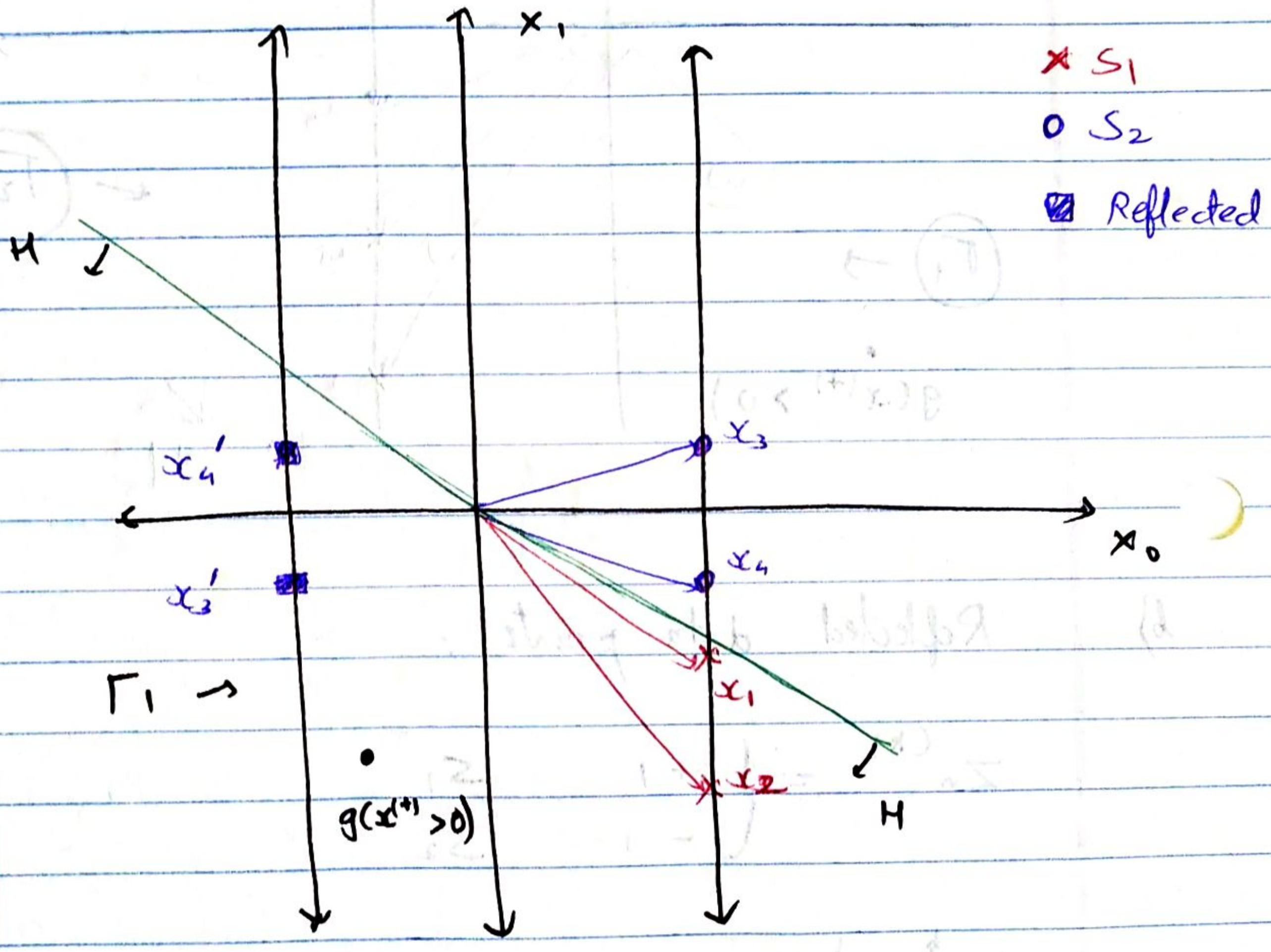
$$z_n^{(k)} = \begin{cases} +1 & S_1 \\ -1 & S_2 \end{cases}$$

$$z_n^k x_n^k = Z_n X_n \rightarrow \text{reflected data points}$$

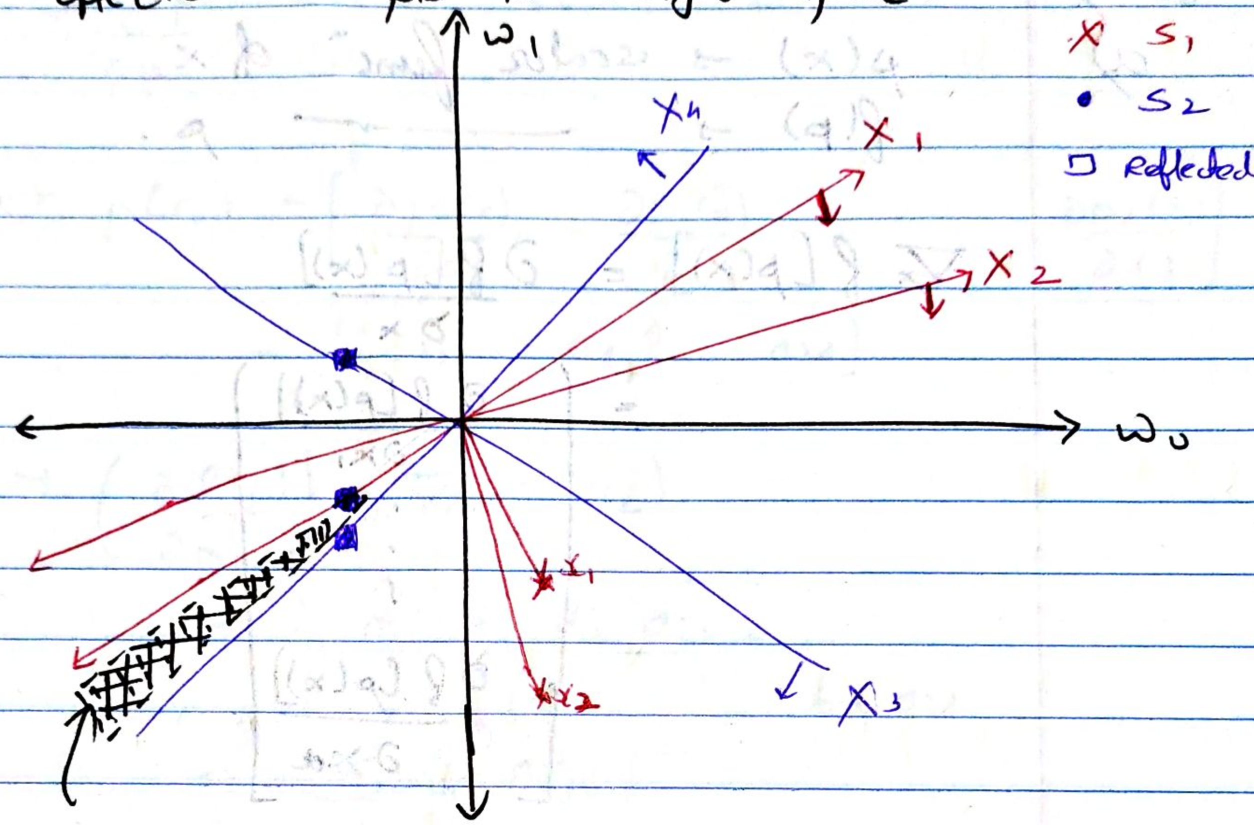
$$\underset{\text{reflected data pt matrix}}{X} = \begin{bmatrix} (1, -3) \\ (1, -5) \\ (-1, -1) \\ (-1, 1) \end{bmatrix}$$

Reflected data pt lie on +ve side of hyperplane.

$\Rightarrow$ Reflected data pt are correctly classified.

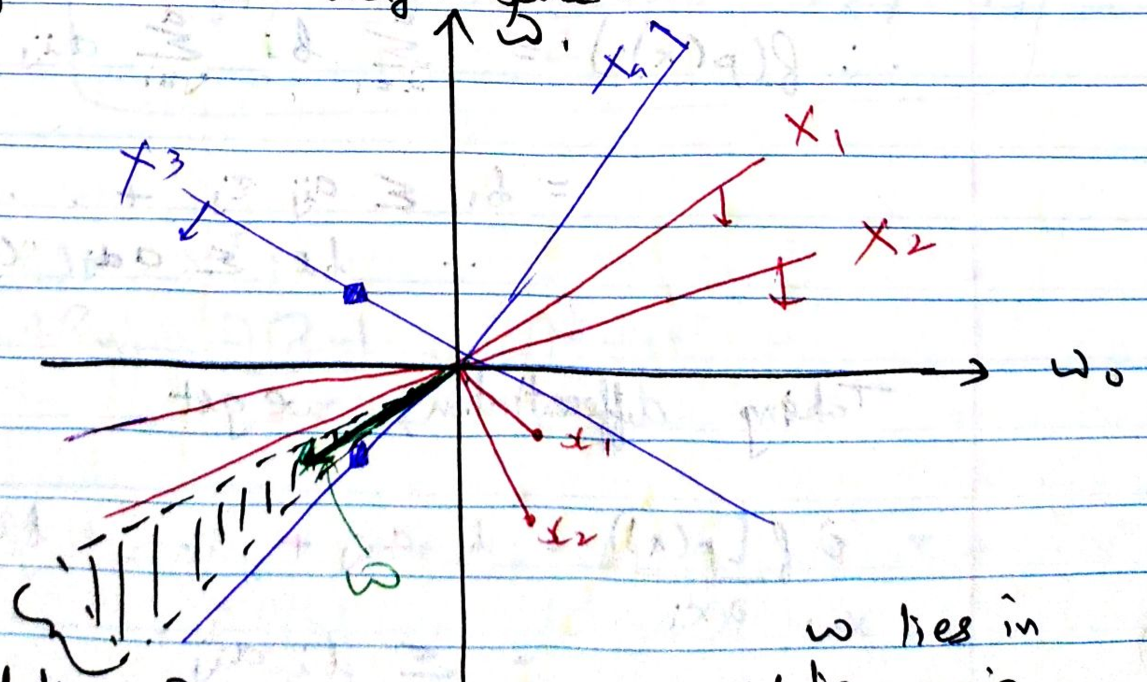


c) Reflected data pts in weight space.



Solution Region

d) Weight vector in weight space:



Solution Region

w lies in solution region

Q.3) a)

$p(x) \rightarrow$ scalar func^n of x
$f(p) \rightarrow$ ———— p.

$$\nabla_x f[p(x)] = \frac{\partial f[p(x)]}{\partial x}$$

$$= \begin{bmatrix} \frac{\partial f[p(x)]}{\partial x_1} \\ \vdots \\ \frac{\partial f[p(x)]}{\partial x_d} \end{bmatrix}$$

if $p_j(x) = \sum a_{ji} x_i$
& $f(p) = \sum b_i p_i$

$$\therefore f(p(x)) = \sum_{i=1}^{d} b_i \sum_{j=1}^{d} a_{ij} x_j$$

$$= b_1 \sum a_{ij} x_j + \dots$$
$$\dots b_d \sum a_{dj} x_j$$

Taking differentiation, we get

$$\frac{\partial f[p(x)]}{\partial x_j} = b_1 a_{ij} + \dots \quad b_d a_{ij}$$

$$= \sum \underline{b_i a_{ij}}$$

$$\frac{\partial f(p)}{\partial p} = [b_1 \quad b_2 \quad \ldots \quad b_d]^T$$

$$\nabla_x \, p(x_i) = \left[\frac{\partial p_1(x)}{\partial x_i} \quad \frac{\partial p_2(x)}{\partial x_i} \quad \ldots \quad \frac{\partial p_d(x)}{\partial x_i} \right]$$

$$= [a_{i1} \quad a_{ij} \ldots \quad a_{id}]$$

$$\Rightarrow \left(\frac{\partial f(p)}{\partial p}\right) \cdot \nabla_x \, p(x_i)$$

$$= b_1 a_{i1} + b_2 a_{i2} + \ldots b_d a_{id}$$

$$= \sum b_j a_{ij}$$

$$= \nabla_x f[p(x)]$$

$$\therefore \boxed{\nabla_x f[p(x)] = \left(\frac{d}{dp} f(p)\right) \nabla_x \, p(x_i)}$$

b) $f(x) = x^T M x$

then $\dfrac{\partial f(x)}{\partial x} = (M + M^T) x$.

If $M = I$, then $f(x) = x^T I x = x^T x$

$$\frac{\partial f(x)}{\partial x} = (I + I^T) x$$

$= 2Ix$

$= 2x$

$\therefore \boxed{\nabla_x (x^T x) = 2x}$

c) $f(x) = x^T x = x^T I x$

$= \sum x_i \left[\sum I_{ij} x_j \right]$

$= \sum x_i (x_i)$

$= \sum x_i^2$

$\therefore f(x) = \sum x_i^2$

$$\frac{\partial f(x)}{\partial x_j} = \frac{\partial \sum x_i^2}{\partial x_j} = 2x_j$$

$$\boxed{\frac{\partial f(x)}{\partial x} = 2x}$$

d) $\nabla_x [(x^T x)^3] = \nabla_x f[p(x)]$

$p(x) = x^T x$ & $f(p) = p^3$

$$\nabla_x f[p(x)] = \left(\frac{d}{dp} f(p)\right) \nabla_x p(x)$$

$$\frac{d}{d} f(p) = 3p^2 = 3(x^T x)^2 \quad \Big| \quad \nabla_x p(x) = 2x$$

$$\nabla_x f(p(x)) = 3(x^T x)^2 \cdot 2x$$
$$= 6(x^2)^2 x$$

$$\therefore \boxed{\nabla_x f(p(x)) = 6x^5}$$

4 a) $\|w_2\| = \sqrt{\sum_1 w_i^2}$

$= \sqrt{w^T w} = \sqrt{p}$

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

$$\nabla_w \|w\|_2 = \nabla_w f(p(w))$$

$$= \left(\frac{d}{dp} f(p)\right) \nabla_w p(w)$$

$$\frac{d}{dp} f(p) = \frac{1}{2} p^{-1/2} = \frac{1}{2\sqrt{p}} \quad \Big| \quad \nabla_\omega p(\omega) = 2\omega$$

$$\nabla_\omega \|\omega\|_2 = \frac{1}{2\sqrt{p}} \cdot 2\omega$$

$$= \frac{\omega}{\sqrt{p}} = \frac{\omega}{\|\omega\|_2}$$

$$\nabla_\omega \|\omega\|_2 = \frac{\omega}{\|\omega\|_2}$$

4\. b) $\nabla_\omega \|M\omega - b\|_2$

$$\|M\underline{\omega} - b\|_2$$

$$= \sqrt{(M\omega - b)^T (M\omega - b)}$$

$$= \sqrt{p(g(\omega))} = f\{p[g(\omega)]\}$$

$f(p) = \sqrt{p}$ & $p(g) = g^T g$
& $g(\omega) = M\omega - b$

$$\nabla_\omega f\{p[g(\omega)]\}$$

$$= \left[\frac{d}{dp} f(p)\right] \nabla_\omega p[g(\omega)]$$

$$= \frac{1}{2\sqrt{p}} \; 2gM = \frac{gM}{\sqrt{p}}$$

$$\nabla_w \, p(g(w)) = \left(\frac{d}{dg} p(g)\right) * \nabla_w \, g(w)$$

$$= (2g)\, M.$$

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

Q. 5) Total linear separability

$\Rightarrow \forall \, x_n^{(i)} \in \Gamma_i \quad \forall i = 1, 2, \ldots c$

x_n is data pt of class S_i

Meaning:

Every data pt should lie in the decision region of its corresponding class.

The 2 class should be separated by a linear decision boundary.

So, linear separable means data pts are correctly classified using a linear discriminant function g(x).

For multi-class (for eg - c class) there will be c decision boundaries.
These will separate pts into c classes.

A. Total linearly sepearable.

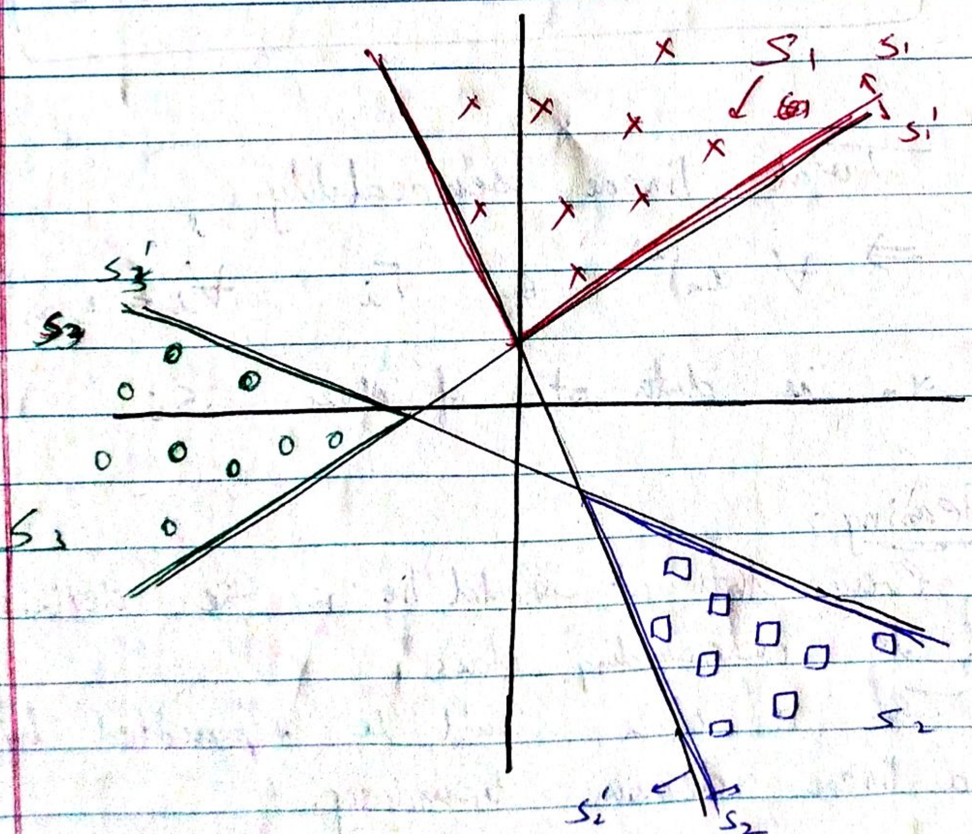


Fig 1

(O V R)
3 class problem.
3 decision boundaries

Using MVM :

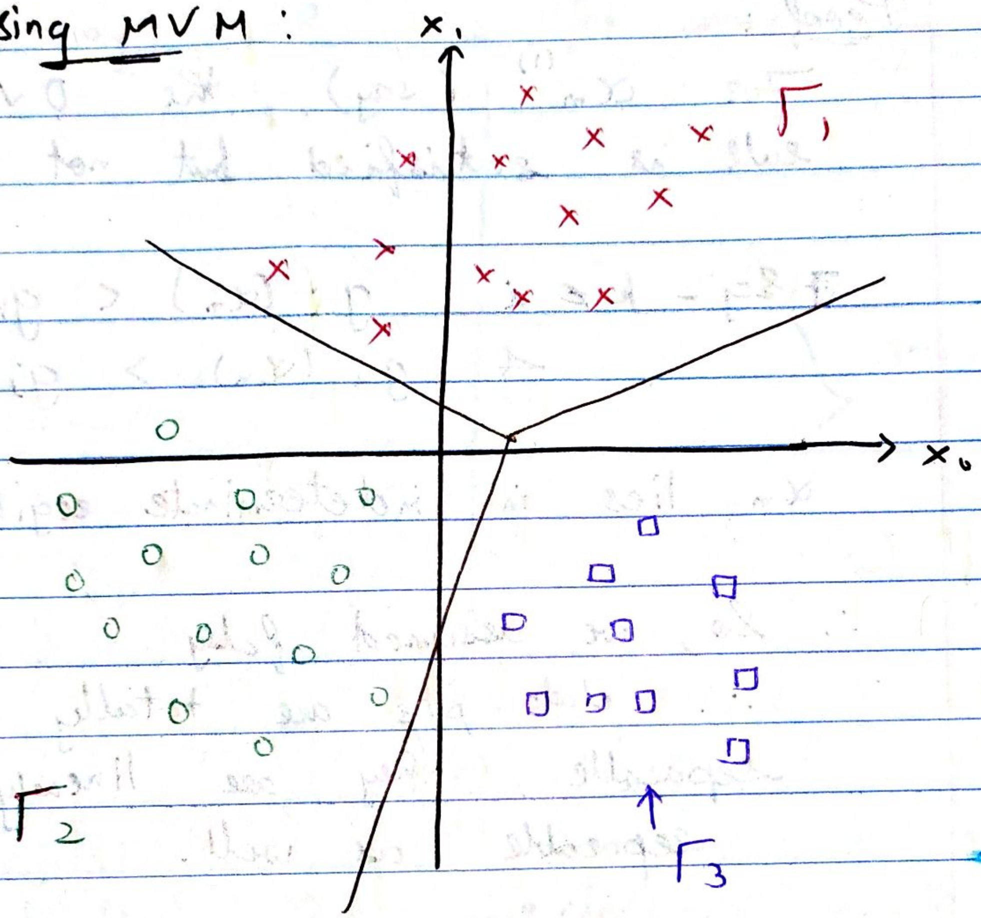


Fig 2

Decision Rule for OVR

$x^{(k)} \in S_k$ iff $x \in S_k$ &

$x \in S_j'$ $\forall_{j \neq k}$.

Decision Rule for MVM :

$x^{(k)} \in S_k$ iff $g_k(x) > g_j(x)$

$\forall_{j \neq k}$

Proof :

For $x_n^{(1)}$ (say), the OVR decision rule is satisfied but not for MVM.

Eg - $k \neq i$ $\quad g_i(x_n) < g_k(x_n)$

& $\quad g_i(x_n) > g_j(x_n)$

x_n lies in indeterminate region in MVM.

∴ So, we assumed falsy

∴ data pts are totally linearly separable, they are linearly separable as well.

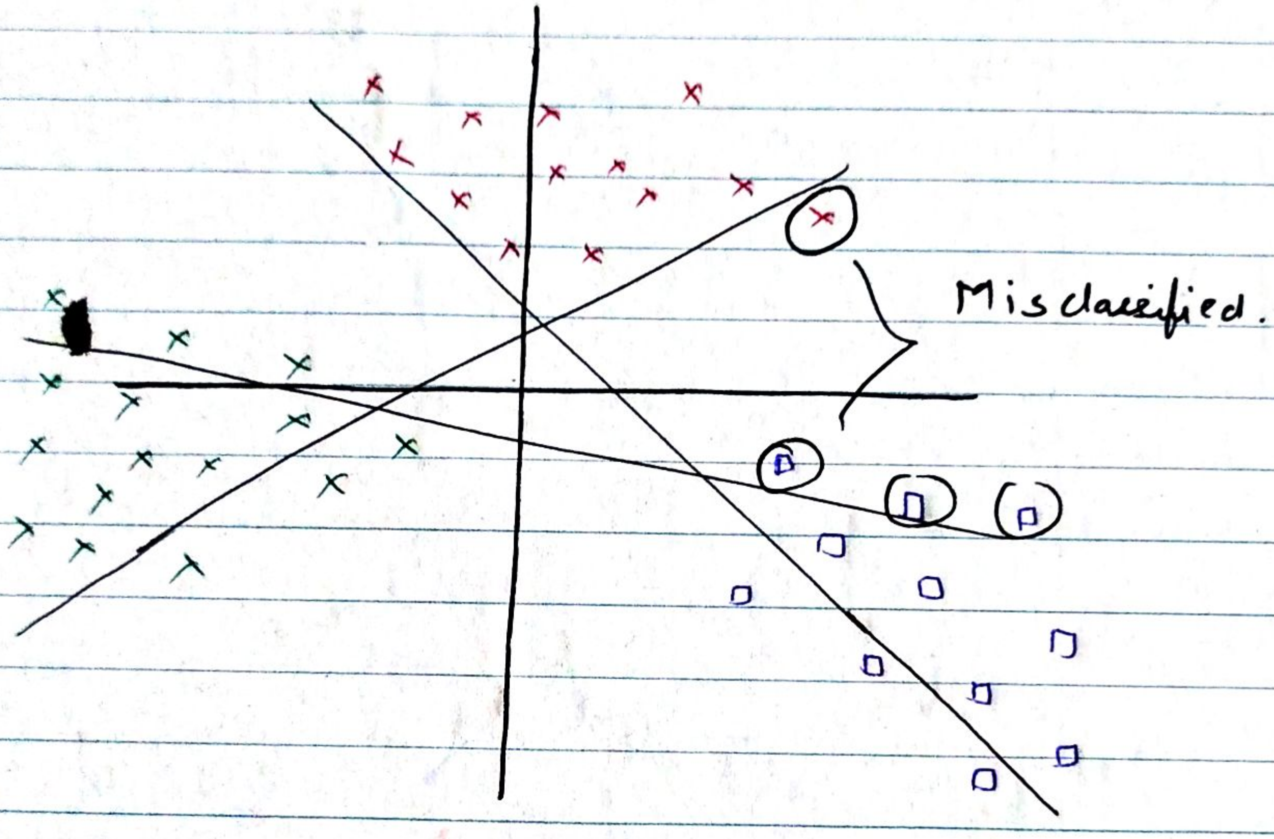
LINEAR SEPARABLE

Taking fig 2, we see that using MVM, the classes are correctly classified.

MVM - linear separable.

We have 3 decision boundaries.

Using OVR show for some data pts.



We can see misclassified points
Data is not totally linearly separable.

$$\boxed{\begin{array}{c}\text{Total linear separable} \\ \Rightarrow \\ \not\Leftarrow \text{ linear separable}\end{array}}$$