Q1. $J(\underline{w}) = \frac{1}{N} \| \underline{\underline{X}}\, \underline{w} - \underline{b} \|_2^2 + \lambda \| \underline{w} \|_2^2$

the purpose of new term is to prefer small $\|w\|_2$ if $\lambda > 0$

a)

$$\nabla_{\underline{w}} J(\underline{w}) = \frac{1}{N} 2 \underline{\underline{X}}^T (\underline{\underline{X}}\, \underline{w} - \underline{b}) + \lambda \nabla_{\underline{w}} (\underline{w}^T \underline{w})$$

$$= \frac{1}{N} 2 \underline{\underline{X}}^T (\underline{\underline{X}}\, \underline{w} - \underline{b}) + \lambda \cdot 2 \underline{w}$$

$$= \frac{2}{N} \underline{\underline{X}}^T \underline{\underline{X}}\, \underline{w} - \frac{2}{N} \underline{\underline{X}}^T \underline{b} + 2 \lambda \underline{w}$$

b)

$$\nabla_{\underline{w}} = 0$$

$$\therefore \frac{2}{N} \underline{\underline{X}}^T \underline{\underline{X}}\, \hat{\underline{w}} - \frac{2}{N} \underline{\underline{X}}^T \underline{b} + 2 \lambda \hat{\underline{w}} = 0$$

$$\therefore \underline{\underline{X}}^T \underline{\underline{X}}\, \hat{\underline{w}} - \underline{\underline{X}}^T \underline{b} + \lambda N \hat{\underline{w}} = 0$$

$$\therefore (\underline{\underline{X}}^T \underline{\underline{X}} + \lambda N) \hat{\underline{w}} = \underline{\underline{X}}^T \underline{b}$$

$$\therefore \boxed{\hat{\underline{w}} = (\underline{\underline{X}}^T \underline{\underline{X}} + \lambda N)^{-1} \underline{\underline{X}}^T \underline{b}}$$

Whereas, the pseudoinverse solution is given by,

$$\hat{\underline{w}} = (\underline{\underline{X}}^T \underline{\underline{X}})^{-1} \underline{\underline{X}}^T \underline{b}$$

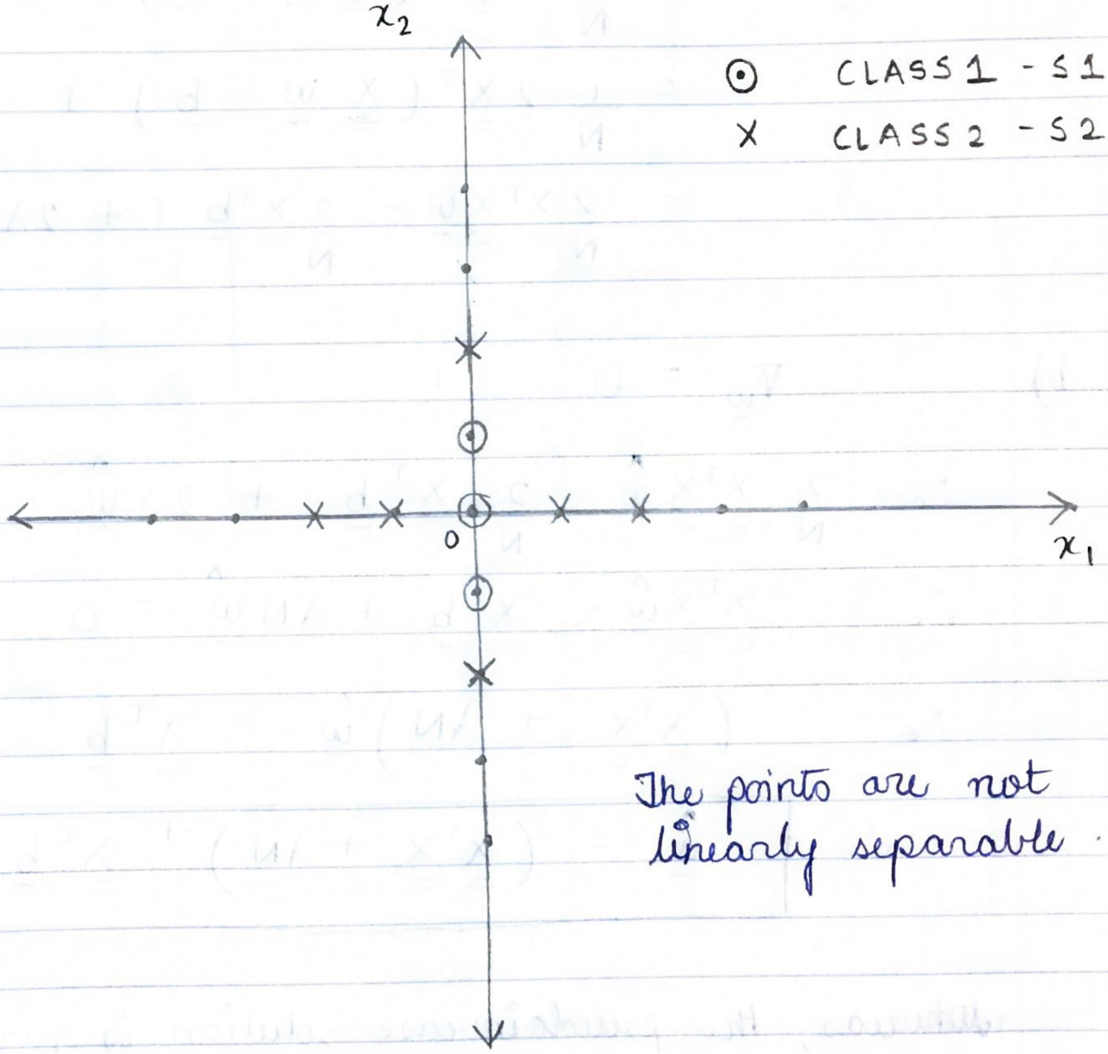
So due to the new term, we get λN inside the inverse.

Q3. $S_1 : (0,0)^T \quad (0,1)^T \quad (0,-1)^T$

$S_2 : (-2,0)^T \quad (-1,0)^T \quad (0,2)^T$

$(0,-2)^T \quad (1,0)^T \quad (2,0)^T$

a)



The points are not linearly separable.

b) $\underline{u} = \phi(\underline{x}) \triangleq \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1 x_2 \\ x_2^2 \end{bmatrix} \qquad \underline{w}^{(k)} \triangleq \begin{bmatrix} w_{00}^{(k)} \\ w_{01}^{(k)} \\ w_{02}^{(k)} \\ w_{11}^{(k)} \\ w_{12}^{(k)} \\ w_{22}^{(k)} \end{bmatrix}$

So, in expanded feature space,

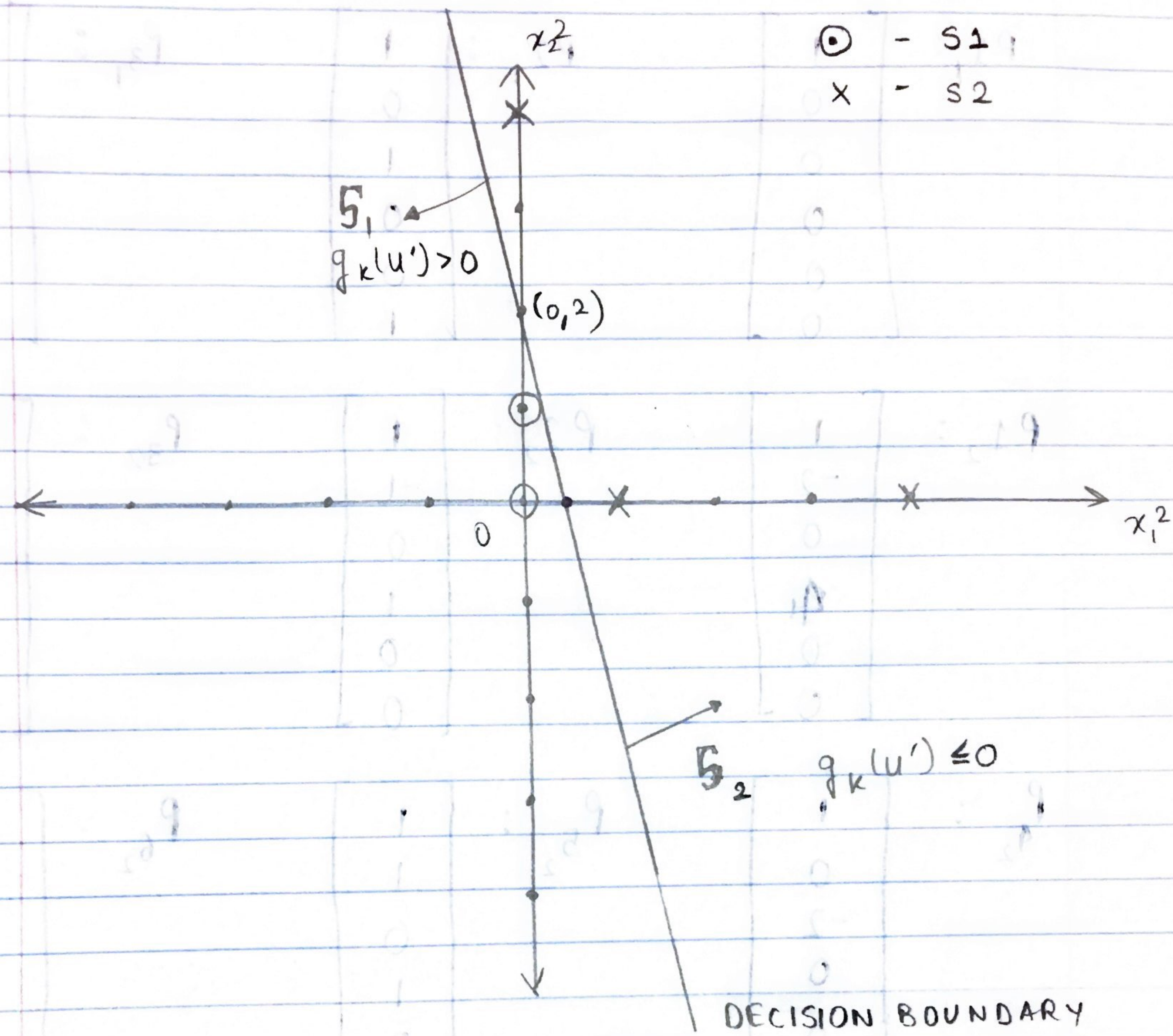
$P1_1 : \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$ $P2_1 : \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$ $P3_1 : \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$

$P1_2 : \begin{bmatrix} 1 \\ -2 \\ 0 \\ 4 \\ 0 \\ 0 \end{bmatrix}$ $P2_2 : \begin{bmatrix} 1 \\ -1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$ $P3_2 : \begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \\ 0 \\ 4 \end{bmatrix}$

$P4_2 : \begin{bmatrix} 1 \\ 0 \\ -2 \\ 0 \\ 0 \\ 4 \end{bmatrix}$ $P5_2 : \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$ $P6_2 : \begin{bmatrix} 1 \\ 2 \\ 0 \\ 4 \\ 0 \\ 0 \end{bmatrix}$

These are the points from both classes in expanded 6-feature space.

c) To plot decision boundary, we have to use (x_1^2, x_2^2) space. Then,



So the decision boundary is given by the line

$$4x_1^2 + x_2^2 = 2$$

and it passes through the points,

$$(0,2) \text{ and } (0.5, 0)$$

So, $\quad 4x_1^2 + x_2^2 - 2 = 0$

Hence,

corresponding weight vector,

$$\underset{\sim}{w}^{(k')} = \left[+2 \; 0 \; 0 \; 0 \; -4 \; 0 \; -1 \right]^T$$

$$= \left[2 \; 0 \; 0 \; -4 \; 0 \; -1 \right]^T$$

For this $\underline{w}^{(k')}$; S_1 : $g(\underline{u}^T \underline{w}^{(k')}) > 0$

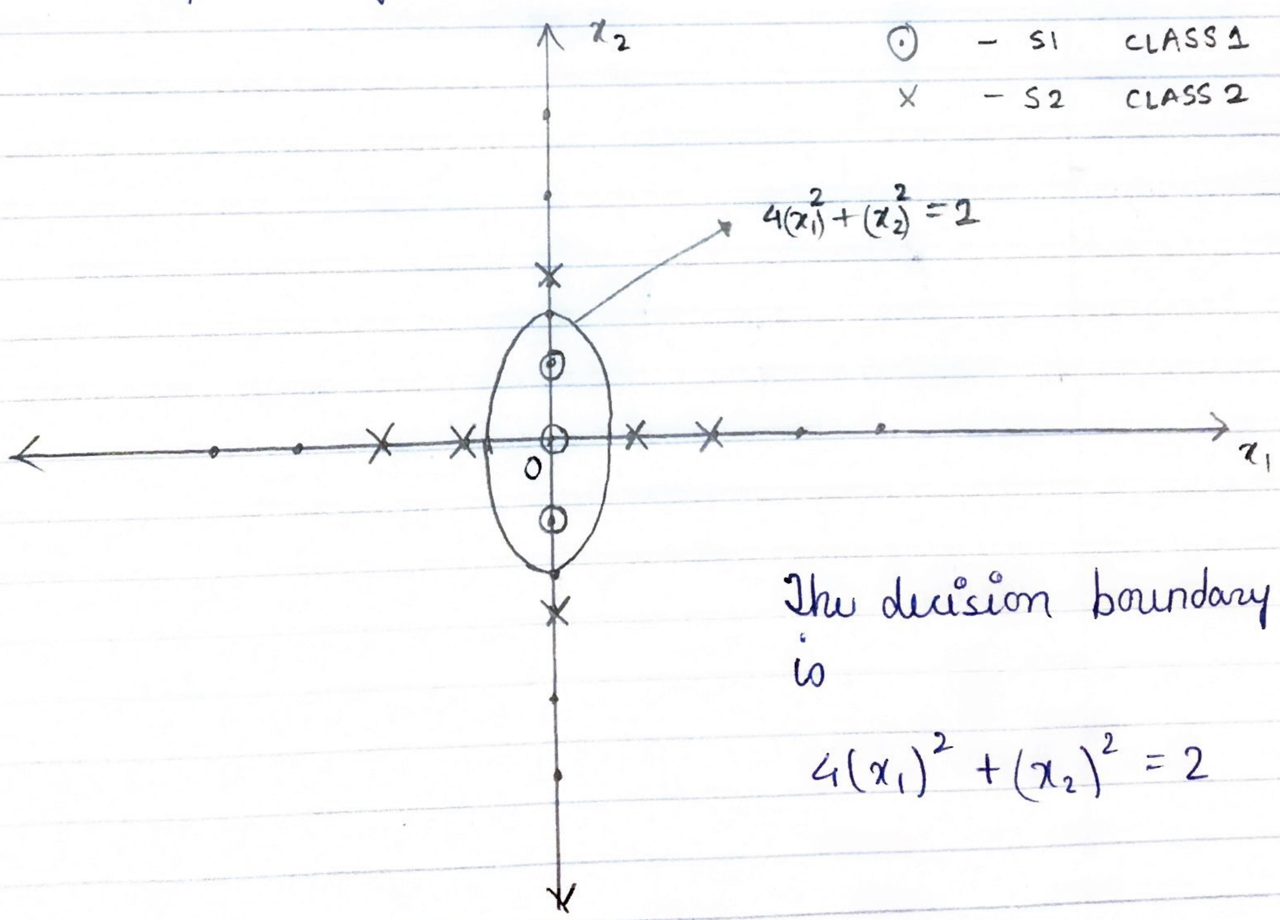
S_2 : $g(\underline{u}^T \underline{w}^{(k')}) \leq 0$

Hence this weight vector correctly classifies.

d) Now, $4x_1^2 + x_2^2 = 2$

$\therefore \; 2x_1^2 + \frac{x_2^2}{2} = 1$

This is equation of ellipse.



The decision boundary is

$$4(x_1)^2 + (x_2)^2 = 2$$

It separates the two classes S_1 and S_2.