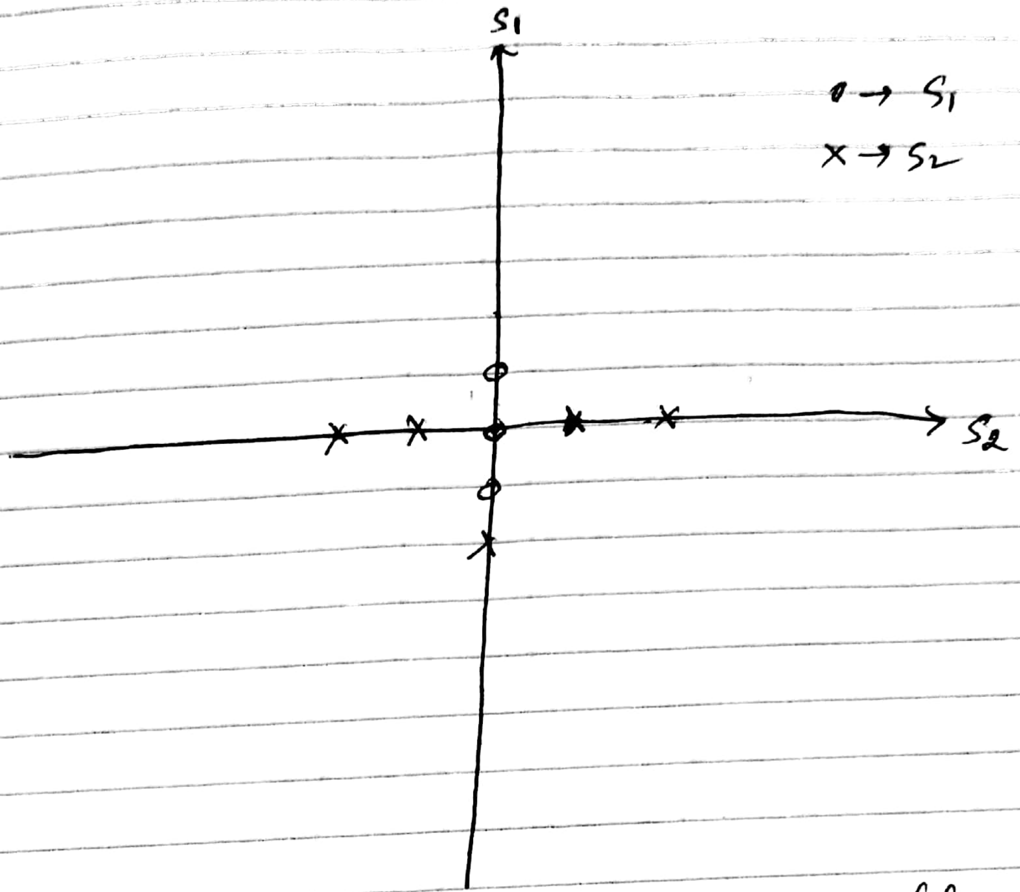
Problem-2

a)



No, they are not linearly separable.

$$\underline{u} = \phi(\underline{x}) = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_1 x_2 \\ x_1^2 \\ x_2^2 \end{bmatrix} \quad [(D'+1)\times 1] \quad (6 \times 1)$$

Total no. of terms $= \dfrac{D^2+3D}{2} + 1 = \dfrac{2^2 + 2\times 3}{2} + 1$

$= \dfrac{4+6}{2} + 1 = 6$

b) Applying the $\phi(x)$ function,

<u>Datapoint</u> (0, 0)

$$\underline{u}_1 = \phi(\underline{x}_1) = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

<u>Datapoint</u> (0, 1)

$$u_2 = \phi(\underline{x}_2) = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

<u>Datapoint</u> (0, -1)

$$u_3 = \phi(\underline{x}_3) = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

<u>Datapoint</u> (-2, 0)

$$u_4 = \phi(\underline{x}_4) = \begin{bmatrix} 1 \\ -2 \\ 0 \\ 4 \\ 0 \\ 0 \end{bmatrix}$$

Datapoint
(-1, 0)

$$u_5 = \phi(\underline{x}_5) = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

Datapoint
(0, 2)

$$u_6 = \phi(\underline{x}_6) = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \\ 0 \\ 4 \end{bmatrix}$$

Datapoint
(0, -2)

$$u_7 = \phi(\underline{x}_7) = \begin{bmatrix} 1 \\ 0 \\ -2 \\ 0 \\ 0 \\ 4 \end{bmatrix}$$

Datapoint
(1, 0)

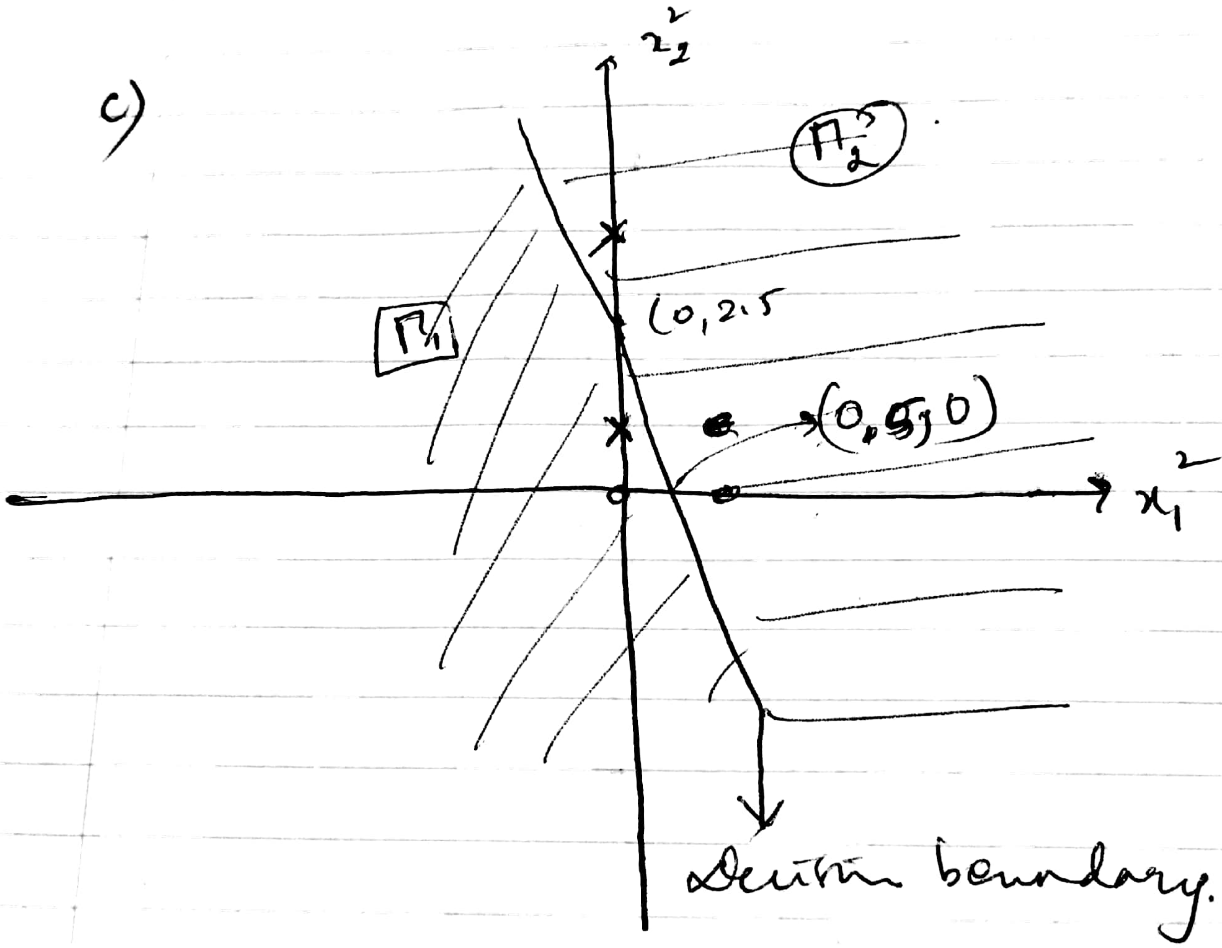
$$u_8 = \phi(\underline{x}_8) = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

Datapoint
(2, 0)

$$u_9 = \phi(\underline{x}_9) = \begin{bmatrix} 1 \\ 2 \\ 0 \\ 4 \\ 0 \\ 0 \end{bmatrix}$$

$$\underline{u} = \begin{bmatrix} [1 & 0 & 0 & 0 & 0 & 0] \\ [1 & 0 & 1 & 0 & 0 & 1] \\ [1 & 0 & -1 & 0 & 0 & 1] \\ [1 & -2 & 0 & 4 & 0 & 0] \\ [1 & -1 & 0 & 1 & 0 & 0] \\ [1 & 0 & 2 & 0 & 0 & 4] \\ [1 & 0 & -2 & 0 & 0 & 4] \\ [1 & 1 & 0 & 1 & 0 & 0] \\ [1 & 2 & 0 & 4 & 0 & 0] \end{bmatrix}$$

c)



Equation →

$$\frac{x_1^2}{0.5} + \frac{x_2^2}{2.5} = 1$$

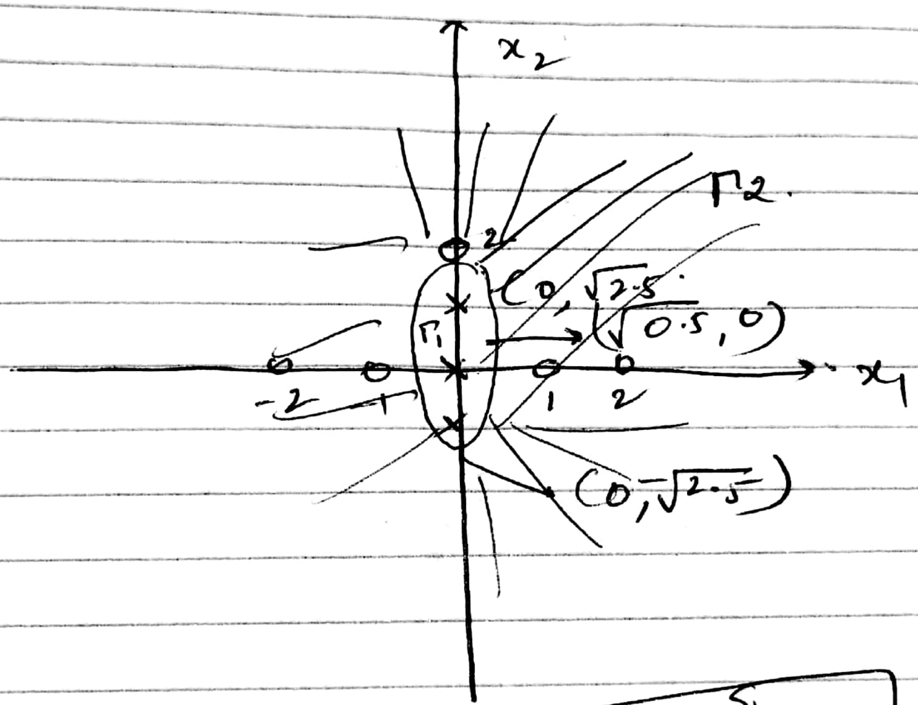
$$2x_1^2 + 0.4x_2^2 - 1 = 0$$

$$\underline{w} = [-1\ 0\ 0\ 2\ 0\ 0.4]$$

$$g(\underline{u}_i) = \underline{w}'^T u_i \underset{s_1}{\overset{s_2}{\gtrless}} 0$$

$$\boxed{g(\underline{u}_i) = 2x_1^2 + 0.4x_2^2 - 1 \underset{s_2}{\overset{s_1}{\lessgtr}} 0}$$

d)



$$\boxed{g(\underline{x}) = 2x_1^2 + 0.4x_2^2 - 1 \underset{s_2}{\overset{s_1}{\lessgtr}} 0.}$$

as mention above

1)
c)(i) According to the docs the initial weight vector (by default) is $\underline{w} = \underline{1}$ for all classes.

1) a) Python [illegible], Scikit-learn.

1) b)
→ The testing is not available during model building,
∴ To get a good estimate of the quality of the model, we use normalization parameters on training data

→ Avoiding small model weights in order to get numerical stability.

→ To achieve quick convergence of optimization algorithms.

(ii)