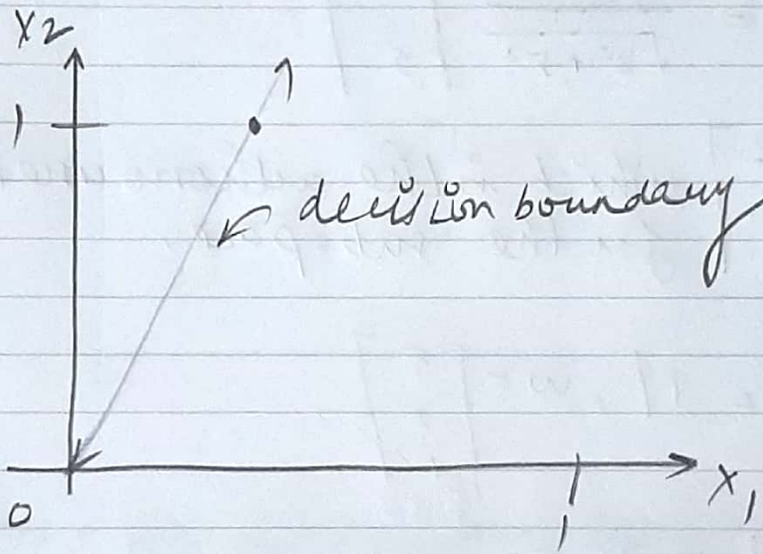Ayan Deep Hazra, ECE 532, Activity 8
(.ipynb pdf at end)

1. a) $x^T = [x_1, x_2]$ & $w = \begin{bmatrix} 5 \\ -2 \end{bmatrix}$

i) $x^T w = 0 \implies 5x_1 - 2x_2 = 0$

decision boundary

ii) * Yes, the decision boundary represents a subspace in $\mathbb{R}^2$.

* Since, the decision boundary is a line that passes through the origin, we know that it is a subspace in $\mathbb{R}^2$.

If we take $x_1 = 2, x_2 = 5$, we find a solution to the system.

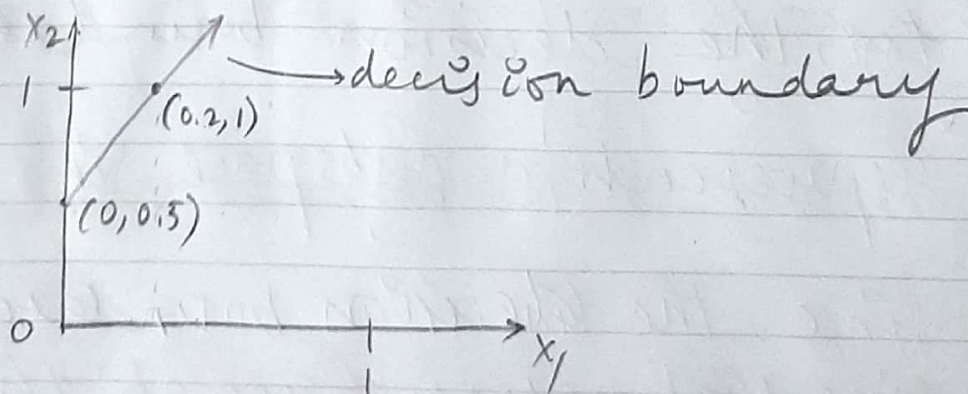Let $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$

By Gram-Schmidt orthogonalization, we have

$$\vec{u_1} = \frac{\vec{x}}{\|\vec{x}\|_2} = \frac{1}{\sqrt{2^2+5^2}}\begin{bmatrix} 2 \\ 5 \end{bmatrix}$$

$\vec{u_1} = \frac{1}{\sqrt{29}}\begin{bmatrix} 2 \\ 5 \end{bmatrix}$ which is the orthonormal basis for the subspace.

b) $x^T = \begin{bmatrix} x_1 & x_2 & 1 \end{bmatrix}$, $w = \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}$

i) $x^T w = 0 \implies 5x_1 - 2x_2 + 1 = 0$

ii) The decision boundary does not represent a subspace in R^2 as it does not pass through the origin.

c) $X^T = \begin{bmatrix} x_1^2 & x_2 & 1 \end{bmatrix}$, $w = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$

i) $X^T w = 0 \Rightarrow$

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

(1,1)

→ decision boundary

(0, 0.5)
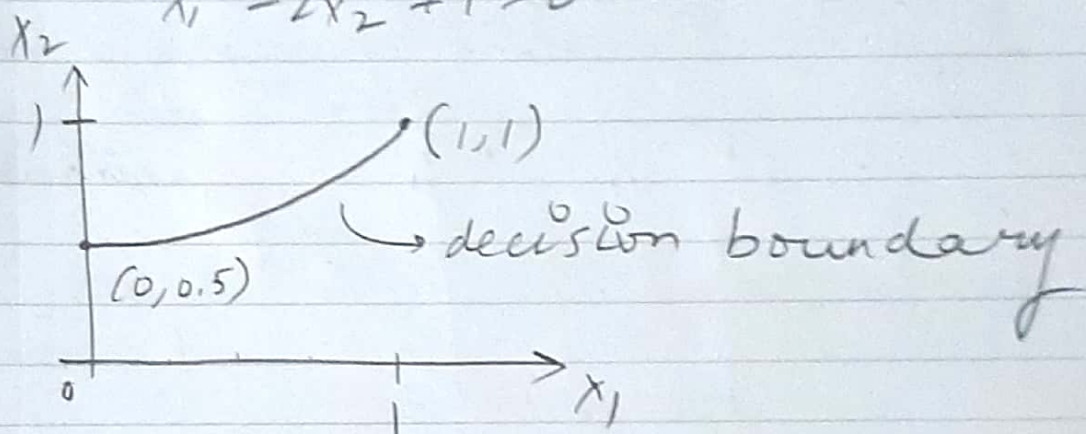
ii) The decision boundary does not represent a subspace in $\mathbb{R}^2$.

This is because the boundary is not a straight line, instead it is a parabola!

2.a) The actual evaluation data seems to have a more non linear curved decision boundary.

The classifier however fits a linear decision boundary.

$$\% \text{ Error} = \frac{\text{Number of errors}}{\text{Total datapoints}}$$

$$= \frac{1102}{10000}$$

$$= 11.02\%$$

b) The curved decision boundary fits much better than the linear classifier. It captures the non-linear characteristic of the dataset much better. The error percent is now,

$$= \frac{\text{Number of errors}}{\text{Total datapoints}}$$

$$= \frac{542}{10000} = 5.42\%$$

c) Adding 1000 datapoints at $(x_1, x_2) = (0, 3)$ skews the dataset and makes the classifier predict wrong and set the decision boundary much more steep with respect to previous.

a collection of 1000 datapoints misclassified datapoints
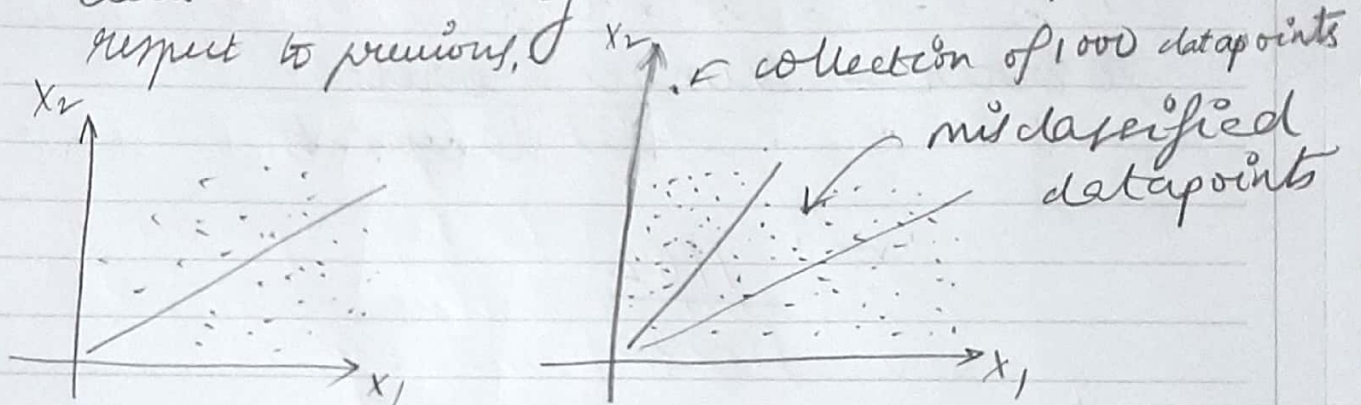
The error rate at $x_1 = 0, x_2 = 3$ is 21.34%. & the error rate at $v_1 = 0, x_2 = 10$ is 32.77%

This happens because the farther these select datapoints are, from the actual detaset, the more they affect the remaining data and thus skewing the decision boundary.

In this case, the decision boundary becomes more steep & thus classifying more red, datapoints as blue.

3. a to e) on the pdf

3. f) We know that the classifier defined in (e) is the worst as it has the most number of errors.

This is because it overfits the data and this is bad for unseen/new data.

It overfits the data by capturing noise/disturbances in the solution for the decision boundary.

4. If we have

$$x^T = [x_1 \quad x_2 \quad x_3] \quad \& \quad \text{some} \quad w = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$$

we see that $x^T w = 0$,

$$[x_1 \quad x_2 \quad x_3] \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = 0 \quad , \quad \text{gives us}$$

$$x_1 w_1 + x_2 w_2 + x_3 w_3 = 0$$

the equation of a 2-D plane in the (x_1, x_2, x_3) coordinate system.

5. If we define $w = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix}$

Given $v^T = [x_1 \ x_2 \ x_3 \ 1]$

Thus the decision boundary $x^T w = 0$ is

$[x_1 \ x_2 \ x_3 \ 1] \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix} = 0 \implies x_1 w_1 + x_2 w_3 + x_3 w_3 + w_4 = 0$

Since the plane is parallel to $x_1 - x_2$ plane we know it does not depend on x_1 & x_2 coordinates in x^T.

Thus, $w_1 = w_2 = 0$

thus, $x_1 w_1 + x_2 w_2 + x_3 w_3 + w_4 = 0,$

$\implies$ (1) $w_3 = -w_4$ let $w_3 = -w_4 = k$

Thus, we have

$w = \begin{bmatrix} 0 \\ 0 \\ k \\ -k \end{bmatrix}$ for some $k \in \mathbb{R}$