HW3 Math Section

Question 1

(a) Sketching the Optimal Separating Hyperplane:

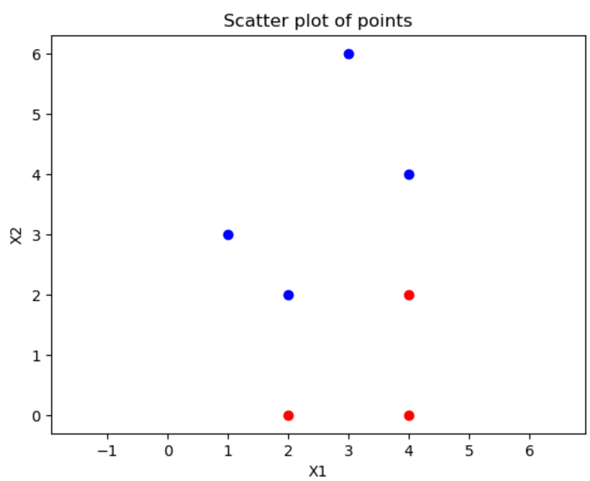
Given the observations:

Index	X_1	X_2	Y
1	3	6	Blue
2	2	2	Blue
3	4	4	Blue
4	1	3	Blue
5	2	0	Red
6	4	2	Red
7	4	0	Red

We want to find the optimal separating hyperplane that maximizes the margin between the two classes. This hyperplane can be represented by the equation:

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 = 0$$

where $\beta_0, \beta_1, \beta_2$ are coefficients to be determined.



$$SV1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$SV2 = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$$

$$SV3 = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

Bias = 1
Now,

$$SV1 = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}$$

$$SV2 = \begin{bmatrix} 4 \\ 4 \\ 1 \end{bmatrix}$$

$$SV3 = \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix}$$

Blue Class is +ve and Red class is -ve

$$\alpha_1 SV_1 SV_1 + \alpha_2 SV_2 SV_1 + \alpha_3 SV_3 SV_1 = -1$$
$$\alpha_1 SV_1 SV_2 + \alpha_2 SV_2 SV_2 + \alpha_3 SV_3 SV_2 = -1$$
$$\alpha_1 SV_1 SV_3 + \alpha_2 SV_2 SV_3 + \alpha_3 SV_3 SV_3 = 1$$

Solving the equations, we get,

$$\alpha_1 = -\frac{5}{2}$$

$$\alpha_2 = \frac{1}{2}$$

$$\alpha_3 = 1$$

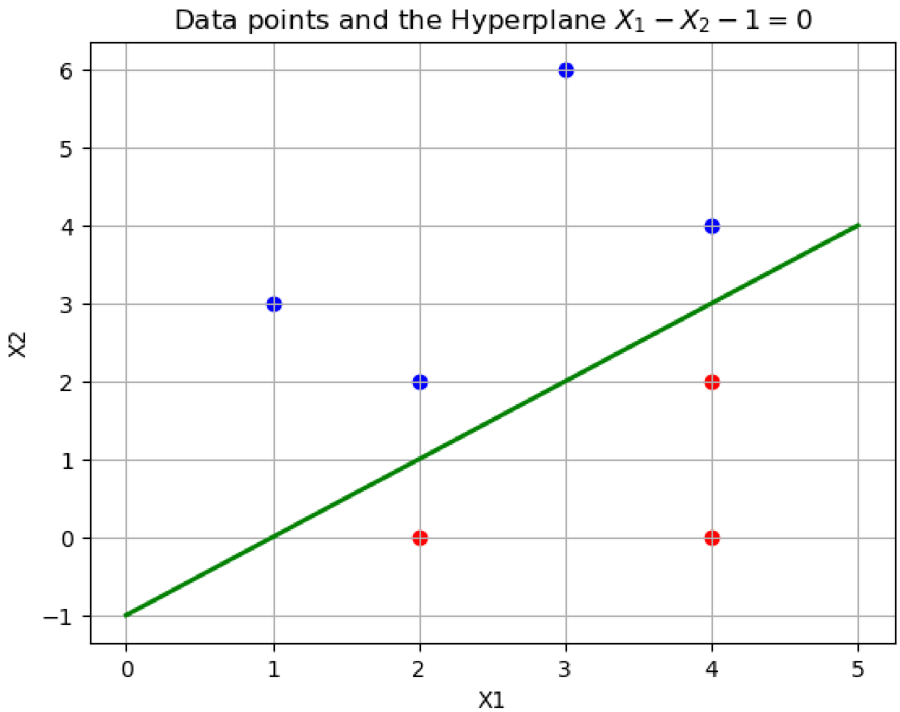
$$\bar{w} = \sum \alpha_i SV_i = \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}$$

So, in the equation $y = wx + b$,

$$w = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

and $b = -1$ (the bias)

The equation is, $Y = X_1 - X_2 - 1 = 0$
So, our Optimal Hyperplane is $X_1 - X_2 - 1 = 0$.



(b) Classification Rule for the Maximal Margin Classifier:

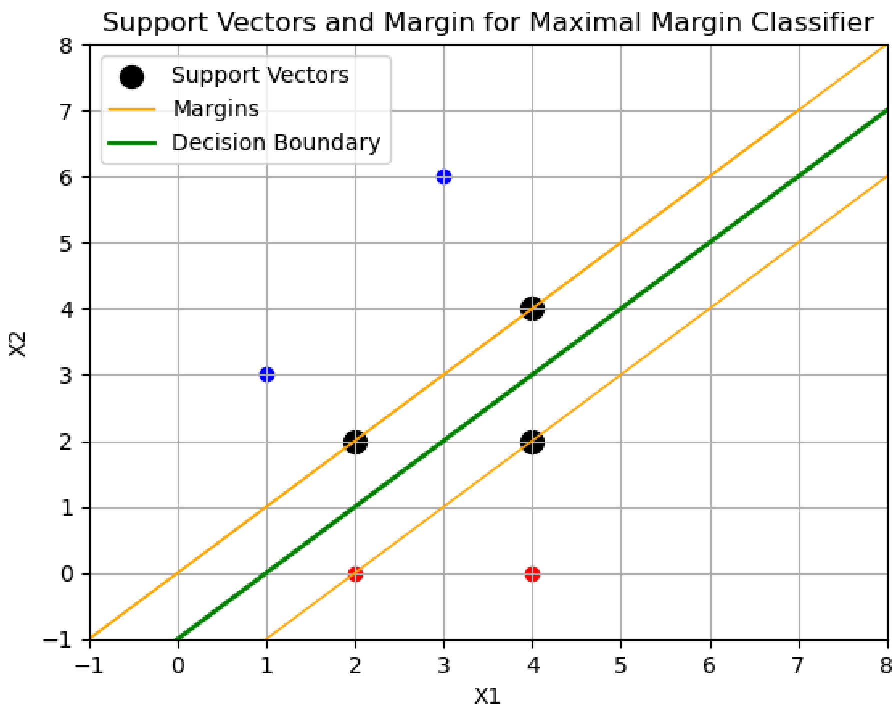
The classification rule for the maximal margin classifier is as follows:

- Classify to Red if $\beta_0 + \beta_1 X_1 + \beta_2 X_2 \geq 0$
- Classify to Blue otherwise

From the equation above, $\beta_0 = -1$, $\beta_1 = 1$, and $\beta_2 = -1$.
So, the rule is:

- Classify to Red if $X_1 - X_2 - 1 \geq 0$
- Classify to Blue otherwise

(c) & (d) Margin for the Maximal Margin Hyperplane and its Support Vectors:



(e) Does a slight movement of the seventh observation affect the maximal margin hyperplane?

No, the 7th data point is neither a support vector nor on the margin, so a slight movement of that data point will not affect the decision boundary of a maximal margin classifier because,

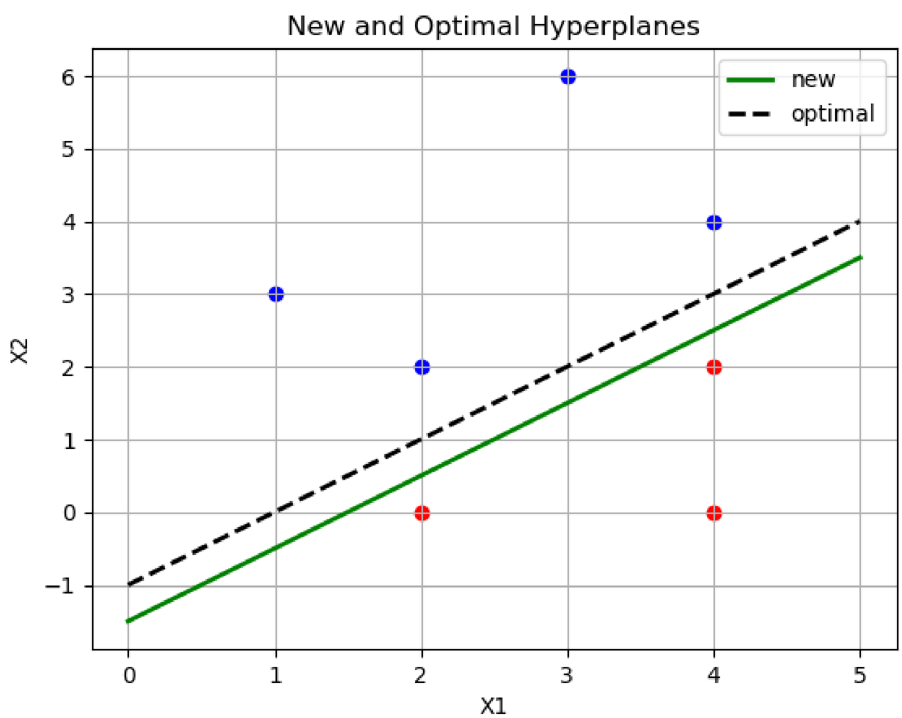
- **Support Vectors Define the Margin:** The decision boundary is primarily determined by the support vectors. These are the data points closest to the hyperplane from each class and define the maximum margin.
- **Non-Support Vectors Have No Influence:** Points further away from the hyperplane (not support vectors) do not directly influence the position of the decision boundary. They only affect the classifier if they are misclassified (fall on the wrong side of the margin).
- **Margin Absorbs Slight Movements:** As long as the movement of a non-support vector stays within the same side of the current margin, it

won't affect the support vectors or the optimal hyperplane. The existing margin can absorb slight variations in these points.

So, a slight movement of such a data point wouldn't necessitate a change in the decision boundary. However, for significant movements that push the point across the margin or cause a misclassification, the classifier might need to be re-evaluated.

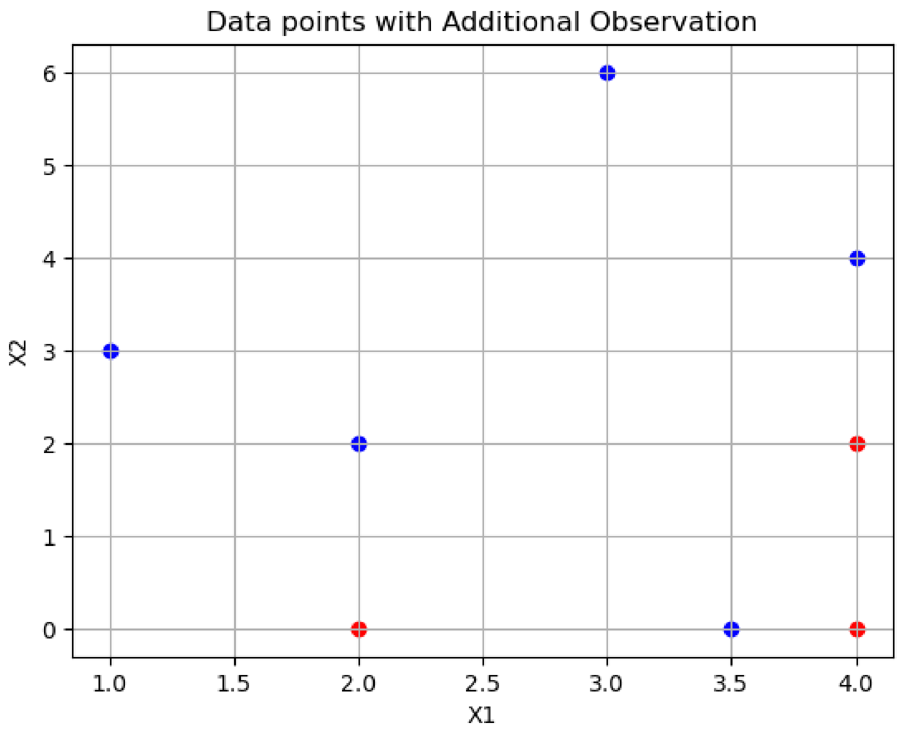
(f) Draw an alternative hyperplane that is not the optimal one, and provide the equation for this hyperplane.

To choose an alternative hyperplane that still classifies the data correctly but is not the optimal separating hyperplane, we can choose a hyperplane parallel to the optimal hyperplane but shifted by a certain amount. Since a margin is $X_1 - X_2 - 2 = 0$ and the optimal hyperplane is $X_1 - X_2 - 1 = 0$, we can take a plane between them like $X_1 - X_2 - 1.5 = 0$.



(g) Draw an additional observation on the plot so that the two classes are no longer separable by a hyperplane

To make the two classes no longer separable by a hyperplane, we need to add an additional observation such that it is incorrectly classified. In this case, if we add a point belonging to the Blue class within the region dominated by the Red class, it would make the separation impossible by a hyperplane.



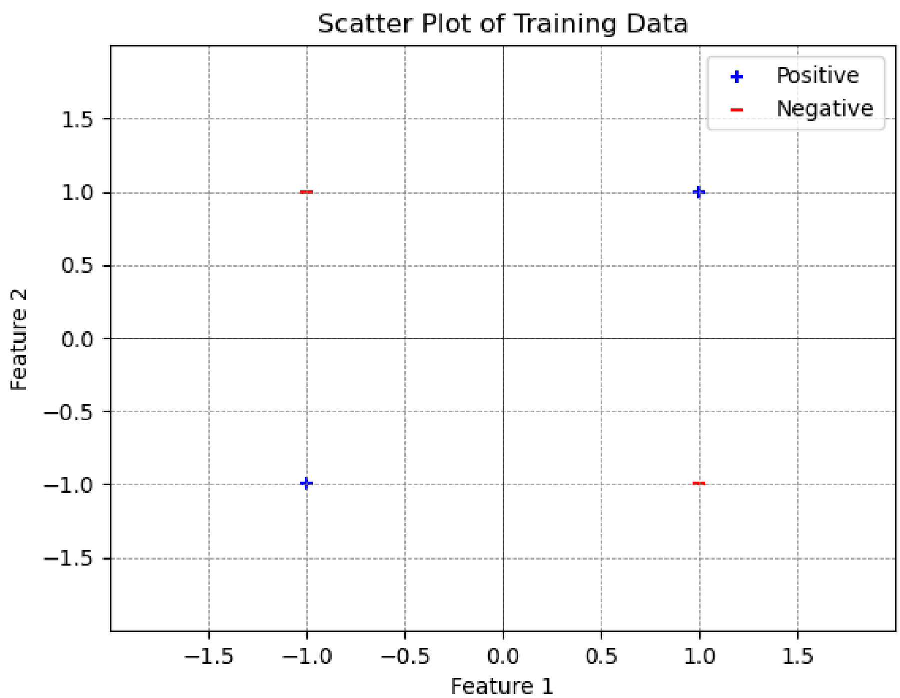
Question 2

(a) The table representing the training set:

Sample	Feature 1 (X1)	Feature 2 (X2)	Class (y)
1	1	1	Positive
2	-1	-1	Positive
3	1	-1	Negative
4	-1	1	Negative

This truth table represents the **XNOR** (exclusive NOR) logic gate.

(b) Let's plot these four points on the x-y plane:



(c) Feature transformation and Drawing the transformed points

These points are **not linearly separable** because there is no single straight line that can be drawn to separate the positive examples from the negative ones.

The feature transformation function is:

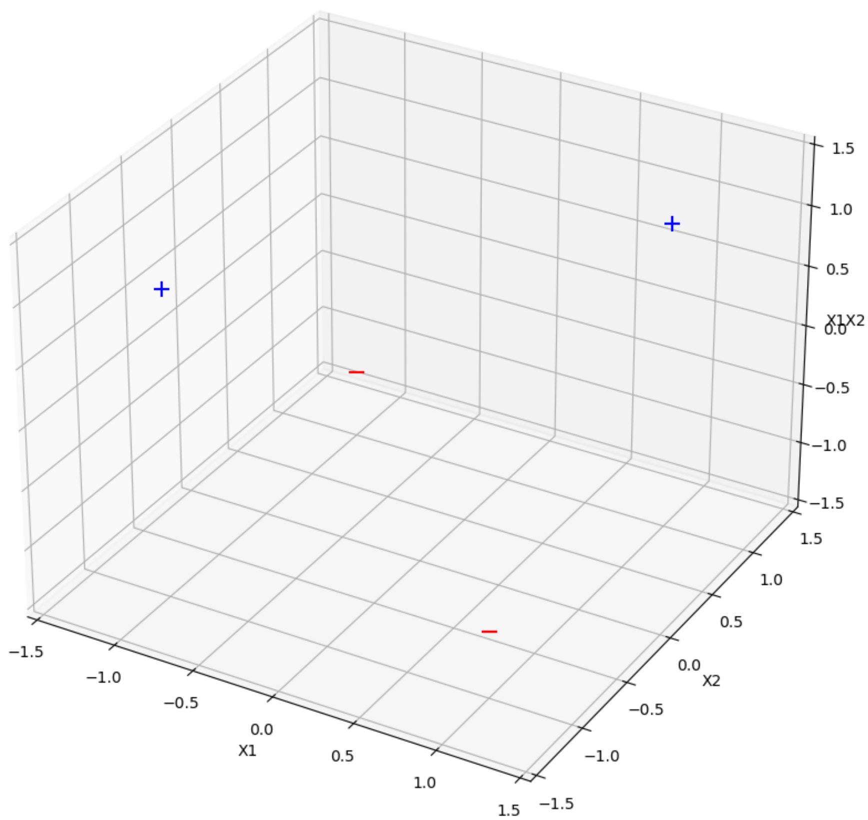
$$\phi(x) = [x_1, x_2, x_1 \cdot x_2]$$

For the given data points:

$$\begin{aligned}(1,1) &: \phi(1,1) = [1,1,1\cdot 1] = [1,1,1]\\(-1,-1) &: \phi(-1,-1) = [-1,-1,-1\cdot -1] = [-1,-1,1]\\(1,-1) &: \phi(1,-1) = [1,-1,1\cdot -1] = [1,-1,-1]\\(-1,1) &: \phi(-1,1) = [-1,1,-1\cdot 1] = [-1,1,-1]\end{aligned}$$

Now, we have the transformed points:

$$
\begin{aligned}
\textbf{(1,1)} &: [\textbf{1, 1, 1}] \\
\textbf{(-1,-1)} &: [\textbf{-1, -1, 1}] \\
\textbf{(1,-1)} &: [\textbf{1, -1, -1}] \\
\textbf{(-1,1)} &: [\textbf{-1, 1, -1}]
\end{aligned}
$$



Linearly Separable

Yes, they are linearly separable as they can now be separated by the plane $X_1X_2 = 0$

While this is one possible solution, any weight vector and bias term that achieves this separation would be valid.

(d) Margin Size and Support Vectors after transformation

In this case, **all four data points** we were given will be support vectors.

Positive Class: [1, 1, 1] and [-1, -1, 1]

Negative Class: [1, -1, -1] and [-1, 1, -1]

We can intuitively see why all points are equidistant to the hyperplane $X_1X_2 = 0$.

Margin size is calculated as the perpendicular distance between the hyperplane and the closest points from each group. So, the margin size here is 1.