Machine Learning Assignment 3

Ritika Nigam, UIN: 633002197

March 2024

1 Math-Question 1

a)

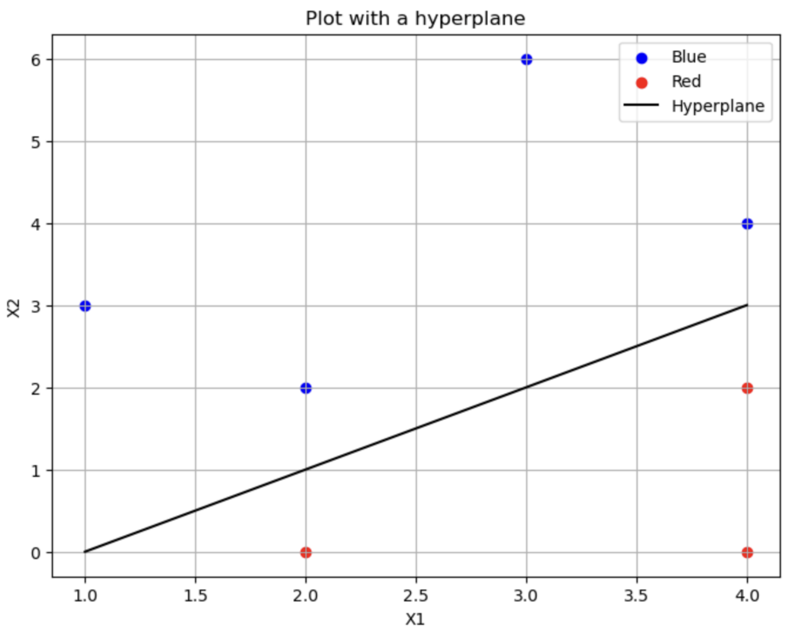

Figure 1: Plotted points and the Hyperplane

The plot above represents the observations and the hypothesized optimal separating hyperplane with the equation $X1 - X2 - 1 = 0$, as determined from our initial assessment.

b) For the classification rule based on the maximal margin classifier. The rule, derived from the equation of the hyperplane, is:
Classify to Blue if $X1 - X2 - 1 = 0$ and
Classify to Red otherwise
This implies that the coefficients are $\beta0$ = -1, $\beta1$ = 1 and $\beta2$ = -1

c)

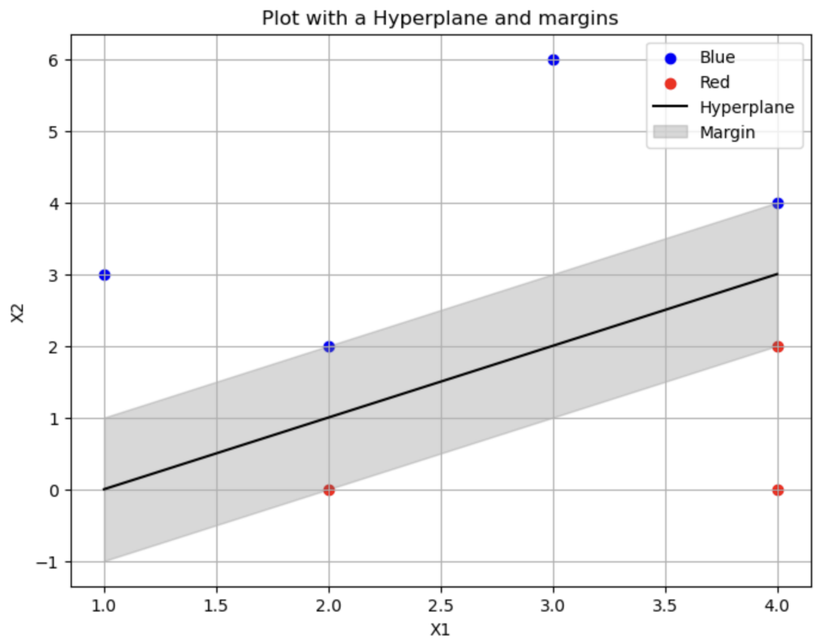

Figure 2: Hyperplane with margins

d)

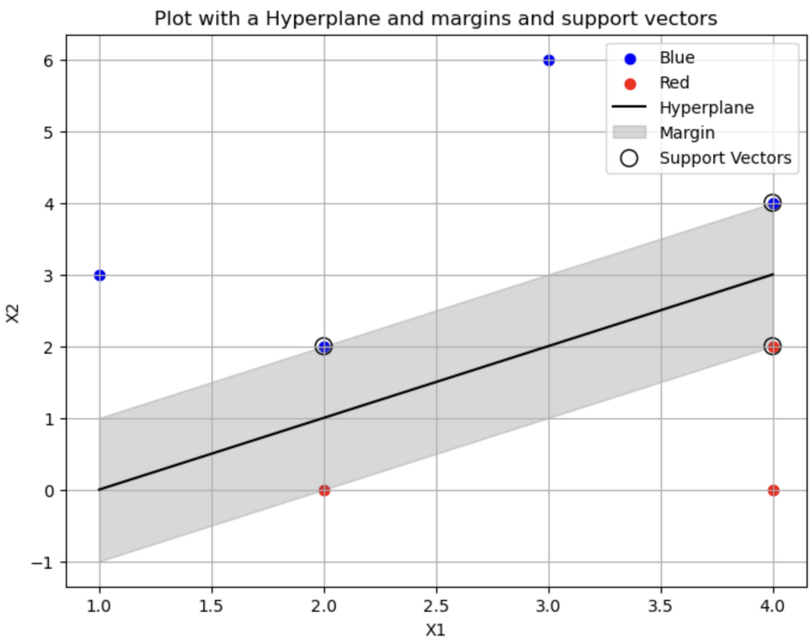

Figure 3: Hyperplane with margins and support vectors

e)Given that the seventh observation is not a support vector, its slight movement would not affect the position of the hyperplane or its margins.

f)

The plot now includes an alternative hyperplane (in green) represented by the equation $X1 - X2 + 0 = 0$. Although it still separates the classes, it is clear that this hyperplane does not offer the widest margin between the classes, highlighting its non-optimality compared to the dashed hyperplane.

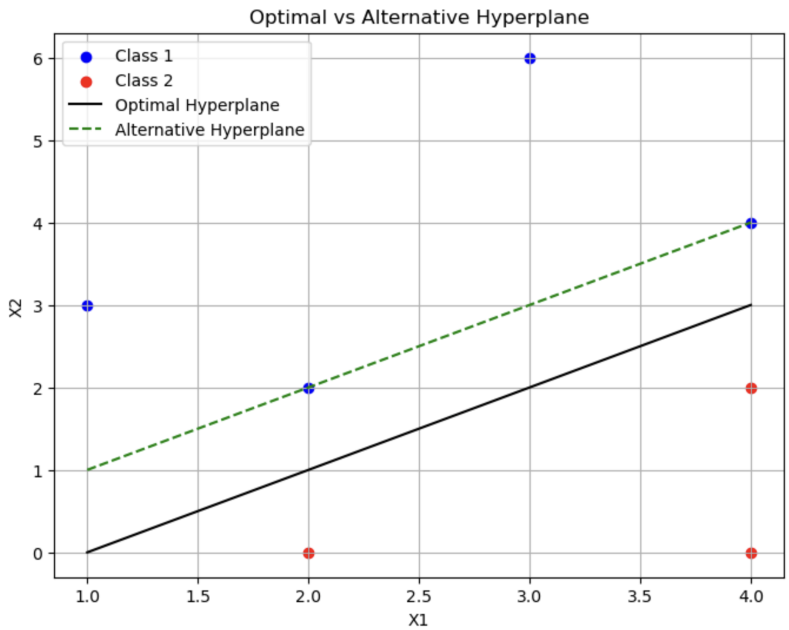

Figure 4: Alternative hyperplane

g) The final plot introduces a new observation, marked with a green 'X', positioned such that it disrupts the clear separability between the two classes by a single hyperplane. This scenario illustrates how the introduction of certain data points can make linear separation impossible without adjusting the classification model to accommodate more complex boundaries or adopting different strategies.

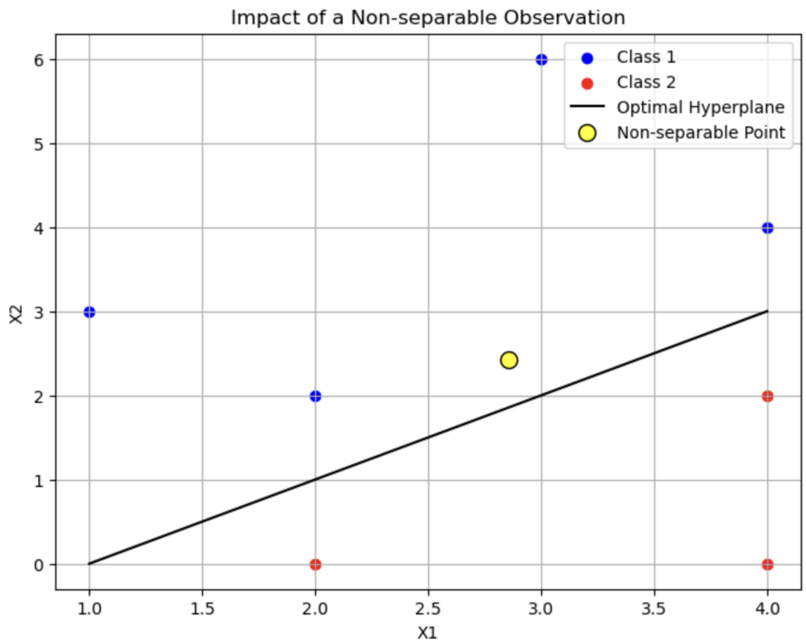

Figure 5: Additional Observation

2 Math-Question 2

a) Training Data Table

Index	X1	X2	Class
1	1	1	+
2	-1	-1	+
3	1	-1	-
4	-1	1	-

Figure 6: Training Data Table

The shape of X (the feature matrix) is 4x2 as there are four samples and each sample has two features. The shape of y (the class labels) is 4x1 since there's one label per sample.
a.bonus This table resembles a truth table for the XOR (Exclusive OR) logic gate. The XOR gate outputs true only when the inputs differ.

b)

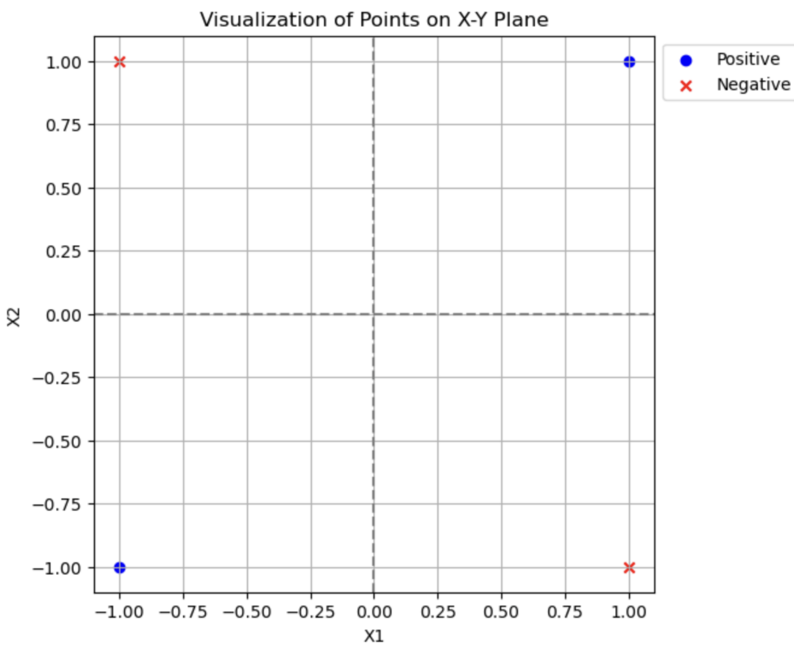

Figure 7: Plotting the points on x-y plane

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

c) Given the feature transformation (x) = [x1, x2, x1x2], we'll transform the coordinates of each of the four samples. Here are the original points for reference:

Positive examples: (1, 1), (-1, -1)
Negative examples: (1, -1), (-1, 1)
Applying the transformation (x) = [x1, x2, x1x2] to each point:
For (1, 1): (1, 1) = [1, 1, 1 * 1] = [1, 1, 1]
For (-1, -1): (-1, -1) = [-1, -1, -1 * -1] = [-1, -1, 1]
For (1, -1): (1, -1) = [1, -1, 1 * -1] = [1, -1, -1]
For (-1, 1): (-1, 1) = [-1, 1, -1 * 1] = [-1, 1, -1]

The plot below represents the four points after being transformed by the function (x) = [x1, x2, x1x2]. From the visualization, it's clear that the transformed points are linearly separable in 3D space. A plane can be drawn that separates the positive examples (in blue) from the negative examples (in red).

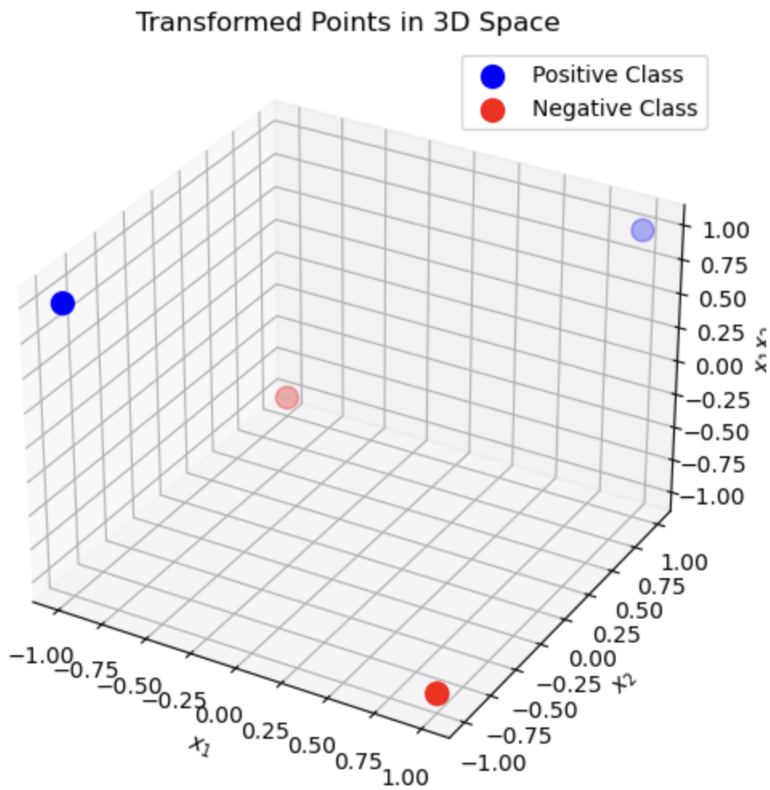

Figure 8: Points in 3-D space

d) Given the transformed points in the 3D space:

Positive examples: (1, 1, 1), (-1, -1, 1)

Negative examples: (1, -1, -1), (-1, 1, -1)

And considering a linear separator in this space, we can imagine, due to the symmetry of the points around the origin and given their transformation, that a plane such as x3 = 0 could act as a separating hyperplane, where x3 corresponds to the transformed feature x1x2. This is not the only possible hyperplane but is a simple one that clearly separates the positive and negative examples based on the sign of x3.

However, to calculate the margin, we need to consider the distance from the points to this hyperplane. The margin is defined as the distance from the closest points (support vectors) to the hyperplane. Given the simplicity of our hyperplane x3=0 and the transformed space, the support vectors are essentially all the points, as they all lie at the same distance from the hyperplane due to the transformation

The distance d from a point (x1,x2,x3) to the plane x3=0 is given by:

d = $\frac{|ax1+bx2+cx3+d|}{\sqrt{a^2+b^2+c^2}}$

In this formula: a, b, c are the coefficients of the plane equation $ax1 + bx2 + cx3 + d = 0$

For the plane x3 = 0, we have a=0, b=0, c=1 and d=0, putting these values in the equation above we get:

d = $\frac{|x3|}{\sqrt{1}}$

Since the transformed feature x3=x1x2 is either 1 or -1 for all points, the distance d from any point to the plane x3=0 is 1.

The margine size is equal to 1 and all four points are the support vectors.