CSCE 633 : Machine Learning - Assignment 3

Neha Joshi

29 March 2024

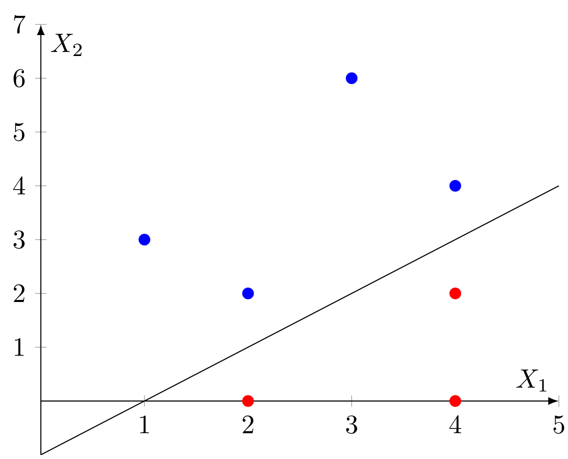
1 Problem 1

Given a dataset with $n = 7$ observations across $p = 2$ dimensions and associated class labels, we analyze the construction of a maximal margin classifier, which is crucial in support vector machine models. The dataset is given as:

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

(a) Sketch the optimal separating hyperplane, and provide the equation for this hyperplane.

Optimal hyperplane would be the one which separates the points of different class the best way. that is the distance between the points (support vectors especially) is largest from the line from boht the classes. As per the graph, support vectors as 2,2 and 4,2 and we need some line between them. Starting with a random line, we change slope and intercepts to find a line that separates the two classes. once the line is between the two classes, we make sure it is in between the support vectors. This way we get the best hyper plane.



(b) Describe the classification rule for the maximal margin classifier. It should be something along the lines of "classify to Red if $\beta_0 + \beta_1 X_1 + \beta_2 X_2 \geq 0$, and classify to Blue otherwise." Provide the values for β_0, β_1, and β_2.

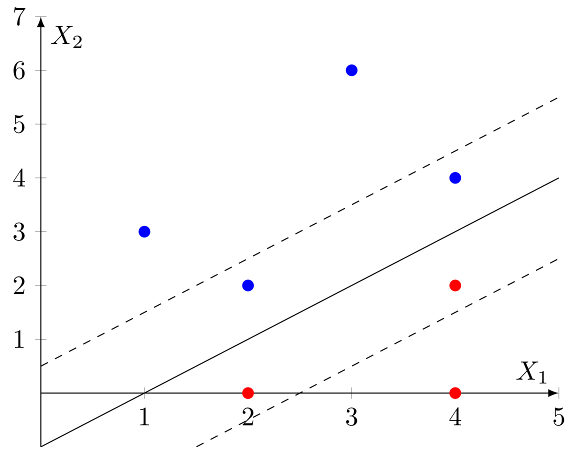
Red if: $-1 + X_1 - X_2 \leq 0$;Blue otherwise.

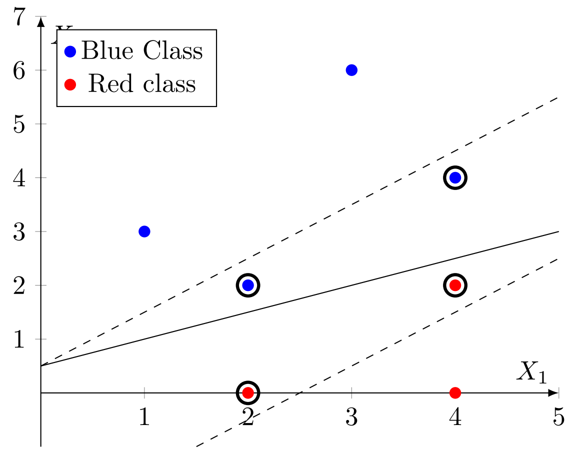
$$\beta_0 = -1$$
$$\beta_1 = 1$$
$$\beta_2 = -1$$

(c) On your sketch, indicate the margin for the maximal margin hyperplane

The margin is a set of parallel lines across the hyperplane that basically consists of all support vectors. Given the slope of the hyperplane is 1, the margin lines will also have a slope of 1 to remain parallel.

(d) Indicate the support vectors for the maximal margin classifier

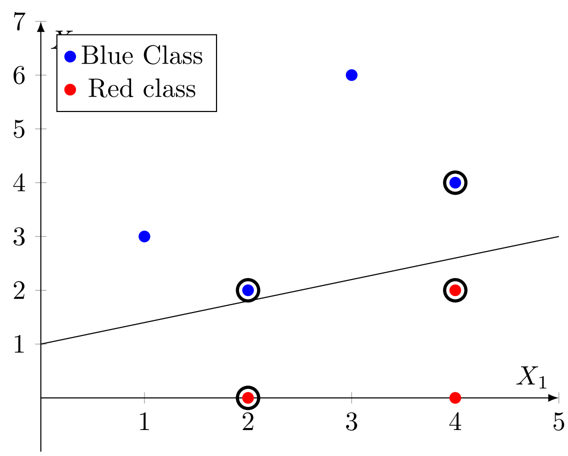




(e) Does a slight movement of the seventh observation affect the maximal margin hyperplane?Justify your answer (4,0) will not show any apparent significant difference until enters the margin to become the support vector. Since (4,0) is not a support vector it wont be significant. Since in SVM only support vectors matter for calculations so we see no differnece.

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

Choosing a line which is close to one part than the other and not the middle area of the classification

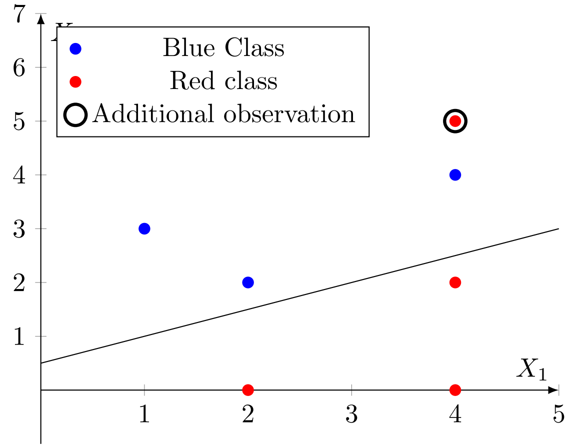


(g) Draw an additional observation on the plot so that the two classes are no longer separable by a hyperplane

Choosing a point such that red classification lies in the blue side of the graph thus defying the hyperplane categorisation i.e (4,5)

2 Problem 2

Assume that we have training data with four samples, 2 features, and 2 classes. The positive examples are $(1, 1)$ and $(-1, -1)$. The negative examples are $(1, -1)$ and $(-1, 1)$.



(a) Draw a table that represents this training set. What is the shape of X and y?

Answer: The training data table can be represented as follows:

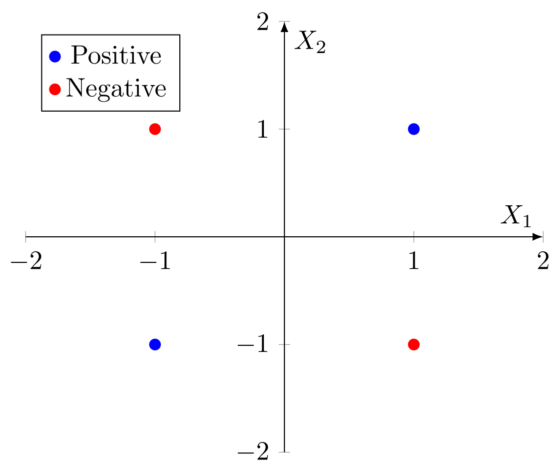
x_1	x_2	Class
1	1	Positive
-1	-1	Positive
1	-1	Negative
-1	1	Negative

The shape of X is 4×2 since there are four samples and each sample has two features. The shape of y is 4×1 since there is one class label for each of the four samples.

(a.bonus) The table you just drew is called a truth table. Do you know the logic gate representing this truth table?

Answer: The truth table represents the XNOR (exclusive NOR) logic gate. This is because the output (Class) is Positive only when the inputs (x_1 and x_2) are different.

(b) Draw these four points on x-y plane. Are these points linearly separable



The points are not linearly separable as no straight line can be drawn on the co-ordinate system to separate the points.

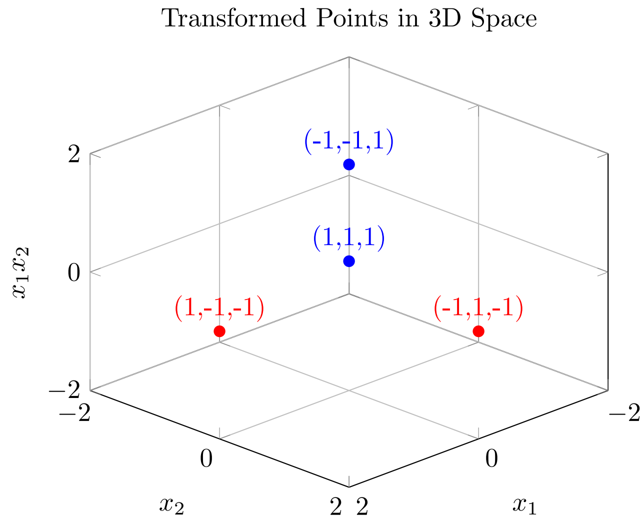
(c) Consider the feature transformation $(x) = [x1, x2, x1x2]$, where x1 and x2 are, respectively, the first and second coordinates of a generic example x. Draw these four points when transformed by the function $\phi(x)$. Are these four transformed points now linearly separable?

After applying the feature transformation $\phi(x) = [x_1, x_2, x_1x_2]$, the coordinates of the points in the transformed space are:

- Positive class become: $(1, 1, 1)$ and $(-1, -1, 1)$
- Negative class become: $(1, -1, -1)$ and $(-1, 1, -1)$

We consider a hyperplane that could separate them in the 3-D space defined by x_1, x_2, and x_1x_2. A simple consideration for a separating hyperplane in this context is one based on the third feature, z, given our transformation.

Given that:



- Third feature of all positive examples (x_1x_2) is positive
- Third feature of all negative examples (x_1x_2) is negative

it is evident that the points are linearly separable in this transformed feature space.

(d) What is the margin size after the transformation? Which points are support vectors ?

- Positive examples: $(1, 1, 1)$ and $(-1, -1, 1)$
- Negative examples: $(1, -1, -1)$ and $(-1, 1, -1)$

In the transformed space, where these points demonstrate linear separability, our aim is to locate a hyperplane capable of separating them. This hyperplane can be generally described as: $w_1x_1 + w_2x_2 + w_3x_1x_2 + b = 0$, where w_1, w_2, and w_3 are the weights and b is the bias. Our hyperplane $x_1x_2 = 0$, thus $w_1 = 0$, $w_2 = 0$, $w_3 = 1$, and $b = 0$.

The margin is the distance from the hyperplane to the nearest data point. For a hyperplane defined by $w^T\phi(x) + b = 0$, the distance of a point x from the hyperplane is given by:

$$\text{distance} = \frac{|w^T\phi(x) + b|}{\|w\|}$$

In our case, with $w = [0, 0, 1]$ and $b = 0$, the norm of w is $\|w\| = \sqrt{0^2 + 0^2 + 1^2} = 1$. Thus, the distance of any point $\phi(x)$ from the hyperplane is:

$$\text{distance} = \frac{|1 \cdot x_1x_2|}{1} = |x_1x_2|$$

$|x_1x_2| = 1$ for all points, the margin (the shortest distance from the points to the hyperplane) is 1*2 = 2. This calculation implies that all four points are equidistant from the hyperplane and thus serve as support vectors. The margin size would be 2.