

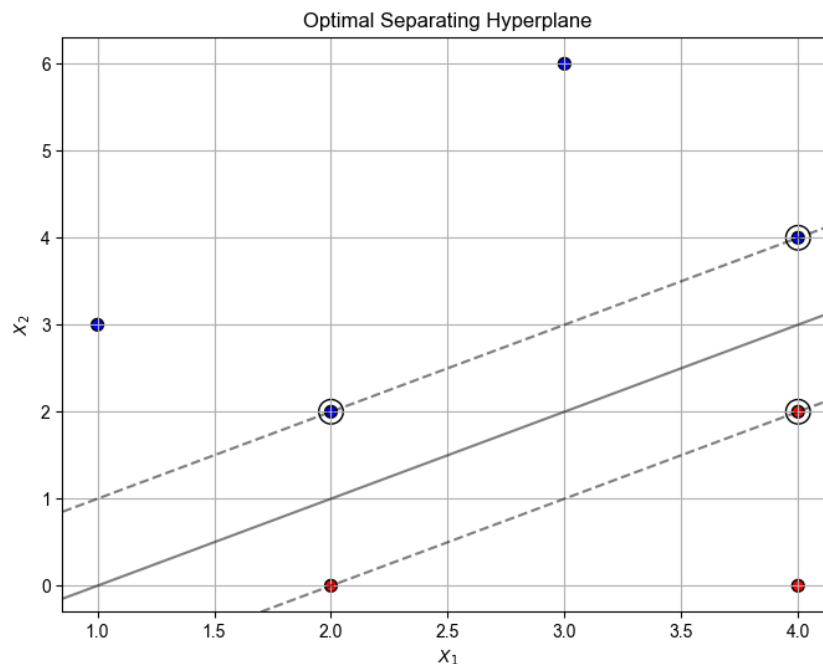
Homework Submission

Math: Question 1 - 30 points

We are given $n = 7$ observations in $p = 2$ dimensions. For each observation, there is an associated class label.

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.



The equation of the optimal separating hyperplane is: $X_1 - X_2 - 1 = 0$

- (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 .

The classification rule for the maximal margin classifier is to classify an observation as Red if $\beta_0 + \beta_1 X_1 + \beta_2 X_2 \geq 0$ and classify an observation as Blue if $\beta_0 + \beta_1 X_1 + \beta_2 X_2 < 0$. The values of the coefficients in the hyperplane equation are $\beta_0 = -1, \beta_1 = 1$, and $\beta_2 = -1$.

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

It has been shown in Figure 1.

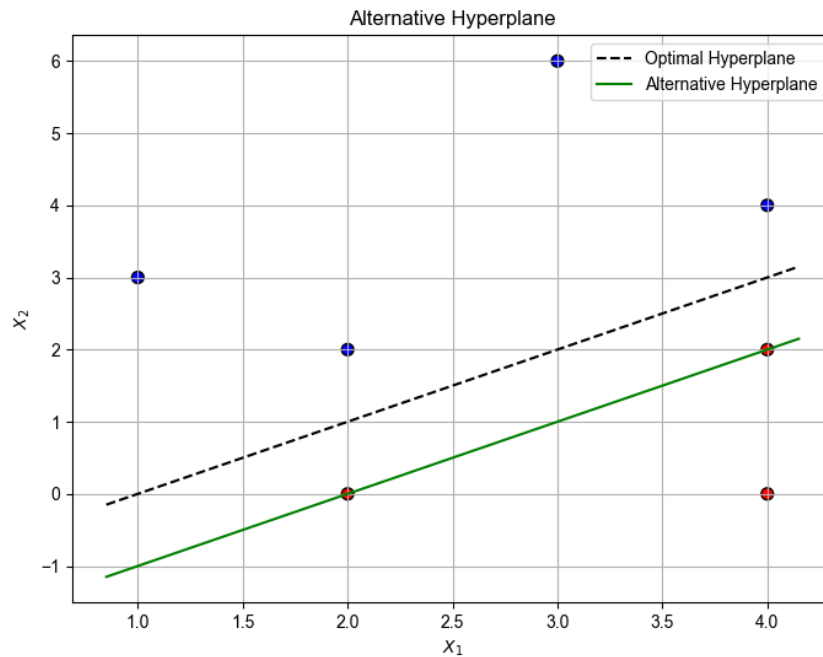
- (d) Indicate the support vectors for the maximal margin classifier.

The support vectors for the maximal margin classifier are the observations with coordinates (2, 2), (4, 4), and (4, 2).

- (e) Does a slight movement of the seventh observation affect the maximal margin hyperplane? Justify your answer.

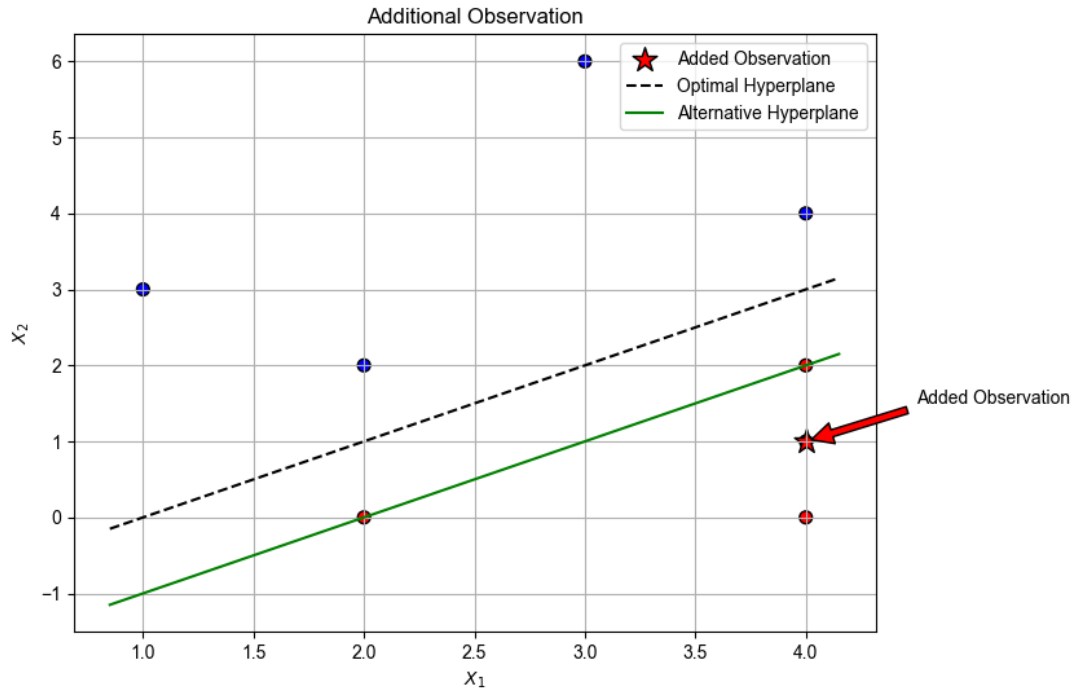
It does not affect the maximal margin hyperplane because it is not a support vector, and its movement does not change the position of the existing support vectors.

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



An alternative hyperplane that separates the classes but is not optimal is given by the equation: $-X_1 + X_2 + 2 = 0$

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



A new observation with coordinates (4,1) and assigning it to the Blue class, the two classes become no longer separable by a hyperplane. This new point lies within the cluster of Red points, making it impossible for a single hyperplane to perfectly separate the Blue and Red classes without misclassification.

Math: Question 2 - 20 points

Assume that we have a 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).

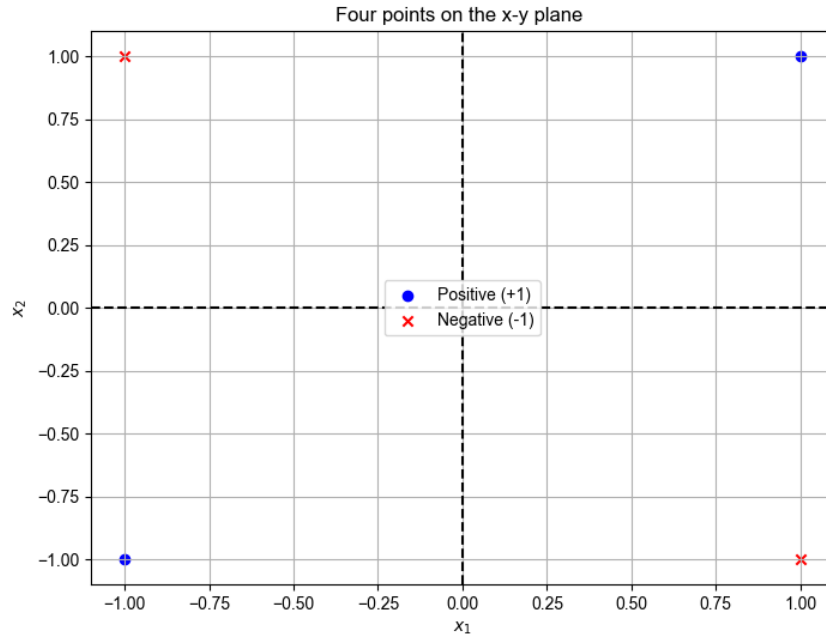
- (a) Draw a table that represents this training set. What is the shape of X and y ?
- (a.bonus) The table you just drew is called a truth table. Do you know the logic gate representing this truth table?

x_1	x_2	y
1	1	1
-1	-1	1
1	-1	-1
-1	1	-1

The shape of X is (4×2) , as there are 4 samples and each sample has 2 features. The shape of y is (4×1) , as there are 4 samples and each sample has a corresponding label.

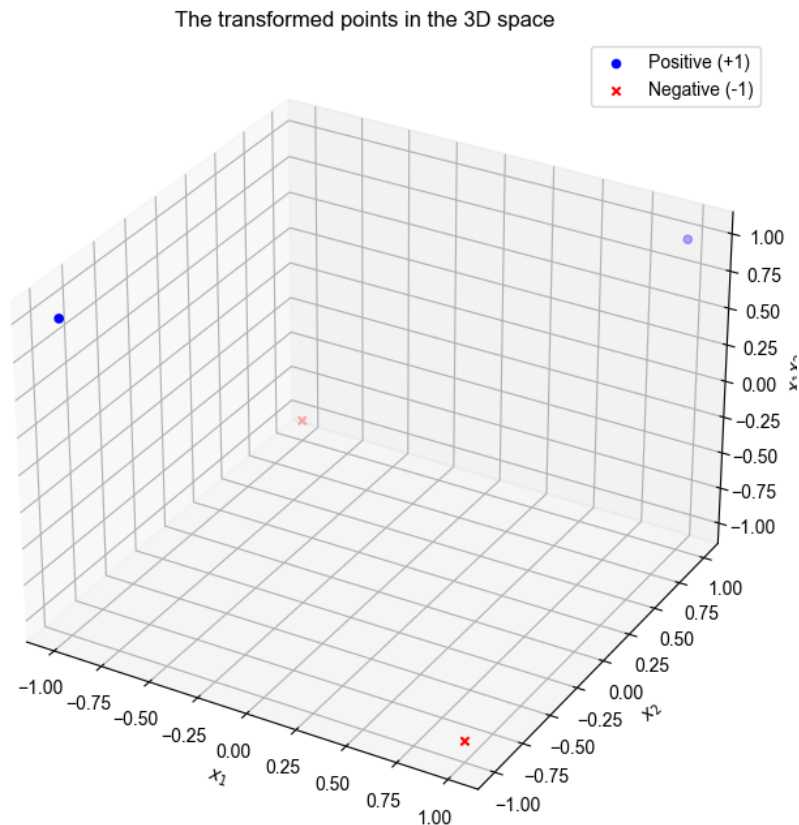
The truth table represents the XOR (exclusive OR) logic gate, where the output is true if and only if the inputs are different.

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



The four points are plotted on the $x - y$ plane as shown in the figure above. It can be found that the points are not linearly separable in the original $2D$ space, as there is no single straight line that can separate the positive examples from the negative examples.

- (c) Consider the feature transformation $\phi(x) = [x_1, x_2, x_1x_2]$, where x_1 and x_2 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 points are mapped into a $3D$ space. In this new space, the transformed points become linearly separable, as there exists a plane that can perfectly separate the positive examples from the negative examples.

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

After applying the feature transformation, all four transformed points have an equal distance to the optimal separating hyperplane. They all contribute to defining the maximum margin between the two classes.

x_1	x_2	x_1x_2	Support Vector
1	1	1	Yes
-1	-1	1	Yes
1	-1	-1	Yes
-1	1	-1	Yes

Since all points are support vectors, it demonstrates the perfect linear separability achieved after the feature transformation. The margin size is calculated to be 2.0, based on the distance between the support vectors and the separating hyperplane.