Homework 3

Mu-Ruei Tseng

March 30, 2024

1 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

1. Sketch the optimal separating hyperplane, and provide the equation for this hyperplane.
2. 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.
3. On your sketch, indicate the margin for the maximal margin hyperplane.
4. Indicate the support vectors for the maximal margin classifier.
5. Does a slight movement of the seventh observation affect the maximal margin hyperplane? Justify your answer.
6. Draw an alternative hyperplane that is not the optimal separating hyperplane, and provide the equation for this hyperplane.
7. Draw an additional observation on the plot so that the two classes are no longer separable by a hyperplane.

1.1 Solution

1.1.1 Sketch the optimal separating hyperplane, and provide the equation for this hyperplane.

To find the optimal separating hyperplane given $p = 2$, we can assume that the hyperplane has the form $\beta_0 + \beta_1 X_1 + \beta_2 X_2 = 0$ and we want to find the hyperplane that maximizes the margin between the two classes. The calculated optimal separating hyperplane is

$$-1 + X_1 - X_2 = 0 \tag{1}$$

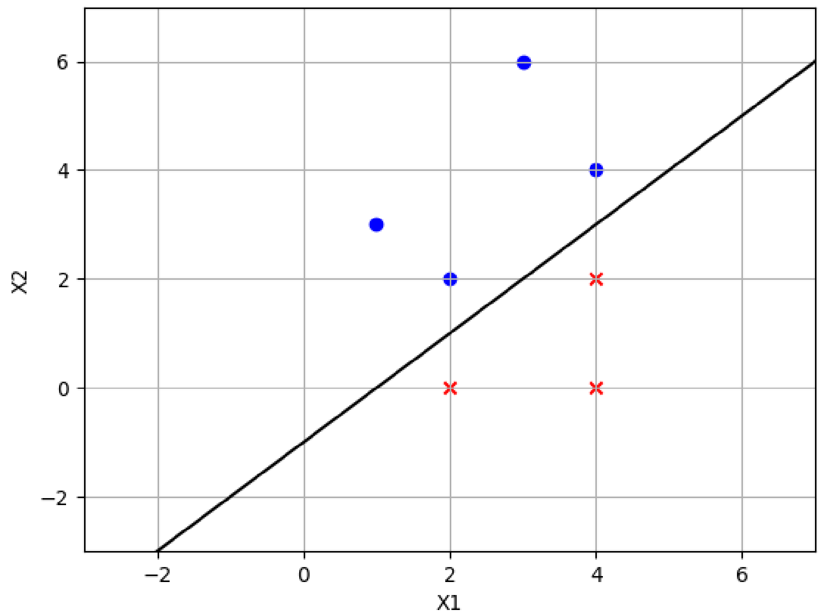

Figure 1: Optimal Separating hyperplane

1.1.2 Describe the classification rule for the maximal margin classifier.

Given the optimal separating hyperplane $-1 + X_1 - X_2 = 0$, we classify Red if $-1 + X_1 - X_2 \geq 0$ and classify Blue otherwise. We have $(\beta_0, \beta_1, \beta_2) = (-1, 1, -1)$. Red: Positive class, Blue: Negative class.

1.1.3 On your sketch, indicate the margin for the maximal margin hyperplane

Margin is the distance between the hyperplane and the observations closest to the hyperplane. We can calculate the distance for every point to the hyperplane. Given a function $L : \beta_0 + \beta_1 X_1 + \beta_2 X_2 = 0$ and a point (x_1, x_2), the distance from (x_1, x_2) to the L is:

$$d = \frac{|\beta_0 + \beta_1 x_1 + \beta_2 x_2|}{\sqrt{\beta_1^2 + \beta_2^2}} \tag{2}$$

Since have $(\beta_0, \beta_1, \beta_2) = (-1, 1, -1)$, the distance for each points to the plane is $d(x_1, x_2)$:

- $d(3,6)$: $\frac{|-1+3-6|}{\sqrt{1^2+(-1)^2}} = \frac{4}{\sqrt{2}} = 2\sqrt{2}$
- $d(2,2)$: $\frac{|-1+2-2|}{\sqrt{1^2+(-1)^2}} = \frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2}$
- $d(4,4)$: $\frac{|-1+4-4|}{\sqrt{1^2+(-1)^2}} = \frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2}$
- $d(1,3)$: $\frac{|-1+1-3|}{\sqrt{1^2+(-1)^2}} = \frac{3}{\sqrt{2}} = \frac{3\sqrt{2}}{2}$
- $d(2,0)$: $\frac{|-1+2-0|}{\sqrt{1^2+(-1)^2}} = \frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2}$

- $d(4,2)$: $\frac{|-1+4-2|}{\sqrt{1^2+(-1)^2}} = \frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2}$
- $d(4,0)$: $\frac{|-1+4-0|}{\sqrt{1^2+(-1)^2}} = \frac{3}{\sqrt{2}} = \frac{3\sqrt{2}}{2}$

Since we can see that the closest distance from the observation to the hyperplane is $\frac{\sqrt{2}}{2}$, the **margin** is $2 \times \frac{\sqrt{2}}{2} = \sqrt{2}$, which we can see also equals $\frac{2}{||w||}$ where $w = (\beta_1, \beta_2) = (1, -1)$.

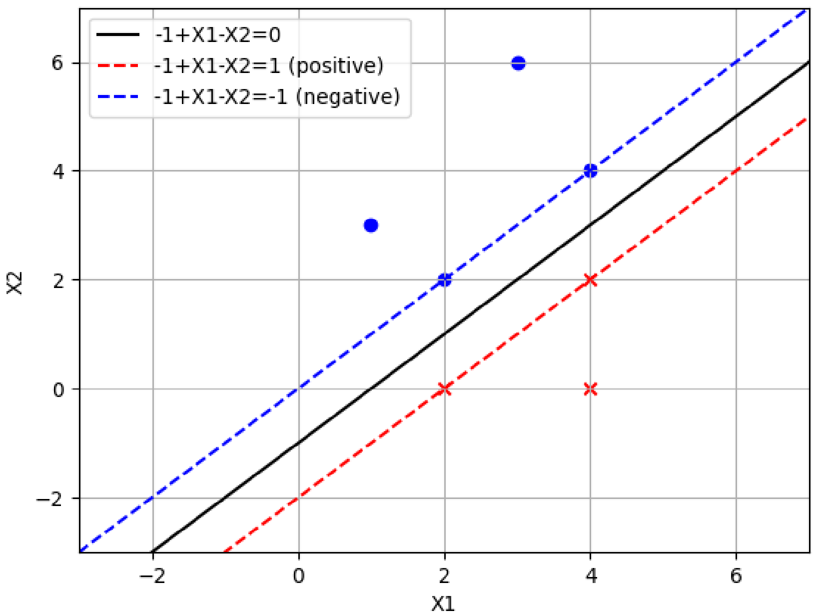

Figure 2: Optimal Separating hyperplane (with its positive/negative hyperplane)

1.1.4 Indicate the support vectors for the maximal margin classifier

The support vectors are the observation (points) that are closest to the hyperplane. It will lie on the positive hyperplane or negative hyperplane (The support vectors of the positive class touch the positive hyperplane, and vice versa). In our case, the positive hyperplane is $-1 + X_1 - X_2 = 1$ and point $(2, 0)$ and $(4, 2)$ lie on it. For the negative hyperplane, it is $-1 + X_1 - X_2 = -1$ and point $(2, 2)$ and $(4, 4)$ lie on it.

Therefore, the support vectors are $(2, 0)$, $(4, 2)$, $(2, 2)$, and $(4, 4)$.

1.1.5 Does a slight movement of the seventh observation affect the maximal margin hyperplane?

The seventh observation is $(X_1, X_2) = (4, 0)$. Since it is not a support vector and is farther from the optimal separating hyperplane, a slight movement of the seventh observation shouldn't affect the maximal margin.

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

Here we define a new hyperplane: $L' = -1 + X_1 - 1.2X_2 = 0$. Following the same procedure in section 1.3, we can calculate the minimum distance from the points to this new hyperplane is:

$$d = \frac{|-1+4-2\times 1.2|}{\sqrt{1^2+(1.2)^2}} \approx 0.384 \tag{3}$$

Therefore, the margin for this hyperplane is $2 * 0.384 \approx 0.77$, which is smaller than the margin of the optimal separating hyperplane (≈ 1.41).

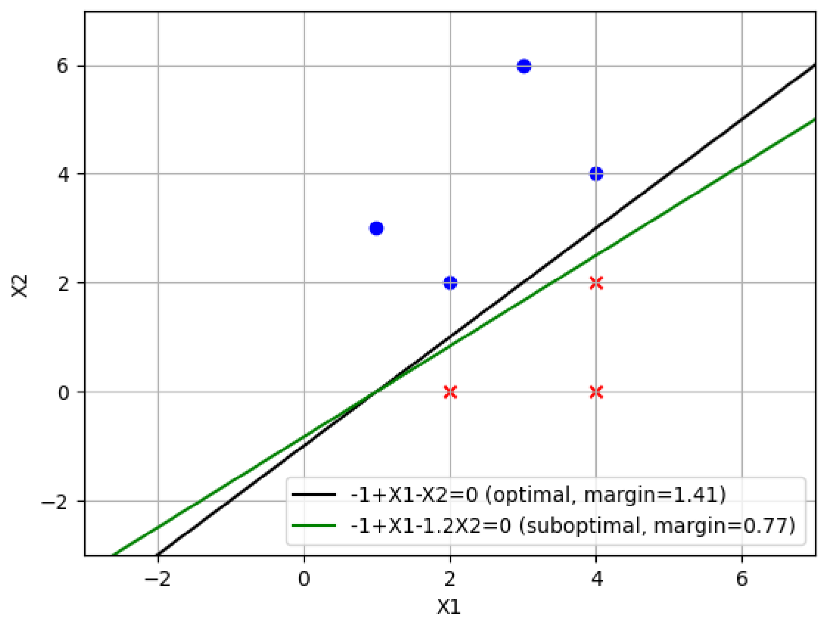

Figure 3: Alternative hyperplane: $-1 + X_1 - 1.2X_2 = 0$

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

For this example, we can add a new observation: $(X_1, X_2, Y) = (0, 6, \text{Red})$. In this case, the Red and Blue points are not linearly separable with a hyperplane in 2D space.

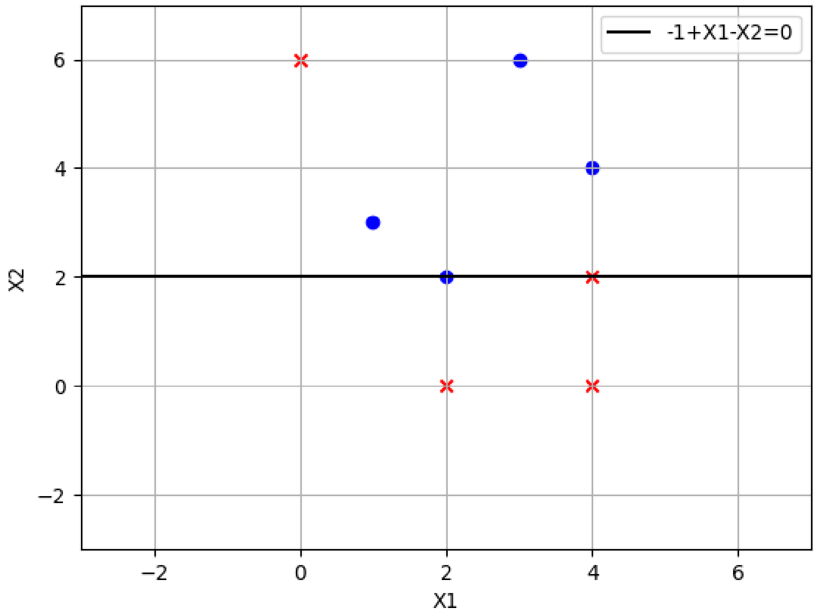

Figure 4: Adding a point at $(0, 6)$ with label Red. This causes the two classes to no longer be separable by a hyperplane in 2D.

2 Math: Question 2 - 20 points

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).

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

2. Draw these four points on x-y plane. Are these points linearly separable?

3. 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?

4. What is the margin size after the transformation? Which points are support?

2.1 Solution

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

We have training data, assume the label for positive examples are 1 and negative examples are -1:

X_1	X_2	y
1	1	1
-1	-1	1
1	-1	-1
-1	1	-1

The shape of X is 4×2 (four samples and two features) while the shape of y is 4×1 (one class label for each sample).

Bonus: The table you just drew is called a truth table. Do you know the logic gate representing this truth table? If we substitute 1 as T and -1 as F in the previous table, we will have:

X_1	X_2	y
T	T	T
F	F	T
T	F	F
F	T	F

, which is indeed a truth table. The logic gate representing the table is an **XNOR** gate.

2.1.2 Draw these four points on x-y plane. Are these points linearly separable?

As we can see from Figure 5, given the four points, the positive and negative classes are not linearly separable by a line (hyperplane in 2D space).

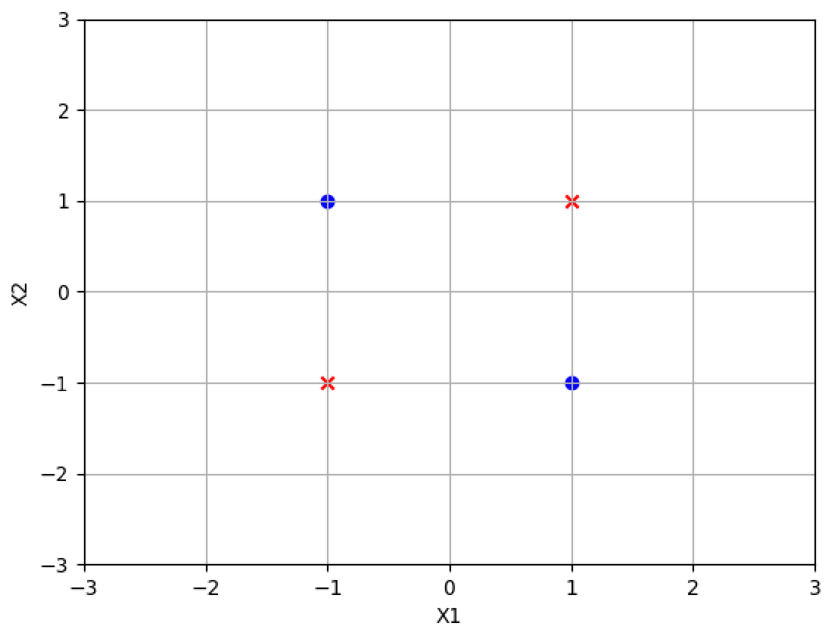

Figure 5: Training data points on the x-y plane.

2.1.3 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?

If we have a transformation function $\phi(x) = [x_1, x_2, x_1x_2]$, our new training data become (assume $X_3 = x_1x_2$):

X_1	X_2	X_3	y
1	1	1	1
-1	-1	1	1
1	-1	-1	-1
-1	1	-1	-1

Now when we plot the data in 3D, we can see that it is linearly separable, with the optimal separating hyperplane $\beta_0 + \beta_1 \times X_1 + \beta_2 \times X_2 + \beta_3 X_1 X_2 = 0$ being where $(\beta_0, \beta_1, \beta_2, \beta_3) = (0, 0, 0, 1)$, which is $X_1 X_2 = 0$

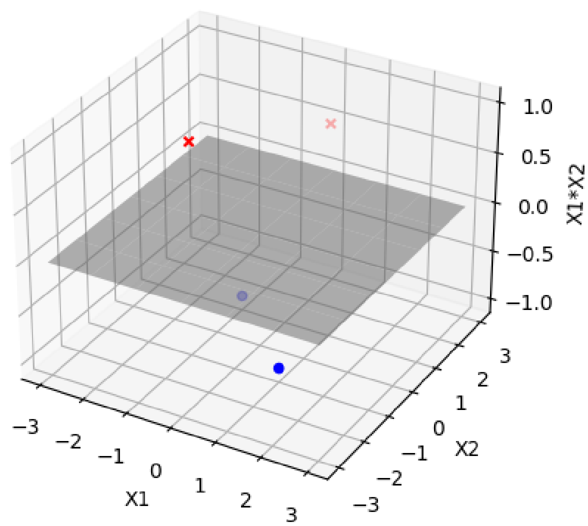

Figure 6: Training data points on the x-y plane.

2.1.4 What is the margin size after the transformation? Which points are support vectors?

Since our optimal separating hyperplane is $X_1 X_2 = 0$ and all the distance from the transformed point to the plane is 1, the margin size should be $1 \times 2 = 2$. Points $(1, 1, 1)$ and $(-1, -1, 1)$ lie on the positive hyperplane ($X_1 X_2 = 1$) and $(1, -1, -1)$ and $(-1, 1, -1)$ lie on the negative hyperplane ($X_1 X_2 = -1$). Therefore, all the points are support vectors.