STATS 415: Homework 9 Solutions

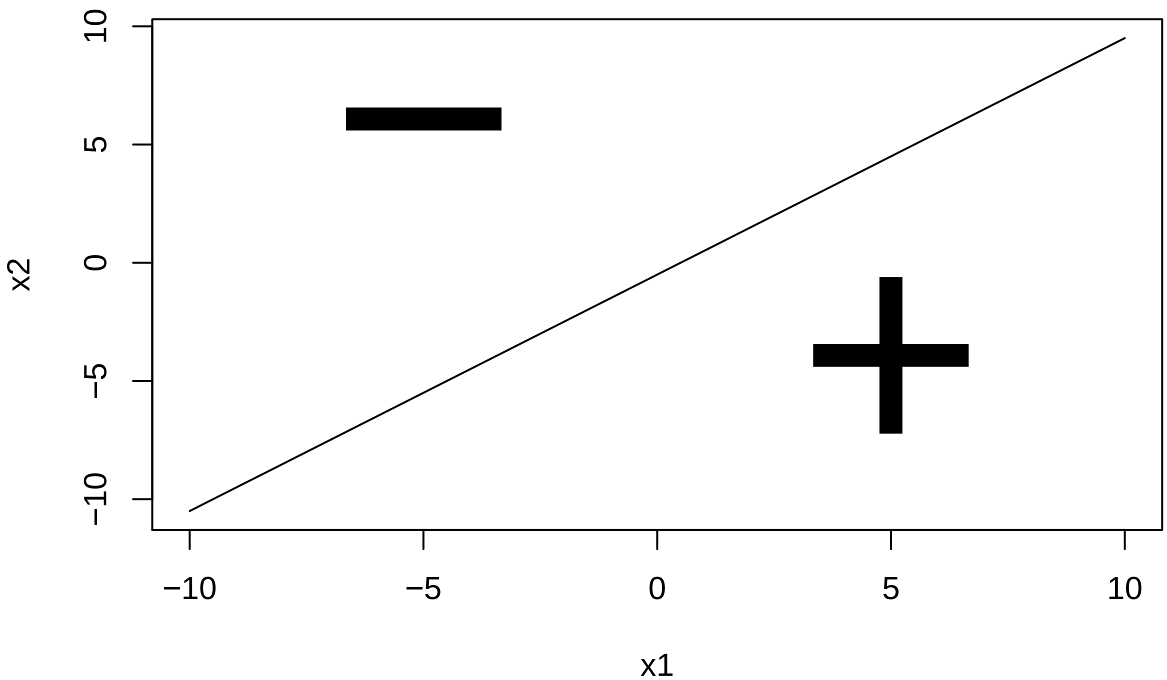
Problem 1:

1. a)

$$2X_1 - 2X_2 - 1 = 0 \implies X_2 = X_1 - \frac{1}{2}$$

$$2X_1 - 2X_2 - 1 > 0 \implies X_1 > X_2 + \frac{1}{2}$$

```
x1 = seq(-10, 10, length.out = 100)
x2 = x1 - 1/2
plot(x1, x2, type = 'l')
points(x = c(-5, 5), y = c(5, -5), pch = c('-', '+'), cex = 10)
```



1. b)

First, define the following:

- $\beta = \begin{pmatrix} 2 \\ -2 \end{pmatrix}$ is a vector perpendicual to the hyperplane
- $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ is an arbitrary point that lies on the desired margin
- x_0 is an arbitrarily chosen point on the hyperplane (here the point corresponding to $X_1 = 0$ is used)

Then the points lying on the $\sqrt{2}$ margin on the positive side will satisfy:

$$\sqrt{2} = \left\langle \frac{\beta}{||\beta||}, x - x_0 \right\rangle \implies \sqrt{2} = \frac{2x_1 - 2x_2 - \beta^T x_0}{\sqrt{8}} \implies 4 = 2x_1 - 2x_2 - (1)$$

This gives the formula of the line as: $x_2 = x_1 - 5/2$.

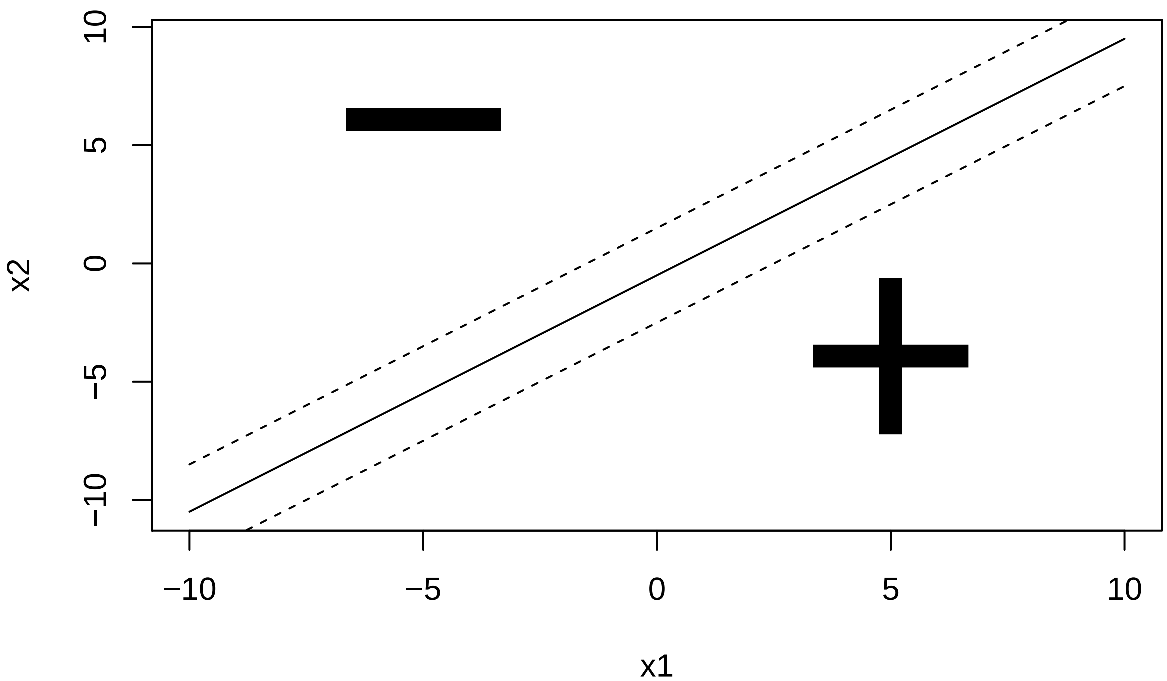
Following similar logic, the equation of the line corresponding to the $-\sqrt{2}$ margin will be given by solving:

$$-\sqrt{2} = \left\langle \frac{\beta}{||\beta||}, x - x_0 \right\rangle \implies -\sqrt{2} = \frac{2x_1 - 2x_2 - \beta^T x_0}{\sqrt{8}} \implies -4 = 2x_1 - 2x_2 - (1),$$

so the equation of the other line is $x_2 = x_1 + 3/2$.

Note: An easier way to compute the equations of the margin boundaries is to notice that the decision boundary makes an angle of $\pi/4$ with the y-axis, which implies that shifting the line up/down by 2 corresponds to a perpendical distiance of $\sqrt{2}$.

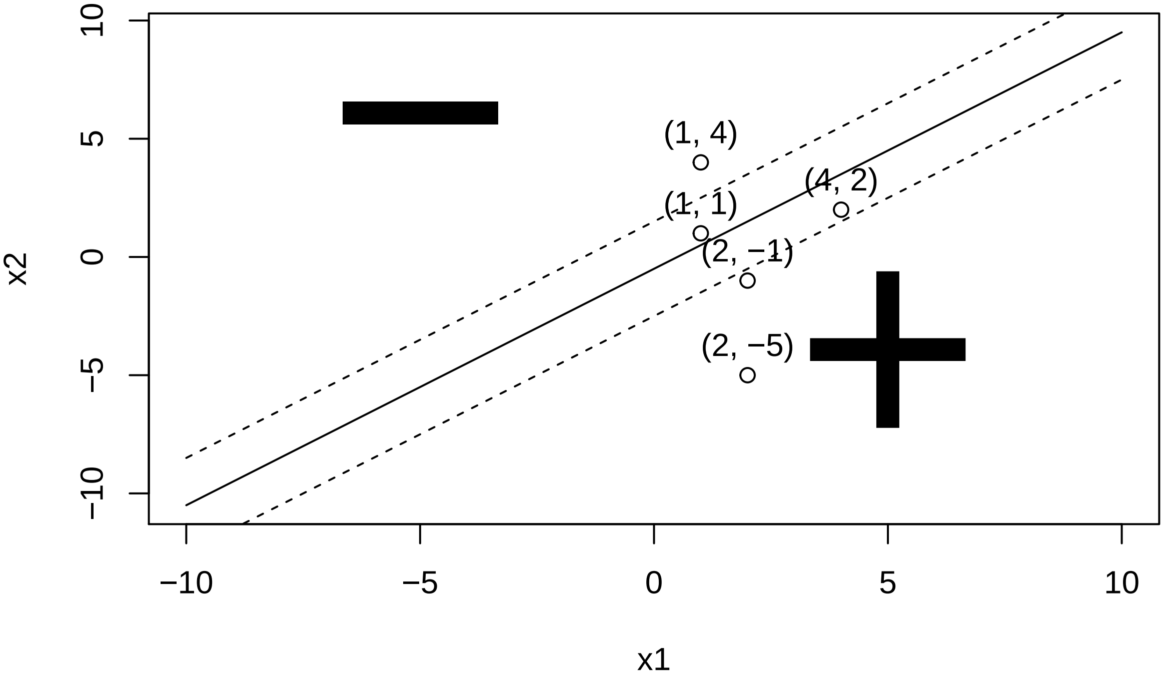
```
x1 = seq(-10, 10, length.out = 100)
x2 = x1 - 1/2
plot(x1, x2, type = 'l')
points(x = c(-5, 5), y = c(5, -5), pch = c('-', '+'), cex = 10)
lines(x = x1, y = x1 + 3/2, lty = 'dashed')
lines(x = x1, y = x1 - 5/2, lty = 'dashed')
```



1. c)

```
x1 = seq(-10, 10, length.out = 100)
x2 = x1 - 1/2
plot(x1, x2, type = 'l')
points(x = c(-5, 5), y = c(5, -5), pch = c('-', '+'), cex = 10)
lines(x = x1, y = x1 + 3/2, lty = 'dashed')
lines(x = x1, y = x1 - 5/2, lty = 'dashed')
```

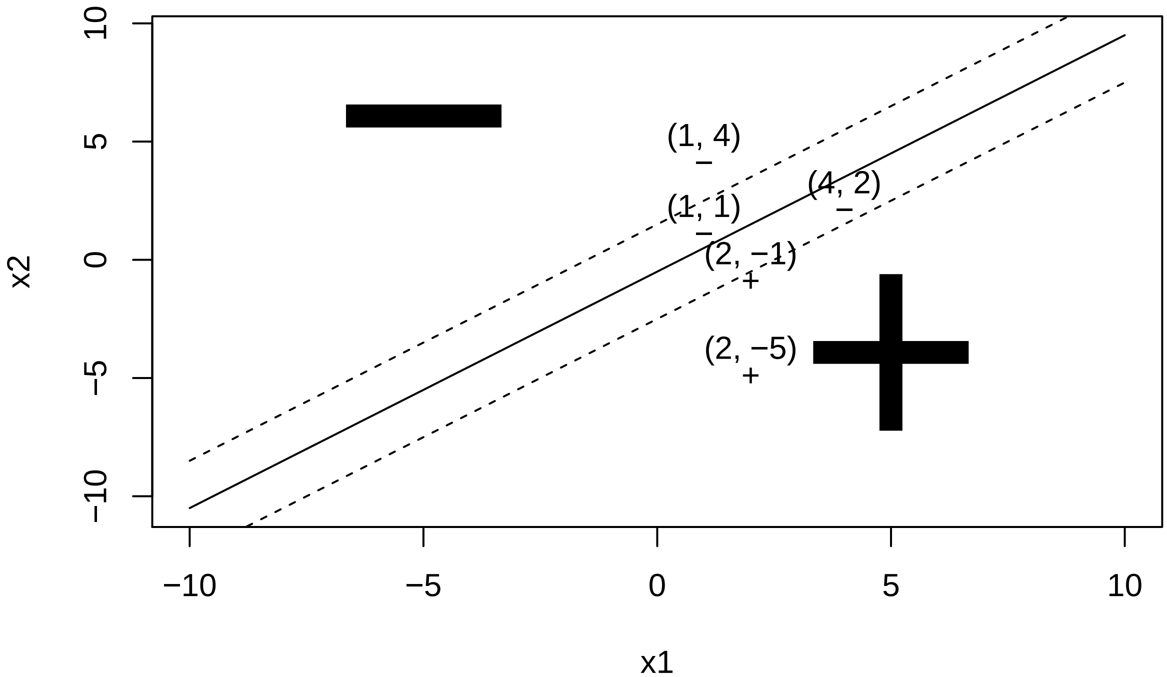
```r
points(x = c(1, 1, 2, 2, 4), y = c(4, 1, -5, -1, 2))
text(x = c(1, 1, 2, 2, 4), y = c(4, 1, -5, -1, 2),
       labels = c('(1, 4)', '(1, 1)', '(2, -5)', '(2, -1)', '(4, 2)'),
     pos = 3)
```



The classifier predicts the + class for the following points: $(2, -5), (2, -1), (4, 2)$. The classifier predicts the - class for the following points: $(1, 4), (1, 1)$.

1. d)

```r
x1 = seq(-10, 10, length.out = 100)
x2 = x1 - 1/2
plot(x1, x2, type = 'l')
points(x = c(-5, 5), y = c(5, -5), pch = c('-', '+'), cex = 10)
lines(x = x1, y = x1 + 3/2, lty = 'dashed')
lines(x = x1, y = x1 - 5/2, lty = 'dashed')
points(x = c(1, 1, 2, 2, 4), y = c(4, 1, -5, -1, 2),
       pch = c('-', '-', '+', '+', '-'))
text(x = c(1, 1, 2, 2, 4), y = c(4, 1, -5, -1, 2),
       labels = c('(1, 4)', '(1, 1)', '(2, -5)', '(2, -1)', '(4, 2)'),
     pos = 3)
```



The picture shows that the points $(2,-5)$, $(2,-1)$, and $(1,4)$ are classified correctly and outside the margins. Therefore their slack values are zero.

To find the slack values for the other two points, compute their respective distances from the boundary:

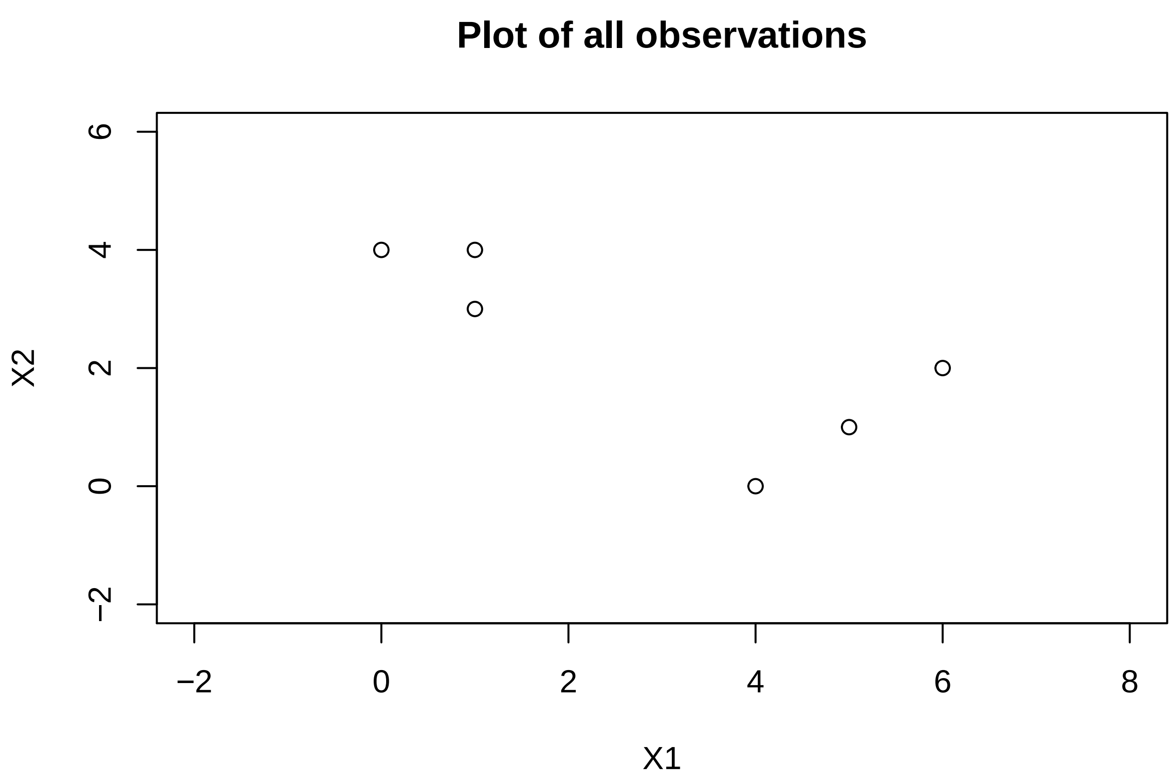
$$\sqrt{2}(1-\xi_{(1,1)}) = -\frac{1}{||\beta||}\left(\langle(1,1),(2,-2)\rangle + \beta_0\right) \implies \xi_{(1,1)} = 1 + \frac{-1}{2\sqrt{2}}\left(\frac{1}{\sqrt{2}}\right) = \frac{3}{4}$$

$$\sqrt{2}(1-\xi_{(4,2)}) = -\frac{1}{||\beta||}\left(\langle(4,2),(2,-2)\rangle + \beta_0\right) \implies \xi_{(4,2)} = 1 + \frac{4-1}{2\sqrt{2}}\left(\frac{1}{\sqrt{2}}\right) = \frac{7}{4}$$

Problem 2:

2. a)

```
X=matrix(c(1,1,0,5,6,4,4,3,4,1,2,0), byrow =F, ncol=2)
plot(X, xlab = 'X1', ylab = 'X2', main = 'Plot of all observations',
     ylim=c(-2,6), xlim = c(-2,8))
```



2. b)

```
set.seed(1)
label=sample(c(1,2), size=6, replace = T, prob = c(0.5,0.5))
XWithLabel=cbind(1:6, X, label)
colnames(XWithLabel)=c('Obs','X1','X2','Class')
XWithLabel
```

```
##      Obs X1 X2 Class
## [1,]   1  1  4     2
## [2,]   2  1  3     2
## [3,]   3  0  4     1
## [4,]   4  5  1     1
## [5,]   5  6  2     2
## [6,]   6  4  0     1
```

2. c)

```
ind1=which(XWithLabel[,4]==1)
ind2=which(XWithLabel[,4]==2)
centroid1=apply(X[ind1,], 2, mean)
centroid2=apply(X[ind2,], 2, mean)
centroid=rbind(centroid1,centroid2)
colnames(centroid)=c('X1','X2')
centroid
```

```
##                 X1       X2
```

```
## centroid1 3.000000 1.666667
## centroid2 2.666667 3.000000
```

2. d)

```r
cluster=apply(X, 1, function(x) which.min(c(sum((x-centroid1)^2),sum((x-centroid2)^2))))
cluster
```

```
## [1] 2 2 2 1 1 1
```

2. e)

After only one iteration, the class labels start to remain unchanged.

```r
centroid1=apply(X[which(cluster==1),], 2, mean)
centroid2=apply(X[which(cluster==2),], 2, mean)
cluster1=apply(X, 1, function(x) which.min(c(sum((x-centroid1)^2),sum((x-centroid2)^2))))
cluster1
```

```
## [1] 2 2 2 1 1 1
```

2. f)

```r
plot(X, col=cluster1+1, xlab = 'X1', ylab = 'X2', main = 'Plot of clustered observations',
     ylim=c(-2,6), xlim = c(-2,8))
```