CMDA 4654: Intermed Data Analytics & ML
Homework 5
Christopher J. Mobley
Due: Monday, 1MAY16

1. (9.7 – 1) This problem involves hyperplanes in two dimensions.
   a. Sketch the hyperplane $1 + 3X_1 - X_2 = 0$. Indicate the set of points for which $1 + 3X_1 - X_2 > 0$, as well as the set of points for which $1 + 3X_1 - X_2 < 0$.

$$1 + 3X_1 - X_2 = 0$$

$$X_2 = 1 + 3X_1$$

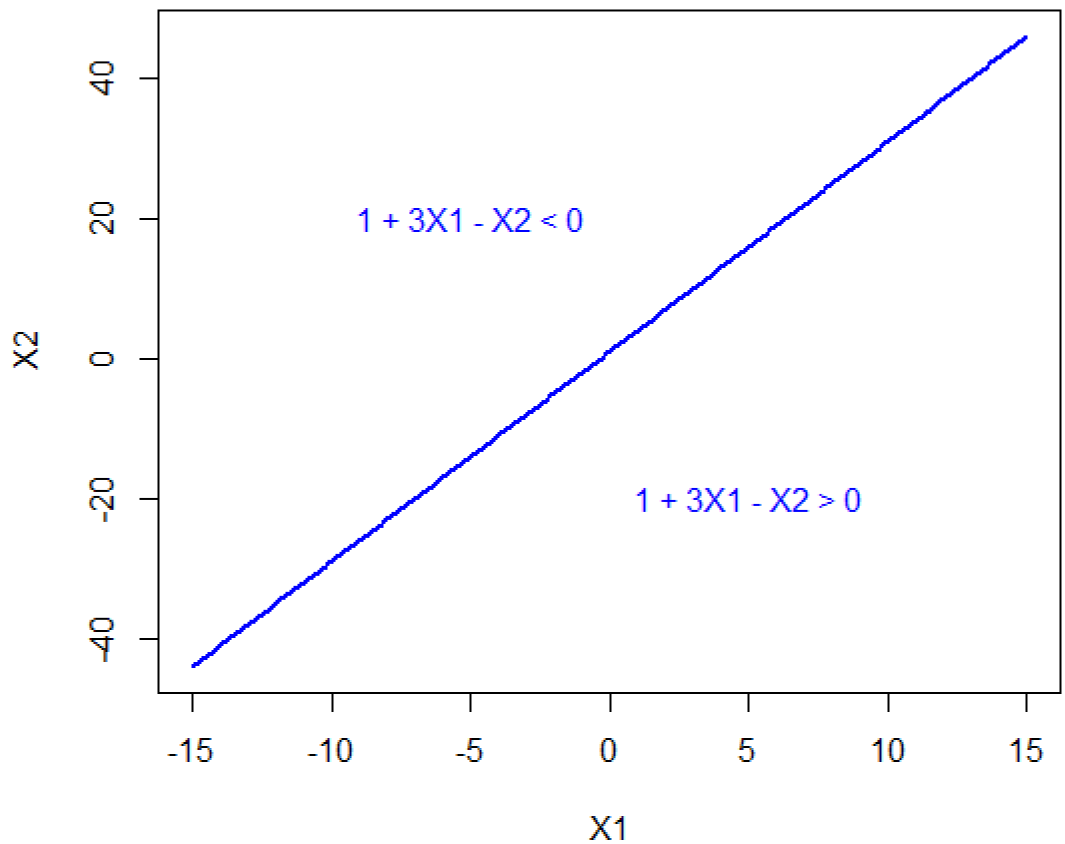

Figure 1: Plotted line of $1 + 3X_1 - X_2 = 0$ with >0 and <0 regions marked

Figure 1 shows that the points above the line, $1 + 3X_1 - X_2 = 0$, are < 0 while the point below the line are > 0.

b. On the same plot, sketch the hyperplane $-2 + X_1 + 2X_2 = 0$. Indicate the set of points for which $-2 + X_1 + 2X_2 > 0$, as well as the set of points for which $-2 + X_1 + 2X_2 < 0$.

$$-2 + X_1 + 2X_2 = 0$$

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

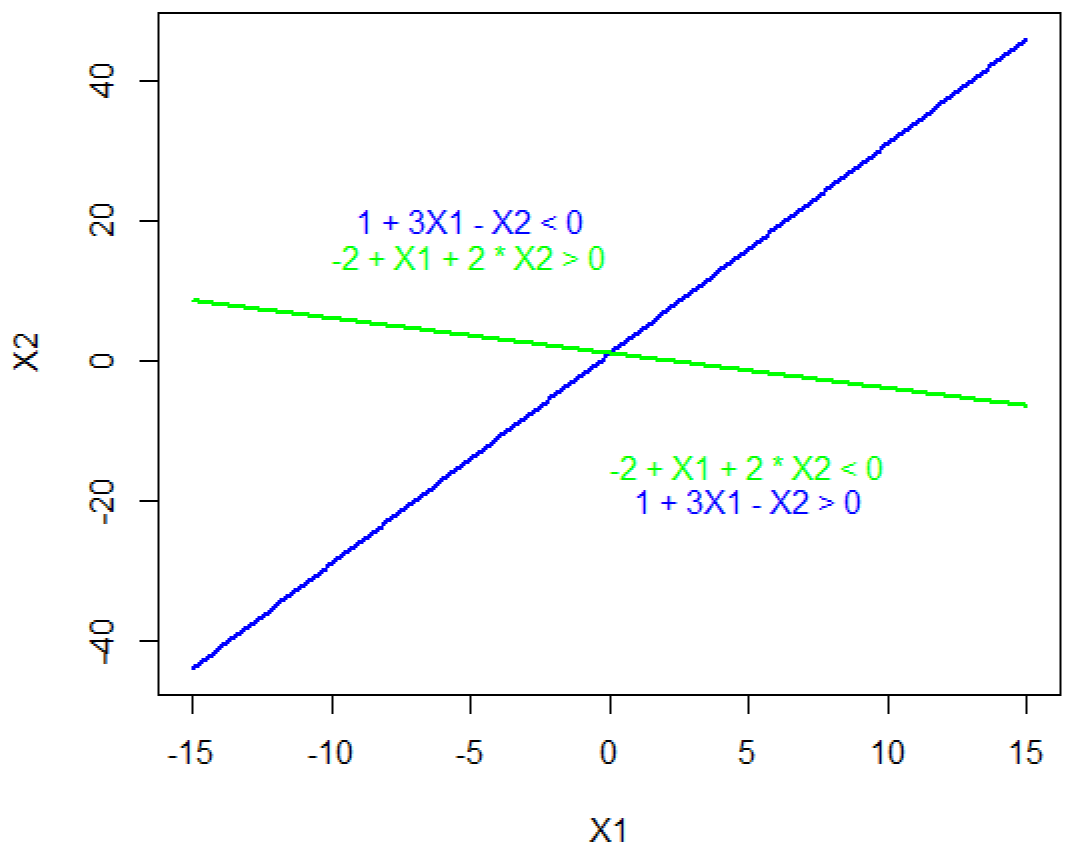

Figure 2: Plotted line of $-2 + X_1 + 2X_2 > 0$ with >0 and <0 regions marked

Figure 2 shows that the points above the line, $-2 + X_1 + 2X_2 = 0$, are > 0 while the point below the line are < 0.

2. (9.7 – 2) We have seen that in p = 2 dimensions, a linear decision boundary takes the form $\beta_0+\beta_1X_1+\beta_2X_2 = 0$. We now investigate a non-linear decision boundary.
   a. Sketch the curve $(1 + X_1)^2 + (2 - X_2)^2 = 4$.

$(1 + X_1)^2 + (2 - X_2)^2 = 4$ is a circle centered at (-1, 2) with a radius of 2. So,

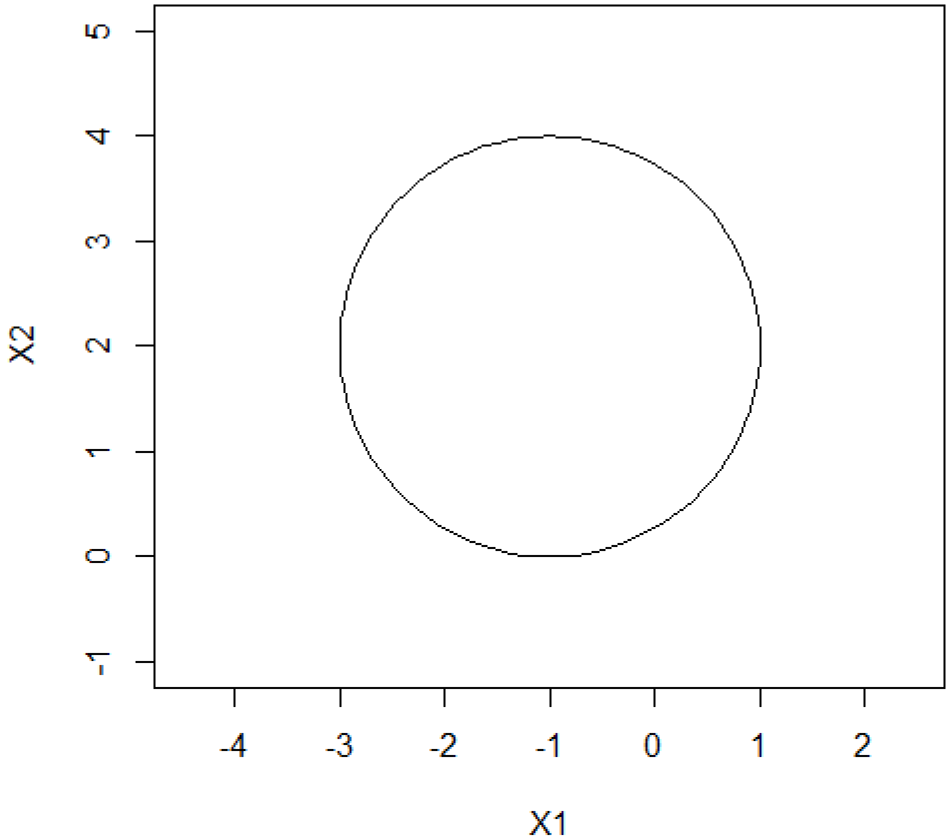

Figure 3: Plotted curve $(1 + X_1)^2 + (2 - X_2)^2 = 4$

b. On your sketch, indicate the set of points for which

$$(1 + X_1)^2 + (2 - X_2)^2 > 4$$

as well as the set of points for which

$$(1 + X_1)^2 + (2 - X_2)^2 \leq 4.$$

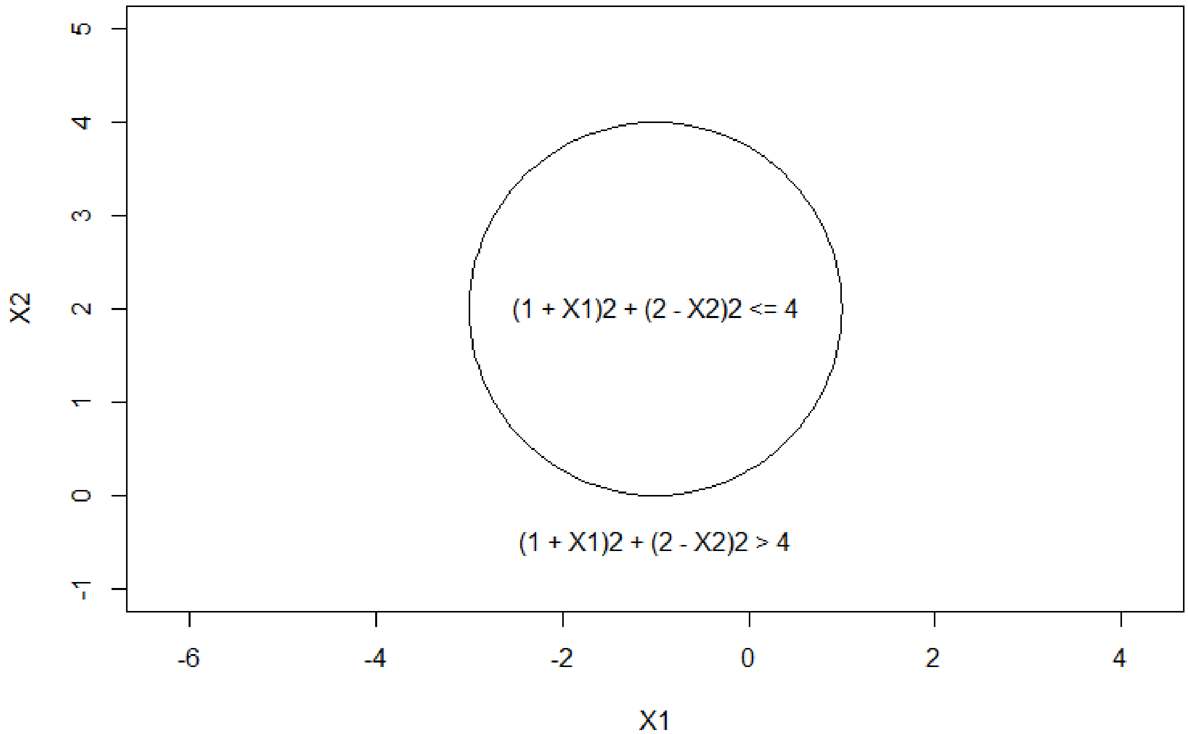

Figure 4: Plotted curve $(1 + X_1)^2 + (2 - X_2)^2 = 4$ with > 4 and ≤ 4 regions marked

Figure 4 shows that points inside the curve, $(1 + X_1)^2 + (2 - X_2)^2 = 4$, are ≤ 4 while the point outside the curve are > 4.

c. Suppose that a classifier assigns an observation to the blue class if

$$(1 + X1)^2 + (2 - X2)^2 > 4,$$

and to the red class otherwise. To what class is the observation (0, 0) classified? (−1, 1)? (2, 2)? (3, 8)?

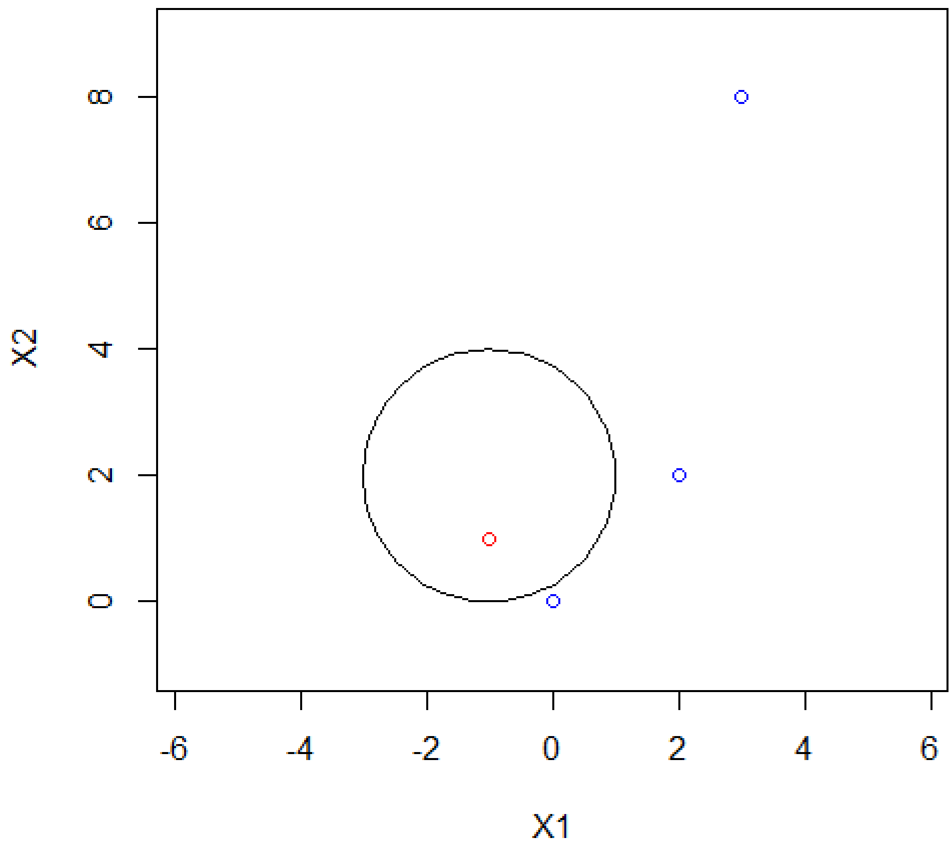

Figure 5: Plotted curve $(1 + X_1)^2 + (2 - X_2)^2 = 4$ with classified points

The points would be marked blue, red, blue and blue respectively as shown in Figure 5.

d. Argue that while the decision boundary in (c) is not linear in terms of X_1 and X_2, it is linear in terms of X_1, X_1^2, X_2, and X_2^2.

After expanding each quadratic expression, the left-hand-side of the equation become $5 + 2X_1 - 4X_2 + X_1^2 + X_2^2$ which is simply a linear equation in terms of X_1, X_1^2, X_2, and X_2^2. The steps to get to the aforementioned equation are below.

$$(1 + X_1)^2 + (2 - X_2)^2 > 4$$

$$1 + 2X_1 + X_1^2 + 4 - 4X_2 + X_2^2 > 4$$

$$5 + 2X_1 - 4X_2 + X_1^2 + X_2^2 > 4$$

3. (9.7 – 4) Generate a simulated two-class data set with 100 observations and two features in which there is a visible but non-linear separation between the two classes. Show that in this setting, a support vector machine with a polynomial kernel (with degree greater than 1) or a radial kernel will outperform a support vector classifier on the training data. Which technique performs best on the test data? Make plots and report training and test error rates in order to back up your assertions.

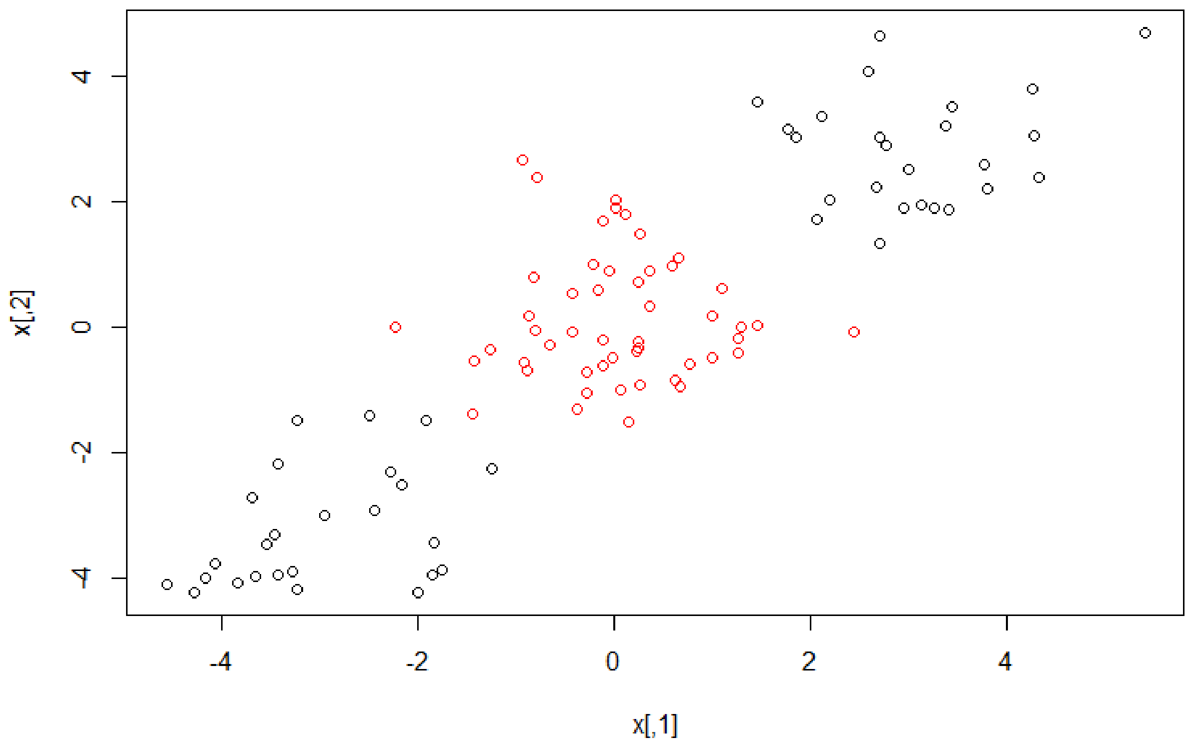

Figure 6: Simulated two-class data set with non-linear separation between the two classes

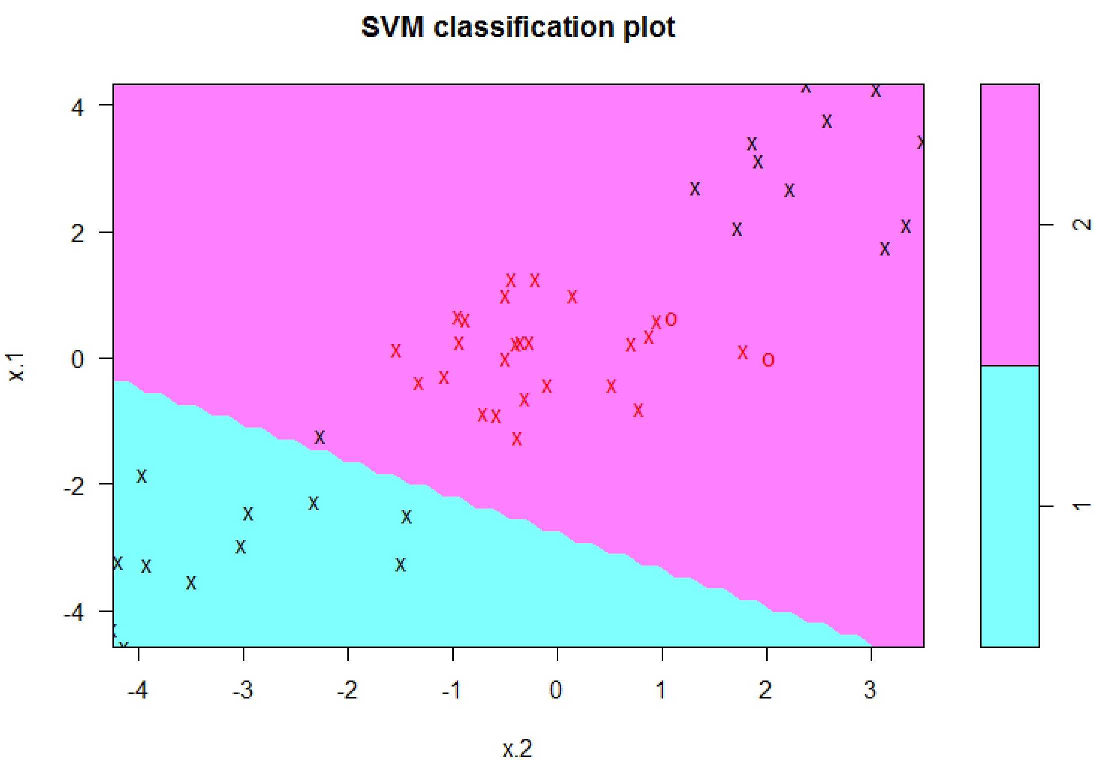

Figure 7: Linear SVC classification plot

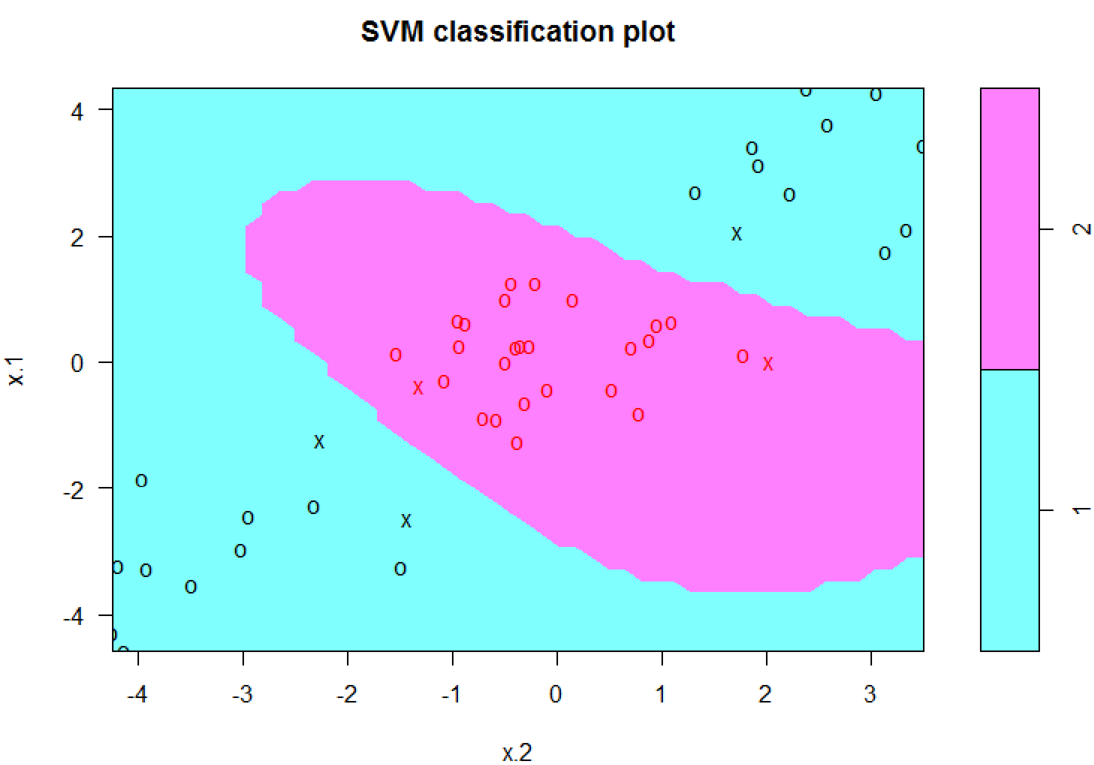

Figure 8: Radial SVM classification plot

Linear SVC Train Confusion Matrix		
	1	2
1	11	12
2	0	27

Linear SVC Test Confusion Matrix		
	1	2
1	12	15
2	0	23

Radial SVM Train Confusion Matrix		
	1	2
1	23	0
2	0	27

Radial SVM Test Confusion Matrix		
	1	2
1	27	0
2	2	21

Figure 9: Linear SVC and Radial SVM Confusion Matrices

Figure 6 shows a simulated two-class data set with non-linear separation between the two classes. Figures 7 and 8 show how both a linear support vector classifier and a radial support vector machine will attempt to classify the data. Figures 7 and 8 support our assumption that the radial support vector machine will outperform the linear support vector classifier. This assumption is verified in Figure 8 which show the train and test confusion matrices for each classifier. The linear support vector classifier train and test error are 0.24 and 0.30 respectively. While the radial support vector machine train and test accuracy are 0.00 and 0.04 respectively.

4. (9.7 – 8) This problem involves the OJ data set which is part of the ISLR package.
   a. Create a training set containing a random sample of 800 observations, and a test set containing the remaining observations.
   b. Fit a support vector classifier to the training data using cost=0.01, with Purchase as the response and the other variables as predictors. Use the summary() function to produce summary statistics, and describe the results obtained.

```
Call:
svm(formula = Purchase ~ ., data = train, kernel = "linear", cost = 0.01)


Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  linear
       cost:  0.01
      gamma:  0.05555556

Number of Support Vectors:  441

 ( 219 222 )


Number of Classes:  2

Levels:
 CH MM
```

Figure 10: Linear SVC summary

Figure 10 shows a summary for the linear svc, which includes the number of support vectors.

c. What are the training and test error rates?

The linear support vector classifier training and test error rates are 0.16625 and 0.1407407 respectively.

d. Use the tune() function to select an optimal cost. Consider values in the range 0.01 to 10.

The optimal cost is 1.1.

e. Compute the training and test error rates using this new value for cost.

The tuned linear support vector classifier training and test error rates are 0.1625 and 0.1481481 respectively.

f. Repeat parts (b) through (e) using a support vector machine with a radial kernel. Use the default value for gamma.

```
Call:
svm(formula = Purchase ~ ., data = train, kernel = "radial", cost = 0.01)


Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  radial
       cost:  0.01
      gamma:  0.05555556

Number of Support Vectors:  620

 ( 309 311 )


Number of Classes:  2

Levels:
 CH MM
```

Figure 11: Radial SVM summary

Figure 11 shows a summary for the radial svm, which includes the number of support vectors.

The radial support vector machine training and test error rates are 0.38625 and 0.40 respectively.

The optimal cost is 0.5.

The tuned radial support vector machine training and test error rates are 0.1525 and 0.1518519 respectively.

g. Repeat parts (b) through (e) using a support vector machine with a polynomial kernel. Set degree=2.

```
Call:
svm(formula = Purchase ~ ., data = train, kernel = "polynomial", cost = 0.01, degree = 2)


Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  polynomial
       cost:  0.01
     degree:  2
      gamma:  0.05555556
     coef.0:  0

Number of Support Vectors:  624

 ( 309 315 )


Number of Classes:  2

Levels:
 CH MM
```

Figure 12: Polynomial SVM summary

Figure 12 shows a summary for the degree 2 polynomial svc, which includes the number of support vectors.

The degree 2 polynomial support vector machine training and test error rates are 0.38625 and 0.40 respectively.

The optimal cost is 8.

The tuned degree 2 polynomial support vector machine training and test error rates are 0.14125 and 0.1592593 respectively.

h. Overall, which approach seems to give the best results on this data?

While each tuned model performed within test error rates of .01 of each other, the linear support vector classifier had the best test error rate of 0.1481481. Therefore, giving the best results.