**R Lab #7 – Solutions**

## 1. 9.6.1 Support Vector Classifier

Work through the lab from pages 359-360. Confirm that the support vector classifier with cost=10 uses seven support vectors and the classifier with cost=0.1 uses 16 support vectors. What happens to the decision boundary in the plot when we change the cost from 10 to 0.1? (If you note where the support vectors are for each plot you'll see why it moves this way.)

a) It moves down and to the left.
b) It moves up and to the right.

**Solution:** B

```
svmfit <- svm(y ~., data = dat, kernel = "linear", cost = 10, scale = F)
plot(svmfit, dat)
summary(svmfit)

svmfit <- svm(y ~., data = dat, kernel = "linear", cost = 0.10, scale = F)
summary(svmfit)
plot(svmfit, dat)
```

## 2. 9.6.1 Support Vector Classifier

Continue with this section of the lab on page 361 and the top of page 362. Confirm that when using cost=0.1 only 1 test observation is incorrectly classified, while when cost=0.01 then 2 test observations are incorrectly classified. How many test observations are incorrectly classified when cost=1?

a) 0
b) 1
c) 2
d) 3

**Solution:** B

```
svmfit <- svm(y ~., data = dat, kernel = "linear", cost = 0.1, scale = F)
ypred <- predict(svmfit, testdat)
table(predict = ypred, truth = testdat$y)

svmfit <- svm(y ~., data = dat, kernel = "linear", cost = 0.01, scale = F)
ypred <- predict(svmfit, testdat)
table(predict = ypred, truth = testdat$y)

svmfit <- svm(y ~., data = dat, kernel = "linear", cost = 1.00, scale = F)
```

```
ypred <- predict(svmfit, testdat)
table(predict = ypred, truth = testdat$y)
```

### 3. 9.6.1 Support Vector Classifier

Work through the remaining part of this section of the lab from pages 362-363. Confirm that the support vector classifier with cost=100000 uses three support vectors and the classifier with cost=1 uses seven support vectors. What happens to the slope of the decision boundary in the plot when we change the cost from 100000 to 1? (If you note where the support vectors are for each plot you'll see why it moves this way.)

a) The slope decreases (becomes shallower)
b) The slope increases (becomes steeper)

**Solution:** A

```
svmfit <- svm(y ~., data = dat, kernel = "linear", cost = 1e5)
summary(svmfit)
plot(svmfit, dat)

svmfit <- svm(y ~., data = dat, kernel = "linear", cost = 1)
summary(svmfit)
plot(svmfit, dat)
```

### 4. 9.6.2 Support Vector Machine

Generate and plot the data for this section as instructed on page 363. What is the best description of the appearance of the data?

a) One class dominates the lower-center portion of the graph with the other class dispersed to the upper-left and upper-right.
b) One class dominates the upper-center portion of the graph with the other class dispersed to the lower-left and lower-right.
c) One class dominates the center part of the graph with the other class dispersed to the lower-left and upper-right.
d) One class dominates the center part of the graph with the other class dispersed to the upper-left and lower-right.

**Solution:** C

```
plot(x, col = y)
```

**5. 9.6.2 Support Vector Machine**

Continue with this section of the lab on pages 363 and 364. Confirm that the SVM with gamma=1 and cost=1 results in an approximate oval shape for the decision boundary. True or false? The SVM with gamma=1 and cost=100000 results in a slightly irregular oval shape for the decision boundary.

**Solution:** False

```
svmfit <- svm(y ~., data = dat[train,], kernel = "radial", gamma = 1, cost = 1)
plot(svmfit, dat[train,])

svmfit <- svm(y ~., data = dat[train,], kernel = "radial", gamma = 1, cost = 1e5)
plot(svmfit, dat[train,])
```

**6. 9.6.2 Support Vector Machine**

Complete this section of the lab by assessing the SVM with cost=1 and gamma=2 on the test set. Match the counts of test observations to each prediction/truth combination. (Note for those using the fourth edition of the book: On page 364, the "newx=dat[-train,]" in the R command at the bottom of the page is incorrect. It should read "newdata=dat[-train,]". Correspondingly the 39% at the bottom of the page should be 10%. The latest version (sixth edition) contains these corrections.).

a) 16
b) 7
c) 3
d) 74

- True class 1, predicted class 1 (D)
- True class 1, predicted class 2 (C)
- True class 2, predicted class 1 (B)
- True class 2, predicted class 2 (A)

**Solution:**

```
table(true = dat[-train, "y"], pred = predict(tune.out$best.model, newdata = dat[-train,]))
```

**7. 9.6.3 ROC Curves**

Work through this section of the lab on pages 365-6. What is the best description of the ROC plots produced?

a) For the training data the curve for gamma=2 is closer to the top-left corner than the curve for gamma=50; For the test data the curve for gamma=2 is closer to the top-left corner than the curve for gamma=50.
b) For the training data the curve for gamma=50 is closer to the top-left corner than the curve for gamma=2; For the test data the curve for gamma=2 is closer to the top-left corner than the curve for gamma=50.
c) For the training data the curve for gamma=2 is closer to the top-left corner than the curve for gamma=50; For the test data the curve for gamma=50 is closer to the top-left corner than the curve for gamma=2.
d) For the training data the curve for gamma=50 is closer to the top-left corner than the curve for gamma=2; For the test data the curve for gamma=50 is closer to the top-left corner than the curve for gamma=2.

**Solution:** B

**8. 9.6.4 SVM with Multiple Classes**

Work through this section of the lab on page 366. What is the best description of the resulting plot of the SVM regions?

a) One region covers the center-left part of the graph, another region covers the upper-right part of the graph, and the remainder covers the upper-left and lower-right parts of the graph.
b) One region covers the center-left part of the graph, another region covers the lower-right part of the graph, and the remainder covers the lower-left and upper-right parts of the graph.
c) One region covers the upper-left part of the graph, another region covers the center-right part of the graph, and the remainder covers the lower-left and upper-right parts of the graph.
d) One region covers the lower-left part of the graph, another region covers the center-right part of the graph, and the remainder covers the upper-left and lower-right parts of the graph.

**Solution:** B

```
svmfit <- svm(y ~., data = dat, kernel = "radial", cost = 10, gamma = 1)
plot(svmfit, dat)
```