p8106_stl2137_hw5

```
data("OJ")
dat_oj <- OJ %>%
  janitor::clean_names()

set.seed(1)

rowTrain <- createDataPartition(y = dat_oj$purchase,
                                p = 799/1070,
                                list = FALSE)
```
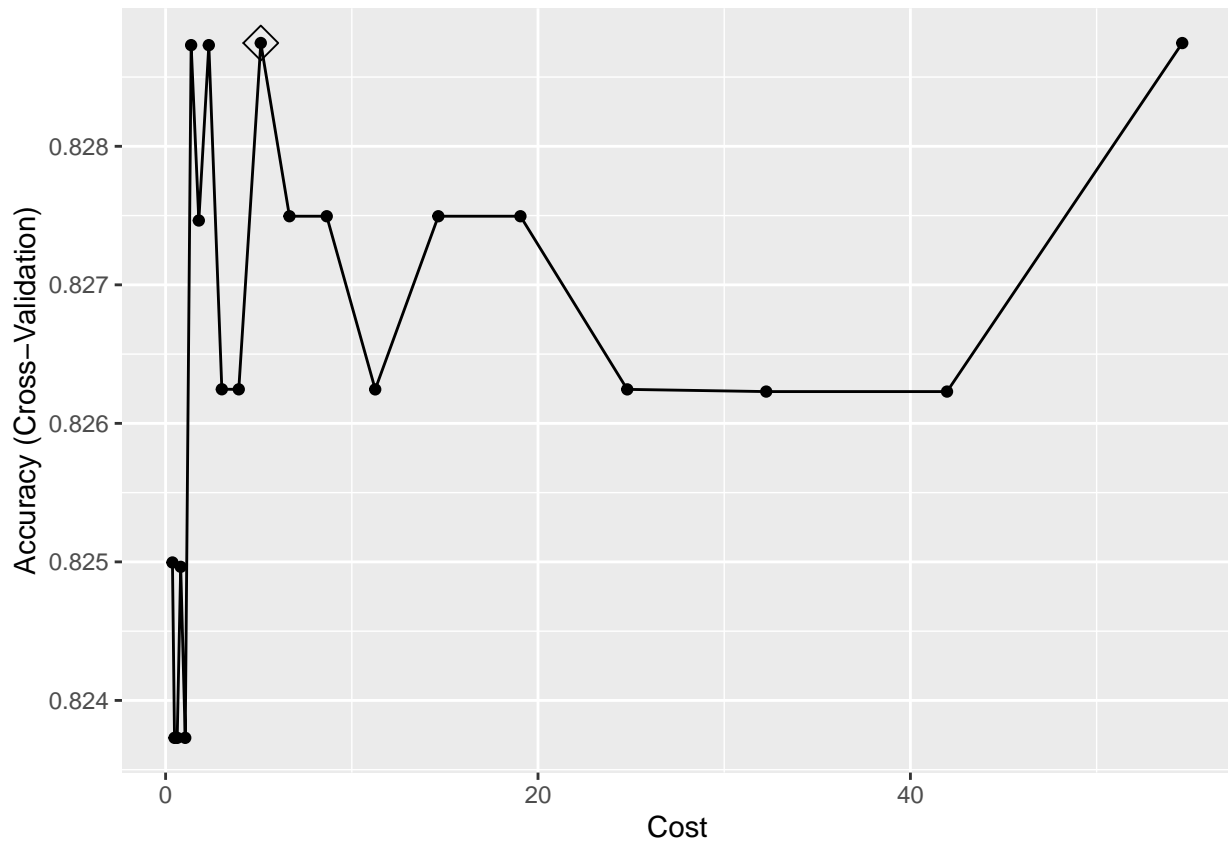
Part A

What are the training and test error rates?

Linear Training Data Performance

```
control1 <- trainControl(method = "cv")

set.seed(1)
svm_linear_fit <- train(purchase ~ .,
                 data = dat_oj[rowTrain,],
                 method = "svmLinear2",
                 preProcess = c("center", "scale"),
                 tuneGrid = data.frame(cost = exp(seq(-1,4,len=20))),
                 trControl = control1)

ggplot(svm_linear_fit, highlight = TRUE)
```

The training error rate for the linear kernel support vector classifier is 0.1712543.

Linear Test Data Performance

```
pred_svm_linear <- predict(svm_linear_fit, newdata = dat_oj[-rowTrain,])


confusion_matrix_svm_linear <- confusionMatrix(data = pred_svm_linear,
                reference = dat_oj$purchase[-rowTrain])
```

The test error rate for the linear kernel support vector classifier is 0.1555556.
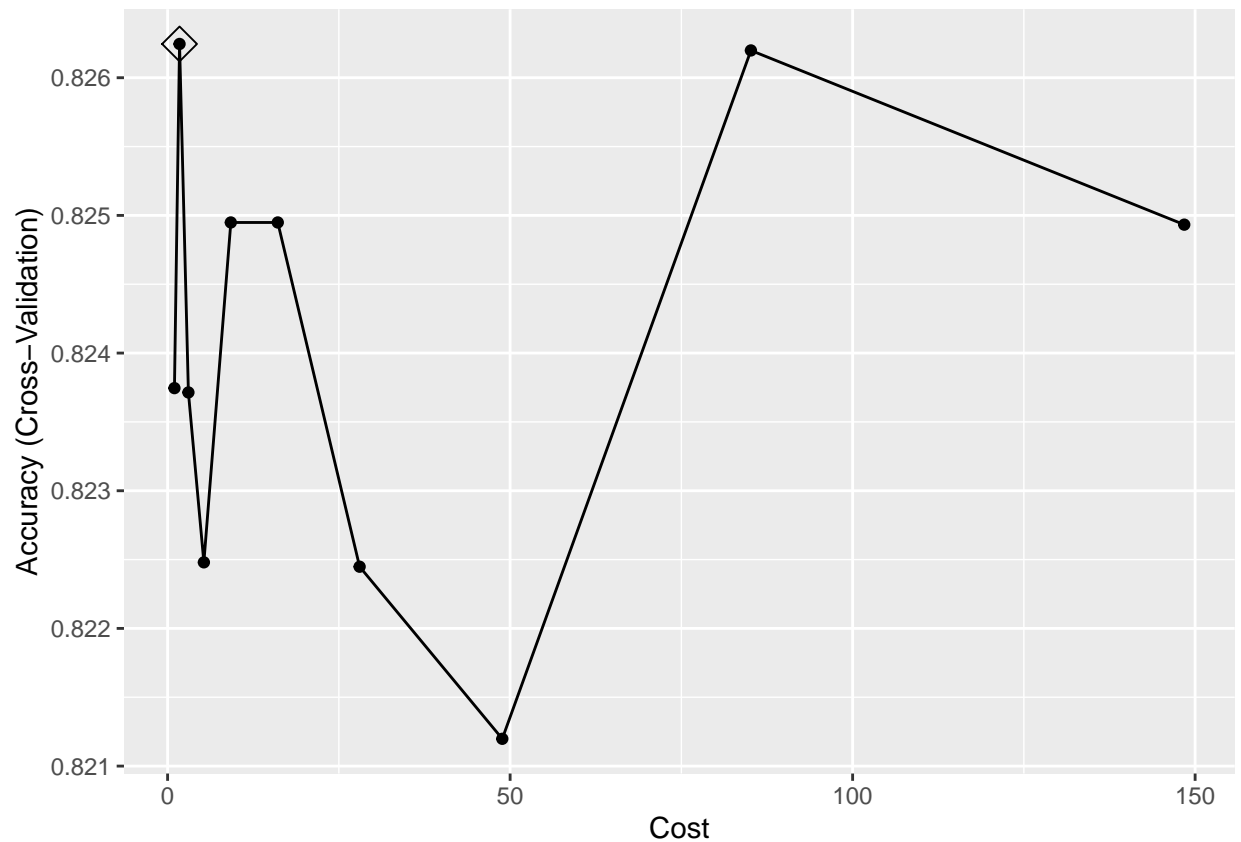
Part B

What are the training and test error rates?

```
svm_radial_grid <- expand.grid(C = exp(seq(0,5,len=10)),
                        sigma = exp(seq(-5,-5,len=5)))
set.seed(1)
svm_radial_fit <- train(purchase ~ ., dat_oj,
                   subset = rowTrain,
                   method = "svmRadial",
                   preProcess = c("center", "scale"),
                   tuneGrid = svm_radial_grid,
                   trControl = control1)
```

```r
ggplot(svm_radial_fit, highlight = TRUE)
```


The training error rate for the support vector machine with a radial kernel is 0.1737547.

Radial Test Data Performance

```r
pred_svm_radial <- predict(svm_radial_fit, newdata = dat_oj[-rowTrain,])


confusion_matrix_svm_radial <- confusionMatrix(data = pred_svm_radial,
              reference = dat_oj$purchase[-rowTrain])
```

The test error rate for the support vector machine with a radial kernel is 0.1518519.