Homework 5

Yanhao Li

Contents

```
library(tidyverse)
library(ISLR)
library(caret)
library(e1071)
```

Load, clean, and tidy data

```
data("OJ")

set.seed(1)

oj = OJ %>%
  janitor::clean_names()

rowTrain <- createDataPartition(y = oj$purchase,
                                p = 799/1070,
                                list = FALSE)

train <- oj[rowTrain, ]

test <- oj[-rowTrain, ]
```

Question a

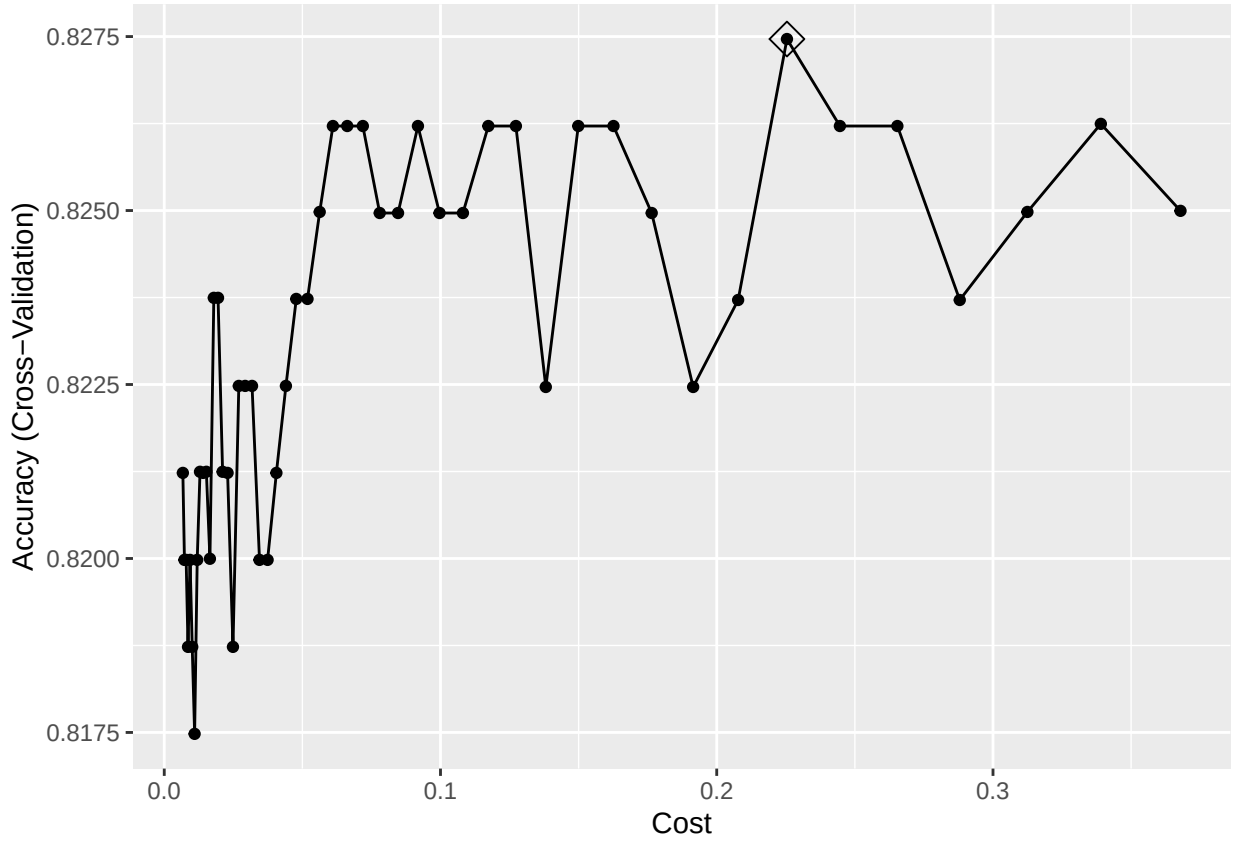
```
ctrl <- trainControl(method = "cv")

set.seed(1)

svml.fit <- train(purchase ~ .,
                  data = train,
                  method = "svmLinear2",
                  preProcess = c("center", "scale"),
                  tuneGrid = data.frame(cost = exp(seq(-5,-1,len = 50))),
                  trControl = ctrl)

svml.fit$finalModel
```

```
##
## Call:
## svm.default(x = as.matrix(x), y = y, kernel = "linear", cost = param$cost,
##     probability = classProbs)
##
##
## Parameters:
##    SVM-Type:  C-classification
##  SVM-Kernel:  linear
##        cost:  0.2254187
##
## Number of Support Vectors:  352
```

```
svml.fit$bestTune
```

```
##         cost
## 44 0.2254187
```

```
ggplot(svml.fit, highlight = TRUE)
```




```
pred_train = predict(svml.fit)

mean(train$purchase != pred_train)
```

```
## [1] 0.17
```

```
pred_test = predict(svml.fit, newdata = test, type = "raw")

mean(test$purchase != pred_test)
```

```
## [1] 0.1518519
```

The training error rate is 0.17.

The test error rate is 0.15.

Question b