

Homework 5

Yanhao Li

Contents

Question a	2
Question b	4

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

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)

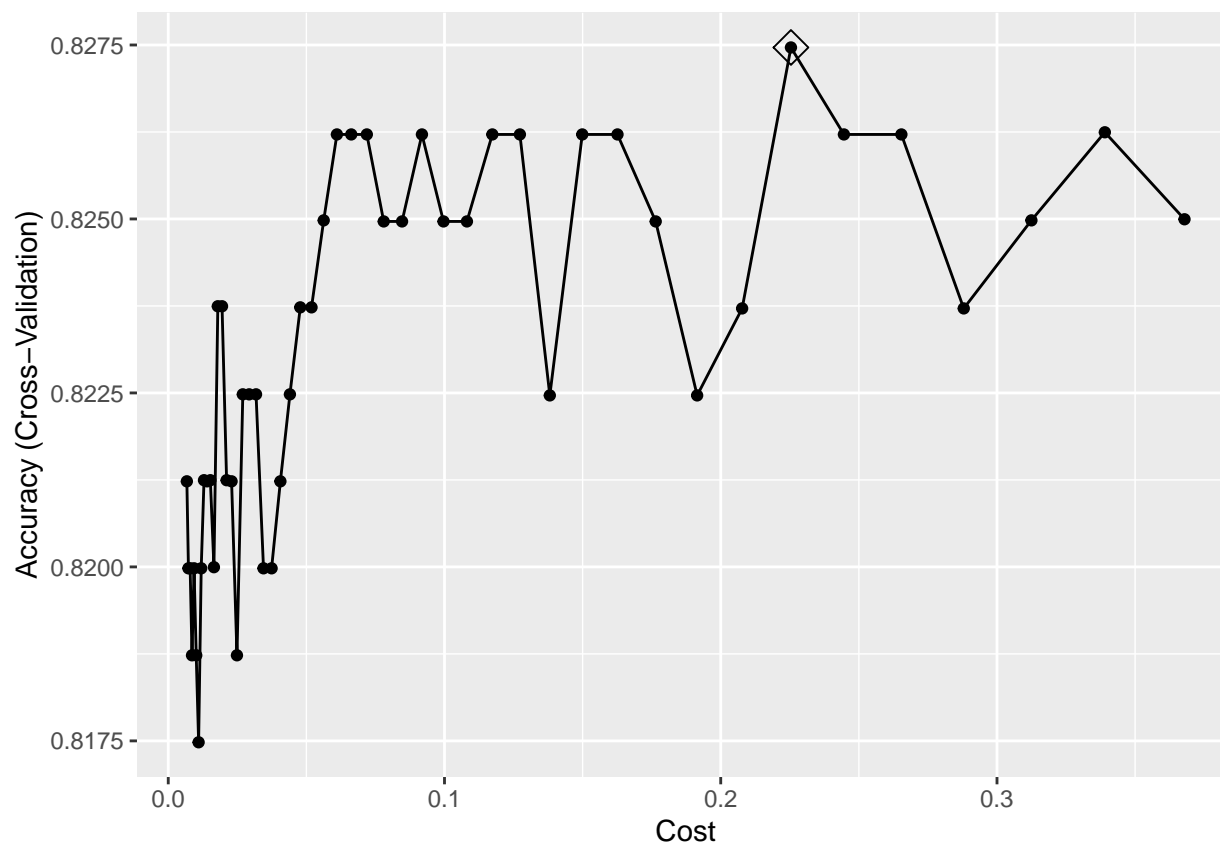
svml_fit$finalModel

##
## 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 0.152.

Question b

```
svmr_grid = expand.grid(C = exp(seq(-5,1,len = 20)),
                        sigma = exp(seq(-5,0,len = 10)))
```

```
set.seed(1)
```

```
svmr_fit <- train(purchase ~ .,
                 data = train,
                 method = "svmRadial",
                 preProcess = c("center", "scale"),
                 tuneGrid = svmr_grid,
                 trControl = ctrl)
```

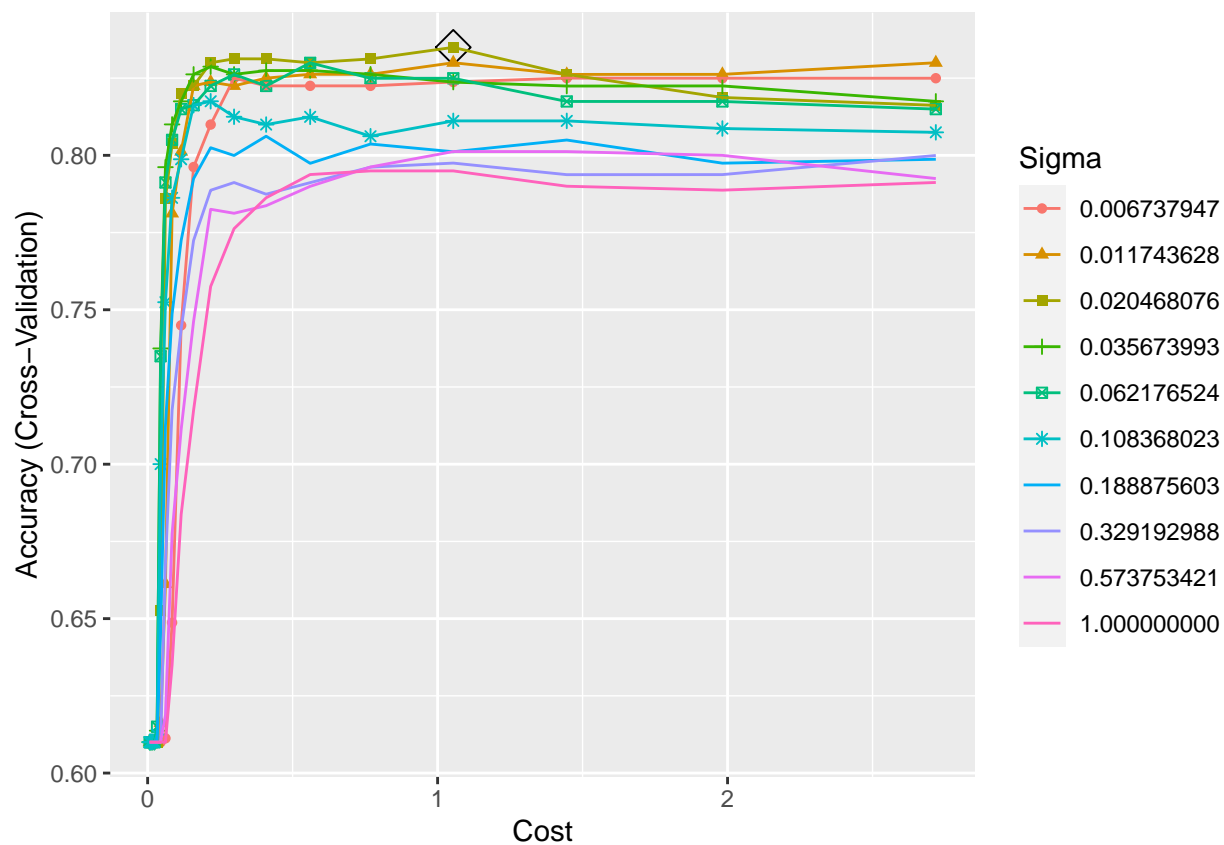
```
svmr_fit$finalModel
```

```
## Support Vector Machine object of class "ksvm"
##
## SV type: C-svc (classification)
## parameter : cost C = 1.0540412425918
##
## Gaussian Radial Basis kernel function.
## Hyperparameter : sigma = 0.0204680757143505
##
## Number of Support Vectors : 389
##
## Objective Function Value : -363.5749
## Training error : 0.1575
```

```
svmr_fit$bestTune
```

```
##          sigma          C
## 163 0.02046808 1.054041
```

```
ggplot(svmr_fit, highlight = TRUE)
```



```
pred_train1 = predict(svmr_fit)
mean(train$purchase != pred_train1)
```

```
## [1] 0.1575
```

```
pred_test1 = predict(svmr_fit, newdata = test, type = "raw")
mean(test$purchase != pred_test1)
```

```
## [1] 0.1555556
```

The training error rate is 0.1575.

The test error rate is 0.1556.