

Homework 5

Xinyi Lin

4/26/2019

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

Question 1

Get training data and test data

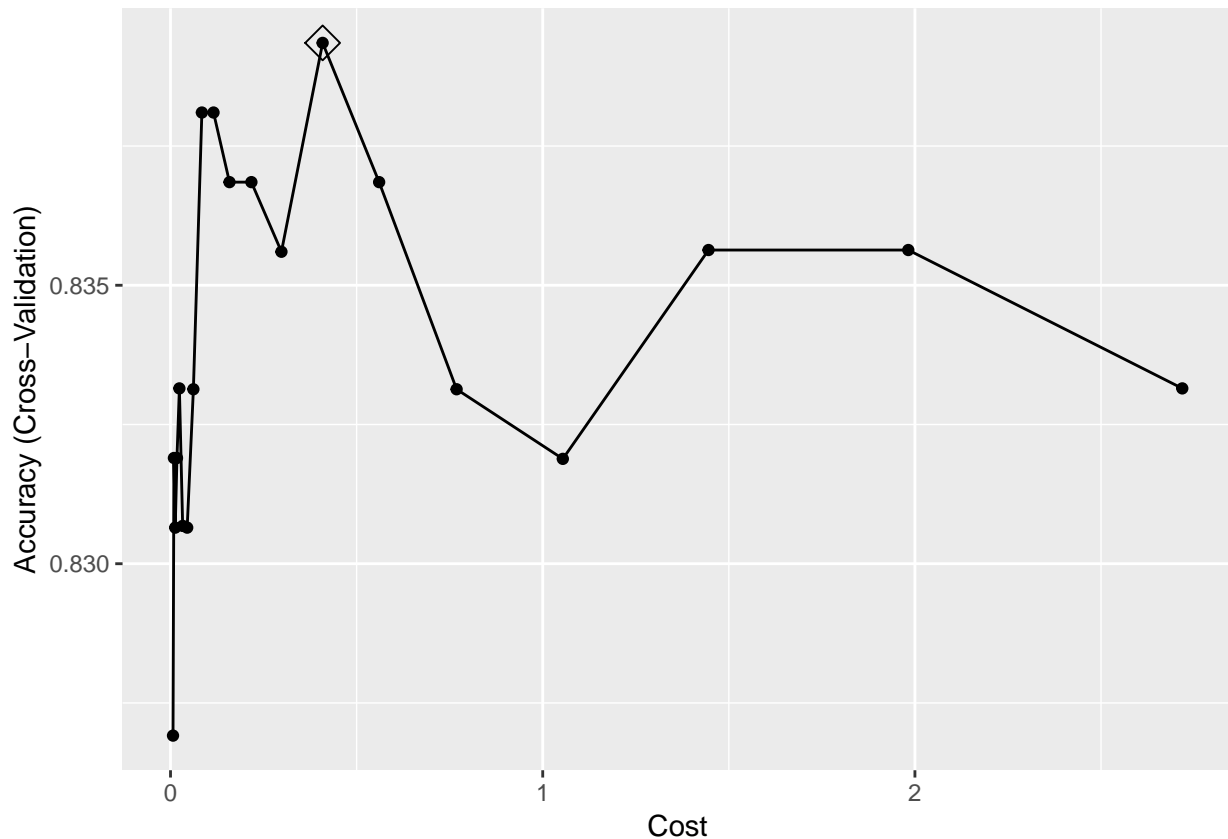
```
#head(OJ)
set.seed(1)
rowTrain <- createDataPartition(y = OJ$Purchase,
                                p = 0.75,
                                list = FALSE)
```

Fit the support vector classifier to the training data.

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

set.seed(1)
svml.fit <- train(Purchase~.,
                  data = OJ[rowTrain,],
                  method = "svmLinear2",
                  preProcess = c("center", "scale"),
                  tuneGrid = data.frame(cost = exp(seq(-5,1,len=20))),
                  trControl = ctrl)

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



Training and test error rates

```
#training error rate
pred.svm1 <- predict(svm1.fit)
mean(OJ[rowTrain,]$Purchase != pred.svm1)
```

```
## [1] 0.1581569
```

```
#test error rate
pred.svm1 <- predict(svm1.fit, newdata = OJ[-rowTrain,])
mean(OJ[-rowTrain,]$Purchase != pred.svm1)
```

```
## [1] 0.164794
```

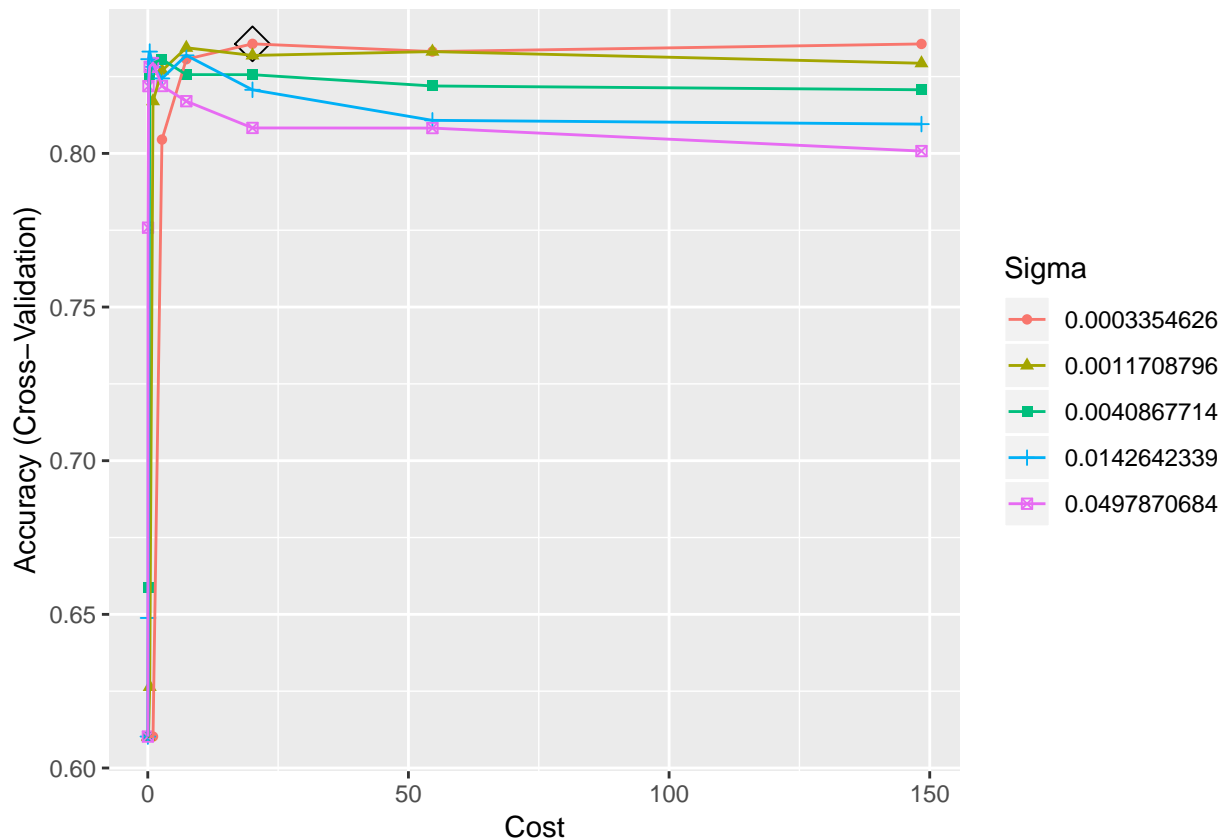
The error rate of training data is 0.158. The error rate of test data is 0.165.

Question 2

Fit the support vector machine model.

```
svmr.grid <- expand.grid(C = exp(seq(-4,5,len=10)),
                        sigma = exp(seq(-8,-3,len=5)))
set.seed(1)
svmr.fit <- train(Purchase~.,
                  data = OJ[rowTrain,],
                  method = "svmRadial",
                  preProcess = c("center", "scale"),
                  tuneGrid = svmr.grid,
                  trControl = ctrl)
```

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



Training and test error rates

```
#training error rate
pred.svmr <- predict(svmr.fit)
mean(OJ[rowTrain,]$Purchase != pred.svmr)
```

```
## [1] 0.1606476
```

```
#test error rate
pred.svmr <- predict(svmr.fit, newdata = OJ[-rowTrain,])
mean(OJ[-rowTrain,]$Purchase != pred.svmr)
```

```
## [1] 0.164794
```

According to results, the error rate of training data is 0.161. The error rate of test data is 0.165.

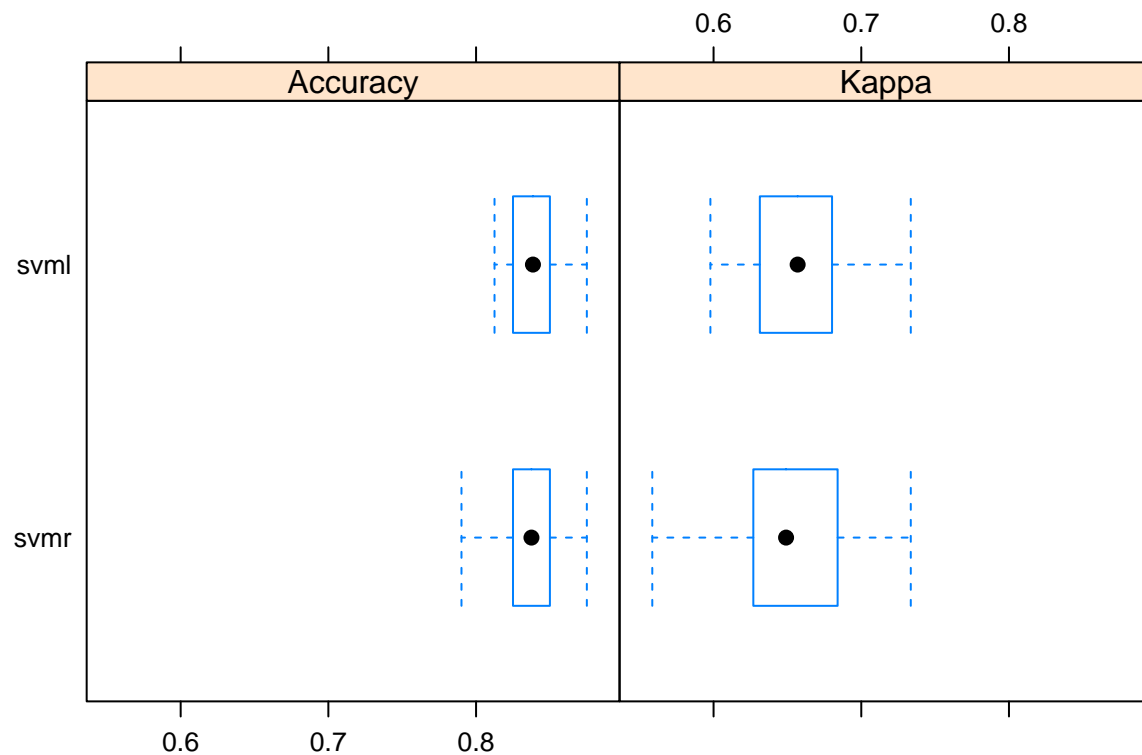
Question 3

```
resamp <- resamples(list(svmr = svmr.fit, svml = svml.fit))
summary(resamp)
```

```
##
## Call:
## summary.resamples(object = resamp)
##
## Models: svmr, svml
```

```
## Number of resamples: 10
##
## Accuracy
##      Min.    1st Qu.    Median      Mean   3rd Qu.    Max. NA's
## svmr 0.7901235 0.8250000 0.8375000 0.8356481 0.8473765 0.875    0
## svm1 0.8125000 0.8255401 0.8385031 0.8393519 0.8500000 0.875    0
##
## Kappa
##      Min.    1st Qu.    Median      Mean   3rd Qu.    Max. NA's
## svmr 0.5585123 0.6269154 0.6490981 0.6492723 0.6769093 0.733511    0
## svm1 0.5978552 0.6321097 0.6569332 0.6577232 0.6800676 0.733511    0
```

`bwplot(resamp)`



According to results, these two models have similar performance. However, the variance of accuracy is smaller and the mean and median are higher in support vector classifier, so support vector classifier performs slightly better.