

housingPriceAna

Chengjia Wang

19 June 2016

Here we use e1701, one of the mostly used reg/clas lib

```
library(e1071)
```

1. load and clean data

Set Paths

```
dataDir <- "/home/cwang/Desktop/amazonInterview/RCode/"  
trStruc.name <- "trainFile.csv"  
teStruc.name <- "testFile.csv"
```

Read data

```
trStruc.Data <- read.csv(paste(dataDir, trStruc.name, sep=""), header = FALSE)  
teStruc.Data <- read.csv(paste(dataDir, teStruc.name, sep=""), header = FALSE)
```

Clean data: sort data based on date, normalize date as number of days

```
colnames(trStruc.Data) <- c("Price", "Date", "Type", "London", "Lease")  
colnames(teStruc.Data) <- c("Price", "Date", "Type", "London", "Lease")  
trStruc.Data <- trStruc.Data[order(trStruc.Data$Date), ]  
teStruc.Data <- teStruc.Data[order(teStruc.Data$Date), ]  
teStruc.Data$Date <- teStruc.Data$Date - min(trStruc.Data$Date)  
trStruc.Data$Date <- trStruc.Data$Date - min(trStruc.Data$Date)
```

Set feature and value

```
trStruc.X = trStruc.Data[, 2:5]  
trStruc.Y = trStruc.Data[, 1]  
teStruc.X = teStruc.Data[, 2:5]  
teStruc.Y = teStruc.Data[, 1]
```

2. fit the model

```
svStruc.model <- svm(trStruc.X, trStruc.Y)
svStruc.trY <- predict(svStruc.model, trStruc.X)
```

calculate the error

```
svStruc.trErrors <- trStruc.Y-svStruc.trY
svStruc.trError <- sqrt(mean(svStruc.trErrors^2))
print (paste('Training RMS error is: ', svStruc.trError, sep = ""))
```

```
## [1] "Training RMS error is: 251758.148286254"
```

3. test the model

```
svStruc.teY = predict(svStruc.model, teStruc.X)
svStruc.teErrors <- teStruc.Y-svStruc.teY
svStruc.teError <- sqrt(mean(svStruc.teErrors^2))
print (paste('Testing RMS error is: ', svStruc.teError, sep = ""))
```

```
## [1] "Testing RMS error is: 264255.007692482"
```

Looks better than in Matlab

4. we can tune the model using R function `tune(svm, ...)`