

STAT542 - Coding Assignment 2

Xiaoming Ji

1. BostonHousing1.Rdata

```
set.seed(6682)

method_names = c("Full", "AIC.F", "AIC.B", "BIC.F", "BIC.B", "R.min", "R.1se",
                  "L.min", "L.1se", "L.Refit")

load("BostonHousing1.Rdata")
HousingData = Housing1

iterations = 50
n = nrow(Housing1)
ntest = round(n*0.25)

#Prepare test ids for all iterations
ALL_TEST_IDS = matrix(0, iterations, ntest)
for(r in 1:iterations){
  ALL_TEST_IDS[r, ] = sample(1:n, ntest)
}

#Make matrix to store MSPE and DF info for each iteration
MSPE_Stat = matrix(0, r, length(method_names))
colnames(MSPE_Stat) = method_names

MODEL_SIZE_Stat = matrix(0, r, length(method_names))
colnames(MODEL_SIZE_Stat) = method_names
```

- Method 1: Full Model

```
eval_full_model = function(data, all_test_ids) {
  r = dim(all_test_ids)[1]
  mspe = rep(0, r)
  model_size = rep(ncol(data) - 1, r)

  for (i in 1:r){
    test_ids = all_test_ids[i,]

    full_model = lm(Y ~ ., data = data[-test_ids, ])
    Ytest_pred = predict(full_model, newdata = data[test_ids,])
    mspe[i] = mean((data[test_ids,]$Y - Ytest_pred)^2)
  }

  list(mspe=mspe, model_size = model_size)
}

start.time = proc.time()
```

```
e = eval_full_model(Housing1, ALL_TEST_IDS)
MSPE_Stat[,1] = e$mspe
MODEL_SIZE_Stat[,1] = e$model_size

proc.time() - start.time
```

```
##    user  system elapsed
##    0.21    0.01    0.22
```

Method 2: Full Model

2.BostonHousing2.Rdata

3.BostonHousing3.Rdata