Advanced Statistical Methods Homework 7

Brandon Hosley

University of Illinois - Springfield

Advanced Statistical Methods Homework 7

Introduction to Statistical Learning
Chapter 8.4 : Problem 12

Apply ~~boosting, bagging,~~ and random forests to a data set of your choice. Be sure to fit the models on a training set and to evaluate their performance on a test set. How accurate are the results compared to simple methods like linear or logistic regression? Which of these approaches yields the best performance?

Prepare the data set:

```
library(MASS)
attach(Boston)
library(randomForest)
set.seed(1234)
```

With the Boston dataset we will train a random forest.

```
train = sample(1:nrow(Boston), nrow(Boston)/2)
test=Boston[-train ,"crim"]
rf.boston= randomForest(crim~.,data=Boston , subset=train,
        mtry=6, importance =TRUE)
yhat.rf = predict(rf.boston ,newdata=Boston[-train ,])
mean((yhat.rf-test)^2)
```

```
Out: 49.66785
```

The Mean-Squared error provided by the Random Forest model is comparable but slightly worse than the MSE given by the Lasso and Ridge Regression models in Module 6.

While the performance is slightly behind those other methods it is still better than simple linear or logistic regression models, as seen in Module 3.