Advanced Statistical Methods Homework 3

Brandon Hosley

University of Illinois - Springfield

Advanced Statistical Methods Homework 3

Introduction to Statistical Learning
Chapter 4.7 : Problem 13

Using the ***Boston*** *data set, fit classification models in order to predict whether a given suburb has a crime rate above or below the median. Explore logistic regression, LDA, and KNN models using various sub-sets of the predictors. Describe your findings.*

Prepare the data set:

```
library(MASS)
attach(Boston)
library(Metrics)

dim(Boston)
cor(Boston[,-14])

summary(crim)
b <- Boston
b$crim = b$crim/max(b$crim)
summary(b$crim)

set.seed(123)
train_ind <- sample(seq_len(nrow(b)), size = floor(0.8 * nrow(b)))
train <- b[train_ind, ]
test <- b[-train_ind, ]
```

(a)

Logistic Regression

```
glm.fits=glm(crim~rad+tax+lstat, data=train, family=binomial)
summary(glm.fits)
```

Train the model on the training data. Then we will test it against the test data.

```
glm.probs=predict(glm.fits, test, type="response")
mse(test$crim,glm.probs)
[1] 0.006966051
```

(b)

LDA

```
glm.fits=glm(crim~rad+tax+lstat, data=b, family=binomial)
summary(glm.fits)
```

(c)

KNN