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)

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

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

(a)

Logistic Regression

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

```
Deviance Residuals:
Min        1Q    Median        3Q       Max
-0.38107  -0.06578  -0.04532   0.00305   2.02780

Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept) -7.546943   2.215077  -3.407 0.000657 ***
rad          0.121906   0.121641   1.002 0.316256
tax          0.002598   0.007140   0.364 0.715925
lstat        0.056465   0.033871   1.667 0.095507 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 62.565  on 505  degrees of freedom
Residual deviance: 16.928  on 502  degrees of freedom
AIC: 69.318

Number of Fisher Scoring iterations: 8
```

(b)

LDA

(c)

KNN