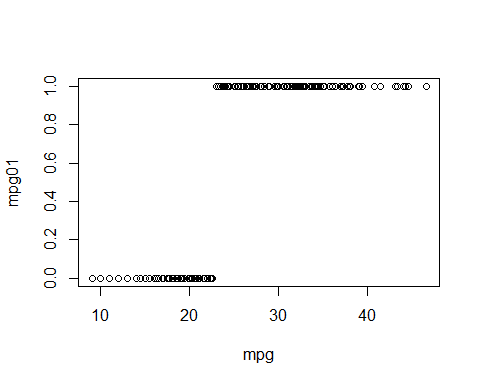
l11ch4.R

7

Sat Apr 11 15:02:03 2015

rm(list=ls())  
library(ISLR)  
attach(Auto)  
#a   
mpg01=ifelse(mpg<median(mpg),0,1)  
plot(mpg,mpg01)



nAuto=data.frame(mpg01,Auto[-1])  
#b  
par(mfrow=c(2,4))  
boxplot(cylinders~mpg01, col=8)  
boxplot(displacement~mpg01, col=2)  
boxplot(horsepower~mpg01, col=3)  
boxplot(weight~mpg01, col=4)  
boxplot(acceleration~mpg01, col=5)  
boxplot(year~mpg01, col=6)  
boxplot(origin~mpg01, col=7)  
#c  
ind=(year<80);  
train=nAuto[ind,]  
test=nAuto[!ind,]  
################  
#d  
library(MASS)  
mod=lda(mpg01~cylinders+displacement+horsepower+weight+year,data = train)  
summary(mod)

## Length Class Mode   
## prior 2 -none- numeric   
## counts 2 -none- numeric   
## means 10 -none- numeric   
## scaling 5 -none- numeric   
## lev 2 -none- character  
## svd 1 -none- numeric   
## N 1 -none- numeric   
## call 3 -none- call   
## terms 3 terms call   
## xlevels 0 -none- list

pred=predict(mod,test,type = "response")  
head(pred$class)

## [1] 1 1 1 1 1 1  
## Levels: 0 1

class\_pred=pred$class  
table(class\_pred,mpg01[!ind])

##   
## class\_pred 0 1  
## 0 4 4  
## 1 1 76

mean(class\_pred!=mpg01[!ind])

## [1] 0.05882353

################  
#e  
mod=qda(mpg01~cylinders+displacement+horsepower+weight+year,data = train)  
summary(mod)

## Length Class Mode   
## prior 2 -none- numeric   
## counts 2 -none- numeric   
## means 10 -none- numeric   
## scaling 50 -none- numeric   
## ldet 2 -none- numeric   
## lev 2 -none- character  
## N 1 -none- numeric   
## call 3 -none- call   
## terms 3 terms call   
## xlevels 0 -none- list

pred=predict(mod,test,type = "response")  
head(pred$class)

## [1] 1 1 1 1 1 1  
## Levels: 0 1

class\_pred=pred$class  
table(class\_pred,mpg01[!ind])

##   
## class\_pred 0 1  
## 0 5 11  
## 1 0 69

mean(class\_pred!=mpg01[!ind])

## [1] 0.1294118

################  
#f  
mod=glm(mpg01~cylinders+displacement+horsepower+weight+year,data = train)  
summary(mod)

##   
## Call:  
## glm(formula = mpg01 ~ cylinders + displacement + horsepower +   
## weight + year, data = train)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.8963 -0.2213 0.1045 0.2156 0.9935   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.024e-01 5.177e-01 1.164 0.24546   
## cylinders -9.744e-02 3.405e-02 -2.862 0.00451 \*\*   
## displacement -6.038e-05 7.310e-04 -0.083 0.93423   
## horsepower 2.155e-03 1.065e-03 2.024 0.04381 \*   
## weight -3.082e-04 5.858e-05 -5.262 2.71e-07 \*\*\*  
## year 1.440e-02 6.841e-03 2.105 0.03614 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for gaussian family taken to be 0.09939563)  
##   
## Null deviance: 72.169 on 306 degrees of freedom  
## Residual deviance: 29.918 on 301 degrees of freedom  
## AIC: 170.41  
##   
## Number of Fisher Scoring iterations: 2

pred=predict(mod,test,type = "response")  
pred[pred<=0.5]=0  
pred[pred>0.5]=1  
table(pred,mpg01[!ind])

##   
## pred 0 1  
## 0 4 5  
## 1 1 75

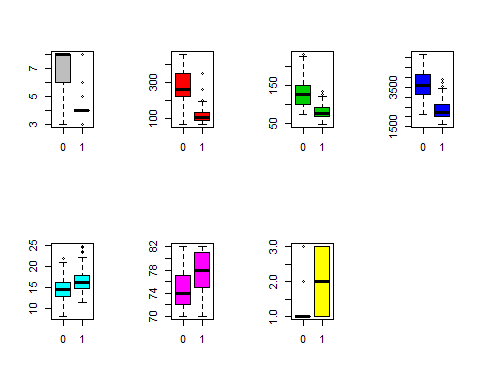
mean(pred!=mpg01[!ind])

## [1] 0.07058824

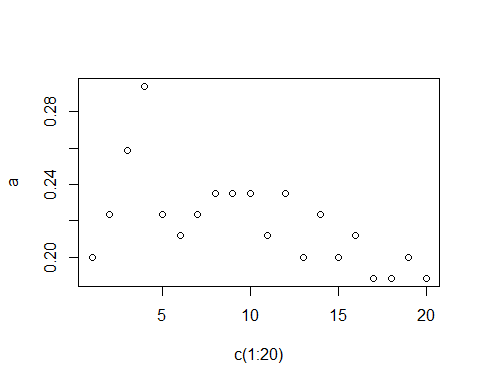
summary(pred)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0000 1.0000 1.0000 0.8941 1.0000 1.0000

################  
#g  
library(class)  
train.Y=mpg01[ind]  
test.Y=mpg01[!ind]  
set.seed(10)  
a={};K=1;  
pred=knn(train[,-9],test[,-9],mpg01[ind],k=1)  
a[1]=mean(test.Y!=pred)  
for (i in 2:20)  
{pred=knn(train[,-9],test[,-9],mpg01[ind],k=i)  
a[i]=mean(test.Y!=pred)  
K=ifelse(a[i]<min(a[-i]),i,K)  
}  
par(mfrow=c(1,1))



plot(c(1:20),a)



K

## [1] 17

pred=knn(train[,-9],test[,-9],mpg01[ind],k=K)  
knn\_er=mean(test.Y!=pred)  
knn\_er

## [1] 0.1882353