l13ch4.R

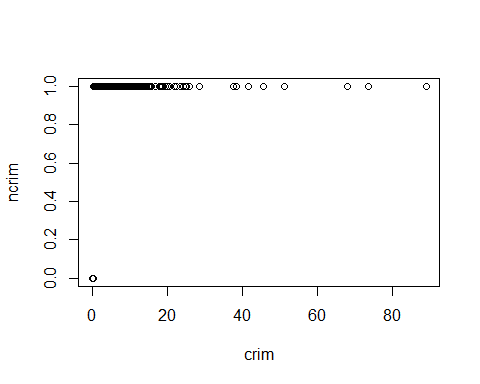
mh

Sat Apr 04 02:34:56 2015

rm(list=ls())  
library(MASS)  
attach(Boston)  
names(Boston)

## [1] "crim" "zn" "indus" "chas" "nox" "rm" "age"   
## [8] "dis" "rad" "tax" "ptratio" "black" "lstat" "medv"

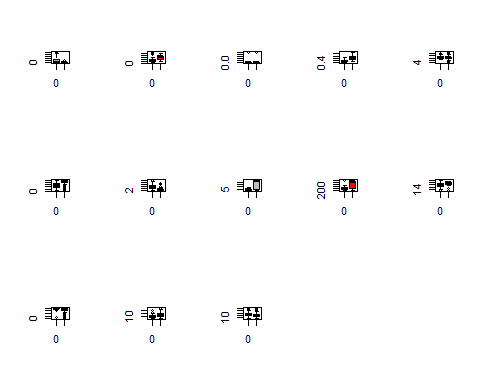
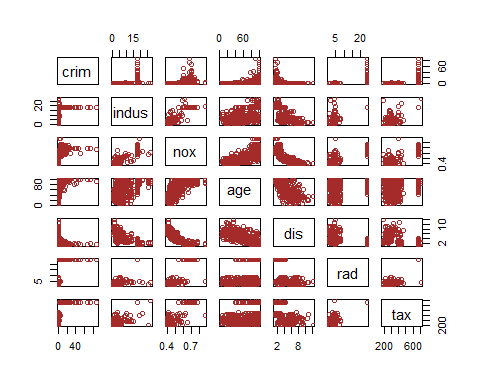
ncrim=ifelse(crim<median(crim),0,1)  
plot(crim,ncrim)



nBoston=data.frame(ncrim,Boston[-1])  
par(mfrow=c(3,5))  
cor(ncrim,medv)

## [1] -0.2630167

boxplot(zn~ncrim, col=8)  
boxplot(indus~ncrim, col=2)  
boxplot(chas~ncrim, col=3)  
boxplot(nox~ncrim, col=4)  
boxplot(rm~ncrim, col=5)  
boxplot(age~ncrim, col=6)  
boxplot(dis~ncrim, col=7)  
boxplot(rad~ncrim, col=8)  
boxplot(tax~ncrim, col=2)  
boxplot(ptratio~ncrim, col=3)  
boxplot(black~ncrim, col=4)  
boxplot(lstat~ncrim, col=5)  
boxplot(medv~ncrim, col=6)  
#indus-nox-age-dis-rad-tax  
pairs(crim~indus-nox-age-dis-rad-tax,col="brown")

ind=(rm<7);  
train=nBoston[ind,]  
test=nBoston[!ind,]  
##################  
#predicting  
mod=lda(ncrim~indus+nox+age+dis+rad+tax,data = train)  
summary(mod)

## Length Class Mode   
## prior 2 -none- numeric   
## counts 2 -none- numeric   
## means 12 -none- numeric   
## scaling 6 -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] 0 0 0 0 0 0  
## Levels: 0 1

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

##   
## class\_pred 0 1  
## 0 35 12  
## 1 0 17

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

## [1] 0.1875

################  
mod=qda(ncrim~indus+nox+age+dis+rad+tax,data = train)  
summary(mod)

## Length Class Mode   
## prior 2 -none- numeric   
## counts 2 -none- numeric   
## means 12 -none- numeric   
## scaling 72 -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] 0 0 0 0 0 0  
## Levels: 0 1

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

##   
## class\_pred 0 1  
## 0 35 6  
## 1 0 23

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

## [1] 0.09375

################  
mod=glm(ncrim~indus+nox+age+dis+rad+tax,data = train)  
summary(mod)

##   
## Call:  
## glm(formula = ncrim ~ indus + nox + age + dis + rad + tax, data = train)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.6155 -0.1806 -0.0509 0.1027 0.8959   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.7068608 0.1585024 -4.460 1.05e-05 \*\*\*  
## indus 0.0051373 0.0043130 1.191 0.23426   
## nox 1.6035450 0.2486188 6.450 2.99e-10 \*\*\*  
## age 0.0028177 0.0008612 3.272 0.00115 \*\*   
## dis -0.0063852 0.0130372 -0.490 0.62455   
## rad 0.0215304 0.0042866 5.023 7.45e-07 \*\*\*  
## tax -0.0003269 0.0002565 -1.275 0.20308   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for gaussian family taken to be 0.1014715)  
##   
## Null deviance: 110.48 on 441 degrees of freedom  
## Residual deviance: 44.14 on 435 degrees of freedom  
## AIC: 252  
##   
## Number of Fisher Scoring iterations: 2

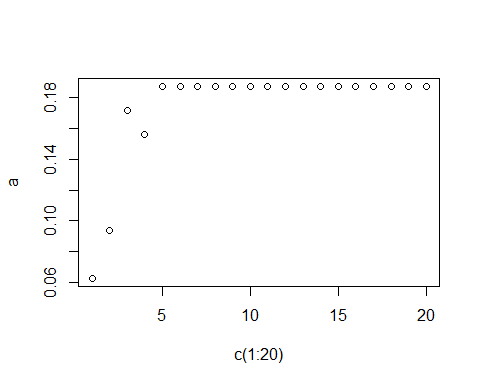
pred=predict(mod,test,type = "response")  
pred[pred<=0.5]=0  
pred[pred>0.5]=1  
table(class\_pred,ncrim[!ind])

##   
## class\_pred 0 1  
## 0 35 6  
## 1 0 23

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

## [1] 0.09375

################  
library(class)  
train.Y=ncrim[ind]  
test.Y=ncrim[!ind]  
set.seed(10)  
a={};K=1;  
pred=knn(train[,c(3,5,7,9,10)],test[,c(3,5,7,9,10)],ncrim[ind],k=1)  
a[1]=mean(test.Y!=pred)  
for (i in 2:20)  
{pred=knn(train[,c(3,5,7,9,10)],test[,c(3,5,7,9,10)],ncrim[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] 1

pred=knn(train[,c(3,5,7,9,10)],test[,c(3,5,7,9,10)],ncrim[ind],k=K)  
knn\_er=mean(test.Y!=pred)  
knn\_er

## [1] 0.0625