Ch 4 Lab

Conner McCloney

The model below can be written as = + Mean.of.the.integrated.profile + SD.of.the.integrated.profile + Excess.kurtosis.of.the.integrated.profile + Skewness.of.the.integrated.profile + Mean.of.the.DM.SNR.curve + SD.of.the.DM.SNR.curve + Excess.kurtosis.of.the.DM.SNR.curve + Skewness.of.the.DM.SNR.curve

pulsar\_data <- read.csv("pulsar\_stars.csv")  
names(pulsar\_data)

## [1] "Mean.of.the.integrated.profile"   
## [2] "Standard.deviation.of.the.integrated.profile"  
## [3] "Excess.kurtosis.of.the.integrated.profile"   
## [4] "Skewness.of.the.integrated.profile"   
## [5] "Mean.of.the.DM.SNR.curve"   
## [6] "Standard.deviation.of.the.DM.SNR.curve"   
## [7] "Excess.kurtosis.of.the.DM.SNR.curve"   
## [8] "Skewness.of.the.DM.SNR.curve"   
## [9] "target\_class"

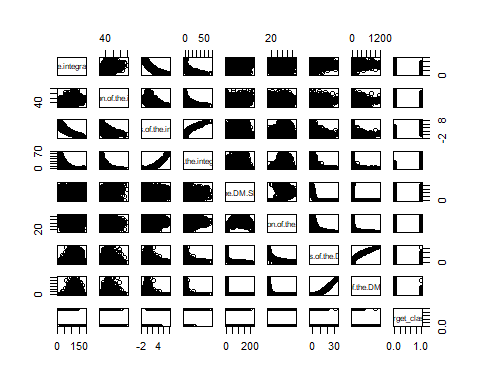
dim(pulsar\_data)

## [1] 17898 9

summary(pulsar\_data)

## Mean.of.the.integrated.profile  
## Min. : 5.812   
## 1st Qu.:100.930   
## Median :115.078   
## Mean :111.080   
## 3rd Qu.:127.086   
## Max. :192.617   
## Standard.deviation.of.the.integrated.profile  
## Min. :24.77   
## 1st Qu.:42.38   
## Median :46.95   
## Mean :46.55   
## 3rd Qu.:51.02   
## Max. :98.78   
## Excess.kurtosis.of.the.integrated.profile  
## Min. :-1.8760   
## 1st Qu.: 0.0271   
## Median : 0.2232   
## Mean : 0.4779   
## 3rd Qu.: 0.4733   
## Max. : 8.0695   
## Skewness.of.the.integrated.profile Mean.of.the.DM.SNR.curve  
## Min. :-1.7919 Min. : 0.2132   
## 1st Qu.:-0.1886 1st Qu.: 1.9231   
## Median : 0.1987 Median : 2.8018   
## Mean : 1.7703 Mean : 12.6144   
## 3rd Qu.: 0.9278 3rd Qu.: 5.4643   
## Max. :68.1016 Max. :223.3921   
## Standard.deviation.of.the.DM.SNR.curve  
## Min. : 7.37   
## 1st Qu.: 14.44   
## Median : 18.46   
## Mean : 26.33   
## 3rd Qu.: 28.43   
## Max. :110.64   
## Excess.kurtosis.of.the.DM.SNR.curve Skewness.of.the.DM.SNR.curve  
## Min. :-3.139 Min. : -1.977   
## 1st Qu.: 5.782 1st Qu.: 34.961   
## Median : 8.434 Median : 83.065   
## Mean : 8.304 Mean : 104.858   
## 3rd Qu.:10.703 3rd Qu.: 139.309   
## Max. :34.540 Max. :1191.001   
## target\_class   
## Min. :0.00000   
## 1st Qu.:0.00000   
## Median :0.00000   
## Mean :0.09157   
## 3rd Qu.:0.00000   
## Max. :1.00000

pairs(pulsar\_data)



cor(pulsar\_data)

## Mean.of.the.integrated.profile  
## Mean.of.the.integrated.profile 1.0000000  
## Standard.deviation.of.the.integrated.profile 0.5471369  
## Excess.kurtosis.of.the.integrated.profile -0.8738984  
## Skewness.of.the.integrated.profile -0.7387748  
## Mean.of.the.DM.SNR.curve -0.2988408  
## Standard.deviation.of.the.DM.SNR.curve -0.3070158  
## Excess.kurtosis.of.the.DM.SNR.curve 0.2343312  
## Skewness.of.the.DM.SNR.curve 0.1440330  
## target\_class -0.6731806  
## Standard.deviation.of.the.integrated.profile  
## Mean.of.the.integrated.profile 0.547136926  
## Standard.deviation.of.the.integrated.profile 1.000000000  
## Excess.kurtosis.of.the.integrated.profile -0.521435272  
## Skewness.of.the.integrated.profile -0.539792970  
## Mean.of.the.DM.SNR.curve 0.006868735  
## Standard.deviation.of.the.DM.SNR.curve -0.047631587  
## Excess.kurtosis.of.the.DM.SNR.curve 0.029429387  
## Skewness.of.the.DM.SNR.curve 0.027691480  
## target\_class -0.363708209  
## Excess.kurtosis.of.the.integrated.profile  
## Mean.of.the.integrated.profile -0.8738984  
## Standard.deviation.of.the.integrated.profile -0.5214353  
## Excess.kurtosis.of.the.integrated.profile 1.0000000  
## Skewness.of.the.integrated.profile 0.9457291  
## Mean.of.the.DM.SNR.curve 0.4143676  
## Standard.deviation.of.the.DM.SNR.curve 0.4328802  
## Excess.kurtosis.of.the.DM.SNR.curve -0.3412090  
## Skewness.of.the.DM.SNR.curve -0.2144909  
## target\_class 0.7915914  
## Skewness.of.the.integrated.profile  
## Mean.of.the.integrated.profile -0.7387748  
## Standard.deviation.of.the.integrated.profile -0.5397930  
## Excess.kurtosis.of.the.integrated.profile 0.9457291  
## Skewness.of.the.integrated.profile 1.0000000  
## Mean.of.the.DM.SNR.curve 0.4120564  
## Standard.deviation.of.the.DM.SNR.curve 0.4151400  
## Excess.kurtosis.of.the.DM.SNR.curve -0.3288433  
## Skewness.of.the.DM.SNR.curve -0.2047825  
## target\_class 0.7095280  
## Mean.of.the.DM.SNR.curve  
## Mean.of.the.integrated.profile -0.298840844  
## Standard.deviation.of.the.integrated.profile 0.006868735  
## Excess.kurtosis.of.the.integrated.profile 0.414367611  
## Skewness.of.the.integrated.profile 0.412056437  
## Mean.of.the.DM.SNR.curve 1.000000000  
## Standard.deviation.of.the.DM.SNR.curve 0.796554844  
## Excess.kurtosis.of.the.DM.SNR.curve -0.615970831  
## Skewness.of.the.DM.SNR.curve -0.354269152  
## target\_class 0.400876093  
## Standard.deviation.of.the.DM.SNR.curve  
## Mean.of.the.integrated.profile -0.30701583  
## Standard.deviation.of.the.integrated.profile -0.04763159  
## Excess.kurtosis.of.the.integrated.profile 0.43288016  
## Skewness.of.the.integrated.profile 0.41513996  
## Mean.of.the.DM.SNR.curve 0.79655484  
## Standard.deviation.of.the.DM.SNR.curve 1.00000000  
## Excess.kurtosis.of.the.DM.SNR.curve -0.80978582  
## Skewness.of.the.DM.SNR.curve -0.57579983  
## target\_class 0.49153506  
## Excess.kurtosis.of.the.DM.SNR.curve  
## Mean.of.the.integrated.profile 0.23433123  
## Standard.deviation.of.the.integrated.profile 0.02942939  
## Excess.kurtosis.of.the.integrated.profile -0.34120902  
## Skewness.of.the.integrated.profile -0.32884331  
## Mean.of.the.DM.SNR.curve -0.61597083  
## Standard.deviation.of.the.DM.SNR.curve -0.80978582  
## Excess.kurtosis.of.the.DM.SNR.curve 1.00000000  
## Skewness.of.the.DM.SNR.curve 0.92374273  
## target\_class -0.39081632  
## Skewness.of.the.DM.SNR.curve  
## Mean.of.the.integrated.profile 0.14403302  
## Standard.deviation.of.the.integrated.profile 0.02769148  
## Excess.kurtosis.of.the.integrated.profile -0.21449091  
## Skewness.of.the.integrated.profile -0.20478247  
## Mean.of.the.DM.SNR.curve -0.35426915  
## Standard.deviation.of.the.DM.SNR.curve -0.57579983  
## Excess.kurtosis.of.the.DM.SNR.curve 0.92374273  
## Skewness.of.the.DM.SNR.curve 1.00000000  
## target\_class -0.25911670  
## target\_class  
## Mean.of.the.integrated.profile -0.6731806  
## Standard.deviation.of.the.integrated.profile -0.3637082  
## Excess.kurtosis.of.the.integrated.profile 0.7915914  
## Skewness.of.the.integrated.profile 0.7095280  
## Mean.of.the.DM.SNR.curve 0.4008761  
## Standard.deviation.of.the.DM.SNR.curve 0.4915351  
## Excess.kurtosis.of.the.DM.SNR.curve -0.3908163  
## Skewness.of.the.DM.SNR.curve -0.2591167  
## target\_class 1.0000000

glm.fits=glm(target\_class∼Mean.of.the.integrated.profile+Standard.deviation.of.the.integrated.profile+  
 Excess.kurtosis.of.the.integrated.profile+Skewness.of.the.integrated.profile+  
 Mean.of.the.DM.SNR.curve+Standard.deviation.of.the.DM.SNR.curve+  
 Excess.kurtosis.of.the.DM.SNR.curve+Skewness.of.the.DM.SNR.curve,data=pulsar\_data ,family =binomial )  
summary(glm.fits)

##   
## Call:  
## glm(formula = target\_class ~ Mean.of.the.integrated.profile +   
## Standard.deviation.of.the.integrated.profile + Excess.kurtosis.of.the.integrated.profile +   
## Skewness.of.the.integrated.profile + Mean.of.the.DM.SNR.curve +   
## Standard.deviation.of.the.DM.SNR.curve + Excess.kurtosis.of.the.DM.SNR.curve +   
## Skewness.of.the.DM.SNR.curve, family = binomial, data = pulsar\_data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -4.3813 -0.1644 -0.1000 -0.0565 3.6178   
##   
## Coefficients:  
## Estimate Std. Error z value  
## (Intercept) -9.019954 0.977017 -9.232  
## Mean.of.the.integrated.profile 0.030260 0.005925 5.107  
## Standard.deviation.of.the.integrated.profile -0.035430 0.010386 -3.411  
## Excess.kurtosis.of.the.integrated.profile 6.577063 0.300110 21.916  
## Skewness.of.the.integrated.profile -0.616243 0.039709 -15.519  
## Mean.of.the.DM.SNR.curve -0.028585 0.003268 -8.747  
## Standard.deviation.of.the.DM.SNR.curve 0.053166 0.007364 7.220  
## Excess.kurtosis.of.the.DM.SNR.curve 0.047749 0.085758 0.557  
## Skewness.of.the.DM.SNR.curve -0.004750 0.003051 -1.557  
## Pr(>|z|)   
## (Intercept) < 2e-16 \*\*\*  
## Mean.of.the.integrated.profile 3.27e-07 \*\*\*  
## Standard.deviation.of.the.integrated.profile 0.000647 \*\*\*  
## Excess.kurtosis.of.the.integrated.profile < 2e-16 \*\*\*  
## Skewness.of.the.integrated.profile < 2e-16 \*\*\*  
## Mean.of.the.DM.SNR.curve < 2e-16 \*\*\*  
## Standard.deviation.of.the.DM.SNR.curve 5.21e-13 \*\*\*  
## Excess.kurtosis.of.the.DM.SNR.curve 0.577676   
## Skewness.of.the.DM.SNR.curve 0.119474   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 10959.5 on 17897 degrees of freedom  
## Residual deviance: 2615.8 on 17889 degrees of freedom  
## AIC: 2633.8  
##   
## Number of Fisher Scoring iterations: 8

coef(glm.fits)

## (Intercept)   
## -9.019954073   
## Mean.of.the.integrated.profile   
## 0.030259971   
## Standard.deviation.of.the.integrated.profile   
## -0.035430322   
## Excess.kurtosis.of.the.integrated.profile   
## 6.577062820   
## Skewness.of.the.integrated.profile   
## -0.616243427   
## Mean.of.the.DM.SNR.curve   
## -0.028584556   
## Standard.deviation.of.the.DM.SNR.curve   
## 0.053165747   
## Excess.kurtosis.of.the.DM.SNR.curve   
## 0.047748688   
## Skewness.of.the.DM.SNR.curve   
## -0.004749673

#summary(glm.fits)$coef  
glm.probs=predict(glm.fits,type="response")  
glm.probs[1:10]

## 1 2 3 4 5   
## 0.0010084627 0.0182637360 0.0091788515 0.0027850119 0.0114443787   
## 6 7 8 9 10   
## 0.0181306352 0.0004293102 0.0006531135 0.0155766130 0.0283063768

pulsar\_data$target\_class <- factor(pulsar\_data$target\_class)  
contrasts(pulsar\_data$target\_class)

## 1  
## 0 0  
## 1 1

glm.pred=rep("0",17898)  
glm.pred[glm.probs>.5]="1"  
table(glm.pred,pulsar\_data$target\_class)

##   
## glm.pred 0 1  
## 0 16173 282  
## 1 86 1357

(16173+1357)/17898

## [1] 0.979439

mean(glm.pred==pulsar\_data$target\_class)

## [1] 0.979439

train=(pulsar\_data[0:1250,])  
test=pulsar\_data[1251:17898,]  
dim(test)

## [1] 16648 9

glm.fits=glm.fits=glm(target\_class∼Mean.of.the.integrated.profile+Standard.deviation.of.the.integrated.profile+  
 Excess.kurtosis.of.the.integrated.profile+Skewness.of.the.integrated.profile+  
 Mean.of.the.DM.SNR.curve+Standard.deviation.of.the.DM.SNR.curve+  
 Excess.kurtosis.of.the.DM.SNR.curve+Skewness.of.the.DM.SNR.curve,data=train ,family =binomial )  
glm.probs=predict(glm.fits,test,type="response")  
glm.pred=rep("0",16648)  
glm.pred[glm.probs>.5]="1"  
table(glm.pred,test$target\_class)

##   
## glm.pred 0 1  
## 0 15008 274  
## 1 85 1281

(15008+1281)/16648

## [1] 0.9784358

mean(glm.pred==test$target\_class)

## [1] 0.9784358

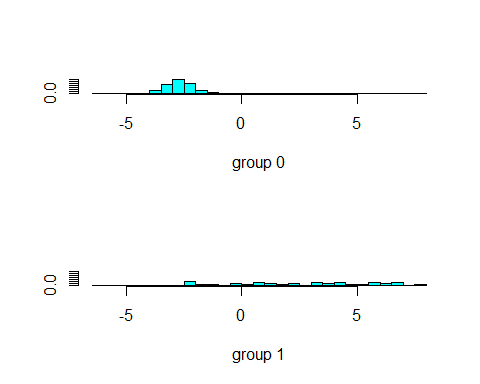
mean(glm.pred!=test$target\_class)

## [1] 0.02156415

#LDA  
library(MASS)  
lda.fit=lda(target\_class∼Mean.of.the.integrated.profile+Standard.deviation.of.the.integrated.profile+  
 Excess.kurtosis.of.the.integrated.profile+Skewness.of.the.integrated.profile+  
 Mean.of.the.DM.SNR.curve+Standard.deviation.of.the.DM.SNR.curve+  
 Excess.kurtosis.of.the.DM.SNR.curve+Skewness.of.the.DM.SNR.curve,data=train)  
lda.fit

## Call:  
## lda(target\_class ~ Mean.of.the.integrated.profile + Standard.deviation.of.the.integrated.profile +   
## Excess.kurtosis.of.the.integrated.profile + Skewness.of.the.integrated.profile +   
## Mean.of.the.DM.SNR.curve + Standard.deviation.of.the.DM.SNR.curve +   
## Excess.kurtosis.of.the.DM.SNR.curve + Skewness.of.the.DM.SNR.curve,   
## data = train)  
##   
## Prior probabilities of groups:  
## 0 1   
## 0.9328 0.0672   
##   
## Group means:  
## Mean.of.the.integrated.profile  
## 0 116.55018  
## 1 57.80887  
## Standard.deviation.of.the.integrated.profile  
## 0 47.51993  
## 1 38.56397  
## Excess.kurtosis.of.the.integrated.profile  
## 0 0.2045606  
## 1 3.0722386  
## Skewness.of.the.integrated.profile Mean.of.the.DM.SNR.curve  
## 0 0.3691994 6.076374  
## 1 15.2117303 45.619854  
## Standard.deviation.of.the.DM.SNR.curve  
## 0 21.78375  
## 1 53.87673  
## Excess.kurtosis.of.the.DM.SNR.curve Skewness.of.the.DM.SNR.curve  
## 0 9.00258 114.14141  
## 1 3.20215 22.97744  
##   
## Coefficients of linear discriminants:  
## LD1  
## Mean.of.the.integrated.profile 0.0266321444  
## Standard.deviation.of.the.integrated.profile -0.0165032466  
## Excess.kurtosis.of.the.integrated.profile 3.1621169508  
## Skewness.of.the.integrated.profile -0.1929183797  
## Mean.of.the.DM.SNR.curve -0.0061082560  
## Standard.deviation.of.the.DM.SNR.curve 0.0316659514  
## Excess.kurtosis.of.the.DM.SNR.curve 0.0249154823  
## Skewness.of.the.DM.SNR.curve -0.0001986053

plot(lda.fit)



lda.pred=predict(lda.fit,test)  
names(lda.pred)

## [1] "class" "posterior" "x"

lda.class=lda.pred$class  
table(lda.class,test$target\_class)

##   
## lda.class 0 1  
## 0 15030 366  
## 1 63 1189

mean(lda.class==test$target\_class)

## [1] 0.9742311

sum(lda.pred$posterior[,1]>=.9)

## [1] 15303

sum(lda.pred$posterior[,1]<.9)

## [1] 1345

#QDA  
qda.fit=qda(target\_class~Mean.of.the.integrated.profile+Standard.deviation.of.the.integrated.profile+  
 Excess.kurtosis.of.the.integrated.profile+Skewness.of.the.integrated.profile+  
 Mean.of.the.DM.SNR.curve+Standard.deviation.of.the.DM.SNR.curve+  
 Excess.kurtosis.of.the.DM.SNR.curve+Skewness.of.the.DM.SNR.curve,data=train)  
qda.fit

## Call:  
## qda(target\_class ~ Mean.of.the.integrated.profile + Standard.deviation.of.the.integrated.profile +   
## Excess.kurtosis.of.the.integrated.profile + Skewness.of.the.integrated.profile +   
## Mean.of.the.DM.SNR.curve + Standard.deviation.of.the.DM.SNR.curve +   
## Excess.kurtosis.of.the.DM.SNR.curve + Skewness.of.the.DM.SNR.curve,   
## data = train)  
##   
## Prior probabilities of groups:  
## 0 1   
## 0.9328 0.0672   
##   
## Group means:  
## Mean.of.the.integrated.profile  
## 0 116.55018  
## 1 57.80887  
## Standard.deviation.of.the.integrated.profile  
## 0 47.51993  
## 1 38.56397  
## Excess.kurtosis.of.the.integrated.profile  
## 0 0.2045606  
## 1 3.0722386  
## Skewness.of.the.integrated.profile Mean.of.the.DM.SNR.curve  
## 0 0.3691994 6.076374  
## 1 15.2117303 45.619854  
## Standard.deviation.of.the.DM.SNR.curve  
## 0 21.78375  
## 1 53.87673  
## Excess.kurtosis.of.the.DM.SNR.curve Skewness.of.the.DM.SNR.curve  
## 0 9.00258 114.14141  
## 1 3.20215 22.97744

qda.class=predict(qda.fit,test)$class  
table(qda.class,test$target\_class)

##   
## qda.class 0 1  
## 0 14497 212  
## 1 596 1343

mean(qda.class==test$target\_class)

## [1] 0.9514656

library(class)  
knn.pred=knn(train,test,train$target\_class,k=3)  
table(knn.pred,test$target\_class)

##   
## knn.pred 0 1  
## 0 14966 425  
## 1 127 1130