banknote_authentication

July 27, 2021

[24]:
```
set.seed(72007)
```

[25]:
```
sdata<-read.csv("BankNote_Authentication.csv")
```

[26]:
```
index=sample(1:nrow(sdata),nrow(sdata)*0.96)
my_data<-sdata[index,]
```

[27]:
```
head(my_data)
```

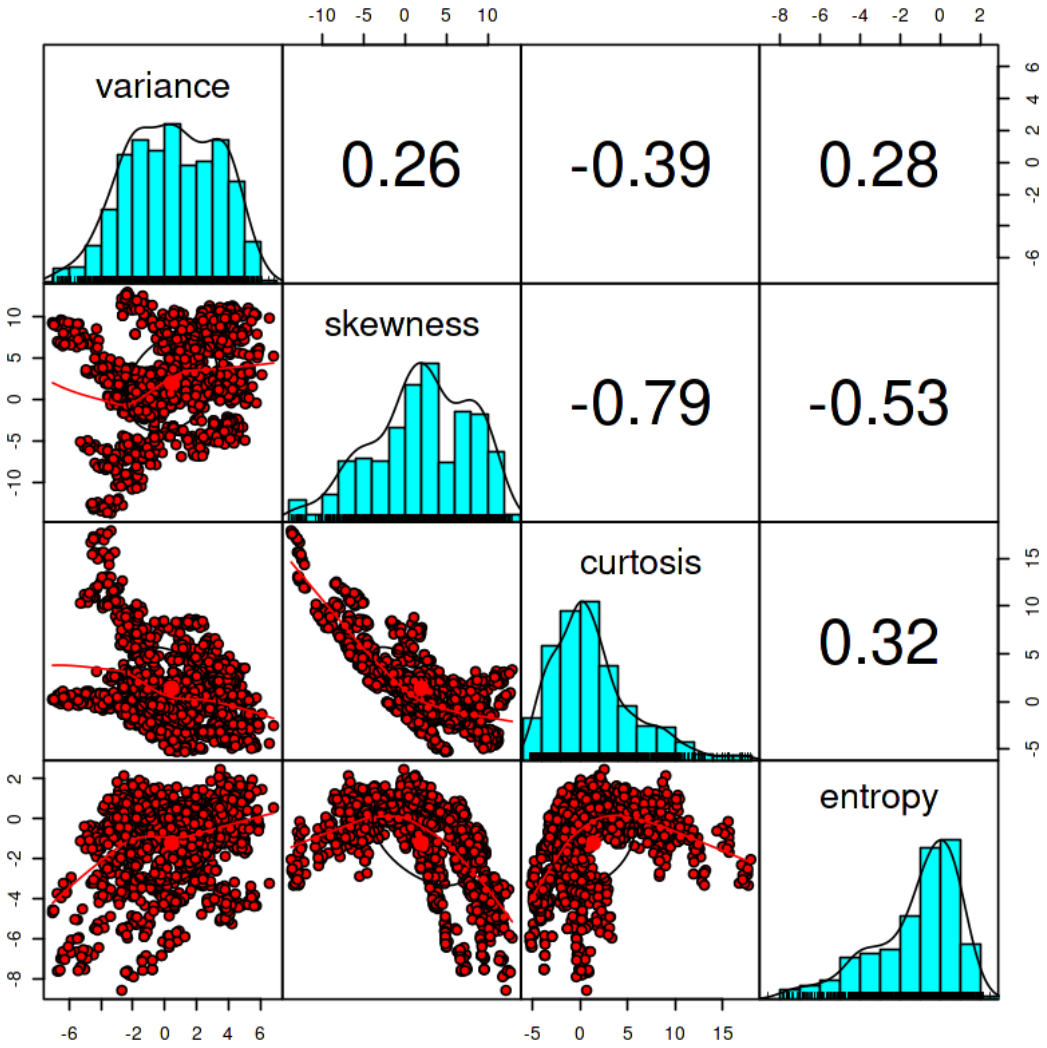
[27]:

A data.frame: 6 × 5

	variance <dbl>	skewness <dbl>	curtosis <dbl>	entropy <dbl>	class <int>
779	-0.77461	-1.8768	2.40230	1.131900	1
650	-0.38388	-1.0471	8.05140	0.495670	0
597	0.65497	5.1815	1.06730	-0.421130	0
762	2.25040	3.5757	0.35273	0.283600	0
1145	-1.39950	-1.9162	2.51540	0.599120	1
480	2.09620	2.4769	1.93790	-0.040962	0

[28]:
```
#constructing scatter plots
library(psych)
pairs.panels(my_data[1:4],
             gap=0,
             bg=c("red","green","blue")[my_data$class],
             pch=21
             )
```

[28]:



[29]:
```
#Data partition with 75% training data
ind<-sample(2,nrow(my_data),
            replace=TRUE,
            prob=c(0.75,0.25))
```

[30]:
```
training<-my_data[ind==1,]
testing<-my_data[ind==2,]
```

[31]:
```
str(training)
```

```
'data.frame':   973 obs. of  5 variables:
 $ variance: num  -0.775 -0.384 2.25 -1.399 2.096 …
 $ skewness: num  -1.88 -1.05 3.58 -1.92 2.48 …
```

```
 $ curtosis: num  2.402 8.051 0.353 2.515 1.938 …
 $ entropy : num  1.132 0.496 0.284 0.599 -0.041 …
 $ class   : int  1 0 0 1 0 1 0 0 0 0 …
```

[32]:
```
str(testing)
```

```
'data.frame':   344 obs. of  5 variables:
 $ variance: num  0.655 -2.076 1.936 -2.567 -1.728 …
 $ skewness: num  5.18 10.82 8.17 -6.88 -6.84 …
 $ curtosis: num  1.0673 2.6439 -0.0234 7.5416 8.9494 …
 $ entropy : num  -0.421 -4.837 -2.259 0.708 0.681 …
 $ class   : int  0 0 0 1 1 1 1 1 0 1 …
```

[33]:
```
#Linear Discriminant Analysis
library(MASS)
linear<-lda(class~.,training)
linear
```

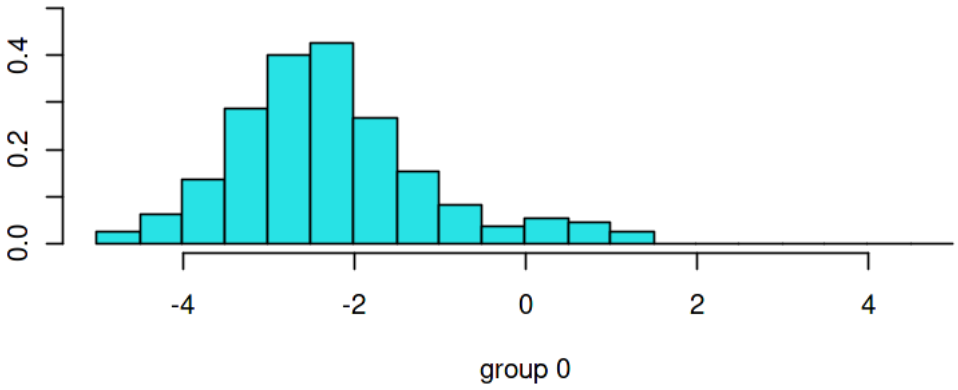
[33]:
```
Call:
lda(class ~ ., data = training)

Prior probabilities of groups:
        0         1
0.5539568 0.4460432

Group means:
   variance  skewness  curtosis   entropy
0  2.240876  4.346292 0.7773169 -1.161729
1 -1.831901 -1.067148 2.1648502 -1.185065

Coefficients of linear discriminants:
                  LD1
variance -0.837747683
skewness -0.466182825
curtosis -0.602815300
entropy   0.007301617
```

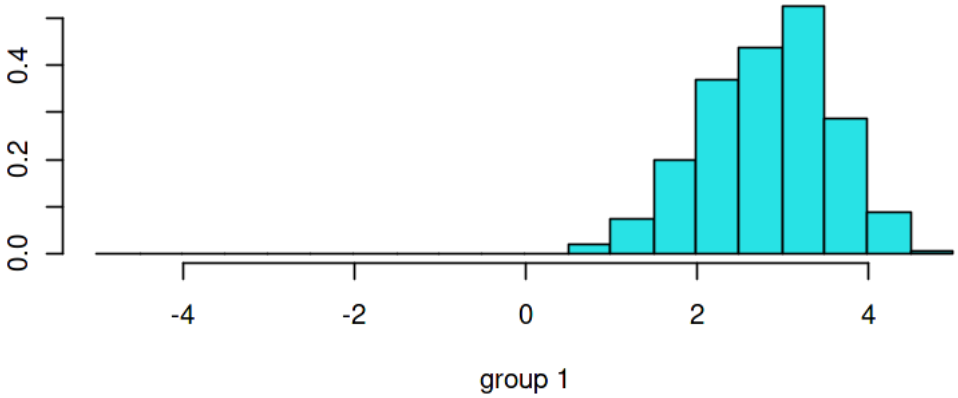
[34]:
```
linear$counts
```

[34]:
0 539 **1** 434

[35]:
```
#Histogram
p<-predict(linear,training)
ldahist(data=p$x[,1],g=training$class)
```

[35]:





[36]:
```
#partition plot
library(klaR)
partimat(class~.,data=training,method="lda")
```

```
    Error in partimat.default(x, grouping, …): at least two classes required
Traceback:


    1. partimat(class ~ ., data = training, method = "lda")

    2. partimat.formula(class ~ ., data = training, method = "lda")
```

```
3. partimat.default(x, grouping, …)

4. stop("at least two classes required")
```

[37]:
```
#confusion matrix and accuracy training data
p1<-predict(linear,training)$class
tab<-table(Predicted=p1, Actual=training$class)
tab
```

[37]:
```
          Actual
Predicted   0   1
        0 515   0
        1  24 434
```

[38]:
```
accuracy<-sum(diag(tab))/sum(tab)
accuracy
```

[38]:
```
0.975334018499486
```

[39]:
```
#The accuracy is 97.53% on training data
```

[40]:
```
#Confusion Matrix and accuracy on testing data
p2<-predict(linear,testing)$class
tab1<-table(Predicted=p2, Actual=testing$class)
tab1
```

[40]:
```
          Actual
Predicted   0   1
        0 187   0
        1   6 151
```

[41]:
```
accuracy1<-sum(diag(tab1))/sum(tab1)
accuracy1
```

[41]:
```
0.982558139534884
```

[42]:
```
#The accuracy is 98.26% on testing data
```

[0]:
```

```