ANÁLISIS DISCRIMINANTE PARA LA CLASIFICACIÓN DEL IRIS DATASET

Abstract

En este informe se estudia la base de datos Iris en la que se almacenan la medida en centímetros del ancho y largo de los sépalos y pétalos de 150 flores de género Iris. Estas ciento cincuenta observaciones se reparten de manera equitativa entre las tres especies (Setosa, Versicolor y Virgínica) presentándose cincuenta ejemplares de cada una. En esta memoria se presentan y comparan dos técnicas de clasificación; análisis discriminante lineal y cuadrático, para la distinción de las especies de las flores según sus medidas.

Se comienza este estudio realizando un breve análisis exploratorio de la base de datos. Para aquellos lectores ajenos al mundo de la botánica se presenta en la siguiente imagen (**Figura 1**) un ejemplar de iris versicolor en el que se distinguen nítidamente el pétalo y el sépalo y se indican como se toman las medidas antes mencionadas:

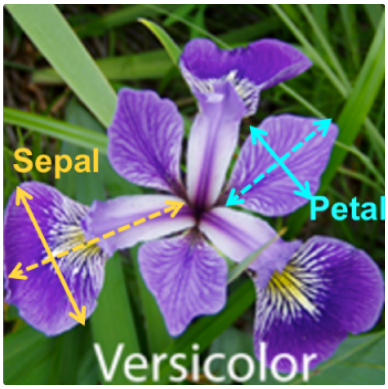


Figura 1. Referencia de toma de medidas.

Se comprueba en primer lugar que la base de datos no posee ningún registro sin definir. Tras ello se realizan una serie de cálculos que se resumen y estructuran en la siguiente tabla:

	Largo del sépalo	Ancho del sépalo	Largo del pétalo	Ancho del pétalo
Mínimo	4,300	2,000	1,000	0,100
Primer Cuartil	5,100	2,800	1,600	0,300
Mediana	5,800	3,000	4,350	1,300
Media	5,843	3,057	3,758	1,199
Tercer Cuartil	6,400	3,300	5,100	1,800
Máximo	7,900	4,400	6,900	2,500
Varianza	0,681122	0,186751	3,092425	0,578532

Para concluir este análisis exploratorio se muestra una gráfica (**Figura 2**) que presenta cómo se distribuyen las especies según las medidas de su pétalo y sépalo (más que sus medidas una aproximación de estas suponiendo pétalos y sépalos rectangulares).

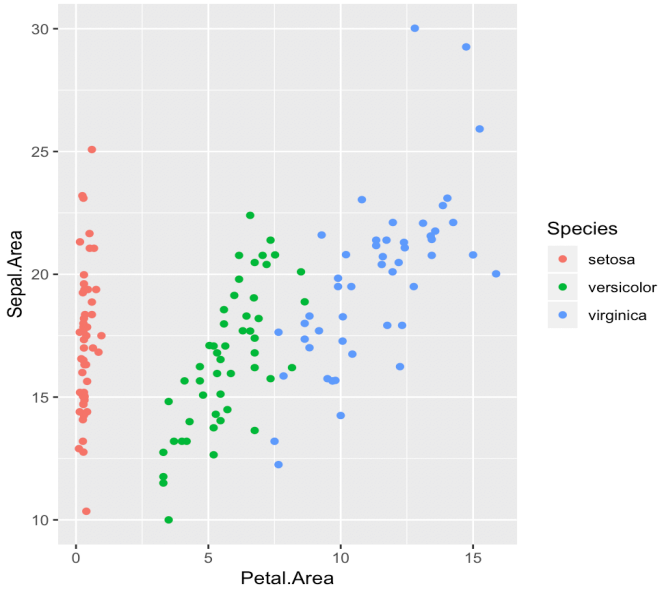


Figura 2. Distribución de las especies según las medidas de superficie de su pétalo y sépalo.

El gráfico muestra que la especie Setosa es fácilmente distinguible de las otras dos mediante las medidas del pétalo. Por su parte los especímenes versicolor y virgínicos son a su vez distinguibles aunque no de una manera tan absoluta, presentan una región común entre los pétalos de 8 cm^2 y los de 10 cm^2 en la que será necesario recurrir a las medidas del sépalo para distinguirlas.

Para confirmar la validez de análisis del discriminante debe poder asegurarse que existe una diferencia de medias estadísticamente significativa y que las variables siguen una distribución normal.

Con el objetivo de comprobar que existe una diferencia de medias estadísticamente significativa en este estudio se emplea el test de Wilks que plantea como hipótesis nula que no existen dichas diferencias entre las medias de cada especie. Al realizar dicho test se obtiene un p-valor inferior a $2 \cdot 10^{-16}$ luego existe una fuerte evidencia empírica en la muestra para rechazar la hipótesis nula por lo que existen diferencias significativas en las medias y por tanto tiene sentido la clasificación obtenida a partir del análisis del discriminante.

Respecto a la normalidad de las variables se observa a partir de distintos contrastes de hipótesis (Jarque-Vera, Shapiro-Wilks…) que en la muestra existe suficiente evidencia empírica para rechazar la hipótesis de normalidad y por ello en este informe se plantean dos variantes. En un primer enfoque se considera que aunque no existe normalidad per se, la distancia de la distribución de las variables a la normal es, aunque existente, aceptablemente leve. En un segundo enfoque se trabaja con las variables normalizadas.

Tras esta breve introducción a los datos se procede al estudio de las dos técnicas antes citadas. Para ambos análisis se usará la biblioteca MASS[1].

[1] El código tanto del análisis exploratorio previo como de los resultados expuestos a continuación se encuentra en el anexo al final de este documento así como en este enlace para aquellos usuarios que deseen interactuar con él.

Análisis lineal del discriminante para variables originales

Al llevar a cabo este análisis se confirma que la muestra está perfectamente balanceada (algo que ya se conocía tras el análisis exploratorio de datos). Además obtenemos los coeficientes para la construcción de los dos discriminantes:

DL1: (0.83 * Largo.Sépalo) + (1.53 * Ancho.Sépalo) - (2.20 * Largo.Pétalo) - (2.81*Ancho.Pétalo)
DL2: (0.02 * Largo.Sépalo) + (2.16 * Ancho.Sépalo) - (0.93 * Largo.Pétalo) + (2.84*Ancho.Pétalo)

El primer discriminante lineal (DL1) explica un 99.12% de la varianza quedando el resto de la varianza (0.88) explicada por el segundo (DL2).

Además, una vez se dispone de este modelo es posible predecir la clasificación de nuevos elementos a considerar. En el código adjunto se puede observar como si ahora se considera una nueva flor con un sépalo de 5.5 x 3.6 y un pétalo de 1.6 x 0.4 (largo x ancho) mediante este discriminante esta se clasifica como setosa. Un resultado consistente pues la gráfica muestra como un pétalo pequeño como es en el caso separa a la setosa de las otras dos especies.

Tras esto se estudia la calidad de este modelo (entendiéndose como un modelo mejor aquel que presenta un menor error). En este caso los errores se penalizan de manera equitativa (en ocasiones existen errores que se pueden considerar más dañinos que otros). Para ello, se plantea el siguiente experimento:

Se selecciona una partición de la muestra del 70% que será dirigida al entrenamiento y ajuste del modelo. Tras ello se empleará el 30% de muestra restante para comprobar la corrección de las predicciones realizadas mediante el modelo entrenado. Dicho experimento se repetirá un determinado número de veces (en este caso una centena) calculando el error en cada caso. Finalmente se computará el error medio que será el indicador de bondad del modelo calculado. En el experimento realizado en el código anexo se obtiene un error medio para el modelo lineal de 0.02289.

Análisis lineal del discriminante para variables normalizadas

La cantidad de ejemplares no varía por normalizar las medidas de sus variables luego la muestra sigue estando balanceada. En esta ocasión los coeficientes para la construcción de los dos discriminantes son:

DL1*: (0.69 * Largo.Sépalo) + (0.67 * Ancho.Sépalo) - (3.89 * Largo.Pétalo) - (2.14*Ancho.Pétalo)
DL2*: (0.02 * Largo.Sépalo) + (0.94 * Ancho.Sépalo) - (1.65 * Largo.Pétalo) + (2.16*Ancho.Pétalo)

El primer discriminante lineal (DL1) explica un 99.12% de la varianza quedando el resto de la varianza (0.88) explicada por el segundo (DL2).

De nuevo se procede al estudio de la calidad del modelo para ello se realiza la misma comprobación que en el apartado anterior repitiendo el mismo experimento con la misma muestra el mismo número de veces. En esta ocasión el experimento devuelve para el modelo lineal con variables normalizadas un error medio (0.02289) idéntico al del modelo lineal con las variables originales.

Análisis cuadrático del discriminante para variables originales

Empleando de nuevo la función de librería MASS se estima el modelo cuadrático. Si se toma de nuevo la flor mencionada en el apartado anterior y se pide al modelo que la clasifique también la distingue como un ejemplar de setosa.

De nuevo se procede al estudio de la calidad del modelo para ello se realiza la misma comprobación que en el apartado anterior repitiendo el mismo experimento con la misma muestra el mismo número de veces. El error medio es en esta ocasión de 0.023111 (ligeramente superior al de los modelos lineales).

Análisis cuadrático del discriminante para variables normalizadas

Al realizar el modelo sobre los datos normalizados se obtienen resultados idénticos a los del apartado anterior.

Conclusiones

En este informe se estudian de manera exhaustiva dos técnicas de clasificación: los análisis del discriminante lineal y cuadrático. En primer lugar se garantiza que esta clasificación sea posible comprobando que existen diferencias estadísticamente significativas entre las medias de cada población lo que garantiza la distinción de las especies. Tras ello se estudia la normalidad de las variables mediante distintos test. Todos los resultados parecen indicar que no se puede garantizar una distribución normal de las variables por lo que se plantean dos estudios paralelos: análisis del discriminante sobre las variables originales y análisis del discriminante sobre las variables normalizadas. Al concluir el estudio se deduce que realmente no es una diferencia tan grave pues los errores medios de los modelos obtenidos para unas y otras variables coinciden. En esta línea destaca que el mejor modelo para este análisis es el lineal presentando un error medio del 2.2%.

Código

En el anexo se presenta el código sobre el que se fundamenta este informe. Con el objetivo de no sobrecargar el informe con gráficos se han mostrado en este solo los más importantes y mantenido en el código otros (como los gráficos de clasificación de las variables enfrentadas dos a dos) que puedan resultar de interés para el lector.

Además para aquellos lectores interesados en el interactuar con el código pueden acceder a él a través del siguiente *enlace*:

(**Nota:** Distintos profesores nos han recomendados que elaboremos todos nuestros códigos en inglés, si le supone algún problema hágamelo saber y haré en español los de las siguientes prácticas.)

Iris_Discriminant_Analysis.R

arturosanchezpalacio

Fri Nov 23 15:35:31 2018

```
### PRÁCTICA 1. Discriminant Analysis for the Iris Dataset

# Author: Arturo Sánchez Palacio
# Date: 21/XI/18
# Role: Student of a Data Science for Finance Master (CUNEF)


#These are the libraries we are going to use in this study. Make sure you have them installed before going o
n.

library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────────────────────────────────────── tid
yverse 1.2.1 ──
```

```
## ✔ ggplot2 3.1.0     ✔ purrr   0.2.5
## ✔ tibble  1.4.2     ✔ dplyr   0.7.8
## ✔ tidyr   0.8.1     ✔ stringr 1.3.1
## ✔ readr   1.1.1     ✔ forcats 0.3.0
```

```
## ── Conflicts ────────────────────────────────────────────────────────────────────────── tidyverse
_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

```
library(MASS)
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
library(klaR)
library(mvnormtest)

# Firstly, we have a glance over the dataframe and check its quality:

head(iris)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

```
str(iris)
```

```
## 'data.frame':    150 obs. of  5 variables:
##  $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
##  $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
##  $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
##  $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
##  $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...
```

```
sum(is.na(iris$Sepal.Length))
```

```
## [1] 0
```

```
sum(is.na(iris$Sepal.Width))
```

```
## [1] 0
```

```
sum(is.na(iris$Petal.Length))
```

```
## [1] 0
```

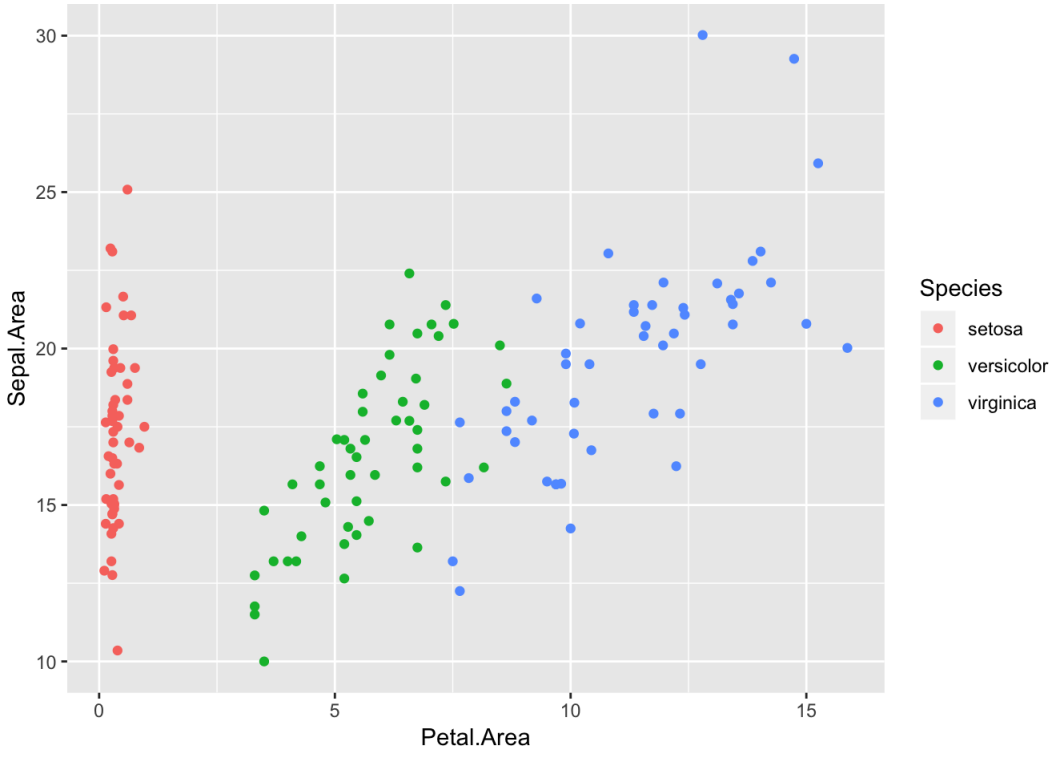
```
sum(is.na(iris$Petal.Width))
```

```
## [1] 0
```

```
sum(is.na(iris$Species))
```

```
## [1] 0
```

```
#The dataset is complete. It is composed by four quantitative variables about the size of the flower and a c
ategoric variable
#specifying the species of each specimen.

#First we are going to check how the sample is distributed. In order to do that we are going to plot the dim
ension of the petal and the sepal
#in a dispersion graph where each colour corresponds the species of each flower:

data <- iris
extradata <- mutate(data, Petal.Area = Petal.Width * Petal.Length ) #Calculates the variable Petal.Area (we
assume petals are close to the shape of a rectangle)
extradata <- mutate(extradata, Sepal.Area = Sepal.Width * Sepal.Length) #Calculates the variable Sepal.Area
(we assume sepals are close to the shape of a rectangle)

ggplot(data = extradata, aes(x = Petal.Area, y = Sepal.Area, col = Species)) + geom_point() #Interpretation
of the graph available on the report.
```



```r
#To finish this exploration we are going to test if the variables follow a normal distribution. In order to do that we are going to use
# the Shapiro-Wilks test through the function mshapiro.test available in the library mvnormtest:

setosa <- data[1:50,1:4]
versicolor <- data[51:100, 1:4]
virginica <- data[101:150, 1:4]

setosa <- t(setosa)
versicolor <- t(versicolor)
virginica <- t(virginica)
mshapiro.test(setosa)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Z
## W = 0.95878, p-value = 0.07906
```

```r
mshapiro.test(versicolor)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Z
## W = 0.93043, p-value = 0.005739
```

```r
mshapiro.test(virginica)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Z
## W = 0.93414, p-value = 0.007955
```

```r
#From the Shapiro test we deduca that the distribution doesn't necessarily follow a normal distribution so we present
# two choices:

#CHOICE 1
#We assume that the difference is not significative and that we can work with the data.

#CHOICE 2
#We scale the varibles in order to guarantee that the distribution are normal and the Discriminant Analysis is effitient.









### CHOICE 1 ####

#Once we have done a small exploration of the sample we are going to check whether there is or not a significal difference between
# each population's means.

fit.manova <- manova(data = data, cbind(data$Sepal.Length, data$Petal.Length, data$Sepal.Width, data$Petal.Width)~data$Species)
summary((fit.manova),test = 'Wilks')
```

```
##               Df    Wilks approx F num Df den Df    Pr(>F)
## data$Species   2 0.023439   199.15      8    288 < 2.2e-16 ***
## Residuals    147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# After this comprobation we proceed to the Linear Discriminant Analysis. In order to conduct this analysis
we are going to employ the library MASS.

(linear_analysis <- lda(Species ~., data = data ))
```

```
## Call:
## lda(Species ~ ., data = data)
##
## Prior probabilities of groups:
##     setosa versicolor  virginica
##  0.3333333  0.3333333  0.3333333
##
## Group means:
##            Sepal.Length Sepal.Width Petal.Length Petal.Width
## setosa            5.006       3.428        1.462       0.246
## versicolor        5.936       2.770        4.260       1.326
## virginica         6.588       2.974        5.552       2.026
##
## Coefficients of linear discriminants:
##                     LD1         LD2
## Sepal.Length  0.8293776  0.02410215
## Sepal.Width   1.5344731  2.16452123
## Petal.Length -2.2012117 -0.93192121
## Petal.Width  -2.8104603  2.83918785
##
## Proportion of trace:
##    LD1    LD2
## 0.9912 0.0088
```

```
#The sample is perfectly balanced so there is no need to add prior probabilities.


# We also conduct the Quadratic Discriminant Analysis:

(quadratic_analysis <- qda(Species ~., data = data))
```

```
## Call:
## qda(Species ~ ., data = data)
##
## Prior probabilities of groups:
##     setosa versicolor  virginica
##  0.3333333  0.3333333  0.3333333
##
## Group means:
##            Sepal.Length Sepal.Width Petal.Length Petal.Width
## setosa            5.006       3.428        1.462       0.246
## versicolor        5.936       2.770        4.260       1.326
## virginica         6.588       2.974        5.552       2.026
```

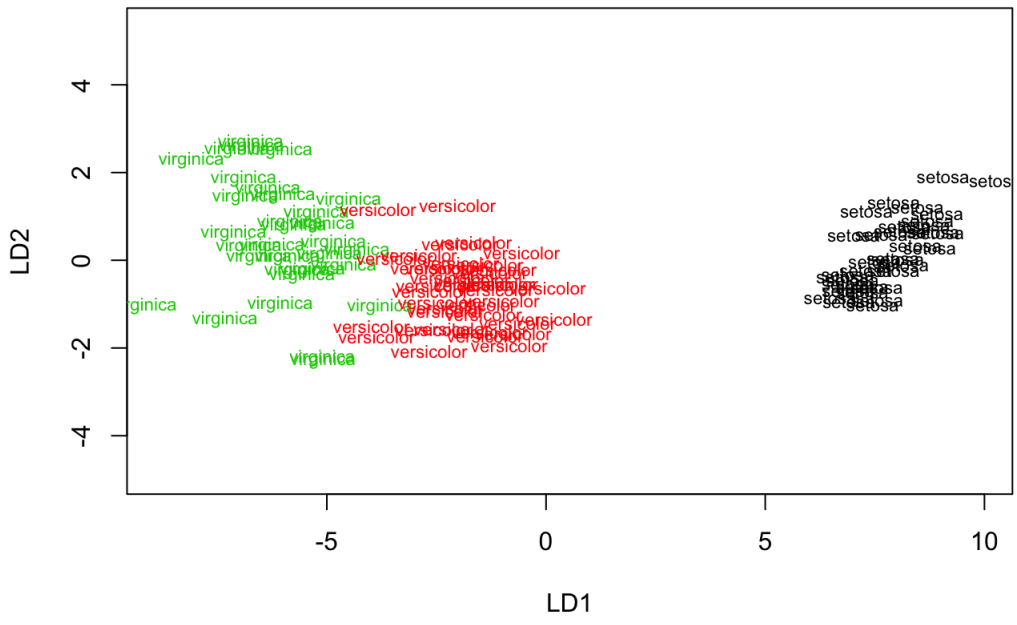
```
#Now that we have both analysis we can check prediction for new elements. Imagin we have a new flower with m
easurements: Petal: 1.6 x 0.4 and Sepal: 5.5 x 3.6 (length x width):

predict(linear_analysis,newdata = data.frame(Sepal.Length = 5.5,Sepal.Width = 3.6, Petal.Length = 1.6, Petal
.Width = 0.4))$class
```

```
## [1] setosa
## Levels: setosa versicolor virginica
```

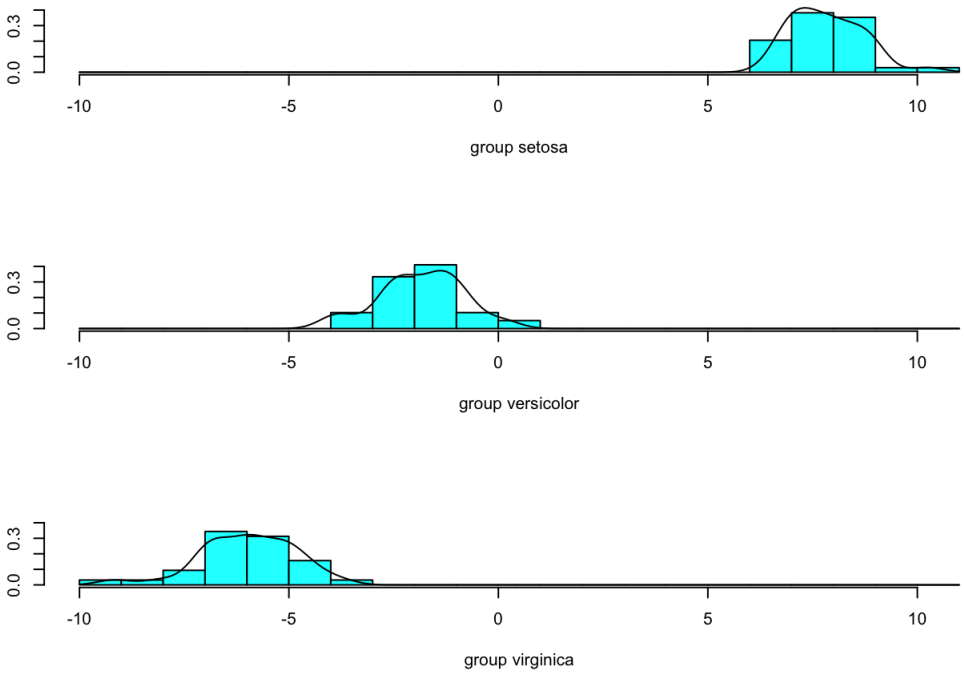
```
predict(quadratic_analysis,newdata = data.frame(Sepal.Length = 5.5,Sepal.Width = 3.6, Petal.Length = 1.6, Pe
tal.Width = 0.4))$class
```

```
## [1] setosa
## Levels: setosa versicolor virginica
```

```
# Both prediction classify this flower as setosa (which seems consistent since its petals and sepals are qui
te tiny)


#Once this is done it is time to check which model gives better predictions. In order to do that the sample
is going to be
#splitted in a training sample and a testing sample. We are going to repeat this experiment several times es
timating in each the error.
#After this we are going to compare the mean error of both models. (Obviously) the one witht the smallest er
ror will be the best one.

sample_size <- dim(iris)[1]
training_size <- sample_size * 0.7 #We take 80% from the original sample to train the model and 20% to test
it.
testing_size <- sample_size - training_size

iterations <- 100 #We are going to repeat the experiment 100 times.

set.seed(12345) #We set the seed in order to obtain consistent results in the experiment

# First, we conduct the experiment over the Linear Analysis

linear_error <- dim(iterations) #We initialize a vector where we are going to store the error

for (k in 1:iterations) {
  train <- sample(1:sample_size,training_size) #We choose the training sample.
  m1 <- lda(Species~.,data[train,]) #We apply the Linear Discriminant Analysis over the training sample.
  predict(m1,data[-train,])$class #We use this model to predict the testing set.
  tablin <- table(data$Species[-train],predict(m1,data[-train,])$class) #This gives a table with the results
from the analysis
  linear_error[k] <- (testing_size - sum(diag(tablin)))/testing_size #We calculate the error
}
specie_train <- data[train, 5]
plot(m1, col = as.integer(specie_train))
```




```
plot(m1, dimen = 1, type = "b")
```




```
(linear_error_mean <- mean(linear_error))
```

```
## [1] 0.02288889
```

```
tablin
```

```
##
##              setosa versicolor virginica
##   setosa         16          0         0
##   versicolor      0         10         1
##   virginica       0          1        17
```

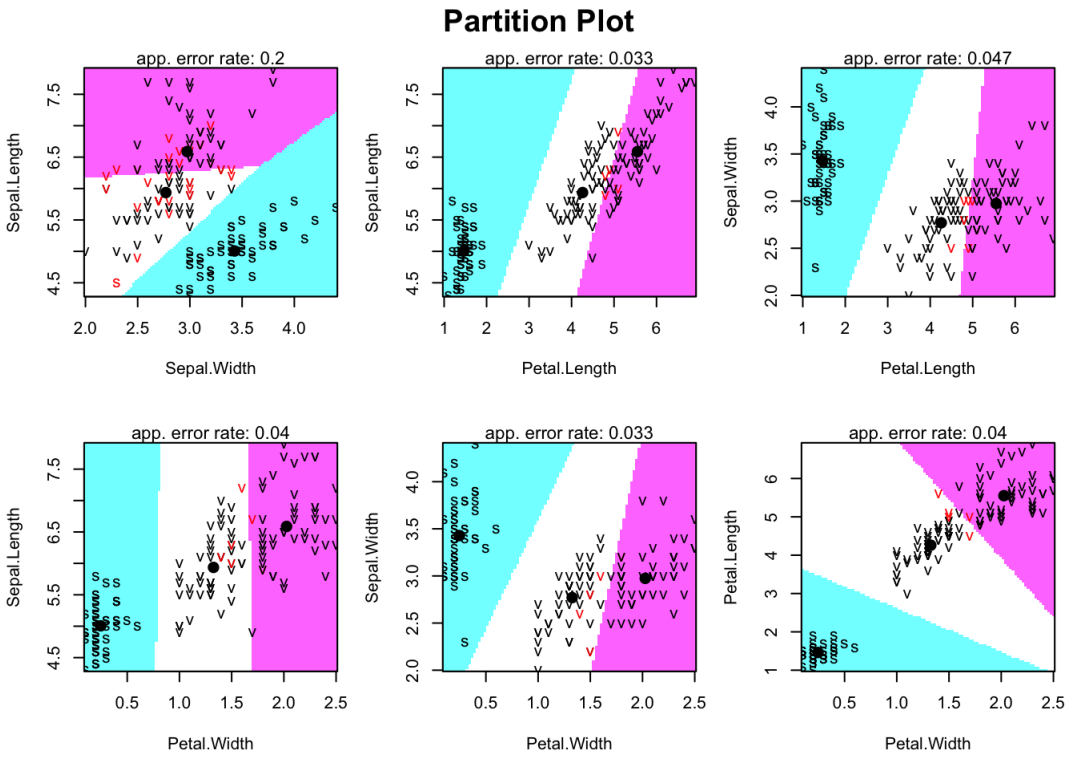
```
# Secondly, we conduct the experiment over the Quadratic Analysis

#We are going to use the same sample as in the first experiment:

quadratic_error <- dim(iterations)

for (k in 1:iterations) {
  train <- sample(1:sample_size,training_size) #We choose the training sample.
  m2 <- qda(Species~.,data[train,]) #We apply the Quadratic Discriminant Analysis over the training sample.
  predict(m2,data[-train,])$class #We use this model to predict the testing set.
  tablin <- table(data$Species[-train],predict(m2,data[-train,])$class) #This gives a table with the results
from the analysis
  quadratic_error[k] <- (testing_size - sum(diag(tablin)))/testing_size #We calculate the error
}
(quadratic_error_mean <- mean(quadratic_error))
```

```
## [1] 0.02311111
```

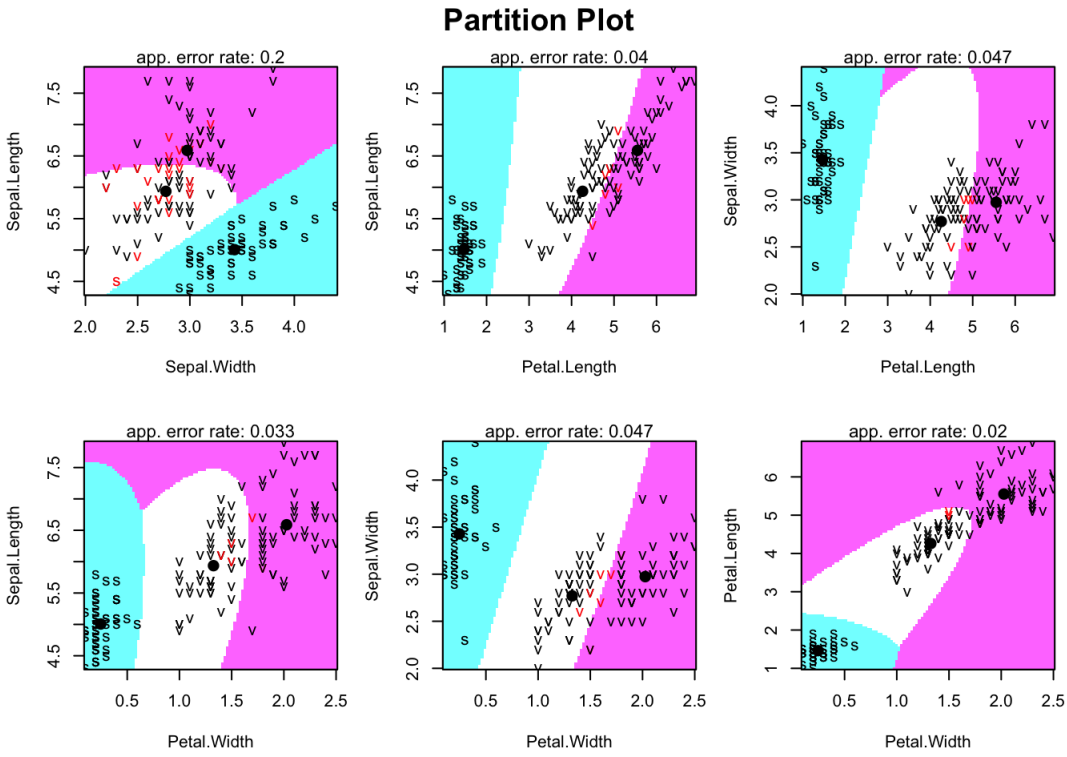
```
tablin
```

```
##
##              setosa versicolor virginica
##   setosa         13          0         0
##   versicolor      0         15         1
##   virginica       0          0        16
```

```
# Finally we are going to plot the results of both classifications (in order to do this qe use the function
partimat included
# in the library klaR):


partimat(Species~.,data = data,method = "lda")
```




```
partimat(Species~.,data = data,method = "qda")
```



```
### CHOICE 2 ####

#Reminder: in this choice we decide to scale the variable to ensure that the follow a normal distribution:

scaled_data <- as.data.frame(scale(data[,-5]))
scaled_data <- mutate(scaled_data, Species = iris$Species)

#Once we have done a small exploration of the sample we are going to check whether there is or not a signifi
cal difference between
# each population's means.

fit.manova <- manova(data = scaled_data, cbind(scaled_data$Sepal.Length, scaled_data$Petal.Length, scaled_da
ta$Sepal.Width, scaled_data$Petal.Width)~scaled_data$Species)
summary((fit.manova),test = 'Wilks')
```

```
##                      Df    Wilks approx F num Df den Df    Pr(>F)
## scaled_data$Species   2 0.023439   199.15      8    288 < 2.2e-16 ***
## Residuals           147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

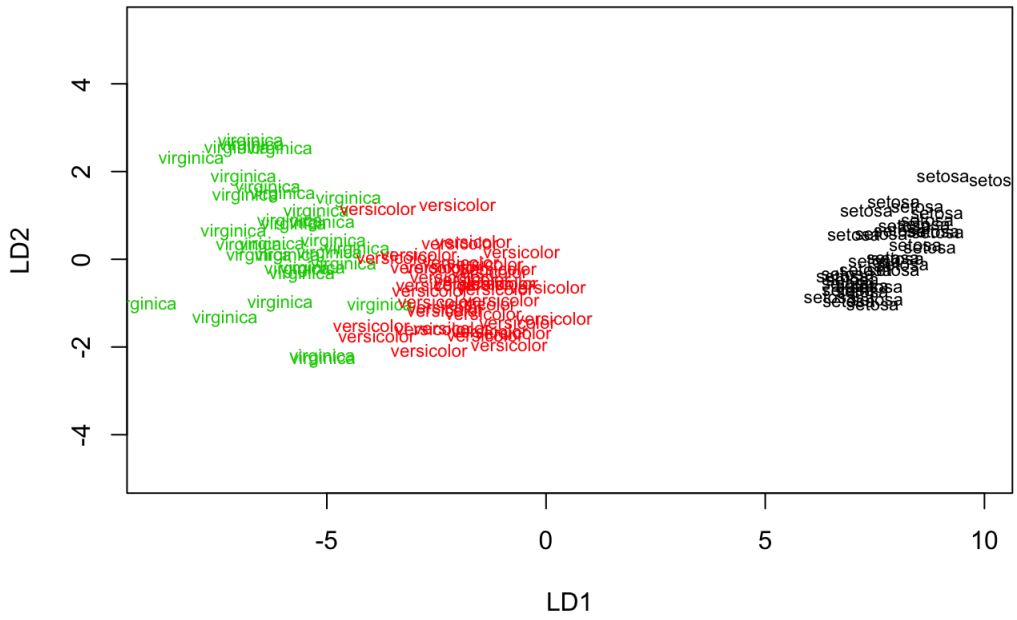
```
# After this comprobation we proceed to the Linear Discriminant Analysis. In order to conduct this analysis
we are going to employ the library MASS.

(linear_analysis <- lda(Species ~., data = scaled_data ))
```

```
## Call:
## lda(Species ~ ., data = scaled_data)
##
## Prior probabilities of groups:
##     setosa versicolor  virginica
##  0.3333333  0.3333333  0.3333333
##
## Group means:
##            Sepal.Length Sepal.Width Petal.Length Petal.Width
## setosa       -1.0111914   0.8504137   -1.3006301  -1.2507035
## versicolor    0.1119073  -0.6592236    0.2843712   0.1661774
## virginica     0.8992841  -0.1911901    1.0162589   1.0845261
##
## Coefficients of linear discriminants:
##                     LD1         LD2
## Sepal.Length  0.6867795  0.01995817
## Sepal.Width   0.6688251  0.94344183
## Petal.Length -3.8857950 -1.64511887
## Petal.Width  -2.1422387  2.16413593
##
## Proportion of trace:
##    LD1    LD2
## 0.9912 0.0088
```

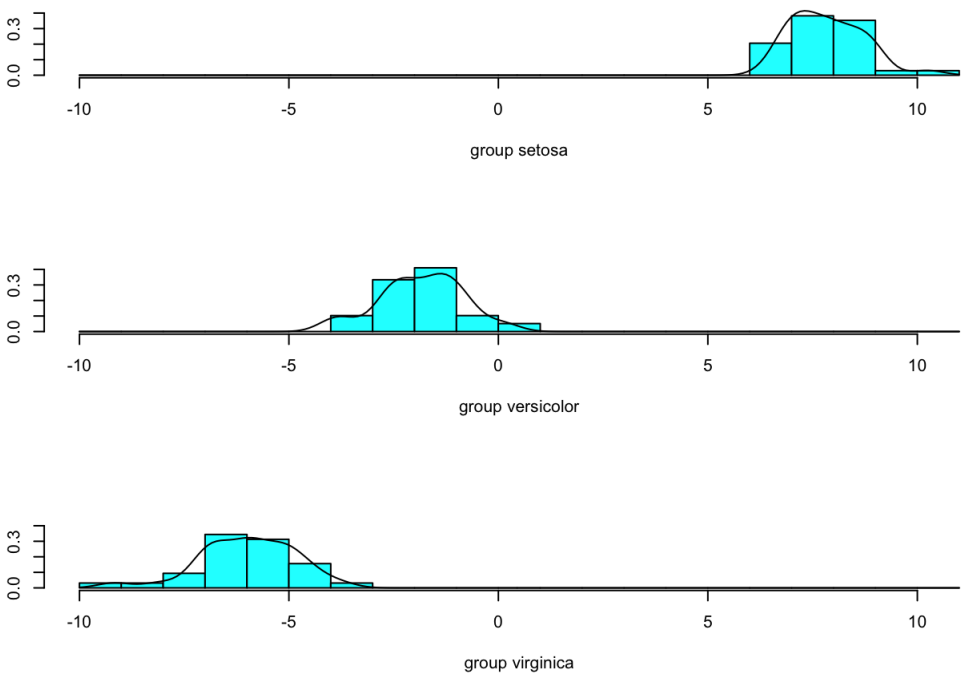
```
#The sample is perfectly balanced so there is no need to add prior probabilities.


# We also conduct the Quadratic Discriminant Analysis:

(quadratic_analysis <- qda(Species ~., data = scaled_data))
```

```
## Call:
## qda(Species ~ ., data = scaled_data)
##
## Prior probabilities of groups:
##     setosa versicolor  virginica
##  0.3333333  0.3333333  0.3333333
##
## Group means:
##            Sepal.Length Sepal.Width Petal.Length Petal.Width
## setosa       -1.0111914   0.8504137   -1.3006301  -1.2507035
## versicolor    0.1119073  -0.6592236    0.2843712   0.1661774
## virginica     0.8992841  -0.1911901    1.0162589   1.0845261
```

```
#Once this is done it is time to check which model gives better predictions. In order to do that the sample
is going to be
#splitted in a training sample and a testing sample. We are going to repeat this experiment several times es
timating in each the error.
#After this we are going to compare the mean error of both models. (Obviously) the one witht the smallest er
ror will be the best one.

sample_size <- dim(iris)[1]
training_size <- sample_size * 0.7 #We take 70% from the original sample to train the model and 30% to test
it.
testing_size <- sample_size - training_size

iterations <- 100 #We are going to repeat the experiment 100 times.

set.seed(12345) #We set the seed in order to obtain consistent results in the experiment

# First, we conduct the experiment over the Linear Analysis

linear_error <- dim(iterations) #We initialize a vector where we are going to store the error

for (k in 1:iterations) {
  train <- sample(1:sample_size,training_size) #We choose the training sample.
  m1 <- lda(Species~.,scaled_data[train,]) #We apply the Linear Discriminant Analysis over the training samp
le.
  predict(m1,scaled_data[-train,])$class #We use this model to predict the testing set.
  tablin <- table(scaled_data$Species[-train],predict(m1,scaled_data[-train,])$class) #This gives a table wi
th the results from the analysis
  linear_error[k] <- (testing_size - sum(diag(tablin)))/testing_size #We calculate the error
}
specie_train <- scaled_data[train, 5]
plot(m1, col = as.integer(specie_train))
```




```
plot(m1, dimen = 1, type = "b")
```




```
(linear_error_mean <- mean(linear_error))
```

```
## [1] 0.02288889
```

```
tablin
```

```
##
##              setosa versicolor virginica
##   setosa         16          0         0
##   versicolor      0         10         1
##   virginica       0          1        17
```

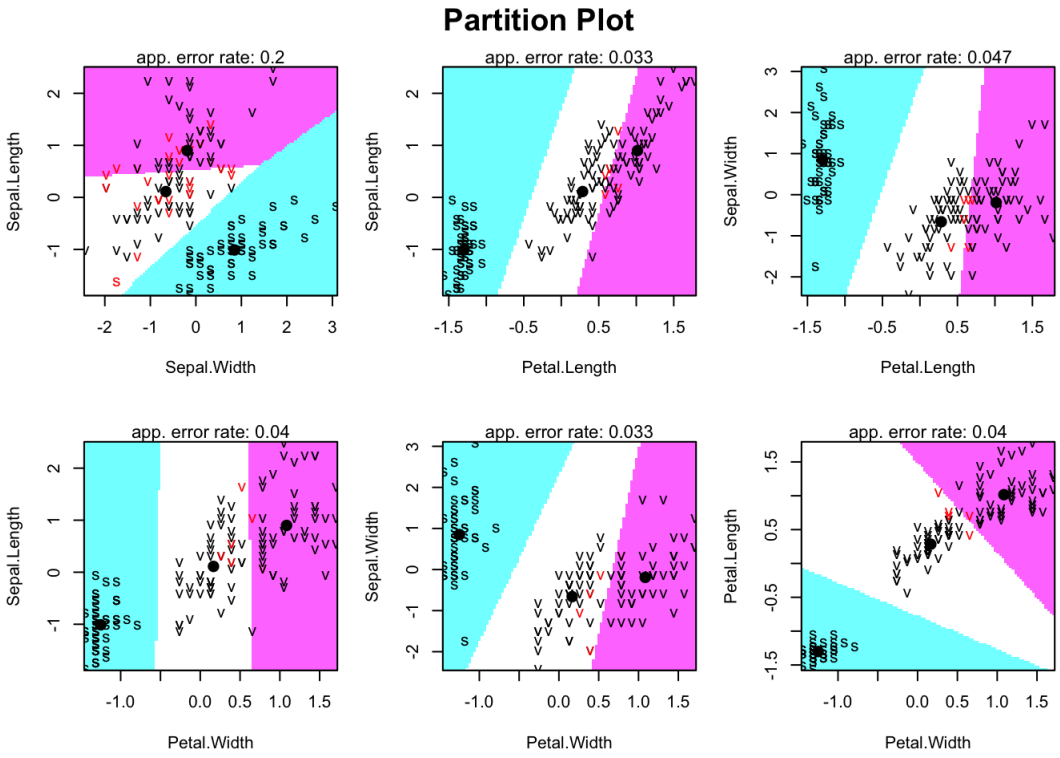
```
# Secondly, we conduct the experiment over the Quadratic Analysis

#We are going to use the same sample as in the first experiment:

quadratic_error <- dim(iterations)

for (k in 1:iterations) {
  train <- sample(1:sample_size,training_size) #We choose the training sample.
  m2 <- qda(Species~.,scaled_data[train,]) #We apply the Quadratic Discriminant Analysis over the training s
ample.
  predict(m2,scaled_data[-train,])$class #We use this model to predict the testing set.
  tablin <- table(scaled_data$Species[-train],predict(m2,scaled_data[-train,])$class) #This gives a table wi
th the results from the analysis
  quadratic_error[k] <- (testing_size - sum(diag(tablin)))/testing_size #We calculate the error
}
(quadratic_error_mean <- mean(quadratic_error))
```

```
## [1] 0.02311111
```

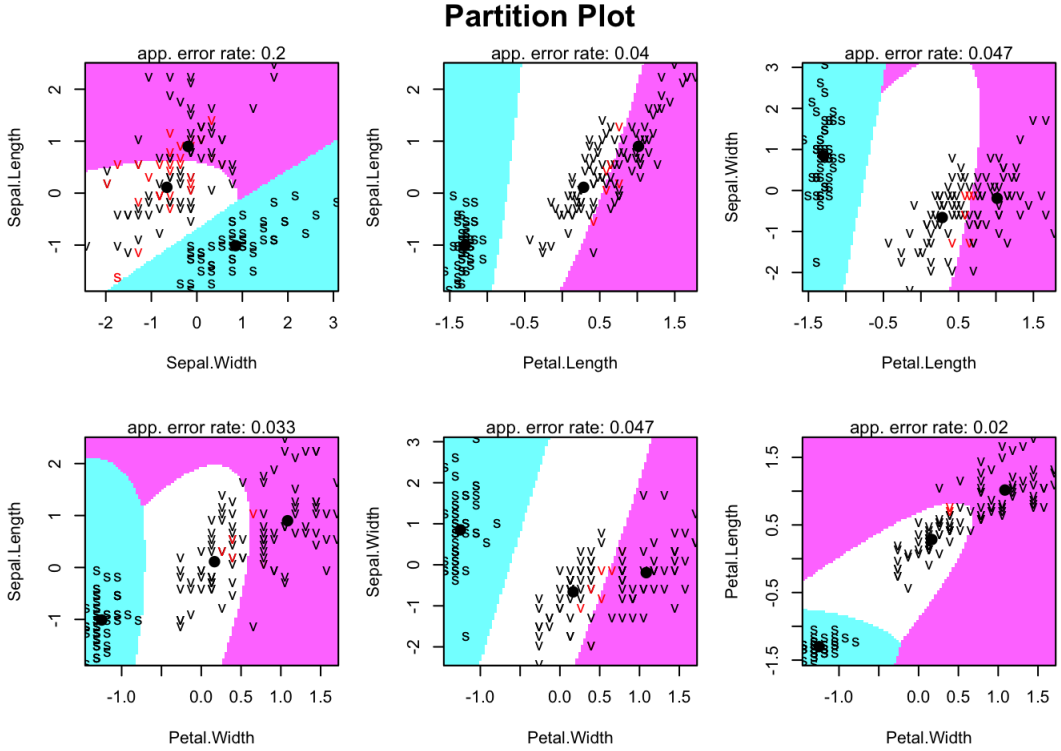
```
tablin
```

```
##
##              setosa versicolor virginica
##   setosa         13          0         0
##   versicolor      0         15         1
##   virginica       0          0        16
```

```
# Finally we are going to plot the results of both classifications (in order to do this qe use the function
partimat included
# in the library klaR):


partimat(Species~.,data = scaled_data,method = "lda")
```




```
partimat(Species~.,data = scaled_data,method = "qda")
```


Partition Plot
app. error rate: 0.2
Sepal.Length
Sepal.Width
app. error rate: 0.04
Sepal.Length
Petal.Length
app. error rate: 0.047
Sepal.Width
Petal.Length
app. error rate: 0.033
Sepal.Length
Petal.Width
app. error rate: 0.047
Sepal.Width
Petal.Width
app. error rate: 0.02
Petal.Length
Petal.Width