Análisis Discriminante Ejemplo

Oscar Elí Bonilla Morales

2022-05-25

##Obtención de los datos Los datos utilizados para este ejemplo fueron obtenidos de la página kaggle https://www.kaggle.com/datasets/bertiemackie/sloth-species

Cargamos la base de datos

Descripción

Esta base cuenta con datos sobre direntes especies de perezosos, de igual manera, contiene datos como largo de la garra, peso, largo de cola, etc.

Preparación de la matriz

```
library(readr)
install.packages("MASS")
library(MASS)
perezoso <- read_csv("perezoso/perezoso.csv")
head(perezoso)
```

```
## # A tibble: 6 x 5
##   claw_length_cm tail_length_cm weight_kg size_cm specie
##            <dbl>          <dbl>     <dbl>   <dbl> <chr>
## 1           6.82           4.45      3.57    52.0 three_toed
## 2           8.26           6.29      2.84    50.1 three_toed
## 3           8.66           4.55      1.26    51.5 three_toed
## 4           8.47           6.98      2.39    50.1 three_toed
## 5           7.10           5.41      3.16    51.4 three_toed
## 6           7.27           3.67      3.30    50.5 three_toed
```

ANALISIS DISCRIMINANTE LINEAl

```
z <- as.data.frame(perezoso)
```

Se define la matriz de datos y la variable

```
x<-z[,1:2]

y<-z[,5]
```

Definir como n y p el numero de flores y variables

```
n<-nrow(x)
p<-ncol(x)
```

Se aplica el Analisis discriminante lineal (LDA)

Cross validation (cv): clasificacion optima

```
lda.pere<-lda(y~.,data=x, CV=TRUE)
```

lda.iris$class contiene las clasificaciones hechas por CV usando LDA.

```
head(lda.pere$class)
```

```
## [1] three_toed three_toed three_toed three_toed three_toed three_toed
## Levels: three_toed two_toed
```

Creacion de la tabla de clasificaciones buenas y malas

```
table.lda<-table(y,lda.pere$class)
table.lda
```

```
##
## y            three_toed two_toed
##   three_toed       2570       92
##   two_toed           98     2240
```

En esta tabla es posible apreciar que nuestra variable "three_toed" (Tres dedos) existen **2570** datos los cuales se encuentran bien clasificados,por otro lado obtenemos **92** de la otra especie

En nuestra variable "two:toed" (Dos dedos) tenemos **2240** bien clasificadas y **98** los cuales nos fueron bien clasificadas.

Proporcion de errores

```
mis.pere<- n-sum(y==lda.pere$class)
mis.pere/n
```

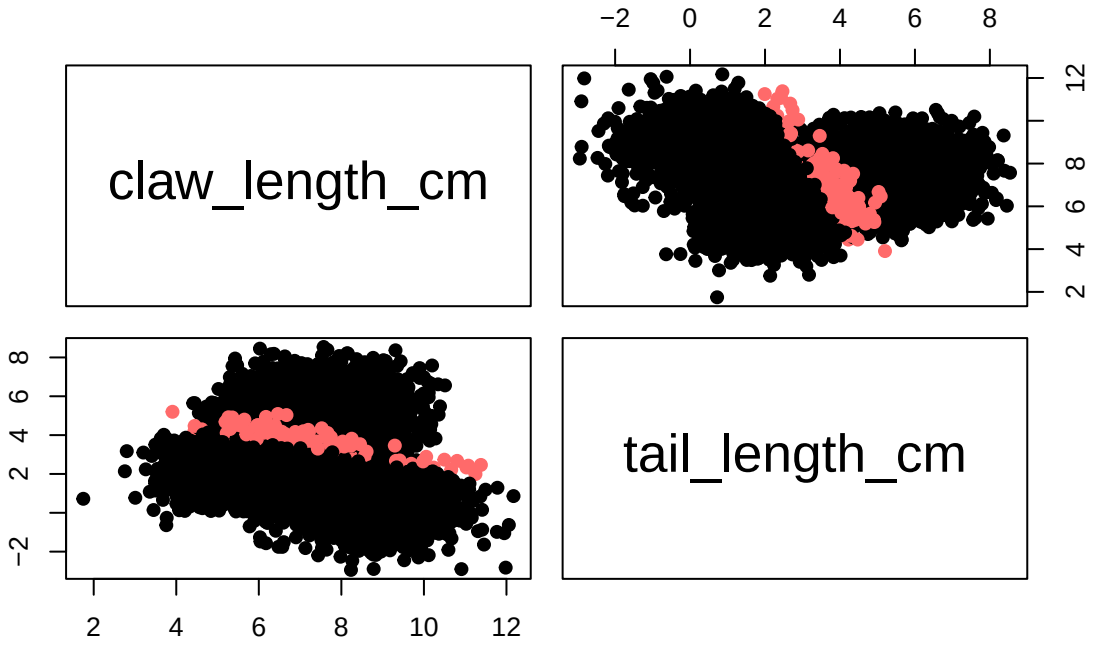
```
## [1] 0.038
```

Este valor representa que por cada 5000 perezosos clasificados, nos equivocaremos un 0.03% de las veces, ya que 0.038 es una probabilidad pequeña podemos decir que tenemos un metódo adecuado

scater plot

Buenas clasificaciones en negro y malas en rojo

```
col.lda.pere<-c("indianred1","black")[1*(y==lda.pere$class)+1]
pairs(x,main="Buena Clasificacion (negro), Mala Clasificacion (rojo)",
      pch=19,col=col.lda.pere)
```



En este gráfico se muestran tanto los datos bien clasificados al igual que los malos, es posible observar que son pocos aquellos datos que estan mal clasificados, por lo que podemos cocluir que las probabilidades de tener buenas clasificaciones es mayor

Probabilidad de pertenencia a uno de los 2 grupos

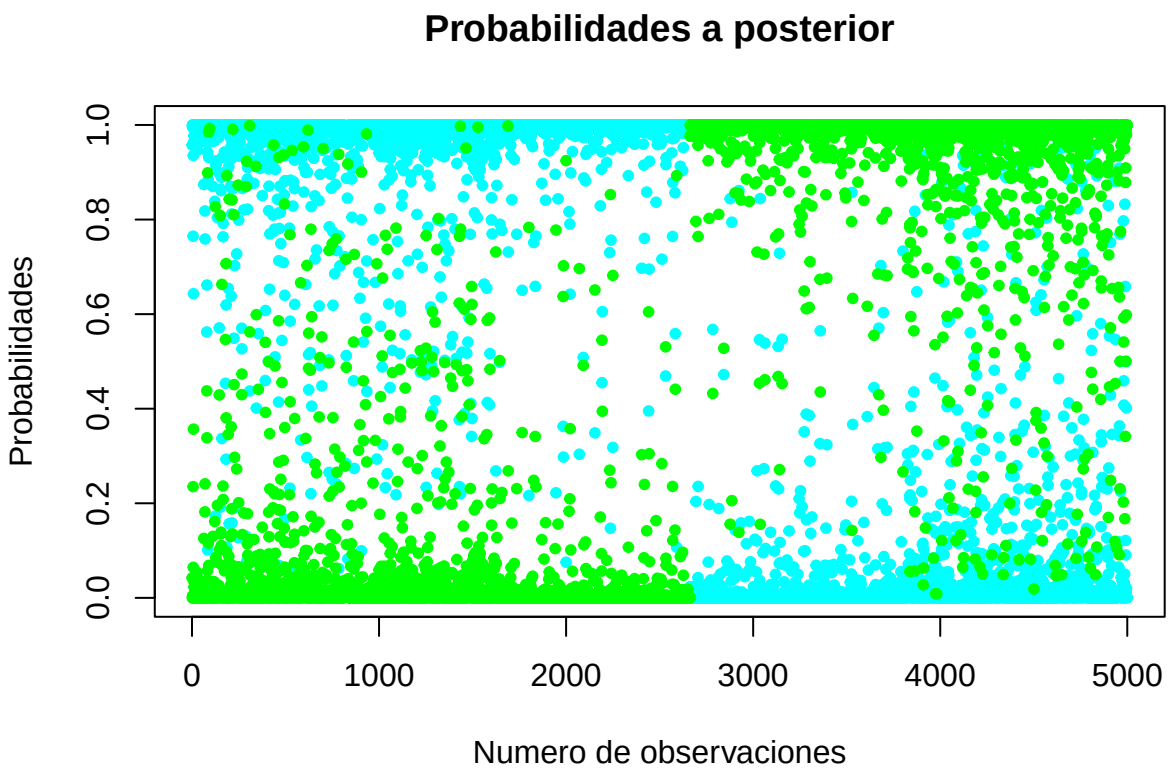
```
head(lda.pere$posterior)
```

```
##   three_toed     two_toed
## 1  0.9575466 4.245342e-02
## 2  0.9999671 3.289133e-05
## 3  0.9953704 4.629594e-03
## 4  0.9999970 3.039415e-06
## 5  0.9983281 1.671895e-03
## 6  0.7648394 2.351606e-01
```

Grafico de probabilidades

```
plot(1:n, lda.pere$posterior[,1],
     main="Probabilidades a posterior",
```

```
     pch=20, col="cyan",
     xlab="Numero de observaciones", ylab="Probabilidades")
points(1:n,lda.pere$posterior[,2],
       pch=20, col="green")
```



En este gráfico es posible apreciar que nuestras probabilidades se encuentran muy dispersas por lo que será dificl identificar los grupos de las clasificaciónes