

Análisis Discriminante (LDA)

Saraí Campos Varela

2022-05-24

Análisis Discriminante Lineal

Introducción

El Análisis Discriminante Lineal o Linear Discriminant Analysis (LDA) es un método de clasificación supervisado de variables cualitativas en el que dos o más grupos son conocidos a priori y nuevas observaciones se clasifican en uno de ellos en función de sus características.

Importación de la matriz

1.- Se cargan los datos de la base de Iris

```
library(MASS)
Z<-as.data.frame(iris)
head(Z)
```

##	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

2.- Se define la matriz de datos y la variable respuesta con las categorías.

```
x<-Z[,1:4]
y<-Z[,5]
```

3.- Definir como n y p el número de flores y variables

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

4.- Se aplica el Análisis discriminante lineal (LDA) con Cross validation (cv): clasificación optima

```
lda.iris<-lda(Z$Species~.,data=Z,CV=TRUE)
```

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

```
lda.iris$class
```

```
## [1] setosa setosa setosa setosa setosa setosa
## [7] setosa setosa setosa setosa setosa setosa
## [13] setosa setosa setosa setosa setosa setosa
## [19] setosa setosa setosa setosa setosa setosa
## [25] setosa setosa setosa setosa setosa setosa
## [31] setosa setosa setosa setosa setosa setosa
## [37] setosa setosa setosa setosa setosa setosa
## [43] setosa setosa setosa setosa setosa setosa
## [49] setosa setosa versicolor versicolor versicolor versicolor
## [55] versicolor versicolor versicolor versicolor versicolor versicolor
## [61] versicolor versicolor versicolor versicolor versicolor versicolor
## [67] versicolor versicolor versicolor versicolor virginica versicolor
## [73] versicolor versicolor versicolor versicolor versicolor versicolor
## [79] versicolor versicolor versicolor versicolor versicolor virginica
## [85] versicolor versicolor versicolor versicolor versicolor versicolor
## [91] versicolor versicolor versicolor versicolor versicolor versicolor
## [97] versicolor versicolor versicolor versicolor virginica virginica
## [103] virginica virginica virginica virginica virginica virginica
## [109] virginica virginica virginica virginica virginica virginica
## [115] virginica virginica virginica virginica virginica virginica
## [121] virginica virginica virginica virginica virginica virginica
## [127] virginica virginica virginica virginica virginica virginica
## [133] virginica versicolor virginica virginica virginica virginica
## [139] virginica virginica virginica virginica virginica virginica
## [145] virginica virginica virginica virginica virginica virginica
## Levels: setosa versicolor virginica
```

6.- Creación de la tabla de clasificaciones buenas y malas

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

```
##
## y      setosa versicolor virginica
## setosa    50         0         0
## versicolor  0        48         2
## virginica  0         1        49
```

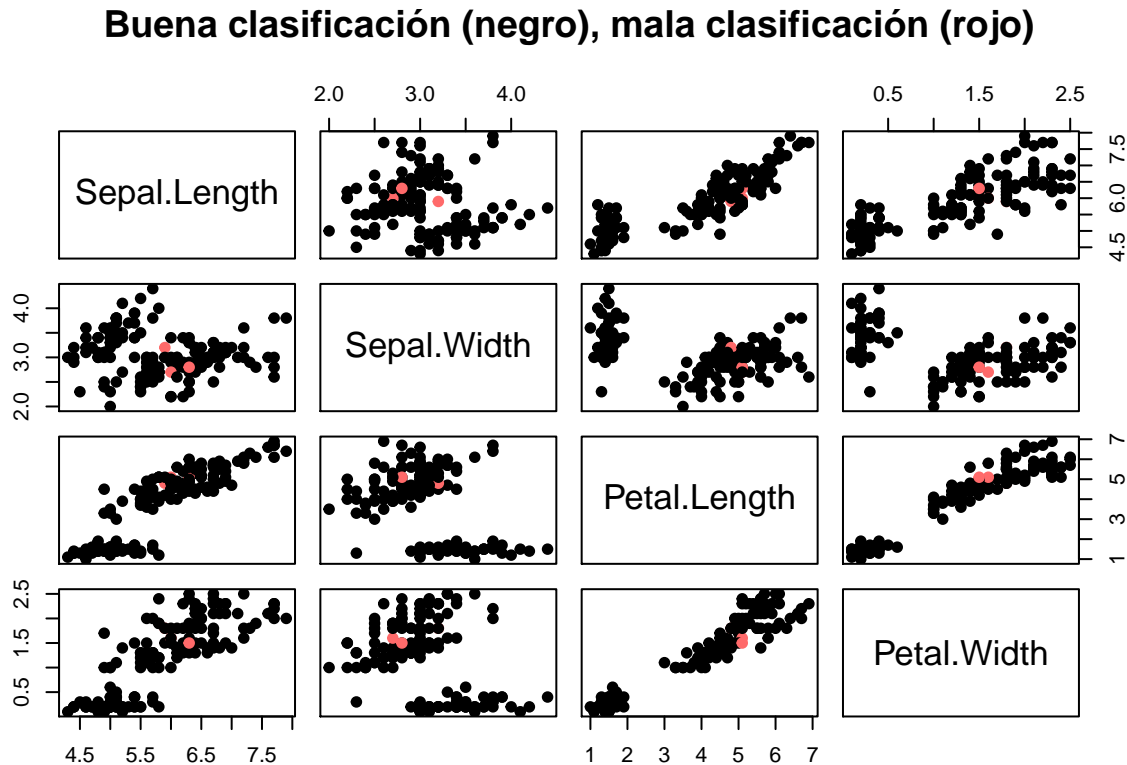
7.- Proporción de errores

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

```
## [1] 0.02
```

8.- scater plot (Buenas clasificaciones en negro y malas en rojo)

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



9.- Probabilidad de pertenencia a uno de los tres grupos

```
lda.iris$posterior[1:10,]
```

```
##      setosa  versicolor  virginica
## 1         1 5.087494e-22 4.385241e-42
## 2         1 9.588256e-18 8.888069e-37
## 3         1 1.983745e-19 8.606982e-39
## 4         1 1.505573e-16 5.101765e-35
## 5         1 2.075670e-22 1.739832e-42
## 6         1 5.332271e-21 8.674906e-40
## 7         1 1.498839e-18 3.999205e-37
## 8         1 5.268133e-20 1.983027e-39
## 9         1 2.280729e-15 1.293376e-33
## 10        1 1.504085e-18 5.037348e-38
```

10.- Gráfico de probabilidades

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

```

    pch=20, col="blue",
    xlab="Número de observaciones", ylab="Probabilidades")
points(1:n,lda.iris$posterior[,2],
      pch=20, col="green")
points(1:n,lda.iris$posterior[,3],
      pch=20, col="orange")

```

Probabilidades a posteriori

