MAE: ENTREGA IV

Jon Zorrilla Gamboa

1. Ejercicio 1

En primer lugar, cargamos los datos.

1.1.

Introducimos el siguiente código:

```
load(url('https://matematicas.uam.es/~joser.berrendero/datos/lirios.RData'))
library(MASS)

lirios1 <- data.frame(lirios[,0:2])
lirios1

#Calculamos los coeficientes de la funcion discriminante de Fisher
resultado.lda <- lda(clases ~ ., data=lirios1, prior = c(0.5, 0.5))
resultado.lda
```

Y obtenemos lo siguiente:

```
Call:
lda(clases ~ ., data = lirios1, prior = c(0.5, 0.5))

Prior probabilities of groups:
  0   1
0.5 0.5

Group means:
  Sepal.Length Sepal.Width
```

```
0          5.936        2.770
1          6.588        2.974

Coefficients of linear discriminants:
                   LD1
Sepal.Length 1.6271842
Sepal.Width  0.3435524
```

Donde, como podemos ver, los coeficientes de la función discriminante de Fisher son 1,6271842 para la longitud del sépalo y 0,3435524 para la anchura del sépalo.

La probabilidad de error mediante riesgo empírico es:

```
predicciones.lda <- predict(resultado.lda)$class
table(clases, predicciones.lda)
      predicciones.lda
clases  0  1
     0 38 12
     1 13 37

mean(clases != predicciones.lda)
```

Obteniendo una tasa de error de 0,25.

Por otro lado, la tasa de error por validación cruzada es:

```
predicciones.lda.cv <- lda(clases ~ ., data = lirios1, prior=c(0.5, 0.5), CV=TRUE)$class
table(clases, predicciones.lda.cv)

      predicciones.lda.cv
clases  0  1
     0 38 12
     1 15 35

mean(clases != predicciones.lda.cv)
```

Donde obtenemos una tasa de error de 0,27.

Ahora, hacemos el mismo cálculo mediante la ecuación el ejercicio 4.

```
datos0 <- lirios[0:50,1:2]
datos1 <- lirios[51:100,1:2]

s0 <- cov(datos0)
s1 <- cov(datos1)

S <- 0.5*(s0 + s1)

m1 <- c()
for (i in 1:50){
  m1 <- append(m1, lirios[,1][i])
}

m2 <- c()
for (i in 1:50){
  m2 <- append(m2, lirios[,2][i])
}

medias0 <- c(mean(m1), mean(m2))

m21 <- c()
for (i in 51:100){
  m21 <- append(m21, lirios[,1][i])
}

m22 <- c()
for (i in 51:100){
  m22 <- append(m22, lirios[,2][i])
}

medias1 <- c(mean(m21), mean(m22))
```

```
distancias <- mahalanobis(medias0, medias1, S)
estimacion <- 1 - pnorm(sqrt(distancias)/2, mean=0, sd=1)
```

Donde obtenemos una tasa de error de 0,2858654, un valor cercano el obtenido mediante validación cruzada.

1.2.

Ahora, hacemos lo mismo pero haciendo uso de las cuatro variables del conjunto de datos.

En primer lugar, los coeficientes de la función discriminante de Fisher:

```
lirios2 <- data.frame(lirios[,0:4])
resultado.lda <- lda(clases ~ ., data=lirios2, prior = c(0.5, 0.5))
resultado.lda

Call:
lda(clases ~ ., data = lirios2, prior = c(0.5, 0.5))

Prior probabilities of groups:
  0   1
0.5 0.5

Group means:
  Sepal.Length Sepal.Width Petal.Length Petal.Width
0        5.936       2.770        4.260       1.326
1        6.588       2.974        5.552       2.026

Coefficients of linear discriminants:
                    LD1
Sepal.Length -0.9431178
Sepal.Width  -1.4794287
Petal.Length  1.8484510
Petal.Width   3.2847304
```

Se obtienen los siguientes coeficientes: $-0{,}9431178$ para la longitud del sépalo, $-1{,}4794287$ para la anchura del sépalo, 1,8484510 para la longitud del pétalo y 3,2847304 para la anchura del pétalo.

La probabilidad de error mediante riesgo empírico:

```
predicciones.lda <- predict(resultado.lda)$class
table(clases, predicciones.lda)

      predicciones.lda
clases  0  1
     0 48  2
     1  1 49

mean(clases != predicciones.lda)
```

Obteniendo una tasa de error de 0,03. Mientras que mediante validación cruzada:

```
predicciones.lda.cv <- lda(clases ~ ., data = lirios2, prior=c(0.5, 0.5), CV=TRUE)$class
table(clases, predicciones.lda.cv)

      predicciones.lda.cv
clases  0  1
     0 48  2
     1  1 49
mean(clases != predicciones.lda.cv)
```

Se obtiene también una tasa de error de 0,03.

Ahora, haremos el cálculo mediante la fórmula del ejercicio anterior.

```
datos0 <- lirios[0:50,1:4]
datos1 <- lirios[51:100,1:4]

s0 <- cov(datos0)
s1 <- cov(datos1)

S <- 0.5*(s0 + s1)

m1 <- c()
for (i in 1:50){
  m1 <- append(m1, lirios[,1][i])
}
```

```r
m2 <- c()
for (i in 1:50){
  m2 <- append(m2, lirios[,2][i])
}

m3 <- c()
for (i in 1:50){
  m3 <- append(m3, lirios[,3][i])
}

m4 <- c()
for (i in 1:50){
  m4 <- append(m4, lirios[,4][i])
}

medias0 <- c(mean(m1), mean(m2), mean(m3), mean(m4))

m21 <- c()
for (i in 51:100){
  m21 <- append(m21, lirios[,1][i])
}

m22 <- c()
for (i in 51:100){
  m22 <- append(m22, lirios[,2][i])
}

m23 <- c()
for (i in 51:100){
  m23 <- append(m23, lirios[,3][i])
}

m24 <- c()
```

```r
for (i in 51:100){
  m24 <- append(m24, lirios[,4][i])
}

medias1 <- c(mean(m21), mean(m22), mean(m23), mean(m24))

distancias <- mahalanobis(medias0, medias1, S)
estimacion <- 1 - pnorm(sqrt(distancias)/2, mean=0, sd=1)
```

Y, obtenemos una tasa de error de 0,02968814; de nuevo, cercano al obtenido previamente.