

# (KNN) K-vecinos mas cercanos

Oscar Elí Bonilla Morales

2022-05-26

## Ejercicio Penguins

- 1.- Descargar la matriz penguins.
- 2.- Copiar y pegar el script.
- 3.- Adaptar el script. (elige la semilla)
- 4.- Generar resultados (activar comandos)
- 5.- Responder las siguientes preguntas:
  - 5.1.- ¿Cual es numero optimo de k-vecinos cercanos?
  - 5.2.- ¿Cual es la cantidad de observaciones mal clasificadas?
  - 5.3.- ¿Cual es el ratio de mala clasificación (MR)?
- 6.- Generar el gráfico de buena y mala clasificacion.

```
#_____ kNN_____ #K-vecinos próximos: Matriz Penguins
```

```
library(readxl)
library(readxl)
penguins_1_ <- read_excel("PENGUIN/penguins (1).xlsx")
head(penguins_1_)
```

```
## # A tibble: 6 x 9
##   ID     especie isla     largo_pico_mm grosor_pico_mm largo_aleta_mm
##   <chr> <chr>   <chr>         <dbl>         <dbl>         <dbl>
## 1 i1    Adelie  Torgersen     39.1          18.7          181
## 2 i2    Adelie  Torgersen     39.5          17.4          186
## 3 i3    Adelie  Torgersen     40.3          18            195
## 4 i4    Adelie  Torgersen     37.8          18.1          190
## 5 i5    Adelie  Torgersen     36.7          19.3          193
## 6 i6    Adelie  Torgersen     39.3          20.6          190
## # ... with 3 more variables: masa_corporal_g <dbl>, genero <chr>, año <dbl>
```

```
library(MASS)
```

## Cargar los datos Penguins

```
Z<-as.data.frame(penguins_1_)
colnames(Z)
```

```
## [1] "ID"          "especie"      "isla"         "largo_pico_mm"
## [5] "grosor_pico_mm" "largo_aleta_mm" "masa_corporal_g" "genero"
## [9] "año"
```

## Definir la matriz de datos y la variable respuesta

### Con las clasificaciones

```
x<-Z[,4:5]  
y<-Z[,8]
```

### Se definen las variables y observaciones

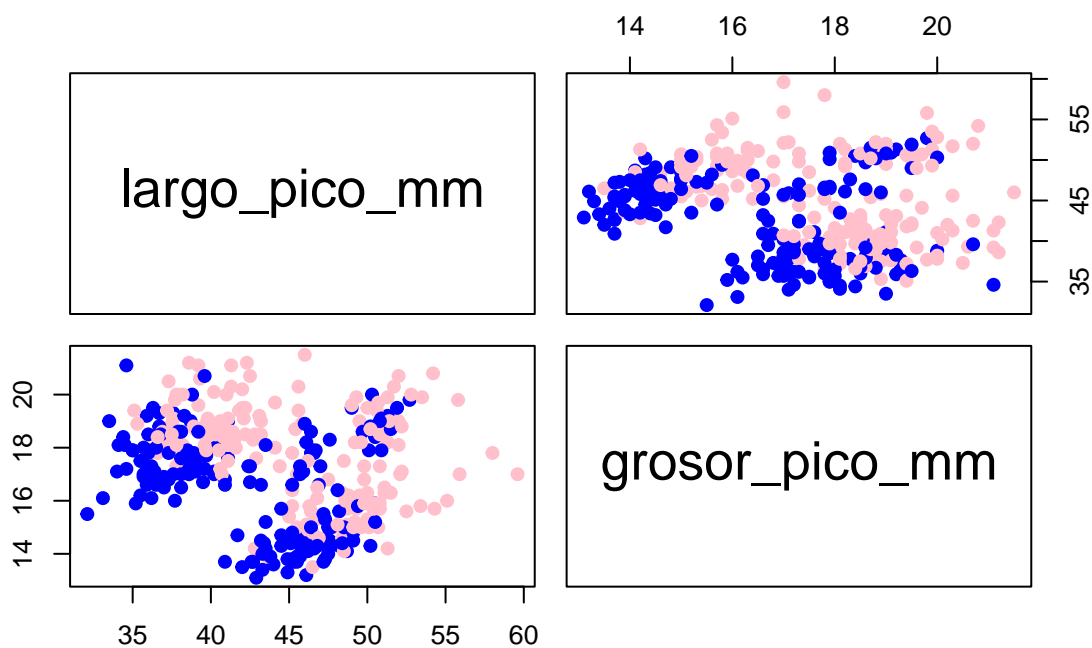
```
n<-nrow(x)  
p<-ncol(x)
```

### Grafico scatter plot

### Creacion de un vector de colores

```
head(y)  
  
## [1] "male" "female" "female" "female" "female" "male"  
  
col.peng<- c("Blue", "pink")  
col.peng  
  
## [1] "Blue" "pink"  
  
pairs(x, main="Data set Penguins, Male(Blue), Female(Pink)",  
      pch=19, col = col.peng)
```

### Data set Penguins, Male(Blue), Female(Pink)



## kNN

```
library(class)
```

Se fija una “semilla” para tener valores iguales

```
set.seed(1002)
```

Creación de los ciclos para k=1 hasta k=20

Selecciona el valor de k que tenga el error mas bajo.

Inicialización de una lista vacia de tamaño 20

```
knn.class<-vector(mode="list",length=20)  
knn.tables<-vector(mode="list", length=20)
```

## Clasificaciones erroneas

```
knn.mis<-matrix(NA, nrow=20, ncol=1)  
knn.mis
```

```
##      [,1]  
## [1,] NA  
## [2,] NA  
## [3,] NA  
## [4,] NA  
## [5,] NA  
## [6,] NA  
## [7,] NA  
## [8,] NA  
## [9,] NA  
## [10,] NA  
## [11,] NA  
## [12,] NA  
## [13,] NA  
## [14,] NA  
## [15,] NA  
## [16,] NA  
## [17,] NA  
## [18,] NA  
## [19,] NA  
## [20,] NA  
  
for(k in 1:20){  
  knn.class[[k]]<-knn.cv(x,y,k=k)  
  knn.tables[[k]]<-table(y,knn.class[[k]])  
  # la suma de las clasificaciones menos las correctas  
  knn.mis[k]<- n-sum(y==knn.class[[k]])  
}
```

```
knn.mis
```

```
##      [,1]
## [1,]  73
## [2,]  76
## [3,]  67
## [4,]  64
## [5,]  60
## [6,]  57
## [7,]  58
## [8,]  55
## [9,]  59
## [10,] 56
## [11,] 61
## [12,] 56
## [13,] 55
## [14,] 58
## [15,] 56
## [16,] 59
## [17,] 56
## [18,] 57
## [19,] 58
## [20,] 59
```

## Numero optimo de k-vecinos

```
which(knn.mis==min(knn.mis))
```

```
## [1]  8 13
```

```
knn.tables[[6]]
```

```
##
## y      female male
## female  150   24
## male    33  137
```

## Nota

Al realizar el análisis encontramos que los valores k son k= 6 y 13, recordemos que se toma el menor para continuar con el análisis

## Se señala el k mas eficiente

```
k.opt<-8
```

```
knn.cv.opt<-knn.class[[k.opt]]
head(knn.cv.opt)
```

```
## [1] male  female male  female female male
## Levels: female male
```

## Tabla de contingencia con las clasificaciones buenas y malas

```
knn.tables[[k.opt]]
```

```
##  
## y          female male  
## female    147    27  
## male      28    142
```

## Cantidad de observaciones mal clasificadas

```
knn.mis[k.opt]
```

```
## [1] 55
```

- *Cantidad de observaciones mal clasificada = 55*

## Error de clasificacion (MR)

```
knn.mis[k.opt]/n
```

```
## [1] 0.1598837
```

- *¿Cual es el ratio de mala clasificación (MR)? = 0.1598*

## Grafico de clasificaciones correctas y erroneas

```
col.knn.peng<-c( "blue", "pink")[1*(y==knn.cv.opt)+1]  
pairs(x, main="Clasificación kNN de Penguin",  
      pch=19, col=col.knn.peng)
```

## Clasificación kNN de Penguin

