

R-PL4

Gabriel López, Sergio Sanz, Álvaro Zamorano

26 de noviembre de 2019

1. Ejercicio realizado en clase.

1.1. K-vecinos

A partir del siguiente conjunto de calificaciones académicas, formados por dos notas: teoría y laboratorio, que tendrán valores entre 0 y 5, realizar un análisis de detección de datos anómalos utilizando el algoritmo **K-vecinos**.

Alumno	Teoría	Laboratorio
A1	4	4
A2	4	3
A3	5	5
A4	1	1
A5	5	4

En primer lugar se introducirán los datos en forma de matriz y se hará la traspuesta de esta.

```
> muestra = matrix(c(4,4,4,3,5,5,1,1,5,4),2,5)
> muestra = t(muestra)
```

En segundo lugar, calculamos las distancias euclídeas entre todos los puntos y los almacenamos en una matriz. El cálculo de las distancias lo realizamos mediante la función **as.matrix**.

```
> distancias = as.matrix(dist(muestra))
> distancias = matrix(distancias,5,5)
```

En tercer lugar, ordenamos las columnas de la matriz de distancias por los valores de cada una de las filas de menor a mayor. Se tiene en cuenta el tercer vecino, por lo que debemos estudiar la cuarta fila de la matriz ya que se tiene en cuenta la distancia de un punto consigo mismo. Después se muestran las muestras cuyo suceso es anómalo o outlier.

```

> for(i in 1:5){
+   distancias[,i] = sort(distancias[,i])
+   if(distancias[4,i] > 2.5) {
+     print(i);print("es un dato anómalo o outlier")
+   }
+ }

[1] 4
[1] "es un dato anómalo o outlier"

```

1.2. Caja y bigotes.

A partir del siguiente conjunto de valores de resistencia y densidad para diferentes tipos de hormigón, se hará un análisis para la detección de outliers utilizando medidas de ordenación sobre la resistencia con el método de **Caja y Bigotes**.

Hormigón	Resistencia	Densidad
H1	3	2
H2	3.5	12
H3	4.7	4.1
H4	5.2	4.9
H5	7.1	6.1
H6	6.2	5.2
H7	14	5.3

En primer lugar, introducimos los datos en una matriz y lo convertimos a dataframe para que el trabajo con las columnas sea más cómodo.

```

> muestra = t(matrix(c(3,2,3.5,12,4.7,4.1,5.2,4.9,7.1,6.1,6.2,5.2,14,5.3),2,7,dimnames = 1
> muestra = data.frame(muestra)

```

En segundo lugar:

1. Se determina el grado de outlier de forma arbitraria. El valor dado es 1.5.
2. Se ordenan los datos y se obtienen los cuartiles.

```

> cuar1 <- quantile(muestra$r,0.25)
> cuar3 <- quantile(muestra$r,0.75)

```

3. Se calculan los límites del intervalo.

```

> int = c(cuar1 - 1.5 * (cuar3 - cuar1),cuar3 + 1.5*(cuar3-cuar1))

```

4. Se identifican como outliers los valores que quedan fuera del intervalo.

```

> for(i in 1:length(muestra$r)){
+   if(muestra$r[i] < int[1] || muestra$r[i] > int[2]){
+     print("El suceso "); print(i);print(" con valor ");print(muestra$r[i]);
+   }
+ }

[1] "El suceso "
[1] 7
[1] " con valor "
[1] 14
[1] " es un suceso anomalo o outiler"

```

Por último, mostramos el diagrama de caja y bigotes donde se identifica el valor anómalo.

```

> source("./Funciones/Bigotes.R")
> bigotes(muestra$r,"bigotes.png",1.5)

```

