



FUNDAMENTOS DE LA CIENCIA DE DATOS

---

## Prueba de Laboratorio 2 (PL2)

---

**Jorge Revenga Martín de Vidales**  
**Ángel Salgado Aldao**  
**Adrián García**

Grado en Ingeniería Informática  
Universidad de Alcalá

5 de diciembre de 2023

# Índice

<b>1. Introducción - Consideraciones previas</b>	<b>2</b>
1.1. Uso de RStudio . . . . .	2
1.2. Introducción de datos de Excel/CSV en R . . . . .	2
<b>2. Ejercicios con ayuda del profesor</b>	<b>4</b>
2.1. Análisis de clasificación no supervisada . . . . .	4
2.1.1. k-Means . . . . .	4
2.1.2. Clusterización Jerárquica Aglomerativa . . . . .	12
2.2. Análisis de clasificación supervisada . . . . .	13
2.2.1. Árboles de decisión . . . . .	13
2.2.2. Regresión . . . . .	13
<b>3. Ejercicios de forma autónoma</b>	<b>14</b>
3.1. Análisis de clasificación no supervisada . . . . .	14
3.1.1. K-means . . . . .	14
3.1.2. Clusterización Jerárquica Aglomerativa . . . . .	14
3.2. Análisis de clasificación supervisada . . . . .	17
3.2.1. Árboles de decisión . . . . .	17
3.2.2. Regresión . . . . .	17

# 1. Introducción - Consideraciones previas

## 1.1. Uso de RStudio

Para utilizar una función en R se escribe el nombre de la función, seguido de los parámetros de entrada entre paréntesis e.g.: `función(parámetros)`

- Función `contributors()`: Muestra los creadores del programa (R)
- Función `help()`: Abre un HTML con información sobre la función `help()` o de la función entre paréntesis de haberla. Para todas las funciones que programemos (para todas las que existan) en R debe poder usarse la función `help()`.

En el archivo HTML se distinguen varios elementos:

- función {paquete}: la función de la que se obtiene información seguida del paquete al que pertenece.
  - Description: descripción de la función.
  - Usage: aparece la función y todos los argumentos que se le pueden introducir.
  - Arguments: Explicación de los argumentos o parámetros.
  - Details: Detalles adicionales de la función.
  - Offline help: Ayuda sin conexión.
  - Note: Nota del autor.
  - References: Referencias.
  - Examples: Ejemplos de uso de la función.
- Función `getwd()` se utiliza para obtener el directorio de trabajo actual (working directory).
  - Función `setwd("C:/...")` permite cambiar el nuevo directorio de trabajo en el que queramos trabajar.
  - `help.start()`: Manda a un compendio de todas las ayudas disponibles para trabajar con R.
  - Función `list.files()`: Muestra todos los archivos en el directorio. `dir()` hace lo mismo.

## 1.2. Introducción de datos de Excel/CSV en R

Para poder leer archivos .xlsx y .csv seguimos un par de pasos:

- Archivos .csv:
  - Primero se crea el archivo .csv y se introducen los datos por filas (los datos de la fila i, se escriben en la celda (i,1)) y separando dichos datos con un delimitador (", ' ' o "; ' ').

- Después se usa la función `read.csv('ruta del archivo', sep = ';')`, la cual es parte del paquete base de R, para leer los datos del archivo. Recibe como parámetros la ruta del archivo y el delimitador que usa para separar los datos.
- Archivos .xlsx:
- Primero se crea el archivo .xlsx y se introducen los datos en forma de matriz, escribiendo cada dato en una celda.
  - Después se carga la librería *openxlsx* y se usa la función `read_excel('ruta del archivo')`, introduciendo la ruta del archivo por parámetro.

## 2. Ejercicios con ayuda del profesor

Realización de cuatro ejercicios con ayuda del profesor en los que se van a realizar, utilizando el entorno R, dos análisis de clasificación no supervisada y dos análisis de clasificación supervisada, aplicando todos los conceptos teóricos vistos en cada lección.

### 2.1. Análisis de clasificación no supervisada

#### 2.1.1. k-Means

El primer conjunto de datos, que se empleará para realizar el análisis de clasificación no supervisada con k-Means, estará formado por las siguientes 8 calificaciones de estudiantes: 1.{4, 4}; 2.{3, 5}; 3.{1, 2}; 4.{5, 5}; 5.{0, 1}; 6.{2, 2}; 7.{4, 5}; 8.{2, 1}, donde las características de las calificaciones son: {Teoría, Laboratorio}.

**Solución:**

```
■ m<-matrix(c(4,4, 3,5, 1,2, 5,5, 0,1, 2,2, 4,5, 2,1),2,8): Explicacion
```

```
> m<-matrix(c(4,4, 3,5, 1,2, 5,5, 0,1, 2,2, 4,5, 2,1),2,8)
> (m<-t(m))
```

```
      [,1] [,2]
[1,]    4    4
[2,]    3    5
[3,]    1    2
[4,]    5    5
[5,]    0    1
[6,]    2    2
[7,]    4    5
[8,]    2    1
```

```
■ c<-matrix(c(0,1,2,2),2,2): Explicacion
```

```
> c<-matrix(c(0,1,2,2),2,2)
> (c<-t(c))
```

```
      [,1] [,2]
[1,]    0    1
[2,]    2    2
```

```
■ (clasificacionns=(kmeans(m,c,4))): Explicacion
```

```
> (clasificacionns=(kmeans(m,c,4)))
```

```
K-means clustering with 2 clusters of sizes 4, 4
```

```
Cluster means:
```

```
      [,1] [,2]
1 1.25 1.50
2 4.00 4.75
```

Clustering vector:

```
[1] 2 2 1 2 1 1 2 1
```

Within cluster sum of squares by cluster:

```
[1] 3.75 2.75
(between_SS / total_SS = 84.8 %)
```

Available components:

```
[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
[6] "betweenss"    "size"         "iter"         "ifault"
```

■ : Explicacion

```
> (m=cbind(clasificacionns$cluster,m))
```

```
      [,1] [,2] [,3]
[1,]     2     4     4
[2,]     2     3     5
[3,]     1     1     2
[4,]     2     5     5
[5,]     1     0     1
[6,]     1     2     2
[7,]     2     4     5
[8,]     1     2     1
```

■ : Explicacion

```
> mc1=subset(m,m[,1]==1)
> mc2=subset(m,m[,1]==2)
> mc1
```

```
      [,1] [,2] [,3]
[1,]     1     1     2
[2,]     1     0     1
[3,]     1     2     2
[4,]     1     2     1
```

```
> mc2
```

```
      [,1] [,2] [,3]
[1,]     2     4     4
[2,]     2     3     5
[3,]     2     5     5
[4,]     2     4     5
```

■ : Explicacion

```
> (mc1=mc1[,-1])
```

```

      [,1] [,2]
[1,]    1    2
[2,]    0    1
[3,]    2    2
[4,]    2    1

```

```
> (mc2=mc2[, -1])
```

```

      [,1] [,2]
[1,]    4    4
[2,]    3    5
[3,]    5    5
[4,]    4    5

```

■ : Explicacion

```
> install.packages("LearnClust")
```

```
package 'LearnClust' successfully unpacked and MD5 sums checked
```

```
The downloaded binary packages are in
```

```
C:\Users\Jorge\AppData\Local\Temp\RtmpcxDkm3\downloaded_packages
```

```
> library(LearnClust)
```

```
> search()
```

```

[1] ".GlobalEnv"          "package:LearnClust" "package:magick"
[4] "package:arules"       "package:Matrix"     "package:stats"
[7] "package:graphics"    "package:grDevices"  "package:utils"
[10] "package:datasets"    "package:methods"    "Autoloads"
[13] "package:base"

```

■ : Explicacion

```
> m<-matrix(c(0.89,2.94, 4.36,5.21, 3.75,1.12, 6.25,3.14, 4.1,1.8, 3.9,4.27),2,6)
```

```
> (m<-t(m))
```

```

      [,1] [,2]
[1,] 0.89 2.94
[2,] 4.36 5.21
[3,] 3.75 1.12
[4,] 6.25 3.14
[5,] 4.10 1.80
[6,] 3.90 4.27

```

■ : Explicacion

```
> agglomerativeHC(m,'EUC','MIN')
```

```
$dendrogram
Number of objects: 6
```

```
$clusters
$clusters[[1]]
      X1  X2
1 0.89 2.94
```

```
$clusters[[2]]
      X1  X2
1 4.36 5.21
```

```
$clusters[[3]]
      X1  X2
1 3.75 1.12
```

```
$clusters[[4]]
      X1  X2
1 6.25 3.14
```

```
$clusters[[5]]
      X1  X2
1 4.1 1.8
```

```
$clusters[[6]]
      X1  X2
1 3.9 4.27
```

```
$clusters[[7]]
      X1  X2
1 3.75 1.12
2 4.10 1.80
```

```
$clusters[[8]]
      X1  X2
1 4.36 5.21
2 3.90 4.27
```

```
$clusters[[9]]
      X1  X2
1 3.75 1.12
2 4.10 1.80
3 4.36 5.21
4 3.90 4.27
```

```
$clusters[[10]]
      X1  X2
```



```

1 6.25 3.14
2 3.75 1.12
3 4.10 1.80
4 4.36 5.21
5 3.90 4.27

```

```

$clusters[[11]]
      X1  X2
1 0.89 2.94
2 6.25 3.14
3 3.75 1.12
4 4.10 1.80
5 4.36 5.21
6 3.90 4.27

```

```

$groupedClusters
  cluster1 cluster2
1         3         5
2         2         6
3         7         8
4         4         9
5         1        10

```

■ : Explicacion

```
> agglomerativeHC.details(m, 'EUC', 'MIN')
```

```

[[1]]
      [,1] [,2] [,3]
[1,] 0.89 2.94    1

```

```

[[2]]
      [,1] [,2] [,3]
[1,] 4.36 5.21    1

```

```

[[3]]
      [,1] [,2] [,3]
[1,] 3.75 1.12    1

```

```

[[4]]
      [,1] [,2] [,3]
[1,] 6.25 3.14    1

```

```

[[5]]
      [,1] [,2] [,3]
[1,] 4.1  1.8    1

```

```

[[6]]

```

```

    [,1] [,2] [,3]
[1,]   3.9 4.27   1

```

```

    [,1] [,2] [,3] [,4] [,5] [,6]
[1,] 0.000000 4.146541 3.3899853 5.363730 3.4064204 3.290745
[2,] 4.146541 0.000000 4.1352388 2.803034 3.4198977 1.046518
[3,] 3.389985 4.135239 0.0000000 3.214094 0.7647876 3.153569
[4,] 5.363730 2.803034 3.2140940 0.000000 2.5333969 2.607566
[5,] 3.406420 3.419898 0.7647876 2.533397 0.0000000 2.478084
[6,] 3.290745 1.046518 3.1535694 2.607566 2.4780839 0.000000

```

```

    X1  X2
1 3.75 1.12
2 4.10 1.80

```

```

    [,1] [,2] [,3] [,4] [,5] [,6] [,7]
[1,] 0.000000 4.146541 0 5.363730 0 3.290745 3.389985
[2,] 4.146541 0.000000 0 2.803034 0 1.046518 3.419898
[3,] 0.000000 0.000000 0 0.000000 0 0.000000 0.000000
[4,] 5.363730 2.803034 0 0.000000 0 2.607566 2.533397
[5,] 0.000000 0.000000 0 0.000000 0 0.000000 0.000000
[6,] 3.290745 1.046518 0 2.607566 0 0.000000 2.478084
[7,] 3.389985 3.419898 0 2.533397 0 2.478084 0.000000

```

```

    X1  X2
1 4.36 5.21
2 3.90 4.27

```

```

    [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,] 0.000000 0 0 5.363730 0 0 3.389985 3.290745
[2,] 0.000000 0 0 0.000000 0 0 0.000000 0.000000
[3,] 0.000000 0 0 0.000000 0 0 0.000000 0.000000
[4,] 5.363730 0 0 0.000000 0 0 2.533397 2.607566
[5,] 0.000000 0 0 0.000000 0 0 0.000000 0.000000
[6,] 0.000000 0 0 0.000000 0 0 0.000000 0.000000
[7,] 3.389985 0 0 2.533397 0 0 0.000000 2.478084
[8,] 3.290745 0 0 2.607566 0 0 2.478084 0.000000

```

```

    X1  X2
1 3.75 1.12
2 4.10 1.80
3 4.36 5.21
4 3.90 4.27

```

```

    [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
[1,] 0.000000 0 0 5.363730 0 0 0 0 3.290745
[2,] 0.000000 0 0 0.000000 0 0 0 0 0.000000
[3,] 0.000000 0 0 0.000000 0 0 0 0 0.000000
[4,] 5.363730 0 0 0.000000 0 0 0 0 2.533397
[5,] 0.000000 0 0 0.000000 0 0 0 0 0.000000
[6,] 0.000000 0 0 0.000000 0 0 0 0 0.000000
[7,] 0.000000 0 0 0.000000 0 0 0 0 0.000000
[8,] 0.000000 0 0 0.000000 0 0 0 0 0.000000
[9,] 3.290745 0 0 2.533397 0 0 0 0 0.000000

```

```

      X1    X2
1 6.25 3.14
2 3.75 1.12
3 4.10 1.80
4 4.36 5.21
5 3.90 4.27

      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
[1,] 0.000000 0 0 0 0 0 0 0 0 3.290745
[2,] 0.000000 0 0 0 0 0 0 0 0 0.000000
[3,] 0.000000 0 0 0 0 0 0 0 0 0.000000
[4,] 0.000000 0 0 0 0 0 0 0 0 0.000000
[5,] 0.000000 0 0 0 0 0 0 0 0 0.000000
[6,] 0.000000 0 0 0 0 0 0 0 0 0.000000
[7,] 0.000000 0 0 0 0 0 0 0 0 0.000000
[8,] 0.000000 0 0 0 0 0 0 0 0 0.000000
[9,] 0.000000 0 0 0 0 0 0 0 0 0.000000
[10,] 3.290745 0 0 0 0 0 0 0 0 0.000000

      X1    X2
1 0.89 2.94
2 6.25 3.14
3 3.75 1.12
4 4.10 1.80
5 4.36 5.21
6 3.90 4.27

```

■ : Explicacion

```
> agglomerativeHC.details(m, 'EUC', 'MAX')
```

```
[[1]]
      [,1] [,2] [,3]
[1,] 0.89 2.94 1
```

```
[[2]]
      [,1] [,2] [,3]
[1,] 4.36 5.21 1
```

```
[[3]]
      [,1] [,2] [,3]
[1,] 3.75 1.12 1
```

```
[[4]]
      [,1] [,2] [,3]
[1,] 6.25 3.14 1
```

```
[[5]]
      [,1] [,2] [,3]
[1,] 4.1 1.8 1
```

[[6]]

[,1] [,2] [,3]

[1,] 3.9 4.27 1

[,1] [,2] [,3] [,4] [,5] [,6]

[1,] 0.000000 4.146541 3.3899853 5.363730 3.4064204 3.290745

[2,] 4.146541 0.000000 4.1352388 2.803034 3.4198977 1.046518

[3,] 3.389985 4.135239 0.0000000 3.214094 0.7647876 3.153569

[4,] 5.363730 2.803034 3.2140940 0.000000 2.5333969 2.607566

[5,] 3.406420 3.419898 0.7647876 2.533397 0.0000000 2.478084

[6,] 3.290745 1.046518 3.1535694 2.607566 2.4780839 0.000000

X1 X2

1 3.75 1.12

2 4.10 1.80

[,1] [,2] [,3] [,4] [,5] [,6] [,7]

[1,] 0.000000 4.146541 0 5.363730 0 3.290745 3.406420

[2,] 4.146541 0.000000 0 2.803034 0 1.046518 4.135239

[3,] 0.000000 0.000000 0 0.000000 0 0.000000 0.000000

[4,] 5.363730 2.803034 0 0.000000 0 2.607566 3.214094

[5,] 0.000000 0.000000 0 0.000000 0 0.000000 0.000000

[6,] 3.290745 1.046518 0 2.607566 0 0.000000 3.153569

[7,] 3.406420 4.135239 0 3.214094 0 3.153569 0.000000

X1 X2

1 4.36 5.21

2 3.90 4.27

[,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]

[1,] 0.000000 0 0 5.363730 0 0 3.406420 4.146541

[2,] 0.000000 0 0 0.000000 0 0 0.000000 0.000000

[3,] 0.000000 0 0 0.000000 0 0 0.000000 0.000000

[4,] 5.363730 0 0 0.000000 0 0 3.214094 2.803034

[5,] 0.000000 0 0 0.000000 0 0 0.000000 0.000000

[6,] 0.000000 0 0 0.000000 0 0 0.000000 0.000000

[7,] 3.406420 0 0 3.214094 0 0 0.000000 4.135239

[8,] 4.146541 0 0 2.803034 0 0 4.135239 0.000000

X1 X2

1 6.25 3.14

2 4.36 5.21

3 3.90 4.27

[,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]

[1,] 0.000000 0 0 0 0 0 3.406420 0 5.363730

[2,] 0.000000 0 0 0 0 0 0.000000 0 0.000000

[3,] 0.000000 0 0 0 0 0 0.000000 0 0.000000

[4,] 0.000000 0 0 0 0 0 0.000000 0 0.000000

[5,] 0.000000 0 0 0 0 0 0.000000 0 0.000000

[6,] 0.000000 0 0 0 0 0 0.000000 0 0.000000

[7,] 3.40642 0 0 0 0 0 0.000000 0 4.135239

[8,] 0.000000 0 0 0 0 0 0.000000 0 0.000000

[9,] 5.36373 0 0 0 0 0 4.135239 0 0.000000

```

      X1    X2
1 0.89 2.94
2 3.75 1.12
3 4.10 1.80
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]      [,9]     [,10]
[1,]    0    0    0    0    0    0    0    0    0 0.00000 0.00000
[2,]    0    0    0    0    0    0    0    0    0 0.00000 0.00000
[3,]    0    0    0    0    0    0    0    0    0 0.00000 0.00000
[4,]    0    0    0    0    0    0    0    0    0 0.00000 0.00000
[5,]    0    0    0    0    0    0    0    0    0 0.00000 0.00000
[6,]    0    0    0    0    0    0    0    0    0 0.00000 0.00000
[7,]    0    0    0    0    0    0    0    0    0 0.00000 0.00000
[8,]    0    0    0    0    0    0    0    0    0 0.00000 0.00000
[9,]    0    0    0    0    0    0    0    0    0 0.00000 5.36373
[10,]   0    0    0    0    0    0    0    0    0 5.36373 0.00000
      X1    X2
1 6.25 3.14
2 4.36 5.21
3 3.90 4.27
4 0.89 2.94
5 3.75 1.12
6 4.10 1.80

```

■ : Explicacion

>

### 2.1.2. Clusterización Jerárquica Aglomerativa

El segundo conjunto de datos, que se empleará para realizar el análisis de clasificación no supervisada con Clusterización Jerárquica Aglomerativa, estará formado por 6 calificaciones de estudiantes: 1.{0.89, 2.94}; 2.{4.36, 5.21}; 3.{3.75, 1.12}; 4.{6.25, 3.14}; 5.{4.1, 1.8}; 6.{3.9, 4.27}.

### Solución

## 2.2. Análisis de clasificación supervisada

### 2.2.1. Árboles de decisión

El tercer conjunto de datos, que se empleará para realizar el análisis de clasificación supervisada utilizando árboles de decisión, estará formado por las siguientes 9 calificaciones de estudiantes: 1. {A,A,B,Ap}; 2. {A,B,D,Ss}; 3. {D,D,C,Ss}; 4. {D,D,A,Ss}; 5. {B,C,B,Ss}; 6. {C,B,B,Ap}; 7. {B,B,A,Ap}; 8. {C,D,C,Ss}; 9. {B,A,C,Ss}, donde las características de las calificaciones son: {Teoría, Laboratorio, Prácticas, Calificación Global).

### 2.2.2. Regresión

El cuarto conjunto de datos, que se empleará para realizar el análisis de clasificación supervisada utilizando regresión, estará formado por los siguientes 4 radios ecuatoriales y densidades de los planetas interiores: {Mercurio,2.4,5.4; Venus,6.1,5.2; Tierra,6.4,5.5; Marte,3.4,3.9}

### 3. Ejercicios de forma autónoma

Realización de cuatro ejercicios de forma autónoma por cada grupo de estudiantes en los que se van a realizar, utilizando el entorno R, dos análisis de clasificación no supervisada y dos análisis de clasificación supervisada, aplicando todos los conceptos teóricos vistos en cada lección.

#### 3.1. Análisis de clasificación no supervisada

##### 3.1.1. K-means

El primer conjunto de datos, que se empleará para realizar el análisis de clasificación no supervisada con K-means, estará formado por los siguientes 15 valores de velocidades de respuesta y temperaturas normalizadas de un microprocesador {Velocidad, Temperatura}: 1.{3.5, 4.5}; 2.{0.75, 3.25}; 3.{0, 3}; 4.{1.75, 0.75}; 5.{3, 3.75}; 6.{3.75, 4.5}; 7.{1.25, 0.75}; 8.{0.25, 3}; 9.{3.5, 4.25}; 10.{1.5, 0.5}; 11.{1, 1}; 12.{3, 4}; 13.{0.5, 3}; 14.{2, 0.25}; 15.{0, 2.5}. Del análisis visual de los datos se ha concluido que hay una alta probabilidad que sean tres clusters.

**Solución:**

##### 3.1.2. Clusterización Jerárquica Aglomerativa

El segundo conjunto de datos, que se empleará para realizar el análisis de clasificación no supervisada con Clusterización Jerárquica Aglomerativa, será el mismo que el utilizado en el ejercicio anterior, por lo tanto estará formado por los siguientes 15 valores de velocidades de respuesta y temperaturas normalizadas de un microprocesador {Velocidad, Temperatura}: 1.{3.5, 4.5}; 2.{0.75, 3.25}; 3.{0, 3}; 4.{1.75, 0.75}; 5.{3, 3.75}; 6.{3.75, 4.5}; 7.{1.25, 0.75}; 8.{0.25, 3}; 9.{3.5, 4.25}; 10.{1.5, 0.5}; 11.{1, 1}; 12.{3, 4}; 13.{0.5, 3}; 14.{2, 0.25}; 15.{0, 2.5}. Del análisis visual de los datos se ha concluido que hay una alta probabilidad que sean tres clusters.

**Solución:**

- `datos <- read.csv("Datos2.2.csv")`: Se leen los datos del archivo .csv y se guardan en la variable "datos".
- `(datos<-t(datos))`: Se hace la traspuesta de la matriz datos para un mejor tratamiento.
- `(clMin<-agglomerative_clustering(datos,"single",FALSE,FALSE))`: Clusterización usando como definición de proximidad, la distancia entre los puntos más cercanos de los clusters.
  - Parámetros:
    - `datos`: Variable que contiene los puntos a evaluar.
    - `"single"`: Algoritmo de clusterización, en este caso usando minimum linkage.
    - `details="FALSE"`: Booleano para determinar si se imprimen o no los logs con la explicación de todo el proceso.

- `waiting="FALSE"`: Booleano para determinar si tiene que esperar input del usuario para ir imprimiendo los logs.
- Salida: Matriz con los clusters y sus distancias.
- Explicación: Realiza un análisis jerárquico aglomerativo de los datos introducidos por parámetro. Repite una serie de pasos hasta que solo quede un cluster. Inicialmente cada punto se asigna a su propio cluster y se calcula la matriz de distancias entre ellos. Después se calcula la proximidad entre dos clusters con la distancia entre los 2 puntos más cercanos de dichos clusters (falta formula)

```
> library("clustlearn")
> datos <- read.csv("Datos2.2.csv")
> (datos<-t(datos))
```

```
      [,1]
X3.5  4.50
X0.75  3.25
X0     3.00
X1.75  0.75
X3     3.75
X3.75  4.50
X1.25  0.75
X0.25  3.00
X3.5.1 4.25
X1.5   0.50
X1     1.00
X3.1   4.00
X0.5   3.00
X2     0.25
X0.1   2.50
```

```
> (clMin<-agglomerative_clustering(datos,"single",FALSE,FALSE))
```

```
Cluster method : single
Distance       : Euclidean
Number of objects: 15
```

- `(clMax<-agglomerative_clustering(datos,"complete",FALSE,FALSE))`: Clusterización usando como definición de proximidad, la distancia que haya entre los dos puntos más lejanos de los clusters.

- Parámetros:
  - `datos`: Variable que contiene los puntos a evaluar.
  - `"single"`: Algoritmo de clusterización, en este caso usando maximum linkage.
  - `details="FALSE"`: Booleano para determinar si se imprimen o no los logs con la explicación de todo el proceso.
  - `waiting="FALSE"`: Booleano para determinar si tiene que esperar input del usuario para ir imprimiendo los logs.
- Salida: Matriz con los clusters y sus distancias.



- Explicación: Realiza un análisis jerárquico aglomerativo de los datos introducidos por parámetro. Repite una serie de pasos hasta que solo quede un cluster. Inicialmente cada punto se asigna a su propio cluster y se calcula la matriz de distancias entre ellos. Después se calcula la proximidad entre dos clusters con la distancia entre los puntos más lejanos de dichos clusters (falta formula)

```
> install.packages("clustlearn")
> library("clustlearn")
> datos <- read.csv("Datos2.2.csv")
> (datos<-t(datos))

      [,1]
X3.5  4.50
X0.75  3.25
X0     3.00
X1.75  0.75
X3     3.75
X3.75  4.50
X1.25  0.75
X0.25  3.00
X3.5.1 4.25
X1.5   0.50
X1     1.00
X3.1   4.00
X0.5   3.00
X2     0.25
X0.1   2.50

> (clMin<-agglomerative_clustering(datos,"complete",FALSE,FALSE))

Cluster method : complete
Distance       : Euclidean
Number of objects: 15
```

- (clAvg<-agglomerative\_clustering(datos,"average",FALSE,FALSE)): Clusterización usando como definición de proximidad, la media de distancias entre todas las parejas de puntos de los dos clusters.

- Parámetros:

- `datos`: Variable que contiene los puntos a evaluar.
- `.average`: Algoritmo de clusterización, en este caso usando average linkage.
- `details="FALSE"`: Booleano para determinar si se imprimen o no los logs con la explicación de todo el proceso.
- `waiting="FALSE"`: Booleano para determinar si tiene que esperar input del usuario para ir imprimiendo los logs.

- Salida: Matriz con los clusters y sus distancias.

- Explicación: Realiza un análisis jerárquico aglomerativo de los datos introducidos por parámetro. Repite una serie de pasos hasta que solo quede un cluster. Inicialmente cada punto se asigna a su propio cluster y se calcula la matriz de

distancias entre ellos. Después se calcula la proximidad entre dos clusters con la media de todos los puntos de ambos (falta formula)

```
> library("clustlearn")
> datos <- read.csv("Datos2.2.csv")
> (datos<-t(datos))

      [,1]
X3.5  4.50
X0.75  3.25
X0     3.00
X1.75  0.75
X3     3.75
X3.75  4.50
X1.25  0.75
X0.25  3.00
X3.5.1 4.25
X1.5   0.50
X1     1.00
X3.1   4.00
X0.5   3.00
X2     0.25
X0.1   2.50

> (clMin<-agglomerative_clustering(datos,"average",FALSE,FALSE))

Cluster method    : average
Distance          : Euclidean
Number of objects: 15
```

## 3.2. Análisis de clasificación supervisada

### 3.2.1. Árboles de decisión

El tercer conjunto de datos, que se empleará para realizar el análisis de clasificación supervisada utilizando árboles de decisión, estará formado por el siguiente conjunto de 10 sucesos constituidos por los valores de cuatro características de vehículos: 1.{B,4,5,Coche}; 2.{A,2,2,Moto}; 3.{N,2,1,Bicicleta}; 4.{B,6,4,Camión}; 5.{B,4,6,Coche}; 6.{B,4,4,Coche}; 7.{N,2,2,Bicicleta}; 8.{B,2,1,Moto}; 9.{B,6,2,Camión}; 10.{N,2,1,Bicicleta}, donde las características de cada suceso son: {TipoCarnet, NúmeroRuedas, NúmeroPasajeros, TipoVehículo}. Se debe clasificar el tipo de vehículo en función del resto de características. TipoCarnet, es el tipo de carnet necesario para conducir el vehículo.

### 3.2.2. Regresión

El cuarto conjunto de datos, que se empleará para realizar el análisis de clasificación supervisada utilizando regresión, estará formado por los siguientes 4 subconjuntos de datos: 1.{10, 8.04; 8, 6.95; 13, 7.58; 9, 8.81; 11, 8.33; 14, 9.96; 6, 7.24; 4, 4.26; 12, 10.84; 7, 4.82; 5, 5.68}; 2.{10, 9.14; 8, 8.14; 13, 8.74; 9, 8.77; 11, 9.26; 14, 8.1; 6, 6.13; 4, 3.1; 12, 9.13; 7, 7.26; 5, 4.74}; 3.{10, 7.46; 8, 6.77; 13, 12.74; 9, 7.11; 11, 7.81; 14, 8.84; 6, 6.08; 4, 5.39; 12, 8.15; 7, 6.42; 5, 5.73}; 4.{8, 6.58; 8, 5.76; 8, 7.71; 8, 8.84; 8, 8.47; 8, 7.04; 8, 5.25;

19, 12.5; 8, 5.56; 8, 7.91; 8, 6.89}. Se deben calcular las rectas de regresión de los cuatro subconjuntos y sus parámetros de ajuste.

**Solución:** Algoritmo de minería de reglas de asociación (apriori): Su objetivo principal es descubrir patrones de asociación entre diferentes conjuntos de datos.

- Parámetros:

- **transacciones:** Lista de sucesos que conforman la muestra.
- **soporte:** Umbral mínimo de soporte que deben superar las asociaciones.
- **confianza:** Umbral mínimo de confianza que deben superar las asociaciones.

- Retorno:

- Explicación:

>