

Dendrograma

Ana Karen Martínez Marín

12/5/2022

Dendrograma

Es un diagrama de árbol que muestra los grupos que se forman al crear conglomerados de observaciones en cada paso y sus niveles de similitud.

Para esta práctica se ocupó una base de datos de la librería **cluster.datasets**

Se cargan las librerías que se necesitan.

```
library(cluster.datasets)
library(dendextend)
```

Base de datos

```
data("all.mammals.milk.1956")
AMM=all.mammals.milk.1956
```

Se usará el data.set de “**all.mammals.milk.1956**”, el cual contiene datos sobre la leche de diferentes especies de animales.

Revisión de la base de datos

Dimensión

```
dim(AMM)
```

```
## [1] 25 6
```

Esta base contiene 25 observaciones y 6 variables

- Datos faltantes

```
anyNA(AMM)
```

```
## [1] FALSE
```

La búsqueda arroja que no hay datos nulos, así que se prosigue con el análisis.

- Tipo de variables

```
str(AMM)
```

```
## 'data.frame': 25 obs. of 6 variables:
## $ name : chr "Horse" "Orangutan" "Monkey" "Donkey" ...
## $ water : num 90.1 88.5 88.4 90.3 90.4 87.7 86.9 82.1 81.9 81.6 ...
## $ protein: num 2.6 1.4 2.2 1.7 0.6 3.5 4.8 5.9 7.4 10.1 ...
## $ fat : num 1 3.5 2.7 1.4 4.5 3.4 1.7 7.9 7.2 6.3 ...
## $ lactose: num 6.9 6 6.4 6.2 4.4 4.8 5.7 4.7 2.7 4.4 ...
```

```
## $ ash : num 0.35 0.24 0.18 0.4 0.1 0.71 0.9 0.78 0.85 0.75 ...
```

En la base hay 5 variables numéricas y una de tipo carácter, en donde se encuentra registrado el nombre de los animales. En las numéricas está la cantidad de proteína, nivel de agua, grasa, lactosa y los minerales de la leche.

Cálculo de la matriz de distancias de Mahalanobis.

```
dist.AMM<-dist(AMM[,2:6])
```

Se calcula la distancia de Mahalanobis para las variables dos a la seis, que son variables numéricas.

Con la distancia de Mahalanobis se calcula la similitud que existe entre las variables teniendo en cuenta la correlación que hay entre ellas.

Redondeo

```
round(as.matrix(dist.AMM)[1:6, 1:6],3)
```

```
##      1      2      3      4      5      6
## 1 0.000 3.327 2.494 1.226 4.759 4.107
## 2 3.327 0.000 1.206 2.794 2.798 2.592
## 3 2.494 1.206 0.000 2.375 3.716 2.348
## 4 1.226 2.794 2.375 0.000 3.763 4.007
## 5 4.759 2.798 3.716 3.763 0.000 4.176
## 6 4.107 2.592 2.348 4.007 4.176 0.000
```

Se realiza un redondeo de los cálculos de la distancia de Mahalanobis y se convierte a una matriz, se proyecta e indica que sólo se usarán los primeros 6 individuos, así que se especifica la selección de las 6 filas y 6 columnas pertenecientes a dichos individuos.

Calculo del dendrograma

```
dend.AMM<-as.dendrogram(hclust(dist.AMM))
```

Se calcula el Dendrograma, donde se hace uso del método de agrupación por Clústers “**hclust**”, el cual ofrece una agrupación jerárquica.

Gráfico del dendrograma

Se hace un vector para las etiquetas que le asignarán al Dendrograma.

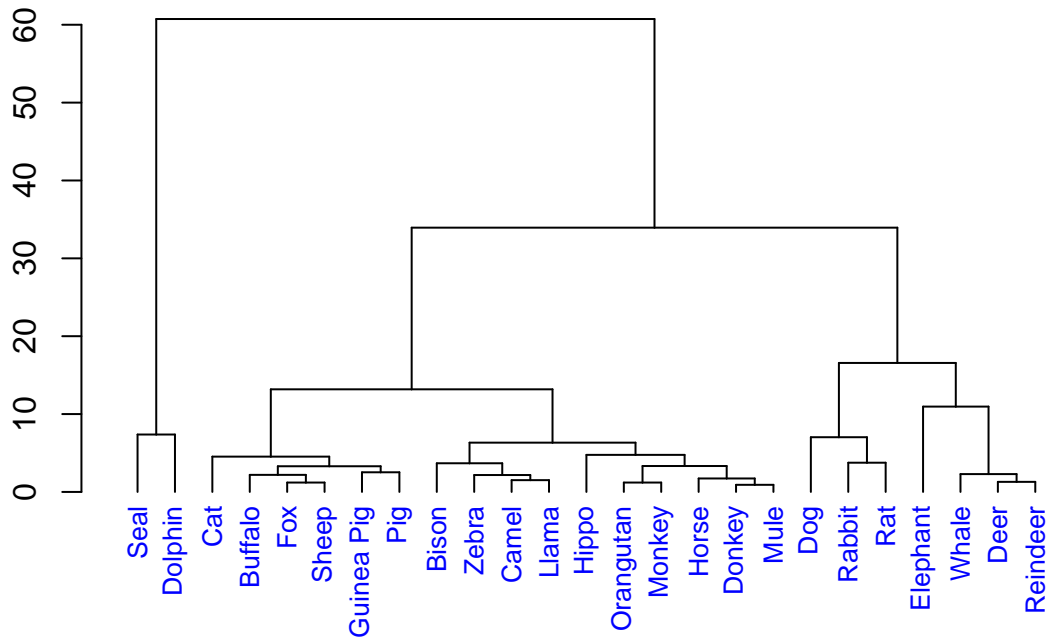
```
L=labels(dend.AMM)
```

```
labels(dend.AMM)=AMM$name[L]
```

Se grafica el Dendrograma añadiendo las etiquetas anteriormente realizadas.

```
dend.AMM %>%
  set(what="labels_col", "blue") %>% #Colores etiqueta
  set(what="labels_cex", 0.8) %>%
  plot(main="Dendrograma de mamíferos")
```

Dendrograma de mamíferos



Visualizando el gráfico permite observar cómo se forman dos grupos princialmente, en donde la leche de la foca y el delfín es muy distinta al resto de los animales. El otro grupo está sub-dividido por otros dos grandes grupos, donde éstos se van ramificando y clasificando por las características muy similares y específicas que tiene la leche de cada uno de los animales, por ejemplo la leche de la ballena(whale), reno(reindeer) y ciervo(deer), la leche es muy parecida.