

TAREA NÚMERO 4

1. (30 puntos) En este ejercicio usaremos la tabla de datos `EjemploAlgoritmosRecomendación.csv`, la cual contiene los promedios de evaluación de 100 personas que adquirieron los mismos productos o muy similares en la tienda AMAZON. La idea consiste en recomendar a un cliente los productos que ha comprado otra persona que pertenece al mismo clúster.
 - a) Ejecute un Clustering Jerárquico con la distancia euclídea y la agregación del Salto Máximo, Salto Mínimo, Promedio y Ward. Guarde la tabla de datos en el archivo `AlgoritmosRecomendación2.csv` con el clúster al que pertenece cada individuo para el caso de la agregación de Ward usando 2 clústeres.
 - b) “Corte” el árbol anterior usando 2 clústeres y la agregación de Ward, interprete los resultados usando interpretación usando gráficos de barras (Horizontal-Vertical) y usando gráficos tipo Radar.
 - c) Si se tienen 4 clústeres usando agregación de Ward ¿Qué productos recomendaría a Teresa, a Leo y a Justin?, es decir, ¿los productos que compra cuál otro cliente? Usando distancia euclídea ¿cuál es la mejor recomendación de compra que le podemos hacer a Teresa, a Leo y a Justin?
 - d) Construya un clustering jerárquico sobre las componentes principales del ACP.
2. (30 puntos) La tabla de datos `VotosCongresoUS.csv` la cual contiene 16 votos (y=Sí, n=No, NS=No votó) dados por los congresistas de Estados Unidos respecto a 16 temáticas diferentes, además en la primera columna aparece el partido al que pertenecen (Republicano o Demócrata).

- a) Ejecute una clasificación jerárquica sobre esta tabla de datos usando la función `daisy` ya que los datos son cualitativos. Use métrica `euclidean` y método `complete` (deje el resultado en la variable `jer`). Cargue los datos con la siguiente instrucción:

```
Datos <- read.csv("VotosCongresoUS.csv",header=TRUE, sep=",", dec=".")
```

- b) Luego “corte” el árbol usando 3 clústeres y ejecute el siguiente código:

```
grupo<-cutree(jer, k = 3)
NDatos<-cbind(Datos,grupo)

cluster<-NDatos$grupo
sel.cluster1<-match(cluster,c(1),0)
Datos.Cluster1<-NDatos[sel.cluster1>0,]
dim(Datos.Cluster1)

sel.cluster2<-match(cluster,c(2),0)
Datos.Cluster2<-NDatos[sel.cluster2>0,]
dim(Datos.Cluster2)
```

```
sel.cluster3<-match(cluster,c(3),0)
Datos.Cluster3<-NDatos[sel.cluster3>0,]
dim(Datos.Cluster3)
```

Explique qué hace el código anterior. Luego ejecute el siguiente código:

```
plot(Datos$Party,col=c(4,6),las=2,main="Party",xlab="Todos los Datos")
plot(Datos.Cluster1$Party,col=c(4,6),las=2,main="Party",xlab="Cluster-1")
plot(Datos.Cluster2$Party,col=c(4,6),las=2,main="Party",xlab="Cluster-2")
plot(Datos.Cluster3$Party,col=c(4,6),las=2,main="party",xlab="Cluster-3")
```

Con ayuda de los gráficos anteriores y tomando en cuenta el tamaño de cada cluster interprete los 3 clústeres formados.

3. (30 puntos) Realice un análisis similar al del ejercicio anterior con la tabla de datos `CompraBicicletas.csv`.
4. (10 puntos) Dada la siguiente matriz de disimilitudes entre cuatro individuos $A1$, $A2$, $A3$ y $A4$, construya “a mano” una Jerarquía Binaria usando la agregación del Salto Máximo y del Promedio, dibuje el dendograma en ambos casos:

$$D = \begin{pmatrix} 0 & & & \\ 5 & 0 & & \\ 2 & 1 & 0 & \\ 3 & 7 & 6 & 0 \end{pmatrix}$$

Verifique los resultados con `hclust`.

Entregables: Incluya en documento autoreproducible (HTML) todas las instrucciones y códigos **R** utilizados en cada ejercicio, incluya los resultados de los cálculos, los gráficos generados y las respuestas a las preguntas.