

Data Mining

Algoritmos de recomendación

Laboratorio II

Ejercicio 1:

Dada la siguiente matriz de distancias entre cuatro individuos A1, A2, A3 y A4, construya una clasificación binaria utilizando primero la agregación del salto máximo y luego la del promedio:

	A1	A2	A3	A4
A1	0	5	2	3
A2		0	1	7
A3			0	6
A4				0

Ejercicio 2:

Describa lo más detalladamente posible el siguiente código, incluyendo resultados o imágenes para aclarar su explicación.

```
wss <- function(d) {  
  sum(scale(d, scale = FALSE)^2)  
}  
wrap <- function(i, hc, x) {  
  cl <- cutree(hc, i)  
  spl <- split(x, cl)  
  wss <- sum(sapply(spl, wss))  
  wss  
}  
iris2 <- iris[, 1:4]  
cl <- hclust(dist(iris2), method = "ward.D")  
res <- sapply(seq.int(1, nrow(iris2)), wrap, h = cl, x = iris2)  
plot(seq_along(res), res, type = "b", pch = 19)  
plot(seq_along(res[1:50]), res[1:50], type = "o", pch = 19)
```

Ejercicio 3:

Trabajaremos con el archivo “EjemploAlgoritmosRecomendacion.csv” que contiene datos reales de las evaluaciones hechas por 100 clientes de la tienda Amazon que adquirieron los mismos productos o muy similares. La tabla de datos se muestra parcialmente a continuación:

	Velocidad Entrega	Precio	Durabilidad	Imagen Producto	Valor Educativo	Servicio Retorno	Tamano Paquete	Calidad Producto	Numero Estrellas
Adam	2,05	0,3	3,45	2,35	2,4	2,3	2,6	2,1	1,7
Anna	0,9	1,5	3,15	3,3	2,5	4	4,2	2,15	2,8
Bernard	1,7	2,6	2,85	3	4,3	2,7	4,1	2,6	3,3
Edward	1,35	0,5	3,55	2,95	1,8	2,3	3,9	1,95	1,7
Emilia	3	0,45	4,8	3,9	3,4	4,6	2,25	3,4	4,3
Fabian	0,95	1,65	3,95	2,4	2,6	1,9	4,85	2,2	3
Philip	2,3	1,2	4,75	3,3	3,5	4,5	3,8	2,9	3,1
Frank	0,65	2,1	3,1	2,55	2,8	2,2	3,45	2,15	2,9
Xavier	2,75	0,8	4,7	2,35	3,5	3	3,8	2,7	4,8
Gabriel	2	1,75	3,25	3	3,7	3,2	4,35	2,7	3,9
Marisol	1,2	0,8	4,4	2,4	2	2,8	2,9	2,15	1,7
Henry	1,95	1,1	4,55	2,3	3	2,5	4,15	2,5	3,2
Irene	1,4	0,7	4,05	1,9	2,1	1,4	3,3	2,2	2,4
Isabelle	1,85	0,75	4,3	2,85	2,7	3,7	3,35	2,5	2,3
Isidore	2,35	0,65	4,95	3,35	3	2,6	3,4	2,95	3,9
Joseph	1,7	1	4,85	2,35	2,7	1,7	2,4	2,35	3,4

1. Cargue la tabla de datos en R (recuerde definir adecuadamente los separadores);
2. Utilice análisis en componentes principales para detectar 2 y tres grupos. Caracterice los grupos utilizando el círculo de correlación. ¿Cuántas componentes utilizó? ¿Por qué?
3. Explore el uso de la función “hclust” y utilícela en la tabla de datos para generar un árbol de clasificación binaria;
4. Descubra qué función le permite realizar cortes en el árbol de modo de elegir el número de grupos a considerar;
5. Elija el número más adecuado de conglomerados a utilizar. ¿En base a qué criterio basó su decisión?
6. Agregue a la tabla de datos una columna nueva que contenga la información del grupo al que pertenece el individuo y guárdela con el nombre “EjemploAlgoritmosRecomendacionHCLUST.csv”;
7. Repita los puntos 2, 4 y 5 intercambiando “hclust” por “kmeans”;
8. En cada caso, grafique e interprete los centros de gravedad de cada conglomerado;
9. En base a los resultados obtenidos mediante los análisis anteriores, ¿qué productos recomendaría a Teresa, Leo y Justin? Justifique.