

Capítulo 4

Análisis de Conglomerados

4.1. Aspectos teóricos

1. Mostrar que las siguientes funciones:

$$a) \ d_{ij} = d(x_i, x_j) = \sum_{k=1}^p w_k |x_{ik} - x_{jk}|, \quad w_k > 0$$

$$b) \ d_{ij} = d(x_i, x_j) = [\sum_{k=1}^p |x_{ik} - x_{jk}|^2]^{1/2}$$

$$c) \ d_{ij} = d(x_i, x_j) = [\sum_{k=1}^p w_k (x_{ik} - x_{jk})^2]^{1/2}, \quad w_k > 0$$

$$d) \ d_{ij} = d(x_i, x_j) = \sum_{k=1}^p \frac{|x_{ik} - x_{jk}|}{(x_{ik} + x_{jk})}$$

$$e) \ d_{ij} = d(x_i, x_j) = [(x_i - x_j)' \Sigma^{-1} (x_i - x_j)]^{1/2}$$

para $i, j = 1, 2, \dots, n$; cumplen las propiedades especificadas para una métrica.

2. Recuerde que la distancia entre conglomerados está definida como:

$$d(R, PQ) = \delta_1 d(R, P) + \delta_2 d(R, Q) + \delta_3 d(P, Q) + \delta_4 |d(R, P) - d(R, Q)|$$

Simplificar esta definición utilizando los valores de los ponderadores para:

a) Enlace Simple

b) Enlace Completo

c) Enlace Promedio no ponderado

- d) Enlace Promedio ponderado
- e) Centroide
- f) Mediana
- g) Ward

4.2. Aspectos prácticos

3. *Algoritmos Jerárquicos aglomerativos* (Agglomerative Nesting - AGNES)

Considere el conjunto de $n = 8$ vectores bivariantes:

$$\begin{aligned}x_1 &= (1, 2)', & x_2 &= (1, 4)', & x_3 &= (2, 2)', \\x_4 &= (2, 4)', & x_5 &= (4, 2)', & x_6 &= (4, 4)', \\x_7 &= (5, 2)', & x_8 &= (5, 4)', & x_9 &= (6, 3)'\end{aligned}$$

- a) Graficar el diagrama de puntos
 - b) Construir la matriz de distancias utilizando la distancia de Mahalanobis
 - c) Construir las matrices de distancias reducida y la representación gráfica de la conformación de los conglomerados en cada paso utilizando segmentos, aplicando los algoritmos del:
 - i) Enlace Simple
 - ii) Enlace Completo
 - iii) Enlace Promedio no ponderado
 - iv) Enlace Promedio ponderado
 - v) Centroide
 - vi) Mediana
 - vii) Ward
 - d) Construir el dendograma para cada tipo de algoritmo utilizando en el ejercicio anterior. Comente los resultados obtenidos.
- ### 4. *Algoritmo Jerárquico divisivo* (Divisive Analysis - DIANA)
- Utilizando el conjunto de datos definido en 3:

- a) aplicar el algoritmo DIANA paso a paso
- b) graficar el dendograma resultante
- c) Comparar los conglomerados hallados, en relación a los obtenidos en la sección anterior.

5. ***Algoritmo de particionamiento*** (K-medias)

Utilizando el conjunto de datos definido en 3, aplicar el algoritmo de las k-medias considerando los siguientes parámetros:

- a) $k = 2$ con $\{5, 4, 3, 9\}$ y $\{1, 7, 2, 6, 8\}$ conglomerados iniciales
- b) $k = 3$ con $\{7, 8, 5\}$, $\{1, 3, 2\}$ y $\{9, 4, 6\}$ conglomerados iniciales

6. ***Algoritmo de particionamiento*** (K-medoides)

Utilizando el conjunto de datos definido en 3, aplicar el algoritmo de las k-medias considerando los siguientes parámetros:

- a) $k = 2$ con $\{5, 4, 3, 9\}$ y $\{1, 7, 2, 6, 8\}$ conglomerados iniciales
- b) $k = 3$ con $\{7, 8, 5\}$, $\{1, 3, 2\}$ y $\{9, 4, 6\}$ conglomerados iniciales