

1. ALGORITMO AGLOMERATIVO

Algoritmo 1

Pseudocódigo del agrupamiento jerárquico aglomerativo (AHC)

Entrada: D (conjunto de datos individuales)

Salida : Estructura jerárquica de clústeres

Iniciarizar $\ell \leftarrow 0$;

Iniciarizar el conjunto de clústeres $C_\ell \leftarrow D$;

mientras $|C_\ell| > 1$ **hacer**

Encontrar par $c_i, c_j \in C_\ell$ tal que $L(c_i, c_j)$ sea mínima;

Crear nuevo clúster $c_* \leftarrow c_i \cup c_j$;

Actualizar $C_{\ell+1} \leftarrow (C_\ell \setminus \{c_i, c_j\}) \cup \{c_*$ };

$\ell \leftarrow \ell + 1$;

devolver C_ℓ

1.1 Tipos de enlace

DEFINICIÓN 1: Función de enlace.

Una función de enlace es una función

$$L: \mathcal{P}(X) \times \mathcal{P}(X) \rightarrow \mathbb{R}$$

que asigna una distancia entre dos clústeres A y B, dada una distancia d entre puntos de X.

DEFINICIÓN 2: Enlace simple.

La distancia entre dos clústeres es la distancia mínima entre los puntos de ambos clústeres. Es decir, dados A y B dos conjuntos, la distancia entre ellos es:

$$L(A, B) = \min_{x \in A, y \in B} d(x, y)$$

DEFINICIÓN 3: Enlace completo.

La distancia entre dos clústeres es la distancia máxima entre los puntos de ambos clústeres. Es decir, dados A y B dos conjuntos, la distancia entre ellos es:

$$L(A, B) = \max_{x \in A, y \in B} d(x, y)$$

DEFINICIÓN 4: Enlace promedio.

La distancia entre dos clústeres es el promedio de las distancias entre los puntos de ambos clústeres. Es decir, dados A y B dos conjuntos, la distancia entre ellos es:

$$L(A, B) = \frac{1}{|A||B|} \sum_{x \in A} \sum_{y \in B} d(x, y)$$

DEFINICIÓN 5: Enlace de centroide.

La distancia entre dos clústeres es la distancia entre los centroides de ambos clústeres. Es decir, dados A y B dos conjuntos, la distancia entre ellos es:

$$L(A, B) = d\left(\frac{1}{|A|} \sum_{x \in A} x, \frac{1}{|B|} \sum_{y \in B} y\right)$$

2. ALGORITMO DIVISIVO

Algoritmo 2

Pseudocódigo del agrupamiento jerárquico divisivo (DHC)

Entrada: D (conjunto de datos individuales)

Salida : Estructura jerárquica de clústeres

Iniciarizar $\ell \leftarrow 0$;

Iniciarizar el conjunto de clústeres $C_\ell \leftarrow \{D\}$;

mientras exista $c \in C_\ell$ que se pueda dividir **hacer**

Seleccionar clúster $c \in C_\ell$ para dividir;

Dividir c en subclústeres c_1, c_2, \dots, c_k ;

Actualizar $C_{\ell+1} \leftarrow (C_\ell \setminus \{c\}) \cup \{c_1, c_2, \dots, c_k\}$;

$\ell \leftarrow \ell + 1$;

devolver C_ℓ
