

Obteniendo jerarquías en la agrupación

Los algoritmos de agrupación jerárquicos, como su nombre lo indica, son capaces de crear una jerarquía en la que los objetos se agrupan formando un árbol que muestra cómo se relacionan entre sí. Estos algoritmos crean grupos solapados y la similitud se determina con medidas de distancia.

Existen dos tipos de algoritmos de agrupación jerárquica: aglomerativos y divisivos (ver Fig. 1). Veamos cada uno de ellos:

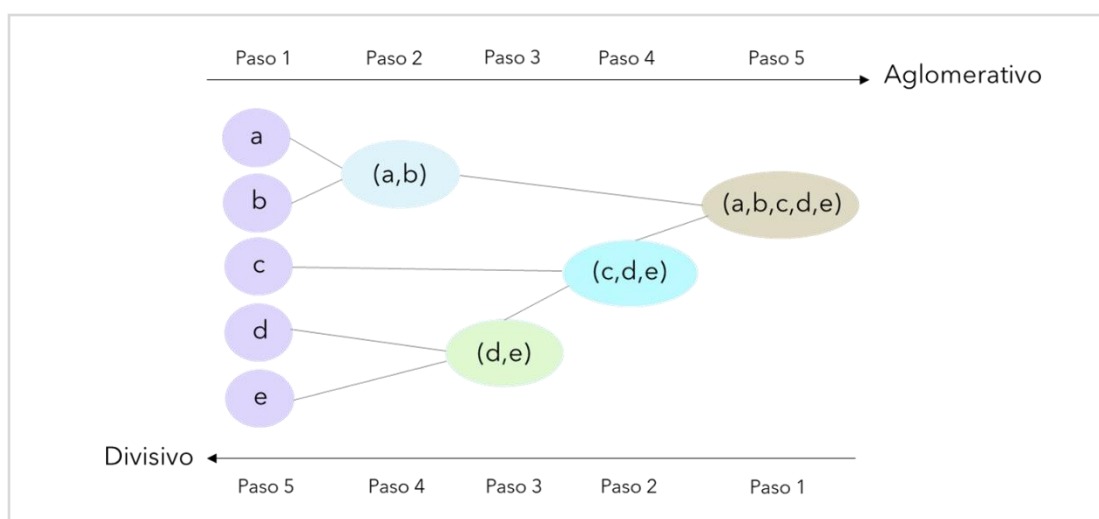


Fig. 1. Construcción de grupos por los algoritmos aglomerativos y divisivos.

Aglomerativos. Comienzan considerando cada objeto como un grupo individual y luego van fusionando los grupos similares entre sí en pasos sucesivos (esquema *bottom-up*). En cada paso, se unen los dos grupos más cercanos. Este proceso continúa hasta que todos los datos se encuentran en un solo grupo o cuando se haya alcanzado alguna otra condición de parada. Estos métodos utilizan medidas de enlace para decidir cuáles grupos se combinan en cada paso del algoritmo. Algunos procedimientos de enlace son: enlace simple (*single linkage*), el cual calcula la similitud entre los grupos con base en la distancia mínima entre sus elementos más cercanos; enlace

completo (*complete linkage*), donde se utiliza la distancia máxima entre los elementos más alejados de los grupos para determinar la similitud entre ellos, y enlace promedio (*average linkage*), con el cual se determina la distancia promedio entre todos los pares de elementos de los grupos para evaluar su similitud.

Divisivos. Inician con un grupo que contiene todos los objetos y luego lo dividen en subgrupos más pequeños con base en la disimilitud entre ellos (esquema *top-down*). En cada paso, se separa un grupo en dos subgrupos. Este proceso continúa hasta que cada grupo contiene un solo dato o se haya alcanzado alguna otra condición de parada. Estos métodos emplean medidas de disimilitud para decidir cómo dividir un grupo grande en subgrupos.

Tanto en los métodos aglomerativos como en los divisivos, es posible representar la jerarquía de la agrupación mediante un dendrograma (Fi. 2). Un dendrograma es una representación gráfica en forma de árbol que muestra la manera como los grupos se van fusionando o dividiendo a lo largo del proceso de agrupación. En un dendrograma, el eje vertical muestra la medida de distancia o similitud entre grupos u objetos, y en el eje horizontal se muestran los datos individuales o las fusiones/divisiones.

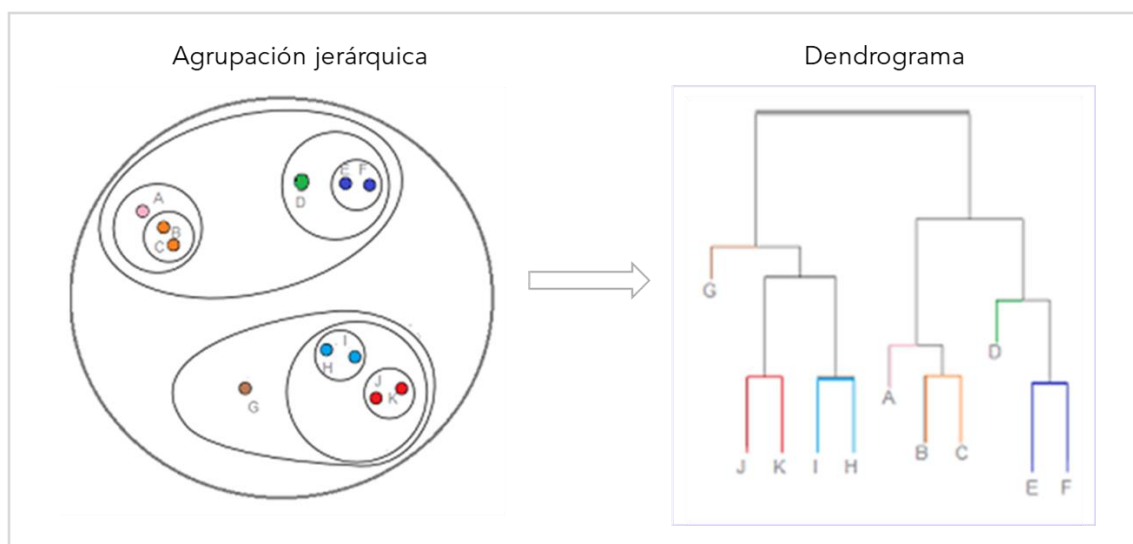


Fig. 2. Dendrograma asociado a un modelo de agrupación jerárquico.

Una ventaja importante de estos algoritmos radica en su capacidad para representar las relaciones jerárquicas entre los datos, lo que permite analizarlos a diferentes niveles de granularidad. Además, no requieren que se les especifique previamente el número de grupos a generar.

Bibliografía

Bonaccorso, G. (2019). *Hands-On Unsupervised learning with Python*. Packt Publishing

© - **Derechos Reservados:** la presente obra, y en general todos sus contenidos, se encuentran protegidos por las normas internacionales y nacionales vigentes sobre propiedad Intelectual, por lo tanto su utilización parcial o total, reproducción, comunicación pública, transformación, distribución, alquiler, préstamo público e importación, total o parcial, en todo o en parte, en formato impreso o digital y en cualquier formato conocido o por conocer, se encuentran prohibidos, y solo serán lícitos en la medida en que se cuente con la autorización previa y expresa por escrito de la Universidad de los Andes.

De igual manera, la utilización de la imagen de las personas, docentes o estudiantes, sin su previa autorización está expresamente prohibida. En caso de incumplirse con lo mencionado, se procederá de conformidad con los reglamentos y políticas de la universidad, sin perjuicio de las demás acciones legales aplicables.
