

INTRODUCCIÓN

Definición

El **Análisis no supervisado** consiste en separar en grupos, categorizar y segmentar datos basados en características comunes. Cada observación se parece lo más posible a las otras observaciones de su mismo grupo y lo menos posible a las observaciones de los otros grupos.



Incluso en casos supervisados (con etiquetas), se emplea el Análisis No Supervisado, como etapa preliminar, para comprender/segmentar los datos.

Objetivos de negocio

- Segmentación del mercado
- Perfilado de clientes
- Detección de anomalías
- Gestión de inventarios
- Evaluación de riesgos
- Sistemas de recomendación.

TÉCNICAS

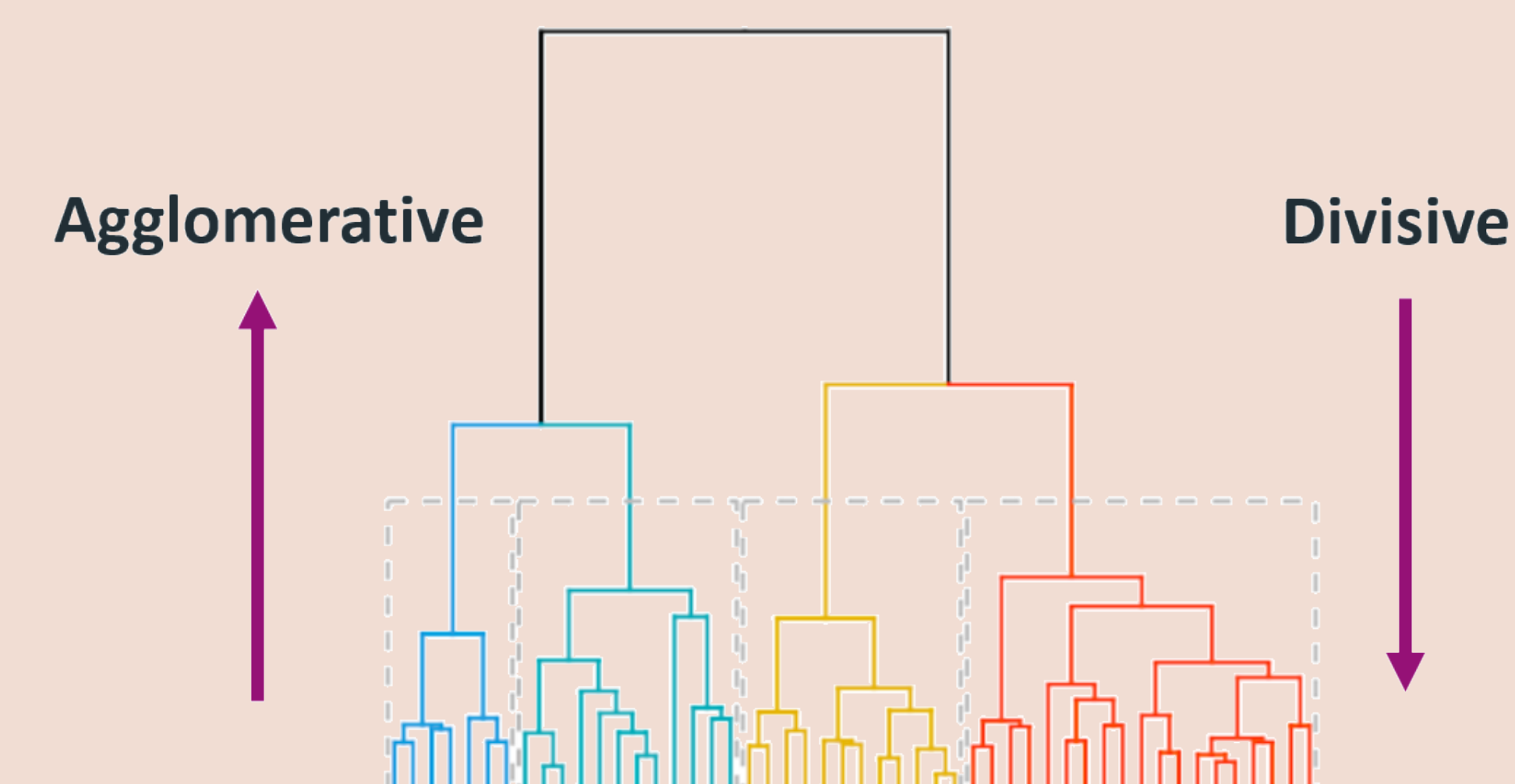
K-means

Dividir un conjunto de datos en k conglomerados, donde cada conglomerado se define por la media de los puntos del conglomerado. El proceso de encontrar los clusters se realiza reasignando iterativamente cada punto de datos al cluster con el centroide más cercano.

Agrupamiento Jerárquico

Se construye un árbol jerárquico de clusters, donde cada nodo del árbol representa un cluster y las hojas representan observaciones individuales.

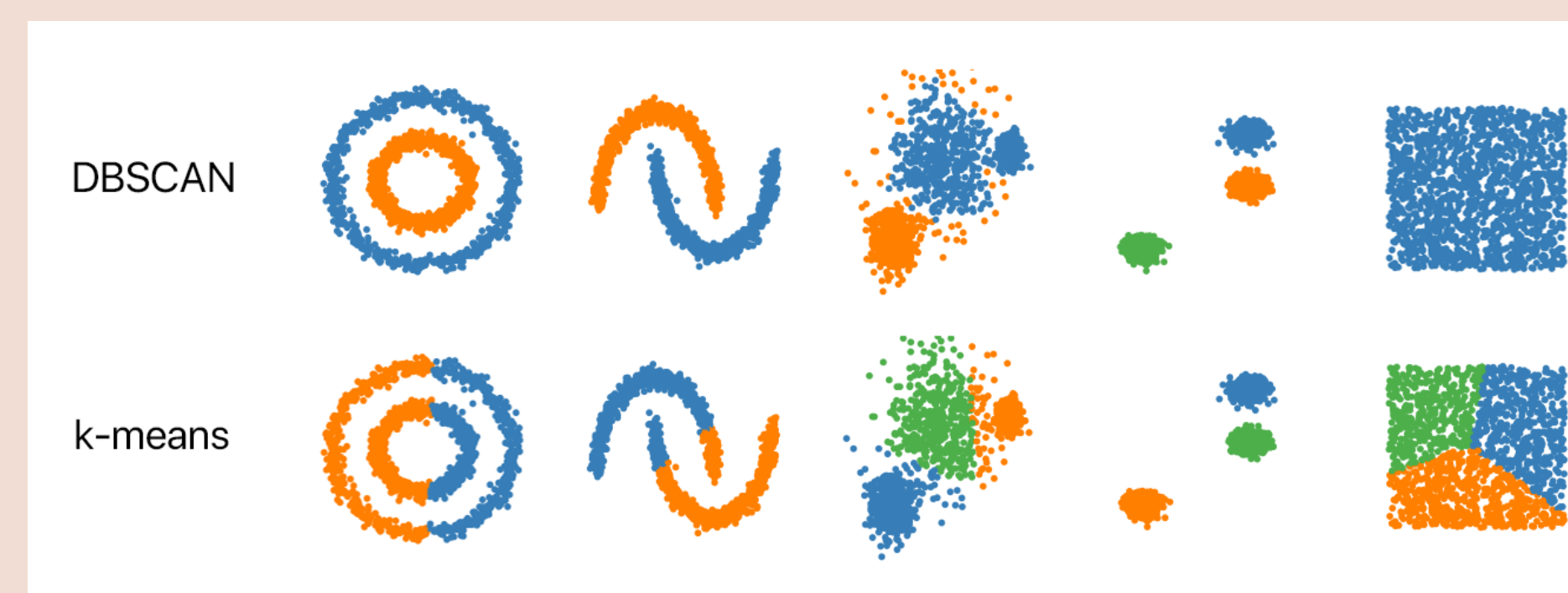
- Aglomerativo: Cada observación comienza como su propio conglomerado y los conglomerados más cercanos se fusionan de forma iterativa.
- Divisivo: todas las observaciones comienzan en el mismo conglomerado y éste se divide iterativamente en conglomerados más pequeños.



Agrupamiento basado en densidad

DBSCAN: agrupa los puntos que están muy juntos y separa los puntos que están más alejados. No requiere que se especifique de antemano el número de clusters y puede manejar clusters de diferentes formas y densidades. Requiere dos parámetros:

- eps: La distancia máxima entre dos puntos para ser considerados parte del mismo cluster.
- MinPts: El número mínimo de puntos necesarios para formar una región densa.



Librerías y Paquetes



scikit-learn, hdbscan, kneed



cluster, fpc, mclust, dbscan, factoextra

PASO A PASO FUNCIONAL

1. Preparar los datos. EDA.
2. Crea una métrica de similitud.
 - Distancia Euclídea.
 - Distancia pre-definida.
 - Distancia propia.
3. Ejecutar el algoritmo de agrupamiento en clusters.
 - No jerárquico. Elegir el número de clusters.
 - Jerárquico.
 - Basado en densidades.
4. Evaluar el modelo.
5. Interpretar los resultados.
 - Nombrar los clusters.
 - Construir clusters con baja explicabilidad dificulta la tarea de interpretabilidad de los modelos.
 - Identificar patrones o tendencias en los grupos.
6. Visualizar los resultados del análisis.
7. Generar conocimiento a partir de las lecciones aprendidas.

