



# No Supervisado Clustering





# Clustering



# Clustering

Mediante las técnicas de clustering se pretende agrupar los datos **sin etiquetar** en diferentes grupos (o **clusters**), de manera automática. Estos algoritmos buscan patrones y similitudes en los datos y los agrupan en  $k$  grupos, siendo  $k$  un parámetro dado o autoajustado por el modelo (no supervisado).



# Clustering

Mediante las técnicas de clustering se pretende agrupar los datos **sin etiquetar** en diferentes grupos (o **clusters**), de manera automática. Estos algoritmos buscan patrones y similitudes en los datos y los agrupan en  $k$  grupos, siendo  $k$  un parámetro dado o autoajustado por el modelo (no supervisado).

## Clasificación vs Clustering

En clasificación tenemos los datos etiquetados y por tanto acudimos a los algoritmos supervisados que ya conocemos (Regresión logística, SVC, DecisionTree...) para entrenar el modelo y predecir después las clases. En clustering no tenemos el dato etiquetado, y por tanto es el modelo el que clasifica los datos en  $k$  clases. En ningún momento sabe qué es cada clase, pero agrupa los datos según similitudes.

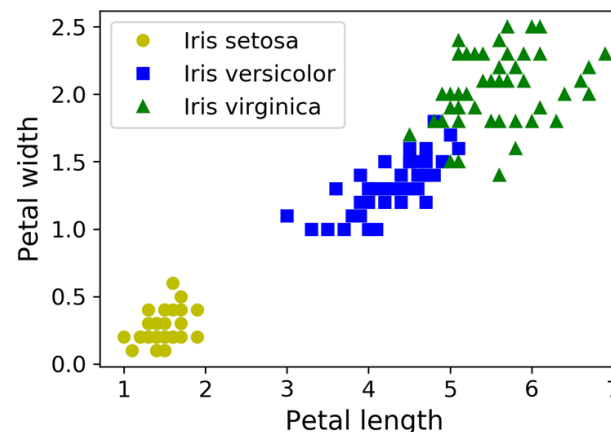


# Clustering

Mediante las técnicas de clustering se pretende agrupar los datos **sin etiquetar** en diferentes grupos (o **clusters**), de manera automática. Estos algoritmos buscan patrones y similitudes en los datos y los agrupan en  $k$  grupos, siendo  $k$  un parámetro dado o autoajustado por el modelo (no supervisado).

## Clasificación vs Clustering

En clasificación tenemos los datos etiquetados y por tanto acudimos a los algoritmos supervisados que ya conocemos (Regresión logística, SVC, DecisionTree...) para entrenar el modelo y predecir después las clases. En clustering no tenemos el dato etiquetado, y por tanto es el modelo el que clasifica los datos en  $k$  clases. En ningún momento sabe qué es cada clase, pero agrupa los datos según similitudes.

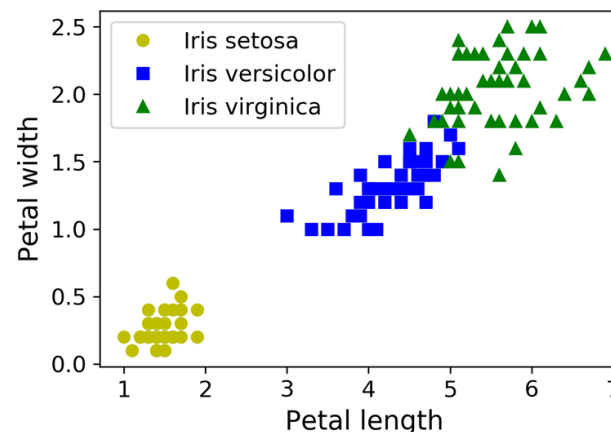


# Clustering

Mediante las técnicas de clustering se pretende agrupar los datos **sin etiquetar** en diferentes grupos (o **clusters**), de manera automática. Estos algoritmos buscan patrones y similitudes en los datos y los agrupan en  $k$  grupos, siendo  $k$  un parámetro dado o autoajustado por el modelo (no supervisado).

## Clasificación vs Clustering

En clasificación tenemos los datos etiquetados y por tanto acudimos a los algoritmos supervisados que ya conocemos (Regresión logística, SVC, DecisionTree...) para entrenar el modelo y predecir después las clases. En clustering no tenemos el dato etiquetado, y por tanto es el modelo el que clasifica los datos en  $k$  clases. En ningún momento sabe qué es cada clase, pero agrupa los datos según similitudes.

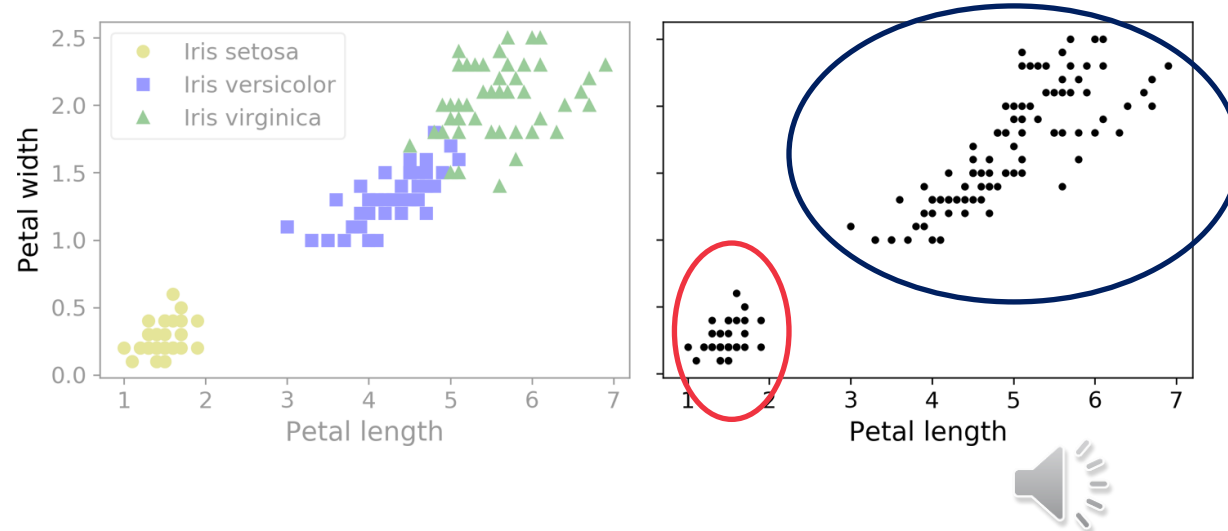


# Clustering

Mediante las técnicas de clustering se pretende agrupar los datos **sin etiquetar** en diferentes grupos (o **clusters**), de manera automática. Estos algoritmos buscan patrones y similitudes en los datos y los agrupan en  $k$  grupos, siendo  $k$  un parámetro dado o autoajustado por el modelo (no supervisado).

## Clasificación vs Clustering

En clasificación tenemos los datos etiquetados y por tanto acudimos a los algoritmos supervisados que ya conocemos (Regresión logística, SVC, DecisionTree...) para entrenar el modelo y predecir después las clases. En clustering no tenemos el dato etiquetado, y por tanto es el modelo el que clasifica los datos en  $k$  clases. En ningún momento sabe qué es cada clase, pero agrupa los datos según similitudes.

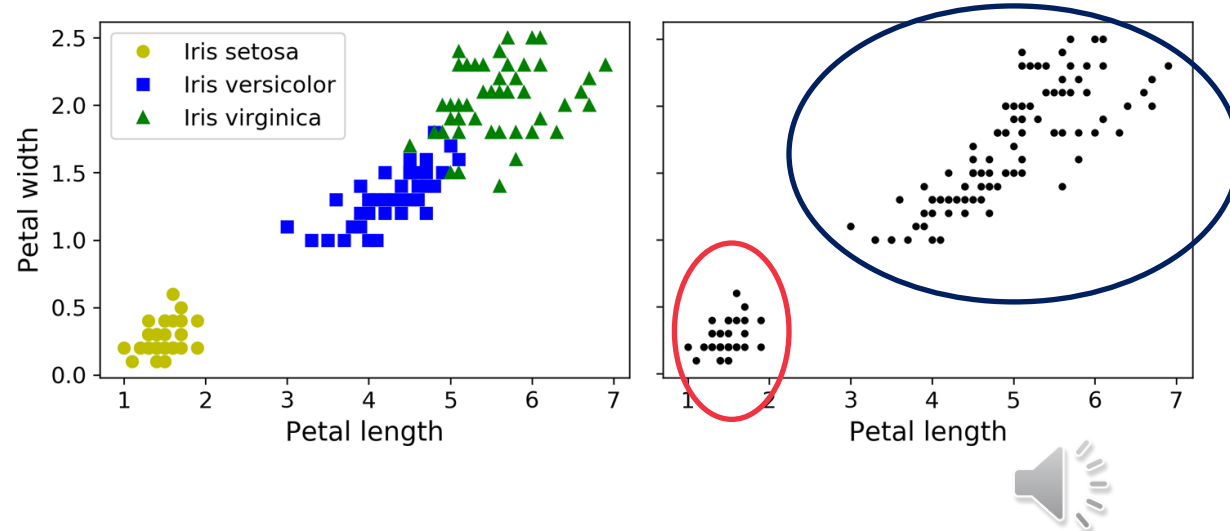


# Clustering

Mediante las técnicas de clustering se pretende agrupar los datos **sin etiquetar** en diferentes grupos (o **clusters**), de manera automática. Estos algoritmos buscan patrones y similitudes en los datos y los agrupan en  $k$  grupos, siendo  $k$  un parámetro dado o autoajustado por el modelo (no supervisado).

## Clasificación vs Clustering

En clasificación tenemos los datos etiquetados y por tanto acudimos a los algoritmos supervisados que ya conocemos (Regresión logística, SVC, DecisionTree...) para entrenar el modelo y predecir después las clases. En clustering no tenemos el dato etiquetado, y por tanto es el modelo el que clasifica los datos en  $k$  clases. En ningún momento sabe qué es cada clase, pero agrupa los datos según similitudes.





# Aplicaciones



# Aplicaciones

## Segmentación de clientes

*Clasificar clientes en grupos para sistemas  
recomendadores de productos*



# Aplicaciones

## Segmentación de clientes

*Clasificar clientes en grupos para sistemas  
recomendadores de productos*

## Data Analysis

*Resulta útil aplicar clustering cuando  
analicemos un dataset para descubrir  
posibles agrupaciones en los datos*



# Aplicaciones

## Segmentación de clientes

*Clasificar clientes en grupos para sistemas recomendadores de productos*

## Data Analysis

*Resulta útil aplicar clustering cuando analicemos un dataset para descubrir posibles agrupaciones en los datos*

## Dimensionality Reduction

*Después de aplicar clustering podemos ver la afinidad de cada observación con sus grupos y que esas  $k$  medidas sean las features.*



# Aplicaciones

## Segmentación de clientes

*Clasificar clientes en grupos para sistemas recomendadores de productos*

## Data Analysis

*Resulta útil aplicar clustering cuando analicemos un dataset para descubrir posibles agrupaciones en los datos*

## Dimensionality Reduction

*Después de aplicar clustering podemos ver la afinidad de cada observación con sus grupos y que esas  $k$  medidas sean las features.*

## Anomaly Detection

*Toda observación con poca afinidad respecto a su grupo es susceptible de ser una anomalía o un outlier. Detección de comportamientos inusuales de los usuarios*





# Aplicaciones

## Segmentación de clientes

*Clasificar clientes en grupos para sistemas recomendadores de productos*

## Data Analysis

*Resulta útil aplicar clustering cuando analicemos un dataset para descubrir posibles agrupaciones en los datos*

## Dimensionality Reduction

*Después de aplicar clustering podemos ver la afinidad de cada observación con sus grupos y que esas  $k$  medidas sean las features.*

## Anomaly Detection

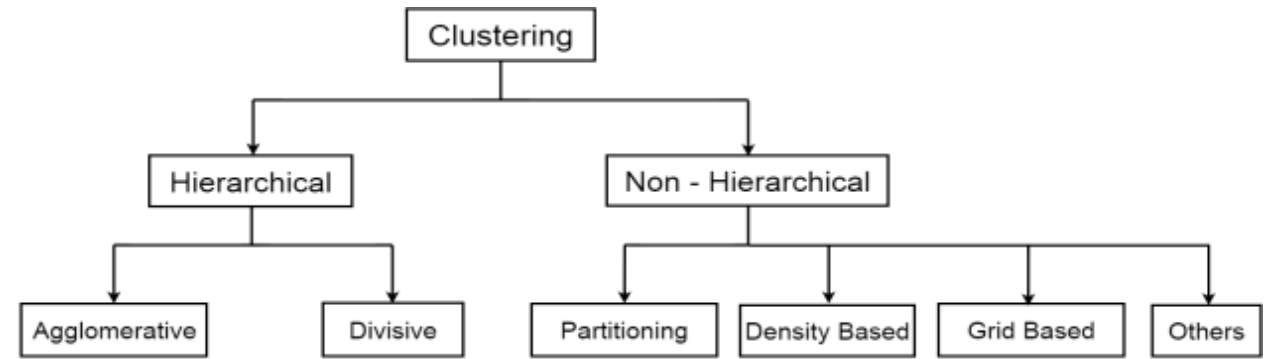
*Toda observación con poca afinidad respecto a su grupo es susceptible de ser una anomalía o un outlier. Detección de comportamientos inusuales de los usuarios*

## Algoritmos de búsqueda

*Algoritmos de búsqueda de imágenes similares a una referencia. Como si de un sistema de recomendación se tratase.*



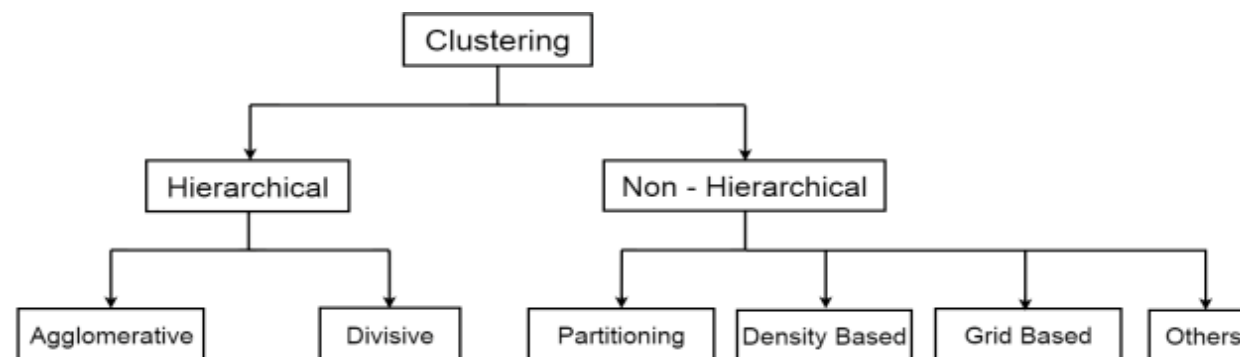
# Tipos



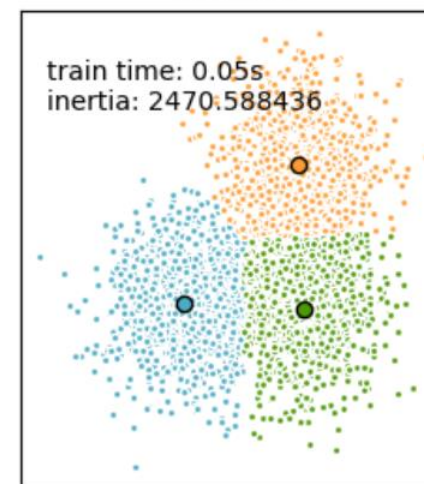
## Tipos

### NO JERÁRQUICOS

- Se optimiza una función de “distancia” a partir de unos “prototipos”
- El DS establece unos puntos de partida e iterativamente el algortimo mejora el *clustering*



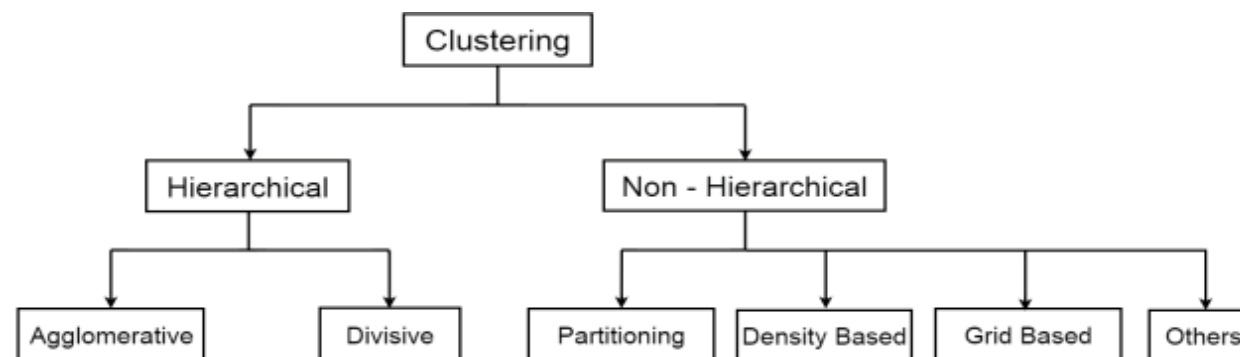
KMeans



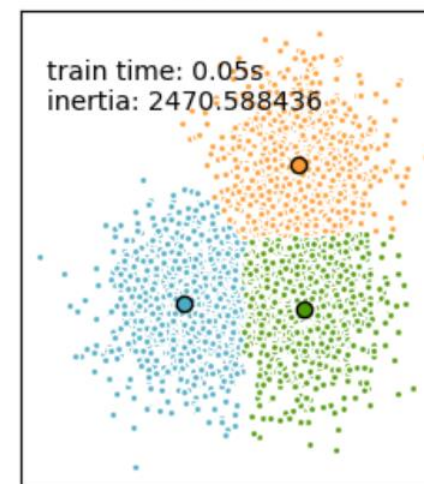
## Tipos

### NO JERÁRQUICOS

- Se optimiza una función de “distancia” a partir de unos “prototipos”
- El DS establece unos puntos de partida e iterativamente el algortimo mejora el *clustering*



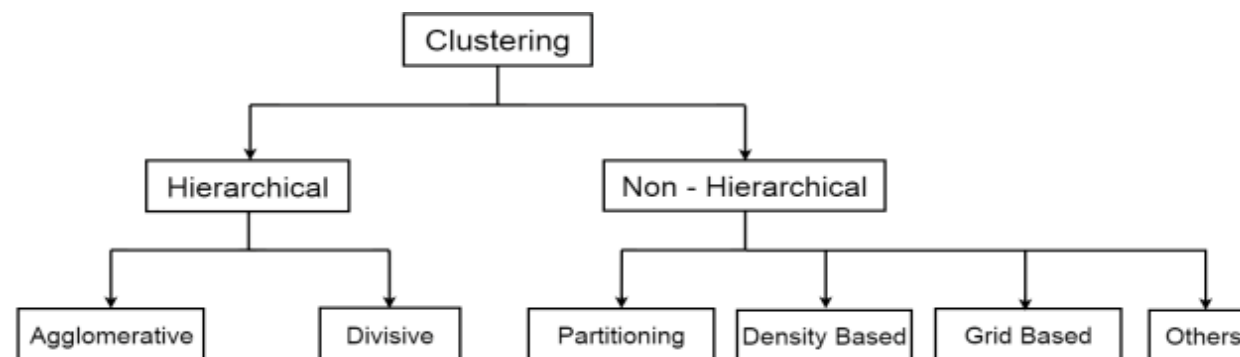
KMeans



## Tipos

### NO JERÁRQUICOS

- Se optimiza una función de “distancia” a partir de unos “prototipos”
- El DS establece unos puntos de partida e iterativamente el algoritmo mejora el *clustering*



### JERÁRQUICOS

- Se construye un dendrograma (de arriba abajo o de abajo arriba) y se puede poner un punto de corte y generar tantos clusters como se crea necesario

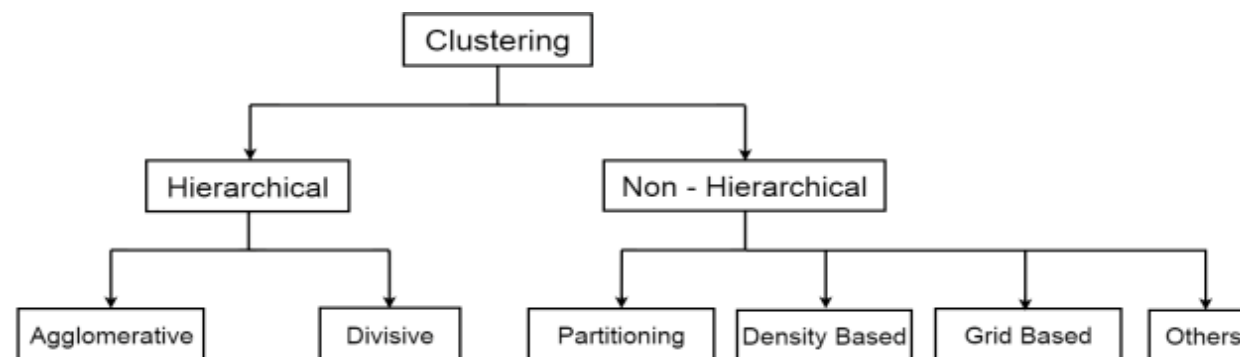




## Tipos

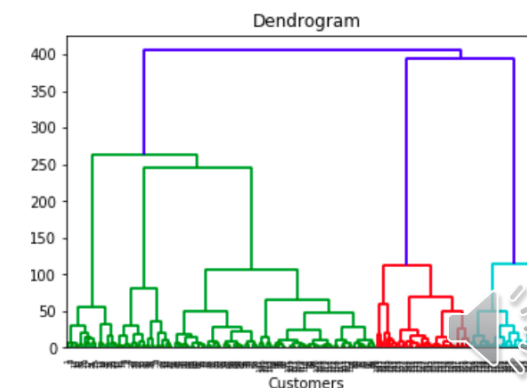
### NO JERÁRQUICOS

- Se optimiza una función de “distancia” a partir de unos “prototipos”
- El DS establece unos puntos de partida e iterativamente el algoritmo mejora el *clustering*



### JERÁRQUICOS

- Se construye un dendrograma (de arriba abajo o de abajo arriba) y se puede poner un punto de corte y generar tantos clusters como se crea necesario



# NO JERÁRQUICOS

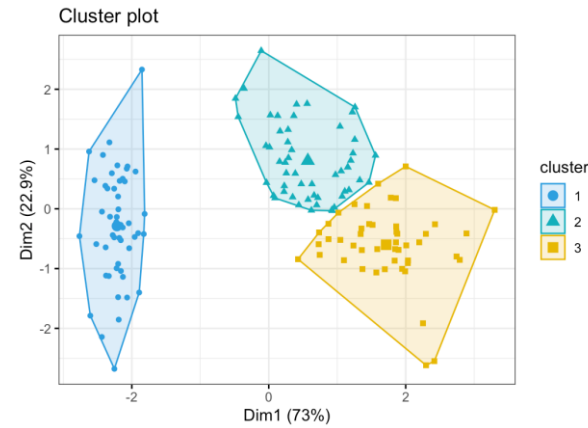
# JERÁRQUICOS

- Se construye un dendrograma (de arriba abajo o de abajo arriba) y se puede poner un punto de corte y generar tantos clusters como se crea necesario

## Tipos

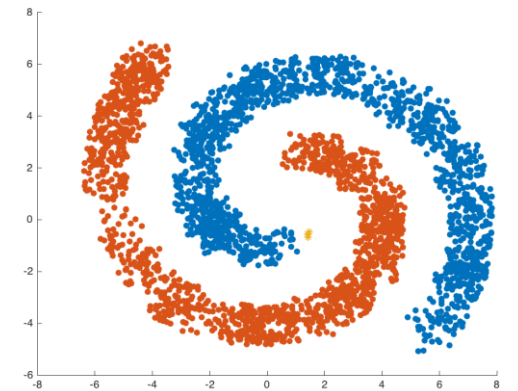
### K-Means

*Algoritmo de clustering basado en distancias*



### DBSCAN

*Define los clusters como regiones continuas de alta densidad de observaciones.*



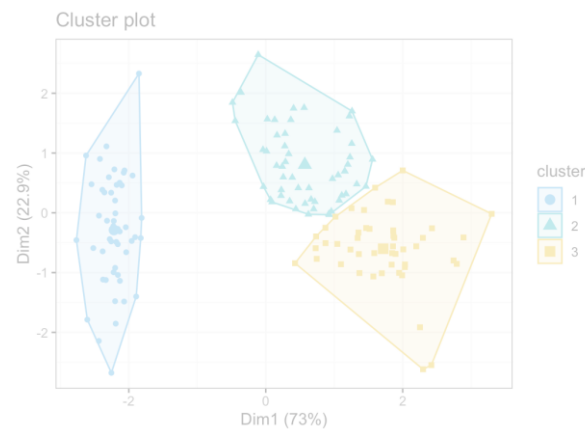
NO JERÁRQUICOS

JERÁRQUICOS

## Tipos

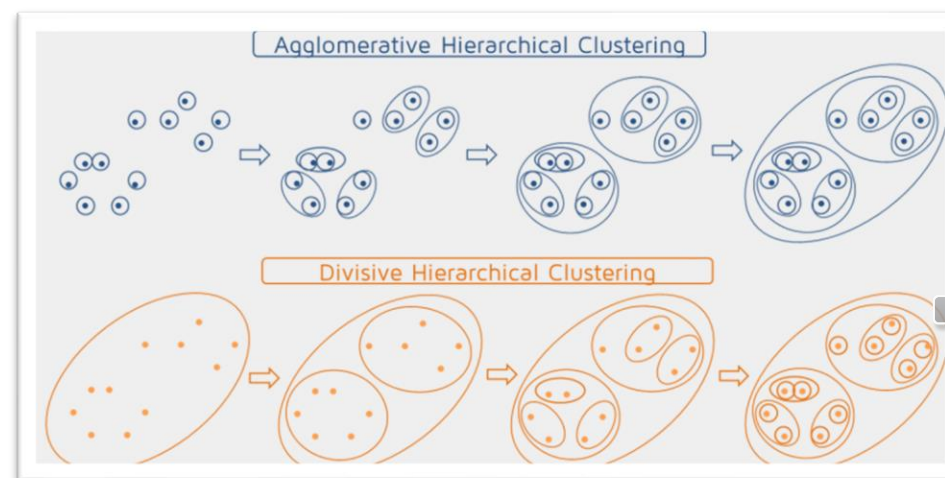
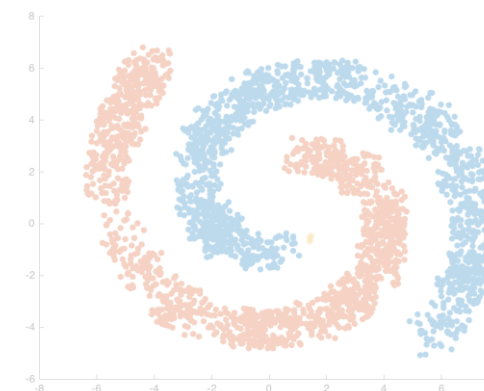
### K-Means

Algoritmo de clustering basado en distancias



### DBSCAN

Define los clusters como regiones continuas de alta densidad de observaciones.



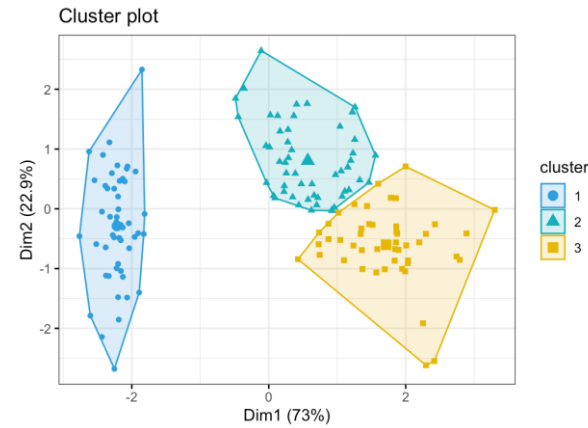
# NO JERÁRQUICOS

# JERÁRQUICOS

## Tipos

### K-Means

Algoritmo de clustering basado en distancias



### DBSCAN

Define los clusters como regiones continuas de alta densidad de observaciones.

