Aprendizaje no supervisado

3.1. Agrupamiento jerárquico: Divisivo

Javier Sevilla

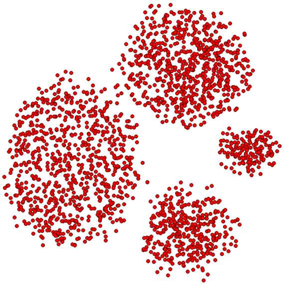
Universidad Internacional de Valencia

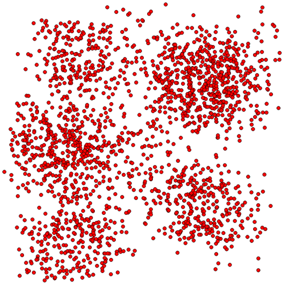
27.02.2020

Agrupamiento

Tipos de algoritmos de agrupamiento

- Basados en particiones
- **Jerárquicos**
- Espectrales
- Basados en densidad
- Probabilísticos





Agrupamiento Jerárquico

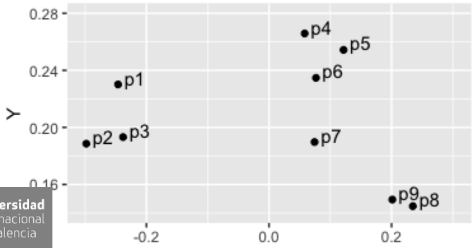
Un continuo de particiones de los datos

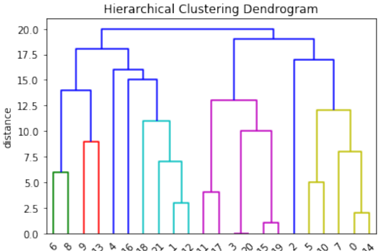
Se particiona el dataset desde $K = 1$ hasta $K = n$

** ¿Cuál es la mejor partición?

Algoritmos:

- Aglomerativo
- **Divisivo**



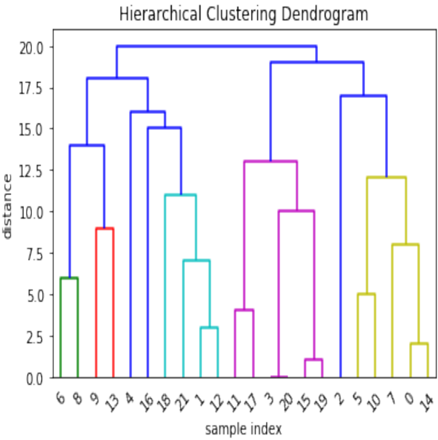


Agrupamiento Jerárquico

Dendrograma

Representación gráfica de un agrupamiento jerárquico

- Cada nodo, es un conjunto de ejemplos (clúster)
- Los clústeres se van uniendo/separando según criterios de distancia
- La longitud de las líneas verticales indica la distancia entre los clústeres que se unen/separan



División

Partiendo de $K = 1$, se va dividiendo (en dos) iterativamente un clúster hasta $K = n$, de manera voraz

0. Al principio, sólo existe un clúster ($K = 1$) con todos los ejemplos
1. Tras la primera división, existen $K = 2$ clústeres
   (se reparten entre los dos todos los ejemplos)

...

i. Tras la i-ésima división, existen $K = i + 1$ clústeres

...

n-1. El algoritmo acaba cuando $K = n$
   (se divide el único clúster que no es unitario y cada ejemplo tiene su propio clúster)

Agrupamiento Jerárquico

Divisivo

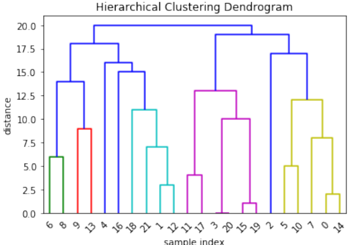
Tres cuestiones

A medida que avanza el algoritmo...

¿qué clúster se debe dividir en cada paso?

Al final del algoritmo, si queremos un partición concreta,

¿con qué partición nos quedamos?



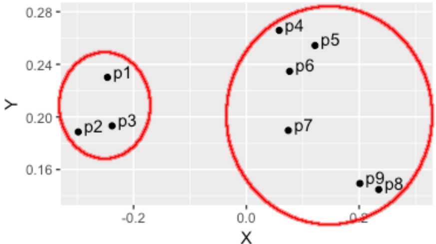
Tres cuestiones

A medida que avanza el algoritmo...

¿qué clúster se debe dividir en cada paso?

Al final del algoritmo, si queremos un partición concreta,

¿con qué partición nos quedamos?



Tres cuestiones

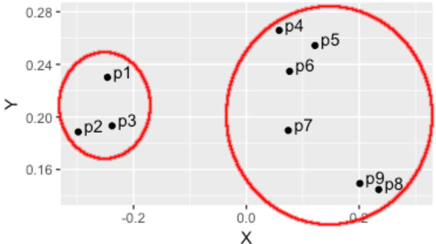
A medida que avanza el algoritmo...

¿qué clúster se debe dividir en cada paso?

¿cómo se divide el clúster seleccionado?

Al final del algoritmo, si queremos un partición concreta,

¿con qué partición nos quedamos?



Primera cuestión

A medida que avanza el algoritmo...

¿qué clúster se debe dividir en cada paso?

El clúster, S_K^*, con mayor disimilitud intraclúster:

$$S_K^* = \operatorname*{argmáx}_{S_K} d(S_K)$$

Primera cuestión

A medida que avanza el algoritmo...

¿qué clúster se debe dividir en cada paso?

El clúster, S_K^*, con mayor disimilitud intraclúster:

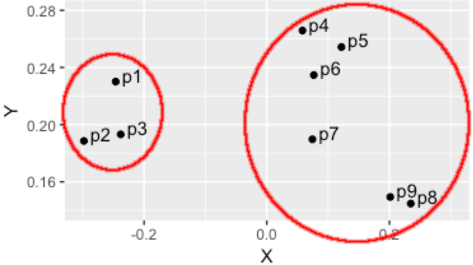
$$S_K^* = \operatorname*{argmáx}_{S_K} d(S_K)$$

¿cómo se mide la disimilitud intraclúster?

Criterios de división

$$d(S_K) = \max_{x_i, x_j \in S_K} d(x_i, x_j)$$

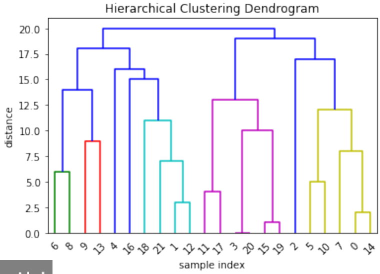
Diámetro

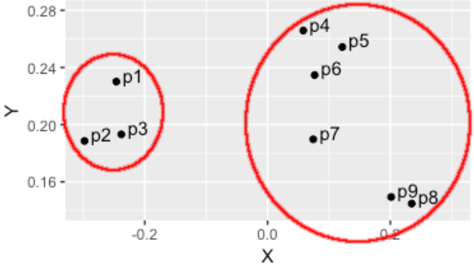


Criterios de división

$$d(S_K) = \max_{x_i, x_j \in S_K} d(x_i, x_j)$$

Diámetro

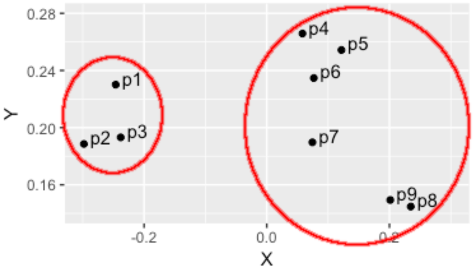




Criterios de división

$$d(S_K) = \frac{1}{|S_K|^2} \sum_{x_i \in S_K} \sum_{x_j \in S_K} d(x_i, x_j)$$

Disimilitud media



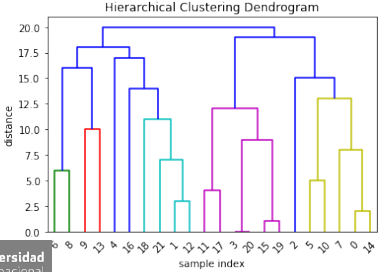
Agrupamiento Jerárquico

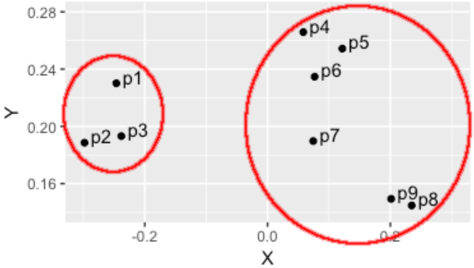
Divisivo

Criterios de división

$$d(S_K) = \frac{1}{|S_K|^2} \sum_{x_i \in S_K} \sum_{x_j \in S_K} d(x_i, x_j)$$

Disimilitud media





Segunda cuestión

A medida que avanza el algoritmo, una vez se ha elegido un clúster a dividir...

¿cómo se divide el clúster seleccionado?

La mejor separación del conjunto de datos:

$$\{S_A^*, S_B^*\} = \arg \max_{S_A, S_B \subset S_K} sep(S_A, S_B)$$

Segunda cuestión

A medida que avanza el algoritmo, una vez se ha elegido un clúster a dividir...

¿cómo se divide el clúster seleccionado?

La mejor separación del conjunto de datos:

$$\{S_A^*, S_B^*\} = \arg \max_{S_A, S_B \subset S_K} sep(S_A, S_B)$$

¿cómo se encuentra la mejor separación o división?

Criterios de división

K-means

Aplicar el popular algoritmo K-means con $K = 2$

Criterios de división

K-means

Aplicar el popular algoritmo K-means con $K = 2$

Criterios de división

Técnica de Macnaughton-Smith

- Elegir el ejemplo $\boldsymbol{x}$ que en media está más lejano del resto del clúster, S_K
- $S_B = S_K \backslash \{\boldsymbol{x}\}$; $S_A = \{\boldsymbol{x}\}$
- Ir añadiendo casos $\boldsymbol{x}^*$
$$\boldsymbol{x}^* = \text{argmáx}_{x_i \in S_B} \left(\frac{1}{|S_B|-1} \sum_{x_{i'} \in S_B : x_i \neq x_{i'}} d(x_i, x_{i'}) - \frac{1}{|S_A|} \sum_{x_j \in S_A} d(x_i, x_j) \right)$$
- Hasta que el valor de la función *argmax* sea inferior a 0

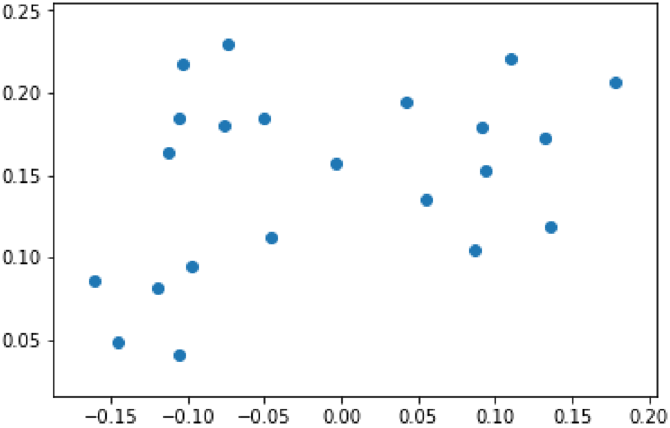
Criterios de división

Técnica de Macnaughton-Smith

- Elegir el ejemplo $\boldsymbol{x}$ que en media está más lejano del resto del clúster, S_K
- $S_B = S_K \setminus \{\boldsymbol{x}\}$; $S_A = \{\boldsymbol{x}\}$
- Ir añadiendo casos $\boldsymbol{x}^*$
  $$\boldsymbol{x}^* = \text{argmáx}_{x_i \in S_B} \Big(\tfrac{1}{|S_B|-1} \textstyle\sum_{x_{i'} \in S_B : x_i \neq x_{i'}} d(x_i, x_{i'}) - \tfrac{1}{|S_A|} \textstyle\sum_{x_j \in S_A} d(x_i, x_j) \Big)$$
- Hasta que el valor de la función *argmax* sea inferior a 0

Agrupamiento Jerárquico

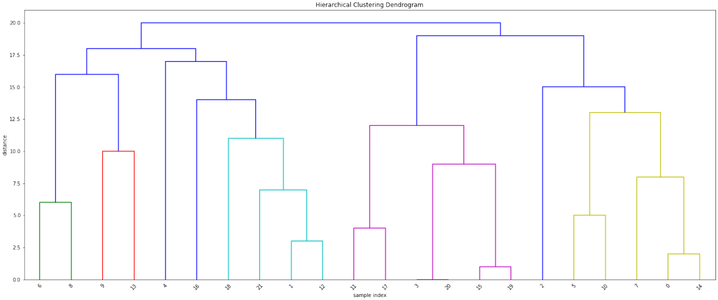
Divisivo



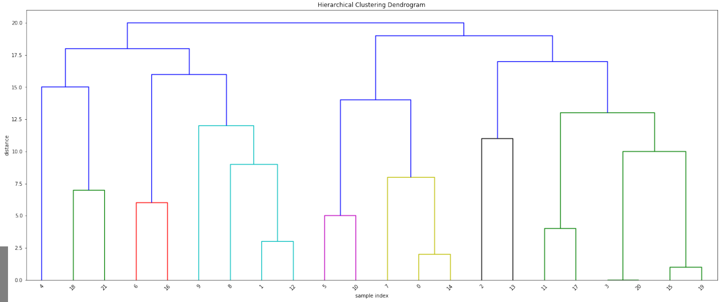
Agrupamiento Jerárquico

Divisivo

Mac.-Smith



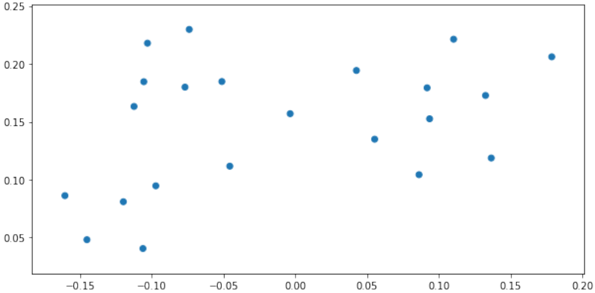
K-means



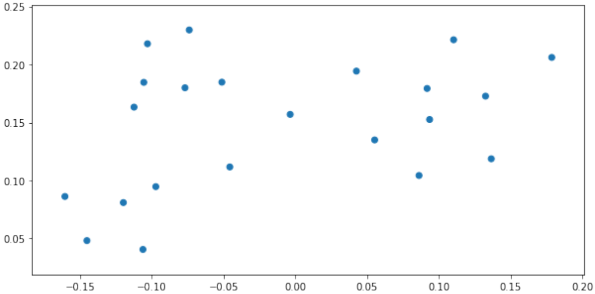
Agrupamiento Jerárquico

Divisivo

Mac.-Smith



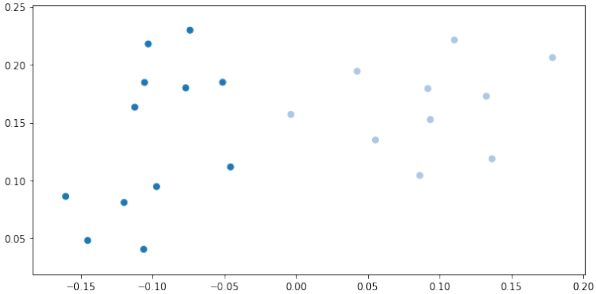
K-means



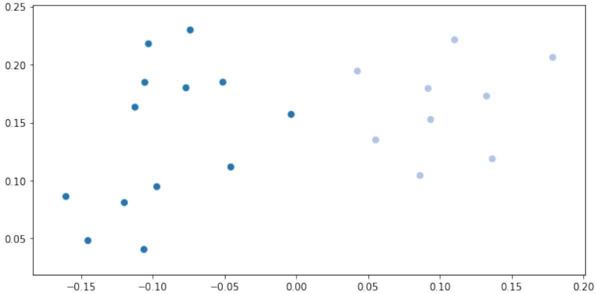
Agrupamiento Jerárquico

Divisivo

Mac.-Smith



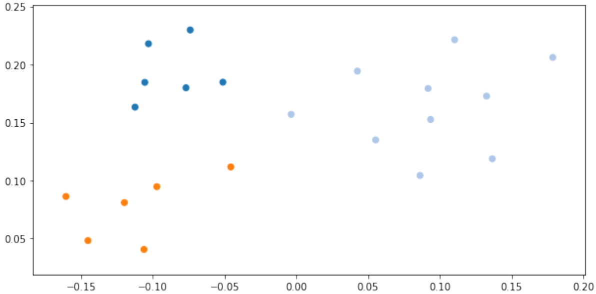
K-means



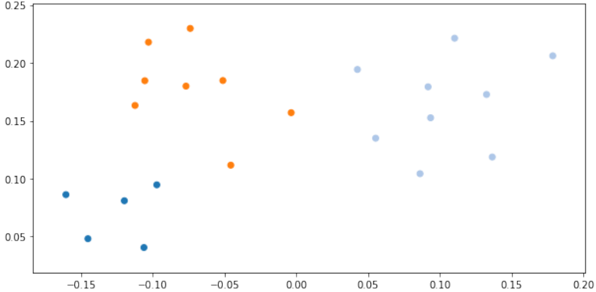
Agrupamiento Jerárquico

Divisivo

Mac.-Smith



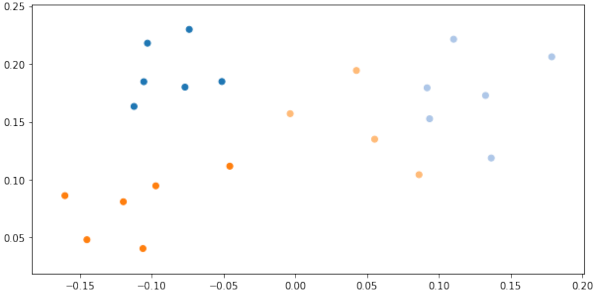
K-means



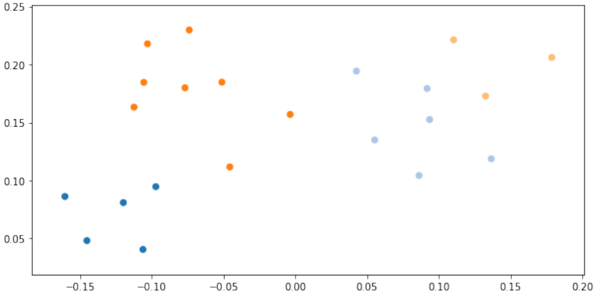
Agrupamiento Jerárquico

Divisivo

Mac.-Smith

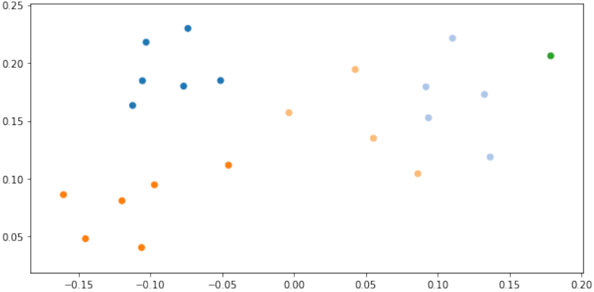


K-means

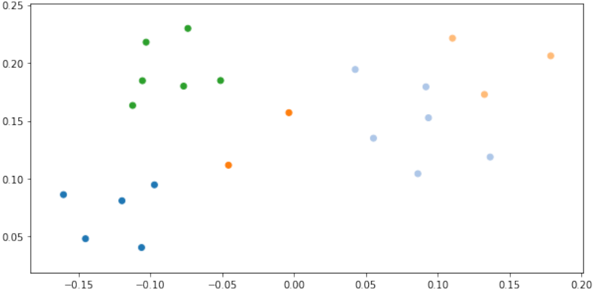


Agrupamiento Jerárquico
Divisivo

Mac.-Smith

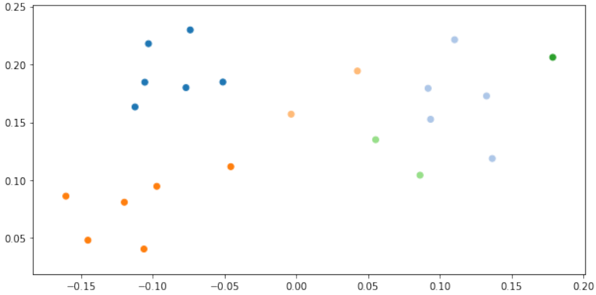


K-means

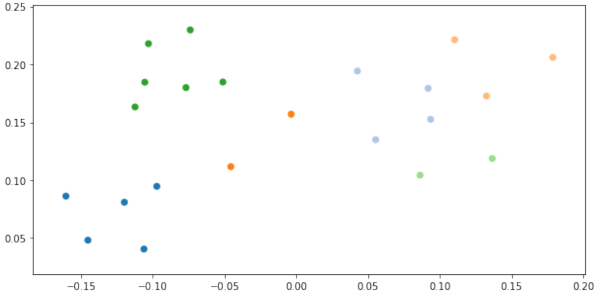


Agrupamiento Jerárquico
Divisivo

Mac.-Smith

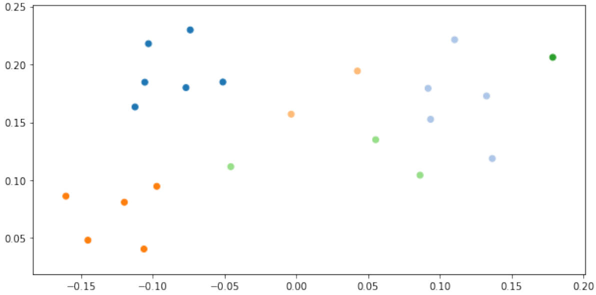


K-means

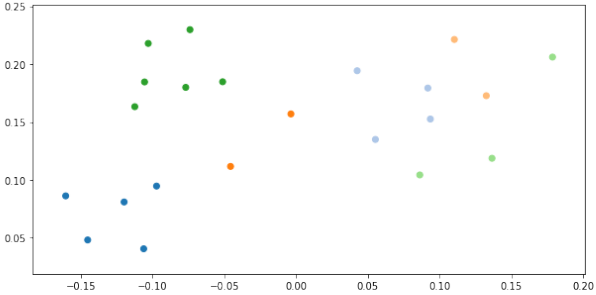


Agrupamiento Jerárquico
Divisivo

Mac.-Smith



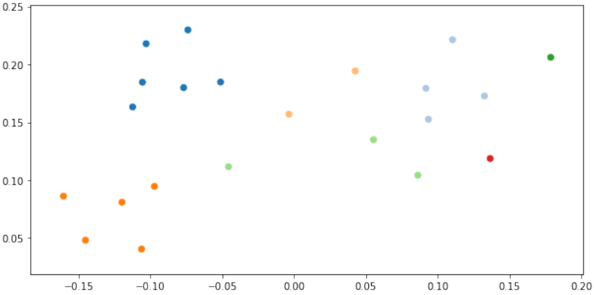
K-means



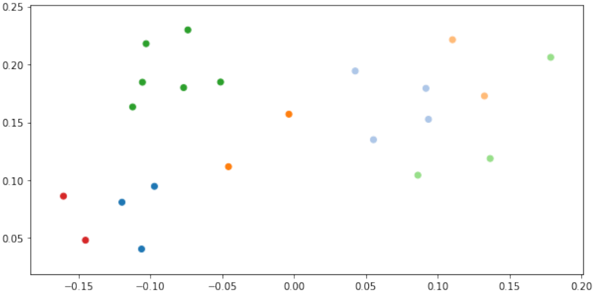
Agrupamiento Jerárquico

Divisivo

Mac.-Smith



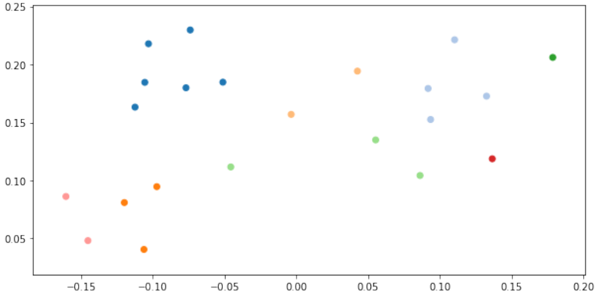
K-means



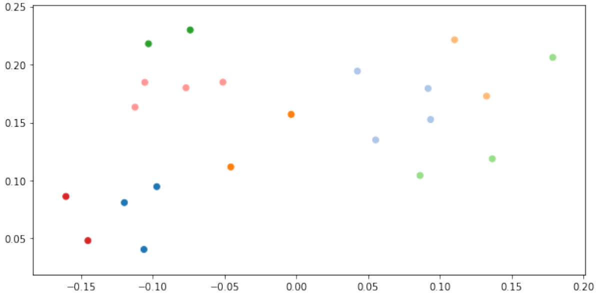
Agrupamiento Jerárquico

Divisivo

Mac.-Smith

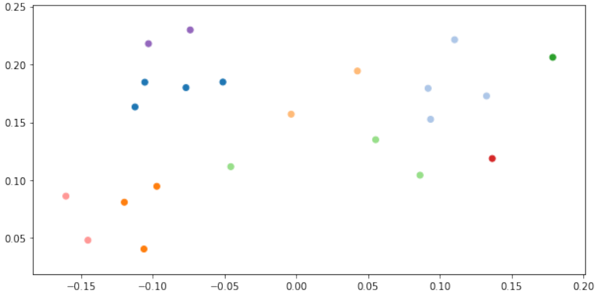


K-means

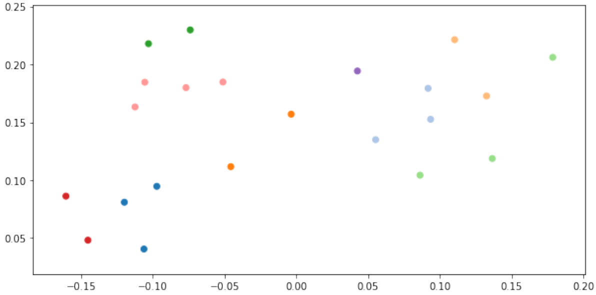


Agrupamiento Jerárquico
Divisivo

Mac.-Smith



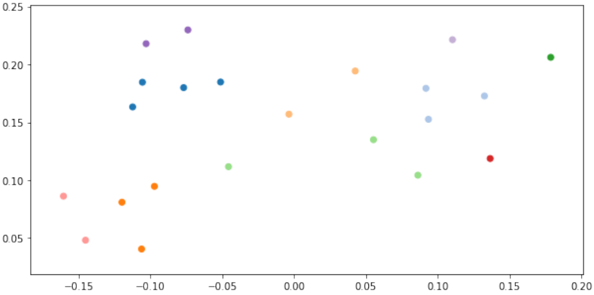
K-means



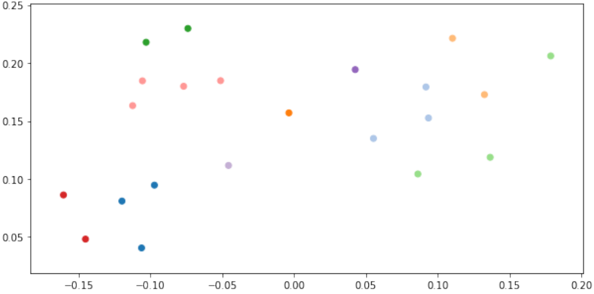
Agrupamiento Jerárquico

Divisivo

Mac.-Smith



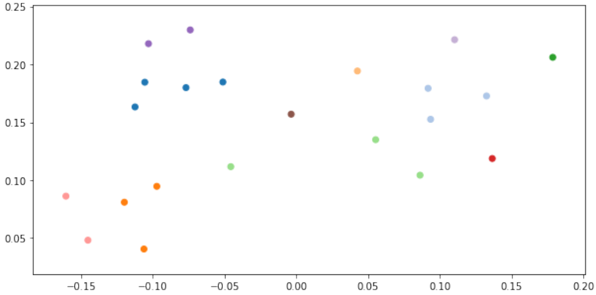
K-means



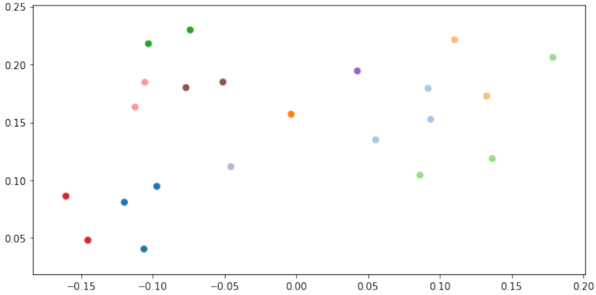
Agrupamiento Jerárquico

Divisivo

Mac.-Smith



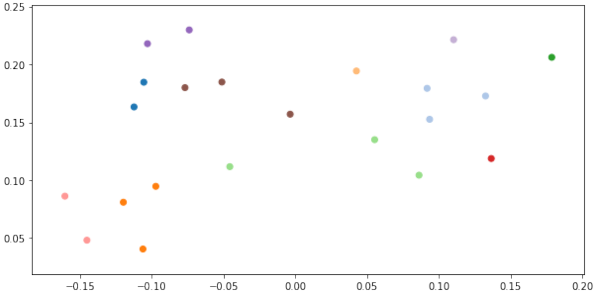
K-means



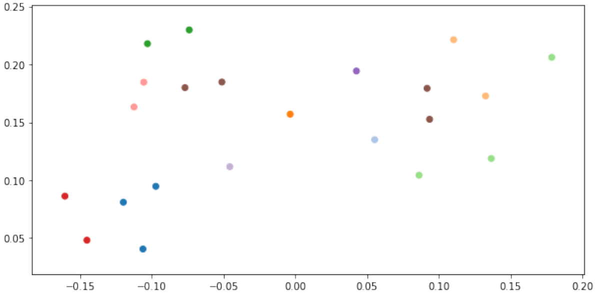
Agrupamiento Jerárquico

Divisivo

Mac.-Smith



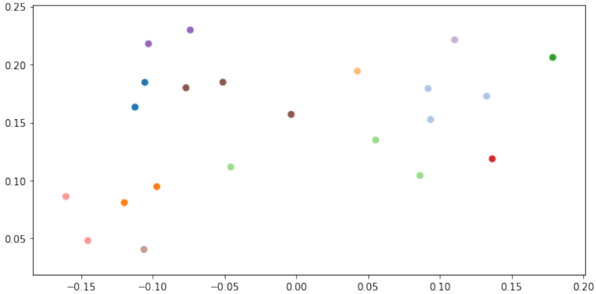
K-means



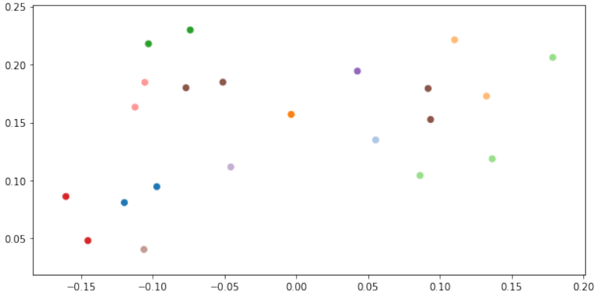
Agrupamiento Jerárquico

Divisivo

Mac.-Smith



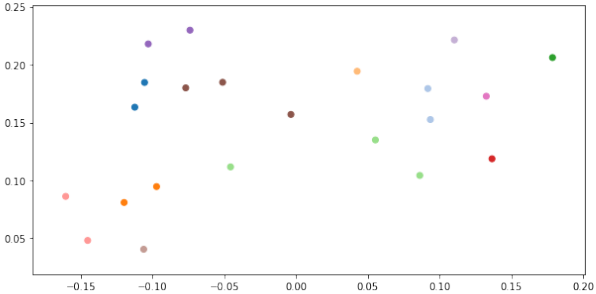
K-means



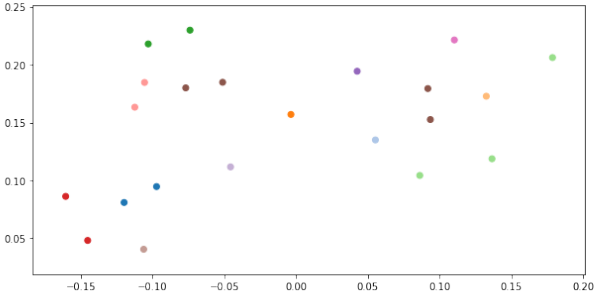
Agrupamiento Jerárquico

Divisivo

Mac.-Smith



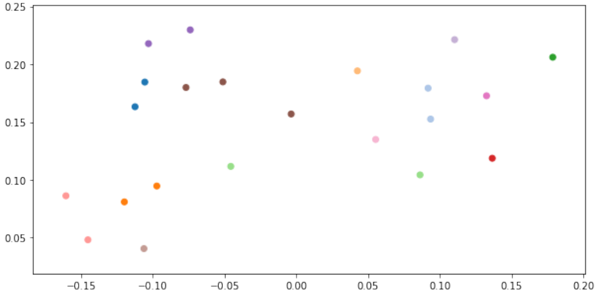
K-means



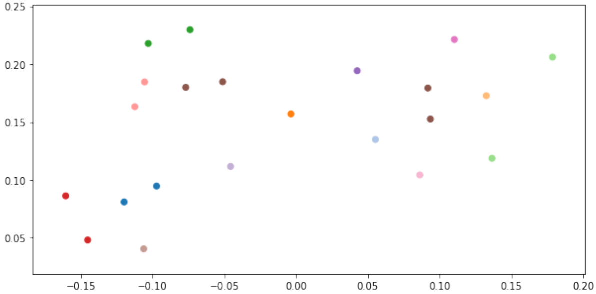
Agrupamiento Jerárquico

Divisivo

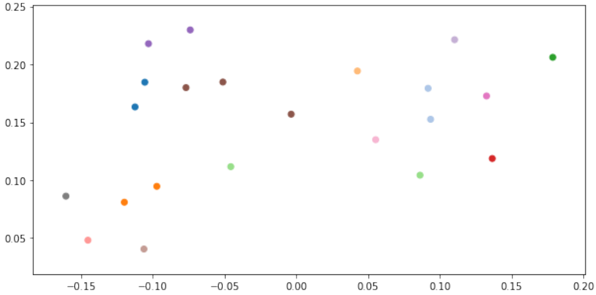
Mac.-Smith



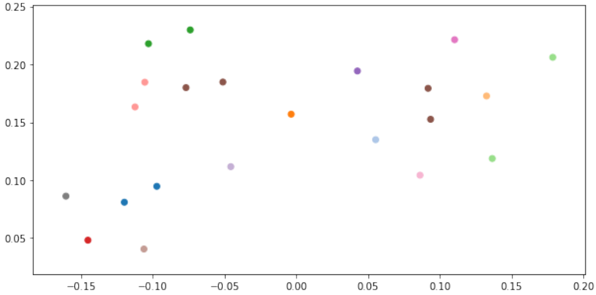
K-means



Mac.-Smith



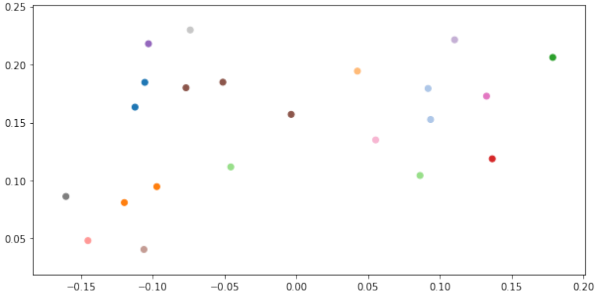
K-means



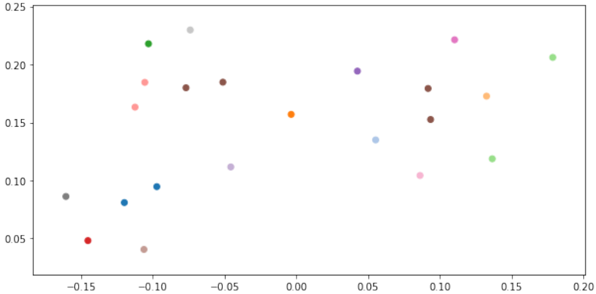
Agrupamiento Jerárquico

Divisivo

Mac.-Smith



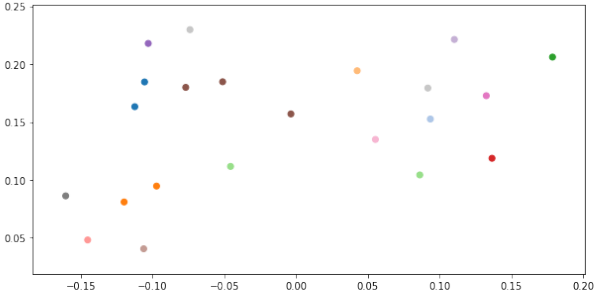
K-means



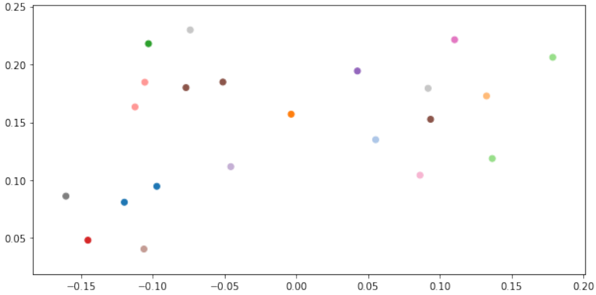
Agrupamiento Jerárquico

Divisivo

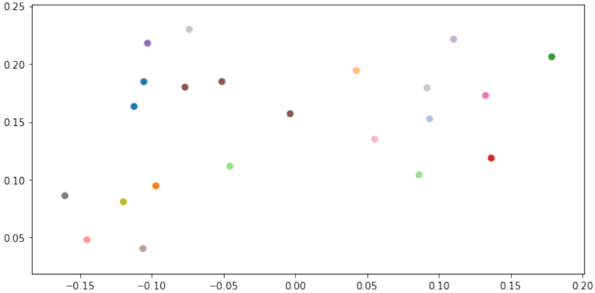
Mac.-Smith



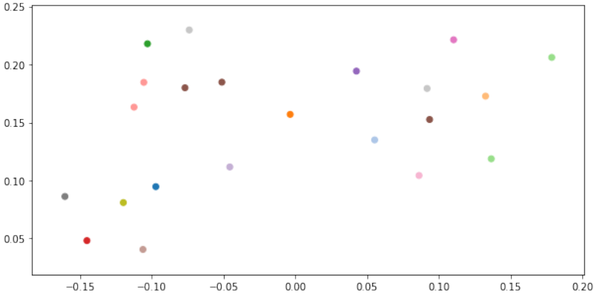
K-means



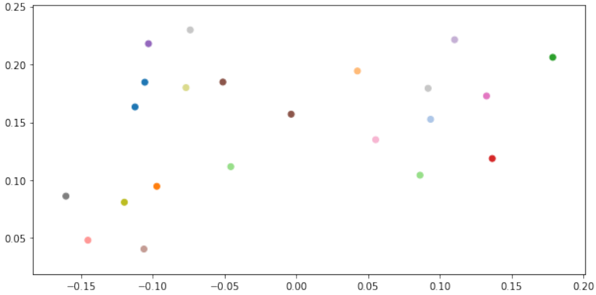
Mac.-Smith



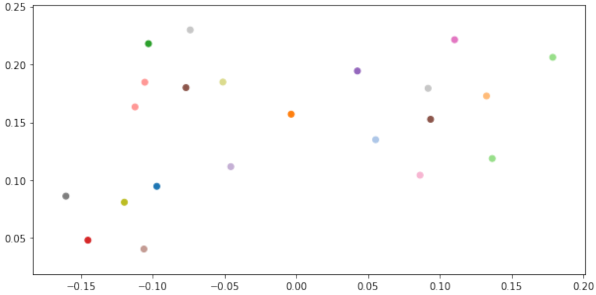
K-means



Mac.-Smith



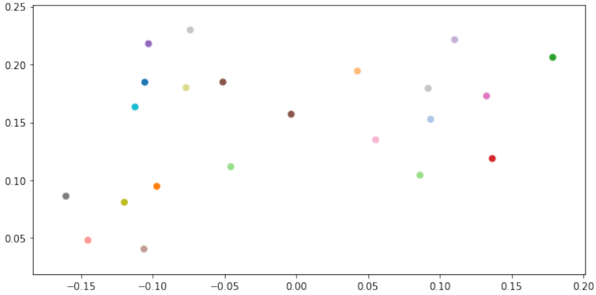
K-means



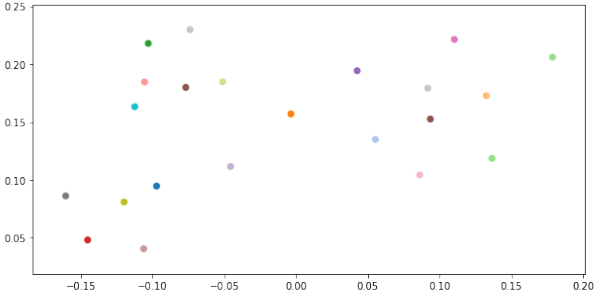
Agrupamiento Jerárquico

Divisivo

Mac.-Smith



K-means



Agrupamiento Jerárquico

Divisivo

Ventajas

- Intuitivo
- Funciona con clústeres de diferente tamaño
- Diferentes criterios y maneras de dividir
- Puede funcionar con diferentes medidas de distancia

Agrupamiento Jerárquico

Divisivo

Desventajas

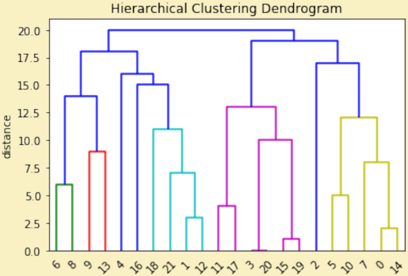
- Muy lento
- Problemas al lidiar con clústeres de diferente densidad
- Dos decisiones de división

Agrupamiento Jerárquico

Divisivo

Desventajas

- Muy lento
- Problemas al lidiar con clústeres de diferente densidad
- Dos decisiones de división
- ¿Qué partición elegir?
  - Número de clústeres concreto (fijando K)
  - Máxima distancia en la unión de clústeres



Gracias

Universidad
Internacional
de Valencia