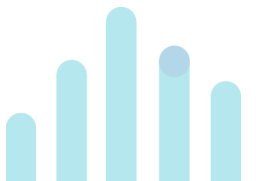
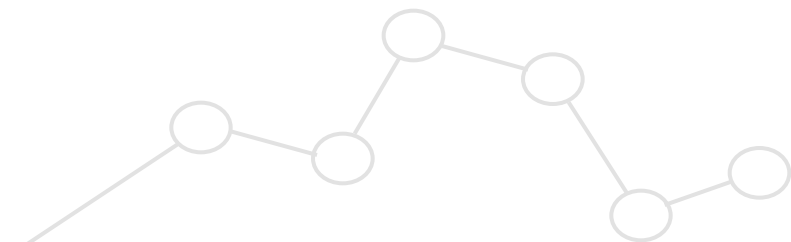


# MACHINE LEARNING INMERSION

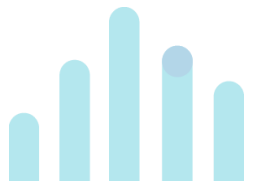
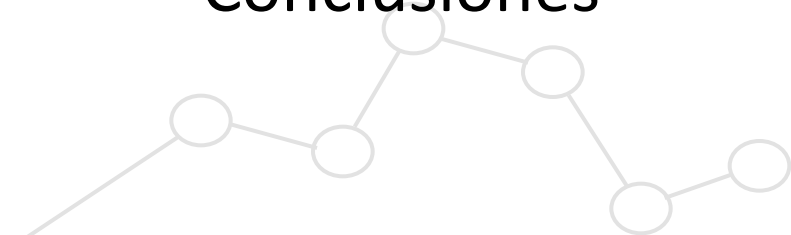
ANDRÉ OMAR CHÁVEZ PANDURO

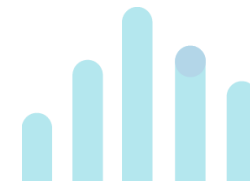
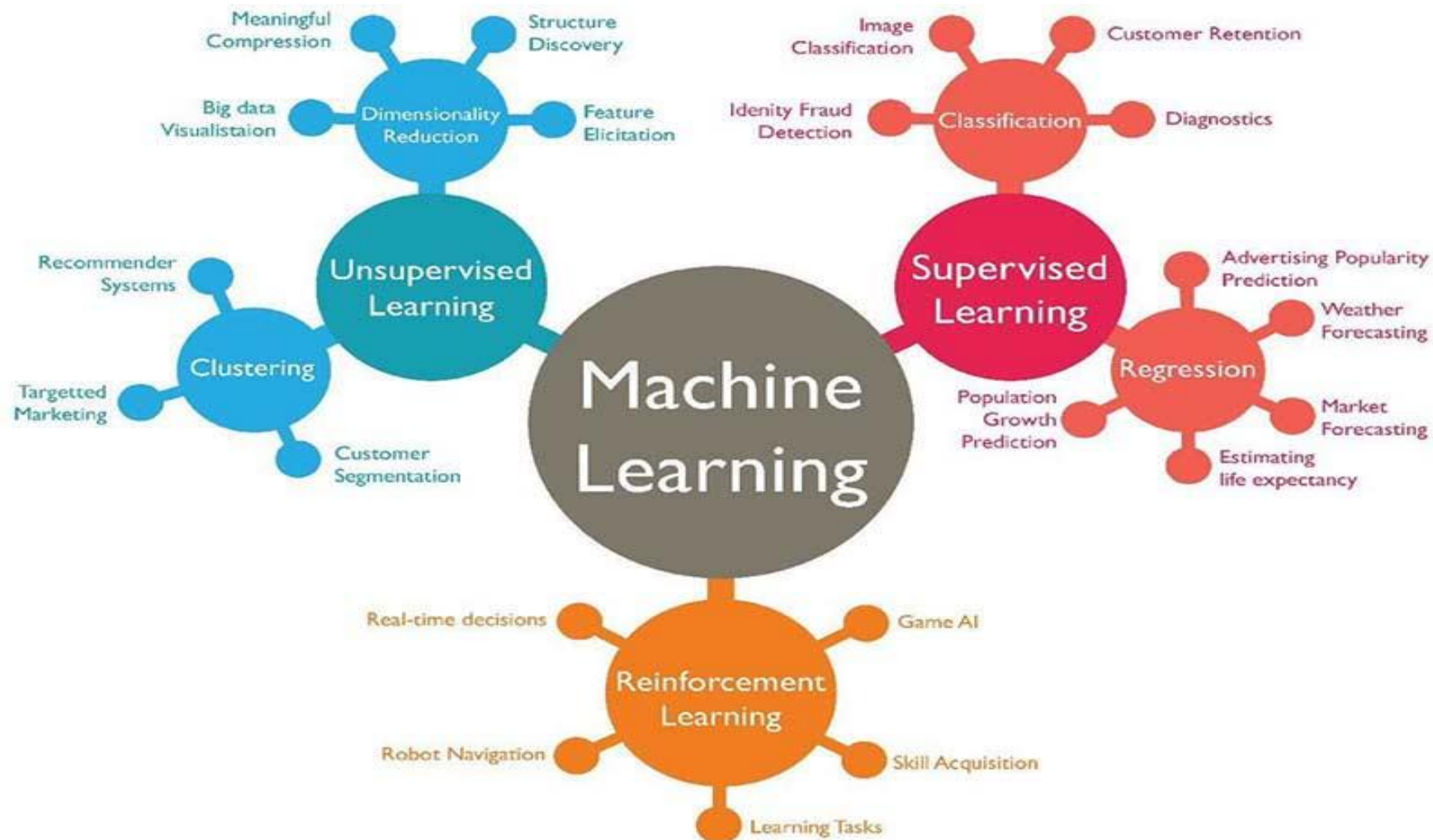
« Si el **Plan** no funciona ,  
cambia el **Plan** pero **No Cambies** los  
**OBJETIVOS NI METAS**»



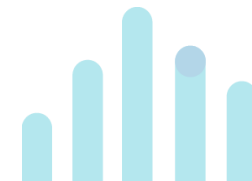
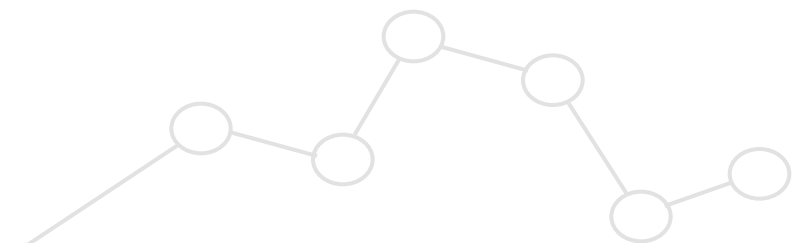
# Agenda

- Segmentación de Clientes
- Modelos No Supervisados
- Análisis de Conglomerados :Objetivos
- Análisis de Conglomerados : Criterio de Inercia
- Tipos de Algoritmos de Segmentación.
- Cluster Jerárquicos.
- Cluster No Jerárquicos.
- Algoritmo de K-Means: Objetivo, Método y Elección de k.
- Conclusiones



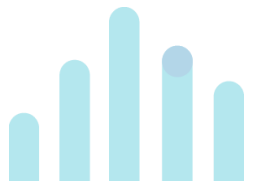
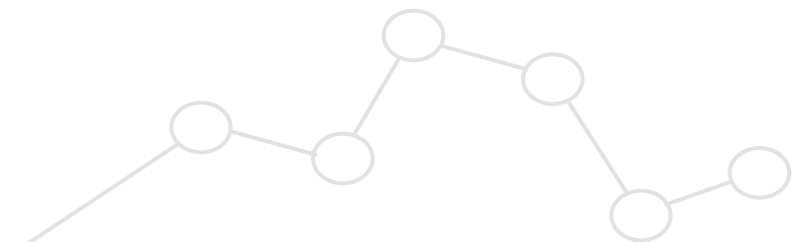


# SEGMENTACIÓN DE CLIENTES



# SEGMENTACIÓN DE CLIENTES

- Es el proceso de dividir **clientes** en grupos basados en características comunes para que las compañías puedan mercadear cada grupo efectiva y apropiadamente.
- Los grupos o segmentos deben ser homogéneos intragrupos y heterogéneos intergrupos.





# Modelos No Supervisados

- No hay una variable objetivo (Variable de Salida).
- No hay variables que ayudan a predecir a la variable de salida.



- Todas las variables tienen la misma importancia.
- Se busca la interdependencia de las variables.

# Modelos no Supervisados



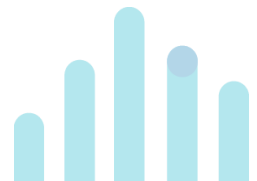
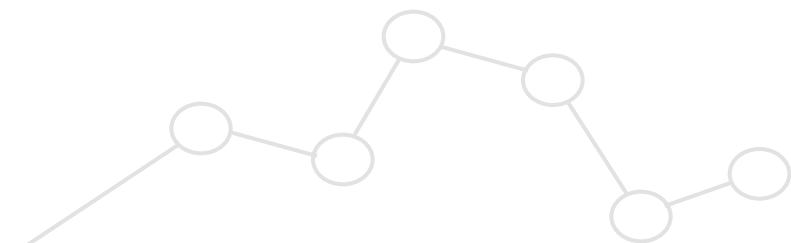


# Métodos de agrupamiento o clustering

- **“Clustering”**: (Clasificación no supervisada, aprendizaje no supervisado): Es similar a la clasificación (discriminación), excepto que los grupos no son predefinidos. El objetivo es **particionar** o **segmentar** un conjunto de datos o individuos en grupos que pueden ser disjuntos o no. Los grupos se forman basados en la similaridad de los datos o individuos en ciertas variables. Como los grupos no son dados a priori el experto debe dar una interpretación de los grupos que se forman.

- Métodos:

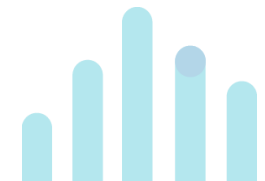
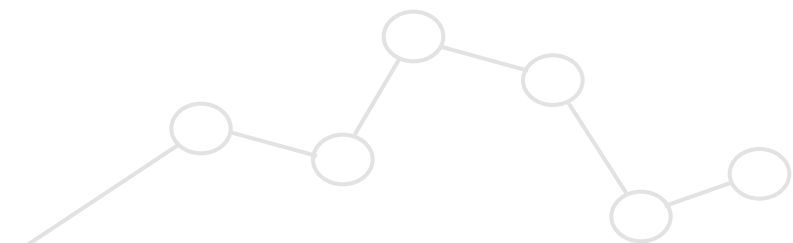
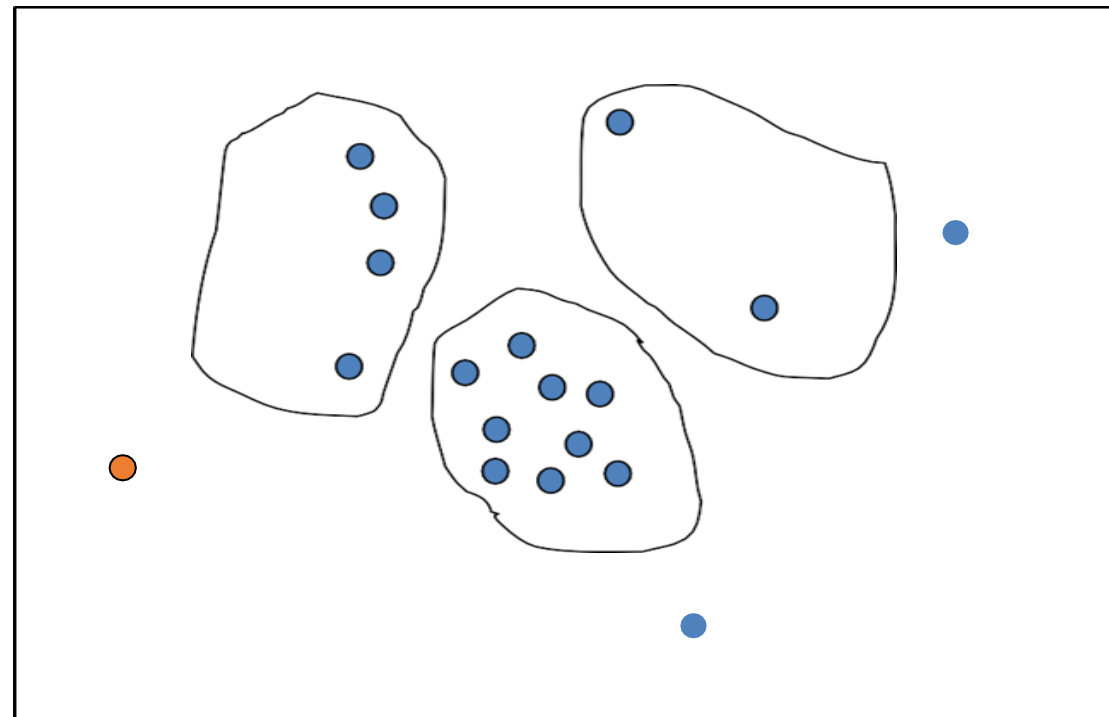
- Clasificación Jerárquica (grupos disjuntos).
- Nubes Dinámicas – k-means (grupos disjuntos).
- Clasificación Piramidal (grupos NO disjuntos).



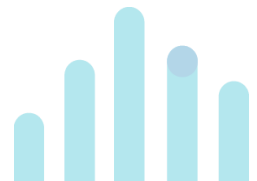
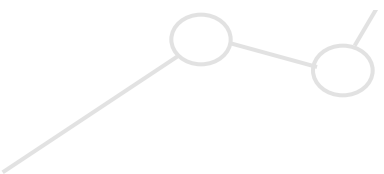
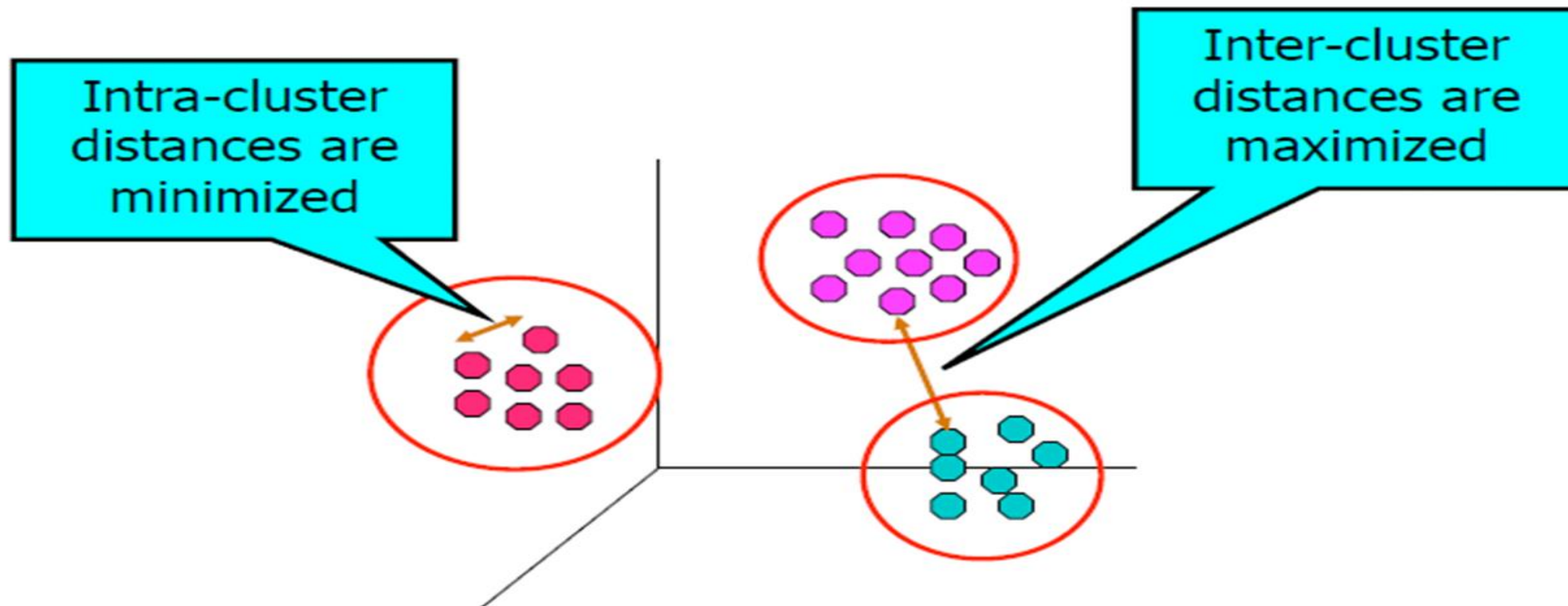
# Análisis de Conglomerados : Objetivo

## Objetivo:

Obtener clases lo más homogéneas posibles y tal que estén suficientemente separadas.



# Análisis de conglomerados: Criterio de la Inercia



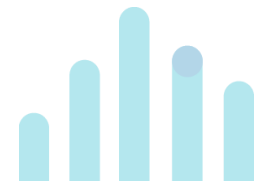
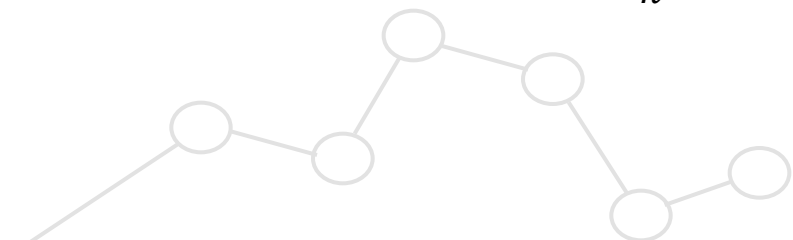
# Análisis de conglomerados: Criterio de la Inercia

- Como se ha mencionado , se quiere obtener clases lo más homogéneas posibles y que estén suficientemente separadas. Este objetivo se puede concretar numéricamente a partir de la siguiente propiedad :
- Supóngase que se está en presencia de una partición  $P=(C_1, C_2, C_3, \dots, C_k)$  de  $\Omega$  , donde  $g_1, g_2, g_3, g_4, \dots, g_k$  son los centros de gravedad de las clases:

$$g_k = \frac{1}{|C_k|} \sum_{i \in C_k} x_i$$

- $g$  es el centro de gravedad total:

$$g = \frac{1}{n} \sum_{i=1}^n x_i$$



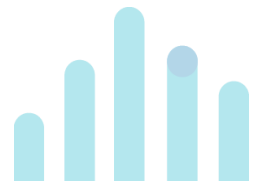
# Análisis de conglomerados: Criterio de la Inercia

- **Inercia total** de la nube de puntos:

$$I = \frac{1}{n} \sum_{i=1}^n ||\mathbf{x}_i - \mathbf{g}||^2$$

- **Inercia inter-clases**, es decir la inercia de los centros de gravedad respecto al centro de gravedad total:

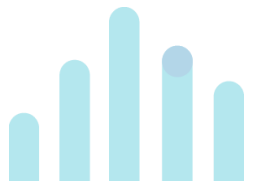
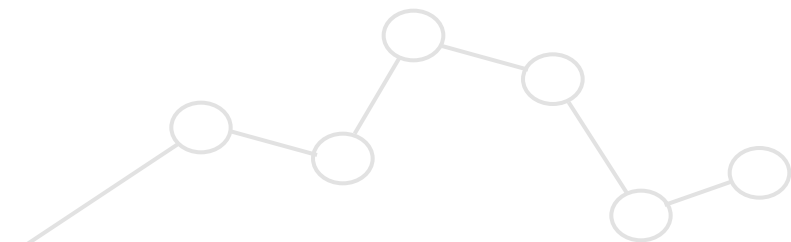
$$B(P) = \sum_{k=1}^K \frac{|C_k|}{n} ||\mathbf{g}_k - \mathbf{g}||^2$$



# Análisis de conglomerados: Criterio de la Inercia

- *INERCIA INTRA-CLASES, ES DECIR LA INERCIA AL INTERIOR DE CADA CLASE :*

$$W(P) = \sum_{k=1}^K I(C_k) = \frac{1}{n} \sum_{k=1}^K \sum_{i \in C_k} \|\mathbf{x}_i - \mathbf{g}_k\|^2$$



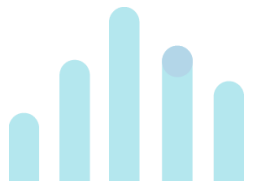


# Análisis de conglomerados: Criterio de la Inercia

## Teorema: Igualdad de Fisher

$$\begin{aligned} \text{Inercia total} &= \text{Inercia inter - clases} \\ &+ \\ &\text{Inercia intra-clases} \end{aligned}$$

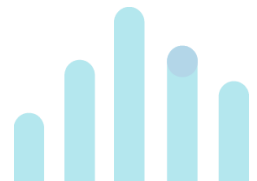
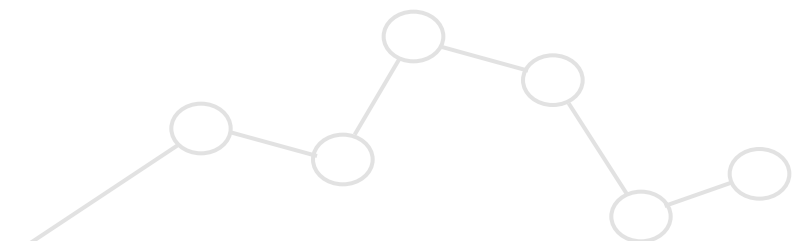
$$I = B(P) + W(P)$$



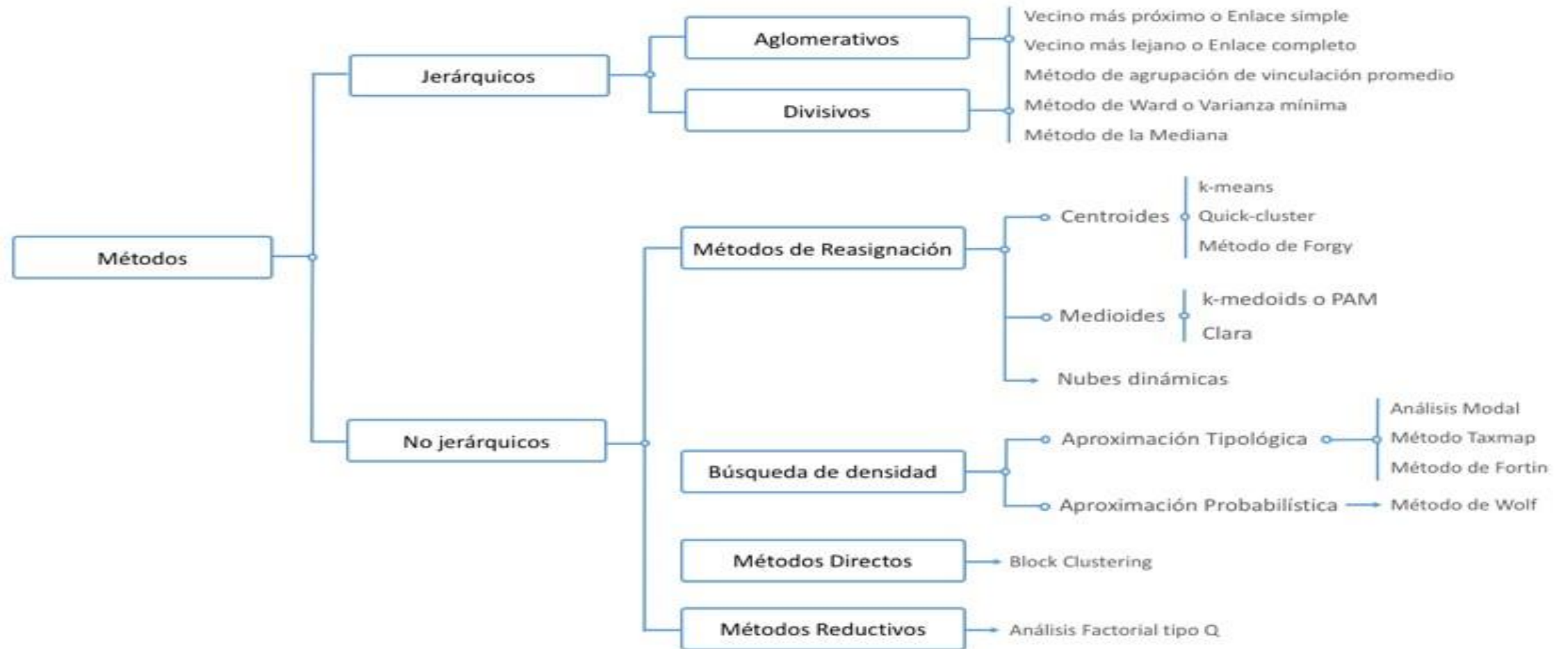
# Objetivos del Clustering

**Objetivo:** Se quiere que  **$B(P)$**  sea máxima y  **$W(P)$**  sea mínima.

- ✓ Como la inercia  $I(P)$  es fija, dada la nube de puntos, entonces al maximizar  $B(P)$  se minimiza automáticamente  $W(P)$ .
- ✓ Por lo tanto, los dos objetivos (homogeneidad al interior de las clases y separación entre las clases) se alcanzan al mismo tiempo al querer minimizar  $W(P)$ .



# Tipos de Clustering

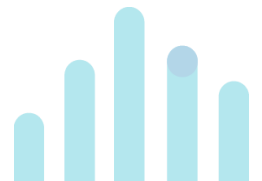
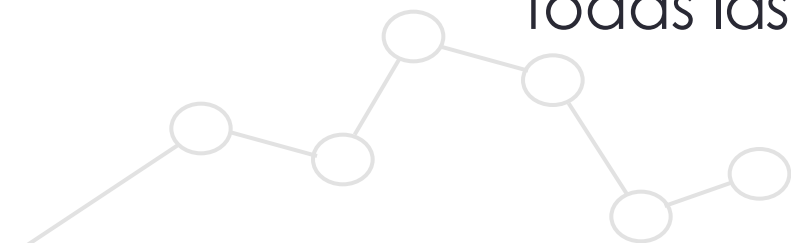


# CLUSTER JERARQUICOS

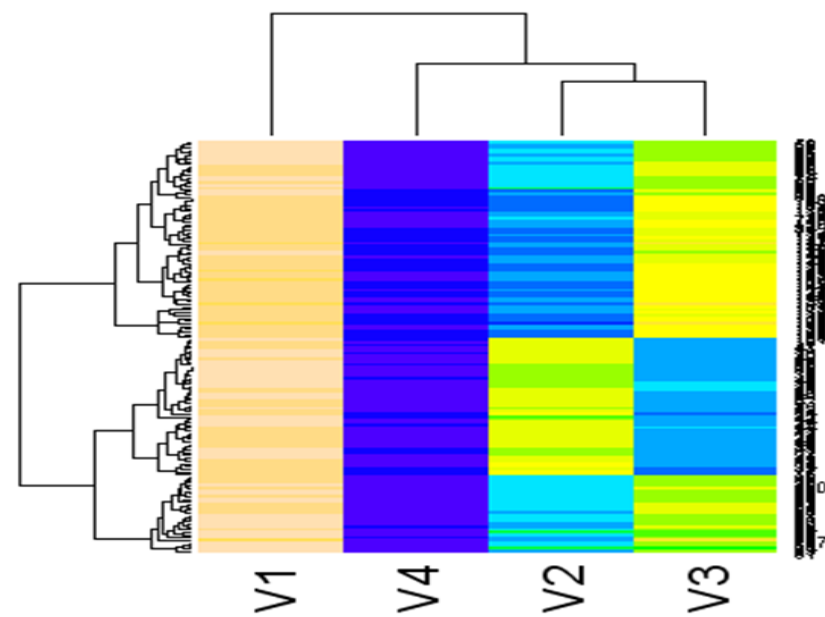
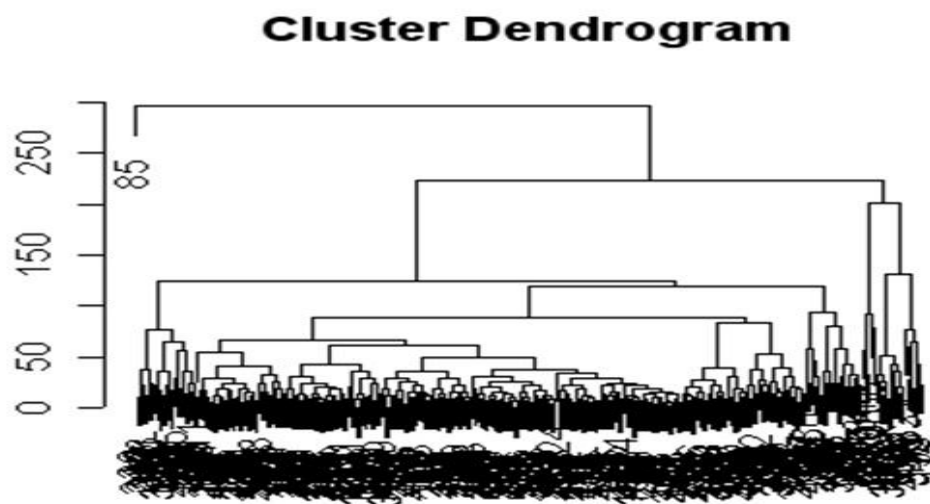
- En estos algoritmos se generan sucesiones ordenadas (jerarquias) de conglomerados. Puede ser juntando cluster pequeños en mas grande o dividiendo grandes clusters en otros mas pequeños. La estructura jerárquica es representada en forma de un árbol y es llamada **Dendograma**.

Se dividen en dos tipos:

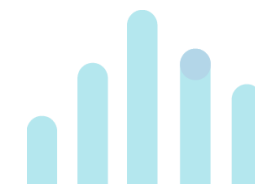
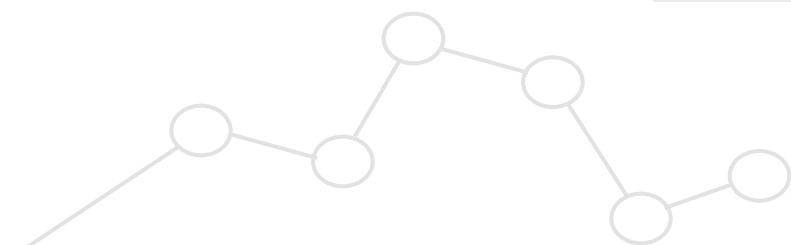
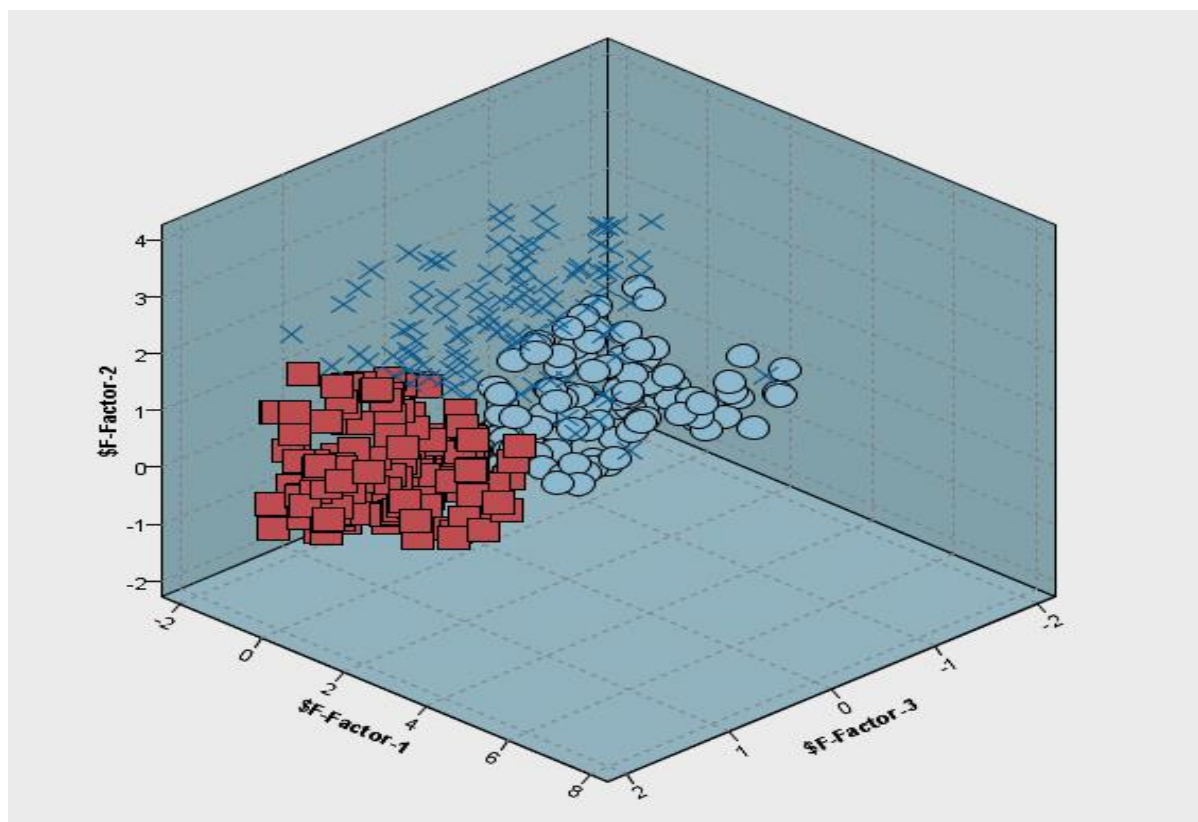
- **Algoritmos jerárquicos aglomerativos** (bottom-up, inicialmente cada instancia es un cluster). AGNES
- **Algoritmos jerárquicos divisivos** (top-down, inicialmente todas las instancias estan en un solo cluster). DIANA.



# Dendrograma



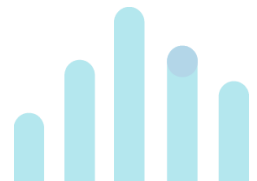
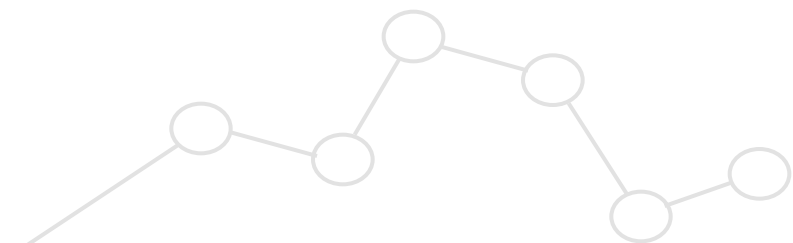
# ALGORITMO DE K - MEANS





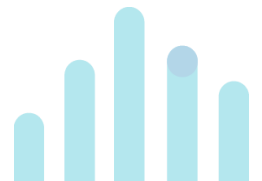
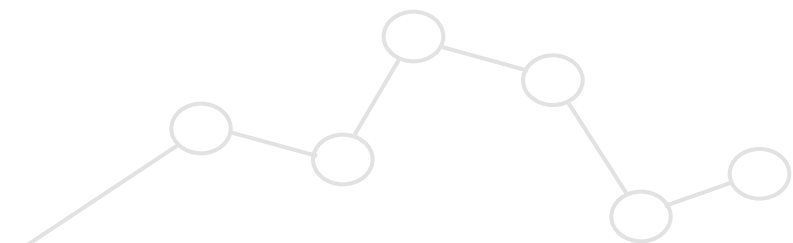
# Objetivo del Método K-means

- Así, el objetivo en el método de K-means es encontrar una partición  $P$  de  $\square$  y representantes de las clases, tales que  $W(P)$  sea mínima.



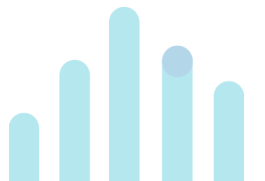
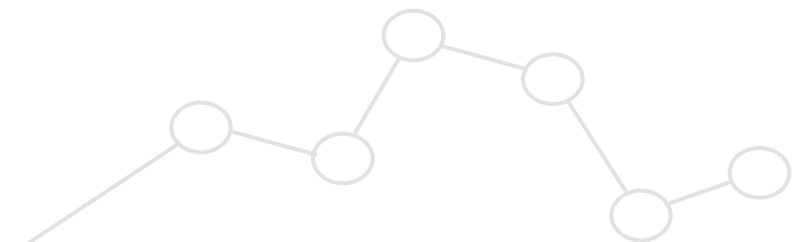
# Método de K - Means

- ✓ Existe un poco de confusión en la literatura acerca del método de las k-medias, ya que hay dos métodos distintos que son llamados con el mismo nombre.
- ✓ Originalmente, Forgy propuso en 1965 un primer método de reasignación-recentraje que consiste básicamente en la iteración sucesiva, hasta obtener convergencia, de las dos operaciones siguientes:

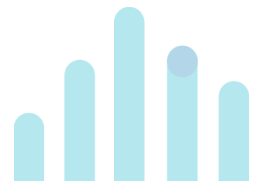
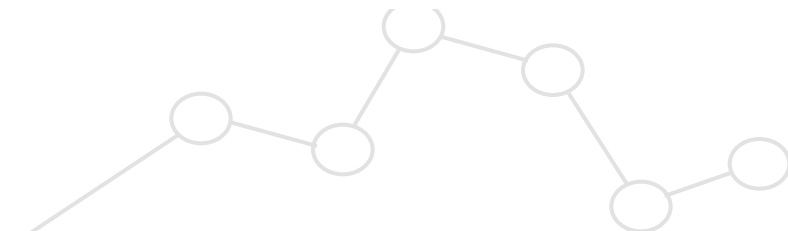
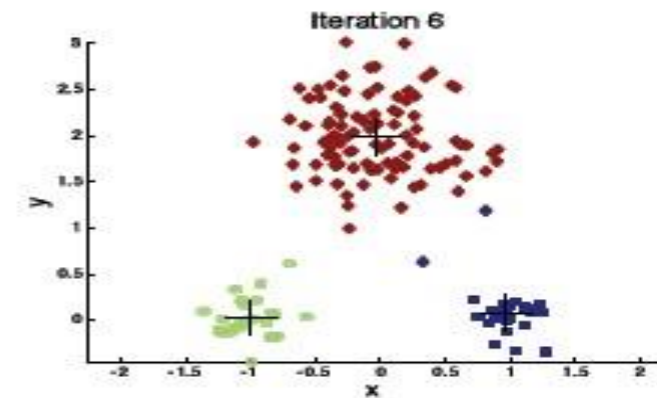
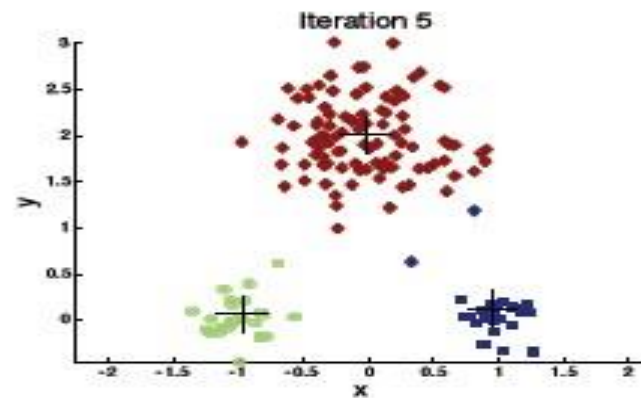
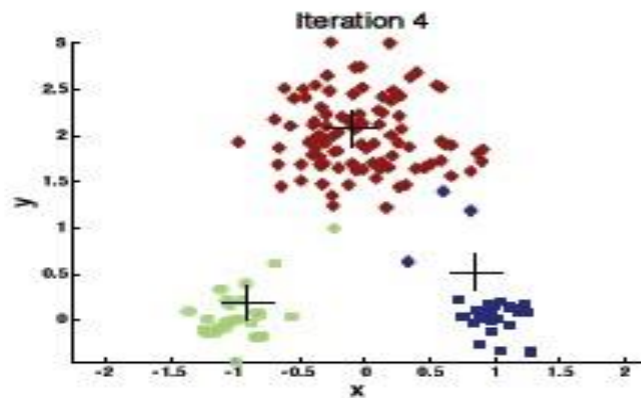
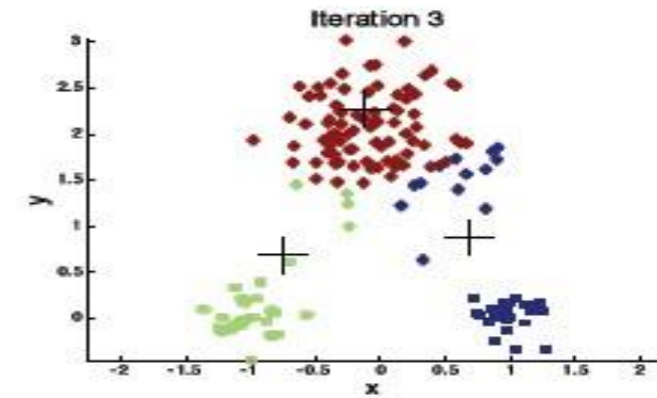
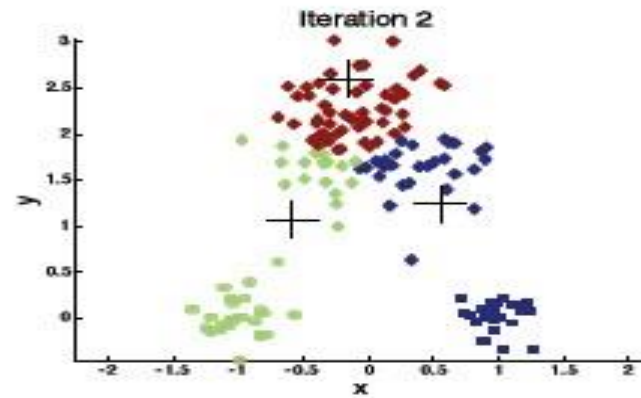
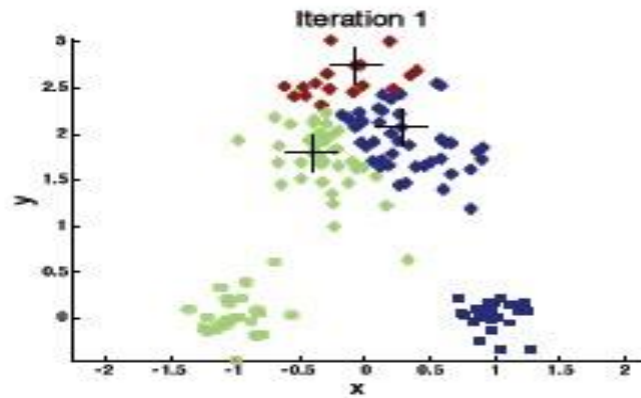


# Método de K – Means : Proceso

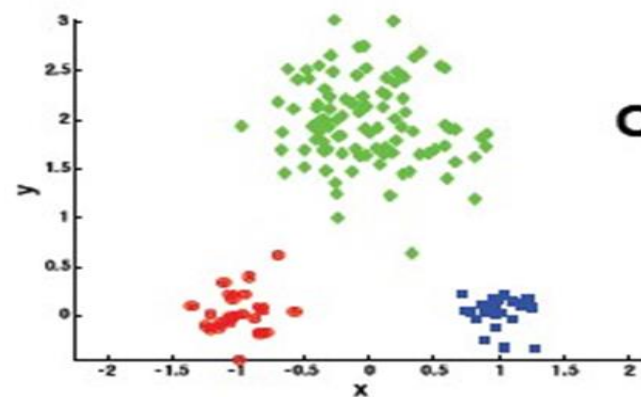
- Representar una clase por su centro de gravedad, esto es, por su vector de promedios.
- Asignar los objetos a la clase del centro de gravedad más cercano.



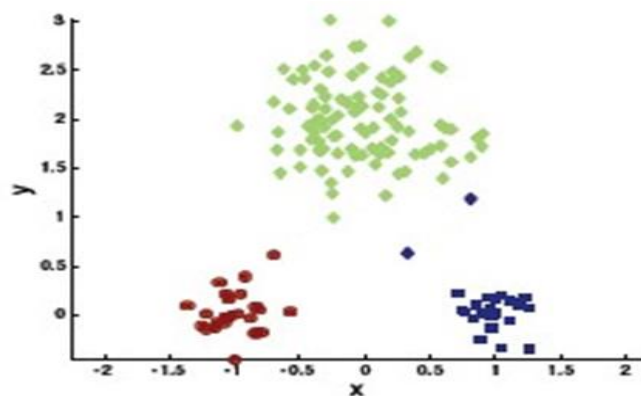
# Método de K - Means



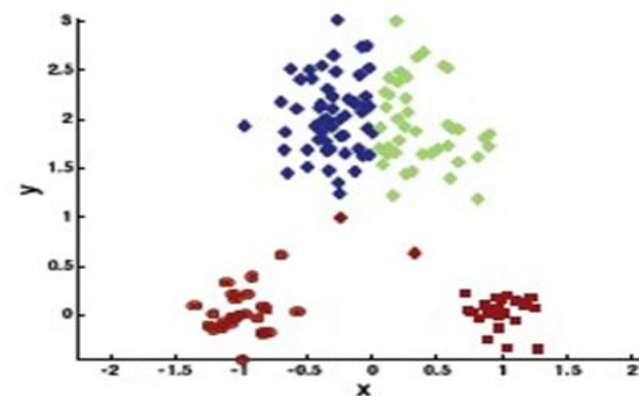
# Método de K - Means



**Original Points**

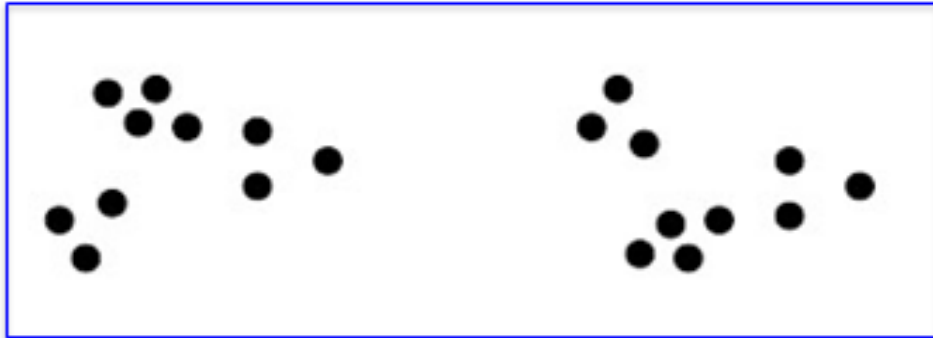


**Optimal Clustering**



**Sub-optimal Clustering**

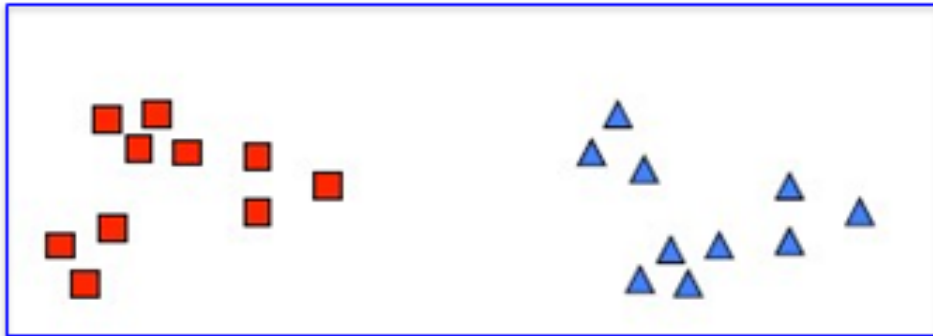
# ¿ Cuántos clústeres?



**Datos originales**



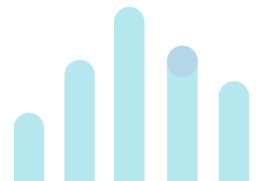
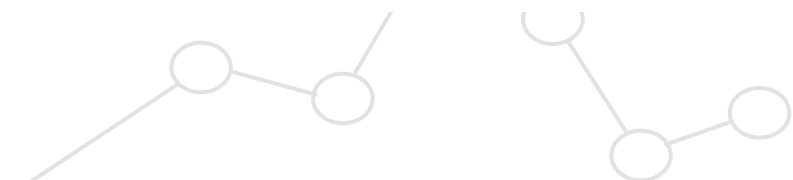
**6 clústeres**



**2 clústeres**



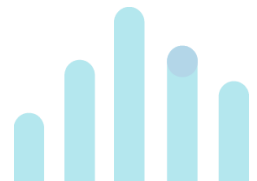
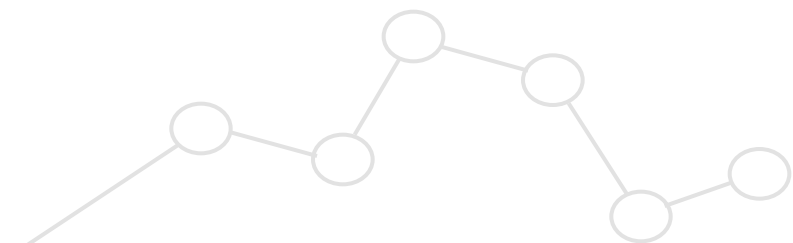
**4 clústeres**





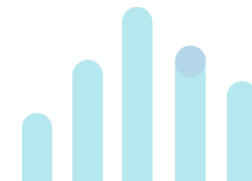
# Elección de $K$ : Problema Combinatorio

- Es necesario hacer notar que, cuando se quiere obtener una partición en  $K$  clases de un conjunto con  $n$  individuos, no tiene sentido examinar *todas* las posibles particiones del conjunto de individuos en  $K$  clases.
- En efecto, se está en presencia de un problema combinatorio muy complejo; sólo para efectos de ilustración, mencionemos que el número de particiones en 2 clases de un conjunto de 60 elementos es aproximadamente  $10^{18}$  y para 100 elementos en 5 clases anda por  $10^{68}$ .



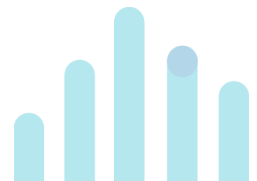
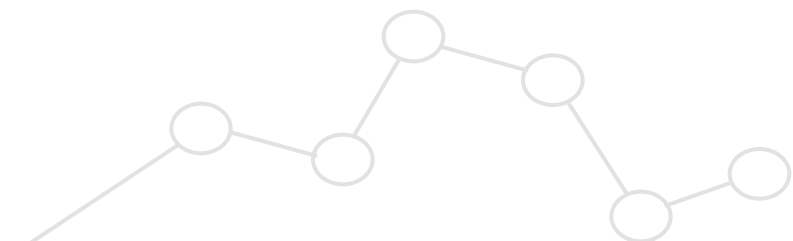
# Conclusiones

- Aunque la selección inicial en el algoritmo toma tiempo extra, k-means converge muy rápidamente después de la selección de puntos iniciales.
- Los autores probaron su método con conjuntos de datos reales y sintéticos y obtuvieron mejoras de 2-veces en la velocidad, y para ciertos conjuntos de datos, cerca de 1000 veces mejoras en error.



# A seguir estudiando.....

- **Fuzzy C-Means Clustering** es una versión difusa del K-means, donde cada punto tiene un grado difuso de pertenecía a cada grupo.
- **Modelos de Mezclas Gaussianas** entrenadas con el algoritmo de esperanza-maximización presentan una asignación probabilística a cada grupo, en vez de asignaciones deterministas.
- El algoritmo **Spherical K-means** es bastante usado para datos direccionales.
- El algoritmo **Minkowski Metric Weighted K-means** trata el problema del ruido asignando pesos a las componentes de los vectores por grupos.
- Algoritmos de Clustering basados en medianas en lugar de medias: **PAM, CLARA, CLARANS.**



**GRACIAS  
POR SU ATENCIÓN**