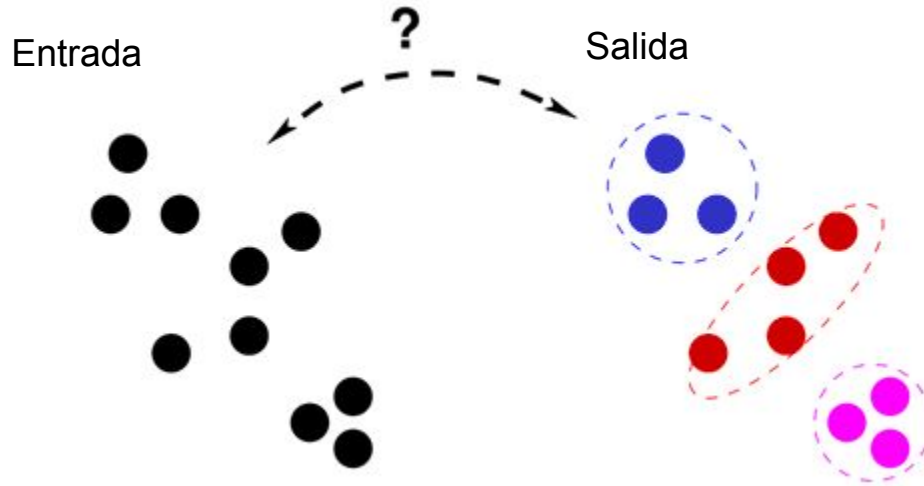


Aprendizaje no supervisado

Aprendizaje no supervisado

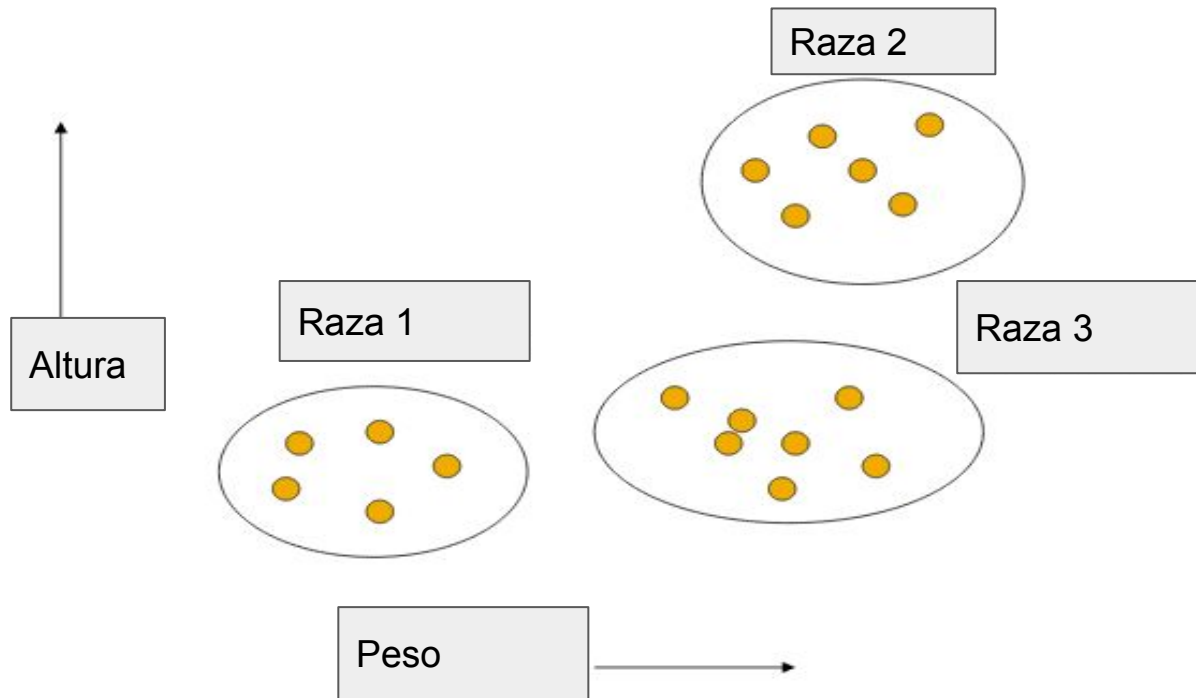
- No hay etiquetas (no hay variables de salida), sólo variables de entrada.
- Agrupamiento de datos



Clustering - enunciado del problema

- Dado un conjunto de puntos (datos de entrenamiento) agruparlo en clusters de tal forma que:
 - Puntos en un cluster son similares unos a otros.
 - Puntos en diferentes clusters son disimilares.
- Los puntos están generalmente en un espacio de alta dimensionalidad, por lo que se definen métricas de distancia para medir la similaridad.
 - Distancia euclidiana
 - Distancia de coseno
 - Edit distance

Ejemplo:



Métodos de clustering

1. Jerárquicos o de aglomeración

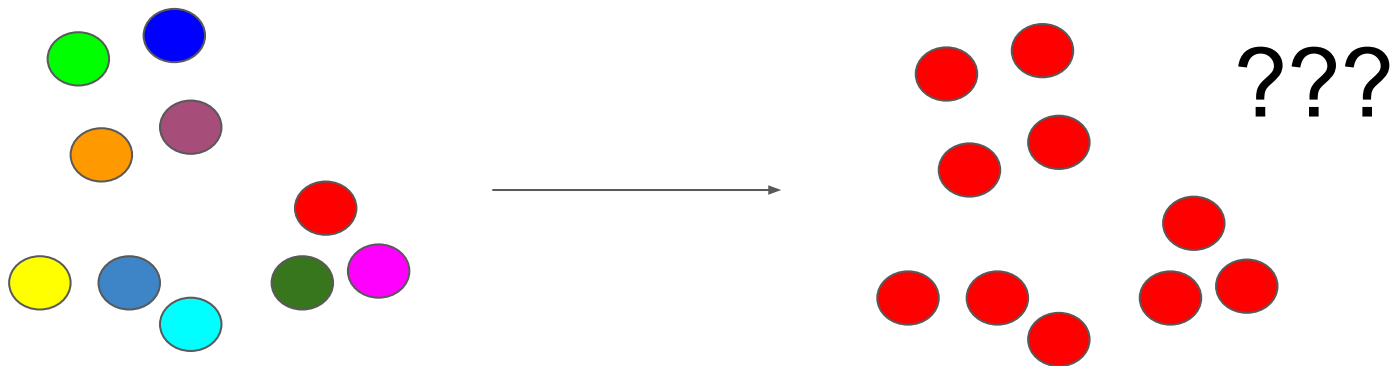
- Cada punto inicialmente es un cluster
- A cada paso se van juntando dos cluster en uno basados en alguna medida de similaridad

2. Por clusters definidos (asignación de puntos)

- Se tiene un número fijo de clusters
- Puntos se van asignando al cluster que se encuentra más cercano a ellos.

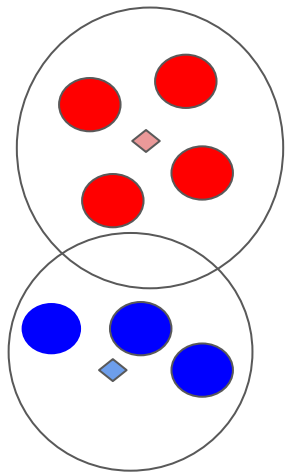
Clustering Jerárquico

- Preguntas a responder:
 - Cómo representar un cluster de más de un punto ?
 - Cuál medida de similaridad utilizar?
 - Cuando parar de combinar clusters?



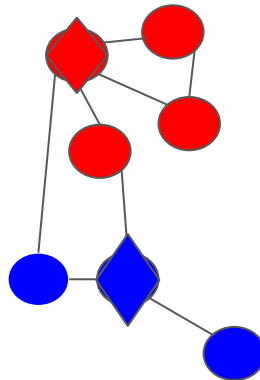
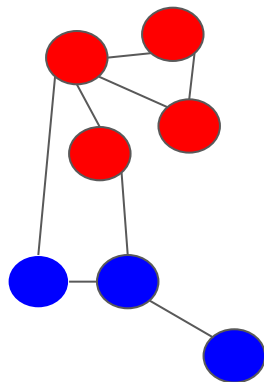
Representación de clusters

- En el espacio Euclidiano -> centroides
 - Cada cluster se puede representar por su centroide.
 - Distancia entre clusters = distancia entre sus centroides.



Representación de clusters

- En el espacio no Euclidiano -> clustroid
 - Como no hay média de puntos, un punto tiene que ser el “centroide”
 - Se toma el punto que se encuentra “más cercano” a los otros puntos de su cluster



Medidas de similaridad

- Estas distancias se calculan entre clusters
 - Distancia euclidiana
 - Distancia de coseno
 - Edit distance
 - Etc, etc, etc

Cohesión

- Se busca la mayor cohesión entre puntos de un mismo cluster
- Fusionar clusters cuya unión tengan la mayor cohesión.

Criterio de parada

1. Parar cuando se alcanza un mínimo número de clusters
2. Parar cuando la mejor cohesión de dos clusters que se van a combinar sea menor a un umbral determinado.
3. Parar cuando la cohesión sufre un cambio muy notorio de una iteración a otra.

Implementación de clustering jerárquico.

1. Cada punto es un cluster.
 2. Computar las distancias entre cada par de clusters
 3. Juntar aquel o aquellos que presenten la mejor cohesión.
 4. Ejecutar 2. y 3. hasta que un criterio de parada sea encontrado.
-
- Implementación fuerza bruta (n^3)
 - Utilizando colas de prioridad ($n^2 \log(n)$)
 - Son lentos en general.

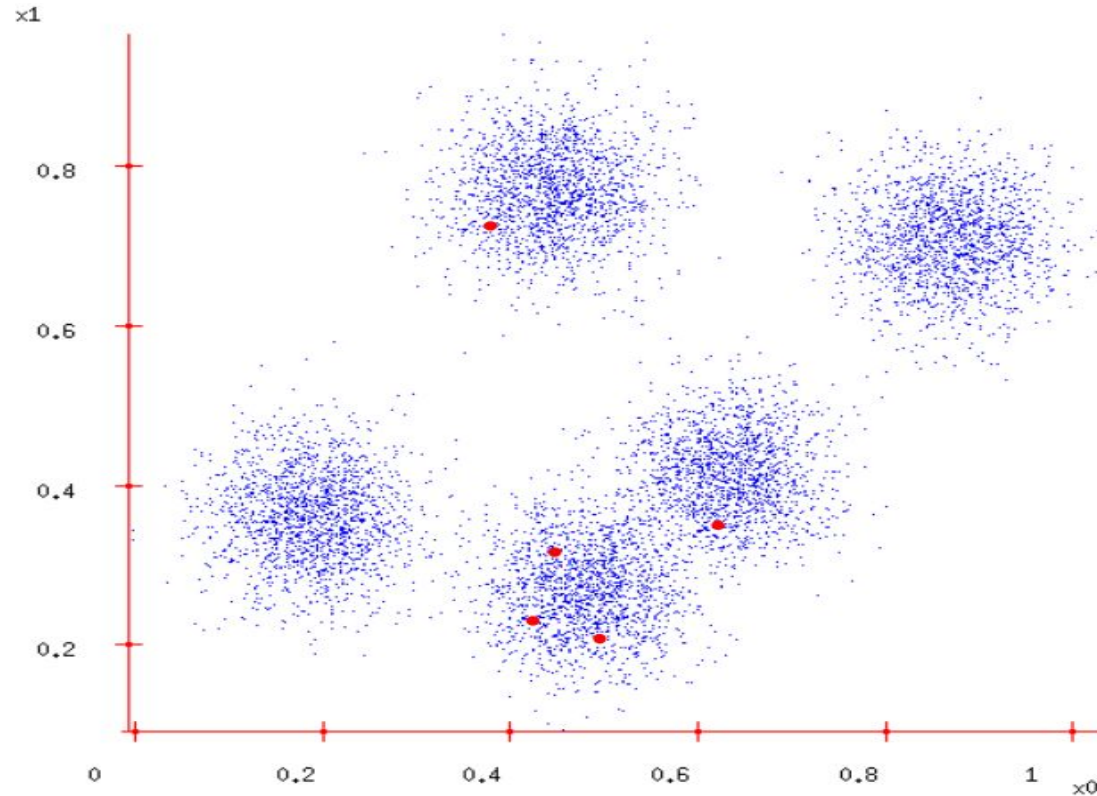
Por asignación de puntos (k-means)

- K-means es el algoritmo más conocido en clustering.
- Asume que se está trabajando en un espacio euclidiano.
- Recibe como entrada el número K de clusters.
- Los pasos en las siguientes diapositivas.

1) Escoger los K-centroides iniciales

- Enfoques:
 - Escoger aleatoriamente K puntos de la data de entrada.
 - Escoger un punto aleatoriamente de la data de entrada y luego K-1 puntos que se encuentren los más alejado posible de los puntos previos escogidos.

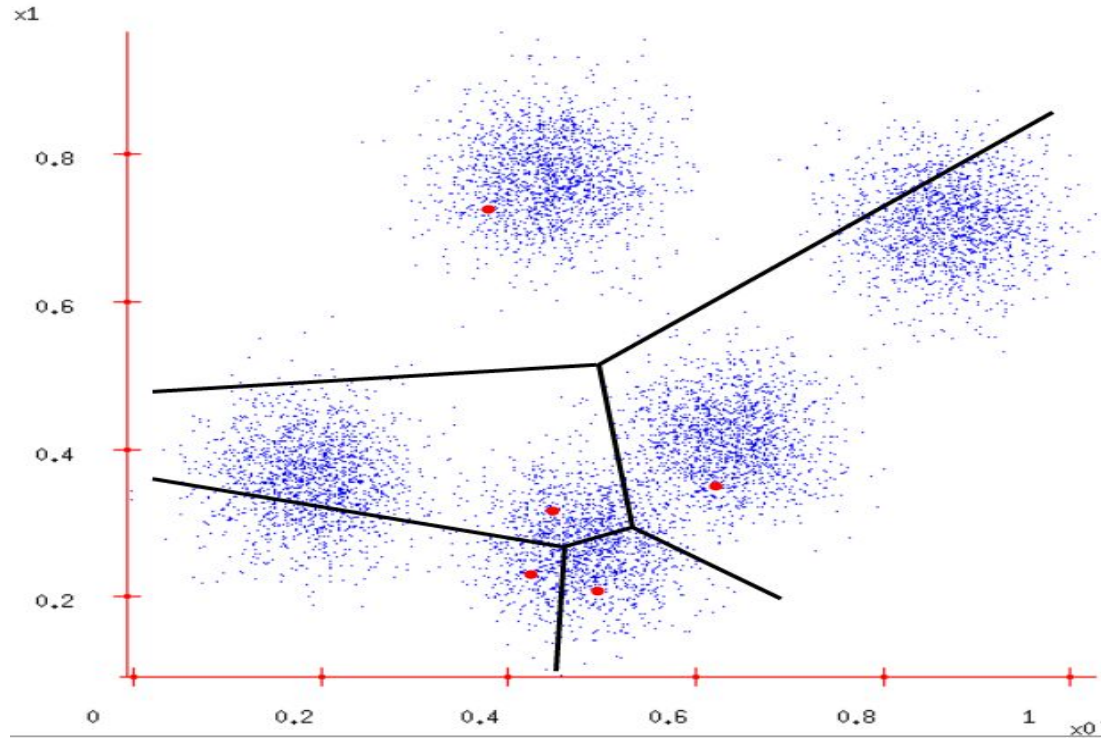
1) Escoger los K-centroides iniciales



2) Asignar cada punto al centroide más cercano

- Utilizando distancia euclidiana o alguna otra medida, cada punto será asignado al centroide más cercano

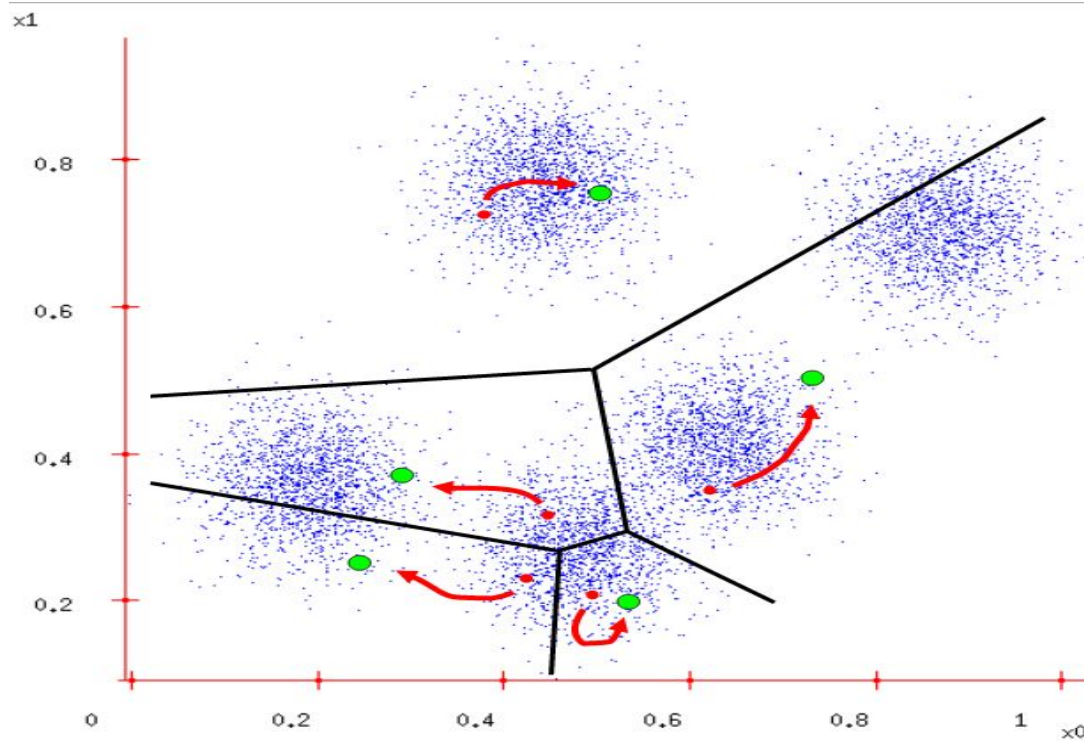
2) Asignar cada punto al centroide más cercano



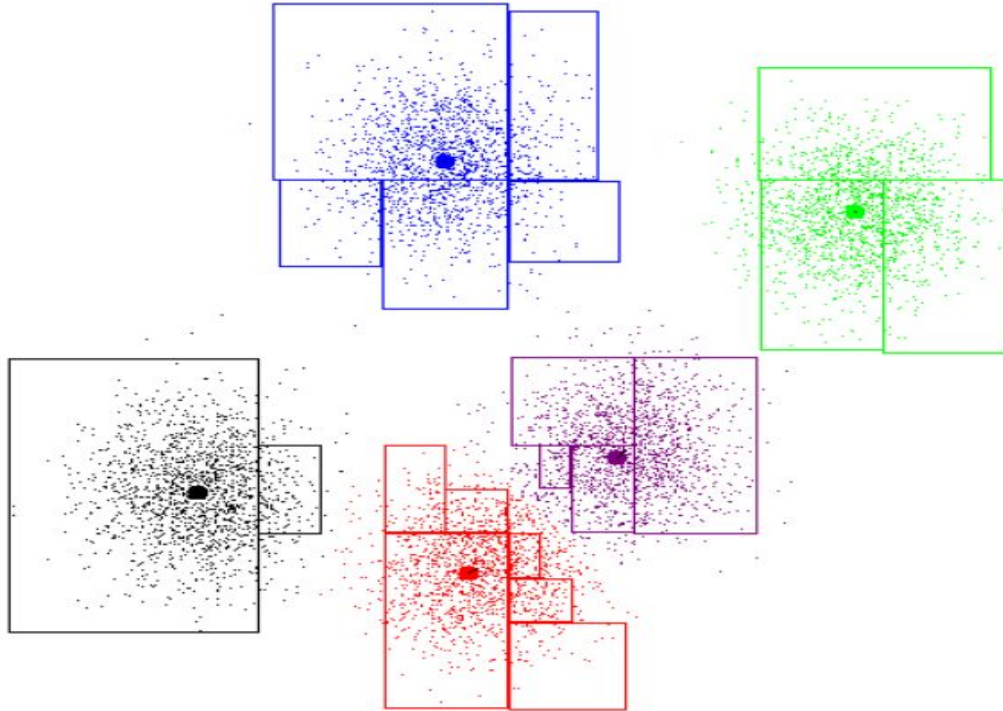
3) Actualizar los K centroides

- Calcular la media aritmética de todos los puntos de un cluster.
- El nuevo centroide del cluster será esa media.

3) Actualizar los K centroides



4) Iterar entre los pasos 2 y 3 hasta convergencia



Función Costo (J) - (Distorsión)

$c^{(i)}$: Índice del cluster al que pertenece $x^{(i)}$ $x^{(i)} \rightarrow 5$ $c^{(i)} = 5$

μ_k : Centroide del cluster k (k = 1 K)

$\mu_{c^{(i)}}$: Centroide del cluster para el cual $x^{(i)}$ ha sido asignado $\mu_{c^{(i)}} = \mu_5$

$$J(c^{(1)}, \dots, c^{(m)}, \mu_1, \dots, \mu_K) = \frac{1}{m} \sum_{i=1}^m \|x^{(i)} - \mu_{c^{(i)}}\|^2$$

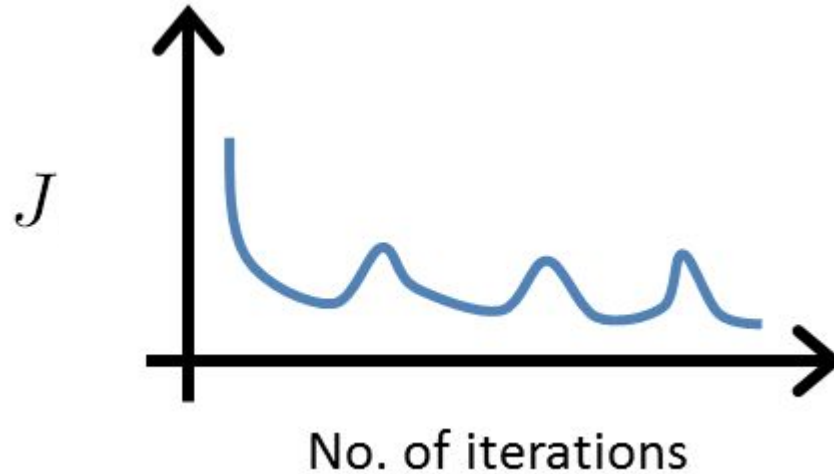
Función Costo (J)

$$\min_{\substack{c^{(1)}, \dots, c^{(m)}, \\ \mu_1, \dots, \mu_K}} J(c^{(1)}, \dots, c^{(m)}, \mu_1, \dots, \mu_K)$$

- J es minimizado con respecto a $c^{(1)}, \dots, c^{(m)}$ en el paso 2 del algoritmo
- J es minimizado con respecto a μ_1, \dots, μ_K en el paso 3 del algoritmo

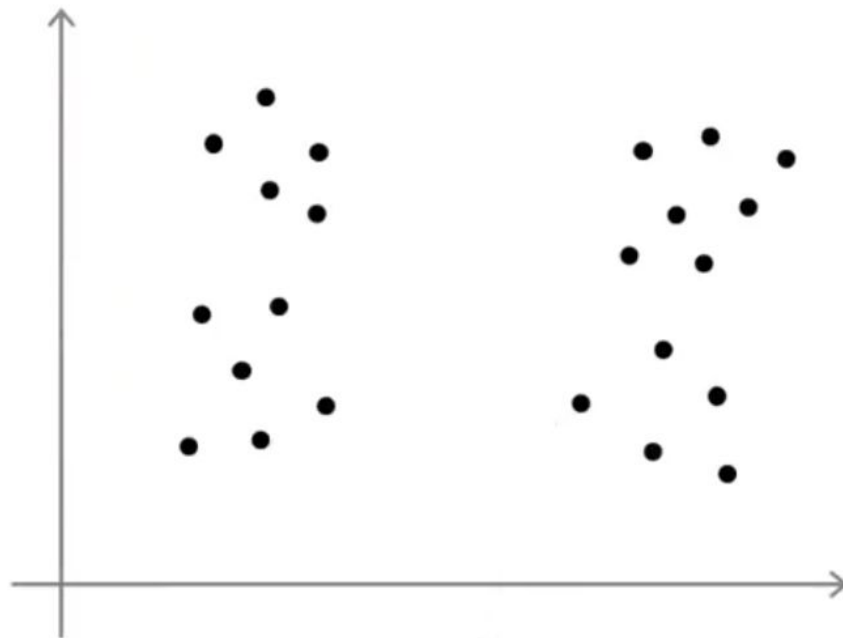
Función Costo J

- **Situación imposible:** Error en la implementación.



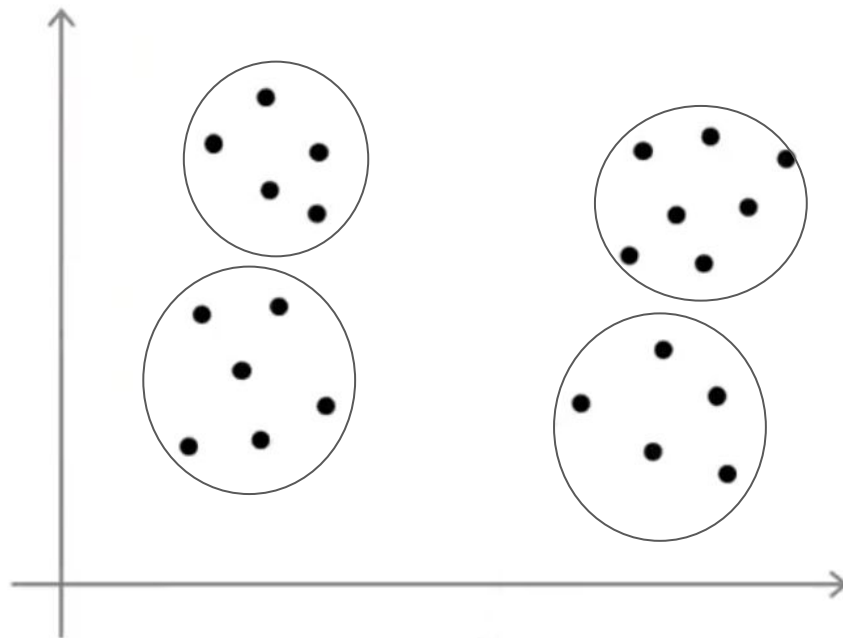
Encontrando K

- Escoger manualmente es la mejor opción.
- Es un problema ambiguo.



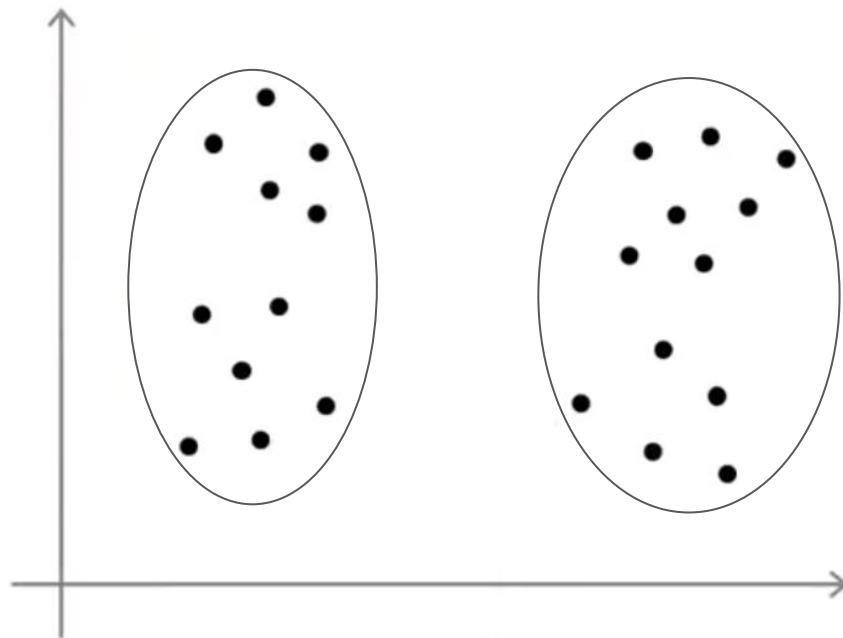
Encontrando K

- Escoger manualmente es la mejor opción.
- Es un problema ambiguo.

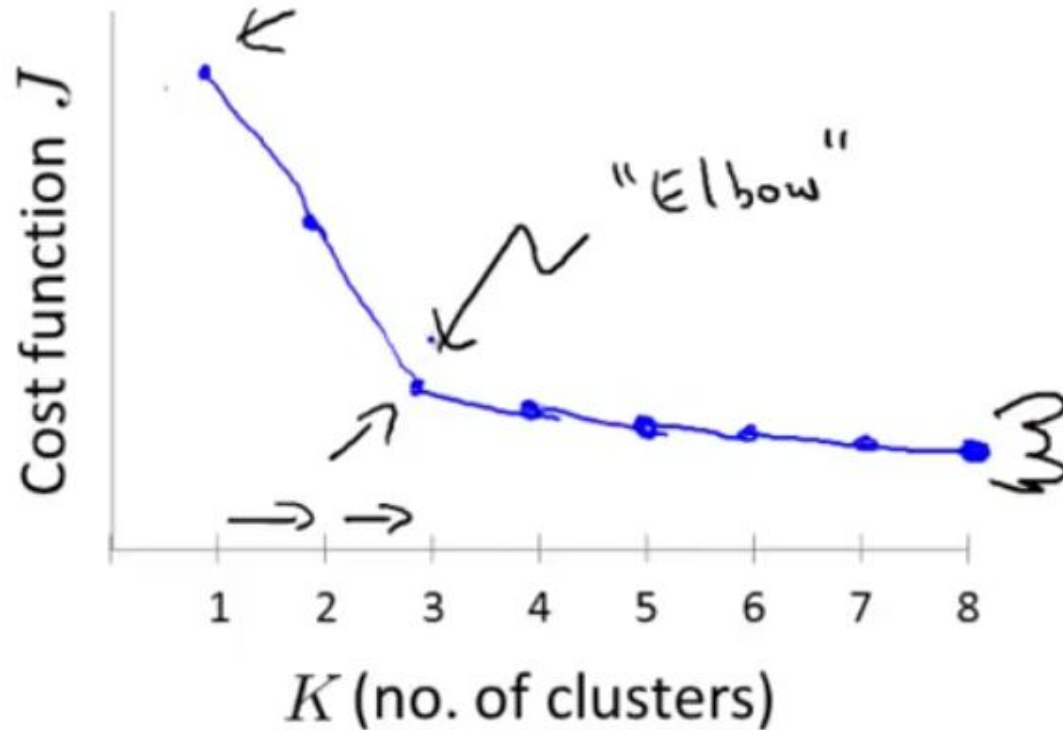


Encontrando K

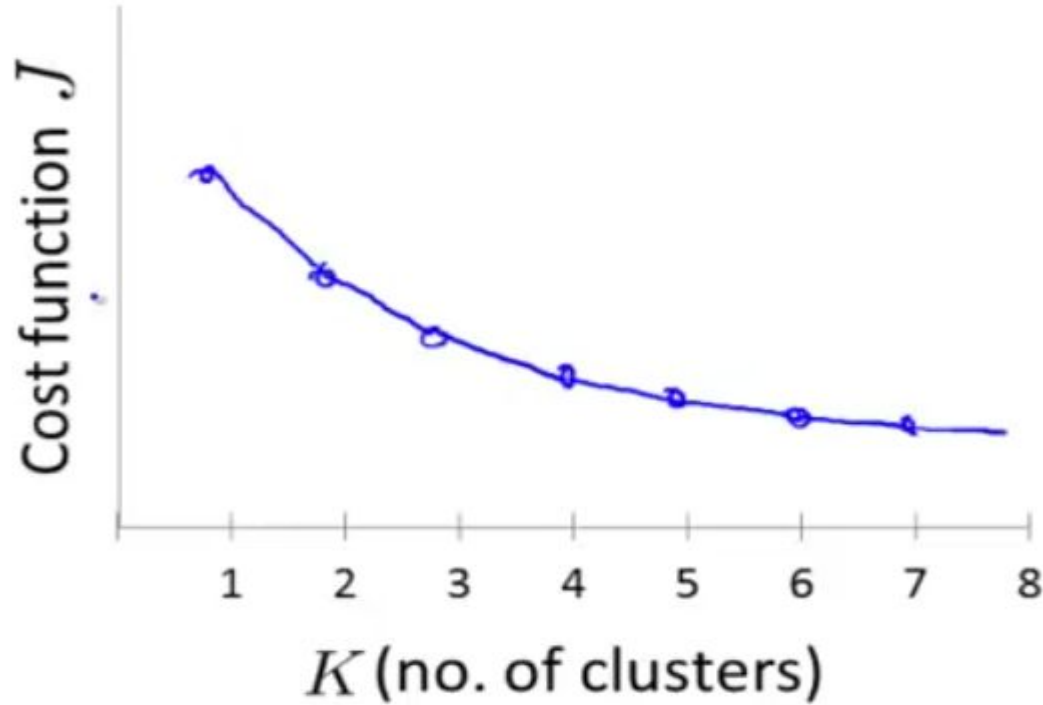
- Escoger manualmente es la mejor opción.
- Es un problema ambiguo.



Elbow method (método del codo)

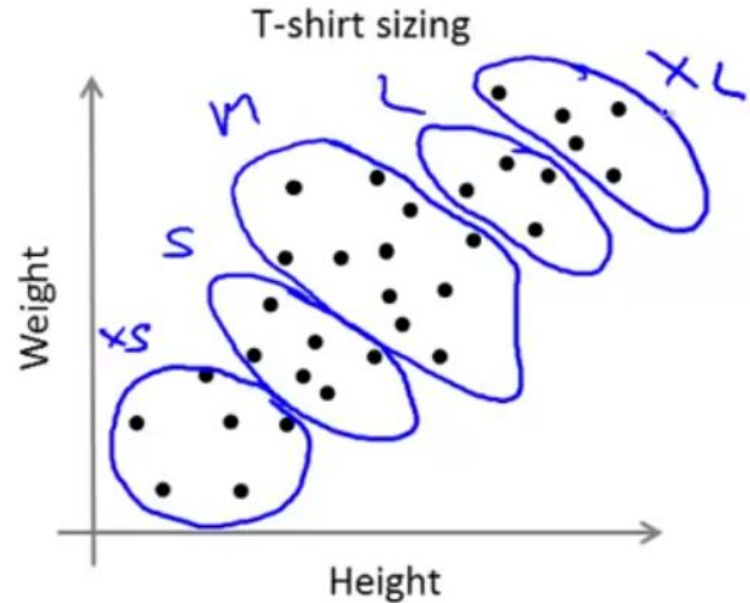
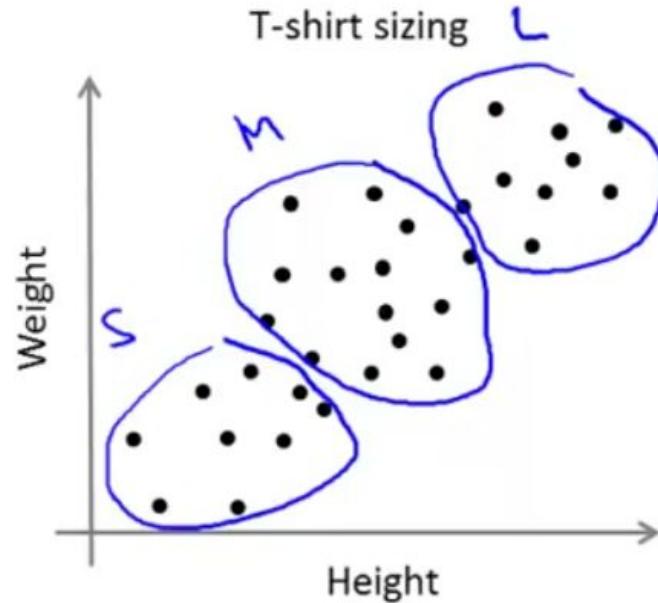


Elbow method - Ambigüedad!



Mejor forma de encontrar K

- Analizando K en base a la información del problema que se está resolviendo.



Mas algoritmos - Sklearn

