K-means : Algoritmo no supervisado.

===================================

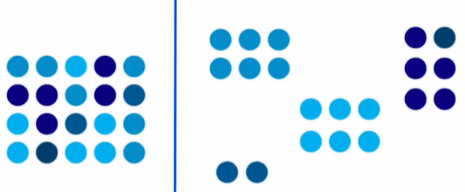
Agrupar los datos (crea k grupos) a partir de las similitudes que tenga.

Trata información que no tiene etiquetas asignadas y busca asignarla a un grupo segun caracteristicas

Estos datos se agrupan alrededor de centroides (k centroides). (el centro)

Para calcular el k adecuado recurrimos al viejo método de prueba y error.

Aproximación a K: método del codo



**QUE HACE K-MEANS?**

Te dice, con los datos de test, si dices, K=3 supones que hay 3 grupos, digame que exactitude tiene?

De modo que por prueba y error evaluas k=2,3,4,5,6,7,8…. Y evaluas si aumenta o no la exactitude

En el K que da mayor exactitude es el NUMERO de grupos mas similares, y con la exactitud dice que tan seguro es tener ese numero de grupos

Por ejemplo, problemos con K=2 , y nos da una exactitude baja, luego con K=3 nos da mayor exactitude, entonces nos va diciendo que no eran 2 grupos, sino 3, pero luego probemos con K=4, K=5, para evaluar cual es el que tiene mayor exactitud.

Eso implica, para una tienda, cuantos grupos similares de clientes se podria tener.

**LO COMPLICADO, CREAR EL ARRAY TARGET, ME PARECE.**

1) Selecciona un valor para K(centroides) metodo del codo

2) Asignamos a c/u de los elementos el centro mas cercano (con caracteristicas mas similares)

Aplicaciones:

============

Segmentación por comportamiento:

------------------------------------------------

por historial de compras

actividad en una aplicación móvil, web

Definir personas basadas en sus intereses.

Crear perfiles basado en el monitoreo de actividad.

Ordenando medidas de sensores:

------------------------------

Detecta tipos de actividades en sensores de movimiento.

Grupos de imágenes.

Separar audio.

Identificar grupos en monitoreo de salud.

Nota:

Existe un método gráfico conocido como el método del codo (elbow), en el que elegimos el punto de inflexión de la grafica de errores obtenidos en cada valor de k

A picture containing chart

Description automatically generated

Se asignan 3 centroides, personas que son

* Compran ropa
* Compran electronicos
* Compran musica

Son los centroides, y se van agrupando los demas a estos 3 centroides.

Algunos pueden ser ambos compradores, pero pueden crear una similitud mas cercana a uno de los centroides

Metodo del codo

- Calcular agrupamiento de diferentes de K

- El error al cuadrado para cada punto es el cuadrado de la distancia del punto desde su centro