## K-Means.

## Introducción:

K\_Means es un algoritmo de aprendizaje no supervisado (Es decir desconocemos los clases), para la creación de clusters. Este algoritmo se apoya en el uso de centroides K.

La principal idea de esta algoritmo es particionar las entradas de datos en k clusters. K means trabaja por iteraciones siempre modificando las posiciones de los centroides, los cuales son los identificadores de los centros de cada cluster.

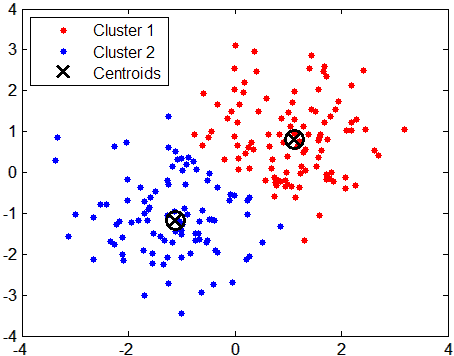


Imagen 1. Ejemplo Clusters dos dimensiones.

## Desarrollo:

-Para el desarrollo de este algoritmo se utiliza la base de datos Iris con las dos clases linealmente separables.

Utilizando los cuatro atributos de cada individuo se crea un HiperPlano de cuatro dimensiones.

1-Posicionamos todos los individuos es sus posiciones correspondientes dentro del HiperPlano.

2-Creamos n números de centroides K (Dependiendo del número de clases que deseemos obtener) en la misma posición de algún individuo (Este individuo es tomado al azar).

3-Una vez que tenemos los centroides k y los individuos posicionados, obtenemos las distancias de cada individuo con respecto a los centroides. Ejemplo: Tenemos 50 individuos y dos centroides k. Obtendremos 50 distancias para el centroide uno y 50 para el centroide dos.

4-se clasifican los individuos en una clase, dependiendo de que tan cerca se encuentren de un centroide . Ejemplo

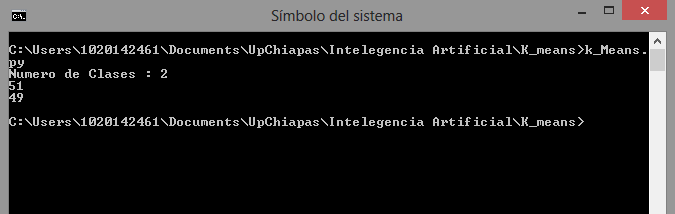
Individuo A : Distancias 50 al Centroide uno , Distancia 80 al Centroide dos. Siendo el individuo A perteneciente a la clase uno.

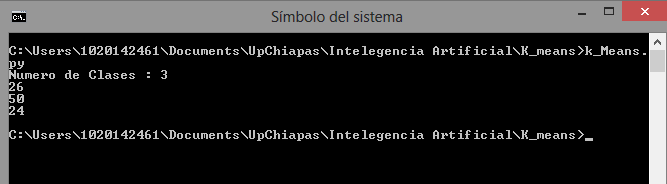
5-Una vez clasificados todos los individuos en una clase, se procede a sacar un promedio de todos los individuos por cada atributo que ellos tengan,. Obteniendo n número de resultados correspondientes a los n números de atributos.

6-Cuando el promedio fue obtenido se recalcula la posición de los centroides con respecto a los promedios obtenidos.

7-se repiten los pasos del 3 al 6 siempre que los centroides cambien de posición. El algoritmo termina cuando los centroides se establecen en una posición.

## Evidencias.





### Conclusión:

Como conclusión podemos destacar la importancia de este algoritmo para la clasificación de datos cuando estos son en grandes cantidades. Un ejemplo práctico es la toma de decisiones más personalizada desde escoger tallas de camisa para un grupo de individuos hasta ventas de productos en ciertas zonas, tipos de consumidores etc...

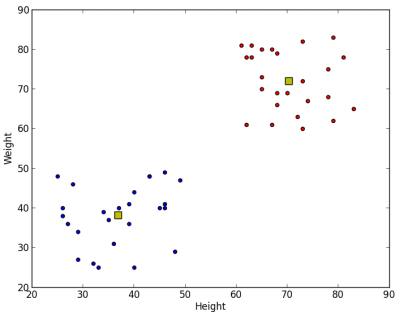


Imagen 2. Ejemplo de camisas.