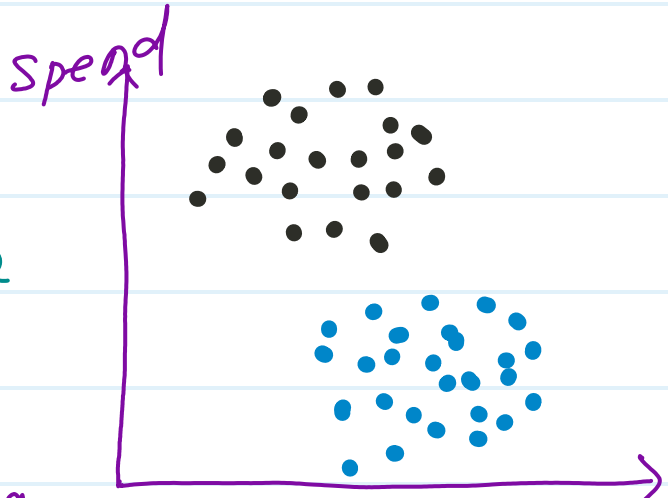
Unsupervised Learning

It does not required labled data.

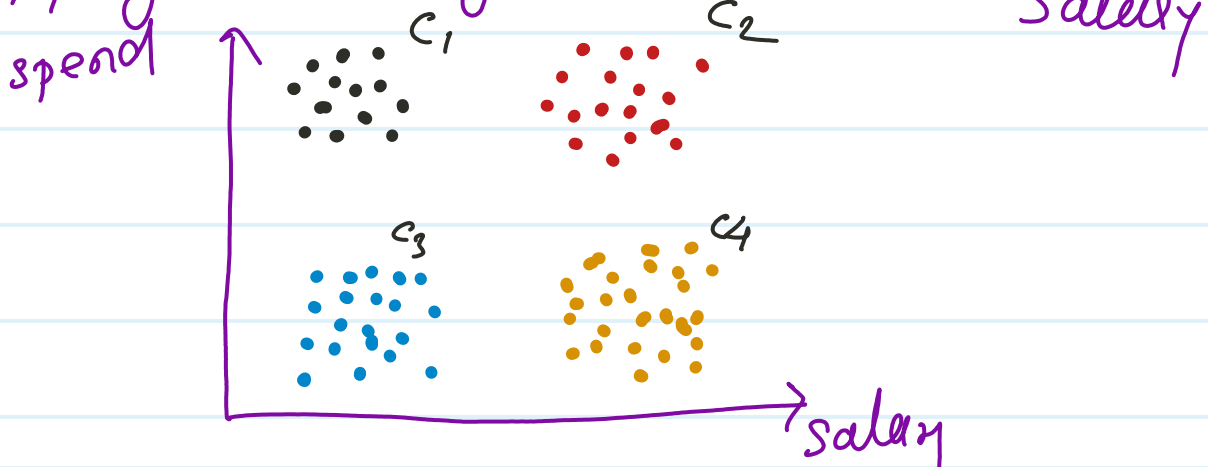
It can be work on structure and unstructure data.

① k-mean clustering
② Hierarchical clustering
③ DBSCAN clustering
④ PCA
⑤ Apriori

x_1	x_2
Income	Spending score
—	—
—	—

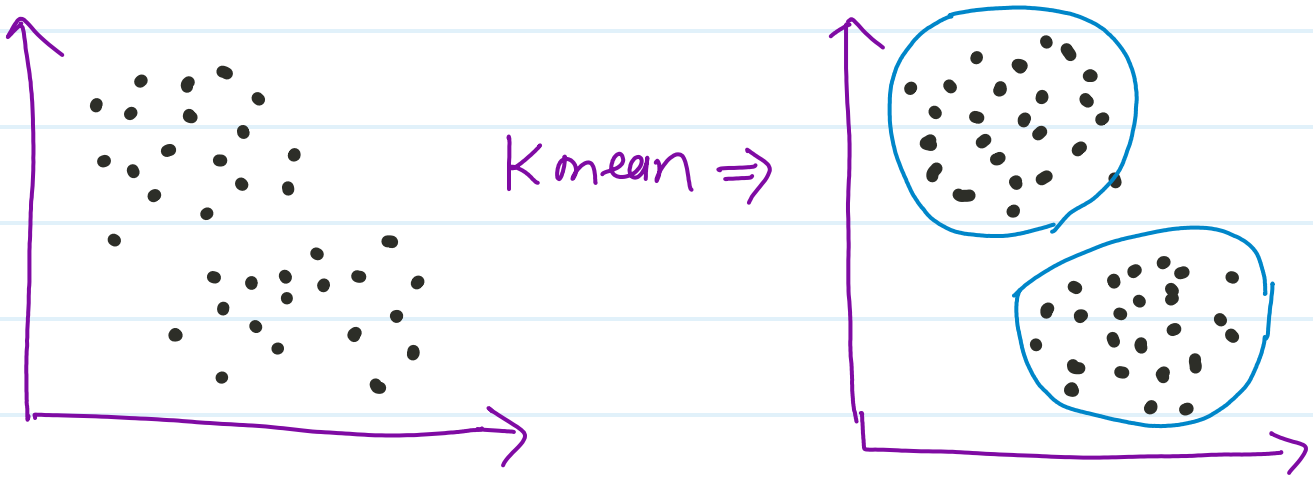


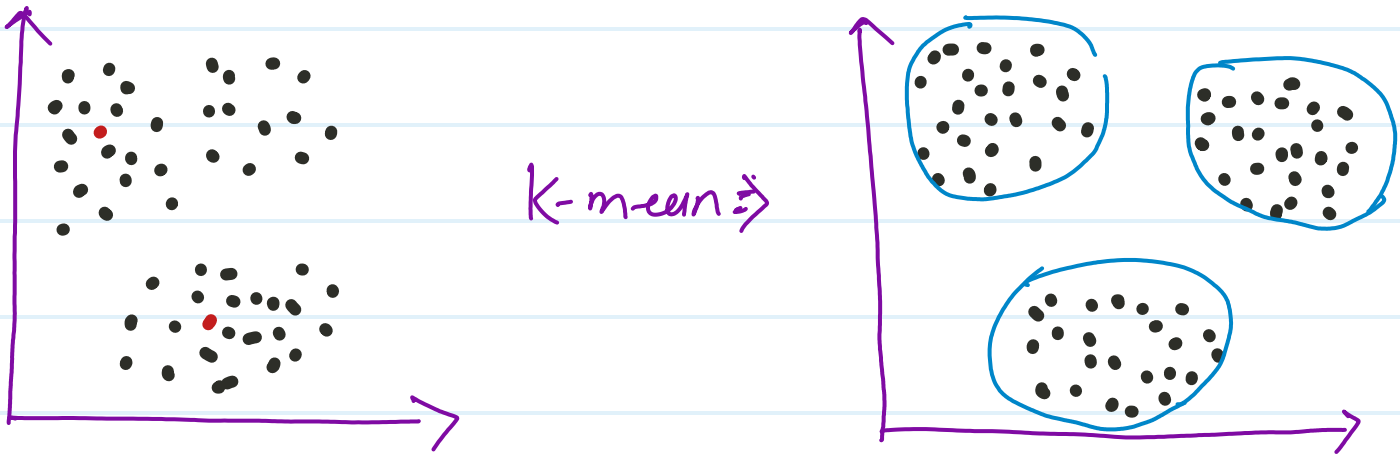
After Applying clustering



K-mean clustering

Geometric Intution :-



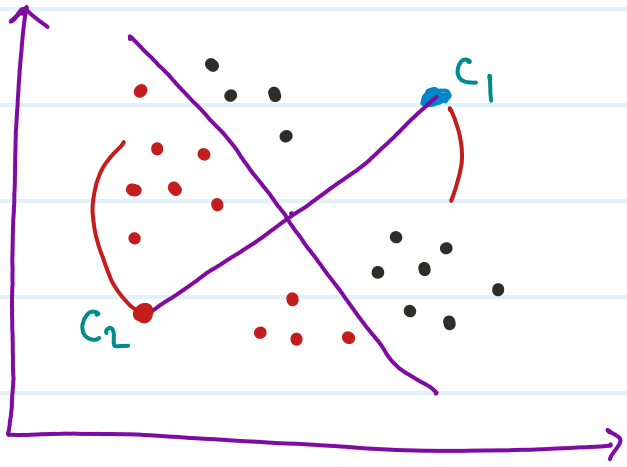


steps -

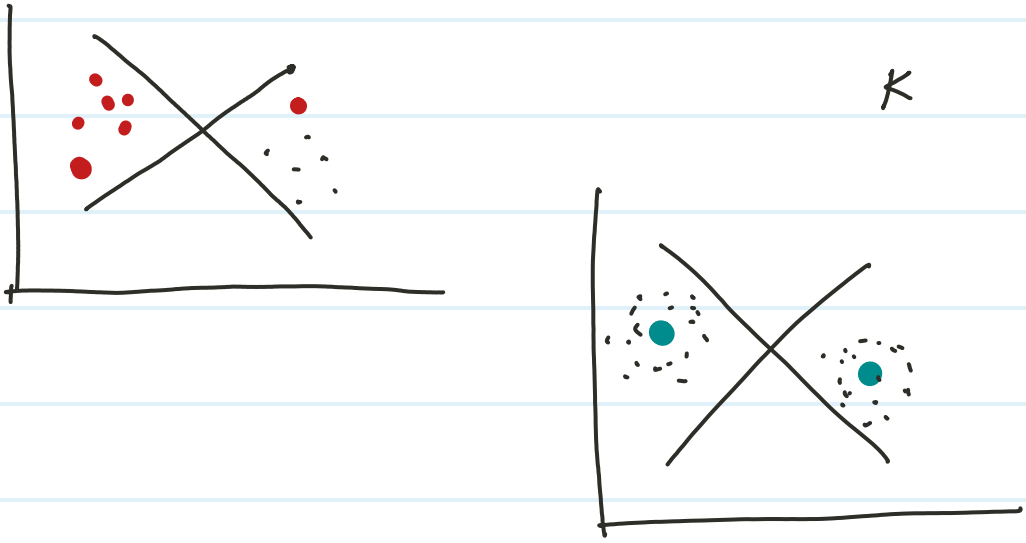
① Initialize some new centroid (K)

② point that are nearest to the centroid group them.

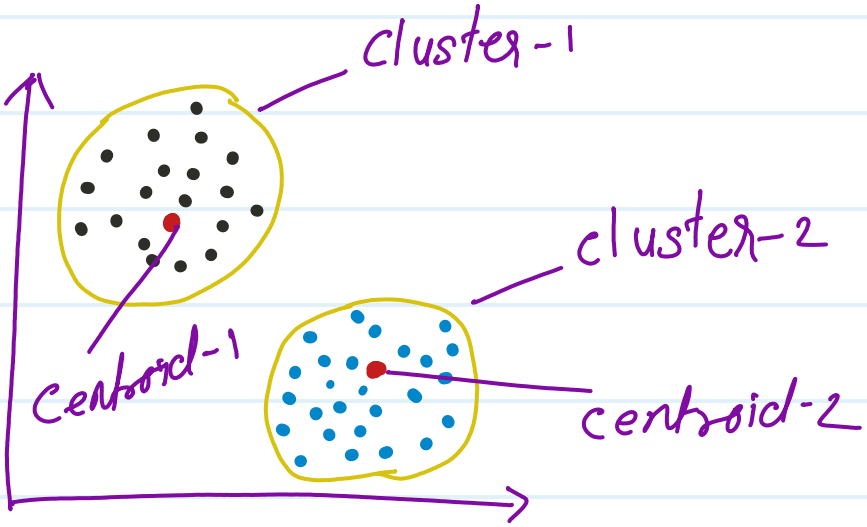
③ move centroid by calculating the mean of points



centroid point will sift toward.....

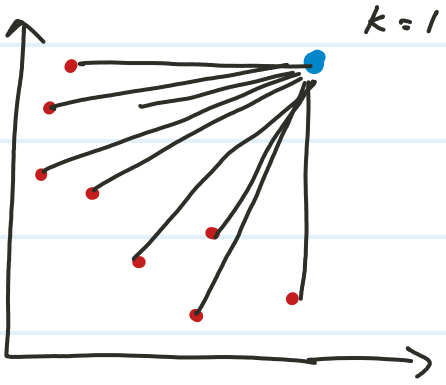


Final output



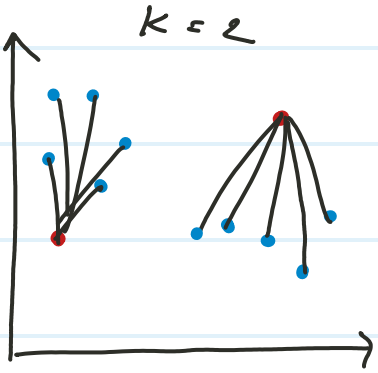
How to select k-value?

Wcss – withing cluster sum of square

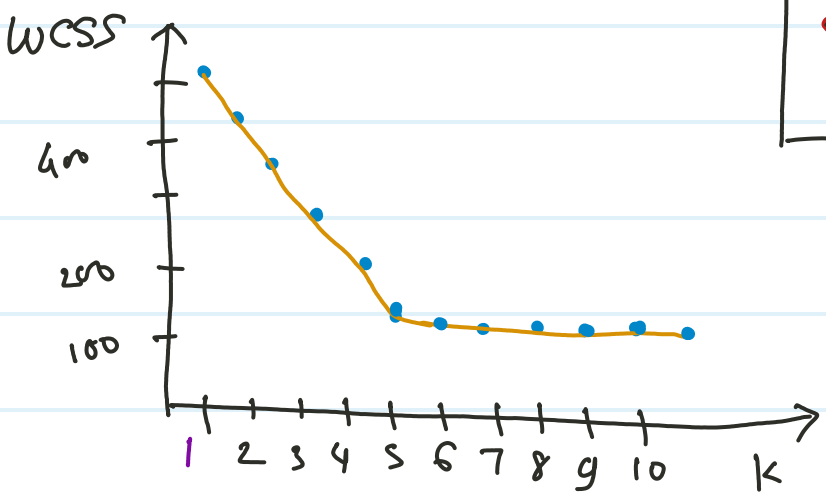


$$wcss = \sum_{i=0}^{k} (\text{Distance b/w point to nearest Centroid})^2$$

suppose k = 1-10



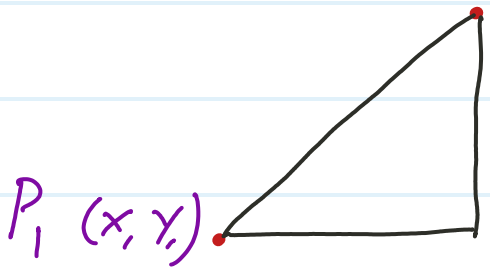
* Elbo method -



when wcss calculate with k=1, wcss value is high, when we increase k=2, it will reduce, when k=3 it will reduce. when k=3 it will reduce and after

Some point it will be straight in graph it is called elbo method.
When curv become straight at that point we get k value. like above graph, 5 will be our k value.

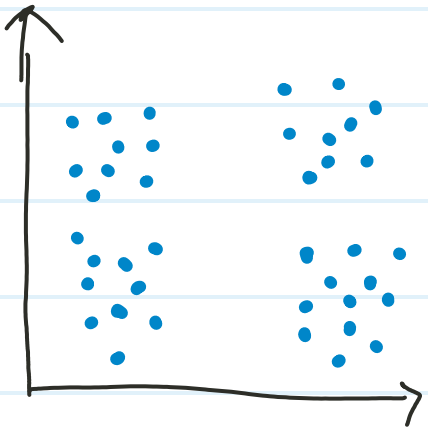
To calculate distance we neeed.
euclidean distance or Manhattan distance



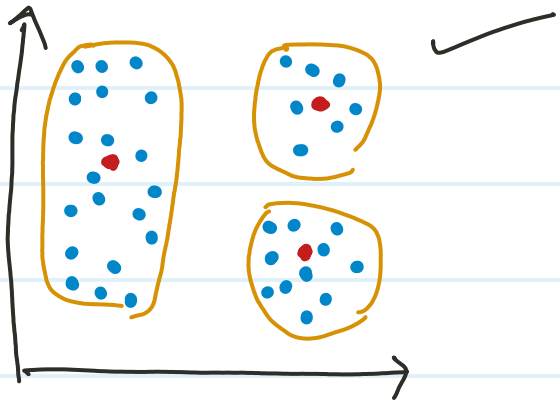
$$ED = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

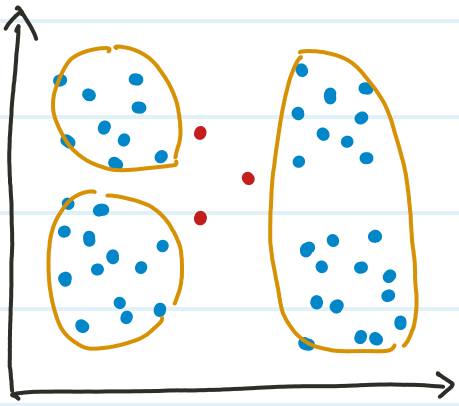
$$MD = |x_2 - x_1| + |y_2 - y_1|$$

Random Initialization trap (K-means++)



K=3





When centroid form very nearest to each other it stuck in Random trap.

To handle this kind of trap we use K-means++ method.

With help of it centroid form far from each other. and prevent from random initialization trap.