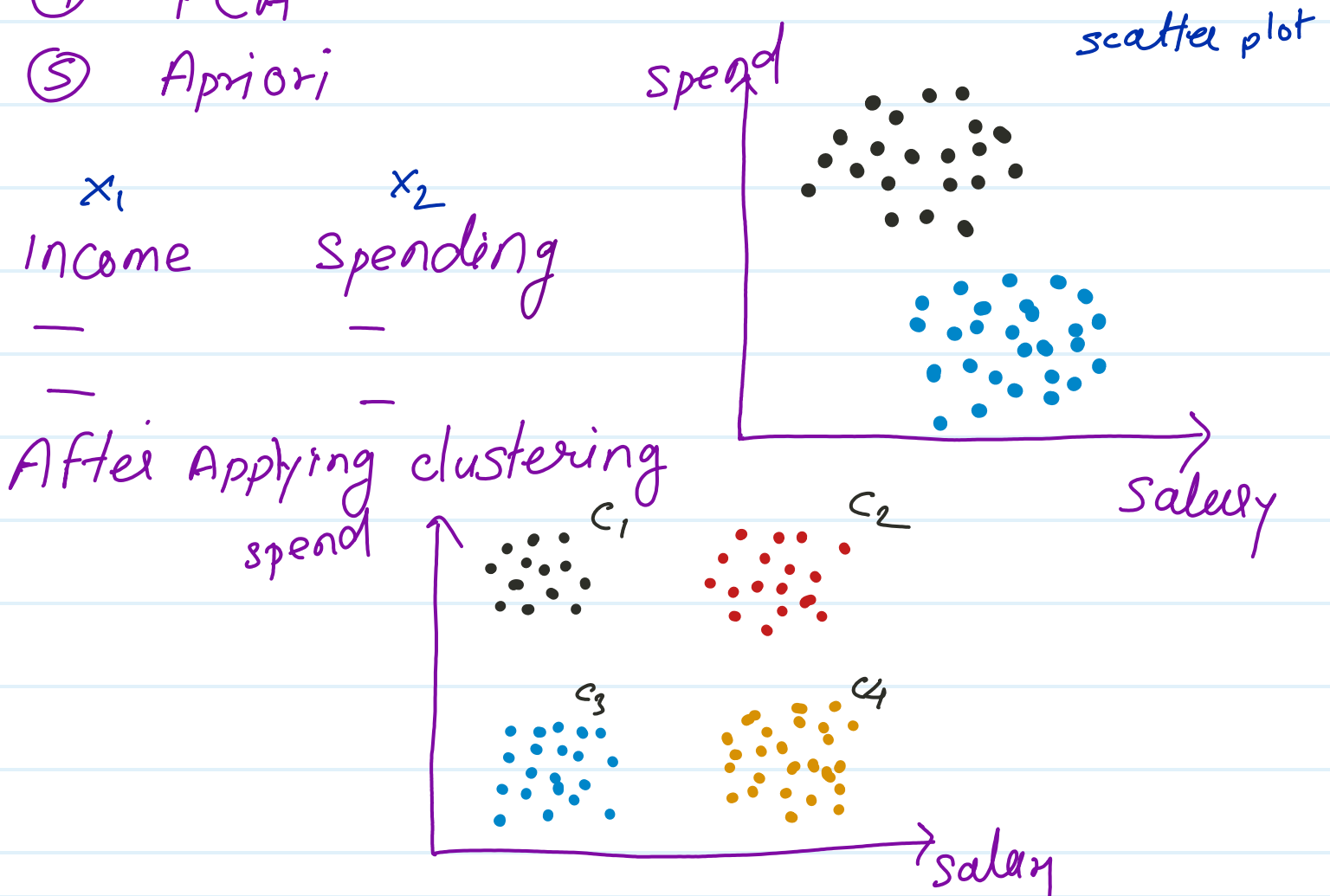


Unsupervised Learning

It does not required lab's data.

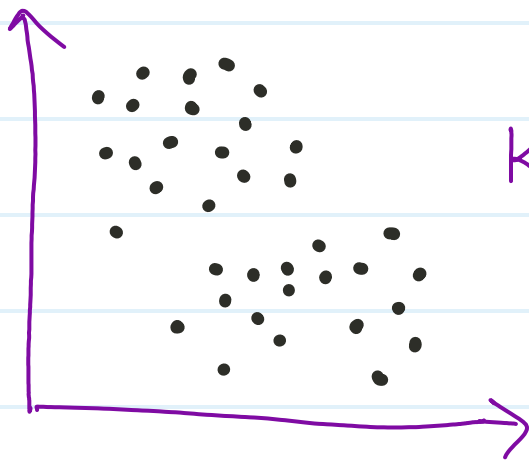
It can be work on structure and unstructure data.

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

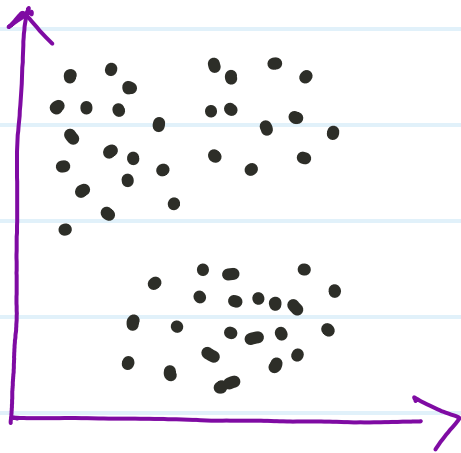
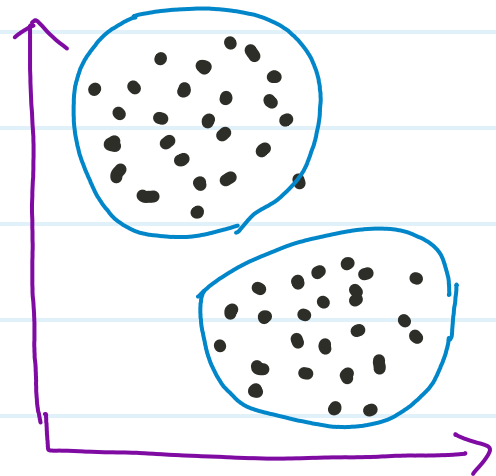


k-mean clustering

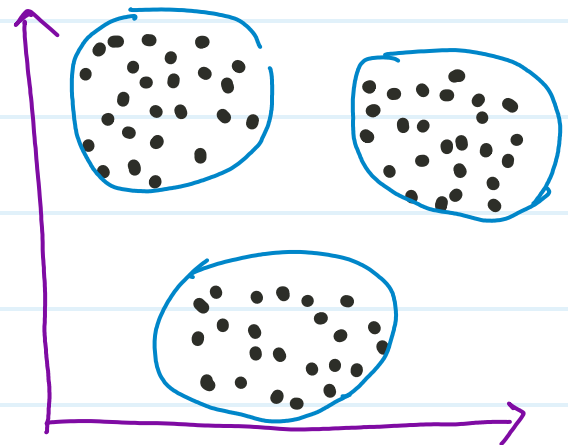
Geometric Intuition:-



K mean \Rightarrow

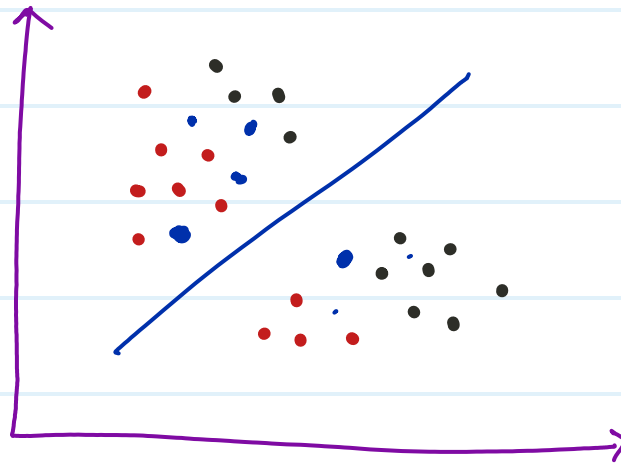


K-mean \Rightarrow



steps -

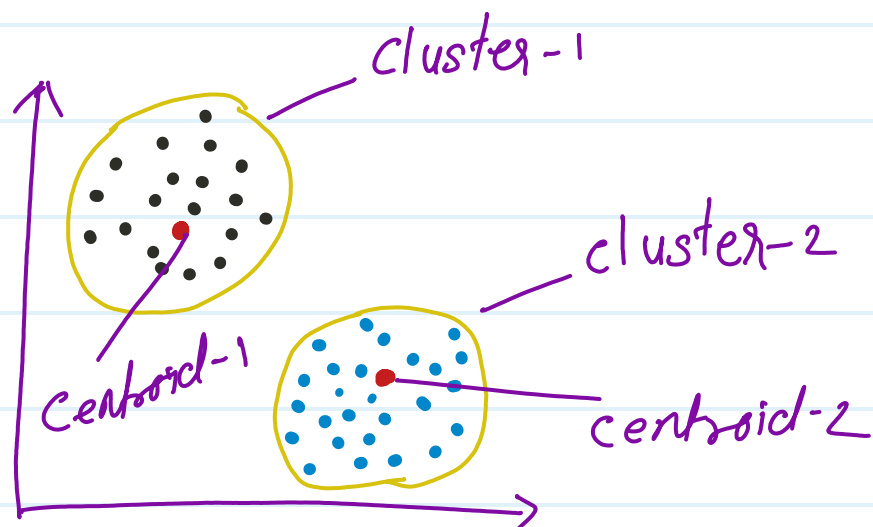
- ① Initialize some new centroid ^(K)
- ② point that are nearest to the centroid group them.
- ③ move centroid by calculating the mean of points



centroid point will shift toward....

x_1	x_2	x_3
82	1	10.2
79	2	12.1
68	1.5	9.0
83	2	11
89	3	15
65	1	8.5

Final
output



x_1 x_2

Random $k = 2$

180 70

170 60

c_1 180 70

c_2 170 60

1 | 178 54

2 | 166 58

3 | 182 71

4 | 175 68

$$P_1 = c_1 \quad 180 - 178, \quad 70 - 54 \Rightarrow 2, 6$$

$$P_1 = c_2 \quad 170 - 178, \quad 60 - 54 \Rightarrow -8, 6$$

$$C_{1, \text{new}} = \frac{180 + 178}{2}, \quad \frac{70 + 54}{2} \Rightarrow 179, 62$$

P_2 166, 58

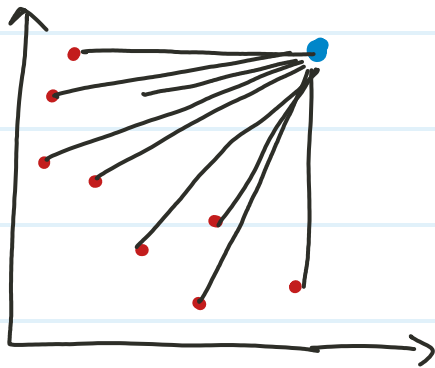
$$C_1 = 179 - 166, \quad 62 - 58 \Rightarrow 13, 4$$

$$C_2 = 170 - 166, \quad 60 - 58 \Rightarrow 4, 2$$

$$C_{2, \text{new}} = \frac{170 + 166}{2}, \quad \frac{60 + 58}{2} \Rightarrow 168, 59$$

How to select k-value?

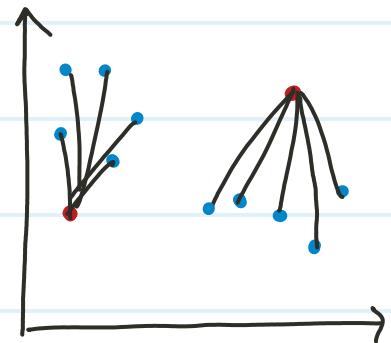
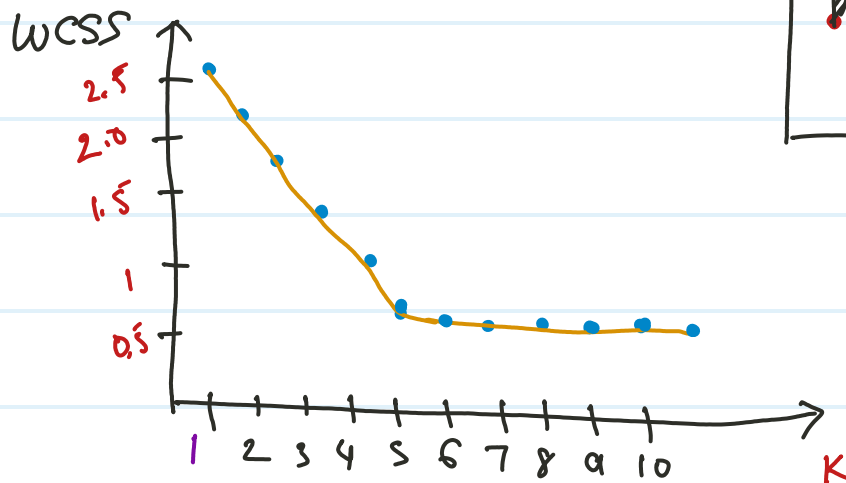
WCSS - within cluster sum of square



$$WCSS = \sum_{i=0}^k \left(\text{Distance b/w point to nearest centroid} \right)^2$$

suppose $k = 1 - 20$

* Elbo method -

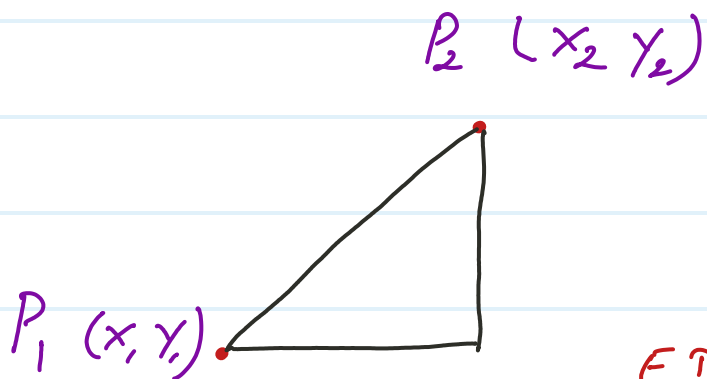


when wcss calculate with $k=1$, wcss value is high, when we increase $k=2$, it will reduce, when $k=3$ it will reduce. when $k=3$ it will reduce and after

Some point it will be straight in graph it is called elbo method.

When curve become straight at that point we got k value. like above graph, 5 will be our k value.

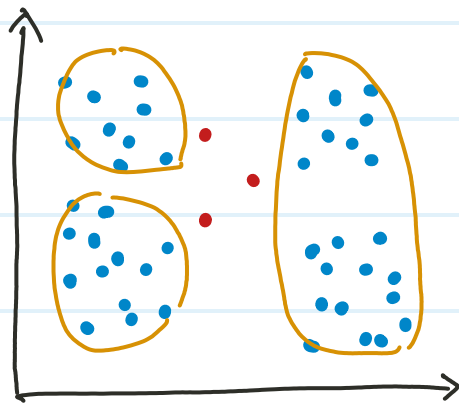
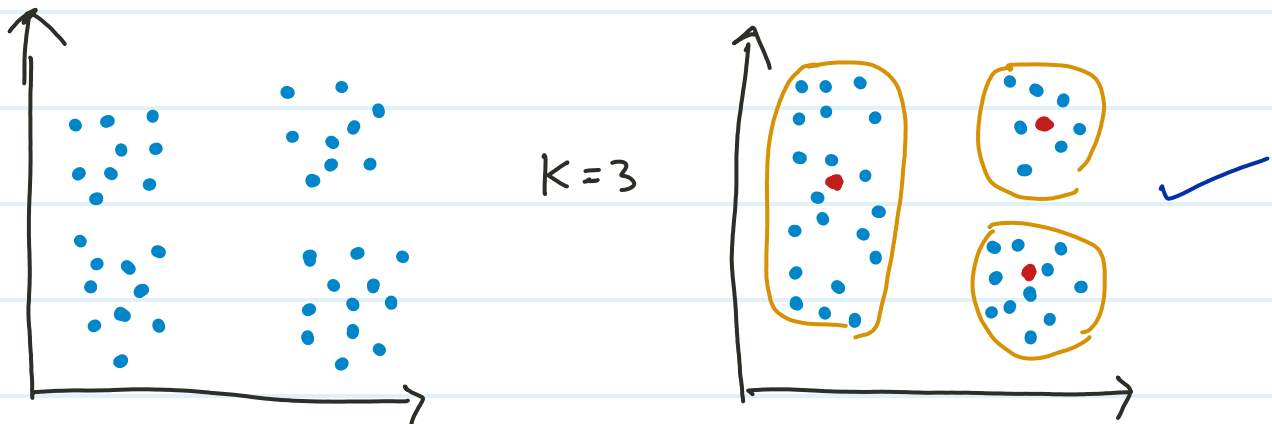
To calculate distance we need. to use
Euclidean distance or Manhattan distanc.



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

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

✓ Random Initialization trap (K-mean++)



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

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

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

