

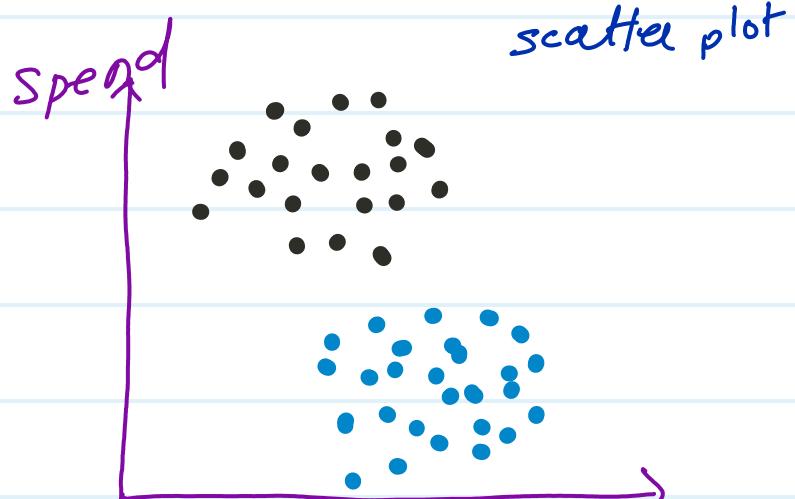
# Unsupervised Learning

It does not require label data.

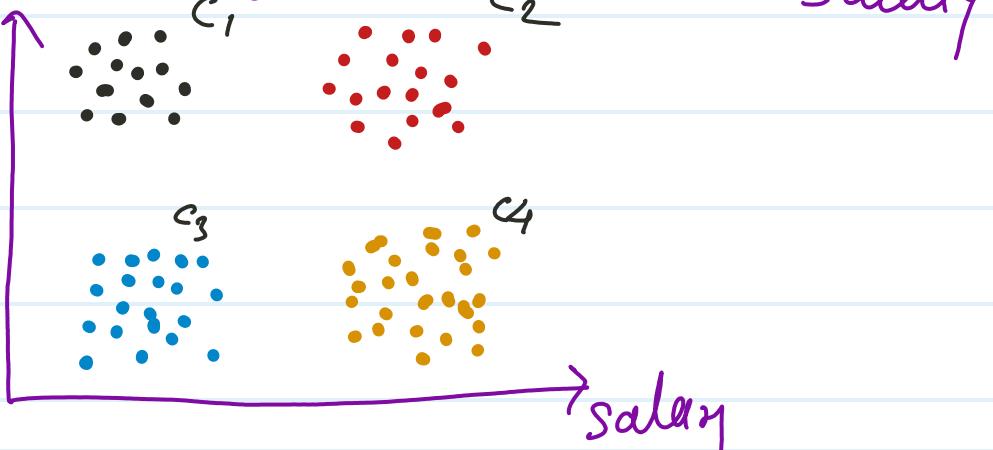
It can be work on structure and unstructure data.

- ① K-mean clustering
- ② Hierarchical clustering
- ③ DB Scan clustering
- ④ PCA
- ⑤ Apriori

$x_1$   
 Income       $x_2$   
 Spending  
 —  
 —

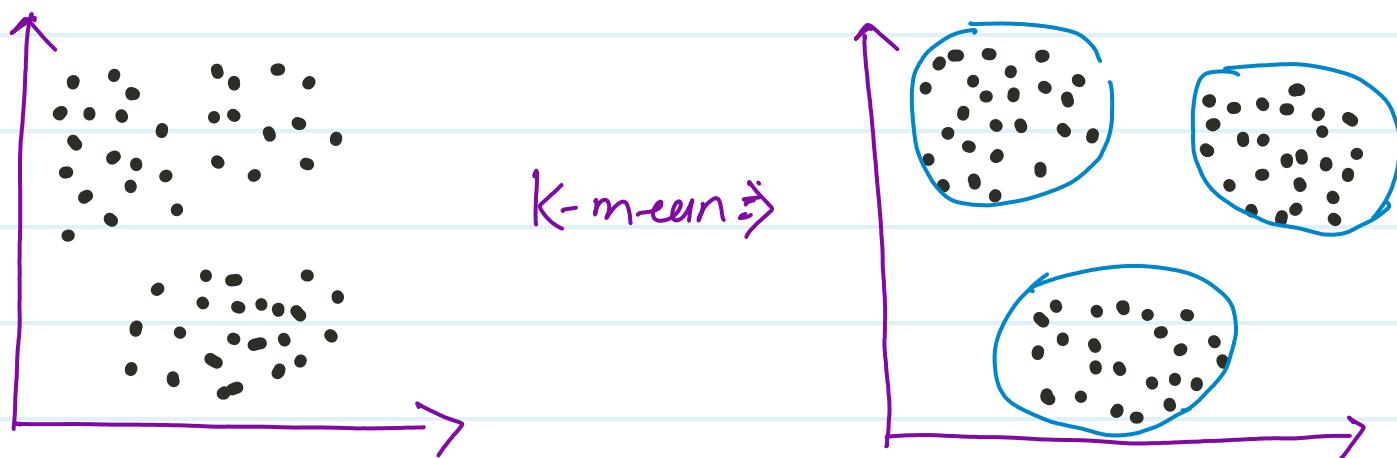
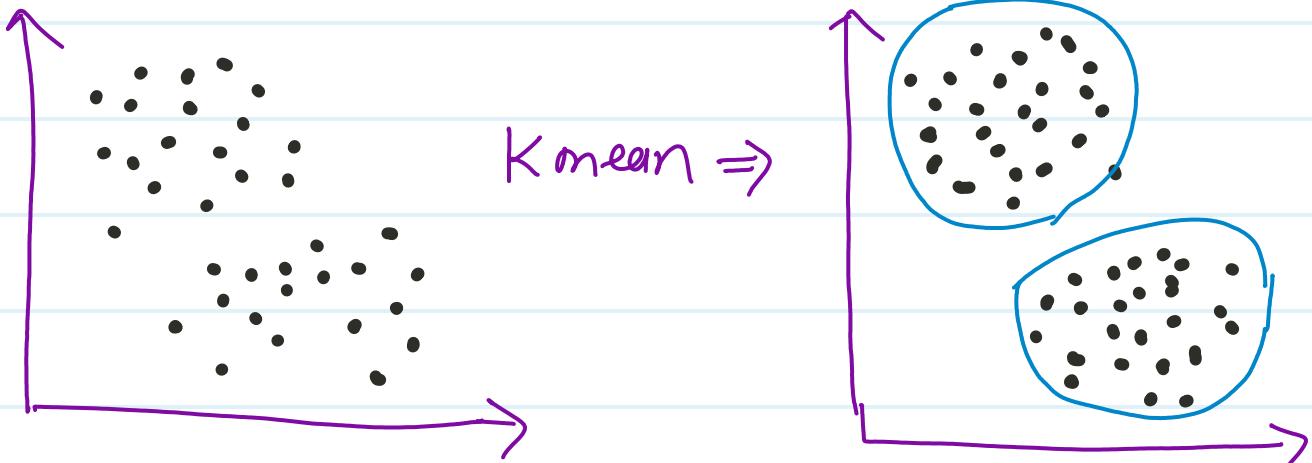


After Applying clustering



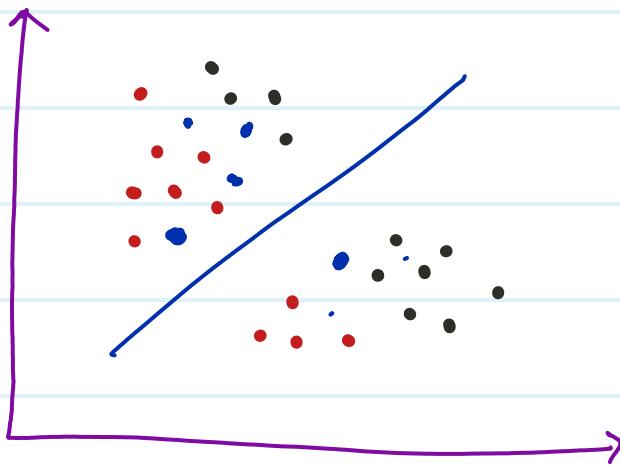
# K - mean clustering

Geometric Intuition:-



steps -

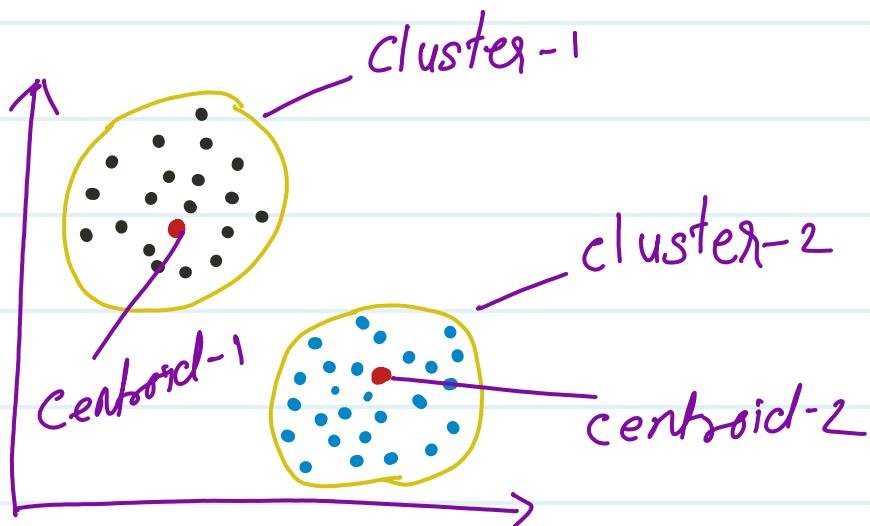
- ① Initialize some new <sup>(K)</sup> centroid
- ② point that are nearest to the centroid group them.
- ③ move centroid by calculating the mean of points



centroid point will sift 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	6.5

Final output



$x_1 \quad x_2$

Random  $k = 2$

180 70

170 60

$c_1 \quad 180 \quad 70$

1	178	59
2	166	58
3	182	71
4	175	68

$c_2 \quad 170 \quad 60$

$$P_1 = c_1 \quad 180 - 178, \quad 70 - 59 \Rightarrow 2, 6$$

$$P_1 = c_2 \quad 170 - 178, \quad 60 - 59 \Rightarrow -8, 6$$

$$c_{1, \text{new}} = \frac{180 + 178}{2}, \quad \frac{70 + 59}{2} \Rightarrow 179, 62$$

$$P_2 \quad 166, 58$$

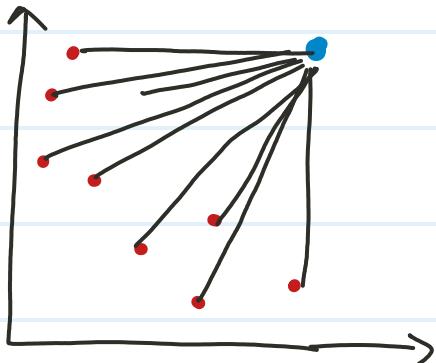
$$c_1 = 179 - 166, \quad 62 - 58 \Rightarrow 13, 4$$

$$c_2 = 170 - 166, \quad 60 - 58 \Rightarrow 4, 2$$

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

# How to select k-value?

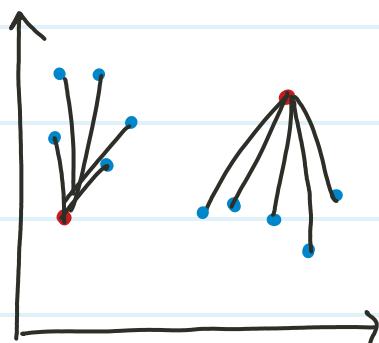
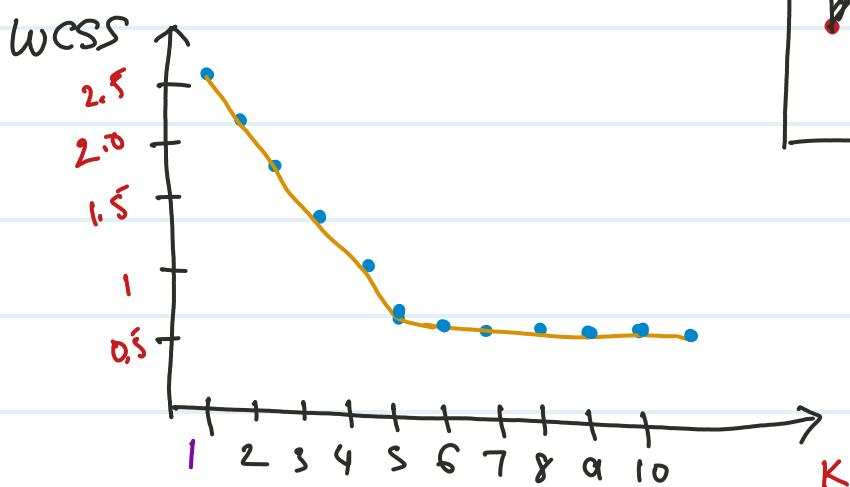
# WCSS - within cluster sum of square



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

suppose  $k = 1 - 20$

# Elbo method -



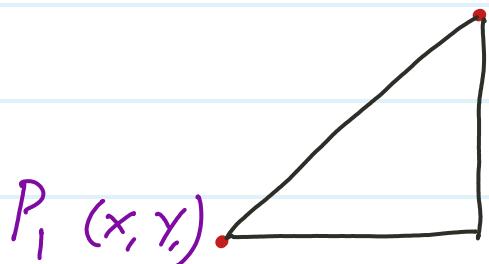
When WCSS calculate with  $k=1$ , WCSS value is high, when we increase  $k=2$ , it will reduce, when  $k=3$  it will reduce. When  $k=3$  it will reduce and after

some point it will be straight in graph  
it is called elbow 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.

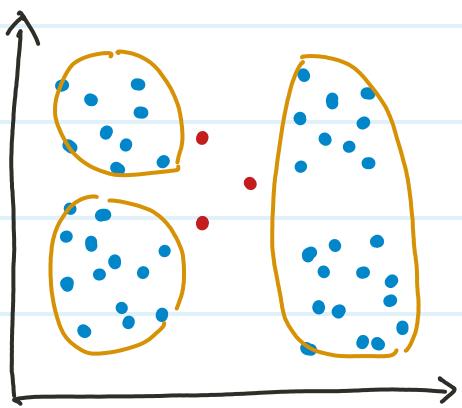
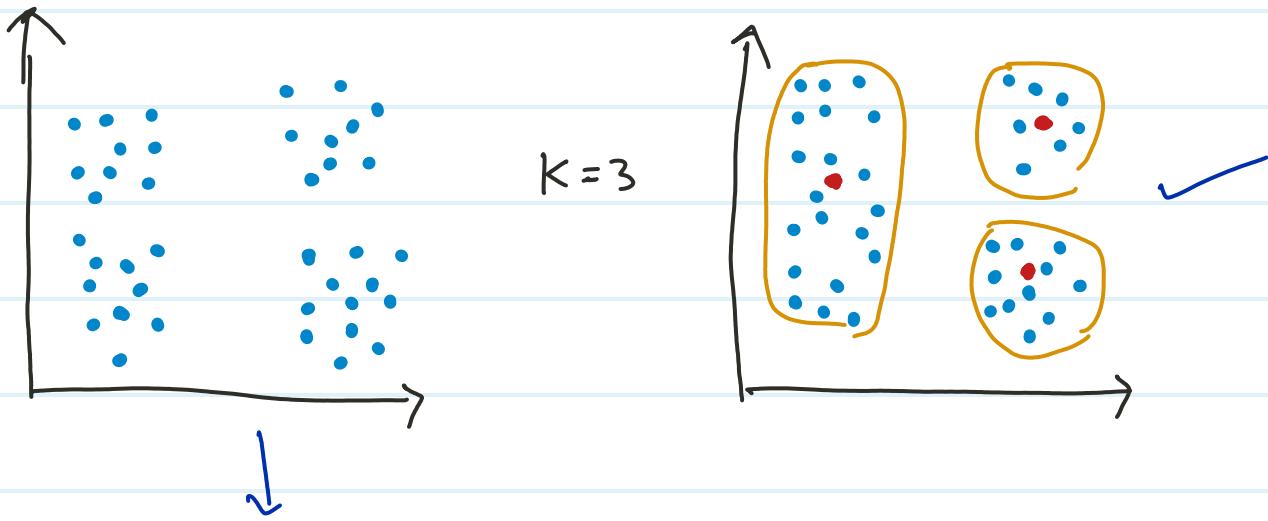
$P_2 (x_2, y_2)$



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

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

## 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.

