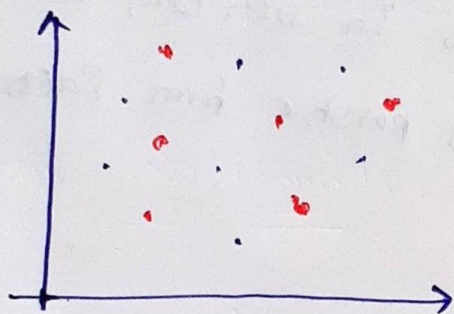
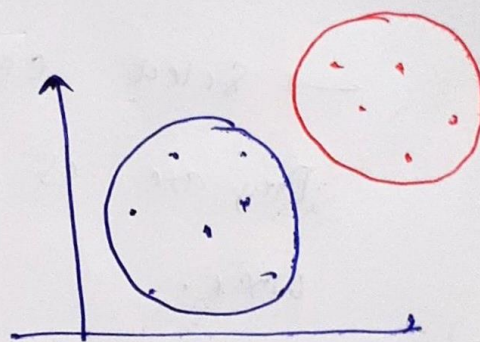


## K-Means Clustering Algorithm

- It is a unsupervised Algorithm
- It will use to cluster data
- Cluster is defined as a collection of datapoints exhibiting certain similarity



Before K-Means



After K-Means

K-Means Partition the data such that

- Each data point belongs to a cluster with nearest mean
- Data points belongs to one cluster have high degree of similarity.
- Data points belongs to different clusters have high degree of dissimilarity.



## Algorithm    (K-Means)

Step-①    Choose number of clusters 'K'

Step-② :    → Randomly choose any K-data points as cluster centers.

→ Select cluster centers in such way that they are as far as possible from each other.

Step-③    → Calculate the distance between each data point and each cluster center, either distance function  
(b) Euclidean distance formula

$$A(a_1, a_2, \dots, a_n), \quad C(c_1, c_2, \dots, c_n)$$

$$\text{Distance} = \sum_{i=1}^n |a_i - c_i|$$

$$ED = \sqrt{\sum_{i=1}^n (a_i - c_i)^2}$$

Step-④ .    Assign each data point to a cluster which is nearest based on distance



Step 5 → Re compute the centers of newly formed clusters

→ Cluster center calculated by taking mean of all the data points in that cluster.

Step 6! Repeat Step 3, 4, 5 till reaching termination criteria.

Termination Criteria

- Centers of newly formed clusters don't change
- Data points remain in same cluster
- Maximum number of iterations reached;



Using K-means Algorithm create 2 clusters

Sample	feature 1	feature 2
A	2	2
B	3	2
C	1	1
D	3	1
E	1.5	0.5

Choose 2 centroids

A(2,2) & C(1,1)  
(BD)

Data point	A(2,2)	C(1,1)	cluster
A	0	1.414	A
B	1	2.23	A
C	1.414	0	C
D	1.414	2	A
E	1.58	0.7	C

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



cluster - (1)

A (2, 2)

B (3, 2)

D (3, 1)

$$\text{centroid} = \left[ \frac{2+3+3}{3}, \frac{2+2+1}{3} \right]$$

$$= [2.66, 1.66]$$

cluster (2)

C (1, 1)

E (1.5, 0.5)

$$\text{centroid} = \left[ \frac{1+1.5}{2}, \frac{1+0.5}{2} \right]$$

$$= [1.25, 0.75]$$

Iteration : 2

Sample	C1 (2.66, 1.66)	C2 (1.25, 0.75)	centroid
A (2, 2)	0.742	1.45	C1
B (3, 2)	0.48	2.15	C1
C (1, 1)	1.78	0.35	C2
D (3, 1)	0.742	1.76	C1
E (1.5, 0.5)	1.64	0.35	C2

cluster - (1)

A (2, 2)

B (3, 2)

D (3, 1)

$$\text{centroid 1} = [2.66, 1.66]$$

cluster - (2)

C (1, 1)

E (1.5, 0.5)

$$\text{centroid 2} = [1.25, 0.75]$$