Ex# 1 Suppose you have the following points:

(2, 3), (3, 2), (2, 2), (7, 8), (8, 7), (9, 8)

Let's say we want to group these points into two clusters (K=2) using K-means clustering.

1. **Initialization**: Randomly place two centroids on the graph. Let's start with (3, 3) and (8, 8). These are our initial centroids.
2. **Assignment**: Measure the distance of each point from these centroids and assign each point to the nearest centroid.

Points closer to (3, 3) might be: (2, 3), (3, 2), (2, 2) Points closer to (8, 8) might be: (7, 8), (8, 7), (9, 8)

These assignments create two initial clusters.

3. **Update**: Recalculate the centroids by taking the mean of all the points in each cluster.

New centroid for the first cluster (center of points (2, 3), (3, 2), (2, 2)) might be (2.33, 2.33). New centroid for the second cluster (center of points (7, 8), (8, 7), (9, 8)) might be (8, 7.67).

4. **Repeating Steps 2 and 3**: Repeat the process by reassigning points to the closest centroid and updating the centroids. Continue this iteration until the centroids stabilize (they stop moving significantly) or after a fixed number of iterations.

The Euclidean distance between two points (x1, y1) and (x2, y2) in a 2D space is given by:

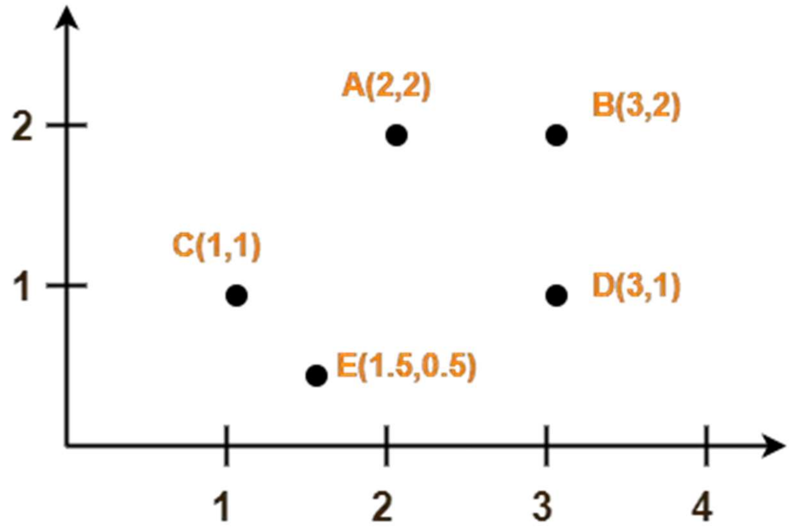
$$\text{Distance} = \sqrt{(x2 - x1)^2 + (y2 - y1)^2}$$

For example, if you want to find the distance between the point (3, 3) and the centroid (5, 7), you'd plug these values into the formula:

$$\text{Distance} = \sqrt{(5-3)^2 + (7-3)^2}$$
$$\text{Distance} = \sqrt{(2)^2 + (4)^2}$$
$$\text{Distance} = \sqrt{4 + 16}$$
$$\text{Distance} = \sqrt{20}$$
$$\text{Distance} = 4.47$$

Problem-02:

Use K-Means Algorithm to create two clusters-



Solution-

We follow the above discussed K-Means Clustering Algorithm.

Assume A(2, 2) and C(1, 1) are centers of the two clusters.

Iteration-01:

- We calculate the distance of each point from each of the center of the two clusters.
- The distance is calculated by using the euclidean distance formula.

The following illustration shows the calculation of distance between point A(2, 2) and each of the center of the two clusters-

Calculating Distance Between A(2, 2) and C1(2, 2)-

P(A, C1)

= sqrt [$(x2 – x1)^2$ + $(y2 – y1)^2$]

= sqrt [$(2 – 2)^2$ + $(2 – 2)^2$]

= sqrt [0 + 0]

= 0

Calculating Distance Between A(2, 2) and C2(1, 1)-

P(A, C2)

= sqrt [$(x2 – x1)^2 + (y2 – y1)^2$]

= sqrt [$(1 – 2)^2 + (1 – 2)^2$]

= sqrt [1 + 1]

= sqrt [2]

= 1.41

In the similar manner, we calculate the distance of other points from each of the center of the two clusters.

Next,

- We draw a table showing all the results.
- Using the table, we decide which point belongs to which cluster.
- The given point belongs to that cluster whose center is nearest to it.

Given Points	Distance from center (2, 2) of Cluster-01	Distance from center (1, 1) of Cluster-02	Point belongs to Cluster
A(2, 2)	0	1.41	C1
B(3, 2)	1	2.24	C1
C(1, 1)	1.41	0	C2
D(3, 1)	1.41	2	C1
E(1.5, 0.5)	1.58	0.71	C2

From here, New clusters are-

Cluster-01:

First cluster contains points-

- A(2, 2)
- B(3, 2)
- E(1.5, 0.5)
- D(3, 1)

Cluster-02:

Second cluster contains points-

- C(1, 1)
- E(1.5, 0.5)

Now,

- We re-compute the new cluster clusters.
- The new cluster center is computed by taking mean of all the points contained in that cluster.

For Cluster-01:

Center of Cluster-01

= ((2 + 3 + 3)/3, (2 + 2 + 1)/3)

= (2.67, 1.67)

For Cluster-02:

Center of Cluster-02

= ((1 + 1.5)/2, (1 + 0.5)/2)

= (1.25, 0.75)

This is completion of Iteration-01.

Next, we go to iteration-02, iteration-03 and so on until the centers do not change anymore.