

## **K-means Clustering**

### **Example**

Cluster the following eight points (with (x, y) representing locations) into three clusters:

A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9)

### **Solution**

Initial cluster centers are: A1(2, 10), A4(5, 8) and A7(1, 2).

### **Iteration-01:**

We calculate the distance of each point from each of the center of the three clusters.

The distance is calculated by using the Euclidean distance formula.

| Given Points | Distance from<br>center (2, 10) of<br>Cluster-01 | Distance from<br>center (5, 8) of<br>Cluster-02 | Distance from<br>center (1, 2) of<br>Cluster-03 | Point belongs<br>to Cluster |
|--------------|--|---|---|-----------------------------|
| A1(2, 10)    | 0  | 5   | 9   | C1                          |
| A2(2, 5)     | 5  | 6   | 4   | C3                          |
| A3(8, 4)     | 12   | 7   | 9   | C2                          |
| A4(5, 8)     | 5  | 0   | 10  | C2                          |
| A5(7, 5)     | 10   | 5   | 9   | C2                          |
| A6(6, 4)     | 10   | 5   | 7   | C2                          |
| A7(1, 2)     | 9  | 10  | 0   | C3                          |
| A8(4, 9)     | 3  | 2   | 10  | C2                          |

From here, New clusters are-

Cluster-01:

First cluster contains points-

A1(2, 10)

Cluster-02:

Second cluster contains points-

A3(8, 4)

A4(5, 8)

A5(7, 5)

A6(6, 4)

A8(4, 9)

Cluster-03:

Third cluster contains points-

A2(2, 5)

A7(1, 2)

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:

We have only one point A1(2, 10) in Cluster-01.

So, cluster center remains the same.

For Cluster-02:

Center of Cluster-02

$$= ((8 + 5 + 7 + 6 + 4)/5, (4 + 8 + 5 + 4 + 9)/5)$$

$$= (6, 6)$$

For Cluster-03:

Center of Cluster-03

$$= ((2 + 1)/2, (5 + 2)/2)$$

$$= (1.5, 3.5)$$

This is completion of Iteration-01.

### **Iteration-02:**

We calculate the distance of each point from each of the center of the three clusters.

The distance is calculated by using the given distance function.

| Given Points | Distance from<br>center (2, 10) of<br>Cluster-01 | Distance from<br>center (6, 6) of<br>Cluster-02 | Distance from<br>center (1.5, 3.5) of<br>Cluster-03 | Point belongs to<br>Cluster |
|--------------|--|---|---|-----------------------------|
| A1(2, 10)    | 0  | 8   | 7   | C1                          |
| A2(2, 5)     | 5  | 5   | 2   | C3                          |
| A3(8, 4)     | 12   | 4   | 7   | C2                          |
| A4(5, 8)     | 5  | 3   | 8   | C2                          |
| A5(7, 5)     | 10   | 2   | 7   | C2                          |
| A6(6, 4)     | 10   | 2   | 5   | C2                          |
| A7(1, 2)     | 9  | 9   | 2   | C3                          |
| A8(4, 9)     | 3  | 5   | 8   | C1                          |

From here, New clusters are-

Cluster-01:

First cluster contains points-

A1(2, 10)

A8(4, 9)

Cluster-02:

Second cluster contains points-

A3(8, 4)

A4(5, 8)

A5(7, 5)

A6(6, 4)

Cluster-03:

Third cluster contains points-

A2(2, 5)

A7(1, 2)

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

$$= ((2 + 4)/2, (10 + 9)/2)$$

$$= (3, 9.5)$$

For Cluster-02:

Center of Cluster-02

$$= ((8 + 5 + 7 + 6)/4, (4 + 8 + 5 + 4)/4)$$

$$= (6.5, 5.25)$$

For Cluster-03:

Center of Cluster-03

$$= ((2 + 1)/2, (5 + 2)/2)$$

$$= (1.5, 3.5)$$

This is completion of Iteration-02.

After second iteration, the center of the three clusters are-

$$C1(3, 9.5)$$

$$C2(6.5, 5.25)$$

$$C3(1.5, 3.5)$$