

(*)

K-Means clustering.

PARTITIONAL CLUSTERING:-

Problem:-

Cluster the following 5 datapoints, representing as (x, y)

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

Apply K-Means clustering algorithm to create 2 clusters.

Solution:-

STEP-1:-

Let us assume datapoint $A(2, 2)$ is the centre for cluster-1 $\rightarrow C_1(2, 2)$

Let us assume datapoint $C(1, 1)$ is the centre for cluster-2 $\rightarrow C_2(1, 1)$

Iteration 1:-

calculate the distance of datapoint $A(2, 2)$ from the 2 cluster's centres which is C_1 and C_2 .

Where C_1 is $(2,2)$ and C_2 is $(1,1)$
Apply Euclidean Distance formula,

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

$(x_1, y_1) \quad (x_2, y_2)$

$$d(A, C_1) = \sqrt{(2-2)^2 + (2-2)^2} = \sqrt{0} = 0$$

$(2,2) \quad (2,2)$

$$d(A, C_2) = \sqrt{(1-2)^2 + (1-2)^2} = \sqrt{1^2 + 1^2}$$

$(2,2) \quad (1,1)$

$$= \sqrt{2} = \underline{\underline{1.414}}$$

Iteration 2:

$$d(B, C_1) = \sqrt{(2-3)^2 + (2-2)^2} = \sqrt{1}$$

$(3,2) \quad (2,2)$

$$= \underline{\underline{1}}$$

$$d(B, C_2) = \sqrt{(1-3)^2 + (1-2)^2} = \sqrt{2^2 + 1^2}$$

$(3,2) \quad (1,1)$

$$= \sqrt{4+1} = \sqrt{5} = \underline{\underline{2.236}}$$

Iteration 3:

$$d(C, C_1) = \sqrt{(2-1)^2 + (2-1)^2} = \sqrt{1+1}$$

$(1,1) \quad (2,2)$

$$= \sqrt{2} = \underline{\underline{1.414}}$$

$$d(C, C_2) = \sqrt{(1-1)^2 + (1-1)^2} = 0$$

$(1,1) \quad (1,1)$

Iteration 4:

$$d(D, C_1) = \sqrt{(2-3)^2 + (2-1)^2}$$

$(3,1) \quad (2,2)$

$$= \sqrt{1+1} = \sqrt{2}$$
$$= \underline{\underline{1.414}}$$

$$d(D, C_2) = \sqrt{(1-3)^2 + (1-1)^2}$$

$$(3,1) (1,1)$$

$$= \sqrt{2^2} = \sqrt{4} = \underline{2}$$

Iteration 5 :-

$$d(E, C_1) = \sqrt{(2-1.5)^2 + (2-0.5)^2}$$

$$(1.5, 0.5) (2, 2)$$

$$= \sqrt{(0.5)^2 + (1.5)^2}$$

$$\sqrt{0.25 + 2.25} = \sqrt{2.5} = \underline{1.58}$$

$$d(E, C_2) = \sqrt{(1-1.5)^2 + (1-0.5)^2}$$

$$(1.5, 0.5) (1, 1)$$

$$\sqrt{(0.5)^2 + (0.5)^2}$$

$$= \sqrt{0.25 + 0.25} = \sqrt{0.5} = \underline{0.70}$$

So, here we are calculating the distance from datapoint $E(1.5, 0.5)$ from the 2 clusters' centers C_1 and C_2 . \Rightarrow Take the lowest value

STEP - 2 :-

Datapoints	Distance from cluster center C_1	Distance from cluster center C_2	cluster of Datapoint
A(2,2)	0	1.41	C_1
B(3,2)	1	2.36	C_1
C(1,1)	1.41	0	C_2
D(3,1)	1.41	2	C_1
E(1.5, 0.5)	1.58	0.70	C_2

STEP-3:-

So, here the cluster 1 consist of

$A(2,2)$ $B(3,2)$ $D(3,1)$

and the cluster 2 consist of $C(1,1)$ and

$E(1.5, 0.5)$

k-Means Algorithm:-

STEP-1:-

select the points (datapoints) as initial centroids.

STEP-2:- Then, calculate the distance from each datapoint to the cluster's centers.

STEP-3:- Find the Minimal distance and group the datapoint into the corresponding cluster.