

K Means Clustering for multi-Dimensional Data

Example : Given Data and k =2

Sample No.	X	Y
1	185	72
2	170	56
3	168	60
4	179	68
5	182	72
6	188	77

Step 1: Initial Centroid

Cluster	X	Y
k1	185	72
k2	170	56

Step 2: Consider Euclidean distance for calculating distances.

$$\text{Distance}[(x,y),(a,b)] = \sqrt{(x-a)^2 + (y-b)^2}$$

Iteration 1:

Step 3: Find the Euclidean distance for the Sample no 1 (185,72)

$$\text{Distance from Cluster 1}(185,72) = \sqrt{(185-185)^2 + (72-72)^2} = \underline{0}$$

$$\begin{aligned}\text{Cluster-2}(170,56) &= \sqrt{(170-185)^2 + (56-72)^2} = \sqrt{(-15)^2 + (-16)^2} = \sqrt{255 + 256} \\ &= \sqrt{481} = \underline{\mathbf{21.93}}\end{aligned}$$

Find the Euclidean distance for the Sample no 2 (170,56)

$$\begin{aligned}\text{Cluster-1}(185,72) &= \sqrt{(185-170)^2 + (72-56)^2} = \sqrt{(15)^2 + (16)^2} = \sqrt{255 + 256} \\ &= \sqrt{481} = \underline{\mathbf{21.93}}\end{aligned}$$

$$\text{Cluster-2}(170,56) = \sqrt{(170-170)^2 + (56-56)^2} = \underline{0}$$

Step 4 : Cluster Assignment

Sample	Euclidean Distance		Cluster Assignment
	Cluster 1	Cluster 2	
(185,72)	0	21.93	1
(170,56)	21.93	0	2

Step 5: Centroid Update

Cluster	Centroid	
	X	Y
k1	185	72
k2	170	56

Iteration 2: repeat (step 3, 4 & 5)

Step 3 : Calculate Euclidean distance for the next dataset --Sample 3(168,60)

Distance from Cluster 1(185,72) = $\sqrt{(168-185)^2+(60-72)^2}$ = 20.808

Distance from Cluster 2(170,56) = $\sqrt{(168-170)^2+(60-56)^2}$ = 4.472

Step 4: Cluster Assignment

Sample	Euclidean Distance		Cluster Assignment
	Cluster 1	Cluster 2	
(168,60)	20.808	4.472	2

Step 5: Centroid Update

Cluster	Centroid	
	X	Y
k1	185	72
k2	$=(170+168)/2$ =169	$=(56+60)/2$ =58

Iteration 3: repeat (step 3, 4 & 5)

Step 3: Calculate Euclidean distance for the next dataset- sample 4(179,68)

Distance from Cluster 1(185,72) = $\sqrt{(179-185)^2+(68-72)^2}$ = 7.211103

Distance from Cluster 2(169,58) = $\sqrt{(179-169)^2+(68-58)^2}$ = 14.14214

Step 4: Cluster Assignment

Sample	Euclidean Distance		Cluster Assignment
	Cluster 1	Cluster 2	
(179,68)	7.211103	14.14214	1

Step 5: Centroid Update

Cluster	Centroid	
	X	Y
k1	$= (185+179)/2$ =182	$= (72+68)/2$ =70
k2	169	58

Iteration 4: repeat (step 3, 4 & 5)

Step 3: Calculate Euclidean distance for the next dataset- sample 5(182, 72)

Distance from Cluster 1(182,70) = $\text{sqrt}[(182-182)^2+(72-70)^2] = 2$

Distance from Cluster 2(169,58) = $\text{sqrt}[(182-169)^2+(72-58)^2] = 19.10$

Step 4: Cluster Assignment

Sample	Euclidean Distance		Cluster Assignment
	Cluster 1	Cluster 2	
(182,72)	2	19.10	1

Step 5: Centroid Update

Cluster	Centroid	
	X	Y
k1	$= (182+182)/2$ =182	$= (70+72)/2$ =71
k2	169	58

Iteration 5: repeat (step 3, 4 & 5)

Step 3: Calculate Euclidean distance for the next dataset- sample 5(188, 77)

Distance from Cluster 1(182,71) = $\text{sqrt}[(188-182)^2+(77-71)^2] = 8.4852$

Distance from Cluster 2(169,58) = $\text{sqrt}[(188-169)^2+(77-58)^2] = 26.87$

Step 4: Cluster Assignment

Sample	Euclidean Distance		Cluster Assignment
	Cluster 1	Cluster 2	
(188,77)	8.4852	26.87	1

Step 5: Centroid Update

Cluster	Centroid	
	X	Y

k1	$\frac{182+188}{2}$ =185	$\frac{71+77}{2}$ =74
k2	169	58

Step 6: FINAL ASSIGNMENT TABLE

Sample No.	X	Y	Assignment
1	185	72	1
2	170	56	2
3	168	60	2
4	179	68	1
5	182	72	1
6	188	77	1