K-Mean Clustering

Sr.	X	Y
1	1	1
2	1.5	2
3	3	4
4	5	7
5	3.5	5
6	4.5	5
7	3.5	4.5

Here, K=2
and we assume
cluster 1 center is (1, 1)
and cluster 2 center
is (5, 7)

Initial centroid
$K_1 = (1, 1)$
$K_2 = (5, 7)$

Data points		Distance to Center:			Cluster
		1	1	5, 7	
1	1	0	~~7.2~~	7.2	1
1.5	2	1.11		6.10	1
3	4	3.6		3.6	1
5	7	7.2		0	2
3.5	5	4.7		2.5	2
4.5	5	5.3		2.06	2
3.5	4.5	4.3		2.9	2

$$ED = \sqrt{(X_o - X_c)^2 - (Y_o - Y_c)^2}$$

for cluster 1:
$ED(1,1) = \sqrt{(1-1)^2 - (1-1)^2} = 0$
$ED(1.5,2) = \sqrt{(1.5-1)^2 - (2-1)^2} = 1.11$

for cluster 2:
$ED(1,1) = \sqrt{(1-5)^2 - (1-7)^2} = 7.2$
$ED(1.5,2) = \sqrt{(1.5-5)^2 - (2-7)^2} = 6.10$

⇒ New centroid

$K_1 = \left(\frac{1+1.5+3}{3}, \frac{1+2+4}{3}\right) = (1.83, 2.33)$

$K_2 = \left(\frac{5+3.5+4.5+3.5}{4}, \frac{7+5+5+4.5}{4}\right) = (4.12, 5.37)$

Data Points		Distance To Center		Cluster	New Cluster
		1.83 | 2.3	4.12 | 5.37		
1	1	1.54	5.36	1	1
1.5	2	0.44	4.26	1	1
3	4	2.06	1.76	1	2
5	7	5.66	1.8	2	2
3.5	5	3.17	0.72	2	2
4.5	5	3.79	0.53	2	2
3.5	4.5	2.762	1.06	2	2

$ED = \sqrt{(X_o - X_c)^2 - (Y_o - Y_c)^2}$

- for cluster 1,

$ED(1,1) = \sqrt{(1-1.83)^2 - (1-2.3)^2} = 1.54$
$ED(1.5,2) = \sqrt{(1.5-1.83)^2 - (2-2.3)^2} = 0.44$

- for cluster 2,

$ED(1,1) = \sqrt{(1-4.12)^2 - (1-5.37)^2} = 5.36$
$ED(3,4) = \sqrt{(3-4.12)^2 - (4-5.37)^2} = 1.76$

⇒ New Centroid

$K_1 = \left(\frac{1 + 1.5}{2}, \frac{1+2}{2} \right) = (1.25, 1.5)$

$K_2 = \left(\frac{3+5+3.5+4.5+3.5}{5}, \frac{4+7+5+5+4.5}{5} \right) = (3.9, 5.1)$

Data points		Distance to Center		Cluster	New Cluster
		1.25, 1.5	3.9, 5.1		
1	1	0.55	5.02	1	1
1.5	2	0.55	3.92	1	1
3	4	3.05	1.42	2	2
5	7	6.65	2.19	2	2
3.5	5	4.16	0.41	2	2
4.5	5	4.77	0.60	2	2
3.5	4.5	3.75	0.72	2	2

- for cluster 1,

$ED(1,1) = \sqrt{(1-1.25)^2 - (1-1.5)^2} = 0.55$

$ED(1.5,2) = \sqrt{(1.5-1.25)^2 - (2-1.5)^2} = 0.55$

- for cluster 2,

$ED(3,4) = \sqrt{(3-3.9)^2 - (4-5.1)^2} = 1.42$

$ED(5,7) = \sqrt{(5-3.9)^2 - (7-5.1)^2} = 2.19$

⇒ Here no change in cluster so we terminate the Algorithm.

K-Medoids clustering

Sr.	X	Y
0	8	7
1	3	7
2	4	9
3	9	6
4	8	5
5	5	8
6	7	3
7	8	4
8	7	5
9	4	5

Here, $K=2$

let, $C_1 = (4,5)$

$C_2 = (8,5)$

Sr.	X	Y	Dissimilarity from C_1	Dissimilarity from C_2
0	8	7	$\lvert(8-4)\rvert+\lvert(7-5)\rvert = 6$	$\lvert(8-8)\rvert+\lvert(7-5)\rvert = \underline{2}$
1	3	7	$\lvert(3-4)\rvert+\lvert(7-5)\rvert = \underline{3}$	$\lvert(3-8)\rvert+\lvert(7-5)\rvert = 7$
2	4	9	$\lvert(4-4)\rvert+\lvert(9-5)\rvert = \underline{4}$	$\lvert(4-8)\rvert+\lvert(9-5)\rvert = 8$
3	9	6	$\lvert(9-4)\rvert+\lvert(6-5)\rvert = 6$	$\lvert(9-8)\rvert+\lvert(6-5)\rvert = \underline{2}$
4	8	5	-	-
5	5	8	$\lvert(5-4)\rvert+\lvert(8-5)\rvert = \underline{4}$	$\lvert(5-8)\rvert+\lvert(8-5)\rvert = 6$
6	7	3	$\lvert(7-4)\rvert+\lvert(3-5)\rvert = 5$	$\lvert(7-8)\rvert+\lvert(3-5)\rvert = \underline{3}$
7	8	4	$\lvert(8-4)\rvert+\lvert(4-5)\rvert = 5$	$\lvert(8-8)\rvert+\lvert(4-5)\rvert = \underline{1}$
8	7	5	$\lvert(7-4)\rvert+\lvert(5-5)\rvert = 3$	$\lvert(7-8)\rvert+\lvert(5-5)\rvert = \underline{1}$
9	4	5	-	-

The cost $= (3+4+4) + (2+2+3+1+1)$

$= 20$

cluster $C_1 \longrightarrow 1, 2, 5$
cluster $C_2 \longrightarrow 0, 3, 6, 7, 8$

→ randomly select non-medoid point & recalculate cost
let,
$C1 = (4, 5)$
$C2 = (8, 4)$

Sr.	X	Y	Dissimilarity from C_1	Dissimilarity from C_2
0	8	7	$\|8-4\|+\|7-5\| = 6$	$\|8-8\|+\|7-4\| = \underline{3}$
1	3	7	$\|3-4\|+\|7-5\| = \underline{3}$	$\|3-8\|+\|7-4\| = 8$
2	4	9	$\underline{4}$	9
3	9	6	6	$\underline{3}$
4	8	5	4	$\underline{1}$
5	5	8	$\underline{4}$	7
6	7	3	5	$\underline{2}$
8	7	5	3	$\underline{2}$

The cost $= (3+4+4) + (3+3+1+2+2)$
$= 22$

cluster $C_1 \longrightarrow 1, 2, 5$
cluster $C_2 \longrightarrow 0, 3, 4, 6, 8$

→ Swap cost = New Cost − Previous Cost
$= 22 - 20 = 2$

∴ So, $2 > 0$ that is possitive, now our previous medoid is best.