K-Medoid Clustering – Solved Example – 1

i	x	y
X1	2	6
X2	3	4
X3	3	8
X4	4	7
X5	6	2
X6	6	4
X7	7	3
X8	7	4
X9	8	5
X10	7	6

- Apply K-Medoid clustering algorithm to form two clusters.
- Use Manhattan distance to find the between data point and medoid.

K-Medoid Clustering – Solved Example – 1

Step 1

- Select two medoids
- C1=(3, 4)
- C2=(7, 4)
- $\boldsymbol{Manhattan\ Dist} = |\boldsymbol{x_1} - \boldsymbol{x_2}| + |\boldsymbol{y_1} - \boldsymbol{y_2}|$
- $\boldsymbol{Mdist}[(\mathbf{2},\mathbf{6}),(\mathbf{3},\mathbf{4})] = |\mathbf{2}-\mathbf{3}| + |\mathbf{6}-\mathbf{4}| = \mathbf{3}$
- $\boldsymbol{Mdist}[(\mathbf{3},\mathbf{4}),(\mathbf{3},\mathbf{4})] = |\mathbf{3}-\mathbf{3}| + |\mathbf{4}-\mathbf{4}| = \mathbf{0}$

i	x	y	C1	C2	Cluster
X1	2	6	3	7	C1
X2	3	4	0	4	C1
X3	3	8	4	8	C1
X4	4	7	4	6	C1
X5	6	2	5	3	C2
X6	6	4	3	1	C2
X7	7	3	5	1	C2
X8	7	4	4	0	C2
X9	8	5	6	2	C2
X10	7	6	6	2	C2

K-Medoid Clustering – Solved Example – 1

Step 2

- **Cluster are**
- **C1: {(2,6), (3,4), (3,8), (4,7)}**
- **C2: {(6, 2), (6, 4), (7, 3), (7, 4), (8, 5), (7,6)}**

i	x	y	C1	C2	Cluster
X1	2	6	3	7	C1
X2	3	4	0	4	C1
X3	3	8	4	8	C1
X4	4	7	4	6	C1
X5	6	2	5	3	C2
X6	6	4	3	1	C2
X7	7	3	5	1	C2
X8	7	4	4	0	C2
X9	8	5	6	2	C2
X10	7	6	6	2	C2

- C1: {(2,6), **(3,4),** (3,8), (4,7)}
- C2: {(6, 2), (6, 4), (7, 3), **(7, 4),** (8, 5), (7,6)}
- **Calculate the Total Cost**
- $\boldsymbol{Cost(c,x)} = \sum_i |\boldsymbol{c_i} - \boldsymbol{x_i}|$ 3 1+2 0+4
- $Total\ Cost = \{Cost((3,4),(2,6)) + Cost((3,4),(3,8)) + Cost((3,4),(4,7)) + Cost((7,4),(6,2)) + Cost((7,4),(6,4)) + Cost((7,4),(7,3)) + Cost((7,4),(8,5)) + Cost((7,4),(7,6))\}$
- $\boldsymbol{Total\ Cost = 3 + 4 + 4 + 2 + 3 + 1 + 1 + 2 = 20}$

K-Medoid Clustering – Solved Example – 1

Step 3

- **Randomly select one non-medoid point and recalculate the cost.**
- C1=(3, 4) and C2=(7, 4)
- **O=(7, 3)**
- **Swap C2 with O**
- **New Medoids**
- **C1=(3, 4) and O=(7, 3)**

i	x	y	C1	O	Cluster
X1	2	6			
X2	3	4			
X3	3	8			
X4	4	7			
X5	6	2			
X6	6	4			
X7	7	3			
X8	7	4			
X9	8	5			
X10	7	6			

Step 3

- **New Medoids**
- **C1=(3, 4) and O=(7, 3)**
- $\boldsymbol{Manhattan\ Dist} = |\boldsymbol{x_1} - \boldsymbol{x_2}| + |\boldsymbol{y_1} - \boldsymbol{y_2}|$
- $\boldsymbol{Mdist}[(\boldsymbol{2},\boldsymbol{6}),(\boldsymbol{7},\boldsymbol{3})] = |\boldsymbol{2}-\boldsymbol{7}| + |\boldsymbol{6}-\boldsymbol{3}| = \boldsymbol{8}$

i	x	y	C1	O	Cluster
X1	2	6	3	8	C1
X2	3	4	0	5	C1
X3	3	8	4	9	C1
X4	4	7	4	7	C1
X5	6	2	5	2	O
X6	6	4	3	2	O
X7	7	3	5	0	O
X8	7	4	4	1	O
X9	8	5	6	3	O
X10	7	6	6	3	O

K-Medoid Clustering – Solved Example – 1

- C1: {(2,6), **(3,4),** (3,8), (4,7)}
- O: {(6, 2), (6, 4), **(7, 3),** (7, 4), (8, 5), (7,6)}
- **Calculate the Total Cost**
- $\boldsymbol{Cost}(\boldsymbol{c},\boldsymbol{x}) = \sum_i |\boldsymbol{c_i} - \boldsymbol{x_i}|$
- $Current\ Total\ Cost = \{Cost\big((3,4),(2,6)\big) + Cost\big((3,4),(3,8)\big) + Cost\big((3,4),(4,7)\big) + Cost\big((7,3),(6,2)\big) + Cost\big((7,3),(6,4)\big) + Cost\big((7,3),(7,4)\big) + Cost\big((7,3),(8,5)\big) + Cost\big((7,3),(7,6)\big)\}$
- $\boldsymbol{Current\ Total\ Cost} = \boldsymbol{3+4+4+2+2+1+3+3 = 22}$

Step 4

- Cost of Swapping of medoid C2 with O
- $S = Current\ Total\ Cost - Previous\ Total\ Cost$
- $S = 22 - 20 = 2 > 0$
- Hence Swapping C2 with O is not a good Idea.
- Final Medoids are C1 =(3, 4) and C2=(7, 4)
- Clusters are
- C1: {(2,6), (3,4), (3,8), (4,7)}
- C2: {(6, 2), (6, 4), (7, 3), (7, 4), (8, 5), (7,6)}

K-Medoid Clustering – Solved Example – 2

Point	X	Y
1	1	4
2	5	1
3	5	2
4	5	4
5	10	4
6	25	4
7	25	6
8	25	7
9	25	8
10	29	7

- Apply K-Medoid clustering algorithm to form two clusters.
- Use Euclidean distance to find the distance between data point and medoid.

Step 1

- Let us choose that (1, 4) and (10, 4) are the medoids
- C1=(1, 4) and C2=(10, 4)
- (x_1, y_1) *and* (x_2, y_2) *are data points*
- $\boldsymbol{Eucladian\ Dist} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$
- $\boldsymbol{EDist}[(\mathbf{1,4}),(\mathbf{1,4})] = \sqrt{(\mathbf{1}-\mathbf{1})^2 + (\mathbf{4}-\mathbf{4})^2} = \mathbf{0}$
- $\boldsymbol{EDist}[(\mathbf{1,4}),(\mathbf{5,1})] = \sqrt{(\mathbf{5}-\mathbf{1})^2 + (\mathbf{1}-\mathbf{4})^2} = \mathbf{5}$

Point	X	Y	C1	C2	Cluster
1	1	4	0	9	C1
2	5	1	5	5.83	C1
3	5	2	4.47	5.39	C1
4	5	4	4	5	C1
5	10	4	9	0	C2
6	25	4	24	15	C2
7	25	6	24.08	15.13	C2
8	25	7	24.19	15.30	C2
9	25	8	24.33	15.52	C2
10	29	7	28.16	19.24	C2

Step 1

- **Cluster are**
- **C1: {(1,4), (5,1), (5,2), (5,4)} and**
- **C2: {(10,4), (25,4), (25,6), (25,7), (25,8), (29,7)}.**
- **Now calculating the cost which is nothing but the sum of distance of each non-medoid point from the medoid of the cluster it belongs to.**

Point	X	Y	C1	C2	Cluster
1	1	4	0	9	C1
2	5	1	5	5.83	C1
3	5	2	4.47	5.39	C1
4	5	4	4	5	C1
5	10	4	9	0	C2
6	25	4	24	15	C2
7	25	6	24.08	15.13	C2
8	25	7	24.19	15.30	C2
9	25	8	24.33	15.52	C2
10	29	7	28.16	19.24	C2

- **C1: {(1,4), (5,1), (5,2), (5,4)}** **and** **C2: {(10,4), (25,4), (25,6), (25,7), (25,8), (29,7)}.**
- **Calculate the Total Cost**
- $\boldsymbol{Cost}(\boldsymbol{c}, \boldsymbol{x}) = \sum_i |\boldsymbol{c}_i - \boldsymbol{x}_i|$
- $Total\ Cost = cost((1,4),(5,1)) + cost((1,4),(5,2)) + cost((1,4),(5,4)) + cost((10,4),(25,4)) + cost((10,4),(25,6)) + cost((10,4),(25,7)) + cost((10,4),(25,8)) + cost((10,4),(29,7))$
- $Total\ Cost = \{|1-5| + |4-1|\} + \{|1-5| + |4-2|\} + \{|1-5| + |4-4|\} + \{|10-25| + |4-4|\} + \{|10-25| + |4-6|\} + \{|10-25| + |4-7|\} + \{|10-25| + |4-8|\} + \{|10-29| + |4-7|\}$
- $\boldsymbol{Total\ Cost} = \mathbf{4+3+4+2+4+0+15+0+15+2+15+3+15+4+19+3 = 108}$

Step 2

- C1=(5, 4) and C2=(25, 7)
- $Eucladian\ Dist = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$
- $EDist[(1,4),(5,4)] = \sqrt{(5-1)^2 + (4-4)^2} = 4$
- $EDist[(1,4),(25,7)] =$

  $\sqrt{(25-1)^2 + (7-4)^2} = 24.19$

Point	X	Y	C1	C2	Cluster
1	1	4	4	24.19	C1
2	5	1	3	20.8	C1
3	5	2	2	20.62	C1
4	5	4	0	20.22	C1
5	10	4	5	15.30	C1
6	25	4	20	3	C2
7	25	6	20.10	1	C2
8	25	7	20.22	0	C2
9	25	8	20.40	1	C2
10	29	7	24.19	4	C2

K-Medoid Clustering – Solved Example – 2

- C1: {(1,4), (5,1), (5,2), (5,4), (10, 4)} and
- C2: {(25,4), (25,6), (25,7), (25,8), (29,7)}.
- Calculate the Total Cost
- $Cost(c,x) = \sum_i |c_i - x_i|$
- $Total\ Cost = cost((5,4),(1,4)) + cost((5,4),(5,1)) + cost((5,4),(5,2)) + cost((5,4),(10,4)) + cost((25,7),(25,4)) + cost((25,7),(25,6)) + cost((25,7),(25,8)) + cost((25,7),(29,7))$
- $Total\ Cost = 23$

Step 2

- $Swapping\ cost = Current\ Total\ Cost - Previous\ Total\ Cost$
- $Swapping\ cost = 23 - 108 = -85 \nless 0$
- If swapping cost > 0, we stop the algorithm here, and clusters formed in **Step 1** is the final answer

Step 3

- **Randomly select two non-medoid point and recalculate the cost.**
- C5=(10, 4) and C6=(29, 7)
- **Swap C1 with C5 and C2 with C6**
- **New Medoids**
- **C1=(10, 4) and C2=(29, 7)**

Point	X	Y	C1	C2	Cluster
1	1	4			
2	5	1			
3	5	2			
4	5	4			
5	10	4			
6	25	4			
7	25	6			
8	25	7			
9	25	8			
10	29	7			

K-Medoid Clustering – Solved Example – 2

Step 3

- **C1=(10, 4) and C2=(29, 7)**
- $\boldsymbol{Eucladian\ Dist} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$

Point	X	Y	C1	C2	Cluster
1	1	4	9	28.16	C1
2	5	1	5.8	24.73	C1
3	5	2	5.3	24.51	C1
4	5	4	5	24.18	C1
5	10	4	0	19.23	C1
6	25	4	15	5	C2
7	25	6	15.13	4.12	C2
8	25	7	15.29	4	C2
9	25	8	15.52	4.12	C2
10	29	7	19.23	0	C2

Step 3

- Cluster are
- C1: {(1,4), (5,1), (5,2), (5,4), (10, 4)} and
- C2: {(25,4), (25,6), (25,7), (25,8), (29,7)}.

- $\text{Total cost} = 9+5+3+5+2+5+0+4+3+4+1+4+0+4+1 = 46$

Point	X	Y	C1	C2	Cluster
1	1	4	9	28.16	C1
2	5	1	5.8	24.73	C1
3	5	2	5.3	24.51	C1
4	5	4	5	24.18	C1
5	10	4	0	19.23	C1
6	25	4	15	5	C2
7	25	6	15.13	4.12	C2
8	25	7	15.29	4	C2
9	25	8	15.52	4.12	C2
10	29	7	19.23	0	C2

K-Medoid Clustering – Solved Example – 2

Step 3

- $\textit{Swapping cost} = \textit{Current Total Cost} - \textit{Previous Total Cost}$
- $\textit{Swapping cost} = 46 - 23 = 23 > 0$
- Swapping cost > 0, we stop the algorithm here.

- Our final clusters are
- C1: {(1,4), (5,1), (5,2), (5,4), (10,4)} and
- C2: {(25,4), (25,6), (25,7), (25,8), (29,7)}

K-Medoid vs K-Means Clustering

- Partitioning Around Medoids or the K-medoids algorithm is a partitional clustering algorithm which is slightly modified from the K-means algorithm.
- In K-means algorithm, they choose means as the centroids but in the K-medoids, data points are chosen to be the medoids.