Digital Image Processing

Dr. Mubashir Ahmad (Ph.D.)

Solved Example 1 K-Mean clustering

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.

Solved Example 1 K-Mean clustering

Step 1

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

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

Solved Example 1 K-Mean clustering

Step 1

- Select two medoids
- C1=(3, 4)
- C2=(7, 4)
- $\boldsymbol{Manhattan\ Dist} = |x_1 - x_2| + |y_1 - y_2|$

$|2-7| + |6-4| = 5+2 = 7$

$|3-7| + |4-4| = 4+0 = 4$

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

Solved Example 1 K-Mean clustering

If C1 < C2 then result is C1 otherwise C2

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

Solved Example 1 K-Mean clustering

- 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**
- $\boldsymbol{Cost(c,x)} = \sum_i |\boldsymbol{c_i} - \boldsymbol{x_i}|$

$= |3-2| + |4-6| = 1+2 = 3$

- $Total\ Cost = \{Cost\big((3,4),(2,6)\big) + Cost\big((3,4),(3,8)\big) + Cost\big((3,4),(4,7)\big) +$

  $Cost\big((7,4),(6,2)\big) + Cost\big((7,4),(6,4)\big) + Cost\big((7,4),(7,3)\big) + Cost\big((7,4),(8,5)\big) +$

  $Cost\big((7,4),(7,6)\big)\}$

$\boldsymbol{Total\ Cost} = \mathbf{3+4+4+2+3+1+1+2} = \mathbf{20}$

Solved Example 1 K-Mean clustering

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	C2	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			

Solved Example 1 K-Mean clustering

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

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

Solved Example 1 K-Mean clustering

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

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

Solved Example 1 K-Mean clustering

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

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

Solved Example 1 K-Mean clustering

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

Solved Example 1 K-Mean clustering

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.**

Solved Example 2 K-Mean clustering

- Suppose that the data mining task is to cluster points into three clusters,
- where the points are
- $A1(2, 10)$, $A2(2, 5)$, $A3(8, 4)$, $B1(5, 8)$, $B2(7, 5)$, $B3(6, 4)$, $C1(1, 2)$, $C2(4, 9)$.
- The distance function is Euclidean distance.
- Suppose initially we assign $A1$, $B1$, and $C1$ as the center of each cluster, respectively.

Solved Example 2 K-Mean clustering

Initial Centroids:
A1: (2, 10)
B1: (5, 8)
C1: (1, 2)

Data Points			Distance to						Cluster	New Cluster
			2	10	5	8	1	2		
A1	2	10								
A2	2	5								
A3	8	4								
B1	5	8								
B2	7	5								
B3	6	4								
C1	1	2								
C2	4	9								

$$d(p_1, p_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

ED

Solved Example 2 K-Mean clustering

Initial Centroids:
A1: (2, 10)
B1: (5, 8)
C1: (1, 2)

Data Points			Distance to						Cluster	New Cluster
			2	10	5	8	1	2		
A1	2	10	0.00							
A2	2	5	5.00							
A3	8	4	8.49							
B1	5	8	3.61							
B2	7	5	7.07							
B3	6	4	7.21							
C1	1	2	8.06							
C2	4	9	2.24							

$$d(p_1, p_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Solved Example 2 K-Mean clustering

Initial Centroids:
A1: (2, 10)
B1: (5, 8)
C1: (1, 2)

Data Points			Distance to						Cluster	New Cluster
			2	10	5	8	1	2		
A1	2	10	0.00		3.61		8.06			
A2	2	5	5.00		4.24		3.16			
A3	8	4	8.49		5.00		7.28			
B1	5	8	3.61		0.00		7.21			
B2	7	5	7.07		3.61		6.71			
B3	6	4	7.21		4.12		5.39			
C1	1	2	8.06		7.21		0.00			
C2	4	9	2.24		1.41		7.62			

$$d(p_1, p_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Solved Example 2 K-Mean clustering

Initial Centroids:

A1: (2, 10)

B1: (5, 8)

C1: (1, 2)

Data Points			Distance to						Cluster	New Cluster
			2	10	5	8	1	2		
A1	2	10	0.00		3.61		8.06		1	
A2	2	5	5.00		4.24		3.16		3	
A3	8	4	8.49		5.00		7.28		2	
B1	5	8	3.61		0.00		7.21		2	
B2	7	5	7.07		3.61		6.71		2	
B3	6	4	7.21		4.12		5.39		2	
C1	1	2	8.06		7.21		0.00		3	
C2	4	9	2.24		1.41		7.62		2	

New Centroids:

A1: (2, 10)

B1: (6, 6)

C1: (1.5, 3.5)

$\frac{30}{5} : 6$

Solved Example 2 K-Mean clustering

Current Centroids:

A1: (2, 10)

B1: (6, 6)

C1: (1.5, 3.5)

Data Points			Distance to			Cluster	New Cluster
			2 10	6 6	1.5 3.5		
A1	2	10	0.00	5.66	6.52	1	1
A2	2	5	5.00	4.12	1.58	3	3
A3	8	4	8.49	2.83	6.52	2	2
B1	5	8	3.61	2.24	5.70	2	2
B2	7	5	7.07	1.41	5.70	2	2
B3	6	4	7.21	2.00	4.53	2	2
C1	1	2	8.06	6.40	1.58	3	3
C2	4	9	2.24	3.61	6.04	2	1

$$d(p_1, p_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Solved Example 2 K-Mean clustering

Current Centroids:
A1: (3, 9.5)
B1: (6.5, 5.25)
C1: (1.5, 3.5)

New Centroids:
A1: (3.67, 9)
B1: (7, 4.33)
C1: (1.5, 3.5)

Data Points			Distance to						Cluster	New Cluster
			3	9.5	6.5	5.25	1.5	3.5		
A1	2	10	1.12		6.54		6.52		1	
A2	2	5	4.61		4.51		1.58		3	
A3	8	4	7.43		1.95		6.52		2	
B1	5	8	2.50		3.13		5.70		1	
B2	7	5	6.02		0.56		5.70		2	
B3	6	4	6.26		1.35		4.53		2	
C1	1	2	7.76		6.39		1.58		3	
C2	4	9	1.12		4.51		6.04		1	

Solved Example 2 K-Mean clustering

Current Centroids:

A1: (3.67, 9)

B1: (7, 4.33)

C1: (1.5, 3.5)

Data Points			Distance to						Cluster	New Cluster
			3.67	9	7	4.33	1.5	3.5		
A1	2	10	1.94		7.56		6.52		1	1
A2	2	5	4.33		5.04		1.58		3	3
A3	8	4	6.62		1.05		6.52		2	2
B1	5	8	1.67		4.18		5.70		1	1
B2	7	5	5.21		0.67		5.70		2	2
B3	6	4	5.52		1.05		4.53		2	2
C1	1	2	7.49		6.44		1.58		3	3
C2	4	9	0.33		5.55		6.04		1	1

$$d(p_1, p_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$