Nitte Meenakshi Institute of Technology
Department of Computer Science & Engineering
Course Coordinator: Dr.Vani V

Clustering: K-Means

Exercise:

1. Consider the 1-dimensional data set with 10 data points {1, 2, 3, . . . 10}. Show three iterations of the k-means algorithms when k = 2, and the random seeds are initialized to {1,2}. Repeat the problem with random seeds {2,9}. How did the different choice of the seed set affect the quality of the results?

Use Manhattan $|m_1-x_i|$ $|m_2-x_i|$ i=1 to 10

Solution:

x_i	m1(1)	m2(2)	MANHATTAN		
1	0	1			
2	1	0			
3	2	1			
4	3	2			
5	4	3			
6	5	4			
7	6	5			
8	7	6			
9	8	7			
10	9	8			
INITIAL:	C1={1}	C2={2,3,4,5,6,7,8,9,10}		6	
dp	m1(1)	m2(6)			
1	0	5			
2	1	4			
3	2	3			
4	3	2			
5	4	1			
6	5	0			
7	6	1			
8	7	2			
9	8	3			
10	9	4			
ITERATION 1	C1={1,2,3}	C2={4,5,6,7,8,9,10}		7	

dp	m1(2)	m2(7)			
1	1	6			
2	0	5			
3	1	4			
4	2	3			
5	3	2			
6	4	1			
7	5	0			
8	6	1			
9	7	2			
10	8	3			
ITERATION 2:	C1={1,2,3,4}	2.5	C2={5,6,7,8,9,10}		7.5
dp	m1(2.5)	m2(7.5)			
1	1.5	6.5			
2	0.5	5.5			
3	0.5	4.5			
4	1.5	3.5			
5	2.5	2.5			
6	3.5	1.5			
7	4.5	0.5			
8	5.5	0.5			
9	6.5	1.5			
10	7.5	2.5			
ITERATION 3: c1={1,2,3,4,5}					
	c2=(6,7,8,9,10}				

- √ With the initial seed as {2,9}; the quality of the algorithm is improvised, and the algorithm converges in the first iteration itself.
- √ Initial:C1(m1=2) = {1,2,3,4,5}
  C2(m2=9) = {6,7,8,9,10}
- √ Iteration 1: C1(m1=3) = {1,2,3,4,5}
  C2(m2=8) = {6,7,8,9,10}