Assignment #06 18F-0326

Q.No 1

C1 ⟹ (1, 4)
C2 ⟹ (3, 8)
C3 ⟹ (7, 2)

Iteration 1st

M1:

$C1 - M_1 = |1-2| + |4-10|$
$= |-1| + |-6|$
$= 7$

$C2 - M_1 = |3-2| + |8-10|$
$= |1| + |-2|$
$= 3$

$C3 - M_1 = |7-2| + |2-10|$
$= |5| + |-8|$
$= 13$

M2:

$C_1M = |1-2| + |4-5|$; $C_2M = |3-2| + |8-5|$; $C_3M = |7-2| + |2-5|$
$C_1 = 2$ $C_2 = 4$ $C_3 = 5$

M_8:

$C_1M = |1-8| + |4-4|$; $C_2M = |3-8| + |8-4|$; $C_3M = |7-8| + |2-4|$

$C_1 = 7$; $C_2 = 9$; $C_3 = 3$

M_4:

$C_1M = |1-5| + |4-8|$; $C_2M = |3-5| + |8-8|$; $C_3M = |7-5| + |2-8|$

$C_1 = 8$; $C_2 = 2$; $C_3 = 8$

M_5:

$C_1M = |1-7| + |4-5|$; $C_2M = |3-7| + |8-5|$; $C_3M = |7-7| + |2-5|$

$C_1 = 7$; $C_2 = 7$; $C_3 = 3$

M_6:

$C_1M = |1-6| + |4-4|$; $C_2M = |3-6| + |8-4|$; $C_3M = |7-6| + |2-4|$

$C_1 = 5$; $C_2 = 7$; $C_3 = 3$

M_7:

$C_1M = |1-1| + |4-2|$; $C_2M = |3-1| + |8-2|$; $C_3M = |7-1| + |2-2|$

$C_1 = 2$; $C_2 = 8$; $C_3 = 6$

M_8:

$C_1M = |1-4| + |4-9|$; $C_2M = |3-4| + |8-9|$; $C_3M = |7-4| + |2-9|$

$C_1 = 7$ $C_2 = 2$ $C_3 = 10$

Clusters

$C_1 = \{M_2, M_7\}$

$C_2 = \{M_1, M_4, M_8\}$

$C_3 = \{M_3, M_5, M_6\}$

centroid calculation:

$$C_1 = \left(\frac{|2|+|11|}{2}, \frac{|5|+|2|}{2}\right)$$

$$C_1 = (1.5, 3.5)$$

$$C_2 = \left(\frac{|2|+|5|+|4|}{3}, \frac{|10|+|8|+|9|}{3}\right)$$

$$C_2 = (3.66, 9)$$

$$C_3 = \left(\frac{|8|+|7|+|6|}{3}, \frac{|4|+|5|+|4|}{3}\right)$$

$$C_3 = (7, 4.33)$$

Iteration 2:

	C_1-M	C_2-M	C_3-M
M_1 :	7	2.66	10.67
M_2 :	2	5.66	5.67
M_3 :	7	9.34	1.33
M_4 :	8	2.34	5.67
M_5 :	7.0	7.34	0.66
M_6 :	5.0	7.34	1.33
M_7 :	2	9.66	8.33
M_8 :	8.0	0.33	7.67

Clustering:

$$C_1 = \{M_2, M_7\}$$
$$C_2 = \{M_1, M_4, M_8\}$$
$$C_3 = \{M_3, M_5, M_6\}$$

Centroid calculation:

$$C_1 = (1.5, 3.5)$$
$$C_2 = (3.66, 9)$$
$$C_3 = (7, 3.33)$$

as centroid values remains same so, we found our K-means clusters.

Q.No 2:

K-mean:

K-mean is a classical Partitioning technique of clustering that clusters the data set of n object into K-clusters with K known a prior.

Time complexities:

K-mean $O((n+k)d)$

Advantages:
- Simple to implement
- Scales to large dataset
- Guarantees convergence
- Generalize to clusters

K-medoid:

K-medoid is related to K-mean. where K-mean minimize the total squared error while K-medoids minimize the sum of dissimilarities between points labeled to be in a cluster; K-medoids chooses datapoints as centers.

Time complexities:

K-medoid $O(k(n-k)^2)$

Advantages:
- Easy to understand and execute
- Quick and convergent
- Allow using general dissimilarities of
- Normally less delicate to outlier than K-means

Disadvantages:

- choose K manually
- ~~dependent~~ on initial values
- Clustering data of varying size and densities
- scaling with number of dimensia

Disadvantages:

- different initial sets of medoids can lead to different final clustering
- Resulting ~~units~~ clusters depends on the unit of measurement.