1/9

Q2]

D]

K-means algorithm

{2, 3, 6, 8, 9, 12, 15, 18, 22}

Break into 3 clusters (Randomly assign data to 3 clusters) and calculate mean value.

① K1 = 2, 8, 15 mean = 8.3
K2 = 3, 9, 18 mean = 10
K3 = 6, 12, 22 mean = 13.3

② Re-assign
K1 = 2, 3, 6, 8, 9 mean = 5.6
K2 = mean = 0
K3 = 12, 15, 18, 22 mean = 16.75

③ Reassign
K1 = 3, 6, 8, 9 mean = 6.5
K2 = 2 mean = 2
K3 = 12, 15, 18, 22 mean = 16.75

④ K1 = 6, 8, 9 mean = 7.6
K2 = 2, 3 mean = 2.5
K3 = 12, 15, 18, 22 mean = 16.75

⑤ K1 = 6, 8, 9 mean = 7.6 K3 = 12, 15, 18, 22
K2 = 2, 3 mean = 2.5 mean = 16.75

last 2 groups are same so we have
cluster 1 = {6, 8, 9} cluster 2 = {2, 3} cluster 3 = {12, 15, 18, 22}

Q2]

F]

Spatial clustering Technique: CLARANS.

① CLARANS known as clustering large Applications based upon RANdomized search

② CLARANS like PAM starts with a randomly selected set of k-mediods (or few pairs) instead of examing all pairs, for swapping at the current state.

③ It check at most the "maxneighbour" number of pairs for swapping and if a pair with negative cost is found, it update the mediods as a local optimum and restarts with a new randomly selected mediod, set to search for another local optimum.

④ CLARANS stops after the "numlocal" number of local optimal mediod sets are determined, and return the best among these.

⑤ Algorithm steps

i) Input parameters numlocal and maxneighbour. Initialize i to 1 and mincost to a large number

ii) Set current to an arbitary node in Gn,k

iii) set j to 1

iv) consider a random neighbours of current and cal the cost differential of the two nodes

v) If S has a lower cost, set current to S and go to step 3

vi) otherwise increment j by 1 If j ≤ maxneighbour

vii) otherwise when j > maxneighbour, compare the cost of current with mincost - If the previous is less than mincost, set mincost to the cost of

current and set bestnode to current

viii) Increment i by 1. If i > numlocal, output bestnode and halt. Otherwise go to step 2.

Q2]

A]

Metadata can also be considered as an equivalent of Amazon book store. If we consider each data element as a book, the meta-data will contain

- Name of the book :
- Summary of the book :
- Assessments about the book -
- Date of publication
- High level description of what it contains :
- Publishers of the book
- How can you find the book
- Author of the book
- Whether the book is available OR not
- This information helps you to :
  - Search for the book
  - Access the book
  - Understand about the book befor you access OR buy it
- Responsible User
- Data Quality Revieved
- Last update Date
- Last Refresh cycle
- Last Duplication

B2020
SEM:06 Rebecca Dsuza

4/9

Q2]
E]

Min support = 2

Item	Support Count
I_2	7
I_1	6
I_3	6
I_4	2
I_5	2

Since the transactions satisfy the min support consider the following transactions for FP Tree

FP Tree :-