Name: AMEY THAKUR COMPS Class: TE Div: B Roll No: 50

Subject: DWM Topic: IAT-2 Date: Page No:

Q6 A)

Hierarchical Clustering

- A hierarchical clustering method works via grouping data into a tree of cluster.
- Hierarchical clustering begins by treating every data point as separate cluster.
- Then it is repeatedly executes the subsequent steps.

(1) Identify the two clusters which can be closest together, and
(2) Merge 2 maximum comparable clusters.
We need to continue these steps until all clusters are merged together.

- In Hierarchical clustering, the aim is to produce a hierarchical series of nested clusters.
- A diagram called dendrogram graphically represents this Hierarchy and is an inverted tree that describe the order in which factors are merged or cluster are break up

Method to generate hierarchical clustering are
(1) Agglomerative
(2) Divisive

TU3F1819127

Amey

Name: AMEY THAKUR COMPS Class: TE Div: B Roll No: 50

Subject: DWM Topic: IAT-2 Date: Page No:

6 A)

K-means Algorithm

Method 1

(1) Randomly assign means : $m_1 = 3$, $m_2 = 4$

(2) The numbers which are close to mean $m_1 = 3$ are grouped into cluster $K_1$ and members which are close to mean $m_2 = 4$ are grouped into cluster $K_2$.

(3) Again calculate the new mean for new cluster groups

(4) $K_1 = \{2, 3\}$, $K_2 = \{4, 10, 12, 20, 30, 11, 25\}$,
$m_1 = 2.5$, $m_2 = 16$

(5) $K_1 = \{2, 3, 4\}$, $K_2 = \{10, 12, 20, 30, 11, 25\}$,
$m_1 = 3$, $m_2 = 18$

(6) $K_1 = \{2, 3, 4, 10\}$, $K_2 = \{12, 20, 30, 11, 25\}$
$m_1 = 4.75$, $m_2 = 19.6$

(7) $K_1 = \{2, 3, 4, 10, 11, 12\}$, $K_2 = \{20, 30, 25\}$,
$m_1 = 7$, $m_2 = 25$

(8) $K_1 = \{2, 3, 4, 10, 11, 12\}$, $K_2 = \{20, 30, 25\}$

TU3F1819127 Amey

Name: AMEY THAKUR COMPS Class: TE Div: B Roll No: 50

Subject: DWM Topic: IAT-2 Date: Page No:

(9) Stop as the clusters which means (in step 7 & 8) are the same.
The clusters in last 2 groups are identical.

(10) So the final answer is $K_1$ = {2, 3, 4, 10, 11, 12}
$K_2$ = {20, 30, 25}

AMEY B 50 Amey

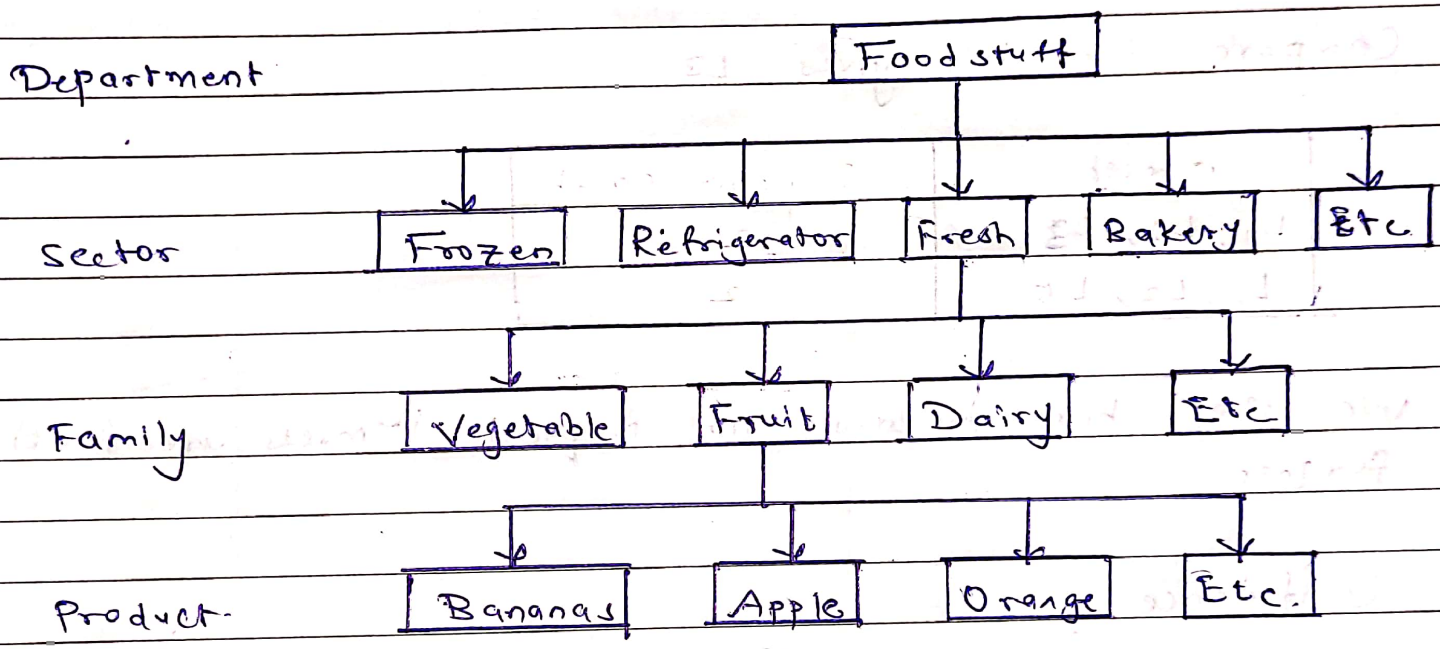
Page No.: Date: YOUVA

Q6.B — Association — Rules

Ans:

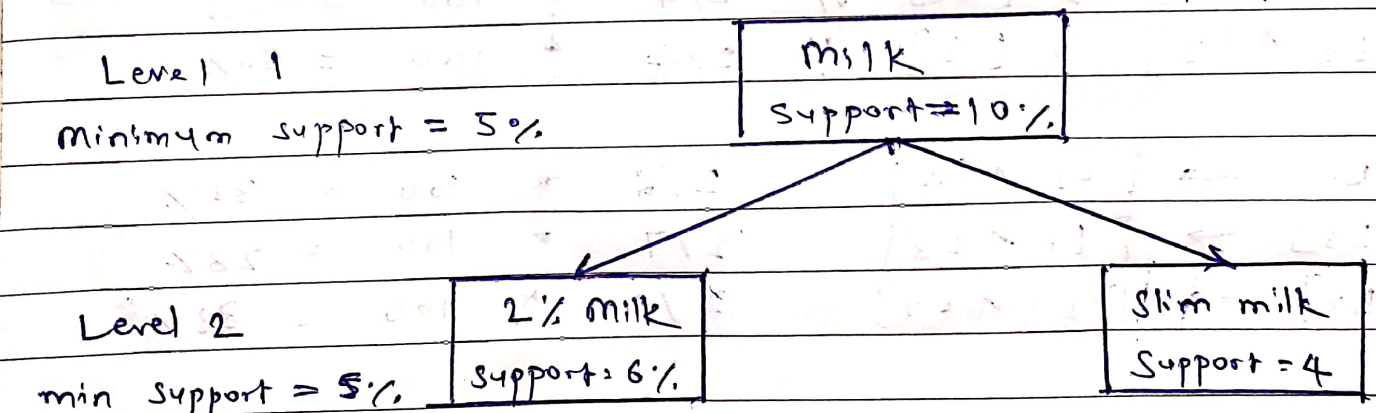
Multilevel Mining Association rules

- Items often form hierarchy.
- Items of the lower has lower support.
- Rules which contain associations with hierarchy of concepts are called multilevel association rules.




(1) Using uniform support level for all levels

- The same minimum support for all levels
- There is only one minimum support threshold no need to examine itemsets
- If support threshold is too low → generate too many high level association.

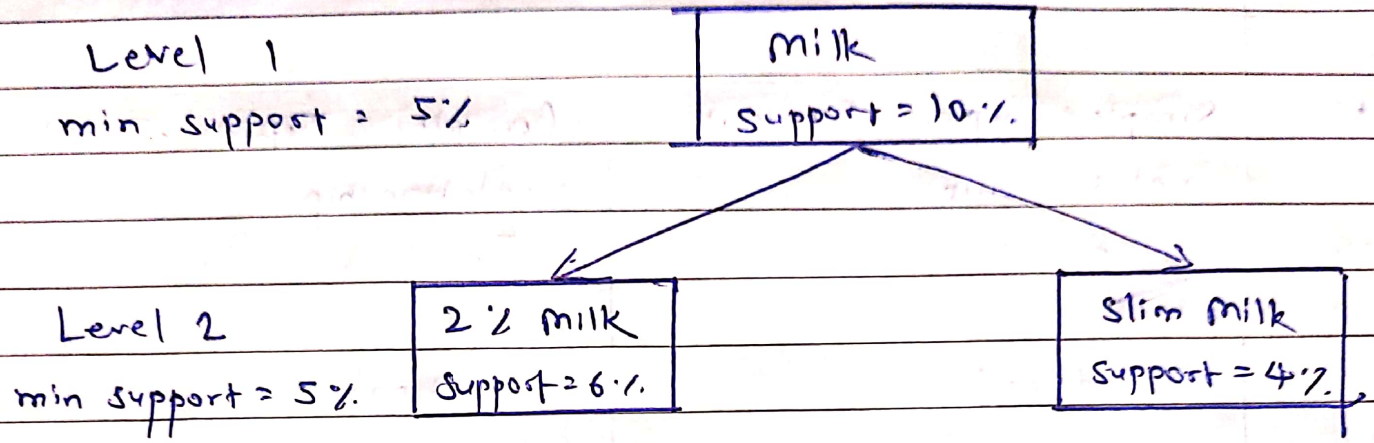


AMEY B. 50 Amey

Page No.: Date: YOUVA

(2) Using reduced minimum support level for all levels

- At every level of abstraction, there is its own minimum support threshold
- So minimum support at lower level reduced.




Multidimensional Association Rule.

(1) In multidimensional association.

- Attribute can be categorical or quantitative
- Quantitative attributes are numeric and incorporates hierarchy
- Numeric attributes must be discretized.
- It consists of more than one dimension.

(2) Three approaches in mining multi dimensional association rule.

(i) Using static criticization of quantitative attributes
(ii) Using dynamic discriticization of quantitative attributes
(iii) Using distance based discriticization with clustering

AMEY B 50 Amey

Page No.: YOUVA
Date:

① Using static discritization of quantitative attributes

- Discritization is static and occurs prior to mining
- Discritized attributes are treated as categorical
- Use apriori algorithm to find all k-frequent predicate sets. (this requires 'k or k+1 table scans.
- Every subset of frequent predicate set must be frequent.
- Eg If in a data cube the 3D cuboid (age, income, buys) is frequent implies (age, income), (age, buys), (income, buys) are also frequent.
- Data cubes are well suited for mining since they make mining faster.
- The cells of an n dimensional data cuboid correspond to the predicate cells.

② Using dynamic discritization of quantitative attribute

- Known as mining quantitative association rules
- Numeric attributes are dynamically discreticized
- Eg: age(X, "20..25") ∧ income(X, "30k..41.k") buys (X, "Laptop Computer")

| | Age=20 | Age=21 | Age=22 | Age=23 |
|---|---|---|---|---|
| income, 38 to 41 | | | | |
| income, 34 to 37 | | | | |
| income, 30 to 33 | | | | |

Grid for Tuples.

AMEY B 50 Amey

Page No.:
Date:
YOUVA

(3) Using distance based discritization with clustering

- It involves a two step mining process
  - → Perform clustering to find the interval of attributes involved.
  - → Object association rules by searching for groups of clusters that occur together.

- The resultant rules may satisfy
  - → Clusters in the rule antecedent are strongly associated with clusters of rules in the consequent
  - → Clusters in the antecedent occur together
  - → Clusters in the consequent occur together.

Name: AMEY THAKUR Comps Class: TE Div: B Roll No: 50

Subject: DWM Topic: IAT-2 Date: Page No:

Q6 B.

Step 1:

| Items | Frequency | Support. |
|---|---|---|
| 1 | 3 | 3/5 = 60% |
| 2 | 3 | 3/5 = 60% |
| 3 | 3 | 3/5 = 60% |
| 4 | 2 | 2/5 = 40% |
| 5 | 3 | 3/5 = 60% |
| 6 | 1 | 1/5 = 20% |
| 7 | 1 | 1/5 = 20% |
| 8 | 1 | 1/5 = 20% |
| 9 | 1 | 1/5 = 20% |
| 10 | 1 | 1/5 = 20% |

Step 2:

| Item | Frequency | Support. |
|---|---|---|
| 1,2 | 1 | 20% |
| 1,3 | 2 | 40% |
| 1,4 | 2 | 40% |
| 1,5 | 1 | 20% |
| 2,3 | 2 | 40% |
| 2,4 | 0 | 0 |
| 2,5 | 3 | 60% |
| 3,4 | 1 | 20% |
| 3,5 | 2 | 40% |
| 4,5 | 0 | 0 |

TV3F1819127

Amey