

(AN AUTONOMOUS INSTITUTION UNDER MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WEST BENGAL)

The figures in the margin indicate full marks
Candidates are required to give their answers in their own words as far as practicable

a) Database b) Knowledge c) Queries d) Transactions

(vii) Identify the type of learning in which labelled training data is used.

- a) Semi-supervised Learning
- b) Supervised Learning
- c) Reinforcement Learning
- d) Unsupervised Learning

(viii) Modeling ordered values comes under

- a) Classification
- b) Prediction
- c) Supervised
- d) Reinforcement

(ix) Euclidean Distance can only be used if attributes are

- a) Continuous
- b) Finite
- c) Discrete
- d) Infinite

(x) Learning by Observation is also known as

- a) Classification
- b) Clustering
- c) Regression
- d) Prediction

(xi) Simple Matching Coefficient is used when binary variable is

- a) Asymmetric
- b) Symmetric
- c) in-variant
- d) non-invariant

(xii) k-medoids comes under which method

- a) Partitioning
- b) Hierarchical
- c) Temporal
- d) Pattern

(xiii) Based on the precision of the periodicity, a pattern can be divided into

- a) 2 types
- b) 3 types
- c) 4 types
- d) 5 types

(xiv) Long-Memory Time-Series Modeling is a method for

- a) Time-Series
- b) Periodicity Analysis
- c) Trend Analysis
- d) Similarity Search

(xv) Under similarity search, dimensionality reduction techniques include

- a) DAT
- b) DCT
- c) SVD
- d) SVM

(xvi) Full form of FAISS is

- a) Facebook AI Similarity Search
- b) Facebook Augmented Similarity Search
- c) Facebook AI Similarity Scheme
- d) Facebook AI Similarity Scan

(xvii) Similarity Search finds which type of data

- a) Exact
- b) Similar
- c) Mixed
- d) None of the above

(xviii) Full form of ARIMA is

- a) Auto-Regressive Integrated Moving Average
- b) Auto-Regressive Installed Moving Average

- c) Auto-Regressive Integrated Mining Average
- d) All-Regressive Integrated Moving Average

(xix) Which of the following is not the property of stream data?

- a) Vast Volume
- b) Static Nature
- c) Possibly Infinite
- d) Multidimensional Features

(xx) Which of the following are not the methods for stream data processing & systems?

- a) Reservoir Sampling
- b) Multi-resolution Models
- c) Sketches
- d) Static Algorithms

(xxi) Bucket width in Lossy Counting Algorithm is

- a) $\text{ceil}(1/\epsilon)$
- b) $\text{floor}(1/\epsilon)$
- c) $\text{ceil}(1*\epsilon)$
- d) $\text{floor}(1*\epsilon)$

(xxii) In lossy counting algorithm, frequency list is stored in

- a) Secondary Memory
- b) Cache Memory
- c) Primary Memory
- d) All of the above

(xxiii) Hoeffding tree algorithm runs in

- a) Linear Time
- b) Exponential Time
- c) Sublinear Time
- d) Can't say about time

(xxiv) What is the full form of VFDT?

- a) Very Frequent Decision Tree
- b) Volatile Frequent Decision Tree
- c) Very Fast Decision Tree
- d) Very Far Decision Tree

(xxv) The Polysemy Problem is very common in

- a) Frequent Pattern Mining
- b) Data Warehousing
- c) Graph Mining
- d) Web Mining

(xxvi) In social network analysis, the out-degree and in-degree both follows

- a) Heavy-tailed Distribution
- b) Light-tailed Distribution
- c) Middle-tailed Distribution
- d) None of the above

(xxvii) Full form of DOM is

- a) Data Object Model
- b) Document Offer Model
- c) Document Object Model
- d) Document Object Mining

(xxviii) In class imbalance problems the numbers of positive and negative class are

- a) Equal
- b) Not equal
- c) Depends on situations
- d) No relation

(xxix) Full form of FSG is

- a) Frequent Sub-graph Mining
- b) Fast Sub-graph Mining
- c) Fast Sub-graph Method
- d) Frequency Sub-graph Method

(xxx) Removal of noisy and irrelevant data is known as

- a) Data Cleaning
- b) Data Aggregation
- c) Extraction of Data
- d) Compression of Data

Group – B

(Multiple Choice Type Questions)

2. Choose the correct alternatives from the followings:

10 x 2 = 20

(i) A sequence α is frequent in sequence database S if

- a) $\text{Support}_S(\alpha) > \text{min_sup}$
- b) $\text{Support}_S(\alpha) = \text{min_sup}$
- c) $\text{Support}_S(\alpha) \geq \text{min_sup}$
- d) $\text{Support}_S(\alpha) \leq \text{min_sup}$

(ii) Which type of clustering technique is applied in BIRCH?

- a) Multi Order
- b) Single Order
- c) Single Phase
- d) Multi Phase

(iii) In trend analysis, Moving Average comes under which of the following characteristics

- a) Cyclic Movements
- b) Seasonal Variations
- c) Trend Movements
- d) Random Movements

(iv) In AGM algorithm, two size- k frequent graphs are joined only if they have

- a) same size- $(k-1)$ subgraph
- b) different size- $(k-1)$ subgraph
- c) same size- $(1-k)$ subgraph
- d) optimal size- $(k+1)$ subgraph

(v) Compared with DOM-based methods, the segments obtained by VIPS are

- a) more semantically spreaded.
- b) less semantically aggregated.
- c) more systemetically aggregated.
- d) more semantically aggregated.

(vi) gSpan algorithm is a

- a) Apriori-based Approach
- b) Pattern Growth Approach
- c) Classification
- d) Clustering

(vii) Which is the correct limitation of Hoeffding Tree Algorithm

- a) The algorithm cannot handle concept drift
- b) It takes exponential time
- c) A node can't be splitted more than once
- d) None of the above

(viii) The accuracy of ensemble approach is

- a) greater than single classifier
- b) lesser than single classifier
- c) greater than multiple classifier
- d) No relation

(ix) Drawback of using probabilistic hierarchical clustering is

- a) It gives detailed analysis than other clustering approach which may be problematic
- b) Various probability models are used
- c) It cannot handle the uncertainty of cluster hierarchies
- d) All of the above

(x) In K-Means the quality of cluster C_i can be measured by

- a) Sum of squared error
- b) Absolute squared error
- c) Squared error
- d) Any one of the above

Group – C

(Short Answer Type Questions)

3. Attempt any four from the followings:

4 x 5 = 20

- i) Explain Data Mining Architecture with proper diagram.
- ii) (a) Explain Agglomerative Clustering with diagram. (2 Marks)
(b) Explain Divisive Clustering with diagram. (3 Marks)
- iii) What are the major components of trend analysis?
- iv) Explain Lossy Counting Algorithm.
- v) What are the major characteristics of Social Network Analysis?
- vi) How to mine multimedia data present on the web?

Santanu koley

Signature of the Paper Setter/Moderator

Answer Key:

Group – A

- | | | | | | |
|-----------------|----------|-----------|------------|-----------|-----------|
| 1. (i) a | (ii) d | (iii) d | (iv) d | (v) d | (vi) b |
| (vii) b | (viii) b | (ix) a | (x) b | (xi) b | (xii) a |
| (xiii) a | (xiv) c | (xv) c | (xvi) a | (xvii) b | (xviii) a |
| (xix) b | (xx) d | (xxi) a | (xxii) c | (xxiii) c | (xxiv) c |
| (xxv) d | (xxvi) a | (xxvii) c | (xxviii) b | (xxix) a | (xxx) a |

Group – B

- | | | | | | |
|-----------------|----------|---------|--------|-------|--------|
| 2. (i) c | (ii) d | (iii) c | (iv) a | (v) d | (vi) b |
| (vii) a | (viii) a | (ix) c | (x) a | | |