Roll No……

Plot No. 2, Knowledge Park-III, Greater Noida (U.P.) –201306

POST GRADUATE DIPLOMA IN MANAGEMENT (2024-26)

MID TERM QUIZ EXAMINATION (TERM -V)

Subject Name: Data Mining for Decision Making **Time: 00.30 hrs**

Sub. Code: PGIT 51 **Max Marks: 20**

Note:

1. Writing anything except Roll Number on Quiz paper will be deemed as an act of indulging in unfair means and action shall be taken as per rules.

2. There is no negative marking for wrong answers.

3. Tick marks the correct answer.

Attempt all questions. All questions are compulsory. **40×0.5 = 20 Marks**

1. Which of the following processes is not a part of data mining?

A Knowledge extraction

B Data archaeology

C Data transformation

D Data exploration

Ans: **B**

2. **What is not a data mining task?**

A Classification
B Linear Programming
C Clustering
D Regression

Ans: **B**

3. **Which is a classification algorithm?**

A K-means
B DBSCAN
C Apriori
D Decision tree

Ans: **D**

4. Which is not a type of regression analysis?

A Naive Bayes regression
B Polynomial regression
C Linear regression
D Logistic regression

Ans: **A**

5. **Which is not a classification algorithm?**

A Decision tree
B Linear regression
C K-nearest neighbors
D Naive Bayes

Ans: **B**

6. **Which of the following helps to decide a model from labeled data?**

A Reinforcement learning
B Hybrid learning
C Unsupervised learning
D Supervised learning

Ans: **D**

7. Which is true for Classification?

A A measure of the accuracy
B Merging of subsets
C The task of assigning a classification
D subdivision of a set

Ans: **D**

8. **The telecommunication companies desire to segment their clients into distinct groups in order to send suitable and**

related subscription offer. This can be considered as an example of which of the following methods?

A Supervised learning
B Data extraction
C Regression
D Unsupervised learning

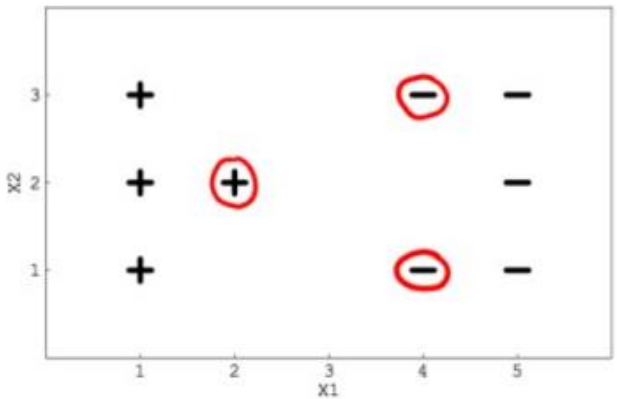
Ans: **D**

9. You are given data about seismic activity in India, and you want to predict the magnitude of the upcoming earthquake. Which method will you go with?

A Supervised learning
B Unsupervised learning
C Dimensionality reduction
D Clustering

Ans: **D**

10. In SVM classifier with 2 class classification problem, you have been given the following data in which some points are circled red that represent support vectors. If you remove the following, anyone red points from the data. Will the decision boundary change?



A Yes
B No

Ans:**B**

11. Which is real-world application of the SVM?

A Image Classification
B Clustering of Research Articles
C Text and Hypertext Categories
D All the above

Ans: **D**

12. **What does KDD stand for in data mining process?**

A Knowledge data house
B Knowledge discovery data
C Knowledge data definition
D Knowledge discovery database
Ans: D

13. What is the sequence of events in data mining process?

A Exploration, Infrastructure, interpretation, analysis, and exploitation
B Exploration, Infrastructure, analysis, interpretation, and exploitation
C Exploration, Infrastructure, analysis, exploitation, and analysis
D Infrastructure, exploration, analysis, interpretation, and exploitation
Ans: D

14. which is an application of data mining?

A Fraud Detection
B Risk Management & Corporate Analysis
C Market Management and Analysis
D All the above
Ans: **D**

15. In a medical application domain, suppose we build a classifier for patient screening (True means patient has cancer). Suppose that the confusion matrix is from testing the classifier on some test data. (FN= False negative, TN= True negative, TP= True Positive, TN =True Negative)

A FP >> FN
B FN >> FP
C FN = FP × TP
D All the above
Ans: **A**

16. What is a common application of clustering algorithms in unsupervised learning?

A Customer segmentation
B Text classification
C Image recognition
D Speech recognition
Ans: **A**

17. **Choose the unsupervised learning algorithm from the following**.

A K-Means Clustering
B Decision Tree
C Support Vector Machine
D Random Forest
Ans: **A**

18. What is the purpose of feature scaling?

A To remove outliers from data
B To select the most important features
C To normalize feature values for better model performance
D To increase model complexity

Ans: External influences

19. K-Means algorithm is sensitive to

A Initial centroids
B Data scaling
COutliers
D all the above

Ans: **A**

20. What does the Silhouette score measure in clustering algorithms

A variance within clusters
B clustering quality
C dimensionality reduction
D accuracy of the model

Ans: **A**

21. **Which technique is commonly used for market segmentation?**

A PCA
B Regression
C Clustering
D Logistic regression

Ans: **C**

22. What is the value of K in K-Means?

A Maximum iterations
B Number of clusters
C Number of features
D None

Ans: **B**

23. An e-commerce website wants to suggest "Users who bought this also bought..." to its customers. Which technique is best suited for finding these association rules?

A K-means
B Apriori algorithm
C PCA
D Linear regression

Ans: **B**

24. A security system captures images of faces but does not have labeled data to identify individuals. Which technique can be used to group

these images based on facial features?

A Classification
B Clustering
C Regression
D Reinforcement learning
Ans: **B**

25. **When applying K-Means clustering, what method is used to determine the optimal number of clusters?**

A Confusion method
B Elbow method
C ROC Curve
D AUC
Ans: **B**

26. The correlation coefficient for two real-valued attributes tells

A The attributes are not linearly related.
B As the value of one attribute decreases the value of the second attribute increases.
C As the value of one attribute increases the value of the second attribute also increases.
D The attributes show a linear relationship
Ans: **C**

27. Which of the following are interesting measures for association rules?

A recall
B lift
C accuracy
D compactness
Ans: **B**

28. Frequency of occurrence of an itemset is called as _____

A Support
B Confidence
C Support Count
D Rules
Ans: **A**

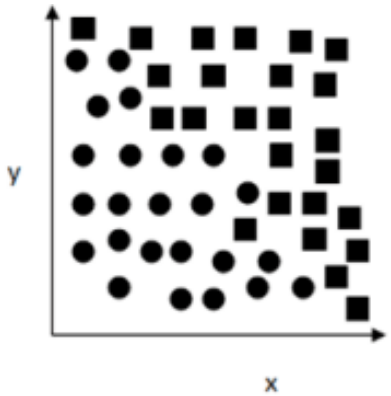
29. **Some fields where Apriori algorithm is used ____**

A In Education Field: Extracting association rules in data mining of admitted students through characteristics and specialties.
B In the Medical field: For example Analysis of the patient's database
C In Forestry
D All of the above
Ans: **D**

30. Classification accuracy is

A A subdivision of a set of examples into a number of classes
B Measure of the accuracy, of the classification of a concept that is given by a certain theory.
C The task of assigning a classification to a set of examples
D None of these
Ans: **C**

31. **Which of the following methods can be used to solve the below classification problem?**

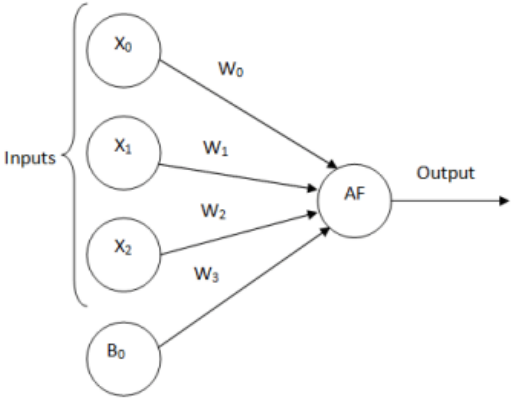


A Non linear logistic regression
B Linear regression
C Simple logistic regression
D Multivariate linear regression
Ans: **A**

32. **Which of the following statements is not true about neural networks?**

A Neural networks are a class of computationally inexpensive algorithms
B They are class of very powerful data mining classifiers
C These are capable of fitting almost any hypotheses
D NN has lots of interconnected nodes which are organized in layers
Ans: **A**

33. Which of the following statements is not true about the given figure?



A We can treat a node as an Activation Function, which can output a computation value based on activation function
B The terminology of parameters W in NN is called weights
C Bias unit is mandatory in neural networks
D The input values can be the input features
Ans: **C**

34. Which of the following statements is not true about the structure of a neural network?
A It consists of a set of nodes connected by links
B A set of thresholds or activation is used
C Not every link in a neural network is associated with a weight
D The processing elements in the neural network are known as neurons
Ans: **External influences**
35. The initialization and training of the weights on the interconnections through the training set require in the design of a neural network.

A True
B False
Ans: **A**
36. A marketing team uses a neural network to predict whether a customer will respond to a promotional email. After training, the model performs extremely well on training data but poorly on new customers.

A Underfitting
B High bias
C Overfitting
D Poor data visualization
Ans: **A**

37. A supply chain manager notices that increasing the number of hidden layers improves accuracy initially but later worsens performance What managerial decision should be taken?

A Increase learning rate indefinitely
B Add more training data or apply regularization
C Remove all hidden layers
D Ignore validation results
Ans: **B**

38. A subscription-based OTT platform wants to predict customer churn using viewing time, genre preferences, subscription tenure, payment delays, and device usage. The analytics team recommends a neural network over logistic regression. What is the strongest managerial justification for this choice?

A Logistic regression cannot handle categorical data
B Neural networks can model complex, non-linear customer behavior
C Neural networks require less data cleaning
D Logistic regression cannot perform classification
Ans: **B**

39. A financial services firm builds a neural network to classify loan applicants into approve or reject categories. Which activation function is most appropriate in the output layer?

A ReLU
B Sigmoid
C Tanh
D None of the above
Ans: **A**

40. A retail firm predicts whether a customer will purchase a product based on age group, income category, browsing history, and previous purchases using Naïve Bayes. What is the key assumption of the model?

A Attributes are linearly related
B Attributes are independent given the class
C Output variable is continuous
D Data must be normally distributed
Ans: **B**

Mapping of Questions with Course Learning Outcome

Question Number	COs	Bloom's taxonomy level	Marks Allocated
Q. 1-40	CO1, CO2, CO3, CO4	L3,L4,L5	0.5*40