



# BRAINWARE UNIVERSITY

Term End Examination 2022

Programme – M.Tech.(CSE)-AIML-2022

Course Name – Data Warehousing and Data Mining

Course Code - PCC-MCSM103

( Semester I )

Full Marks : 60

Time : 2:30 Hours

[The figure in the margin indicates full marks. Candidates are required to give their answers in their own words as far as practicable.]

## Group-A

(Multiple Choice Type Question)

1 x 15=15

1. Choose the correct alternative from the following :

(i) Express the following is a valid logical rule for the decision tree below

- |  |  |
|--|--|
| a) IF Business Appointment = No & Temp above 70 = No THEN Decision = wear slacks | b) IF Business Appointment = Yes & Temp above 70 = Yes THEN Decision = wear shorts |
| c) IF Temp above 70 = No THEN Decision = wear shorts                             | d) IF Business Appointment= No & Temp above 70 = No THEN Decision = wear jeans     |

(ii) A decision tree is pruned in order to:

- |  |                                       |
|--|---------------------------------------|
| a) improve classification accuracy on the training set | b) improve generalization performance |
| c) reduce the dimensionality of the data               | d) make the tree balanced             |

(iii) Bayes classifier also act as

- |                                   |                               |
|-----------------------------------|-------------------------------|
| a) Maximum aposteriori classifier | b) Maximum apriori classifier |
| c) Most likely classifier         | d) Maximum margin classifier  |

(iv) There are two binary attributes A and B. We know that the attributes are independent and probability  $P(A=1) = 0.7$ , and  $P(B=0) = 0.3$ . Calculate the probability that both A and B have values 1?

- |         |         |
|---------|---------|
| a) 0.11 | b) 0.33 |
| c) 0.40 | d) 0.49 |

(v) We want to classify instance x into one of two classes C1 and C2. Class labels of ten other training set instances sorted in increasing order of their distance to x are as follows: {C1, C2, C1, C2, C2, C2, C1, C2, C1, C2}. Calculate how will a K=5 nearest neighbor classifier classify x.

- |                        |                                       |
|------------------------|---------------------------------------|
| a) There will be a tie | b) C1                                 |
| c) C2                  | d) Not enough information to classify |

(vi) Classify the following statement is true for a multilayered perceptron

- |   |   |
|---|---|
| a) Output of all the nodes of a layer is input to | b) Output of all the nodes of a layer is input to |
|---|---|

- all the nodes of the next layer
- c) Output of all the nodes of a layer is input to all the nodes of the previous layer
- d) Output of all the nodes of a layer is input to all the nodes of the same layer
- d) Output of all the nodes of a layer is input to all the nodes of the output layer
- (vii) Classify the statement is true for the backpropagation learning algorithm
- a) It always converges to global minima
- b) Convergence is independent of the initial weight values
- c) It may converge to local minima
- d) Learning time decreases with an increase in the number of hidden layers
- (viii) Regression is concluded in:
- a) Predicting independent variables using dependent variables
- b) Predicting dependent variables using independent variables
- c) Predicting independence of independent variables
- d) Predicting dependent coefficients
- (ix) Suppose you are working on stock market prediction. You would like to predict whether the US Dollar will go up against the Euro tomorrow (i.e., whether a dollar will be worth more euros tomorrow than it is worth today). Compare this as a classification or a regression problem
- a) Regression
- b) Classification
- c) All of the above
- d) None of these
- (x) Express these is a reasonable definition of machine learning
- a) Machine learning means from labeled data.
- b) Machine learning is the science of programming computers.
- c) Machine learning is the field of allowing robots to act intelligently
- d) Machine learning is the field of study that gives computers the ability to learn without being explicitly programmed.
- (xi) A computer program is said to learn from experience E with respect to some task T and some performance measure P if its performance on T, as measured by P, improves with experience E. Suppose we feed a learning algorithm a lot of historical weather data, and have it learn to predict weather. In this setting, express E?
- a) The probability of it correctly predicting a future date's weather
- b) The weather prediction task.
- c) The process of the algorithm examining a large amount of historical weather data
- d) None of these.
- (xii) Data mining is the process of finding \_\_\_\_\_, novel, useful, and actionable, patterns in a large volume of data. Identify the following terms best fills the gap above.
- a) Voluminous
- b) Heterogeneous
- c) Valid
- d) Invalid
- (xiii) Recognize the following operations can be performed on nominal attributes.
- a) Distinctness
- b) Order
- c) Addition
- d) Multiplication
- (xiv) height of a person, can be recalled as an attribute of type.
- a) Nominal
- b) Ordinal
- c) Interval
- d) Ratio
- (xv) Internal nodes of a decision tree refer to:
- a) Attributes
- b) Classes
- c) Data instances
- d) None of the above

### Group-B

(Short Answer Type Questions)

3 x 5=15

2. Discuss the various process of the predictive data mining approach with proper examples. (3)

3. Develop the formulation of class conditional distribution from length vs count histogram. (3)
4. Compute the Apriori probability for Bayes Classifier. (3)
5. Analysis of the necessity of activation function and categorizing the various types of activation functions (3)

**OR**

- Distinguish between intra-cluster and inter-cluster distance with the help of the similarity and distance principle in the clustering approach. (3)
6. Justify the multivariate linear regression is more optimal with respect to single variant linear regression. (3)

**OR**

Propose a greedy approach to prepare a plan in the dimensionality reduction principle. (3)

### Group-C

(Long Answer Type Questions)

5 x 6=30

7. Observe the architecture of typical Data Mining System. (5)
8. Identify the confusion Matrix along with the formulation of various Cost-Sensitive Measures. (5)
9. Establish the Hessian Matrix to optimize the dual problem to identify good decision boundaries in support vector machine-based classifiers. (5)
10. Devise the neural network architecture and calculate the output of the net based on the following input and weight values.  $[X_1, X_2] = [0.2, 0.6]$  ,  $[W_1, W_2] = [0.3, 0.7]$  and  $b = 0.45$  (5)

**OR**

- Consider a Kohonen net with two clusters unit and five input units. The weight vectors for the cluster units are  $W_1 = (0.1, 0.9, 0.7, 0.3, 0.2)$  and  $W_2 = (0.6, 0.7, 0.5, 0.4, 1.0)$ . Use the square of Euclidian distance to find the winning cluster unit for the input pattern  $X = (0.0, 0.2, 0.1, 0.2, 0.0)$ . Using a learning rate of 0.2, Estimate the new weights for the winning unit. (5)
11. Illustrate the dendrogram in the Hierarchical clustering principle. (5)

**OR**

- Deduce the EPS and MinPts in the DBSCAN Algorithm with a proper example. (5)
12. Decide the utility of the dimensionality reduction Principle in classification approaches. (5)

**OR**

Compare the Principal Component Analysis Algorithm with Random Forest Dimensionality reduction algorithm. (5)

\*\*\*\*\*