BRAINWARE UNIVERSITY

Term End Examination 2024-2025
Programme – B.Tech.(CSE)-AIML-2021/B.Tech.(CSE)-DS-2021/B.Tech.(CSE)-AIML-2022/B.Tech.(CSE)-DS-2022
Course Name – Machine Learning for Real World Application
Course Code - PCC-CSM601/PCC-CSD601
(Semester VI)

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) Identify the mathematical concept fundamental for understanding optimization algorithms in machine learning.

a) Calculus   b) Linear Algebra
c) Probability Theory   d) Number Theory

(ii) Select the gradient of a function in machine learning.

a) The rate of change of the function   b) The area under the curve of the function
c) The value of the function at a specific point   d) The integral of the function

(iii) Name the matrix operation that is used to find the eigenvalues and eigenvectors of a square matrix.

a) Transposition   b) Inversion
c) Singular Value Decomposition   d) Diagonalization

(iv) Select the definition of standard deviation.

a) The difference between the highest and lowest values
b) The sum of all data values divided by the count
c) The measure of spread showing how much data varies from the mean
d) The middle value of a dataset

(v) Identify the key difference between supervised and unsupervised learning.

a) Supervised uses labelled data, while Unsupervised uses unlabelled data.
b) Supervised learns classification, while Unsupervised learns regression
c) Supervised is used for prediction, while Unsupervised is used for analysis.
d) Supervised uses Unlabelled data, while Unsupervised uses labelled data

(vi) Select the statement that is not true about supervised learning.

a) It requires labeled data for training.
b) It aims to learn a mapping from inputs to desired outputs.

c) It can be used for both classification and regression tasks

d) It can automatically discover hidden patterns in data.

(vii) Identify example of an unsupervised learning task.

a) Predicting house prices based on square footage and neighbourhood

b) Grouping customers into different segments based on their purchase history

c) Detecting fraudulent transactions in financial data

d) Identifying sentiment (positive, negative) from customer reviews

(viii) Identify out the primary goal of an agent in a reinforcement learning environment

a) To minimize the number of actions taken

b) To accurately predict the environment state.

c) To maximize the long-term expected rewards received

d) To learn the exact dynamics of the environment.

(ix) Predict the first phase of the CRISP-DM process

a) Data Preparation

b) Business Understanding

c) Modeling

d) Evaluation

(x) Identify the primary objective of EDA

a) Cleaning the dataset

b) Summarizing the dataset and finding patterns

c) Training the model

d) Evaluating model performance

(xi) Identify the following is NOT a common EDA technique

a) Scatter plots

b) Summary statistics

c) Data encryption

d) Histogram

(xii) Predict the EDA is primarily used to:

a) Train a machine learning model

b) Interpret model predictions

c) Understand the dataset before modeling

d) Optimize hyperparameters

(xiii) Relate R-Squared to regression model performance.

a) Measures goodness-of-fit in regression models

b) Always indicates better performance at higher values

c) Used only for classification problems

d) Reduces model complexity

(xiv) Identify the type of function used in logistic regression.

a) Sigmoid

b) Linear

c) Polynomial

d) Exponential

(xv) Select the scenario logistic regression is preferred over linear regression.

a) Predicting continuous values

b) Classifying email as spam or not spam

c) Finding correlation between two numeric variables

d) Predicting housing prices

Group-B
(Short Answer Type Questions)   3 x 5=15

2. Observe the mean , variance and standard deviation for the following data :   (3)
   2,4,5,6,8,17,23,25,27,29
3. Discuss features of unsupervised learning.   (3)
4. Determine the significance of feature scaling in Data Preparation.   (3)
5. Compare precision, recall, and F1-score in classification model evaluation.   (3)
6. Compare supervised and unsupervised learning in machine learning.   (3)

OR

Differentiate between classification and regression.   (3)

Group-C
(Long Answer Type Questions)   5 x 6=30

7. Discuss working methodology of supervised learning.   (5)

8. Explain the concepts of underfitting and overfitting in machine learning and compare their (5) causes and effects.
9. Evaluate the role of distance metrics in hierarchical clustering with examples from a real (5) dataset.
10. Explain the impact of data imbalances on model training and evaluation and also discover (5) the techniques that can be used during Data Preparation to address class imbalances.
11. Explain the key evaluation metrics used for logistic regression models. (5)
12. Illustrate the types of unsupervised learning. (5)

OR

(5)

Illustrate the seven major steps involved in the machine learning life cycle.

****