BRAINWARE UNIVERSITY

Term End Examination 2023-2024
Programme – B.Tech.(CSE)-AIML-2021/B.Tech.(CSE)-DS-2021
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) Name the mathematical concept which is fundamental for understanding the optimization algorithms used in machine learning.
   a) Calculus
   b) Linear Algebra
   c) Probability Theory
   d) Number Theory

(ii) In machine learning, identify the purpose of regularization?
   a) To increase the complexity of the model
   b) To reduce the variance of the model
   c) To increase overfitting
   d) To decrease the bias of the model

(iii) Predict the main advantage of using kernel methods in machine learning.
   a) They are computationally less expensive
   b) They are only applicable to linearly separable data
   c) They implicitly map data into higher-dimensional spaces
   d) They require less training data

(iv) Choose that is NOT a type of Machine Learning?
   a) Supervised learning
   b) Unsupervised learning
   c) Transferred learning
   d) Reinforcement learning

(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) Give 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

(vii) Choose the correct option In PCA (Principal Component Analysis), the number of input dimensions is equal to the principal components.

a) True   b) False

(viii) Identify which of the following statements is not true about boosting

a) It mainly increases the bias and the variance

b) It tries to generate complementary base-learners by training the next learner on the mistakes of the previous learners

c) It is a technique for solving two-class classification problems

d) It uses the mechanism of increasing the weights of misclassified data in preceding classifiers

(ix) Examine which of the following statements is true about stochastic gradient descent

a) It processes one training example per iteration

b) It processes all the training examples for each iteration of gradient descent

c) It is not preferred, if the number of training examples is large

d) It is computationally very expensive, if the number of training examples is large

(x) Select the correct statement:

a) Asymptotics are used for inference usually

b) caret includes several functions to pre-process the predictor data

c) The function dummyVars can be used to generate a complete set of dummy variables from one or more factors

d) All of the Mentioned

(xi) Select the class of the situation on "A feature F1 can take certain value: A, B, C, D, E, & F and represents grade of students from a college."

a) Feature F1 is an example of nominal variable.

b) Feature F1 is an example of ordinal variable.

c) It doesn't belong to any of the above category.

d) All of these

(xii) Select What hyperparameters can be tuned in decision trees to control model complexity?

a) Maximum depth   b) Minimum samples per leaf

c) Minimum samples split   d) All of the above

(xiii) Which hyperparameter helps prevent decision trees from overfitting by limiting the maximum number of features considered for splitting at each node?

a) Max features   b) Min samples split

c) Max depth   d) Min samples per leaf

(xiv) select Which is not numerical functions in the various function representation of Machine Learning?

a) Neural Network   b) Case-based

c) Linear Regression   d) Support Vector Machines

(xv) Any Machine learning methods does not observed more about _____. Identify which one of the following is matched.

a) Memorization   b) Analogy

c) Deduction   d) Introduction

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

2. Identify the potential benefits and risks associated with the widespread adoption of Machine Learning in various domains.   (3)

3. Define precision and recall.   (3)

4. Define mean,mode,median of the following data set? 32, 6, 21, 10, 8, 11, 12, 36, 17, 16, 15, 18, 40, 24, 21, 23, 24, 24, 29, 16, 32, 31, 10, 30, 35, 32, 18, 39, 12, 20,23,25,18,   (3)

5. Explain linear regression (3)
6. Justify the choice of distance metric affect the performance of the k-NN algorithm (3)

OR

justify can clustering algorithms be applied in real-life situations? (3)

Group-C
(Long Answer Type Questions)

5 x 6=30

7. Define different types of hypothesis with example. (5)
8. Explain the Sigmoid curve with an example. (5)
9. Establish that linear regression is suitable for any given data (5)
10. Select the cut-off with the variable (X or Y) for splitting the node of decision tree. (5)

X	Y	Result
1	1	0
1	2	1
2	1	1
2	2	0
2	3	1
3	3	1

11. Analyse the role of unsupervised algorithms in real-world applications. (5)
12. Evaluate the impact of regularization on logistic regression. (5)

OR

Justify the selection of ensemble methods for improving model performance. (5)

****