Seat No: _____ **Enrollment No: _____**

PARUL UNIVERSITY

FACULTY OF ENGINEERING & TECHNOLOGY

B.Tech. Winter 2023 - 24 Examination

Semester: 7 **Date: 04-11-2023**
Subject Code: 203105515 **Time: 10:30 am to 1:00 pm**
Subject Name: Machine Learning **Total Marks: 60**

Instructions:
1. All questions are compulsory.
2. Figures to the right indicate full marks.
3. Make suitable assumptions wherever necessary.
4. Start new question on new page.

Q.1	**Objective Type Questions** - (Fill in the blanks, one word answer, MCQ-not more than Five in case of MCQ) (All are compulsory) (Each of one mark)	**(15)**
	1. Which of the following Machine learning algorithm is based on the idea of bagging? a) Decision tree b) Random forest c) Classification d) Regression	**1**
	2. PAC stands for_____. a) Probably Approx cost b) Probably Approximate correct c) Probability Approx communication d) Probably Approximate compute	**1**
	3. Adaptive boosting, assigns _____ to data points that are misclassified by the previous weak learners.	**1**
	4. _____ are the simplest form of the neural network used for binary classification problem.	**1**
	5. K-mode clustering is designed for? a) Numerical data b) Text data c) Image data d) Categorical data	**1**
	6 One of the common form of normalization is _____ normalization, where data is transformed into mean of zero and standard deviation of one.	**1**
	7_____ is used to display Correlation matrix to show relationship between matrix.	**1**
	8. What does CART stands for? a) Classification and Regression Trees b) Cluster Analysis and Regression Trees c) Categorical Analysis and Regression Trees d) None of the above	**1**
	9. In _____ classification the class of the data-point is determined by a majority of vote of its nearest value.	**1**
	10. In supervised learning data-points in the training dataset is associated with _____ variable.	**1**
	11 The goodness of fit in linear regression is quantifies using _____of determination.	**1**
	12. Which of the following significance level is commonly used in Hypothesis testing? a) 0.1 b) 0.5 c) 0.05	**1**

	d) None of above	
	13. Logistic regression can be extended to handle _____ classification by using techniques like One-Vs-All or softmax regression.	**1**
	14. In Ensemble learning goal of Bagging is to reduce_____ by averaging the predictions of models.	**1**
	15. PCA is used for_____.	**1**
Q.2	**Answer the following questions**. (Attempt any three)	**(15)**
	A) Differentiate between Supervised Learning and Unsupervised Learning.	**5**
	B) Define Data-preprocessing. How to deal with missing values in Machine learning?	**5**
	C) KNN is a lazy learner algorithm. Justify and class value of sample 5 using KNN algorithm. Consider k= 3.	**5**
	D) Briefly explain Perceptron and Mention its limitation.	**5**
Q.3	A) What is kernel function in SVM? Why it is required? Explain in brief.	**(07)**
	B) Explain different Evaluation measures for ML algorithms. Calculate Precision, Recall and Accuracy from the following Confusion Matrix.	**(08)**
	OR	
	B) Explain the concept of Information Gain and Gini Impurity in Decision tree. How are these criteria used to make decisions when splitting nodes in decision tree?	**(08)**
Q.4	A) Explain the concept of Bagging and Boosting.	**(07)**
	OR	
	A) Define Normalization. Calculate Z score for following data: 15,18,20,25,30.	**(07)**
	B) Enlist types of Clustering algorithms. Explain K-means clustering and Hierarchical clustering in detail.	**(08)**

Sr.No	X1	X2	Y(class)
1	7	7	BAD
2	7	4	BAD
3	3	4	GOOD
4	1	4	GOOD
5	3	7	?

	True Positive	True Negative
Predicted Positive	60	15
Predicted Negative	10	15