

Roll No



**PRESIDENCY UNIVERSITY
BENGALURU**

**SCHOOL OF INFORMATION SCIENCE
END TERM EXAMINATION - JUN 2023**

Semester : Semester VI - 2020

Course Code : BCA215

Course Name : Sem VI - BCA215 - Machine Learning

Program : BCA

Date : 7-JUN-2023

Time : 1.00PM - 4.00PM

Max Marks : 100

Weightage : 50%

Instructions:

- (i) Read all questions carefully and answer accordingly.*
- (ii) Question paper consists of 3 parts.*
- (iii) Scientific and non-programmable calculator are permitted.*
- (iv) Do not write any information on the question paper other than Roll Number.*

PART A

ANSWER ALL THE QUESTIONS

(10 X 2 = 20M)

1. What is cost function?
(CO3) [Knowledge]
2. Define Gini Index
(CO2) [Knowledge]
3. What are the essential qualities of high quality clusters?
(CO4) [Knowledge]
4. What is the version space in machine learning, and how is it used for model selection?
(CO2) [Knowledge]
5. Mention the difference between bias and variance
(CO1) [Knowledge]
6. What is boosting?
(CO3) [Knowledge]
7. What is a dendrogram?
(CO4) [Knowledge]
8. What are the two most popular forms of collaborative filtering?
(CO4) [Knowledge]
9. What is hyper parameter tuning?
(CO3) [Knowledge]
10. Define Bayes Theorem
(CO3) [Knowledge]

PART B

ANSWER ALL THE QUESTIONS

(5 X 10 = 50M)

11. You are using the candidate elimination algorithm to build a model that predicts whether a fruit is an apple or an orange based on its color and shape. The training dataset contains examples of different fruits along with their corresponding color, shape, and the label indicating whether it is an apple or an orange.
- (CO2) [Comprehension]
12. A grocery store wants to understand the buying patterns of its customers and identify any interesting associations between different products. They have a transaction dataset that records the items purchased by each customer in their shopping cart. How would you apply association rule mining to uncover interesting associations between products based on the given transaction dataset?
- (CO4) [Comprehension]
13. A real estate agency wants to predict the sale prices of houses based on various features such as the area of the house, the number of bedrooms, and the distance to the nearest school. They have collected a dataset consisting of the sale prices and corresponding features of a set of houses in a particular neighborhood. Using linear regression, how would you approach this scenario to predict the sale prices of houses based on the given features?
- (CO1) [Comprehension]
14. A machine learning engineer is developing a model to predict housing prices based on various features such as the size of the house, number of bedrooms, location, and age. The engineer decides to use linear regression and wants to optimize the model's parameters using gradient descent. How would you apply the gradient descent algorithm to optimize the parameters of the linear regression model for predicting housing prices based on the given dataset?
- (CO3) [Comprehension]
15. A credit card company wants to predict whether a credit card applicant is likely to default on their payments or not. They have collected a dataset containing various attributes of applicants such as age, income, credit score, employment status, and previous payment history. The dataset also includes information on whether each applicant has defaulted or not. How would you apply the Random Forest algorithm to predict the likelihood of default for credit card applicants based on the given dataset?
- (CO3) [Comprehension]

PART C

ANSWER ALL THE QUESTIONS

(2 X 15 = 30M)

16. A credit card fraud detection system is being developed to identify fraudulent transactions and minimize financial losses for a banking institution. The system has collected a dataset of labeled credit card transactions, where each transaction is labeled as either fraudulent or legitimate. How would you apply ensemble learning algorithms to enhance the credit card fraud detection system based on the given dataset?
- (CO3) [Application]
17. Scenario: A retail company wants to segment its customer base to better understand their purchasing behavior and tailor marketing strategies accordingly. The company has collected a dataset containing various customer attributes such as age, gender, income, and purchase history. How would you apply two different types of clustering algorithms to segment the customer base of the retail company based on the given dataset?
- (CO4) [Application]