Printed Page:- 05 **Subject Code:- ACSML0401**

Roll. No:

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA
(An Autonomous Institute Affiliated to AKTU, Lucknow)
B.Tech
SEM: IV - THEORY EXAMINATION (2023 - 2024)
Subject: Machine Learning

Time: 3 Hours **Max. Marks: 100**

General Instructions:

IMP: *Verify that you have received the question paper with the correct course, code, branch etc.*

1. *This Question paper comprises of* ***three Sections -A, B, & C.*** *It consists of Multiple Choice Questions (MCQ's) & Subjective type questions.*

2. *Maximum marks for each question are indicated on right -hand side of each question.*

3. *Illustrate your answers with neat sketches wherever necessary.*

4. *Assume suitable data if necessary.*

5. *Preferably, write the answers in sequential order.*

6. *No sheet should be left blank. Any written material after a blank sheet will not be evaluated/checked.*

SECTION-A 20

1. Attempt all parts:-

1-a. What is bias and variance in the context of machine learning? (CO1) 1

(a) Bias is the error due to overfitting, and variance is the error due to underfitting

(b) Bias is the error due to overly simplistic models, and variance is the error due to overly complex models.

(c) Bias and variance are the same thing in machine learning.

(d) Bias is the error due to noisy data, and variance is the error due to outliers

1-b. What is the primary role of a confusion matrix in machine learning? (CO1) 1

(a) To confuse the model during training

(b) To measure the amount of noise in the data

(c) To evaluate the performance of a classification model

(d) To visualize the data distribution

1-c. What type of regression can be used to model relationships with multiple independent variables? (CO2) 1

(a) Linear Regression

(b) Polynomial Regression

(c) Multiple Linear Regression

(d) Logistic Regression

1-d. In the context of decision trees, what does "CART" stand for? (CO2) 1

(a) Classification and Regression Trees

(b) Categorization and Regression Trees

(c) Clustering and Rule Trees

(d) Classification and Rule Trees

1-e. K-Nearest Neighbor (K-NN) is primarily used for which type of machine learning task? (CO3) 1

(a) Clustering

(b) Regression

(c) Classification

(d) Dimensionality reduction

1-f. Which clustering algorithm forms clusters by connecting data points based on their relative distances and density? (CO3) 1

(a) K-Means clustering

(b) Hierarchical clustering

(c) DBSCAN

(d) K-Mode clustering

1-g. The Bayes Optimal Classifier makes decisions based on: (CO4) 1

(a) The prior probability of classes

(b) Random guessing

(c) The posterior probability of classes

(d) The majority class in the dataset

1-h. What is the primary difference between boosting and bagging in ensemble methods?(CO4) 1

(a) Boosting uses randomization, while bagging does not

(b) Bagging combines weak learners sequentially, while boosting does not

(c) Bagging combines weak learners independently, while boosting combines them sequentially

(d) Boosting always outperforms bagging

1-i. In Reinforcement Learning, what is the primary objective of the learning task? (CO5) 1

(a) Minimize the error between predicted and actual outputs

(b) Maximize the cumulative reward over time

(c) Discover hidden patterns in the data

(d) Optimize feature selection

1-j. Which of the following is part of the Q Learning function? (CO5) 1

(a) The environment's reward function

(b) The state transition probabilities

(c) The agent's memory size

(d) The gradient descent algorithm

2. Attempt all parts:-

2.a.	Discuss the need of data preprocessing. (CO1)	2
2.b.	Differentiate linear and logistic regression in terms of task accomplished by these algorithms. (CO2)	2
2.c.	Why K Nearest neighbor algorithm is called lazy learning algorithm? (CO3)	2
2.d.	What is the fundamental concept behind Bayesian Learning? (CO4)	2
2.e.	Describe the primary objective of a learning task in Reinforcement Learning. (CO5)	2

SECTION-B 30

3. Answer any five of the following:-

3-a.	Summarize the history of machine learning, highlighting significant milestones and developments. (CO1)	6
3-b.	Describe the process of model building in machine learning, including data preprocessing and evaluation. (CO1)	6
3-c.	Explain the main differences between classification and regression problems in machine learning. Provide examples to illustrate each. (CO2)	6
3-d.	Discuss the concept of overfitting in regression models. How can overfitting be recognized, and what strategies can be employed to prevent it? (CO2)	6
3.e.	Discuss the role of clustering in unsupervised machine learning. (CO3)	6
3.f.	Explore the strengths and weaknesses of the Naïve Bayes Classifier. (CO4)	6
3.g.	Discuss the concept of the Q Learning function in Reinforcement Learning (CO5)	6

SECTION-C 50

4. Answer any one of the following:-

4-a.	Outline the essential steps in designing a learning system, from data collection to model deployment. Highlight key considerations. (CO1)	10
4-b.	Apply find S algorithm to convert the most specific hypothesis into most general hypothesis on given dataset. (CO1)	10

Example	*Sky*	*AirTemp*	*Humidity*	*Wind*	*Water*	*Forecast*	*EnjoySport*
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

5. Answer any <u>one</u> of the following:-

5-a. Describe logistic regression model along with its advantages and disadvantages. (CO2) 10

5-b. Explain the concept of "support vectors" in the context of Support Vector Machines, and their role in creating decision boundaries. (CO2) 10

6. Answer any <u>one</u> of the following:-

6-a. What is K-Nearest Neighbor (K-NN) clustering, and when is it useful? (CO3) 10

6-b. For a given dataset apply CART algorithm to find optimal decision tree. (CO3) 10

Outlook	Temp	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

7. Answer any <u>one</u> of the following:-

7-a.	What are Bagging and boosting, and how do they enhance model performance? (CO4)	10
7-b.	Compare traditional decision trees and C5.0 boosting in terms of accuracy, interpretability, and computational demands. (CO4)	10

8. Answer any <u>one</u> of the following:-

8-a.	Explain various real-world examples of Reinforcement Learning. (CO5)	10
8-b.	What is the value function in Reinforcement Learning, and how does it impact decision-making? (CO5)	10