Printed Page:- Subject Code:- AMICA0611

Roll. No:

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA

(An Autonomous Institute Affiliated to AKTU, Lucknow)

MCA (Integrated)

SEM: VI - THEORY EXAMINATION (20..... - 20.....)

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. In machine learning, what does "overfitting" refer to: (CO1, K2) 1

(a) Creating a model that perfectly fits the training data but generalizes poorly to new data.

(b) Creating a model that underestimates the data distribution.

(c) Creating a model that has low variance.

(d) Creating a model that is too simple to capture the data patterns.

1-b. In the context of machine learning, what does "feature selection" refer to: (CO1, K2) 1

(a) Selecting the best machine learning algorithm for a problem

(b) Selecting the most relevant features from the dataset to improve model performance

(c) Removing all features from the dataset

(d) Adding more features to the dataset

1-c. In machine learning, what task involves assigning an input data point to a category: (CO2, K2) 1

(a) Classification

(b) Regression

(c) Clustering

(d) Association

1-d. The primary goal of linear regression is: (CO2, K2) 1

(a) Minimize the sum of squared errors
(b) Maximize the number of features
(c) Maximize the complexity of the model
(d) All

1-e. Partitional clustering algorithms, like K-means, aim to: (CO3, K2) 1

(a) Form hierarchical clusters
(b) Partition data into subsets without a predefined number of clusters
(c) Merge existing clusters into larger ones
(d) Discover clusters of varying shapes

1-f. In the context of clustering, what does the acronym "DBSCAN" stand for: (CO3, K1) 1

(a) Density-Based Spatial Clustering of Applications with Noise
(b) Data-Based Segmentation and Clustering for Analysis
(c) Distance-Based Similarity Clustering and Numbering
(d) Diverse Biologically Structured Clustering and Annotation Network

1-g. C5.0 boosting is a technique used in: (CO4, K2) 1

(a) Decision tree modeling
(b) Linear regression
(c) Neural networks
(d) Clustering

1-h. Gradient Boosting Machines (GBM) aim to: (CO4, K2) 1

(a) Combine weak learners into a strong learner
(b) Randomize the feature selection process
(c) Build multiple deep decision trees
(d) Use a single decision tree for classification

1-i. The common Machine Learning task in marketing is: (CO5, K2) 1

(a) Medical image classification
(b) Stock price prediction
(c) Loan approval
(d) Customer segmentation

1-j. In financial data, overfitting is especially dangerous because: (CO5, K2) 1

(a) It reduces model complexity
(b) It results in high training error
(c) It doesn’t generalize to future data
(d) It improves predictions

2. Attempt all parts:-

2.a. Name two types of machine learning approaches. (CO1, K2) 2

2.b. Differentiate between linear and logistic regression in terms of task accomplished 2

.

by these algorithms. (CO2, K2)

2.c.	In what situations is hierarchical clustering preferred over partitional clustering? (CO3, K3)	2
2.d.	Describe the advantages and disadvantages of deleting rows with missing values. (CO4, K2)	2
2.e.	Explain the type of machine learning technique is used in anomaly detection for fraud. (CO5, K3)	2

SECTION-B 30

3. Answer any five of the following:-

3-a.	Describe the concepts of bias and variance in machine learning and strategies for balancing them. (CO1, K2)	6
3-b.	Differentiate between categorical and numerical features with examples. How should these types of features be handled during the data preprocessing stage in machine learning? (CO1, K2)	6
3-c.	Explain the main differences between classification and regression problems in machine learning. Provide examples to illustrate each. (CO2, K2)	6
3-d.	Describe the idea behind polynomial regression and when it is advantageous to use higher-degree polynomial terms in your model. Provide an example. (CO2, K2)	6
3.e.	Discuss the K-Nearest Neighbor (K-NN) algorithm in detail. (CO3, K2)	6
3.f.	Compare and contrast bagging, boosting, and stacking in terms of approach, advantages, and use cases. (CO4, K2)	6
3.g.	Explain how recommendation engines work and their significance in e-commerce. (CO5, K3)	6

SECTION-C 50

4. Answer any one of the following:-

4-a.	Explain the fundamental concepts of learning in machine learning, emphasizing differences between supervised and unsupervised learning. (CO1, K2)	10
4-b.	Explain the process of building machine learning models, from data preprocessing to model evaluation. Highlight the importance of each step. (CO1, K3)	10

5. Answer any one of the following:-

5-a.	Explain the process by which decision tree algorithms, such as ID3, construct decision trees from training data. (CO2, K3)	10
5-b.	Define the core concept of a support vector machine (SVM) and its application in binary classification problems. Offer a practical illustration. (CO2, K3)	10

6. Answer any one of the following:-

6-a.	Explain the working of Principal Component Analysis (PCA). How does PCA reduce dimensionality, and what are its limitations? Illustrate your answer with an example. (CO3, K3)	10
6-b.	Discuss hierarchical clustering and its types with the help of real-world example.	10

(CO3, K3)

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

7-a.	Explain the bias-variance tradeoff with simple examples of high bias and high variance models. (CO4, K2)	10
7-b.	Design a strategy for feature engineering in a predictive model for loan approval. Include feature creation, encoding categorical data, interaction features, and handling skewed data. Justify your choices. (CO4, K3)	10

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

8-a.	Explain various real-world examples of Reinforcement Learning. (CO5, K3)	10
8-b.	Describe how clustering helps in identifying communities within a social network. (CO5, K3)	10

.