

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA
(An Autonomous Institute Affiliated to AKTU, Lucknow)

B.Tech

SEM: V - THEORY EXAMINATION (2025 - 2026)

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. Which of the following is NOT an approach to machine learning? (CO1, K2) 1
- (a) Descriptive learning
- (b) Supervised learning
- (c) Unsupervised learning
- (d) Reinforcement learning
- 1-b. 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-c. Which decision tree algorithm uses information gain as the splitting criterion?(CO2,K2) 1
- (a) C4.5
- (b) CART
- (c) ID3
- (d) Perceptron
- 1-d. What machine learning algorithm is known for its effectiveness in both classification and regression tasks with complex decision boundaries?(CO2, K2) 1
- (a) Linear Regression
- (b) Logistic Regression
- (c) Perceptron

- (d) Support Vector Machine (SVM)
- 1-e. How does the K-Mode clustering algorithm differ from K-Means clustering?(CO3,K2) 1
- (a) K-Mode uses the Euclidean distance metric, while K-Means uses a different metric.
- (b) K-Mode is primarily used for text data, while K-Means is used for numerical data.
- (c) K-Mode is non-hierarchical, while K-Means is a hierarchical clustering algorithm.
- (d) K-Mode is sensitive to outliers, while K-Means is not.
- 1-f. Which clustering algorithm is known for forming clusters of arbitrary shapes and is sensitive to noise?(CO3, K2) 1
- (a) K-Means clustering
- (b) K-Mode clustering
- (c) DBSCAN
- (d) Expectation Maximization (EM)
- 1-g. In Bayesian Belief Networks, what are nodes used to represent?(CO4 , K2)) 1
- (a) Probabilistic dependencies between variables
- (b) Data points in a dataset
- (c) Decision boundaries
- (d) Clustering results
- 1-h. What does the "X" in XGBoost stand for?(CO4,K1) 1
- (a) Extreme
- (b) Extra
- (c) Ensemble
- (d) Exponential
- 1-i. What is the Q Learning algorithm primarily used for?(CO5,K2) 1
- (a) Unsupervised learning
- (b) Reinforcement Learning, particularly for solving problems with discrete action spaces
- (c) Feature engineering
- (d) Text classification
- 1-j. In Reinforcement Learning, what is the primary application of Q Learning?(CO5,K2) 1
- (a) Image recognition
- (b) Game-playing agents that learn optimal strategies
- (c) Natural language processing
- (d) Stock market prediction

2. Attempt all parts:-

- 2.a. List three primary learning categories in ML(CO1,K2) 2
- 2.b. State the purpose of linear regression.(CO2, K2) 2
- 2.c. Provide a short definition of K-Nearest Neighbor.(CO3,K2) 2
- 2.d. Give a short definition of a Bayesian Belief Network.(CO4,K2) 2
- 2.e. List two application areas of RL.(C05,K2) 2

SECTION-B

30

3. Attempt all parts:-

3.a. Answer any one of the following:-

3.a.(i) Write the Find-S Algorithm and to find the Maximally Specific Hypothesis for the given data set.(CO1,K3) 6

Example	Sky	Air Temp	Humidity	Wind	Water	Forecast	Enjoy The Sport
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

3.a.(ii) State difference between: (CO1, K3) 6

1. Underfitting and Overfitting
2. Bias and Variance

3.b. Answer any one of the following:-

3.b.(i) In what manner do classification and regression differ from each other?(CO2,K2) 6

3.b.(ii) Explain the characteristics of the decision boundary generated by a Support Vector Machine (SVM). (CO2,K3) 6

3.c. Answer any one of the following:-

3.c.(i) Describe the concept of density reachability in density-based clustering.(CO3,K2) 6

3.c.(ii) Explain the role and significance of modes in K-Mode clustering.(CO3,K3) 6

3.d. Answer any one of the following:-

3.d.(i) Discuss the role of priors in Bayesian Learning and their impact on modeling uncertainty (CO4,K2) 6

3.d.(ii) What advantages does Random Forest offer for handling overfitting, and can you provide an exemplary use case?(CO4,K2) 6

3.e. Answer any one of the following:-

3.e.(i) Define reinforcement learning and its essential components (CO5,K2) 6

3.e.(ii) Describe the interaction between agent and environment in RL.(CO5,K2) 6

SECTION-C

50

4. Answer any one of the following:-

4-a. Discuss the process of model building in data science, highlighting the key steps, challenges, and considerations involved in developing an effective predictive model.(CO1, K3) 10

4-b. Discuss the role played by performance measures in evaluating and guiding learning.(CO1,K3) 10

5. Answer any one of the following:-

5-a. Build a decision tree using ID3 algorithm for the given training data in the table (Buy Computer data), and predict the class of the following new example: age<=30, income=medium, student=yes, credit-rating=fair (CO2,K3) 10

age	income	student	Credit rating	Buys computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31...40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31...40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31...40	medium	no	excellent	yes
31...40	high	yes	fair	yes
>40	medium	no	excellent	no

- 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. Describe the concept of Gaussian Mixture Models and their role in clustering analysis.(CO3,K3) 10
- 6-b. Use density-based clustering techniques to identify and detect outliers in a dataset.(CO3, K3) 10
7. Answer any one of the following:-
- 7-a. Describe the difference between bagging and boosting.(CO4,K2) 10
- 7-b. Explain the mechanisms by which Random Forest mitigates overfitting, along with a real-world use case.(CO4,K4) 10
8. Answer any one of the following:-
- 8-a. Differentiate reinforcement learning from supervised learning in terms of goal-oriented behavior.(CO5,K2) 10
- 8-b. Develop a Q-Learning strategy for personalized e-commerce offers.(CO5,K2) 10