

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

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

BCA

SEM: IV - 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:-

- | | | |
|-------|---|---|
| (1.1) | High variance leads to what? (CO1, K3) | 1 |
| | (a) Overfitting | |
| | (b) Underfitting | |
| | (c) Accuracy | |
| | (d) Stability | |
| (1.2) | Candidate elimination uses what? (CO1, K3) | 1 |
| | (a) S and G boundaries | |
| | (b) Only S boundary | |
| | (c) Only G boundary | |
| | (d) Only data | |
| (1.3) | Regression predicts (CO2, K1) | 1 |
| | (a) Category | |
| | (b) Label | |
| | (c) Continuous value | |
| | (d) Class | |
| (1.4) | C4.5 improves (CO2, K2) | 1 |
| | (a) ID3 | |
| | (b) CART | |
| | (c) SVM | |
| | (d) KNN | |
| (1.5) | Which distance is commonly used in K-means? (CO3, K1) | 1 |

- (a) Hamming
 - (b) Euclidean
 - (c) Manhattan
 - (d) Cosine
- (1.6) t-SNE is mostly used for (CO3, K1) 1
- (a) Visualization
 - (b) Sorting
 - (c) Encryption
 - (d) Prediction
- (1.7) Which of the following best describes the benefit of conditional independence in probability models? (CO4, K2) 1
- (a) It increases the amount of data required for computation.
 - (b) It reduces computation complexity.
 - (c) It makes all variables dependent on each other.
 - (d) It eliminates the need for probability calculations.
- (1.8) XGBoost is known for which of the following features? (CO4, K2) 1
- (a) Absence of regularization
 - (b) Support only for unsupervised learning
 - (c) Inclusion of regularization to reduce overfitting
 - (d) No optimization
- (1.9) Through which components an agent in reinforcement learning improves its behavior? (CO5, K1) 1
- (a) Rewards and penalties
 - (b) Random guessing only
 - (c) Fixed programming without feedback
 - (d) Supervised labels only
- (1.10) Exploitation in machine learning refers to which of the following? (CO5, K2) 1
- (a) Using known best actions based on current knowledge
 - (b) Using only random actions
 - (c) Ignoring previously learned information
 - (d) Exploring unknown actions only

2. Attempt all parts:-

- (2.1) Define sensitivity analysis. (CO1, K1) 2
- (2.2) State difference between classification and regression. (CO2, K2) 2
- (2.3) State the role of distance metrics in KNN. (CO3, K1) 2
- (2.4) State independence assumption in Naïve Bayes. (CO4, K1) 2
- (2.5) Define environment in RL. (CO5, K1) 2

SECTION-B

30

Answer any one of the following:-

- (3.1) Compare underfitting and overfitting. (CO1, K2) 6

(3.2)	Explain training and testing process in ML. (CO1, K2)	6
Answer any one of the following:-		
(3.3)	Demonstrate polynomial regression with graphical interpretation. (CO2, K3)	6
(3.4)	Explain association rule mining with support and confidence. (CO2, K2)	6
Answer any one of the following:-		
(3.5)	Discuss challenges and solutions for handling continuous values in K-means clustering. (CO3, K3)	6
(3.6)	Justify the importance of dimensionality reduction before applying visualization techniques. (CO3, K5)	6
Answer any one of the following:-		
(3.7)	Describe ensemble learning approach with advantages. (CO4, K2)	6
(3.8)	Explain probabilistic reasoning in Bayesian Learning. (CO4, K2)	6
Answer any one of the following:-		
(3.9)	Explain learning rate and discount factor. (CO5, K2)	6
(3.10)	Describe policy and value function. (CO5, K2)	6
SECTION-C		50
4. Answer any <u>one</u> of the following:-		
(4.1)	Explain Version Space representation and boundaries with diagram. (CO1, K2)	10
(4.2)	Explain hypothesis generalization and specialization process in concept learning. (CO1, K2)	10
5. Answer any <u>one</u> of the following:-		
(5.1)	Describe logistic regression model with decision boundary explanation. (CO2, K4)	10
(5.2)	Describe CART algorithm with Gini index and pruning techniques. (CO2, K3)	10
6. Answer any <u>one</u> of the following:-		
(6.1)	How does the silhouette score help in evaluating clustering performance? Explain with an example. (CO3, K4)	10
(6.2)	Explain Expectation Maximization algorithm with detailed steps and its role in probabilistic clustering. (CO3, K2)	10
7. Answer any <u>one</u> of the following:-		
(7.1)	Describe bagging algorithm with advantages and limitations. (CO4, K2)	10
(7.2)	Explain XGBoost algorithm with improvements over traditional boosting. (CO4, K2)	10
8. Answer any <u>one</u> of the following:-		
(8.1)	Describe learning task in RL with components and formulation. (CO5, K2)	10
(8.2)	Explain RL application in smart cities case study. (CO5, K3)	10