Printed Page:-

Subject Code:- ACSBS0513

Roll. No:

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA

(An Autonomous Institute Affiliated to AKTU, Lucknow)

B.Tech.

SEM: V - THEORY EXAMINATION (2022 - 2023)

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 the application of machine learning methods to a large database called? (CO1) 1

(a) Big data computing

(b) Internet of Things

(c) Data mining

(d) Artificial Intelligence

1-b Identify the type of learning in which labeled training data is used. (CO1) 1

(a) Semi Supervised learning

(b) Supervised Learning

(c) Reinforcement Learning

(d) Unsupervised Learnng

1-c Which Regression technique uses F-test or T-test? (CO2) 1

(a) Ridge Regression

(b) Linear regression

(c) Stepwise Regression

(d) None of them

1 d In terms of bias and variance. Which of the following is true when you fit degree 2 polynomial? (CO2) 1

(a) Bias will be high, variance will be high

(b) Bias will be low, variance will be high

(c) Bias will be high, variance will be low

(d) Bias will be low, variance will be low

1 c Sentiment Analysis is an example of: 1

1. Regression
2. Classification
3. Clustering
4. Reinforcement Learning (CO3)

(a) 1,2,4

(b) 1

(c) 1,2

(d) 1,2,3

1 b The final output of Hierarchical clustering is- (CO3) 1

(a) The number of cluster centroids

(b) The tree representing how close the data points are to each other

(c) A map defining the similar data points into individual groups

(d) All

1 a Naive Bayes requires? (CO4) 1

(a) Categorical Values

(b) Numerical Values

(c) Either a or b

(d) Both a and b

1 b Increase in Training time will tends to (CO4) 1

(a) Decreased of Size

(b) Constant Size

(c) Increased of Size

(d) None of them

1 c _____Reinforcement is defined as when an event, occurs due to a particular behavior. 1

(CO5)

(a) Negative

(b) Positive

(c) neutral

(d) None of these

1 ز The A Priori algorithm uses _____approach (CO5) 1

(a) Dynamic approach

(b) BFS

(c) DFS

(d) None of these

2. Attempt all parts:-

2.a. Define the term Hypothesis testing (CO1) 2

2.b. Describe Linear and Polynomial Kernal. (CO2) 2

2.c. Define term DBSCAN (CO3) 2

2.d. Is Naive Bayes is a discriminative classifier or generative classifier? (CO4) 2

2.e. Define the Bellman Equation (CO5) 2

SECTION B 30

3. Answer any five of the following:-

3.a How Do You Handle Missing or Corrupted Data in a Dataset? (CO1) 6

3.b What is 'training Set' and 'test Set' in a Machine Learning Model? How Much Data Will You Allocate for Your Training, Validation, and Test Sets? (CO1) 6

3.c What is Gini Index and how is it used in Decision Trees? (CO2) 6

3.d Explain the concept of CARTtrees (CO2) 6

3.e. Describe the concept of Gaussian Mixture Model with equations derivations in details? (CO3) 6

3.f. What are the differences between GBDTs and Random Forests? (CO4) 6

3.g. Compare the Model based learning and Value based learning (CO5) 6

SECTION C 50

4. Answer any one of the following:-

4.a Describe the Find-S Algorithm. Write the steps involve in Find-S algorithm. (CO1) 10

4.b Discuss the following : 1)Random Forest 2)Artifical Neural network (CO1) 10

5. Answer any one of the following:-

5.a Differentiate between Gradient Descent and Perceptron training rule (CO2) 10

5.b Discuss the concept of decision tree.Discuss the following terms: 1) Root node 2) Splitting 3)Branch or subtree 4)Leaf node 5)Pruning 6) Decision node (CO2) 10

6. Answer any one of the following:-

6.a Perform the DBSCAN on the below dataset with core =2 , and minimum point =3 (CO3) 10

	A	B	C	D	E	F
A	0					
B	0.7	0				
C	5.7	4.9	0			
D	3.6	2.9	2.9	0		
E	4.2	3.5	1.4	1	0	
F	3.2	2.5	2.5	0.5	1.1	0

6.b Describe K means Clustering Algorithm with the help of Example. (CO3) 10

7. Answer any one of the following:-

7.a Describe the Bayesian optimal classifier in detail with example. (CO4) 10

7.b How Bagging is different from boosting. (CO4) 10

8. Answer any one of the following:-

8.a Construct the FP growth algorithm for the below item set. Support threshold =50% and Confidence threshold =60% (CO5) 10

Transaction ID	Items
T1	{E,K,M,N,O,Y}
T2	{D,E,K,N,O,Y}
T3	{A,E,K,M}
T4	{C,K,M,U,Y}
T5	{C,E,I,K,O,O}

8.b Differentiate between Q learning and Markov Decision Process (CO5) 10