Printed Page:-04 Subject Code:- ACSE0515 /ACSEH0515

Roll. No:

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA

(An Autonomous Institute Affiliated to AKTU, Lucknow)

B.Tech

SEM: V - THEORY EXAMINATION (2024 - 2025)

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 Step is not involved in Data Cleaning (CO1, K1) 1

(a) Removal of unwanted observations

(b) Fixing Structural errors

(c) Binarization Data

(d) Handling missing data

1-b. What are some examples of data quality problems. (CO1, K1): 1

(a) Noise and outliers

(b) Duplicate data

(c) Missing values

(d) All of the Above

1-c. _____ calculated as the sum of absolute differences between the coordinates of data point and centroid of each class. (CO2, K1) 1

(a) Euclidean Distance

(b) Feature Scaling

(c) Manhattan Distance

(d) Minkowski Distance

1-d. Regression can be define as . (CO2, K1) 1

(a) It is a technique to predict values

(b) It is a technique to find outliers

(c) It is a technique to fix data

(d) It is a Machine Learning algorithm

1-e. Which of the following statements is FALSE about Ridge and Lasso Regression? (CO3, K1) 1

(a) These are types of regularization methods to solve the overfitting problem.

(b) Lasso Regression is a type of regularization method

(c) Ridge regression shrinks the coefficient to a lower value.

(d) Ridge regression lowers some coefficients to a zero value.

1-f. In a naive Bayes algorithm, when an attribute value in the testing record has no example in the training set, then the entire posterior probability will be zero. (CO3, K1) 1

(a) True

(b) FALSE

(c) Can't determined

(d) None of these

1-g. Reinforcement learning is- (CO4, K1) 1

(a) Unsupervised learning

(b) Supervised learning

(c) Award based learning

(d) None

1-h. Which of the following algorithms cannot be used for reducing the dimensionality of data? (CO4, K1) 1

(a) t-SNE

(b) PCA

(c) LDA False

(d) None of these

1-i. Which of the following is a mandatory data pre-processing step(s) for XGBOOST? i. Impute Missing Values
ii. Remove Outliers
iii. Convert data to numeric array / sparse matrix
iv.Input variable must have normal distribution
v.Select the sample of records for each tree/ estimators (CO5, K1) 1

(a) i and ii

(b) I, ii and iii

(c) iii, iv and v

(d) iii

1-j. For parameter tuning in a boosting algorithm, which of the following search strategies may give best tuned model: (CO5, K1) 1

(a) Random Search.

(b) Grid Search

(c) 1 or 2

(d) Can't say

2. Attempt all parts:-

2.a.	Explain how do you select important variables while working on a data set. (CO1, K1)	2
2.b.	What is the difference between overfitting and underfitting? (CO2, K4)	2
2.c.	Explain how the Random Forests give output for Classification, and Regression problems.(CO3, K2)	2
2.d.	Discuss the chalange of Q-Learning. (CO4, K2)	2
2.e.	Write Features of Recommender Systems. (CO5, K1)	2

SECTION-B 30

3. Answer any five of the following:-

3-a.	List the issues in Machine Learning. (CO1, K1)	6
3-b.	Explain how do you encode the catogorical data which consists of both ordinal and nominal data type.(CO1, K2)	6
3-c.	What's the difference between hard and soft clustering. (CO2, K4)	6
3-d.	Explain why can't we use the mean square error cost function used in linear regression for logistic regression. (CO2, K2)	6
3.e.	Compare Random forest and Decision Tree in detail. (CO3, K4)	6
3.f.	Explain Principal Component Analysis algorithm with its Benefits and limitations. (CO4, K2)	6
3.g.	Explain LightGBM and XGBoost parameters. (CO5, K2)	6

SECTION-C 50

4. Answer any one of the following:-

4-a.	Discuss the Three Stages of Building a Model in Machine Learning. (CO1, K2)	10
4-b.	Describe Semi-supervised machine learning with real world example. (CO1, K2)	10

5. Answer any one of the following:-

5-a.	Explain K-means clustering? How does K-means clustering work? (CO2, K2)	10
5-b.	Explain principle of Logistic Regression. State types of Logistic Regression.(CO2, K2)	10

6. Answer any one of the following:-

6-a.	Describe the Limitations, advantages and application of Naïve bayes.(CO3, K2)	10
6-b.	Discuss Regularization. What kind of problems does regularization solve? (CO3, K2)	10

7. Answer any one of the following:-

7-a.	Discuss some approaches or algorithms you know in to solve a problem in Reinforcement Learning. (CO4, K2)	10

7-b.	Explain the Q-function and Q-Learning Algorithm. (CO4, K2)	10
8. Answer any <u>one</u> of the following:-		
8-a.	Explain Model-Based Collaborative approach. (CO5, K2)	10
8-b.	Explain in detail how Knowledge-based Recommender Systems different from Collaborative and Content-based Recommender Systems. (CO5, K2)	10