.

Printed Page:-4 **Subject Code:- BCSAI0516**

Roll. No:

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA
(An Autonomous Institute Affiliated to AKTU, Lucknow)
B.Tech
SEM: V - THEORY EXAMINATION (2025 - 2026)
Subject: Predictive Analytics

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. If the slope of the regression equation y = bo + b1x is positive, then (CO1, K1) 1

(a) as x increases y decreases
(b) as x increases so does y
(c) Either a or b is correct
(d) as x decreases y increases

1-b. Least square method calculates the best-fitting line for the observed data by minimizing the sum of the squares of the _____ deviations. (CO1, K1) 1

(a) Vertical
(b) Horizontal
(c) Both of these
(d) None of these

1-c. In regression, the equation that describes how the response variable (y) is related to the explanatory variable (x) is: (CO2, K1) 1

(a) the correlation model
(b) the regression model
(c) used to compute the correlation coefficient
(d) None of these alternatives is correct.

1-d. _____ is rapidly being adopted for computing descriptive and query types of analytics on Big data. (CO2, K1) 1

(a) EDR
(b) Hadoop
(c) Azure

(d) InfoSight

1-e. You are predicting whether an email is spam or not. Based on the features, you obtained an estimated probability to be 0.75. What's the meaning of this estimated probability? The threshold to differ the classes is 0.5 (CO3, K1) 1

(a) The email is not spam
(b) The email is spam
(c) Can't determine
(d) both (A) and (B)

1-f. _____ are defined as the ratio of the probability of an event occurring to the probability of the event not occurring (CO3, K1) 1

(a) Simple
(b) Even
(c) Regex
(d) Odds

1-g. Which of the following step is performed by data scientist after acquiring the data (CO4, K1) 1

(a) Data Cleansing
(b) Data Integration
(c) Data Replication
(d) All of the mentioned

1-h. Adjacent observations in time series data (excluding white noise) are independent and identically distributed (IID). (CO4, K1) 1

(a) TRUE
(b) FALSE

1-i. Which of the following statements is true about the arithmetic mean of two regression coefficients? (CO5, K1) 1

(a) It is less than the correlation coefficient
(b) It is equal to the correlation coefficient
(c) It is greater than or equal to the correlation coefficient
(d) It is greater than the correlation coefficient

1-j. List the meaning of the testing of the hypothesis? (CO5, K1) 1

(a) It is a significant estimation of the problem
(b) It is a rule for acceptance or rejection of the hypothesis of the research problem
(c) It is a method of making a significant statement
(d) None of the above

2. Attempt all parts:-

2.a. state the curse of dimensionality. Can you give an example? (CO1, K1) 2

2.b. State the differnces between Ridge and elastic net. (CO2, K2) 2

2.c. Explain Sigmoid Function. (CO3, K2) 2

2.d. State some advantage of trend analysis(CO4, K1) 2

2.e. State various some common ways to gather domain knowledge? (CO5, K1) 2

SECTION-B 30

3. Attempt all parts:-

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

3.a.(i)	Explain how does predictive analysis work?(CO1, K1)	6
3.a.(ii)	Explain heteroscedastic approach with example (CO1, K2)	6

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

3.b.(i)	Illustrate K-fold Cross-Validation. Explain using diagram about 5 fold cross validation? (CO2, K2)	6
3.b.(ii)	What do you understand by One Hot Encoding? (CO2, K1)	6

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

3.c.(i)	Define which cost function should be used in Logistic Regression? (CO3, K1)	6
3.c.(ii)	Define logistic function? What is the range of values of a logistic function? (CO3, K1)	6

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

3.d.(i)	State how would you calculate the standard error of the regression? What does it measure? (CO4, K1)	6
3.d.(ii)	Explain Moving Average along with an example (CO4, K2)	6

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

3.e.(i)	Explain the following: a)Mean Absolute Difference (MAD) b)Dispersion Ratio c)Mutual Dependence d)Relief (CO5, K2)	6
3.e.(ii)	Discuss the key techniques used in EDA for numerical and categorical data. Provide examples. (CO5, K2)	6

SECTION-C 50

4. Answer any one of the following:-

4-a.	Explain the terms Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning. (CO1, K2)	10
4-b.	State Bias, Variance and what do you mean by Bias-Variance Tradeoff (CO1, K1)	10

5. Answer any one of the following:-

5-a.	Explain the advantage of Ridge Regression over linear Regression? Also define the Loss Function for Ridge along with an example. (CO2, K1)	10
5-b.	Explain how would you validate a model you created to generate a predictive model of a quantitative outcome variable using multiple regression? (CO2, K1)	10

6. Answer any one of the following:-

6-a.	Critically evaluate the use of ROC (Receiver Operating Characteristic) and AUC (Area Under Curve) for assessing the performance of Logistic Regression models in classification tasks. (CO3, K2)	10
6-b.	You are building a Logistic Regression model to predict whether a loan applicant will default. Describe how you would interpret the model's output, particularly the coefficients of the predictor variables. (CO3, K3)	10

7. Answer any one of the following:-

7-a.	Explain what is meant by seasonal fluctuations of a time series. A company	10

manufactures bicycles. Given the quarterly production figures of the company for the last 4 years, explain the procedure to compute seasonal indices by the 'link relatives' method. Use link- relatives method to compute seasonal indices from the recorded production figures given below: (CO4, K2)

YEAR	Q1	Q2	Q3	Q4
2016	420	414	502	365
2017	491	456	516	337
2018	496	365	487	310
2019	502	487	436	404

7-b. Explain the difference between simple moving average (SMA) and weighted moving average (WMA). When would you use each? (CO4, K1) 10

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

8-a. How can imbalanced datasets affect machine learning models?cexplain with suitable example(CO5, K2) 10

8-b. How do feature encoding and imputation impact the performance of predictive models? Illustrate with examples. (CO5, K2) 10