Printed Page:-04 Subject Code:- ACSAI0516

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: 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. State what does the following expression (H0: 1 = 0) mean? (CO1, K1) 1

(a) Mean-Y changes as as a result of a change in X

(b) Mean-Y does not change as a result of a change in X

(c) Mean-Y value becomes 0 as a result of a change in X

(d) Mean-Y value is equal to 0 when X=0

1-b. State whether ratio of MSR/MSE yields? (CO1, K1) 1

(a) the t statistic

(b) SST

(c) the F statistic

(d) the Z statistic

1-c. State whether regression modelling is a statistical framework for developing a mathematical equation that describes how? (CO2, K1) 1

(a) one explanatory and one or more response variables are related

(b) several explanatory and several response variables response are related

(c) one response and one or more explanatory variables are related

(d) All of these are correct.

1-d. State in least squares regression, which of the following is not a required assumption about the error term ε? (CO2, K1) 1

(a) The expected value of the error term is one.

(b) The variance of the error term is the same for all values of x.

(c) The values of the error term are independent.

(d) The error term is normally distributed.

1-e. State whether Logistic Regression is a Machine Learning algorithm that is used to predict the probability of a ___? (CO3, K1) 1

(a) categorical independent variable

.

(b) categorical dependent variable.

(c) numerical dependent variable

(d) numerical independent variable

1-f. State the the hypothesis of logistic regression? (CO3, K1) 1

(a) to limit the cost function between 0 and 1

(b) to limit the cost function between -1 and 1

(c) to limit the cost function between -infinity and +infinity

(d) to limit the cost function between 0 and +infinity

1-g. State which of the following can't be a component for a time series plot? (CO4, K1) 1

(a) Seasonality

(b) Trend

(c) Cyclical

(d) None of the above

1-h. State what Autocovariance measures _____ ? (CO4, K1) 1

(a) Linear dependence between multiple points on the different series observed at different times

(b) Quadratic dependence between two points on the same series observed at different times

(c) Linear dependence between two points on different series observed at same time

(d) Linear dependence between two points on the same series observed at different times

1-i. Identify the kind of learning algorithm for "facial identities for facial expressions"? (CO5,K1) 1

(a) Prediction

(b) Recognizing Patterns

(c) Recognizing anomalies

(d) Generating patterns

1-j. Identify which of the following is true for the coefficient of correlation? (CO5,K1) 1

(a) The coefficient of correlation is not dependent on the change of scale

(b) The coefficient of correlation is not dependent on the change of origin

.

(c) The coefficient of correlation is not dependent on both the change of scale and change of origin

(d) None of the above

2. Attempt all parts:-

2.a.	Define the least square method? State the two bias categories of least square problems and state how is sigma square calculated? (CO1, K1)	2
2.b.	Describe the coefficient of determination with an example? (CO2, K1)	2
2.c.	Explain the use of ROC curve and the AUC of a ROC Curve? (CO3, K2)	2
2.d.	State what is residual analysis and its types? (CO4, K1)	2
2.e.	Explain the difference between feature extraction and feature selection? When should each one be used? (CO5, K2)	2

SECTION-B 30

3. Answer any five of the following:-

3-a.	Explain what do you understand by likelihood profiles for 95% confidence intervals in detail? (CO1,K2)	6
3-b.	Obtain the equations of line of regression for the following data : ? (CO1, K2) X: 1 2 3 4 5 6 7 8 9 Y: 9 8 10 12 11 13 14 16 15 Also obtain an estimate of y for x = 6.2	6
3-c.	Describe what is Regularization and all it's types in detail with diagram? (CO2, K2)	6
3-d.	State wehether "Is it true that the L1 term in Lasso has the following purposes: performing feature selection, compensating for overfitting, and smoothing",Which regularization is used to reduce the over fit problem? (CO2, K1)	6
3.e.	Explain do you implement multinomial logistic regression? Suppose that you are trying to predict whether a consumer will recommend a particular brand of chocolate or not. Let us say your hypothesis function outputs h(x)=0.55 where h(x) is the probability that y=1 (or that a consumer recommends the chocolate) given any input x. Does this mean that the consumer will recommend the chocolate? (CO3, K2)	6
3.f.	Applying the equation of two regression lines obtained in a correlation analysis of 60 observations are : 5x=6y+24 and 1000y=768x-3608 ,What is the correlation coefficient, Show that the ratio of coefficient of variability of x to that of y is 5/24 , and what is the ration of variance of x and y? (CO4, K3)	6
3.g.	Explain all patterns emerging in time series data with an example? (CO5, K2)	6

SECTION-C 50

4. Answer any one of the following:-

4-a.	Find the correlation coefficient for the following data and intrepret the result? (CO1,K1) X: 23 27 28 28 29 30 31 33 35 36	10

Y: 18 20 22 27 21 29 27 29 28 29

4-b. Explain what do you understand by maximum likelihood criterian? Explain with an example? (CO1, K2) 10

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

5-a. Discuss how "BIC is some what similar to AIC but with slight change" also discuss how do we measure the quality of statistical model in detail? (CO2, K2) 10

5-b. Explain Null Hypothesis in terms of Whites test also define the following terms, with an example ? (CO2, K1) 10

1. Homoscedasticity
2. Independent Test

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

6-a. Explain how to interpret the results of a Logistic regression model, discuss some features of logistic regression with diagram and also state the meanings of alpha and beta in a logistic regression model with equations? (CO3, K1) 10

6-b. Explain the following with the help of an example? (CO3, K1) 10

1. Precision
2. Recall
3. F1 Score
4. Accuracy

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

7-a. Explain the equation of an AR(p), an MA(q), and an ARIMA(p, 1, q) model and explain context in detail? (CO4, K1) 10

7-b. Explain Time Series Analysis in Detail, also state applications of univariate time series in terms of forecasting? (CO4,K1) 10

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

8-a. Explain why do we need dimensionality reduction, what are its drawbacks? Explain the Curse of Dimensionality in detail? (CO5, K1) 10

8-b. Explain how does feature scaling impact the performance of PCA and LDA? (CO5, K1) 10