

23P363

(Pages: 2)

Name:

Reg. No:

THIRD SEMESTER M.Sc. DEGREE EXAMINATION, NOVEMBER 2024

(CBCSS-PG)

(Regular/Supplementary/Improvement)

CC22P MST3 E19 – STATISTICAL MACHINE LEARNING

(Statistics)

(2022 Admission onwards)

Time: Three Hours

Maximum: 30 Weightage

PART A

Answer any *four* questions. Each question carries 2 weightage.

1. What are the key features of Jupyter Notebook?
2. Describe different strategies to handle missing values in a dataset.
3. What is skewed data, and why is it problematic in machine learning. Describe methods to handle skewed data. How does handling skewed data impact model performance.
4. Distinguish parametric and non-parametric method of modelling.
5. Describe training set and test set.
6. Explain a decision tree.
7. How do you assess the goodness-of-fit of a Cox Proportional Hazards model?

(4 × 2 = 8 Weightage)

PART B

Answer any *four* questions. Each question carries 3 weightage.

8. Explain the application of codes remove and append used in python and illustrate with example.
9. Discuss the primary methods used for binning numeric variables in data analysis. Give an example.
10. Describe the process of EDA. Explain the steps involved in EDA and how they contribute to understanding the dataset.
11. Explain prediction and inference models followed in machine learning.
12. Explain what it means that adjusted R square is much less than R square.
13. Discuss KNN algorithm.
14. Write a short note on survival analysis and censoring time.

(4 × 3 = 12 Weightage)

PART C

Answer any *two* questions. Each question carries 5 weightage.

15. a) Elaborate on both graphical methods and numerical methods for identifying outliers.

b) You are given a dataset consisting of two features, X_1 and X_2 with different ranges.

Perform Min-Max Transformation and Z-score Standardization on these features.

X_1	10	15	8	12	20
X_2	500	750	600	700	900

16. Compare and contrast supervised and unsupervised learning in statistical learning.

Provide examples of each

17. Discuss simple regression model. Explain why an analyst may prefer a confidence interval to a hypothesis test.

18. Explain linear discriminant analysis for $p > 1$.

(2 × 5 = 10 Weightage)
