THIRD SEMESTER M.Sc. DEGREE EXAMINATION, NOVEMBER 2025

(CBCSS-PG)

(Regular/Supplementary/Improvement)

CC22PMST3E19 – 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. Explain the core Python libraries and their functionalities

2. Explain the data preparation phase and modelling phase of CRISP DM

3. Discuss the role of pie chart in exploring categorical variables

4. Describe various supervised learning methods.

5. What is extrapolation?

6. Define distance function.

7. Write a short note on survival analysis.

(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 different methods of handling missing values during data cleaning, and analyze their advantages and disadvantages.

10. Distinguish percentage bar chart and clustered bar chart with examples.

11. The exam scores of 15 students are given below: 45, 50, 54, 43, 55, 67, 54, 56, 46, 78, 100, 108, 78, 12, 69, 58. Use Box plot (IQR method) to identify if there are any outliers.

12. Explain the scope of flexible models of statistical learning. Justify whether   a flexible statistical learning method to be better or worse than an inflexible method when the sample size n is extremely large, and the number of predictors p is small.

13. When using indicator variables, explain the meaning and interpretation of the indicator variable coefficients, graphically and numerically.

14. Discuss KNN algorithm.

(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. a) Why assessing model accuracy is essential statistical learning methods?

b) Explain mean square error method for quality of fit.

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

18. a) What do you understand by confusion matrix?

b) Explain the terms True positive, True Negative, False Positive and False Negative

with the help of an example

(2 × 5 = 10 Weightage)

****