

Date:8-03-2022

Registration number:

ST. JOSEPH’S COLLEGE (AUTONOMOUS), BENGALURU-27

M.Sc. Statistics - III SEMESTER

SEMESTER EXAMINATION: OCTOBER 2021

(Examination conducted in March 2022)

**ST 9220 - Machine Learning Techniques**

Time- 2 ½ hrs Max Marks-70

This question paper contains two printed pages and two parts

**Part A**

 **Answer any SIX questions 3X6 = 18**

1. Define Feature. Define Feature Engineering.
2. Differentiate between Supervised learning and Unsupervised learning with an example.
3. Define Numpy and Scipy.
4. Name any 3 values that would be converted to “False” when using the “bool” function.
5. Briefly describe K fold cross validation.
6. What is interaction effect in Multiple Regression? Given an example.
7. What is “Wisdom of the crowd”? Give 2 examples for the same.
8. Write the mathematical construction of the Maximal Margin Classifier.

**Part B**

 **Answer any FOUR questions 13X 4 = 52**

1. a) Define Machine Learning according to Tom M Mitchell and define the components of the same (3)

b) Draw the box of learning system and explain the same. (5)

c) What are the rules one must follow while naming a variable in Python. (3)

d) Briefly describe any two data types in Python. (2)

1. a) Write a note on Bias Variance trade off with an example. (8)

b) Write a note on standardization of data. (2)

c) Write the code for standardization of a column. (3)

1. a) Mathematically model qualitative data using logistic regression. (8)

b) What is odds ratio? How do we interpret the same? (2.5)

c) Briefly describe the testing of hypothesis for the parameter of logistic regression. (2.5)

1. a) What is Curse of Dimensionality. Illustrate with an example. (3)

b) What are Shrinkage Methods. Briefly describe the three shrinkage methods. (7)

c) What is a Kernel. Define any 2 popular kernel functions. (3)

1. a) Find the best split among the following attributes from the data given below. (8)



b) Write the algorithm of Ada Boost. (5)

1. a) Give a brief overview of Gaussian Mixture Model and EM method for estimating the parameters for the same. (7)

b) What is the difference between Linear Regression and SVM Regression. (2)

c) Write the Algorithm for K means Clustering. (4)