

M.Sc. AIML & AIML (DS) Semester 2 Examination

Advance Python

Time : 3-00 Hours]

June-2024

[Max. Marks : 100

Instructions:

- Write both the Sections in the separate answer book.
- Both Sections having equal weightage.
- Draw Diagrams wherever necessary.
- Make Assumptions wherever necessary.

SECTION – I

- Q-1 Explain the following terms with an appropriate example. (10)
- Features and observations
 - numpy and array
 - True Positive and RElu
 - Classification and Regression
 - Recall and Precision

- Q-2 Attempt the following: (20)
- Give the importance of good quality data. Give the importance of data preprocessing.
 - What are the reasons of the missing values in data set? What are the different types of missing values?

OR

- Q-2 Attempt the following:
- Define Outliers. Why do they occur in data? What are side effects of outliers in data
 - Compute the following:
 - Give importance of splitting data. How is it done in sci-kit learn. What problem may arise if we randomly split a dataset that has 99% negative class and 1% positive class?
 - What is overfitting and underfitting
 - What is correlation? In which type of ML problem, we use correlation and why do we use it? How to calculate correlation values using any library of python?
- Q-3 Attempt the following: (20)
- Write the Fuzzy C means algorithm and apply on the data points step by step below and create a table finding the new membership values.

| X | Y | C1 | C2 |
|---|---|-----|-----|
| 1 | 3 | 0.8 | 0.2 |
| 2 | 5 | 0.7 | 0.3 |
| 4 | 8 | 0.2 | 0.8 |
| 7 | 9 | 0.1 | 0.9 |

24/3 2

- b. How can you handle the following data?
1. Inconsistence records
 2. Duplicate records
 3. Null records
 4. Unbalanced data

OR

Q-3 Attempt the following:

- a. What are outliers? Explain different methods for handling outliers.
- b. What is Data Preprocessing? Explain major tasks in Data Preprocessing in detail with examples.

SECTION – II

Q-4 Attempt the following. (18)

- a. Name three metrics to evaluate a classification model and explain them.
- b. What is Hierarchical Clustering? How do we compute distance between two clusters? Explain all with formulas, diagrams and examples.

Q-5 What are the different ways to create a pandas data frame. State how to perform following functionalities using pandas: (16)

- a. Fill null values
 - b. Remove irrelevant column
 - c. Find maximum and minimum values
 - d. Create new column with existing ones
 - e. Count number of occurrences in individual column
 - f. Generate heat map
 - g. Remove duplicate observations
- Rename the columns

Q-6 Attempt the following: (16)

- a. What are the different types of charts plotting facility available in MATPLOTLIB? Explain where each is applicable.
- b. How do you adjust more than one figures on single plot area. How are images plotted using MATPLOLIB? In which color spaces they are shown by default?

OR

Q-6 Attempt the following: (16)

- a. How to identify if your trained machine learning model is getting overfitted? What cross validation technique is used to solve the problem?
- b. How do you fine best parameters of ML model to avoid overfitting

—X—