Roll No. | | | | | | | | | | | | Total No. of Pages : 03

Total No. of Questions : 09

MCA (Sem.–4)

MACHINE LEARNING AND DATA ANALYTICS USING PYTHON

Subject Code : PGCA-1976

Paper ID : 91855

Date of Examination : 22-05-2023

Time : 3 Hrs. Max. Marks : 70

INSTRUCTIONS TO CANDIDATES :

1. SECTION-A is COMPULSORY consisting of TEN questions carrying TWO marks each.
2. SECTION - B & C have FOUR questions each.
3. Attempt any FIVE questions from SECTION B & C carrying TEN marks each.
4. Select atleast TWO questions from SECTION - B & C.

SECTION-A

1. Write short notes on :

a) Explain the difference between simple and multiple linear regression.

b) Name all models which can be implemented for using the linear classification.

c) Can we do multiclass classification using logistic regression? Explain.

d) Differentiate between KNN for classification versus regression.

e) Explain the working of decision trees briefly.

f) What are the drawbacks of a decision tree?

g) What is a DataFrame in Pandas? Write the command to import a CSV file.

h) What is the full form of Numpy?

i) How to decide whether the PCA will be employed?

j) What is unsupervised learning? Explain with an example.

SECTION-B

2. a) Explain the principle of bagging in Random Forest with an example. Define the out-of-bag error.

   b) What are the advantages of Random Forests over Decision Trees?

3. a) How is the principal component analysis method helpful in reducing the complexity of the model being trained?

   b) What is the role of the Sigmoid function in logistic regression?

4. a) What is the functionality of hidden layers in a neural network?

   b) What is a perceptron? Explain the hyperparameter tuning in neural network training.

5. Consider the following training set that classifies the output variable play tennis as Yes or No depending upon weather conditions such as Outlook, Temperature, Humidity, and Windy Status.

S. No.	Outlook	Temperature	Humidity	Windy	Play Tennis
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rainy	Mild	High	Weak	Yes
5	Rainy	Cool	Normal	Weak	Yes
6	Rainy	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rainy	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rainy	Mild	High	Strong	No

Using Naive Bayes Classifier, classify whether, on a Rainy, Mild, High Humidity, and Windy day, we can play tennis or not.

SECTION-C

6. Write the Python code for printing the first N prime numbers.

7. a) What are the various functionalities of Pandas library?

   b) What is the difference between series and Data Frames? Explain using examples.

8. a) Explain the usage of the Matplotlib library using examples.

   b) When do we need to plot histograms versus pie charts? Discuss with examples.

9. a) How to handle missing data using Pandas?

   b) Discuss the usage of joining, concatenation, and merging in Pandas.

NOTE : Disclosure of Identity by writing Mobile No. or Marking of passing request on any paper of Answer Sheet will lead to UMC against the Student.