(b) Write short notes on the following : (CO5)

(i) Data manipulation

(ii) Matplotlib

(iii) File handling

(iv) Built in functions

(c) Explain Scikit-Learn and Pandas in detail with examples. (CO5)

420

H Roll No. .....

TMC–302

MCA (THIRD SEMESTER) END SEMESTER EXAMINATION, Dec., 2022

MACHINE LEARNING USING PYTHON

Time : Three Hours

Maximum Marks : 100

Note : (i) All questions are compulsory.

(ii) Answer any *two* sub-questions among (a), (b) and (c) in each main question.

(iii) Total marks in each main question are **twenty**.

(iv) Each sub-question carries 10 marks.

1. (a) What is Machine Learning ? Explain its various approaches. What are the differences between machine learning and traditional programming ? (CO1)

(b) Explain the properties of matrix multiplication. Find out the inverse of the given matrix : (CO1)

$$A = \begin{bmatrix} 1 & -1 & 2 \\ 4 & 0 & 6 \\ 0 & 1 & -1 \end{bmatrix}$$

(c) Define Outliers. Discuss the IQR method of outlier detection. Find the outliers for the following data set : (CO1)

3, 10, 14, 22, 19, 29, 70, 49, 36, 32

2. (a) Consider a tuple with values Hindi, English, Mathematics, Science, EVS and Drawing. Convert tuple into string and list. Write a program to make a list by extracting first letter of each word from the above given words. (CO2)

(b) Write short notes on the following : (CO2)

(i) Variable

(ii) Vector and Scalar

(iii) Average Deviation and Range

(iv) Command Line Argument

(c) List and explain different arithmetic operators supported by Python. Discuss about their precedence and associativity. (CO2)

3. (a) Suppose that the data mining task is to cluster points (with (x, y) representing location) into three clusters were the points are : (CO3)

A1 (2, 10), A2 (2, 5), A3 (8, 4), B_1 (5, 8), B_2 (7, 5), B_3 (6, 4), C1 (1, 2), C2 (4, 9).

The distance function is Euclidean distance. Suppose initially we assign A1, B1, and C1 as center of each cluster, respectively. Use K-means algorithm to show only :

(i) The three cluster centers after the first round of execution.

(ii) The final three clusters.

(b) What is Dimensionality Reduction ? What are the benefits of applying dimensionality reduction ? Explain common techniques of dimensionality reduction in brief. (CO3)

(c) Discuss hierarchical clustering. Find the clusters single link technique. Use Euclidean distance and draw the dendrogram : (CO3)

	X	Y
P1	0.40	0.53
P2	0.22	0.38
P3	0.35	0.32
P4	0.26	0.19
P5	0.08	0.41
P6	0.45	0.30

4. (a) What is training and validation set ? Explain the various validation techniques with example. (CO4)

(b) Define Regression. Explain linear and logistic regression with code snapshots. (CO4)

(c) Write short note on the following : (CO4)

(i) KNN

(ii) NumPy

(iii) SVM

(iv) ANN

5. (a) Given:

Day	Temperature	Wind speed	Event
01-01-2017	32	6	Rain
01-04-2017		9	Sunny
01-05-2017	28		Snow
01-06-2017		7	
01-07-2017	32		Rain
01-08-2017			Sunny
01-09-2017			
01-10-2017	34	8	Cloudy
01-11-2017	40	12	Sunny

Based on the above data, write code for the following operations : (CO5)

(i) Read csv file

(ii) Fill all the missing values

(iii) Drop NA

(iv) Find out the maximum temperature

(v) Plot bar chart showing day and temperature