[5919]-32

M.Sc. (Computer Science)
CSUT - 232 : MACHINE LEARNING
(2019 Pattern) (Semester - III)

Time : 3 Hours] *[Max. Marks : 70*

Instructions to the candidates:

1) *Quesiton 1 is compulsory.*

2) *Solve any Five questions from Q.2 to Q.7.*

3) *Quesiton Q.2 to Q.7 carry equal marks.*

Q1) **Solve any Five of the following :** **[10]**

a) Define Entropy.

b) Define information Gain.

c) List out the libraries of python which support ML?

d) List out the steps of data preprocessing in Machine Learning?

e) What is feature scaling?

f) What is Under fitting?

Q2) **Solve the following :** **[12]**

a) Describe Classification in brief? **[7]**

b) How is KNN different from k-means clustering? **[5]**

Q3) **Solve the following :** **[12]**

a) Define and explain precision and recall. **[7]**

b) Find the Probability of Fruit with Yellow color, Sweet Taste and Long in Size. (Naïve Bay's Theorem). **[5]**

Fruit	Yellow	Sweet	Long	Total
Orange	350	450	0	650
Banana	400	300	350	400
Others	50	100	50	150
Total	800	850	400	1200

P.T.O.

Q4) **Solve the following :**   **[12]**

a) What is Bayes' Theorem? Discuss how it is useful in a machine learning context.   **[7]**

b) Perform KNN algorithm & predict the type of a fruit or food type to which tomato (Sweet = 6, crunch = 4) belongs.   **[5]**

Ingredient	Sweet	Crunch	Food Type
Grape	8	5	Fruit
Green Bean	3	7	Vegetable
Nuts	3	6	Protein
Orange	7	3	Fruit

Q5) **Attempt the following :**   **[12]**

a) Write an algorithm of an agglomerative clustering and Perform Agglomerative algorithm on the following data and plot a dendrogram using **complete linkage** approach.   **[7]**

Item	P1	P2	P3	P4	P5
P1	0				
P2	9	0			
P3	3	7	0		
P4	6	5	9	0	
P5	11	10	2	8	0

b) How do you handle missing or corrupted data in a dataset?   **[5]**

Q6) **Solve the following :**   **[12]**

a) Explain Association Rules in ML?   **[7]**

b) Consider following Data Points, Using Linear SVM plot the graph. (Select 3 Support Vectors).

   Positively Labeled Data = {(3, 1), (3, -1), (6, 1), (6, -1)}

   Negatively Labeled Data = {(1, 0), (0, 1), (0, -1), (-1, 0)}   **[5]**

Q7) **Write short notes on any Two of following.**   **[12]**

a) Q learning.

b) PCA (Principle Component Analysis).

c) Cross-Validation.

▽ ▽ ▽ ▽