Total No. of Questions : 12]

SEAT No. :

PA-672

[Total No. of Pages : 4

[5928]-118

M.E. (A.I & D.S./Computer)

MATHEMATICAL FOUNDATION FOR DATA SCIENCE (2017 Pattern) (Semester - I) (510301)

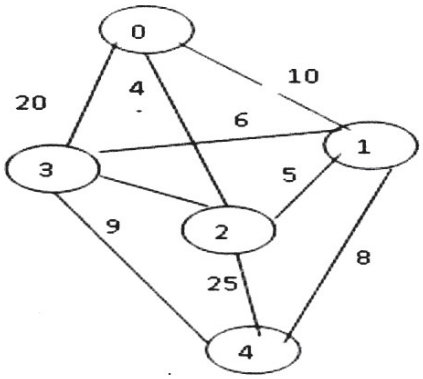
Time : 3 Hours] *[Max. Marks : 50*

Instructions to the candidates:

1) ***Answer Q.1 or Q.2, Q.3 or Q.4, Q.5 or Q.6, Q.7 or Q.8, Q.9 or Q.10, Q.11 or Q.12.***
2) ***Neat diagrams must be drawn wherever necessary.***
3) ***Figures to the right indicate full marks.***
4) ***Assume suitable data, if necessary.***

Q1) a) In a group of 100 persons, 72 people can speak English and 43 can speak French. How many can speak English only? How many can speak French only and how many can speak both English and French? **[3]**

b) Represent the given Graph using adjacency list or Adjacency Matrix. Write Depth First Search Traversal of given graph considering 2 as starting vertex. **[6]**



OR

Q2) a) Prove the following by using Venn diagram. **[6]**

i) $(A \cap B) \cap C = A \cap (B \cap C)$

ii) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

b) The following numbers are inserted into an empty binary search tree in the given order: 20, 10, 30, 50, 15, 12, 16, 40, 14. Construct tree and write its Inorder traversal. **[3]**

P.T.O.

Q3) a) For given attribute marks values: **[4]**

13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70. Compute mean, median, mode, standard deviation.

b) Compare Poisson distribution and Binomial distribution. **[4]**

OR

Q4) a) Explain following for numerical data using example: **[6]**

i) Quantile

ii) Five Number summary

b) A bag contains 2 white balls, 3 black balls and 4 red balls. In how many ways can 3 balls be drawn from the bag, if at least one black ball is to be included in the draw? **[2]**

Q5) a) For given attribute marks values: **[5]**

10, 90, 30, 20, 50, 30, 60, 40, 70, 40, 30, 60, 80, 20

Compute standard deviation, Range, Inter Quartile Range (IQR), five number summary plot it using boxplot.

b) Explain concept and application of Skewness & Kurtosis. **[4]**

OR

Q6) a) Compute covariance of age and glucose values given below: **[4]**

(Age - 1.5, 2, 1.6, 1.2, 1.1)

(Glucose - 1.7, 1.9, 1.8, 1.5, 1)

b) Explain any two graphical representation methods for qualitative data. **[5]**

Q7) a) Use these methods to normalize the following group of data: **[4]**

20, 30, 40, 60, 100

i) z-score normalization

ii) z-score normalization using the mean absolute deviation instead of standard deviation

b) Explain any one Probabilistic models with hidden variables using example. **[4]**

OR

Q8) a) Find correlation of following data set $x = \{2, 5, 6, 8, 9\}, y = \{4, 3, 7, 5, 6\}$. **[3]**

b) Consider following dataset, predict the class label using Naive Bayesian classification for tuple (Yes, No, Male, Yes, B). **[5]**

Owns home	Married	Gender	Employed	Credit rating	Risk class
Yes	Yes	Male	Yes	A	B
No	No	Female	Yes	A	A
Yes	Yes	Female	Yes	B	C
Yes	No	Male	No	B	B
No	Yes	Female	Yes	B	C
No	No	Female	Yes	B	A
No	No	Male	No	B	B
Yes	No	Female	Yes	A	A
No	Yes	Female	Yes	A	C
Yes	Yes	Female	Yes	A	C

Q9) a) Solve the following system of equations using matrix Decomposition method: **[4]**

$$X_1 + X_2 + X_3 = 1,\ 4X_1 + 3X_2 - X_3 = 6,\ 3X_1 + 5X_2 + 3X_3 = 4$$

b) Explain one application of Jacobian Matrix. **[4]**

OR

Q10)a) Find the eigenvalues of the matrix: **[4]**

$$\begin{bmatrix} 2 & 2 \\ 5 & -1 \end{bmatrix}$$

b) Explain two forms of Chain Rule. **[4]**

Q11)a) Overfitting and Multicollinearity with respect to regression. **[4]**

b) Find linear regression equation for the following two sets of data: **[4]**

X	1	2	3	4	5	6	7
Y	9	8	10	12	11	13	14

OR

Q12)Suppose we have the following dataset with one response variable Y and two predictor variables X1 and X2. Fit a multiple linear regression model to this dataset. **[8]**

X1	60	62	67	70	71	72	75	78
X2	22	25	24	20	15	14	14	11
Y	140	155	159	179	192	200	212	215