

[6005]-661

M.E. (Computer Engineering & Artificial Intelligence & Data Science)

MATHEMATICAL FOUNDATION FOR DATA SCIENCE

(2020 Pattern) (Semester - I) (510301)

Time : 3 Hours]

[Max. Marks : 50

Instructions to the candidates:

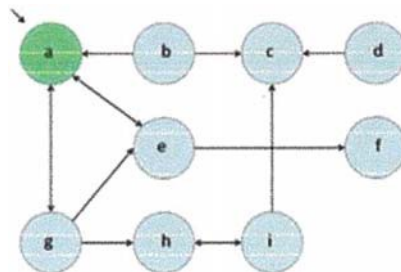
- 1) Solve Q1 or Q2, Q3 or Q4, Q5 or Q6, Q7 or Q8, Q9 or Q10, Q11 or Q12
- 2) Neat diagrams must be drawn wherever necessary.
- 3) Figures to the right indicate full marks.
- 4) Assume suitable data, if necessary.

Q1) a) Determine which of the following sets are countable or uncountable and justify. [3]

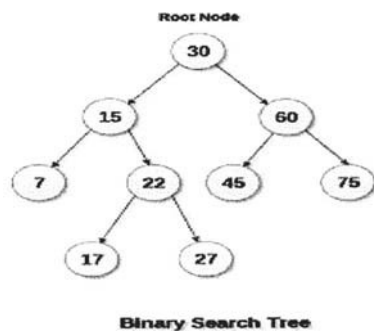
$$A = \{x \in \mathbb{Q} \mid -100 \leq x \leq 100\}$$

$$B = (0, 0.1]$$

b) Represent given graph using adjacency matrix. [3]



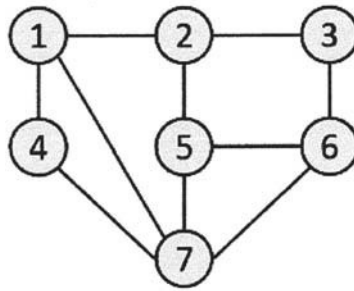
c) Write Inorder traversal of given Binary Search Tree. [3]



OR

P.T.O.

- Q2)** a) The following numbers are inserted into an empty binary search tree in the given order : 10, 1, 3, 5, 15, 12, 16, 4, 14. Construct tree step by step. [3]
- b) In a group of 60 people, 27 like cold drinks and 42 like hot drinks and each person likes at least one of the two drinks. How many like both coffee and tea? [3]
- c) Write Breadth First Search Traversal of given graph considering 2 as starting vertex. [3]



- Q3)** a) For given attribute marks values: [4]
 10, 90, 30, 20, 50, 30, 60, 40, 70, 40, 30, 60, 80, 20.
 Compute mean, median, mode. Also compute five number summary.
- b) There are 35 students in art class and 57 students in dance class. Find the number of students who are either in art class or in dance class. When two classes meet at different hours and 12 students are enrolled in both activities. [4]

OR

- Q4)** a) Explain any two Central tendency and dispersion measures of numerical data with example. [4]
- b) A machinist produces 22 items during a shift. Three of the 22 items are defective and the rest are not defective. In how many different orders can the 22 items be arranged if all the defective items are considered identical and all the non-defective items are identical of a different class? [4]

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

50, 40, 30, 30, 40, 50, 50, 30, 40, 50, 60, 60, 50, 50

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) What is correlation coefficient 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 correlation analysis of Nominal attribute using Chi-square. [5]

Q7) a) Find covariance for 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]

Owens 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

OR

Q8) a) Use these methods to normalize the following group of data : [4]
200, 300, 400, 600, 1000

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]

Q9) a) Solve the following system of equations using LU Decomposition method:

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

b) List the applications of chain rule and discuss any one in detail. [4]

OR

Q10)a) Find the eigenvalues and associated eigenvectors of the matrix. [4]

$$\begin{bmatrix} -1 & 2 \\ 0 & -1 \end{bmatrix}$$

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

Q11)a) Discuss advantages and disadvantages of Multivariate Regression. [4]

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

x	2	4	6	8
y	3	7	5	10

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]

X ₁	60	62	67	70	71	72	75	78
X ₂	22	25	24	20	15	14	14	11
Y	140	155	159	179	192	200	212	215

x x x