

Title : Question Paper

FF No. 868

Reg.No. 12220156

Bansilal Ramnath Agarwal Charitable Trust's  
VISHWAKARMA INSTITUTE OF TECHNOLOGY, PUNE – 411037.  
( An Autonomous Institute Affiliated to Savitribai Phule Pune University)  
Examination: ESE

Year: S.Y. Common

Branch:

Subject: Data science

Subject Code: MD 2201

Max. Marks:60

Total Pages of Question Paper: 1 + 1

Day & Date: Thursday, 11/05/23

Time: 11.00 am -1.00 pm

### Instructions to Candidate

1. All questions are compulsory.
2. Neat diagrams must be drawn wherever necessary.
3. Figures to the right indicate full marks.

Q.No.	CO No	BT No		Max marks
Q.1.	1	1	Match the pair and explain each term on written on left side (shown in numbers) with an example – 1. Validity                      A. Degree to which data is consistent 2. Accuracy                    B. Degree to which data conforms to constraints 3. Uniformity                C. Degree to which data is close to true values 4. Consistency              D. Degree to which data is specified using same Units	8
Q. 2. (A)	2	2	What is a 'Choose function' in Binomial distribution. What is the probability of getting a desired outcome of '3', 4 times when an unbiased dice is rolled 10 times?	6
(B)	2	2	A group of researchers want to test the possible effect of an epilepsy medication taken by pregnant mothers on the cognitive development of their children. As evidence, they want to estimate the IQ scores of 3-yr old children born to mothers who were on this medication during pregnancy. Previous studies suggest that the SD of IQ scores is 18 points. How many such children should the researchers sample in order to obtain a 95% confidence interval with a margin of error $\leq 4$ points?  (Z* for 95% Confidence Interval = 1.96)	6
Q. 3. (A)	3	1	Calculate the distance between points A (4,5,6) and B (7.2,2.8,6.8) using i. Manhattan Distance metric and ii. Euclidean Distance metric	6
(B)	3	2	For a given univariate function $f(x) = 3x^4 - 4x^3 - 12x^2 + 3$ find out the optimal local minimum and global minimum value.	4
Q. 4. (A)	4	4	Two wheeler sales of a leading brand are modelled by a Multiple Linear Regression equation – Sales = - 22036.74 – 1007.61 Unit Cost + 358.88 Ad Exp + 674. 92 Prom Exp. Unit Cost of the Vehicle is in 'Thousand Rs. Ad Exp. is Rs. in Lac and Prom Exp. is also Rs. in Lac. Interpret the four coefficients.	6
(B)	4	1	How is the figure of merit $R^2$ calculated in regression? Explain with a neat sketch. Also justify that value of $R^2$ tending to 1 indicates a good fit of the regression model and a value tending to 0 indicates a bad fit	4
Q. 5. (A)	5	4	The training data for a supervised classification is as follows – $X_1 = (1.8, 1.4, 1)$ , $X_2 = (2.2, 1.8, 1)$ , $X_3 = (3.0, 2.4, 1)$ , $X_4 = (2.2, 2.4, 1)$ , $X_5 = (6.2, 4.4, 2)$ , $X_6 = (7.2, 4.4, 2)$ , $X_7 = (6.6, 4.4, 2)$ , $X_8 = (7.0, 4.6, 2)$ . The test datapoint is at (4.4, 3.3). Use K-nn approach with $k = 3$ to assign appropriate class to the test point	6



Q. 5. (B)	5	4	Consider 40 patterns that are split in 3 classes with 8, 20 and 12 samples respectively. Calculate the Entropy impurity, Gini impurity and Misclassification impurity at the node.	6																																																																																																
			OR alternate option for Q.5 (A) AND Q. 5 (B) together as Q. 5 (C)																																																																																																	
Q. 5 (C)	5	4	<p>The table below gives the following training data –</p> <table><tr><th>Sr. No.</th><th>Cook</th><th>Mood</th><th>Cuisine</th><th>Time</th><th>Tasty</th></tr><tr><td>1</td><td>Sita</td><td>Bad</td><td>Indian</td><td>Lunch</td><td>Yes</td></tr><tr><td>2</td><td>Sita</td><td>Good</td><td>Mexican</td><td>Dinner</td><td>Yes</td></tr><tr><td>3</td><td>Asha</td><td>Bad</td><td>Indian</td><td>Lunch</td><td>No</td></tr><tr><td>4</td><td>Asha</td><td>Bad</td><td>Mexican</td><td>Dinner</td><td>No</td></tr><tr><td>5</td><td>Asha</td><td>Good</td><td>Thai</td><td>Lunch</td><td>Yes</td></tr><tr><td>6</td><td>Sita</td><td>Good</td><td>Thai</td><td>Dinner</td><td>Yes</td></tr><tr><td>7</td><td>Sham</td><td>Good</td><td>Indian</td><td>Lunch</td><td>Yes</td></tr><tr><td>8</td><td>Sham</td><td>Bad</td><td>Mexican</td><td>Dinner</td><td>No</td></tr><tr><td>9</td><td>Asha</td><td>Bad</td><td>Thai</td><td>Lunch</td><td>No</td></tr><tr><td>10</td><td>Asha</td><td>Good</td><td>Thai</td><td>Dinner</td><td>Yes</td></tr><tr><td>11</td><td>Sita</td><td>Good</td><td>Indian</td><td>Dinner</td><td>Yes</td></tr><tr><td>12</td><td>Sham</td><td>Bad</td><td>Thai</td><td>Lunch</td><td>No</td></tr><tr><td>13</td><td>Sham</td><td>Good</td><td>Thai</td><td>Dinner</td><td>Yes</td></tr><tr><td>14</td><td>Sita</td><td>Bad</td><td>Mexican</td><td>Lunch</td><td>Yes</td></tr><tr><td>15</td><td>Sham</td><td>Good</td><td>Indian</td><td>Lunch</td><td>No</td></tr></table> <p>Using Naïve Bays Classification, estimate the class for 'Tasty' for the given feature vector <math>X = \{ \text{Cook} = \text{Sham}, \text{Mood} = \text{Bad}, \text{Cuisine} = \text{Indian}, \text{Time} = \text{Lunch} \}</math></p>	Sr. No.	Cook	Mood	Cuisine	Time	Tasty	1	Sita	Bad	Indian	Lunch	Yes	2	Sita	Good	Mexican	Dinner	Yes	3	Asha	Bad	Indian	Lunch	No	4	Asha	Bad	Mexican	Dinner	No	5	Asha	Good	Thai	Lunch	Yes	6	Sita	Good	Thai	Dinner	Yes	7	Sham	Good	Indian	Lunch	Yes	8	Sham	Bad	Mexican	Dinner	No	9	Asha	Bad	Thai	Lunch	No	10	Asha	Good	Thai	Dinner	Yes	11	Sita	Good	Indian	Dinner	Yes	12	Sham	Bad	Thai	Lunch	No	13	Sham	Good	Thai	Dinner	Yes	14	Sita	Bad	Mexican	Lunch	Yes	15	Sham	Good	Indian	Lunch	No	12
Sr. No.	Cook	Mood	Cuisine	Time	Tasty																																																																																															
1	Sita	Bad	Indian	Lunch	Yes																																																																																															
2	Sita	Good	Mexican	Dinner	Yes																																																																																															
3	Asha	Bad	Indian	Lunch	No																																																																																															
4	Asha	Bad	Mexican	Dinner	No																																																																																															
5	Asha	Good	Thai	Lunch	Yes																																																																																															
6	Sita	Good	Thai	Dinner	Yes																																																																																															
7	Sham	Good	Indian	Lunch	Yes																																																																																															
8	Sham	Bad	Mexican	Dinner	No																																																																																															
9	Asha	Bad	Thai	Lunch	No																																																																																															
10	Asha	Good	Thai	Dinner	Yes																																																																																															
11	Sita	Good	Indian	Dinner	Yes																																																																																															
12	Sham	Bad	Thai	Lunch	No																																																																																															
13	Sham	Good	Thai	Dinner	Yes																																																																																															
14	Sita	Bad	Mexican	Lunch	Yes																																																																																															
15	Sham	Good	Indian	Lunch	No																																																																																															
Q. 6. (A)	6	2	Explain the 'k-fold cross validation' approach used in planning the training and testing data	4																																																																																																
(B)	6	3	A Confusion matrix for a classification exercise returns the following values – TP = 0.942, TN = 0.912, FP = 0.14, FN = 0.06. Calculate Accuracy, precision, recall and f-score	4																																																																																																