

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -3 EXAMINATION- 2024

VI Semester (IT/BT/BI)

COURSE CODE (CREDITS): 18B11CI613/ 18B1WBI632 (3)

MAX. MARKS: 35

COURSE NAME: Data Mining

COURSE INSTRUCTORS: Dr. Aman Sharma

MAX. TIME: 2 Hours

Note: (a) All questions are compulsory. (b) Marks are indicated against each question in square brackets. (c) The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems

Q1. For the given data, compute two clusters using K-means algorithm for clustering where initial cluster centers are (1.0, 1.0) and (5.0, 7.0). Execute for two iterations. **[3 Marks, CO-4]**

Record Number	A	B
R1	1.0	1.0
R2	1.5	2.0
R3	3.0	4.0
R4	5.0	7.0
R5	3.5	5.0
R6	4.5	5.0
R7	3.5	4.5

Q2. Use the distance matrix in Table below to perform single link and complete link hierarchical clustering. Show your results by drawing a dendrogram. The dendrogram should clearly show the order in which the points are merged. **[5 Marks, CO-4]**

	P1	P2	P3	P4	P5
P1	0.00	0.10	0.41	0.55	0.35
P2	0.10	0.00	0.64	0.47	0.98
P3	0.41	0.64	0.00	0.44	0.85
P4	0.55	0.47	0.44	0.00	0.76
P5	0.35	0.98	0.85	0.76	0.00

Q3. Anomalies can sometimes reveal valuable insights or trends that were previously unnoticed. Discuss five different case studies where anomalies can be used for revealing valuable insights or trends. **[5 Marks, CO-1]**

Q4. How we can perform hyper parameter tuning in Decision tree, Random Forest, SVM, Neural Networks. Write the hyper parameters of all the mentioned algorithms. **[4 Marks, CO-2]**

Q5. Write the equations for five activation functions used in neural networks along with their significance and graphical representation. **[5 Marks, CO-3]**

Q6. Calculate recall, precision, sensitivity, specificity, f-1 score and accuracy for the below mentioned confusion matrix. **[6 Marks, CO-2]**

n=165	Predicted: NO	Predicted: YES
	Actual: NO	Actual: YES
	50	10
	5	100

Q7. Consider the following incomplete dataset about users of a web site and its document search and retrieval history. The dataset includes the following attributes:

User ID	Document ID	Search Query	Clicked	Timestamp	Duration (seconds)
1	201	"Data Science"	Yes	2024-04-01 10:00:00	180
1	202	"Data Science"	No	2024-04-01 10:10:00	45
2	203	"Machine Learning"	Yes	2024-04-02 11:30:00	120
2	204	"Big Data"	Yes	2024-04-02 11:45:00	150
3	205	"Data Mining"	No	2024-04-03 09:00:00	30
3	201	"Data Science"	Yes	2024-04-03 09:15:00	200

1. Identify the equivalents of items and item-grouping attribute (e.g., transaction) in the context of this dataset for applying association rule mining.
2. Explain how you would apply association rule mining to this dataset. Include steps such as data preparation, choice of item and transaction definitions, and any preprocessing required.
3. List potential uses of the patterns obtained from the data set. [2+3+2 Marks, CO-3]

*****Best of Luck*****