Reg. No. \_\_\_\_\_\_\_\_\_\_\_\_\_



**End Semester Examination – Nov / Dec – 2019**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| **Code :** | **14CS3074** | **Duration :** | **3hrs** |
| **Sub. Name :** | **ADVANCED DATA MINING** | **Max. Marks :** | **100** |

**ANSWER ALL QUESTIONS (5 x 20 = 100 Marks)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Q. No.** | **Sub Div.** | **Questions** | **Course**  **Outcome** | **Marks** |
| 1. | a. | State the purpose of sparse data. How will you prepare a Library dataset with the following attributes in ARFF format  Library(Book\_title, Book\_id, Author\_name, Publisher, Year, No\_of\_copies) | CO1 | 8 |
| b. | Explain the concept of Replicated subtree with suitable illustration | CO2 | 8 |
| c. | Describe the steps involved in the process of Knowledge Discovery. | CO1 | 4 |
| **(OR)** | | | | |
| 2. | a. | List the various output knowledge representations. Explain any four representations in detail with suitable illustrstions. | CO2 | 15 |
| b. | Distinguish the classification and clustering tasks of datamining technique and mention some of its related applications. | CO2 | 5 |
|  |  |  |  |  |
| 3. | a. | Apply 1D Haar Wavelet Transform and convert the following dataset to wavelet transformed data. Explain the steps involved. Reconstruct the data to its original form and calculate the reconstruction error.   |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | 15 | 30 | 25 | 36 | 49 | 52 | 63 | 17 | | CO2 | 10 |
| b. | Consider the following dataset: 11, 14, 20, 15, 25, 5, 45, 50, 17, 51, 29, 10, 72, 33, 75. Partition them into three bins using the following methods: (a) equal-frequency partitioning (b) equal-width partitioning. | CO2 | 10 |
| **(OR)** | | | | |
| 4. | a. | What is data reduction? Explain in detail about principal component analysis. | CO2 | 10 |
| b. | Use the following dataset: 27, 14, 22, 21, 33, 15, 18, 9, 26, 40, 36, 7. (a) Use min-max normalization to transform the value 35 into the range [0.0, 1.0], (b) Use normalization by decimal scaling to transform the value 35, (c) Use Z-score normalization to transform the value 35. | CO2 | 10 |
|  |  |  |  |  |
| 5. | a. | Consider the following table and derive the classification rules using covering algorithm.   | Age | Job | Marital status | Education | Salary | loan | Potential Customer | | --- | --- | --- | --- | --- | --- | --- | | 20-30 | unemployed | married | primary | no | no | no | | 30-40 | self employed | single | secondary | yes | yes | yes | | 40-50 | management | single | tertiary | yes | no | yes | | 40-50 | management | married | tertiary | yes | yes | no | | 50-60 | self employed | married | primary | yes | no | No | | CO2 | 10 |
| b. | Identify the best attribute of the dataset given in question 5a using IR algorithm. | CO2 | 10 |
| **(OR)** | | | | |
| 6. | a. | Construct a decision tree for the dataset shown in question 5a. | CO2 | 15 |
| b. | State the significance of incorporating exceptions in classification rules with suitable example. | CO2 | 5 |
|  |  |  |  |  |
| 7. | a. | Describe the concepts of learning in Artificial Neural Network. Explain Perceptron learning law and learn the patterns of ‘AND’ gate and analyze the updated weight values after one iteration.  Use Net = 1 if Net >= 0.7  = 0 if Net < 0.7  The initial weight vector = [0.1, 0.3], c = 0.1 | CO2 | 10 |
| b. | State the purpose of Multiple Layer Perceptron (MLP) algorithm. Draw the architecture and explain the working principle of the algorithm. | CO2 | 10 |
| **(OR)** | | | | |
| 8. | a. | Demonstrate the working principle of K-Means algorithm in clustering the following data points into 2 clusters:  A1(1,4), A2(2,8), A3(3,4), A4(4,2), A5(5,6), A6(7,2), A7(10,4), A8(12,10), A9(5,5)  Use Euclidean distance measure. Intially, assign A2 and A8 as the centres of each cluster. Show the new cluster centers after the first iteration. | CO2 | 10 |
| b. | Summarize the working principle of cluster formation in BIRCH hierarchical clustering algorithm and explain using suitable illustration. | CO2 | 10 |
|  | | **Compulsory**: |  |  |
| 9. | a. | Discuss on the following methods:   * Cross-validation * “0.632 bootstrap” method | CO3 | 10 |
| b. | List all the performance metrics. Compute the metrics for the following data.   |  |  |  |  | | --- | --- | --- | --- | |  |  | **Predicted Class** | | | **A** | **B** | | **Actual Class** | **A** | 170 | 20 | | **B** | 57 | 110 | | CO3 | 10 |