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



**UNIVERSITY**

****

(Karunya Institute of Technology & Sciences)

(Declared as Deemed-to-be University under Sec.3 of the UGC Act, 1956)

**End Semester Examination – Nov/Dec – 2017**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| **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. | An application has to be implemented to categorize the employees of a company based on their profile. Suggest the input and output parameters of the application. Recommend atleast 2 data mining techniques to categorize the employees and state their corresponding output knowledge representation. Create ARFF file for the employee data. | CO1 | 8 |
| b. | Compare the different means of representing the knowledge extracted from the dataset by data mining techniques. | CO1 | 12 |
| (OR) | | | | |
| 2. | 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.   |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | 25 | 30 | 15 | 16 | 19 | 22 | 23 | 17 | | CO2 | 10 |
| b. | Analyze the following nominal attributes using chi-square test and determine whether students and hobby are correlated.   |  |  |  |  | | --- | --- | --- | --- | |  | **College Students** | **School Students** | **Total** | | Sports | 400 | 1150 | 1550 | | Music | 600 | 850 | 1450 | | **Total** | 1000 | 2000 | 3000 |   Note: Use chi-squared value of 10.828 for df=1. | CO2 | 10 |
| 3. | a. | Construct the first two levels of the decision tree (i.e. the root node and the first level of the tree) using ‘DIVIDE-and-CONQUER’ approach using the following dataset.   |  |  |  |  |  | | --- | --- | --- | --- | --- | | **Age** | **Prescription** | **Astigmatism** | **Tear Rate** | **Recommendations** | | Young | Myope | No | Reduced | None | | Pre-presbyopic | Myope | No | Reduced | None | | Pre-presbyopic | Myope | No | Normal | Soft | | Pre-presbyopic | Myope | Yes | Reduced | None | | Young | Hypermetrope | Yes | Reduced | None | | Young | Hypermetrope | Yes | Normal | Hard | | Presbyopic | Myope | No | Reduced | Hard | | Presbyopic | Myope | Yes | Normal | Hard | | Presbyopic | Hypermetrope | No | Normal | Soft | | CO2 | 15 |
|  | b. | Explain the concept of “instance based representation”. What are the ways in which the instances can be compared? How KD-tree will be useful in searching for an instance? | CO2 | 5 |
| (OR) | | | | |
| 4. | a. | Identify the best attribute using 1R algorithm from the dataset given in question 3a. | CO2 | 10 |
|  | b. | Design Naive Bayes classifier for the dataset given in question 3a. | CO2 | 10 |
|  |  |  |  |  |
| 5. | a. | Model “OR” logical gate using Single Layer Perceptron (SLP) concept. Initialize all the weight links with a value of 0.1 and learning rate parameter = 1. Use the following threshold details.    Show the updated weight vector after 2 iterations. | CO2 | 12 |
|  | b. | Show how the weight updation in SLP is different from Balanced Winnow technique? | CO2 | 4 |
|  | c. | Propose a solution to learn “XOR” gate using neural network. | CO2 | 4 |
| (OR) | | | | |
| 6. | a. | Compare “association” and “classification” rules. Use the dataset in question 3a and generate few 2-itemsets and 3-itemsets with coverage >= 2. Use one of the 3-itemsets and produce the potential association rules. | CO2 | 10 |
|  | b. | Explain the steps involved in training Bayesian Belief Network. | CO2 | 10 |
|  |  |  |  |  |
| 7. | a. | List the hierarchical clustering methods and illustrate the principle of cluster formation of these methods | CO2 | 14 |
|  | b. | Summarize the four widely used measures for calculating distance between two clusters. | CO2 | 6 |
| (OR) | | | | |
| 8. | a. | How can we ﬁnd dense regions in density-based clustering? How does DBSCAN quantify the neighborhood of an object? | CO2 | 10 |
|  | b. | Outline the algorithm of K-Medoids clustering algorithm. | CO2 | 10 |
|  | | **Compulsory**: |  |  |
| 9. | a. | Summarize the different cross-validation techniques with examples | CO2 | 8 |
|  | b. | Evaluate the output (shown below) of 2-class predictor using the following metrics: TP rate, FP rate, Precision, Recall, Accuracy, F-measure and Overall success rate.   |  |  |  |  | | --- | --- | --- | --- | |  | | Predicted Class | | | a | b | | Actual Class | a | 90 | 20 | | b | 15 | 21 | | CO3 | 7 |
|  | c. | Justify the statement “Bootstrap” method is called as “0.632 bootstrap”. | CO2 | 5 |

ALL THE BEST