****

**UNIVERSITY**

(Karunya Institute of Technology & Sciences)

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

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

**End Semester Examination – Nov/Dec – 2016**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **Semester :** | **2016-17 ODD** |
| **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** | **Marks** |
| 1. | a. | What is sparse data? | 2 |
| b. | Is it necessary to use only ARFF format to read data in Weka? What are the other formats that it can support? How will you prepare input data in ARFF format? Provide an example. | 8 |
| c. | What is instance based representation of data mining output? Explain using an example | 10 |
| (OR) | | | |
| 2. | a. | What are the different ways of representing output data? Explain ‘REPLICATED SUBTREE PROBLEM’ with example | 10 |
| b. | Explain the concepts of linear classification using Perceptron and Winnow | 10 |
| 3. | a. | How will you construct rules using “Covering Algorithm” approach? Show the steps of deriving the rules for the class = ‘None’ 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 | None | | Presbyopic | Myope | Yes | Normal | Hard | | Presbyopic | Hypermetrope | No | Normal | Soft | | 10 |
|  | b. | Explain the concept of Naive Bayesian classifier. | 10 |
| (OR) | | | |
| 4. | a. | Calculate information gain using ‘DIVIDE-and-CONQUER’ approach? Show the steps of determining the root node and one more subsequent node using the dataset from question 3a. | 12 |
|  | b. | Explain the concept of 1R algorithm. Use 1R algorithm and determine the best representative attribute of the dataset shown in question 3a | 8 |
| 5. | a. | “Bootstrap” method is called as “0.632 bootstrap”. Why? How will you estimate the error in this method | 6 |
|  | b. | How will you evaluate different data mining schemes using graphical techniques like ROC, life charts and recall-precision curves? | 14 |
| (OR) | | | |
| 6. | a. | What is cross-validation? Write about the different types of cross-validation with examples | 12 |
|  | b. | Evaluate the output (shown below) of 2-class predictor using the following metrics: TP rate, FP rate, Precison, Recall, F-measure and Overall success rate.   |  |  |  |  | | --- | --- | --- | --- | |  |  | Predicted Class | | | a | b | | Actual Class | a | 90 | 20 | | b | 15 | 21 | | 8 |
| 7. |  | Write about different types of clustering and the metrics used to measure the quality of clusters formed. | 20 |
| (OR) | | | |
| 8. |  | Write about Bayesian Network learning. Construct Bayesian Networks using K2 algorithm. How AD tree is able to support fast learning? | 20 |
|  | | **Compulsory:** |  |
| 9. | a. | What are the different types of classifiers does Weka support? Explain atleast 2 classifiers in detail. | 10 |
|  | b. | Weka manipulates attributes/instances using unsupervised filter operations. Explain atleast ten functions. | 10 |

ALL THE BEST