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



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

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| **Code :** | **18CS2013** | **Duration :** | **3hrs** |
| **Sub. Name :** | **MACHINE LEARNING TECHNIQUES** | **Max. Marks :** | **100** |

|  |  |  |  |
| --- | --- | --- | --- |
| **Q. No.** | **Questions** | **Course Outcome** | **Marks** |
| **PART – A (10X1 = 10 MARKS)** | | | |
| 1. | Mention few applications of machine learning techniques in the medical domain. | CO1 | 1 |
| 2. | Consider ‘N’ number of 1-dimensional data points. Determine the number of hypotheses possible. | CO1 | 1 |
| 3. | Differentiate between hard classifier and probabilistic classifier. | CO1 | 1 |
| 4. | Determine the number of possible concepts of an application with five features each holding three possible values. | CO1 | 1 |
| 5. | Define R squared statistical measure of regression model. | CO2 | 1 |
| 6. | State the purpose of support vectors. | CO1 | 1 |
| 7. | List the distance based classification and clustering models. | CO1 | 1 |
| 8. | What are the limitations of K-medoids clustering approach? | CO3 | 1 |
| 9. | What is entropy? | CO2 | 1 |
| 10. | Which impurity meaure is used for CART approach? | CO2 | 1 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **PART – B (6 X 3 = 18 MARKS)** | | | | | |
| 11. | | Calculate the Interquartile range for the following dataset.   1. 12 10 1 7 9 5 16 11 | | CO2 | 3 |
| 12. | | List the measures available for probability estimates. | | CO2 | 3 |
| 13. | | Compare multiple layer perceptron and support vector machine. | | CO3 | 3 |
| 14. | | Write the algorithm of K-Means algorithm. | | CO4 | 3 |
| 15. | | Discuss the separate-and-conquer rule learning strategy in more detail. | | CO4 | 3 |
| 16. | | State the difference between AdaBoost and Random Forest ensemble methods. | | CO3 | 3 |
|  | | | | | |
| **PART – C (6 X 12 = 72 MARKS)**  **(Answer any five Questions from Q.no 17 to 23. Q.No 24 is a Compulsory Question)** | | | | | |
| 17. | a. | | Discuss the concept of learning. Explain the characteristics of three machine learning models with suitable illustrations. | CO3 | 7 |
| b. | | Consider the following dataset: 15, 25, 10, 20, 55, 75, 12, 29, 11, 72, 13, 45. Partition them into ‘k’ bins using the following methods:  (i) equal-frequency partitioning.  (ii) equal-width partitioning. | CO3 | 5 |
|  |  | |  |  |  |
| 18. |  | | Define hypothesis. Explain the ‘Least general generalization’ algorithm. Derive the hypothesis for the following dataset using the algorithm. Explain the intermediate steps.   |  |  |  |  | | --- | --- | --- | --- | | **Color** | **Type** | **Origin** | **Stolen** | | Red | Sports | Domestic | Yes | | Red | General | Domestic | No | | Yellow | Sports | Domestic | Yes | | Yellow | Sports | Imported | Yes | | Yellow | SUV | Imported | No | | CO4 | 12 |
|  |  | |  |  |  |
| 19. | a. | | Apply ‘Naïve Bayes’ approach to the following dataset and prepare the frequency and likelihood tables.   |  |  |  |  |  | | --- | --- | --- | --- | --- | | **Age** | **Income** | **Student** | **Laptop Rate** | **Buy** | | Youth | High | No | Fair | No | | Youth | High | No | High | No | | middle aged | High | No | Fair | Yes | | Senior | Low | Yes | Fair | Yes | | Senior | Medium | No | Fair | Yes | | Youth | Medium | Yes | High | Yes | | middle aged | Medium | No | High | Yes | | Senior | Low | Yes | High | No |   Analyze the performance of the model using the following test dataset.   |  |  |  |  |  | | --- | --- | --- | --- | --- | | **Age** | **Income** | **Student** | **Laptop Rate** | **Buy** | | Senior | High | No | Fair | ? | | middle aged | Low | Yes | Fair | ? | | CO4 | 7 |
| b. | | List the kernel functions that could be used to transform the original input data into a higher dimensional space. Write in detail the Kernel function in Perceptron. | CO3 | 5 |
|  |  | |  |  |  |
| 20. | a. | | In unsupervised learning, there are no target output values. So, what does it learn? List the examples of such networks. Describe the architecture of the Self Organizing Map (SOM) known as a Kohonen network. Explain the algorithm of how the weights in a Kohonen network (Self Organizing Map) updated. | CO3 | 6 |
| b. | | Apply K-Medoids clustering algorithm to the following data points : (3,6), (2,4), (4,7), (5,6), (7,2), (8,4), (9,3), (9,2), (8,6), (3,3) and compute the absolute error. Consider (3,6) and (9,3) as initial medoids. Replace (9,3) with (8,6) and calculate the absolute error. Which are the best medoids? | CO4 | 6 |
|  |  | |  |  |  |
| 21. | a. | | Apply association rule mining approach to the dataset shown in question 19a. Analyze the performance by framing the association rules based on  2- itemsets. | CO4 | 6 |
| b. | | Explain the principle of decision tree construction with suitable illustration. | CO4 | 6 |
|  |  | |  |  |  |
| 22. | a. | | State the significance of performance metrics, ROC and AUC. Using the following output from a probabilistic classifier,  (i) evaluate all the performance measures  (ii) draw the ROC curve.   |  |  |  | | --- | --- | --- | | **Tuple #** | **Actual Class** | **Probability output** | | 1 | P | 0.92 | | 2 | N | 0.80 | | 3 | N | 0.71 | | 4 | N | 0.62 | | 5 | N | 0.60 | | 6 | P | 0.54 | | 7 | N | 0.49 | | 8 | P | 0.45 | | 9 | P | 0.41 | | 10 | P | 0.40 | | CO6 | 8 |
| b. | | What are the approaches available to combine the results of binary classifiers? Discuss in detail. | CO3 | 4 |
|  |  | |  |  |  |
| 23. | a. | | Explain the categories of statistical calculations on features. | CO2 | 6 |
| b. | | Use the following dataset: 17, 4, 12, 21, 13, 25, 8, 9, 16, 10, 30, 1.  (i) Use min-max normalization to transform the value 27 into the range  [0.0, 1.0],  (ii) Use normalization by decimal scaling to transform the value 27,  (iii) Use Z-score normalization to transform the value 27. | CO3 | 6 |
|  |  | |  |  |  |
| **Compulsory:** | | | | | |
| 24. |  | | State why the ensemble methods may improve the accuracy of the classifier. Write the pseudocode and explain the working principle of the following mehods with suitable illustrations.   1. Bagging 2. AdaBoost 3. Random Forests. | CO3 | 12 |