|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Roll No. |  |  |  |  |  |  |  |  |  |  |  |  |



**PRESIDENCY UNIVERSITY**

**Bengaluru**

|  |
| --- |
| **End - Term Examinations – JANUARY 2025** |
| **Date:** 03-01-2025 **Time:** 09:30 am – 12:30 pm |

|  |  |  |
| --- | --- | --- |
| **School:** SOCSE | **Program:** B. Tech(CSE/CCS/CIT/CDV/CBC /CSD/CBD/ISE/ISR/ISB) | |
| **Course Code :** CSE3001 | **Course Name :** Artificial Intelligence and Machine Learning | |
| **Semester**: V | **Max Marks**: 100 | **Weightage**: 50% |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **CO - Levels** | **CO1** | **CO2** | **CO3** | **CO4** | **CO5** |
| **Marks** | **14** | **29** | **16** | **41** |  |

**Instructions:**

1. *Read all questions carefully and answer accordingly.*
2. *Do not write anything on the question paper other than roll number.*

**Part A**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Answer ALL the Questions. 10 x 2 Marks =20 Marks** | | | | |
| **1** | Define the concept of an intelligent agent in AI and its key components. | **2 Marks** | **L1** | **CO1** |
| **2** | List two common knowledge representation techniques used in AI. | **2 Marks** | **L1** | **CO1** |
| **3** | What happens if the max\_depth **parameter** in a decision tree is set to False how does it affect the training data and test data? | **2 Marks** | **L1** | **CO2** |
| **4** | Define Logistic regression. | **2 Marks** | **L1** | **CO2** |
| **5** | Define a cost function in machine learning. Give an example of a cost function used in regression problems. | **2 Marks** | **L1** | **CO3** |
| **6** | Why is minimizing the cost function crucial for model performance in machine learning? | **2 Marks** | **L2** | **CO3** |
| **7** | Define trend in time series data. | **2 Marks** | **L1** | **CO4** |
| **8** | List and explain any two patterns in time series data. | **2 Marks** | **L1** | **CO4** |
| **9** | Define gradient descent and why it is been used in machine learning. | **2 Marks** | **L1** | **CO3** |
| **10** | Differentiate Internal Cluster and External Cluster validation. | **2 Marks** | **L2** | **CO4** |

**Part B**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Answer the Questions Total 80 Marks** | | | | | |
| **11.** | **a.** | Explain the Knowledge Representation techniques with Suitable Example. | **10 Marks** | **L2** | **CO1** |
| **Or** | | | | | |
| **12.** | **a.** | “Imagine you are designing an AI system for a hospital to assist doctors in diagnosing diseases. The system needs to store and reason with complex medical knowledge, including symptoms, diseases, treatments, and patient information.”  Discuss how you would use two different knowledge representation techniques (from the list provided) to represent the medical knowledge in this AI system. Provide specific examples of how each technique would be used to represent different types of medical information. Additionally, compare and contrast the advantages and disadvantages of the two chosen techniques in this specific context. | **10 Marks** | **L2** | **CO1** |
|  |  |  |  |  |  |
| **13.** | **a.** | A dataset contains the following information about whether students will "Play Outside" based on two features: Weather and Homework.   | **Weather** | **Homework** | **Play Outside** | | --- | --- | --- | | Sunny | Done | Yes | | Sunny | Not Done | No | | Cloudy | Done | Yes | | Cloudy | Not Done | Yes | | Rainy | Done | No | | Rainy | Not Done | No |  1. Calculate the entropy of the target variable "Play Outside".   **(b)** Calculate the information gain for both features, "Weather" and "Homework", and determine which feature is the most suitable root node for the decision tree. | **10 Marks** | **L3** | **CO2** |
| **Or** | | | | | |
| **14.** | **a.** | A dataset contains the following information about whether an email is Spam or Not Spam, based on two features: "Contains 'Buy'" and "Contains 'Discount'".   | **Contains 'Buy'** | **Contains 'Discount'** | | **Spam** | | --- | --- | --- | --- | | Yes | Yes | Yes | | | Yes | No | Yes | | | No | Yes | No | | | No | No | No | | | Yes | Yes | Yes | | | No | Yes | No | |  1. Calculate the prior probabilities for the classes Spam and Not Spam. 2. Calculate the conditional probabilities for the features:   P (Contains 'Buy' | Spam), P(Contains 'Buy' | Not Spam)  P (Contains 'Discount' | Spam), P (Contains 'Discount' | Not Spam). Assume Laplace Smoothing with α = 1.  **(c)** Use the Naive Bayes Classifier to determine whether a new email containing 'Buy' = Yes and 'Discount' = No should be classified as Spam or Not Spam. Show all calculations and justify your final answer. | **10 Marks** | **L3** | **CO2** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **15.** | **a.** | Calculate K-Nearest Neighbor (KNN) Algorithm: A new data point z=(30,60) needs to be classified using the KNN algorithm. Given the following data points, classify the new data point if k=3:  • Brightness: 40, Saturation: 20, Color: Red  • Brightness: 50, Saturation: 50, Color: Blue  • Brightness: 60, Saturation: 90, Color: Blue  • Brightness: 10, Saturation: 25, Color: Red  • Brightness: 70, Saturation: 70, Color: Blue | **10 Marks** | **L3** | **CO3** |
| **Or** | | | | | |
| **16.** | **a.** | Examine the common issues faced during the application of Gradient Descent, such as the effect of a poor learning rate, local minima, and convergence speed. Propose solutions to mitigate these issues. | **10 Marks** | **L3** | **CO3** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **17.** | **a.** | Explain forecasting methods with any one example. | **15 Marks** | **L2** | **CO4** |
| **Or** | | | | | |
| **18.** | **a.** | Using sample data set, Examine K means Clustering with its accuracy.(Write Python code for K- means Clustering) | **15 Marks** | **L3** | **CO4** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **19.** | **a.** | Illustrate Decision tree with its execution using Python code. | **15 Marks** | **L2** | **CO2** |
| **Or** | | | | | |
| **20.** | **a.** | Summarize the role of the **cost function** in Logistic Regression. Write down the formula for the cost function and explain why it is preferred over Mean Squared Error (MSE) for classification problems | **15 Marks** | **L2** | **CO2** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **21.** | **a.** | Demonstrate Ensemble Learning using sample data set in detail. | **20 Marks** | **L3** | **CO4** |
| **Or** | | | | | |
| **22.** | **a.** | Explain K means Clustering Algorithm with Procedures using an example. | **20 Marks** | **L2** | **CO4** |

**\*\*\*\*\* BEST WISHES \*\*\*\*\***