End Semester Examination 2023

Name of the program: MCA **Semester: 2**

Name of the course: Machine Learning Using Python **Course Code: TMC-205**

Time : 3 Hour **Maximum Marks: 100**

Note:

i. All questions are compulsory.
ii. Answer any two sub questions among a, b and c in each main question.
iii. Total marks in each main question are twenty.
iv. Each question carries 10 marks.

Q1. **20 marks**

a. Describe the steps involved in the machine learning workflow with an explanation of the types of Machine Learning. — CO1

b. The height details of the boys and girls are given in the table 1: — CO1

Table 1: Sample Data

Height of the Boys(x_i)	65	70	75	78
Height of the Girls(y_i)	63	67	70	73

Fit a Suitable line of best fit for the above data.

c. . Construct a regression tree using the following table 2 which comprises of 10 instances and 3 attributes 'Assessment', 'Assignment' and 'Project'. The target attribute 'Result' is a continuous attribute.

Table 2: Training Dataset

S.No.	Assessment	Assignment	Project	Result(%)
1	Good	Yes	Yes	95
2	Average	Yes	No	70
3	Good	No	Yes	75
4	Poor	No	No	45
5	Good	Yes	Yes	98
6	Average	No	Yes	80
7	Good	No	No	75
8	Poor	Yes	Yes	65
9	Average	No	No	58
10	Good	Yes	Yes	89

Q2. **20 marks**

a. What is regression analysis and how is it different from classification? Explain the concept of dependent and independent variables in regression using an example. — CO2

b. Consider a feed-forward neural network with the following details: — CO2
- Input layer: 3 neurons (x1, x2, x3)
- Hidden layer: 4 neurons (h1, h2, h3, h4)
- Output layer: 2 neurons (o1, o2)

Given the following inputs and biases:

Inputs:
x1 = 0.2, x2 = 0.5, x3 = 0.8

Biases:
Bias_h1 = -0.3, Bias_h2 = 0.1, Bias_h3 = 0.5,Bias_h4 = -0.2
Bias_o1 = 0.4 ,Bias_o2 = 0.6

Assume the following weights :

Weights:
w1_h1 = 0.6, w1_h2 = 0.3, w1_h3 = 0.8, w1_h4 = -0.5
w2_h1 = -0.2, w2_h2 = 0.7, w2_h3 = 0.9, w2_h4 = 0.1
w3_h1 = 0.4 ,w3_h2 = -0.6, w3_h3 = 0.2, w3_h4 = 0.5
w1_o1 = 0.3 ,w1_o2 = -0.1 ,w2_o1 = -0.4 ,w2_o2 = 0.2
w3_o1 = 0.7 ,w3_o2 = 0.5 ,w4_o1 = -0.6 ,w4_o2 = 0.8

Calculate the output values (o1, o2) for the given inputs.

Note: The activation function used in this question is the sigmoid function.

c. Consider a dataset containing the following values: 10, 15, 18, 22, 25, 30, 35, 40
- Normalize the dataset using min-max normalization, where the normalized values will be scaled between 0 and 1.
- Suppose we have a new data point with a value of 28. Normalize this data point using the same min-max normalization technique.

Q3. **20 marks**

a. What is an agent in the context of Reinforcement Learning? Describe the algorithm of Q-Learning. — CO3

b. Consider the following training dataset of 10 instances shown in the Table 3 which describes the award performance of the individual students based on the GPA and no. of projects done. The target variable is 'Award' which is a discrete valued variable that takes 2 values 'Yes' or 'No'. — CO3

Table 3

S.NO.	GPA	No. of Projects Done	Award
1.	9.5	5	Yes
2.	8.0	4	Yes
3.	7.2	1	No

4.	6.5	5	Yes	
5.	9.5	4	Yes	
6.	3.2	1	No	
7.	6.6	1	No	
8.	5.4	1	No	
9.	8.9	3	Yes	
10.	7.2	4	Yes	

Given a test instance (GPA- 7.8, No. of projects done- 4) , use the training set to classify the test instance. Choose k=3

- k-Nearest Neighbor Classifier
- Weighted k-Nearest Neighbor Classifier
- Nearest Centroid Classifier

c. Describe the concepts of support vectors and margin in SVM and Discuss the difference between hard margin and soft margin SVM.

Q4. **20 marks**

a. CO4

(i)Consider the following data and calculate the Euclidean, Manhattan and Chebyshev distances. CO4

(a) (2 3 4) and (1 5 6) and (b) (2 2 9) and (7 8 9)

(ii) Find the distance between

(a) Employee ID: 1000 and 1001 (b) Employee Name: John & John and John & Joan

b. Define bagging and explain how it works in the context of ensemble learning and describe the key steps involved in the bagging algorithm.

C, Explain SARSA learning and SARSA algorithm .

Q5. **20 marks**

a. Consider the following dataset with five data points: CO5

Data points: (2, 4), (3, 6), (4, 8), (7, 3), (8, 5) CO5

Perform k-means clustering with k = 2. Initialize the cluster centroids randomly. Iterate until convergence and assign each data point to its closest centroid. Calculate the final cluster centroids.

b. Consider the following dataset with seven data points:
Data points: (1, 1), (2, 2), (2, 4), (3, 3), (4, 3), (4, 4), (5, 4)
Perform agglomerative clustering step by step, merging the two closest clusters based on the complete linkage criterion

c. Consider the following dataset S={12,14,19,22,24,26,28,31,34}. Apply various binning techniques and show the result.