

**BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI
(END SEMESTER EXAMINATION)**

CLASS: MTech
BRANCH: AIML

SEMESTER : I
SESSION : MO/2025

SUBJECT: AI502 ADVANCED CONCEPTS OF SUPERVISED LEARNING

TIME: 3 Hours

FULL MARKS: 50

INSTRUCTIONS:

1. The question paper contains 5 questions each of 10 marks and total 50 marks.
2. Attempt all questions.
3. The missing data, if any, may be assumed suitably.
4. Before attempting the question paper, be sure that you have got the correct question paper.
5. Tables/Data hand book/Graph paper etc. to be supplied to the candidates in the examination hall.

- | | | CO | BL |
|--|---|-----|----|
| Q.1(a) A hospital uses Naive Bayes to predict whether a patient has a disease (D = Yes) based on symptoms. Historical data shows $P(D = \text{Yes}) = 0.3$, $P(D = \text{No}) = 0.7$, with $P(\text{Cough} = \text{Yes} \mid D = \text{Yes}) = 0.8$ and $P(\text{Cough} = \text{Yes} \mid D = \text{No}) = 0.2$. A new patient presents with a cough. Use Bayes' Theorem to calculate $P(D = \text{Yes} \mid \text{Cough} = \text{Yes})$ and $P(D = \text{No} \mid \text{Cough} = \text{Yes})$, and identify the likely diagnosis. [5] | 1 | 3,5 | |
| Q.1(b) A machine learning model is used to classify whether emails are spam or not. The following confusion matrix summarizes the classification results: Compute the accuracy, precision, recall, and F1-Score. Which one is the best measure between Recall and Precision? Justify your answer. [5] | 1 | 3,5 | |

	Predicted: Spam	Predicted: Not spam
Actual: Spam	80	20
Actual: Not spam	10	90

- | Q.2(a) Construct a Decision Tree using ID3 algorithm with the data given: [5] | 2 | 6 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|------------|--------|--------|-------|-----------|----------|------|-----|---------|---------|------|-----|---------|----------|------|----|---------|------------|------|----|-----------|----------|--------|-----|-----------|---------|--------|----|---------|----------|--------|-----|--|--|
| <table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th style="width: 25%;">City</th> <th style="width: 25%;">Job</th> <th style="width: 25%;">Gender</th> <th style="width: 25%;">Class</th> </tr> </thead> <tbody> <tr><td>Bangalore</td><td>Employed</td><td>Male</td><td>Yes</td></tr> <tr><td>Chennai</td><td>Student</td><td>Male</td><td>Yes</td></tr> <tr><td>Chennai</td><td>Employed</td><td>Male</td><td>No</td></tr> <tr><td>Chennai</td><td>UnEmployed</td><td>Male</td><td>No</td></tr> <tr><td>Bangalore</td><td>Employed</td><td>Female</td><td>Yes</td></tr> <tr><td>Bangalore</td><td>Student</td><td>Female</td><td>No</td></tr> <tr><td>Chennai</td><td>Employed</td><td>Female</td><td>Yes</td></tr> </tbody> </table> | City | Job | Gender | Class | Bangalore | Employed | Male | Yes | Chennai | Student | Male | Yes | Chennai | Employed | Male | No | Chennai | UnEmployed | Male | No | Bangalore | Employed | Female | Yes | Bangalore | Student | Female | No | Chennai | Employed | Female | Yes | | |
| City | Job | Gender | Class | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Bangalore | Employed | Male | Yes | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Chennai | Student | Male | Yes | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Chennai | Employed | Male | No | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Chennai | UnEmployed | Male | No | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Bangalore | Employed | Female | Yes | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Bangalore | Student | Female | No | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Chennai | Employed | Female | Yes | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

- | | | |
|---|---|-----|
| Q.2(b) Considering Support Vector Machine concept, with the following: Positively labelled data points (2,2)(2,-2)(-2,-2)(-2,2) and Negatively labelled data points (1,1)(1,-1)(-1,-1)(-1,1). Given the Kernel function, find the separating hyperplane (i) [5] | 2 | 1,3 |
|---|---|-----|

$$\phi \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{cases} \begin{pmatrix} 4 - x_2 + |x_1 - x_2| \\ 4 - x_1 + |x_1 - x_2| \end{pmatrix}, & \text{if } \sqrt{x_1^2 + x_2^2} > 2 \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, & \text{Otherwise.} \end{cases}$$

- | | | |
|--|---|-----|
| Q.3(a) Consider a 3-layer MLP with a single hidden layer. The hidden layer has 2 neurons, and the output layer has 1 neuron. The input to hidden layer weight set is [0.2, 0.3; 0.4, 0.5] and hidden to output layer weight set is [0.2, 0.3]. The activation function used is ReLU. [8] | 3 | 2,3 |
| <p>(i) Given an input vector $x = [0.4, 0.2]$ and given weights, walk through the forward pass to calculate the output.</p> <p>(ii) Now, perform the backward pass (backpropagation) to calculate the gradients with respect to the weights using the mean squared error (MSE) loss function.</p> | | |

Q.3(b) Describe the role of activation functions in MLP. Why can't we use linear activation in all layers? [2] 3 2,4

Q.4(a) What is semi supervised learning? Explain its importance with any real-world application. [5] 4 2

Q.4(b) A dataset is assumed to come from a mixture of two coins: [5] 4 2,4

- Coin A has unknown probability of heads: θ_A
- Coin B has unknown probability of heads: θ_B
- Each trial uses coin A with probability π , coin B with probability $1 - \pi$

You observe the following incomplete data (you do NOT know which coin was used):

Trial	Heads (H)	Tails (T)
1	5	5
2	9	1
3	8	2

Initial parameter guesses:

$$\pi^{(0)} = 0.5, \quad \theta_A^{(0)} = 0.6, \quad \theta_B^{(0)} = 0.4$$

Perform ONE iteration of the EM algorithm and compute updated values of

$$\pi^{(1)}, \theta_A^{(1)}, \theta_B^{(1)}.$$

Q.5(a) What is Random Forest and How it can be used in supervised learning? Explain the reasons for underfitting and overfitting. Describe the techniques to reduce underfitting and overfitting. [5] 5 1,2

Q.5(b) Find the final classification for the data provided using stacking. The Metamodel (SVM) learns the pattern and provides the classification. [5] 5 2

Sample	LR	DT	RF	True Label
1	0	1	1	1
2	1	0	1	1
3	1	1	1	1
4	0	0	0	0
5	1	1	0	0

Weights are given as: $W_{LR} = 0.4, W_{RF} = 0.3, W_{DT} = 0.3$.