33

ISLAMIC UNIVERSITY OF TECHNOLOGY (IUT)

ORGANISATION OF ISLAMIC COOPERATION (OIC)

Department of Computer Science and Engineering (CSE)

MID SEMESTER EXAMINATION **WINTER SEMESTER, 2022-2023**

DURATION: 1 HOUR 30 MINUTES **FULL MARKS: 75**

CSE 4553: Machine Learning

Programmable calculators are not allowed. Do not write anything on the question paper.

Answer **all 3 (three)** questions. Figures in the right margin indicate full marks of questions whereas corresponding CO and PO are written within parentheses.

---

1. a) In the process of assessing loan applicants for eligibility, consider the decision-making process is based on text answers provided in a fixed questionnaire form. To determine the deserving candidates, which approach would you recommend: a rule-based approach or a machine-learning approach? Provide a comprehensive justification for your chosen approach. 7 (CO2) (PO1)

   b) Given training examples x^n, y^n where, $n = 1, 2, \ldots, N$, and $x, y \in \mathbb{R}$, a linear regression fit is: $y(x) = a + bx$. To determine the best parameters a, b we measure the discrepancy between the observed outputs and the linear regression fit function called Ordinary Least Squares (OLS) and minimizes the average vertical projects of the points y to the fitted line.

$$E(a, b) = \sum_{n=1}^{N} [y^n - y(x^n)]^2$$

   Answer the followings:

   i. If OLS is used as the optimizer for regression, what problems may arise in terms of outliers, overfitting, and model complexity? Explain. 7 (CO2) (PO2)

   ii. How can the L2 regularization (Ridge regression) and L1 regularization (Lasso regression) be applied in OLS optimizer and what are the regularization effects on the coefficients, overfitting, and feature selection in linear regression? 7 (CO2) (PO2)

   c) What is the use of the Sigmoid function in determining cross-entropy loss (log loss) of logistic regression? 4 (CO1) (PO1)

2. a) Consider a dataset in Table 1 to learn a decision tree that predicts if students pass a machine learning course (Yes (T) or No (F)), based on their previous GPA (High (H), Medium (M), or Low (L)) and whether or not they studied (Studied (T), not studied (F)). 10 (CO1) (PO1)

Table 1: Dataset for Question 2.

GPA	Studied	Passed
L	F	F
L	T	T
M	F	F
M	T	T
H	F	T
H	T	T

   i. What is the Entropy, $H(Passed)$?

   ii. What is the Entropy, $H(Passed \mid GPA)$?

iii. What is the Entropy, $H(Passed \mid Study)$?

iv. Draw the full decision tree that would be learned for this dataset.

b) Explain the performance of the decision tree for the following cases: 3×2 (CO2) (PO2)

i. Capturing linear relationships between features and the target variable

ii. Handling high-dimensional data

iii. The tree is deep and complex

c) Suppose Table 2 consists of the result of a binary classification model showing the true labels and the predicted probability scores. 9 (CO1) (PO1)

Table 2: Performance measure of a Model for Question 2.c)

Sample	True Label	Predicted Probability
1	Positive	0.8
2	Negative	0.2
3	Positive	0.6
4	Negative	0.4
5	Positive	0.7

i. Construct three confusion matrix of this binary classification model on three threshold values of 0.3, 0.5, and 0.7 and show True Positive Rate and False Positive Rate corresponds to the three threshold values.

ii. Draw the Receiver Operating Characteristic (ROC) curve at those three different threshold values.

3. a) Assume that you are a data scientist at a social media platform aiming to categorize user posts into topics like 'Technology', 'Sports', 'Entertainment' and 'Travel' based on their content. You have two model choices: a discriminative model like logistic regression that estimates topic probabilities directly and a generative model like Naive Bayes that models word-topic joint probabilities. Which model would you recommend for this text classification task? Justify your choice. 10 (CO2) (PO2)

b) Consider the dataset in Table 3 . It consists of five patient samples with different symptoms and class labels. 9 + 6 (CO1) (PO1)

Table 3: Dataset for Question 3.b)

Patient ID	Fever	Cough	Fatigue	Headache	Class
1	Yes	Yes	No	Yes	Positive
2	Yes	No	Yes	No	Negative
3	No	Yes	Yes	Yes	Positive
4	Yes	No	Yes	No	Negative
5	No	No	No	No	Negative

i. Construct a Bayesian network that represents all attributes, assuming that the predicting attributes are pairwise independent. Provide the probability table for each of the predicting attributes.

ii. Show how a Naïve Bayesian classifier would classify the following test sample:

Fever='Yes', Cough='No', Fatigue='Yes', Headache='Yes', Class=?