COMSATS University, Islamabad.

Department of Computer Sciences

Mid Term Examination FALL-2025

Subject: AIC354- Machine Learning Fundamental
Semester: BS(CS) –V, BS(AI)-IV,V, BDS-IV
Dated: 31st October, 2025

Marks: 25
Time: 90 Minutes
Instructor: Dr. Javed Iqbal, Dr. Gulmina Rextina, Mr. Umar Nauman

CLO1: Explain learning paradigms along with task and applications.

Question 1 (CLO 1) (5 Marks)

Classify the following tasks into supervised, unsupervised, or reinforcement learning and explain how?

a. Spam email detection
b. Customer segmentation
c. Game playing by AI agents
d. Credit score prediction
e. Self-driving car

Question 2 (CLO 1) (5 Marks)

A binary classification model was tested on 10 samples to predict whether an email is **Spam (1)** or **Not Spam (0)**.
The following table shows the **actual** and **predicted** labels:

Email	Actual (y)	Predicted (ŷ)
1	1	1
2	0	0
3	1	0
4	0	0
5	1	1
6	0	1
7	1	1
8	0	0
9	1	0
10	0	1

a) Construct the **Confusion Matrix**
b) Compute the following metrics (show your working):

i. **Accuracy**
ii. **Precision**
iii. **Recall**
iv. **F1-score**

CLO2: Apply supervised learning techniques to solve classification problems.

Question 3 (CLO 2) (5 Marks)

You're training a **KNN** model and a **linear regression (with gradient descent)** model on a housing dataset with features such as `LotSize` (1,000–50,000 sq ft), `NumBedrooms` (1–7), and `DistanceToCityCenter` (0.5–40 km). The KNN model's predictions seem dominated by `LotSize`, and gradient descent converges very slowly. What steps should you apply to fix these issues, and why would it help both models?

Question 4 (CLO 2) (5 Marks)

A data science team at a company is developing a machine learning model to predict **house prices** based on features such as size of the house, age of the house and so on. After collecting 1000 samples and training a polynomial regression model on it the team observed the following. What can you say about the performance of the model? Furthermore, how and what steps would you take to improve the performance?

Dataset	RMSE (Error)
Training Set	0.02
Test Set	0.85

Question 5 (CLO 2) (5 Marks)

A **fitness analytics company** uses the **K-Nearest Neighbors (KNN)** algorithm to classify users' workout sessions into: **Class A – Cardio Training and Class B – Strength Training**
Each session is characterized by two features collected from a smartwatch:

- x_1: Average heart rate (beats per minute)
- x_2: Average motion intensity (arbitrary scale)

Historical data for previously labeled sessions is shown below:

Point	x_1	x_2	Class
P1	2	4	A
P2	4	2	A
P3	4	4	B
P4	6	4	B

Now, two new workout sessions are recorded:

- **Session S_1:** ($x_1 = 5, x_2 = 3$)
- **Session S_2:** ($x_1 = 3, x_2 = 3$)

Answer the following questions:

1. **Apply KNN with K = 3 and Euclidean distance** to classify both new sessions.
2. **Discuss** how the classification might change if K = 5.
3. **Explain** how normalizing features or feature scaling could affect the classification decision.