Sri Sivasubramaniya Nadar College of Engineering, Chennai
(An autonomous Institution affiliated to Anna University)
Department of Computer Science and Engineering

CAT Assignment 1

Degree & Branch	M. Tech (Integrated) Computer Science & Engineering	Semester	V
Subject Code & Name	ICS1502 & Introduction to Machine Learning		
Academic year	2025-2026 (Odd)	Batch:2023-2028	**Due date:28-08-2025**

- **Instructions:** You are expected to complete the assignment independently. While you are free to engage in discussions with your peers, do not copy from one another. If we identify submissions from two or more students that are significantly similar in content, all students involved will receive zero marks for this assignment. All submissions will be checked for plagiarism using standard tools. Please adhere to the following submission guidelines:

  - Use Python for all coding tasks.
  - Submit your code in the form of a Jupyter Notebook (.ipynb).
  - Submit your answers to the theory questions in a separate document (PDF or Word format).

1 Programming Assignment

1.1 Regression

Mobile Phone Price Prediction **dataset**. Critically evaluate the implementation of Linear Regression using the following methods. Use the matrices approach to examine the suitability of this technique for price prediction. Divide the data into train and test sets and assess how this split impacts the learning process. Use training samples to construct the data matrix and label matrix, and justify their configuration. Based on the dimension of the data, assume an appropriate parameter matrix and interpret its influence on the prediction outcome. Use error analysis and performance measures to evaluate the accuracy of your predictions and determine the effectiveness of the matrix-based approach in modeling mobile phone prices [CO2, K5, 2.3.1, 3.4.3, 5.1.2, 6.2.1, 8.2.1 10.1.1,10.3.1, 10.3.2, 12.2.2].

1. Using closed-form solution (**4 marks**)   [10.1.1, 2.3.1]

2. Using gradient descent (**4 marks**)   [5.1.2, 10.3.2]

3. Plot predicted vs. actual values on test data for (1) and (2) (**2 marks**)   [3.4.3]

4. Repeat (1), (2), and (3) with l_2 regularization (**3 marks**)   [10.3.1, 10.3.2]

5. Compare results in (4) with and without data standardization (**3 marks**)   [6.2.1]

6. Plot predicted vs. actual on test data for different λ values in (4) (**2 marks**)   [8.2.1]

7. Analyze feature importance from l_2 weights (**2 marks**)   [12.2.2]

1.2 Linear Classification

- Fit a Linear Classification Model or any single-layered or multi-layered neural network (2-class classification model) using the given dataset: Bank Note Authentication **dataset** Evaluate the suitability and performance of a linear classification model for solving this binary classification problem [CO2, CO3, K5, 2.3.1, 3.4.3, 5.1.2, 6.2.1, 8.2.1 10.1.1,10.3.1, 10.3.2, 12.2.2].

  1. Divide dataset into train and test sets (**3 mark**)   [2.3.1]
  2. Fit classification model with and without l_2 regularization; compare accuracies (**6 marks**) [10.3.1, 10.3.2]
  3. Plot training and test accuracy vs. λ (**3 marks**)   [8.2.1]
  4. Visualize 3D plot using three important features (**3 mark**)   [3.4.3]
  5. Intentionally introduce outliers by shifting data points (**3 mark**)   [5.1.2]
  6. Fit classifier on outlier-injected data and comment on its impact (**2 marks**)   [6.2.1, 12.2.2]

Evaluation scheme:

Metrics	Marks
Design & Coding	15
Executing all Test cases	15
Report	10