Birla Institute of Technology & Science, Pilani
Work Integrated Learning Program Division
First Semester 2025-2026

Assignment 2

Course No.	:	AIMLC ZC416/ DSECLZC416
Course Title	:	Mathematical Foundations for Machine Language/ Mathematical Foundations for Data Science
Weightage	:	10%

Instructions to Students:

1. All parts of a question should be answered consecutively.
2. You must submit the scanned copy of your ***hand written assignment in pdf format.*** Assignments containing types answers will not be accepted.
3. The last date to submit the Assignment is 07-02-2026. No extension will be given after that.

Question:

An electricity utility company wants to predict the weekly electricity consumption (in kWh) of small commercial shops based on their operating hours per week. A linear regression model with bias is assumed:

$$\hat{y}_i = w_0 + w_1 x_i$$

The following data are collected:

Shop Number	Operating Hours x	Consumption y
1	1	3
2	2	5
3	4	9
4	6	13
5	8	17

The model parameters are learned by minimizing the Squared Error loss function $J(w_0, w_1) = \frac{1}{2}\sum_{i=1}^{5}(\hat{y}_i - y_i)^2$ using gradient descent.

(a) Formulate the Squared Error loss function $J(w_0, w_1)$ explicitly for the given dataset.
Compute the Hessian matrix of J with respect to w_0 and w_1, and use it to prove that J is a convex function of the parameters (w_0, w_1).

(b) Compute the gradient of the Squared Error loss function: $\nabla J = \begin{bmatrix} \frac{\partial J}{\partial w_0} \\ \frac{\partial J}{\partial w_1} \end{bmatrix}$.

For problems (c) and (e), use the initial parameter values $w_0^{(0)} = 2,\ w_1^{(0)} = 2$.

(c) Perform two iterations of gradient descent using:

1. A constant learning rate $\eta = 0.05$

2. An exponentially decaying learning rate

$\eta_t = \eta_0 e^{-kt}, \eta_0 = 0.1$, k=0.4, where t=1, 2,.. denotes the iteration index.

Show the updated parameter values after each iteration.

(d) Comment on whether the learning-rate decay used in Part (c) improves the stability and convergence behaviour compared to the constant learning rate.

(e) For a fixed descent direction $d = -\nabla J(w)$ at the current iterate w, the optimal step size η is chosen by minimizing the one-dimensional function

$$\phi(\eta) = J(w + \eta d)$$

Explain how the step size η can be selected using

(i) Binary search: Use binary search with the initial interval $[0, 1]$.
At each iteration, compare the function values $\phi(m)$ and $\phi(m + 10^{-3})$, where $m$is the midpoint of the current interval.
Perform two iterations and show the reduced interval after each iteration.

(ii) Golden-Section Search: Use the Golden-Section Search with the initial interval $[0, 1]$, and take the interior points as $M_1 = \frac{1}{4}$and $M_2 = \frac{3}{4}$ (for the notation M_1 and M_2 refer the slides). Determine the reduced interval in the first iteration. For the second iteration, specify any valid choice of $M_1$and M_2within the reduced interval.
