Homework #1, CS6140 Fall 2025*(Due Friday 19, and 11:59pm PST)*

Problem: Regression with gradient descent and the closed-form solution

Given the dataset data.csv, use regression to find the relation between the features and the target.

Tasks

1. Load the data (10 pts)

- Use pandas to read data.csv into a DataFrame.
- Display the first five rows to confirm correct loading.

2. Visualize the data (10 pts)

- Create a scatterplot of x vs. y using matplotlib.
- Based on the plot, briefly describe what type of relationship seems appropriate.

3. Construction of the Feature Matrix (20 pts)

- Build a NumPy matrix Φ for a regression that can model the relationship you observed in point 2.
- Include a bias (column of 1's) and explain the columns you chose.

4. Closed-Form Solution (35 pts)

- Derive the normal equation $w = (\Phi^T \Phi)^{-1} \Phi^T y$ for your model.
- Implement this equation in Python (use np.linalg.pinv for stability).
- Print the weights.
- Overlay the fitted curve on the scatterplot.

5. Gradient descent (15 pts)

- Compute the Mean Squared Error (MSE) as the loss (use gradient descent)
- Loss = (1/m) Σ (y_i - ŷ_i)^2.
- Report this loss value and comment briefly on the quality of the fit.

6. Discussion (10 pts)

- Explain why the regression you chose was appropriate for this dataset.
- Discuss any potential limitations or sources of error.

7. What you should deliver for this HW:

- Colab/VS etc. notebook (.ipynb) containing all Python code with explanatory comments
- Plots for data visualization and the final fit
- Include markdown (text box) when you see it fit as well