

Linear Regression

The Education Board wants to develop a predictive model to estimate **students' final exam scores** based on various **academic and behavioral factors** such as study hours, attendance, parental education level, and test preparation.

You are provided with the dataset **student_performance.csv**, which contains information about students' demographics, study habits, and exam results.

| Column | Description |
|-----------------------------|---|
| gender | Gender of the student |
| race/ethnicity | Group classification of the student |
| parental_level_of_education | Highest education level achieved by the student's parent(s) |
| lunch | Type of lunch (standard/free-reduced) |
| test_preparation_course | Whether the student completed a test preparation course |
| math_score | Marks obtained in mathematics |
| reading_score | Marks obtained in reading |
| writing_score | Marks obtained in writing |

| Column | Description |
|---------------|---|
| average_score | Average of math, reading, and writing scores (<i>Target variable</i>) |

- **Your Task:**

Your task is to build a Linear Regression model from scratch using NumPy to predict the average_score of students based on other academic and categorical factors.

- **What to Do**
 - **Load and explore the dataset**
 - **Preprocess the data — handle categorical variables and normalize numeric columns.**
 - **Implement Linear Regression from scratch (do not use sklearn).**
 - **Train your model to predict average_score (target variable).**
 - **Visualize the regression results or residuals.**
-
- **Allowed Libraries :**
 - **Numpy**
 - **Pandas**
 - **Matplotlib**

Logistic Regression

The National Socioeconomic Research Institute (NSRI) aims to develop a predictive model to determine whether an individual earns **more than \$50,000** per year based on their demographic and employment attributes.

You are provided with the dataset `adult.csv`, which contains census data including age, education, occupation, and other socioeconomic indicators.

Your task is to implement Logistic Regression from scratch using NumPy to predict whether a person earns >50K or ≤50K annually. You will preprocess the data, train your model, visualize results, and evaluate its performance.

| Column | Description |
|---------------|--|
| age | Age of the individual |
| workclass | Type of employment (e.g., Private, Self-emp, Govt) |
| fnlwgt | Final weight assigned to the record |
| education | Highest level of education attained |
| education-num | Numeric representation of education level |

| Column | Description |
|-----------------------|--|
| marital-status | Marital status of the individual |
| occupation | Type of occupation |
| relationship | Relationship status (Husband, Wife, etc.) |
| race | Race of the individual |
| sex | Gender of the individual |
| capital-gain | Capital gains from investments |
| capital-loss | Capital losses from investments |
| hours-per-week | Average working hours per week |
| native-country | Country of origin |
| income | Target variable ($>50K = 1$, $\leq 50K = 0$) |

• What to Do

- Implement Logistic Regression from scratch (do not use sklearn or other ML libraries).
- Start by loading and preprocessing the dataset
- Train your Logistic Regression model and visualize important aspects (such as loss over iterations or feature relationships)

- **Output the accuracy your model achieves on the test set.**

- **Allowed libraries:**
 - **Numpy**
 - **Pandas**
 - **Matplotlib**