

Product Requirements Document: Multiple Linear Regression Simulation

1. Introduction

This document outlines the product requirements for a Python script that simulates multiple linear regression. The script will generate a dataset, perform linear regression to model the data, and visualize the results to compare the actual and predicted values.

2. Functional Requirements

2.1. Data Generation

- The script must generate a synthetic dataset with a configurable number of samples and features.
- The dataset will be generated based on a linear equation with known coefficients and a random error term.
- The user shall be able to set a seed for the random number generator to ensure reproducibility.

2.2. Linear Regression Analysis

- The script must perform multiple linear regression on the generated dataset.
- It shall use the Normal Equation method to estimate the model coefficients, including an intercept term.
- The script will use the estimated coefficients to make predictions on the input data.

2.3. Visualization

- The script must generate a scatter plot to visualize the relationship between the actual and predicted output values.
- The plot must include an ideal regression line (a 45-degree line) for reference.
- The plot must display the original and estimated regression equations for comparison.
- The plot must have a clear title, axis labels, and a legend.
- The generated plot must be saved as a PNG file.

3. Non-Functional Requirements

3.1. Dependencies

- The script will rely on the following Python libraries:
 - `numpy` for numerical operations.
 - `matplotlib` for data visualization.

3.2. Output

- The script will save the visualization as a PNG image file named `simulation6.png` in the `Documentation` directory.
- The script will print a confirmation message to the console upon successful completion.

4. Future Enhancements (Optional)

- Allowing the user to input their own dataset.
- Implementing other regression techniques for comparison (e.g., Ridge, Lasso).
- Adding more detailed model evaluation metrics (e.g., R-squared, Mean Squared Error).