

Visualizing SAT and GPA Scores and Performing Line Search

Objective: To understand the relationship between SAT scores and GPA, and to determine the best-fit line using the least squares method.

Procedure:

1. Data Loading and Visualization:
 - a. Load the CSV file into your programming environment.
 - b. Create a scatter plot with SAT scores on the x-axis and GPA scores on the y-axis. Label the axes appropriately.
 - c. Display the scatter plot to visualize the distribution of scores.
2. Initial Line Plotting:
 - a. Choose initial values for $GPA = \alpha \cdot SAT + \beta$
 - b. For simplicity, you can start with
 - c. Using the equation $GPA = \alpha \cdot SAT + \beta$, compute the GPA values for each SAT score in your dataset.
 - d. Plot this initial line on the same scatter plot to visualize how well it fits the data.
3. Cost Function and Line Search:
 - a. Define the cost function as the mean squared error (MSE) between the predicted GPA values (using the line) and the actual GPA values from the dataset.
 - b. Initialize a learning rate (e.g., $\alpha = 0.0001$) and a set number of iterations (e.g., 1000).
 - c. Perform a line search for both α and β .
 - d. Store the cost function values for each α and β .

Visualizing the Best-Fit Line:

- a. Determine the best values of α and β .
- b. For α and β obtained from the line search, compute the predicted GPA values for each SAT score in your dataset.
- c. Plot this best-fit line on the same scatter plot.
- d. Display the scatter plot with the best-fit line to visualize how well it fits the data.