Linear Regression: Fundamentals, Loss, and Optimization

Naman Goel

October 9, 2025

Abstract

This report presents an extensive introduction to linear regression, one of the most fundamental methods in statistics and machine learning. Key aspects discussed include the mathematical foundation, visualization of regression, the concept and formula for mean squared error (MSE) as the cost function, and the optimization process through gradient descent. Images and plots are included to aid understanding.

1 Introduction

Linear regression is a statistical approach to modeling the linear relationship between a dependent variable and one or more independent variables [1, 2]. Its primary purpose is to understand, quantify, and predict trends in data using a simple yet powerful mathematical framework. This model is a starting point for many complex statistical and machine learning methods.

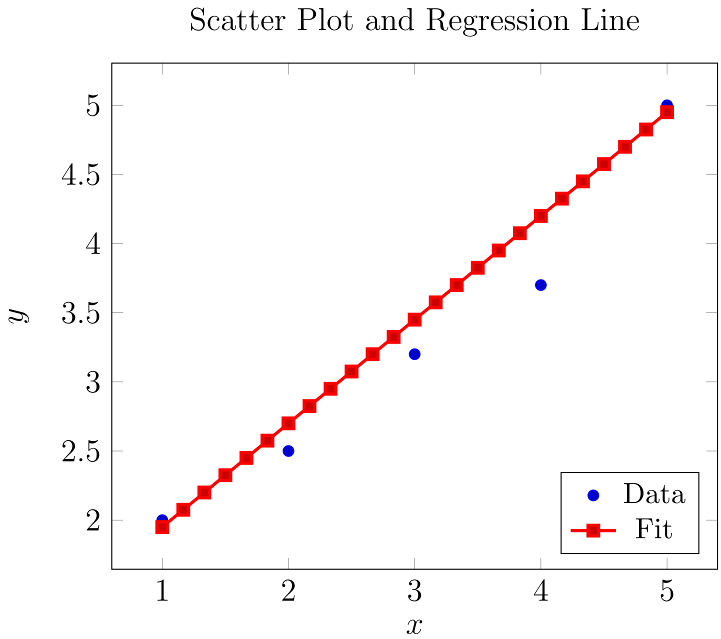

Figure 1: Scatter plot of data with best-fit regression line (red).

2 Mathematical Foundation

The general form of a simple linear regression model is:

$$y = wx + b$$

where y is the dependent variable, x is the independent variable, w (weight/slope) and b (bias/intercept) are the parameters learned from the data [1]. The objective is to determine the parameters that yield the closest possible prediction to the true data values.

3 Visualization in Linear Regression

Visualization helps to understand the quality of fit and the magnitude of prediction errors (residuals).

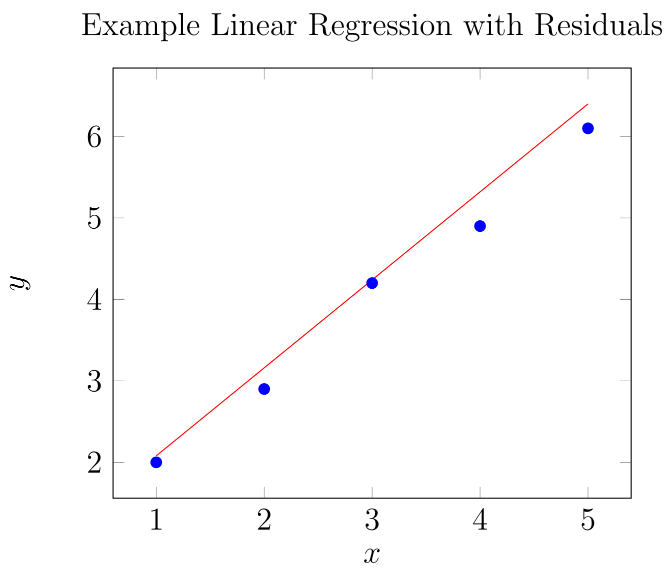

Figure 2: Synthetic data and the best-fit linear regression line.

4 Mean Squared Error (MSE) Loss Function

A key aspect of linear regression is quantifying prediction error. The most common metric used is the *mean squared error*:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

where n is the number of data points, y_i is the actual value, and $\hat{y}_i$ is the predicted value from the regression line [5, 3]. Minimizing MSE ensures the "best" fit for the data in terms of least squares error.

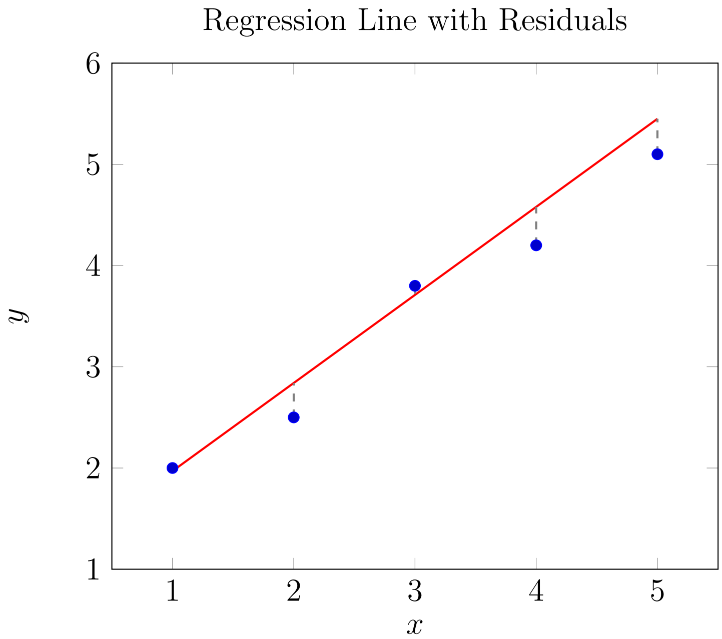

Figure 3: Regression line (red), data points (blue), and residuals (vertical gray dashed lines).

5 Gradient Descent for Parameter Optimization

Finding the optimal parameters (w, b) is usually done by minimizing the MSE cost function. While closed-form solutions exist, gradient descent is a popular iterative optimization technique, especially for large datasets:

5.1 Algorithm Steps

1. Initialize weights and biases (often randomly).
2. Compute the gradient of the loss (MSE) with respect to each parameter.
3. Update parameters in the direction of negative gradient, proportionally to the learning rate α:

$$w \leftarrow w - \alpha \frac{\partial \text{MSE}}{\partial w}$$

$$b \leftarrow b - \alpha \frac{\partial \text{MSE}}{\partial b}$$

4. Iterate until convergence (loss no longer decreases significantly) [4, 6].

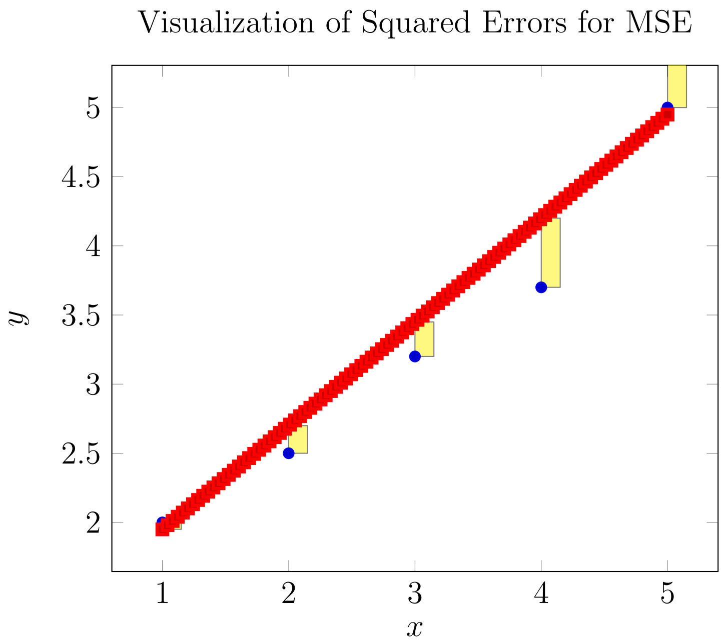

Figure 4: Squares (yellow) show squared error for calculation of MSE.

6 Conclusion

Linear regression is foundational in statistics and machine learning, providing a simple yet effective means of modeling relationships in data. Its strength lies in mathematical clarity and ease of computation when combined with error minimization (MSE) and robust optimization (gradient descent).

References

[1] Linear Regression. *Wikipedia.* `https://en.wikipedia.org/wiki/Linear_regression`

[2] An Introduction to Linear Regression. *Towards Data Science.* `https://towardsdatascience.com/an-introduction-to-linear-regression-9cbb64b52d23/`

[3] Mean Squared Error: Definition and Example. *StatisticsHowTo*. https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/mean-squared-error/

[4] Linear Regression: Gradient Descent. *Google Machine Learning Crash Course*. https://developers.google.com/machine-learning/crash-course/linear-regression/gradient-descent

[5] Mean Squared Error (MSE). *Statistics By Jim*. https://statisticsbyjim.com/regression/mean-squared-error-mse/

[6] Gradient Descent in Linear Regression. *GeeksforGeeks*. https://www.geeksforgeeks.org/machine-learning/gradient-descent-in-linear-regression/