Linear Regression by Numerical Optimisation

leta199

10 October 2025

1 Setup of Theory

Numerical Optimisation

Line equation:

$$\hat{y} = m\hat{x} + b$$

- $\hat{y}$: predicted dependent variable
- $\hat{x}$: predicted independent variable

Error in regression:

$$e_i = y_i - \hat{y}_i = y_i - (m\hat{x}_i + b)$$

Total error:

$$\sum_{i=1}^{n} (y_i - (m\hat{x}_i + b))$$

Mean Squared Error (MSE):

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} [y_i - (m\hat{x}_i + b)]^2$$

We must minimise the MSE to determine the line of best fit.

2 Gradient Descent

We treat the MSE as a function:

$$f(m, b) = \frac{1}{n} \sum_{i=1}^{n} [y_i - (m\hat{x}_i + b)]^2$$

To minimise this function, we take partial derivatives with respect to m and b:

$$\frac{\partial f}{\partial m} = -\frac{2}{n}\sum_{i=1}^{n} \hat{x}_i \left[y_i - (m\hat{x}_i + b)\right]$$

$$\frac{\partial f}{\partial b} = -\frac{2}{n}\sum_{i=1}^{n} \left[y_i - (m\hat{x}_i + b)\right]$$

To find the optimal m and b, start with initial (e.g., random) values and update them using gradient descent iteratively by taking steps (L) in the negative direction of maximum gradient:

$$m_{\text{new}} = m - L \cdot \frac{\partial f}{\partial m}$$

$$b_{\text{new}} = b - L \cdot \frac{\partial f}{\partial b}$$

Where L is the learning rate (step size).