Least Squares Derivation - Plain Equations

We start with the cost function used in linear regression:

$$E = \sum (y_i - (a * x_i + b))^2$$

Our goal is to minimize the error E by finding the values of a and b.

Step 1: Compute Partial Derivatives

$$dE/da = -2 * \sum [x_i * (y_i - (a * x_i + b))]$$

$$dE/db = -2 * \sum [y_i - (a * x_i + b)]$$

Step 2: Set the derivatives to zero to minimize E

$$\sum [x_i * (y_i - a * x_i - b)] = 0$$

$$\sum [y_i - a * x_i - b] = 0$$

Step 3: Expand both equations

Let:

$S_x = \sum x_i$

$S_y = \sum y_i$

$S_{xx} = \sum x_i^2$

$S_{xy} = \sum x_i * y_i$

N = number of data points

Then the two equations become:

$$S_{xy} - a * S_{xx} - b * S_x = 0 \qquad \text{(Equation 1)}$$

$$S_y - a * S_x - b * N = 0 \qquad \text{(Equation 2)}$$

Step 4: Solve Equation 2 for b

$$b = (S_y - a * S_x) / N$$

Step 5: Substitute this expression for b into Equation 1 and solve for a

$$a = (N * S_{xy} - S_x * S_y) / (N * S_{xx} - S_x^2)$$

Step 6: Plug a back into the equation for b

$$b = (S_y - a * S_x) / N$$

Final expressions to compute a and b (Least Squares Solution):

$$a = (N * \text{sum}(x_i * y_i) - \text{sum}(x_i) * \text{sum}(y_i)) / (N * \text{sum}(x_i^2) - (\text{sum}(x_i))^2)$$

$$b = (\text{sum}(y_i) - a * \text{sum}(x_i)) / N$$