Solving the Regression Problem Step by Step

Step 1: Loss Function (Squared Error)

$$L = \sum_i (\hat{y}_i - y_i)^2 = \sum_i (X_{ij} w_j - y_i)^2$$

The Einstein summation convention implies a sum over repeated indices j.

Step 2: Take the Derivative with Respect to w_k

Vector Notation (Scarry and IMO error-prone)

$$\nabla_w L = \nabla_w \sum_i (Xw - y)$$

We apply the chain rule.

$$\frac{\partial L}{\partial w_k} = 2 \sum_i (X_{ij} w_j - y_i) \underbrace{\frac{\partial}{\partial w_k} (X_{ij} w_j - y_i)}_{\delta_{kj} X_{ij}}$$

Solving the Regression Problem Step by Step

Step 3: Set the Gradient Equal to Zero

$$\frac{\partial L}{\partial w_k} = 0 \implies 2(X_{ij}w_j - y_i)X_{ik} = 0$$

$$X_{ik}X_{ij}w_j = X_{ik}y_i$$
$$(X^TX)_{kj}w_j = (X^Ty)_k$$

Step 4: Solve for the Weight Vector

$$w = (X^TX)^{-1}X^Ty$$

This is the final formula for the weight vector w, derived by solving the normal equation.