Step 4: Combining the Derivatives

Now, we combine the derivatives of each term:

$$\frac{\partial}{\partial \boldsymbol{\beta}} RSS = -2\mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}$$

Step 5: Set the Derivative Equal to Zero

To find the $\boldsymbol{\beta}$ that minimizes the RSS, we set the derivative equal to zero:

$$-2\mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = 0$$

Step 6: Solve for $\boldsymbol{\beta}$

Now, we can solve for $\boldsymbol{\beta}$. First, cancel out the 2 from both sides:

$$-\mathbf{X}^T\mathbf{y} + \mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = 0$$

Next, solve for $\boldsymbol{\beta}$:

$$\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^T\mathbf{y}$$

Finally, if $\mathbf{X}^T\mathbf{X}$ is invertible, multiply both sides by $(\mathbf{X}^T\mathbf{X})^{-1}$:

$$\boldsymbol{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

Summary of the Derivative Rules Used

- **For a scalar term** (independent of $\boldsymbol{\beta}$): The derivative is zero.
- **For a linear term $\mathbf{a}^T\mathbf{b}$**: The derivative with respect to $\mathbf{b}$ is $\mathbf{a}$.