Step 3: Taking the Derivative with Respect to β

Now we differentiate the above expression with respect to the coefficient vector $\boldsymbol{\beta}$. We will apply matrix calculus rules.

Derivative of $\mathbf{y}^T\mathbf{y}$:

This term is independent of $\boldsymbol{\beta}$, so its derivative with respect to $\boldsymbol{\beta}$ is zero.

$$\frac{\partial}{\partial \boldsymbol{\beta}}(\mathbf{y}^T\mathbf{y}) = 0$$

Derivative of $-2\mathbf{y}^T\mathbf{X}\boldsymbol{\beta}$:

This term is linear in $\boldsymbol{\beta}$. The derivative of a linear term $\mathbf{a}^T\mathbf{b}$ with respect to $\mathbf{b}$ is just $\mathbf{a}$. In this case, $\mathbf{a} = -2\mathbf{X}^T\mathbf{y}$, so the derivative is:

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

Derivative of $\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}$:

This term is quadratic in $\boldsymbol{\beta}$. The derivative of a quadratic form $\boldsymbol{\beta}^T\mathbf{A}\boldsymbol{\beta}$ with respect to $\boldsymbol{\beta}$ is $2\mathbf{A}\boldsymbol{\beta}$ if $\mathbf{A}$ is symmetric. In this case, $\mathbf{A} = \mathbf{X}^T\mathbf{X}$, which is symmetric. So the derivative is:

$$\frac{\partial}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}) = 2\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}$$