Source task and conventions

Prove the augmentation identity and its computational implications.

Conventional solution and proof

For centred data and $\lambda \geq 0$, define $\widetilde{X} = \begin{pmatrix} X \\ \sqrt{\lambda} I_p \end{pmatrix}$ and $\widetilde{y} = \begin{pmatrix} y \\ 0 \end{pmatrix}$. Then

$$\|\widetilde{y} - \widetilde{X}\beta\|^2 = \|y - X\beta\|^2 + \lambda\|\beta\|^2.$$

The objectives agree for every coefficient vector, so their minimizer sets agree. If $\lambda > 0$, $\widetilde{X}$ has full column rank, and its normal equations are $(X^\top X + \lambda I)\beta = X^\top y$. If $\lambda = 0$, augmentation adds zero rows and does not remove any nonuniqueness. An unpenalized intercept must be excluded from the identity block, or eliminated by centring.

Structural or analytic derivation

The artificial observations encode a quadratic penalty as Euclidean residuals. Orthogonal QR on the augmented matrix computes the same solution without explicitly forming $X^\top X$, avoiding the condition-number squaring of normal equations.

Comparative cost and preferred method

Naively storing the added rows costs p^2 entries; structured implementations can account for the identity without materializing it. QR costs $O((N + p)p^2)$ classically. A dual solve is preferable when $p \gg N$, while augmentation can improve numerical robustness in other regimes. Objective equality, not a universal speedup, is the proved result.