Source task and conventions

Derive the centred ridge and lasso formulations and their coefficient correspondence.

Conventional solution and proof

Write $Z_{ij} = x_{ij} - \bar{x}_j$. The bijection between parameters is $\beta_j^c = \beta_j$ for slopes and $\beta_0^c = \beta_0 + \sum_j \bar{x}_j \beta_j$. It gives the same fitted value $\beta_0 + x_i^\top \beta = \beta_0^c + z_i^\top \beta^c$ and leaves the slope penalty unchanged. Because every centred column sums to zero, differentiating with respect to the unpenalized intercept yields $\widehat{\beta}_0^c = \bar{y}$. The ridge slopes solve

$$(Z^\top Z + \lambda I)\widehat{\beta}^c = Z^\top (y - \bar{y}\mathbf{1}),$$

so for $\lambda > 0$ the unique slopes are its inverse applied to the right side. The original intercept is $\bar{y} - \bar{x}^\top \widehat{\beta}^c$. Replacing the squared penalty by $\lambda \sum_j |\beta_j|$ preserves the same bijection and the same intercept calculation. The lasso slopes are the minimizers of the centred penalized criterion; they need not be unique without further design assumptions. Penalizing the intercept would invalidate this equivalence as stated.

Structural or analytic derivation

The centring matrix J is the orthogonal projector off the constant vector. Eliminating the intercept gives $\min_\beta \|Jy - JX\beta\|^2 + \lambda P(\beta)$ with the original slope penalty P. This is an exact elimination of a nuisance coordinate.

Comparative cost and preferred method