Source task and conventions

Derive every entry of Table 3.4, including ties and penalty normalization.

Conventional solution and proof

Assume $X^\top X = I$ and set $b = X^\top y$, the unrestricted least-squares coefficients. Orthogonal decomposition gives

$$\|y - X\beta\|^2 = \|y - Xb\|^2 + \sum_j (\beta_j - b_j)^2.$$

For best subset of size at most M, fixing a support makes $\beta_j = b_j$ on that support and zero elsewhere. The residual reduction is $\sum_{j\in S} b_j^2$, maximized by selecting M largest absolute coefficients. At a boundary tie, choose any required number of tied coordinates; the table's indicator with $\geq$ can include too many unless ties are resolved.

For ridge with objective SSE$+\lambda\sum_j \beta_j^2$, each coordinate minimizes $(\beta_j - b_j)^2 + \lambda\beta_j^2$, giving $\widehat{\beta}_j = b_j/(1+\lambda)$ by setting its derivative to zero; the second derivative is positive.

For the lasso normalization $\frac{1}{2}\text{SSE} + \lambda\sum_j |\beta_j|$, a positive solution has derivative $\beta_j - b_j + \lambda = 0$, and a negative solution has derivative $\beta_j - b_j - \lambda = 0$. At zero the subgradient condition is $|b_j| \leq \lambda$. Thus

$$\widehat{\beta}_j = \text{sign}(b_j)(|b_j| - \lambda)_+.$$

These three cases exhaust the convex one-dimensional problem. With SSE$+\lambda\|\beta\|_1$ the threshold is $\lambda/2$; Table 3.4 absorbs this scaling into its tuning parameter.

Structural or analytic derivation