Conventional solution and proof

Assume $\lambda > 0$ and centred one-predictor data. Put $s = X^\top X$ and $c = X^\top y$, so the original coefficient is $a = c/(s + \lambda)$. With two identical columns, strict convexity and exchange symmetry force $b_1 = b_2 = b$. Differentiation gives $s(2b) - c + \lambda b = 0$, hence

$$\boxed{b = \frac{c}{2s + \lambda} = a\frac{s + \lambda}{2s + \lambda}.}$$

It is generally not $a/2$. More generally, for m duplicate coefficients with aggregate $t = \sum_j b_j$, the prediction depends only on t and

$$\sum_j b_j^2 \geq t^2/m,$$

by Cauchy–Schwarz, with equality exactly when every $b_j = t/m$. This remains true in the presence of other predictors because their coefficients can be held fixed in this argument. Every ridge minimizer therefore has equal duplicate coefficients. In the one-variable case $b_j = c/(ms + \lambda)$ and $t = c/(s + \lambda/m)$. At $\lambda = 0$, equality is a minimum-norm choice but is not required of every least-squares minimizer.

Structural or analytic derivation

Duplicating a feature divides its effective quadratic penalty by the duplication count when aggregate prediction is held fixed. The norm-minimization identity gives the result directly and identifies why duplicate columns change the fitted aggregate.

Comparative cost and preferred method