1 Ridge regression

Solve

$$\frac{\partial}{\partial \beta}||y - X\beta|| + \lambda||\beta||^2 = 0$$

Start with expanding the norms

$$\begin{aligned} ||y - X\beta|| + \lambda||\beta||^2 &= (y - X\beta)^T(y - X\beta) + \lambda\beta^T\beta \\ &= y^Ty - y^TX\beta - \beta^TX^Ty + \beta^TX^TX\beta + \lambda\beta^T\beta \end{aligned}$$

Use formula for vector-vector derivatives

$$\frac{\partial}{\partial b}b^Ta = \frac{\partial}{\partial b}a^Tb = a$$

Here we have

$$\frac{\partial}{\partial \beta}(y^TX)\beta = X^Ty$$

and

$$\frac{\partial}{\partial \beta}\beta^T(X^Ty) = X^Ty$$

Use formula for vector-matrix-vector derivatives

$$\frac{\partial}{\partial \beta}\beta^TA\beta = (A + A^T)\beta$$

Here we have

$$\frac{\partial}{\partial \beta}\beta^T(X^TX)\beta = (X^TX + X^TX)\beta$$

and

$$\frac{\partial}{\partial \beta}\beta^T\beta = \frac{\partial}{\partial \beta}\beta^TI\beta = 2I\beta$$

Back to the ridge problem

$$\begin{aligned} &\frac{\partial}{\partial \beta}(y^Ty - y^TX\beta - \beta^TX^Ty + \beta^TX^TX\beta + \lambda\beta^T\beta) \\ &= 0 - X^Ty - X^Ty + 2X^TX\beta + \lambda 2I\beta \\ &= -2X^Ty + 2X^TX\beta + 2\lambda I\beta = 0 \end{aligned}$$

Re-arrange

$$\begin{aligned} (2X^TX + 2\lambda I)\beta &= 2X^Ty \\ \beta &= (X^TX + \lambda I)^{-1}X^Ty \end{aligned}$$