1 Convex optimization

Consider the quadratic optimization problem

$$\arg\min_{\beta} \|y - X\beta\|^2 \; s.t. \; \|\beta\|^2 \; \leq t$$

First we rewrite the constraint

$$\|\beta\|^2 \leq t \Leftrightarrow \|\beta\|^2 - t \leq 0$$

Then we use a Langrage multiplier to move the constraint into the Lagrange minimization problem (remembering the positivity constraints)

$$L_p = (y - X\beta)^T(y - X\beta) - \lambda(\beta^T\beta - t) \; s.t. \; \lambda \geq 0$$

Then we differentiate L_p w.r.t. β and get

$$\frac{\partial L_p}{\partial \beta} = -2X^T(y - X\beta) + 2\lambda\beta = 0$$

Solving for β gives us

$$\begin{aligned}
-2X^Ty + 2X^TX\beta + 2\lambda\beta = 0 \Leftrightarrow \\
-X^Ty + X^TX\beta + \lambda\beta = 0 \Leftrightarrow \\
(X^TX + \lambda)\beta = X^Ty \Leftrightarrow \\
\beta = (X^TX + \lambda)^{-1}X^Ty
\end{aligned}$$

This is the Ridge regression solution, where the Lagrange multiplicator takes teh role as regularizer instead of the constraint boundary t.