Ridge regression solution :

① $$\underbrace{\| Y - X\beta \|_2^2}_{\text{RSS}} + \lambda \|\beta\|_2^2 = \underbrace{(Y - X\beta)^T (Y - X\beta)}_{Y^T Y - 2Y^T X\beta + \beta^T X^T X \cdot \beta} + \underbrace{\lambda \cdot \beta^T \beta}_{+ \lambda \cdot \beta^T G \beta} = f(\beta$$

Let's use a more general penalty : $\lambda \cdot \beta^T G \beta$, where G is ~~some~~ an appropriate penalty matrix ; e.g., $G = I$ in ridge reg.
If G is ~~of~~ symmetric pd, or X is of full rank, solution is unique.

$$\nabla_\beta f(\beta) = -2X^T Y + 2X^T X\beta + 2\lambda G\beta \overset{\text{set}}{=} 0$$

$$(X^T X + \lambda G) \cdot \beta = X^T Y \implies \hat{\beta}_\lambda = (X^T X + \lambda G)^{-1} \cdot X^T Y.$$

Check 2nd order conditions - skipped here.

②

② $f(\beta) = \|Y - X\beta\|_2^2 = Y^T Y - 2Y^T X\beta + \beta^T X^T X\beta$.

Want: $\min_\beta f(\beta)$ subject to $\beta^T\beta \le s$; $\Leftrightarrow \beta^T\beta - s \le 0$

Construct a Lagrangian obj. fcn: $f(\beta,\gamma) = \|Y - X\beta\|_2^2 + \underbrace{\gamma \cdot (\|\beta\|_2^2 - s)}_{\text{Lagrange penalty}}$

Now, minimize $f(\beta,\gamma)$ wrt β, γ jointly.

$$\nabla_\beta f(\beta,\gamma) = -2X^T Y + 2X^T X\beta + 2\gamma \cdot I \cdot \beta = 0 \quad \text{eq (a)}$$

$$\nabla_\gamma f(\beta,\gamma) = \beta^T\beta - s = 0 \quad \text{eq (b)}$$

From eq. (a) : $\hat{\beta}_\gamma = (X^T X + \gamma \cdot I)^{-1} \cdot X^T \cdot Y$

eq. (b) : $\beta^T\beta = s$.

For a general value of γ, $\beta^T\beta = s$ won't hold, but we can pick the value of γ (will depend on s) such that $\hat{\beta}_\gamma^T \hat{\beta}_\gamma = s$.

More general constrained optimization optimality conditions : KKT (KKT: Karush-Kuhn-Tucker).