Ridge regression derivation

ref: Ordinary Least Squares derivation

To add ridge penalty to ordinary least squares, we're going to make a small change to the loss function, then go through all the same steps

$$L = (Y - X\omega)^T(Y - X\omega) + \omega^T\omega$$

$$\begin{aligned}\frac{\partial L}{\partial \omega} &= 2(Y - X\omega)^T \frac{\partial (Y - X\omega)}{\partial \omega} + 2\omega^T \\ &= 2(Y - X\omega)^T(-X) + 2\omega^T \\ &= 0\end{aligned}$$

$$\begin{aligned}2(Y - X\omega)^T(-X) + 2\omega^T &= 0 \\ (Y - X\omega)^T(-X) + \omega^T &= 0 \\ (-X)^T(Y - X\omega) + \omega &= 0 \qquad \text{took transpose of both sides} \\ -X^TY + X^TX\omega + \omega &= 0 \\ -X^TY + (X^TX + I)\omega &= 0 \\ (X^TX + I)\omega &= X^TY \\ \omega &= (X^TX + I)^{-1}X^TY\end{aligned}$$