Ridge Regression

- How to derive the solution $\hat{\boldsymbol{\beta}}^{\text{ridge}}$?
- Understand the shrinkage effect of Ridge.
- Why we want to do shrinkage?
- How to quantify the dimension (or df) of a ridge regression model?
- How to select the tuning parameter λ? (see R page)

- In Ridge regression, the criterion we want to minimize is

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \lambda\boldsymbol{\beta}^T\boldsymbol{\beta}.$$

- The solution

$$\hat{\boldsymbol{\beta}}^{\text{ridge}} = (\mathbf{X}^T\mathbf{X} + \lambda I)^{-1}\mathbf{X}^T\mathbf{y}.$$

  Later we'll see that ridge coefficients can be computed efficiently for all λ using SVD.

- Compared to the OLS estimate $\hat{\boldsymbol{\beta}}^{\text{LS}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$, the ridge regression solution adds a non-negative constant to the diagonal of $\mathbf{X}^T\mathbf{X}$, so we can take the inversion even if $\mathbf{X}^T\mathbf{X}$ is not of full rank and it was the initial motivation for ridge regression (Hoerl and Kennard, 1970).