Ridge Regression

Mike Bowles,PhD and Patricia Hoffman, PhD

Normal Equation Solution

Find best function ($\mathbf{X}$ is the input $p \times N$ matrix)

$$f(\mathbf{X}) = \beta_0 + \Sigma_j \, \mathbf{X}\beta_j$$

Oservations: $\mathbf{X}^T = (\mathbf{X}_1, \mathbf{X}_2, ...\mathbf{X}_p)$

Regression Coef: $\widehat{\beta} = (\beta_0 ... \beta_p)$

minimize RSS $\sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2$

Solution: $\widehat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$

Ridge Regression

$\lambda \geq 0$ complexity parameter controls shrinkage.
$\lambda = 0 \Rightarrow$ solution is the same as regular regression.
λ penalizes the sum-of-squares of the parameters.

Minimize

$$\sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$