STA4241 Interactive Lab Week 8: Ridge regression

Today we are going to focus on ridge regression and dig deeper into why it can outperform standard least squares in a somewhat simplified scenario. We will make the simplifying assumption throughout that the covariates are orthonormal, which implies that $\boldsymbol{X}^T\boldsymbol{X} = \boldsymbol{I}_p$

(1) Derive the bias of the ridge regression estimator for a chosen λ value, i.e $E(\widehat{\boldsymbol{\beta}}_\lambda^R - \boldsymbol{\beta})$

(2) Derive the variance of the same ridge regression estimator

(3) Now we will show that the mean squared error of the ridge regression estimator can be given by

$$E\left[(\widehat{\boldsymbol{\beta}}_\lambda^R - \boldsymbol{\beta})^T(\widehat{\boldsymbol{\beta}}_\lambda^R - \boldsymbol{\beta})\right] = \frac{p\sigma^2}{(1+\lambda)^2} + \frac{\lambda^2\boldsymbol{\beta}^T\boldsymbol{\beta}}{(1+\lambda)^2}$$

Note that we need to use a result on quadratic forms, which tells us that

$$E\left[(\widehat{\boldsymbol{\beta}}_\lambda^R - \boldsymbol{\beta})^T(\widehat{\boldsymbol{\beta}}_\lambda^R - \boldsymbol{\beta})\right] = Tr(\text{Var}[\widehat{\boldsymbol{\beta}}_\lambda^R]) + E[(\widehat{\boldsymbol{\beta}}_\lambda^R - \boldsymbol{\beta})]^T E[(\widehat{\boldsymbol{\beta}}_\lambda^R - \boldsymbol{\beta})]$$

Don't worry about why this result holds true. If you want more information on it, check out the quadratic form page on Wikipedia: `https://en.wikipedia.org/wiki/Quadratic_form_(statistics)`

(4) Now we will show that the value of λ that minimizes this MSE is given by

$$\lambda = \frac{p\sigma^2}{\boldsymbol{\beta}^T\boldsymbol{\beta}}$$

(5) What does this tell us about when ridge regression is most useful relative to ordinary least squares?

(6) Let's look at some R code to see how this plays out empirically