6.2 Shrinkage Methods

Zongyi Liu

2023-07-19

6.2 Shrinkage Methods

This section provides an alternative to fit a model by constraining or regularizing the coefficient estimates, in other words, shrinks the coefficient estimates toward zero.

The two most famous techniques to shrink the regression coefficients are

6.2.1 Ridge Regression

From previous chapters we know that coefficients can be estimated using the values that minimize

$$\mathrm{RSS} = \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 .$$

In ridge regression, the regression coefficient estimates $\hat{\beta}^R$ for the value that minimize

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}\beta_j^2 = \mathrm{RSS} + \lambda\sum_{j=1}^{p}\beta_j^2,$$

Here $\lambda \geq 0$ is a tuning parameter, which will need to be determined separately.

the equation above trades off two different criteria. As with least squares, ridge regression seeks coefficient estimates that fit the data well, by making the RSS small. However, the second term, $\lambda\sum_j \beta_j^2$, called a shrinkage penalty, is small when βs are close to zero, and so it has the effect of shrinking the estimates of β_j towards zero. The tuning parameter λ serves to control the relative impact of these two terms on the regression coefficient estimates. When $\lambda = 0$, the penalty term has no effect, and ridge regression will produce the least squares estimates.

6.2.2 The Lasso

You can also embed plots, for example:

6.2.3 Selecting the Tuning Parameter