More regression methods

Our text discusses best subset selection and ridge regression, but there are many more methods available. I will summarize a couple of other methods that are also in common use. For an excellent account of all of these methods and much more, refer to *Elements of Statistical Learning: Data Mining, Inference, and Prediction* (ESL) by Hastie, Friedman, and Tibshirani.

Ridge Regression

Our text covers ridge regression and shows the solution for the coefficient estimates given the ridge constant d. Another characterization of ridge regression is that it finds the least squares solution subject to a constraint. Specifically, the ridge solution minimizes squared error subject to

$$\sum_j \beta_j^2 \leq t$$

LASSO Regression

Lasso regression also minimizes squared error subject to a constraint, but in this case the constraint is:

$$\sum_j |\beta_j| \leq t$$

This seemlingly minor difference from ridge regression leads to many differences, from the actual estimator (which is no longer closed form) to its behavior in reducing the number of non-zero coefficient estimates.

Partial Least Squares

Like principal component regression, partial least squares proceeds by creating new variables and using them as covariates in the regression model. In each procedure, successive variables are created to satisfy an orthogonality condition with respect to previously derived variables. In partial least squares, these new variables involve correlations with the dependent variable y. For example, the first new variable z_1 is derived via:

$$z_1 = \sum_j \widehat{\varphi}_j x_j, \text{ where } \widehat{\varphi}_j = \sum_i x_{ij} y_i$$

ESL has an excellent discussion comparing these different approaches. One point is that best subset, PCR, and PLS methods discretize the selection problem by retaining a subset of variables or new variables, while ridge and LASSO regression more smoothly downweight coefficients, with the LASSO tending to set some coefficients to zero.