Regression, Lasso Regression and Ridge Regression

In linear regression with p explanatory variable and vector of 1's to fit the mean of the dependent variable, the fitting algorithm finds coefficients β_i for i= 0, 1, ...,p to minimize the residual sum of squares, RSS.

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

Lasso regression adds a shrinkage penalty to the residual sum of squares and fitting algorithm minimizes the sum of the residual sum of squares and the shrinkage penalty.

$$RSS + \lambda \sum_{j=1}^{p} |\beta_j|$$

The shrinkage penalty consists of λ, a tuning parameter, times the sum of the absolute values of the coefficients with β_0 excluded. As the tuning parameter gets larger the minimization will have to increasingly shrink the magnitude of coefficients at the expense of increasing the residual sum of squares. Often lasso shrinks many coefficients to zero, so lasso effectively provides variable selection via variable removal. In the assignment we try positive values for λ that vary over many orders of magnitude. Then we use 10 fold cross validation to help us pick a good tuning parameter to using in our model.

The R script briefly mention to ridge regression which, while not discussed or illustrated in the script, is another good option to have in our tool kit of models. Ridge regression has a different shrinkage penalty. It does not shrink coefficients to zero.

$$RSS + \lambda \sum_{j=1}^{p} \beta_j^2$$