Regularized Regression

In the case of linear regression, we might consider four different regularization functions:

$$f_R(\underset{\sim}{\beta}) = \sum_{m=1}^{M} \beta_m^2 \quad \text{(ridge regression)}$$

$$f_R(\underset{\sim}{\beta}) = \sum_{m=1}^{M} |\beta_m| \quad \text{(lasso regression)}$$

$$f_R(\underset{\sim}{\beta}) = (1-\alpha)\sum_{m=1}^{M} \beta_m^2 + \alpha \sum_{m=1}^{M} |\beta_m| \quad \text{(elastic net regression)}$$

$$f_R(\underset{\sim}{\beta}) = \sum_{m=1}^{M} (1, \text{ if } \beta_m \neq 0;\ 0, \text{ if } \beta_m = 0)$$

Note that in all cases the intercept term is not part of the sum. It is more related to the level of the target value, not to the value of any of the features. The fourth function presented does not have a name and is rarely used in this context. It looks simple, adding a penalty of 1 for each coefficient used. But it is difficult to implement as the regression must be done using each possible subset of features to perform the minimization.

We will begin with ridge regression on the next page.