Ridge and Lasso

Hypothesis for OLS

OLS minimizes $\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_p x_{ip})^2$

While Ridge Regression minimizes what OLS minimizes with the constraint $\sum_{j=1}^{p} \beta_j^2 \leq T$, put it the other way, it's minimizing $\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_p x_{ip})^2 + \lambda \sum_{j=1}^{p} \beta_j^2,$ called a tuning parameter (shrinkage parameter, regularization).

Lasso minimizes what OLS minimizes with the constraint

$\sum_{j=1}^{p}|\beta_j| \leq T$, put it the other way, it's minimizing $\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_p x_{ip})^2 + \lambda \sum_{j=1}^{p}|\beta_j|$

Geographically, Ridge is the intercept between the ellipse and the circle; Lasso is with the square. $\lim_{\lambda \to 0}$, T is approximating ∞, then Ridge and Lasso is estimating OLS.

Non-scale invariant & Standardization

Larger s are penalized.

After standardization, if all predictors are standardized and uncorrelated

$\hat{\beta}_{j,\lambda,Ridge} = \frac{1}{1+\lambda}\beta_{j,OLS}$

When there's multicollinearity, lambda is large, LFS is much smaller than the variance of OLS

- Comparison between Ridge and Lasso: Ridge deals with **multicollinearity**, Lasso deals with **sparsity (lots of variables have the coefficient 0)**

Lambda

1. training data, compute the smallest rooted mean square error (RMSE) (model fitting)
2. test data

Performing tests in R

```r
EEO = read.table("P236.txt", h = T)
```

```r
library(MASS)
lm.ridge(ACHV ~ FAM + PEER + SCHOOL, data = EEO, lambda = c(1, 5, 10, 15, 20))
```

```
                     FAM      PEER     SCHOOL
 1 -0.04055397 0.3768768 1.3205433 -0.6276745
 5 -0.02708020 0.2318348 0.7229545  0.0419616
10 -0.02354968 0.2383780 0.5567732  0.1624044
15 -0.02168803 0.2429370 0.4847180  0.2015936
20 -0.02032593 0.2440326 0.4422964  0.2187849
```

```r
# note: the output is the Ridge regression coefficient, as lambda growths larger, the coeffi
```

```r
meatspec = read.table("http://www.stat.uchicago.edu/~yibi/s224/data/meatspec.txt", header = T
```

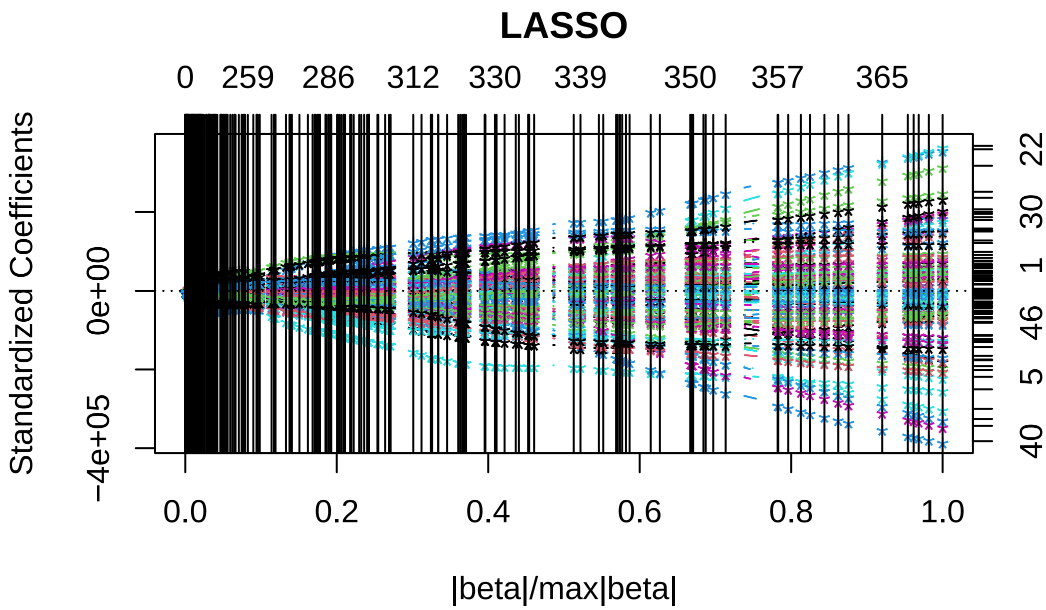
```r
#library(lars) #installation required for the function
#dataset: 100 predictors in total
trainmeat = meatspec[1:172, ]
testmeat = meatspec[173:215, ]
```

```r
trainy = trainmeat$fat
trainx = as.matrix(trainmeat[, -101]) # converts the predictor variables (excluding column 10
library(lars)
```

```
Loaded lars 1.3
```

```r
lassomod = lars(trainx, trainy) # fits the Lasso regression model to the training data
```

```r
plot(lassomod)
```



X-axis: At 0: all coefficients are **zero** (maximum regularization); at 1: OLS

Y-axis: represents the **standardized coefficients** for each predictor variable

- some coefficient starts at 0
- black vertical dashes indicates when the predictor enter the model
- those enter earliest are the most important predictors