

Ch 6.2: Shrinkage - The Lasso

Lecture 18 - CMSE 381

Prof. Mengsen Zhang

Michigan State University

::

Dept of Computational Mathematics, Science & Engineering

Fri, Feb 28, 2025

Announcements

Last time:

- Ridge Regression

This time:

- The Lasso

Announcements:

- HW4 due Sunday
- HW5 released
- Think about the project, choose partner in crime

	W	2/12	Midterm #1		
12	F	2/14	Leave one out CV	5.1.1, 5.1.2	
13	M	2/17	k-fold CV	5.1.3	
14	W	2/19	More k-fold CV	5.1.4-5	
15	F	2/21	k-fold CV for classification	5.1.5	
16	M	2/24	Subset selection	6.1	
17	W	2/26	Shrinkage: Ridge	6.2.1	
18	F	2/28	Shrinkage: Lasso	6.2.2	HW #4 Due Sun 3/2
	M	3/3	Spring Break		
	W	3/5	Spring Break		
	F	3/7	Spring Break		
19	M	3/10	PCA	6.3	
20	W	3/12	PCR	6.3	
	F	3/14	Review		HW #5 Due Sun 3/16
	M	3/17	Midterm #2		

Section 1

Last time

- Fit model using all p predictors
- Aim to constrain (regularize) coefficient estimates
- Shrink the coefficient estimates towards 0

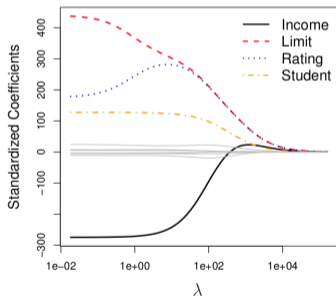
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$

Ridge regression

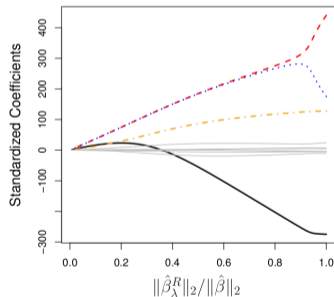
Before:

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

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



After:



Scale equivariance (or lack thereof)

Scale equivariant: Multiplying a variable by c (cX_i) just returns a coefficient multiplied by $1/c$ ($1/c\beta_i$)

Solution: standardize predictors

$$\tilde{x}_{ij} = \frac{x_{ij}}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}}$$

- Least squares is scale equivariant
- Ridge regression is not

Section 2

The Lasso

Same goal as before

- Fit model using all p predictors
- Aim to constrain (regularize) coefficient estimates
- Shrink the coefficient estimates towards 0

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$

Least Squares:

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

Ridge:

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

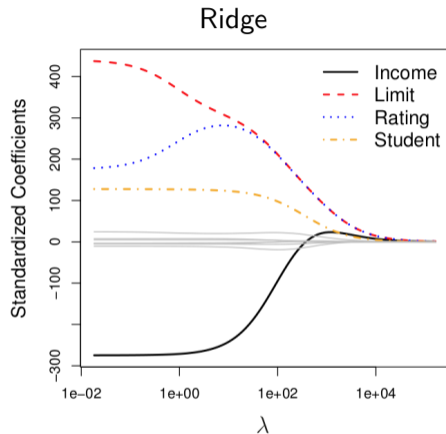
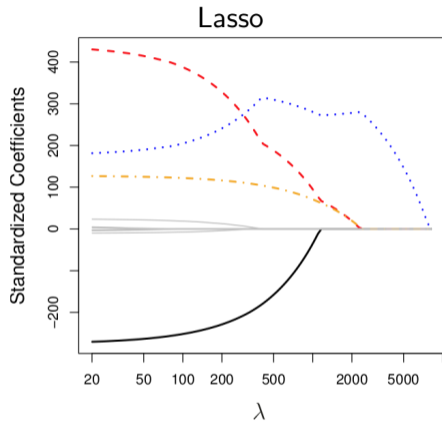
The Lasso:

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

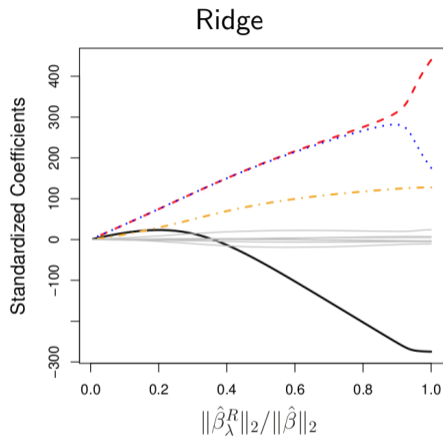
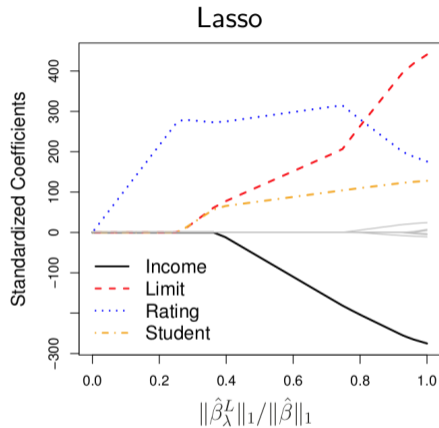
Subsets with lasso

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

An example on Credit data set



More example on Credit data set

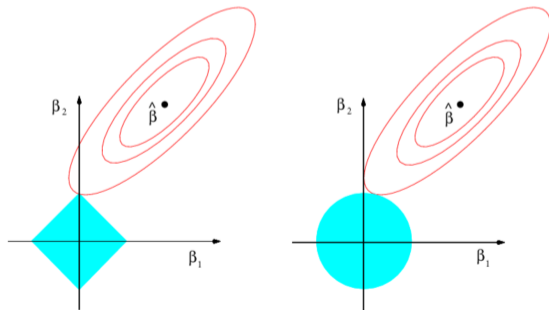


Why the hell lasso can select variable (while ridge cannot)?

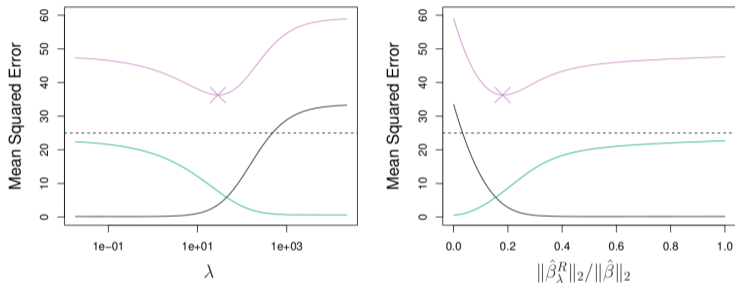
Alternative formulation of lasso & ridge regression ([play more with \$\ell_p\$](#))

$$\min_{\beta} \sum (y_i - \hat{y}_i)^2 \text{ where } \sum |\beta_j| \leq s$$

$$\min_{\beta} \sum (y_i - \hat{y}_i)^2 \text{ where } \sum |\beta_j|^2 \leq s$$



Bias-Variance tradeoff

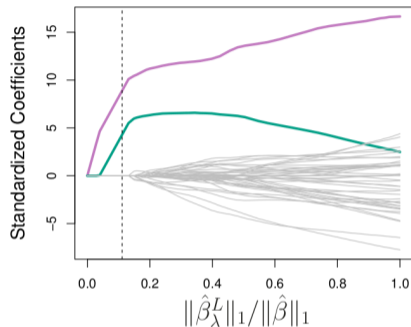
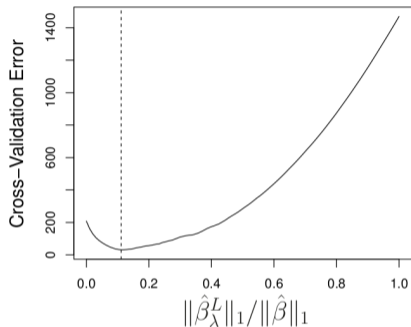


Squared bias (black), variance (green), and test mean squared error (purple) for simulated data.

Using Cross-Validation to find λ

- Choose a grid of λ values
- Compute the (k -fold) cross-validation error for each value of λ
- Select the tuning parameter value λ for which the CV error is smallest.
- The model is re-fit using all of the available observations and the selected value of the tuning parameter.

10-fold CV choice of λ for lasso and simulated data



Coding example

Ridge Regression:

Lasso:

Least Squares:

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

Ridge:

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

The Lasso:

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

Next time

	W	2/12	Midterm #1		
12	F	2/14	Leave one out CV	5.1.1, 5.1.2	
13	M	2/17	k-fold CV	5.1.3	
14	W	2/19	More k-fold CV	5.1.4-5	
15	F	2/21	k-fold CV for classification	5.1.5	
16	M	2/24	Subset selection	6.1	
17	W	2/26	Shrinkage: Ridge	6.2.1	
18	F	2/28	Shrinkage: Lasso	6.2.2	HW #4 Due Sun 3/2
	M	3/3	Spring Break		
	W	3/5	Spring Break		
	F	3/7	Spring Break		
19	M	3/10	PCA	6.3	
20	W	3/12	PCR	6.3	
	F	3/14	Review		HW #5 Due Sun 3/16
	M	3/17	Midterm #2		