18) Ridge Regression and Least Absolute Shrinkage and Selection Operator (LASSO)

Vitor Kamada

January 2020

Reference

Tables, Graphics, and Figures from:

1) An Introduction to Statistical Learning

James et al. (2017): Ch 6.2

2) The Elements of Statistical Learning

Hastie et al. (2017): Ch 3.3, and 3.4

Stamey et al. (1989)

Y: log of Prostate-Specific Antigen

lcavol: log cancer volume

lweight: log prostate weight

lbph: log of the amount of benign prostatic hyperplasia

svi: seminal vesicle invasion

lcp: log of capsular penetration

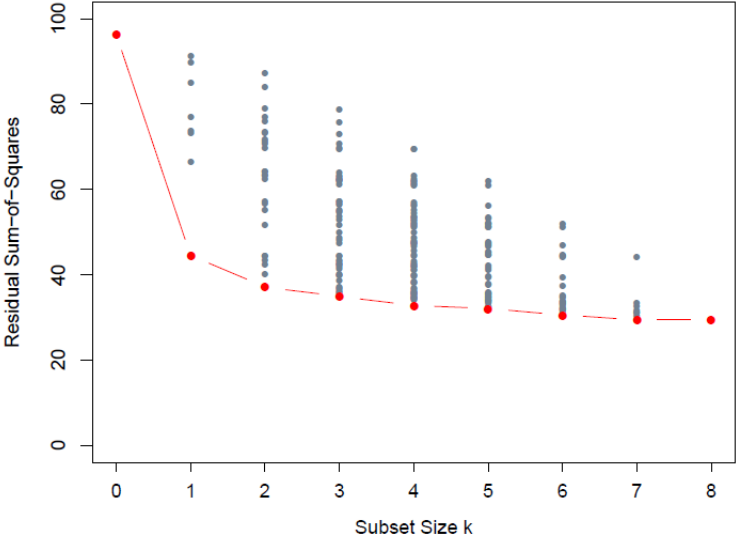
gleason: Gleason score

pgg45: Gleason scores 4 or 5

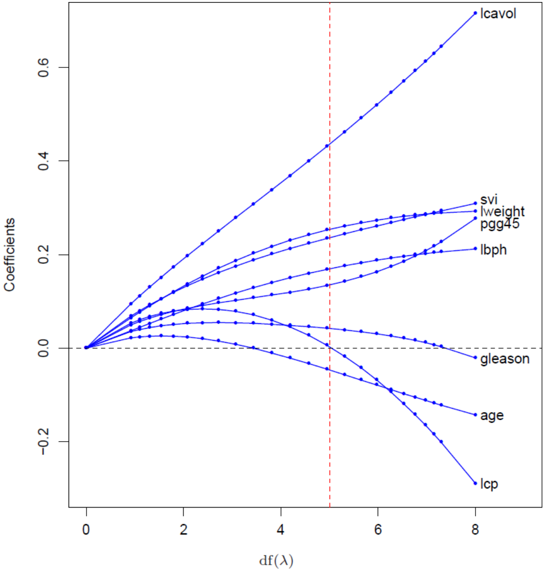
Tenfold Cross-Validation

Term	LS	Best Subset	Ridge	Lasso
`Intercept`	2.465	2.477	2.452	2.468
`lcavol`	0.680	0.740	0.420	0.533
`lweight`	0.263	0.316	0.238	0.169
`age`	−0.141		−0.046	
`lbph`	0.210		0.162	0.002
`svi`	0.305		0.227	0.094
`lcp`	−0.288		0.000	
`gleason`	−0.021		0.040	
`pgg45`	0.267		0.133	
Test Error	0.521	0.492	0.492	0.479
Std Error	0.179	0.143	0.165	0.164

Best-Subset Selection (Prostate Cancer)



Ridge Coefficients for the Prostate Cancer



Ridge Regression

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

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

$$\frac{||\hat{\beta}_\lambda^R||_2}{||\hat{\beta}||_2}$$

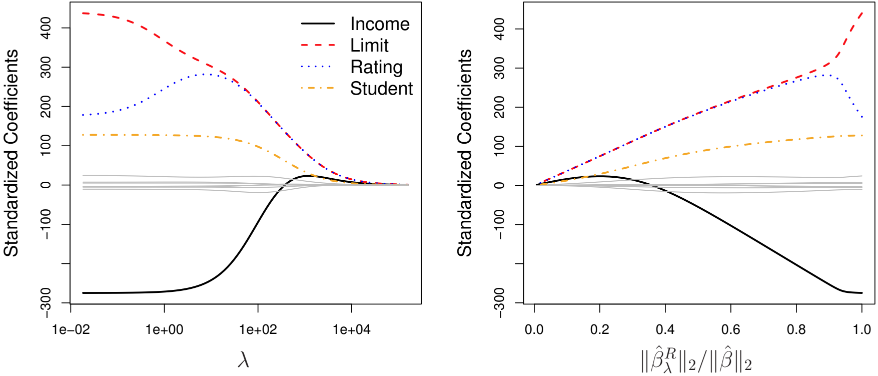
$$||\beta||_2=\sqrt{\sum_{j=1}^{p}\beta_j^2}$$

Ridge Regression - Matrix Form

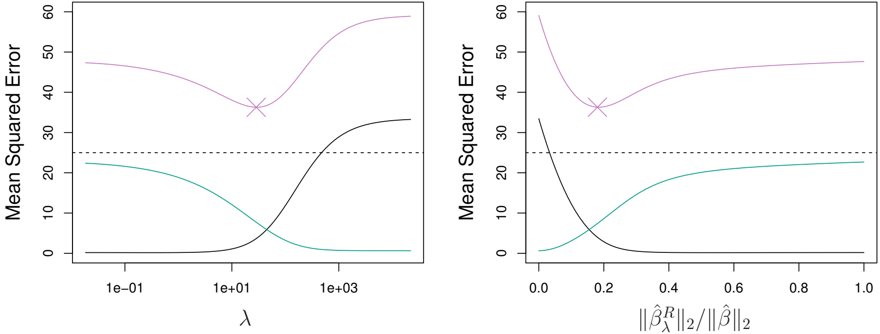
$$RSS(\lambda) = (y - X\beta)^T(y - X\beta)^T + \lambda\beta^T\beta$$

$$\hat{\beta}^{ridge} = (X^TX + \lambda I)^{-1}X^Ty$$

Credit Data Set



Ridge: Squared Bias (Black), Variance (Green), and Test Mean Squared Error (Pink)



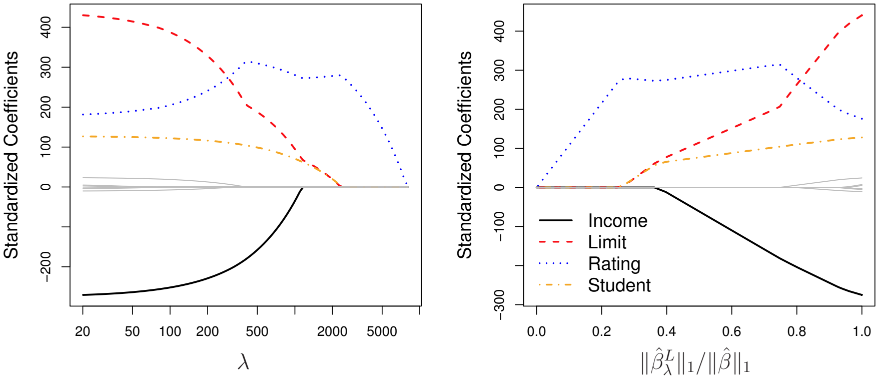
Least Absolute Shrinkage and Selection Operator (LASSO)

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

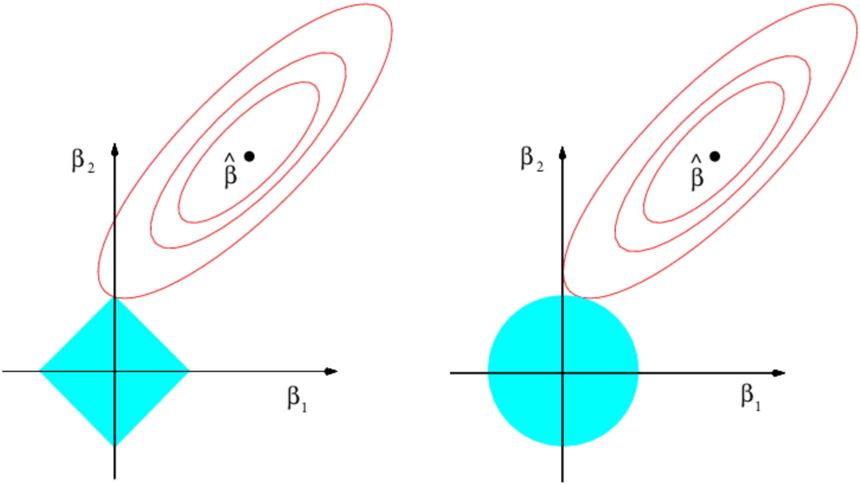
$$\frac{\|\hat{\beta}_\lambda^L\|_1}{\|\hat{\beta}\|_1}$$

$$\|\beta\|_1 = \Sigma|\beta_j|$$

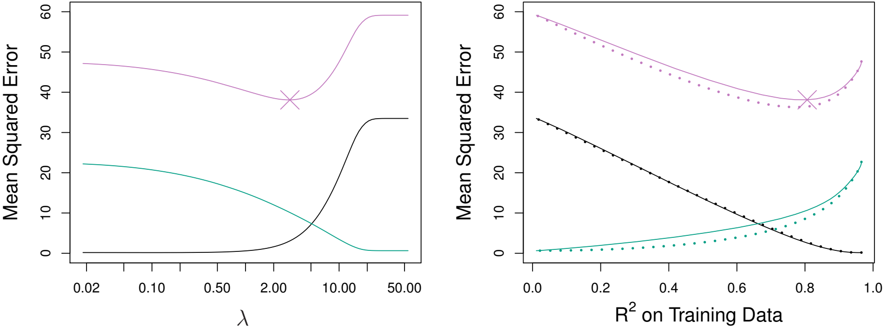
The Standardized Lasso Coefficients



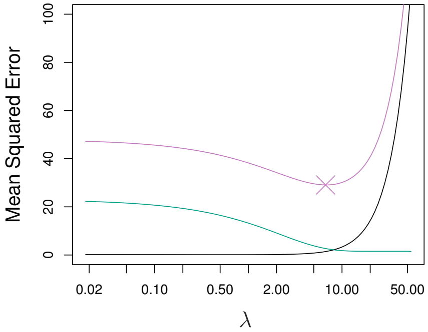
$|\beta_1| + |\beta_2| \leq s$ and $\beta_1^2 + \beta_2^2 \leq s$

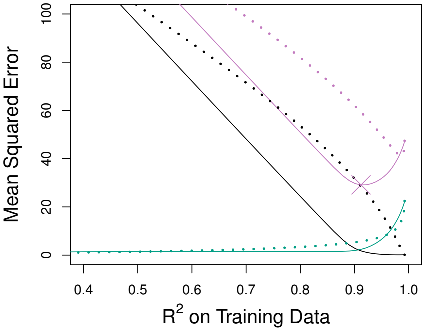


Squared Bias (Black), Variance (Green), and Test MSE (Pink)



Squared Bias (Black), Variance (Green), and Test MSE (Pink)





$n = p$ and X a Diagonal Matrix with 1's

$$\sum_{j=1}^{p} (y_j - \beta_j)^2$$

$$\hat{\beta}_j = y_j$$

$$\sum_{j=1}^{p} (y_j - \beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

$$\hat{\beta}_j^L = \begin{cases} y_j - \lambda/2 & \text{if} \quad y_j > \lambda/2 \\ y_j + \lambda/2 & \text{if} \quad y_j < -\lambda/2 \\ 0 & \text{if} \quad |y_j| \leq \lambda/2 \end{cases}$$

Ridge and Lasso Regression

$$\sum_{j=1}^{p} (y_j - \beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

$$\hat{\beta}_j^R = y_j/(1+\lambda)$$