Regression Shrinkage and Selection via the Lasso

Section 7. Simulations

Robert Tibshirani

Oct. 19, 2006

Objective

- Comparisons of the following approaches
  - least square estimate with all variable
  - lasso (five-fold cross validation)
  - lasso (Stein)
  - lasso (GCV)
  - non-negative garotte
  - best subset selection
  - ridge regression

Data generation

- Procedure of the simulation
  - set $\boldsymbol{\beta} = (3, 1.5, 0, 0, 2, 0, 0, 0)$
  - generate variables $\mathbf{x} \sim N(\mathbf{0}, \Sigma(\rho))$ such that $corr(x_i, x_j) = \rho^{|i-j|}$ where $\rho = .5$.
  - set $\sigma = 3$ and generate error $\epsilon \sim N(0, 1)$
  - signal-to-noise ratio: $\frac{\boldsymbol{\beta}^T \Sigma(\rho) \boldsymbol{\beta}}{\sigma^2} \approx 5.7$
  - 20 observations from the model

$$y = \boldsymbol{\beta}^T \mathbf{x} + \sigma \epsilon \text{ with } \mathbf{x} = (x_1, \cdots, x_8)'$$

- Repeat the procedure 50 (? 200) times to output 50 (? 200) datasets with 20 observations for each

Results

Table. 3: Results for example 1

Method	Median mean -squared error	Average number of 0 coefficients	Average $\hat{s}$
Least squares	2.79(0.12)	0.0	—
Lasso (CV)	2.43(0.14)	3.3	0.63(0.01)
Lasso (Stein)	2.07(0.10)	2.6	0.69(0.02)
Lasso (GCV)	1.93(0.09)	2.4	0.73(0.01)
Garotte	2.29(0.16)	3.9	—
Best subset	2.44(0.16)	4.8	—
Ridge	3.21(0.12)	0.0	—

(Standard errors are given in parentheses)

Lasso (GCV) wins in terms of Median MSE!

Results

Table. 4 and 5: Most frequent models selected by lasso (GCV) and best subset in example 1

Models slected by Lasso (GCV)	Proportion	Models slected by Best subset	Proportion
12∅45678	0.055	12∅∅5∅∅∅	0.24
123456∅∅	0.050	1∅∅∅5∅∅∅	0.20
12∅∅5∅∅8	0.045	1∅∅∅∅∅∅∅	0.095
12∅45∅∅∅	0.045	12∅∅5∅7∅	0.04
12∅∅5∅∅∅	0.025	12∅∅5∅∅∅	0.24

Recall $\beta = (3, 1.5, 0, 0, 2, 0, 0, 0)$, so the true model is of the form $12\emptyset\emptyset5\emptyset\emptyset\emptyset$. Among all the models selected by lasso (GCV), 95.5% of them contains all the (1,2,5) variables. However, for best subset method, the percentage is 53.5%.

Data generation

- Procedure of the simulation
  - $\surd$ set $\boldsymbol{\beta} = (0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85)$
  - generate variables $\mathbf{x} \sim N(\mathbf{0}, \Sigma(\rho))$ such that $corr(x_i, x_j) = \rho^{|i-j|}$ where $\rho = .5$.
  - $\surd$ set $\sigma = 3$ and generate error $\epsilon \sim N(0, 1)$
  - $\surd$ signal-to-noise ratio: $\frac{\boldsymbol{\beta}^T \Sigma(\rho) \boldsymbol{\beta}}{\sigma^2} \approx 1.8$
  - 20 observations from the model

$$y = \boldsymbol{\beta}^T \mathbf{x} + \sigma\epsilon \text{ with } \mathbf{x} = (x_1, \cdots, x_8)^T$$

- Repeat the procedure 50 times to output 50 datasets with 20 observations for each

Results

Table. 6: Results for example 2

Method	Median mean -squared error	Average number of 0 coefficients	Average $\hat{s}$
Least squares	6.50(0.64)	0.0	—
Lasso (CV)	5.30(0.45)	3.0	0.50(0.03)
Lasso (Stein)	5.85(0.36)	2.7	0.55(0.03)
Lasso (GCV)	4.87(0.35)	2.3	0.69(0.23)
Garotte	7.40(0.48)	4.3	—
Best subset	9.05(0.78)	5.2	—
Ridge	2.30(0.22)	0.0	—

(Standard errors are given in parentheses)

Ridge regression wins in terms of Median MSE and Average number of 0 coeeficients! All three lasso outperform the least squares.

Data generation

- Procedure of the simulation
  - $\surd$ set $\boldsymbol{\beta} = (5, 0, 0, 0, 0, 0, 0, 0)$
  - generate variables $\mathbf{x} \sim N(\mathbf{0}, \Sigma(\rho))$ such that $corr(x_i, x_j) = \rho^{|i-j|}$ where $\rho = .5$.
  - $\surd$ set $\sigma = 2$ and generate error $\epsilon \sim N(0, 1)$
  - $\surd$ signal-to-noise ratio: $\frac{\boldsymbol{\beta}^T \Sigma(\rho) \boldsymbol{\beta}}{\sigma^2} \approx 7$
  - 20 observations from the model

$$y = \boldsymbol{\beta}^T \mathbf{x} + \sigma\epsilon \text{ with } \mathbf{x} = (x_1, \cdots, x_8)^T$$

- Repeat the procedure 50 times to output 50 datasets with 20 observations for each

Results

Table. 7: Results for example 3

Method	Median mean -squared error	Average number of 0 coefficients	Average $\hat{s}$
Least squares	2.89(0.04)	0.0	—
Lasso (CV)	0.89(0.01)	3.0	0.50(0.03)
Lasso (Stein)	1.26(0.02)	2.6	0.70(0.01)
Lasso (GCV)	1.02(0.02)	3.9	0.63(0.04)
Garotte	0.52(0.01)	5.5	—
Best subset	0.64(0.02)	6.3	—
Ridge	3.53(0.05)	0.0	—

(Standard errors are given in parentheses)

Garotte wins in terms of Median MSE, and best subset wins in terms of Average number of 0 coeeficients! All three lasso outperform the least squares.

Data generation

- Procedure of the simulation
  - $\surd$ set 40×1 vector $\boldsymbol{\beta} = (0, \cdots, 0, 2, \cdots, 2, 0, \cdots, 0, 2, \cdots, 2)$
  - $\surd$ generate $x_{ij} = z_{ij} + z_i$ such that $corr(x_{ij}, x_{il}) = 0.5$ for $j \neq l$ where

$$z_{ij}, z_i;\ \text{for } i = 1, \cdots, 200 \text{ and } j, l = 1, \cdots, 40 \overset{i.i.d.}{\sim} N(0,1).$$

  - $\surd$ set $\sigma = 15$ and generate error $\epsilon \sim N(0,1)$
  - $\surd$ signa-to-noise ratio: $\frac{\boldsymbol{\beta}^T \Sigma \boldsymbol{\beta}}{\sigma^2} \approx 9$ where $\Sigma = \mathbf{I} + \mathbf{J} = \mathbf{1}\mathbf{1}^T + diag(\mathbf{1})$
  - $\surd$ 100 observations from the model

$$y = \boldsymbol{\beta}^T \mathbf{x} + \sigma\epsilon \text{ with } \mathbf{x} = (x_1, \cdots, x_{40})^T$$

- Repeat the procedure 50 times to output 50 datasets with 100 observations for each

Results

Table. 8: Results for example 4

Method	Median mean -squared error	Average number of 0 coeeficients	Average $\hat{s}$
Least squares	137.3(7.3)	0.0	—
Lasso (Stein)	80.2(4.9)	14.4	0.55(0.02)
Lasso (GCV)	64.9(2.3)	13.6	0.60(0.88)
Garotte	94.8(3.2)	22.9	—
Ridge	57.4(1.4)	0.0	—

(Standard errors are given in parentheses) Lasso (CV) and best subset is **impractical** for large size of observations and variables respectively. Ridge regression wins in terms of Median MSE. Both two lasso methods outperform the least squares.

$\boldsymbol{\beta}$ \ lasso	CV	Stein	GCV
$(3, 1.5, 0, 0, 2, 0, 0, 0)$	4/7	2/7	1/7
$(0.85, \cdots, 0.85)$	3/7	4/7	2/7
$(5, 0, 0, 0, 0, 0, 0, 0)$	3/7	5/7	4/7
$(0, \cdots, 0, 2, \cdots, 2, 0, \cdots, 0, 2, \cdots, 2)$	–	3/5	2/5

Lasso is a robust method for regression and variable selection!