

BIOS:7600 Homework 5

Chuan Lu

March 16, 2019

1. Problem 3.1

The plot for two-stage adaptive lasso is shown in Figure 1, where the initial estimate of β is the OLS solution. The fit plot is generated by `ncvreg`.

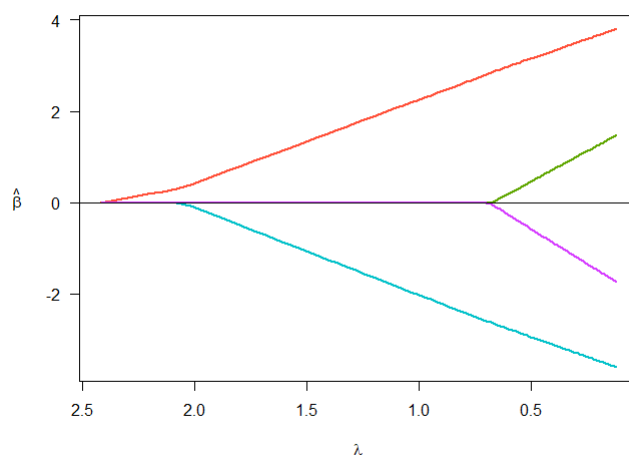


Figure 1: two stage adaptive lasso

Since the weights are constant across all values of λ , the path looks different with pathwise adaptive lasso, but similar with original lasso paths, but all four variables are selected with a smaller λ since the initial weights are 1.231007, 2.067553, 1.310293, 2.252761, larger than 1 used in lasso.

2. Problem 3.2(b)

Proof. First, when $|x| \leq \lambda$, SCAD penalty is just the $\|\cdot\|_1$ penalty for lasso, so the thresholding operator is the same with lasso. When $|x| \geq \gamma\lambda$, the penalty is a constant, so the operator does not change the value. Now

consider the case when $2\lambda < z < \gamma\lambda$; the other part when $z < 0$ is just the same. The subgradient of Q in the univariate problem is

$$\partial Q = -z + \beta + \frac{\gamma\lambda - \partial(|\beta|)}{\gamma - 1}.$$

If $\beta \neq 0$, then let $\partial Q = 0$, we have

$$\beta = \frac{\gamma - 1}{\gamma - 2} \left(z - \frac{\lambda\gamma}{\gamma - 1} \right).$$

If $\beta = 0$, then $\partial(|\beta|) = [-1, 1]$, and by $0 \in \partial Q$ we have

$$|z| \leq \frac{\gamma\lambda}{\gamma - 1}.$$

So

$$\beta = \begin{cases} \frac{\gamma - 1}{\gamma - 2} \left(z - \frac{\lambda\gamma}{\gamma - 1} \right), & |z| > \frac{\gamma\lambda}{\gamma - 1}, \\ 0, & \text{otherwise.} \end{cases}$$

Notice $\gamma > 2$, we have $\frac{\gamma}{\gamma - 1} < 2$. By combining the second case with the lasso operator, we can show the result. \square

3. Problem 2.7

Proof. First, the lasso approximation is

$$f(y_i) = x_i^\top S(\beta_i | \lambda) = x_i^\top S\left(\frac{1}{n} x_i^\top y\right),$$

which is piecewise linear and continuous. Then f is Lipschitz, which means f is absolutely continuous. Besides,

$$f'(y) = \frac{1}{n} x^\top x \cdot 1_{|\beta| > \lambda} = \frac{1}{n} I_n \cdot 1_{|\beta^{OLS}| > \lambda},$$

which is bounded, so

$$df = \sum_{i=1}^n f'(y_i) = \sum_{i=1}^n 1_{|\beta_i^{OLS}| > \lambda},$$

which proves the result. \square

4. Problem 3.3

(a) By the same process with the last problem, we get

$$f'(y) = \begin{cases} 0, & |\beta^{OLS}| < \lambda \\ \frac{\gamma}{\gamma-1}, & \lambda \leq |\beta^{OLS}| \leq \gamma\lambda \\ 1, & |\beta^{OLS}| > \gamma\lambda \end{cases}$$

Then

$$df = \sum_{i=1}^n f'(y_i) = \#\{|\beta^{OLS}| > \gamma\lambda\} + \frac{\gamma}{\gamma-1} \#\{\lambda \leq |\beta^{OLS}| \leq \gamma\lambda\}.$$

(b) When $\gamma = 3$,

$$df_{MCP} = \#\{|\beta^{OLS}| > 3\lambda\} + \frac{3}{2} \#\{\lambda \leq |\beta^{OLS}| \leq 3\lambda\} \geq df_{LASSO}.$$

5. Problem 3.7

	forward	lasso	ridge	MCP	SCAD
1	0.3645032	0.1436703	0.6538278	0.1218058	0.1318939
2	NA	0.1680950	2.0420509	0.1434068	0.1540713
3	0.7072784	0.8207570	0.6495371	0.7677788	0.7481251
4	NA	2.258974	2.028072	2.178437	2.140876

6. Problem 3.8

(a) The plot is show in Figure 2.

(b) The graph of p' is shown in Figure 3, where p' is

$$p'(\theta) = \begin{cases} e^{-\frac{\tau}{\lambda}\theta}, & |\theta| > 0 \\ [-1, 1], & |\theta| = 0 \end{cases}$$

(c) Based on the result of p' , the estimates would be similar with MCP when β is not large; however, when β is large, since $p' > 0$, it should be a combination of MCP and Lasso, while MCP has a larger weight.

(d) TBD.

7. Problem 3.10

$\gamma = 2.3$ by minimizing `fit$cve`.

(a) The plots of R^2 by MCP and Lasso are shown in Figure 4. The max R^2 by MCP is about 0.5, while it is about 0.25.

(b) At the max R^2 , 5 variables are chosen from MCP, while 47 variables are chosen from lasso.

(c) The gene enters MCP model first is **C16ORF30**, enters lasso first is **GNPDA1**. I think there must be something wrong with the code since the genes are different. The paths are shown in Figure 5:

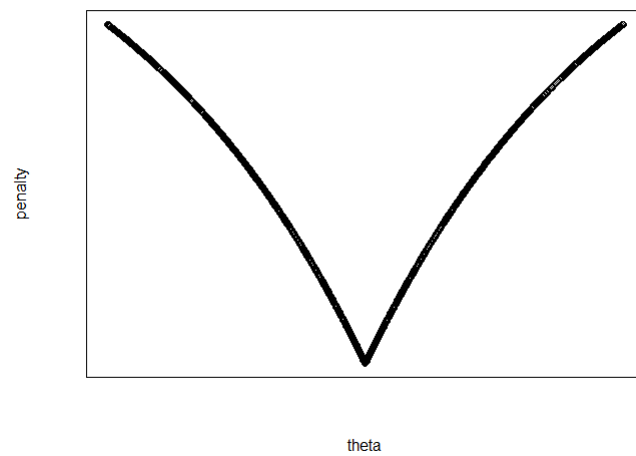


Figure 2: exponential penalty

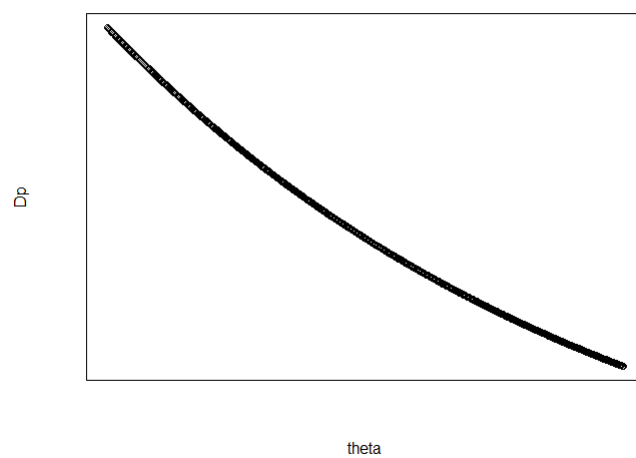


Figure 3: derivative of exponential penalty

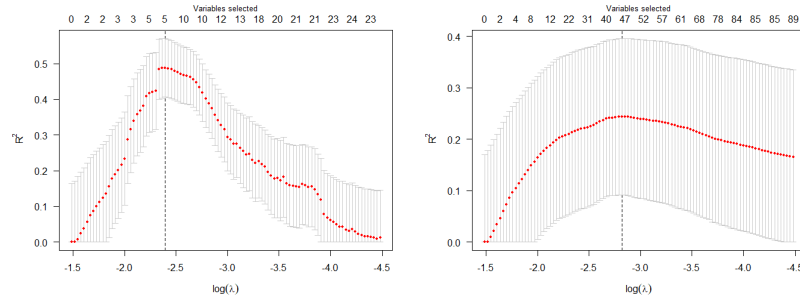


Figure 4: Estimates of R^2 with MCP (left) and Lasso (right).

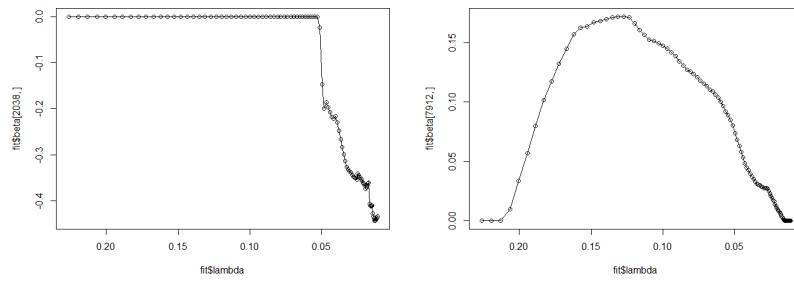


Figure 5: Fitting paths with MCP (left) and Lasso (right).