

Challenge: Multi-Stage Lasso and Post-Selection Inference

Made by 郑丽珊&宋悦溪

本部分基于一个真实模型，分别采用 Lasso 和 relaxed Lasso 对数据进行拟合，并根据拟合的结果评价两种方法。

(1) 设定真实模型

模型一：

$$Y = AX + 0.001b \text{ (总变量数}=50, \text{非零系数个数}=25)$$

$$b \sim N(0, 1),$$

$$A_1, A_3, \dots, A_{19} \sim U(-2, 2),$$

$$A_{21}, A_{23}, \dots, A_{49} \sim N(0, 1),$$

$$A_2, A_4, \dots, A_{50} = 0$$

产生系数：[0.1818806, 0., -1.31253874, 0., 1.41172724, 0.,
-1.27910636, 0., 0.3548002, 0., -1.5169004, 0., 0.218795, 0.,
-1.67189529, 0., 1.9708207, 0., -0.98912295, 0., 0.54547015, 0.,
0.17186532, 0., 0.85293181, 0., 0.18022341, 0., 0.58870005, 0.,
0.1207749, 0., 0.55469875, 0., 0.08202618, 0., 0.99270517, 0.,
0.25271926, 0., 0.53296001, 0., 0.8921391, 0., 0.24209363, 0.,
0.41762074, 0., 0.86794548, 0.]

模型二：

$$Y=AX+0.001b(\text{总变量数}=60, \text{非零系数个数}=30)$$

$$b \sim N(0, 1),$$

$$A_1, A_3, \dots, A_{19} \sim U(-2, 2),$$

$$A_{21}, A_{23}, \dots, A_{59} \sim N(0, 1)$$

$$A_2, A_4, \dots, A_{60} = 0$$

系数: [0.1818806, 0., -1.31253874, 0., 1.41172724, 0.,
-1.27910636, 0., 0.3548002, 0., -1.5169004, 0., 0.218795, 0.,
-1.67189529, 0., 1.9708207, 0., -0.98912295, 0., 0.54547015, 0.,
0.17186532, 0., 0.85293181, 0., 0.18022341, 0., 0.58870005, 0.,
0.1207749, 0., 0.55469875, 0., 0.08202618, 0., 0.99270517, 0.,
0.25271926, 0., 0.53296001, 0., 0.8921391, 0., 0.24209363, 0.,
0.41762074, 0., 0.86794548, 0., 0.28633024, 0., 0.13769966, 0.,
0.0507936, 0., 0.92829087, 0., 0.77995014, 0.]

模型三：

$$Y=AX+0.001b \text{ (总变量数}=50\text{,非零系数个数}=10\text{)}$$

$$b \sim N(0,1), A_1, A_3, \dots, A_{19} \sim U(-2,2), A_2, A_4, \dots, A_{20}, A_{21}, A_{22}, \dots, A_{50} = 0$$

系数：[0.1818806, 0., -1.31253874, 0., 1.41172724, 0.,
-1.27910636, 0., 0.3548002, 0., -1.5169004, 0., 0.218795, 0.,
-1.67189529, 0., 1.9708207, 0., -0.98912295, 0., 0., 0., 0., 0., 0., 0., 0.,
0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.]

(2) 代码 (见附件)

Lasso.py

(3) 比较 Lasso 和 relaxed Lasso

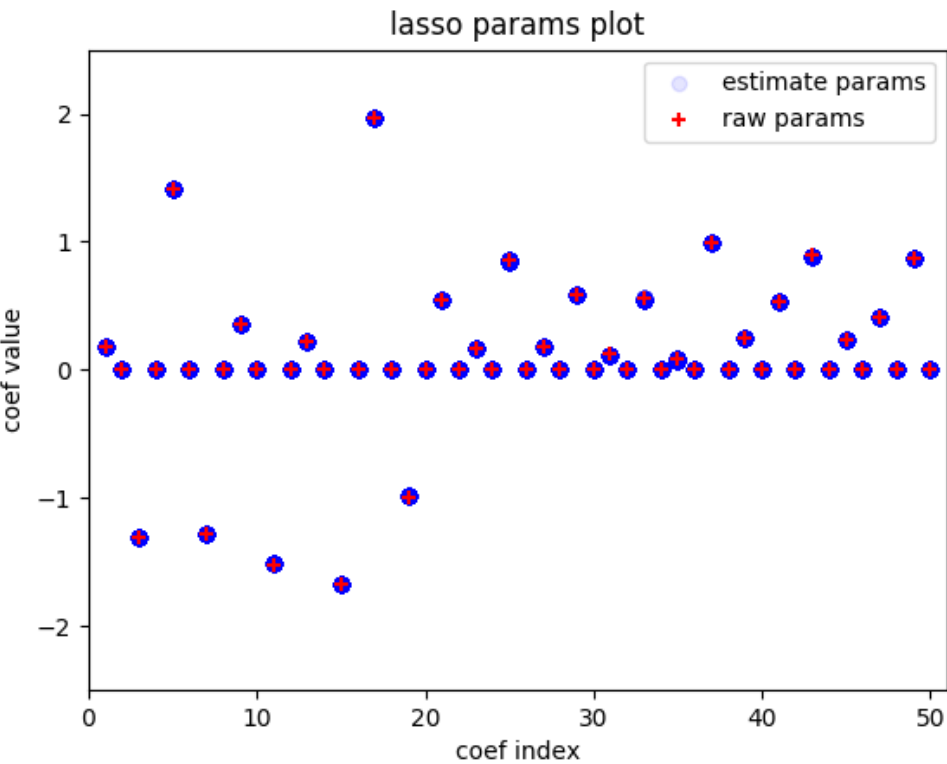
	真实模型设定	Lasso(alone) (mean test error)	Relaxed lasso (mean test error)	拟合结果较 优的模型
模型一	Size=50, 非零系数个数=25	0.000081022	0.000000272	Relaxed lasso
模型二	Size=60, 非零系数个数=30	0.000100388	0.000000272	Relaxed lasso
模型三	Size=50, 非零系数个数=10	0.000030786	0.000000257	Relaxed lasso

从表格可以看出，当 size 增加，通过模型一和模型二的对比可以得知 Relaxed lasso 方法拟合的效果更好；当 beta 的稀疏性降低，通过模型一和模型三的对比 Relaxed lasso 方法拟合的效果更好。

附件： $\hat{\beta}$ 的分布图

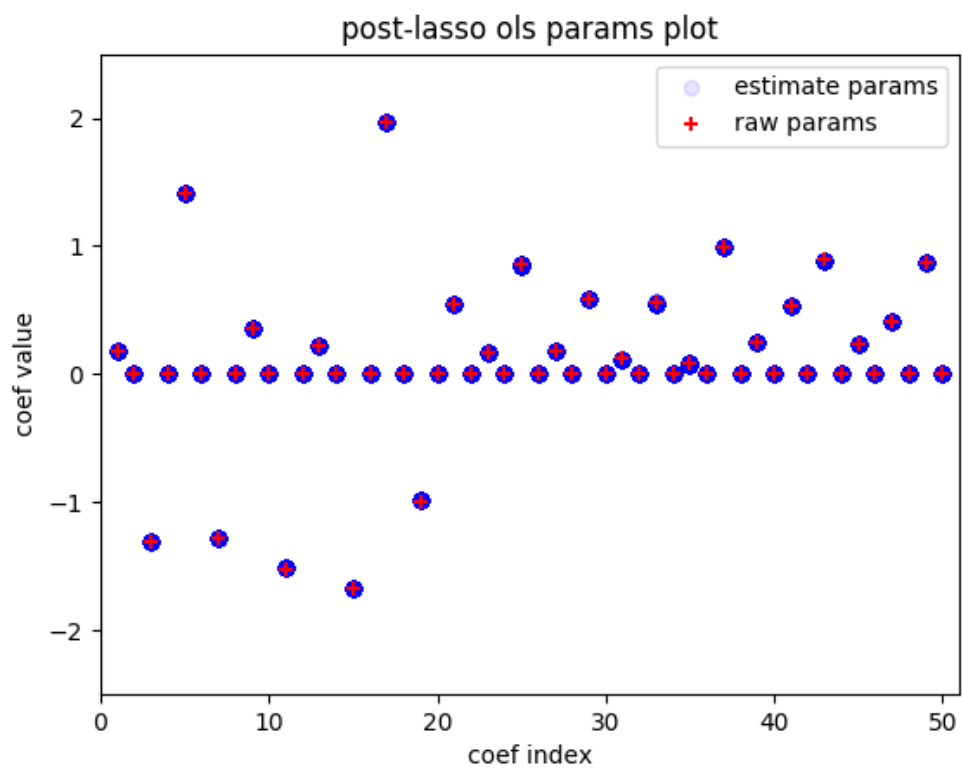
(1) 模型一

lasso 最优参数: [0.17988671 0. -1.31077005 -0. 1.40984824 0.
-1.27718015 -0. 0.35308116 0. -1.51488992 -0.
0.2170569 0. -1.67017142 0. 1.96899674 0.
-0.98792226 -0. 0.54373816 0. 0.17023463 0.
0.85099623 -0. 0.17857303 -0. 0.58691637 0.
0.11924367 -0. 0.5529339 0. 0.0798056 -0.
0.99108747 -0. 0.25060935 0. 0.53138003 -0.
0.89016251 -0. 0.24015096 0. 0.41636022 -0.
0.86590163 -0.]



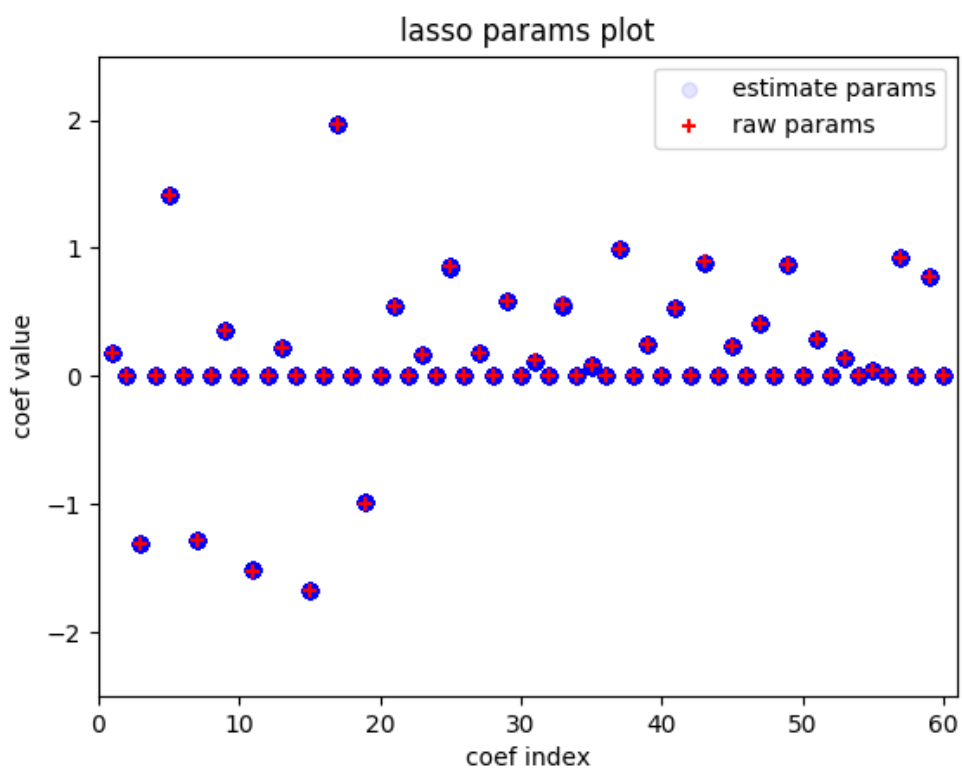
Post-lasso ols 最优参数:

```
[ 0.18188701  0.        -1.31252981  0.         1.41171863  0.
 -1.27910768  0.         0.35484739  0.        -1.51684499  0.
  0.21885817  0.        -1.67187965  0.         1.97084173  0.
 -0.98909687  0.         0.54547374  0.         0.17189064  0.
  0.85287276  0.         0.18020672  0.         0.58870699  0.
  0.12081204  0.         0.55471412  0.         0.08199522  0.
  0.99265355  0.         0.25273476  0.         0.53295756  0.
  0.892127    0.         0.24206291  0.         0.41762993  0.
  0.86790947  0.         ]
```

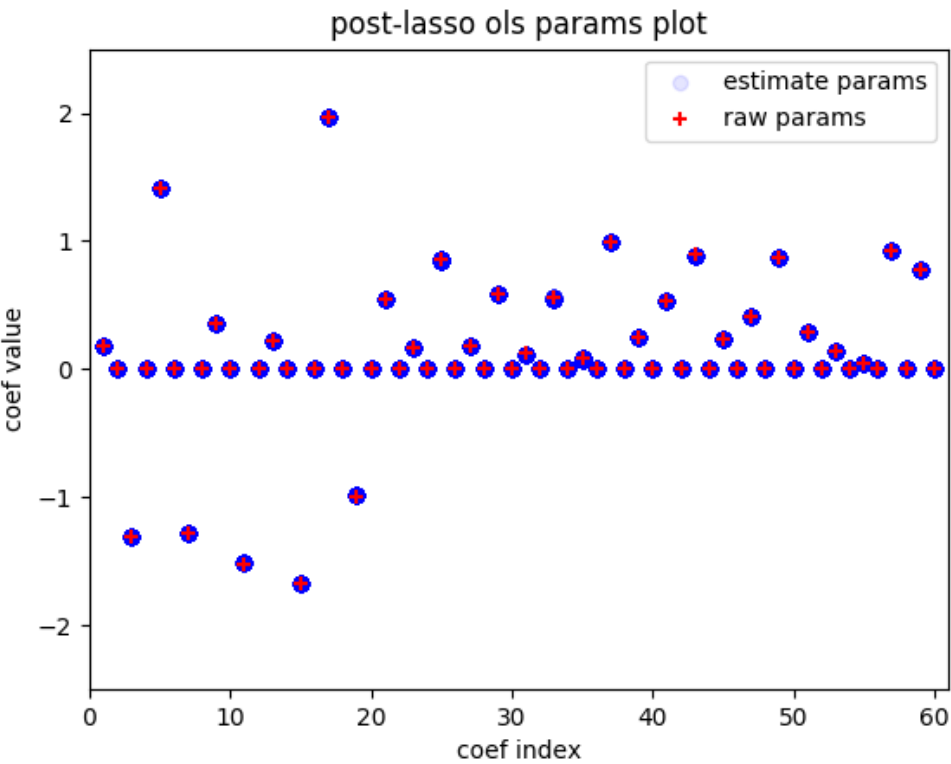


(2) 模型二

lasso 最优参数: [0.1801571 0. -1.31139 -0. 1.41010818 -0.
-1.27746943 -0. 0.3531744 -0. -1.51461514 -0.
0.21717148 0. -1.67041278 -0. 1.96899467 -0.
-0.98712032 0. 0.54361288 0. 0.16991217 -0.
0.85096305 -0. 0.178269 -0. 0.58701307 -0.
0.11880912 0. 0.55300308 -0. 0.08010884 0.
0.99050093 0. 0.25096023 -0. 0.5309818 -0.
0.89056753 -0. 0.24030336 -0. 0.41596118 -0.
0.86593893 0. 0.28515513 -0. 0.13556217 -0.
0.0486784 0. 0.92646244 0. 0.77791295
-0.]

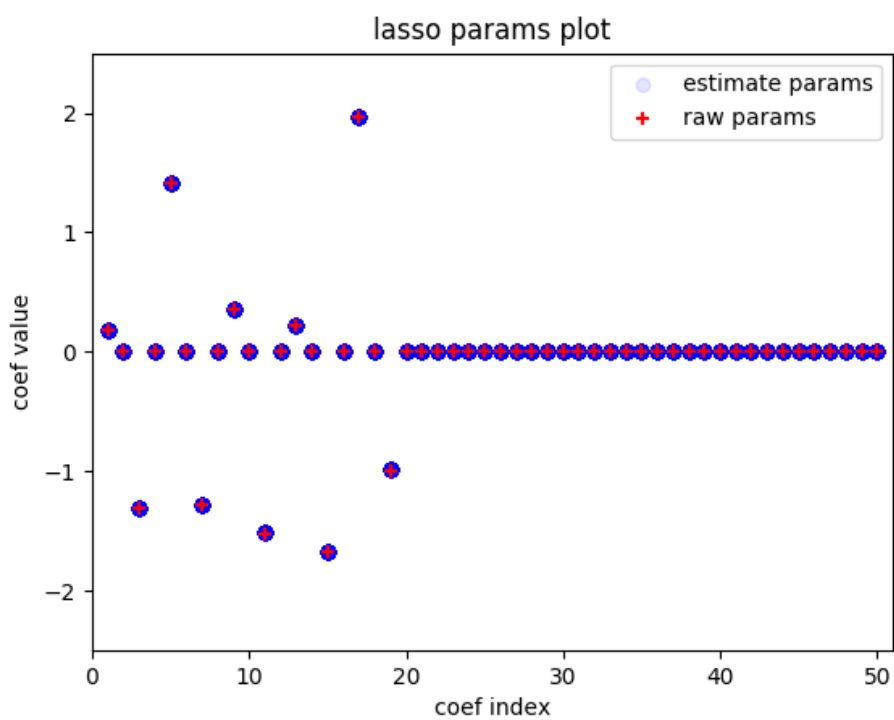


Post-lasso ols 最优参数: [0.18189983 0. -1.31257536 0. 1.4117433 0.
-1.27912017 0. 0.35483469 0. -1.51690463 0.
0.21876408 0. -1.67188295 0. 1.9708285 0.
-0.98917356 0. 0.54543665 0. 0.17186088 0.
0.85291713 0. 0.18020135 0. 0.5886292 0.
0.12080875 0. 0.55464975 0. 0.08203822 0.
0.99275004 0. 0.25270528 0. 0.53297321 0.
0.89214998 0. 0.24207581 0. 0.41761584 0.
0.86790691 0. 0.28632084 0. 0.13769229 0.
0.05083624 0. 0.92825443 0. 0.78000667 0.]



(3) 模型三

```
lasso 最优参数: [ 0.18011583 -0.          -1.31118699 -0.          1.40978615 -0.
-1.27703523  0.          0.35292332  0.          -1.51507578  0.
 0.21722016  0.          -1.67054891 -0.          1.96901942 -0.
-0.98736244 -0.          -0.          -0.          0.          0.
 0.          -0.          -0.          0.          0.          0.
-0.          -0.          -0.          -0.          0.          -0.
-0.          -0.          0.          0.          0.          -0.
-0.          0.          -0.          -0.          -0.          -0.
 0.          -0.          ]
```



Post-lasso ols 最优参数: [0.18185899 0. -1.31256331 0. 1.41174429 0.
-1.27909596 0. 0.35480425 0. -1.51692603 0.
0.21884589 0. -1.67186121 0. 1.9707994 0.
-0.98914936 0. 0. 0. 0. 0.
0. 0. 0. 0. 0. 0.
0. 0. 0. 0. 0. 0.
0. 0. 0. 0. 0. 0.
0. 0. 0. 0. 0. 0.
0. 0.]

