

基于 SCAD 正则最小一乘回归问题研究

罗孝敏,彭定涛*

(贵州大学 数学与统计学院,贵州 贵阳 550025)

摘要:针对损失函数为最小一乘,惩罚项由基数函数定义的稀疏回归问题,用 SCAD(smoothly clipped absolute deviation)罚来连续逼近基数罚,得到一个连续的松弛问题,研究 SCAD 罚问题与原基数罚问题之间解的等价性。首先,证明了 SCAD 罚松弛模型的下界性质,并借助此下界性质分析了原问题与松弛问题之间解的等价性,证明了在一定条件下两个问题具有相同的全局最优解以及最优值。此外,证明了松弛模型的局部最优解是原问题的局部最优解并且在局部极小值点处松弛模型与原问题的目标值相等。

关键词:基数罚问题;SCAD;解的等价性

中图分类号:O224 **文献标识码:**A

首先,考虑以下的 L_0 正则最小一乘问题:

$$\min_{\mathbf{x} \in \mathbf{R}^n} G(\mathbf{x}) := f(\mathbf{x}) + \lambda \|\mathbf{x}\|_0. \quad (1)$$

这里, $\|\mathbf{x}\|_0 = \#\{x_i, i = 1, \dots, n; x_i \neq 0\}$, #表示

基数, $f(\mathbf{x})$ 是损失函数, $\mathbf{x} \in \mathbf{R}^n$ 是稀疏向量, 罚参数 $\lambda > 0$ 。问题(1)可以应用在很多领域上,例如,2006年,Donoho D、Candès E、Tao T等^[1-2]针对信号重构问题,可以通过求解下面的模型:

$$\min_{\mathbf{x} \in \mathbf{R}^n} F(\mathbf{x}) := \frac{1}{2} \|\mathbf{Ax} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{x}\|_0. \quad (2)$$

由于这个问题是 NP 难的,Candès、Tao 等将 L_0 范数松弛为 L_1 范数^[2]。 L_1 罚优化问题为:

$$\min_{\mathbf{x} \in \mathbf{R}^n} F(\mathbf{x}) := \frac{1}{2} \|\mathbf{Ax} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{x}\|_1. \quad (3)$$

2001年,Fan J Q等^[3]指出 L_1 正则近似逼近 L_0 正则,得到的结果常有偏估计量。另外,提出用 SCAD 罚来近似逼近 L_0 罚,并证明了 SCAD 正则松弛优化问题具有很好的统计性质,例如有 oracle 性质,可以用于高维非参数建模;SCAD 罚能够同时实现系数估计和变量选择,其罚函数是对称的,在原点处是奇异的,在确定的条件下,其解是连续的,从而具有稀疏性和稳定性;对于较大的系数,SCAD 罚函数以确定的常数界,从而产生无偏估

计量^[4-5]。2017年,Soubies E等^[6]提出了用精确连续 L_0 (CE L_0) 罚松弛模型(4)来连续逼近问题(2)。精确连续 L_0 罚松弛模型为

$$\min_{\mathbf{x} \in \mathbf{R}^n} F(\mathbf{x}) = \frac{1}{2} \|\mathbf{Ax} - \mathbf{b}\|_2^2 + \Phi(\mathbf{x}). \quad (4)$$

这里, $\Phi(\mathbf{x})$ 为 CE L_0 罚函数。最后证明了模型(4)与问题(2)具有相同的全局最优解,局部最优解具有包含关系。因此,我们提出一个新的问题:当损失函数 $f(\mathbf{x})$ 为非光滑凸函数时是否还有相同的结论,例如 $f(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|_1$ 。这个问题是很有意义的,因为 $f(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|_1$ 广泛地应用于非均匀噪声的情况。在噪声不服从正态分布的情况下,以及对离群点数据的处理,最小一乘损失比最小二乘损失更具有鲁棒性。故针对损失函数为最小一乘,惩罚项由基数函数定义的稀疏回归问题,用 SCAD 罚来连续逼近基数罚,得到一个连续的松弛问题,研究这两个问题解的关系。

原始基数罚问题:

$$\min_{\mathbf{x} \in \mathbf{R}^n} G(\mathbf{x}) := \|\mathbf{Ax} - \mathbf{b}\|_1 + \lambda \|\mathbf{x}\|_0. \quad (5)$$

其中, $\mathbf{A} = [\mathbf{A}_1, \dots, \mathbf{A}_n] \in \mathbf{R}^{m \times n}$, $\mathbf{A}_i \in \mathbf{R}^m$ 表示矩阵 $\mathbf{A}$ 的第 i 列元素组成的向量, $\mathbf{b} \in \mathbf{R}^m$ 。

SCAD 罚松弛优化问题为:

收稿日期:2020-01-14

基金项目:国家自然科学基金资助项目(11861020);贵州省高层次留学人才创新创业择优资助重点项目([2018]03);贵州省科技计划资助项目([2018]5781);贵州省青年科技人才成长资助项目([2018]121)

作者简介:罗孝敏(1993-),女,在读硕士,研究方向:稀疏优化,Email:lxm2440775499@163.com.

* 通讯作者:彭定涛,Email:dingtaopeng@126.com.

$$\min_{x \in \mathbf{R}^n} F(x) := \|Ax - b\|_1 + \Phi(x). \quad (6)$$

其中, $\Phi(x) = \sum_{i=1}^n \varphi(\theta, \lambda; x_i)$, 对参数 $\theta > 2$, $\lambda > 0$, SCAD 罚函数为:

$$\varphi(\theta, \lambda, x_i) = \begin{cases} \lambda |x_i| & |x_i| \leq \lambda \\ \frac{2\theta\lambda|x_i| - x_i^2 - \lambda^2}{2(\theta - 1)} & \lambda < |x_i| \leq \theta\lambda \\ \frac{\lambda^2(\theta + 1)}{2} & |x_i| > \theta\lambda \end{cases}.$$

SCAD 罚函数满足下列的性质:

- (i) 当 $x_i > 0$, $\varphi(\theta, \lambda; x_i)$ 是不减的、凹的函数;
- (ii) 对 $x_i \in \mathbf{R}$, 有 $\varphi(\theta, \lambda; x_i) > 0$, 及 $\varphi(\theta, \lambda; 0) = 0$;
- (iii) 对 $\forall x_i \in \mathbf{R} \setminus \{0\}$, $\varphi(\theta, \lambda; x_i)$ 是可微的且有

$$0 \leq \varphi'(\theta, \lambda; x_i) \leq \lambda,$$

$$\varphi'(\theta, \lambda, x_i) = \begin{cases} \lambda \operatorname{sign}(x_i) & |x_i| \leq \lambda \\ \frac{\theta\lambda \operatorname{sign}(x_i) - x_i}{\theta - 1} & \lambda < |x_i| \leq \theta\lambda \\ 0 & |x_i| > \theta\lambda \end{cases}.$$

这里

$$\operatorname{sign}(x_i) = \begin{cases} -1 & x_i < 0 \\ 0 & x_i = 0 \\ 1 & x_i > 0 \end{cases}.$$

图1是一维情形下 L_0, L_1, SCAD 罚函数的图像。由 SCAD 函数的定义及图1可知: 当 $|x_i| \leq \lambda$ 时, SCAD 罚与 L_1 罚是相同的; 当 $\lambda < |x_i| \leq \theta\lambda$ 时, 随着特征系数 $|x_i|$ 的增大而减小压缩的程度; 当 $|x_i| > \theta\lambda$ 时, SCAD 罚对特征系数不再进行压缩。SCAD 罚克服了 L_1 罚有偏估计的缺点^[3]。

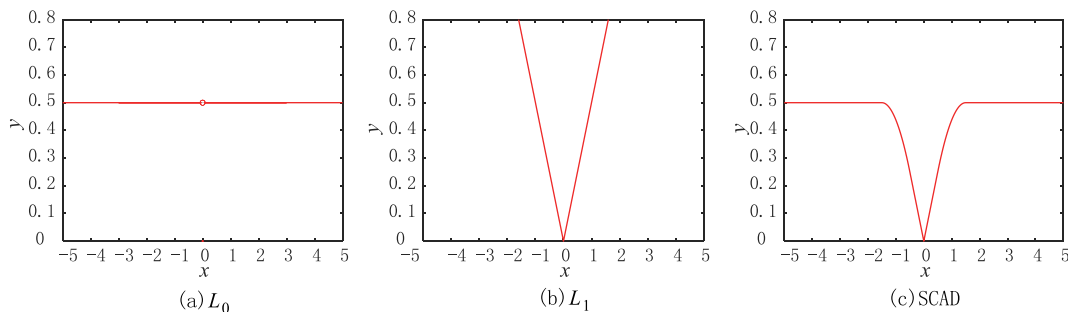


图1 L_0, L_1, SCAD 函数图像

Fig.1 Function images of L_0, L_1, SCAD

本文主要研究原始基数罚问题(5)和 SCAD 罚松弛优化问题(6)之间解的关系。第一部分主要介绍需要用到的预备知识; 第二部分主要对 SCAD 罚问题证明了下界理论性质, 并在一定条件下, 证明了问题(6)与模型(5)有相同的全局最优解以及最优值, 松弛模型(6)的局部最优解是问题(5)的局部最优解, 且在局部极小值点处问题(5)和问题(6)的最优值是相等的; 最后是一个简单的总结。

1 记号

对给定的 $x^* \in \mathbf{R}^n$, x^* 的支撑集为

$$T(x) := \{i \in \{1, \dots, n\} : x_i^* \neq 0\}.$$

此外, $\|x\|_1 := \sum_{i=1}^n |x_i|$, $\|x\|_2 := (\sum_{i=1}^n |x_i|^2)^{1/2}$, $\|x\|_\infty = \max_i \{|x_i|\}$ 依次表示向量 x 的 1-范数、2-范数、 ∞ -范数。

次微分 $\partial \|x\|_1 = (\partial \|x_1\|_1, \dots, \partial \|x_n\|_1)$, 其中

$$\partial \|x_i\|_1 = \begin{cases} 1 & x_i > 0 \\ [-1, 1] & x_i = 0 \\ -1 & x_i < 0 \end{cases}$$

SCAD 罚函数的 Clarke 次微分为:

$$\partial \varphi(\theta, \lambda, x_i) = \begin{cases} \varphi'(\theta, \lambda, x_i) & x_i \neq 0 \\ [-\theta\lambda, \theta\lambda] & x_i = 0 \end{cases}.$$

定义1 称 $x^* \in \mathbf{R}^n$ 是问题(6)的一阶稳定点, 若

$$0 \in (A^T \operatorname{sgn}(Ax^* - b))_i + \partial \varphi(x_i^*).$$

命题1 若 $x^* \in \mathbf{R}^n$ 是问题(6)的局部最优解, 则

$x^* \in \mathbf{R}^n$ 是问题(6)的一阶稳定点。

证明 注意到最小一乘损失函数 $\|Ax - b\|_1$ 是关于 x 的 Lipschitz 函数, 由文[7]中的 Exercise 10.10, 命题成立。

2 SCAD 罚松弛问题与原问题解的关系

在这一节里, 建立了 SCAD 松弛问题的下界理

论性质,并分析了问题(5)和(6)在一定条件下解(包括局部解和全局解)的关系。

引理 1 设 $\lambda \geq \max_{1 \leq i \leq n} \|A_i\|_1$, 如果 x^* 是问题(6)的稳定点, 则有

$$x_i^* \in [-\lambda, \lambda] \Rightarrow x_i^* = 0.$$

证明 假设 $x_i^* \in [-\lambda, \lambda] \setminus \{0\}$, 则由定义 1 可知

$$0 \in (A^T \operatorname{sgn}(Ax^* - b))_i + \lambda \operatorname{sgn}(x_i^*).$$

因为 $\operatorname{sgn}(x_i^*) \neq 0$, 且 $|\operatorname{sgn}(x_i^*)| = 1$, 则上式可为

$$-\lambda \operatorname{sgn}(x_i^*) \in (A^T \operatorname{sgn}(Ax^* - b))_i,$$

且

$$|(A^T \operatorname{sgn}(Ax^* - b))_i| \leq \|A_i\|_1,$$

故

$$|-\lambda \operatorname{sgn}(x_i^*)| \leq \|A_i\|_1.$$

则 $|\lambda| \leq \|A_i\|_1 \leq \max_{1 \leq i \leq n} \|A_i\|_1$ 与假设矛盾, 故有

$$x_i^* \in [-\lambda, \lambda] \Rightarrow x_i^* = 0.$$

引理 2 对 $\lambda > 0, \theta > 2, x \in \mathbf{R}^n, (\theta + 1)\lambda = 2$, 有 $\Phi(x) \leq \lambda \|x\|_0$.

证明 由 $\Phi(x) = \sum_{i=1}^n \varphi(\theta, \lambda, x_i)$, 以及 $\varphi(\theta, \lambda, x_i)$ 的定义可知:

$$\begin{aligned} \varphi(\theta, \lambda, 0) &= 0 \leq \varphi(\theta, \lambda, x_i) \\ &\leq \frac{(\theta + 1)\lambda^2}{2}. \end{aligned}$$

又由 $(\theta + 1)\lambda = 2$ 可知 $\frac{(\theta + 1)\lambda^2}{2} = \lambda$, 即

$$\varphi(\theta, \lambda, 0) = 0 \leq \varphi(\theta, \lambda, x_i) \leq \lambda.$$

因此

$$\varphi(\theta, \lambda, x_i) \leq \lambda |\operatorname{sign}(x_i)|.$$

对上式两边分别求和, 有

$$\sum_{i=1}^n \varphi(\theta, \lambda, x_i) \leq \lambda \sum_{i=1}^n |\operatorname{sign}(x_i)|,$$

故 $\Phi(x) \leq \lambda \|x\|_0$.

定理 1 设 $\lambda \geq \max_{1 \leq i \leq n} \|A_i\|_1, \|b\|_1 < \lambda^2, (\theta + 1)\lambda = 2, x^* \in \mathbf{R}^n$ 是问题(5)的全局解, 当且仅当 $x^* \in \mathbf{R}^n$ 时是问题(6)的全局解; 且问题(5)与(6)具有相同的最优值。

证明 令 x^* 为问题(6)的全局最优解, 则由命题 1 可知 x^* 是问题(6)的稳定点。若 $x_i^* \in [-\lambda, \lambda]$, 则由引理 1 可知 $x_i^* = 0$ 。

对 $x_i \in [-\theta\lambda, \theta\lambda] \setminus [-\lambda, \lambda]$, 由 $\varphi(\theta, \lambda, x_i^*)$ 的定义可知:

$$\varphi(\theta, \lambda, x_i^*) > \varphi(\theta, \lambda, \lambda) = \lambda^2,$$

所以

$$\Phi(x^*) \geq \varphi(\theta, \lambda, x_i^*) > \lambda^2.$$

故

$$\begin{aligned} F(0) &= \|b\|_1 \leq \|Ax^* - b\|_1 + \lambda^2 \\ &< \|Ax^* - b\|_1 + \Phi(x^*) = F(x^*). \end{aligned}$$

这与 x^* 是问题(6)的全局最优解矛盾。因此, 对 $x_i^* \in [-\theta\lambda, \theta\lambda]$ 有 $x_i^* = 0$ 。

若 $|x_i^*| > \theta\lambda$, 由 $(\theta + 1)\lambda = 2$ 以及 $\varphi(\theta, \lambda, x_i^*)$ 的定义可知, $\varphi(\theta, \lambda, x_i^*) = \lambda$ 。综上可知

$$\Phi(x^*) = \lambda \|x^*\|_0.$$

由引理 2 可知, $\forall x \in \mathbf{R}^n$, 有 $\Phi(x) \leq \lambda \|x\|_0$, 则

$$\begin{aligned} \|Ax^* - b\|_1 + \lambda \|x^*\|_0 &= \|Ax^* - b\|_1 + \Phi(x^*) \\ &\leq \|Ax - b\|_1 + \Phi(x) \\ &\leq \|Ax - b\|_1 + \lambda \|x\|_0, \end{aligned}$$

故 x^* 是问题(5)的全局最优解。

反之, 设 $x^* \in \mathbf{R}^n$ 是问题(5)的全局最优解, 但不是问题(6)的全局最优解, 则问题(6)存在一个全局最优解 $\hat{x} \in \mathbf{R}^n$, 使得

$$\|A\hat{x} - b\|_1 + \Phi(\hat{x}) < \|Ax^* - b\|_1 + \Phi(x^*).$$

同理可得 $\Phi(\hat{x}) = \lambda \|\hat{x}\|_0$, 及 $\Phi(x^*) \leq \lambda \|x^*\|_0$, 故

$$\begin{aligned} \|A\hat{x} - b\|_1 + \lambda \|\hat{x}\|_0 &= \|A\hat{x} - b\|_1 + \Phi(\hat{x}) \\ &< \|Ax^* - b\|_1 + \Phi(x^*) \\ &\leq \|Ax^* - b\|_1 + \lambda \|x^*\|_0. \end{aligned}$$

这与 x^* 是问题(5)的全局最优解矛盾, 所以 x^* 是问题(5)的全局最优解, 并且由于 $\Phi(x^*) = \lambda \|x^*\|_0$, 从而有 $G(x^*) = F(x^*)$ 。即问题(5)与(6)具有相同的最优值。此定理表明了问题(5)与(6)具有相同的全局最优解集合以及最优值。

下面研究问题(5)与(6)局部最优解之间的关系。

定理 2 设 $\lambda \geq \max_{1 \leq i \leq n} \|A_i\|_1, \|b\|_1 < \lambda^2, (\theta + 1)\lambda = 2$, 若 $x^* \in \mathbf{R}^n$ 是模型(6)的一个局部最优解, 则 x^* 是问题(5)的局部解且在 x^* 处与问题(6)有相同的最优值。即

$$G(x^*) = F(x^*).$$

证明 设 $x^* \in \mathbf{R}^n$ 是问题(6)的局部最优解, 由命题 1 可知 x^* 是问题(6)的稳定点。因此, 与证明定理 1 类似可得:

$$\Phi(x^*) = \lambda \|x^*\|_0.$$

由于 x^* 是问题(6)的局部最优解, 故存在 x^* 的邻域 $N(x^*)$, 使得对 $\forall x \in N(x^*)$, 有

$$F(\mathbf{x}^*) \leq F(\mathbf{x}).$$

又由引理2可知,对 $\forall \mathbf{x} \in \mathbf{R}^n$, 有 $\Phi(\mathbf{x}) \leq \lambda \|\mathbf{x}\|_0$, 故对 $\forall \mathbf{x} \in N(\mathbf{x}^*)$, 有

$$\begin{aligned} \|\mathbf{Ax}^* - \mathbf{b}\|_1 + \lambda \|\mathbf{x}^*\|_0 &= \|\mathbf{Ax}^* - \mathbf{b}\|_1 + \Phi(\mathbf{x}^*) \\ &\leq \|\mathbf{Ax} - \mathbf{b}\|_1 + \Phi(\mathbf{x}) \\ &\leq \|\mathbf{Ax} - \mathbf{b}\|_1 + \lambda \|\mathbf{x}\|_0. \end{aligned}$$

因此可知, $\mathbf{x}^*$ 是问题(5)的局部最优解且 $G(\mathbf{x}^*) = F(\mathbf{x}^*)$, 即问题(5)与(6)在 $\mathbf{x}^*$ 处有相同的目标值。

定理2表明了问题(6)的局部最优解是问题(5)的局部最优解,在局部最优值点处问题(5)与问题(6)的最优值相等。

最后,在满足一定条件下针对问题(5)与问题(6)解的关系,用图2做一个简短的总结。

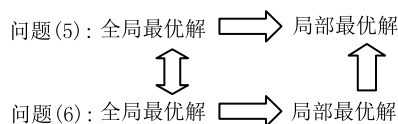


图2 问题(5)与(6)的关系

Fig2 Relation of between problems(5)and(6)

3 总结

本文主要介绍基于SCAD罚的最小一乘问题(6)与原问题(5)之间解的等价性。证明了松弛模型的下界理论性质,并在此下界理论性质下分析了原问题与松弛问题之间解的等价性。在一定条件

下,证明了问题(5)与问题(6)具有相同的全局最优解以及最优值。此外,松弛模型的局部最优解是原问题的局部最优解,在局部极小值点处松弛模型与原问题的最优值是相等的。这为进一步设计有效的算法提供了理论基础。

参考文献:

- [1] DONOHO D L. Compressed sensing[J]. IEEE Transactions Information Theory, 2006, 52(4): 1289-1306.
- [2] CANDÈS E J, ROMBERG J, TAO T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information[J]. IEEE Transactions on Information and Theory, 2006, 52(2): 489-509.
- [3] FAN J Q, LI R Z. Variable selection via nonconcave penalized likelihood and its oracle properties[J]. Amer Statist Assoc, 2001, 9: 1348-1360.
- [4] FAN J Q, PENG H. Nonconcave penalized likelihood with a diverging number of parameters[J]. The Annals of Statistics, 2004, 32(3): 928-961.
- [5] LIU H C, YAO T, LI R Z, et al. Folded concave penalized sparse linear regression: sparsity, statistical performance, and algorithmic theory for local solution[J]. Mathematical Programming, 2017, 166(1-2): 207-240.
- [6] SOUBIES E, BLANC-FERAUD L, AUBERT G. A unified view of exact continuous penalties for l2-l0 minimization[J]. SIAM Journal on Optimization, 2017, 27(3): 2034-2060.
- [7] ROCKAFWLLAR R T, WETS R J B. Variational analysis[M]. Berlin: Springer, 1998.

(责任编辑:曾 晶)

Smoothly Clipped Absolute Deviation (SCAD) for Least Absolute Deviation Regression Regularized Problem

LUO Xiaomin, PENG Dingtao*

(School of Mathematics and Statistics, Guizhou University, Guiyang 550025, China)

Abstract: For the sparse regression problem where the loss function is the least absolute deviation and the penalty term is the cardinal penalty, we use SCAD (smoothly clipped absolute deviation) penalty to relax the cardinal penalty. We focus on the equivalence of solutions between the relaxed problem and the original problem. Firstly, the lower bound theory property of the relaxed model is proved, and the equivalence between the original problem and the relaxed problem is analyzed under the lower bound property. It is proved that the two problems have the same global optimal solution and optimal value under certain conditions. In addition, it is proved that the local optimal solution of the relaxed model is the local optimal solution of the original problem, and the relaxed module is at the local minimum point. The optimal value of type A is equal to that of the original problem.

Key words: cardinal penalty problem; SCAD; equivalence of solutions