

含高相关协变量的纵向广义线性模型变量选择研究

范涛^{1*}, 赵培信^{1,2}

¹重庆工商大学数学与统计学院, 重庆

²统计智能计算与检测重庆市重点实验室, 重庆

收稿日期: 2024年6月15日; 录用日期: 2024年7月9日; 发布日期: 2024年7月17日

摘要

结合半标准偏协方差函数(SPAC)和Lasso惩罚估计方法, 对一类纵向数据下的广义线性模型的变量选择问题提出一种基于SPAC-Lasso惩罚的变量选择方法。在一些正则性条件下证明了所提出的变量选择方法的相合性, 并给出了所得正则估计的收敛速度。所提出的变量选择方法允许协变量之间存在高相关性, 改进并推广了已有变量选择方法的应用领域。

关键词

广义线性模型, 高相关协变量, 纵向数据, 变量选择

Variable Selection in Longitudinal Generalized Linear Models with Highly Correlated Covariates

Tao Fan^{1*}, Peixin Zhao^{1,2}

¹School of Mathematics and Statistics, Chongqing Technology and Business University, Chongqing

²Chongqing Key Laboratory of Statistical Intelligent Computing and Monitoring, Chongqing

Received: Jun. 15th, 2024; accepted: Jul. 9th, 2024; published: Jul. 17th, 2024

Abstract

Combining the semi-parametric approximate covariance function (SPAC) and Lasso penalized es-

*第一作者。

文章引用: 范涛, 赵培信. 含高相关协变量的纵向广义线性模型变量选择研究[J]. 应用数学进展, 2024, 13(7): 3175-3181. DOI: 10.12677/aam.2024.137303

timization method, we propose a variable selection approach based on SPAC-Lasso penalty for a class of longitudinal generalized linear models. Under some regularity conditions, we demonstrate the consistency of the proposed variable selection method and provide the convergence rate of the resulting regularized estimates. The proposed variable selection method allows for high correlations among covariates, improving and extending the applicability of existing variable selection methods.

Keywords

Generalized Linear Models, Highly Correlated Covariates, Longitudinal Data, Variable Selection

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

纵向数据常常出现在生物医学、社会经济以及金融工程等领域的跟踪调查研究中。纵向数据下的广义线性模型可以对跟踪调查中的连续性协变量与离散响应变量之间有效地建立一种量化关系, 因此具有较好的适应性和广泛的应用领域。具体地, 纵向数据下的广义线性模型具有如下结构:

$$E(Y_{ij} | X_{ij}) = g(X_{ij}^T \beta), i = 1, \dots, n; j = 1, \dots, m \quad (1.1)$$

其中 Y_{ij} 是响应变量, $X_{ij} = (X_{ij1}, \dots, X_{ijp})^T$ 是一个 p 维协变量, $\beta = (\beta_1, \dots, \beta_p)^T$ 为未知回归系数向量, $g(\cdot)$ 为一个已知的连接函数的逆。不失一般性, 本文考虑正则链接函数, 即假定 $g(\cdot) = b'(\cdot)$, 其中 $b'(\cdot)$ 是连接函数 $b(\cdot)$ 的一阶导数。

对于纵向数据下的广义线性模型, Avella-Medina 和 Ronchetti [1] 给出了一个稳健的惩罚拟似然估计, 并证明了所提出的方法的 oracle 性质。WANG 和 TIAN [2] 提出了一种自适应分组 Lasso 变量选择过程, 并证明了相应的选择一致性。但是在这些变量选择方法中, 均假定模型中的重要变量与不重要变量之间是相互独立的, 当重要协变量与不重要协变量之间存在高度相关性时, 上述给出的 Lasso 惩罚方法则不能有效地识别出重要协变量以及不重要协变量。

为此, Bühlmann [3] 等人引入了部分信赖度的概念, 并提出了基于部分相关性的 PC-SIMPLE 算法来选择重要协变量。JIA 和 Rohe [4] 研究了 Puffer 变换方法, 并证明了改进后的 Pre-Lasso 变量选择过程可以满足变量选择的一致性。但是 JIA 和 Rohe [4] 中的预处理技术可能会夸大误差项的方差, 使变换后的观测值往往包含相关误差, 从而导致偏相关方法对获得相关协变量的信号强度无效 [3]-[5]。而 XUE 和 QU [6] 提出的半标准偏协方差函数 (SPAC) 方法则可以在结合系数强度的同时, 减少其他协变量相关性的影响。

基于此, 本文将进一步基于 SPAC 方法的变量选择方法, 对纵向数据下的广义线性模型, 提出一种基于 SPAC-Lasso 惩罚的变量选择方法, 并且在一些正则性条件下, 证明了所提出的 SPAC-Lasso 变量选择方法的相合性, 并给出了正则估计量的收敛速度。

2. 基于 SPAC-Lasso 变量选择过程

假设模型 (1.1) 中的回归系数 β 是稀疏的, 即与响应变量相关的重要协变量很少, 其对应的系数不为零。进一步设重要协变量的个数为 q , 则当 $1 \leq k \leq q$ 时 $\beta_k \neq 0$, 当 $q < k \leq p$ 时 $\beta_k = 0$ 。类似 XUE 和 QU [6],

定义响应变量 Y 与第 k 个协变量 X_k 之间的 SPAC 如下:

$$\gamma_k = \beta_k / \omega_{kk}^{1/2}, k=1, \dots, p \quad (2.1)$$

其中 ω_{kk} 是矩阵 $\Omega = \Sigma^{-1}$ 的第 k 个对角线元素, $\Sigma = \text{Var}(X_{ij})$ 表示 X_{ij} 的协方差矩阵。注意到 $\gamma_k = 0$ 当且仅当 $\beta_k = 0$, $k=1, \dots, p$, 所以可以通过识别 γ_k 是否为零来进行选择模型的重要协变量。由(2.1)可知 $\beta_k = \gamma_k \omega_{kk}^{1/2}$, $k=1, \dots, p$, 进而再由模型(1.1)可得:

$$E(Y_{ij} | X_{ij}) = g\left(\sum_{k=1}^p X_{ijk} \gamma_k \omega_{kk}^{1/2}\right), i=1, \dots, n; j=1, \dots, m \quad (2.2)$$

记 $\gamma = (\gamma_1, \dots, \gamma_p)^T$, 则 γ 的惩罚估计 $\hat{\gamma} = (\hat{\gamma}_1, \dots, \hat{\gamma}_p)^T$ 可以通过最小化如下带惩罚的目标函数得到:

$$L(\gamma, \hat{\omega}) = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m \left\{ Y_{ij} \sum_{k=1}^p x_{ijk} \gamma_k \sqrt{\hat{\omega}_{kk}} - b\left(\sum_{k=1}^p x_{ijk} \gamma_k \sqrt{\hat{\omega}_{kk}}\right) \right\} + \sum_{k=1}^p p_\lambda(\gamma_k) \hat{\omega}_{kk}, \quad (2.3)$$

其中 $\hat{\omega} = (\hat{\omega}_{11}, \dots, \hat{\omega}_{pp})^T$ 是 $\omega = (\omega_{11}, \dots, \omega_{pp})^T$ 的相合估计量。注意到 $\Omega = \Sigma^{-1}$ 以及 $\Sigma = \text{Var}(X_{ij})$, 则可以利用矩估计方法给出 Σ 的估计, 进而得到 ω 的估计。进一步, 如果惩罚函数 $p_\lambda(\cdot)$ 取为 Lasso 惩罚, 则 γ 的 Lasso 惩罚估计量定义为:

$$\hat{\gamma} = \arg \min_{\gamma} \left[-\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m \left\{ Y_{ij} \sum_{k=1}^p x_{ijk} \gamma_k \sqrt{\hat{\omega}_{kk}} - b\left(\sum_{k=1}^p x_{ijk} \gamma_k \sqrt{\hat{\omega}_{kk}}\right) \right\} + 2\lambda \sum_{k=1}^p \hat{\omega}_{kk} |\gamma_k| \right] \quad (2.4)$$

进一步记 $\tilde{X}_{ijk} = X_{ijk} / \sqrt{\hat{\omega}_{kk}}$, $\gamma_k^* = \gamma_k \hat{\omega}_{kk}$, 则(2.3)式中的惩罚似然函数可进一步改写为:

$$\mathbb{P}\ell_*(\gamma^*) = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m \left\{ Y_{ij} \sum_{k=1}^p \tilde{X}_{ijk} \gamma_k^* - b\left(\sum_{k=1}^p \tilde{X}_{ijk} \gamma_k^*\right) \right\}$$

并且结合(2.4)式可得 γ_k^* SPAC-Lasso 估计量如下:

$$\hat{\gamma}^* = \arg \min_{\gamma^*} \left(\mathbb{P}\ell_*(\gamma^*) + 2\lambda \sum_{k=1}^p |\gamma_k^*| \right)$$

3. 渐进性质及证明

为表述方便, 设 X_{ij} 的前 q 个元素为模型的重要协变量, 记为 X_I , 并且 X_{ij} 的后 $p-q$ 个元素为不重要协变量, 记为 X_{II} 。进一步, 记 $\hat{C}(\omega) = (1/n) X_{ij}^T H(\gamma_0, \omega) X_{ij}$, $\hat{V}^{-1} = \text{diag}\{\omega_{11}^{1/2}, \dots, \omega_{pp}^{1/2}\}$, 且矩阵 $\hat{C}(\omega)$ 可以划分为以下块结构:

$$\hat{C}(\omega) = \begin{pmatrix} \hat{C}_{11}(\omega) & \hat{C}_{12}(\omega) \\ \hat{C}_{21}(\omega) & \hat{C}_{22}(\omega) \end{pmatrix}, \quad (3.1)$$

类似 XUE 和 QU [6], $\hat{C}(\omega)$ 假定满足如下条件, 即存在一个正常向量 η 使得:

$$|V_{II} \hat{C}_{21}(\omega) \hat{C}_{11}^{-1}(\omega) V^{-1} \text{sign}(\gamma_1)| \leq 1 - \eta, \quad (3.2)$$

其中 $V_I = \text{diag}\{1/\omega_{11}^{1/2}, \dots, 1/\omega_{qq}^{1/2}\}$, $V_{II} = \text{diag}\{1/\omega_{q+1,q+1}^{1/2}, \dots, 1/\omega_{pp}^{1/2}\}$, 1 是 1 的 $(p-q) \times 1$ 向量, $||$ 表示取每个元素的绝对值。Bunea [7] 和 Blazere [8] 在设计矩阵上提出了如下 Stabil 条件: 对于 $c, \varepsilon > 0$, 记限制集为 $C(c, \varepsilon) = \{\zeta \in \mathbb{R}^p : \|\zeta_{II}\|_1 \leq c \|\zeta_I\|_1 + \varepsilon\}$, 那么对于任何 $\zeta \in C(c, \varepsilon)$, 则存在 $0 < k < 1$ 使得协方差矩阵满足:

$$\zeta^T \Sigma \zeta \geq k \|\zeta_I\|_2^2 - \varepsilon, \quad (3.3)$$

另外为了得到 SPAC-Lasso 的变量选择相合性和正则估计的收敛速度, 还需要一些正则性条件。

(C1) 存在常数 $M_1 > 0$ 使得:

$$\max_{1 \leq ij \leq n} \max_{1 \leq k \leq p} |X_{ijk}| \leq M_1$$

(C2) 存在一个常数 $B > 0$ 使得 $\|\beta_0\|_1 \leq B$ 。

(C3) $\hat{C}(\omega)$ 的特征值有界且均大于零。

(C4) 存在一个常数 $M_2 > 0$ 和一个包含 γ_0^* 的开子集 $\mathbb{B}_n$, 使得 $|b^{(3)}(\tilde{X}_{ij}^T \gamma^*)| \leq M_2$ 对所有 $\gamma^* \in \mathbb{B}_n$ 和几乎所有 $\tilde{X}_{ij}$ 都成立, 其中 $b^{(3)}(\cdot)$ 表示 $b(\cdot)$ 的三阶导数。

(C5) 当 $n \rightarrow \infty$, 假设 $q\sqrt{\log p/n} \rightarrow 0$ 和 $\sqrt{\log n/n} = o(\lambda)$ 。

(C6) 对于任何常数 $A > \sqrt{2}$, 存在常数 $M_3 > 0$ 使得:

$$\Pr\left(\max_{1 \leq k \leq p} |\hat{\omega}_{kk} - \omega_{kk}| \leq M_3 \sqrt{\log n/n}\right) \geq 1 - (2p)^{-A^2}$$

条件(C1)~(C4)是广义线性模型的正则性条件[3]-[5], 详见 WANG 和 TIAN [2]、CUI [5]、Bunea [7]、Blazere [8]、FAN 和 PENG [9]。条件(C5)是证明主要定理所必需的, 在其它文献中也很常见, 如 PENG [10] 和 VandeGeer [11]等人的条件(C6)假定了精度矩阵元素的估计量 $\hat{\omega}_{kk}$ 的收敛速度, 该条也类似 XUE 和 QU [6]和 PENG [8]的情况。如下定理 3.1 表明所提出的 SPAC-Lasso 方法可以得到参数 β 的相合估计, 并给出了正则估计的收敛速度。

定理 3.1 假设正则性条件(C1)~(C6)成立, 那么在条件(3.2)和(3.3)下, 估计量 $\hat{\beta}$ 满足:

$$\|\hat{\beta} - \beta_0\|_1 \leq \frac{4}{c_n k \sqrt{c_0}} \lambda q + \left(1 + \frac{1}{\lambda}\right) \frac{1}{n \sqrt{c_0}},$$

其中 $c_n = \min_{|x| < M_1 \left(9B + \frac{2}{n}\right)} \{b''(x)/2\}$ 且 $c_0 > 0$ 。

证明: 根据 $\hat{\gamma}^*$ 的定义和文献[8], 我们有:

$$\mathbb{P}_n \ell_*(\hat{\gamma}^*) + 2\lambda \|\hat{\gamma}^*\|_1 \leq \mathbb{P}_n \ell_*(\gamma_0^*) + 2\lambda \|\gamma_0^*\|_1,$$

通过将 $\mathbb{P}_n(\ell_*(\hat{\gamma}^*) - \ell_*(\gamma_0^*)) + \lambda \|\hat{\gamma}^* - \gamma_0^*\|_1$ 加到上述不等式的两侧并重新排列, 我们得到:

$$\begin{aligned} & \mathbb{P}_n(\ell_*(\hat{\gamma}^*) - \ell_*(\gamma_0^*)) + \lambda \|\hat{\gamma}^* - \gamma_0^*\|_1 \\ & \leq (\mathbb{P}_n - \mathbb{P})(\ell_*(\gamma_0^*) - \ell_*(\hat{\gamma}^*)) + \lambda \|\hat{\gamma}^* - \gamma_0^*\|_1 + 2\lambda \|\gamma_0^*\|_1 - 2\lambda \|\hat{\gamma}^*\|_1 \end{aligned} \quad (3.4)$$

根据文献[3]引理 3.2, 我们可以证明:

$$(\mathbb{P}_n - \mathbb{P})(\ell_*(\gamma_0^*) - \ell_*(\hat{\gamma}^*)) \leq \lambda \|\hat{\gamma}^* - \gamma_0^*\|_1 + \lambda \varepsilon_n$$

在事件 A_n 上成立。然后, 可以将不等式(3.4)重写为:

$$\mathbb{P}_n(\ell_*(\hat{\gamma}^*) - \ell_*(\gamma_0^*)) + \lambda \|\hat{\gamma}^* - \gamma_0^*\|_1 \leq 2\lambda \|\hat{\gamma}^* - \gamma_0^*\|_1 + 2\lambda \|\gamma_0^*\|_1 - 2\lambda \|\hat{\gamma}^*\|_1 + \lambda \varepsilon_n, \quad (3.5)$$

请注意, $\|\hat{\gamma}_n^* - \hat{\gamma}_{0n}^*\|_1 + \|\hat{\gamma}_{0n}^*\|_1 - \|\hat{\gamma}_n^*\|_1 = 0$ 。否则, $\|\gamma_{01}^*\|_1 - \|\hat{\gamma}_1^*\|_1 \leq \|\hat{\gamma}_1^* - \gamma_{01}^*\|_1$ 基于三角不等式成立。因此, (3.5)式的不等式界可以是:

$$4\lambda \|\hat{\gamma}_1^* - \gamma_{01}^*\|_1 + \lambda \varepsilon_n, \quad (3.6)$$

根据第二节中 γ_0^* 的定义, 我们有 $\mathbb{P}_n(\ell_*(\hat{\gamma}^*) - \ell_*(\gamma_0^*)) > 0$, 并且:

$$\|\hat{\gamma}^* - \gamma_{01}^*\|_1 \leq 4\|\hat{\gamma}_1^* - \gamma_{01}^*\|_1 + \varepsilon_n$$

通过简单的代数, 我们有:

$$\|\hat{\gamma}_{11}^* - \gamma_{011}^*\|_1 \leq 3\|\hat{\gamma}^* - \gamma_{01}^*\|_1 + \varepsilon_n \quad (3.7)$$

也就是说, $\hat{\gamma}^* - \gamma_0^* \in C(3, \varepsilon_n)$ 。

接下来, 我们将给出 $\mathbb{P}_n(\ell_*(\hat{\gamma}^*) - \ell_*(\gamma_0^*))$ 的下界。结合二阶泰勒展开式和文献[5]中 $\mathbb{P}_n \ell_*(\hat{\gamma}^*)$ 的定义, 以及条件(C6), 我们得到了:

$$\begin{aligned} \mathbb{P}_n(\ell_*(\hat{\gamma}^*) - \ell_*(\gamma_0^*)) &= \mathbb{E}[-Y_{ij} \tilde{X}_{ij}^T \hat{\gamma}^* + b(\tilde{X}_{ij}^T \hat{\gamma}^*) + Y_{ij} \tilde{X}_{ij}^T \gamma_0^* - b(\tilde{X}_{ij}^T \gamma_0^*)] \\ &= \mathbb{E}[-Y_{ij} \tilde{X}_{ij}^T \hat{\gamma}^* + Y_{ij} \tilde{X}_{ij}^T \gamma_0^*] + \mathbb{E}[b(\tilde{X}_{ij}^T \hat{\gamma}^*) - b(\tilde{X}_{ij}^T \gamma_0^*)] \\ &= -\mathbb{E}[Y_{ij} | X_{ij}](\tilde{X}_{ij}^T \hat{\gamma}^* - \tilde{X}_{ij}^T \gamma_0^*) + \mathbb{E}[b'(\tilde{X}_{ij}^T \gamma_0^*)(\tilde{X}_{ij}^T \hat{\gamma}^* - \tilde{X}_{ij}^T \gamma_0^*)] \\ &\quad + \mathbb{E}\left[\frac{1}{2} b''(\tilde{X}_{ij}^T \tilde{\gamma}^*)(\tilde{X}_{ij}^T \hat{\gamma}^* - \tilde{X}_{ij}^T \gamma_0^*)^2\right] \\ &\geq c_n \mathbb{E}(\tilde{X}_{ij}^T \hat{\gamma}^* - \tilde{X}_{ij}^T \gamma_0^*)^2, \end{aligned}$$

其中 $\tilde{X}_{ij}^T \tilde{\gamma}^*$ 是 $\tilde{X}_{ij}^T \hat{\gamma}^*$ 和 $\tilde{X}_{ij}^T \gamma_0^*$ 之间的中间点, $c_n = \min_{|x| < M_1(9B + \frac{2}{n})} \{b''(x)/2\}$, $|\tilde{X}_{ij}^T \tilde{\gamma}^*| \leq M_1(9B + 2/n)$ 。

我们得到了:

$$\lambda \|\hat{\gamma}^* - \gamma_0^*\|_1 + c_n k \|\hat{\gamma}_1^* - \gamma_{01}^*\|_2^2 \leq 4\lambda \sqrt{q} \sqrt{\sum_{k=1}^q (\hat{\gamma}_k^* - \gamma_{0k}^*)^2} + (\lambda + 1)\varepsilon_n, \quad (3.8)$$

结合对于所有 $t > 0$ 的 $2xy \leq tx^2 + y^2/t$ 的事实, (3.8)更进一步展示:

$$\lambda \|\hat{\gamma}^* - \gamma_0^*\|_1 + c_n k \|\hat{\gamma}_1^* - \gamma_{01}^*\|_2^2 \leq 4t\lambda^2 q + \frac{1}{t} \|\hat{\gamma}_1^* - \gamma_{01}^*\|_2^2 + (\lambda + 1)\varepsilon_n, \quad (3.9)$$

在(3.9)式中用 $c_n k$ 替换 t , 我们有:

$$\|\hat{\gamma}^* - \gamma_0^*\|_1 \leq \frac{4}{c_n k} \lambda q + \left(1 + \frac{1}{\lambda}\right) \varepsilon_n, \quad (3.10)$$

根据条件(C2)、(C5)和(C6), 存在常数 $c_B > 0$ 使得 $\|\gamma_0\|_1 \leq c_B q$, 则:

$$\begin{aligned} \|\hat{\beta} - \beta_0\|_1 &= \|\hat{V}^{-1} \hat{\gamma} - V^{-1} \gamma_0\|_1 \\ &= \|\hat{V}^{-1} \hat{\gamma} - V^{-1} \gamma_0\|_1 + \|\hat{V}^{-1} \gamma_0 - V^{-1} \gamma_0\|_1 \\ &\leq \frac{1}{\sqrt{c_0}} \left[\frac{4}{c_n k} \lambda q + \left(1 + \frac{1}{\lambda}\right) \varepsilon_n \right] + c_B q M_3 \sqrt{\log n / n} \\ &= \frac{4}{c_n k \sqrt{c_0}} (\lambda q) + \left(1 + \frac{1}{\lambda}\right) \frac{1}{n \sqrt{c_0}}. \end{aligned}$$

至此, 我们完成了定理 3.1 的证明。

如下定理 3.2 表明所提出的 SPAC-Lasso 方法可以依概率趋于 1, 正确地选择出真模型。

定理 3.2 假设正则性条件(C1)~(C6)成立, 那在(3.2)式和(3.3)式中, 则以至少 $1 - (2p)^{-A_2}$ 的概率有:
 $\hat{\gamma}^s = \gamma_0^s$ 。

证明: 以至少 $1-(2p)^{-A^2}$ 的概率存在下列受限问题的解 $\hat{\gamma}_R = \hat{\gamma}(\lambda, \hat{\omega})$ 。

$$\min_{\gamma: \gamma_{\Pi}=0} \left\{ \tilde{L}(\gamma, \hat{\omega}) + 2\lambda \sum_{k=1}^p \omega_{kk} |\gamma_k| \right\}, \quad (3.11)$$

其中 $\tilde{L}(\gamma, \hat{\omega}) = -n^{-1} \sum_{i=1}^n \sum_{j=1}^m \left\{ Y_{ij} \sum_{k=1}^p x_{ijk} \gamma_k \sqrt{\hat{\omega}_{kk}} - b \left(\sum_{k=1}^p x_{ijk} \gamma_k \sqrt{\hat{\omega}_{kk}} \right) \right\}$ 。

事实上, 对任意常数 M_0 和向量 $h = (h_1, \dots, h_p)^T$, $\|h\|_1 = M_0$, 最后 $p-q$ 个元素 $h_{\Pi} = 0$ 。

设 h_1 和 $\frac{\partial \tilde{L}}{\partial \gamma}(\gamma_0, \hat{\omega})_1$ 分别是 h 和 $\frac{\partial \tilde{L}}{\partial \gamma}(\gamma_0, \hat{\omega})$ 的第 q 元素。根据伯恩斯坦不等式, 存在一个常数 $c_3 > 0$ 使得:

$$\begin{aligned} \Pr \left(\left\| \frac{\partial \tilde{L}}{\partial \gamma}(\gamma_0, \hat{\omega})_1 \right\|_{\infty} > c_3 \sqrt{\frac{\log(n)}{n}} \right) &= \Pr \left(\max_{1 \leq k \leq q} \left| \frac{\partial \tilde{L}}{\partial \gamma_k}(\gamma_0, \hat{\omega})_1 \right| > c_3 \sqrt{\frac{\log(n)}{n}} \right) \\ &\leq q \Pr \left(\left| \frac{\partial \tilde{L}}{\partial \gamma_k}(\gamma_0, \hat{\omega})_1 \right| > c_3 \sqrt{\frac{\log(n)}{n}} \right) \leq (2p)^{-A^2}. \end{aligned}$$

因此, 通过文献[7]引理 6.3 和条件(C3)和(C6), 存在常数 $C_M > 0$, 使得:

$$\begin{aligned} \tilde{L}(\gamma_0 + a_n h, \hat{\omega}) - L(\gamma_0, \hat{\omega}) &= \tilde{L}(\gamma_0 + a_n h, \hat{\omega}) - \tilde{L}(\gamma_0, \hat{\omega}) + 2\lambda \sum_{k=1}^p \hat{\omega}_{kk} |\gamma_{0k} + a_n h_k| - 2\lambda \sum_{k=1}^p \hat{\omega}_{kk} |\gamma_{0k}| \\ &\geq \frac{1}{4} a_n^2 h_1^T V_1^{-1} \hat{C}_{11}(\hat{\omega}) V_1^{-1} h_1 - 2\lambda a_n \sum_{k=1}^p \hat{\omega}_{kk} |h_k| \\ &\geq \frac{1}{4} a_n^2 C_M M_0^2 / q - C_0 M_0 a_n^2 / q. \end{aligned}$$

如果我们取上述不等式的 $M_0 = 4C_0/C_M + c_6$, 其中 $c_6 > 0$, 则:

$$\min_{\gamma: \gamma_{\Pi}=0, \|h\|_1=M_0} L(\gamma_0 + a_n h, \hat{\omega}) \geq L(\gamma_0, \hat{\omega})$$

保持概率至少为 $1-(2p)^{-A^2}$, 也就是说, 存在 $L(\gamma, \hat{\omega})$ 的局部极小值。

接下来证明以至少 $1-(2p)^{-A^2}$ 的概率有: $\hat{\gamma} \stackrel{s}{=} \gamma_0$, 据定理 3.2 和条件(C5), 我们有:

$$\begin{aligned} \Pr \left(\hat{\gamma}_R \stackrel{s}{=} \gamma \right) &\geq \Pr \left(\left\| \hat{\gamma}_R - \gamma_0 \right\|_1 \leq M \left\{ \lambda q + (1 + \lambda^{-1}) n^{-1} \right\}, \min_{1 \leq k \leq q} |\gamma_{0k}| \geq 2M \left\{ \lambda q + (1 + \lambda^{-1}) n^{-1} \right\} \right) \\ &\geq 1 - (2p)^{-A^2}. \end{aligned}$$

在定理 3.2 相同的条件下, 则(3.11)式的解满足:

$$\Pr \left(\max_{q+1 \leq k \leq p} \left\| \frac{1}{\hat{\omega}_{kk}} \frac{\partial \tilde{L}}{\partial \gamma_k}(\hat{\gamma}_R, \hat{\omega}) \right\| \leq 2\lambda \right) \geq 1 - (2p)^{-A^2}$$

我们有:

$$\left| \frac{1}{\hat{\omega}_{kk}} \frac{\partial \tilde{L}}{\partial \gamma_k}(\gamma_0, \hat{\omega}) \right| \leq \frac{1}{c_0} \left\| \frac{\partial \tilde{L}}{\partial \gamma_k}(\gamma_0, \hat{\omega})_1 \right\|_{\infty} \leq \frac{c_3}{c_0} \sqrt{\frac{\log(n)}{n}} = o(\lambda) \quad (3.12)$$

保持概率至少为 $1-(2p)^{-A^2}$ 。根据条件(C3)~(C6)和三角不等式, 存在一个常数 $c_7 > 0$, 使得(3.12)式中第三项的绝对值在 $c_7 c_3 \sqrt{\log(n)/n} = o(\lambda)$ 上有界, 概率至少为 $1-(2p)^{-A^2}$ 。

至此, 我们完成了定理 3.2 的证明。

基金项目

国家社会科学基金一般项目(18BTJ035); 重庆市自然科学基金面上项目(CSTC2020JCYJ-msxm0006); 重庆工商大学研究生创新型科研项目(yjscxx2023-211-193)。

参考文献

- [1] Avella-Medina, M. and Ronchetti, E. (2017) Robust and Consistent Variable Selection in High-Dimensional Generalized Linear Models. *Biometrika*, **105**, 31-44. <https://doi.org/10.1093/biomet/asx070>
- [2] Wang, M. and Tian, G. (2017) Adaptive Group Lasso for High-Dimensional Generalized Linear Models. *Statistical Papers*, **60**, 1469-1486. <https://doi.org/10.1007/s00362-017-0882-z>
- [3] Buhlmann, P., Kalisch, M. and Maathuis, M.H. (2010) Variable Selection in High-Dimensional Linear Models: Partially Faithful Distributions and the Pc-Simple Algorithm. *Biometrika*, **97**, 261-278. <https://doi.org/10.1093/biomet/asq008>
- [4] Jia, J. and Rohe, K. (2015) Preconditioning the Lasso for Sign Consistency. *Electronic Journal of Statistics*, **9**, 1150-1172. <https://doi.org/10.1214/15-ejs1029>
- [5] Cui, Y., Chen, X. and Yan, L. (2017) Adaptive Lasso for Generalized Linear Models with a Diverging Number of Parameters. *Communications in Statistics—Theory and Methods*, **46**, 11826-11842. <https://doi.org/10.1080/03610926.2017.1285926>
- [6] Xue, F. and Qu, A. (2023) Semi-standard Partial Covariance Variable Selection When Irrepresentable Conditions Fail. *Statistica Sinica*, **32**, 1881-1909. <https://doi.org/10.5705/ss.202020.0495>
- [7] Bunea, F. (2008) Honest Variable Selection in Linear and Logistic Regression Models via ℓ_1 and $\ell_1+\ell_2$ Penalization. *Electronic Journal of Statistics*, **2**, 1153-1194. <https://doi.org/10.1214/08-ejs287>
- [8] Blazere, M., Loubes, J. and Gamboa, F. (2014) Oracle Inequalities for a Group Lasso Procedure Applied to Generalized Linear Models in High Dimension. *IEEE Transactions on Information Theory*, **60**, 2303-2318. <https://doi.org/10.1109/tit.2014.2303121>
- [9] Fan, J. and Peng, H. (2004) Nonconcave Penalized Likelihood with a Diverging Number of Parameters. *The Annals of Statistics*, **32**, 928-961. <https://doi.org/10.1214/009053604000000256>
- [10] Peng, J., Wang, P., Zhou, N. and Zhu, J. (2009) Partial Correlation Estimation by Joint Sparse Regression Models. *Journal of the American Statistical Association*, **104**, 735-746. <https://doi.org/10.1198/jasa.2009.0126>
- [11] van de Geer, S.A. (2008) High-Dimensional Generalized Linear Models and the Lasso. *The Annals of Statistics*, **36**, 614-645. <https://doi.org/10.1214/009053607000000929>