·论 著·

基于正则化回归的变量选择方法在高维数据中的应用

荣雯雯, 张奇, 刘艳

哈尔滨医科大学卫生统计学教研室,黑龙江 哈尔滨 150081

摘要: 变量筛选和模型估计一直是高维数据的研究热点,而高维数据的维度灾难问题日渐突出,传统的统计分析方法因模型不稳定不再适用,本文对高维数据中基于正则化回归的变量选择方法的原理、适用的数据类型及优缺点、调整参数的选择进行综述。

关键词: 高维数据; 正则化; 惩罚项; 调整参数

中图分类号:R195.1 **文献标识码:**A **文章编号:**1006-3110(2018)06-0645-04 DOI:10.3969/j.issn.1006-3110.2018.06.002

Application of variable selection method based on regularized regression to high dimensional data

RONG Wen-wen, ZHANG Qi, LIU Yan

Department of Health Statistics, Harbin Medical University, Harbin, Heilongjiang 150081, China

Corresponding author: LIU Yan, E-mail:liuyan@ems.hrbmu.edu.cn

Abstract: Variable filtering and model estimation have been the hotspot of high dimensional data, and the dimensionality problem of high dimensional data is becoming more and more prominent. The traditional statistical analysis method is no longer applicable due to the instability of the model. In this paper, we review the principle of variable selection method based on regularized regression in high dimensional data, the data type and the advantages and disadvantages, and the selection of adjustment parameters.

Key words: high dimensional data; regularization; penalty item; adjustment parameter

随着科技的不断发展,人类对数据的收集和存储能力不断增强,海量且高维的数据开始出现[1-4]。高维数据中,参数或协变量个数 P 远大于样本观测 n,因此急需解决这种维度灾难问题,变量选择是有效的方法。进行变量筛选,传统的方法如逐步回归法,模型会随着观测数据的微小改变而发生很大的变化,预测精度不理想,不能适用于“大 P,小 n”数据[5],而基于惩罚函数的正则化回归方法开始受到关注。

正则化回归的方法,不仅能进行变量选择,还能进行参数估计,方法的基本原理是:在损失函数的基础上加上惩罚项,得到新的惩罚目标函数。令 $\hat{\beta}=(\beta_1,\beta_2,\beta_3,\cdots,\beta_i)$ 为回归系数,i 是变量的总个数,对于线性回归模型来说,参数 β 的最小二乘估计表示为:$\hat{\beta}=arg\ min\|Y-X\beta\|^2$,其中 $\|\cdot\|$ 表示 L_2 范数,但是最小二乘估计容易出现过拟合的现象,加入惩罚项可以避免过拟合现象,具体形式为:$\hat{\beta}=arg\ min\{\|Y-X\beta\|^2+\sum_{i=1}^{p}P_\lambda(|\beta_i|)\}$,其中 $\sum_{i=1}^{p}P_\lambda(|\beta_i|)$ 就是惩罚项,λ 是调整参数,$P(|\beta|)$ 是变量回归系数的惩罚函数,这种方法对较小的回归系数进行很大压缩甚至压缩至 0 来筛除掉,对较大的回归系数进行很小的压缩从而留在模型里,最终实现变量筛选和参数估计。本文就适用于不同数据类型的基于正则化回归的方法进行阐述。

基金项目:国家自然基金(81172741,30972537)

作者简介:荣雯雯(1991-),女,安徽省淮北市人,硕士在读,主要从事医学领域统计学方法的应用与研究工作。

通信作者:刘艳,E-mail:liuyan@ems.hrbmu.edu.cn。

1 选择单变量的正则化回归方法

1.1 Lasso Lasso 是由 Tibshirani(1996)[7]提出的一种同时进行变量选择和参数估计的方法,它是在最小二乘的基础上对 L_1 范数进行惩罚,可以对系数连续压缩。Lasso 的惩罚函数为:$\sum_{i=1}^{p}P_\lambda(|\beta_i|)=\lambda\sum_{i=1}^{p}|\beta_i|$,因此 *Lasso* 惩罚系数估计为:

$$\hat{\beta}(Lasso)=arg\ min\{\|Y-X\beta\|^2+\lambda\sum_{i=1}^{p}|\beta_i|\} \quad (1)$$

其中 λ≥0 为调整参数,λ 越大,对变量回归系数的压缩越大甚至压缩为 0,从而实现变量选择。最常用的估计其回归系数的算法是最小角算法(least angle regression,LARS)[8],Zou(2005)[9]提出的自适应 Lasso(adaptive Lasso),它的基本思想是对显著变量的系数实施较大的权重,对不显著变量的系数实施较小的权

重,因此它的惩罚函数为:$\sum_{i=1}^{p}P_\lambda(|\beta_i|)=\lambda\sum_{i=1}^{p}\hat{\omega}|\beta_i|$,自适应 Lasso 惩罚估计系数估计为:

$$\hat{\beta}(Ada\text{-}Lasso)=arg\ min\{\ \|Y-X\beta\|^2+\lambda\sum_{i=1}^{p}\hat{\omega}_i|\beta_i|\} \tag{2}$$

其中,$\hat{\omega}_i=1/|\beta_i|^\gamma$,$\gamma>0$,且只要权重 $\hat{\omega}_i$ 是$\sqrt{n}$相合的,参数估计就具有选择一致性。自适应 Lasso 方法得到的模型比 Lasso 方法更加稀疏,但是它在某些方面只是 Lasso 方法的一种改善,对于解决共线性问题依然不理想。

1.2　SCAD　SCAD 是由 Fan and Li(2001)[10]提出的,是在 LASSO 的基础上提出的对 L_1 惩罚过度压缩系数的惩罚函数项的弥补。它的惩罚函数形式如下:

$$P_\lambda(|\beta_i|)=\begin{cases}\lambda|\beta_i|, & 0\leqslant|\beta_i|\leqslant\lambda;\\ -(|\beta_i|^2-2a\lambda|\beta_i|+\lambda^2)/\{2(a-1)\}, & \lambda<|\beta_i|\leqslant a\lambda;\\ (a+1)\lambda^2/2, & |\beta_i|>a\lambda;\end{cases} \tag{3}$$

其中,α>2,λ>0 为调整参数,Fan 建议 α 取 3.7,SCAD 惩罚估计结果为:

$$\hat{\beta}(SCAD)=arg\ min\{\ \|Y-X\beta\|^2+\sum_{i=1}^{p}P_\lambda(|\beta_i|)\} \tag{4}$$

这种非凸的惩罚函数可以将较小的回归系数进行较大压缩,减小模型的复杂性,但是对较大的回归系数压缩很小。

1.3　MCP　MCP 是由 Zhang[11]提出的一种新的惩罚方法,这一方法解决了近似无偏估计和如何找到凹度最小的惩罚的计算困难问题。它的惩罚函数为:

$$P_\lambda(|\beta_i|)=\begin{cases}\lambda|\beta_i|-\beta_i^2/2a, & |\beta_i|\leqslant a\lambda;\\ a\lambda^2/2, & |\beta_i|>a\lambda;\end{cases} \tag{5}$$

其中,$\lambda>0$ 且 α>1,所以 MCP 惩罚系数估计为:

$$\hat{\beta}(MCP)=arg\ min\{\ \|Y-X\beta\|^2+\sum_{i=1}^{p}P_\lambda(|\beta_i|)\} \tag{6}$$

MCP 相当于先应用和 Lasso 的保持一致的惩罚,然后不断的对惩罚进行放松,直到当$|\beta_i|>\alpha\lambda$,惩罚速率降到 0。MCP 的原理可以理解为单变量的解决方案,采用的是坐标下降算法[12]来获得目标函数的最小值。见表 1。

表 1　单变量选择方法的适用类型、惩罚项目、参数选取和优缺点比较

方法	适用类型	惩罚项目	参数选取	优点	缺点
Lasso	单变量选择	L_1	λ≥0	连续且稳健,对高维数据可降维	不能实现组效应,不具有 Oracle 性质
SCAD	单变量选择	L_1	α>2,λ>0	继承 Lasso 优点,同时具有 Oracle 性质	不能处理"大 P,小 n"和具有组效应的数据
MCP	单变量选择	L_1	λ>0 且 α>1	实现变量选择,具有估计一致性和 Oracle 性质	不能处理具有组效应的数据

2　识别高度相关数据的正则化回归方法

2.1　Ridge　Ridge[6,13]是由 Hoerl 和 Kennard(1970)提出的能够处理多重共线性问题的方法,是在传统最小化残差平方和基础上加上一个 L_2 惩罚来对回归系数进行收缩。是一种有偏估计,通过损失部分信息、降低精度来获得变量的回归系数。Ridge 的惩罚函数为:$\sum_{i=1}^{p}P_\lambda(|\beta_i|)=\lambda\sum_{i=1}^{p}\beta_i^2$,因此 Ridge 惩罚系数估计为:

$$\hat{\beta}(Ridge)=arg\ min\{\ \|Y-X\beta\|^2+\lambda\sum_{i=1}^{p}\beta_i^2\} \tag{7}$$

其中 λ≥0 为调整参数,参数 λ 用来控制压缩量,不过只能使系数趋于 0,而不会等于 0。

2.2　Elastic Net　为了既能对 Lasso 方法进行改进同时又能保持其优点,Zou 和 Hastie(2005)[14]提出了 Elastic Net 方法,这种方法能够解决组效应问题,不仅适用于 n 大于 p 的模型,也适用于 n 小于 p 的模型。Elastic Net 的惩罚函数为:$\sum_{i=1}^{p}P_\lambda(|\beta_i|)=\lambda_1\sum_{i=1}^{p}|\beta_i|+\lambda_2\sum_{i=1}^{p}\beta_i^2$,因此 Elastic Net 惩罚估计系数估计为:

$$\hat{\beta}(Elastic\text{-}net)=arg\ min\{\ \|Y-X\beta\|^2+\lambda_1\|\beta\|_1+\lambda_2\|\beta\|_2\} \tag{8}$$

其中,$\|\beta\|_1=\sum_{i=1}^{p}|\beta_i|$ 和 $\|\beta\|_2=\sum_{i=1}^{p}\beta_i^2$,若令 $\alpha=\lambda_2/(\lambda_1+\lambda_2)$,参数估计的结果也可以写成:$\hat{\beta}(Elastic\text{-}net)=arg\ min\{\ \|Y-X\beta\|^2+(1-\alpha)\|\beta\|_1+\alpha\|\beta\|_2\}$,且$(1-\alpha)\|\beta\|_1+\alpha\|\beta\|_2\leqslant t$。

其中 t 为常数界值,$\alpha\in[0,1]$。当 $\alpha=0$,Elastic Net 方法就变为 Lasso 方法,当 α=1,Elastic Net 方法又变成 Ridge 方法。Ghosh(2007)[17]对弹性网络中的惩罚项进行改进,对 L_1 的各个参数施加不同的权重而其他部分不变得到的变量选择方法叫做 Adaptive Elastic Net 方法,同时 Zou 和 Zhang(2009)[18]进一步研究,发现 Adaptive Elastic Net 方法不但继承了 Elastic Net 方法的优点,同时也具有 Oracle 性质。

2.3　Mnet　Mnet 方法是由 Huang 和 Breheny 等(2010)[19]提出的,是把 L_2 惩罚和 MCP 惩罚方法结合起来的方法,它能够处理 $P\geqslant n$ 时的高度相关性数据,Mnet 的惩罚函数为:$\sum_{i=1}^{p}P_\lambda(|\beta_i|)=\sum_{i=1}^{p}P_\lambda^{MCP}(|\beta_i|)+\frac{1}{2}\lambda(1-c)\sum_{i=1}^{p}\beta_i^2$,其中,$\sum_{i=1}^{p}P_\lambda^{MCP}(|\beta_i|)$是 MCP 惩罚,对于 α 的取值,Zhang(2007)[11]建议:线性回归模型时 α 取 3,广义线性模型时 α 取 30;对于 c 的取值,为了防止由于二次惩罚对系数惩罚过大,Huang[19]建议 c 取 0.2。因此 Mnet 惩罚估计系数估计为:

$$\hat{\beta}(Mnet)=arg\ min\{\frac{1}{2}\|Y-X\beta\|^2+\sum_{i=1}^{p}P_\lambda(|\beta_i|)\} \tag{9}$$

Huang 和 Breheny 等(2010)[19]通过模拟发现,在

高度相关性的数据处理上 Mnet 方法优于 Elastic Net。

3　选择群组变量的正则化回归方法

3.1　Group Lasso　目前,变量之间一般都具有群组效应,如生物学通路之间,但是很多时候,变量的分组情况并不清楚。不过可以通过一些先验信息来获得变量分组情况,从而在变量筛选时使同组变量全部进入或者剔除模型。于是 Yuan 和 Lin(2006)[20]提出一种在已知变量的分组因子信息的基础上加入惩罚项进行变量选择的方法,叫做 Group Lasso。惩罚函数为:$\sum_{i=1}^{p} P_\lambda (|\beta_i|)=\lambda\sum_{i=1}^{p}\sqrt{P_i}\,\|\beta_i\|_2$,因此 Group Lasso 的惩罚系数估计为:

$$\hat{\beta}(Grplasso)=arg\ min\{\ \|Y-X\beta\|^2+\lambda\sum_{i=1}^{p}\sqrt{P_i}\,\|\beta_i\|_2\} \tag{10}$$

Group Lasso 对变量集 β_i 加入二次惩罚,调整 λ 的值可以控制组别个数,λ 的值越大,对各组的压缩力度就越大,因此整组特征变量可以全部筛除或进入模型,实现组特征选择的功能。为了对 Group Lasso 进行求解,Yuan 和 Lin 于 2006 年[20]在 Lasso 的 LARS 算法上提出了 GroupLARS 算法。Breheny 和 Huang(2009)[21]提出了基于坐标下降法的局部坐标下降(Locally Coordinate Descent,LCD)的方法。

3.2　*Group SCAD 和 Group MCP 方法*　由于前面提出的 SCAD 方法和 MCP 方法都不能对组变量进行选择,Wang 等在 2007 年[22]提出了 Group SCAD 方法,Huang 等(2012)[23]基于 MCP 方法的基础上提出了 Group MCP 方法,其对应的惩罚可以从 SCAD 和 MCP 方法中得到。Group SCAD 和 Group MCP 方法的惩罚系数估计为:

$$\hat{\beta}(Grp\ SCAD/MCP)=arg\ min\{\ \|Y-X\beta\|^2+\sum_{i=1}^{p}P_\lambda\,\|\beta_i\|_2\} \tag{11}$$

其中 P_λ 是对应的 SCAD 和 MCP 的惩罚函数。

4　选择双层变量的正则化回归方法

前面所述的组变量筛选方法都是基于 L_2 的,其缺点是无法实现组内的特征选择。统计学者们继续研究,随后一些既能对组间进行特征选取又能在组内进行特征选取的方法开始出现,也就是双层变量选择的方法。Huang 等(2012)[23]将现有双层变量选择的方法分为两类:复合惩罚类和稀疏组惩罚类。

4.1　*复合惩罚类*　复合惩罚类是组内和组间惩罚的复合函数,其惩罚函数可以表示为:

$$P_{outer}\Big[\sum_{k=1}^{G_i}P_{inner}(|\beta_k{}^i|)\Big] \tag{12}$$

其参数估计系数可以表示为:

$$\hat{\beta}(Grp\ SCAD/MCP)=arg\ min\{\ \|Y-X\beta\|^2+\sum_{i=1}^{p}P_\lambda\,\|\beta_i\|_1\} \tag{13}$$

其中 P 分别为 SCAD 和 MCP 惩罚,组内惩罚是 Lasso,组间惩罚是 Group MCP。因为 Lasso 在单变量选择中筛选过多变量,又不具有 Oracle 性质,性能和效果不如 MCP 惩罚,于是 Liu 等(2014)[24]提出了复合 MCP 惩罚方法(Composite MCP),其组内和组间惩罚都是 MCP,其惩罚系数估计为:

$$\hat{\beta}(cMCP)=arg\ min\{\ \|Y-X\beta\|^2+\sum_{i=1}^{p}P_{MCP}\Big[\sum_{m=1}^{M}P_{MCP}(\beta_i{}^m)\Big]\} \tag{14}$$

Zhang 等(2015)指出在满足一定条件下,复合 MCP 方法在组间和组内都具有选择一致性,而 L_1 Group MCP 只满足组选择一致性,因此复合 MCP 性质更稳定。

4.2　*稀疏组惩罚类*　稀疏组惩罚方法是通过将两个惩罚函数结合来实现双层变量选择的功能的,其中一个是能对组特征进行选取的惩罚,另一个是能对单个特征进行选取的惩罚。其惩罚函数为:

$$\sum_{i=1}^{p}P_\lambda(|\beta_i|)=\lambda_1\sum_{i=1}^{p}P_1(\ \|\beta_i\|\ )+\lambda_2\sum_{i=1}^{p}\sum_{m=1}^{M}P_2(|\beta_i{}^{(m)}|) \tag{15}$$

其中,函数 $P_1(\cdot)$是控制系数组,能实现组选择,函数 $P_2(\cdot)$是对每一个系数都起作用,能实现单特征的选取。例如 Zhang 等(2015)提出的 Sparse Group MCP 函数,其中的 $P_1(\cdot)$和 $P_2(\cdot)$都是 MCP 惩罚方法。再比如将 Group Lasso 和 Lasso 组合就形成了 Sparse Group Lasso 方法(SGL)[25-27],其参数估计系数可表示为:

$$\hat{\beta}(SGL)=arg\ min\{\ \|Y-X\beta\|^2+\lambda_1\sum_{i=1}^{p}\|\beta_i\|_2+\lambda_2\|\beta_i\|_1\} \tag{16}$$

SGL 是 Group Lasso 和 Lasso 的结合,由于两者都不具有 Oracle 性质,因此 SGL 也不具有 Oracle 性质,于是 Fang 等(2014)[28]提出了 Adaptive Sparse Group Lasso(AdSGL)。通过增加组权重和单个系数权重来提高估计一致性。其参数估计系数可表示为:

$$\hat{\beta}(AdSGL)=arg\ min\{\ \|Y-X\beta\|^2+\lambda_1\sum_{i=1}^{p}\omega_i\|\beta_i\|_2+\lambda_2\xi_i\|\beta_i\|_1\} \tag{17}$$

其中 ω 是组权重,ξ 是单个系数权重,两个权重的大小和系数值成反比,系数值越大,权重就越小,对系数的压缩就越弱,估计结果也就更接近系数值。

对于 SGL 和 AdSGL 的求解,Friedman 等[26]提出了类似的坐标下降算法,但是当组数特别大时,这些算法因为产生过多的矩阵运算过程而计算效率不高。因此,Simon 等(2013)[25]提出了广义梯度下降算法,该算法首先把似然函数扩展为二阶函数,然后加入惩罚函数求出最优参数估计值,最终确定下降的方向和步长。见表 2。

表 2 多变量选择方法的适用类型、参数选取和优缺点比较

方法	适用类型	参数选取	优点	缺点
Ridge	高度相关数据	$\lambda \geqslant 0$	模型稳定,高度相关的变量存在时,能够得到唯一解[5]	不能实现变量选择
Elastic Net	高度相关数据	$\alpha \in [0,1]$	能处理高度相关数据也能进行变量选择[15]	选择过多的群组,满足极强条件时才具有选择一致性[16]
Mnet	高度相关数据	$\lambda>0$ 且 $\alpha>1$	具有 Oracle 性质	不具有组效应
Group Lasso	群组变量选择	$\lambda \geqslant 0$	目标函数为凸函数,因此存在全局最小值	不具有 Oracle 性质,无法进行群组内重要特征选择
Group SCAD, Group MCP	群组变量选择	$\alpha>2$ 且 $\lambda>0$, $\lambda>0$ 且 $\alpha>1$	组间能实现变量选择	无法在组内进行变量选择
复合惩罚类	双层变量选择	$\lambda>0$ 且 $\alpha>1$	组间和组内都具有选择一致性	
稀疏组惩罚类	双层变量选择	$\lambda>0$ 且 $\alpha>1$	既能选择组特征,又能选择单个特征	

对于惩罚模型的估计,调整参数不同其对应的模型也不一样。调整参数的选择有 3 个步骤:①先列出所有调整参数并对对应的模型进行估计;②选定一个准则来评价模型的优劣,计算①模型中的准侧值;③比较准则值大小来获得最优调整参数。因此,要确定调整参数的大小,首先要选用合适的评价方法,评价方法中常用的有赤池信息准则(AIC)、贝叶斯信息准则(BIC)、Cp 准则、交叉验证(cross validation, CV)、广义交叉验证(GCV)等。实际应用中,CV 方法思想简单,而且其他准则的效果不如它,因此一般选用 CV 计算各调整参数对应的准则值[29]。

综上所述,基于正则化回归的变量选择方法,可根据不同的研究目的通过调整参数来获取不同惩罚项,弥补了传统的特征变量选择方法的缺点,且使用灵活。通过对上述四种类型的正则化方法的阐述可以发现,实际中没有哪一种是适合所有数据类型的最优特征变量选取方法,每种方法都有各自的适用特征和范围,研究者应根据实际需要进行合理选择。随着大数据时代的到来,基于正则化回归的变量选择方法应用于更多的组学融合的特征变量选择中,为疾病的发生机制研究提供更为全面的分析方法。

参考文献

[1] 王巧智, 黄强, 黄河,等. 大数据下结核患者诊疗质量控制系统架构设计探讨[J]. 实用预防医学, 2016, 23(10):1280-1283.
[2] 顾星博, 温琪, 史晓雯,等. 随机森林的并行运算方法及适用条件[J]. 实用预防医学, 2016, 23(2):129-132.
[3] 李仲达, 林建浩, 王美今. 大数据时代的高维统计:稀疏建模的发展及其应用[J]. 统计研究, 2015, 32(1):3-11.
[4] 邱东. 大数据时代对统计学的挑战[J]. 统计研究, 2014, 31(1):16-22.
[5] Breiman L. Heuristics of instability and stabilization in model selection [J]. Ann Stat, 1996, 24(6):2350-2383.
[6] Hoerl AE, Kennard RW. Ridge regression: applications to nonorthogonal problems[J]. Technometrics, 1970, 12(1):69-82.
[7] Tibshirani R. Regression shrinkage and subset selection with the lasso [J]. J Roy Stat Soc, 1996,58(1):267-288.
[8] Efron B, Hastie T, Johnstone I, et al. Least angle regression[J]. Ann Stat, 2004, 32(2):407-451.
[9] Zou H. The adaptive lasso and its oracle properties[J]. J Am Stat Assoc, 2006, 101(476):1418-1429.
[10] Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties[J]. J Am Stat Assoc, 2001, 96(456):1348-1360.
[11] Zhang CH. Penalized linear unbiased selection[J]. Dept Stat, 2007,3.
[12] Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent.[J]. J Stat Softw, 2010, 33(1):1-22.
[13] Hoerl AE, Kennard RW. Ridge regression: biased estimation for non-orthogonal problems[J]. Technometrics, 2000, 42(1):55-67.
[14] Zou H, Hastie T. Regularization and variable selection via the elastic net [J]. J Roy Stat Soc, 2005, 67(2):301-320.
[15] 刘匆提, 李昂, 门志红,等. 惩罚 logistic 回归方法在 SNPs 数据变量筛选研究中的应用[J]. 实用预防医学, 2016, 23(11):1395-1399.
[16] Jia JZ, Yu B. On model selection consistency of the elastic net when p》n[J]. Stat Sin, 2008, 24(2):595-611.
[17] Ghosh S. Adaptive elastic net: an improvement of elastic net to achieve oracle properties: IUPUI tech report No. pr07-01[R]. Indianapolis, USA: Department of Mathematical Sciences, Indiana University-Purdue University, 2007.
[18] Zou H, Zhang HH. On the adaptive elastic-net with a diverging number of parameters[J]. Ann Stat, 2009, 37(4):1733-1751.
[19] Huang J, Breheny P, Ma S, et al. The mnet method for variable selection[J]. Stat Sin, 2010,26(3):718-721.
[20] Yuan M, Lin Y. Model selection and estimation in regression with grouped variables[J]. J Roy Stat Soc, 2006, 68(1):49-67.
[21] Breheny P, Huang J. Penalized methods for bi-level variable selection [J]. Stat Intfc, 2009, 2(3):369-380.
[22] Wang L, Chen G, Li H. Group SCAD regression analysis for microarray time course gene expression data[J]. Bioinformatics, 2007, 23(12):1486-1494.
[23] Huang J, Breheny P, Ma S. A selective review of group selection in high-dimensional models[J]. J Inst Math Stat, 2013, 27(4):481-499.
[24] Liu J, Huang J, Ma S. Integrative analysis of multiple cancer prognosis datasets under the heterogeneity model[M]. Springer New York, 2013, 32 (20):3509-3521.
[25] Noah S, Jerome F, Trevor H, et al. A sparse-group lasso[J]. J Comput Graph Stat, 2013, 22(2):231-245.
[26] Friedman J, Hastie T, Tibshirani R. A note on the group lasso and a sparse group lasso[J]. Statistics, 2010.
[27] Wu TT, Lange K. Coordinate descent algorithms for lasso penalized regression[J]. Ann Appl Stat, 2008, 2(1):224-244.
[28] Fang K, Wang X, Zhang S, et al. Bi-level variable selection via adaptive sparse group lasso[J]. J Stat Comput Sim, 2014, 85(1):1-11.
[29] 胡局新, 张功杰. 基于 K 折交叉验证的选择性集成分类算法[J]. 科技通报, 2013,33(12):115-117.

收稿日期:2017-05-17