中国科学: 信息科学 2016 年 第 46 卷 第 7 期: 870–882
SCIENTIA SINICA Informationis
《中国科学》杂志社 SCIENCE CHINA PRESS

论 文

基于无标度先验的图模型结构学习

郭骁[1], 张海[1,2]*, 吴奖伦[3]

① 西北大学数学学院, 西安 710127
② 中国科学院数学与系统科学研究院应用数学所, 北京 100190
③ Department of Mathematics, College of Science, Swansea University, Swansea SA2 8PP, UK
* 通信作者. E-mail: zhanghai@nwu.edu.cn

收稿日期: 2015–10–07; 接受日期: 2015–11–26
国家自然科学基金 (批准号: 11171272, 11571011) 资助项目

摘要 本文研究在无标度先验下, 图模型的结构学习问题. 提出新的正则化模型, 其惩罚项为 Log 型和 L_q 型惩罚函数的复合, 该模型包含无标度先验. 本文使用重赋权迭代算法求解该模型. 实验表明, 所提出的新模型有效、实用, 其在参数估计和模型选择方面均有良好效果.

关键词 稀疏 无标度 图模型 重赋权算法 正则化

1 引言

复杂网络作为一种数据可视化和提取数据特征的方法成为了近期机器学习、统计学和统计物理学领域的研究热点. Barabási 和 Albert [1] 研究发现, 不同于随机网络, 基因网络、神经网络、社交网络和社会网络等大多数真实网络为无标度网络 (scale-free network), 即网络节点的度服从幂律分布. 无标度网络的典型特征是网络中大部分节点度很小, 存在度非常大的少数节点. 因此, 在无标度的假设下研究网络的结构, 从而理解真实网络的功能和特征不仅具有理论意义, 更具有实际价值.

图模型 (graphical model) [2] 作为一种分析网络结构的方法被广泛应用. 它是一类用图来表示随机变量联合概率分布, 并基于此分布研究变量之间相互关系的方法. 一般地, 图模型分为连续型和离散型两种类型: 连续型图模型的主要代表为高斯图模型 (Gaussian graphical model, GGM); 离散型图模型的主要代表为 Ising 模型. 本文基于 GGM 开展真实网络结构分析. GGM 的数学模型如下, 假设 $X=(X^{(1)},\ldots,X^{(p)})$ 服从 p 维正态分布 $N_p(\mu,\Sigma)$, 其中, μ 为均值, Σ 为协方差矩阵. 利用 n 个独立同分布于 $N_p(\mu,\Sigma)$ 的样本 $X_1,\ldots,X_n$ 估计 $\Theta=(\theta_{ij})_{1\leqslant i,j\leqslant p}=\Sigma^{-1}$(协方差矩阵的逆) 被 Dempster [3] 称为协方差选择问题. 众所周知, $\theta_{ij}=0$ 当且仅当 $X^{(i)}$ 与 $X^{(j)}$ 条件独立, 即给定除 $X^{(i)}$ 与 $X^{(j)}$ 之外的变量, $X^{(i)}$ 与 $X^{(j)}$ 相互独立. 通常地, 一个随机向量可以用一个无向图 $G=(V,E)$ 表示, 其中, V 代表节点, 每个节点分别对应随机向量的一个分量. E 代表边, 两个节点之间无边当且仅当这两个节点对应的随机变量条件独立, 即 $\theta_{ij}=0$. 因此, 研究网络的结构等价于判断协方差矩阵的逆 Θ 中每个元素是否为 0. 经典的估计方法是首先用样本协方差矩阵 $\hat{\Sigma}$ 估计, 然后令 $\hat{\Theta}=(\hat{\Sigma})^{-1}$. 存在的问题在于,

引用格式: 郭骁, 张海, 吴奖伦. 基于无标度先验的图模型结构学习. 中国科学: 信息科学, 2016, 46: 870–882, doi: 10.1360/N112015-00136

此时 $\hat{\Theta}$ 往往对应一个完全连接的图, 从而不能有效发现网络的结构信息. 另一种研究方法为引入稀疏性先验, 对 Θ 直接估计从而抽取网络的结构信息.

众多学者通过引入 L_1 正则子 [4] 研究图模型的模型选择和参数估计问题. Meinshausen 和 Bühlmann [5] 提出了邻域选择 (neighborhood selection) 方法, 将 Θ 的估计转化为 p 个 L_1 正则化 [4] 问题, 即分别以每个变量作为因变量, 其他变量为自变量, 基于 L_1 正则化研究因变量与自变量的关系. 在一系列假设之下, 他们证明了邻域选择的变量选择一致性. 但是, 该方法存在两个主要缺点: 其一, 该方法得到的 Θ 的估计不具有对称性; 其二, 作者建议 p 个 L_1 正则化问题的正则化参数取相同数值, 也就是近似认为网络的度的分布是均匀分布, 显然, 这是不符合实际情况的. 2007 年, Yuan 和 Lin [6] 建立了一种罚似然函数理论框架, 即所谓的 Graphical Lasso, 来克服 Meinshausen 和 Bühlmann 所提方法的不稳定性和非对称性, 并分析了该估计在低维情形下的 Oracle 性质, 同时指出该模型可以由 Maxdet 算法有效求解. 对于 Graphical Lasso 模型, Friedman 等 [7] 随后提出了更高效的分块坐标下降法来求解 Graphical Lasso, 使其有更广泛的应用. Ravikumar 等 [8] 分析了 Graphical Lasso 的高维统计性质. 另外, 2009 年, Peng 等 [9] 提出了联合回归模型 SPACE 来估计偏相关系数以克服邻域选择的缺点. 通过调整不同节点对应的权重来增加其模型的灵活性是该方法的一大亮点. 特别地, 可以调整每个节点的权重使之与对应节点的度成比例, 试图得到更接近真实网络的无标度网络.

上述方法大多通过在模型中加入 L_1 正则子来引入稀疏性先验, 从 Bayes 角度来看, 即认为每个参数 θ_{ij} 是独立的且均服从 Laplace 分布. 然而, 如本文开头所述, 真实网络大多具有无标度的特点. 因此研究在无标度先验下, 图模型的结构学习问题具有理论和实际双重价值. 近年来已有部分学者关注这一问题. 注意到无标度网络是具有中心点 (Hubs) 的, Hero 和 Rajaratnam [10], Tan 等 [11] 先后利用网络具有中心点的信息建立了模型并用凸优化方法求解. 不同于 Peng 等 [9] 的启发式思想, 2011 年, Liu 和 Ihler [12] 直接利用网络节点的度服从幂律分布这一先验信息, 提出了如下模型:

$$\hat{\Theta} \in \arg\max\left\{S(X,\Theta) - \alpha\sum_i \log(\|\theta_{-i}\|_1 + \epsilon_i) - \beta\sum_i |\theta_{ii}|\right\},$$

其中, $S(X,\Theta)$ 为损失函数, α, β 为参数, $\theta_{-i} = (\theta_{i1}, \ldots, \theta_{i,i-1}, \theta_{i,i+1}, \ldots, \theta_{ip})'$, $\|\theta_{-i}\|_1 = \sum_{j\neq i}|\theta_{ij}|$. 此外, 他们提出用重赋权迭代算法来求解这一模型, $\epsilon_i > 0$ 是正值小常数, 以保证算法的可行性. 可以看到, Liu 和 Ihler 用 θ_{-i} 的 l_1 范数 $\|\theta_{-i}\|_1$ 近似替代节点 i 的度 d_i 即 $\sum_{j:j\neq i} I_{\{\theta_{ij}\neq 0\}}$, 来克服 l_0 带来的组合优化问题. 然而, 在经典的稀疏正则化理论中, 众多研究表明, $L_q(0<q<1)$ 比 L_1 具有更优良的理论性质: 如 Knight 和 Fu [13] 证明了 L_q 在低维情形下的渐近性质如 Oracle 性质 [14]; Huang 等 [15] 进一步给出了 L_q 在高维情形下的估计误差界; Xu 等 [16] 提出了 $L_{1/2}$ 正则化理论. 另一方面, Liu 和 Ihler [12] 所提模型缺乏理论支撑, 从而不能从理论上刻画估计结果.

本文直接利用网络无标度的信息结合非凸正则化成果研究图模型的模型选择与参数估计问题. 通过引入网络中节点度的分布信息和边的稀疏先验, 提出新的正则化模型, 其正则化项为 Log 型和 L_q 型惩罚函数的复合. 本文使用重赋权迭代算法求解该模型, 将问题转化为重赋权的 L_1 问题. 在一定条件下, 从理论上证明了重赋权算法的优良性. 最后, 实验表明所提出的模型在参数估计和模型选择方面均有良好效果.

2 无标度图模型

本节讨论无标度先验下 GGM 的建立.

对于无标度网络 $G=(V,E)$, 节点的度服从幂律分布, 即

$$P(d) \propto d^{-\alpha}, \tag{1}$$

其中, $\alpha>0$ 为尺度参数, d 为节点的度. 反映到图模型中, 节点 $i\,(i=1,\ldots,p)$ 的度 d_i 为 $\sum_{j:j\neq i} I_{\{\theta_{ij}\neq 0\}}$, 即 $\theta_{-i}=(\theta_{i1},\ldots,\theta_{i,i-1},\theta_{i,i+1},\ldots,\theta_{ip})'$ 的 l_0 拟范数. 考虑到 l_0 的不连续性及其带来的组合优化问题, 可用 $\|\theta_{-i}\|_1$ 代替 d_i. 实际应用中, 节点的度往往需要更稀疏的解释, 因此基于 $L_q(0<q<1)$ 比 L_1 具有更优良的性质 [13,15,16], 用 $\|\theta_{-i}\|_q^q$ 代替 d_i, 其中 $\|\theta_{-i}\|_q^q=|\theta_{i1}|^q+\cdots+|\theta_{i,i-1}|^q+|\theta_{i,i+1}|^q+\cdots+|\theta_{ip}|^q$. 假设每个节点的度是独立的, 则

$$\log P(G=(V,E)) \propto \prod_{i=1}^{p}(\|\theta_{-i}\|_q^q)^{-\alpha}. \tag{2}$$

基于 X 的 n 次观测 $X_1,\ldots,X_n$, GGM 的对数似然函数为

$$P(X_1,\ldots,X_n|G)=\log|\Theta|-\mathrm{tr}(\Theta\hat{\Sigma}), \tag{3}$$

其中, $|\cdot|$ 为矩阵的行列式, $\mathrm{tr}(\cdot)$ 为矩阵的迹, $\hat{\Sigma}$ 为样本协方差矩阵. 结合网络节点度的分布信息以及边的稀疏先验, 提出以下正则化模型:

$$\max_{\Theta} l(\Theta),$$

其中

$$l(\Theta)=\log|\Theta|-\mathrm{tr}(\Theta\hat{\Sigma})-\lambda\sum_{i=1}^{p}\log(\|\theta_{-i}\|_q^q+\epsilon_i), \tag{4}$$

最大化 $l(\Theta)$ 可得到无标度先验下图模型的协方差矩阵的逆的估计.

注 1 $\epsilon_i>0\ (i=1,\ldots,p)$ 保证 Log 的指数部分大于 0. $\lambda>0$ 为正则化参数. 由于对角线元素的值不影响网络的结构, 本文仅对 Θ 的非对角线元素进行惩罚. 式 (4) 正则化项为 Log 型和 L_q 型惩罚函数的复合, 相当于引入网络节点度的分布信息以及边的稀疏先验.

3 重赋权迭代算法

本节给出重赋权算法求解模型 (4) 并证明在一定条件下, 重赋权算法一步迭代求得的解具有 Oracle 性质 [14],

$$l(\Theta)=\log|\Theta|-\mathrm{tr}(\Theta\hat{\Sigma})-\lambda\sum_{i=1}^{p}\log(\|\theta_{-i}\|_q^q+\epsilon_i).$$

由于 $l(\Theta)$ 对应的惩罚项为 Log 与 L_q 型惩罚函数的复合, 所以直接求解 $l(\Theta)$ 存在困难. 基于 Minorize-Maximization (MM) 算法 [17] 的思想, 构造 $l(\Theta)$ 的 Minorization 函数, 设为 $g(\Theta,\Theta^n)$, 其中 Θ^n 为算法的第 n 步估计. Minorization 函数需要满足以下两条性质:

$$\mathrm{c}_1.\quad l(\Theta)\geqslant g(\Theta,\Theta^n);$$

$$\mathrm{c}_2.\quad l(\Theta^n)=g(\Theta^n,\Theta^n).$$

对于算法的第 n 步估计 Θ^n, 由于 $x>0$ 时, $\log(1+x)\leqslant x$, 故

$$\sum_i\log(\|\theta_{-i}\|_q^q+\epsilon_i)-\sum_i\log(\|\theta_{-i}^n\|_q^q+\epsilon_i)$$

$$\leqslant \sum_i \left(\frac{\|\theta_{-i}\|_q^q+\epsilon_i}{\|\theta_{-i}^n\|_q^q+\epsilon_i}-1\right)=\sum_{i\neq j}\lambda_{ij}|\theta_{ij}|^q+\text{const},$$

令

$$g(\Theta,\Theta^n)=\log|\Theta|-\operatorname{tr}(\Theta\hat{\Sigma})-\lambda\sum_{i\neq j}\left(\frac{1}{\|\theta_{-i}^n\|_q^q+\epsilon_i}+\frac{1}{\|\theta_{-j}^n\|_q^q+\epsilon_j}\right)|\theta_{ij}|^q+\text{const}. \tag{5}$$

注意到上式右端 const 代表常数, 实际上为 $l(\Theta^n)$. 容易看出 $g(\Theta,\Theta^n)$ 满足 $\mathrm{c}_1,\mathrm{c}_2$ 两条性质, 故为 $l(\Theta)$ 的 Minorization 函数. 下面采用重赋权方法逼近式 (5), 将式 (5) 问题转化为重赋权的 L_1 问题. 令

$$h(\Theta,\Theta^n)=\log|\Theta|-\operatorname{tr}(\Theta\hat{\Sigma})-\lambda\sum_{i\neq j}\left(\frac{1}{\|\theta_{-i}^n\|_q^q+\epsilon_i}+\frac{1}{\|\theta_{-j}^n\|_q^q+\epsilon_j}\right)\frac{q}{|\theta_{ij}^n|^{1-q}+\epsilon_{ij}}|\theta_{ij}|+\text{const}, \tag{6}$$

其中, ϵ_{ij} 为大于 0 的数. 按下面公式更新 Θ:

$$\Theta^{n+1}=\arg\max_{\Theta} h(\Theta,\Theta^n). \tag{7}$$

算法的具体步骤如下:

步 1. 输入一系列 $\lambda>0$, $\epsilon_i>0$ $(i=1,\ldots,p)$, $\epsilon_{ij}>0$ $(i\neq j,\ i,j=1,\ldots,p)$ 及初始值 Θ^0, $k=0$;

步 2. 对于每个 λ, 对于 $k=1,2,\ldots$, 重复下式直至收敛或达到要求的迭代次数:

$$\Theta^{k+1}=\arg\max_{\Theta}\log|\Theta|-\operatorname{tr}(\Theta\hat{\Sigma})-\lambda\sum_{i\neq j}\left(\frac{1}{\|\theta_{-i}^k\|_q^q+\epsilon_i}+\frac{1}{\|\theta_{-j}^k\|_q^q+\epsilon_j}\right)\frac{q}{|\theta_{ij}^k|^{1-q}+\epsilon_{ij}}|\theta_{ij}|;$$

步 3. 使用 BIC 准则选择 λ, 并输出相应的 Θ 的估计值.

为了保证算法可实施, 步 1 中需提前设定 ϵ_i 和 ϵ_{ij} 的值. 取其与 θ_{ij} 的阶数相同. 步 2 相当于求解一个 Graphical Lasso 问题, 关于此问题, Yuan 和 Lin [6], Friedman 等 [7], Witten 等 [18] 相继提出了高效的算法. 注意到节点 i 在第 k 步的权重与其在 $k-1$ 步对应的度呈负相关关系, 即节点在 $k-1$ 步的度数越大, 在第 k 步对其惩罚越小, 因此更容易得出具有幂律分布的网络.

下面分析算法的理论性质. 首先给出符号说明, 假设 S 为 Θ 中不为 0 的元素下标集合, S^c 为 Θ 中为 0 的元素下标集合. 相应地, Θ_S, Θ_{S^c} 分别表示其对应元素. 令

$$\hat{\Theta}^{\mathrm{t}}=\arg\min_{\Theta_{S^c}=0}\{-\log|\Theta|+\operatorname{tr}(\Theta\hat{\Sigma})\}.$$

$\hat{\Theta}^{\mathrm{t}}$ 是所谓 Oracle 解, 即知道真实模型 (不为 0 元素) 时的最大似然估计. 下面定义以 $\hat{\Sigma}^{-1}$ 为初值的一步重赋权的解 $\hat{\Theta}$. 令

$$\hat{\Theta}=\arg\min J(\Theta),$$

其中,

$$J(\Theta)=-\log|\Theta|+\operatorname{tr}(\Theta\hat{\Sigma})+\lambda\sum_{i\neq j}\left(\frac{1}{\|\theta_{-i}^0\|_q^q+\epsilon_i}+\frac{1}{\|\theta_{-j}^0\|_q^q+\epsilon_j}\right)\frac{q}{|\theta_{ij}^0|^{1-q}}|\theta_{ij}|, \tag{8}$$

$\Theta^0=(\theta_{ij}^0)_{1\leqslant i,j\leqslant p}=\hat{\Sigma}^{-1}$. 由于 θ_{ij}^0 取值不为 0, 故式 (8) 的惩罚项中, $|\theta_{ij}^0|^{1-q}$ 替代了 $|\theta_{ij}^0|^{1-q}+\epsilon_{ij}$. 有如下定理.

定理1 当 $\sqrt{n}\lambda \to 0$, $n^{1-q/2}\lambda \to \infty$ 时, $\hat{\Theta}$ 满足如下 Oracle 性质:

(1) 变量选择一致性, $P(\hat{\Theta}_{S^c} = 0) \to 1$;

(2) 渐近正态性, 当 $n \to \infty$ 时, $\sqrt{n}(\hat{\Theta} - \Theta) =_d \sqrt{n}(\hat{\Theta}^{\mathrm{t}} - \Theta)$, 且均服从正态分布. 其中, "$=_d$" 表示分布相同.

证明 令

$$\begin{aligned} V_n(U) &= J\left(\Theta + \frac{U}{\sqrt{n}}\right) - J(\Theta) \\ &= -\log\left|\Theta + \frac{U}{\sqrt{n}}\right| + \log|\Theta| + \mathrm{tr}\left\{\left(\Theta + \frac{U}{\sqrt{n}}\right)\hat{\Sigma}\right\} - \mathrm{tr}(\Theta\hat{\Sigma}) \\ &\quad + \lambda\sum_{i\neq j}\left(\frac{1}{\|\theta^0_{-i}\|_q^q + \epsilon_i} + \frac{1}{\|\theta^0_{-j}\|_q^q + \epsilon_j}\right)\frac{q}{|\theta^0_{ij}|^{1-q}}\left(\left|\theta_{ij} + \frac{u_{ij}}{\sqrt{n}}\right| - |\theta_{ij}|\right), \end{aligned}$$

其中, $U = (u_{ij})_{(1\leqslant i,j\leqslant p)}$. 注意到, Θ 为真实的协方差矩阵的逆. 为书写方便, 记

$$I_1 = -\log\left|\Theta + \frac{U}{\sqrt{n}}\right| + \log|\Theta| + \mathrm{tr}\left\{\left(\Theta + \frac{U}{\sqrt{n}}\right)\hat{\Sigma}\right\} - \mathrm{tr}(\Theta\hat{\Sigma}),$$

$$I_2 = \lambda\sum_{i\neq j}\left(\frac{1}{\|\theta^0_{-i}\|_q^q + \epsilon_i} + \frac{1}{\|\theta^0_{-j}\|_q^q + \epsilon_j}\right)\frac{q}{|\theta^0_{ij}|^{1-q}}\left(\left|\theta_{ij} + \frac{u_{ij}}{\sqrt{n}}\right| - |\theta_{ij}|\right).$$

从而

$$V_n(U) = I_1 + I_2.$$

易知

$$\hat{U} = \arg\min_U V_n(U) = \sqrt{n}(\hat{\Theta} - \Theta).$$

对于 I_1, 一方面

$$\log\left|\Theta + \frac{U}{\sqrt{n}}\right| - \log|\Theta| = \log\left|I + \frac{\Sigma^{1/2}U\Sigma^{1/2}}{\sqrt{n}}\right|,$$

其中, I 为单位矩阵. 将实对称矩阵 $\Sigma^{1/2}U\Sigma^{1/2}$ 特征分解为 $B'(\Sigma^{1/2}U\Sigma^{1/2})B = C$, 其中, $B'B = I$ 且 B 为以 $\Sigma^{1/2}U\Sigma^{1/2}$ 的特征向量为列组成的矩阵, C 为对角矩阵且其对角线元素为 $\Sigma^{1/2}U\Sigma^{1/2}$ 的特征值. 基于上述特征分解, 有

$$\begin{aligned} \log\left|I + \frac{\Sigma^{1/2}U\Sigma^{1/2}}{\sqrt{n}}\right| &= \log\left|B'\left(I + \frac{\Sigma^{1/2}U\Sigma^{1/2}}{\sqrt{n}}\right)B\right| = \log\left|I + \frac{C}{\sqrt{n}}\right| \\ &= \log\prod_{i=1}^p\left\{1 + \frac{\sigma_i(\Sigma^{1/2}U\Sigma^{1/2})}{\sqrt{n}}\right\} = \sum_{i=1}^p\log\left\{1 + \frac{\sigma_i(\Sigma^{1/2}U\Sigma^{1/2})}{\sqrt{n}}\right\}, \end{aligned}$$

其中, $\sigma_i(\cdot)$ 表示矩阵的第 i 个特征值. 由于

$$\log\left\{1 + \frac{\sigma_i(\Sigma^{1/2}U\Sigma^{1/2})}{\sqrt{n}}\right\} = \frac{\sigma_i(\Sigma^{1/2}U\Sigma^{1/2})}{\sqrt{n}} - \frac{\sigma_i^2(\Sigma^{1/2}U\Sigma^{1/2})}{2n} + o\left(\frac{1}{n}\right),$$

并注意到

$$\sum_i\sigma_i^2(\Sigma^{1/2}U\Sigma^{1/2}) = \sum_i\sigma_i(\Sigma^{1/2}U\Sigma^{1/2}\Sigma^{1/2}U\Sigma^{1/2}) = \mathrm{tr}(\Sigma^{1/2}U\Sigma U\Sigma^{1/2}) = \mathrm{tr}(U\Sigma U\Sigma),$$

故

$$\log\left|\Theta+\frac{U}{\sqrt{n}}\right|-\log|\Theta|=\frac{\operatorname{tr}(U\Sigma)}{\sqrt{n}}-\frac{\operatorname{tr}(U\Sigma U\Sigma)}{2n}+o\left(\frac{1}{n}\right). \tag{9}$$

注意到, 上式用到了维数 p 固定这一条件.

另一方面

$$\operatorname{tr}\left\{\left(\Theta+\frac{U}{\sqrt{n}}\right)\hat{\Sigma}\right\}-\operatorname{tr}(\Theta\hat{\Sigma})=\operatorname{tr}\left(\frac{U\hat{\Sigma}}{\sqrt{n}}\right)=\operatorname{tr}\left(\frac{U\Sigma}{\sqrt{n}}\right)+\operatorname{tr}\left(\frac{U(\hat{\Sigma}-\Sigma)}{\sqrt{n}}\right). \tag{10}$$

故结合式 (9) 与 (10), 得到

$$I_1=\frac{1}{2}\operatorname{tr}(U\Sigma U\Sigma)+\operatorname{tr}\{U\sqrt{n}(\hat{\Sigma}-\Sigma)\}+o(1).$$

对于 I_2, 为了书写方便, 记

$$T_{ij}=n\lambda\left(\frac{1}{\|\theta^0_{-i}\|_q^q+\epsilon_i}+\frac{1}{\|\theta^0_{-j}\|_q^q+\epsilon_j}\right)\frac{q}{|\theta^0_{ij}|^{1-q}}\left(\left|\theta_{ij}+\frac{u_{ij}}{\sqrt{n}}\right|-|\theta_{ij}|\right).$$

从而

$$nI_2=\sum_{i\neq j}T_{ij}.$$

当 $\theta_{ij}\neq 0$ 即 $(i,j)\in S$ 时, 易知

$$\left|\theta_{ij}+\frac{u_{ij}}{\sqrt{n}}\right|-|\theta_{ij}|=\frac{u_{ij}}{\sqrt{n}}\operatorname{sign}(\theta_{ij}),$$

其中, $\operatorname{sign}(\theta_{ij})$ 为 θ_{ij} 的符号函数. 记 $M(\Theta^0,\epsilon)=\frac{1}{\|\theta^0_{-i}\|_q^q+\epsilon_i}+\frac{1}{\|\theta^0_{-j}\|_q^q+\epsilon_j}$. 则

$$T_{ij}=n\lambda M(\Theta^0,\epsilon)\frac{q}{|\theta^0_{ij}|^{1-q}}\frac{u_{ij}}{\sqrt{n}}\operatorname{sign}(\theta_{ij})=\sqrt{n}\lambda M(\Theta^0,\epsilon)\frac{q}{|\theta^0_{ij}|^{1-q}}u_{ij}\operatorname{sign}(\theta_{ij}).$$

由文献 [19] 可知,

$$\sqrt{n}(\theta^0-\theta)\to_d N(0,I^{-1}(\theta)), \tag{11}$$

其中, θ^0,θ 分别为矩阵 Θ^0,Θ 向量化后的向量. $I(\theta)$ 为 θ 的 Fisher 信息矩阵. 故由 Slutsky 定理得到当 $\theta_{ij}\neq 0$ 时,

$$\theta^0_{ij}\to_p\theta_{ij}.$$

结合定理 1 条件 $\sqrt{n}\lambda\to 0$ 与 $M(\Theta^0,\epsilon)$ 的有界性, 得到当 $\theta_{ij}\neq 0$ 时, $T_{ij}\to 0$.

当 $\theta_{ij}=0$ 即 $(i,j)\in S^c$ 时,

$$T_{ij}=n\lambda M(\Theta^0,\epsilon)\frac{q}{|\theta^0_{ij}|^{1-q}}\left|\frac{u_{ij}}{\sqrt{n}}\right|=n^{1-q/2}\lambda\sum_{i\neq j}M(\Theta^0,\epsilon)\frac{q}{|\sqrt{n}\theta^0_{ij}|^{1-q}}.$$

由式 (11), 得到当 $\theta_{ij}=0$ 时,

$$\sqrt{n}\theta^0_{ij}=O_p(1).$$

故结合定理 1 条件 $n^{1-q/2}\lambda\to\infty$, 可以得到当 $u_{ij}=0$ 时, $T_{ij}=0$; 当 $u_{ij}\neq 0$ 时, $T_{ij}\to\infty$.

因此, 结合 I_1 与 I_2 的渐近性质, 对于固定的 U, 有

$$nV_n(U) \to_d V(U) = \begin{cases} \frac{1}{2}\mathrm{tr}(U\Sigma U\Sigma) + \mathrm{tr}(UW_n), & U_{S^c} = 0, \\ \infty, & \text{otherwise}, \end{cases}$$

其中, $U_{S^c} = 0$ 表示 U 在下标集合 S^c 上的元素均为 0, $W_n = \sqrt{n}(\hat{\Sigma} - \Sigma)$. 由文献 [13,20] 中结果并注意到

$$\hat{U} = \arg\min nV_n(U) = \sqrt{n}(\hat{\Theta} - \Theta),$$

有

$$\sqrt{n}(\hat{\Theta} - \Theta) \to_d \arg\min_{U_{S^c}=0} \left\{ \frac{1}{2}\mathrm{tr}(U\Sigma U\Sigma) + \mathrm{tr}(UW_n) \right\}.$$

注意到

$$\sqrt{n}(\hat{\Theta}^{\mathrm{t}} - \Theta) \to_d \arg\min_{U_{S^c}=0} \left\{ \frac{1}{2}\mathrm{tr}(U\Sigma U\Sigma) + \mathrm{tr}(UW_n) \right\}.$$

因此, 性质 (2) 成立.

下面证明性质 (1), 即 $(i,j) \in S^c$ 时, $P(\hat{\theta}_{ij} \neq 0) \to 0$. 假设 $\hat{\theta}_{ij} \neq 0$, 则由 KKT 条件可以得到

$$\sqrt{n}(-\hat{\Theta}^{-1} + \hat{\Sigma})_{(i,j)\in S^c} = -\lambda n^{1-q/2}\mathrm{sign}(\hat{\theta}_{ij}) M(\Theta^0, \epsilon) \frac{q}{|\sqrt{n}\theta^0_{ij}|^{1-q}}. \tag{12}$$

由于 $n^{1-q/2}\lambda \to \infty$ 及 $\sqrt{n}\theta^0_{ij} = O_p(1)$, 故式 (12) 右边趋于无穷大. 而

$$\sqrt{n}(-\hat{\Theta}^{-1} + \hat{\Sigma})_{(i,j)\in S^c} = \sqrt{n}(-\hat{\Theta}^{-1} + \Sigma - \Sigma + \hat{\Sigma})_{(i,j)\in S^c},$$

故式 (12) 左边依分布收敛于正态分布. 因此, 等式 (12) 不成立, 与假设矛盾, 所以 $(i,j) \in S^c$ 时,

$$P(\hat{\theta}_{ij} \neq 0) \to 0,$$

性质 (1) 成立.

注 2 以上证明基于 Yuan 和 Lin [6]、Knight 和 Fu [13] 以及 Zou 和 Li [21] 的思想. 其中, Zou 和 Li [21] 研究一系列非凸正则化问题如 SCAD, MCP, LSP, 证明了如果初值满足一定条件, 重赋权算法一步迭代得到的解具有 Oracle 性质.

注 3 定理 1 为模型提供了理论保证, 而 Liu 与 Ihler 提出的模型 [12] 并不具有上述理论性质.

4 实验

本节分别用数值实验和真实数据实验比较 Graphical Lasso [6]、Tan 等提出的 Hub Graphical Lasso [11]、Liu 和 Ihler [12] 的模型及所提出的模型在参数估计、模型选择以及提取网络无标度特征的能力方面的优劣.

4.1 数值实验

模型的生成借鉴 Liu 与 Ihler [12] 的方法. 首先, 根据 Barabási 和 Albert [22] 提出的算法生成无标度网络, 设对应邻接矩阵为 A. 然后, 利用 A 的信息生成协方差矩阵的逆 Θ. 具体地, 令 $L = \eta D - A$, 其

表 1　数值实验结果 n=100, p=20

Table 1　The simulation results with $n=100$, $p=20$

Model	KL loss	FPR	TPR	FP	TP	FN	TN
gLasso	1.113(0.213)	0.23	**1.00**	38.50	**19.00**	**0.00**	132.50
hgLasso	3.196(0.337)	0.21	**1.00**	36.60	**19.00**	**0.00**	134.40
2sLog L_1	1.147(0.230)	0.14	1.00	24.60	**19.00**	**0.00**	146.40
2sLog $L_{0.5}$	**0.767(0.163)**	**0.05**	**1.00**	**9.10**	**19.00**	**0.00**	**161.90**
1sLog L_1	1.124(0.169)	0.17	**1.00**	36.60	**19.00**	**0.00**	134.40
1sLog $L_{0.5}$	0.874(0.140)	0.08	0.98	14.30	18.70	0.30	156.70

中 D 为对角矩阵且每个对角线元素为相应结点的度, η 为使 L 可逆的大于 1 的常数. 令 $\Theta=\Lambda^{\frac{1}{2}}L\Lambda^{\frac{1}{2}}$, 其中 Λ 为 L^{-1} 的对角线元素组成的对角矩阵. 注意到协方差矩阵 $\Sigma=\Theta^{-1}$ 的对角线元素即随机向量的每个分量方差均为 1. 最后生成 n 个服从 p 维正态分布 $N_p(0,\Theta^{-1})$ 的样本 $X_1,\ldots,X_n$. 本文进行了 4 组实验, 样本个数 n 为 100, 维数 p 分别为 20, 30, 50, 100. 对于每组 (n,p), 生成 10 个独立的数据集.

从参数估计和模型选择两方面比较各种方法的优劣. 参数估计的准确性用 Yuan 和 Lin [6] 所采用的 KL 损失来衡量:

$$\mathrm{KL}=-\log|\hat{\Theta}|+\mathrm{tr}(\hat{\Theta}\hat{\Sigma})-(-\log|\Sigma^{-1}|+p).$$

用 FPR (false positive rate), TPR (true positive rate), 及 FP (false positive), TP (true positive), FN (false negative), TN (true negative) 的个数作为比较依据, 检测模型选择的一致性. 其中, FP 为真实网络无边, 模型估计为有边的个数; TP 为真实网络有边, 模型估计为有边的个数; FN 为真实网络有边, 模型估计为无边的个数; TN 为真实网络无边, 模型估计为无边的个数; FPR=FP/(FP+TN); TPR=TP/(TP+FN).

实验中, 使用如下 BIC 准则选择正则化参数 λ:

$$\mathrm{BIC}(\lambda)=-\log|\hat{\Theta}(\lambda)|+\mathrm{tr}\{\hat{\Theta}(\lambda)\hat{\Sigma}\}+\frac{\log n}{n}\sum_{i\neq j}\hat{e}_{ij}(\lambda),$$

其中, 如果 $\hat{\theta}_{ij}=0$, 则 $\hat{e}_{ij}=0$; 此外, $\hat{e}_{ij}=1$. 对于 Graphical Lasso, 用 Friedman 等 [7] 提出的分块坐标下降法求解. 对于本文及 Liu 与 Ihler 提出的模型, 我们用重赋权算法求解. 其中, 仍用分块坐标下降法求解改变权重后的 L_1 正则化问题. 根据诸多文献结果 [12,21], 两步重赋权求解非凸问题效果明显. 因此, 实验中采用了初值为 Graphical Lasso 的两步重赋权求解. 此外, 为了验证本文定理 1 的有效性, 本文也比较了初值为 $\hat{\Sigma}^{-1}$ 的一步重赋权的结果.

表 1∼4 分别为 4 组实验在 10 个独立数据集上平均的结果, 最好的结果已用黑体标出. 其中, 括号中的数字为 10 次计算的标准误差, gLasso 代表 Graphical Lasso, hgLasso 代表 Tan 等提出的 Hub Graphical Lasso, 我们选择 $q=0.5$ 作为本文模型的代表, 1sLog L_q 代表初值为 $\hat{\Sigma}^{-1}$ 的一步重赋权求解本文提出的模型 ($q=0.5$) 及 Liu 与 Ihler 的模型 ($q=1$), 相应地, 2sLog L_q 代表初值为 Graphical Lasso 的两步重赋权求解本文提出的模型 ($q=0.5$) 及 Liu 与 Ihler 的模型 ($q=1$).

从表 1∼4 可以看出, 在参数估计方面, 本文提出的模型具有更小的 KL 损失, 其参数估计的准确率是最高的, 这有助于进一步分析节点之间的关系 (偏相关关系), 为实际应用提供帮助. 由于 Graphical Lasso 以及 Hub Graphical Lasso 并未考虑网络无标度这一先验信息, 所以其参数估计准确率低于本文提出的模型, 以及 Liu 和 Ihler 的模型.

表 2　数值实验结果 n=100, p=30

Table 2　The simulation results with $n = 100$, $p = 30$

Model	KL loss	FPR	TPR	FP	TP	FN	TN
gLasso	1.999(0.406)	0.15	**1.00**	58.90	28.90	0.10	347.10
hgLasso	3.692(0.539)	0.22	**1.00**	90.20	**29.00**	**0.00**	315.80
2sLog L_1	1.786(0.267)	0.10	**1.00**	40.50	**29.00**	**0.00**	365.50
2sLog $L_{0.5}$	**1.049(0.222)**	**0.04**	0.99	**17.10**	28.80	0.20	**388.90**
1sLog L_1	1.890(0.381)	0.11	**1.00**	44.80	**29.00**	**0.00**	361.20
1sLog $L_{0.5}$	1.360(0.320)	0.06	0.99	36.30	28.70	0.30	379.70

表 3　数值实验结果 n=100, p=50

Table 3　The simulation results with $n = 100$, $p = 50$

Model	KL loss	FPR	TPR	FP	TP	FN	TN
gLasso	4.463(0.519)	0.06	0.98	75.80	48.20	0.80	1100.20
hgLasso	6.308(0.845)	0.13	**0.99**	151.60	**48.60**	**0.40**	1024.40
2sLog L_1	3.639(0.515)	0.05	**0.99**	59.80	48.50	0.50	1116.20
2sLog $L_{0.5}$	**1.849(0.372)**	**0.03**	0.98	**38.50**	48.10	0.90	**1137.50**

表 4　数值实验结果 n=100, p=100

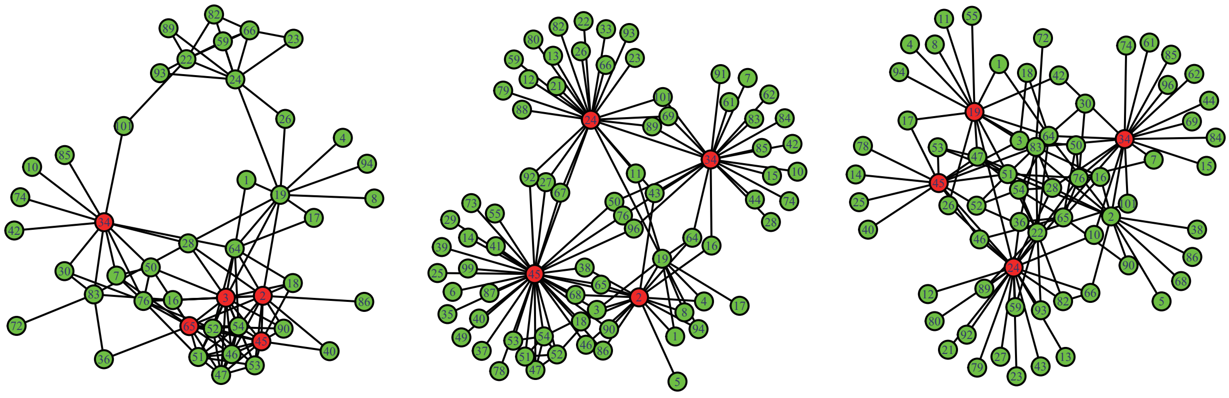
Table 4　The simulation results with $n = 100$, $p = 100$

Model	KL loss	FPR	TPR	FP	TP	FN	TN
gLasso	12.021(1.162)	**0.02**	0.90	98.50	88.70	10.30	4752.50
hgLasso	12.718(0.714)	0.26	**0.98**	1289.00	**97.10**	**1.90**	3562.00
2sLog L_1	10.149(880)	**0.02**	0.91	108.20	89.70	9.30	4742.80
2sLog $L_{0.5}$	**4.918(0.844)**	**0.02**	0.91	**76.90**	90.40	8.60	**4774.10**

在模型选择方面, $p = 20$ 时, 所提出的模型在所有模型中是最优的; $p = 30, 50, 100$ 时, 没有一种模型在所有方面是最优的. 下面用 $p = 30$ 的两步迭代结果为例说明. 节点个数为 30 时, Liu 与 Ihler 的模型和本文的模型是效果最好的两个模型. 下面进一步比较, Liu 与 Ihler 的模型所得的网络 10 次平均的 FP, TP, FN, TN 分别为 40.50, 29.00, 0, 365.50; 本文的模型 ($q = 0.5$) 10 次平均的 FP, TP, FN, TN 分别为 17.10, 28.80, 0.20, 388.90. 在 FP, TN 方面, 本文的模型比 Liu 和 Ihler 的模型少了 23.4 条错误的边, 故所提出的模型明显好于 Liu 和 Ihler 的模型; 而在 TP, FN 方面, 本文的模型比 Liu 和 Ihler 的模型少了平均 0.2 条正确的边. 故两个模型的能力相当, 或者说本文的模型稍弱于 Liu 和 Ihler 的模型. 所以, 虽然没有一个模型在所有方面是最优的, 但是综合来看, 本文的模型在其好的方面具有较为明显的优势, 而在其他方面与最好模型能力相当. 另外可以看到, 本文的模型得出的网络更稀疏, 更符合真实网络的特征.

故综合参数估计和模型选择, 本文的模型优于相关模型.

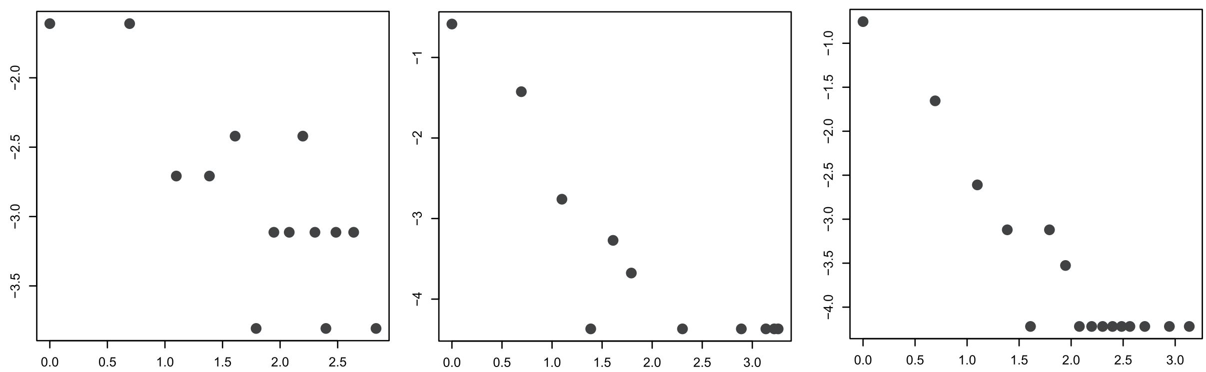
注 4　当 p 较小时, 如 $p = 20, 30$, 一步重赋权得出的解在参数估计和模型选择方面均表现良好. 然而, 当 p 增大时, 如 $p = 50, 100$, 其表现变差 (我们略去其结果). 这是由于定理 1 为低维结果, 而当 $p = 50$ 时, 未知的参数已大约有 1250 个, 导致定理 1 失效. 所以研究模型的高维统计性质是我们未来的工作.



(a)　　(b)　　(c)

图 1　(网络版彩图)3 种模型应用于基因微阵列数据生成的网络

Figure 1　(Color online) Network estimated on gene microarray data. (a) Graphical Lasso; (b) the model proposed by Liu and Ihler; (c) the model proposed in this paper. Each of the three networks has equal number of 120 edges. Red nodes are hubs and non-hub nodes are colored by green. Obviously, in (b) and (c), hub nodes are more distinguishable from non-hub nodes.



(a)　　(b)　　(c)

图 2　3 种模型应用于基因微阵列数据生成的网络对应的节点度 (取对数后) 的分布

Figure 2　Log-log plot of the degree distributions of the networks estimated on gene microarray data, where the horizontal axis stands for the degrees of nodes after log transformations, and the vertical axis represents the corresponding proportions of the number of nodes after log transformations. (a) Graphical lasso; (b) the model proposed by Liu and Ihler; (c) the model proposed in this paper. Noticeably, in (b) and (c), the estimated networks are more scale-free.

4.2　基因微阵列数据

本小节将 Graphical Lasso [6], Liu 和 Ihler [12] 的模型及所提出的模型应用于基因微阵列数据 [23], 比较 3 种方法在模型选择, 参数估计以及生成无标度网络的能力. Chen 等 [24], Liu 和 Ihler [12] 均使用该数据研究无标度网络的重建问题.

该数据为通过微阵列实验获得的酵母细胞的基因表达数据. 包括在 18 个时间点上采集的 102 个基因的表达值, 即样本个数为 18, 维数为 102. 将 3 种模型分别应用于该数据. 与 Liu 和 Ihler 相同, 我们选取边数为 120 的网络. 图 1 为 3 种模型生成的网络, 图 2 为网络对应的节点度 (取对数后) 的分布.

从图 1 可以看到, 与 Graphical Lasso 相比, 本文的模型及 Liu 和 Ihler 的模型生成的网络具有更

明显的中心点 (Hubs). 且本文的模型与 Liu 和 Ihler 的模型生成的网络具有 3 个共同的中心点, 说明二者之间发现中心点的能力相当.

图 2 为网络节点度 (取对数后) 的分布. 注意到当节点的度服从幂律分布, 即 $P(d) \propto d^{-\alpha}$ 时,

$$\log(P(d)) = -\alpha d + c,$$

其中 c 为常数. 故取对数后, 无标度网络节点的度的分布应近似为线性的. 因此, 本文提出的模型及 Liu 和 Ihler 的模型生成的网络更具有无标度的特征.

与 Liu 和 Ihler 的工作相比, 所提模型的一大亮点为其具有良好的理论性质, 即当初值选择合理时, 一步重赋权算法得出的解具有 Oracle 性质. 下面从参数估计角度说明所提出的模型较 Liu 与 Ihler 的模型的优势. 由于真实的协方差矩阵的逆 Θ 未知, 因此不能类似于数值实验用 KL 损失比较参数估计准确率. 比较原始数据的样本协方差矩阵与从估计出分布抽样得出的样本对应的协方差矩阵之间的距离, 从而判断这组数据更像是从哪个总体分布中抽样得出. 具体地, 假设原始数据的样本协方差矩阵

$$\hat{\Sigma} = \frac{1}{n} X' X,$$

其中, $X = (X_1, \ldots, X_{102})$ 为网络节点对应的 102 个变量的取值, n 为样本个数, 这里为 18. 假设 Liu 与 Ihler 的模型估计出的协方差矩阵为

$$\hat{\Sigma}_1 = \hat{\Theta}_1^{-1},$$

所提模型估计出的协方差矩阵为

$$\hat{\Sigma}_2 = \hat{\Theta}_2^{-1}.$$

从 $N(0, \hat{\Sigma}_1)$ 中随机抽出 18 组数据 $X^{(1)}, \ldots, X^{(18)}$, 并计算其样本协方差矩阵 $\hat{\Sigma}_{11}$; 从 $N(0, \hat{\Sigma}_2)$ 中随机抽出 18 组数据 $Y^{(1)}, \ldots, Y^{(18)}$, 并计算其样本协方差矩阵 $\hat{\Sigma}_{22}$. 计算得

$$\|\hat{\Sigma} - \hat{\Sigma}_{11}\|_{\mathrm{F}} = 2.305, \quad \|\hat{\Sigma} - \hat{\Sigma}_{22}\|_{\mathrm{F}} = 1.877.$$

因此, 原始数据的样本协方差矩阵与本文估计模型抽样得出的样本协方差矩阵距离更小. 从而所提出的模型具有更高的准确率.

5 结论

本文基于无标度先验研究图模型的结构学习问题. 通过在模型中引入 Log 型和 L_q 型惩罚函数的复合惩罚, 构造了新的正则化模型. 本文采用了重赋权迭代算法求解这一模型. 通过在每一步迭代中改变相应参数的权重, 使得度越大的节点在下一步迭代中具有更小的惩罚. 进一步, 本文证明了当初值为 $\hat{\Sigma}^{-1}$ 且正则化参数选择合理时, 一步重赋权的解具有 Oracle 性质. 实验表明, 所提出的模型更容易生成具有无标度特征的网络且在参数估计、模型选择方面均有良好的效果.

本文主要关注 GGM, 相应结果可以推广到 Ising 模型 [25]、混合图模型 [26] 及多图模型 [27] 等. 以上模型目前均在研究中.

致谢 感谢 Dai Yang 教授和 Liu Qiang 博士分享基因微阵列数据.

参考文献

1 Barabási A L, Albert R. Statistical mechanics of complex networks. Rev Mod Phys, 2002, 74: 47–97
2 Ewards D M. Introduction to Graphical Modelling. New York: Springer, 2000
3 Dempster A P. Covariance selection. Biometrika, 1972, 32: 95–108
4 Tibshirani R. Regression shrinkage and selection via the lasso. J Royal Stat Soc B, 1996, 58: 267–288
5 Meinshausen N, Bühlmann P. High-dimensional graphs with the lasso. Ann Statist, 2006, 34: 1436–1462
6 Yuan M, Lin Y. Model selection and estimation in the Gaussian graphical model. Biometrika, 2007, 94: 19–35
7 Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. Biostat, 2008, 9: 432–441
8 Ravikumar P, Raskutti G, Wainwright M J, et al. High-dimensional covariance estimation by minimizing L_1-penalized log-determinant. Electron J Stat, 2011, 5: 935–980
9 Peng J, Wang P, Zhou N, et al. Partial correlation estimation by joint sparse regression models. J Am Statist Assoc, 2009, 104: 735–746
10 Hero A, Rajaratnam B. Hub discovery in partial correlation graphs. IEEE Trans Inf Theory, 2012, 58: 6064–6078
11 Tan K M, London P, Mohan K, et al. Learning graphical models with hubs. J Mach Learn Res, 2014, 15: 3297–3331
12 Liu Q, Ihler A T. Learning scale free networks by reweighed L_1 regularization. In: Proceedings of the 14th International Conference on Artificial Intelligence and Statistics, Fort Lauderdale, 2011. 15: 40–48
13 Knight K, Fu W J. Asymptotics for lasso-type estimators. Ann Statist, 2000, 28: 1356–1378
14 Fan J Q, Li R Z. Variable selection via nonconcave penalized likelihood and its oracle properties. J Am Statist Assoc, 2001, 96: 1348–1360
15 Huang J, Horowitz J L, Ma S. Asymptotic properties of bridge estimators in sparse high-dimensional regression models. Ann Statist, 2008, 36: 587–613
16 Zhang H, Wang Y, Chang X Y, et al. $L_{1/2}$ regularization. Sci Sin Inform, 2010, 40: 412–422 [张海, 王尧, 常象宇, 等. $L_{1/2}$ 正则化. 中国科学: 信息科学, 2010, 40: 412–422]
17 Lange K, Hunter D, Yang I. Optimization transfer using surrogate objective functions (with discussion). J Comput Graph Statist, 2000, 9: 1–59
18 Witten D, Friedman J H, Simon N. New insights and faster computations for the graphical lasso. J Comput Graph Statist, 2011, 20: 892–900
19 van der Vaart A W. Asymptotic Statistics. New York: Cambridge University Press, 1998. 61–67
20 Geyer C. On the asymptotics of constrainted M-estimation. Ann Statist, 1994, 22: 1993–2010
21 Zou H, Li R Z. One-step sparse estimates in nonconcave penalized likelihood models. Ann Statist, 2008, 36: 1509–1533
22 Barabási A L, Albert R. Emergence of scaling in random networks. Science, 1999, 286: 509–512
23 Spellman P T, Sherlock G, Zhang M Q, et al. Comprehensive identification of cell cycleregulated genes of the yeast saccharomyces cerevisiae by microarray hybridization. Mol Biol Cell, 1998, 9: 3273–3297
24 Chen G, Larsen P, Almasri E, et al. Rank-based edge reconstruction for scale-free genetic regulatory networks. BMC Bioinformatics, 2008, 9: 75
25 Ravikumar P, Wainwright M J, Lafferty J D. High-dimensional Ising model selection using l_1-regularized logistic regression. Ann Statist, 2010, 38: 1287–1319
26 Chen S, Witten D, Shojaie A. Selection and estimation for mixed graphical models. Biometrika, 2015, 102: 47–64
27 Zhang L J, Zhang H. Joint estimation of multiple graphical models via bridge. Appl Math: J Chinese Univ (Ser A), 2014, 92: 127–137 [张凌洁, 张海. 多图模型的联合估计的群桥方法. 高校应用数学学报, 2014, 92: 127–137]

Structure learning in graphical models incorporating the scale-free prior

Xiao GUO[1], Hai ZHANG[1,2*] & Jianglun WU[3]

1 *School of Mathematics, Northwest University, Xi'an* 710127*, China*;

2 *Institute of Applied Mathematics, Academy of Mathematics and System Sciences, Chinese Academy of Sciences, Beijing* 100190*, China*;
3 *Department of Mathematics, College of Science, Swansea University, Swansea* SA2 8PP*, UK*
*E-mail: zhanghai@nwu.edu.cn

Abstract In this paper, we consider the problem of structure learning in graphical models under the prior that the underlying networks are scale free. We propose a novel regularization model, which incorporates the scale-free prior, with a penalty that is a hybrid of the Log-type and L_q-type penalty functions. An iterative reweighted L_1 algorithm is employed to solve the model. Numerical studies show that our method is both effective and practical and performs well in terms of parameter estimation and model selection.

Keywords sparsity, scale-free, graphical model, reweighted L_1 algorithm, regularization

Xiao GUO was born in 1990. She is currently working toward the Ph.D. degree in Statistics from Northwest University, Xi'an, China. Her current research interests include high-dimensional statistics and statistical inference for graphical models.

Hai ZHANG was born in 1975. He received the Ph.D. degree in applied mathematics from Xi'an Jiaotong University, Xi'an, in 2012. He is currently a professor in the Department of Financial Mathematics and Statistics at Northwest University. His research interests include statistical machine learning, high-dimensional statistics, and social network analysis.

Jianglun WU was born in 1963. He received the Ph.D. degree in probability and mathematical statistics from the Institute of Applied Mathematics, Chinese Academy of Sciences, Beijing, in 1991. Currently, he is professor of mathematics at Swansea University, UK. His research interests include stochastic analysis, partial differential equations, functional analysis, random data analysis, statistical and mathematical physics, and mathematical logic. Dr. Wu is a member of the London Mathematical Society, as well as a regular reviewer for Mathematical Reviews.