Joint Gaussian Graphical Model Series – V

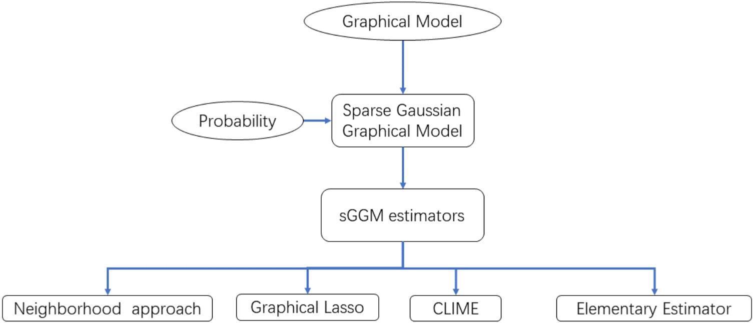
sparse Gaussian Graphical Model estimators

Beilun Wang

[1]Department of Computer Science, University of Virginia
http://jointggm.org/

July 28th, 2017

Road Map


Graphical Model
Probability
Sparse Gaussian Graphical Model
sGGM estimators
Neighborhood approach
Graphical Lasso
CLIME
Elementary Estimator

Outline

1. Notation
2. Review
3. Neighborhood Method
4. Graphical Lasso
5. CLIME
6. Elementary Estimator for Gaussian Graphical Model

Notation

Notation

- Σ The covariance matrix.
- Ω The precision matrix.
- p The number of features.
- n The number of samples.

Review

Review from last talk

- Regularized M-estimator $\underset{\theta}{\operatorname{argmin}} \mathcal{L}(\theta) + \lambda_n \mathcal{R}(\theta)$
- a unified framework to analyze the statistical convergence rate for high-dimensional statistics
- Elementary Estimator

Review of Gaussian Graphical Model

Suppose the precision matrix $\Omega = \Sigma^{-1}$.

The log-likelihood of Ω equals to $\ln\det(\Omega) - \mathrm{tr}\left(\Omega\widehat{S}\right)$

In this talk, we will use this likelihood to derive several estimators of sparse Gaussian Graphical Model (sGGM)

Neighborhood Method

Neighborhood approach

If $X \sim N(0, \Sigma)$ and let $X_1 = X_j$.
$X_j | X_{\setminus j}$ $N(\Sigma_{\setminus j, j} \Sigma^{-1}_{\setminus j, \setminus j} X_{\setminus j}, \Sigma_{jj} - \Sigma_{\setminus j, j} \Sigma^{-1}_{\setminus j, \setminus j} \Sigma_{\setminus j, j})$
Let $\alpha_j := \Sigma_{\setminus j, j} \Sigma^{-1}_{\setminus j, \setminus j}$ and $\sigma_j^2 := \Sigma_{jj} - \Sigma_{\setminus j, j} \Sigma^{-1}_{\setminus j, \setminus j} \Sigma_{\setminus j, j}$. We have that

$$X_j = \alpha_j^T X_{, \setminus j} + \epsilon_j \tag{3.1}$$

where $\epsilon_j \sim N(0, \sigma_j^2)$ is independent of $X_{, \setminus j}$.

Neighborhood approach with sparse assumption

By the sparse assumption, we estimate each α_j by a lasso estimator

$$\alpha_j = \underset{\alpha_j}{\operatorname{argmin}} ||\alpha_j^T X_{,\setminus j} - X_j||_2^2 + \lambda ||\alpha_j||_1 \tag{3.2}$$

Review of Lasso solution

Lasso

$$\beta = \underset{\beta}{\operatorname{argmin}} \, ||\beta^T X - y||_2^2 + \lambda ||\beta||_1 \tag{3.3}$$

subgradient method

$$g(\beta; \lambda) = -2X^T(y - X\beta) + \lambda \text{sgn}(\beta) \tag{3.4}$$

Review of Lasso solution: State of the Art

We see that the proximity operator is important because x^* is a minimizer to the problem $\min_{x \in \mathcal{H}} F(x) + R(x)$ if and only if $x^* = \text{prox}_{\gamma R}\left(x^* - \gamma \nabla F(x^*)\right)$, where $\gamma > 0$. γ is any positive real number.

Proximal gradient method

$$\left(\text{prox}_{\gamma R}(x)\right)_i = \begin{cases} x_i - \gamma, & x_i > \gamma \\ 0, & |x_i| \leq \gamma \\ x_i + \gamma, & x_i < -\gamma, \end{cases} \tag{3.5}$$

By using the fixed point method, you can obtain the estimation of β.

Graphical Lasso

Graphical Lasso[Friedman et al.(2008)Friedman, Hastie, and Tibshirani

We already have the log-likelihood as the loss function. Can we use it to obtain a similar estimator as Lasso?

$$\underset{\Omega}{\operatorname{argmin}} - \ln \det(\Omega) + \operatorname{tr}\left(\Omega \widehat{S}\right) + \lambda_n ||\Omega||_1 \tag{4.1}$$

Proximal gradient method to solve it

Let's do a practice in the white board.

Super Linear algorithm.

$\lim_{k\to\infty} \frac{|x_{k+1}-x^*|}{|x_k-x^*|} = 0.$

State of the art method: Big & QUIC[Hsieh et al.(2011)Hsieh, Sustik, Dhillon, and Ravikum

Parallelized Coordinate descent.

approximated quadratic algorithm.

$\lim_{k \to \infty} \frac{|x_{k+1} - x^*|}{|x_k - x^*|^2} < M$

CLIME

CLIME[Cai et al.(2011)Cai, Liu, and Luo]

CLIME

$$\underset{\Omega}{\operatorname{argmin}} ||\Omega||_1 \ \text{, subject to: } ||\Sigma\Omega - I||_\infty \leq \lambda \tag{5.1}$$

Here $\lambda > 0$ is the tuning parameter.

By taking the first derivative of Eq. (4.1) and setting it equal to zero, the solution $\widehat{\Omega}_{glasso}$ also satisfies:

$$\widehat{\Omega}_{glasso}^{-1} - \widehat{\Sigma} = \lambda \widehat{Z} \tag{5.2}$$

where $\widehat{Z}$ is an element of the subdifferential $\partial ||\widehat{\Omega}_{glasso}||_1$.

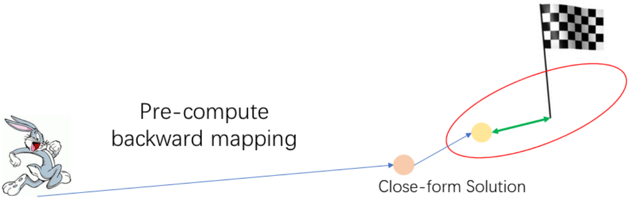
Column-wise estimator

$$\text{argmin}\,||\beta||_1 \quad \text{subject to} \quad ||\Sigma\beta - e_j||_\infty \leq \lambda$$

CLIME can be estimated column-by-column.

Elementary Estimator for Gaussian Graphical Model

Elementary Estimator



Elementary Estimator[Yang et al.(2014b)Yang, Lozano, and Ravikumar]

$$\underset{\theta}{\operatorname{argmin}} \mathcal{R}(\theta) \tag{6.1}$$

$$\text{Subject to: } \mathcal{R}^*(\theta - \mathcal{B}^*(\widehat{\phi})) \le \lambda_n \tag{6.2}$$

Here $\mathcal{B}^*(\widehat{\phi})$ is a backward mapping for $\widehat{\phi}$.

Example: sparse linear regression[Yang et al.(2014a)Yang, Lozano, and Ravikumar]

$$\operatorname*{argmin}_{\theta} ||\theta||_1 \tag{6.3}$$

$$\text{Subject to: } ||\theta - (X^T X + \epsilon I)^{-1} X^T y||_\infty \leq \lambda_n \tag{6.4}$$

Elementary Estimator for sGGM

$$\begin{aligned} &\underset{\Omega}{\operatorname{argmin}} \, |\Omega|_{1,off} \\ \text{subject to:} &|\Omega - [T_v(\Sigma)]^{-1}|_{\infty,off} \leq \lambda_n \end{aligned} \tag{6.5}$$

Summary

- We review most sGGM estimators.

References I

T. Cai, W. Liu, and X. Luo.
A constrained l1 minimization approach to sparse precision matrix estimation.
Journal of the American Statistical Association, 106(494):594–607, 2011.

J. Friedman, T. Hastie, and R. Tibshirani.
Sparse inverse covariance estimation with the graphical lasso.
Biostatistics, 9(3):432–441, 2008.

C.-J. Hsieh, M. A. Sustik, I. S. Dhillon, and P. D. Ravikumar.
Sparse inverse covariance matrix estimation using quadratic approximation.
In *NIPS*, pages 2330–2338, 2011.

References II

E. Yang, A. Lozano, and P. Ravikumar.
Elementary estimators for high-dimensional linear regression.
In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 388–396, 2014a.

E. Yang, A. C. Lozano, and P. K. Ravikumar.
Elementary estimators for graphical models.
In *Advances in Neural Information Processing Systems*, pages 2159–2167, 2014b.