Review of Gaussian Graphical Model

Meilei Jiang[†]

[†]Department of Statistics and Operations Research
University of North Carolina at Chapel Hill

October 27, 2015

Overview

Gaussian Graphical Model

- $\mathbf{X} = (X^{(1)}, X^{(2)}, \cdots, X^{(p)})$ and $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
  - Precision matrix $\mathbf{C} = \boldsymbol{\Sigma}^{-1}$.
  - Partial correlation between $X^{(i)}$ and $X^{(j)}$: $\rho^{ij} = -(c_{ij}/\sqrt{c_{ii}c_{jj}})$.
- Graph $G = (V, E)$, $V = \{X^{(1)}, X^{(2)}, \cdots, X^{(p)}\}, E \in V \times V$.
  - $e_{ij} \in E$ if and only if $c_{ij} \neq 0$
- $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ is a random sample of $\mathbf{X}$.

Log-likelihood Estimation

- Log-likelihood function $l(\mathbf{C}) = \frac{1}{2}\log\det(\mathbf{C}) - \frac{1}{2}\langle\hat{\mathbf{\Sigma}}, \mathbf{C}\rangle$ + Constant.
  - $\hat{\mathbf{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'$
  - $\langle\hat{\mathbf{\Sigma}}, \mathbf{C}\rangle = \mathrm{tr}(\hat{\mathbf{\Sigma}}\mathbf{C})$
- When $n > p$, the global maximizer of $l(\mathbf{C})$ is given by $\hat{\mathbf{C}} = \hat{\mathbf{\Sigma}}^{-1}$.

Penalized Likelihood Estimation

In order to estimate a sparse graph, a penalty term $\sum_{i=1}^{p}\sum_{j=1}^{p} p_{\lambda_{ij}}(|c_{ij}|)$ has been added.

- LASSO: $\max_{\mathbf{C}\in S_p} \log\det(\mathbf{C}) - \langle \hat{\mathbf{\Sigma}}, \mathbf{C}\rangle - \lambda \sum_{i\neq j} |c_{ij}|$
- SCAD: $\max_{\mathbf{C}\in S_p} \log\det(\mathbf{C}) - \langle \hat{\mathbf{\Sigma}}, \mathbf{C}\rangle - \sum_{i=1}^{p}\sum_{j=1}^{p} \mathrm{SCAD}_{\lambda,a}(|c_{ij}|)$
  - $\mathrm{SCAD}'_{\lambda,a}(x) = \lambda\left\{I(|x|\leq\lambda) + \frac{(a\lambda-|x|)_+}{(a-1)\lambda} I(|x|>\lambda)\right\}$
- Adaptive LASSO: $\max_{\mathbf{C}\in S_p} \log\det(\mathbf{C}) - \langle \hat{\mathbf{\Sigma}}, \mathbf{C}\rangle - \lambda \sum_{i=1}^{p}\sum_{j=1}^{p} \tilde{c}_{ij}|c_{ij}|$.
  - $\tilde{\mathbf{C}} = (\tilde{c}_{ij})$ can be any consistent estimate of $\mathbf{C}$.
- Iterative reweighted penalized likelihood: $\max_{\mathbf{C}\in S_p} \log\det(\mathbf{C}) - \langle \hat{\mathbf{\Sigma}}, \mathbf{C}\rangle - \sum_{i=1}^{p}\sum_{j=1}^{p} \tilde{c}_{ij}|c_{ij}|$.
  - $\tilde{c}_{ij} = p'_{\lambda}(|\hat{c}_{ij}^{(k)}|)$

Model Selection in Penalized Likelihood Estimation

- $\text{BIC}(\lambda) = -\log\det(\lambda) + \text{tr}(\hat{\mathbf{C}}\boldsymbol{\Sigma}) + \frac{\log n}{n}\sum_{i\leq j}\hat{e}_{ij}(\lambda)$
  - $\hat{e}_{ij} = 0$, if $\hat{c}_{ij} = 0$, and $\hat{e}_{ij} = 1$ otherwise.
- The K-fold cross validation:
  $\text{CV}(\lambda) = \sum_{k=1}^{K}\left(n_k \log|\hat{\mathbf{C}}_{-k}(\lambda)| - \sum_{i\in T_k}\mathbf{X}_i^T\hat{\mathbf{C}}_{-k}(\lambda)\mathbf{X}_i\right)$
  - T_k is the kth fold of training set with size n_k.
  - $\hat{\mathbf{C}}_{-k}(\lambda)$ is estimated based on the training samples without T_k.

Local Neighborhood Selection

- The neighborhood ne_a of the node $X^{(a)}$ is the smallest subset of V, such that $X^{(a)} \perp \{X^{(k)}; k \in V \setminus \mathrm{cl}_a\}$.
  - $\mathrm{cl}_a = \mathrm{ne}_a \cup a$.
- Optimal prediction: $\theta^a = \arg\min_{\theta:\theta_a=0} \mathbb{E}(X^{(a)} - \sum_{k\in V} \theta_k X^{(k)})$.
  - $\theta^a_b = -c_{ab}/c_{aa}$.
  - $\mathrm{ne}_a = \{b \in V : \theta^a_b \neq 0\}$.
- Neighborhood selection with the Lasso:
  $\hat{\theta}^{a,\lambda} = \arg\min_{\theta:\theta_a=0} n^{-1}\|(\mathbf{X}^{(a)} - \mathbf{X}\theta)\|_2^2 + \lambda\|\theta\|_1$.
  - $\hat{\mathrm{ne}}^\lambda_a = \{b \in V : \hat{\theta}^{a,\lambda}_b \neq 0\}$.

Local Neighborhood selection

- The edge set $E = \{(a, b) : a \in \mathrm{ne}_b \wedge b \in \mathrm{ne}_a\}$
  - $\hat{E}^{\lambda,\wedge} = \{(a, b) : a \in \hat{\mathrm{ne}}_b \wedge b \in \hat{\mathrm{ne}}_a\}$
  - $\hat{E}^{\lambda,\vee} = \{(a, b) : a \in \hat{\mathrm{ne}}_b \vee b \in \hat{\mathrm{ne}}_a\}$
- Under certain assumptions, both $\hat{E}^{\lambda,\wedge}$ and $\hat{E}^{\lambda,\vee}$ are consistent.

Joint Sparse Regression Model

- $L_n(\boldsymbol{\theta}, \mathbf{C}, \mathbf{X}) = \frac{1}{2}\Big(\sum_{i=1}^{p} w_i \|\mathbf{X}^{(i)} - \sum_{j \neq i} \beta_{ij} \mathbf{X}^{(j)}\|^2 \Big)$
  - $\beta_{ij} = \rho^{ij} \sqrt{\frac{c_{jj}}{c_{ii}}}$
  - $\boldsymbol{\theta} = (\rho^{12}, \cdots, \rho^{(p-1)p})^T$
- $\mathcal{L}_n(\boldsymbol{\theta}, \mathbf{C}, \mathbf{X}) = L_n(\boldsymbol{\theta}, \mathbf{C}, \mathbf{X}) + \mathcal{J}(\boldsymbol{\theta})$.
  - $\mathcal{J}(\boldsymbol{\theta}) = \lambda \|\boldsymbol{\theta}\|_1$.
- Solve $\mathcal{L}_n(\boldsymbol{\theta}, \mathbf{C}, \mathbf{X})$: active-shooting

Joint Estimation of Multiple Graphical Models

- Heterogeneous dataset with p variables and K categories.
  - $\mathbf{X}_i^{(k)} = (X_{i,1}^{(k)}, X_{i,2}^{(k)}, \cdots, X_{i,p}^{(k)}) \sim \mathcal{N}(0, \mathbf{\Sigma}^{(k)}), i = 1, \ldots, n_k; k = 1, \ldots, K.$
  - $\mathbf{C}^k = (\mathbf{\Sigma}^{(k)})^{-1}$
- Reparameterization: $c_{jj'}^{(k)} = \theta_{jj'}\gamma_{jj'}^{(k)}, 1 \leq j \neq j' \leq p; k = 1, \ldots, K.$
  - $\theta_{jj'} = \theta_{j'j} \geq 0, \gamma_{jj'}^{(k)} = \gamma_{j'j}^{(k)}, 1 \leq j \neq j' \leq p.$
  - $\theta_{jj} = 1, c_{jj}^{(k)} = \gamma_{jj}^{(k)}.$
- $\min_{\mathbf{\Theta},(\mathbf{\Gamma}^{(k)})_{k=1^K}} \sum_{k=1}^{K} \left[\mathrm{tr}(\hat{\mathbf{\Sigma}}^{(k)}\mathbf{C}^{(k)} - \log\det(\mathbf{C}^{(k)})\right] + \eta_1 \sum_{j\neq j'} \theta_{jj'} + \eta_2 \sum_{j\neq j'} \sum_{k=1}^{K} |\gamma_{jj'}^{(k)}|$

On Time Varying Undirected Graphs

- $\mathbf{x}^i \sim (0, \mathbf{\Sigma}^{t_i}), i = 1, \ldots, n. t_i = i/n.$
- $G^{t_i} = (V, E^{t_i})$
- $\mathbf{C}^{t_i} = (\mathbf{\Sigma}^{t_i})^{-1}$
  - $\hat{\mathbf{C}}^{\tau} = \arg\min_{\mathbf{C}>0}\{\mathrm{tr}\mathbf{C}\hat{\mathbf{\Sigma}}^{\tau} - \log|\mathbf{C}| + \lambda\|\mathbf{C}^{-}\|_1\}$
  - $\hat{\mathbf{\Sigma}}^{\tau} = \sum_i w_i^{\tau}\mathbf{x}^i(\mathbf{x}^i)'$
  - $w_i^{\tau} = \frac{K_h(t_i-\tau)}{\sum_i K_h(t_i-\tau)}$

Time Varying Undirected Graphs: Penalized likelihood estimation

- $\hat{\mathbf{C}}^\tau = \arg\min_{\mathbf{C}>0}\{\mathrm{tr}\mathbf{C}\hat{\boldsymbol{\Sigma}}^\tau - \log|\mathbf{C}| + \lambda\|\mathbf{C}^-\|_1\}$
- $\hat{\boldsymbol{\Sigma}}^\tau = \sum_i w_i^\tau \mathbf{x}^i(\mathbf{x}^i)'$
- $w_i^\tau = \frac{K_h(t_i-\tau)}{\sum_i K_h(t_i-\tau)}$

Time Varying Undirected Graphs: Nearest neighbor selection

- The partial correlation coefficient $\rho^t_{ab} = -c^t_{ab}/\sqrt{c^t_{aa}c^t_{bb}}$
- $X_a = \sum_{b \in V \setminus \{a\}} X_b \theta^t_{ab} + \epsilon^t_a, a \in V$
- $\hat{\boldsymbol{\theta}}^{\tau}_{\setminus a} = \arg\min_{\boldsymbol{\theta} \in \mathbb{R}^{p-1}} \sum_i \left(x^i_a - \sum_{b \neq a} x^i_b \theta_b \right)^2 w^{\tau}_i + \lambda \|\boldsymbol{\theta}\|_1.$
- $\hat{\text{ne}}^{\tau}_a = \text{ne}(\hat{\boldsymbol{\theta}}^{\tau}_{\setminus a}).$
- $\hat{E}^{\tau,\wedge} = \{(a,b) : a \in \hat{\text{ne}}^{\tau}_b \wedge b \in \hat{\text{ne}}^{\tau}_a\},$
  $\hat{E}^{\tau,\vee} = \{(a,b) : a \in \hat{\text{ne}}^{\tau}_b \vee b \in \hat{\text{ne}}^{\tau}_a\}$