Graphic Lasso: two precision matrices

Jiaxin Hu

January 17, 2021

1 Consistency

Suppose K categories are clustered by two groups with precision matrices Θ_1, Θ_2. The model becomes

$$\Omega^k = \boldsymbol{I}_k \Theta_1 + (1 - \boldsymbol{I}_k)\Theta_2, \quad k = 1, ..., K,$$

where $\boldsymbol{I}_k = \boldsymbol{I}(k\text{-th category belongs to group 1})$ are indicator functions. The model is identifiable since the indicator functions can be replaced by a membership matrix. Consider the optimization problem

$$\begin{aligned} \min_{\Theta_1, \Theta_2, \boldsymbol{I}_k} \quad & \sum_{k=1}^{K} \operatorname{tr}(S^k \Omega^k) - \log |\Omega^k| \\ s.t. \quad & \Omega^k = \boldsymbol{I}_k \Theta_1 + (1 - \boldsymbol{I}_k)\Theta_2, \quad k = 1, ..., K, \\ & \|\Theta_i\|_0 \leq b, \quad i = 1, 2. \end{aligned}$$

Theorem 1.1. *Let $(\Theta_1, \Theta_2, \boldsymbol{I}_k)$ be the true precision matrices and the membership. Suppose $0 < \tau_1 < \phi_{min}(\Theta_i) \leq \phi_{max}(\Theta_0) < \tau_2 < \infty$, where $i = 1, 2$ and τ_1, τ_2 are positive constants. For the estimation $(\hat{\Theta}_1, \hat{\Theta}_2, \hat{I}_k)$ such that $\sum_{k=1}^{K} tr(S^k \hat{\Omega}^k) - \log |\hat{\Omega}^k| \leq \sum_{k=1}^{K} tr(S^k \Omega^k) - \log |\Omega^k|$, we have the following accuracy with probability tending to 1*

$$\sum_{k=1}^{K} \left\| \hat{\Omega}^k - \Omega^k \right\| \leq 2\sqrt{2K} C'' \left[C \sqrt{\frac{b \log p}{n}} + C' \sqrt{\frac{p \log p}{n}} \right]. \tag{1}$$

Proof. Let Σ^1, Σ^2 denote the true covariance matrices. Define the sets $A_{11} = \{k : \hat{I}_k = I_k = 1\}, A_{12} = \{k : \hat{I}_k = 1, I_k = 0\}, A_{21} = \{k : \hat{I}_k = 0, I_k = 1\}$ and $A_{22} = \{k : \hat{I}_k = I_k = 0\}$. Correspondingly, we define $\Delta_{11} = \hat{\Theta}_1 - \Theta_1$, $\Delta_{12} = \hat{\Theta}_1 - \Theta_2$, $\Delta_{21} = \hat{\Theta}_2 - \Theta_1$, and $\Delta_{22} = \hat{\Theta}_2 - \Theta_2$. Let $\Delta^k = \hat{\Omega}^k - \Omega \in \{\Delta_{11}, \Delta_{12}, \Delta_{21}, \Delta_{22}\}$. The value of Δ^k depends on the true and estimated membership of k. Consider the function

$$G(\Delta_{11}, \Delta_{12}, \Delta_{21}, \Delta_{22}) = I_1 + I_2,$$

where

$$\begin{aligned} I_1 &= \sum_{k \in A_{11}} \operatorname{tr}((S^k - \Sigma^1)\Delta_{11}) + \sum_{k \in A_{12}} \operatorname{tr}((S^k - \Sigma^2)\Delta_{12}) + \sum_{k \in A_{21}} \operatorname{tr}((S^k - \Sigma^1)\Delta_{21}) + \sum_{k \in A_{22}} \operatorname{tr}((S^k - \Sigma^2)\Delta_{22}) \\ &= I_{11} + I_{12} + I_{21} + I_{22}, \end{aligned}$$

and

$$I_2 = |A_{11}|f(\Delta_{11}, \Theta_1) + |A_{12}|f(\Delta_{12}, \Theta_2) + |A_{21}|f(\Delta_{21}, \Theta_1) + |A_{22}|f(\Delta_{22}, \Theta_2),$$

with $f(\Delta, \Theta) = (\tilde{\Delta})^T \int_0^1 (1-v)(\Theta + v\Delta)^{-1} \otimes (\Theta + v\Delta)^{-1} dv \tilde{\Delta}$.

Recall the result in the common precision matrix case. For each $I_{ij}, i, j = 1, 2$, we have

$$\frac{1}{|A_{ij}|}|I_{ij}| = \mathrm{tr}\left(\left(\frac{1}{|A_{ij}|}\sum_{k \in A_{ij}} S^k - \Sigma^j\right)\Delta_{ij}\right) \le C_{ij}\sqrt{\frac{\log p}{n|A_{ij}|}}|\Delta_{ij}^-|_1 + C'_{ij}\sqrt{\frac{p\log p}{n|A_{ij}|}}\left\|\Delta_{ij}\right\|_F.$$

Let $T_j = \{(k, l) : \Theta_{j,kl} \neq 0\}, j = 1, 2$. We have $|\Delta_{ij}^-|_1 = |\Delta_{T_j,ij}^-|_1 + |\Delta_{T_j^c,ij}^-|_1$. Note that $|\Delta_{T_j,ij}^-|_0, |\Delta_{T_j^c,ij}^-|_0 \le b$ and $|\Delta_{T_j,ij}^-|_1, |\Delta_{T_j^c,ij}^-|_1 \le \sqrt{b}\left\|\Delta_{ij}\right\|_F$. Then, we have

$$|I_{ij}| \le \sqrt{|A_{ij}|}\left[C_{ij}\sqrt{\frac{b\log p}{n}} + C'_{ij}\sqrt{\frac{p\log p}{n}}\right]\left\|\Delta_{ij}\right\|_F.$$

On the other hand, the lower bound for I_2 is

$$I_2 \ge \frac{1}{4\tau_2^2}\sum_{ij}|A_{ij}|\left\|\Delta_{ij}\right\|_F^2.$$

To let $G \le 0$, we have $I_2 \le |I_1| \le \sum_{ij}|I_{ij}|$. Plug the upper bound for $|I_{ij}|$ and the lower bound for I_2, we have

$$\frac{1}{4\tau_2^2}\sum_{ij}|A_{ij}|\left\|\Delta_{ij}\right\|_F^2 \le \left[C\sqrt{\frac{b\log p}{n}} + C'\sqrt{\frac{p\log p}{n}}\right]\sum_{ij}\sqrt{|A_{ij}|}\left\|\Delta_{ij}\right\|_F.$$

By Cauchy Schwartz inequality, we have

$$\sum_{ij}|A_{ij}|\left\|\Delta_{ij}\right\|_F^2 \ge \frac{1}{4}\left(\sum_{ij}\sqrt{|A_{ij}|}\left\|\Delta_{ij}\right\|_F\right)^2. \tag{2}$$

Thus, we have

$$\sum_{ij}\sqrt{|A_{ij}|}\left\|\Delta_{ij}\right\|_F \le 4C''\left[C\sqrt{\frac{b\log p}{n}} + C'\sqrt{\frac{p\log p}{n}}\right].$$

Multiply $\max\sqrt{|A_{ij}|}$ on both sides. We obtain the accuracy

$$\begin{aligned}\sum_{k=1}^K\left\|\hat{\Omega}^k - \Omega^k\right\|_F &= \sum_{ij}|A_{ij}|\left\|\Delta_{ij}\right\|_F \\ &\le \max\sqrt{|A_{ij}|}\sum_{ij}\sqrt{|A_{ij}|}\left\|\Delta_{ij}\right\|_F \\ &\le 4\max\sqrt{|A_{ij}|}C''\left[C\sqrt{\frac{b\log p}{n}} + C'\sqrt{\frac{p\log p}{n}}\right]. \end{aligned} \tag{3}$$

$\square$

Remark 1. In two group case with equal group size, we have $\max\sqrt{|A_{ij}|} \le \sqrt{\frac{K}{2}}$. Then, we obtain the accuracy (1) in Theorem 1.1. If we have r groups and each group has equal number of categories, the number 4 should be replaced by $r(r-1)$ and $\max\sqrt{|A_{ij}|} \le \sqrt{\frac{K}{r}}$. Thus the accuracy is of order $\mathcal{O}(\sqrt{K}r^{3/2})$.

2 Discussion

What's the difference between known $\boldsymbol{I}_k$ and unknown $\boldsymbol{I}_k$?

In two group case with equal sample size, if the membership $\boldsymbol{I}_k$ is already known, we only need to consider A_{11} and A_{22} because $A_{12} = A_{21} = \emptyset$. Then, by Cauchy Schwartz, the inequality (2) becomes

$$\sum_{ij} |A_{ij}| \, \|\Delta_{ij}\|_F^2 \ge \frac{1}{2}\left(\sum_{ij}\sqrt{|A_{ij}|}\, \|\Delta_{ij}\|_F\right)^2,$$

$\max\sqrt{|A_{ij}|} = \sqrt{\frac{K}{2}}$, and thus

$$\sum_{k=1}^{K}\left\|\hat{\Omega}^k - \Omega^k\right\|_F \le \sqrt{2K}C''\left[C\sqrt{\frac{b\log p}{n}} + C'\sqrt{\frac{p\log p}{n}}\right].$$

Further, consider the r group case with equal sample size. If the membership is known, the number 4 in accuracy (3) becomes r and $\max\sqrt{|A_{ij}|} = \sqrt{\frac{K}{r}}$. The accuracy in this case is of order $\mathcal{O}(\sqrt{Kr})$, which is better than the unknown membership case with accuracy of order $\mathcal{O}(\sqrt{K}r^{3/2})$.

How the rate affected by the clustering result?

Let $a = |\{(i,j) : A_{ij} \ne \emptyset\}|$, which is the number of non-empty A_{ij}. The inequality (2) becomes

$$\sum_{ij} |A_{ij}| \, \|\Delta_{ij}\|_F^2 \ge \frac{1}{a}\left(\sum_{ij}\sqrt{|A_{ij}|}\, \|\Delta_{ij}\|_F\right)^2.$$

Thus, the accuracy will be of order $\mathcal{O}(a\max\sqrt{|A_{ij}|}) = \mathcal{O}(aK^{1/2}r^{-1/2})$. The number $a \in [r, r(r-1)]$, and the small a implies a more accurate clustering. Therefore, the accuracy of clustering will affect the estimation accuracy through the factor $a(r)$.

On the other hand, the accuracy of clustering will also affect estimation accuracy through factor $\max\sqrt{|A_{ij}|}$. Consider r group with equal group size. The worst estimation leads to $|A_{i1}| = \cdots = |A_{ir}| = \frac{K}{r^2}$, for all $i = 1, ..., r$. However, the accuracy under the worst case is of $\mathcal{O}(r\sqrt{K})$, which is not the worst accuracy $\mathcal{O}(\sqrt{K}r^{3/2})$. Need to figure out this problem next.

What is the accuracy of $\hat{\Theta}_1, \hat{\Theta}_2$ and $\hat{\boldsymbol{I}}_k$?

Note that

$$\begin{aligned}
\|\Delta_{11}\|_F &= \left\|\hat{\Theta}_1 - \Theta_1\right\|_F, \quad \|\Delta_{21}\|_F \le \left\|\hat{\Theta}_1 - \Theta_1\right\|_F + \left\|\hat{\Theta}_1 - \hat{\Theta}_2\right\|_F \\
\|\Delta_{22}\|_F &= \left\|\hat{\Theta}_2 - \Theta_2\right\|_F, \quad \|\Delta_{12}\|_F \le \left\|\hat{\Theta}_2 - \Theta_2\right\|_F + \left\|\hat{\Theta}_1 - \hat{\Theta}_2\right\|_F
\end{aligned}$$