Graphic Lasso: Misclassification Error

Jiaxin Hu

March 1, 2021

1 Misclassification error

The precision model is stated as

$$\mathbb{E}[S^k] = \Omega^k = \sum_{l=1}^{r} u_{kl}\Theta^l, \quad k \in [K].$$

Consider the following penalized optimization problem

$$\max_{\boldsymbol{U},\Theta^l} \mathcal{L}_S(\boldsymbol{U}, \Theta^l) = -\sum_{k=1}^{K} \operatorname{tr}(S^k\Omega^k) + \log\det(\Omega^k) + \lambda \left\|\Omega^k\right\|,$$

where $\boldsymbol{U}$ is a membership matrix, and $\{\Theta^l\}$ are irreducible and invertible.

Notations.

1. $I'_l = \{k : u'_{kl} \neq 0\}$ is the index set for the l-th group based on the membership $\boldsymbol{U}'$.
2. δ be the minimal gap between $Theta^l$. That is
$$\min_{k,l\in[r]} \left\|\Theta^l - \Theta^k\right\|_F^2 = \delta^2.$$
3. Let $l(\boldsymbol{U}, \Theta^l)$ be the population-based loss function. That is
$$l(\boldsymbol{U}, \Theta^l) = \mathbb{E}_S[\mathcal{L}_S(\boldsymbol{U}, \Theta^l)] = -\sum_{k=1}^{K} \operatorname{tr}(\Sigma^k\Omega^k) + \log\det(\Omega^k) + \lambda \sum_{k=1}^{K} \left\|\Omega^k\right\|_1.$$
4. Given the membership $\boldsymbol{U}'$, let $\hat{\Theta}^l(\boldsymbol{U}') = \arg\max_{\Theta^l} \mathcal{L}_S(\boldsymbol{U}', \Theta)$. Particularly, for each $l \in [r]$, we have
$$\hat{\Theta}^l(\boldsymbol{U}') = \arg\max_{\Theta} -\sum_{k\in I'_l} \langle S^k, \Theta\rangle + |I'_l| \log\det(\Theta) + \lambda |I'_l| \left\|\Theta\right\|_1,$$
5. Given the membership $\boldsymbol{U}'$, let $\tilde{\Theta}^l(\boldsymbol{U}') = \arg\max_{\boldsymbol{U}',\Theta^l}$. Particularly, for each $l \in [r]$, we have
$$\tilde{\Theta}^l(\boldsymbol{U}') = \arg\max_{\Theta} -\sum_{k\in I'_l} \langle \Sigma^k, \Theta\rangle + |I'_l| \log\det(\Theta) + \lambda |I'_l| \left\|\Theta\right\|_1. \tag{1}$$
6. Define functions
$$F(\boldsymbol{U}') = \mathcal{L}_S(\boldsymbol{U}', \hat{\Theta}^l(\boldsymbol{U}')), \quad G(\boldsymbol{U}') = l(\boldsymbol{U}', \tilde{\Theta}^l(\boldsymbol{U}')).$$

Lemma 1 (Self-consistency of $\boldsymbol{U}$)**.** *Suppose $MCR(\boldsymbol{U}', \boldsymbol{U}) \geq \epsilon$ and the minimal gap between $\{\Theta^l\}$ denoted δ is positive. For $\lambda < \frac{\delta}{8\tau^2\sqrt{p}}$, we have the perturbation version of the self-consistency.*

$$G(\boldsymbol{U}') - G(\boldsymbol{U}) \leq \epsilon\delta\left(-\frac{1}{8\tau^2}\delta + \lambda\sqrt{p}\right) < 0.$$

Consider the simplest case
lambda = 0 in the following proof.

Proof. First, we write the explicit form of $G(\boldsymbol{U}')$ and $G(\boldsymbol{U})$.

$$\begin{aligned} G(\boldsymbol{U}') &= l(\boldsymbol{U}', \tilde{\Theta}^l(\boldsymbol{U}')) \\ &= \sum_{l=1}^{r}\left[\sum_{a=1}^{r} D_{al}\left(-\langle \Sigma^a, \tilde{\Theta}^l(\boldsymbol{U}')\rangle + \log\det(\tilde{\Theta}^l(\boldsymbol{U}')) - \lambda\left\|\tilde{\Theta}^l(\boldsymbol{U}')\right\|_1\right)\right], \end{aligned}$$

and

$$\begin{aligned} G(\boldsymbol{U}) &= l(\boldsymbol{U}, \tilde{\Theta}^l(\boldsymbol{U})) \\ &= \sum_{l=1}^{r}\left[\sum_{a=1}^{r} D_{al}\left(-\langle \Sigma^a, \tilde{\Theta}^a(\boldsymbol{U})\rangle + \log\det(\tilde{\Theta}^a(\boldsymbol{U})) - \lambda\left\|\tilde{\Theta}^a(\boldsymbol{U})\right\|\right)\right]. \end{aligned}$$

We have in total n (not r) covariance matrices.
Sigma^k here is the averaged covariance
matrices based on ground truth U.

When U = Utrue:
Sigma^k = Ave{Sigma_i: i belongs to
group k}, where group k depends on U

Define the function

$$h^k(\Theta) = -\langle \Sigma^k, \Theta\rangle + \log\det(\Theta) - \lambda\left\|\Theta\right\|_1.$$

By the definition of (1), we have

$$\tilde{\Theta}^k(\boldsymbol{U}) = \arg\max_{\Theta} h^k(\Theta), k = 1, \ldots, r.$$

Where is the mistake?
1. The form of Sigma^k depends on U.
2. These two h^a are different.

Then, we have

There should be two versions of h^k.
U maximizes h^k. U' maximizes (h')^k.

$$G(\boldsymbol{U}') - G(\boldsymbol{U}) = \sum_{l=1}^{r}\left[\sum_{a=1}^{r} D_{al}\left(h^a(\tilde{\Theta}^l(\boldsymbol{U}')) - h^a(\tilde{\Theta}^a(\boldsymbol{U}))\right)\right] \leq 0,$$

where the equation holds only when $\boldsymbol{U}' = \boldsymbol{U}$ since $h^k(\Theta)$ is concave function and has unique maximizer. Thus we have the point-wise self-consistency of $\boldsymbol{U}$. Next, we develop the perturbation version self-consistency of $\boldsymbol{U}$.

Suppose $MCR(\hat{\boldsymbol{U}}, \boldsymbol{U}) \geq \epsilon$. There exist $l, k \neq k' \in [r]$ such that $\min\{D_{kl}, D_{k'l}\} \geq \epsilon$.
Then, we have

holds when h is defined based on U'

mistake: these two h's are different

$$\begin{aligned} G(\hat{\boldsymbol{U}}) - G(\boldsymbol{U}) &\leq D_{kl}\left(h^k(\tilde{\Theta}^l(\hat{\boldsymbol{U}})) - h^k(\tilde{\Theta}^k(\boldsymbol{U}))\right) + D_{k'l}\left(h^{k'}(\tilde{\Theta}^l(\hat{\boldsymbol{U}})) - h^{k'}(\tilde{\Theta}^{k'}(\boldsymbol{U}))\right) \\ &\leq D_{kl}\left(h^k(\tilde{\Theta}^l(\hat{\boldsymbol{U}})) - h^k(\Theta^k)\right) + D_{k'l}\left(h^{k'}(\tilde{\Theta}^l(\hat{\boldsymbol{U}})) - h^{k'}(\Theta^{k'})\right), \end{aligned} \quad (2)$$

holds when h is defined based on U

where Θ^l are true precision matrices, and the second inequality follows the fact that $h^k(\Theta^k) \leq h^k(\tilde{\Theta}^k(\boldsymbol{U}))$. For simplicity, let $\tilde{\Theta}$ denote $\tilde{\Theta}^l(\boldsymbol{U}')$. Define $\Delta^k = \tilde{\Theta} - \Theta^k$. Consider the function

$$f^k(t) = \log\det(\Theta^k + t\Delta),$$

does not hold b.c. we have two versions
of h.

and by Taylor expansion we have

$$f^k(1) - f^{k'}(0) = \langle \Sigma^k, \Delta^k\rangle - \mathrm{vec}(\Delta^k)^T\int_0^1 (1-v)(\Theta^k + v\Delta^k)^{-1} \otimes (\Theta^k + v\Delta^k)^{-1} dv\mathrm{vec}(\Delta^k).$$

Then, we have

$$h^k(\Theta^k) - h^k(\tilde{\Theta}^k) = \langle \Sigma^k, \Delta^k \rangle - f^k(1) + f^k(0) - \lambda \left(\left\| \Theta^k \right\|_1 - \left\| \hat{\Theta} \right\|_1 \right)$$
$$\geq A_1 - |A_2|,$$

where

$$A_1 = \text{vec}(\Delta^k)^T \int_0^1 (1-v)(\Theta^k + v\Delta^k)^{-1} \otimes (\Theta^k + v\Delta^k)^{-1} dv \text{vec}(\Delta^k)$$
$$A_2 = \lambda \left(\left\| \Theta^k \right\|_1 - \left\| \tilde{\Theta} \right\|_1 \right).$$

By Guo's paper, we know that

$$A_1 \geq \frac{1}{4\tau^2} \left\| \Delta^k \right\|_F^2, \tag{3}$$

where $\max_{k \in [r]} \varphi_{\max}(\Theta^k) \leq \tau < \infty$. Also note that

$$|A_2| \leq \lambda \left\| \Theta^k - \tilde{\Theta} \right\|_1 \leq \lambda \sqrt{p} \left\| \Delta^k \right\|_F. \tag{4}$$

Plug the inequalities (3) and (4) in to the inequality (2), we obtain that

$$\begin{aligned} G(\boldsymbol{U}') - G(\boldsymbol{U}) &\leq D_{kl} \left(-\frac{1}{4\tau^2} \left\| \Delta^k \right\|_F^2 + \lambda\sqrt{p} \left\| \Delta^k \right\|_F \right) + D_{k'l} \left(-\frac{1}{4\tau^2} \left\| \Delta^{k'} \right\|_F^2 + \lambda\sqrt{p} \left\| \Delta^{k'} \right\|_F \right) \\ &\leq \epsilon \left\{ -\frac{1}{4\tau^2} \left[\left\| \Delta^k \right\|_F^2 + \left\| \Delta^{k'} \right\|_F^2 \right] + \lambda\sqrt{p} \left[\left\| \Delta^k \right\|_F + \left\| \Delta^{k'} \right\|_F \right] \right\} \\ &\leq \epsilon \left\{ -\frac{1}{8\tau^2} \left[\left\| \Delta^k \right\|_F + \left\| \Delta^{k'} \right\|_F \right]^2 + \lambda\sqrt{p} \left[\left\| \Delta^k \right\|_F + \left\| \Delta^{k'} \right\|_F \right] \right\}. \end{aligned}$$

where the second inequality follows by the definition of MCR, the third inequality follows by the Cauchy-Schwartz inequality that $(x+y)^2 \leq 2(x^2+y^2)$.

Consider the function $m(x) = -\frac{1}{8\tau^2}x^2 + \lambda\sqrt{p}x$. When $x > 8\tau^2\lambda\sqrt{p}$, $m(x)$ is a decreasing function with negative value. Hence, let λ satisfies the following constrain

$$8\tau^2\lambda\sqrt{p} < \left\| \Delta^k - \Delta^{k'} \right\|_F. \tag{5}$$

Then, we have

$$G(\boldsymbol{U}') - G(\boldsymbol{U}) \leq \epsilon \left\{ -\frac{1}{8\tau^2} \left[\left\| \Delta^k - \Delta^{k'} \right\|_F \right]^2 + \lambda\sqrt{p} \left[\left\| \Delta^k - \Delta^{k'} \right\|_F \right] \right\}, \tag{6}$$

by the triangle inequality that $\|A - B\|_F \leq \|A\|_F + \|B\|_F$. Note that

$$\left\| \Delta^k - \Delta^{k'} \right\|_F = \left\| \Theta^k - \Theta^{k'} \right\|_F \geq \delta,$$

where δ is the minimal gap between true precision matrices. Replacing $\left\| \Delta^k - \Delta^{k'} \right\|$ by δ in the constrain (5) and inequality (6), we obtain the perturbation version of self-consistency. That is

$$G(\boldsymbol{U}') - G(\boldsymbol{U}) \leq \epsilon\delta \left(-\frac{1}{8\tau^2}\delta + \lambda\sqrt{p} \right),$$

and the right hand side is negative when $\lambda < \frac{\delta}{8\tau^2\sqrt{p}}$. □