

# 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

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

where  $\mathbf{U}$  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  $\mathbf{U}'$ .
2.  $\delta$  be the minimal gap between  $\Theta^l$ . That is

$$\min_{k, l \in [r]} \|\Theta^l - \Theta^k\|_F^2 = \delta^2.$$

3. Let  $l(\mathbf{U}, \Theta^l)$  be the population-based loss function. That is

$$l(\mathbf{U}, \Theta^l) = \mathbb{E}_S[\mathcal{L}_S(\mathbf{U}, \Theta^l)] = - \sum_{k=1}^K \text{tr}(\Sigma^k \Omega^k) + \log \det(\Omega^k) + \lambda \sum_{k=1}^K \|\Omega^k\|_1.$$

4. Given the membership  $\mathbf{U}'$ , let  $\hat{\Theta}^l(\mathbf{U}') = \arg \max_{\Theta^l} \mathcal{L}_S(\mathbf{U}', \Theta^l)$ . Particularly, for each  $l \in [r]$ , we have

$$\hat{\Theta}^l(\mathbf{U}') = \arg \max_{\Theta} - \sum_{k \in I'_l} \langle S^k, \Theta \rangle + |I'_l| \log \det(\Theta) + \lambda |I'_l| \|\Theta\|_1,$$

5. Given the membership  $\mathbf{U}'$ , let  $\tilde{\Theta}^l(\mathbf{U}') = \arg \max_{\mathbf{U}', \Theta^l} \mathcal{L}_S(\mathbf{U}', \Theta^l)$ . Particularly, for each  $l \in [r]$ , we have

$$\tilde{\Theta}^l(\mathbf{U}') = \arg \max_{\Theta} - \sum_{k \in I'_l} \langle \Sigma^k, \Theta \rangle + |I'_l| \log \det(\Theta) + \lambda |I'_l| \|\Theta\|_1. \quad (1)$$

6. Define functions

$$F(\mathbf{U}') = \mathcal{L}_S(\mathbf{U}', \hat{\Theta}^l(\mathbf{U}')), \quad G(\mathbf{U}') = l(\mathbf{U}', \tilde{\Theta}^l(\mathbf{U}')).$$

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

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

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

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

and

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

Define the function

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

By the definition of (1), we have

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

Then, we have

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

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

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

Then, we have

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

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

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

and by Taylor expansion we have

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

Then, we have

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

where

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

By Guo's paper, we know that

$$A_1 \geq \frac{1}{4\tau^2} \|\Delta^k\|_F^2, \quad (3)$$

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

$$|A_2| \leq \lambda \|\Theta^k - \tilde{\Theta}^k\|_1 \leq \lambda \sqrt{p} \|\Delta^k\|_F. \quad (4)$$

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

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

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  $\lambda$  satisfies the following constrain

$$8\tau^2\lambda\sqrt{p} < \|\Delta^k - \Delta^{k'}\|_F. \quad (5)$$

Then, we have

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

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

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

$\Delta^k - \Delta^{k'} = (\tilde{\Theta} - \Theta^k) - (\tilde{\Theta} - \Theta^{k'})$ .

$\tilde{\Theta}$  in different  $\Delta^k$  are the same.  $\tilde{\Theta}$  refers to the estimate of  $l$ -th group with membership  $\mathbf{U}'$ .

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

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

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