Misclassification rate

Jiaxin Hu

June 24, 2021

This is a correction of Note 0228. We consider the model and optimization

$$\begin{aligned} \min_{\Theta_r} \quad & \sum_{k=1}^{K} \langle S_k, \Omega_k \rangle - \log\det(\Omega_k) + \lambda \sum_{r=1}^{R} |I_r| \left\| \Theta_r \right\|_1 \\ s.t. \quad & \Omega_k = \sum_{r=1}^{R} u_{kr} \Theta_r, \end{aligned}$$

where $U = [\![u_{kr}]\!] \in \mathbb{R}^{K \times R}$ is a membership matrix with only one positive element in each row and others remain 0, and $I_r = \{k \in [K] : u_{kr} > 0\}$. Let $\Theta_r^*, U^*, u_{kr}^*$ denote the true parameters, and $\Sigma_r^* = (\Theta_r^*)^{-1}$ denote the "slopes" of the true covariance matrices. For sample covariance matrix S_k with $k \in I_r^*$, the corresponding precision matrix

$$\mathbb{E}[S_k] = \Sigma_k = (u_{kr}^* \Theta_r^*)^{-1} = 1/u_{kr}^* \Sigma_r^*.$$

Notations.

1. D_{ar}: the number of categories in true group a and estimated group r, i.e.,

$$D_{ar} = \sum_{k \in [K]} \mathbf{1}\{u_{ka}^*, u_{kr} \neq 0\}.$$

   Correspondingly, the MCR is defined as $MCR(U, U^*) = \max_{r \in [R], a \neq a' \in [R]} \min\{D_{ar}, D_{a'r}\}$ and $I_{ar} = \{k \in [K] : u_{ka}^*, u_{kr} \neq 0\}$.

2. $\ell(\Theta_r, U)$: the population-based loss function, where

$$\ell(\Theta_r, U) = \sum_{r=1}^{R} \sum_{k \in I_r} \langle \Sigma_k, u_{kr} \Theta_r \rangle - \log\det(u_{kr} \Theta_r) + \left\| \Theta_r \right\|_1.$$

3. $\tilde{\Theta}_r(U)$: optimizer of $\ell(\Theta, U)$ with given membership U, where

$$\tilde{\Theta}_r(U) = \arg\min_{\Theta} \sum_{k \in I_r} \langle \Sigma_k, u_{kr} \Theta_r \rangle - \log\det(u_{kr} \Theta_r) + \left\| \Theta_r \right\|_1. \tag{1}$$

   Note that $\tilde{\Theta}_r(U)$ is a function of the given membership U. For simplicity, we use $\tilde{\Theta}_r$ denote the $\tilde{\Theta}_r(U)$, and $\tilde{\Theta}_r^*$ denote the optimizer of $\ell(\Theta, U^*)$ with given true membership U^*.

4. $G(U)$: population-based objective function for the membership, where

$$G(U) = \ell(\tilde{\Theta}_r, U).$$

1 Without intercept case

Lemma 1. *Suppose $MCR(U,U^*) \geq \epsilon$ and the minimal gap between $\{\Theta_r\}$ is $\delta > 0$. For $\lambda \leq \min_{k\in[K],a,r\in[R]} C\frac{\left\|\Delta_{k,ar}\right\|_F}{\sqrt{p}\tau^2}$ with $\Delta_{k,ar} = \frac{u_{kr}}{u^*_{ka}}\tilde{\Theta}_r - \Theta^*_a$, we have*

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

where $e = \max_{k\in I_{ar}\cup I_{a'r}}\left\|\left(1-\frac{u_{kr}}{u^*_{ka}}\right)\tilde{\Theta}_r\right\|_F$ and δ is the minimal gap between different precision matrices. For $e \leq \frac{1}{4}\delta$, and

$$\lambda \leq \min\left\{\frac{1}{\sqrt{p}}\left[\frac{1}{8\tau^2}\delta - \frac{1}{2\tau^2}e\right], \min_{k\in[K],a,r\in[R]} C\frac{\left\|\Delta_{k,ar}\right\|_F}{\sqrt{p}\tau^2}\right\},$$

the right hand side of inequality (2) is negative.

Remark 1. The Lemma 1 implies that when the membership U leads to a good estimation that $\max_{k\in I_{ar},r\in[R]} 1-\frac{u_{kr}}{u^*_{ka}}$ is small enough with a small λ, the true membership U^* is the minimizer of the population based objective function.

Proof. By the definition of $\ell(\Theta,U)$, we have

$$\begin{aligned}
G(U) &= \sum_{r=1}^R\left[\sum_{a=1}^R\sum_{k\in I_{ar}}\langle\Sigma_k, u_{kr}\tilde{\Theta}_r\rangle - \log\det(\tilde{\Theta}_r) + \lambda\left\|\tilde{\Theta}_r\right\|_1 - \log(u_{kr})\right],\\
&= \sum_{r=1}^R\left[\sum_{a=1}^R\sum_{k\in I_{ar}}\langle\Sigma^*_a, \frac{u_{kr}}{u^*_{ka}}\tilde{\Theta}_r\rangle - \log\det\left(\frac{u_{kr}}{u^*_{ka}}\tilde{\Theta}_r\right) + \lambda\left\|\frac{u_{kr}}{u^*_{ka}}\tilde{\Theta}_r\right\|_1 - \log(u^*_{ka}) + \lambda\left(1-\frac{u_{kr}}{u^*_{ka}}\right)\left\|\tilde{\Theta}_r\right\|_1\right]
\end{aligned}$$

$$\begin{aligned}
G(U^*) &= \sum_{r=1}^R\left[\sum_{a=1}^R\sum_{k\in I_{ar}}\langle\Sigma_k, u^*_{ka}\tilde{\Theta}_a\rangle - \log\det(\tilde{\Theta}^*_a) + \lambda\left\|\tilde{\Theta}^*_a\right\|_1 - \log(u^*_{ka})\right]\\
&= \sum_{r=1}^R\left[\sum_{a=1}^R\sum_{k\in I_{ar}}\langle\Sigma^*_a, \tilde{\Theta}^*_a\rangle - \log\det(\tilde{\Theta}^*_a) + \lambda\left\|\tilde{\Theta}^*_a\right\|_1 - \log(u^*_{ka})\right]
\end{aligned}$$

Consider the function

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

By the definition (1), we know that

$$\tilde{\Theta}^*_a = \arg\min_{\Theta} h_a(\Theta),$$

which follows by the fact that u^*_{kr} are given and $f(x), f(x)+c$ share the same optimizer for any function f and constant c. Then, we have

$$\begin{aligned}
G(U^*) - G(U) &= \sum_{r=1}^R\left[\sum_{a=1}^R\sum_{k\in I_{ar}}\left(h_a(\tilde{\Theta}^*_a) - h_a\left(\frac{u_{kr}}{u^*_{ka}}\tilde{\Theta}_r\right)\right) - \lambda\left(1-\frac{u_{kr}}{u^*_{ka}}\right)\left\|\tilde{\Theta}_r\right\|_1\right]\\
&\leq \sum_{r=1}^R\left[\sum_{a=1}^R\sum_{k\in I_{ar}}\left(h_a(\Theta^*_a) - h_a\left(\frac{u_{kr}}{u^*_{ka}}\tilde{\Theta}_r\right)\right) - \lambda\left(1-\frac{u_{kr}}{u^*_{ka}}\right)\left\|\tilde{\Theta}_r\right\|_1\right],
\end{aligned}$$

where the inequality follows by the fact that $h_a(\Theta_a^*) \geq h_a(\tilde{\Theta}_a^*)$ since Θ_a^* is the minimizer. By Taylor Expansion, we have

$$-\log\det(\Theta_a^* + \Delta) + \log\det(\Theta_a^*) \geq \frac{1}{4\tau^2} \|\Delta\|_F^2 - \langle \Sigma_a^*, \Delta \rangle,$$

where τ^2 is the largest singular value of Θ_a^*. Then, let $\Delta_{k,ar} = \frac{u_{kr}}{u_{ka}^*}\tilde{\Theta}_r - \Theta_a^*$. We have

$$\begin{aligned}
h_a(\Theta_a^*) - h_a\left(\frac{u_{kr}}{u_{ka}^*}\tilde{\Theta}_r\right) - \lambda\left(1 - \frac{u_{kr}}{u_{ka}^*}\right)\left\|\tilde{\Theta}_r\right\|_1 &\leq -\frac{1}{4\tau^2}\|\Delta_{k,ar}\|_F^2 + \lambda\left(\|\Theta_a^*\|_1 - \left\|\tilde{\Theta}_r\right\|_1\right) \\
&\leq -\frac{1}{4\tau^2}\|\Delta_{k,ar}\|_F^2 + \lambda\sqrt{p}\,\|\Delta_{k,ar}\|_F.
\end{aligned}$$

Let

$$\lambda \leq \min_{k\in[K],a,r\in[R]} C\frac{\|\Delta_{k,ar}\|_F}{\sqrt{p}\tau^2}, \tag{3}$$

for some constant C small enough. Then for all combination of k, a, r, we have

$$h_a(\Theta_a^*) - h_a\left(\frac{u_{kr}}{u_{ka}^*}\tilde{\Theta}_r\right) - \lambda\left(1 - \frac{u_{kr}}{u_{ka}^*}\right)\left\|\tilde{\Theta}_r\right\|_1 \leq 0,$$

and thus

$$G(U^*) - G(U) \leq 0.$$

Next, we derive a negative upper bound of $G(U^*) - G(U)$ related to the MCR.

By the definition of MCR, there exists a group r and corresponding a, a' such that $D_{ar}, D_{a'r} \geq \epsilon$. In the following proof, we consider only specific difference in the objective function in combinations (a, r) and (a', r). For the other combinations, the difference in objective is negative given λ in (3). Then, we rewrite the difference

$$G(U^*) - G(U) \leq \sum_{k\in I_{ar}}\left[-\frac{1}{4\tau^2}\|\Delta_{k,ar}\|_F^2 + \lambda\sqrt{p}\,\|\Delta_{k,ar}\|_F\right] + \sum_{k\in I_{a'r}}\left[-\frac{1}{4\tau^2}\|\Delta_{k,a'r}\|_F^2 + \lambda\sqrt{p}\,\|\Delta_{k,a'r}\|_F\right].$$

Note that

$$\|\Delta_{k,ar}\|_F = \left\|\tilde{\Theta}_r - \Theta_a^* + \left(1 - \frac{u_{kr}}{u_{ka}^*}\right)\tilde{\Theta}_r\right\|_F \geq \|\Delta_a\|_F - \left\|\left(1 - \frac{u_{kr}}{u_{ka}^*}\right)\tilde{\Theta}_r\right\|_F,$$

where $\Delta_a = \tilde{\Theta}_r - \Theta_a^*$, and then

$$-\frac{1}{4\tau^2}\|\Delta_{k,ar}\|_F^2 + \lambda\sqrt{p}\,\|\Delta_{k,ar}\|_F \leq -\frac{1}{4\tau^2}\|\Delta_a\|_F^2 + \left[\frac{1}{2\tau^2}\left\|\left(1 - \frac{u_{kr}}{u_{ka}^*}\right)\tilde{\Theta}_r\right\|_F + \lambda\sqrt{p}\right]\|\Delta_a\|_F.$$

Thus, we have

$$\begin{aligned}
G(U^*) - G(U) &\leq \epsilon\left\{-\frac{1}{4\tau^2}\left[\|\Delta_a\|_F^2 + \|\Delta_{a'}\|_F^2\right] + \left(\frac{1}{2\tau^2}e + \lambda\sqrt{p}\right)\left[\|\Delta_a\|_F + \|\Delta_{a'}\|_F\right]\right\} \\
&\leq \epsilon\left\{-\frac{1}{8\tau^2}\left[\|\Delta_a\|_F + \|\Delta_{a'}\|_F\right]^2 + \left(\frac{1}{2\tau^2}e + \lambda\sqrt{p}\right)\left[\|\Delta_a\|_F + \|\Delta_{a'}\|_F\right]\right\}
\end{aligned}$$

where $e = \max_{k \in I_{ar} \cup I_{a'r}} \left\| \left(1 - \frac{u_{kr}}{u^*_{ka}}\right) \tilde{\Theta}_r \right\|_F$, and the last inequality follows by the fact that $(x+y)^2 \leq 2(x^2 + y^2)$. Note that

$$\|\Delta_a\|_F + \|\Delta_{a'}\|_F \geq \|\Delta_a - \Delta_{a'}\|_F = \|\Theta^*_a - \Theta^*_{a'}\|_F \geq \delta,$$

where δ is the minimal gap between the true precision matrices. Hence, we have

$$G(U^*) - G(U) \leq \epsilon\delta \left\{ -\frac{1}{8\tau^2}\delta + \left(\frac{1}{2\tau^2} e + \lambda\sqrt{p} \right) \right\}. \tag{4}$$

When $e \leq \frac{1}{4}\delta$ and

$$\lambda \leq \min \left\{ \frac{1}{\sqrt{p}} \left[\frac{1}{8\tau^2}\delta - \frac{1}{2\tau^2} e \right], \min_{k \in [K], a, r \in [R]} C \frac{\|\Delta_{k,ar}\|_F}{\sqrt{p}\tau^2} \right\},$$

the right hand side of inequality (4) is negative.

□

References