Misclassification Error for Intercept Case

Jiaxin Hu

July 1, 2021

Consider the optimization problem

$$
\begin{aligned}
\min_{U,\Theta_r} \quad & \mathcal{Q}(U,\Theta_r) = \sum_{k=1}^{K} \langle S_k, \Omega_k \rangle - \log\det(\Omega_k) + \lambda \left[K \|\Theta_0\|_1 + \sum_{r=1}^{R} |I_r| \, \|\Theta_r\|_1 \right] \qquad (1)\\
s.t. \quad & \Omega_k = \Theta_0 + \sum_{r=1}^{R} u_{kr}\Theta_r,\\
& \|u_{\cdot r}\|_F = 1, \quad \sum_{k=1}^{K} u_{kr} = 0, \quad \text{for all } r \in [R].
\end{aligned}
$$

Notations.

1. Let $\Sigma_k = (\Theta_0^* + u_{kr}^*\Theta_r^*)^{-1}$ be the true precision matrix for $k \in I_r^*$.
2. Let $q(U,\Theta_r)$ denote the population version of $\mathcal{Q}$, i.e.,
$$
q(U,\Theta_r) = \sum_{k=1}^{K} \langle \Sigma_k, \Omega_k \rangle - \log\det(\Omega_k) + \lambda \left[K \|\Theta_0\|_1 + \sum_{r=1}^{R} |I_r| \, \|\Theta_r\|_1 \right].
$$
3. Let U^*, Θ_r^*, I_r^* denote the true parameters and membership.
4. Let $I_r = \{k \in [K] : u_{kr} \neq 0\}$ collects the categories that belong to group r with given membership U, and $I_{ar} = \{k \in [K] : u_{kr}, u_{ka}^* \neq 0\}$ collects the categories that belong to group r and true group a with given membership U and the true membership U^*.
5. Let $0 < \tau_1 < \min_{r\in[R]} \varphi_{\min}(\Theta_r) \leq \max_{r\in[R]} \varphi_{\max}(\Theta_r) < \tau_2$.
6. Let $\Delta_0 = \Theta_0 - \Theta_0^*$, $\Delta_{k,ar} = \Delta_0 + u_{kr}\Theta_r - u_{ka}^*\Theta_a^*$, and $\Delta_{ar} = \Theta_r - \Theta_a^*$

Lemma 1. *Suppose the scalars $|u_{kr}| > m$ for all $k \in [K], r \in [R]$ and factor precision matrices satisfy $\max_{a,a'\in[R]} \cos(\Theta_a^*, \Theta_{a'}^*) < \delta < 1$. Also, suppose the singular values for factor precision matrices are bounded as $0 < \tau_1 < \min_{r\in[R]} \varphi_{\min}(\Theta_r) \leq \max_{r\in[R]} \varphi_{\max}(\Theta_r) < \tau_2$. Assume the λ satisfies the assumption in Lemma 2, note 0626, i.e.,*

$$
\Lambda_1 \max\left\{ \sqrt{\frac{\log p}{nK}}, \max_{a,r\in[R]} \sqrt{\frac{\log p}{n|I_{ar}|}} \right\} \leq \lambda \leq \Lambda_2 \min\left\{ \sqrt{\frac{\log p}{nK}}, \min_{a,r\in[R]} \sqrt{\frac{\log p}{n|I_{ar}|}} \right\},
$$

for some positive constants Λ_1, Λ_2. Consider the local minimizer to the problem (1), (U, Θ_r). If the $MCR(U, U^*) \geq \epsilon$, with high probability, we have

$$q(U^*, \Theta_r^*) - q(U, \Theta_r) \leq A_1 + A_2 \leq \epsilon \left[-\frac{m^2}{8\tau_2^2} F^2 + \lambda p F + 2\lambda p |1-m| \sqrt{p} \tau_2 \right].$$

, where

$$F^2 \geq 2m^2\tau_1^2 - \frac{2\delta\tau_2^2}{m^2}.$$

Remark 1. In Lemma 1, the assumption $|u_{kr}| > m$ can be considered as a condition for identifiability; $\max_{a,a' \in [R]} \cos(\Theta_a^*, \Theta_{a'}^*) < \delta < 1$ requires the angles between the factor precision matrices are far away from 0; the assumption for λ allows us to use the conclusions for local minimizer. The definition of MCR does not change, but the definition of minimal gap has been changed from $\left\| \Theta_a^* - \Theta_{a'}^* \right\|_F$ to the max constraint for the angles. Note that to have the lower bound for $F^2 \geq 2m^2\tau_1^2 - \frac{2\delta\tau_2^2}{m^2}$ strictly larger than 0, we may have some extra conditions on δ, m and the conditional number of the precision matrices. In general, as λ goes to 0, the misclassification error $q(U^*, \Theta_r^*) - q(U, \Theta_r)$ has a negative upper bound with proper choice of the parameters.

Proof. Consider the local minimizer to the problem (1), (U, Θ_r). By the definition, we have

$$q(U^*, \Theta_r^*) - q(U, \Theta_r) = A_1 + A_2,$$

where

$$A_1 = \sum_{r=1}^{R} \sum_{a=1}^{R} \sum_{k \in I_{ar}} -\langle \Sigma_k, \Delta_{k,ar} \rangle - \log\det(\Theta_0^* + u_{ka}^* \Theta_a^*) + \log\det(\Theta_0 + u_{kr}\Theta_r)$$

$$A_2 = \lambda \left[K \left(\|\Theta_0^*\|_1 - \|\Theta_0\|_1 \right) + \sum_{r=1}^{R} \sum_{a=1}^{R} |I_{ar}| \left(\|\Theta_a^*\|_1 - \|\Theta_r\|_1 \right) \right].$$

For the first term, by Taylor expansion, we have

$$\begin{aligned} A_1 &\leq -\frac{1}{4\tau_2^2} \sum_{r=1}^{R} \sum_{a=1}^{R} \sum_{k \in I_{ar}} \|\Delta_{k,ar}\|_F^2 \\ &= -\frac{1}{4\tau^2} \sum_{r=1}^{R} \sum_{a=1}^{R} \sum_{k \in I_{ar}} \left[\|\Delta_0\|_F^2 + \|u_{kr}^* \Delta_{ar} + (u_{ka}^* - u_{kr})\Theta_a^*\|_F^2 \right]. \end{aligned}$$

For the second term, by triangle inequality, we have

$$\begin{aligned} A_2 &\leq \lambda \left[K \|\Delta_0\|_1 + \sum_{r=1}^{R} \sum_{a=1}^{R} |I_{ar}| \|\Delta_{ar}\|_1 \right] \\ &\leq \lambda p \left[K \|\Delta_0\|_F + \sum_{r=1}^{R} \sum_{a=1}^{R} |I_{ar}| \|\Delta_{ar}\|_F \right], \end{aligned}$$

where the second inequality follows by the fact that $\|\Delta\|_1 \le p\|\Delta\|_F, \Delta \in \mathbb{R}^{p\times p}$.

Since (U, Θ_r) is the local minimizer, by the Lemma 2 in Note 0626, suppose

$$\Lambda_1 \max\left\{\sqrt{\frac{\log p}{nK}}, \max_{a,r\in[R]}\sqrt{\frac{\log p}{n|I_{ar}|}}\right\} \le \lambda \le \Lambda_2 \min\left\{\sqrt{\frac{\log p}{nK}}, \min_{a,r\in[R]}\sqrt{\frac{\log p}{n|I_{ar}|}}\right\},$$

we have following inequalities with high probability

$$\|\Delta_0\|_F \le M_0\sqrt{\frac{s_0\log p}{nK}}, \quad \|\Delta_{ar}\|_F \le M_{ar}\sqrt{\frac{s_a \log p}{n|I_{ar}|}}, \quad |u_{kr} - u^*_{ka}| \le M_k\sqrt{\frac{p^2\log p}{n}},$$

for $k \in I_{ar}, a, r \in [R]$ and some large positive constants M_0, M_{ar}, M_k. Therefore, with proper choice of Λ_1, Λ_2, every term in $A_1 + A_2$ is non-positive and $A_1 + A_2 \le 0$ with high probability.

Next, we find a negative upper bound for $A_1 + A_2$. By the definition of MCR and the assumption that $MCR(U, U^*) \ge \epsilon$, there exists a r and a, a' such that $\min |I_{ar}|, |I_{a'r}| \ge \epsilon$. In the following proof, we focus on the estimated group r and true groups a, a'. Also note that

$$-\frac{1}{4\tau_2^2}\|\Delta_0\|_F^2 + \lambda p\|\Delta_0\|_F \le 0,$$

with proper λ. Hence, we focus on the term $\|u^*_{kr}\Delta_{ar} + (u^*_{ka} - u_{kr})\Theta^*_a\|_F^2$. Notice that

$$\|u^*_{kr}\Delta_{ar} + (u^*_{ka} - u_{kr})\Theta^*_a\|_F = \|u_{kr}\Theta_r - u^*_{ka}\Theta^*_a\|_F = |u_{kr}|\left\|\Theta_r - \frac{u^*_{ka}}{u_{kr}}\Theta^*_a\right\|_F \ge m\left\|\Theta_r - \frac{u^*_{ka}}{u_{kr}}\Theta^*_a\right\|_F,$$

where the last inequality follows by the assumption that $|u_{kr}| \ge m$, for $k \in [K], r \in [R]$. Then, we have

$$\begin{aligned} A_1 + A_2 \le &\sum_{k\in I_{ar}} -\frac{m^2}{4\tau_2^2}\left\|\Theta_r - \frac{u^*_{ka}}{u_{kr}}\Theta^*_a\right\|_F^2 + \lambda p|I_{ar}|\|\Delta_{ar}\|_F \\ &+ \sum_{k'\in I_{a'r}} -\frac{m^2}{4\tau_2^2}\left\|\Theta_r - \frac{u^*_{k'a'}}{u_{k'r}}\Theta^*_{a'}\right\|_F^2 + \lambda p|I_{a'r}|\|\Delta_{a'r}\|_F. \end{aligned}$$

For the square terms, note that

$$\begin{aligned} \left\|\Theta_r - \frac{u^*_{ka}}{u_{kr}}\Theta^*_a\right\|_F^2 + \left\|\Theta_r - \frac{u^*_{k'a'}}{u_{k'r}}\Theta^*_{a'}\right\|_F^2 &\ge \frac{1}{2}\left[\left\|\Theta_r - \frac{u^*_{ka}}{u_{kr}}\Theta^*_a\right\|_F + \left\|\Theta_r - \frac{u^*_{k'a'}}{u_{k'r}}\Theta^*_{a'}\right\|_F\right]^2 \\ &\ge \frac{1}{2}\left\|\frac{u^*_{ka}}{u_{kr}}\Theta^*_a - \frac{u^*_{k'a'}}{u_{k'r}}\Theta^*_{a'}\right\|_F^2 \\ &\ge \frac{1}{2}\left[\left(\frac{u^*_{ka}}{u_{kr}}\right)^2\|\Theta^*_a\|_F^2 + \left(\frac{u^*_{k'a'}}{u_{k'r}}\right)^2\|\Theta^*_{a'}\|_F^2 - 2|\frac{u^*_{ka}u^*_{k'a'}}{u_{kr}u_{k'r}}||\langle\Theta^*_a, \Theta^*_{a'}\rangle|\right]. \end{aligned} \tag{2}$$

Notice that for any $k \in I_{ar}, k' \in I_{a'r}$, we have $\|\Theta^*_a\|_F, \|\Theta^*_{a'}\|_F \ge \tau_1$,

$$\left(\frac{u^*_{ka}}{u_{kr}}\right)^2 \ge m^2, \quad \left(\frac{u^*_{k'a'}}{u_{k'r}}\right)^2 \ge m^2, \quad |\frac{u^*_{ka}u^*_{k'a'}}{u_{kr}u_{k'r}}| \le m^{-2},$$

and

$$|\langle \Theta_a^*, \Theta_{a'}^* \rangle| \leq \|\Theta_a^*\|_2 \, \|\Theta_{a'}^*\|_2 \cos(\Theta_a^*, \Theta_{a'}^*) \leq \delta \, \|\Theta_a^*\|_2 \, \|\Theta_{a'}^*\|_2 \leq \tau_2^2 \delta,$$

where the last inequality follows by the assumption that the singular values of Θ_a for $a \in [R]$ are upper bounded by τ_2 and lower bounded by τ_1, and the angle between factor precision matrices is $\max_{a,a' \in [R]} \cos(\Theta_a^*, \Theta_{a'}^*) < \delta$. For simplicity, let $F = \left[\left\| \Theta_r - \frac{u_{ka}^*}{u_{kr}} \Theta_a^* \right\|_F + \left\| \Theta_r - \frac{u_{k'a'}^*}{u_{k'r}} \Theta_{a'}^* \right\|_F \right]$, and the by inequality (2), we have

$$F^2 \geq 2m^2 \tau_1^2 - \frac{2\delta \tau_2^2}{m^2}.$$

Also, note that

$$\|\Delta_{ar}\|_F \leq \left\| \Theta_r - \frac{u_{ka}^*}{u_{kr}} \Theta_a^* \right\|_F + |1 - \frac{u_{ka}^*}{u_{kr}}| \, \|\Theta_a^*\|_F \leq \left\| \Theta_r - \frac{u_{ka}^*}{u_{kr}} \Theta_a^* \right\|_F + |1 - m| \sqrt{p} \tau_2.$$

Then, we have

$$q(U^*, \Theta_r^*) - q(U, \Theta_r) \leq A_1 + A_2 \leq \epsilon \left[-\frac{m^2}{8\tau_2^2} F^2 + \lambda p F + 2\lambda p |1 - m| \sqrt{p} \tau_2 \right].$$

□

References