Precision clustering

Jiaxin Hu

July 11, 2021

1 Preliminary

1.1 Model and Notations

Suppose we have K categories in R groups. Let $z(k) \in [R]^K$ denote the group assignment, and $X_{z(k)} \sim \mathcal{N}_p(0, \Sigma_{z(k)})$, where

$$\Sigma_{z(k)}^{-1} = \Omega_{z(k)} = \Theta_0 + u_k \Theta_{z(k)},$$

where Σ_r, Ω_r are the true covariance and precision matrices, respectively, Θ_0 is denoted as intercept matrix and Θ_r for $r \in [R]$ are denoted as factor matrices.

1. Let $z^*, u^*, \Theta_0^*, \Theta_r^*, I_r^*, \Omega_k^*$ denote the true parameters.
2. Let S_k denote the sample covariance matrix for k-th category with n independent sample $X_{z(k),1}, ..., X_{z(k),n}$.
3. Let $I_r = \{k \in [K] : z(k) = r\}$.

1.2 Parameter Space

Suppose the true parameters $(u^*, \Theta_0^*, \Theta_r^*)$ belongs to the space $\mathcal{P}^*$ with given assignment z^*, where

$$\begin{aligned}
\mathcal{P}^*(z^*, \tau_1, \tau_2, \delta, \beta, m, M) = \Big\{ (u, \Theta_0, \Theta_r) : \quad & \Theta_0, \Theta_r \text{ are positive definite for all } r \in [R]; \\
& 0 < \tau_1 < \min_{r \in \{0\} \cup [R]} \varphi_{\min}(\Theta_r) \leq \max_{r \in \{0\} \cup [R]} \varphi_{\max}(\Theta_r) < \tau_2; \\
& \max_{r, r' \in [R]} \cos(\Theta_r, \Theta_{r'}) < \delta < 1; \\
& \frac{K}{\beta R} \leq |I_r| \leq \frac{K\beta}{R}, r \in [R]; \quad \min_{k \in [K]} |u_k| > m > 0; \\
& \sum_{k \in I_r} u_k^2 = M^2 K, \quad \sum_{k \in I_r} u_k = 0, \text{ for all } r \in [R] \Big\}.
\end{aligned}$$

Suppose we find the estimate in a larger space $\mathcal{P}$ with given assignment z, where

$$\mathcal{P}(z,\delta,\beta,m,M)=\Bigg\{(u,\Theta_0,\Theta_r):\quad \Theta_0,\Theta_r \text{ are positive definite for all } r\in[R];$$

$$\max_{r,r'\in[R]}\cos(\Theta_r,\Theta_{r'})<\delta<1;$$

$$\frac{K}{\beta R}\le|I_r|\le\frac{K\beta}{R}, r\in[R];\quad \min_{k\in[K]}|u_k|>m>0;$$

$$\sum_{k\in I_r}u_k^2=M^2K,\quad \sum_{k\in I_r}u_k=0,\ \text{for all } r\in[R]\Bigg\}.$$

1.3 Oracle Estimation error

Suppose we already know the true assignment z^*. Consider the MLE $(\hat{u}(z^*),\hat{\Theta}_0(z^*),\hat{\Theta}_r(z^*))$ with given z^*, where

$$(\hat{u}(z^*),\hat{\Theta}_0(z^*),\hat{\Theta}_r(z^*))=\operatorname*{arg\,min}_{(u,\Theta_0.\Theta_r)\in\mathcal{P}(z^*,\delta,\beta,m,M)}\mathcal{Q}(z^*,u,\Theta_0,\Theta_r),$$

and

$$\mathcal{Q}(z,u,\Theta_0,\Theta_r)=\sum_{k\in[K]}\mathcal{Q}_k(z(k),u_k,\Theta)=\sum_{k\in[K]}\langle S_k,\Theta_0+u_k\Theta_{z(k)}\rangle-\log\det(\Theta_0+u_k\Theta_{z(k)}).$$

The following Lemma shows the estimation error of the oracle estimator. Let

$$\Delta_0=\hat{\Theta}_0(z^*)-\Theta_0^*,\quad \Delta_r=\hat{\Theta}_r(z^*)-\Theta_r.$$

Lemma 1 (Oracle estimation error). *The optimizer $(\hat{u}(z^*),\hat{\Theta}_0(z^*),\hat{\Theta}_r(z^*))$ satisfies the following inequalities simultaneously with probability at least $1-\exp(-\sqrt{n})$,*

$$\|\Delta_0\|_F\le M_0\sqrt{\frac{p^2\log p}{\sqrt{n}K}},\quad \|\Delta_r\|_F\le M_r\sqrt{\frac{p^2\log p}{\sqrt{n}|I_r^*|}},\quad |\hat{u}_k(z^*)-u_k^*|\le M_k\sqrt{\frac{p^2\log p}{\sqrt{n}}},$$

for some large positive constants M_0,M_r,M_k.

Remark 1. Note that the estimation in Lemma 1 is not optimal, and this version is a compromise for the exponential rate. If we require a slower rate, e.g., $1-\mathcal{O}(1/n)$, then the $\sqrt{n}$ can be replaced by n in the denominator. Also, the p^2 can be reduced to sparsity s if we add more constraints.

Proof. Very similar to the proof in Note 0626. The set $\mathcal{A}$ is convex now since we know the true assignment. □

2 Misclassification rate

For simplicity, we define the function

$$\Omega_k(u,\Theta,z(k))=\Theta_0+u_k\Theta_{z(k)}.$$

Define the misclassification loss

$$
\begin{aligned}
\ell(z, z^*) &= \sum_{k\in[K]} \|\Omega_k(u^*, \Theta^*, z^*(k)) - \Omega_k(u^*, \Theta^*, z(k))\|_F^2 \\
&= \sum_{k\in[K]} \sum_{b\in[R]/z^*(k)} \|\Omega_k(u^*, \Theta^*, z^*(k)) - \Omega_k(u^*, \Theta^*, b)\|_F^2 \, \mathbf{1}\{z(k) = b\}
\end{aligned}
$$

Also define the minimal gap between different groups

$$
\begin{aligned}
\Delta^2_{\min}(p, m, \tau_1, \tau_2, \delta) &= \min_{k\in[K]} \min_{a\neq b\in[R]} \|\Omega_k(u^*, \Theta^*, a) - \Omega_k(u^*, \Theta^*, b)\|_F^2 \\
&\geq m^2 \min_{a\neq b\in[R]} \|\Theta^*_a - \Theta^*_b\|_F^2 \\
&\geq 2m^2 \left[p\tau_1^2 - \tau_2^2\delta\right],
\end{aligned}
$$

where the last inequality follow by the fact that

$$
\|A - B\|_F^2 \geq \|A\|_F^2 + \|B\|_F^2 - 2\langle A, B\rangle \geq p\left[\varphi^2_{\min}(A) + \varphi^2_{\min}(B)\right] - 2\|A\|_2 \|B\|_2 \cos(A, B),
$$

for $A, B \in \mathbb{R}^{p\times p}$. For simplicity, we use $\Delta^2_{\min}$ to denote the minimal gap. Consider the Hamming loss $h(z, z^*) = \sum_{k\in[K]} \mathbf{1}\{z(k) \neq z^*(k)\}$. Then, we have

$$
\ell(z, z^*) \geq \Delta^2_{\min} h(z, z^*).
$$

Our goal is to bound the misclassification error $\ell(z, z^*)$ for the MLE $(\hat{z}, \hat{u}(\hat{z}), \hat{\Theta}(\hat{z}))$. First, we mimic Theorem 3 in (Gao and Zhang, 2019) to decompose the loss, then mimic Lemma 4.1 in (Gao and Zhang, 2019) to find the oracle misclassification rate.

2.1 Error decomposition

Suppose $z^*(k) = a$. We need to analyze the following event to study the misclassification of MLE $\hat{z}$ which implies $\hat{z}(k) = b$.

$$
\mathcal{Q}_k(b, \hat{u}(\hat{z}), \hat{\Theta}(\hat{z})) \leq \mathcal{Q}_k(a, \hat{u}(\hat{z}), \hat{\Theta}(\hat{z})). \tag{1}
$$

Let

$$
\begin{aligned}
\Delta(b, a) &= \Omega_k(\hat{u}(\hat{z}), \hat{\Theta}(\hat{z}), b) - \Omega_k(u^*, \Theta^*, a) \\
\Delta^*(b, a) &= \Omega_k(\hat{u}(z^*), \hat{\Theta}(z^*), b) - \Omega_k(u^*, \Theta^*, a) \\
\hat{\Delta}(b, a) &= \Omega_k(\hat{u}(\hat{z}), \hat{\Theta}(\hat{z}), b) - \Omega_k(\hat{u}(z^*), \hat{\Theta}(z^*), a),
\end{aligned}
$$

and $\tilde{\Delta} = \mathrm{vec}(\Delta)$. Note that by Taylor Expansion we have

$$
\mathcal{Q}_k(b, \hat{u}(z), \hat{\Theta}(z)) - \mathcal{Q}_k(a, u^*, \Theta^*) = \langle S_k - \Sigma_k, \Delta(b, a)\rangle + T_2(b, a), \tag{2}
$$

where

$$
\begin{aligned}
T_2(b, a) &= (\tilde{\Delta}(b, a))^T \int_0^1 (1 - v)(\Omega_k^* + \Delta(b, a))^{-1} \otimes (\Omega_k^* + \Delta(b, a))^{-1} dv \tilde{\Delta}(b, a) \\
&= c\,\|\Delta(b, a)\|_F^2,
\end{aligned}
$$

with a constant c related to the τ_1, τ_2 and the second equation follows by the fact that $\Delta(b,a)$ has bounded Frobenius norm due the property of MLE. (Trivial estimation error for the MLE implies that $\Delta(b,a) \to 0$ as $n \to \infty$. Will add the corresponding Lemma later.)

Plugging the Taylor Expansion (2) into the event (1), the event is upper bounded by the event

$$\langle S_k - \Sigma_k, \Delta(b,a) - \Delta(a,a)\rangle \leq C\left[\|\Delta(a,a)\|_F^2 - \|\Delta(b,a)\|_F^2\right].$$

Rearranging the inequality, we have

$$\langle S_k - \Sigma_k, \Omega_k(\hat{u}(z^*), \hat{\Theta}(z^*), b) - \Omega_k(\hat{u}(z^*), \hat{\Theta}(z^*), a)\rangle \leq -C\bar{\Delta}_k(a,b)^2 + CG_k(a,b,\hat{z}) + CH_k(a,b,\hat{z}) + F_k(a,b,\hat{z})$$

where

$$\begin{aligned}
\bar{\Delta}_k(a,b)^2 &= \|\Omega_k(u^*, \Theta^*, a) - \Omega_k(u^*, \Theta^*, b)\|_F^2 \geq \Delta_{\min}^2. \\
F_k(a,b,\hat{z}) &= \langle S_k - \Sigma_k, \hat{\Delta}(a,a) - \hat{\Delta}(b,b)\rangle. \\
G_k(a,b,\hat{z}) &= \left(\|\Delta(a,a)\|_F^2 - \|\Delta^*(a,a)\|_F^2\right) - \left(\|\Delta(b,a)\|_F^2 - \|\Delta^*(b,a)\|_F^2\right). \\
H_k(a,b) &= \|\Delta^*(a,a)\|_F^2 - \left(\|\Delta^*(b,a)\|_F^2 - \|\Omega_k(u^*, \Theta^*, a) - \Omega_k(u^*, \Theta^*, b)\|_F^2\right).
\end{aligned}$$

Note that F_k, G_k can be controlled the difference between $(\hat{u}(\hat{z}), \Theta(\hat{z}))$ between $(\hat{u}(z^*), \Theta(z^*))$, which further depends on $\ell(z, z^*)$, and H_k can be controlled by the difference between $(\hat{u}(z^*), \Theta(z^*))$ and (u^*, Θ^*), which can be bounded by the conclusion of oracle estimation error. Therefore, we impose three reasonable conditions:

Condition 2.1. *Assume that*

$$\max_{\{z:\ell(z,z^*)\leq\tau\}} \sum_{k\in[K]} \max_{b\in[K]/z^*(k)} \frac{F_k(z^*(k), b, z)^2 \|\Omega_k(u^*, \Theta^*, z^*(k)) - \Omega_k(u^*, \Theta^*, b)\|_F^2}{\bar{\Delta}_k(z^*(k), b)^4 \ell(z, z^*)} \leq C_1\delta^2,$$

holds with probability at least $1 - \eta_1$ *for some* $\tau, \delta, \eta_1 > 0$.

Condition 2.2. *Assume that*

$$\max_{\{z:\ell(z,z^*)\leq\tau\}} \max_{T\subset[K]} \frac{\tau}{4\Delta_{\min}^2|T| + \tau} sum_{k\in[K]} \max_{b\in[K]/z^*(k)} \frac{G_k(z^*(k), b, z)^2 \|\Omega_k(u^*, \Theta^*, z^*(k)) - \Omega_k(u^*, \Theta^*, b)\|_F^2}{\bar{\Delta}_k(z^*(k), b)^4 \ell(z, z^*)} \leq C_2\delta^2,$$

holds with probability at least $1 - \eta_2$ *for some* $\tau, \delta, \eta_2 > 0$.

Condition 2.3. *Assume that*

$$\max_{k\in[K]} \max_{b\in[K]/z^*(k)} \frac{|H_k(z^*(k), b)|}{\bar{\Delta}_k(z^*(k), b)^2} \leq C_3\delta,$$

holds with probability at least $1 - \eta_3$ *for some* $\tau, \delta, \eta_3 > 0$.

Lemma 2 (Condition check)**.** *The MLE* $\hat{z}$ *satisfies the Conditions 2.1, 2.2, 2.3.*

Proof Sketch for Lemma 2. For Conditions 2.1, 2.2, we need to consider the term

$$\left\|\Omega_k(\hat{u}(\hat{z}), \hat{\Theta}(\hat{z}), a) - \Omega_k(\hat{u}(z^*), \hat{\Theta}(z^*), a)\right\|_F^2.$$

To bound this term with the misclassification loss, we consider the following facts

$$(\hat{u}(\hat{z}), \Theta(\hat{z})_a, \Theta(\hat{z})_a) = \underset{(u,\Theta_0,\Theta_a)}{\arg\min} \sum_{k\in\hat{I}_a} \mathcal{Q}_k(a, \hat{u}(\hat{z}), \hat{\Theta}(\hat{z})),$$

and

$$\begin{aligned} 0 &\geq \sum_{k\in\hat{I}_a} \mathcal{Q}_k(a, \hat{u}(\hat{z}), \hat{\Theta}(\hat{z})) - \mathcal{Q}_k(a, \hat{u}(z^*), \hat{\Theta}(z^*)) \\ &= \sum_{k\in I_{aa}} \mathcal{Q}_k(a, \hat{u}(\hat{z}), \hat{\Theta}(\hat{z})) - \mathcal{Q}_k(a, \hat{u}(z^*), \hat{\Theta}(z^*)) + \sum_{b\neq a}\sum_{k\in I_{ab}} \mathcal{Q}_k(a, \hat{u}(\hat{z}), \hat{\Theta}(\hat{z})) - \mathcal{Q}_k(b, \hat{u}(z^*), \hat{\Theta}(z^*)), \end{aligned}$$

where $I_{ab} = \{k \in [K] : \hat{z}(k) = a, z^*(k) = b\}$. The second term is related to the misclassification loss $\ell(\hat{z}, z^*)$, and the objective function $\mathcal{Q}$ is related to the desired Frobenius norm. Therefore, we may proof the model satisfies the first two conditions.

For Conditions 2.3, we may use the results in Lemma 1 directly to find the upper bound for H_k. $\square$

Noticed that F_k, G_k, H_k are bounded, we only need to consider the oracle misclassification rate. Hence, we define the oracle misclassification loss as

$$\begin{aligned} \xi_{\text{ideal}}(\varepsilon) = \sum_{k\in[K]}\sum_{b\in[R]/z^*(k)} &\|\Omega_k(u^*, \Theta^*, z^*(k)) - \Omega_k(u^*, \Theta^*, b)\|_F^2 \cdot \\ &\mathbf{1}\left\{\langle S_k - \Sigma_k, \Omega_k(\hat{u}(z^*), \hat{\Theta}(z^*), b) - \Omega_k(\hat{u}(z^*), \hat{\Theta}(z^*), a)\rangle \leq -C(1-\varepsilon)\bar{\Delta}_k(a,b)^2\right\}. \end{aligned}$$

Lemma 3 (Error Decomposition)**.** *The MLE $\hat{z}$ satisfies following inequality*

$$\ell(\hat{z}, z^*) \leq C'\xi_{ideal}(\varepsilon),$$

with probability at least $1-\eta_1-\eta_2-\eta_3$ for some constant C'.

Proof Sketch of Lemma 3. To bound $\ell(\hat{z}, z^*)$, we need to consider the event

$$\begin{aligned} \mathbf{1}\{\hat{z}(k) = b\} &\leq \mathbf{1}\left\{\mathcal{Q}_k(b, \hat{u}(\hat{z}), \hat{\Theta}(\hat{z})) \leq \mathcal{Q}_k(z^*(k), \hat{u}(\hat{z}), \hat{\Theta}(\hat{z}))\right\} \\ &\leq \mathbf{1}\left\{\langle S_k - \Sigma_k, \Omega_k(\hat{u}(z^*), \hat{\Theta}(z^*), b) - \Omega_k(\hat{u}(z^*), \hat{\Theta}(z^*), z^*(k))\rangle \leq -C(1-\varepsilon)\bar{\Delta}_k(z^*(k),b)^2\right\} \\ &\quad + \mathbf{1}\left\{C\varepsilon\bar{\Delta}_k(z^*(k),b)^2 \leq CG_k(z^*(k),b,\hat{z}) + CH_k(z^*(k),b,\hat{z}) + F_k(z^*(k),b,\hat{z})\right\} \\ &\leq \mathbf{1}\left\{\langle S_k - \Sigma_k, \Omega_k(\hat{u}(z^*), \hat{\Theta}(z^*), b) - \Omega_k(\hat{u}(z^*), \hat{\Theta}(z^*), z^*(k))\rangle \leq -C(1-\varepsilon)\bar{\Delta}_k(z^*(k),b)^2\right\} \\ &\quad + \mathbf{1}\left\{C\varepsilon\bar{\Delta}_k(z^*(k),b)^2/2 \leq CG_k(z^*(k),b,\hat{z}) + F_k(z^*(k),b,\hat{z})\right\} \\ &\leq \mathbf{1}\left\{\langle S_k - \Sigma_k, \Omega_k(\hat{u}(z^*), \hat{\Theta}(z^*), b) - \Omega_k(\hat{u}(z^*), \hat{\Theta}(z^*), z^*(k))\rangle \leq -C(1-\varepsilon)\bar{\Delta}_k(z^*(k),b)^2\right\} \\ &\quad + C'\frac{G_k(z^*(k),b,\hat{z})^2}{\varepsilon^2\bar{\Delta}_k(z^*(k),b)^4} + C''\frac{F_k(z^*(k),b,\hat{z})^2}{\varepsilon^2\bar{\Delta}_k(z^*(k),b)^4}, \end{aligned}$$

where the second inequality follows by the error decomposition, the third inequality follows by Condition (2.3), and the last inequality follows by the fact that if $C'\frac{G_k(z^*(k),b,\hat{z})^2}{\varepsilon^2\bar{\Delta}_k(z^*(k),b)^4}+C''\frac{F_k(z^*(k),b,\hat{z})^2}{\varepsilon^2\bar{\Delta}_k(z^*(k),b)^4} < 1$ then the indicator $\mathbf{1}\left\{C\varepsilon\bar{\Delta}_k(z^*(k),b)^2/2 \leq CG_k(z^*(k),b,\hat{z}) + F_k(z^*(k),b,\hat{z})\right\} = 0$.

Therefore, by the definition of $\ell(\hat{z}, z^*)$, the loss is upper bounded by

$$
\begin{aligned}
&\sum_{k\in[K]}\sum_{b\in[R]/z^*(k)} \|\Omega_k(u^*,\Theta^*,z^*(k)) - \Omega_k(u^*,\Theta^*,z(k))\|_F^2 \cdot \\
&\qquad \mathbf{1}\left\{\langle S_k - \Sigma_k, \Omega_k(\hat{u}(z^*),\hat{\Theta}(z^*),b) - \Omega_k(\hat{u}(z^*),\hat{\Theta}(z^*),z^*(k))\rangle \leq -C(1-\varepsilon)\bar{\Delta}_k(z^*(k),b)^2\right\} \\
&\quad + \sum_{k\in[K]}\sum_{b\in[R]/z^*(k)} \|\Omega_k(u^*,\Theta^*,z^*(k)) - \Omega_k(u^*,\Theta^*,z(k))\|_F^2 \cdot \left[C'\frac{G_k(z^*(k),b,\hat{z})^2}{\varepsilon^2\bar{\Delta}_k(z^*(k),b)^4} + C''\frac{F_k(z^*(k),b,\hat{z})^2}{\varepsilon^2\bar{\Delta}_k(z^*(k),b)^4}\right] \\
\leq\ & \xi_{\text{ideal}}(\varepsilon) + C_0, \ell(\hat{z}, z^*),
\end{aligned}
$$

where the inequality follows by the Condition 2.1 and 2.2 and the definition of $\xi_{\text{ideal}}(\varepsilon)$. Hence, with proper constants, we have

$$\ell(\hat{z}, z^*) \leq C'\xi_{\text{ideal}}(\varepsilon).$$

□

References

Gao, C. and Zhang, A. Y. (2019). Iterative algorithm for discrete structure recovery. arXiv preprint arXiv:1911.01018.