

Complete proof for Precision clustering

Jiaxin Hu

July 14, 2021

1 Definitions

1.1 Model

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. For simplicity, we let $\Theta = \{\Theta_r\}_{r=0}^R$ denote the intercept and the sequence of factor matrices. Let $I_r = \{k \in [K] : z(k) = r\}$, and S_k denote the sample covariance matrix for k -th category with n independent sample $X_{z(k),1}, \dots, X_{z(k),n}$.

1.2 Parameter space

Define the parameter space for the assignment, \mathcal{P}_z as following

$$\mathcal{P}_z(R, \beta) = \left\{ z \in [R]^K : \frac{K}{\beta R} \leq |I_r| \leq \frac{K\beta}{R}, r \in [R] \right\}.$$

With given assignment $z \in \mathcal{P}(R, \beta)$, define the true parameter space $\mathcal{P}^*(z, \tau_1, \tau_2, \delta, m, M)$ and the estimator search space $\mathcal{P}(z, \delta, m, M)$ as following

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

and

$$\mathcal{P}(z, \delta, m, M) = \left\{ (u, \Theta_0, \Theta_r) : \begin{aligned} &\Theta_0, \Theta_r \text{ are positive definite for all } r \in [R]; \\ &\max_{r, r' \in [R]} \cos(\Theta_r, \Theta_{r'}) < \delta < 1; \quad \langle \Theta_0^{-1}, \Theta_r \rangle = 0, r \in [R]; \\ &\min_{k \in [K]} |u_k| > m > 0; \quad \sum_{k \in I_r} u_k = 0, \text{ for all } r \in [R] \end{aligned} \right\}.$$

1.3 Estimators

Let (z^*, u^*, Θ^*) denote the true parameters. Let $\mathcal{Q}(z, u, \Theta)$ denote the negative log-likelihood function, where

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

Then, we let $(\hat{z}, \hat{u}, \hat{\Theta})$ denote the MLE where,

our estimate is penalized MLE, not raw MLE

$$(\hat{z}, \hat{u}, \hat{\Theta}) = \arg \min_{z \in \mathcal{P}_z(R, \beta), (u, \Theta) \in \mathcal{P}(z, \delta, m, M)} \mathcal{Q}(z, u, \Theta), \text{ + lambda * penalty}$$

and let $(\tilde{u}, \tilde{\Theta})$ denote the oracle estimator with given z^* , where

$$(\tilde{u}, \tilde{\Theta}) = \arg \min_{(u, \Theta) \in \mathcal{P}(z^*, \delta, m, M)} \mathcal{Q}(z, u, \Theta).$$

For simplicity, in the following proof, the notation $\hat{\theta}$ implies θ some parameter derived from the MLE, $\tilde{\theta}$ implies some parameters derived from the oracle estimator, and θ^* implies some parameters derived from the true parameters.

1.4 Misclassification loss

Define the function

$$\hat{\Omega}_k(a) = \hat{\Theta}_0 + \hat{u}_k \hat{\Theta}_a.$$

Similar definitions $\tilde{\Omega}_k(a)$ and $\Omega_k^*(a)$ are proposed with $(z^*, \tilde{u}, \tilde{\Theta})$ and (z^*, u^*, Θ^*) . Then, we define the misclassification loss

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

Also define the minimal gap between different groups

$$\begin{aligned}\Delta_{\min}^2(p, m, \tau_1, \tau_2, \delta) &= \min_{k \in [K]} \min_{a \neq b \in [R]} \|\Omega_k^*(a) - \Omega_k^*(b)\|_F^2 \\ &\geq m^2 \min_{a \neq b \in [R]} \|\Theta_a^* - \Theta_b^*\|_F^2 \\ &\geq 2m^2 [p\tau_1^2 - \tau_2^2\delta], \quad \text{wrong}\end{aligned}$$

where the last inequality follow by the fact that

not spectral norm. Should be F-norm (vectorial 2 norm)

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

for $A, B \in \mathbb{R}^{p \times p}$. Note that Δ_{\min} is a increasing function in p, m and a decreasing function in δ . For simplicity, we use Δ_{\min}^2 to denote the minimal gap.

Last, we consider the Hamming loss $h(z, z^*) = \sum_{k \in [K]} \mathbf{1}\{z(k) \neq z^*(k)\}$, where

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

1.5 Error decomposition

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

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

Define the errors

$$\hat{\Delta}(a, b) = \hat{\Omega}_k(a) - \Omega_k^*(b); \quad \tilde{\Delta}(a, b) = \tilde{\Omega}_k(a) - \Omega_k^*(b); \quad \Delta(a, b) = \hat{\Omega}_k(a) - \tilde{\Omega}_k(b).$$

By the Taylor Expansion, we have

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

where

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

with a constant c related to the τ_1, τ_2 and the second equation follows by the Lemma 2 that $\left\| \hat{\Delta}(b, a) \right\|_F$ is bounded for n large enough.

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

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

Rearranging the inequality, we have

$$\langle S_k - \Sigma_k, \tilde{\Omega}_k(b) - \tilde{\Omega}_k(a) \rangle \leq -c\bar{\Delta}_k(a, b)^2 + cG_k(a, b, \hat{z}) + cH_k(a, b) + F_k(a, b, \hat{z}),$$

where

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

Last, we define oracle misclassification loss as

$$\xi_{\text{ideal}}(\epsilon) = \sum_{k \in [K]} \sum_{b \in [R]/z^*(k)} \|\Omega_k^*(z^*(k)) - \Omega_k^*(b)\|_F^2 \cdot \mathbf{1} \left\{ \langle S_k - \Sigma_k, \tilde{\Omega}_k(b) - \tilde{\Omega}_k(z^*(k)) \rangle \leq -c(1 - \epsilon)\bar{\Delta}_k(a, b)^2 \right\}.$$

2 Useful Lemmas

Lemma 1 (Oracle estimation error). *The oracle estimator $(\tilde{u}, \tilde{\Theta})$ satisfy the following inequalities simultaneously with probability at least $1 - \mathcal{O}(1/n)$,*

$$\|\tilde{\Theta}_0 - \Theta_0^*\|_F \leq C_0 p \sqrt{\frac{\log p \log n}{nK}}, \quad \|\tilde{\Theta}_r - \Theta_r^*\|_F \leq C_r p \sqrt{\frac{\log p \log n}{n|I_r^*|}}, \quad |\tilde{u}_k - u_k^*| \leq C_k p \sqrt{\frac{\log p \log n}{n}},$$

for some large positive constants C_0, C_r, C_k and $\min_{r \in [R]} |I_r^*| \geq \frac{K}{\beta R}$.

Lemma 2 (MLE estimation error). *The MLE $(\hat{z}, \hat{u}, \hat{\Theta})$ satisfy the following inequalities simultaneously with probability at least $1 - \mathcal{O}(1/n)$,*

$$\sum_{k \in [K]} \left\| \hat{\Omega}_k(\hat{z}(k)) - \Omega_k^*(z^*(k)) \right\|_F \leq CK \sqrt{\frac{\log n \log p}{n}},$$

and that $\hat{z} \rightarrow z^*$, i.e., $\ell(\hat{z}, z^*) \rightarrow 0$ as $n \rightarrow \infty$.

Lemma 3 (Intercept estimation). *The oracle estimator and MLE of the intercept are equivalent, i.e., $\hat{\Theta}_0 = \tilde{\Theta}_0$.*

Is this an assumption or conclusion? Should we add it as conditions to Thm 3.1-3.2?

Lemma 4 (Conditions check). *For n large enough such that $\ell(\hat{z}, z^*) \leq \tau = \frac{K}{2\beta R}$, we have*

1.

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

holds with probability at least $1 - \eta_1$ for small positive constant C_1 , and $\epsilon, \eta_1 > 0$;

2.

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

holds with probability at least $1 - \eta_2$ for small positive constant C_2 , and $\epsilon, \eta_2 > 0$;

3.

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

holds with probability at least $1 - \eta_3$ for small positive constant C_3 , and $\epsilon, \eta_3 > 0$.

3 Main theorems

Theorem 3.1 (Error decomposition). *The MLE \hat{z} satisfies following inequality*

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

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

Remark 1. The parameter ϵ in $\xi_{\text{ideal}}(\epsilon)$ is the same as the ϵ in Lemma 4.

Proof. Very similar to the Proof of Theorem 3.1 in (Gao and Zhang, 2019). \square

Theorem 3.2 (Oracle misclassification rate). *Assume $\Delta_{\min} = \mathcal{O}(K^\gamma)$ for some $\gamma > 0$, $n = C_n \exp(\Delta_{\min}^2)$ for C_n large enough, and the ϵ in Lemma 4 small enough. With probability $1 - \eta_1 - \eta_2 - \eta_3 - \exp(-\Delta_{\min})$ as $K \rightarrow \infty$*

$$\xi_{\text{ideal}}(\epsilon) \leq K \exp(-(1 - c\epsilon)^2 C \Delta_{\min}^2),$$

where c, C are two positive constants.

Remark 2. In fact, if Lemma 4 holds for the decreasing sequence $\epsilon_K \rightarrow 0$ and $\eta_1, \eta_2, \eta_3 \rightarrow \mathcal{O}(\exp(-\Delta_{\min}))$ as $K \rightarrow \infty$, we have

$$\xi_{\text{ideal}}(\epsilon) \leq K \exp(-(1 + o(1)) C \Delta_{\min}^2),$$

with $1 - \mathcal{O}(\exp(-\Delta_{\min}))$.

References

Gao, C. and Zhang, A. Y. (2019). Iterative algorithm for discrete structure recovery. [arXiv preprint arXiv:1911.01018](https://arxiv.org/abs/1911.01018).