Precision clustering vs Community detection

Jiaxin Hu

July 8, 2021

1 Some thoughts about precision clustering

Intuitively, the misclassification rate of the MLE of the clustering, $\hat{z}$, should decrease as the number of categories K increasing, e.g., $\mathbb{P}(MCR(\hat{z}, z^*) = 1)$ decreases as $K \to \infty$. Note that we have four blocks of parameter in the model: clustering assignment z (or membership matrix M), degree-corrected parameter for each category u_k (or the diagonal matrix U), intercept matrix Θ_0, factor matrix Θ_r. For fixed number of groups R, there two possible explanation for the accuracy improvement:

1. (direct improvement) given other parameters u_k, Θ_0, Θ_r, the accuracy of $\hat{z}$ increases due to the increment of K;
2. (indirect improvement) the increment of K benefits the estimation of u_k, Θ_0, Θ_r, and thereof improves the estimation of $\hat{z}$.

The following example indicates that the community detection (biclustering, or tensor clustering) takes advantage of the "direct improvement" while precision clustering does not enjoy the "direct improvement". Without the loss of generality, we let the dimension of precision matrices is 1 and thus precision clustering is equal to an 1-dimensional clustering.

Example 1. Consider the case $R = 2$ and $2K$ independent categories $X_1, ..., X_K, X_{K+1}, ..., X_{2K}$, where $z^*(k) = 1$ for $k = 1, ..., K$ and $z^*(k) = 2$ for $k = K + 1, ..., 2K$. Suppose we have a new category X_{2K+1}, and we want to estimate the assignment $z^*(2K + 1)$. Assume we know the new category X_{2K+1} belongs to the group 1, i.e., $z^*(2K + 1) = 1$ and the degree-corrected parameters are known.

In **community detection**, suppose the in-community connection p and between-community connection q are known. With new category X_{2K+1}, we have new $4K$ entries in the adjacent matrix, $A_{i,2K+1}$ and $A_{2K+1,i}$, for $i = 1, ..., 2K$. Then, the probability for the MLE $\hat{z}(2K + 1)$ misclassifies is

$$\mathbb{P}_{z^*(2K+1)=1}(\hat{z}(2K + 1) = 2) = \mathbb{P}_{z^*(2K+1)} \left[\mathcal{L}(z(2K + 1) = 2, p, q|A) > \mathcal{L}(z(2K + 1) = 1, p, q|A)\right],$$

where $\mathbb{P}_{z^*(2K+1)=1}$ denote the true probability measure, i.e.,

$$\begin{aligned} A_{i,2K+1} = A_{2K+1,i} &\sim Ber(\theta_i \theta_{2K+1} p), \quad i = 1, ..., K \\ A_{i,2K+1} = A_{2K+1,i} &\sim Ber(\theta_i \theta_{2K+1} q), \quad i = K + 1, ..., 2K, \end{aligned}$$

and $\mathcal{L}(z(2K+1)=r,p,q|A), r=1,2$ denote the likelihood with parameter p,q and the data A if X_{2k+1} belongs to group r. Note that $\mathcal{L}$ is a function of $A_{i,2K+1}$ and $A_{2K+1,i}$ which implies the comparison between $\mathcal{L}(z(2K+1)=2,p,q|A) > \mathcal{L}(z(2K+1)=1,p,q|A)$ is related to the categories $X_1,...,X_{2K}$. Thus, the misclassification probability of the new category depends on previous $2K$ categories in biclustering. Specifically, the equation (21) in Gao et al. (2018) implies that

$$\mathbb{P}_{z^*(2K+1)=1}(z(2K+1)=2) \leq 2\exp\left(-\theta_{2K+1}K(\sqrt{p}-\sqrt{q})^2\right)$$

In **precision clustering**, suppose the group 1 follows $N(0,\sigma_1^2)$ and group 2 follows $N(0,\sigma_2^2)$. With new category X_{2K+1}, we have one new data point X_{2K+1}. The the probability for the MLE $\hat{z}(2K+1)$ misclassifies is

$$\mathbb{P}_{z^*(2K+1)=1}(\hat{z}(2K+1)=2) = \mathbb{P}_{z^*(2K+1)}\left[\mathcal{L}(z(2K+1)=2,\sigma_1,\sigma_2|X_k) > \mathcal{L}(z(2K+1)=2,\sigma_1,\sigma_2|X_k)\right]$$

$$= \mathbb{P}_{z^*(2K+1)}\left[\ell(\sigma_2|X_{2K+1}) > \ell(\sigma_1|X_{2K+1})\right],$$

where $\mathbb{P}_{z^*(2K+1)=1}$ denote the true probability measure, i.e., $X_{2K+1} \sim N(0,\sigma_1^2)$, $\ell(\sigma|X)$ denote the likelihood function for data $X \sim N(0,\sigma^2)$, and the second equation follows by the fact the $X_1,...,X_{2K+1}$ are independent. Thus, the misclassification probability of the new category does not depend on previous $2K$ categories with given true parameters in precision clustering.

Hence, I suspects that precision clustering may not achieve the exponential rate in K as biclustering or tensor clustering, since precision clustering can only reflect the "indirect improvement" from K.

2 Exact clustering in 1 dimensional clustering

Model

Consider $Y_j \sim \mathcal{N}_d(\theta_{z^*(j)}, I_d)$ for $j=1,...,p$, and the least square estimator

$$(\hat{z},\hat{\theta}) = \arg\min_{z,\theta} \left\|Y_j - \theta_{z(j)}\right\|^2.$$

Conclusion

Assume $\frac{\Delta^2_{\min}}{\log k + kd/p} \to \infty, p/k \to \infty$ and $\min_{a\in[k]} \sum_{j=1}^p 1\{z^*(j)=a\} \geq \frac{\alpha p}{k}$. Then

$$\ell(\hat{z},z^*) \leq C\xi_{\text{ideal}} \leq p\exp(-(1-o(1))\frac{\Delta^2_{\min}}{8}),$$

with probability at least $1-\exp(-\Delta_{\min})$.

Proof Sketch To obtain the results, we should combine the proof for Theorem 3 and Lemma 4.2. Here are the key ideas:

1. In Theorem 3, $\ell(\hat{z},z^*)$ refers to the misclassification rate for the estimated assignment $\hat{z}$ when

the true parameters θ_j^* are given. Then, we need to bound the probability

$$\begin{aligned}\mathbb{P}(\hat{z}(j) = b) &= \mathbb{P}\left(\left\|Y_j - \hat{\theta}(\hat{z})_{\hat{z}(j)}\right\|^2 \leq \left\|Y_j - \hat{\theta}(\hat{z})_{\hat{z}^*(j)}\right\|^2\right) \\ &= \mathbb{P}\left(\langle \epsilon, \hat{\theta}(z^*)_{\hat{z}(j)} - \hat{\theta}(z^*)_{z^*(j)} \rangle \leq f(\Delta_{\min})\right) + o(1).\end{aligned}$$

The key idea is still using the likelihood comparison and decompose the probability to the oracle error (the misclassication error when z^* is known) and the other part.

2. In Lemma 4.2, ξ_{ideal} describes the oracle error, which somehow describes the estimation error when the true assignment z^* is known. Note that the estimation error is usually exponential rate due to the Gaussianity. Then the final rate for misclassification rate is exponential.

References

Gao, C., Ma, Z., Zhang, A. Y., and Zhou, H. H. (2018). Community detection in degree-corrected block models. The Annals of Statistics, 46(5):2153–2185.