

Thoughts for iteration in hDCBM

Jiaxin Hu

August 10, 2021

1 Some analysis of clustering with degree

Consider an order- d binary observation $\mathcal{Y} \in \{0, 1\}^{p \times \dots \times p}$ which is generated from the following model

$$\mathcal{Y} = \mathcal{X} + \mathcal{E} = \mathcal{S} \times_1 \Theta \mathbf{M} \times_2 \dots \times_d \Theta \mathbf{M} + \mathcal{E},$$

where $\mathcal{S} \times \mathbb{R}^{r \times \dots \times r}$ is the symmetric group mean tensor.

Note that when $d = 1$, there is no way to do the clustering unless for each node $j \in [p]$ we have a multidimensional representation $y_j \in \mathbb{R}^m, m \geq 2$. Therefore, we start from the network biclustering case with $d = 2$.

Here we discuss how to find the optimal assignment for node j given the membership z for other nodes and the signal \mathcal{S} .

1.1 When $d = 2$

Based on the Neyman-pearson lemma, the optimal assignment should maximize the likelihood for node j with observations $\mathcal{Y}_{ji}, i \in [p]/j$. The likelihood function and log-likelihood function are

$$\mathcal{L}(z_j | \theta, \mathcal{S}, \mathcal{Y}) = \prod_{i \in [p]/j} (\theta_j \theta_i \mathcal{S}_{z_j, z_i})^{\mathcal{Y}_{ji}} (1 - \theta_j \theta_i \mathcal{S}_{z_j, z_i})^{1 - \mathcal{Y}_{ji}},$$

and

$$l(z_j | \theta, \mathcal{S}, \mathcal{Y}) = \sum_{i \in [p]/j} \mathcal{Y}_{ji} \log(\theta_j \theta_i \mathcal{S}_{z_j, z_i}) + (1 - \mathcal{Y}_{ji}) \log(1 - \theta_j \theta_i \mathcal{S}_{z_j, z_i}),$$

where $\mathcal{S}_{z_j, z_i} = \alpha$ if $z_j = z_i$ and $\mathcal{S}_{z_j, z_i} = \beta$ if $z_j \neq z_i$. The optimal rule to update the assignment is

$$\hat{z}_j = \arg \max_{z_j \in [r]} l(z_j | \theta, \mathcal{S}, \mathcal{Y}). \quad (1)$$

We consider $\theta_j \mathcal{S}_{z_j, z_i}$ as a single term since θ_j is determinant and positive even though unknown, and the vectors $\theta_j \mathcal{S}_{z_j, \cdot}$ and $\mathcal{S}_{z_j, \cdot}$ share the same pattern. For simplicity, we ignore θ_j in the following analysis.

Angle approximation

Rewrite the log-likelihood function in terms of inner product

$$\begin{aligned} l(z_j|\theta, \mathcal{S}, \mathcal{Y}) &= \sum_{i \in [p]/j} \mathcal{Y}_{ji} \log(\theta_j \theta_i \mathcal{S}_{z_j, z_i}) + (1 - \mathcal{Y}_{ji}) \log(1 - \theta_j \theta_i \mathcal{S}_{z_j, z_i}) \\ &= \langle \mathcal{Y}_{j\cdot}, \log(\theta_i \mathcal{S}_{z_j, \cdot}) \rangle + \langle 1 - \mathcal{Y}_{j\cdot}, \log(1 - \theta_i \mathcal{S}_{z_j, \cdot}) \rangle, \end{aligned}$$

where $\mathcal{S}_{z_j, \cdot} = \llbracket \mathcal{S}_{z_j, z_i} \rrbracket \in \mathbb{R}^p$. Note that the log-likelihood is the sum of two angles in form $\langle \mathcal{Y}_{j\cdot}, \log(\theta_i \mathcal{S}_{z_j, \cdot}) \rangle$, and the log is monotone function. A natural thoughts is that if $\mathcal{S}_{a\cdot}, a \in [r]$ **are separable enough**, the angle $\langle \mathcal{Y}_{j\cdot}, \theta_i \mathcal{S}_{z_j, \cdot} \rangle$ can be a good approximation of $\langle \mathcal{Y}_{j\cdot}, \log(\theta_i \mathcal{S}_{z_j, \cdot}) \rangle$. Also, this approximation leads to easier computation and link the optimal rule with k -means (discuss later). Therefore, the optimal rule can be approximated by

$$\hat{z}_j \approx \arg \max_{z_j \in [r]} \sum_{i \in [p]/j} \mathcal{Y}_{ji} \theta_i \mathcal{S}_{z_j, z_i} + (1 - \mathcal{Y}_{ji})(1 - \theta_i \mathcal{S}_{z_j, z_i}). \quad (2)$$

Note that the separation condition is necessary for the approximation. A counterexample of optimal rule (1) and approximation (2) is following.

Example 1 (Counterexample of the approximation). Consider the case $d = 2, r = 2, p = 2, \theta_1 = \theta_2 = 1$. Suppose $\mathcal{S}_{1\cdot} = (0.25, 0.8)$ and $\mathcal{S}_{2\cdot} = (0.4, 0.9)$. We have an observation $\mathcal{Y} = (1, 0)$. According to the optimal rule (1)

$$\hat{z} = \arg \max_{1,2} \{l(z = 1) = \log(0.25) + \log(0.2), l(z = 2) = \log(0.4) + \log(0.1)\} = 1.$$

According to the approximate rule (2)

$$\hat{z} = \arg \max_{1,2} \{l(z = 1) = 0.25 + 0.2, l(z = 2) = 0.4 + 0.1\} = 2.$$

Assortative case in Gao et al. (2018)

In this case, we assume \mathcal{S} takes only two distinct values: α for diagonal elements and β for off-diagonal elements, and $\alpha > \beta$.

By the angle approximation, we have

$$\begin{aligned} \hat{z}_j &\approx \arg \max_{z_j \in [r]} \sum_{i \in [p]/j} \mathcal{Y}_{ji} \theta_i \mathcal{S}_{z_j, z_i} + (1 - \mathcal{Y}_{ji})(1 - \theta_i \mathcal{S}_{z_j, z_i}) \\ &= \arg \max_{z_j \in [r]} \sum_{i \in [p]/j} (2\mathcal{Y}_{ji} - 1) \theta_i \mathcal{S}_{z_j, z_i} \\ &= \arg \max_{z_j \in [r]} \sum_{i: z_i = z_j} (2\mathcal{Y}_{ji} - 1) \theta_i \alpha + \sum_{i: z_i \neq z_j} (2\mathcal{Y}_{ji} - 1) \theta_i \beta, \end{aligned} \quad (3)$$

which implies

$$\hat{z}_j = \arg \max_{a \in [r]} \sum_{i: z_i = a} (2\mathcal{Y}_{ji} - 1) \theta_i$$

Compared with non-degree clustering, the main difference comes from θ_i . However, the assumption $\frac{1}{p_a} \sum_{z_i=a} \theta_i \approx 1$ implies each θ_i is around 1, and this assumption leads the non-degree refinement to be a good approximation of refinement for clustering with degrees. Specifically,

$$\begin{aligned}\hat{z}_j &\approx \arg \max_{a \in [r]} \sum_{i: z_i=a} 2\mathcal{Y}_{ji}\theta_i - |\{i : z_i = a\}| \\ &\approx \arg \max_{a \in [r]} \frac{1}{|\{i : z_i = a\}|} \sum_{i: z_i=a} \mathcal{Y}_{ji},\end{aligned}$$

where the first and second approximations follow by the assumption $\frac{1}{p_a} \sum_{z_i=a} \theta_i \approx 1$.

Non-assortative case

In [Gao et al. \(2018\)](#), \mathcal{S} only takes two distinct values. Here, we relax such assumption. To generalize, we start from equation (3).

$$\begin{aligned}\hat{z}_j &= \arg \max_{z_j \in [r]} \sum_{i \in [p]/j} (2\mathcal{Y}_{ji} - 1)\theta_i \mathcal{S}_{z_j, z_i} \\ &\approx \arg \max_{z_j \in [r]} \sum_{i \in [p]/j} (2\mathcal{Y}_{ji} - 1)\mathcal{S}_{z_j, z_i} \\ &= \arg \max_{z_j \in [r]} \langle 2\mathcal{Y}_{j\cdot} - 1, \mathcal{S}_{z_j, \cdot} \rangle.\end{aligned}$$

1.2 Compared with k -means

Here we compare the optimal rule (2) with k -means. The k -means rule is

$$\begin{aligned}\hat{z}_j &= \arg \min_{z_j \in [r]} \|\mathcal{Y}_{j\cdot} - \Theta \mathcal{S}_{z_j, \cdot}\|_F^2 \\ &= \arg \min_{z_j \in [r]} \frac{1}{2} \left[\|\mathcal{Y}_{j\cdot}\|_F^2 + \|\Theta \mathcal{S}_{z_j, \cdot}\|_F^2 - 2\langle \mathcal{Y}_{j\cdot}, \Theta \mathcal{S}_{z_j, \cdot} \rangle \right] \\ &\quad + \frac{1}{2} \left[\|1 - \mathcal{Y}_{j\cdot}\|_F^2 + \|1 - \Theta \mathcal{S}_{z_j, \cdot}\|_F^2 - 2\langle 1 - \mathcal{Y}_{j\cdot}, 1 - \Theta \mathcal{S}_{z_j, \cdot} \rangle \right] \\ &= \arg \max_{z_j \in [r]} \langle \mathcal{Y}_{j\cdot}, \Theta \mathcal{S}_{z_j, \cdot} \rangle + \langle 1 - \mathcal{Y}_{j\cdot}, 1 - \Theta \mathcal{S}_{z_j, \cdot} \rangle - \frac{1}{2} \left[\|\Theta \mathcal{S}_{z_j, \cdot}\|_F^2 + \|1 - \Theta \mathcal{S}_{z_j, \cdot}\|_F^2 \right].\end{aligned}$$

In general, only if $\|\Theta \mathcal{S}_{z_j, \cdot}\|_F^2$ and $\|1 - \Theta \mathcal{S}_{z_j, \cdot}\|_F^2$ has the same value for all $z_j \in [r]$, the k -means is equivalent to the approximate rule (2).

In assortative case, if $\theta_i = 1$, $\mathcal{S}_{a\cdot}$ share the same norm, and thus the optimal rule is equal to k -means. That means, the refinement in Algorithm 2 in [Gao et al. \(2018\)](#) is equivalent to the regular k -means.

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.