Oracle analysis hDCBM algorithm

Jiaxin Hu

July 22, 2021

1 Oracle analysis

Suppose we have p nodes from r communities and observe the adjacent tensor $\mathcal{Y} \in \{0,1\}^{p\times p\times p}$ whose entry $\mathcal{Y}_{ijk}$ refers to the connection of the triplet (i,j,k). Let $\theta = (\theta_1,...,\theta_p) \in \mathbb{R}^p$ denote the degree-corrected parameters and $z = (z_1,...,z_p) \in [r]^p$ denote the clustering assignment. Consider the hDCBM model

$$\mathbb{E}[\mathcal{Y}] = \mathcal{S} \times_1 \Theta \boldsymbol{M} \times_2 \Theta \boldsymbol{M} \times_3 \Theta \boldsymbol{M},$$

where $\mathcal{S} \in \mathbb{R}^{r\times r\times r}$ is the symmetric core tensor, $\Theta = \text{diag}(\theta) \in \mathbb{R}^{p\times p}$ and $\boldsymbol{M} \in \mathbb{R}^{p\times r}$ is the hard membership matrix based on z. Algorithms 1, 2 are modified slightly for this simpler setting.

Suppose we have the noiseless observation

$$\mathcal{Q} = \mathcal{S} \times_1 \Theta \boldsymbol{M} \times_2 \Theta \boldsymbol{M} \times_3 \Theta \boldsymbol{M}.$$

We study the output of the initialization algorithm 1 and the iteration algorithm 2. Without the loss of generality, we use SCORE function $h(x) = \|x\|_F$.

1.1 Initialization

Note that $\mathcal{Q}$ is symmetric. So, the intermediate parameters $\tilde{\boldsymbol{U}}_k, \hat{\boldsymbol{U}}_k, z_k^{(0)}$ in Algorithm 1 are the same on d modes. For simplicity, let $\tilde{\boldsymbol{U}}, \hat{\boldsymbol{U}}, z^{(0)}$ denote the common parameters.

1. The very first step in Algorithm 1 is HOSVD factor $\tilde{\boldsymbol{U}}$. Following the tensor algebra in Han et al. (2020), we have

$$\tilde{\boldsymbol{U}} = \text{SVD}_r\left(\mathcal{M}_1(\mathcal{Q})\right) = \text{SVD}_r\left(\Theta \boldsymbol{M} \mathcal{M}_1(\mathcal{S})(\Theta \boldsymbol{M} \otimes \Theta \boldsymbol{M})^T\right). \tag{1}$$

By the Lemma 3.1 in Ke et al. (2019), we have

$$\tilde{\boldsymbol{U}} = \Theta \boldsymbol{M} \boldsymbol{D}^{-1} \boldsymbol{B},$$

where $\boldsymbol{D} = \text{diag}(d), d = (d_1,...,d_r) \in \mathbb{R}^r$, $d_k = (\sum_{z_i = k} \theta_i^2)^{1/2}$, $\boldsymbol{B}$ is an orthonormal matrix, i.e., $\boldsymbol{B}^T\boldsymbol{B} = \boldsymbol{1}_r$. Thus, we also have $\tilde{\boldsymbol{U}}^T\tilde{\boldsymbol{U}} = \boldsymbol{1}_r$.

2. Secondly, we improve the SVD estimation via

$$\begin{aligned}\hat{\boldsymbol{U}} &= \text{SVD}_r\left(\mathcal{M}_1(\mathcal{Q} \times_2 \tilde{\boldsymbol{U}}^T \times_3 \tilde{\boldsymbol{U}}^T)\right)\\ &= \text{SVD}_r\left(\Theta \boldsymbol{M} \mathcal{M}_1(\mathcal{S})(\tilde{\boldsymbol{U}}^T \Theta \boldsymbol{M} \otimes \tilde{\boldsymbol{U}}^T \Theta \boldsymbol{M})^T\right).\end{aligned} \tag{2}$$

Note that $\tilde{\boldsymbol{U}}, \hat{\boldsymbol{U}}$ refer to the left singular space, both of them are orthonormal, and the right hand side in equation (1) and (2) share the same column space. Therefore, we have $\tilde{\boldsymbol{U}} = \hat{\boldsymbol{U}}$. However, the equivalence is not true for noisy observation.

3. Thirdly, we calculate the projected matrix

$$\begin{aligned}\hat{\boldsymbol{Y}}_1 &= \hat{\boldsymbol{U}}\hat{\boldsymbol{U}}^T \mathcal{M}_1(\mathcal{Q} \times_2 \tilde{\boldsymbol{U}}^T \times_3 \tilde{\boldsymbol{U}}^T)\\ &= \left[\Theta \boldsymbol{M} \boldsymbol{D}^{-1} \boldsymbol{B}\right] \left[\boldsymbol{B}^T \boldsymbol{D}^{-1} \boldsymbol{M}^T \Theta\right] \Theta \boldsymbol{M} \mathcal{M}_1(\mathcal{S})(\tilde{\boldsymbol{U}}^T \Theta \boldsymbol{M} \otimes \tilde{\boldsymbol{U}}^T \Theta \boldsymbol{M})^T\\ &= \Theta \boldsymbol{M} \boldsymbol{V},\end{aligned}$$

where $\boldsymbol{V} = \boldsymbol{B}\boldsymbol{B}^T \mathcal{M}_1(\mathcal{S})(\tilde{\boldsymbol{U}}^T \Theta \boldsymbol{M} \otimes \tilde{\boldsymbol{U}}^T \Theta \boldsymbol{M})^T \in \mathbb{R}^{r \times r^2}$ since $\boldsymbol{M}^T \Theta \Theta \boldsymbol{M} = \boldsymbol{D}^2$.

4. Last, we apply SCORE normalization to $\hat{\boldsymbol{Y}}_1$. Let $\hat{\boldsymbol{Y}}_{1,j}$ denote the j-th row or $\hat{\boldsymbol{Y}}_1$ for $j \in [p]$, and $\boldsymbol{V}_u$ denote the u-th row of $\boldsymbol{V}$ for $u \in [r]$. Then, for $z_j = u$, we have normalized row

$$\hat{\boldsymbol{Y}}_{1,j}^s = \frac{\hat{\boldsymbol{Y}}_{1,j}}{\left\|\hat{\boldsymbol{Y}}_{1,j}\right\|} = \frac{\theta_j \boldsymbol{V}_u}{\theta_j \left\|\boldsymbol{V}_u\right\|_F} = \frac{\boldsymbol{V}_u}{\left\|\boldsymbol{V}_u\right\|_F}.$$

Hence, the normalized $\hat{\boldsymbol{Y}}_1^s$ only has r distinct rows, and thus the k-means clustering exactly covers the true membership, i.e., $z^{(0)} = z^*$.

1.2 Iteration

Take the true assignment $z^{(0)} = z^*$ and the true signal $\mathcal{Q}$ as input of the iteration Algorithm 2.

1. The first step imposes the SCORE normalization on all modes. Note that $\mathcal{Q}$ is symmetric. Let $\boldsymbol{Q}_k$ denote the k-th matrication of $\mathcal{Q}$. We have

$$\hat{\theta}_j = \left\|\boldsymbol{Q}_{1,j}\right\| = \theta_j \left\|\boldsymbol{V}_{z_j^*}\right\|_F,$$

where $\boldsymbol{V} = \mathcal{M}_1(\mathcal{S})(\Theta \boldsymbol{M} \otimes \Theta \boldsymbol{M})^T$. Let $\tilde{\boldsymbol{V}} = \text{diag}(\left\|\boldsymbol{V}_1\right\|_F, ..., \left\|\boldsymbol{V}_r\right\|_F)$.

2. Secondly, we apply the SCORE normalization to $\mathcal{Q}$. Note that

$$\hat{\Theta} = \Theta \text{diag}\left(\left\|\boldsymbol{V}_{z_1^*}\right\|_F, ..., \left\|\boldsymbol{V}_{z_p^*}\right\|_F\right)$$

$$\begin{aligned}\mathcal{Q}^s &= \mathcal{Q} \times_1 \hat{\Theta}^{-1} \times_2 \cdots \hat{\Theta}^{-1} \times_3 \hat{\Theta}^{-1}\\ &= \mathcal{S} \times_1 \boldsymbol{M}\tilde{\boldsymbol{V}} \times_2 \boldsymbol{M}\tilde{\boldsymbol{V}} \times_2 \boldsymbol{M}\tilde{\boldsymbol{V}}\\ &= \tilde{\mathcal{S}} \times_1 \boldsymbol{M} \times_2 \boldsymbol{M} \times_3 \boldsymbol{M},\end{aligned}$$

where $\tilde{\mathcal{S}} = \mathcal{S} \times_1 \tilde{\boldsymbol{V}} \times_2 \tilde{\boldsymbol{V}} \times_3 \tilde{\boldsymbol{V}}$.

3. Last, note that $\mathcal{Q}^s$ is completely separated. The HLloyd algorithm will return the true assignment, i.e., $z^{(T)} = z^*$.

Algorithm 1 High-order degree-corrected spectral clustering (HDCSC)

Input: Observation $\mathcal{Y} \in \mathbb{R}^{p\times\cdots\times p}$, r, relaxation factor in k-means $M > 1$, SCORE normalization function h

Compute $\tilde{\boldsymbol{U}}_k = SVD_{r_k}(\mathcal{M}_k(\mathcal{Y})$ for $k \in [d]$

for $k \in [d]$ **do**

  Estimate the singular space $\hat{\boldsymbol{U}}_k$ via

$$\hat{\boldsymbol{U}}_k = SVD_r(\mathcal{M}_k(\mathcal{Y} \times_1 \tilde{\boldsymbol{U}}_1^T \times \cdots \times_{k-1} \tilde{\boldsymbol{U}}_{k-1}^T \times_{k+1} \tilde{\boldsymbol{U}}_{k+1}^T \times \cdots \times_d \tilde{\boldsymbol{U}}_d^T))$$

end for

for $k \in [d]$ **do**

  Calculate $\hat{\boldsymbol{Y}}_k$ via $\hat{\boldsymbol{Y}}_k = \hat{\boldsymbol{U}}_k \hat{\boldsymbol{U}}_k^T \mathcal{M}_k(\mathcal{Y} \times_1 \hat{\boldsymbol{U}}_1^T \times \cdots \times_{k-1} \hat{\boldsymbol{U}}_{k-1}^T \times_{k+1} \hat{\boldsymbol{U}}_{k+1}^T \times \cdots \times_d \hat{\boldsymbol{U}}_d^T)$

  Let $\hat{\boldsymbol{Y}}_{kj}$ denote the rows of $\hat{\boldsymbol{Y}}_k$ for $j \in [p_k]$. Obtain the SCORE normalized $\hat{\boldsymbol{Y}}_k^s$ via $\hat{\boldsymbol{Y}}_{kj}^s = \frac{\hat{\boldsymbol{Y}}_{kj}}{h(\hat{\boldsymbol{Y}}_{kj})}$

  Find the initial assignment $z_k^{(0)} \in [r_k]^{p_k}$ and centroids $\hat{x}_1, ..., \hat{x}_{rk} \in \mathbb{R}^{r-k}$ such that

$$\sum_{j=1}^{p_k} \left\| (\hat{\boldsymbol{Y}}_{kj}^s)^T - \hat{x}_{(z_k^{(0)})_j} \right\|^2 \leq M \min_{x_1,\ldots,x_{r_k}, z_k} \sum_{j=1}^{p_k} \left\| (\hat{\boldsymbol{Y}}_{kj}^s)^T - x_{(z_k^{(0)})_j} \right\|^2$$

end for

Find the average of $z_k^{(0)}, k \in [d]$, $z^{(0)}$.

Output: $\left\{ z^{(0)} \in [r]^p \right\}$

Algorithm 2 High-order degree-corrected Lloyd Algorithm (HDCLloyd)

Input: Observation $\mathcal{Y} \in \mathbb{R}^{p\times\cdots\times p}$, initialization $\{z^{(0)} \in [r]^p\}$, iteration number T, SCORE normalization function h

Let $\boldsymbol{Y}_k$ denote the k-th model matricization of $\mathcal{Y}$ and $\boldsymbol{Y}_{k,j}$ denote the j-th column of $\boldsymbol{Y}_k$. Find the surrogate estimation of θ via

$$\hat{\theta}_j = \frac{1}{d}\left\|\sum_{k=1}^{d} \boldsymbol{Y}_{k,j}\right\|_F$$

Let $\hat{\Theta} = \mathrm{diag}(\hat{\theta})$. Apply SCORE normalization to $\mathcal{Y}$ via

$$\mathcal{Y}^s = \mathcal{Y} \times_1 \hat{\Theta}^{-1} \times \cdots \times_d \hat{\Theta}^{-1}$$

Apply HLloyd with $\mathcal{Y}^s$, initialization $\{z^{(0)} \in [r]^p\}$, and iteration number T.

Output: $\left\{z^{(T)} \in [r]^p\right\}$

References

Han, R., Luo, Y., Wang, M., and Zhang, A. R. (2020). Exact clustering in tensor block model: Statistical optimality and computational limit. arXiv preprint arXiv:2012.09996.

Ke, Z. T., Shi, F., and Xia, D. (2019). Community detection for hypergraph networks via regularized tensor power iteration. arXiv preprint arXiv:1909.06503.