k-median or k-means

Jiaxin Hu

August 27, 2021

1 Problem

Our model

$$\mathcal{A} = \mathcal{S} \times_1 \Theta \boldsymbol{M} \times_2 \cdots \times_K \Theta \boldsymbol{M} + \mathcal{E},$$

where we let $\mathcal{X} = \mathcal{S} \times_1 \Theta \boldsymbol{M} \times_2 \cdots \times_K \Theta \boldsymbol{M}$ and $\mathcal{E} = [\![\epsilon_{j_1,\ldots,j_K}]\!]$ has independent mean-0 sub-Gaussian noise entries, i.e., where

$$\epsilon_{j_1,\ldots,j_K} \sim \text{subG}(\sigma^2_{i_1,\ldots,i_K}).$$

Particularly, if we have Bernoulli noise,

$$\sigma^2_{i_1,\ldots,i_K} = \mathbb{E}\left[\mathcal{A}_{j_1,\ldots,j_K}\right]\left(1 - \mathbb{E}\left[\mathcal{A}_{j_1,\ldots,j_K}\right]\right) \leq \frac{1}{4}.$$

Now we have two-step estimate of $\mathcal{X}$,

$$\hat{\mathcal{X}} = \mathcal{Y} \times_1 \hat{\boldsymbol{U}}_1 \hat{\boldsymbol{U}}_1^T \times_2 \cdots \times_K \hat{\boldsymbol{U}}_K \hat{\boldsymbol{U}}_K^T.$$

, and normalized rows $\hat{\boldsymbol{X}}^s$ via

$$\hat{\boldsymbol{X}}^s_{j:} = \frac{\hat{\boldsymbol{X}}_{j:}}{\left\|\hat{\boldsymbol{X}}_{j:}\right\|_1},\ j \in [p],$$

where $\boldsymbol{X} = \mathcal{M}_1(\hat{\mathcal{X}})$. We propose k-median (1) or k-means (2) with the normalized rows.

$$\sum_{j=1}^{p} \left\|\hat{\boldsymbol{X}}_{j:}\right\|_1 \left\|\hat{\boldsymbol{X}}^s_{j:} - \hat{x}_{z_j^{(0)}}\right\|_1 \leq M \min_{x'_1,\ldots,x'_r,z'} \sum_{j=1}^{p} \left\|\hat{\boldsymbol{X}}_{j:}\right\|_1 \left\|\hat{\boldsymbol{X}}^s_{j:} - x'_{z'_j}\right\| \tag{1}$$

$$\sum_{j=1}^{p} \left\|\hat{\boldsymbol{X}}_{j:}\right\|_F^2 \left\|\hat{\boldsymbol{X}}^s_{j:} - \hat{x}_{z_j^{(0)}}\right\|_F^2 \leq M \min_{x'_1,\ldots,x'_r,z'} \sum_{j=1}^{p} \left\|\hat{\boldsymbol{X}}_{j:}\right\|_F^2 \left\|\hat{\boldsymbol{X}}^s_{j:} - x'_{z'_j}\right\|_F^2 \tag{2}$$

To prove the accuracy, we need three useful lemmas, which are applicable for both F-norm or ℓ_1 norm cases.

Lemma 1 (Singular-value-gap-free tensor estimation error bound)**.** *Let an order-K tensor $\mathcal{A} = \mathcal{X} + \mathcal{Z} \in \mathbb{R}^{p\times\cdots\times p}$, and $\mathcal{X}$ has tucker rank $(r, ...r)$ and $\mathcal{Z}$ has independent sub-Gaussian entries with parameter σ^2. Let $\hat{\mathcal{X}}$ denote the two-step estimated tensor in* (**??**) *. Then with probability at least $1 - C\exp(-cp)$, we have*

$$\left\|\hat{\mathcal{X}} - \mathcal{X}\right\|_F^2 \leq C\sigma^2\left(p^{K/2}r + pr^2 + r^K\right).$$

Lemma 2 (Upper bound for the sum of degree-corrected parameter of misclassified nodes)**.** *Let z be the true assignment in hDCBM model with parameter space $\mathcal{P}(\delta, \alpha_1, \alpha_2, \beta)$ and X_j denote the rows in $\boldsymbol{X} = \mathcal{M}_1(\mathcal{X})$. Given any estimate $\hat{z}, \{\hat{x}_i\}_{i=1}^r \in \mathbb{R}^{p^{K-1}}$, and $\{\hat{X}_j\}_{j=1}^p$ where $\hat{X}_j = \hat{x}_{z_j}$. Then, for any norm $\|\cdot\|$ satisfying the inequality such that*

$$\min_{z_j \neq z_l} \|X_j - X_l\| \geq 2b,$$

with some constant $b > 0$, we have

$$\min_{\pi\in\Pi} \sum_{j:\hat{z}_j\neq\pi(z_j)} \theta_j \leq (2\beta^2 + 1)\sum_{j\in S}\theta,$$

where Π is the permutation space and

$$S = \left\{j \in [p] : \left\|\hat{X}_j - X_j\right\| \geq b\right\}.$$

Lemma 3 (Difference between normalized vectors)**.** *For any two nonzero vectors v_1, v_2 of same dimension, for any norm $\|\cdot\|$, we have*

$$\left\|\frac{v_1}{\|v_1\|} - \frac{v_2}{\|v_2\|}\right\| \leq \frac{2\|v_1 - v_2\|}{\max(\|v_1\|, \|v_2\|)}.$$

The proof idea is following

1. By Lemma (2), find the corresponding b, i.e., the lower bound for

$$\min_{z_j\neq z_l, j,l\in[p]} \left\|\boldsymbol{X}_{j:}^s - \boldsymbol{X}_{l:}^s\right\| \geq 2b,$$

   thereof upper bound the target quantity as

$$\min_{\pi\in\Pi} \sum_{j:\hat{z}_j\neq\pi(z_j)} \theta_j \leq (2\beta^2 + 1)\sum_{j\in S}\theta.$$

   with

$$S = \left\{j \in [p] : \left\|\hat{x}_{z_j^{(0)}} - \boldsymbol{X}_{j:}^s\right\| \geq b\right\}.$$

2. Find the upper bound

$$C\sum_{j\in S}\theta_j \leq \sum_{j\in S}\|\boldsymbol{X}_{j:}\|,$$

   where C relies on p, δ and other parameters.

3. Note that

$$\sum_{j\in S}\|\boldsymbol{X}_{j:}\| \leq \sum_{j\in S}\left\|\hat{\boldsymbol{X}}_{j:}\right\| + \left\|\hat{\boldsymbol{X}}_{j:}-\boldsymbol{X}_{j:}\right\|.$$

The second term can be directly bounded by $\left\|\hat{\mathcal{X}}-\mathcal{X}\right\|$. For the first term, taking the advantage of S, we have

$$\begin{aligned}
\sum_{j\in S}\left\|\hat{\boldsymbol{X}}_{j:}\right\| &\leq \frac{1}{b}\sum_{j\in S}\left\|\hat{\boldsymbol{X}}_{j:}\right\|\left\|\hat{x}_{z_j^{(0)}}-\boldsymbol{X}^s_{j:}\right\| \\
&\leq \frac{1}{b}\sum_{j\in S}\left\|\hat{\boldsymbol{X}}_{j:}\right\|\left[\left\|\hat{x}_{z_j^{(0)}}-\hat{\boldsymbol{X}}^s_{j:}\right\| + \left\|\hat{\boldsymbol{X}}^s_{j:}-\boldsymbol{X}^s_{j:}\right\|\right] \\
&\leq \frac{(1+M)}{b}\sum_{j\in S}\left\|\hat{\boldsymbol{X}}_{j:}\right\|\left\|\hat{\boldsymbol{X}}^s_{j:}-\boldsymbol{X}^s_{j:}\right\| \\
&\leq \frac{(1+M)}{b}\sum_{j\in S}\left\|\hat{\boldsymbol{X}}_{j:}-\boldsymbol{X}_{j:}\right\|
\end{aligned}$$

where the second and third inequality follows by the the update rule (2) or (1) and the Lemma 3, and the last term can be bounded by $\left\|\hat{\mathcal{X}}-\mathcal{X}\right\|$. Then, we finally obtain that

$$\min_{\pi\in\Pi}\sum_{j:\hat{z}_j\neq\pi(z_j)}\theta_j \leq \frac{(2\beta^2+1)}{C}\left(\frac{1+M}{b}+1\right)\left\|\hat{\mathcal{X}}-\mathcal{X}\right\|$$

4. **So the key is to find b, C with different norms.**

2 k-median

In k-median, we use ℓ_1 norm.

1. First, we find the b in this case. WLOG, assume $z_j = a, z_l = b$.

$$\begin{aligned}
\left\|\boldsymbol{X}^s_{j:}-\boldsymbol{X}^s_{l:}\right\|_1 &= \left\|\frac{\boldsymbol{X}_{j:}}{\|\boldsymbol{X}_{j:}\|_1}-\frac{\boldsymbol{X}_{l:}}{\|\boldsymbol{X}_{l:}\|_1}\right\|_1 \\
&\geq \left\|\frac{\boldsymbol{X}_{j:}/\theta_j}{\|\boldsymbol{S}_{a:}\|_1}-\frac{\boldsymbol{X}_{l:}/\theta_l}{\|\boldsymbol{S}_{b:}\|_1}\right\|_1 \min_{j\in[p]}\frac{\left\|\boldsymbol{S}_{z_j:}\right\|_1}{\|\boldsymbol{X}_{j:}\|_1/\theta_j}.
\end{aligned}$$

Note that

$$\|\boldsymbol{X}_{j:}\|_1/\theta_j \leq \left\|\boldsymbol{S}_{z_j:}\right\|_1\left(\frac{\beta p}{r}(1+\delta)\right)^{K-1},$$

and

$$\left\|\frac{\boldsymbol{X}_{j:}/\theta_j}{\|\boldsymbol{S}_{a:}\|_1}-\frac{\boldsymbol{X}_{l:}/\theta_l}{\|\boldsymbol{S}_{b:}\|_1}\right\|_1 \geq \left\|\frac{\boldsymbol{S}_{a:}}{\|\boldsymbol{S}_{a:}\|_1}-\frac{\boldsymbol{S}_{b:}}{\|\boldsymbol{S}_{b:}\|_1}\right\|_1\left(\frac{p}{\beta r}(1-\delta)\right)^{K-1}.$$

The intuition behind these inequality is that for any cluster $(a_1, ..., a_K)$, the degree parameters satisfies

$$\left[\frac{p}{\beta r}(1-\delta)\right]^K \leq \prod_{k=1}^{K} \sum_{j_k : z_{j_k} = a_k} \theta_{j_k} \leq \left[\frac{\beta p}{r}(1+\delta)\right]^K.$$

Therefore, we obtain

$$\min_{z_j \neq z_l, j, l \in [p]} \left\|\boldsymbol{X}_{j:}^s - \boldsymbol{X}_{l:}^s\right\|_1 = \left\|\frac{\boldsymbol{X}_{j:}}{\|\boldsymbol{X}_{j:}\|_1} - \frac{\boldsymbol{X}_{l:}}{\|\boldsymbol{X}_{l:}\|_1}\right\|_1 \geq \Delta_{\min} \frac{(1-\delta)^{K-1}}{(1+\delta)^{K-1}\beta^{2(k-1)}},$$

and thus we let $b = B\Delta_{\min}$ where $B = \frac{(1-\delta)^{K-1}}{2(1+\delta)^{K-1}\beta^{2(k-1)}}$.

2. Note that

$$\sum_{j\in S} \|\boldsymbol{X}_{j:}\|_1 \geq \sum_{j \in S} \theta_j \left\|\boldsymbol{S}_{z_j:}\right\|_1 \left(\frac{p}{\beta r}(1-\delta)\right)^{K-1} \geq \left(\frac{p}{\beta r}(1-\delta)\right)^{K-1} r^{K-1/2}\alpha_1 \sum_{j\in S}\theta_j,$$

where the second inequality follows by the assumption $\|\boldsymbol{S}_{a:}\|_F^2 \geq r^{K-1}\alpha_1^2$. Thus, $C = \left(\frac{p}{\beta r}(1-\delta)\right)^{K-1} r^{K-1/2}\alpha_1$.

3. Therefore, we have

$$\begin{aligned}\min_{\pi \in \Pi} \sum_{j:\hat{z}_j \neq \pi(z_j)} \theta_j &\leq \frac{(2\beta^2+1)}{C}\left(\frac{1+M}{B\Delta_{\min}} + 1\right)\left\|\hat{\mathcal{X}} - \mathcal{X}\right\|_1 \\ &\lesssim \frac{r^{K-1/2}}{p^{K-1}\alpha_1\Delta_{\min}} p^{K/2}\left\|\hat{\mathcal{X}} - \mathcal{X}\right\|_F \\ &\lesssim \frac{r^{K-1/2}}{p^{K/4-1}\alpha_1\Delta_{\min}}.\end{aligned}$$

3 k-means

In k-means, we use F-norm.

1. First, we find the b in this case. WLOG, assume $z_j = a, z_l = b$.

$$\begin{aligned}\left\|\boldsymbol{X}_{j:}^s - \boldsymbol{X}_{l:}^s\right\|_F^2 &= \left\|\frac{\boldsymbol{X}_{j:}}{\|\boldsymbol{X}_{j:}\|_F} - \frac{\boldsymbol{X}_{l:}}{\|\boldsymbol{X}_{l:}\|_F}\right\|_F^2 \\ &\geq \left\|\frac{\boldsymbol{X}_{j:}/\theta_j}{\|\boldsymbol{S}_{a:}\|_F} - \frac{\boldsymbol{X}_{l:}/\theta_l}{\|\boldsymbol{S}_{b:}\|_F}\right\|_F^2 \min_{j\in[p]} \frac{\left\|\boldsymbol{S}_{z_j:}\right\|_F^2}{\|\boldsymbol{X}_{j:}\|_F^2/\theta_j^2}.\end{aligned}$$

Note that

$$\|\boldsymbol{X}_{j:}\|_F^2/\theta_j^2 \leq \left\|\boldsymbol{S}_{z_j:}\right\|_F^2 \left[\frac{\beta p}{r}\theta_{\max}^2\right]^{K-1}.$$

and

$$\left\| \frac{\boldsymbol{X}_{j:}/\theta_j}{\|\boldsymbol{S}_{a:}\|_F} - \frac{\boldsymbol{X}_{l:}/\theta_l}{\|\boldsymbol{S}_{b:}\|_F} \right\|_F^2 \geq \left\| \frac{\boldsymbol{S}_{a:}}{\|\boldsymbol{S}_{a:}\|_F} - \frac{\boldsymbol{S}_{b:}}{\|\boldsymbol{S}_{b:}\|_F} \right\|_F^2 \left[\frac{p}{\beta r}(1-\delta)^2 \right]^{K-1}.$$

The intuition behind these inequality is that for any cluster $(a_1, ..., a_K)$, the degree parameters satisfies

$$\prod_{k=1}^{K} \sum_{j_k: z_{j_k}=a_k} \theta_{j_k}^2 \leq \left[\frac{\beta p}{r} \theta_{\max}^2 \right]^K$$

and

$$\prod_{k=1}^{K} \sum_{j_k: z_{j_k}=a_k} \theta_{j_k}^2 \geq \prod_{k=1}^{K} \frac{1}{p_{ak}} \left[\sum_{j_k: z_{j_k}=a_k} \theta_{j_k} \right]^2 \geq \left[\frac{p}{\beta r}(1-\delta)^2 \right]^K,$$

where the second inequality follows by Cauchy-Schwartz.

Therefore, we obtain

$$\min_{z_j \neq z_l, j, l \in [p]} \left\| \boldsymbol{X}_{j:}^s - \boldsymbol{X}_{l:}^s \right\|_F^2 = \Delta_{\min}^2 \frac{(1-\delta)^{2(K-1)}}{\theta_{\max}^{2(K-1)} \beta^{2(K-1)}}$$

and thus we let $b = B\Delta_{\min}^2$ where $B = \frac{(1-\delta)^{2(K-1)}}{2\theta_{\max}^{2(K-1)}\beta^{2(K-1)}}$. Note that B is a constant that is independent with p if and only if $\theta_{\max} = \mathcal{O}(1+\delta)$.

2. Note that

$$\sum_{j \in S} \|\boldsymbol{X}_{j:}\|_F^2 \geq \sum_{j \in S} \theta_j^2 \|\boldsymbol{S}_{z_j:}\|_F^2 \left[\frac{p}{\beta r}(1-\delta)^2 \right]^{K-1} \geq \frac{1}{p} \left[\sum_{j \in S} \theta_j \right]^2 r^{K-1} \alpha_1^2 \left[\frac{p}{\beta r}(1-\delta)^2 \right]^{K-1},$$

where the second inequality follows by Cauchy-Schwartz, and the assumption $\|\boldsymbol{S}_{a:}\|_F^2 \geq r^{K-1}\alpha_1^2$. Thus, let $C = \frac{p^{K-2}\alpha_1^2}{\beta^{K-1}}(1-\delta)^{2(K-1)}$.

3. Slightly different with k-median procedure, we finally have

$$\begin{aligned} \min_{\pi \in \Pi} \sum_{j: \hat{z}_j \neq \pi(z_j)} \theta_j &\leq (2\beta^2+1) \sqrt{\frac{1}{C} \left(\frac{1+M}{B\Delta_{\min}^2} + 1 \right) \left\| \hat{\mathcal{X}} - \mathcal{X} \right\|_F^2} \\ &\lesssim \frac{1}{p^{K/4-1} \Delta_{\min} \theta_{\max}^{K-1}}. \end{aligned}$$

Remark 1. Note that both k-median and k-means achieves rate $\mathcal{O}(p^{-K/4+1})$. While k-means requires $\theta_{\max}$ independent with p, which is more restrictive than k-median.

References