Graphic Lasso: Self-Consistency

Jiaxin Hu

February 15, 2021

1 Noiseless case

Consider the noiseless case

$$\mathcal{Y} = f(\Theta),$$

where $\Theta = \mathcal{C} \times_1 \boldsymbol{M}_1 \times_2 \cdots \times_K \boldsymbol{M}_K$, and $f(\cdot)$ is an entry-wise link function. Suppose we have the following optimization problem.

$$\max_{\Theta=\mathcal{C}\times_1\boldsymbol{M}_1\times_2\cdots\times_K\boldsymbol{M}_K} \mathcal{L}_{\mathcal{Y}}(\Theta) = \langle \mathcal{Y}, \Theta \rangle - \sum_{i_1,\ldots,i_K} g(\Theta_{i_1,\ldots,i_K}). \tag{1}$$

Lemma 1 (Noiseless estimation)**.** *Let $\{\mathcal{C}, \boldsymbol{M}_k\}$ denote the true parameters and $\{\hat{\mathcal{C}}, \hat{\boldsymbol{M}}_k\}$ are the estimation which maximizes the loss function. Suppose $g(\cdot)$ is a convex function with bounded second derivative $\sup_x g''(x) \leq a$, and $\max_{r_1,\ldots,r_K} |(g')^{-1}(f(c_{r_1,\ldots,r_K}))| \leq C$, where C is a positive constant depends on $\mathcal{C}$. Assume the minimal gap between blocks is strictly larger than 0, i.e., $\delta > 0$. Then, for any $\epsilon > 0$, we have*

$$\mathbb{P}(MCR(\hat{\boldsymbol{M}}_k, \boldsymbol{M}_k) \geq \epsilon) = 0.$$

Proof. We prove the accuracy in following steps.

1. With given membership matrix $\hat{\boldsymbol{M}}_k$, the estimate $\hat{\mathcal{C}}$ is

$$\hat{c}_{r_1,\ldots,r_K}(\hat{\boldsymbol{M}}_k) = (g')^{-1}\left(\frac{1}{\prod_k d_k \prod_k \hat{p}^{(k)}_{r_k}} [f(\mathcal{C}) \times_1 \boldsymbol{M}_1\hat{\boldsymbol{M}}_1^T \times_2 \cdots \times_K \boldsymbol{M}_K\boldsymbol{M}_K^T]_{r_1,\ldots,r_K}\right).$$

   Note that the estimation $\hat{\mathcal{C}}$ depends on $\hat{\boldsymbol{M}}_k$. Therefore, we denote the estimation as $\hat{\mathcal{C}}(\hat{\boldsymbol{M}}_k) = [\![\hat{c}_{r_1,\ldots,r_K}(\hat{\boldsymbol{M}}_k)]\!]$.

2. We define some useful functions. First, we define

$$F(\hat{\boldsymbol{M}}_k) = \mathcal{L}_{\mathcal{Y}}(\hat{\mathcal{C}}(\hat{\boldsymbol{M}}_k), \hat{\boldsymbol{M}}_k) = \sum_{r_1,\ldots,r_K} \prod_k d_k \prod_k \hat{p}^{(k)}_{r_k} h(g'(\hat{c}_{r_1,\ldots,r_K}(\hat{\boldsymbol{M}}_k))),$$

   where $h(x) = x(g')^{-1}(x) - g((g')^{-1}(x))$.

   Note that $\hat{\mathcal{C}}(\hat{\boldsymbol{M}}_k)$ does not include the randomness. Thus, we have $g'(\hat{c}_{r_1,\ldots,r_K}(\hat{\boldsymbol{M}}_k)) = \mathbb{E}\left[g'(\hat{c}_{r_1,\ldots,r_K}(\hat{\boldsymbol{M}}_k))\right]$, and

$$G(\hat{\boldsymbol{M}}_k) = \sum_{r_1,\ldots,r_K} \prod_k d_k \prod_k \hat{p}^{(k)}_{r_k} h(\mathbb{E}\left[g'(\hat{c}_{r_1,\ldots,r_K}(\hat{\boldsymbol{M}}_k))\right]) = F(\hat{\boldsymbol{M}}_k),$$

which implies that there does not exist the estimation error.

Note that for true membership, we have

$$F(\boldsymbol{M}_k) = G(\boldsymbol{M}_k) = \mathcal{L}_{\mathcal{Y}}(\hat{\mathcal{C}}(\boldsymbol{M}_k), \boldsymbol{M}_k),$$

where $\hat{\mathcal{C}}(\boldsymbol{M}_k) = (g')^{-1}(f(\mathcal{C}))$ **is not equal to the true core tensor** $\mathcal{C}$.

3. We only need to consider the classification error. Under the assumptions of the positive minimal gap and the boundedness of the second derivative of g, when $MCR(\hat{\boldsymbol{M}}_k, \boldsymbol{M}_k) \geq \epsilon$ for any $\epsilon > 0$, we have
$$G(\hat{\boldsymbol{M}}_k) - G(\boldsymbol{M}_k) \leq -\frac{\epsilon}{4a}\tau^{K-1}\delta.$$

4. Since $\{\hat{\mathcal{C}}\hat{\boldsymbol{M}}_k, \hat{\boldsymbol{M}}_k\}$ is the maximizer of the loss function, we have
$$0 \leq F(\hat{\boldsymbol{M}}_k) - F(\boldsymbol{M}_k) = G(\hat{\boldsymbol{M}}_k) - G(\boldsymbol{M}_k).$$
Therefore, we obtain that
$$\mathbb{P}(MCR(\hat{\boldsymbol{M}}_k, \boldsymbol{M}_k) \geq \epsilon) = \mathbb{P}(G(\hat{\boldsymbol{M}}_k) - G(\boldsymbol{M}_k) \leq -\frac{\epsilon}{4a}\tau^{K-1}\delta) = 0.$$

□

Remark 1. The lemma 1 implies that the true membership $\boldsymbol{M}_k$ is the maximizer of the function $G(\boldsymbol{M}'_k)$. Due to the noiselessness, $G(\boldsymbol{M}'_k) = \mathcal{L}_{\mathcal{Y}}(\hat{\mathcal{C}}(\boldsymbol{M}'_k), \boldsymbol{M}'_k)$, and $\{\hat{\mathcal{C}}(\boldsymbol{M}_k), \boldsymbol{M}_k\}$ is the maximizer of the noiseless loss function. However, the true parameter $\{\mathcal{C}, \boldsymbol{M}_k\}$ is not the maximizer of the noiseless loss function, since $\hat{\mathcal{C}}(\boldsymbol{M}_k) \neq \mathcal{C}$. Therefore, we conclude that the loss function (1) is **self-consistent to** $\{\hat{\mathcal{C}}(\boldsymbol{M}_k), \boldsymbol{M}_k\}$ but not **self-consistent to** Θ.

Remark 2. Define
$$\hat{\Theta} = \hat{\mathcal{C}}(\boldsymbol{M}_k) \times_1 \boldsymbol{M}_1 \times_1 \cdots \times_K \boldsymbol{M}_K.$$
Then, $\hat{\Theta}$ is an unbiased estimate of Θ if and only if $g' = f$.

Remark 3. Which assumption in the noisy case corresponds to the self-consistency of $\boldsymbol{M}_k$?

Note that in the noisy case, we have

$$\begin{aligned}
G_{noise}(\hat{\boldsymbol{M}}_k) &= \sum_{r_1,\ldots,r_K} \prod_k d_k \prod_k \hat{p}^{(k)}_{r_k} h(\mathbb{E}\left[g'(\hat{c}_{r_1,\ldots,r_K}(\hat{\boldsymbol{M}}_k))\right]) \\
&= \langle f(\mathcal{C}) \times_1 \boldsymbol{M}_1\hat{\boldsymbol{M}}_1^T \times_2 \cdots \times_K \boldsymbol{M}_K\boldsymbol{M}_K^T, (g')^{-1}\left[f(\mathcal{C}) \times_1 \boldsymbol{M}_1\hat{\boldsymbol{M}}_1^T \times_2 \cdots \times_K \boldsymbol{M}_K\boldsymbol{M}_K^T\right]\rangle \\
&\quad - \sum_{i_1,\ldots,i_K} g\left((g')^{-1}\left[f(\mathcal{C}) \times_1 \boldsymbol{M}_1\hat{\boldsymbol{M}}_1^T \times_2 \cdots \times_K \boldsymbol{M}_K\boldsymbol{M}_K^T\right] \times_1 \boldsymbol{M}_1 \times_2 \cdots \times_K \boldsymbol{M}_K\right)_{i_1,\ldots,i_K} \\
&= F_{noiseless}(\hat{\boldsymbol{M}}_k).
\end{aligned}$$

Therefore, we use the self-consistency when we derive the misclassification error. Note that the result that when $MCR(\hat{\boldsymbol{M}}_k, \boldsymbol{M}_k) \geq \epsilon$,

$$G_{noise}(\hat{\boldsymbol{M}}_k) - G_{noise}(\boldsymbol{M}_k) \leq -\frac{\epsilon}{4a}\tau^{K-1}\delta \tag{2}$$

implies the self-consistency of $\boldsymbol{M}_k$. To obtain the result (2), we require

1. the convexity of g and $\sup_x g''(x) \geq a$;
2. minimal gap strictly larger than 0, i.e., $\delta > 0$.