

Graphic Lasso: Self-Consistency

Jiabin Hu

February 14, 2021

1 Noiseless case

Consider the noiseless case

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

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

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

Lemma 1 (Noiseless estimation). *Let $\{\mathcal{C}, \mathbf{M}_k\}$ denote the true parameters and $\{\hat{\mathcal{C}}, \hat{\mathbf{M}}_k\}$ 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 $\max_{r_1, \dots, r_K} |(g')^{-1}(f(c_{r_1, \dots, r_K}))| \leq C$, 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*

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

Proof. We prove the accuracy in following steps.

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

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

Note that the estimation $\hat{\mathcal{C}}$ depends on $\hat{\mathbf{M}}_k$. Therefore, we denote the estimation as $\hat{\mathcal{C}}(\hat{\mathbf{M}}_k) = \llbracket \hat{c}_{r_1, \dots, r_K}(\hat{\mathbf{M}}_k) \rrbracket$.

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

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

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

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

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

Define $\Theta = \mathcal{C} \times \{M_1, \dots, M_K\}$

What estimand does Θ correspond to?

which implies that there does not exist the estimation error.

Note that for true membership, we have

How to explain $\Theta \neq \Theta$ even in the noiseless case?

$$F(M_k) = G(M_k) = \mathcal{L}_Y(\hat{\mathcal{C}}(M_k), M_k),$$

where $\hat{\mathcal{C}}(M_k) = (g')^{-1}(f(\mathcal{C}))$ 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 $MCR(\hat{M}_k, M_k) \geq \epsilon$ for any $\epsilon > 0$, we have

$$G(\hat{M}_k) - G(M_k) \leq -\frac{\epsilon}{4a} \tau^{K-1} \delta.$$

4. Since $\{\hat{\mathcal{C}}\hat{M}_k, \hat{M}_k\}$ is the maximizer of the loss function, we have

$$0 \leq F(\hat{M}_k) - F(M_k) = G(\hat{M}_k) - G(M_k).$$

Therefore, we obtain that

$$\mathbb{P}(MCR(\hat{M}_k, M_k) \geq \epsilon) = \mathbb{P}(G(\hat{M}_k) - G(M_k) \leq -\frac{\epsilon}{4a} \tau^{K-1} \delta) = 0.$$

□

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

In your earlier note (noisy case), which assumption corresponds to self-consistency to M ?