

Graphic Lasso: Miscellaneous

Jiaxin Hu

February 10, 2021

1 Correct

1.1 Definitions

Consider the optimization

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

Some key definitions in the previous notes are not correct. Here is the correctness.

1. With given membership \mathbf{M}_k , the estimation of $\mathcal{C} = \llbracket c_{r_1, \dots, r_K} \rrbracket$ is of form

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

2. We define the function $F(\mathbf{M}_k) = \mathcal{L}_{\mathcal{Y}}(\hat{\mathcal{C}}, \mathbf{M}_k)$, i.e.,

$$\begin{aligned} F(\mathbf{M}_k) &= \langle \mathcal{Y} \times_1 \mathbf{M}_1^T \times_2 \cdots \times_K \mathbf{M}_K^T, \hat{\mathcal{C}} \rangle - \sum_{i_1, \dots, i_K} g(\hat{\mathcal{C}} \times_1 \mathbf{M}_1 \times_2 \cdots \times_K \mathbf{M}_K)_{i_1, \dots, i_K} \\ &\propto \sum_{r_1, \dots, r_K} \prod_k p_{r_k}^{(k)} (g'(\hat{c}_{r_1, \dots, r_K}) \hat{c}_{r_1, \dots, r_K} - g(\hat{c}_{r_1, \dots, r_K})) \\ &= \sum_{r_1, \dots, r_K} \prod_k p_{r_k}^{(k)} h(g'(\hat{c}_{r_1, \dots, r_K})), \end{aligned}$$

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

3. Correspondingly, we define $G(\mathbf{M}_k)$ as

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

1.2 Loss functions

The sample-based loss function is

$$\mathcal{L}_{\mathcal{Y}}(\mathcal{C}', \mathbf{M}'_k) = \langle \mathcal{Y} \times_1 (\mathbf{M}'_1)^T \times_2 \cdots \times_K (\mathbf{M}'_K)^T, \mathcal{C}' \rangle - \sum_{(i_1, \dots, i_K)} g(\mathcal{C}' \times_1 \mathbf{M}'_1 \times_2 \cdots \times_K \mathbf{M}'_K)_{i_1, \dots, i_K}.$$

Correspondingly, the population-based loss function is

$$\begin{aligned} l(\mathcal{C}', \mathbf{M}'_k) &= \mathbb{E}_{\mathcal{Y}}[\mathcal{L}_{\mathcal{Y}}(\mathcal{C}', \mathbf{M}'_k)] \\ &= \langle f(\mathcal{C}) \times_1 D_1^T \times_2 \cdots \times_K D_K^T, \mathcal{C}' \rangle - \sum_{(i_1, \dots, i_K)} g(\mathcal{C}' \times_1 \mathbf{M}'_1 \times_2 \cdots \times_K \mathbf{M}'_k)_{i_1, \dots, i_K}. \end{aligned}$$

Therefore, we know that

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

where $\hat{\mathcal{C}}$ is the estimate depends on \mathbf{M}_k . On the other hand, the explicit form of $G(\mathbf{M}_k)$ is

$$\begin{aligned} G(\hat{\mathbf{M}}_k) &= \sum_{r_1, \dots, r_K} \prod_k p_{r_k}^{(k)} h \left(\frac{1}{\prod_k d_k \prod_k p_{r_k}^{(k)}} f(\mathcal{C}) \times_1 D_1^T \times_2 \cdots \times_K D_K^T \right) \\ &= \langle f(\mathcal{C}) \times_1 D_1^T \times_2 \cdots \times_K D_K^T, \tilde{\mathcal{C}} \rangle - \sum_{(i_1, \dots, i_K)} g(\tilde{\mathcal{C}} \times_1 \hat{\mathbf{M}}_1 \times_2 \cdots \times_K \hat{\mathbf{M}}_k)_{i_1, \dots, i_K}, \\ &= l(\tilde{\mathcal{C}}, \hat{\mathbf{M}}_k). \end{aligned}$$

where

$$\tilde{c}_{r_1, \dots, r_K} = (g')^{-1} \left(\frac{1}{\prod_k d_k \prod_k p_{r_k}^{(k)}} f(\mathcal{C}) \times_1 D_1^T \times_2 \cdots \times_K D_K^T \right)_{r_1, \dots, r_K}.$$

Also, with the true membership $\{\mathbf{M}_k\}$, the previous definition for $G(\mathbf{M}_k)$ is

$$\begin{aligned} G(\mathbf{M}_k) &= \sum_{r_1, \dots, r_K} \prod_k p_{r_k}^{(k)} h(g'(c_{r_1, \dots, r_K})) \\ &= \langle g'(\mathcal{C}) \times_1 \mathbf{M}_1 \times_2 \cdots \times_K \mathbf{M}_K, \mathcal{C} \times_1 \mathbf{M}_1 \times_2 \cdots \times_K \mathbf{M}_K \rangle - \sum_{i_1, \dots, i_K} g(\mathcal{C} \times_1 \mathbf{M}_1 \times_2 \cdots \times_K \mathbf{M}_K)_{i_1, \dots, i_K}, \end{aligned}$$

which is not equal to the $l(\mathcal{C}, \mathbf{M}_k)$ because the first term in the inner product should be $f(\mathcal{C})$.

1.3 Self-consistency (conjecture)

My conjecture is that the upper bound for misclassification rate uses the self-consistency property implicitly.

Suppose we have self-consistency property. We have

$$(\mathcal{C}, \mathbf{M}_k) = \arg \max_{(\mathcal{C}', \mathbf{M}'_k)} l(\mathcal{C}', \mathbf{M}'_k),$$

which implies $f = g'$. Then, based on the definition of $G(\mathbf{M}_k)$, we have $G(\mathbf{M}_k) = l(\mathcal{C}, \mathbf{M}_k)$. Thus, it is natural that

$$G(\hat{\mathbf{M}}_k) - G(\mathbf{M}_k) = l(\tilde{\mathcal{C}}, \hat{\mathbf{M}}_k) - l(\mathcal{C}, \mathbf{M}_k) < 0.$$

If we do not have self-consistency, $G(\cdot)$ is not the population-based function for the true parameter, and thus the true membership $\{\mathbf{M}_k\}$ may not be the maximizer of $G(\cdot)$.

1.4 Functional properties for the loss function(conjecture)

Since $(\hat{\mathcal{C}}, \hat{\mathbf{M}}_k)$ is the maximizer of the sample-based objective function, we have

$$0 < \mathcal{L}_{\mathcal{Y}}(\hat{\mathcal{C}}, \hat{\mathbf{M}}_k) - l(\hat{\mathcal{C}}, \hat{\mathbf{M}}_k) + l(\hat{\mathcal{C}}, \hat{\mathbf{M}}_k) - l(\mathcal{C}, \mathbf{M}_k) + l(\mathcal{C}, \mathbf{M}_k) - \mathcal{L}_{\mathcal{Y}}(\mathcal{C}, \mathbf{M}_k).$$

We are able to obtain the clustering accuracy result with following properties

1. (Self-consistency) Suppose $MCR(\hat{\mathbf{M}}_k, \mathbf{M}_k) \geq \epsilon$, we have

$$G(\hat{\mathbf{M}}_k) - G(\mathbf{M}_k) \leq -C_1\epsilon,$$

where C_1 is a positive constant may related to the minimal gap and the variance of \mathcal{Y} .

2. (Bounded difference between sample- and population-based objective function) Define the probability

$$p(t) = \mathbb{P}(|\mathcal{L}_{\mathcal{Y}}(\mathcal{C}', \mathbf{M}'_k) - l(\mathcal{C}', \mathbf{M}'_k)| \geq t) \rightarrow 0, \quad t \rightarrow \infty.$$

With the above two properties, we have

$$\mathbb{P}(MCR(\hat{\mathbf{M}}_k, \mathbf{M}_k) \geq \epsilon) \leq \mathbb{P}\left(\sup_{\mathcal{C}', \mathbf{M}'_k} |tL_{\mathcal{Y}}(\mathcal{C}', \mathbf{M}'_k) - l(\mathcal{C}', \mathbf{M}'_k)| \geq \frac{C_1\epsilon}{2}\right) = p\left(\frac{C_1\epsilon}{2}\right).$$