Graphic Lasso: Expectation Issue

Jiaxin Hu

January 22, 2021

1 Bernoulli Data

Suppose $\mathcal{Y}$ is a binary tensor, where $\mathcal{Y}_{i_1,\ldots,i_K} \sim Ber(c_{r_1,\ldots,r_K})$ independently.

1.1 Least squared

Suppose the objective function is the least squared function. With given membership $\{\boldsymbol{M}_k\}$, the estimation of core tensor is

$$\hat{c}_{r_1,\ldots,r_K} = \frac{1}{d_1 \ldots d_K p^{(1)}_{r_1} \cdots p^{(K)}_{r_K}} [\mathcal{Y} \times_1 \boldsymbol{M}_1^T \times_2 \cdots \times_K^T \boldsymbol{M}_K]_{r_1,\ldots,r_K}.$$

The function $F(\boldsymbol{M}_k)$ and $G(\boldsymbol{M}_k)$ are defined as following.

$$F(\boldsymbol{M}_k) = \sum_{r_1,\ldots,r_K} \prod_k p^{(k)}_{r_k} \hat{c}^2_{r_1,\ldots,r_K}, \quad G(\boldsymbol{M}_k) = \mathbb{E}[F(\boldsymbol{M}_k)] = \sum_{r_1,\ldots,r_K} \prod_k p^{(k)}_{r_k} \mathbb{E}(\hat{c}^2_{r_1,\ldots,r_K}). \tag{1}$$

Let $\mu_{r_1,\ldots,r_K} = \mathbb{E}[\hat{c}_{r_1,\ldots,r_K}]$. We have

$$\mu_{r_1,\ldots,r_K} = \mathbb{E}[\hat{c}_{r_1,\ldots,r_K}] = \frac{1}{\prod_k p^{(k)}_{r_k}} [\mathcal{C} \times_1 \boldsymbol{D}^{(1),T} \times_2 \cdots \times_K \boldsymbol{D}^{(K),T}]_{r_1,\ldots,r_K}.$$

Note that $\text{Var}(\mathcal{Y}) = V(\mathcal{C}) \times_1 \boldsymbol{M}_1 \times_2 \cdots \times_K \boldsymbol{M}_K$, where $V(c_{r_1,\ldots,r_K}) = c_{r_1,\ldots,r_K}(1 - c_{r_1,\ldots,r_K})$. Therefore, we have

$$\begin{aligned}
\mathbb{E}[\hat{c}^2_{r_1,\ldots,r_K}] &= \text{Var}(\hat{c}_{r_1,\ldots,r_K}) + [\mathbb{E}(\hat{c}_{r_1,\ldots,r_K})]^2 \\
&= \frac{1}{[\prod_k d_k][\prod_k p^{(k)}_{r_k}]^2} [V(\mathcal{C}) \times_1 \boldsymbol{D}^{(1),T} \times_2 \cdots \times_K \boldsymbol{D}^{(K),T}]_{r_1,\ldots,r_K} + \mu^2_{r_1,\ldots,r_K}.
\end{aligned} \tag{2}$$

Plugging the equation (2) into the definition of $G(\boldsymbol{M}_k)$ (1), we have

$$G(\boldsymbol{M}_k) = \sum_{r_1,\ldots,r_K} \prod_k p^{(k)}_{r_k} \mu^2_{r_1,\ldots,r_K} + \sum_{r_1,\ldots,r_K} \frac{1}{[\prod_k d_k][\prod_k p^{(k)}_{r_k}]} [V(\mathcal{C}) \times_1 \boldsymbol{D}^{(1),T} \times_2 \cdots \times_K \boldsymbol{D}^{(K),T}]_{r_1,\ldots,r_K}.$$

Since the estimation of the core tensor $\mathcal{C}$ is related to both the mean of $\mathcal{Y}$ but also the variance of $\mathcal{Y}$, the error for misclassification can be separated into two parts.