Graphic Lasso: Identifiability of Multi-Layer Model

Jiaxin Hu

January 7, 2021

Suppose we have a dataset with p variables and K categories. In multi-layer model, we assume the rank of decomposition r is known, and the precision matrices are of form

$$\Omega^k = \Theta_0 + \sum_{l=1}^{r} u_{lk}\Theta_l, \quad \text{for} \quad k = 1, ..., K. \tag{1}$$

The identifiability problem for $\{\Theta_0, \Theta_1, ..., \Theta_r, \mathbf{u}_1, ..., \mathbf{u}_r\}$ is actually an identifiability problem for tensor decomposition.

Let $\mathcal{Y} \in \mathbb{R}^{p\times p\times K}$ denote the collection of K networks, where $\mathcal{Y}[,,k] = \Omega^k, k \in [K]$. Let $\mathcal{C} \in \mathbb{R}^{p\times p\times(r+1)}$ denote the collection of "core" networks, where $\mathcal{C}[,,1] = \sqrt{K}\Theta_0$, $\mathcal{C}[,,l] = \Theta_{l-1}, l = 2, ..., (r+1)$. Let $\boldsymbol{U} \in \mathbb{R}^{K\times(r+1)} = (\mathbf{u}_0, \mathbf{u}_1, ..., \mathbf{u}_r)$ denote the factor matrix, where $\mathbf{u}_0 = \mathbf{1}_K/\sqrt{K}$. Rewrite the model (1) in tensor form.

$$\mathcal{Y} = \mathcal{C} \times_3 \boldsymbol{U}. \tag{2}$$

Therefore, the identifiability problem for $\{\Theta_l, \mathbf{u}_l\}$ becomes the identifiability problem for $\{\mathcal{C}, \boldsymbol{U}\}$. Let Unfold$(\cdot)$ denote the unfold representation of a tensor on mode 3. The model (2) is equal to the following matrix factorization model

$$\text{Unfold}(\mathcal{Y}) = \boldsymbol{U}\text{Unfold}(\mathcal{C}).$$

1 No sparsity constrain of U

1.1 Add constrain for distinct singular values

By SVD, we have decomposition $\text{Unfold}(\mathcal{Y}) = \tilde{\boldsymbol{U}}\Sigma\tilde{\boldsymbol{V}}$. If Unfold$(\mathcal{Y})$ has $(r+1)$ distinct singular values, the SVD is unique up to orthogonal rotation by letting the diagonal elements of Σ be in descending order. Further, since the first column of $\boldsymbol{U}$ should be $\mathbf{1}_K/\sqrt{K}$, the orthogonal rotation from $\tilde{\boldsymbol{U}}$ to $\boldsymbol{U}$ is unique, and then $\boldsymbol{U}$ and $\mathcal{C}$ are identifiable.

Next, we discuss the constrain to let Unfold$(\mathcal{Y})$ has $(r+1)$ distinct singular values. For simplicity, let $Y = \text{Unfold}(\mathcal{Y})$. Rewrite Y. We have

$$Y = \begin{pmatrix} \text{vec}^T(\Omega^1) \\ \vdots \\ \text{vec}^T(\Omega^K) \end{pmatrix}, \quad \sigma_i(Y) = \text{eigen}_i\left\{YY^T\right\} = \text{eigen}_i \begin{pmatrix} \left\|\Omega^1\right\|_F^2 & \langle\Omega^1,\Omega^2\rangle & \cdots & \langle\Omega^1,\Omega^K\rangle \\ \langle\Omega^1,\Omega^2\rangle & \left\|\Omega^2\right\|_F^2 & \cdots & \langle\Omega^2,\Omega^K\rangle \\ \vdots & \vdots & \vdots & \vdots \\ \langle\Omega^1,\Omega^K\rangle & \langle\Omega^2,\Omega^K\rangle & \cdots & \left\|\Omega^K\right\|_F^2 \end{pmatrix},$$

for $i = 1$,

1.2 Add sparsity constrain of Θ_ls

2 Membership constrain of U

2.1 Hard membership constrain

We assume $\boldsymbol{U}$ is a hard membership matrix, i.e., for each row of $\boldsymbol{U}$ there is only 1 copy of 1 and massive 0.

2.1.1 Irreducible condition

Without intercept, by the Proposition 1 in Wang et al, $\{\boldsymbol{U}, \Theta_l\}$ are identifiable if $\mathcal{C}$ is irreducible on mode 3. The irreducibility is equivalent to saying that $\Theta_k \neq \Theta_l$, for all $k \neq l, k, l \in [r]$.

2.1.2 Add intercept Θ_0

Suppose we still want to cluster K categories into r groups. We need to add one of following constrains.

1. **Set to 0 constrain**

   Let $\tilde{\Theta}_1 = 0$. Consider the new parameters $\{\tilde{\boldsymbol{U}}, \tilde{\Theta}_0, \tilde{\Theta}_2, ..., \tilde{\Theta}_r\}$. The model (1) becomes

   $$\Omega^k = \tilde{\Theta}_0 + \tilde{\Theta}_{i_k}, \quad \text{for} \quad k = 1, ..., K.$$

   Then, $\tilde{\boldsymbol{U}}$ is identifiable by replacing the first column of $\boldsymbol{U}$ as $\mathbf{1}_K$. The collection of $\{\Theta_l\}$ are also identifiable by replacing $\tilde{\Theta}_0 = \Theta_1$, and $\tilde{\Theta}_l = \Theta_l - \Theta_1, l = 2, ..., r$.

2. **Sum to 0 constrain**

   To let $\{\tilde{\boldsymbol{U}}, \tilde{\Theta}_0, \tilde{\Theta}_1, ..., \tilde{\Theta}_r\}$ identifiable, we need one of the following constrains

   $$\sum_{l=1}^{r} \Theta_l = 0, \tag{3}$$

   or

   $$\sum_{l=1}^{r} m_l \Theta_l = 0, \tag{4}$$

   where $m_l = |\mathbf{u}_l|, l = 1, ..., r$. In some sense, the weighted sum to 0 constrain (4) is better because Θ_0 at this point is the average network of K categories; however, constrain (4) would be more difficult since we do not know the true value of m_l. In contrast, sum to 0 constrain (3) is more computationally convenient, and Θ_0 is the average network of $\{\Theta_l\}$ rather than the average of K categories.

2.2 Mixed membership constrain

We assume $\boldsymbol{U}$ is a mixed membership matrix, i.e., $\sum_{l=1}^{r} u_{kl} = 1, u_{kl} \geq 0$, for all $k \in [K]$.

First, we assume there is no intercept Θ_0. Note that mixed membership is a weaker constrain than hard membership. Therefore, the irreducible assumption on $\{\Theta_l\}$ is necessary.

3 Consistency

3.1 Magnitude constrain

3.2 Transfer the error from Ω^k to $\{\mathbf{u}_l, \Theta_l\}$