Graphic Lasso: Membership Constrain for Multi-layer model

Jiaxin Hu

January 10, 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

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

Let $\mathbf{u}_l = (u_{1l}, ..., u_{Kl}) \in \mathbb{R}^K, l \in [r]$, and $\boldsymbol{U} = [\mathbf{u}_1, ..., \mathbf{u}_r] \in \mathbb{R}^{K\times r}$.

1 Option 2 (Without Intercept)

Let $\boldsymbol{U}$ be positive membership matrix, i.e., for each of $\boldsymbol{U}$ there is only 1 copy of 1 and others remain 0.

Proposition 1. *Assume $\boldsymbol{U}$ is a positive membership matrix, and $\{\Theta_l\}$ are irreducible in the sense that $\Theta_k \neq \Theta_l$, for all $k \neq l$. Then, the membership matrix $\boldsymbol{U}$ and $\{\Theta_l\}$ are identifiable up to permutation.*

Proof. Since $\boldsymbol{U}$ is a positive membership matrix, the model (1) becomes

$$\Omega^k = \Theta_{l(k)}, \quad k = 1, ..., K, \tag{2}$$

where $l(k) \in [r]$ is the group assignment of the k-th category. Let $I_l = \{k | u_{kl} > 0, k \in [K]\}, k \in [r]$.

We prove the identifiability of $(\boldsymbol{U}, \{\Theta_l\})$ by contradiction. Suppose there exist parameters $(\tilde{\boldsymbol{U}}, \{\tilde{\Theta}_l\})$ also satisfy the equation (2) with corresponding $\tilde{I}_l$. Since $\tilde{\boldsymbol{U}}$ is not a permutation of $\boldsymbol{U}$, there exist l, k_1, k_2 such that $\tilde{I}_l \cap I_{k_1} \neq \emptyset$ and $\tilde{I}_l \cap I_{k_2} \neq \emptyset$, where $k_1 \neq k_2$. Without the loss of generality, let $l = 1, k_1 = 1, k_2 = 2$. Then, there exists categories $i_0 \in \tilde{I}_1 \cap I_1$ and $i_1 \in \tilde{I}_1 \cap I_2$. Then, we have

$$\tilde{\Theta}_1 = \Theta_1 = \Theta_2,$$

which contradicts to the irreducibility of Θ_1 and Θ_2. Therefore, the parameters $(\boldsymbol{U}, \{\Theta_l\})$ are identifiable. □

2 Option 3

Let each column of $\boldsymbol{U}$ be non-overlapped and sum to 0, and has a unit norm. However, this non-overlap condition is not sufficient for identifiability if we only have irreducibility for $\{\Theta_l\}$.

Statement: If $\Theta_k = C\Theta_l$ where $C \neq 0, 1$, then group k, l can be merged as new group k_0 with a new vector $\tilde{\mathbf{u}}_{k_0}$ satisfies the non-overlap constraints. Further, a group k can be separated to two new groups k_1, k_2 with $\Theta_{k_1} = C\Theta_{k_2}$ for some constant $C \neq 1$. The two new vectors $\tilde{\mathbf{u}}_{k_1}$ and $\tilde{\mathbf{u}}_{k_2}$ also satisfy the non-overlap constraints. Then, $\{\boldsymbol{U}, \{\Theta_l\}\}$ are not identifiable.

Counter Example: Consider the parameters

$$\Theta_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \Theta_2 = \begin{pmatrix} 100 & 100 \\ 0 & 100 \end{pmatrix}, \quad \Theta_3 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

and

$$\boldsymbol{U} = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & 0 \\ -\frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{6}} \\ 0 & 0 & -\frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{1}{\sqrt{3}} \\ 0 & 0 & -\frac{1}{\sqrt{3}} \end{pmatrix} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{pmatrix}.$$

To merge group 1 and 2, let

$$\mathbf{u}_1\Theta_1 = \mathbf{u'}_1\tilde{\Theta}_1, \quad \mathbf{u}_2\Theta_2 = \mathbf{u'}_2\tilde{\Theta}_1.$$

Assume $a\Theta_1 = \tilde{\Theta}_1$. Then, we have $\mathbf{u'}_1 = \frac{1}{a}\mathbf{u}_1$ and $\mathbf{u'}_2 = \frac{100}{a}\mathbf{u}_2$. Consider the new vector $\tilde{\mathbf{u}}_1 = \mathbf{u'}_1 + \mathbf{u'}_2$. Since $\sum_{k=1}^K u_{k1} = \sum_{k=1}^K u_{k2} = 0$ and $\mathbf{u}_1, \mathbf{u}_2$ are non-overlapped, we have $\sum_{k=1}^K u'_{k1} + u'_{k2} = 0$. Besides, let a satisfy the following equation

$$\left\|\mathbf{u'}_1 + \mathbf{u'}_2\right\|_2^2 = \frac{1}{a^2} + \frac{100^2}{a^2} = 1.$$

We have $a = \sqrt{1 + 100^2}$, and then $\tilde{\mathbf{u}}_1$ satisfies the non-overlapped constraint.

To separate the group 3, let $\mathbf{u}_3^1 = (0, 0, 0, 0, \frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, 0, 0)^T$ and $\mathbf{u}_3^2 = (0, 0, 0, 0, 0, 0, \frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}})^T$. Let

$$\tilde{\mathbf{u}}_2 = \sqrt{3}\mathbf{u}_3^1, \quad \tilde{\Theta}_2 = \begin{pmatrix} \frac{1}{\sqrt{3}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix} \quad \text{and} \quad \tilde{\mathbf{u}}_3 = \frac{3}{\sqrt{6}}\mathbf{u}_3^2, \quad \tilde{\Theta}_3 = \begin{pmatrix} \frac{\sqrt{6}}{3} & 0 \\ \frac{\sqrt{6}}{3} & \frac{\sqrt{6}}{3} \end{pmatrix}.$$

Therefore, the new parameters

$$\tilde{\Theta}_1 = \begin{pmatrix} a & a \\ 0 & a \end{pmatrix}, \quad \tilde{\Theta}_2 = \begin{pmatrix} \frac{1}{\sqrt{3}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix}, \quad \tilde{\Theta}_3 = \begin{pmatrix} \frac{\sqrt{6}}{3} & 0 \\ \frac{\sqrt{6}}{3} & \frac{\sqrt{6}}{3} \end{pmatrix},$$

and

$$\tilde{\boldsymbol{U}} = (\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2, \tilde{\mathbf{u}}_3) = \begin{pmatrix} \frac{1}{\sqrt{2}a} & 0 & 0 \\ -\frac{1}{\sqrt{2}a} & 0 & 0 \\ \frac{100}{\sqrt{2}a} & 0 & 0 \\ -\frac{100}{\sqrt{2}a} & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \\ 0 & 0 & -\frac{1}{\sqrt{2}} \end{pmatrix}$$

also satisfy the model.

Proposition 2. *Assume each column of $\boldsymbol{U}$ be non-overlapped and sum to 0, and has a unit norm. If $\{\Theta_l\}$ are "irreducible" in the sense that $\Theta_l \neq C\Theta_l$, for all $k \neq l, C \in \mathbb{R}$. Then, $\{\boldsymbol{U}, \{\Theta_l\}\}$ are identifiable.*

Proof. Since the columns of $\boldsymbol{U}$ are non-overlapped, the model (1) becomes

$$\Omega^k = u_{k,l(k)}\Theta_{l(k)}, \quad k = 1, ..., K, \tag{3}$$

where $l(k) \in [r]$ is the group assignment of the k-th category. Let $I_l = \{k|u_{kl} \neq 0, k \in [K]\}, k \in [r]$ where $\cup_{l=1}^r I_l = K$, and $\mathbf{u}_l = (u_{1l}, ..., u_{u_{kl}})$ denote the column of $\boldsymbol{U}$.

Suppose there exist parameters $(\tilde{\boldsymbol{U}}, \{\tilde{\Theta}_l\})$ also satisfy the equation (3) with corresponding $\tilde{I}_l$. We prove that $(\boldsymbol{U}, \{\Theta_l\})$ are identifiable up to the sign for each column.

1. Suppose $I_l = \tilde{I}_l$ for all $l = 1, ..., r$. Take an arbitrary $l \in [r]$. For all $k \in I_l$, we have

   $$u_{kl}\Theta_l = \tilde{u}_{kl}\tilde{\Theta}_l,$$

   which implies $\tilde{\mathbf{u}}_l = \frac{\Theta_l}{\tilde{\Theta}_l}\mathbf{u}_l$. Note that $\|\mathbf{u}_l\|_2 = \|\tilde{\mathbf{u}}_l\|_2 = 1$. Then, we obtain that $\Theta_l = \tilde{\Theta}_l$ or $\Theta_l = -\tilde{\Theta}_l$, and thus $\tilde{\mathbf{u}}_l = \mathbf{u}_l$ or $\tilde{\mathbf{u}}_l = -\mathbf{u}_l$, for all $l \in [r]$.

2. Suppose not all $\tilde{I}_l \neq I_l$ and $\tilde{I}_l \neq I_k, k \neq l, k \in [r]$.

   Without the loss of generality, suppose $k_0 \in I_l \cap \tilde{I}_1$ and $k_1 \in I_k \cap \tilde{I}_1$, for $k \neq l$. Then, we have

   $$u_{k_0 l}\Theta_l = \tilde{u}_{k_0 1}\tilde{\Theta}_1, \quad u_{k_1 l}\Theta_k = \tilde{u}_{k_1 1}\tilde{\Theta}_1,$$

   which implies there exists a constant C such that $\Theta_l = C\Theta_k$. This result contradicts to the "irreducible" assumption for $\{\Theta_l\}$. Then, we go back to case 1.

□