Graphic Lasso: Clustering accuracy for precision matrix model

Jiaxin Hu

January 24, 2021

1 Accuracy

Consider the optimization problem.

$$
\begin{aligned}
\min_{\boldsymbol{U},\{\Theta^l\}} \quad & Q(\boldsymbol{U},\{\Theta^l\}) = \sum_{k=1}^{K} \operatorname{tr}(S^k\Omega^k) - \log|\Omega^k|, \\
s.t. \quad & \Omega^k = \sum_{l=1}^{r} u_{kl}\Theta^l, \\
& \boldsymbol{U} \in \{0,1\}^{k\times r} \text{ is a membership matrix}, \\
& \{\Theta^l\} \text{ are irreducible and invertible.}
\end{aligned}
$$

Proof. Recall the objective function

$$
Q(\boldsymbol{U},\{\Theta^l\}) = \sum_{k=1}^{K} \operatorname{tr}(S^k\Omega^k) - \log|\Omega^k| = \sum_{k=1}^{K} \langle S^k,\Omega^k\rangle - \log|\Omega^k|.
$$

The deviation between the true parameters $\{\boldsymbol{U},\{\Theta^l\}\}$ and estimations $\{\hat{\boldsymbol{U}},\{\hat{\Theta}^l\}\}$ comes from two aspects: the estimation of $\{\Theta^l\}$ and the misclassification (estimation of $\boldsymbol{U}$). We tease apart these two parts.

1. First, we suppose the membership $\boldsymbol{U}$ is given. We now assess the stochastic error due to the estimation of $\{\Theta^l\}$, conditional on $\boldsymbol{U}$. Define the index sets $I_l = \{k : u_{kl} = 1\}$ for $l = 1,...,r$. Note that the objective function is convex. The optimal $\{\Theta^l\}$ satisfies the first order condition.

$$
\frac{\partial Q}{\partial \Theta^l_{ij}} = \sum_{k\in I_l} S^k_{ij} - |I_l|\frac{C(\Theta^l)_{ij}}{|\Theta^l|} = 0,
$$

   where $C(A)_{ij}$ is the cofactor of matrix A corresponding to the element A_{ij}. Note that Θ^l should be a symmetric matrix. The cofactor matrix $C(\Theta^l) = C^T(\Theta^l)$. Then, the derivation of the matrix Θ^l is equal to

$$
\frac{\partial Q}{\partial \Theta^l} = \sum_{k\in I_l} S^k - |I_l|\frac{C^T(\Theta^l)}{|\Theta^l|} = 0,
$$

   which implies that

$$
\hat{\Theta}^l = \left(\frac{\sum_{k\in I_l} S^k}{|I_l|}\right)^{-1}, \quad \text{for} \quad l \in [r].
$$

Therefore, the estimation of $\{\Theta^l\}$ is a function of $\boldsymbol{U}$. Consider the function $F(\boldsymbol{U}) = Q(\boldsymbol{U}, \{\hat{\Theta}^l\})$. By a straightforward calculation, we have

$$
\begin{aligned}
F(\boldsymbol{U}) &= \sum_{l=1}^{r} \sum_{k \in I_l} \langle S^k, \left(\frac{\sum_{k \in I_l} S^k}{|I_l|} \right)^{-1} \rangle - |I_l| \log \left| \left(\frac{\sum_{k \in I_l} S^k}{|I_l|} \right)^{-1} \right| \\
&= \sum_{l=1}^{r} \langle \sum_{k \in I_l} S^k, \left(\frac{\sum_{k \in I_l} S^k}{|I_l|} \right)^{-1} \rangle - |I_l| \log \left| \left(\frac{\sum_{k \in I_l} S^k}{|I_l|} \right)^{-1} \right| \\
&= \sum_{l=1}^{r} |I_l| p - |I_l| \log \left| \left(\frac{\sum_{k \in I_l} S^k}{|I_l|} \right)^{-1} \right|.
\end{aligned}
$$

Note that $\sum_{l=1}^{r} |I_l| p = Kp$ is independent with the membership. We only need to consider the second term. For simplicity, we define

$$
F(\boldsymbol{U}) = \sum_{l=1}^{r} |I_l| \log \left| \left(\frac{\sum_{k \in I_l} S^k}{|I_l|} \right)^{-1} \right|.
$$

Correspondingly, we define the population version of $F(\boldsymbol{U})$ as following.

$$
G(\boldsymbol{U}) = \sum_{l=1}^{r} |I_l| \log \left| \left(\frac{\sum_{k \in I_l} \Sigma^k}{|I_l|} \right)^{-1} \right|,
$$

where $\Sigma^k = \mathbb{E}[S^k]$ is the true covariance matrices. Therefore, the deviation $F(\boldsymbol{U}) - G(\boldsymbol{U})$ quantifies the stochastic error due to the estimation of $\{\Theta^l\}$.

2. Next, we free $\boldsymbol{U}$ and quantify the total deviation. Considering the maximizer,

$$
\hat{\boldsymbol{U}} = \arg\max_{\boldsymbol{U}} F(\boldsymbol{U}).
$$

The corresponding $G(\hat{\boldsymbol{U}})$ is

$$
G(\hat{\boldsymbol{U}}) = \sum_{l=1}^{r} |\hat{I}_l| \log \left| \left(\frac{\sum_{k \in \hat{I}_l} \Sigma^k}{|\hat{I}_l|} \right)^{-1} \right|,
$$

and the function $G(\boldsymbol{U})$ with true membership is

$$
G(\boldsymbol{U}) = \sum_{l=1}^{r} |I_l| \log \left| \left(\frac{\sum_{k \in I_l} \Sigma^k}{|I_l|} \right)^{-1} \right| = \sum_{l=1}^{r} |I_l| \log |\Theta^l|.
$$

Then, the deviation $G(\hat{\boldsymbol{U}}) - G(\boldsymbol{U})$ measures the stochastic error of the misclassification.

□

Lemma 1. *Suppose the misclassification rate* $MCR(\hat{\boldsymbol{U}}, \boldsymbol{U}) = \max_{l, a \neq a' \in [r]} \min\{D_{al}, D(a'l)\} \geq \epsilon$.

$$
G(\hat{\boldsymbol{U}}) - G(\boldsymbol{U}) \leq
$$

Proof. Recall the definition of confusion matrix $D_{ll'} = \sum_{k=1}^{K} \boldsymbol{I}\{u_{kl} = 1, \hat{u}_{kl'} = 1\}$. Since $MCR(\hat{\boldsymbol{U}}, \boldsymbol{U}) \geq \epsilon$, without the loss of generality, let D_{11} be the largest element in the first column of D, D_{12} be the second largest element in the column, and $D_{21} \geq \epsilon$. □