Q & A about novelty 2

Jiaxin Hu

May 25, 2022

1 Models

In previous note, we have two representations of the correlated Gaussian model (GM). Here, we unify the representation by introduce a scaling parameter $c > 0$ and recall the GMs in 1-d, 2-d, and m-d cases with $m \geq 3$. A simplified QA section about novelty is based on the unify models.

1.1 1-d correlated Gaussian model (1-d GM)

Consider the feature matrices $\boldsymbol{A} \in \mathbb{R}^{n\times N}$, $\boldsymbol{B} \in \mathbb{R}^{n\times N}$, and the true row permutation $\pi^* : [n] \mapsto [n]$ with corresponding permutation matrix $\Pi^* \in \mathbb{R}^{n\times n}$. The matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ follows the 1-d GM with parameters (π^*, ρ^*, c) for $\rho^* \in [0, 1]$ and $c > 0$ if

$$\boldsymbol{A}_{ij} \sim_{i.i.d.} N(0, 1), \text{ for all } i \in [n], j \in [N], \quad c\boldsymbol{B} = \Pi^*\boldsymbol{A} + \sigma\boldsymbol{Z}, \tag{1}$$

where $\sigma = \frac{\sqrt{1-\rho^*}}{\rho^*}$, $\boldsymbol{Z}_{ij} \sim_{i.i.d.} N(0, 1)$ and $\boldsymbol{A}_{ij} \perp \boldsymbol{Z}_{ij}$ for all $i \in [n], j \in [N]$. The parameter ρ^* is the correlation coefficient between variables $\boldsymbol{A}_{ij}$ and $\boldsymbol{B}_{\pi^*(i)j}$. *Note that ρ^* is not the covariance.*

1.2 2-d correlated Gaussian model (2-d GM)

Consider the adjacency matrices $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{B} \in \mathbb{R}^{n\times n}$, and the true row permutation $\pi^* : [n] \mapsto [n]$ with corresponding permutation matrix $\Pi^* \in \mathbb{R}^{n\times n}$. The matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ follows the 2-d non-symmetric GM with parameters (π^*, ρ^*, c) for $\rho^* \in [0, 1]$ and $c > 0$ if

$$\boldsymbol{A}_{ij} \sim_{i.i.d.} N(0, 1), \text{ for all } i \in [n], j \in [n], \quad c\boldsymbol{B} = \Pi^*\boldsymbol{A}(\Pi^*)^T + \sigma\boldsymbol{Z}, \tag{2}$$

where $\sigma = \frac{\sqrt{1-\rho^*}}{\rho^*}$, $\boldsymbol{Z}_{ij} \sim_{i.i.d.} N(0, 1)$ and $\boldsymbol{A}_{ij} \perp \boldsymbol{Z}_{ij}$ for all $i, j \in [n]$. The parameter ρ^* is the correlation coefficient between variables $\boldsymbol{A}_{ij}$ and $\boldsymbol{B}_{\pi^*(i)\pi^*(j)}$.

The symmetric GM assumes $\boldsymbol{A}$ and $\boldsymbol{Z}$ are symmetric and only the entries in the lower-triangle of the matrices are i.i.d. distributed; i.e., $\boldsymbol{A}_{ij} \sim_{i.i.d.} N(0, 1), \boldsymbol{Z}_{ij} \sim_{i.i.d.} N(0, 1)$ for all $1 \leq i \leq j \leq n$.

1.3 m-d ($m \geq 3$) correlated Gaussian model (m-d GM)

Consider the order-m tensors $\mathcal{A} \in \mathbb{R}^{n^{\otimes m}}$ and $\mathcal{B} \in\in \mathbb{R}^{n^{\otimes m}}$, and the true row permutation $\pi^* : [n] \mapsto [n]$ with corresponding permutation matrix $\Pi^* \in \mathbb{R}^{n \times n}$. The tensors $\mathcal{A}$ and $\mathcal{B}$ follows the non-symmetric m-d GM with parameters (π^*, ρ^*, c) for $\rho^* \in [0, 1]$ and $c > 0$ if

$$\mathcal{A}_{\boldsymbol{\omega}} \sim_{i.i.d.} N(0,1), \text{ for all } \boldsymbol{\omega} \in [n]^m, \quad c\mathcal{B} = \mathcal{A} \times_1 \Pi^* \times_2 \cdots \times_m \Pi^* + \sigma \mathcal{Z}, \tag{3}$$

where $\sigma = \frac{\sqrt{1-\rho^*}}{\rho^*}$, $\mathcal{Z}_{\boldsymbol{\omega}} \sim_{i.i.d.} N(0,1)$ and $\mathcal{A}_{\boldsymbol{\omega}} \perp \mathcal{Z}_{\boldsymbol{\omega}}$ for all $\boldsymbol{\omega} \in [n]^m$. For the indices $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_m) \in [n]^m$ and permutation π on $[n]$, let $\pi \circ \boldsymbol{\omega} = (\pi(\omega_1), \ldots, \pi(\omega_m))$ denote the permuted indices. The parameter ρ^* is the correlation coefficient between variables $\mathcal{A}_{\boldsymbol{\omega}}$ and $\mathcal{B}_{\pi^* \circ \boldsymbol{\omega}}$.

Remark 1 (Special cases)**.** Let $c = 1$. The 2-d GM (2) is equal to the model (20) in Fan et al. (2019). Let $c = \frac{1}{\rho^*}$. The 2-d GM (2) is equal to the model in Ding et al. (2021).

Remark 2 (Notation)**.** In previous note, we use $\pi^* \circ \mathcal{A}$ for $\mathcal{A} \times_1 \Pi^* \times_2 \cdots \times_m \Pi^*$ for simplicity. The order of composition in $\pi^* \circ \mathcal{A}$ is wrong, and we should use $\mathcal{A} \circ \pi^*$ instead. Because the tensor $\mathcal{A}$ maps the index to the real field and the permutation π^* maps index to index.

2 Q&A

2.1 What is the difference between spectral and degree profile algorithm?

The most significant difference between degree profile algorithm (DA) and spectral algorithm (SA) is: DA uses the original observations while SA uses the singular space of observations.

The DA compares the empirical distributions of rows $\boldsymbol{A}_{i:}$ and $\boldsymbol{B}_{k:}$ to match the nodes i and k in $\boldsymbol{A}, \boldsymbol{B}$, respectively. The component correspondence between $\boldsymbol{A}_{i:}$ and $\boldsymbol{B}_{k:}$ is known in 1-d case but unknown in 2-d and m-d cases, due to the permutation on the multiple modes. Therefore, DA is able to use more information in 1-d. In addition, when we assume symmetry in 2-d and m-d, DA should be modified to address the entry dependence issue brought from the symmetry.

In contrast, the SA compares the singular space on the first mode of $\boldsymbol{A}$ and $\boldsymbol{B}$ to reveal the permutation, instead of using the original observations. The singular space on a particular mode is not affected by the permutation on other modes. Therefore, SA works in the same way regardless the order of the observations, though the spectral techniques used for different orders are different. Also, with symmetry assumption, SA does not suffer the entry dependence issue because the matrix and tensor decomposition is symmetric with symmetric input and thereof address the symmetry automatically.

2.2 What is the difference between 1-d and 2-d matching problem? Where does the matrix-specific technique come in?

From the perspective of model formulation, the most significant difference between 1-d matching (1) and 2-d matching (2) is the multiway permutation issue. The multiway permutation issue brings difficulty for the method using the observations directly, e.g., DA is more difficult in 2-d than 1-d case.

From the perspective of problem complexity, 1-d and 2-d matching problems share a similar complexity when the number of features N is of the same order of n; i.e., $N \asymp n$. The number of free parameters in 1-d and 2-d (even in m-d) are the same, which is the number of free parameters in the permutation π^*. The sample size of 1-d and 2-d problem are nN and n^2, respectively. Therefore, the ratios (#free parameter)/(#sample size) for 1-d and 2-d matching are of the same order when $N \asymp n$.

The same matrix technique is employed in SA for 1-d and 2-d cases. The DA is modified to address the multiway permutation issue from 1-d to 2-d, but there is no significant difference between the techniques for 2-d and m-d.

2.3 Relationship between matching versus regression and clustering problems.

The 1-d model (1) is most related to the linear regression model with random effects. Specifically, the variable $c\boldsymbol{B}$ serves as response, $\boldsymbol{A}$ serves as random effect, and Π^* serves as the design matrix. However, our goal is to estimate the design matrix Π^* rather than the variance of random effect $\boldsymbol{A}$. For higher order cases, the matching problem closely related to the tensor-on-tensor regression. To illustrate, we rewrite the m-d model (3) in a marginal form. For an arbitrary indices $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_m)$, we have

$$c\mathcal{B}_{\boldsymbol{\omega}} = \left\langle \mathcal{A}, \Pi^*_{:\omega_1} \otimes \Pi^*_{:\omega_2} \otimes \cdots \otimes \Pi^*_{:\omega_m} \right\rangle + \sigma \mathcal{Z}_{\boldsymbol{\omega}}, \tag{4}$$

where $\Pi^*_{:\omega}$ is the ω-th column of Π^* and $\otimes$ refers to the outer product between column vectors. The model (4) is the tensor-on-tensor regression and our goal is to estimate the signal (predictor) tensor $\Pi^*_{:\omega_1} \otimes \Pi^*_{:\omega_2} \otimes \cdots \otimes \Pi^*_{:\omega_m}$ for all $\boldsymbol{\omega} \in [n]^m$ with given $\mathcal{A}, \mathcal{B}$.

The matching problem also relates to the matching problem by viewing Π^* as the membership matrix for n clusters and $\boldsymbol{A}$ as the known core collecting the cluster means. Unlike traditional clustering problems with cluster size assumption, matching problem assumes that there is only 1 node for each cluster.

2.4 Scaling issue.

The SA in Fan et al. (2019) is proposed with 2-d GM (2) with $c = 1$, and the DA in Ding et al. (2021) is proposed with 2-d GM (2) with $c = 1/\rho^*$. The scaling issue concerns the invariance of SA and DA results with input $(\boldsymbol{A}, \boldsymbol{B})$ when c varies. The SA is invariant to the scaling but needs to tune the bandwidth parameter even under the noiseless case; the DA is not invariant to the scaling issue. See Q4 in note 0523 for examples.

In the view of linear regression, in 1-d model (1), only the parameter $c^{-1}\Pi^*$ is identifiable. Note that Π^* only has entries with value 0 and 1. So, we are able to recover Π^* from $c^{-1}\Pi^*$ when c is unknown. Similarly, in higher order models, the ideal algorithm should be able to estimate $c^{-1}\Pi^*$ and recover Π^* when c is an unknown constant.

2.5 How to highlight the novelty for DA in tensor matching?

Currently, I believe the biggest novelty for m-d DA lies in the symmetry concern. Under the symmetry assumption, the MLE phase transition is more difficult than 2-d case, and the entry

dependence issue is more complicate when establishing the algorithm guarantee.

References

Ding, J., Ma, Z., Wu, Y., and Xu, J. (2021). Efficient random graph matching via degree profiles. *Probability Theory and Related Fields*, 179(1):29–115.

Fan, Z., Mao, C., Wu, Y., and Xu, J. (2019). Spectral graph matching and regularized quadratic relaxations i: The gaussian model. *arXiv preprint arXiv:1907.08880*.