Compare two semi-supervised tensor regression model with covariate matrix

Jiaxin Hu

Email: jhu267@wisc.edu

Date: 2019.4.4

1 Show the equivalence between two models

In the Gaussian-response case, consider the response tensor $Y \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ and a covariate matrix $A \in \mathbb{R}^{d_1 \times p}, d_1 \ll p$. The model 3 in my previous is:

$$\mathbb{E}Y = U \times_1 M_1 \times_2 M_2 \times_3 M_3, A = M_1 W \tag{1}$$

where $U \in \mathbb{R}^{r_1 \times r_2 \times r_3}, M_1 \in \mathbb{R}^{d_1 \times r_1}, M_2 \in \mathbb{R}^{d_2 \times r_2}, M_3 \in \mathbb{R}^{d_3 \times r_3}, W \in \mathbb{R}^{r_1 \times p}$.

WTS model 3 is equal to the model below:

$$\mathbb{E}Y = B \times_1 A = Q \times_1 N_1 \times_2 N_2 \times_3 N_3 \times_1 A \tag{2}$$

where $B \in \mathbb{R}^{p \times d_2 \times d_3}, Q \in \mathbb{R}^{r_1 \times r_2 \times r_3}, N_1 \in \mathbb{R}^{p \times r_1}, N_2 \in \mathbb{R}^{d_2 \times r_2}, N_3 \in \mathbb{R}^{d_3 \times r_3}$.

First: Notice that the estimate of the expectation response tensor would be the same, so to show the equivalence, we should find Q, N_1, N_2, N_3 subject to:

$$Q \times_1 N_1 \times_2 N_2 \times_3 N_3 \times_1 A = U \times_1 M_1 \times_2 M_2 \times_3 M_3,\ where\ A = M_1 W$$

Second: Notice that the model should be the same on mode 2 and mode 3, so we can easily conclude that $M_2 = N_2, M_3 = N_3$.

Third: Notice that in (1), the expectation response tensor is decomposed by tucker with r_1 on mode 1. Therefore, it is reasonable to say that $rank(A)$ should be larger than or equal to r_1. And because $A = M_1 W, rank(M_1) = r_1, rank(A) \leq min\{rank(M_1), rank(W)\}$, so $rank(A)$ should be r_1 , thus $rank(W) = r_1$.

Therefore, we can find a row operation matrix $P \in \mathbb{R}^{p \times p}$ and decompose PW^T as :

$$PW^T = \begin{bmatrix} W_b \\ W_n \end{bmatrix}$$

where $W_b \in \mathbb{R}^{r_1 \times r_1}, W_n \in \mathbb{R}^{(p-r_1) \times r_1}$, and $rank(W_b) = r_1$. Then we can construct N_1^T as:

$$N_1^T = [W_b^{-1}, 0]P$$

Therefore,

$$N_1^T W^T = [W_b^{-1}, 0]PP^{-1} \begin{bmatrix} W_b \\ W_n \end{bmatrix} = I_{r_1 \times r_1}$$

And let $Q = U$, we can get:

$$\begin{aligned}\mathbb{E}Y &= Q \times_1 N_1 \times_2 N_2 \times_3 N_3 \times_1 A \\ &= U \times_1 N_1 \times_2 M_2 \times_3 M_3 \times_1 (M_1 W) \\ &= U \times_1 (M_1 W) N_1 \times_2 M_2 \times_3 M_3 \\ &= U \times_1 M_1 \times_2 M_2 \times_3 M_3,\ where\ A = M_1 W\end{aligned}$$

So we can conclude that these two model are equivalent.

2 Compare model 3 with the model in $Supervised\ multiway\ factorization$

Under the settings in last part:

Model 3:(Consider the error in the linear regression part is 0)

$$Y = U \times_1 M_1 \times_2 M_2 \times_3 M_3 + \mathbb{E}, \quad A = M_1 W$$

where $\mathbb{E} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ is a error tensor with independent normal entries.

Model in the paper(model 4):

$$\begin{aligned} &\mathbb{X} = [U, V_1, \ldots, V_k] + \mathbb{E}, \quad U = YB + F \\ \Leftrightarrow\ &\mathbb{X} = C \times_0 U \times_1 V_1 \cdots \times_k V_k + \mathbb{E}, \quad U = YB + F\end{aligned}$$

Transfer the model in our notation and settings:

$$Y = C \times_1 M_1 \times_2 M_2 \times_3 M_3 + \mathbb{E}, \quad M_1 = AV + F$$

where $C \in \mathbb{R}^{r \times r \times}$ which is a identity tensor and $M_i \in \mathbb{R}^{d_i \times r}, i = 1, 2, 3,\ V \in \mathbb{R}^{p \times r}, F \in \mathbb{R}^{d_1 \times r}$ and each row of F are identically and independently distributed in multivariate normal distribution.

- **Similarity:** In these two models, the factorization parts are all semi-supervised under the covariate matrix A through a linear regression way.

  And as the last section says, the model 3 can be equivalent in another way:

$$\mathbb{E}Y = B \times_1 A$$

And easily, in the model 4, we can get:

$$\begin{aligned} \mathbb{E}Y &= C \times_1 (AB) \times_2 M_2 \times_3 M_3 \\ &= \tilde{B} \times_1 A \end{aligned}$$

- **Difference:** Obviously, the first trivial difference is model 3 use tucker decomposition and model 4 use CP factorization, which leads to the tucker form B and CP factorization form $\tilde{B}$.

  The main difference between the two models should be the different formula of linear regression part. In model 3, we regard the covariate matrix as the response while in model 4 we regard the covariate matrix as the explanatory matrix.