Semi-Supervised Binary Tensor Factorization on dnations data

Zhuoyan Xu

Consider we have an extra covariate matrix $Y^{d1\times p}$(accounting for features), which contains the information of countries. We want to connect the membership matrix(or factor matrix) A and B with the information in Y. We came up with two models.

1 Treat Y as predictors

The first one, we treat Y as predictors. To connect the information of $\mathcal{X}$ and Y, we use matrix product to connect Y and one of $\mathcal{X}$'s factor matrices.

The general form is:

$$logit\left\{\mathbb{E}\left[\mathcal{X}^{d_1d_2d_3}\right]\right\} = \mathcal{G}^{r_1r_2r_3} \times_1 A^{d_1r_1} \times_2 B^{d_2r_2} \times_3 C^{d_3r_3}$$
$$A^{d_1r_1} = Y^{d_1p}W^{pr_1}$$

Where $\mathcal{G}$ is the low rank core tensor of factorization. A, B, C is three factor matrices. W is the regression coefficient matrix for Y on A. Under this scenario, $111 = p > d1 = 14$.

The degree of freedom of this model is

$$r_1r_2r_3 + pr_1 + d_2r_2 + d_3r_3$$

The total sample size is

$$d_1d_2d_3 + d_1p$$

Note that we need to choose $r_1 < rank(Y)$.

1.1 Another view: Bilinear Model

We can write down the model in another view, which helps to compute:

$$logit\left\{\mathbb{E}\left[\mathcal{X}^{d_1d_2d_3}\right]\right\} = \mathcal{C}^{pd_2d_3} \times_1 Y^{d_1p}$$
$$\mathcal{C}^{pd_2d_3} = \mathcal{G}^{r_1r_2r_3} \times_1 W^{pr_1} \times_2 B^{d_2r_2} \times_3 C^{d_3r_3}$$

1.2 Optimization

The optimization is the same as previous with unsupervised model. We first came up with initialization of $\mathcal{C}$ given $\mathcal{X}$ and Y using glm. Then we use tucker decomposition to get initial value of $\mathcal{G}, W, B, C$.

The general form can be written as:

$$logit\left\{\mathbb{E}\left[\mathcal{X}^{d_1d_2d_3}\right]\right\} = \mathcal{G}^{r_1r_2r_3} \times_1 (YW)^{d_1r_1} \times_2 B^{d_2r_2} \times_3 C^{d_3r_3}$$

Once we get the initial value, we can use glm to update $\mathcal{G}, W, B, C$ as we did before. The update of $\mathcal{G}$, B and C are same as unsupervised factorization.

Update of W when fixed $\mathcal{G}, B, C$ Once we fixed $\mathcal{G}, B, C$, consider

$$\begin{aligned}\mathcal{U}^{d_1d_2d_3} =& logit\left\{\mathbb{E}\left[\mathcal{X}^{d_1d_2d_3}\right]\right\} = \mathcal{G}^{r_1r_2r_3} \times_1 (YW)^{d_1r_1} \times_2 B^{d_2r_2} \times_3 C^{d_3r_3} \\ =& \mathcal{G}_{BC}^{r_1d_2d_3} \times_1 (YW)^{d_1r_1}\end{aligned}$$

Use U,G to denote the unfolded matrix of $\mathcal{U}, \mathcal{G}$ along mode 1. Thus we have:

$$U^{d_1\times d_2d_3} = Y^{d_1p}W^{pr_1}G^{r_1\times d_2d_3}$$

Use tensor notation for above matrix product, we have:

$$U^{d_1\times d_2d_3} = W^{pr_1} \times_1 Y^{d_1p} \times_2 t(G)^{d_2d_3\times r_1}$$

Consider element-wise formula for either tensor notation or matrix notation, we have:

$$\begin{aligned}U_{ij} &= \sum_{i'=1}^{p}\sum_{j'=1}^{r_1} W_{i'j'}Y_{ii'}t(G)_{jj'} \\ &= \sum_{i'=1}^{p}\sum_{j'=1}^{r_1} W_{i'j'}Y_{ii'}G_{j'j}\end{aligned}$$

Let

$$[N^{ij}]_{i'j'} = Y_{ii'}G_{j'j} = Y_{ii'}t(G)_{jj'}$$

where $i = 1, \ldots, d_1; j = 1, \ldots, d_2d_3$ There are $d_1 * d_2d_3$ N^{ij} tensors totally.The dimension of N^{ijk} is $p \times r_1$.

Then we have:

$$U_{ij} = \sum_{i'=1}^{p}\sum_{j'=1}^{r_1} W_{i'j'}N_{i'j'}^{ij}$$

Since we have:

$$\text{Vec}\left(N^{ij}\right) = t(G)_{j:} \otimes Y_{i:}$$

Thus:

$$\begin{aligned}U_{ij} &= \sum_{j'=1}^{r_1}\sum_{i'=1}^{p} W_{i'j'}N_{i'j'}^{ij} \\ &= Vec(N^{ij})_{1\times pr_1}^{T}\,\text{Vec}\,(W)^{pr_1\times 1}\end{aligned}$$

First step:

$$U_{:j} = \begin{bmatrix} U_{1j} \\ U_{2j} \\ \vdots \\ U_{d_1 j} \end{bmatrix}_{d_1*1} = \begin{bmatrix} Vec(N^{1j})^T \\ Vec(N^{2j})^T \\ \vdots \\ Vec(N^{d_1 j})^T \end{bmatrix}_{d_1*pr_1} \times Vec(W)_{pr_1*1} = Vec(N_{d_1 j}) \times Vec(W)$$

Second step:

$$U_{::} = \begin{bmatrix} U_{:1} \\ U_{:2} \\ \vdots \\ U_{:d_2 d_3} \end{bmatrix}_{d_1 d_2 d_3*1} = \begin{bmatrix} Vec(N_{d_1 1}) \\ Vec(N_{d_1 2}) \\ \vdots \\ Vec(N_{d_1 d_2 d_3}) \end{bmatrix}_{d_1 d_2 d_3*pr_1} \times Vec(W)_{pr_1*1} = N^{long}_{d_1 d_2 d_3*pr_1} \times Vec(W)$$

The explict steps are shown below:

Algorithm 1 Semi-binary tensor decomposition

Input $\mathcal{X}, Y$, the shape of core tensor r_1, r_2, r_3, simulation times N
Use GLM to get $\mathcal{C}^{(1)}$, then use Tucker decomposition to get $W^{(1)}, B^{(1)}, C^{(1)}$
for $n = 1, 2, \ldots, N$ **do**
    (a) Update B,C use matrix form GLM, $B^{(n)} \leftarrow B^{(n+1)}, C^{(n)} \leftarrow C^{(n+1)}$, then orthogonalize B,C.
    (b) Update $\mathcal{G}$ using vectorization of tensor, $\mathcal{G}^{(n)} \leftarrow \mathcal{G}^{(n+1)}$
    (c) Update W using vectorization of matrix, $W^{(n)} \leftarrow W^{(n+1)}$, then orthogonalize W.
Output $\mathcal{G}, W, B, C$.

1.3 Simulation

Recalling the notarion form of bilinear model:

$$logit\left\{\mathbb{E}\left[\mathcal{X}^{d_1 d_2 d_3}\right]\right\} = \mathcal{C}^{pd_2 d_3} \times_1 Y^{d_1 p}$$
$$\mathcal{C}^{pd_2 d_3} = \mathcal{G}^{r_1 r_2 r_3} \times_1 W^{pr_1} \times_2 B^{d_2 r_2} \times_3 C^{d_3 r_3}$$

I set

$$d_1 = d_2 = d_3 = 20$$
$$p = 5$$
$$r_1 = r_2 = r_3 = 2$$

To compare the estomation error of tensor $\mathcal{C}$ and tensor $\mathcal{U}$. I use the Frobenius norm of the difference of matrix to denote it. As we can see:

As we can see, the error of second simulation is vary large(maybe because not convergence).

The true $\mathcal{C}$ and the $\hat{\mathcal{C}}$ is shown below:

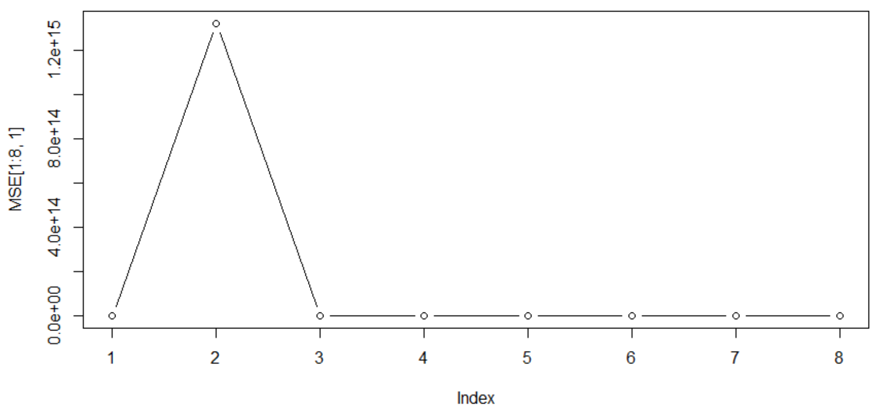


Figure 1: The error of C

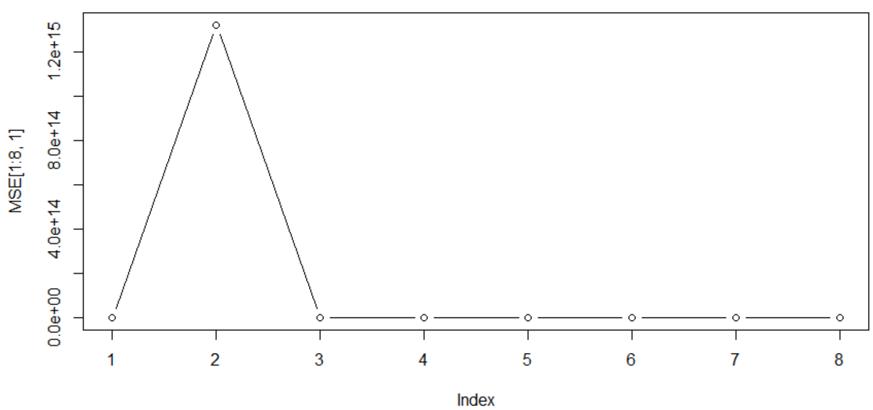
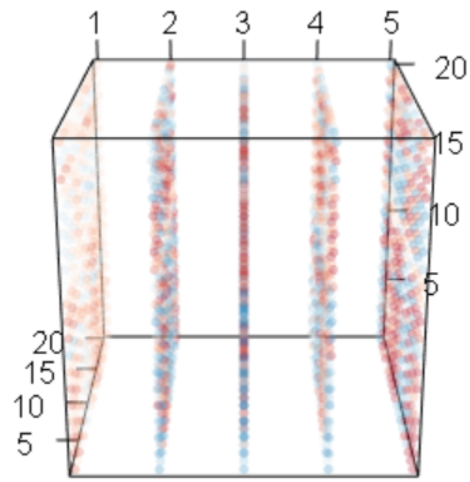


Figure 2: The error of U



(a) True C

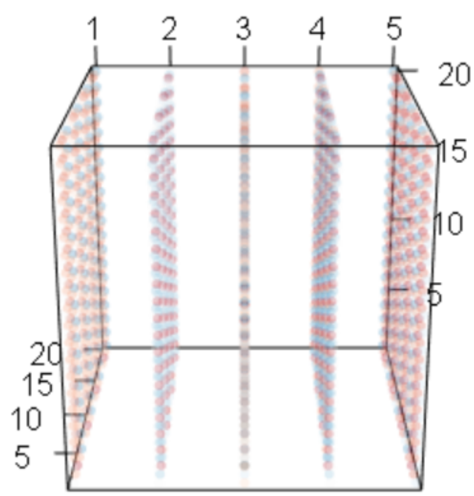


Figure 3: Estimated C

2 Treat Y as responses

The general form is:

$$\begin{aligned} logit\left\{\mathbb{E}\left[\mathcal{X}^{d_1d_2d_3}\right]\right\} &= \mathcal{G}^{r_1r_2r_3} \times_1 A^{d_1r_1} \times_2 B^{d_2r_2} \times_3 C^{d_3r_3} \\ logit\{\mathbb{E}\left[Y^{d_1p}\right]\} &= A^{d_1r_1}W^{r_1p} \end{aligned}$$

Where $\mathcal{G}$ is the low rank core tensor of factorization. A, B, C is three factor matrices. W is the GLM regression coefficient matrix for A on Y. Under this scenario, $111 = p > d1 = 14$.

The degree of freedom of this model is

$$r_1r_2r_3 + d_1r_1 + d_2r_2 + d_3r_3 + pr_1$$

The total sample size is

$$d_1d_2d_3 + d_1p$$

2.1 Optimization

Update B,C as the same as before

Update A when fixed W,B,C Consider

$$\begin{aligned} \mathcal{U}_X^{d_1d_2d_3} =& logit\left\{\mathbb{E}\left[\mathcal{X}^{d_1d_2d_3}\right]\right\} = \mathcal{G}^{r_1r_2r_3} \times_1 A^{d_1r_1} \times_2 B^{d_2r_2} \times_3 C^{d_3r_3} \\ =& \mathcal{G}_{BC}^{r_1d_2d_3} \times_1 A^{d_1r_1} \\ Then : \mathcal{U}_{X(1)}^{d_1\times d_2d_3} =& A^{d_1r_1}\mathcal{G}_{BC(1)}^{r_1\times d_2d_3} \\ U_Y^{d_1p} =& A^{d_1r_1}W^{r_1p} \end{aligned}$$

Where $U_{X(1)}$ is the unfold of $\mathcal{U}_X$ through mode-1. Thus, we have:

$$\left[U_{X(1)}, U_Y\right] = A\left[\mathcal{G}_{BC(1)}, W^{r_1p}\right]$$

We can use matrix form GLM to realize it.