Some Evidence Theory and Checking Algorithm

Zhuoyan Xu

4/18/2019

1 Evidence theory

Consider the semi-supervised binary tensor factorization, under the scenario where treat covariate as predictors. The general model is:

$$U = logit(\mathbb{E}Y) = B \times_1 X$$
$$B = C \times_1 N_1 \times_2 N_2 \times_3 N_3$$

where Y is the given binary tensor, X is the given covariate. N_1, N_2, N_3 are the factor matrices, C is the core tensor. B is the intermediate tensor.

1.1 Conclusion 1 : Each update has unique solution, and each update increases log-likelihood

To write in a general form, we have:

$$U = logit(\mathbb{E}Y) = C \times_1 XN_1 \times_2 N_2 \times_3 N_3$$

1.1.1 Multi-process of GLM

First, we prove that the update of factor matrices or core tensor is just the multi-process of regular GLM.

For updating N_3 when fix C, N_1, N_2(update N_2 is the same) Since we have:

$$U_{(3)} = N_3 C_{(3)} [N_2 \otimes (XN_1)]^T$$

Where $U_{(3)}, C_{(3)}$ are the unfold matrices of tensor U,C through mode-3.
Recalling the matrix form GLM, which is identical to this scenario:
Consider

$$logit[E(Y))_{n*p}] = U_{n*p} = X_{n*R} \times \beta_{R*p}$$

Where $U_{n*p} = (u_1, \ldots, u_p)$, $\beta_{R*p} = (\beta_1, \ldots, \beta_p)$

We have:

$$\begin{aligned}
u_1^{N*1} &= X^{n*R} \times \beta_1^{R*1} \\
u_2^{N*1} &= X^{n*R} \times \beta_2^{R*1} \\
&\vdots \\
u_p^{N*1} &= X^{n*R} \times \beta_P^{R*1}
\end{aligned}$$

Then we implement regular GLM on each equation(i.e. each column of matrix β). These are multiprocessing and mutually independent GLM processes.

For updating N_1 when fix C, N_2, N_3

Since we have:

$$U_{(1)} = XN_1C_{(1)}[N_3 \otimes N_2]^T$$

Where $U_{(1)}, C_{(1)}$ are the unfold matrices of tensor U,C through mode-1.

Recalling the form on note **Semi-Supervised Binary Tensor Factorization on dnations data**:

$$U^{d_1 \times d_2 d_3} = Y^{d_1 p} W^{p r_1} G^{r_1 \times d_2 d_3}$$

we have already proved that through some vectorization of matrix, it can be written as regular GLM form :

$$Y = X\beta$$

where β is the vectorization of N_1 in this scenario.

Thus it's still GLM.

For updating C when fix N_1, N_2, N_3

Recalling the note **Unsupervised Binary Tensor Factorization**, we have already proved that through some vectorization of matrix, it can be written as regular GLM form :

$$Y = X\beta$$

where β is the vectorization of C in this scenario. Thus it's still GLM.

1.1.2 Conclusion 1 on GLM

Then we just need to show the conclusion is satisfied on the GLM, in this case logistic regression.

Consider the form:

$$logit(\mathbb{E}\tilde{Y}) = U = X\beta$$

Recalling the negative log-likehood(i.e. cross-entropy), we have:

$$\begin{aligned}
Q &= -\sum_{i=1}^{n} \{\tilde{y}_i \log(\pi_i) + (1 - \tilde{y}_i)\log(1 - \pi_i)\} \\
\pi_i &= \frac{\exp(\boldsymbol{x}_i' \boldsymbol{\beta})}{1 + \exp(\boldsymbol{x}_i' \boldsymbol{\beta})}
\end{aligned}$$

Where x_i are one observation. If our dimension is p, then $x_i, \beta \in \mathbb{R}^p$

Consider $\tilde{y}_i \in \{0, 1\}$, let

$$y_i = 2\tilde{y}_i - 1 \in \{-1, 1\}$$

Then:

$$\begin{aligned} Q &= -\sum_{i=1}^{n} log\left\{\frac{1}{1+\exp\left(-y_i x_i'\beta\right)}\right\} \\ &= \sum_{i=1}^{n} log\left\{1+\exp\left(-y_i x_i'\beta\right)\right\} \end{aligned}$$

Then we consider the optimization problem:

$$\beta^* = \arg\min_{\beta} Q \tag{1}$$

$$= \arg\min_{\beta} \sum_{i=1}^{n} log\left\{1+\exp\left(-y_i x_i'\beta\right)\right\} \tag{2}$$

Since we have:

$$\begin{aligned} \frac{\partial Q}{\partial \beta} &= -\sum_{i=1}^{n} \frac{y_i}{1+\exp\left(y_i x_i'\beta\right)} x_i \\ \frac{\partial^2 Q}{\partial \beta^2} &= \sum_{i=1}^{n} \left[\frac{y_i}{1+\exp\left(y_i x_i'\beta\right)}\right]^2 x_i x_i^T \end{aligned}$$

Since for $\forall\ z \in \mathbb{R}^p$, we have:

$$z^T \frac{\partial^2 Q}{\partial \beta^2} z = \sum_{i=1}^{n} \left[\frac{y_i}{1+\exp\left(y_i x_i'\beta\right)}\right]^2 (z^T x_i)^2 \geq 0$$

The equal sign holds if and only if z=0. Then we conclude that the Hessian matrix is positive definite.

According to:

$$\nabla^2 Q(\beta) > 0 \text{ for all } \beta \in \mathrm{dom}(Q) = \mathbb{R}^p$$

We can conclude Q is strictly convex with respect to β. Then the optimization problem (2) has unique solution.

Consider the update at t step:

$$\begin{aligned} \beta^{(t)} &= \arg\min_{\beta} Q(\beta^{(t-1)}) \\ &= \arg\min_{\beta} \sum_{i=1}^{n} log\left\{1+\exp\left(-y_i x_i'\beta^{(t-1)}\right)\right\} \end{aligned}$$

According to what we derived, each update has unique solution. And each update won't decrease the log-likelihood. If $\beta^{(t-1)}$ reach the global minimum of opbejctive function, the update will remain the same:$\beta^{(t)} = \beta^{(t-1)}$.

1.2 Conclusion 2 : The log-likelihood is invariant to orthogonalization

Suppose we have:

$$
\begin{aligned}
U &= logit(\mathbb{E}Y) = B \times_1 X \\
B &= C \times_1 N_1 \times_2 N_2 \times_3 N_3
\end{aligned}
$$

where Y is the given binary tensor, X is the given covariate. N_1, N_2, N_3 are the factor matrices, C is the core tensor. B is the intermediate tensor.

When we conduct the orthogonalization and normalization on N_1 (N_2, N_3 are the same), such as SVD:

$$N_1 = U\Sigma V^T$$

Take $\tilde{N}_1 = U$ as orthonormal version of N_1, let $P = \Sigma V^T$, then:

$$N_1 = \tilde{N}_1 P$$

Let $C = \tilde{C}P^{-1}$ Then we have:

$$
\begin{aligned}
B &= C \times_1 N_1 \times_2 N_2 \times_3 N_3 \\
&= \tilde{C}P^{-1} \times_1 \tilde{N}_1 P \times_2 N_2 \times_3 N_3 \\
&= \tilde{C} \times_1 \tilde{N}_1 \times_2 N_2 \times_3 N_3
\end{aligned}
$$

Then the log-likelihood is the same.