Multi-way tensor decomposition from binary data

Miaoyan Wang, Feb 21, 2019

0.1 Model

For the binary tensor $\mathcal{Y} = [\![y_{i_1,\ldots,i_K}]\!] \in \{0,1\}^{d_1\times\cdots\times d_K}$, we assume its entries are realizations of independent (think about why not "i.i.d"? are they *identically* distributed?) Bernoulli random variables, in that

$$\mathcal{Y}|\Theta \sim \text{Bernoulli}\left\{f(\Theta)\right\}, \quad \text{with } \mathbb{P}(y_{i_1,\ldots,i_K}=1)=f(\theta_{i_1,\ldots,i_K}),$$

for all $(i_1,\ldots,i_K)\in[d_1]\times\cdots\times[d_K]$. In this model, $f\colon\mathbb{R}\to[0,1]$ is a strictly increasing function. We further assume that:

1. $f(\theta)$ is twice-differentiable in $\theta\in\mathbb{R}/\{0\}$;
2. $f(\theta)$ is strictly increasing and strictly log-concave;
3. $f'(\theta)$ is unimodal and symmetric with respect to $\theta=0$.

All these assumptions are fairly mild. In the GLM language, f is often referred to as the "inverse link function". When no confusion arises, we also call f the "link function". The parameter tensor $\Theta = [\![\theta_{i_1,\ldots,i_K}]\!] \in \mathbb{R}^{d_1\times\cdots\times d_K}$ is continuous-valued, unknown, and is the main object of interest in our tensor estimation inquiry.

Furthermore, we assume the parameter tensor Θ admits a $(r_1,\ldots,r_K)$-block model:

$$\Theta = \mathcal{U}\times_1 \boldsymbol{M}_1 \times_2 \cdots \times_K \boldsymbol{M}_K,$$

where $\mathcal{U}\in\mathbb{R}^{r_1\times\cdots\times r_K}$ is the low-dimensional core tensor, $\boldsymbol{M}_k\in\{0,1\}^{r_k\times d_k}$ is the mode-k membership matrix, $\boldsymbol{M}_k\boldsymbol{M}_k^T = I_{r_k}$ for all $k=1,\ldots,K$.

0.2 Link

We next consider two common choices of link function f:

Example 1. (Logistic link/Logistic noise). The logistic model is represented by (1) with $f(\theta) = \left(1+e^{-\theta/\sigma}\right)^{-1}$ and the scale parameter $\sigma>0$.

Example 2. (Probit link/Gaussian noise). The probit model is represented by (1) with $f(\theta) = \Phi(\theta/\sigma)$, where Φ is the cumulative distribution function of a standard Gaussian.

0.3 Estimation

We propose to estimate the unknown parameter tensor Θ in model (1) using a likelihood approach. The log-likelihood function for the model is

$$\begin{aligned}
\mathcal{L}_{\mathcal{Y}}(\Theta) &= \sum_{i_1,\ldots,i_K} \left[\mathbb{1}_{\{y_{i_1,\ldots,i_K}=1\}} \log f(\theta_{i_1,\ldots,i_K}) + \mathbb{1}_{\{y_{i_1,\ldots,i_K}=0\}} \log \{1 - f(\theta_{i_1,\ldots,i_K})\} \right] \\
&= \sum_{i_1,\ldots,i_K} \log f(q_{i_1,\ldots,i_K} \theta_{i_1,\ldots,i_K}),
\end{aligned}$$

where $q_{i_1,\ldots,i_K} = 2y_{i_1,\ldots,i_K} - 1$ takes values -1 or 1, and the second equality is due to the symmetry of the link function f. In particular, plugging the standard logistic link $f(x) = \frac{e^x}{1+e^x}$, we obtain

$$\mathcal{L}_{\mathcal{Y}}(\Theta) = \sum_{i_1,\ldots,i_K} \log \left(\frac{e^{q_{i_1,\ldots,i_K} \theta_{i_1,\ldots,i_K}}}{1 + e^{q_{i_1,\ldots,i_K} \theta_{i_1,\ldots,i_K}}} \right).$$