Generalized tensor response regression with multilinear features

1 Preliminaries

Let $\mathcal{Y} = [\![y_{i_1,\ldots,i_K}]\!] \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ be an order-K tensor, and $\boldsymbol{X}_k \in \mathbb{R}^{d_k \times p_k}$ be the features along the k-th mode.

$$f(\mathbb{E}(\mathcal{Y}|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K)) = \mathcal{B} \times_1 \boldsymbol{X}_1 \times_2 \cdots \times_K \boldsymbol{X}_K$$

where $f(\cdot)$ is the inverse link function which is assumed to be applied to tensors in an entry-wise manner. Here $\mathcal{B} \in \mathbb{R}^{p_1 \times \cdots p_K}$ is the unknown parameter tensor of interest.

$$\mathcal{Y}|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K \overset{\text{in D.}}{\sim} \exp\left(\langle \mathcal{Y}, \mathcal{U} \rangle - \sum_{i_1,\ldots,i_K} b(u_{i_1,\ldots,i_K}) \right) \prod_{i_1,\ldots,i_K} h(y_{i_1,\ldots,i_K})$$

where $\mathcal{U} = \mathcal{B} \times_1 \boldsymbol{X}_1 \times_2 \cdots \times_K \boldsymbol{X}_K$

where $b'(\cdot)$ is a link function and $\mathcal{U} = [\![u_{i_1,\ldots,i_K}]\!]$ is the canonical parameter

$$\mathbb{E}(y_{i_1,\ldots,i_K}|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K) = b'(u_{i_1,\ldots,i_K}) = \mathcal{B} \times_1 \boldsymbol{X}_1 \times_2 \cdots \times_K \boldsymbol{X}_K$$

2 Model

Suppose that the response tensor $\mathcal{Y} = [\![Y_{i_1,\ldots,i_K}]\!]$ follows an exponential family distribution:

$$f(\mathcal{Y}|\Theta) = \prod_{i_1,\ldots,i_K} f(y_{i_1,\ldots,i_K}|\Theta) = \prod_{i_1,\ldots,i_K} c(y_{i_1,\ldots,i_K}) \exp\left(\frac{y_{i_1,\ldots,i_K}\theta_{i_1,\ldots,i_K} - b(\theta_{i_1,\ldots,i_K})}{\phi} \right).$$

Here ϕ is a constant, $c(\cdot)$ and $b(\cdot)$ are known functions, and $\Theta = [\![\theta_{i_1,\ldots,i_K}]\!]$ collects the linear predictors

$$\Theta = \Theta(\mathcal{B}, \mathcal{X}) = \mathcal{B} \times_1 \boldsymbol{X}_1 \times_2 \cdots \times_K \boldsymbol{X}_K, \quad \text{rank}(\mathcal{B}) \leq (r_1, \ldots, r_K),$$

where $\mathcal{B} \in \mathbb{R}^{p_1 \times \cdots \times p_K}$ is the unknown coefficient tensor of interest.

	$b(\theta)$	$b'(\theta)$	$c(\theta)$	ϕ
Normal	$\frac{1}{2}\theta^2$	θ	...	σ^2
Bernoulli		$\frac{1}{1+e^{-\theta}}$	...	1
Poisson	$e^(\theta)$	e^{θ}	...	1

The log-likelihood for the coefficient tensor $\mathcal{B}$ is given, up to an affine transformation, by

$$\mathcal{L}_{\mathcal{Y}}(\mathcal{B}) = \langle \mathcal{Y}, \Theta \rangle - \sum_{i_1,\ldots,i_K} b(\theta_{i_1,\ldots,i_K}), \quad \text{where} \quad \Theta = \mathcal{B} \times_1 \boldsymbol{X}_1 \times_2 \cdots \times_K \boldsymbol{X}_K.$$

We propose the constrained MLE:

$$\hat{\mathcal{B}} = \underset{\text{rank}(\mathcal{B}) \leq \boldsymbol{r}}{\arg\max}\ \mathcal{L}_{\mathcal{Y}}(\mathcal{B})$$

Theorem 1. $\|\hat{\mathcal{B}} - \mathcal{B}\|_F^2 \leq w$

Proof. Let $\ell(\mathcal{B}) = \mathbb{E}_{\mathcal{Y} \sim \mathcal{B}_{\text{true}}}(\mathcal{L}_{\mathcal{Y}}(\mathcal{B}))$. We first show that (1) the stochastic deviation $\mathcal{L}_{\mathcal{Y}}(\mathcal{B}) - \ell(\mathcal{B})$ is uniformly small for all $\mathcal{B}$ and (2) $\ell(\hat{\mathcal{B}}) - \ell(\mathcal{B}_{\text{true}}) \approx C\|\hat{\mathcal{B}} - \mathcal{B}_{\text{true}}\|_F^2$.

Note that

$$\begin{aligned}
\mathcal{L}_{\mathcal{Y}}(\mathcal{B}) - \ell(\mathcal{B}) &= \langle \mathcal{Y} - \mathbb{E}(\mathcal{Y}|\mathcal{X}),\ \Theta(\mathcal{B}) \rangle \\
&= \langle \mathcal{Y} - b'(\Theta_{\text{true}}),\ \Theta \rangle \\
&= \langle \mathcal{E} \times_1 \boldsymbol{X}_1^T \times_2 \cdots \times_K \boldsymbol{X}_K^T, \mathcal{B} \rangle,
\end{aligned}$$

where $\mathcal{E} = [\![\varepsilon_{i_1,\ldots,i_K}]\!] \stackrel{\text{def}}{=} \mathcal{Y} - b'(\Theta_{\text{true}})$. Based on assumption 1, $\mathcal{E}$ is sub-Gaussian with parameter bounded by ϕM. Therefore $\mathcal{E} \times_1 \boldsymbol{X}_1^T \times_2 \cdots \times_K \boldsymbol{X}_K^T$ is a sub-Gaussian with parameter bounded by $\phi M \prod_k \sigma_k(\boldsymbol{X}_k)$.

Second,

$$\ell(\mathcal{B}) = \ell(\mathcal{B}_{\text{true}}) - \frac{1}{2}(\mathcal{B} - \mathcal{B}_{\text{true}})^T \mathbb{E}(\mathcal{H}_{\mathcal{Y}}(\check{\mathcal{B}}))(\mathcal{B} - \mathcal{B}_{\text{true}})$$

Note that (non-random) Hession matrix is .... So

$$\ell(\mathcal{B}) - \ell(\mathcal{B}_{\text{true}}) = -\frac{1}{2} \sum_{i_1,\ldots,i_K} b''(\check{\theta}_{i_1,\ldots,i_K})(\theta_{i_1,\ldots,i_K} - \theta_{\text{true},i_1,\ldots,i_K})^2 \leq -\frac{L}{2}\|\Theta - \Theta_{\text{true}}\|_F^2,$$

holds for all $\mathcal{B}$, provided that $b''(\theta) \geq L > 0$.

Now consider $\hat{\mathcal{B}}$. Because $\mathcal{L}_{\mathcal{Y}}(\hat{\mathcal{B}}) - \mathcal{L}_{\mathcal{Y}}(\mathcal{B}_{\text{true}}) \geq 0$,

$$\begin{aligned}
0 &\leq \mathcal{L}_{\mathcal{Y}}(\hat{\mathcal{B}}) - \mathcal{L}_{\mathcal{Y}}(\mathcal{B}_{\text{true}}) \\
&\leq \Big(\mathcal{L}_{\mathcal{Y}}(\hat{\mathcal{B}}) - \ell(\hat{\mathcal{B}})\Big) - (\mathcal{L}_{\mathcal{Y}}(\mathcal{B}_{\text{true}}) - \ell(\mathcal{B}_{\text{true}})) + \Big(\ell(\hat{\mathcal{B}}) - \ell(\mathcal{B}_{\text{true}})\Big) \\
&\leq \langle \mathcal{E}, \hat{\Theta} - \Theta_{\text{true}} \rangle - \frac{L}{2}\|\hat{\Theta} - \Theta_{\text{true}}\|_F^2
\end{aligned}$$

$$\|\hat{\Theta} - \Theta_{\text{true}}\|_F \leq \frac{2}{L}\langle \mathcal{E}, \frac{\hat{\Theta} - \Theta_{\text{true}}}{\|\hat{\Theta} - \Theta_{\text{true}}\|_F} \rangle \leq \frac{2}{L} \sup_{\Theta: \|\Theta\|_F = 1, \Theta = \mathcal{B} \times_1 \boldsymbol{X}_1 \times_2 \cdots \times_K \boldsymbol{X}_K} \langle \mathcal{E}, \Theta \rangle$$

Therefore,

$$\|\hat{\Theta} - \Theta_{\text{true}}\|_F \leq \frac{2R}{L} \sum_{i=1} p_i()$$

$$\|\hat{\Theta} - \Theta_{\text{true}}\|_F^2 \leq \frac{2}{L} \sup_{\mathcal{B}} |\mathcal{L}_{\mathcal{Y}}(\mathcal{B}) - \ell(\mathcal{B})|$$

□

Assumption 1. *Consider the following conditions*

- $0 < L \leq b''(\theta) \leq M \leq \infty$ *for any* $\theta \in \mathbb{R}$