
Nonparametric Tensor Estimation and Completion via Hypergraphon Learning

Miaoyan Wang

University of Wisconsin – Madison
miaoyan.wang@wisc.edu

Abstract

We consider the problem of tensor estimation from noisy observations with possibly missing entries. A nonparametric approach to tensor estimation is developed based on K -uniform hypergraphons. The hypergraphon model captures the key features of conditional independence arising in the generative process of multiway tensor data. The nonparametric hypergraphon representation of tensors encompasses many existing tensor models—such as CP models, Tucker models, shape constrained tensor models—as special examples. We develop a rate-optimal nonparametric tensor estimator using the stochastic blockmodel approximation to the underlying hypergraphon. A surprising distinction between the hypergraphon-based tensor estimation and high-dimensional nonparametric regression is revealed. Furthermore, we establish the integrated risk bound for the hypergraphon estimation in specially structured functional classes. The result uncovers the joint contribution of the statistical bias-variance error and the agnostic discretization error. Numerical results demonstrate the robustness of our proposal over previous tensor methods and the attractive performance as the tensor order increases.

1 Introduction

2 Nonparametric Tensor Model via Hypergraphons

Let $\mathcal{Y} = \llbracket \mathcal{Y}(\mathbf{i}) \rrbracket \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ be an order- K (d_1, \dots, d_K)-dimensional data tensor. We propose a two-stage generative process for the tensor entries. Firstly, we draw a collection of i.i.d. random variables, $x_k(i) \in \mathcal{X}$ following the probability measure $(\mathcal{X}, \mu_{\mathcal{X}})$, for all $i \in [d_k]$, $k \in [K]$. The Cartesian product of the random variables, denoted $(x_1(i_1), \dots, x_K(i_K))$, represents the latent features at position $\mathbf{i} = (i_1, \dots, i_K)$ of the tensor. Second, conditional on the latent features, the tensor entries are independently distributed.

- Nonparametric mean: there exists a unknown function $f: \mathcal{X}^K \mapsto [0, 1]$ such that

$$\mathbb{E}[\mathcal{Y}(\mathbf{i})] = f(x_1(i_1), \dots, x_K(i_K)), \quad \text{for all } \mathbf{i} = (i_1, \dots, i_K) \in [d_1] \times \cdots \times [d_K]. \quad (1)$$

Here $x_k(i_k) \in \mathcal{X}$ denotes the latent feature associated with the i_k -th entry along the k -th mode of the tensor, for $k \in [K]$.

- Conditionally independence: conditional on the latent features $\{x_k(i_k)\}$, the tensor entries $\mathcal{Y}(\mathbf{i})$ are independently distributed, mean-zero, sub-Gaussian random variables; i.e. $\mathbb{E}[e^{t\mathcal{Y}(\mathbf{i})}] \leq \exp(t\sigma^2)$ for all $\mathbf{i} \in [d_1] \times \cdots \times [d_K]$ and $t > 0$.
- We set $\mathcal{X} = [0, 1]$ and $\mu_{\mathcal{X}}$ the Lebesgue measure over $[0, 1]$. The latent features $x_k(i_k)$ are mutually independent for all $i_k \in [d_k]$ and $k \in [K]$.
- Regularity conditions on f . Two possible options: 1. stepwise functions, 2. Holder smooth functions.

(draw a figure...) We call the model (1) the nonparametric tensor model because the mean function f is unknown and thus to be estimated. By specializing the realization of the latent features and the function f , model (1) incorporates several common previously-studied tensor models as special cases.

Example 1 (CP model) Let $\mathcal{X} = \mathbb{R}^R$ and $\mathbf{x}^{(k)} = (x_1^{(k)}, \dots, x_R^{(k)})^T \in \mathbb{R}^R$ for all $k \in [K]$. Define $f: (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(K)}) \mapsto \sum_r \prod_k x_r^{(k)}$. Then model (1) can be written as

$$\mathbb{E}(\mathcal{Y} | \boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_K) = \sum_r \boldsymbol{\xi}_r^{(1)} \otimes \dots \otimes \boldsymbol{\xi}_r^{(K)}, \text{ where } \boldsymbol{\xi}_r^{(k)} := (x_1^{(k)}, \dots, x_{d_k}^{(k)})^T \text{ for } k \in [K].$$

Definition 1 (Uniform random design) A uniform random design over the index set $[d]$ is represented by a stochastic process $\{x_i\}$, where x_i are i.i.d. sampled from $\text{Uniform}[0, 1]$ for all $i \in [d]$.

Definition 2 (Cartesian product of random designs) An order- K random design over the index set $[d_1] \times \dots \times [d_K]$ is defined as the Cartesian product of random designs over each marginal index set $[d_k]$ for $k \in [K]$. Specifically, the order- K uniform random design is represented by a collection of random vectors $\{(x_{i_1}^{(1)}, \dots, x_{i_K}^{(K)})\}$, where $x_{i_k}^{(k)}$ are i.i.d. sampled from $\text{Uniform}[0, 1]$ for all $i_k \in [d_k]$ and $k \in [K]$.

Definition 3 (Hypergraphon) A K -uniform hypergraphon is a measurable function $f: [0, 1]^K \mapsto [0, 1]$.

The function can be seen as a kernel functions for random tensor models.

Acknowledgements