Learning Multiple Networks via Supervised Tensor Decomposition

Jiaxin Hu
jhu267@wisc.edu

Chanwoo Lee
chanwoo.lee@wisc.edu

Miaoyan Wang
miaoyan.wang@wisc.edu

Department of Statistics
University of Wisconsin-Madison

Machine Learning and Physical Science workshop, NeurIPS 2020

Motivation

- Scientific studies (e.g. neuroimaging, social network analysis) often collect the tensor observations with side information.
- Tensor observations may consist of non-Gaussian measurements (e.g. binary, count data).

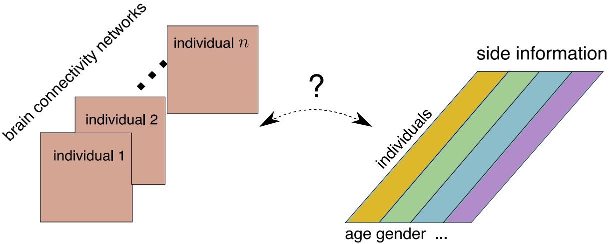


Figure 1: Brain connectivity networks (binary adjacency matrices) with side information.

Our Goal: Identify the structural variation in the data tensor affected by side information.

Model

Notation:

- $\mathcal{Y} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$: an order-$K$ tensor observation.
- $\boldsymbol{X}_i \in \mathbb{R}^{d_i \times p_i}$: feature matrices contain the side information, for $i = 1, ..., K$.
- $\times_i$: the tensor-by-matrix product on i-th mode.

Proposed supervised tensor decomposition:

$$\mathbb{E}[\mathcal{Y}|\boldsymbol{X}_1, ..., \boldsymbol{X}_K] = f\left(\mathcal{C} \times_1 \boldsymbol{X}_1 \boldsymbol{M}_1 \times_2 \cdots \times_K \boldsymbol{X}_K \boldsymbol{M}_K\right), \tag{1}$$

where $f(\cdot)$ is the known link function depending on the data type, $\mathcal{C} \in \mathbb{R}^{r_1 \times \cdots \times r_K}$ is an unknown full-rank core tensor, and $\boldsymbol{M}_i \in \mathbb{R}^{p_i \times r_i}$ are unknown factor matrices consisting of orthonormal columns.

Model

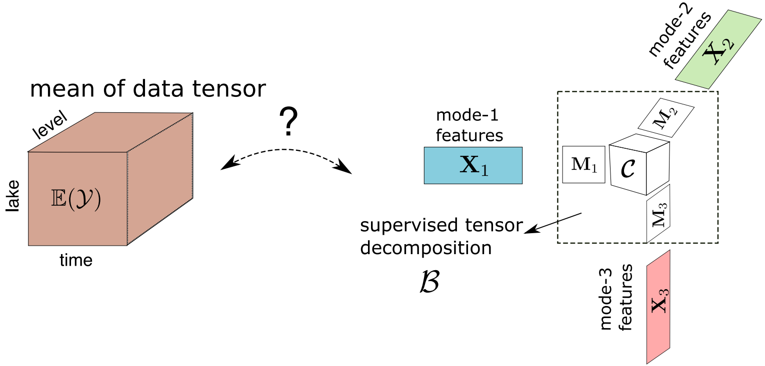


Figure 2: Supervised tensor decomposition for an order-3 tensor with side information.

The features $\boldsymbol{X}_i$ affect the distribution of tensor entries in through "supervised tensor factors" $\boldsymbol{X}_i\boldsymbol{M}_i$. We call $\boldsymbol{M}_i$ the "dimension reduction matrix", and $\mathcal{C}$ collects the interaction effects between sufficient features.

Likelihood-based estimator

$$(\hat{\mathcal{C}}, \hat{\boldsymbol{M}}_1, \ldots, \hat{\boldsymbol{M}}_K) = \underset{(\mathcal{C}, \boldsymbol{M}_1, \ldots, \boldsymbol{M}_K) \in \mathcal{P}}{\arg\max} \mathcal{L}_{\mathcal{Y}}(\mathcal{C}, \boldsymbol{M}_1, \ldots, \boldsymbol{M}_K), \tag{2}$$

where $\mathcal{L}_{\mathcal{Y}}(\cdot)$ is quasi log-likelihood function and $\mathcal{P}$ is the parameter space.

Alternating Algorithm

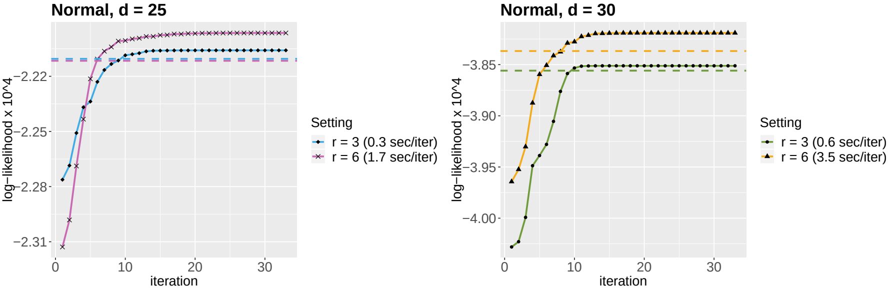


Figure 3: Likelihood trajectory for normal model. Dashed lines are likelihood with true parameters.

Statistical Guarantees

Theorem

Under mild technical assumptions, there exists two positive constants $C_1, C_2 > 0$, such that, with probability at least $1 - \exp\left(-C_1 \sum_k p_k\right)$,

$$\sin^2 \Theta(\boldsymbol{M}_{i,true}, \hat{\boldsymbol{M}}_i) \leq \frac{C_2 \prod_k r_k}{\max_k r_k \sigma^2_{min}(Unfold_i(\mathcal{C}_{true}))} \frac{\sum_k p_k}{\prod d_k}, \tag{3}$$

for $i = 1, ..., K$, and

$$\left\| \mathcal{B}_{true} - \hat{\mathcal{B}} \right\|_F^2 \leq \frac{C_2 \prod_k r_k}{\max_k r_k} \frac{\sum_k p_k}{\prod d_k}, \tag{4}$$

where $\sin \Theta(\boldsymbol{M}_{i,true}, \hat{\boldsymbol{M}}_i)$ is the angle distance between the column spaces, $Unfold_i(\mathcal{C})$ denote the operation that reshapes the tensor along mode i into a matrix of size $r_i \times \prod_{k \neq i} r_k$, and $\mathcal{B} = \mathcal{C} \times_1 \boldsymbol{M}_1 \times_2 \cdots \times_K \boldsymbol{M}_K$. The constants C_1, C_2 are independent of $\{p_i\}$ and $\{d_i\}$.

Application

Our method identifies the global connectivity pattern as well as local regions associated with age and gender in the Human Connectome Project (HCP) data.

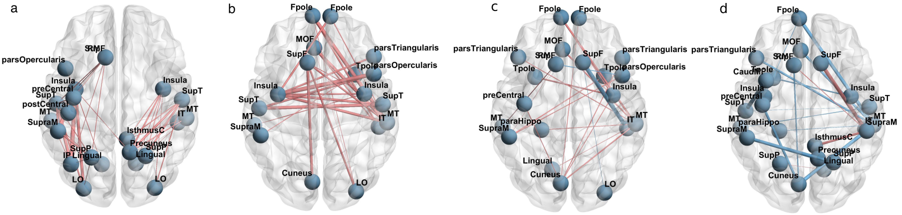
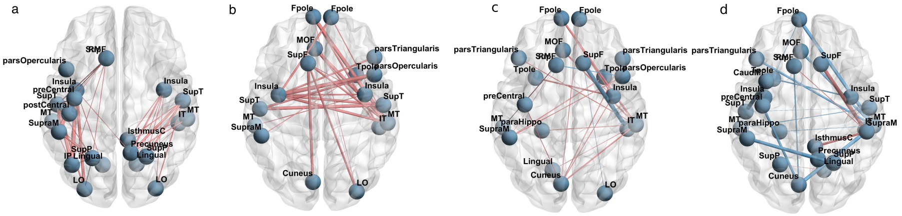


Figure 4: Top edges with large effects. (a) Global effect; (b) Female effect; (c) Age 22-25; (d) Age 31+. Red (blue) edges represent positive (negative) effects. Edge-widths are proportional to the magnitude of effect sizes.



(a) Global effect; (b) Female effect; (c) Age 22-25; (d) Age 31+.

- Global connection exhibits clear spatial separation.
- Female brains display higher inter-hemispheric connectivity.
- Several edges in frontal-pole region have declined connection in the group Age 31+.

More Information

- All data and codes are available in our R package `tensorregress`: `https://CRAN.R-project.org/package=tensorregress`.