MATH-BIOINF-STATS 547: Mathematics of Data

Due Date:

Name:

Collaborators:

Problem Set 4: Hypergraphs

For this problem set, please submit a `.pdf` document with a write-up of your results and observations. We encourage using Overleaf, but the MATLAB Live Editor or other word-processing software is acceptable. We have provided a LaTeX template to help get you started, which is available on Overleaf where you can make a copy of it.

Uniform Hypergraphs. A $k-$uniform hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$ is a set of vertices $\mathcal{V}$ together with a set of hyperedges $\mathcal{E}$ where each hyperedge is a set of k vertices. $k-$uniform hypergraphs with $|\mathcal{V}| = n$ may be represented as $k-$mode adjacency tensors $\mathbf{A} \in \mathbf{R}^{\overbrace{n \times \cdots \times n}^{k}}$ where

$$\mathbf{A}_{i_1,i_2,\ldots,i_k} = \begin{cases} \frac{1}{(k-1)!} & \text{if } (i_1, i_2, \ldots, i_k) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Problem 1: Uniform Hypergraphs from Nonuniform Data

Often data is non-uniform, and we have a set of vertices $\mathcal{V}$ and hyperedges of varying sizes [1]. This occurs in social networks, information systems, chromosome conformation capture data, and many other systems where the size of each group interaction varries. When this occurs, there is not one definitive method for representing the non-uniform data as a tensor or uniform hypergraph.

Propose your own method to construct a tensor representation or $k-$uniform hypergraph representation of nonuniform data. Write cleanly and explain the motivation of your method and the advantages of your representation of nonuniform data.

Remark: You can guidance from Section 6: Discussion from the following paper [1].

Solution:

Multi-correlation. Given n variables $\mathrm{Y}_1, \mathrm{Y}_2, \ldots, \mathrm{Y}_n$, we are interested in finding the correlation of a subset of $k \leq n$ variables $\mathrm{Y}_1, \mathrm{Y}_2, \ldots, \mathrm{Y}_k$ using the measures defined below. Similar to bivariate correlation coefficients, high multi-correlation coefficients indicate strong relationships among the k variables.

Method 1: Multirelation - A Correlation among More Than Two Variables [2]

$$r(\mathrm{Y}_1, \mathrm{Y}_2, \ldots, \mathrm{Y}_k) = 1 - \lambda(\mathbf{R}),$$

where $\lambda(\mathbf{R})$ is the smallest eigenvalue of the correlation matrix $\mathbf{R}$ of variables $\mathrm{Y}_1, \mathrm{Y}_2, \ldots, \mathrm{Y}_k$.
https://www.sciencedirect.com/science/article/pii/0167947393E00467

Method 2: Measures of Correlation for Multiple Variables [3]

$$r(\mathrm{Y}_1, \mathrm{Y}_2, \ldots, \mathrm{Y}_k) = (1 - \det(\mathbf{R}))^{\frac{1}{2}},$$

where $\mathbf{R}$ is the correlation matrix of variables $\mathrm{Y}_1, \mathrm{Y}_2, \ldots, \mathrm{Y}_k$.
https://arxiv.org/abs/1401.4827

Method 3: A Multi-Way Correlation Coefficient [4]

$$r(\mathrm{Y}_1, \mathrm{Y}_2, \ldots, \mathrm{Y}_k) = \frac{1}{\sqrt{k}}\mathrm{std}(\{\lambda_1, \lambda_2, \ldots, \lambda_k\}),$$

where λ_i are the eigenvalues of the correlation matrix of variables $\mathrm{Y}_1, \mathrm{Y}_2, \ldots, \mathrm{Y}_k$, and std = standard deviation.

https://arxiv.org/abs/2003.02561

Problem 2: Multi-correlations

In this problem, we will construct hypergraphs from time series data using the three multi-correlation methods above. The activity of 21 neurons was monitored over the course of a feeding experiment to generate a time series data set. Using the provided time series data, the three methods of multi-correlation, we will construct a series of hypergraphs.

Remark: You can guidance from Section 5C: Mouse Neuron Endomicroscopy from the following paper[1].

(a) Compute all 3, 4, and 5-way multi-correlation coefficients among neurons using the time series data using each multi-correlation coefficient.

(b) Set a threshold for the minimum multi-correlation coefficient that will indicate a hyperedge in your hypergraph.

(c) Construct $k-$uniform hypergraph objects from your 3, 4, and 5-way multi-correlaiton data.

(d) Apply your proposed method from Exercise 1 to construct a single hypergraph or tensor representation from the 3-uniform 4-uniform, and 5-uniform hypergraphs that you constructed in parts (a)-(c).

The data set and complete starter code is provided for parts (a)-(c), and in part (d) you should implement your proposed method from Exercise 1.

Solution:

References

[1] Can Chen, Amit Surana, Anthony M Bloch, and Indika Rajapakse. Controllability of hypergraphs. *IEEE Transactions on Network Science and Engineering*, 8(2):1646–1657, 2021.

[2] Zvi Drezner. Multirelation—a correlation among more than two variables. *Computational Statistics & Data Analysis*, 19(3):283–292, 1995.

[3] Jianji Wang and Nanning Zheng. Measures of correlation for multiple variables. *arXiv preprint arXiv:1401.4827*, 2014.

[4] Benjamin M Taylor. A multi-way correlation coefficient. *arXiv preprint arXiv:2003.02561*, 2020.