

Data Science Seminar

Hosted by Department of Mathematical Sciences

- Date: Tuesday, September 28, 2021
- Time: 12:00pm - 1:00pm
- Room: Via Zoom
- Speaker: Dr. Eric Lock (University of Minnesota)
- Title: Bidimensional Linked Matrix Decomposition for Pan-Omics Pan-Cancer Analysis

Abstract

Several recent methods address the integrative dimension reduction and decomposition of linked high-content data matrices. Typically, these methods consider one dimension, rows or columns, that is shared among the matrices. This shared dimension may represent common features measured for different sample sets (horizontal integration) or a common sample set with features from different platforms (vertical integration). This is limiting for data that take the form of bidimensionally linked matrices, e.g., multiple molecular omics platforms measured for multiple sample cohorts, which are increasingly common in biomedical studies. We propose a flexible approach to the simultaneous factorization and decomposition of variation across bidimensionally linked matrices, BIDIFAC+. This decomposes variation into a series of low-rank components that may be shared across any number of row sets (e.g., omics platforms) or column sets (e.g., sample cohorts). Our objective function extends nuclear norm penalization, is motivated by random matrix theory, and can be shown to give the mode of a Bayesian posterior distribution. We apply the method to pan-omics pan-cancer data from The Cancer Genome Atlas (TCGA), integrating data from 4 different omics platforms and 29 different cancer types.

Biography of the speaker: Dr. Lock is an Associate Professor of Biostatistics at the University of Minnesota. The central theme of his research program is data integration, with applications to systems biology. In particular, he develops methods for the comprehensive analysis of high-content data that are from multiple sources (e.g., gene expression, metabolomics, imaging) or measured in multiple dimensions or ways (e.g., over multiple tissues, regions, or time points). These methods are informed by biomedical collaborations that require multi-source multi-way data integration, and they are widely used. He is leading multiple methodological projects in the area of data integration with funding from the National Institutes of Health (NIH).

From:

<http://www2.math.binghamton.edu/> - **Department of Mathematics and Statistics,
Binghamton University**

Permanent link:

<http://www2.math.binghamton.edu/p/seminars/datasci/092821>



Last update: **2021/11/24 20:32**