Mathematics, Statistics, and Computer Science   **@ UIC**

Statistics and Data Science Seminar

A sparse clustering algorithm for identifying cluster changes across conditions with applications in single-cell RNA-sequencing data

Jun Li (University of Notre Dame)

Abstract: Clustering analysis, in its traditional setting, identifies groupings of samples from a single population/condition. We consider a different setting when the data available are samples from two different conditions, such as cells before and after drug treatment. Cell types in cell populations change as the condition changes: some cell types die out, new cell types may emerge, and surviving cell types evolve to adapt to the new condition. Using single-cell RNA-sequencing data that measure the gene expression of cells before and after the condition change, we propose an algorithm, SparseDC, which identifies cell types, traces their changes across conditions, and identifies genes which are marker genes for these changes. By solving a unified optimization problem, SparseDC completes all three tasks simultaneously. As a general algorithm that detects shared/distinct clusters for two groups of samples, SparseDC can be applied to problems outside the field of biology.

Wednesday, September 18 at 4:00 PM in 636 SEO