

DEPARTMENT OF MATHEMATICAL SCIENCES

Colloquium

Anna Seigal
Harvard
University

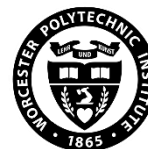
Title: Multi-context principal component analysis

Abstract: Principal component analysis (PCA) finds axes that explain variation in data. Across domains, a current challenge is to understand how data change across contexts (for example, patients across diseases or words across genres). We propose multi-context principal component analysis (MCPCA), a generalization of PCA to find variance-optimal axes of variation and the combinations of contexts in which they appear. Just as usual principal component analysis is a low-rank approximation of the covariance matrix, MCPCA finds a low rank approximation of the third-order tensor obtained by stacking covariance matrices across contexts. I'll describe our algorithm for MCPCA, its theoretical guarantees, and applications to study gene expression across disease types and contextualized word embeddings across genres of text. Based on joint work with Kexin Wang, Salil Bhate, João Pereira, Joe Kileel, and Matylda Figlerowicz.

Friday, September 18

11:00 – 12:00 pm

Stratton 202



WPI