Genevera Allen

Department of Electrical and Computer Engineering, Statistics and Computer Science
Rice University

Data Integration: Data-Driven Discovery from Diverse Data Sources

March 4, 2020
11:45am – 12:45pm
Light refreshments will be served
110 Frelinghuysen Road
Hill Center, Room 552

Abstract: Data integration, or the strategic analysis of multiple sources of data simultaneously, can often lead to discoveries that may be hidden in individual analyses of a single data source. In this talk, we present several new techniques for data integration of mixed, multi-view data where multiple sets of features, possibly each of a different domain, are measured for the same set of samples. This type of data is common in healthcare, biomedicine, national security, multi-senor recordings, multi-modal imaging, and online advertising, among others. In this talk, we specifically highlight how mixed graphical models and new feature selection techniques for mixed, mutli-view data allow us to explore relationships amongst features from different domains. Next, we present new frameworks for integrated principal components analysis and integrated generalized convex clustering that leverage diverse data sources to discover joint patterns amongst the samples. We apply these techniques to integrative genomic studies in cancer and neurodegenerative diseases to make scientific discoveries that would not be possible from analysis of a single data set.

Bio: Genevera Allen is an Associate Professor of Electrical and Computer Engineering, Statistics and Computer Science at Rice University and an investigator at the Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital and Baylor College of Medicine. She is also the Founder and Faculty Director of the Rice Center for Transforming Data to Knowledge, informally called the Rice D2K Lab.

Dr. Allen's research focuses on developing statistical machine learning tools to help scientists make reproducible data-driven discoveries. Her work lies in the areas of interpretable machine learning, optimization, data integration, modern multivariate analysis, and graphical models with applications in neuroscience and bioinformatics. Dr. Allen is the recipient of several honors including a National Science Foundation Career award, the George R. Brown School of Engineering's Research and Teaching Excellence Award at Rice University, and in 2014, she was named to the "Forbes '30 under 30': Science and Healthcare" list. Dr. Allen received her PhD in statistics from Stanford University (2010), under the mentorship of Prof. Robert Tibshirani, and her bachelors, also in statistics, from Rice University (2006).