STATISTICS SEMINAR

Sanvesh Srivastava, Ph.D.
Duke University

Thursday, February 5, 2015
Dickens Hall, Room 207, 4:00-5:00 pm
Refreshments: Dickens 108, 3:30-4:00 pm

Expandable Factor Analysis

Abstract: Modern data are characterized by their large sample-size and complex dependence structure. Bayesian nonparametric methods provide a general probabilistic approach for flexible modeling of such patterns, but they are computationally expensive. In particular, sampling based approaches used for posterior computation (e.g., MCMC methods) scale poorly in the sample-size and parameter dimension. This severely limits the applicability of Bayesian methods in massive data settings. Due to these limitations, Bayesian sparse factor models---a rich class of models that has received much attention recently---faces problems in estimation of high-dimensional loadings matrices and adaptive selection of factors. To address both these issues, we introduce the expandable factor analysis (xFA) framework. Using a novel multiscale generalized double Pareto prior, the xFA framework adaptively selects the required number of factors and enables efficient estimation of low-rank and sparse loadings matrices through weighted L1-regularized regression. Integrated nested Laplace approximations are used for model averaging to accommodate uncertainty in the number of factors and hyperparameters. Theoretical support for the computational algorithm and estimated parameters is discussed, and xFA's performance is demonstrated on both simulated and genomic data.

This talk is based on joint work with David B. Dunson (Duke University) and Barbara E. Engelhardt (Princeton University)