Bayesian Latent Factor Models Recover Gene Networks and Expression QTLs

Dr. Barbara Engelhardt
Princeton University

Abstract

Latent factor models have been the recent focus of much attention in `big data' applications because of their ability to quickly allow the user to explore the underlying data in a controlled and interpretable way. In genomics, latent factor models are commonly used to identify population substructure, identify gene clusters, and control noise in large data sets. In this talk I present a general framework for Bayesian latent factor models. I will motivate some of the structural extensions to these models that have been proposed by my group. I will illustrate the power and the promise of these models for a much broader class of problems in genomics through two specific application to the Genotype-tissue Expression (GTEx) data set. First, I will show how this class of statistical models can be used to identify gene co-expression networks that co-vary uniquely in one tissue type. Second, I will show how these models can be used to identify pleiotropic expression QTLs, or genetic variants that jointly regulate the transcription levels of multiple genes. I will close by describing other uses for these models on genomic applications.

The Johns Hopkins Bloomberg School of Public Health
Department of Biostatistics, Monday, November 10, 2014, 12:15-1:15
Room W3008, School of Public Health (Refreshments: 12:00-12:15)

Note: Taking photos during the seminar is prohibited

For disability access information or listening devices, please contact the Office of Support Services at 410-955-1197 or on the Web at www.jhsph.edu/SupportServices. EO/AA