Statistics Seminar

Wednesday, October 28, 2015
4:15 pm
G01 Biotechnology

Shane Jensen
University of Pennsylvania

Variable Selection with Bayesian Additive Regression Trees

Bayesian Additive Regression Trees (BART, Chipman, George, and McCulloch, 2010) provides a novel nonparametric alternative to parametric regression approaches, such as the lasso or stepwise regression, especially when the number of relevant predictors is sparse relative to the total number of available predictors and the fundamental relationships are nonlinear. However, tree-based approaches such as BART are designed primarily to produce good predictions rather than being used for variable selection. We develop a principled permutation-based inferential approach for determining when the effect of a selected predictor is likely to be real. To demonstrate the potential of our approach in a biological context, we apply it to the task of inferring the gene regulatory network in yeast (Saccharomyces cerevisiae). We find that our BART-based procedure is best able to recover the subset of covariates with the largest signal compared to other variable selection methods.

Refreshments will be served after the seminar in 1181 Comstock Hall.