NYU Stern School of Business
Department of Information, Operations & Management Sciences
STATISTICS RESEARCH SEMINAR

TOPIC: A Lasso for Hierarchical Interactions
SPEAKER: Jacob Bien (Cornell University)
DATE: Friday, November 1st, 2013
TIME: 11:30 AM - 12:30 PM
PLACE: KMC 5-75

ABSTRACT

Building predictive interaction models is an important yet challenging problem, especially when the number of variables is large. Statisticians commonly demand that an interaction only be included in a model if both variables are marginally important. We study the problem of identifying hierarchical two-way interaction models from the viewpoint of the Lasso. By adding a set of convex constraints to the Lasso problem, we produce sparse interaction models that honor the hierarchy restriction. We provide a precise characterization of the effect of the hierarchy restriction and derive an unbiased estimate for the degrees of freedom of our estimator. A simple bound on this estimate gives a sense of the amount of fitting "saved" by the hierarchy constraint. Finally, we distinguish between two types of sparsity: "parameter sparsity" -- the number of nonzero coefficients in the model -- and "practical sparsity" -- the number of raw variables one needs to measure to make predictions in the future. While most statistical procedures focus on the former, the restriction to sparse hierarchical interactions gets at the latter, which is the quantity more closely tied to important data collection concerns such as cost, time, and effort. This is joint work with Jonathan Taylor and Robert Tibshirani.