



UNC
GILLINGS SCHOOL OF
GLOBAL PUBLIC HEALTH

Department of Biostatistics

Bernard G. Greenberg Distinguished Lecture Series

Presented by

Robert John Tibshirani
Professor of Public Health Sciences and Department of
Statistics
Stanford University

Sparse Hierarchical Interactions

ABSTRACT

Building predictive interaction models is a 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 (i.e., L1-penalized regression). We show that by adding a set of convex constraints to the Lasso problem, we can produce sparse interaction models that honor the hierarchy restriction. In contrast to stepwise procedures that are most commonly used for building interaction models, our formulation is convex, and its solution is completely characterized by a set of optimality conditions. This makes it easier to study as a statistical estimator. We argue that restricting to hierarchical interactions can be advantageous both statistically and computationally. We study its properties, give examples and present an efficient computational algorithm. This is based on the PhD thesis work of my student Jacob Bien and is also joint with Jonathan Taylor

Place: Blue Cross Blue Shield Auditorium
Michael Hooker Research Center
Date: Thursday, June 7, 2012
Time: 2:00 – 3:00 PM