MINI-SEMINAR FOR FIRST-YEAR PH.D. STUDENTS

Department of Statistics

Least Angle Regression

by

David Matteson
Department of Statistics, University of Chicago

Wednesday, May 26, 2004, 5:15 pm in Eckhart 110
5734 S. University Avenue

ABSTRACT

The purpose of model selection algorithms such as All Subsets, Forward Selection and Backward Elimination is to choose a linear model on the basis of the same set of data to which the model will be applied. Typically we have available a large collection of possible covariates from which we hope to select a parsimonious set for the efficient prediction of a response variable. Least Angle Regression (LARS), a new model selection algorithm, is a useful and less greedy version of traditional forward selection methods. Three main properties will be discussed: (1) A simple modification of the LARS algorithm implements the Lasso, (2) A different LARS modification efficiently implements Forward Stagewise linear regression, and (3) A simple approximation for the degrees of freedom of a LARS estimate is available, from which we derive a Cp estimate of prediction error; this allows a principled choice among the range of possible LARS estimates. The algorithm described (publicly available) requires only the same order of magnitude of computational effort as ordinary least squares applied to the full set of covariates.