

STANFORD UNIVERSITY
DEPARTMENT OF STATISTICS
DEPARTMENTAL SEMINAR

4:15 p.m., Tuesday, November 9, 2004
Sequoia Hall Room 200
(Cookies at 3:45 in 1st Floor Lounge)

Robert J. Tibshirani
Stanford University

Prediction by Supervised Principal Components

Abstract: In regression problems where the number of predictors greatly exceeds the number of observations, conventional regression techniques may produce unsatisfactory results. We describe a technique called supervised principal components that can be applied to this type of problem. Supervised principal components is similar to conventional principal components analysis except that it uses a subset of the predictors that are selected based on their association with the outcome. Supervised principal components can be applied to regression and generalized regression problems such as survival analysis. It compares favorably to other techniques for this type of problem, and can also account for the effects of other covariates and help identify which predictor variables are most important. We also provide asymptotic consistency results to help support our empirical findings. These methods could become important tools for DNA microarray data, where they may be used to more accurately diagnose and treat cancer.

This is joint work with Eric Bair, Trevor Hastie and Debashis Paul.