

**STANFORD UNIVERSITY
DEPARTMENT OF STATISTICS
DEPARTMENTAL SEMINAR**

4:15 p.m., Tuesday, July 1, 2008
Sequoia Hall Room 200
(Cookies at 3:45 in 1st Floor Lounge)

Bradley Efron
Department of Statistics
Stanford University

LARGE-SCALE PREDICTION PROBLEMS

Classical prediction methods such as Fisher's linear discriminant function were designed for small-scale problems, where the number N of candidate predictors was much smaller than the number of observations n . Modern scientific devices often reverse this situation. A micro- array analysis, for example, might include $n = 100$ subjects measured on $N=10,000$ genes, each of which is a potential predictor. I will discuss "Ebay", an empirical Bayes prediction algorithm designed to handle $N \gg n$ situations. It is closely related to the Shrunken Centroids algorithm of Tibshirani, Hastie, Narasimhan, and Chu.