

STANFORD UNIVERSITY
DEPARTMENT OF STATISTICS
DEPARTMENTAL SEMINAR

4:15 p.m., Tuesday, May 30, 2000
Sequoia Hall Rm. 200
(Cookies at 3:45 in 1st Floor Lounge)

Laura Lazzeroni
Stanford University
Art Owen
Stanford University

Plaid Models

This talk describes the plaid model, a tool for exploratory analysis of multivariate data. The motivating application is the search for interpretable biological structure in gene expression microarray data. Interpretable structure can mean that a set of genes has a similar expression pattern, in the samples under study, or in just a subset of them (such as the cancerous samples).

A set of genes behaving similarly in a set of samples, defines what we call a “layer”. These are very much like clusters, except that: genes can belong to more than one layer or to none of them, the layer may be defined with respect to only a subset of the samples, and the role of genes and samples is symmetric in our formulation.

The plaid model is a superposition of two way anova models, each defined over subsets of genes and samples. We will present the plaid model, an interior point style algorithm for fitting it, and some examples from yeast DNA arrays and other problems.