

PHD SEMINAR ANNOUNCEMENT
Department of Statistics

A Sequential Clustering Algorithm with Applications to Gene Expression Data

Thursday July 31, 2003, 3:00 pm
Eckhart Hall, Room 133, 5734 S. University Avenue

Jongwoo Song
Department of Statistics, University of Chicago

ABSTRACT

Microarrays are part of a new class of biotechnologies which allow the monitoring of expression levels for thousands of genes simultaneously. Gene profile data come from experiments that investigate the behavior of genes over several time points. Biologists are interested in these gene expression profiles because it is believed that genes in the same functional pathway have similar profiles of gene expression.

In the microarray experiments, most of the unsupervised learning processes involve three steps: standardization, defining a dissimilarity measure, and applying a clustering algorithm. We will discuss the issues involved in these steps, and we will propose new methods. A clustering algorithm that finds clusters sequentially and that allows for sporadic objects will be discussed. A new distance measure that takes into account the time order and the time distance between experiments will be introduced.

We will discuss a semiparametric mixture model which is motivated by the sequential clustering algorithm. An EM type estimator for the mixing proportion in semiparametric mixture model is proposed, and its properties are investigated using simulations.