

⇒⇒⇒ PLEASE POST ⇐⇐⇐

COLLOQUIUM ANNOUNCEMENT

Department of Statistics
Virginia Tech

Thursday, September 3, 2009
204 Hutcheson Hall
3:30 p.m.

SPEAKER: Andrew Nobel, Dept. of Statistics & Operations
Research, University of North Carolina, Chapel Hill

TITLE: Finding Significant Large-Average Submatrices in High
Dimensional Data

ABSTRACT:

Exploratory analysis of high dimensional data often begins with independent clustering of samples and variables, yielding a partition of the data matrix into disjoint row-column blocks (submatrices). Of particular interest in practice are submatrices whose entries are large on average. In conjunction with clinical and functional annotation, large average submatrices are frequently the starting point for subsequent analyses, such as the identification of genetic pathways and new disease subtypes.

This talk describes a simple algorithm, belonging to the general category of biclustering methods, for identifying large average submatrices in high dimensional data. Like other biclustering methods, the algorithm improves on independent sample variable clustering in several respects: the submatrices it identifies can overlap and they need not cover the entire data matrix (features that better reflect underlying biology), and the inclusion of samples and variables in a submatrix does not depend on their expression values outside the submatrix. The algorithm seeks to maximize a simple measure of statistical significance, which also provides an objective basis for comparing and selecting among submatrices of different sizes and average intensities. I will discuss the applications of the algorithm to a recent gene-expression based cancer study, and will provide a detailed comparison of its performance with several other biclustering methods, including its application to semi-supervised classification.



☺ Refreshments will be available at Top of the Stairs following the seminar.☺