Statistics Seminar Series

Bayesian Classification of Tumors Using Gene Expression Data

Bani K. Mallick

Department of Statistics, Texas A&M University

Thursday, March 27, 2003, 4:00-5:00pm
Mathematics and Computer Science (MCS) Building, Room 149
111 Cummington Street, Boston

Tea and Cookies at 3:30pm in MCS 153

Abstract: Precise classification of tumors is critical for cancer diagnosis and treatment. Diagnostic pathology has traditionally relied on macro and microscopic histology and tumor morphology as the basis for tumor classification. Current classification frameworks, however, are unable to discriminate among tumors with similar histopathologic features, which vary in clinical course and in response to treatment. In recent years, there has been a move towards the use of cDNA microarrays for tumor classification. These high-throughput assays provide relative mRNA expression measurements simultaneously for thousands of genes. A key goal statistical task is to perform classification via different expression patterns. Gene expression profiles may offer more information than classical morphology and may potentially provide an alternative to classical tumor diagnosis schemes.

First we will select significant genes via expression patterns and use a linear classifier based on these genes for tumor classification. Owing to small sample size and a large number of variables (genes), the selection process can be unstable. We propose a hierarchical Bayesian model for gene (variable) selection. Assignment of a prior distribution over the dimension (number of significant genes) of the model keeps the dimension small.

Next we will consider several nonlinear Bayesian classification methods for the analysis of microarray data based on kernel reproducing Hilbert spaces. We consider the logistic likelihood as well as likelihoods related to the Support Vector Machine (SVM) models. It is shown through three examples that SVM models with multiple shrinkage parameters produce the least amount of misclassification errors in comparison to several existing classical methods as well as Bayesian methods based on the logistic likelihood or those involving only one shrinkage parameter.

For directions and maps, please see *http://math.bu.edu/research/statistics/statseminar.html.*