Bayesian Classification of Tumors Using Gene Expression Data

Bani K. Mallick, Texas A & M University
Debashis Ghosh, University of Michigan
Malay Ghosh, University of Florida

May 6, 2002

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.

This paper considers several Bayesian classification methods for classification 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.

KEYWORDS: Reproducing kernel Hilbert space, microarrays, support vector machines, Markov chain Monte Carlo, Gibbs sampling, Metropolis-Hastings algorithm, shrinkage parameters.