

Bayesian Classification of Tumors Using Gene Expression Data

Bani K. Mallick, Department of Statistics, TAMU 3143, Texas A & M University
College Station, TX 77843-3143, bmallick@stat.tamu.edu

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 talk will consider several Bayesian classification methods for the analysis of microarray data based on reproducing kernel 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.

REFERENCES

- Brown, M.P.S., Grundy, W.N., Lin, D., Cristianini, N., Sugnet, C.W., Furey, T.S. and Haussler, D. (2000) Knowledge-Based analysis of microarray gene expression data by using support vector machines. *Proc Natl Acad Sci*, **97**, 262-267.
- Denison, D., Holmes, C., Mallick, B and Smith, A.F.M. (2002) *Bayesian Methods for non-linear classification and regression*. London: Wiley.
- Efron, B., Tibshirani, R., Storey, J. D. and Tusher, V. (2001) Empirical Bayes analysis of a microarray experiment. *J. Am. Statist. Ass.*, **96**, 1151 –1160.
- Wahba, G. (1999) Support vector machines, reproducing kernel hilbert spaces and the randomized GACV. In Schölkopf, B., Burges, C. and Smola, A. (Eds), *Advances in Kernel Methods*, 69-88.
- West, M., Nevins, J.R., Marks, J.R., Spang, R. and Zuzman, H. (2000) Bayesian regression analysis in the ‘Large p, small n’ paradigm with application in DNA Microarray studies. Technical Report, Duke University.