

On building robust classifiers based on gene-expression

Elier Broche Cristo

Instituto de Matemática e Estatística - USP
Rua do Matão 1010 - Cidade Universitária
05508-090 - São Paulo - SP - Brazil
elier@ime.usp.br

Roberto Hirata Jr.

SENAC College of Computer Science and Technology
Rua Tito, 54 - São Paulo - SP 05051-000 - Brazil
roberto.hirata@sp.senac.br

Eduardo Jordão Neves

Instituto de Matemática e Estatística - USP
Rua do Matão 1010 - Cidade Universitária
05508-090 - São Paulo - SP - Brazil
neves@ime.usp.br

Abstract

The cDNA array technology [1] made the assay of thousands of genes possible and through a well experiment design one can compare which genes are differentially expressed or which genes or group of genes can distinguish between two or more conditions, for instance, normal against cancer [1,2]. In this work we compare two approaches to build classifiers based on the expression of the assayed genes, Linear Discrimination Analysis [3] and Support Vector Machines [4], and how we use bootstrap [5] to have a confidence measurement for the classifier.

- [3] Mosteller, Frederick, Tukey, John W. Data Analysis and Regression: A Second Course in Statistics
- [4] Hastie T., Tibshirani R., Friedman, J. H. The Elements of Statistical Learning: Data Mining, Inference, and Prediction (Springer Series in Statistics)
- [5] Efron, Bradley, Tibshirani, Robert J. An Introduction to the Bootstrap, CRC Press

References

- [1] Schena, Mark Microarray Analysis, John Wiley.
- [2] Parmigiani C., Garrett, E. S., Irizarry, R. A., Zeger S., Clark, G. The Analysis of Gene Expression Data (Statistics for Biology and Health), Springer Verlag