

Models for cancer classification by gene expression data

Ancona N.(1), Maglietta R.(1), D'Addabbo A.(1), Liuni S.(2), Pesole G.(2,3)

(1) Istituto di Studi sui Sistemi Intelligenti per l'Automazione - C.N.R., Bari. (2) Istituto di Tecnologie Biomediche - Sezione di Bari - C.N.R., Bari. (3) Dipartimento Scienze Biomolecolari e Biotecnologie, Università di Milano, Milano.

Motivation

The advent of the technology of DNA microarrays constitutes an epochal change in the study, treatment, analysis, classification and discovery of different types of cancer. The information provided by DNA microarrays allows of approaching to the problem of cancer diagnosis and treatment from a quantitative rather than qualitative point of view.

Methods

Support Vector Machines (SVMs) are the state-of-the-art supervised learning techniques for cancer classification. Other machine learning approaches such as Regularized Least Square (RLS) classifiers may represent a highly suitable alternative for their simplicity and reliability. We then compared the performance of the RLS classifiers, originally proposed in regularization theory, with SVM. The performances of both approaches have been also investigated on three different benchmark datasets, also with respect to the number of selected genes and different gene selection strategies.

Results

We show that RLS classifiers have performances comparable to SVM classifiers expressed in terms of the Leave-One-Out(LOO) error. The main advantage of RLS machines is that for solving a classification problem they use a linear system of order equal to the number of training examples. Moreover, RLS machines allow to get an exact measure of the LOO error with just one training.

Contact email: ancona@ba.issia.cnr.it