

The analysis of molecular profiles and microarray data using Genetic Algorithms

F.Stefanini

Dipartimento di Statistica "G. Parenti", Università di Firenze

Recent techniques of molecular biology have made available a wealth of molecular data that may be used in fingerprinting and gene expression analysis.

The analysis of molecular data, whether they are obtained using simple electrophoresis gels or DNA microarrays, typically involve the use of algorithms to perform feature extraction or methods for pattern recognition.

Even when the quality of data has been tested, the analysis of molecular data remains challenging for the amount of data and the huge size of the features space.

Several techniques have been proposed to achieve a successful analysis, from classical statistical techniques, like principal components, singular value decomposition, block clustering, K-means clustering, neighbor discriminant algorithms, to the more recent Support Vector Machines, Self Organizing Maps, and Genetic Algorithms (GA).

Genetic Algorithms seem particularly interesting because they assist, not replace, more traditional model-based statistical methods.

Here GAs are presented as a tool to perform discriminant and cluster analysis of molecular data.

The core of this approach is based on the selection of an objective function to be optimized.

This step reveals several issues that deserve more general consideration.