

# Soft Computing and Bioinformatics: Principles and Applications

*Sanghamitra Bandyopadhyay and Ujjwal Maulik*

## Abstract

Soft computing is a consortium of methodologies that work synergistically and provides, in one form or another, flexible information processing capabilities for handling real life ambiguous situations. Its aim, unlike conventional (hard) computing, is to exploit the tolerance for imprecision, uncertainty, approximate reasoning and partial truth in order to achieve tractability, robustness, low solution cost, and close resemblance with human like decision-making. At this juncture, Fuzzy Sets (FS), Artificial Neural Networks (ANN), Evolutionary Algorithms (EAs) (including genetic algorithms (GAs), genetic programming (GP), evolutionary strategies (ES)), Support Vector Machines (SVM), Wavelets, Rough Sets (RS), Simulated Annealing (SA), Swarm Optimization (SO), Memetic Algorithms (MA), Ant Colony Optimization (ACO) and Tabu Search (TS) are the major components of Soft Computing.

Over the past few decades, major advances in the field of molecular biology, coupled with advances in genomic technologies, have led to an explosive growth in the biological information generated by the scientific community. Bioinformatics, an area that has evolved in response to this deluge of information, can be viewed as the *use of computational methods to handle biological data*. It is an interdisciplinary field involving biology, computer science, mathematics and statistics to analyze biological sequence data, genome content & arrangement, and to predict the function and structure of macromolecules. The ultimate goal of the field is to enable the discovery of new biological insights as well as to create a global perspective from which unifying principles in biology can be derived. One important sub-discipline within bioinformatics involves the development of new algorithms and models to extract new, and potentially useful, information from various types of biological data including DNA (nucleotide sequences) and proteins (amino acid sequences). Analysis of these macromolecules is performed both structurally and functionally.

The tutorial will be in three parts. In the first part, we will describe the basic principles of some soft computing techniques. In this regard, particular emphasis will be placed on evolutionary algorithms, fuzzy theory and neural network. Their basic concepts will be described, and their implementation details will be discussed with several illustrative examples. Real-life applications will be demonstrated.

In the second part of the tutorial, the basic principles of Bioinformatics and molecular biology will be described with explanatory figures. In particular the basics of genomics and proteomics will be covered. The tutorial will be designed from first principles, and will cover sufficient preliminary matter so as to be understandable to the layman, as well as be of interest to the practitioner.

Recently, use of soft computing tools for solving bioinformatics problems have been gaining the attention of researchers because of their ability to handle imprecision, uncertainty and near optimality in large and complex search spaces. The last part of the tutorial will describe how application of soft computing becomes relevant for solving some Bioinformatics and molecular biology problems. Two recent applications, viz., in the domains of protein classification and structure-based ligand design, will be discussed in detail. Protein classification leads to identification and proper functional assignment of uncharacterized proteins with a final goal towards finding homologies and drug discovery. Again, structure based ligand design is one of the crucial steps in rational drug discovery, where a small molecule is designed by targeting the structure and biochemical properties of the target. The applications of soft computing to the problems of protein classification and structure based ligand design correspond to some of the recent works by the speakers.