

Format of Registration Form

Name (in BLOCK Letters) WORKSHOP - Data mining in Biology

Designation (Check your category) Researcher in Academia Researcher in Industry

Teacher Institute _____

Research Scholars

Field _____

Institute _____

Address _____

City _____

Status _____

ZIP/PIN _____ Phone _____ Fax _____

E-mail _____

Registration Fee(s): INR Rs.1500.00 (Academics)
INR Rs.3000.00 (Industry)
* Accommodation will not be provided.

Mode of Payment Bank Draft in favour of ' Finance Officer, JNU '
payable at New Delhi.
Enclosed Bank Draft No. _____ Dated _____
Drawn on _____

Date: _____ Signature of the Applicant _____

* Kindly post the registration form along with fee and recommendation letter. Advance copy may be sent by email to nsrao@mail.jnu.ac.in

Maximum Seats 20
Eligibility criteria

Completed M.Sc/ M.Phil in Biology/Physics/Chemistry/Bioinformatics/ Biotechnology/B.Pharma/M.Pharma/BE/B.Tech. Involved in teaching & research work related to Computational Biology and Bioinformatics. Preference will be given to the Teachers & Research Scholars. Basic Knowledge in Statistics and computers is required. Purpose or Motivation for applying the workshop. Recommendation letter from faculty/Head of the Institute. Only one person from each institute will be entertained. Only selected candidate will be informed via email.

Last Date of registration : 30th Sept. 2009
Last Date of Confirmation of Participation : 6th Oct.,2009

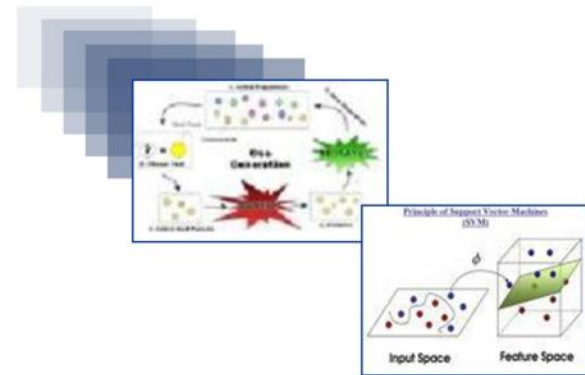
No Accommodation will be provided to the participants.

Correspondence **Dr. N.Subbarao**, Workshop Coordinator
Centre of Excellence
Centre for Computational Biology & Bioinformatics
School of Information Technology
Jawaharlal Nehru University
New Delhi-1100677.
Tel. : 011-26704082 Fax :26741586

E-mail:nsrao@mail.jnu.ac.in
nsrao.jnu@gmail.com



Centre for Computational Biology
& Bioinformatics



WORKSHOP ON

Data mining in biology Applications of Clustering, SVM, Genetic algorithms & Ant Colony Optimization

October 20-25 , 2009

Venue : School of Information Technology , Room 9
JNU, New Delhi - 110067.

WORKSHOP ON

" Data mining in biology: Applications of Clustering, SVM, Genetic algorithms & Ant Colony Optimization ".

SHORT DESCRIPTION

Genetic algorithms (GA) are based on the principle of evolution. Starting from a population of potential solutions, the algorithm evolves better solutions by means of natural selection and genetic operations like cross-over and mutation that are akin to biological systems. Support vector machines (SVM) is a powerful learning algorithm firmly based on statistical learning theory and structural risk minimization principle. The effectiveness of GA is mainly because of its derivative free search and ease of modeling, and that of SVM is due to its excellent, learning and generalization capabilities. Fuzzy logic is a precise logic of imprecision and approximate reasoning, an effective technique in decision analysis in fuzzy environment.

The workshop will introduce the participants to clustering GA, SVM and their applications in the areas of chemo- and bio- informatics. The application case studies would include gene identification, cancer detection, and identifications of protein functions, Quantitative Structure-Activity Relationships, protein structure prediction similarity search and phylogenetic analysis, etc.

COURSE CONTENTS

INTRODUCTION TO BASIC STATISTICS

CLUSTER ANALYSIS

- Different measures of similarity/ Dissimilarity
- Clustering
 - Hierarchical (agglomerative & Devisive)
 - KNN
 - PCA
- Supervised /Unsupervised Clustering

SUPPORT VECTOR MACHINES

- Introduction
- Introduction to Pattern recognition
- Illustrations of Pattern recognition
- Supervised and Unsupervised Learning
- Binary Classification
- Introduction to Support Vector Machines
- Introduction to Hyperplanes and Concept of Margin
- Optimal Separating Hyperplane: Linearly Separable Problems
- Formulation of and Solution Methodology of QP Problem
- The Concept of Support Vectors
- The SVM Decision Function
- Linearly Non-separable Problems: Incorporation of Misclassification
- Errors
- Generalization in Higher Dimensional Feature Space
- Introduction to Kernel Functions
- Explanation of SVM = Linear Hyperplane in the feature space, but computations enabled in input space itself by virtue of Kernel functions

- Final Decision Function in feature space
- Selection of Kernels
- Selection of Model Parameters
- Cross Validation Error
- Leave-One-Out-Error
- Introduction to Multi-Class SVM Classification
- Introduction to Single Class SVM Classification
- Introduction to SVM Regression
- Case Studies in Bioinformatics and Chemo- informatics
- Cancer Detection
- Familiarization with data from microarrays and next generation sequencing
- Identification of Protein Functions
- QSAR/qSAR
- Gene Identification
- Arrhythmia Classification

GENETIC ALGORITHMS

- Introduction, brief history, ideas of population and evolution
- Simple genetic algorithm, definition and mechanism
- Representation, selection, crossover, mutation, objective function evaluation
- An example of how GA works
- Similarity search as a case study
 - Formulation of the problem
 - Representation and Objective function
 - Discussion
- Protein structure Prediction as a case study
 - Problem definition
 - Representation
 - Selection and genetic operations
 - Energy as Objective function
 - Analysis of results
 - Results for villin, crambin
 - Results of hybrid algorithms
- Discussion of variations in algorithms

RESOURCE PERSONS

Dr. Narinder Sahni School of Information Technology, JNU
Dr. Rashi Gupta School of Information Technology, JNU
Dr. V. K. Jayaraman Former Scientist in NCL, C-DAC, Pune.
Dr. V. Sundararajan C-DAC, Pune.

Workshop Coordinator

Dr. N.Subbarao School of Information Technology, JNU