THE UNIVERSITY   **OF HONG KONG**

Department of Mathematics

Numerical Mathematics and Applied Analysis Group Seminar (NMAA)

Tutorial on Bioinformatics and Machine Learning Methods

Chris Ding

Lawrence Berkeley National Laboratory

June 28, 2004 (Monday)
T3, Meng Wah Complex, HKU

Session 1:   10:30am – 12:00noon
Session 2:   2:30pm – 4:00pm

Abstract

The fast evolving trends in bioinformatics and computational genomics are to use machine learning methods to computationally determine functions, structures, interactions, among DNAs and proteins with biological significance. For example, using classification methods, one can predict protein 3D structures, RNA coding regions, binding/non-bind active sites, etc. In this tutorial, I will cover several areas where machine learning methods are most widely and fruitfully adopted.

The tutorial will have 5 sections:

(1) Genomic Basics: DNA, Proteins, annotation, sequence alignment, hidden Markov models

(2) Function and Structure predictions using classification methods. This is the most widely adopted in current research and applications. Protein fold recognition. RNA coding region detection. Binding site detection. Cancer type detection.

(3) Feature extraction and selection. This is necessary for effective application of classification methods. More important, most relevant features often have direct biological significance regarding to the biological system in question. Filter methods. Wrapper methods.

(4) Unsupervised class, feature, phenotype discovery through data clustering. This supplements biological expertise and help to achieve new and deeper insights.

(5) Biological networks: gene regulation networks, protein interaction networks. The fastest growing research area. Network structure deduction. Bayesian networks. Linear models. Multi-protein complexes identification. Connectivity analysis and clustering.

Contact: Michael Ng (mng@maths.hku.hk)

All are welcome