Bioinformatics II

Instructor:

Asst. Prof. Stephen Winters-Hilt
Phone: (504) 896-2761; (504) 280-2407
E-mail: winters@cs.uno.edu

Time/Location:

Lecture Location: Math 229; Lecture Hours: MW 1:00-2:15
Office Location: CERM 217; Office Hours: MW 10:30-12:00

Prerequisites: CSCI 2125, MATH 2314, or permission of instructor.

Textbooks (required):

Biological Sequence Analysis by R. Durbin *et al*. Cambridge University Press (1999). ISBN 0-521-62971-3.

Class notes

Reference Books (optional):

Programming Perl (3rd ed.) by Larry Wall *et al*. O'Reilly Media (2000). ISBN 0-596-00027-8.

Background:

The Cheminformatics & Protein Channel Biophysics course introduces channel current cheminformatics, the biophysics of protein channels, and single molecule biochemistry/biophysics. Single molecule biochemistry and protein channel biophysics are at a multidisciplinary nexus that is seeing rapid growth in scientific and technological applications. Nanopore-based single molecule detection is a promising new "nanobiotech" application. As with many of the new technologies at the nanometer-scale, signal analysis and adaptive pattern recognition needs *require* that the multidisciplinary collaboration also include *extensive use of modern informatics methods to fully enable the new device physics*. The course focuses on real-world examples and offers bioinformatics methods for channel current analysis as well as lab work options involving single protein channel studies and nanopore-based single-biomolecule studies (for those with sufficient lab skills). This course involves introductory material similar to that given in CSCI 4567, but turns its focus to channel current cheminformatics and to the interdisciplinary Biomedical engineering and Biophysics issues that the informatics tools will help to address, while CSCI 4567 focuses more on computational genomics applications. Either this course, or CSCI 4567, serve as prerequisites to the introductory and advanced CSCI projects courses: 4589, 4590, 4595; 6589, 6590, 6595. This course will include program writing for data mining, as well as electrical signal analysis for biomedical informatics. There will be a large project component to the course with a wide selection of problems, from programming intensive informatics solutions to data gathering optimizations in biophysics experiments.

Course Abstract:

An introduction to the algorithms and theory used in bioinformatics and cheminformatics, with current applications in computational genomics and biomedical informatics Covers statistical methods for identifying motifs in biosequences and identifying features in channel currents. Includes hidden Markov models for identifying structure in stochastic sequential data (for gene finding and for feature extraction from protein-channel ionic current measurements) and discriminative methods for use in informatics, particularly Support Vector Machine approaches.

Course Objectives:

Task decomposition.

Students should understand how to decompose a complex informatics task into a collection of standard informatics tasks: feature identification and knowledge discovery, signal acquisition and filtering, feature extraction, classification, and data-rejection.

Method selection.

Students should understand how to analyze the general properties of their data and factor in their computational limitations in order to select the most efficient informatics method at each stage of the task decomposition.

Real-world deployment.

Students should be familiar with training and testing in a real computational environment (including simple distributed computational arrangements on a networked cluster of computers to the extent that time permits).

Performance optimization.

Students should understand how to obtain statistically valid (objective) scores of performance and how to use that information for performance optimization.

Grading:

(A) 90-100; (B) 75-89; (C) 65-74; (D) 55-64; (F) below 55.

Homework assignments	50%
Midterm	20%
Final Project	30%

Students will have opportunities to do the following:

1. Follow the most recent research in the field covered in the course
2. Solve the real-world informatics problem using techniques covered in the class
3. Provide incisive critiques to current research and point out some potential research directions
4. Final project is mature enough for journal paper submission

Policies:

- Most of the assignments can done with others
- Final Projects must be done individually
- Homework is due in class on due date specified
- Omit documentation in your code at your own risk

Topics Covered:

Shannon information theory and applications	*week 1*
Power Signal Analysis (Wavelet & Fourier analysis)	*week 2*
Hidden Markov Models (HMMs)	*week 3*
HMM theory and applications to biological sequence analyses	*week 4*
Unsupervised Learning	*week 5*
Model based clustering and heuristic clustering	*week 6*
Midterm	
Supervised Learning	*week 7*
Feature selection and access classification error	*week 7*
Support Vector Machines	*week 8*
SVM models with frequently used kernels, feature selection etc.	*week 9*
Tree-based machine learning methods	*week 10*
CART, Random Forest and applications	*week 11*
Cheminformatics: Single Molecule Classification	*week 12*
Final projects presentations	*week 13-14*