EECS800 -- Special Topics in Mining Biological Data
Fall, 2006

Course Goals and Requirements: The analysis of large volume of data has been playing a central role in the exciting Bioinformatics and Computational Biology research. The field of biological data analysis has evolved rapidly and has built connections to various research fields in Computer Science including data mining, database, and machine learning, and fields in biology also.

The primary goal of this seminar course is to survey current data analysis techniques in Bioinformatics and Computational Biology research. We will discuss research papers covering a wide range of topics bridging Computer Science research and Biology research. The common theme of these papers is to develop or utilize computational techniques to understand the structure of data. Our understanding of biological data may come from recognizing patterns in the data, from constructing descriptive models of the data, or from building predictive models for the data. We will also try to organize the field of biological data analysis e.g., by rigorously defining problems that arise in the setting, by identifying core algorithmic techniques useful in the domain, and by proposing guidelines for future systems and software development.

Instructor: Jun Huan, assistant professor
Office: 2034 Eaton Hall
Phone: (785) 864-4620 (department phone number)
Email: jhuan@eecs.ku.edu
Office Hours: I will be in my office after each class meeting for at least an hour (unless otherwise announced). Please feel free to stop by my office anytime. If you need an appointment, send me an email/call.

Class Info: This class will meet M/W 9:00 – 10:15 am at Eaton Hall, Room 2001

There is the class web page:
http://people.eecs.ku.edu/~jhuan/fall06.html
Please check it often!

Prerequisites:
No. General knowledge of algorithm is assumed.

References (not required textbooks):

[1] *Data Mining: Concepts and Techniques, by Jiawei Han & Micheline Kamber, Morgan Kaufmann, 2001. (ISBN: 1-55860-489-8)*

[2] *The Elements of Statistical Learning --- Data Mining, Inference, and Prediction, by Hastie, Tibshirani, and Friedman, Springer, 2001. (ISBN: 0-387-95284-5)*

[3] *Bioinformatics: Genes, Proteins, and Computers, Christine Orengo, David Jones, Janet Thornton edit, Bios Scientific Publishers, 2003. (ISBN: 1-85996-0545)*

Grading:
Class Presentation: 45%
Projects: 45%
Class participation: 10%

Special Needs: If you need special accommodation for any reason, I will make every reasonable attempt to meet your needs. However, it is your responsibility to discuss this with me in the first few days of class.

Tentative Course Schedule (subject to change according to class progresses)**:**

August 21	First day of class
August 23	Association rules, basics
August 28 – September 1	Association rules, advanced techniques
September 4 – September 8	Mining Microarray data (I)
September 11 – September 15	Analyzing sequences, trees, and graphs
September 18 – September 22	Motifs in gene sequence and protein structures
September 25 – September 29	Descriptive models of data: K-mean, PCA
October 2 - October 6	Subspace clustering
October 9 – October 13	Mining Microarray data (II)
October 16 – October 20	Predictive models of data: Decision tree, LDA
October 23 – October 27	Kernel methods and applications
October 30 – November 3	Text mining, gene ontology, & bio-data management
November 6 – November 10	Selected topics I: Systems biology, proteomics
November 13 – November 17	Selected topics II: Analyzing biological networks
November 20 – November 24	Selected topics III: Data integration
November 27 – December 1	Class presentation
December 4	Class presentation
December 6	Last day of class