Syllabus Fall 2008
BINF/STAT 5354 Post-Genomic Analysis
BINF 5113 Mathematics Seminar for Bioinformatics
Multivariate Data Analysis and Probabilistic Modeling
T 9:00 - 11:20am, R 9:00am - 12:20pm Bell Hall 130A

INSTRUCTORS: Naijun Sha and Ming-Ying Leung

OFFICE: Bell Hall 203 (Sha) and 225 (Leung)

PHONE: 747-6844 (Sha) and 747-6836 (Leung)

E-MAIL: nsha@utep.edu; mleung@utep.edu

WEBPAGE: http://www.math.utep.edu/Faculty/naijun/sha.html
http://www.math.utep.edu/Faculty/mleung.html

OFFICE HOURS: MW: 1:00-2:00pm (Sha) or by appointment;
Dr. Leung will post her office hours for each week outside her office.

DEADLINE for Automatic W: Oct. 31, 2008

Text: Applied Multivariate Statistical Analysis, 6th edition by R. A. Johnson and D. W. Wichern

Course Description: This course and seminar focus on statistical methods and probability models for analyzing post-genomics data. It consists of two modules:

1. Statistical analysis of a multivariate response. Topics covered: Descriptive multivariate statistics, multivariate normal distribution, principal component analysis, classification and clustering analysis. Applications with the use of statistical packages will be considered. Prerequisite: STAT 5380, STAT 5385 or equivalent, or consent of instructor.
2. Probabilistic modeling for nucleic and amino acid sequences. Topics covered: Markov chains and Hidden Markov Models (HMM). Probabilistic approaches to sequence alignment, phylogeny, and RNA structure analysis. Prerequisite: Instructor approval.

Course Objectives: Students will learn to apply appropriate probabilistic models and statistical techniques to address bioinformatics problems such as biomolecular sequence and microarray data analyses. Through assigned projects, class discussions, hands-on labs, as well as oral presentations in the seminar, students will also acquire skills to identify bioinformatics problems that require advanced mathematical and statistical knowledge, and to describe and discuss these issues with individuals with suitable expertise.

Assignment: Homework and data analysis projects will be assigned throughout the semester. You may do the exercises with your partner, but each student must answer the questions individually. ZERO grades will be obtained for those whose solutions are the exact copies of someone else's. NO LATE HOMEWORK WILL BE ACCEPTED.

Attendance: Class attendance is required and helpful to decide borderline grades. If a student has to be absent from a particular lecture or lab, he/she will be responsible for catching up with the rest of the class and for arranging with the instructors regarding making up missed work.

Grading:

BINF 5354: Homework 30%
Midterm Exam 30%
Final/Project 40%

BINF 5113: Each student will be evaluated by the quality of his/her own assigned presentations as well as their contribution to the discussions during other students' presentations.

The final grade is based on a scale of 90 − 100 = A, 80 − 89 = B, 70 − 79 = C, 60 − 69 = D, below 60 = F. Attendance and class participation will be used to help decide borderline grades.

Suggested References:

1. An Introduction to Multivariate Statistical Analysis by T. W. Anderson. Wiley.
2. Methods of Multivariate Analysis by Alvin C. Rencher. Wiley.
3. Multivariate Data Analysis by J. F. Hair, R. L. Tatham, R. E. Anderson and W. Black.
4. Data Mining: Concepts and Techniques by J. Han and M. Kamber. Morgan Kaufmann.
5. Methods of Microarray Data Analysis eds by J. Showmaker and S. Lin. Wiley.
6. Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids by R. Durbin, S. Eddy, A. Krogh and G. Mitchison.