Course data:
Department: The University of Texas at Austin, Department of Computer Sciences
Instructor: Professor William H. Press
Now being given: Spring, 2011
Previously offered: Spring, 2008; Spring, 2009; Spring, 2010.
Meets for lectures: MW 1:30 - 3:00 PM in PAI 3.14, and occasionally F 1:30 - 3:00 in WEL 4.224
Meets for discussion section and computational workshop: F 1:30 - 3:00 in WEL 4.224
Current Information for Spring 2011:
For up-to-date information see the CS395T/CAM383M Discussion Forum. The rest of this page has information about the course in general.
Course description:
This is a practical course in applying (mostly) modern statistical techniques to (mostly) real data, particularly bioinformatic data and large data sets. There is only a small amount of theorem proving; the emphasis is on efficient computation and concise coding, mostly in MATLAB (where we learn various data-parallel language idioms) and C++ (which we learn to interface seamlessly to MATLAB for convenience and computational power).
Topics covered:
Topics covered include probability theory and Bayesian inference; univariate distributions; Central Limit Theorem; generation of random deviates; tail (p-value) tests; multiple hypothesis correction; empirical distributions; model fitting; error estimation; contingency tables; multivariate normal distributions; phylogenetic clustering; Gaussian mixture models; EM methods; maximum likelihood estimation; Markov Chain Monte Carlo; principal component analysis; dynamic programming; hidden Markov models; performance measures for classifiers; support vector machines; Wiener filtering; wavelets; multidimensional interpolation; information theory.
Lecture notes and concept list:
A detailed course outline, with links to complete lecture notes (PDF slide files) is here, from Spring, 2010. The list of concepts associated with each lecture is provided as a study guide. Earlier versions from previous years are also available: 2009, and 2008.

(Instructors at other institutions may obtain PowerPoint versions of these files on request.)
Prerequisites:
Graduate standing, or upper-division undergraduate with consent of instructor. Mathematics at least including undergraduate multivariable calculus and linear algebra is assumed, as well as some programming experience in MATLAB, C++, and/or Java (or, possibly, Mathematica, Fortran, or C). A previous course in undergraduate level probability and statistics is helpful, but not required.
Texts:
There is no required text. However, many lectures will utilize methods in Numerical Recipes, Third Edition. Enrolled students will be provided with a free electronic subscription to this book, as well as access to its source code.

Some other relevant books are:

- Bishop, Pattern Recognition and Machine Learning
- Hastie, Tibshirani, and Friedman, The Elements of Statistical Learning
Course requirements:
Enrolled students are expected to attend lectures. Occasional (not regular) problem sets or computer exercises will be assigned. Students are expected to contribute to the course wiki. A student project or paper is required (may be collaborative). No written exams, but there will be individual final oral interviews (20-30 min.) covering the lecture material.

2010 course forum:
The 2010 course discussion forum has additional materials, links to references, and discussion contributed by students. (Also includes archived 2009 course forum.)

2008 course wiki:
The 2008 course wiki has an earlier version of the lecture notes, with some discussion threads and contributions by students. (Unfortunately, some of these are hard to find, since they are attached to individual slide links in the lectures.)