Computational Statistics

Spring 2016

Instructor:	Emre Neftci	**Time:**	Fr 1:00 – 3:40
Email:	eneftci@uci.edu	**Place:**	SBSG 2200.

Course Overview: Computational statistics is an approach in statistics that focuses on computer intensive algorithms to process data into knowledge. "The idea is to rely on computer algorithms to solve problems in statistical analysis that would be too difficult to solve using analytic techniques by themselves. By using a computational approach, the researcher can often approximate solutions to complex problems without being forced to make unrealistic assumptions (e.g. normality and independence) required for some analytic techniques."

This course will mostly focus on the part of computational statistics that is concerned with latent probabilistic models, such as Bayesian networks (belief nets). The goal here is to develop probabilistic models for the data that can be used to explain and describe the data as well as make predictions about future data. In Bayesian model, it can often be challenging to analyze the posterior distributions over the variables of interest. A number of computational procedures have been developed to approximate posterior distributions using sampling techniques. One such approach is known as Markov chain Monte Carlo.

In this course, the goal is to give students a practical, programming-based introduction to latent probabilistic modeling. The course will focus on sampling techniques such Metropolis-Hasting and Gibbs sampling, both forms of Markov chain Monte Carlo.

This is a programming-based course, with programming-based assignments interleaved with lectures. Students will be required to attend a weekly three hour laboratory session and completing assigned programming exercises in class. In the first weeks, the class will be taught with 30-minute lectures followed by in-class programming exercises. The preferred programming language used throughout the course is Python. However, all course material will also be available in MATLAB and students may use any other language for their projects. Students will need a laptop preferably with Python or MATLAB already installed (including the statistics toolbox for MATLAB; and numpy, scipy, matplotlib libraries for Python). During the remaining weeks the students will work on a computational statistics project of their choice (ideally a topic related to their research focus) and present it orally on week 10 to the class. Projects will be judged on the results, difficulty of the problem and quality of the oral presentation.

This course will follow Mark Steyvers' textbook.

Key References: http://www.scipy-lectures.org/intro/index.html (sections 1.1 through 1.4)

Course Pages: http://nmi-lab.org/teaching

Office Hours: Fr 4::00 - 5:00 PM at SBSG 2308 or by appointment.

Tentative Course Schedule:

Course Introduction ..... Apr 1
Random Variables – Sampling ..... Apr 8
Monte Carlo Sampling – Rejection Sampling and Importance Sampling ..... Apr 15
Markov Chain Monte Carlo Sampling – Metropolis-Hastings and Gibbs Sampling . Apr 22
Bayesian Data Analysis ..... Apr 29
Directed Graphical Models (Belief Networks) ..... May 6
Approximate Inference in Graphical Models ..... May 13
Projects ..... May 20, 27
Project presentations and Q&A ..... Jun 3

Grading: Assignments (20%), Projects (40%), Final (40%).

Important Dates:

Midterm ..... May 3, 2016
Final Exam ..... May 17, 2016