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. 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 in-class programming-based exercises based on the textbook. 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 and present it orally on week 10 to the class. Ideally, this project should related to the students' research and with data already collected.

This course will follow Mark Steyvers' textbook, which can be found on the Canvas website.

Key References:

[1] Geof H Givens and Jennifer A Hoeting. *Computational statistics*. Vol. 710. John Wiley & Sons, 2012.

[2] Ioana A Cosma and Ludger Evers. "Markov chains and Monte carlo methods". **African Institute for Mathematical Sciences, Cape Town** (2010).

[3] C.M. Bishop. *Pattern recognition and machine learning*. Springer-Verlag New York, Inc. Secaucus, NJ, USA, 2006.

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

Course Pages:

Office Hours: Tu Th 11::00 - 12:30 PM at SBSG 2308.

Tentative Course Schedule:

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

Grading: Assignments (30%), Project (70%).

Projects: Students will chose and complete a computational statistics project of their choice and present it orally on week 10 to the class. Ideally, this project should related to the students' research and with data already collected. Students must write a report using a research article template (e.g., Abstract, Methods, Results, Discussion), on the project and present it to the class. Grading of the projects will be based on the report and the oral presentation.

Important Dates:

Final Presentations ..... Jun 3, 2016