NYU COURANT INSTITUTE OF MATHEMATICAL SCIENCES

Computational Statistics
MA-GY 6973
Fall 2021

Course Syllabus

Cross listing	MATH-GA 2080
Lecture	Wednesday 11:00am - 1:30pm 2 MTC 804 (Brooklyn)
Instructor	Mike O'Neil CIWW 1119 and 2 MTC 854 cims.nyu.edu/~oneil oneil@cims.nyu.edu 212-998-3125
Course website	cims.nyu.edu/~oneil/compstat21

Prerequisites
Undergraduate-level proficiency in linear algebra and multivariable calculus; undergraduate-level proficiency in probability and statistics; programming experience required.

Description
Computation plays a central role in modern statistics and machine learning. This course aims to cover topics needed to develop a broad working knowledge of modern computational statistics. We seek to develop a practical understanding of how and why existing methods work, enabling effective use of modern statistical methods. Achieving these goals requires familiarity with diverse topics in statistical computing, computational statistics, computer science, and numerical analysis. Specific topics include: intro to numerical linear algebra, regression and Gaussian processes, Newton's method and optimization, numerical integration, random variable generation, Markov chain Monte Carlo (MCMC) and variance reduction, the Bootstrap, density estimation, and an introduction to modern methods in machine learning (neural networks and deep learning).

Materials
Much of the course material will be drawn from the following texts, both of which are available for free via NYU Libraries:

- Efron and Hastie, *Computer Age Statistical Inference*, Cambridge, 2016.
- Gentle, *Computational Statistics*, Springer, 2009.

Assignments
There will be a mix of homework assignments, an in-class midterm, and a final project. There will be four homework assignments during the semester that will consist of both written and computational

(computer programming) work. Some of the assignments may require the use of statistical software or libraries, such as STAN, julia, etc.

Grading

The overall course grade will be determined from a final numerical weighted average. The following breakdown will be used to compute an overall numerical grade:

40% Homework (10% each)
30% Midterm
30% Final project

The final project (possibly done in collaboration with another student) will require a short presentation to the class, as well as submission of a report composed in LaTeX. There must be a programming component to the project, and you must submit your code along with the final report.

Weekly schedule

1. Probability and random variables review
2. Root finding and optimization
3. Numerical linear algebra
4. Dimensionality reduction and principal component analysis
5. Regression and Gaussian processes
6. Function approximation
7. *Midterm*
8. Numerical integration, random variable generation
9. Markov chain Monte Carlo Methods
10. The Bootstrap
11. Density fitting
12. Expectation maximization
13. The LASSO
14. Introduction to neural networks and deep learning
15. *Project presentations*

Academic Integrity

This course will adhere to the academic integrity statement issued by the NYU Tandon School of Engineering, available at: https://engineering.nyu.edu/campus-and-community/student-life/office-student-affairs/policies/student-code-conduct