
ISyE 6416 - Computational Statistics - Spring 2020

Tentative Syllabus

Computational statistics is an interface between statistics and computing. We will cover algorithms for solving statistical problems in terms of estimation and model fitting, hypothesis test/classification, prediction/generalization, and uncertainty quantification. We will talk about how to use optimization and other modern computational tools to develop algorithms. Course will cover the following cluster of topics:

Deterministic algorithms

- Optimization methods for statistics, solving MLE
- EM algorithms for computing MLE with “unknowns”: Gaussian mixture model (GMM), Hidden Markov models (HMM) for sequential data
- LDA/QDA, SVM, Logistic regression
- Linear models and splines
- Model selection and cross-validation
- Low-rank models, PCA, and recommender systems

“Random” algorithms and probabilistic methods

- Bootstrap
- Random forest
- Monte Carlo methods
- Markov Chain Monte Carlo

Class Time and Location: MWF, 11:15-12:05pm, MRDC 2404.

Instructor: Prof. Yao Xie, Groseclose #339, email: yao.xie@isye.gatech.edu

Instructor Office Hour: Monday 12:05pm-1:05pm, Groseclose 339.

Class TA: Rui Zhang (ruizhang_ray@gatech.edu)

TA Office Hour: TBA.

Class Website: Canvas

Class material available on our website includes

- Announcements
- Course syllabus
- Homework assignments and solutions
- Slides and other lecture material
- Practical exams
- Your course grades on exams and homework
- Any important announcements

Important: All homework submissions will be done electronically via Canvas. Make sure the scores in Canvas are consistent with what you got. We will not make any change in grading for works older than 2 weeks.

Class Mailing List: Registered students are automatically subscribed to the class mailing list.

Textbook: the course material will be based on lectures and slides posted on Canvas.

References:

The elements of Statistical Learning: Data Mining, Inference, and Predictions, 2nd edition, Trevor Hastie, Robert Tibshirani, and Jerome Friedman.

Machine learning: A probabilistic perspective, K. P. Murphy.

Foundations of Machine Learning, M. Mohri, A. Rostamizadeh, A. Talwalkar.

Computational statistics. G. H. Givens and J. A. Hoeting, 2013.

Computational Statistics. James E. Gentle. 2009.

Numerical Analysis for Statisticians, Kenneth Lange.

An introduction to statistical learning: with applications in R, G. James, D. Witten, T. Hastie, R. Tibshirani.

Computational Statistics handbook with MATLAB, W. L. Martinez and A. R. Martinez.

Elements of Computational Statistics, J. E. Gentle.

Numerical linear algebra, Lloyd N. Trefethen and David Bau III.

Numerical Recipes, 3rd edition, William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery.

Prerequisites: Undergraduate level basic probability, linear algebra, and statistics.

Software: MATLAB and/or R. You will use this software for homework assignment.

Grading Policy: Class Participation 9% (there will be a few quizzes), Submitting Teaching Evaluation - 1%, Homework - 15%, Midterm 1 - 15%, Midterm 2 - 15%, Project - 15%, Final - 30%.

Homework: The homework should be handed in electronically on canvas before the end of the class on the due date. The lowest homework-score will be dropped. If you submit the homework late: one day late the grade will be discount to 75% of your total, two days late the grade will discount to 50% of your total, three days late the grade will discount to 25% of your total. Past three days, your homework will not be accepted.

Assignments will include both exercises and computer problems; the computer problems will ask you to carry out statistical analysis using computer statistical software. Keep in mind that you should not hand in raw computer output. Conclusions and interpretation of results are more important than good printouts.

You are allowed to work together with other students on homework, as long as you write up and turn in your own solutions. You are also allowed (and encouraged) to ask me questions, although you should try to think about the problems before asking. Request for re-grading the Homework/Exams/Quizzes should be made within a week of returning Homework/Exams/Quizzes.

Important: Make sure the scores in Canvas are consistent with what you got. We will not make any change in grading for works older than 2 weeks.

The following grading scale, with scores rounded to the nearest whole number, will be used in the course:

- 90-100%: A
- 80-89%: B

- 70-79%: C
- 60-69%: D
- below 60%: F

PLAGIARISM

Plagiarism is considered a serious offense. You are not allowed to copy and paste or submit materials created or published by others, as if you created the materials. All materials submitted and posted must be your own original work.

STUDENT HONOR CODE

You are responsible for completing your own work. All students are expected and required to abide by the *letter* and the *spirit* of the Georgia Tech Honor Code. The teaching assistants and I will also abide by these honor codes. I am very serious about this expectation because ethical behavior is extremely important in all facets of life. To review the Georgia Tech Honor Code, please visit <http://osi.gatech.edu/content/honor-code> . Any OMS Analytics degree student suspected of behavior in violation of the Georgia Tech Honor Code will be referred to Georgia Tech's Office of Student Integrity.

ACCOMMODATIONS FOR STUDENTS WITH DISABILITIES

If you are a student with learning needs that require special accommodation, contact the Office of Disability Services at (404) 894-2563 or <http://disabilityservices.gatech.edu/>, as soon as possible, to make an appointment to discuss your special needs and to obtain an accommodations letter. Please also e-mail me as soon as possible in order to set up a time to discuss your learning needs.

Student-Faculty expectations <http://www.catalog.gatech.edu/rules/22/> This spells out what the students and instructor are expected to do, including what you can require of the students, such as attendance.

Formal excuses for absence: Officially, there are 2 channels a student can get formal excuses. A student can request from Dean's office for illness or other personal issues by filling out a form at https://gatech-advocate.symplicity.com/care_report/index.php/pid122802?. The other way is to request from registrar for institute approved such as attending conference, religious observances (within first 2 weeks) <https://registrar.gatech.edu/info/institute-approved-absence-form-for-students>.

Exams: There will be two in-class midterm exams during the class and one final exam during the exam period. All exams are open book. You are not allowed to use your cell phone nor laptop during the exam. There are **no make-ups**; if there is any conflict, please let us know ASAP and you may take the exam before the assigned period.

Midterm 1: week of Feb 3, in class. Midterm 2: week of March 9, in class.

Final Exam: Friday, April 24, 11:20am-2:10pm.

Project: by group, each group consists of 2-4 students. Details see project guideline. There will be a required short presentation towards the end of the semester.