ECE 4803: Mathematical Foundations of Data Science
Fall 2020 Syllabus

Summary

ECE 4803 is an introduction to the mathematical foundations of data science and machine learning. The central theme of the course is the use of linear algebra and optimization in posing and solving modern problems leveraging data focusing on applications in ECE. Upon successful completion of this course you should be able to:

- pose real-world engineering tasks as optimization problems and recognize common kinds of optimization problems when you encounter them,
- efficiently compute the solutions to large-scale optimization problems, and
- understand how to combine data with mathematical models to solve practical engineering problems.

Prerequisites

Introductory courses in linear algebra and multivariable calculus are the main pre-requisites. Most of the course will use the language of matrices and vectors. Students should be comfortable with the use of matrices to represent systems of equations and the notion of taking a gradient of a function of many variables – some existing familiarity with eigenvalues, eigenvectors, and eigenvalue decompositions will be extremely helpful. Familiarity with the basics of probability (especially random variables, expected value, and related notions) will also be useful in understanding some of the applications covered in the course. Finally, students should also have basic Python programming skills.

Instructor

Mark Davenport
Email: mdav@gatech.edu
Office hours: Tuesdays/Thursdays from 9:30-10:45. Additional availability by appointment.

Instruction Modality

In Fall 2020, ECE 4803 will be taught in a hybrid mode. What this will mean in this course is that, while we do have a room on campus reserved and available (Howey N210 on Tuesdays and Thursdays from 9:30-10:45am), I plan to actually deliver all lectures remotely. These will be pre-recorded and available to watch at whatever time is most convenient for you. I will reserve our regularly scheduled class hours for (initially online) discussions, working extra examples, answering questions about the homeworks, etc. These discussions will also be recorded for those who cannot attend.

I expect that the course will be fully online (meaning that our discussions and office hours will be conducted via BlueJeans) for at least the first 6 weeks of class (September 24). After that point, I hope to move these discussions to the campus during our regularly scheduled class time. In the event that we do this, I still plan to pre-record my technical lectures and use the class time for discussions and student presentations. I can continue to record these in-person meetings for those who cannot or do not want to attend. For those students who are uncomfortable attending in-person activities, it will be possible to complete this course fully online.

(Note that the current estimate for the capacity of Howey N210 with social distancing is 8 – if we move to the campus and there are more than 8 of us who wish to attend in person, we may have to create a rotating schedule for in-person attendance.)

Grading

Your grade will be based on the following factors, with the following allocations:

- **Homeworks (35%):** There will be ≈ 10 homework assignments. See further details below.
- **Quizzes (35%):** There will be ≈ 4 regularly spaced quizzes every ≈ 3 weeks to be completed remotely. These are tentatively scheduled for September 15, October 1, October 20, and November 12.
- **Semester project (25%):** In lieu of a final exam, we will have an in-depth project on a topic of your choosing. These will be conducted either alone or in small groups of 2-3 students. The final project will include a short presentation/video due by the last day of class (November 24), and a short write-up due at the end of finals period (December 8). Additional details will be provided later in the semester.
- **Participation (5%):** This is based on my assessment of your engagement in the course. This will be based on factors such as participation in online/classroom discussions and engagement during office hours and/or on Piazza.

Your final grade will be assigned as a letter grade according to the scale:

A: 90-100% B: 80-89% C: 70-79% D: 60-69% F: $\leq 59\%$

I may exercise the option to "curve" quiz scores as necessary (by adjusting the grades higher, but not lower) if I determine that a quiz was more difficult than I intended.

Homework

Homework will be assigned weekly (approximately). **Homework will be turned in via canvas. Unless you have made prior arrangements with me, late submissions will get zero credit.** Each homework assignment will be graded out of 100 points. Over the course of the semester, **the maximum number of homework points that you can earn is** $(N-1)\cdot 100$**, where** N **is the number of assignments** – this serves a similar role to allowing you to drop one homework assignment, but should encourage you to still submit a partially completed one (and avoid panic if there is occasionally a problem that you do not finish in time.)

The homework assignments will be hard; many of them will require significant amounts of time and effort to complete. But this is really where most of the learning takes place. You will get out of the assignments what you put into them. Students who complete all of the assignments in full will be rewarded with a deep understanding of the role that linear algebra and optimization play in data science, machine learning, and signal processing (among other things). Effectively, homework is worth much more than 35% of your grade. In teaching many courses over the years, **neither instructor has ever seen a case where a student does not put effort into the homework assignments but does well on the exams**.

Course materials

The course webpage is at: mdav.ece.gatech.edu/ece-4803-fall2020. This page will provide general course information, links to lecture notes/videos, and homework assignments. Homework assignments and solutions will also be posted in canvas, as they become available.

I also plan to post recordings of our discussions that we will hold in lieu of a traditional lecture. Since, if you speak up, you will be included in this recording, I will post these only to canvas.

I also plan to make exclusive use of Piazza to make announcements and answer questions. This site can be accessed via: piazza.com/gatech/fall2020/ece4803/home. Piazza is a great platform to discuss problems, find study groups, etc. Please direct any questions you might have to Piazza. Unless your questions are personal in nature, please do not make private posts – if you have a question you are probably not the only one, and other students may benefit from seeing the discussion.

There is no required text. Below is a list of books that the instructors have found helpful over the years for learning (and teaching) the material in this class.

Linear algebra

- Strang: *Linear Algebra and its Applications* [amazon]
- Trefethen and Bau: *Numerical Linear Algebra* [amazon]
- Horn and Johnson: *Matrix Analysis* [amazon]

Probability and statistics

- Bertsekas and Tsitsiklis: *Introduction to Probability* [amazon]
- Durrett: *Elementary Probability for Applications* [amazon]
- Wasserman: *All of Statistics* [amazon]

Optimization

- Boyd and Vanderberghe: *Convex Optimization* [Available as a free pdf from author's website]
- Calafiore and El Ghaoui: *Optimization Models* [amazon]
- Nocedal and Wright: *Numerical Optimization* [amazon]
- Ben-Tal and Nemirovski: *Lectures on Modern Convex Optimization* [amazon]
- Luenberger: *Optimization by Vector Space Methods* [amazon]

Machine learning

- Watt, Borhani, and Katsaggelos: *Machine Learning Refined* [amazon]
- Strang: *Linear Algebra and Learning from Data* [amazon]
- Hastie, Tibshirani, and Friedman: *The Elements of Statistical Learning*, [Available as a free pdf from author's website]
- Abu-Mostafa, Magdon-Ismail, and Lin: *Learning from Data.* [amazon]

Potpourri

- Mlodinow: *The Drunkard's Walk: How Randomness Rules our Lives* [amazon]
- Bernstein: *Against the Gods: The Remarkable Story of Risk* [amazon]
- Taleb: *Fooled by Randomness* [amazon]
- Silver: *The Signal and the Noise* [amazon]
- Domingos: *The Master Algorithm* [amazon]
- Schneier: *Data and Goliath* [amazon]
- O'Neil: *Weapons of Math Destruction* [amazon]

Course Expectations and Guidelines

COVID-19 considerations

All Georgia Tech faculty and students are required to wear face coverings while inside campus facilities/buildings. This includes classrooms and offices. If/when we transition to on-campus meetings, you must wear an appropriate face covering. This, however, is not a substitute for other social distancing measures.

As noted above, you may complete this course fully online if you have any discomfort with attending in-person activities.

Academic integrity

Georgia Tech aims to cultivate a community based on trust, academic integrity, and honor. Students are expected to act according to the highest ethical standards. For information on Georgia Tech's Academic Honor Code, please visit www.catalog.gatech.edu/policies/honor-code. Any student suspected of cheating or plagiarizing on a quiz, exam, or assignment will be reported to the Office of Student Integrity, who will investigate the incident and identify the appropriate penalty for violations.

Redistributing materials from this course and/or using external sites for assistance (e.g., contributing to test banks, CourseHero, Chegg, or similar sites) is prohibited.

Collaboration and group work

Students are *strongly* encouraged to discuss homework problems with one another. However, **each student must write up and turn in their own solutions written in their own words**. **Cases where solutions appear to be identical or nearly identical will be immediately referred to the Office of Student Integrity.**

Absences, late assignments, and missed quizzes

Active participation in the class discussions is a factor in your grade. You are expected to attend these discussions unless you have a compelling reason not to do so. However, you will not be penalized for any excused absences (e.g., due to illnesses, religious observances, career fairs, job interviews, etc.) I plan to discuss the homework assignments in class the day after they are due, and thus **I cannot accept late homeworks**. In the event that an excused absence prevents you from submitting an assignment, your homework grade will be calculated on a pro-rated basis. **Quizzes will be completed remotely, but during normal class hours. If you expect to miss a quiz, please contact me as soon as you realize this so we can make alternative arrangements.** We may consider options to take the quiz at an alternate time or instead may adjust the grading allocation to place more emphasis on other quizzes, depending on the circumstances.

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 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 agreement

At Georgia Tech we believe that it is important to strive for an atmosphere of mutual respect, acknowledgement, and responsibility between faculty members and the student body. In the end, simple respect for knowledge, hard work, and cordial interactions will help build the environment we seek. Therefore, I encourage you to remain committed to the ideals of Georgia Tech while in this class. See www.catalog.gatech.edu/rules/22 for an articulation of some basic expectation that you can have of me and that I have of you.

Digital etiquette

Most of this course will occur online. Active participation in the Piazza forum and in remote discussions will be critical to replicating the in-person experience we would all prefer to be having. Please come to our discussion sessions prepared with questions to ask. In general, I would prefer if everyone leaves their video on. It is much easier to have a conversation when we can see each other, and I often rely on this feedback to know when I say something that doesn't actually make sense. Ultimately, however, I do understand that internet connections do not always allow for this, and I will not penalize you if you must disable your video.

Outline

The outline below should be treated as an approximation; it is subject to changes.

1. The method of least squares
   (a) Applications and formulation: Regression and interpolation
   (b) Solving least squares problems (Review of multivariable calculus)
   (c) Understanding least squares problems (Review of linear algebra)
   (d) Computing least squares solutions
2. Unconstrained optimization
   (a) Convex optimization
   (b) Gradient descent
   (c) Conjugate gradients
   (d) Acceleration: The heavy ball method and Nesterov's optimal method
   (e) Newton's method and quasi-Newton methods
   (f) Non-smooth optimization
   (g) Stochastic gradient descent
   (h) Applications: Approximation, filter design, tracking, logistic regression, neural networks
3. Constrained optimization
   (a) Lagrange duality
   (b) The KKT conditions
   (c) Algorithms for constrained optimization
   (d) Linear programming
   (e) The simplex algorithm
   (f) Second order cone programs
   (g) Semidefinite programs
   (h) Applications: Support vector machines, portfolio optimization, feature selection, optimal power flow, recommendation systems
4. Beyond convex optimization
   (a) Integer programming
   (b) Dynamic programming
   (c) Optimization on graphs
   (d) Optimization in game theory
   (e) Applications: Error correction, optimal control, reinforcement learning, generative adversarial networks