

Topics in Data Science

ECE 512

Fall 2023

Instructor. Dror Baron
e-mail: dzbaron AT ncsu DOT edu
Office hour (online): Mondays 2 PM - 3 PM
<https://barondror.com>

Teaching assistant: Hangjin Liu
e-mail: hliu25 AT ncsu DOT edu
Office hour (online): Friday 11 AM - 12 PM

Course Text. The course borrows material from various sources, and there is no single textbook. The following may be useful at times.

- C. M. Bishop, Pattern Recognition and Machine Learning, 2006.
- D. MacKay, Information Theory, Inference, and Learning Algorithms, 2003.
- M. Mohri, A. Rostamizadeh, and A. Talwalkar, Foundations of Machine Learning, 2012.
- T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, 2001.
- T. H. Cormen, C. E. Leiserson, and R. L. Rivest, Introduction to Algorithms, 1990.
- S. Mallat, A Wavelet Tour of Signal Processing, 1999.

Some of you may want to download electronic versions of these books.

Prerequisites.

In addition to eagerness to learn about data science, students must have taken ECE 301 or equivalent (Fourier transforms), ECE 410 or 510 (analog to digital conversion, filters), probability (e.g., ST 371), linear algebra, and calculus. Students should also be comfortable with programming (we will be using Matlab and/or Python; details on page 4). Past experience suggests that students who wanted to take the course because it's "timely" or "cool" but lacked the mathematical background did not end up satisfied. On the other hand, former students have informed me that the course provides an overview of the data science area, and helped them connect dots in other courses.

Course purpose. ECE 512 (Data Science from a Signal Processing Perspective) will acquaint students with some core topics in data science. Specific topics covered will include:

1. **Scientific computing** - concepts related to efficient scientific programming such as computational complexity and basic data structures.
2. **Optimization** - we'll skim topics such as dynamic programming, linear programming, convex optimization, integer programming, and the expectation-maximization (EM) algorithm.
3. **Machine learning basics** - some of the topics we'll discuss are classification, decision theory, regression, clustering, and subset selection.
4. **Sparse signal processing** - wavelets and compressed sensing.
5. **Dimensionality reduction** - including principle components analysis (PCA).

Finally, you will learn to solve data science problems numerically using software. In particular, you will be able to apply a methodology to data science problems that involves looking at the problem, translating it to mathematics, proposing an algorithm, and implementing it in software.

Course Objectives. Upon completion of this course, students will be able to:

1. Apply probabilistic tools to select proper models for data.
2. Analyze the computational complexity of an algorithm.
3. Produce efficient scientific code that uses core data structures, if needed, and confirm that the code is efficient using profiling.
4. Apply standard optimization tools such as linear programming and convex optimization to data.
5. Apply machine learning techniques such as classification, regression, and clustering to data.
6. Separate data into training and test data to validate algorithms' usefulness.
7. Apply sparsifying transforms such as Fourier and wavelets to data.
8. Acquire and recover sparse signals.

Policies and Procedures. Academic integrity is important in ECE 512, because many of you will be running algorithms on large data sets in several years. If you cut corners in your future professional work, it could lead to lawsuits, you could be fired, and in extreme cases people could be injured or die. (If these comments seem far fetched, recall that multiple autonomous car makers have had to explain to the public why their vehicles crashed.)

Students should refer to the University policy on academic integrity. Here are some specific expectations we have.

- When working on homeworks and the final project (you are encouraged to submit in pairs or triples), students can certainly work together and submit together; each student should make sure that they understand different aspects of the problems being worked through, and that the assignment helped them grasp concepts taught in class.
- We expect your final projects to include a survey of related techniques and papers you may have gone through. These works should be referenced with a citation.¹
- Our tests (including quizzes) will be administered electronically through Moodle. When working on tests, no cooperation or “collaboration” between students is allowed. While it could be tempting to text or email a friend during a test that is administered electronically, this is not allowed. You will be allowed to use your notes, books, a browser, and software such as Matlab and/or Python.² However, while working on the test you should not text, email, or communicate with other people (certainly not other students) in any way, unless you are consulting with the course staff. **By submitting the test, you will be acknowledging that you completed the work on your own without the help of others in any capacity.** Any such aid would be unauthorized and a violation of the academic integrity policy.

All cases of academic misconduct will be submitted to the Office of Student Conduct; the recommended penalty will be a failing grade for the entire course.

Homework. We expect homeworks due during the semester roughly every 1–2 weeks; a tentative schedule appears on the course webpage. Students will submit homework individually, in pairs, or in triples. Assignments and the schedule for submitting them will be posted on the course web site.

Final projects. The *final project* will involve a topic that pairs or triples of students choose to work on. The project may involve reading a paper and summarizing it for the entire class, or perhaps working on some data set using an algorithm that wasn’t covered in depth in class; an excellent project will involve novel work by the group. The final project will require the group to submit a report and present the work to the entire class; the formats required for the report and presentation will be specified during the semester. The project will be evaluated in part based on peer grading. Overall, the objective of the final project is to give students a personalized learning experience and an opportunity to present the findings to the entire class and receive ample feedback. Students will submit final project reports in pairs or triples.

Lead class discussions. At the beginning of each class, one of the students (schedule forthcoming) will lead a discussion about possible applications of the material (covered in Youtube modules) that will appear in that class. This is not intended to be a fancy presentation. Get up in front of the class, explain possible applications for 5 minutes, and collect your 5% (the grade structure appears on page 5). Basically, start a conversation about why the material is potentially useful.

¹If students are unsure how to cite, Dr. Baron will be glad to post examples on the course webpage.

²You can use the browser to access Moodle, the course webpage, and look up technical topics. Similar to a normal test, you must not communicate with other people.

Quizzes. Academic research has shown that studying consistently leads to better retention and mastery of the material than crashing for an exam. And because the material is available in video modules on Youtube, you can indeed pace yourself through the semester. However, in previous years, many students did not watch modules ahead of class, and class attendance was low. To incentivize you to study consistently and watch the modules ahead of time, most weeks there will be a short quiz on Moodle about that week's material. Our objective is for these quizzes to be easy; watch the modules ahead of time, take the test, and collect your 5% "brownie points." And **please** do attend class, it'll help create a positive atmosphere. **To offset the possibility that you will have a problem that prevents you from taking some quiz, the lowest grade among the quizzes will be dropped.**

Keep in mind that homework will be 30% of the grade, and final projects 20% (the grade structure appears on page 5). Homeworks and projects will be submitted electronically via Moodle; the schedule for submitting them will be posted on the course web site.

Late submissions. Unless you received permission in advance (see below), homeworks and projects should be turned in an electronic copy by midnight on the due date, unless authorized ahead of time (see the next paragraph). In case of an electronic malfunction, we will give you a grace period of 10 minutes to email your submission to the TA. After that, late submissions will be penalized 20%; and each hour, the grade will dip by another 2%. To offset the possibility that such a penalty may impact your grade, the lowest grade among the homeworks will be dropped.

Requesting to submit something late 2–3 days ahead of the deadline is absolutely reasonable, but doing so the morning it's due isn't. **As a guiding principle, the nicer you are to the course staff, the nicer they're likely to be to you.**

Programming. The projects will involve Matlab and/or Python programming. A free Matlab download is available on the EOS website:

<http://www.eos.ncsu.edu/software/downloads/>

And a link to a tutorial on Python:

<https://docs.python.org/3/tutorial/>

Tests. We will have multiple tests during the semester. Details about the test schedule will be published on the course webpage. The tests will be open-book, open-notes. Owing to the anticipated online nature of tests, computers will be allowed. See comments about intellectual integrity on page 2. (The aforementioned guiding principle also applies here. Students who are taking multiple courses with tests scheduled at inconvenient times must consult with Dr. Baron early on during the semester. Realizing that you have another test the same day in another course 3 days before that test is inappropriate. Of course, situations such as being sick or going to a conference will be handled with greater flexibility. We do not want you to take a test if you are truly feeling bad.)

Tests will be on Moodle, and available during a 24 hour window. Once you start working on a test, you should submit your work within 90 minutes (class is 75 minutes, plus 15 minutes to upload your work and so on). In case you have problems submitting the test, contact

us immediately. A student who submits late (after more than 90 minutes) without having contacted us will be penalized 1% per minute late. For example, a student who submitted after 2 hours (30 minutes late) and scored 90% will get $90-30=60\%$.

Extra credit. Extra credit of up to 2–3% will be allowed. Extra credit will be allocated based on factors such as class participation, feedback about assignments, attendance of office hours, and overall contributions to the course. Details about how much extra credit is allocated to different activities will not be published. Note that Dr. Baron has typically allocated roughly 1% extra credit, on average. The bottom line is that you are encouraged to contribute to a pleasant course experience!

Grading.

Tests	40%
Homework	30 %
Final project	20%
Quizzes	5%
Lead class discussion	5%

Weighted averages of 90, 80, and 70 will guarantee *minimal* letter grades of A-, B-, and C-, respectively. During semesters when grades were lower, “curving” was sometimes used.

Instructors’ commitment. You can expect your instructor to be courteous, punctual, well organized, and prepared for class activities; to answer questions clearly and in a non-negative fashion; to be available during office hours or to notify you beforehand if they are unable to keep them; and to grade uniformly and consistently according to the posted guidelines.

Disabled students. North Carolina State University is subject to the Department of Health, Education, and Welfare regulations implementing Section 504 of the Rehabilitation Act of 1973. Section 504 provides that: “No otherwise qualified handicapped individual in the United States . . . shall, solely by reason of his handicap be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance.” This regulation includes students with hearing, visual, motor, or learning disabilities and states that colleges and universities must make “reasonable adjustments” to ensure that academic requirements are not discriminatory. Modifications may require rescheduling classes from inaccessible to accessible buildings, providing access to auxiliary aids such as tape recorders, special lab equipment, or other services such as readers, note takers, or interpreters. It further requires that exams actually evaluate students’ progress and achievement rather than reflect their impaired skills. This may require oral or taped tests, readers, scribes, separate testing rooms, or extension of time limits.

Schedule. A tentative schedule appears on the course webpage. As we progress through the semester, the schedule will be updated periodically. While the university has scheduled a final exam, we will have 3 tests that follow the format described above. Barring major disruptions, there will not be a final exam.

Class Evaluations. Online class evaluations will be available for students to complete; this will happen toward the end of the semester. Students will receive an email message directing them to a website where they can login using their Unity ID and complete evaluations. All evaluations are confidential; instructors will never know how any one student responded to any question, and students will never know the ratings for any particular instructors.