Topics in Data Science
ECE 592
Fall 2022

Instructor: Dror Baron
e-mail: dzbaron AT ncsu DOT edu
Office hour (online): Mondays 2 PM - 3 PM
http://people.engr.ncsu.edu/dzbaron

Teaching assistant: Hangjin Liu
e-mail: hliu25 AT ncsu DOT edu
Office hour (online): Friday 2 PM - 3 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.
- M. Mohri, A. Rostamizadeh, and A. Talwalkar, Foundations of Machine Learning,
  2012.

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 a firm command of
undergraduate signal processing (ECE 421) and probability (ST 371), and be comfortable
with math (linear algebra, calculus, multi-dimensional spaces) and programming (we will
be using Matlab and/or Python). 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 592 (Topics in Data Science) 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, and in particular we will use Matlab and/or Python. 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:** By the end of the semester, the student should be able to:

- Analyze the computational complexity of an algorithm.
- Familiarity with key data structures including graphs.
- Produce efficient scientific code, and make sure that it works well using profiling.
- Apply standard optimization tools such as linear programming and convex optimization.
- Apply standard machine learning techniques such as classification, regression, and clustering to data.
- Know how to develop a model with training data and validate its usefulness on test data.
- Use sparsifying transforms such as Fourier and wavelets on data.
- Acquire and recover sparse signals.
- Apply principle components analysis to data sets.
- Develop software (and in particular using Matlab and/or Python) for solving data science problems.
Policies and Procedures: Academic integrity is important in ECE 592, 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 will likely 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. 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. All the students will be providing feedback (including the grade) to each other. Overall, the objective of the final project is to give students a personalized learning experience while providing an opportunity to present the findings to the entire class and receive ample feedback. Students will submit final project reports in pairs or triples.

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

\(^2\)You 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.
Keep in mind that homework will be 40% of the grade, and final projects 15%. Homeworks and final projects will be submitted electronically via Moodle. Assignments and 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. Late submissions will immediately be penalized 50%; after 24 hours, no credit will be given.

Requesting to submit something late 2–3 days ahead of the deadline is 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. Exceptions (permitting late submissions without advance notice) will only be made in emergency situations such as medical problems.

*Matlab:* 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 3. (The aforementioned guiding principle also applies here. Students who are taking multiple courses with tests scheduled at inconvenient times should consult with Dr. Baron early on during the semester. In contrast, special circumstances such as a student going to a conference will be handled with greater flexibility.)

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 (a standard class is 75 minutes, plus 15 minutes to scan your work, upload it, 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 lose 30%, resulting in a grade of 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, however, that Dr. Baron has typically allocated roughly 1% extra credit for typical students. The bottom line is that you are encouraged to contribute to a pleasant course experience!

*Grading:*

<table>
<thead>
<tr>
<th>Component</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework</td>
<td>40%</td>
</tr>
<tr>
<td>Final project</td>
<td>15%</td>
</tr>
<tr>
<td>Tests</td>
<td>45%</td>
</tr>
</tbody>
</table>
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 detailed tentative schedule appears on the course webpage. As we progress through the semester, the schedule will be updated periodically. The final exam is scheduled by the university for December 9, 2022, 12:00–2:30 PM. That said, it is likely that the last among our tests will have the same grade weight, length, and electronic format as other tests.

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.