Topics in Data Science
ECE 592
Fall 2019

Instructor: Dror Baron
e-mail: dzbaron AT ncsu DOT edu
Office hour: Mondays 1-2 PM in EB2 2097
http://people.engr.ncsu.edu/dzbaron

Teaching assistant: Hangjin Liu
e-mail: hliu25 AT ncsu DOT edu
Office hour: Tuesday 10 AM - 12 PM in EB2 3034

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.

Some of you may want to download electronic versions of these books. Standard web searches
often lead to pdf files 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).
Course purpose: ECE 592 (Topics in Data Science) will acquaint students with some core basic topics in data science. Specific topics covered will include concepts related to efficient scientific programming such as computational complexity and basic data structures; optimization; machine learning basics such as classification, regression, and clustering; wavelets and sparse signal processing; and dimensionality reduction and principle components analysis.

Finally, you will learn to solve data science problems numerically using software, and in particular we will use the 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.

Detailed objectives that will help prepare for the midterm and final exam will be posted on the course website prior to these tests.

Policies and Procedures:

Homework: 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.

Projects: We expect 4 “homework style” projects during the semester, and one final project.
• **Homework style projects** will require you to derive some math, working out software solutions, and look at data. Each such homework style project will involve an application, and we hope that you will be able to better appreciate how data science is used in many real world settings. Students will submit homework style projects individually, in pairs, or in triples.

• The **final project** will be a topic that pairs or triples of students choose to work on. This could be 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. The ideal 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 presentation style will be a brief lecture in class. We envision having two classes dedicated to presentations. 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.

Keep in mind that projects will be **40%** of the grade, with half (20%) for “homework style” projects, and the other half for the final project.

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 Dr. Baron and Ms. Liu, the nicer they’re likely to be to you. Exceptions (permitting late submissions without advance notice) will only be made in emergency situations.

**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:** There will be a final exam at the end of the semester. The test will be open-book, open-notes. Computers will not be allowed; calculators are allowed. Students who are unable to take the test at the time determined by the university (8-11 AM on December 13) should inform the instructor at the beginning of the semester (no later than August 31, 2019), and an alternate arrangement may be formulated. (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.)

Extra credit: Extra credit of 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. The bottom line is that you are encouraged to contribute to a pleasant and energetic atmosphere in class!

Grading:

- Homework 15%
- Projects 20%
- Final project 20%
- Midterm 20%
- Final 25%

Weighted averages of 90, 80, and 70 will guarantee minimal letter grades of A-, B-, and C-, respectively. Grades are often somewhat lower, and “curving” will be used to provide a reasonable average GPA.

Instructors’ commitment: You can expect your instructor to be courteous, punctual, well organized, and prepared for lecture and other 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; to provide a suitable guest lecturer or post pre-recorded lectures online when they are traveling; 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 13, 2019, 8:00–11:00 AM.
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.