Topics in Data Science
ECE 592
Fall 2016

Instructor: Dror Baron
e-mail: dzbaron@ncsu.edu
Office hours: Wednesdays 1-2 PM in EB2 2097
http://people.engr.ncsu.edu/dzbaron

Teaching assistant: none

Message Board: The course message board is an excellent resource for understanding problems that come up during class, understanding homeworks, providing feedback to the instructor, etc. Students are strongly encouraged to use it.

Course Text: As we are developing the course this semester and it borrows material from various sources, there is no single textbook. The following may be useful.

- 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. Standard web searches often lead to pdf files of these books.

Prerequesites:
The main prerequisite is eagerness to learn about data science. True technical prerequisites are somewhat informal, and include comfort in math (especially linear algebra and probability) and comfort with computers (specifically, we will be using Matlab).

Course purpose: ECE 592 (Topics in Data Science) will acquaint students with some core basic topics in data science. Specific topics covered will include computational complexity, basic data structures, scientific programming, optimization, wavelets, sparse signal processing, 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 software package. 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:

- Apply standard machine learning techniques such as classification, regression, and clustering to data.
- 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.
- 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) for solving data science problems.

More detailed objectives that are relevant to specific chapters covered in the final exam will be posted on the course website prior to these tests.

Policies and Procedures:
Academic integrity: Students should refer to the University policy on academic integrity found in the Code of Student Conduct. Authorized aid on an individual assignment includes discussing the interpretation of the problem statement, sharing ideas or approaches for solving the problem, and explaining concepts involved in the problem. Any other 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.

Homework: Students will submit homework individually or in pairs. Assignments and the schedule for submitting them will be posted on the course web site.

Projects: We expect 2-3 "homework style" projects during the semester, and one individual project.

- *Homework style projects* will be a comprehensive assignment that combines deriving some math, working out some Matlab solutions, and possibly looking at data. Each such homework style project will involve some 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 or in pairs.
- The *individual project* will be a topic that individual students (possibly pairs 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 project will involve submitting a (brief!) report and presenting the project to the entire class. The presentation style will be to either provide a brief lecture in class, or using a poster. We envision having one class dedicated to presentations, and another dedicated to posters. Each student can choose which session they want to present their project in, and all the students will be providing feedback (including the grade) to each other. Overall, the objective of the individual project to give the student a personalized learning experience while providing an opportunity to present the findings to the entire class and receive ample feedback. Students will submit individual projects individually (or possibly in pairs - TBD).

Keeping in mind that projects will be 50% of the grade, half of this (50%) will be for "homework style" projects, and the other half (25%) for the individual project.

Projects will be submitted in hard copies to the instructor in class. Assignments and the schedule for submitting them will be posted on the course web site.

Late submissions: Homeworks and projects should be turned in a hard copy by the end of class on the due date. Late submissions will immediately be penalized 50%; after 24 hours, no credit will be given. Exceptions will only be made in emergency situations.

Matlab: The projects will involve Matlab programming. A free Matlab download is available on the EOS website:

`http://www.eos.ncsu.edu/software/downloads/`

Tests: There will be a comprehensive final exam. The test will be open-book, open-notes. Computers are absolutely not allowed; calculators are allowed. Students who are unable to take the test at those times should inform the instructor at the beginning of the semester (no later than September 10, 2016), and an alternate arrangement may be formulated.

Extra credit: Up to 2% of extra credit will be allowed. Extra credit will be allocated based on factors such as class participation, message board participation, and feedback about assignments. The bottom line is that you are encouraged to contribute to a pleasant and energetic atmosphere in class!

Grading:

Homework	25%
Projects	50% (half each for HW-style & individual)
Final	25%
Extra credit	2%

Weighted averages of 90, 80, and 70 will guarantee *minimal* letter grades of A-, B-, and C-, respectively.

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:
Final: December 9, 2016, 8:00–11:00 AM

Class Evaluations: Online class evaluations will be available for students to complete between November 18 and December 4.

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.