ECE 6254: Statistical Signal Processing
Spring 2016 Syllabus

January 11, 2016

Summary

This course will explore a range of modern "data driven" approaches to signal processing. In contrast to most traditional approaches to signal processing, in this course we will focus on *how to learn effective models from data* and *how to apply these models to practical signal processing problems*. We will approach these problems from the perspective of statistical inference. We will study both practical algorithms for statistical inference and theoretical aspects of how to reason about and work with probabilistic models. We will consider a variety of applications, including classification, prediction, regression, clustering, modeling, and data exploration/visiualization.

Prerequisites

Throughout this course we will take a statistical perspective, which will require familiarity with basic concepts in probability (random variables, expectation, joint distributions, independence, conditional distributions, Bayes rule, etc.). We will also be using the language of linear algebra to describe the algorithms and carry out any analysis, so you should be familiar with concepts such as norms, inner products, orthogonality, linear independence, eigenvalues/vectors, eigenvalue decompositions, etc. If you have had courses on these topics as an undergraduate (or more recently) you should be able to fill in any gaps in your understanding as the semester progresses. Finally, many of the homework assignments and the course projects will require the use MATLAB (or a similar programming language), so some familiarity with MATLAB or an equivalent will be very helpful.

Instructor

Justin Romberg
Email: jrom@ece.gatech.edu
Office: Centergy 5227
Office phone: 404-894-3930
Office hours: I will typically hold scheduled office hours the day before homework is due. More details will be provided soon. I am also available to meet in Centergy 5227 by appointment.

Teaching Assistant

TBA
Email: TBA
Office hours: TBA

Grading

Your grade will be based on the following factors:

- **Pre-test (5%):** During the first week of class there will be a take-home "pre-test" which will review the basic concepts from linear algebra and probability theory that we will be using in this course. This is an open-book/internet test, so you should feel free to consult whatever outside resources you like, but **you must work through this test on your own.** The purpose of this pre-test is to help everyone get on the same page in terms of what you need to know in order to succeed in this course.

- **Homework (25%):** There will be 6 ± 1 homework assignments, concentrated towards the first half of the semester. They will consist of exercises, proofs, and MATLAB implementations. I expect your write-ups to be very clear; I do not just want you to produce correct answers — I want you to demonstrate that you understand the material and write your solutions as if you were explaining your answer to a colleague. Style matters and will be a factor in the grade.

  Homework will be turned in at the beginning of lecture. **Late homework will get zero credit.**

  You are encouraged to discuss the homework with other members of the class. However, everything that you turn in must be your own work. **You must write up the assignments (and accordingly the MATLAB code) by yourself, citing any outside references you use to arrive at your solution. Failure to do so will be considered a violation of the GT Academic Honor Code.** For any questions involving these or any other Academic Honor Code issues, please consult me or www.honor.gatech.edu.

- **Midterm exam (20%):** The midterm exam will occur in-class at about mid-semester (tentatively scheduled for March 16) and will cover the same material as the homeworks.

- **Final project (25%):** During the second half of the semester, we will shift from traditional homework assignments towards an in-depth project on a topic of your choosing. These projects will be done in groups of 3–5 students. The project will have several graded components, including a proposal, a presentation, and written report. The presentations will be in the format of a poster session, tentatively scheduled for the second-to-last day of class. The written report will be due at the end of the finals period. Further details about the final project will be provided later in the semester.

- **Final Exam (20%):** The final exam will occur at the designated time during the final period.

- **Participation (5%):** This part of your grade is based on my assessment of your engagement in the course. This will be based on **attendance**, participation in classroom discussions, engagement outside of the classroom (such as during office hours and/or on Piazza), and on peer evaluations from study/project groups.

Study groups and peer grading

Students are required to form small study groups of 3–5 students each. There will be at least one "group question" in each homework assignment for which the group only needs to submit a single answer. For the rest of the questions, each student must submit their own work, but you should still feel free to discuss the assignment as a group if you would like. The groups can change throughout the semester if necessary, but for each assignment you must collaborate with a group to answer the group question. You can use Piazza to help you find a group if you need one.

The study groups will also serve an additional purpose. Each homework assignment will be graded for completeness. In addition, a randomly selected subset of problems will be graded in detail. In general, however, for any given assignment, many of the problems will not be graded by the instructor/TA. In order to improve the quality of the feedback you are receiving on your assignments, after a homework assignment is returned you must exchange your assignments among the members of the group to be graded. You can arrange to do the swap according to whatever scheme your group agrees upon, provided that: *(i)* you cannot grade the same person's assignment two weeks in a row and *(ii)* in any given week, you cannot grade for the same person who is grading you.

I will provide a detailed rubric with solutions to guide your grading. The grades assigned by your peers *will not directly count towards your grade*, but you must still report them so that I can get some feedback as to how the class is doing. I am choosing to not factor the grades assigned by your peers into your overall course grade to encourage you to provide honest, quality feedback without any of the baggage normally associated with assigning a grade. In grading your peers' work, don't be critical for no reason, but try not to feel bad about pointing out any mistakes or offering suggestions for improvement. Regarding your grade, try to stay positive and view any criticism as useful feedback that will help you focus on areas you haven't quite mastered yet.

You will also be asked to submit "peer evaluations" where you evaluate the quality of the feedback you received, e.g., was the feedback accurate? helpful? The peer evaluations could be a factor in determining your participation grade, so you should take your grading responsibilities seriously.

Web page, T-square, and Piazza

The course webpage is located at
http://users.ece.gatech.edu/~justin/ECE-6254-Spring-2016/

Course information, notes for the lectures, homework assignments, and supplemental materials will be posted here.

This course will also make use of Piazza:
https://piazza.com/gatech/spring2016/ece6254/home

Please use this as a resource to post questions about lectures and homework assignments. I will also use it for general course announcements. I also encourage students to answer questions. The enrollment in this class is fairly large, so there will be a great economy of scale here. Contributions to the discussions on Piazza will be looked on favorably when determining the Participation grade.

Finally, as this course also has an online enrollment, all of my lectures are being recorded. These lectures will be made available to the on-campus students as well and can be accessed through the GT Courses website:
https://gtcourses.gatech.edu

Distance learning

Students enrolled in the "video section" (ECE 6254Q) will have a one-week lag for turning in assignments and tests. I will also be available to answer questions via email or the telephone. For basic procedures and policies, please visit the DLPE website at
`http://www.dlpe.gatech.edu/` .

Distance learning students will generally be required to complete the same assignments as the on-campus students, but with a one-week delay. Distance students must also form study/project groups, but all group assignments should be easy to coordinate online. All lecture materials/handouts will be available through some combination of GT Courses, T-Square, and the course website. Alternate arrangements for handling the group project presentations may be considered. All technical/logistical questions about accessing online materials should be directed to cdlops@pe.gatech.edu.

Text

There is no required text. The material for this course comes from a number of different sources, including:

- Hastie, Tibshirani, and Friedman, *The Elements of Statistical Learning*, 2011.
  (pdf available online, free and legal).
- Abu-Mostafa, Magdon-Ismail, Lin, *Learning from Data*, 2012.
  (Hardback available for $28 on Amazon. Text for an online course.)
- Bishop, *Pattern Recognition and Machine Learning*, 2007.

Outline

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

Supervised Learning

- The Bayes classifier and the likelihood ratio test
- Nearest neighbor classification
- Linear classifiers
  - plugin classifiers (LDA, logistic regression, Naive Bayes)
  - the perceptron learning algorithm
  - maximum margin principle, separating hyperplanes, and support vector machines (SVMs)
- From linear to nonlinear: feature maps and the kernel trick
- Kernel-based SVMs
- Generalization Theory I
  - concentration inequalities
  - VC dimension
  - VC generalization
- Neural networks
- Linear regression
  - least-squares
  - regularization
  - the LASSO
  - empirical risk minimization
- Generalization Theory II
  - overfitting
  - bias-variance trade-off
  - error estimation and validation

Unsupervised Learning

- Density estimation
- Linear dimensionality reduction, PCA
- Clustering

  - k-means
  - GMMs and the EM algorithm
  - spectral clustering
- Euclidean embeddings
  - mutltidimensional scaling
  - manifold learning
- Latent variables and matrix factorization
  - dictionary learning
  - non-negative matrix factorization
  - blind source separation
- Feature selection

Further topics (Important things that we will probably not get time to cover.)

- Reinforcement learning
  - Markov decision processes
  - optimal planning
  - learning policies
- Ensemble methods
  - boosting
  - random forests
- Graphical models