

Selected Topics in Applied Mathematics

Applied and Comp Math 990

Section: G100

Term: 2010 Spring

Instructor: Paul Tupper

Discussion Topics: What is Learning Theory? I quote Cosma Shalizi: The basic set-up is as follows. We have a bunch of inputs and outputs, and an unknown relationship between the two. We do have a class of hypotheses describing this relationship, and suppose one of them is correct. (The hypothesis class is always circumscribed, but may be infinite.) A learning algorithm takes in a set of inputs and outputs, its data, and produces a hypothesis. Generally we assume the data are generated by some random process, and the hypothesis changes as the data change. (<http://cscs.umich.edu/~crshalizi/>) What Shalizi describes is sometimes called supervised learning, since you are told what is input and what is output. Another aspect of learning is unsupervised learning, where you are just given a bunch of data and you try to find some sort of regularity in it that is useful or interesting to you.

The Final Project: Students will write a short paper where techniques in the course (or in the text) are applied to some real data. Here are some examples of projects you could pursue:

- * computer identification of hand-written characters
- * predicting a viewers response to a movie given past ratings of movies
- * automatic classification of images, webpages, text documents, e-mail (e.g. spam or not)
- * building decision trees for medical diagnosis
- * interpolation of noisy data from wherever you like

Grading: Homework: 40%
Midterm: 30%
Final Project: 30%

Required Texts: Text: The text for this course will be The Elements of Statistical Learning: Data Mining, Inference Prediction by Hastie, Tibshirani, and Friedman, 2nd edition. The chapters I will base the class on are

- * Ch. 1. Introduction.
- * Ch. 2. Overview of Supervised Learning.
- * Ch. 3. Linear Methods for Regression.
- * Ch. 4. Linear Methods for Classification.
- * Ch. 7. Model Assessment and Selection.

We will cover other chapters in the book according to what students want to do for their projects. (Chapter 5 on Splines, Chapter 9 on Tree-Based methods, and Chapter 14 on Unsupervised Learning all look good.) The text is available online from the library website and I recommend looking at the first chapter.

Another source that we might use is the review paper by O. Bousquet, S. Boucheron, and G. Lugosi, Introduction to Statistical Learning Theory (<http://www.stat.cmu.edu/~larry/=sml2008/BBL.pdf>). This introduces the more theoretical end of the subject.

Recommended Texts:

Selected Topics in Applied Mathematics

Materials/Supplies:

Prerequisite/Corequisite: Basic probability, proficiency with Matlab or R or a similar programming environment.

You do not need to know any statistics.

Notes:

THE INSTRUCTOR RESERVES THE RIGHT TO CHANGE ANY OF THE ABOVE

INFORMATION.

Students should be aware that they have certain rights to confidentiality concerning the return of course papers and the posting of marks. Please pay careful attention to the options discussed in class at the beginning of the semester.

This outline is derived from a course outline repository database that was maintained by SFU Student Services and the University's IT Services Department. The database was retired in 2014 and the data migrated to SFU Archives in 2015.