CS 7641 Machine Learning

Instructor:
Charles Isbell, isbell@cc.gatech.edu
259, College of Computing Building

Creators of Online Material:
Prof. Charles Isbell, isbell@cc.gatech.edu
Prof. Michael Littman, littman@cs.brown.edu

TAs:
Amir Afsharinejad, a.afsharinejad@gatech.edu
Alex (Sasha) Lambert , alambert6@gatech.edu

Office Hours:
TBD

Required Text:

Machine Learning
by Tom Mitchell, McGraw Hill, 1997

General Information

Machine Learning is a three-credit course on, well, Machine Learning. Machine Learning is that area of Artificial Intelligence that is concerned with computational artifacts that modify and improve their performance through experience. The area is concerned with issues both theoretical and practical. This particular class is a part of a series of classes in Machine Learning, and takes care to present algorithms and approaches in such a way that grounds them in larger systems. We will cover a variety of topics, including: statistical supervised and unsupervised learning methods, randomized search algorithms, Bayesian learning methods, and reinforcement learning. The course also covers theoretical concepts such as inductive bias, the PAC and Mistake-bound learning frameworks, minimum description length principle, and Ockham's Razor. In order to ground these methods the course includes some programming and involvement in a number of projects.

Objectives

There are four primary objectives for the course:

• To provide a broad survey of approaches and techniques in ML

- To develop a deeper understanding of several major topics in ML

- To develop the design and programming skills that will help you to build intelligent, adaptive artifacts

- To develop the basic skills necessary to pursue research in ML

The last objective is the core one: you should develop enough background that you can pursue any desire you have to learn more about specific techniques in ML, either to pursue ML as a research career, or to apply ML techniques in other research areas in interesting (as opposed to uninteresting) ways.

Prerequisites

The official prerequisite for this course is an introductory course in artificial intelligence. In particular, those of you with experience in general representational issues in AI, some AI programming, and at least some background (or barring that, willingness to pick up some background) in statistics and information theory should be fine. Any student who did well in an AI course like this one should be fine. You will note that the syllabus for that particular course suggests at least some tentative background in some machine learning techniques as well. Having said all that, the most important prerequisite for enjoying and doing well in this class is your interest in the material. I say that every semester and I know it sounds trite, but it's true. In the end it will be your own motivation to understand the material that gets you through it more than anything else. If you are not sure whether this class is for you, please talk to me.

Resources

- **Readings.** The textbook for the course is *Machine Learning* by Tom Mitchell. We will follow the textbook quite closely for most of the semester, so it is imperative that you have a copy of the book. We will also use supplemental readings as well, but those will be provided for you.
- **Computing.** You will have access to CoC clusters for your programming assignments. You are free to use whatever machines you want to do your work; however, the final result will have to run on the standard CoC boxes. Exactly what this means will be spelled out. This shouldn't be much of a restriction for you.
- **Web.** We will use the class web page to post last-minute announcements, so check it early and often.

Statement of Academic Honesty

At this point in your academic careers, I feel that it would be impolite to harp on cheating, so I won't. You are all adults, more or less, and are expected to follow the university's code of academic conduct (you know, the honor code). Furthermore, at

least some of you are researchers-in-training, and I expect that you understand proper attribution and the importance of intellectual honesty.

Please note that unauthorized use of any previous semester course materials, such as tests, quizzes, homework, projects, videos, and any other coursework, is prohibited in this course. In particular, you are not allowed to use old exams. Using these materials will be considered a direct violation of academic policy and will be dealt with according to the GT Academic Honor Code. Furthermore, I do not allow copies of my exams out in the ether (so there should not be any out there for you to use anyway). Just as you are not to use the previous material you are not to share current material—including lecture material—with others either now or in the future. My policy on that is strict. If you violate the policy in any shape, form or fashion you will be dealt with according to the GT Academic Honor Code. I also have several... friends... from Texas who will help me personally deal with you.

Readings and Lectures

My research area is machine learning, and I'm deeply into the area. Given that and my enormous lung capacity, and my tendency to get distracted, it turns out that I can ramble on about the material for days on end, even with an editor to try to make me concise; however, that rather misses the point.

The online lectures are meant to summarize the readings and stress the important points. You are expected to critically read any assigned material. Your active participation in the material, the lectures, and various forums are crucial in making the course successful. This is less about my teaching than about your learning. My role is merely to assist you in the process of learning more about the area.

You can watch the videos at:

Udacity: https://classroom.udacity.com/courses/ud129

EdX: https://courses.edx.org/courses/course-v1:GTx+CS7641+1T2018/course/

To help you to pace yourself, I have provided a nominal schedule (see the Schedule link on the left) that tells you when we would be covering material if we were meeting twice a week during the term. I recommend you try to keep that pace.

Grading

Your final grade is divided into three components: assignments, a midterm and a final exam.

- **Assignments.** There will be four graded assignments, one for the first section, two for the second, and one for the third. They will be about programming and analysis. Generally, they are designed to give you deeper insight into the material and to

prepare you for the exams. The programming will be in service of allowing you to run and discuss experiments, do analysis, and so on. In fact, the programming is incidental, as you shall see.

- **Midterm.** There will be a written, closed-book midterm roughly halfway through the term. The midterm will be administered via Verificient's [Proctortrack](#) software.

- **Final Exam.** There will be a written, closed-book final exam at the end of the term. The final exam will also be administered via Verificient's [Proctortrack](#) software.

Due Dates

All graded assignments are due by the time and date indicated. I will not accept late assignments or makeup exams. You will get zero credit for any late assignment. The only exceptions will require: a **note** from an appropriate authority and **immediate notification** of the problem when it arises. Naturally, your excuse must be acceptable. If a meteor landed on your bed and destroyed your assignment, I need a signed note from the meteor.

Numbers

Component	
Assignments	50%
Midterm	25%
Final	25%

We intend to provide additional problem sets that you will be able to turn in. We will provide answers but will not grade them. So why would you ever bother? First, they will help prepare you for the exams, which will be graded. Second, in case you are right on the edge of two grades, we will use these problem sets to determine whether to lift you to the next grade.

Disclaimer

I reserve the right to modify any of these plans as need be during the course of the class; however, I won't do anything capriciously, anything I do change won't be too drastic, and you'll be informed as far in advance as possible.

Reading List

Required Text:

- [Tom Mitchell, Machine Learning. McGraw-Hill, 1997.](#)

Optional Text:

- [Larry Wasserman, All of Statistics. Springer, 2010.](#) (Read Part 1 for an intro to Probability Theory)
- [Richard Sutton and Andrew Barto, Reinforcement Learning: An introduction. MIT Press, 1998.](#) (for Reinforcement Learning)

A List:

- Linear Algebra
  - o [Linear Algebra and Eigenproblems](#)
- [Overview of Machine Learning](#)
  - o Mitchell Ch 1
- [Decision Trees](#)
  - o Mitchell Ch 3
- [Regression and Classification](#)
- [Neural Networks](#)
  - o Mitchell Ch 4
- [Instance Based Learning](#)
  - o Mitchell Ch 8
- [Ensemble Learning](#)
  - o [Schapire's Introduction](#)
  - o [Jiri Matas and Jan Sochman's Slides](#)
- [Kernel Methods and SVMs](#)
  - o [An ICML tutorial on SVMs](#)
  - o [Christopher Burges tutorial on SVMs for pattern recognition](#)
  - o [Scholkopf's NIPS tutorial slides on SVMs and kernel methods](#)
- [Computational Learning Theory](#)
  - o Mitchell Ch 7
- [VC Dimensions](#)
  - o Mitchell Ch 7
- [Bayesian Learning](#)
  - o Mitchell Ch 6
- [Bayesian Inference](#)
- [Randomized Optimization](#)
  - o Mitchell Ch 9
  - o [No Free Lunch Theorem](#)
- [Clustering](#)
  - o Mitchell Ch 6
  - o [Intuitive Explanation of EM](#)
  - o [Statical View of EM](#)
  - o [Jon Kleinberg's Impossibility Theorem for Clustering](#)
- [Feature Selection](#)
  - o [ICA: Algorithms and Applications](#)
- [Feature Transformation](#)
- [Information Theory](#)

- - [Charles Isbell's Note on Information Theory](#)
  - [An Introduction to Information Theory and Entropy](#)
- [Markov Decision Processes](#)
- [Reinforcement Learning](#)
  - Mitchell Ch 13
  - [Richard Sutton and Andrew Barto, Reinforcement Learning: An introduction. MIT Press, 1998.](#)
  - [Reinforcement Learning: A Survey](#)
- [Game Theory](#) and [Continued](#)
  - [Andrew Moore's slides on Zero-Sum games](#)
  - [Andrew Moore's slides on Non-zero-Sum games](#)

Software

- [WEKA](#) Machine learning software in JAVA that you can use for your projects
- [Data Mining with Weka](#) A MOOC Course
- [ABAGAIL](#) Machine learning software in JAVA. This is hosted on my github, so you can contribute too
- [scikit-learn](#) A popular python library for supervised and unsupervised learning algorithms
- [pybrain](#) A popular python library for artifical neural networks
- [MATLAB NN Toolbox](#) The toolbox supports supervised learning with feedforward, radial basis, and dynamic networks and unsupervised learning with self-organizing maps and competitive layers.
- [Murphy's MDP Toolbox for Matlab](#)
- [MATLAB Clustering Package](#) By [Frank Dellaert](#)

Datasets

- [UCI Machine Learning Repository](#) An online repository of data sets that can be used for machine learning experiments.
- [Stanford Large Network Dataset](#) Dataset of large social and information networks.
- [Vision Benchmark Suite](#) Autonomous car dataset
- [Other datasets](#)