

Syllabus CS 4640 Machine Learning

Frank Dellaert, January 6 2002

1. Scope

Machine Learning is concerned with computer programs that automatically improve their performance through experience. This course covers the theory and practice of machine learning from a variety of perspectives. We cover topics such as learning decision trees, neural network learning, statistical learning methods, genetic algorithms, Bayesian learning methods, explanation-based learning, and reinforcement learning. The course covers theoretical concepts such as inductive bias, the PAC and Mistake-bound learning frameworks, minimum description length principle, and Occam's Razor. Programming assignments include hands-on experiments with various learning algorithms. Typical assignments include neural network learning for face recognition, and decision tree learning from databases of credit records.

This course is heavily modeled on the Machine Learning course originally taught by Tom Mitchell, and later by Sebastian Thrun, at Carnegie Mellon University.

2. Logistics

Web Site

The class web site is at http://www.cc.gatech.edu/classes/AY2002/cs4640_spring

It contains the information in this syllabus, as well as links to course notes (PowerPoint files), assignment handouts etc. Please check the announcements page regularly for updates and news.

Time and Location

Tuesday, Thursday 9:30-11am, IC 219

Instructor

Frank Dellaert

Email: frank@cc.gatech.edu

Office: CoC Bldg (CCB) 254

Office hours: 4-5pm Tuesday and Thursday

Phone: 404-385-2923 (email preferred)

Textbook

"Machine Learning" by Tom Mitchell.

It should be available for sale in the Georgia Tech Book Store.

Important Dates

- First day of class Jan 8
- Last day to drop a course Feb 15
- **Midterm Feb 26**
- Last day to withdraw from school with "W" grades Mar 13
- Spring Break March 4-8
- Last day of class Apr 25
- **Final exam (tentative) Friday May 3 11.30-2.20**

3. Grading & Policy

Course Requirements and Grading

The grade will break down as follows:

- Exercises (10%)
- Assignments (30%)
- Midterm (15%)
- Course Project (30%)
- Final (15%)

Students are expected to abide by the Georgia Tech Honor Code. Honest and ethical behavior is expected at all times. All incidents of suspected dishonesty will be reported to and handled by the office of student affairs. You are to do all assignments yourself, unless explicitly told otherwise. You may discuss the assignments with your classmates, but you may not copy any solution (or part of a solution) from a classmate.

Exercises

There will be 10 small (almost trivial) exercises that will be handed out on Thursday and are due on the next Tuesday. Their purpose is twofold: (a) they will allow you to get acquainted with Haskell/Hugs, (b) make sure that you think about the material at least once by the next class period. Grading of exercises is pass/fail and each exercise counts for 1% of your grade.

Assignments

There will be three assignments that require a more substantial amount of work. These will be due approximately two weeks after they are handed out. For the assignments themselves and submission instructions see the Assignments page

http://www.cc.gatech.edu/classes/AY2002/cs4640_spring/html/assignments.html

Course Project

To pass CS 4640, you are required to do a course project in teams of 2-3 students. See the last section below for important deadlines and requirements. Note that the web site always contains the latest information in response to your questions/suggestions.

Lateness and Absence

Homework is worth full credit at the beginning of class on the due date.

It is worth half credit for the next 48 hours.

It is worth zero credit after that.

There is no penalty for absence (other than: you miss an opportunity to learn something).

4. Schedule

8-Jan Introduction
10-Jan Concept Learning – Version Spaces
15-Jan Decision Trees
17-Jan Neural Networks
11-Jan Evaluating Hypotheses
24-Jan Bayesian Learning - Naïve Bayes
29-Jan Text Learning
31-Jan Instance-based Learning
5-Feb Instance-based Learning
7-Feb PAC Learning
12-Feb Guest Lecture
14-Feb Support-Vector Machines
19-Feb Bayesian Belief Networks - Inference
21-Feb Bayesian Belief Networks - Learning
26-Feb Midterm
28-Feb Hidden Markov Models
5-Mar SPRING BREAK
7-Mar SPRING BREAK
12-Mar Simulated Annealing - Genetic Algorithms
14-Mar Learning Rules
19-Mar Explanation-based Learning
21-Mar Markov Decision Processes & Reinforcement Learning
26-Mar Markov Decision Processes & Reinforcement Learning
28-Mar Partially Observable Markov Decision Processes
2-Apr Bagging and Boosting
4-Apr Advanced Topics
9-Apr Advanced Topics
11-Apr Advanced Topics
16-Apr Advanced Topics
18-Apr Project Presentations
23-Apr Project Presentations
25-Apr Review
3-May Final Exam

5. Course Project

The goal of the course project is to apply machine learning to interesting problem domains, thereby gaining in-depth experience with specific algorithms and problem domains. To root your project in reality you are required to find a faculty mentor that will sponsor your project, similar to the 7001 course.

- Projects must be done in teams of two or three students.
- You must form a project team by the end of class on Jan 17
- A single project proposal must be submitted at the beginning of class on Feb 7. Before writing the proposal you should find and meet with the faculty mentor. The project proposal should be 3-5 pages long (see below). Before submitting the proposal you should also meet with me to have the project approved. You are responsible for scheduling that meeting.
- The final project report is due at the beginning of class on April 11. It should be eight to twelve pages long. Your report should at a minimum include the following sections.
 1. Problem definition. Explain why the problem is interesting and open.
 2. Prior work (discuss at least 5 related papers)
 3. Your approach(es). This is the major part of the report
 4. Results, experimental and/or theoretical.
 5. Conclusions and ideas for future work.
 6. References

Guidelines for the Project Proposal

Each team has to submit one project proposal by Feb 7. By this date, each team must have met with the faculty mentor and with me. The team should have familiarized itself with the data and the problem and should also have a clear grasp of what the problem is.

The Project proposal should be 3-5 pages long and clearly address the following

1. **Problem description.** A concise problem description, with a description of the data sets (if applicable). Make sure you communicate why this project is interesting.
2. **Research goals.** A section on what the different outcomes might be. Tell us: when do you judge the project to be successful?
3. **Project Plan.** A project plan, outlining in detail what type experiments will be performed and when (take your best guess). Also describe other means of achieving the project goal, such as consulting specific literature, talking to specific experts.
4. **Individual Tasks.** A description of the tasks that will be performed by the individual team members. Our requirement: Each team member has to engage in interesting machine learning work, so having one student program the software interface while another runs machine learning experiments is not a good division of labor.

Be as concise as possible. Don't stay in the abstract. Be concrete.

Also, several of the items above clearly require input by the faculty mentor. You should therefore be in close contact to your faculty mentor.

Writing a good proposal is a lot of work, since you have to think through the problem, speculate as to what the results might be, and come up with a work plan in advance. However, it's a great exercise for your job once you graduate from Georgia Tech.