EECS 440: Machine Learning (Fall 2014)

Instructor: Soumya Ray
Office: Olin 516
Office Hours: M 9:30-11

Meeting Time and Location: TR 10-11:15, Sears 354

Textbooks: Machine Learning by Tom M. Mitchell
http://www.cs.cmu.edu/~tom/mlbook.html
Pattern Recognition and Machine Learning by Christopher M. Bishop (books reserved at KSL)

------------------------------------------------------------------------

Overview

This course is focused on algorithms for machine learning: their design, analysis and implementation. We will study different learning settings, including supervised, semi-supervised and unsupervised learning. We will study different ways of representing the learning problem, using propositional, multiple-instance and relational representations. We will study the different algorithms that have been developed for these settings, such as decision trees, neural networks, support vector machines and Bayesian methods. We will learn about the theoretical tradeoffs in the design of these algorithms, and how to evaluate their behavior in practice.

At the end of the course, you should be able to:

* Recognize situations where machine learning algorithms are applicable
* Understand, represent and formulate the learning problem
* Apply the appropriate algorithm(s), or if necessary, design your own, with an understanding of the tradeoffs involved
* Correctly evaluate the behavior of the algorithm when solving the problem.

------------------------------------------------------------------------

Grading

Evaluation: 5 written homeworks, 4 programming assignments, final discussion based exam, course project
Homework: 25%
Programming Assignments: 30%
Final Discussion-based exam: 15%
Class Project: 25% (5% presentation, 20% writeup)
Class Participation: 5%

* Written assignments are due by midnight on the due date, and can be turned in either in class or on Blackboard. Programming assignments are due by midnight on the due date in Blackboard.

* Assignments submitted late are penalized @ 10% for each extra day after the due date. If you have a very good reason, such as illness, please see me.

* Collaboration policy:

You can submit all assignments in pairs. You do not have to pair up with the same person for all assignments. Each assignment must contain the names of both persons and the statement "We certify that we have contributed equally towards this assignment."

You are welcome to discuss assignments with each other and with me, but do not copy solutions from any source, including the web. Any assignment with your name on it must be substantially your work.

*Case Academic Integrity Policy

<http://studentaffairs.case.edu/groups/aiboard/policy.html>

Topics

1. Introduction and Background (Statistics, Optimization)
   *Part 1*  *Algorithms*
2. Foundations/ Propositional Supervised Learning (Chapter 1 and 2)
3. Decision Tree Induction (Chapter 3)
4. Overfitting, Empirical Methodology (Chapter 3)
5. Metrics for Classification
6. Artificial Neural Networks (Chapter 4)
7. Comparing Learning Algorithms and Hypothesis Testing (Chapter 5)
8. Support Vector Machines (Chapter 7, Bishop)
9. Probabilistic Classification (Chapter 6)
7. Naive Bayes and Logistic Regression (Chapter 6)

*Part 2*  
7. Naive Bayes and Logistic Regression (Chapter 6)

*Issues in Machine Learning*

8. Handling Missing Data, Expectation Maximization (Ch 9, Bishop)
9. Ensemble Methods
10. Learning with Prior Knowledge (Chapter 12)

11. Bias Variance Analysis
12. Computational Learning Theory (Chapter 7)
13. Feature Selection and Dimensionality Reduction (Bishop)

*Part 3*  
*Other learning settings*

14. Sequential Supervised Learning
15. Unsupervised and semi-supervised Learning