

# CS4780/5780 - Machine Learning

Fall 2012

Thorsten Joachims  
Cornell University  
Department of Computer Science

# Outline of Today

- Who we are?
  - Prof: Thorsten Joachims
  - TAs: Joshua Moore, Igor Labutov, Moontae Lee
  - Consultants: Declan Boyd, Harry Terkelsen, Jason Zhao, Joe Mongeluzzi, Kyle Hsu, Emma Kilfoyle, ...
- What is learning?
  - Why should a computer be able to learn?
  - Examples of machine learning.
  - What it takes to build a learning system?
- Syllabus
- Administrivia

# (One) Definition of Learning

- Definition [Mitchell]:
  - A computer program is said to learn from
    - experience  $E$  with respect to some class of
    - tasks  $T$  and
    - performance measure  $P$ ,
  - if its performance at tasks in  $T$ , as measured by  $P$ ,
  - improves with experience  $E$ .

# Syllabus

- Concept Learning : Hypothesis space, version space
- Instance-Based Learning : k-nearest neighbor, collaborative filtering
- Decision Trees : TDIDT, attribute selection, pruning and overfitting
- ML Experimentation: hypothesis tests, resampling estimates
- Linear Rules : Perceptron, duality, mistake bound
- Support Vector Machines : optimal hyperplane, kernels, stability
- Generative Models : Naïve Bayes, linear discriminant analysis
- Hidden Markov Models : probabilistic model, estimation, Viterbi
- Structured Output Prediction : predicting sequences, rankings, etc.
- Learning Theory : PAC learning, mistake bounds
- Clustering : HAC Clustering, k-means, mixture of Gaussians
- Recommendation: similarity-based methods, matrix factorization

# Textbook and Course Material

- Main Textbooks
  - Tom Mitchell, "Machine Learning", McGraw Hill, 1997.
  - CS4780 Course Pack from Campus Store
- Additional References (optional)
  - Ethem Alpaydin, "Introduction to Machine Learning", MIT Press, 2004.
  - See other references on course web page.
- Course Notes
  - Slides available on course homepage
  - Material on blackboard

# Pre-Requisites and Related Courses

- Pre-Requisites
  - Programming skills (e.g. CS 2110)
  - Basic linear algebra (e.g. MATH2940)
  - Basic probability theory (e.g. CS 2800)
  - Short exam to test prereqs
- Related Courses
  - CS4700: Foundations of Artificial Intelligence
  - CS4758: Robot Learning
  - CS4300: Information Retrieval
  - CS6780: Advanced Machine Learning
  - CS6784: Advanced Topics in Machine Learning
  - CS6740: Advanced Language Technologies

# Homework Assignments

- Assignments
  - 5 homework assignments
  - Some problem sets, some programming and experiments
- Policies
  - Assignments are due at the beginning of class on the due date in hardcopy. Code must be submitted via CMS by the same deadline.
  - Assignments turned in late will be charged a 1 percentage point reduction of the cumulated final homework grade for each period of 24 hours for which the assignment is late.
  - Everybody had 5 “free” late days. Use them wisely.
  - No assignments will be accepted after the solutions have been made available (typically 3-4 days after deadline).
  - Typically collaboration of two students (see each assignment for detailed collaboration policy).
  - We run automatic cheating detection. Must state all sources of material used in assignments or project. Please review Cornell Academic Integrity Policy!

# Exams and Quizzes

- In-class Quizzes
  - A few per semester
  - No longer than 5 minutes
- Exams
  - Two Prelim exams
    - October 16 (week after fall break)
    - November 20 (week of thanksgiving break)
  - In class
  - No final exam



# Final Project

- Organization
  - Self-defined topic related to your interests and research
  - Groups of 3-4 students
  - Each group has TA as advisor
- Deliverables
  - Project proposal (~ 2 weeks after fall break)
  - Meetings with TA to discuss progress
  - Short presentation (last week of classes)
  - Project report (~ exam period)

# Grading

- Deliverables
  - 2 Prelim Exams (40% of Grade)
  - Final Project (15% of Grade)
  - Homeworks (~5 assignments) (35% of Grade)
  - Quizzes (in class) (5% of Grade)
  - PreReq Exam (2% of Grade)
  - Participation (3% of Grade)
- Outlier elimination
  - For homeworks and quizzes, the lowest grade is replaced by the second lowest grade.

# How to Get in Touch

- Online
  - <http://www.cs.cornell.edu/Courses/cs4780/2012fa/>
  - Piazza forum
  - Videonote (Fall 2011)
- Email Addresses
  - Thorsten Joachims: [tj@cs.cornell.edu](mailto:tj@cs.cornell.edu)
  - Igor Labutov: [iil4@cornell.edu](mailto:iil4@cornell.edu)
  - Moontae Lee: [ml2255@cornell.edu](mailto:ml2255@cornell.edu)
  - Joshua Moore: [jlm434@cornell.edu](mailto:jlm434@cornell.edu)
  - Declan Boyd, Harry Terkelsen, Jason Zhao, Joseph Mongeluzzi, Kyle Hsu, Emma Kilfoyle
- Office Hours
  - Thorsten Joachims:
    - Thursdays 2:40pm – 4:00pm, 4153 Upson Hall
  - Other office hours:
    - TBD