Outline of Today

• Who we are?
  – Prof: Thorsten Joachims
  – TAs: Igor Labutov, Ian Lenz, Karthik Raman, Tobias Schnabel, Emma Kilfoyle
  – Consultants: Darren Voon, Ben Shulman, Wenhai Yang, Anthony Fu, Brook Du, Detian Shi, Steve Mandl
• What is learning?
  – Why should a computer be able to learn?
  – Examples of machine learning (ML).
  – What drives research in and use of ML today?
• 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

• Instance-Based Learning: k-nearest neighbor, collaborative filtering
• Decision Trees: TDIDT, attribute selection, pruning and overfitting
• Linear Rules: Perceptron, logistic regression, linear regression, duality
• Support Vector Machines: optimal hyperplane, margin, kernels, stability
• Generative Models: naïve Bayes, linear discriminant analysis
• Hidden Markov Models: probabilistic model, estimation, Viterbi
• Structured Output Prediction: predicting sequences, rankings, etc.
• Statistical Learning Theory: PAC learning, VC dimension, error bounds
• Online Learning: experts, bandits, online mistake bounds
• Clustering: HAC Clustering, k-means, mixture of Gaussians
• Recommendation: similarity-based methods, matrix factorization, etc.
• ML Experimentation: hypothesis tests, cross validation, resampling

Textbook and Course Material

• Main Textbooks
  – CS4780 Course Pack from Campus Store
• Additional References (optional)
  – See other references on course web page
• Course Notes
  – Writing on blackboard
  – Slides available on course homepage
  – Video of lecture available on course homepage

Pre-Requisites and Related Courses

• Pre-Requisites
  – Programming skills (e.g. CS 2110)
  – Basic linear algebra (e.g. MATH 2940)
  – Basic probability theory (e.g. CS 2800)
  → Short exam to test prereqs (via CMS)
• Related Courses
  – CS4700: Foundations of Artificial Intelligence
  – CS4758: Robot Learning
  – CS4300: Information Retrieval
  – CS4740: Natural Language Processing
  – CS6780: Advanced Machine Learning
  – CS6784: Advanced Topics in Machine Learning
  – CS6740: Advanced Language Technologies
  – CS6782: Probabilistic Graphical Models
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 has 5 “free” late days. Use them wisely.
  - No assignments will be accepted after the solutions have been made available (typically 3-5 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 17 (week of fall break)
    - November 26 (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 (week after fall break)
  - Meetings with TA to discuss progress
  - Poster presentation (last week of classes)
  - Project report (December 11)
  - Peer review (December 18)

Grading

- Deliverables
  - 2 Prelim Exams (50% of Grade)
  - Final Project (15% of Grade)
  - Homeworks (~5 assignments) (25% 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
  - Course Homepage (slides, video, references, policies, office hours)
  - Piazza forum (questions and comments)
  - CMS (homeworks and grades)
- Email Addresses
  - Thorsten Joachims: tj@cs.cornell.edu
  - Igor Labutov: iil4@cornell.edu [homework and solutions]
  - Karthik Raman: kr339@cornell.edu [projects]
  - Tobias Schnabel: tbs49@cornell.edu [office hours, piazza, video]
  - Ian Lenz: ianl@cornell.edu [late submissions, CMS, regrades]
- Office Hours
  - Thorsten Joachims:
    - Thursdays 2:40pm – 4:00pm, 4153 Upson Hall
  - Other office hours:
    - See course homepage