## CS6780
## Advanced Machine Learning

Spring 2015

Thorsten Joachims
Cornell University
Department of Computer Science

## Outline of Today

- Who we are?
  - Prof: Thorsten Joachims
  - TAs: Daniel Sedra, Adith Swaminathan
- What is learning?
  - 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

- Supervised Batch Learning: model, decision theoretic foundation, model selection, model assessment, empirical risk minimization
- Decision Trees : TDIDT, attribute selection, pruning and overfitting
- Statistical Learning Theory : generalization error bounds, VC dimension
- Large-Margin Methods: linear Rules, margin, Perceptron, SVMs
- Kernels: duality, non-linear rules, non-vectorial data
- Probabilistic Models: generative vs. discriminative, maximum likelihood, Bayesian inference
- Sequence Prediction : hidden Markov model, Viterbi
- Structured Output Prediction : undirected graphical models, structural SVMs, conditional random fields
- Latent Variable Models: k-means clustering, mixture of Gaussians, expectation-maximization algorithm, matrix factorization, embeddings
- Online Learning : experts, bandits, online convex optimization
- Other topics: neural nets, ensemble methods, sparsity, …

## Secondary Syllabus

- Practice "soft skills" needed to be a successful researcher
  - Pitch ideas
  - Present your work
  - Write convincing papers
  - Work in groups
  - Give constructive feedback to others
  - Use feedback constructively

## Textbook and Course Material

- Main Textbooks
  - Kevin Murphy, "Machine Learning – a Probabilistic Perspective", MIT Press, 2012.
  - See other references on course web page
- Course Notes
  - Writing on whiteboard
  - Slides available on course homepage

## Pre-Requisites

- Pre-Requisites
  - Programming skills (e.g. CS 2110)
  - Basic linear algebra (e.g. MATH 2940)
  - Basic probability theory (e.g. MATH 4710)
- Not required
  - Previous ugrad machine learning course

## Homework Assignments

- Assignments
  - 4 homework assignments
  - Some problem sets, some programming and experiments
- Policies
  - Assignments are due at the beginning of class on the due date in hardcopy.
  - 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).
  - Please review Cornell Academic Integrity Policy!

## Exam

- Exam
  - April 16
  - In class
  - No final exam

## Project

- Organization
  - Self-defined topic related to your interests and research
  - Groups of 2-3 students
- Deliverables
  - Pitch (Feb 3)
  - Proposal (Feb 12)
  - Presentation (last two weeks of classes)
  - Report (May 11)
  - Peer review (May 14)
  - Author rebuttal (May 15)

## Grading

- Deliverables
  - Exam                          (35% of Grade)
  - Project                       (35% of Grade)
  - Homeworks                     (20% of Grade)
  - Participation                 (10% of Grade)
- Outlier elimination
  - For homeworks, the lowest grade is replaced by the second lowest grade.
- Grade Options
  - Letter grade
  - S/U: a grade of at least D. Excludes project.
  - Audit: attend lectures. Excludes project, homeworks, exam.

## Enrolling

- You can enroll in the class only
  - If you are a PhD student, and
  - If you have not previously taken CS4780 or CS5780.
- Enrollment Process
  - get manual enrollment form from me today after class.
- Enrollment Priorities
  - CS PhD Students
  - Other PhD Students in the order that I have received email from asking for enrollment permission
  - Other PhD Students that have not contacted me before

# How to Get in Touch

- Online
  - Course Homepage (slides, references, policies, office hours)
    - http://www.cs.cornell.edu/Courses/cs6780/2015sp/
  - Piazza forum (questions and comments)
  - CMS (homeworks and grades)
- Email Addresses
  - Thorsten Joachims: tj@cs.cornell.edu
  - Daniel Sedra: dms422@cornell.edu
  - Adith Swaminathan: fa234@cornell.edu
- Office Hours
  - Thorsten Joachims:
    - Fridays 1:30pm – 2:30pm, 418 Gates Hall
    - Exception: on Friday Jan 23 from 12:30-1:30
  - Other office hours:
    - See course homepage