CS6780
Advanced Machine Learning

Spring 2019

Thorsten Joachims
Cornell University
Department of Computer Science

Outline of Today

- Who we are?
  - Prof: Thorsten Joachims
  - TAs: Aman Agarwal, Ashudeep Singh
- What is learning?
  - Examples of machine learning (ML).
  - What drives research in and use of ML today?
- Syllabus
  - Topics and Methods
  - Themes
- 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.

What is the goal of CS6780?

- PhD-level introduction to machine learning
  - First or second ML class
- Broad, but deep along several key themes
- Enable your research in or with machine learning
- Practice "soft" skills you need as researcher

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
- Deep Networks: multi-layer perceptrons, convolutions, pooling
- Structured Output Prediction: hidden Markov model, Viterbi, structural SVMs, conditional random fields
- Probabilistic Models: generative vs. discriminative, maximum likelihood, Bayesian inference
- Latent Variable Models: k-means clustering, mixture of Gaussians, expectation-maximization algorithm, matrix factorization, embeddings
- Online Learning: experts, bandits, online convex optimization
- Causal Inference: interventional vs. observational data, treatment effects, policy learning

Theme: Prediction and Action

- Building intelligent systems vs. analyzing existing systems
  - Prediction
  - Intelligent action
  - Guarantees on prediction/action quality
  - Causality

Theme: Bias vs. Variance

- Fundamental trade-off in learning
  - Training error vs. prediction error
  - Model capacity
  - Statistical learning theory
  - Empirical risk minimization

Theme: Massive Overparameterization

- The success story of machine learning
  - Sparse linear models
  - Kernels
  - Deep networks
  - → Number of parameters ≫ number of examples

Theme: Theoretical Underpinning

- Theory for understanding sake
  - Identify the mechanisms at play in ML
  - Understand model complexity
  - Understand common themes between algorithms

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 blackboard
  - 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)
  - Basic multivariable calculus (e.g. MATH 1920)
- 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.
  - Everybody has 5 “free” late days. Use them wisely.
  - Beyond that, 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.
  - 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 25
  - In class
  - No final exam

Project

- Organization
  - Self-defined topic related to your interests and research
  - Groups of 2-3 students
- Deliverables
  - Proposal (March 12)
  - Poster Presentation (May 2, evening)
  - Report (May 13)
  - Peer review (May 15)
  - Author rebuttal (May 17)

Grading

- Deliverables
  - Exam (40% of Grade)
  - Project (35% of Grade)
  - Homeworks (20% of Grade)
  - Participation (5% 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.
- Enrollment Process
  - open enrollment via studentcenter.
- Enrollment “Deadline”
  - first homework will come out Feb 5.

Audio/Video

- Live stream to Cornell Tech
- Recordings available after class

How to Get in Touch

- Online
  - Course Homepage (slides, references, policies, office hours)
    - http://www.cs.cornell.edu/Courses/cs6780/2019sp/
  - Piazza forum (questions and comments)
  - CMS (homeworks and grades)
  - CMT (projects)
- Email Addresses
  - Thorsten Joachims: tj@cs.cornell.edu
  - Aman Agarwal: aa2398@cornell.edu
  - Ashudeep Singh: as3354@cornell.edu
- Office Hours
  - Thorsten Joachims:
    - Fridays 11:00am – 12:00pm, 418 Gates Hall
  - Other office hours:
    - See course homepage
  - Zoom for CT students