CS6780
Summary and Conclusions

CS6780 – Advanced Machine Learning
Spring 2019

Thorsten Joachims
Cornell University

Theme: Prediction and Action

- <u>Building intelligent systems</u> 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
  - Regularized 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

Batch Learning Approaches

- Empirical Risk Minimization (ERM)
  - Fixed at training time: class of decision rules h: X → Y, loss, x and y
  - Strategy: minimize training loss
- Conditional Probability Models
  - Fixed at training time: class of models for P(Y|X), x and y
  - Strategy: max conditional likelihood or MAP (or Bayes)
- Generative Models
  - Fixed at training time: class models for P(Y,X)
  - Strategy: max likelihood or MAP (or Bayes)

Not covered: Bayesian ML perspective
→ ORIE 6741, Andrew Wilson

Batch Learning for Classification

- ERM
  - Decision Trees
  - Perceptron
  - Linear SVMs
  - Kernel SVMs
  - Neural Networks
- Conditional Probability
  - Logistic Regression
  - Conditional Random Fields
  - Ridge Regression
- Generative
  - Multinomial Naïve Bayes
  - Multivariate Naïve Bayes
  - Linear Discriminant

- Other Methods
  - Logical rule learning
  - Gaussian Processes
  - RBF Networks
  - Boosting
  - Bagging
  - Parametric (Graphical) Models

  - *-Regression
  - *-Multiclass

→Methods + Theory + Algorithms + Practice

Batch Learning for Struct Prediction

- ERM
  - Structural SVMs

- Conditional Probability
  - Conditional Random Fields

- Generative
  - Hidden Markov Model

- Other Methods
  - Maximum Margin Markov Networks
  - Markov Random Fields
  - Bayesian Networks
  - Statistical Relational Learning
  - Markov Logic Networks
  - Encoder/Decoder Networks

NLP classes
→ Claire Cardie

Online Learning

- Expert Setting
  - Halving Algorithm
  - Weighted Majority
  - Exponentiated Gradient
- Bandit Setting
  - EXP3
  - UCB1

- Other Methods
  - Hedge
  - Follow the Leader
  - Zooming
  - Partial Monitoring
  - Contextual Bandits
  - Dueling Bandits
  - Coactive Learning

CS6783 Machine Learning Theory

→ Karthik Sridharan

Unsupervised Learning

- Clustering
  - K-Means
  - Mixture of Gaussians and EM-Algorithm

- Other Methods
  - Spectral Clustering
  - Multi-Dimensional Scaling
  - Latent Dirichlet Allocation
  - Semantic Embeddings
  - Deep Auto-Encoders
- Other Tasks
  - Outlier Detection
  - Novelty Detection
  - Dimensionality Reduction
  - Non-Linear Manifold Detection
  → CS4786 Machine Learning for Data Science
  → CS4850 Math Found for the Information Age
  → INFO 6150 Advanced Topic Models

ML in Computer Visions

- Covered

  Convolutional Neural
  Networks (LeNet)

- Other

  More Deep Learning

  Even more Deep Learning

  CS6670 Computer Vision
  → Bharath Hariharan

  CS4670 Intro Computer Vision
  → Serge Belongie

  → Bharath Hariharan

Learning to Act

- Covered
  - Off-policy policy learning
  - Atomic actions
  - Combinatorial actions

- Other
  - Sequential decisions
  - Markov Decision Processes
  - Reinforcement learning
  - On policy vs. off policy
  - Policy gradient

ML and Causality

- Covered
  - Potential outcomes model
  - Treatment effect estimation

- Other
  - Observational setting
  - Instrumental variables
  - Continuous treatments
  - Longitudinal treatments
  - Causal discovery
  - Parameter inference
  - Causal networks
  - Structural equation models

ML and Fairness

- Covered

- Other
  - Fairness
  - Accountability
  - Transparency
  - Criteria and policy
  - Algorithms and guarantees
  - Classification
  - Ranking

INFO4270: Ethics and Policy in DS
→ Solon Barocas

Other Machine Learning Courses at Cornell

- INFO 3300 - Data-Driven Web Pages
- CS 4700 - Introduction to Artificial Intelligence
- CS 4780/5780 - Machine Learning for Intelligent Systems
- CS 4786/5786 - Machine Learning for Data Science
- CS 4787 - Principles of Large-Scale Machine Learning
- OR 4740 - Statistical Data Mining
- CS 6780 - Advanced Machine Learning
- CS 6783 - Machine Learning Theory
- CS 6784 - Advanced Topics in Machine Learning
- CS 6787 - Advanced Machine Learning Systems
- ORIE 6740 - Statistical Learning Theory for Data Mining
- ORIE 6741 - Bayesian Machine Learning
- ORIE 6750 - Optimal learning
- ORIE 6780 - Bayesian Statistics and Data Analysis
- INFO 6150 - Advanced Topic Models
- MATH 7740 - Statistical Learning Theory
- CS 7790 - AI Seminar
- CS 7792 - Special Topics in Machine Learning