CS6780
Summary and Conclusions

CS6780 – Advanced Machine Learning
Spring 2015

Thorsten Joachims
Cornell University

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)

Batch Learning for Classification

- ERM
  - Decision Trees
  - Perceptron
  - Linear SVMs
  - Kernel SVMs
- 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
  - Neural Networks
  - RBF Networks
  - Boosting
  - Bagging
  - Parametric (Graphical) Models
  - *-Regression
  - *-Multiclass

→ Methods + Theory + Algorithms + Practice

→ CS6784 Advanced Topics in ML
→ Kilian Weinberger

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

→ CS4782 Probabilistic Graphical Models

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
  - Latent Dirichlet Allocation
  - Latent Semantic Analysis
  - Multi-Dimensional Scaling
- 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

→ David Mimno

Other Learning Problems and Applications

- Recommender Systems, Search Ranking, etc.
  - CS4300 Information Retrieval
    → Cristian Niculescu-Danescu-Mizil
- Reinforcement Learning and Markov Decision Processes
  - CS4758 Robot Learning
- Computer Vision
  - CS4670 Intro Computer Vision
    → Serge Belongie
- Natural Language Processing
  - CS4740 Intro Natural Language Processing
    → Lillian Lee
    → Claire Cardie

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 4758 - Robot Learning
- CS 4782 - Probabilistic Graphical Models
- OR 4740 - Statistical Data Mining
- CS 6780 - Advanced Machine Learning
- CS 6783 - Machine Learning Theory
- CS 6784 - Advanced Topics in Machine Learning
- CS 6756 - Advanced Topics in Robot Learning
- ORIE 6740 - Statistical Learning Theory for Data Mining
- ORIE 6750 - Optimal learning
- ORIE 6780 - Bayesian Statistics and Data Analysis
- MATH 7740 - Statistical Learning Theory