# CS6375: Recap

Nicholas Ruozzi
University of Texas at Dallas

# Supervised Learning

- Regression & classification
- Discriminative methods
  - k-NN
  - Decision trees
  - Perceptron
  - SVMs & kernel methods
  - Logistic regression
- Parameter learning
  - Maximum likelihood estimation
  - Expectation maximization

# Bayesian Approaches

- MAP estimation
- Prior/posterior probabilities
- Bayesian networks
  - Naive Bayes
  - Hidden Markov models
  - Structure learning via Chow-Liu Trees
- Latent Dirichlet Allocation (LDA)

# Unsupervised Learning

- Clustering
  - $k$-means
  - Spectral clustering
  - Hierarchical clustering
- Expectation maximization
  - Soft clustering
  - Mixtures of Gaussians

# Learning Theory

- PAC learning
- VC dimension
- Bias/variance tradeoff
- Chernoff bounds
- Sample complexity

# Optimization Methods

- Gradient descent
  - Stochastic gradient descent
  - Subgradient methods
- Coordinate descent
- Lagrange multipliers and duality

# Matrix Based Methods

- Dimensionality Reduction
  - PCA
  - Matrix Factorizations
- Collaborative Filtering
  - Semisupervised learning

# Ensemble Methods

- Bootstrap sampling
- Bagging
- Boosting

# Other Learning Topics

- Active learning
- Reinforcement learning
- Learning to rank
- Neural networks
  - Perceptron and sigmoid neurons
  - Backpropagation

# Questions about the course content?

(Reminder: I do not have office hours this week)

# For the final…

- You should understand the basic concepts and theory of all of the algorithms and techniques that we have discussed in the course
- There is no need to memorize complicated formulas, etc.
  - For example, if I ask for the sample complexity of a scheme, I will give you the generic formula
- However, you should be able to derive the algorithms and updates
  - E.g., Lagrange multipliers and SVMs, the EM algorithm, etc.

# For the final…

- No calculators, books, notes, etc. will be permitted
  - As before, if you need a calculator, you have done something terribly wrong
- The exam will be in roughly the same format
  - Expect true/false questions, short answers, and two-three long answer questions
- Exam will emphasize the new material, but ALL material will be tested
- Take a look at the practice exams!

# Final Exam

Wednesday, 12/16/2015

11:00AM - 1:45PM

ECSS 2.410

# Related Courses at UTD

- Natural Language Processing (CS 6320)
- Statistical Methods in Artificial Intelligence and Machine Learning (CS 6347)
- Artificial Intelligence (CS 6364)
- Information Retrieval (CS 6322)
- Intelligent Systems Analysis (ACN 6347)
- Intelligent Systems Design (ACN 6349)

# ML Related People

- Vincent Ng (NLP)
- Yang Liu (NLP)
- Vibhav Gogate (MLNs, Sampling, Graphical Models)
- Sanda Harabagiu (NLP & Health)
- Dan Moldovan (NLP)
- Nicholas Ruozzi (Graphical Models & Approx. Inference)

# Matrix Decomposition

- PCA is a dimensionality reduction technique that is based on matrix factorizations
  - Drawback: PCA returns the eigenvectors of a matrix as the most relevant vectors (many applications need subsets of the data that best describe it)
- Feature selection / matrix factorization using Bayesian networks
- Input: data points as rows of a $m \times n$ matrix $X$
- Output: $X \sim CU$ where $C$ is a $m \times k$ matrix of columns selected from $X$ and $U$ is an arbitrary matrix

# Airplane Health

- Collaboration with Southwest airlines
  - Pilots/maintenance crews perform physical inspections of planes and are asked to translate observations into maintenance codes
  - The observations (symptoms) and the codes (diagnoses) typically are mismatched (inspections performed quickly and too expensive to train everyone)
  - Multiclass classification problem: given as input correctly labeled training data, learn to predict the codes for new symptoms

# Parameter Tying

- We saw $l_2$ regularization as a way to prefer simpler models
- Another type of simple model might be a Bayesian network in which many of the parameters (i.e., the conditional probability distributions) are the same
- This type of parameter tying is used in neural networks as well (though it is typically done by hand)
- Study the design of regularization based methods for parameter tying and improved inference/sampling methods for models with tied parameters

# Graphical Models

- Generalization of Bayesian networks very popular in the machine learning community (take the class!)
- Lower bounds for continuous "partition functions"
- Theoretical guarantees on the exactness of inference in continuous graphical models
- Faster algorithms (via Frank-Wolfe) for learning in latent variable models

# Please evaluate the course!

eval.utdallas.edu