ENEE759G: Advanced Topics in Computer Engineering - Unsupervised Learning

Fall 2014 (TuTh 3:30 – 4:45; EGR 3106)
Instructor: Joseph JaJa
Course Syllabus

Course Objectives: The course will cover core topics in unsupervised learning with a focus on related statistical techniques. Topics covered will include: density estimation, latent variable models, mixture models, clustering, graphical models and inference, learning graphical models, and association rules. A significant background in probability and statistics, linear Algebra, and data structures and algorithms is required for this course.

Course prerequisites: Graduate standing

Prerequisite topics: Strong background in probability and statistics, Linear Algebra, and data structures and algorithms.

Textbooks: No textbook is required for this course but the following can be used as references:

T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, second edition, Springer, 2009.

K. Murphy, Machine Learning: A Probabilistic Perspective, The MIT Press, 2012.

Core Topics:

1. **Introduction**
    - Basic framework and concepts
    - Probability Review
    - Probability density estimates for low-dimensional data
    - Maximum Likelihood Estimation and Bayesian Estimation.
2. **Latent Variable Models**
    - Principal Component Analysis
    - Gaussian Mixture Models and the EM algorithm
    - Introduction to Factor Analysis
    - Independent Component Analysis
3. **Clustering**
    - Proximity measures and evaluation methodologies
    - The k-means algorithm and its variant the k-medoid algorithm
    - Hierarchical clustering
    - Spectral clustering
4. **Directed Graphical Models**
    - Basic definitions and properties
    - Naïve Bayesian networks
    - Inference: Exact and Approximate
    - Learning Bayesian networks

5. **Undirected Graphical Models**
    - Basic definitions and concepts
    - Markov properties
    - Factor graphs
    - Inference algorithms
6. **Association Rules**
    - Market Basket analysis and overall strategy
    - The Apriori Algorithm
    - Rule Generation

Midterm: October 23 – 40%; Final: Date TBA – 40%; Homeworks: 20%

Homeworks: For each of *eight* of the techniques listed below, you are supposed to describe a specific real-world application that has been addressed using that technique. Your description must be *less than a 2-page pdf document* using the following format:

- *Statement of the Problem*
- *Data Used*
- *Technique Used*
- *Summary of Results*
- *Reference(s)*

Select *eight* out of the following list of techniques:

1. Parzen density estimate
2. Maximum Likelihood estimation
3. Principal Component Analysis
4. Factor Analysis
5. Independent Component Analysis
6. K-means clustering
7. Hierarchical clustering
8. Spectral clustering
9. Naïve Bayesian networks
10. Bayesian networks
11. Undirected Graphical Models

The due date for each technique will be within a week after the topic is covered in class.

Contact Information

Instructor:	Joseph JaJa
Office:	3433 A. V. Williams
Office Hours:	M 4-5 and Th 1:30-3 or by appointment
Email:	Joseph@umiacs.umd.edu
Phone:	405-1925