

ENEE752: Computational Intelligence and Knowledge Engineering

Fall 2009 (MW 2:00 – 3:15)

Instructor: Joseph JaJa

Course Syllabus

Course Objectives: The course will cover core topics in machine learning and their applications to data mining and knowledge discovery. Topics covered will include: statistical learning models; decision trees; rule-based and nearest neighbor classifiers; neural networks; Bayesian networks; support vector machines; clustering; and ensemble methods. Students are expected to develop a good understanding of a *web data mining application* and its relationship to the techniques covered in the course.

Course prerequisites: Graduate standing

Prerequisite topics: Basic concepts in probability and statistics, data structures and algorithms, Boolean logic, and object oriented programming.

Textbook: Introduction to Data Mining, Pang-Ning Tan, Michael Steinback, and Vipin Kumar, Pearson, Addison-Wesley, 2006 (required).

Course Web Page: <http://www.umiacs.umd.edu/~joseph/classes/enee752/Fall09/index.htm>

Core Topics:

1. Introduction

- Machine Learning
- Basic Terminology and Sample Applications
- Brief Introduction to Web Data Mining Applications

2. Machine Learning Framework

- Data Representation and Basic Concepts
- Statistics
- Quantization and Similarity Measures
- Principal Component Analysis

3. Classification Concepts and Decision Trees

- Introduction to Classification Concepts
- Strategies for Building Decision Trees
- The Overfitting Problem
- Evaluation Strategies in Classification

4. Rule-Based and Nearest-Neighbor Classifiers

- Basic strategy
- Building Rule-Ordering and Nearest-Neighbor Classifiers
- Characteristics of these classifiers

5. Neural Networks

- Basic Concepts and Methodology
- A Basic Learning Algorithm
- Multi-layered Networks and the Back-Propagation Algorithm

6. Support Vector Machines

- Maximum Marginal Classifiers
- Brief Overview of Related Nonlinear Optimization Techniques
- Nonlinear Support Vector Machines

7. Bayesian Networks

- Maximum Likelihood and Maximum A Posteriori Estimates
- Joint Probability, Bayes Theorem, and Conditional Independence
- Naïve Bayes Classifiers
- Basic Concepts of Bayesian Networks and Related Inferencing

8. Association Rules

- Basic Concepts
- Apriori Algorithm
- Rule Generation

9. Ensemble Methods

- Overall Strategy
- Bagging and Boosting
- Evaluation of Different Ensemble Methods

10. Clustering Techniques

- Basic Concepts
- The k-Means Algorithm
- Hierarchical Clustering
- Introduction to Probabilistic Techniques (Maximum Likelihood and Mixture Modeling)

Homework (20%) – almost weekly except for exam weeks

Two Midterms (25% each) – October 7, November 18

Project (30%) – December 15

Project:

Each student is expected to select a web data mining application and write a paper (no more than 20 typewritten pages) describing the application and the best known data mining techniques used to handle that application. Each paper is supposed to contain the following three elements:

- A brief description of the application and its importance;
- Best known techniques for handling the application; and
- Overall evaluation of the effectiveness and scalability of these techniques.

Proposals covering the first item and a preliminary list of references are **due September 30, 2009**.

Contact Information

Instructor:	Joseph JaJa
Office:	3433 A. V. Williams
Office Hours:	M,W 4-5:15 or by appointment
Email:	Joseph@umiacs.umd.edu
Phone:	405-1925