ENEE759G: Data Mining and Knowledge Discovery

Instructor: Joseph JaJa

Fall 2005 Course Syllabus

Course Objectives: The course will cover fundamental techniques used for analyzing and classifying large scale scientific and business data. These techniques, primarily based on machine learning and statistical methodologies, will include: statistical models and patterns, supervised learning, Bayesian and neural networks, support vector machines, search and optimization, finding patterns and rules, anomaly detection, and content based retrieval.

Course prerequisites: Graduate standing

Prerequisite topics: Basic algorithms and optimization techniques, and a good background in statistics. A background in nonlinear optimization is desirable.


References:

Core Topics:

1. Data Preprocessing and Exploration (Chapters 2 and 3)
   - Sampling
   - Principal Component Analysis and Singular Value Decomposition
   - Feature Extraction
   - Exploratory Data Analysis

2. Fundamental Classification Strategies (Chapters 4 and 5)
   - Decision Trees
   - Bayesian Networks
   - Neural Networks
   - Support Vector Machines
3. Clustering Techniques (Chapter 8)

- Basic Clustering Techniques
- Probabilistic Techniques (Maximum Likelihood and Mixture Modeling)
- Cluster Evaluation

4. Mining for Rules (Chapter 6)

- Association Rules
- Generalized Association Rules
- Sequential Patterns

6. Anomaly Detection (Chapter 10)

- Statistical Approaches
- Proximity-Based and Density Based Outlier Detection
- Clustering-Based Techniques

Course Grade: Midterm (30%); Final (30%); Project (40%)

Project

Each student is expected to define a project that involves two of the general techniques discussed in class and explain in some details their performances on at least three significant data sets. A proposal explaining the project and the experimental work to be carried out is due on September 29, 2005.

Information about software and data sets related to all the topics covered in class can be found at the textbook web site: www-users.cs.umn.edu/~kumar/dmbook

Each student is supposed to make a presentation about her project during the last two weeks of the class. Final reports are due December 8.

Contact Information: joseph@umiacs.umd.edu; 301-405-1925.

Office: 3433 A.V. Williams Bldg; Office Hours: Monday, Wednesday 3-4:30

Midterm: Tuesday, October 18; Final: Wednesday, Dec. 21, 10:30-12:30 (may be rescheduled)