

Data and Visual Analytics
Spring 2009: CSE8803DVA, CSE4803DVA
Instructor: Guy Lebanon

Overview

Data and visual analytics is an emerging field concerned with analyzing and visualizing complex high dimensional data. The course will introduce students to the field by covering state-of-the-art analysis and visualization techniques. It will emphasize practical challenges involving complex real world data and include case studies involving text documents, social networks, election data, shopping patterns, movie preferences, and disease surveillance.

Prerequisites

- Programming: knowledge of Java, Matlab or R
- Mathematics: basic algebra, calculus, introductory probability

Grade Composition

33% homework, 33% project, 34% exam

Topics

- 1. Introduction**
motivation, a taxonomy of data, data analysis tasks, visualization tasks
- 2. Data Preprocessing**
noise removal, imputation, outlier removal, feature extraction
- 3. Visualization Techniques and Exploratory Data Analysis**
line drawings, scatter plots, contour plots, trellis displays, parallel coordinates, interactive visualization, evaluation
- 4. Density Estimation and Clustering**
parametric and non-parametric estimation, bias-variance tradeoff, k-means, hierarchical clustering, distributional clustering
- 5. Regression and Classification**
linear and local regression, linear classifiers, kernel machines, regularization
- 6. Dimensionality Reduction**
PCA, MDS, non-negative matrix factorization, manifold learning
- 7. Discovering and Visualizing Graph Data**
link analysis, power laws, case studies: visualizing social networks, mining the internet
- 8. Discovering and Visualizing Logic Clauses and Relations**
association rules, logic learning, case study: mining and visualizing shopping patterns
- 9. Analyzing and Visualizing Text**
search engines, text visualization, topic models, case study: visualizing large corpora
- 10. Analyzing and Visualizing Spatial and Temporal Data**
smoothing and clustering across time and space, case study: disease surveillance
- 11. Analyzing and Visualizing non-Metric Data**
non-metric multidimensional scaling, case study: election data and movie preferences