

CSC 391/691 **Spring 2008**
Data Mining

Professor: Torgersen

Office Phone: 758-5536

Office Hours: MW 2:00 – 3:00 and by appointment

Text: Introduction to Data Mining, by Pang-Ning Tan, Michael Steinbach, and Vipin Kumar

Goals:

1. Survey of the issues and techniques used in Data Mining
 - (a) Types of data, data quality, data pre-processing, and measures of similarity
 - i. Principal component analysis
 - (b) Exploring data and data visualization
 - (c) Classification
 - i. Decision Trees
 - ii. Models and model fitting
 - iii. Comparing classifiers
 - iv. Rule-based classifiers
 - v. Bayesian classifiers
 - vi. Neural net classifiers
 - vii. Support vector machines
 - (d) Clustering
 - i. Cluster analysis
 - ii. K-means clustering
 - iii. DBSCAN
 - (e) Data mining of image data
 - i. Morphological image processing
 - ii. Feature recognition
 - iii. Wavelets, iris recognition and the Daugman algorithm
 - iv. Image segmentaion
 - (f) Linear Algebraic techniques for Data mining
 - i. Correlation
 - ii. Term document matrices
 - iii. Non-negative matrix factorization
 - iv. Application: spectral data analysis

Expectations:

1. Attendance
2. Communicate if something is confusing
3. Your best effort

Grading: 2 exams (65%). As many problem sets/projects as we can schedule (35%).

Graduate Student Distinction:

Graduate students will be required to answer more advanced questions on take-home problem sets, and will be required to include advanced features on their data mining computing assignments.

Disability Notice:

If you have a disability that may require an accommodation for taking this course, then please contact the Learning Assistance Center (758-5929) within the first two weeks of the semester.

Pandemic Planning Notice:

The University has requested that faculty collect personal contact information as part of emergency planning and preparation. The information you provide is strictly confidential.