

CSE 494/598: Introduction to Data Mining

Fall 2009

Time: T, Th: 10:30—11:45am
Place: BYAC 260

Instructor:

Dr Jieping Ye
Office: Brickyard 568
Office Hours: Tuesday 1-3pm
Phone: 480-727-7451
Email: Jieping.Ye@asu.edu

TA:

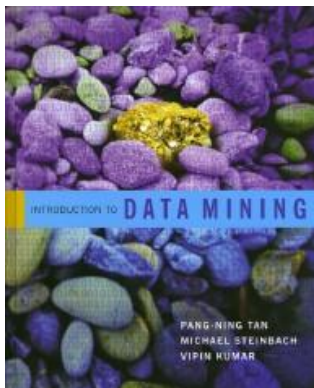
Liang Sun
Office Brickyard 584 AB
Office hours: Wednesday 10am-12noon
Email: sun.liang@asu.edu

Course Description

Recent advances in technology along with the phenomenal growth of the Internet have resulted in an explosion of data collected, stored, and disseminated by various organizations. Because of its massive size, it is difficult for analysts to sift through the data even though it may contain useful information. Data mining holds great promise to address this problem by providing efficient techniques to uncover useful information hidden in the large data repositories.

The key objectives of this course are two-fold: (1) to teach the fundamental concepts of data mining and (2) to provide extensive hands-on experience in applying the concepts to real-world applications. The core topics to be covered in this course include classification, association analysis, clustering, anomaly detection, and semi-supervised clustering. At the end of this course, students are expected to possess the fundamental skills needed to conduct research in data mining or to apply data mining techniques to real-world applications.

Textbook



Introduction to Data Mining (2005)

By Pang-Ning Tan, Michael Steinbach, Vipin Kumar

Addison Wesley

ISBN: 0-321-32136-7

Available at Amazon

Course Outline

1. Introduction
 - What is data mining?
2. Data Preprocessing
 - Data sampling, data cleaning, feature selection, and dimensionality reduction
3. Classification
 - Tree-based, rule-based, and instance-based methods
 - Bayesian methods (naive Bayes and Bayesian belief networks)
 - Neural networks, linear discriminant analysis, support vector machines, and ensemble methods
 - Model evaluation
4. Association Analysis
 - Apriori algorithm and its extensions
 - Pattern evaluation (subjective and objective interestingness measures)
 - Sequential patterns and graph mining
5. Clustering
 - Partitional and hierarchical clustering methods
 - Graph-based and density-based methods
 - Cluster evaluation
6. Anomaly Detection
 - Statistical-based and density-based methods
7. Advanced topic
 - Semi-supervised clustering
8. Case study
 - Gene expression pattern image mining

Grading

- Homework (5-6) 30%
 - CSE 494: 5 (top five scores count)
 - CSE 598: 6
- Project (2) 20%
 - Group project (up to two group members)
- Exam (2) 40%
 - Exam I: closed-book exam
 - Exam II: take-home exam
- Quiz (2) 10%

Class Policy

Assignments are due at the beginning of the lecture. Late assignments will not be accepted. Attendance to lecture is mandatory.

Academic Integrity

Violations of the University Academic Integrity policy will not be ignored. Penalties include reduced or no credit for submitted work, a failing grade in the class, a note on your official transcript that shows you were punished for cheating, suspension, expulsion and revocation of already awarded degrees. The university requires that should I implement any of these penalties, I must report the matter to the Dean's office. The university academic integrity policy can be found at <http://provost.asu.edu/academicintegrity>.