

CS378: Introduction to Data Mining

Fall 2006

Time: TTh: 9:30-11 am

Location: BUR 208

Instructor: Inderjit Dhillon

Homepage: <http://www.cs.utexas.edu/users/inderjit>

E-mail: inderjit@cs.utexas.edu

Office: ACES 2.332, Ph: 471-9725

Office Hours: Tue 11am-noon. Other times by appointment.

Class URL: <http://www.cs.utexas.edu/users/inderjit/courses/cs378/>

TA: Jason Davis

E-mail: jdavis@cs.utexas.edu

Prerequisites: Basic knowledge of programming, data structures, and algorithms.

Course Description:

Recent times have seen a massive growth in the raw data that is available electronically. Data mining is the *automatic discovery* of interesting patterns and relationships in *massive data sets*. This undergraduate course will provide an introduction to the topic of data mining. Areas covered include data preprocessing, visualization, predictive modeling, association analysis, clustering and anomaly detection. Basic concepts, representative algorithms and evaluation techniques will be covered.

Textbook:

Introduction to Data Mining by P. Tan, M. Steinbach, V. Kumar, Addison Wesley, 2006.

Grading:

20 pts: Homeworks

25+25 pts: Two Midterm Exams

25 pts: Class Project

5 pts: Attendance and participation in class discussions

Disabilities statement: “The University of Texas at Austin provides upon request appropriate academic accommodations for qualified students with disabilities. For more information, contact the Office of the Dean of Students at 471-6259, 471-4641 TTY.”