

CS395T/CAM395T (Spring 2008)

Data Mining: A Statistical Learning Perspective

Class times and location: MW: 9:30-11am, PAR 303

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: Mon 11am-noon. Other times available by appointment.

Class URL: <http://www.cs.utexas.edu/users/inderjit/courses/dm2008/>

Prerequisites: Basics (undergraduate level) of linear algebra (M341 or equivalent) and some mathematical sophistication.

Course Description: Recent times have seen an explosive growth in the amount of raw data available electronically. Data mining is the automatic discovery of interesting patterns and relationships in massive data sets. This graduate course will focus on various mathematical and statistical aspects of data mining. Topics covered include supervised learning (regression, classification, support vector machines) and unsupervised learning (clustering, principal components analysis, dimensionality reduction). The technical tools used in the course draw from linear algebra, multivariate statistics and optimization. The main concepts used from these technical areas will be covered in class, but undergraduate level linear algebra is a pre-requisite (see above).

A substantial portion of the course will concentrate on research projects, where students will choose a well-defined research problem. Projects can vary in their theoretical/mathematical content, and in the amount of implementation/programming involved. Projects will be conducted by teams of up to 2-3 students.

Grading:

10 + 30 pts: Class Project (First submission + Final submission)

20 pts: Homeworks

25 pts: Midterm

10 pts: Class presentation of a research paper or book chapter/section

5 pts: Attendance and participation in class discussions

Textbook:

Pattern Recognition and Machine Learning by C. Bishop, Springer, 2006.

Other References:

Elements of Statistical Learning: Data Mining, Inference, and Prediction by T. Hastie, R. Tibshirani, J. Friedman, Springer-Verlag, 2001.

Pattern Classification by R. Duda, P. Hart and D. Stork, John Wiley and Sons, 2000.

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.”