

Statistical Learning and Data Mining
CS 363D/ SDS 358
Unique: 51975/57460

When/Where

WEL 1.316
Spring 2015
Mon. & Wed., 3:30 – 5:00

Instructors

Instructor: Prof. Pradeep Ravikumar
GDC 4.808, pradeepr@cs.utexas.edu, 8–9136
Office Hours: Tues., 4:00 – 5:00

TAs: Ian En-Hsu Yen <a061105@gmail.com>
Xueyu Mao <xmao@cs.utexas.edu>
Office Hours: TBD

I. Course Objectives:

Overview:

In recent years, rapid developments in data collection and storage technologies have led to data sets that are “big” in many senses of the word. Data mining is the automatic discovery of interesting patterns and relationships in such “big data”. This undergraduate course will provide an introduction to the topic of data mining, and some statistical principles underlying its key methods. Topics covered will include data preprocessing, regression, classification, clustering, dimensionality reduction, and association analysis.

II. Course Schedule (Tentative):

Note that these are tentative and are subject to minor changes (including homework release dates).

Date	Title	Readings	Misc.
Jan 21	Data	Chap 1	
Jan 26	Data Contd	Chap 1	
Feb 02	Exploratory Data Analysis	Chap 2	HW1 out
Feb 04	Classification: Decision Trees	Chap 4	
Feb 09	Classification: Decision Trees Contd.	Chap 4	
Feb 11	Classification: Practical Issues	Chap 4	HW2 out
Feb 16	Classification: Evaluation	Chap 4	
Feb 18	Adv. Classification: Nearest Neighbor	Chap 5.2	
Feb 23	Linear Algebra Review; SVD	Slides; Appendix A, B.1	
Feb 25	Linear Algebra Review; SVD	Slides; Appendix A, B.1	HW3 out
Mar 02	Probability Theory Review	Slides; Appendix C	
Mar 04	Probability Theory Review	Slides; Appendix C	
Mar 09	Classification: Naïve Bayes	Slides; Chap 5.3.1-5.3.3	
Mar 11	Midterm		In Class
Mar 16, 18	Spring Break		
Mar 23	Regression	Slides; Appendix D	
Mar 25	Regression	Slides; Appendix D	HW4 out
Mar 30	Clustering		
Apr 01	Clustering: Kmeans	Chap 8	
Apr 06	Clustering: Hierarchical Clustering	Chap 8	
Apr 08	Clustering: Contd	Chap 8	HW5 out
Apr 13	Association Rules	Chap 6	
Apr 15	Association Rules Contd.	Chap 6	
Apr 20	Adv. Classification: Rule Based Classifier	Chap 5.1	HW6 out
Apr 22	Adv. Classification: Rule Based Classifier	Chap 5.1	
Apr 27	Anomaly Detection	Chap 10	
Apr 29	Anomaly Detection		
May 04	Class Review		
May 06	Class Review		
TBD	Final Exam		Exam Time: TBD

III. Course Requirements:

1. Course Readings/Materials:

Textbooks:

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

Textbook Website: <http://www-users.cs.umn.edu/~kumar/dmbook/index.php>.

2. Assignments, Assessment, and Evaluation

a. Grading Policy:

1. 6 Homeworks (60%)

2. Exams

a. 1 Midterm (15%)

b. 1 Final: (25%)

b. Homework Policy: Each student is expected to submit an individually written homework. When using information from papers, or other external sources, please cite this information. The homeworks will be due at the beginning of class on the due date. There will be two “free” late days, that you could use either all on one homework, or on two different homeworks. Otherwise, a homework will be worth 50% if one day late, and 0% if it is two or more days late. It is required to submit all homeworks even if after two days, if you do not want an incomplete grade.

c. Class attendance and participation policy:

I expect students to participate actively in the class.

IV. Academic Integrity

University of Texas Honor Code

Each student in this course is expected to abide by the University of Texas Honor Code.

The core values of The University of Texas at Austin are learning, discovery, freedom, leadership, individual opportunity, and responsibility. Each member of the university is expected to uphold these values through integrity, honesty, trust, fairness, and respect toward peers and community.

Collaboration Policy

You are encouraged to study together and to discuss information and concepts covered in lecture with other students, especially on Piazza. However, this cooperation should never involve one student having possession of a copy of all or part of work done by someone else.

Should copying occur, both the student who copied work from another student and the student who gave material to be copied will both automatically receive a zero for the assignment. Penalty for violation of this Code can also be extended to include failure of the course and University disciplinary action.

V. Other University Notices and Policies

Documented Disability Statement

Any student with a documented disability who requires academic accommodations should contact Services for Students with Disabilities (SSD) at (512) 471-6259 (voice) or 1-866-329-3986 (video phone). Faculty are not required to provide accommodations without an official accommodation letter from SSD.

- Please notify me as quickly as possible if the material being presented in class is not accessible (e.g., instructional videos need captioning, course packets are not readable for proper alternative text conversion, etc.).
- Contact Services for Students with Disabilities at 471-6259 (voice) or 1-866-329-3986 (video phone) or reference SSD's website for more disability-related information: http://www.utexas.edu/diversity/ddce/ssd/for_cstudents.php

Behavior Concerns Advice Line (BCAL)

If you are worried about someone who is acting differently, you may use the Behavior Concerns Advice Line to discuss by phone your concerns about another individual's behavior. This service is provided through a partnership among the Office of the Dean of Students, the Counseling and Mental Health Center (CMHC), the Employee Assistance Program (EAP), and The University of Texas Police Department (UTPD). Call 512-232-5050 or visit <http://www.utexas.edu/safety/bcal>.

Q drop Policy

The State of Texas has enacted a law that limits the number of course drops for academic reasons to six (6). As stated in Senate Bill 1231:

“Beginning with the fall 2007 academic term, an institution of higher education may not permit an undergraduate student a total of more than six dropped courses, including any course a transfer student has dropped at another institution of higher education, unless the student shows good cause for dropping more than that number.”

Emergency Evacuation Policy

Occupants of buildings on the UT Austin campus are required to evacuate and assemble outside when a fire alarm is activated or an announcement is made. Please be aware of the following policies regarding evacuation:

- Familiarize yourself with all exit doors of the classroom and the building. Remember that the nearest exit door may not be the one you used when you entered the building.

- If you require assistance to evacuate, inform me in writing during the first week of class.
- In the event of an evacuation, follow my instructions or those of class instructors.

Do not re-enter a building unless the Austin Fire Department, the UT Austin Police Department, or the Fire Prevention Services office gives you instructions.
