George Mason University

Department of Computer Science

CS 484: Data Mining

Spring 2017

Professor Jessica Lin

News & Announcements
1/19: Welcome to class!
2/5: HW1 posted. Due date 2/16 at 4:30pm.
2/21: HW2 posted. Due date 3/2 at 4:30pm.
3/7: HW3 posted. Due date 3/26 at 11:59pm.
3/10: Project guidelines posted (on Blackboard). Due date 3/19 at 11:59pm.
3/16: Project proposal deadline extended to 3/25 at 11:59pm.
4/7: HW4 posted. Due date 4/20 at 4:30pm.

Course Description

Basic principles and methods for data analysis and knowledge discovery. Emphasizes developing basic skills for modeling and prediction and performance evaluation. Topics include system design; data quality, preprocessing, and association; event classification; clustering; biometrics; business intelligence; and mining complex types of data.

Class Time and Location

Thursday, 4:30-7:10 pm
Planetary Hall 122

Instructor

Dr. Jessica Lin
Office: Engineering Building 4419
Phone: 703-993-4693
Email: jessica [AT] cs [DOT] gmu [DOT] edu
Office Hours: Wednesday & Thursday 3-4pm

Teaching Assistant

Monjura Rumi
Email: mrumi [AT] gmu [DOT] edu
Office Hours: Tuesday 4-6pm at Engineering Building 5321

Prerequisites

Grade of C or better in CS 310 and STAT 344

Course Outcomes

- The ability to apply computing principles, probability and statistics relevant to the data mining discipline to analyze data.
- A thorough understanding of model programming with data mining tools, algorithms for estimation, prediction, and pattern discovery.
- The ability to analyze a problem, identifying and defining the computing requirements appropriate to its solution: data collection and preparation, functional requirements, selection of models and prediction algorithms, software, and performance evaluation.
- The ability to understand performance metrics used in the data mining field to interpret the results of applying an algorithm or model, to compare methods and to reach conclusions about data.
- The ability to communicate effectively to an audience the steps and results followed in solving a data mining problem (through a term project)

Grading

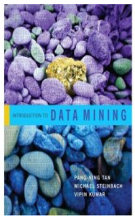
Assignments: 25%
Project: 20%
Midterm: 25%
Final: 30%

Exams

There will be one midterm exam and a final exam covering lectures and readings (both will be in class, closed book). The final exam is comprehensive. Exams must be taken at the scheduled time and place, unless prior arrangement has been made with the instructor. Missed exams cannot be made up.

Textbooks

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



Topics

- Ch.1: Introduction
- Ch.2: Data
- Ch.4: Classification
- Ch.5: Classification: Alternative Techniques
- Ch.6: Association Analysis: Basic Concepts and Algorithms
- Ch.7: Association Analysis: Advanced Concepts
- Ch.8: Cluster Analysis: Basic Concepts and Algorithms
- Ch.9: Cluster Analysis: Additional Issues and Algorithms
- Ch.10: Anomaly Detection

Honor Code Statement

The GMU Honor Code is in effect at all times. In addition, the CS Department has further honor code policies regarding programming projects, which are detailed here. Any deviation from the GMU or the CS department Honor Code is considered an Honor Code violation. All assignments for this class are **individual** unless otherwise specified.

Learning Disability Accommodation

If you have a documented learning disability or other condition which may affect academic performance, make sure this documentation is on file with the Office of Disability Services and then discuss with the professor about accommodations.

Tentative Schedule

Week	Date	Topic	Slide	Assigned	Due	Note
1	1/26	Introduction (Ch. 1)	Introduction			
2	2/2	Data (Ch. 2)	Data	HW1		
3	2/9	Data Classification (Ch. 4-5)	Classification 1			
4	2/16	Classification	Classification 2	HW2	HW1	
5	2/23	Classification	Classification 3			
6	3/2	Classification	Classification 4		HW2	
7	3/9	Classification		HW3		

				Project		
8	3/16	Spring Break				
9	3/23	Midterm			Project proposal (3/25) HW3 (3/26)	
10	3/30	Clustering (Ch. 8-9)	Clustering 1			
11	4/6	Clustering	Clustering 2	HW4		
12	4/13	Clustering Association Analysis	Clustering 3 Association Rules 1			
13	4/20	Anomaly Detection			HW4	
14	4/27	Project presentation				
15	5/4	Project presentation				
16	5/11	Final Exam (4:30-7:15pm)				