Csci 581: Data Mining

Course Overview

This course is designed to teach data mining techniques for analyzing large amounts of data. The topics include exploratory data analysis, classification, clustering, text mining, web mining, recommender systems, neural network as well as advanced topics covering the applications of data mining in cross-cutting domains.

Learning Objectives

- A solid understanding of the basic concepts, principles, and techniques in data mining.
- An ability to analyze real-world applications, to model data mining problems, and to assess different solutions.
- An ability to design, implement, and evaluate data mining software.
- Be able to model, evaluate, visualize and communicate statistical models.

Class Information

Time: Tuesday/Thursday 4:00 PM – 5:15 PM

Location: 235 Weir Hall

Instructor Information

Name: Naeemul Hassan

Office: 216 Weir Hall

Office Hour: Tuesday/Thursday 2:00 PM – 4:00 PM

Contact: nhassan@olemiss.edu

Prerequisites

Pre-requisite: (CSCI 211 and CSCI 223) or Graduate Standing

Textbook

- (Required) [TSK] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. Introduction to Data Mining, Addison-Wesley, 2006. ISBN 0-321-32136-7. (Sample chapters http:/www-users.cs.umn.edu/~kumar/dmbook/index.php)
- (Required for relevant chapters) [MRS] Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze. Introduction to Information Retrieval, Cambridge University Press. 2008. (Free book at http://nlp.stanford.edu/IR-book/)
- (Reference) [MMDS] Jure Leskovec, Anand Rajaraman and Jeff Ullman. Mining of Massive Datasets, 2nd ed., Cambridge University Press. (Free book at http://www.mmds.org/#ver21)

- (Reference) [DMA] Mohammed Zaki and Wagner Meira. Data Mining and Analysis: Fundamental Concepts and Algorithms. Cambridge University Press. (Free book at http://www.dataminingbook.info/pmwiki.php/Main/BookDownload)

Besides these, we will use other online materials from time to time. References of the materials will be provided to the students.

Grading Policy

Midterm Exam: 25% (Date: Tuesday, October 04, 2016, Weir Hall 235)
Final Exam: 30% (Date: Tuesday, December 06, 2016, Weir Hall 235)
Homework (HW): 15% (Must be done independently)
Project (P): 30%
- P1: 7.5% (Must be done independently)
- P2: 7.5% (Must be done independently)
- P3: 15% (Group Project. At most three students per group)

The final letter grades will be based on the curve of students' performance. We will follow the Plus/Minus grading scale.

Assignments and Deadlines

- All the assignments must be submitted through Blackboard.
- We will **NOT** take hardcopy or email submission, unless the university verifies that Blackboard was malfunctioning or unavailable. If you are not able to submit through Blackboard due to its technical failure, you can email your assignment to us, together with a screenshot showing the technical failure. We will verify with the university.
- Everything is due by 11:59pm on the due date. The deadline is automatically managed by Blackboard. You can still turn in assignment after the deadline. However, you automatically lose 5 points per hour after the due time, till you get 0. We cannot waive the penalty, unless there was a case of illness or other substantial impediment beyond your control, with proof in documents.

Announcements

We will use Blackboard for this course. All lecture slides, homework and project documents will be available here at appropriate times. If there is any announcement, that will be posted in Blackboard. You will also be notified by email. We will use the Discussion Board for having healthy discussions regarding the homework and projects. **Solutions to any homework/project should not be posted there.**

Course Schedule

Date	Topic	Assignment	
		Out	Due
Tuesday, August 23, 2016	Introduction- Course, Language, Tools.		
Thursday, August 25, 2016	KD Process: Data Munging, Integration	P1	
Tuesday, August 30, 2016	Data Visualization		
Thursday, September 01, 2016	Classification Overview, Decision Tree		
Tuesday, September 06, 2016	Decision Tree Cont., Random Forest		P1
Thursday, September 08, 2016	Nearest Neighbor	HW1	
Tuesday, September 13, 2016	Bayesian Classifier		
Thursday, September 15, 2016	Support Vector Machine (SVM)		
Tuesday, September 20, 2016	SVM Cont., Kernel	HW2	HW1
Thursday, September 22, 2016	Model Evaluation		
Tuesday, September 27, 2016	Model Evaluation Cont.		
Thursday, September 29, 2016	Vector Space Model		HW2
Tuesday, October 04, 2016	Midterm Exam		
Thursday, October 06, 2016	Word2Vec	P2	
Tuesday, October 11, 2016	Similarity Measures		
Thursday, October 13, 2016	Clustering Overview, K-means		
Tuesday, October 18, 2016	K-means Cont., Hierarchical Clustering		
Thursday, October 20, 2016	Hierarchical Clustering Cont.,	HW3	
Tuesday, October 25, 2016	Web Mining Overview		
Thursday, October 27, 2016	Recommendation Techniques	P3	P2
Tuesday, November 01, 2016	Recommendation Techniques Cont.,		
Thursday, November 03, 2016	Neural Network		
Tuesday, November 08, 2016	Deep Learning, RNN, CovNet		P3.1
Thursday, November 10, 2016	Large Scale Data Analysis, Map Reduce		HW3
Tuesday, November 15, 2016	Map Reduce Cont., Hadoop, Spark		
Thursday, November 17, 2016	Computational Journalism		P3.2
Tuesday, November 22, 2016	Thanksgiving Holiday		
Thursday, November 24, 2016	Thanksgiving Holiday		
Tuesday, November 29, 2016	Project Presentation		P3.3
Thursday, December 01, 2016	Review Class		
Tuesday, December 06, 2016	Final Exam		
Thursday, December 08, 2016	Reserved For Final Exam		

Regrading

A regrading request must be made within 7 days after the score is posted on Blackboard. The instructor will regrade based on his own discretion.

Attendance

It is strongly recommended to attend all the class lectures. Moreover, the University requires the instructor to verify the attendance of students within the first two weeks of a semester. I will collect attendance rolls in the first three classes and submit an attendance verification report based on those rolls.

Students with Disabilities

University policy provides for reasonable accommodations to be made for students with verified disabilities on an individualized and flexible basis as specified under Section 504 of the Rehabilitation Act of 1973 and the Americans with Disabilities Act of 1990 (ADA).

Students with disabilities who believe they may benefit from classroom or other accommodations should contact the Office of Student Disability Services for information at 234 Martindale, 662-915-7128 (Phone), 662-915-7907 (TTY Phone), or sds@olemiss.edu (Email).

Copyrighted Materials

Materials used in connection with this course may be subject to copyright protection under Title 17 of the United States Code. Under certain Fair Use circumstances specified by law, copies may be made for private study, scholarship, or research.

Students should not share electronic copies of copyrighted materials with unauthorized users. Violations of copyright laws could subject individuals to federal and state civil penalties and criminal liability as well as disciplinary action under University policies.

Appropriate Use of Information Technology

The Information Technology (IT) Appropriate Use Policy sets forth the privileges of and restrictions on students, faculty, staff, and other users with respect to the computing and telecommunications systems offered by the University of Mississippi (UM). This policy is designed to protect the University community from illegal or damaging actions by individuals, either knowingly or unknowingly. Inappropriate use exposes the University to risks, including virus attacks, compromise of network systems and services, and legal issues. This policy directly addresses copyright issues related to illegal downloads and peer-to-peer file sharing. For questions about the Appropriate Use Policy, send an email to aup@olemiss.edu.

Academic Integrity

The University of Mississippi is dedicated to supporting and sustaining a safe and scholarly community of learning dedicated to nurturing excellence inside and outside of the classroom. Each student has a duty to become familiar with University values and standards reflected in University policies, and each student has a duty to honor University values and standards reflected in University policies. These policies are outlined in the M Book. For a complete listing of policies, please visit the University Policy Directory.

Verification of Student Attendance

The University must abide by federal guidelines to verify the participation of online students. For all course types, including thesis, internships, labs, online courses, etc., the instructor must verify your participation based on some type of participation. This may include submission of an online assignment or other course related contact with the instructor. (However, simply logging into Blackboard will not count as an academically related activity.) See the Attendance Policy for Online Education for more information.

Student Identity Policy

Federal regulations, our accrediting agency (SACS), and university policies require that safeguards are used to ensure that the student who receives the academic course credit is actually the person doing the work.
Students must present their student IDs before taking proctored exams. The course instructor may verify a student's identity through live or virtual meetings or by using an identity verification program.

Student Privacy Policy

The University of Mississippi protects the privacy of all students, including online and distance learning students, through adherence to the Family Educational Rights and Privacy Act of 1974 (FERPA) through compliance with other institutional policies and procedures governing the management and security of protected information of faculty, staff, and students, and by outlining the expectations of privacy for the university community as regards to electronic information. See the Student Information and Privacy Policy for more information.

As the instructor of this course, I reserve the right to adjust this schedule, assignments, grading policy, due dates and syllabus contents in any way that serves the educational needs of the students enrolled in this course.