

CSE 5243-0020 – Autumn 2019

Introduction to Data Mining

Time/Place: 12:45 PM-02:05 PM, Wed/Fri Caldwell Lab 177

Instructor: Dr. Ping Zhang; 310A LT; zhang.10631@osu.edu; <http://pingzhang.net>

Teaching Assistant (TA): Fang Zhou; 574 DL; zhou.1250@osu.edu

Level and credits: U/G, 3

Office Hours: (Instructor) Wed 10:30 AM – 11:30 AM; (TA) Mon 3:30 PM – 4:40 PM

Course Description: Knowledge discovery, data mining, data preprocessing, data transformations; clustering, classification, frequent pattern mining, anomaly detection, avoiding false discoveries, graph and network analysis; applications.

Prerequisites: Introduction to Databases, Introduction to Algorithms, or grad standing or permission of instructor

Textbooks:

- (Required) Jiawei Han, Micheline Kamber and Jian Pei, Data Mining: Concepts and Techniques (3rd ed), 2011
- (Required) Pang-Ning Tan, Michael Steinbach, and Vipin Kumar, Introduction to data Mining (1st or 2nd ed), 2006
- (Optional) Charu C. Aggarwal, Data Mining: The Textbook, Springer, 2015.
- (Optional) Mohammed J. Zaki and Wagner Meira Jr., Data Mining and Analysis: Fundamental Concepts and Algorithms, 2014.

Grading Plan:

- Quizzes and participation: 20%
 - 12 Quizzes
 - 8 attendances
- Homework: 30%
 - 3 programming assignments
- Midterm exam: 25% (Oct 9 in class)
- Final exam: 25% (Dec 9 at 4pm)

Academic Integrity Policy: Academic integrity is essential to maintaining an environment that fosters excellence in teaching, research, and other educational and scholarly activities. Thus, The Ohio State University and the Committee on Academic Misconduct (COAM) expect that all students have read and understand the University's Code of Student Conduct, and that all students will complete all academic and scholarly assignments with fairness and honesty. Students must recognize that failure to follow the rules and guidelines established in the University's Code of Student Conduct and this syllabus may constitute Academic Misconduct.

Course Syllabus:

Week	Date	Topic	Assignment
1	08/21	Introduction	
1	08/23	Introduction	
2	08/28	Data & Data Preprocessing	
2	08/30	Data & Data Preprocessing	
3	09/04	Classification: Basic Concepts/Methods	
3	09/06	Classification: Basic Concepts/Methods	
4	09/11	Classification: Basic Concepts/Methods	
4	09/13	Classification: Basic Concepts/Methods	
5	09/18	Classification: Advanced Methods	
5	09/20	Classification: Advanced Methods	
6	09/25	Association Analysis	
6	09/27	Association Analysis	
7	10/02	Association Analysis	
7	10/04	Association Analysis	Assignment 1 Due
8	10/09	Midterm Exam	
8	10/11	Autumn Break	
9	10/16	Association Analysis: Advanced Concepts	
9	10/18	Association Analysis: Advanced Concepts	
10	10/23	Association Analysis: Advanced Concepts	
10	10/25	Association Analysis: Advanced Concepts	
11	10/30	Cluster Analysis: Basic Concepts and Algorithms	
11	11/01	Cluster Analysis: Basic Concepts and Algorithms	Assignment 2 Due
12	11/06	Cluster Analysis: Basic Concepts and Algorithms	
12	11/08	Cluster Analysis: Basic Concepts and Algorithms	
13	11/13	Cluster Analysis: Additional Issues and Algorithms	
13	11/15	Cluster Analysis: Additional Issues and Algorithms	
14	11/20	Anomaly Detection	
14	11/22	Avoiding False Discoveries	
15	11/27	Thanksgiving Break	
15	11/29	Thanksgiving Break	Competition Due
16	12/04	Data Competition and AI in Medicine	Assignment 3 Due
16	12/06	Exam Week Begin	
17	12/09	Final Exam	