CS565 Algorithmic Data Mining Syllabus
Spring 2025

Course description

This course is targeted at students who require an advanced level of proficiency in data science concepts and techniques. The course emphasizes on the theoretical skills as well as the working/practical knowledge. It examines at depth concepts such as clustering, network analysis, dimensionality reduction, matrix decomposition and submodular optimization.

Prerequisites: Students taking this class must have some prior familiarity with programming, at the level of CS 112, or equivalent. CS 132, CS 131, CS 237 and CS 330 and CS 365 are also required. Graduate students should have taken courses equivalent to the ones listed.

Learning outcomes: Students who successfully complete this course will be proficient in the theoretical aspects of unsupervised data analysis techniques. They will also understand the efficiency issues and systems issues related to working on very large datasets.

Instructors:
Prof. Dora Erdos edori@bu.edu office: CDS910
TA Anming Gu agu2002@bu.edu

Lectures: Mon/Wed 3:00-4:15 HAR212
Discussions:
Fri 1:25-2:15 IEC B04
2:30-3:20 CAS 208
3:35-4:25 CAS 212

Office hours: see on Piazza

Textbook

There is no required textbook for this course. All material that we use will be shared with the class. Some recommended reading:
Foundations of Data Science by Blum, Hopcroft, Kannan
Mining Massive Datasets by Leskovec, Rajaraman and Ullman

Communication: We will be using **Piazza for all discussion** pertaining to the class. You should post your questions about the material, lectures, homework or course logistics here. Piazza is highly preferred over sending the course staff emails. Most often your question and the answer will be just as useful to your fellow students as yourself. We encourage you to respond to questions. The course staff will be monitoring Piazza and chime in as needed.

Piazza https://piazza.com/bu/spring2025/cascs565
Homework: There will be three homework assignments throughout the semester. The assignments will take 2 weeks each and are to be submitted through Gradescope. Homeworks are to be completed individually.

https://www.gradescope.com/courses/969098 code: G34XKG

Quizzes: There will be 6 quizzes throughout the semester, roughly every other Monday at the start of class (the exact dates will be posted on Piazza). The quiz will take about 10-15 minutes. The content will be based on questions at the end of each slide deck and the lab problems. None of the quiz grades will be dropped. If you have to miss a quiz for a well-documented reason, e.g. illness, contact the instructor asap to set a makeup date.

Project: This course has a substantial semester long project performed individually or in pairs. you will pursue a project that shows the process on how a basic, foundational question evolves into the state-of-the-art exciting new algorithms and insights we have today. We will approach this from a top-down perspective. That is, you will select a current paper or research topic that you are interested in and will then backtrack to its origins. In the course of this project you will implement some of the old and new methods relevant to your paper and experiment with them. The project will culminate in a project report and a class presentation. For details please see the separate document about the project.

Exams: there will be no exams this semester.

Participation: Students have to attend at least 70% of the lectures to get a passing grade, attendance will be taken at the beginning of each class. If for any reason you are absent for more than 30% of the lectures, please contact the instructor asap.

Grading

15% Homework
35% Quizzes
50% Project

Academic Honesty

Homework: You may discuss homework assignments with classmates, but you are solely responsible for what you turn in. Collaboration in the form of discussion is allowed, but the work submitted has to be your own.

Quiz: Quizzes are closed book, no collaboration is allowed.

Project: By the nature of the project you will be using material from many sources, which is allowed and encouraged. However, anything you use that was not created by you has to be properly cited. Using material without citation constitutes to plagiarism.

We -- both teaching staff and students -- are expected
to abide by the guidelines and rules of the Academic Code of Conduct
(which is at
www.bu.edu/academics/policies/academic-conduct-code/).

Graduate students must also be aware of and abide by the GRS Academic Conduct code
www.bu.edu/cas/students/graduate/forms-policies-procedures/academic-discipline-procedures
.