

CSCI4390-6390 Syllabus

Introduction

This course will provide an introduction to the main topics in data mining and knowledge discovery, including: algebraic, geometric and probabilistic foundations, classification, regression, and clustering. Emphasis will be laid on the algorithmic approach.

Learning Objectives

For CSCI-4390: After taking this course students will be

- able to describe the fundamental data mining tasks like pattern mining, classification, regression and clustering
- able to analyze the key algorithms for the main tasks
- able to implement and apply the techniques to real world datasets

For CSCI-6390: After taking this course students will be

- able to describe the fundamental data mining tasks like pattern mining, classification, regression and clustering
- able to analyze the key algorithms for the main tasks
- able to implement and apply the techniques to real world datasets
- able to demonstrate understanding of more advanced topics in data mining
- able to implement more advanced algorithms

Prerequisites

You need a minimum of CS2300: Introduction to Algorithms. Linear algebra forms the foundation of data mining and machine learning, and therefore prior exposure to linear algebra is essentially a prerequisite. A good knowledge of probability and statistics is also a plus.

You are expected to know how to program. Class assignments will require the use of [Python](#), especially using NumPy (and PyTorch if needed). All assignments will be submitted as [JupyterLab Notebooks](#).

Textbook

The following textbook is **required** for the course:

[Data Mining and Machine Learning: Fundamental Concepts and Algorithms](#) (2nd Edition), Mohammed J. Zaki and Wagner Meira, Jr, Cambridge University Press, 2020.

Grading Policy

Your grade will be a combination of the following items.

- **Assignments (40%):** Assignments and HWs will be given throughout the semester. These will include an implementation component and can also have written questions. You can expect about 8-10 assignments over the semester.
- **Exams (50-55%):** There will be three exams covering the main topics of the course. The tentative exam dates are noted on the class schedule table. There is no comprehensive final exam. I may consider take-home exams, otherwise they will be in-person during class hours.
- **Attendance and Quizzes (5-10%):** Attendance is required. There may be pop-quizzes given throughout the semester. Students are also encouraged to participate in the class via discussions and engagement on the Campuswire forum.
- **Late Submissions:** Most assignments will be due just before midnight on the due date. Students can get an automatic one day extension for a 15% grade penalty. No further late assignments will be accepted.

The grading for CSCI4390 and CSCI6390 will be done separately taking into account the more advanced material required for CSCI6390 -- this includes extra/in-depth questions on the exam, and implementation of more advanced algorithms for the assignments. The letter grades typically are also based on different ranges for the two sections.

All assignments will be submitted online via **Submitty**, and discussions will be conducted via **Campuswire**. See the main page for the links.

Health Issues

Students should follow all guidelines from RPI related to health and safety for themselves and other campus members. All illness related accommodations will require officially approved excuse from RPI.

Students who are ill, under quarantine for COVID-19, or suspect they are ill will report that to Student Life. Student Life will verify and notify all faculty who have that student. Once notification is made, all faculty will make every reasonable effort to accommodate the student's absence and will communicate that accommodation directly to the student. Failure to make an appropriate accommodation for a verified or reasonably suspected case of illness may be appealable under the student grade appeal process. Students who need to report an illness should contact the Student Health Center via email or call 518-276-6287. For student seen off campus, a student may request an excused absence via <http://www.bit.ly/rpiabsence> with an uploaded doctor's note that excuses them.

Academic Integrity

Students must work independently on all course assignments. You may consult other members of the class on the assignments, **but you must submit your own work**. For instance you may discuss general approaches to solving a problem, but you must implement the solution on your own (similarity detection software may be used). **Anytime you borrow material from the web or elsewhere, you must acknowledge the source.** Copying and pasting from published sources or the internet is considered plagiarism and is not acceptable. Plagiarized work will receive an automatic grade of zero.

Student-teacher relationships are built on trust. Acts which violate this trust undermine the educational process. The Rensselaer Handbook of Student Rights and Responsibilities and The Rensselaer Graduate Student Supplement define various forms of Academic Dishonesty and procedures for responding to them. **Submission of any assignment that is in violation with these policies will result in a penalty that is deemed by the instructor to be appropriate to the infraction ranging from a grade of zero on the assignment in question, to failure of the class as a whole. The student will also be reported to the Dean of Students or the Dean of Graduate Education as appropriate.** Note that academic dishonesty will be dealt with severely and will be reported to the Dean of Students. If you have any questions concerning this policy before submitting an assignment, please ask for clarification.

Contents © 2022 [Mohammed J. Zaki](#) - Powered by [Nikola](#)