Course Description

This course will focus on understanding the structure of high-dimensional data and the mathematical tools we can use to characterize and reshape it for computational analysis. Several major threads will be woven throughout the course:

- What is the geometry of high-dimensional space?
- What are meaningful structures and dynamics present in high-dimensional data generated from real-world phenomenon?
- How can we reshape the data so that it is more amenable to analysis?

This is a survey course, and we intend to cursorially cover a range of techniques and applications. Topics may vary depending on instructor and student interest, so check the course website: https://www.math.toronto.edu/ywyu/MAT1841-2021-Fall/

Time and Location

MAT1841 (Fall 2021) Tuesday, Wednesday, Friday 15:00–16:00 in KP113

Course Instructors

Name	*E-mail*	*Office*	*Office Hours*
Yun William Yu	ywyu@math.toronto.edu	UTSG: BA 6252 / Zoom	By Appointment Only

Textbook

Foundations of Data Science, Blum, Hopcroft, and Kannan (2020) Cambridge University Press

The full text is available from the University of Toronto libraries as an online downloadable resource. There also exist partial earlier drafts of the textbook elsewhere on the internet, but I recommend you use the published version from the library.

Websites

The course website will be updated with the lecture schedule, notes, and problem sets:
https://www.math.toronto.edu/ywyu/MAT1841-2021-Fall/

Announcements will be posted to the Quercus page. Some homework assignments will also be submitted through Quercus (others will be submitted through Crowdmark).

Prerequisites

Linear algebra, probability, algorithms

Minor Assessments

There will be 11 homework assignments that will be due roughly weekly and will be assigned and collected online. There will be no extensions to posted homework due dates. However, the lowest homework mark will be dropped.

The homework problems will be a mix of theory and implementation. I recommend using Python for implementation, but will accept R, Julia, or C/C++. If you wish to do an implementation in any other language or framework, please clear it with me beforehand. Notably, I will not be accepting MATLAB implementations in this course, unless you present a very compelling reason.

Note that solutions should be as short, clear, and concise as possible. I will be taking marks off for long, meandering solutions to otherwise short problems, even if all of the reasoning is technically correct. Brevity is the soul of wit.

Major Assessments

There will be a final project. You will either (A) design and implement a data analysis method for a problem of your choosing, (B) prove a new result about one of the methods we covered, or (C) perform a survey of a group of relevant academic articles. You will submit a written report in the format of a conference proceedings or journal article and deliver a short presentation to your peers.

There will be no final exam or midterm.

Marking Scheme

Your final grade is determined in the following way:

Assignments	50 %
Term paper	30 %
In-class Presentation	20 %

Communication

When emailing the instructor, please use your official University of Toronto email address and mention MAT1850 in the subject line. Note that a response will be forthcoming during normal business hours.

Lecture Recordings

Please be aware that some of the lectures may be recorded and shared online. For any lectures that are recorded, I will announce before starting the recording in class.

Code of Behaviour / Plagiarism

The University of Toronto treats cases of academic misconduct very seriously. Academic integrity is a fundamental value of learning and scholarship at the UofT. Participating honestly, respectfully, responsibly, and fairly in this academic community ensures that your UofT degree is valued and respected as a true signifier of your individual academic achievement.

The University of Toronto's Code of Behaviour on Academic Matters outlines the behaviours that constitute academic misconduct, the processes for addressing academic offences, and the penalties that may be imposed. You are expected to be familiar with the contents of this document. Potential offences include, but are not limited to:

- In papers and assignments:
  - Using someone else's ideas or words without appropriate acknowledgement.
  - Submitting your own work in more than one course without the permission of the instructor.
  - Making up sources or facts.
  - Obtaining or providing unauthorized assistance on any assignment (this includes working in groups on assignments that are supposed to be individual work).
- On tests and exams:
  - Using or possessing any unauthorized aid, including a cell phone.
  - Looking at someone else's answers.
  - Letting someone else look at your answers.
  - Misrepresenting your identity.
  - Submitting an altered test for re-grading.
- Misrepresentation:
  - Falsifying or altering any documentation required by the University, including (but not limited to) doctor's notes.
  - Falsifying institutional documents or grades.

All suspected cases of academic dishonesty will be investigated following the procedures outlined in the *Code of Behaviour on Academic Matters*. If you have any questions about what is or is not permitted in this course, please do not hesitate to contact an instructor. If you have questions about appropriate

research and citation methods, you are expected to seek out additional information from an instructor or other available campus resources like the College Writing Centres, the Academic Success Centre, or the U of T Writing Website.

Students must not distribute, in any form, any course materials to any third parties.

Accessibility

Students with diverse learning styles and needs are welcome in this course. Please feel free to approach the instructor or contact Accessibility Services (accessibility.services@utoronto.ca) so we can assist you in achieving academic success in this course.