CSC 200/200H - Spring 2022

Instructor:	Fatemeh Nargesian	**Time:**	MW 16:50-18:05
Email:	fnargesian@rochester.edu	**Location:**	Gavett Hall Rm 301
Office:	WEGMN 3015	Office Hours:	M 14:15-15:30

Course Description: This course is intended to provide students with a hands-on feel for research in computer science. A significant component of the course will revolve around a set of semester-long open and computationally difficult projects. These projects will exercise the following abilities:

- researching, understanding, and presenting the state of the art in a specified semi-open problem
- grounding and solidifying this understanding by implementing a solution that approaches the state of the art
- attempting to identify areas where the state of the art might be pushed and pushing on it

Team members will present status reports to the rest of the class on a regular basis, via semi-formal discussion sessions in which team understanding, approaches, and progress will be critiqued and potentially modified. Teams will also prepare weekly written reports.

Several different projects will be pursued in parallel by teams of 3–5 members. Teams and topics will be selected by the instructor and TA a couple of weeks into the course on the basis of survey information and class discussion. Our hope is that most students who are going to drop the course will have done so by that time, so we can have stable teams. Depending on class size, we may have more than one team on a project, in semi-competition. The projects are likely to be related to data intelligence topics.

A second component of the course consists of reading, summarizing, and discussing research papers. This component will be a mix of lectures to teach you the expected background for these reading papers.

Another component of the course consists of guest lectures by researchers (faculty and/or senior graduate students) in the computer science department. Participation in critique sessions is an essential part of the experience, as is attendance at guest lectures. There may be 5-minute quizzes on material occurring in the previous class as a means of documenting attendance and attention. This could include the content of unscripted discussions.

Note that CSC 200 is no longer required for the CS degree. Students should not undertake this course lightly. It is expected to be a significant amount of work. Research is that way. It is also likely to be the first course where the professor is not providing material and answers, or even detailed instructions. That is also how research is.

Topics include: Search in high dimensional data, dimensionality reduction, sampling, time-series, web and data set search, multi-arm-bandit, knowledge bases, fairness in data collection.

Prerequisites: Students are expected to have the following background:

- Knowledge of basic computer science principles and skills, at a level sufficient to write a reasonably non-trivial computer program.
- Good knowledge of Python, Go, C/C++, or any programming language will be extremely helpful.
- Familiarity with basic probability theory, basic linear algebra, and algorithmic analysis.

Project: The project deliverables are: (1) weekly project reports; (2) a mid-way progress presentation; (3) a 6-page paper in https://www.acm.org/publications/proceedings-template format; (4) a presentation of the project during the last week of the semester; (5) code and plots.

Breakdown of Marks:

Participation	15%
Assignments	15%
Paper Presentation	10%
Project	60%

Tentative Schedule:

- Lecture 1 (1/12): nearest neighbor search problem and dimensionality reduction
- Lecture 2 (1/19): locality-sensitive Hashing
- Lecture 3 (1/24): paper reading
- Lecture 4 (1/26): inverted index and applications
- Lecture 3 (1/31): paper reading
- Lecture 5 (2/2): introduction to projects
- Lecture 6 (2/7): guest lecture
- Lecture 7 (2/9): randomized selection, two-point sampling
- Lecture 8 (2/14): the stable marriage problem, the coupon collector's problem
- Lecture 9 (2/16): project selection
- Lecture 10 (2/21): guest lecture
- Lecture 11 (2/23): multi-arm-bandit problem
- Lecture 12 (3/2): paper reading
- Lecture 13 (3/14): project discussion
- Lecture 14 (3/16): project discussion
- Lecture 15 (3/21): time-series and climate data analysis
- Lecture 16 (3/23): paper reading
- Lecture 17 (3/28): sampling algorithms and applications to approximate query answering
- Lecture 18 (3/30): project discussion
- Lecture 19 (4/4): project discussion
- Lecture 20 (4/6): sampling from streams
- Lecture 21 (4/11): core-set selection
- Lecture 22 (4/13): paper reading
- Lecture 22 (4/18): guest lecture
- Lecture 22 (4/20 & 4/25 & 4/27): project presentation

Policies:

- Students are responsible for understanding and upholding the university academic policy. Projects and assignments can be done in groups of size 3-5. Team members are expected equal effort in accomplishing a project. Incorporating existing tools or open source code in projects is allowed as long as students identify clearly the new work to be accomplished for a project.
- Late project and assignment submissions are not graded.
- In the event you encounter any barrier(s) to full participation in this course due to the impact of a disability please contact me or the Office of Disability Resources.