Intro To Computational Social Science

Tyler Horan

horat351@newschool.edu

Computer Science

Tuesdays/Thursdays 09:00-10:00

Course Description

With an increasing amount of data on every aspect of our daily activity—from what we buy, to where we travel, to whom we know—we are able to measure human behavior with precision largely thought impossible just a decade ago, creating an unprecedented opportunity to address longstanding questions in the social sciences. Leveraging this information, however, requires both scalable computational tools, and understanding how the substantive scientific questions should drive the data analysis. Lying at the intersection of applied mathematics, statistics, computer science, and the social sciences, the emerging field of computational social science fills this role, using large-scale demographic, behavioral and network data to investigate human activity and relationships. This class will introduce key tools and techniques of computational social science, consistently framing these methods in terms of motivating research questions. In addition to core mathematical concepts, the course will also address practical issues in working with real datasets, including familiarity with APIs, Unix tools, and statistical libraries to collect, manipulate, and analyze data.

Prerequisites

Linear algebra, Probability & Statistics. Previous exposure to a high-level programming language such as Python, MATLAB, Ruby, Perl, R or similar is recommended.

Course Materials

All course materials are available via the course website, which is accessible through your university portal. The assignments, handouts and lecture slides are available corresponding to each lecture on the syllabus.

Requirements & Expectations

You are required to attend and participate in class. There will be daily in-class quizzes, take-home assignments for discussion sections, and a final exam.

1. Lecture Attendance and Participation. We will start each lecture with a short quiz that you will take on your phone. You must be physically present in the room to take the quiz. Taking the quiz remotely is a form of academic dishonesty. For each question, you get half credit for answering incorrectly and full credit for answering correctly. You will also receive points for participating in classroom activities. After dropping the worst 10 percent of your responses, these points will account for 50 percent of your grade.

2. Discussion Section. You will do either readings, short assignments or both for discussion section each week. Any assignments must be turned in on time at section. Things happen so one missed assignment will not count against you. However, no late assignments will be accepted. Section will count for 30 percent of your grade in total, divided between attendance (10 percent), participation (10 percent), and assignments (10 percent).

3. Final Exam. The final will be held on Thursday, December 11, from 7:00pm to 10:00pm. If you cannot make this date and time, please drop the class. No alternate times or dates will be given. The final will count for 20 percent of your grade.

Policies

There are some University and classroom policies that you need to keep in mind.

1. Phones may only be used for designated class activities. I reserve the right to penalize your grade for misusing your phone in class.

2. Be honest. It is important that all students, and especially Freshers, understand what constitutes plagiarism in a written assignment or cheating on an exam. It is your responsibility to be aware of Duke's standards. Cheating or other violations of academic integrity will be dealt with according to University policy. You should bear in mind that, while the University may punish violations of the community standard as it sees fit, final grades for classes are assigned by the faculty member teaching the course, not the University. I do not tolerate cheating.

3. Inform me in advance of excused absences. By design, attendance and participation are the most vital components of this course. If you have a University excused reason to miss class, please follow the proper procedure for letting me (or your TA, or both if necessary) know about it. You are responsible for knowing and acting in accordance with University policy.

4. Sign up for office hours if you have a concern. Section is for weekly discussion and review of the course material. If you have other concerns, I hold office hours on Mondays from 1pm to 2:30pm when classes are in session.

Schedule

Lecture 1: Introduction: Case Studies in Computational Social Science
Readings:

    (a) Anatomy of the Long Tail: Ordinary People with Extraordinary Tastes, S. Goel, A. Broder, E. Gabrilovich & B. Pang.

    (b) Who Does What on the Web: A Large-scale Study of Browsing Behavior, S. Goel, J.M. Hofman & M.I. Sirer.

    (c) Learning from Seller Experiments in Online Markets, L. Einav, T. Kuchler, J. Levin & N. Sundaresan.

    (d) The Structure of Online Diffusion Networks, S. Goel, D.J. Watts & D.G. Goldstein.

Lecture 2: Counting: An Introduction

Lecture 3: Counting at Scale: MapReduce, Part I

Lecture 4: Counting at Scale: MapReduce, Part II

Lecture 5: Networks and Diffusion: Theory

Lecture 6: Networks and Diffusion: Case studies

Lecture 7: Complexity of Counting, Part I

Lecture 8: Complexity of Counting, Part II

Midterm Exam

Lecture 13: Data Wrangling

Lecture 14: Online Experiments

Lecture 15: Regression: Case Studies

Lecture 16: Regression: Theory and Practice

Lecture 17: Classification: Theory and Practice

Final Exam