Syllabus

Instructor: Jeff M. Phillips. | 3442 MEB | <http://www.cs.utah.edu/~jeffp>

Class Meetings: Mondays and Wednesdays, 1:25pm – 2:45pm, Zoom/YouTube.
Zoom: <https://utah.zoom.us/j/95074583347?pwd=RURrYUl4a2hpdmFEY2RxdzNhZ1c3dz09>
YouTube: <https://www.youtube.com/channel/UCDUS80bdunpmvWVPyFRPqFQ/live>

Course Web Page: <http://www.cs.utah.edu/~jeffp/teaching/FoDA.html>

This class will be an introduction to computational data analysis, focusing on the mathematical foundations, but providing some basic experience in analysis techniques. The goal will be to carefully develop and explore several core topics that form the backbone of modern data analysis topics, including Machine Learning, Data Mining, Artificial Intelligence, and Visualization. This will include some background in probability and linear algebra, and then various topics including Bayes Rule and its connection to inference, linear regression and its polynomial and high dimensional extensions, principal component analysis and dimensionality reduction, as well as classification and clustering. We will also focus on modern PAC (probably approximately correct) and cross-validation models for algorithm evaluation.

The course has the following **expected learning outcomes**:

- Students should have a basic understanding of the way modern data analysis goals are discussed, and be comfortable discussing them in the language of linear algebra and probability.

- Students should understand the most standard approaches within the broad topics covered: regression, clustering, dimensionality reduction, and classification.

- Students who start with decent programming skills, should be able to implement one or more algorithms within each topic.

- Students should be able to assess a supervised learning task using cross-validation.

The goal of this course is not to teach the state of the art, or the full array of techniques within these areas. Those goals are deferred to more advanced courses.

Book
The course text book will be Mathematical Foundations for Data Analysis by Jeff M. Phillips. A preprint is available here: <http://mathfordata.github.io>.

Lecture Availability and Procedure
Lectures will be given over Zoom. I will use this link (<https://utah.zoom.us/j/95074583347?pwd=RURrYUl4a2hpdmFEY2RxdzNhZ1c3dz09>) for all lectures. Students will be able to ask questions either through chat or voice channels – the lag here is minimal.

All students are recommended to turn on their video cameras during the lecture. This helps everyone stay engaged, and will help the instructor give a mode dynamic lecture. Virtual backgrounds (as long as they are not obscene) are perfectly fine.

Lectures will also be live-streamed to YouTube, available here (<https://www.youtube.com/channel/UCDUS80bdunpmvWVPyFRPqFQ/live>). This has a slight delay of about 15-20 seconds, and although you might be able to ask questions in the chat, it will not be monitored.

All lectures will also be archived on YouTube in the following playlist (<https://www.youtube.com/playlist?list=PLbuogVdPnkCqB1sx1eheVmLtp2EN7osYt>)

Getting Help

Take advantage of the instructor and TA office hours over Zoom (posted on course web page). We will work hard to be accessible to students. Please send us email if you need to meet outside of office hours. *Don't be shy if you don't understand something: come to office hours, send email, or ask/post questions in class!*

Students are encouraged to use a discussion group for additional questions outside of class and office hours. The class will rely on the Canvas discussion group. Feel free to post questions regarding any questions related to class: homeworks, schedule, material covered in class. Also feel free to answer questions, the instructors and TAs will also actively be answering questions. But, **do not post potential homework answers**. Such posts will be immediately removed, and not answered.

All important announcements will be made through the discussion group, there is otherwise no class mailing list.

Prerequisits

Some of the topics in this course are often taught as a "treat" at the end of a probability or linear algebra class, and then are often assumed knowledge in advanced data mining or machine learning classes. This class aims to fill these gap. This course will be a pre-requisite for advanced data analysis courses like Data Mining (CS 5140) and Machine Learning (CS5350).

We will begin setting prerequisites as CS 2100 (Discrete Math), CS 2420 (Algorithms & Data Structures), Math 2270 (Linear Algebra). And will set as a co-requisite CS 3130 (Probability and Statistics for Engineers) or Math 3070 (Applied Statistics 1). Exceptions can be made for equivalent experience.

We will show and discuss code examples in python. No prior knowledge is expected.

Grading

There will be one final exam worth 20% of the grade. There will be 6 or 7 quizzes, in total worth 20% of the grade. These will be given as canvas quizzes in times that overlaps with the end of class on that day, but with a generous time window. They are not meant to be challenging, rather they are meant to keep students on schedule for the course. Homeworks will be worth 60% of the grade. There will be 5 homeworks – the lowest one will be dropped.

The homeworks will usually consist of an analytical problems set, and sometimes light programming exercises in python. When python will be used, we typically will work through examples in class first.

Letter Grade Mapping: I will plan to map numerical grades to letter grades at the standard scale:

- 90-100 : A- to A
- 80-90 : B- to B+
- 70-80 : C- to C+
- 60-70 : D- to D+
- below 60 : E

The G- to G to G+ breakdown (for grade G = {A,B,C,D}) will probably align along:

- N0 to N3 : G-
- N3 - N7 : G
- N7 - N9.99 : G+

but I will reserve the right to shift this slightly.

Late Policy

To get full credit for an assignment, it must be turned in through Canvas by the start of class, specifically 1:10pm. Once the 1:10pm deadline is missed, those turned in late will lose 10%. Every subsequent 24 hours until it is turned another 10% is deducted. Assignments will not be accepted more than 48 hours late, and will be given a 0. This will consistently be enforced by if Canvas marks the assignment late (Canvas has a small buffer on the timing).

If you believe there is an error in grading (homeworks or quizzes), you may request a regrading within **one week** of receiving your grade. Requests must be made by email to instructor, explaining clearly why you think your solution is correct. You may consult with the instructor/TA first, but the formal request must always be made by email.

Collaboration Policy

For assignments, you may *discuss* solutions in general with anyone, including problem approach, proofs, and code. Students should not share specific numeric value of solutions with any other students or any other detail which may allow that other student to directly copy the answers. All students *must write their own* code, proofs, and write-ups.

For quizzes and the exams, you must work by yourself. Students discovered discussing with other students the contents of a quiz or test before they have both turned in their quiz will constitute cheating for both students.

Discussion threads, chat areas, and emails are all considered to be equivalent to the classroom, and your behavior in all these venues should conform to the university's student code.

School of Computing Cheating Policy

The School of Computing has instituted a two strikes and youre out cheating policy, meaning if you get caught cheating twice in any SoC classes, you will be unable to take any future SoC courses. https://handbook.cs.utah.edu

If a student is caught cheating on a homework or the project, they will receive a failing grade for the course. For a detailed description of the university policy on cheating, please see the University of Utah Student Code: http://www.regulations.utah.edu/academics/6-400.html.

Students with Disabilities

The University of Utah seeks to provide equal access to its programs, services, and activities for people with disabilities. If you need accommodations in this class, reasonable prior notice needs to be given to the Center for Disability Services, 162 Olpin Union Building, 581-5020 (V/TDD). CDS will work with you and the instructor to make arrangements for accommodations.

Safety and Addressing Sexual Misconduct

The University of Utah values the safety of all campus community members. Title IX makes it clear that violence and harassment based on sex and gender (which Includes sexual orientation and gender identity/expression) is a civil rights offense subject to the same kinds of accountability and the same kinds of support applied to offenses against other protected categories such as race, national origin, color, religion, age, status as a person with a disability, veterans status or genetic in-formation. If you or someone you know has been harassed or assaulted, you are encouraged to come speak to the School of Computing Advisors and/or to the Title IX Coordinator in the Office of Equal Opportunity and Affirmative Action, 135 Park Building, 801-581-8365, or the Office of the Dean of Students, 270 Union Building, 801-581-7066. For support and confidential consultation, contact the Center for Student Wellness, 426 SSB, 801-581-7776.

To report suspicious activity or to request a courtesy escort, call campus police at 801-585-COPS (801-585-2677). You will receive important emergency alerts and safety messages regarding campus safety via text message. For more information regarding safety and to view available training resources, including helpful videos, visit safeu.utah.edu. To report to the police, contact the Department of Public Safety, 801-585-2677(COPS). More information is available at `https://safeu.utah.edu`.

Latex

I recommend using LaTex for writing up homeworks. It is something that everyone should know for research and writing scientific documents. This linked directory (`http://www.cs.utah.edu/~jeffp/teaching/latex/`) contains a sample .tex file, as well as what its .pdf compiled outcome looks like. It also has a figure .pdf to show how to include figures.

Overleaf (`https://www.overleaf.com`) is a free online editor that makes all of this much easier.