Syllabus

Instructor: Jeff M. Phillips | 3404 MEB or 3470 MEB| http://www.cs.utah.edu/~jeffp

Class Meetings: Tuesdays and Thursdays, 10:45am – 12:05pm, WEB L104
Zoom: https://utah.zoom.us/j/98104516374?pwd=V2NiSDJZZ05GdU5hQnFPRXdOTktidz09
YouTube Live: https://www.youtube.com/watch?v=krlty7IcTWA
YouTube Playlist:
https://www.youtube.com/playlist?list=PLbuogVdPnkCpPYXq21ofChxEjiSzrJO79

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**. On completion of the course, students should be able to:

- To represent data points as vectors and data sets as matrices, and manipulate them with tool from linear algebra.

- To understand how to express a model to fit data as geometric object represented by a small number of parameters, with the goal of minimizing sum of squared errors, and motivated by probability assuming iid data.

- To understand basic formulations, models, and algorithms for goals in linear regression, dimensionality reduction, clustering, and classification.

- To be able to optimize a convex function with gradient descent, and how to apply these tools to optimize model parameters with respect to a cost function derived from data.

- To evaluate supervised learning problems (regression and classification), by how well they generalize to new data, with cross-validation.

The goal of this course is not to teach the rapidly-evolving state of the art, or the full array of techniques within these areas. Those goals are deferred to more advanced courses, and this course prepares students to succeed in those courses.

Book

The course text book will be Mathematical Foundations for Data Analysis by Jeff M. Phillips: http://mathfordata.github.io.

Lecture Availability and Procedure

Lectures will be delivered live and in WELB L104 on the University of Utah campus.

To provide an option for students advised to or deciding to quarantine or social distance, lectures will also be simulcast over Zoom, using this link (https://utah.zoom.us/j/98104516374?pwd=V2NiSDJZZ05GdU5hQnFPRXdOTktidz09) for all lectures. Students will be able to ask questions either through chat or voice.

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.

Pre-requisites

Some of the topics in this course are often (but not always) 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 pre-requisites 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, or in the undergraduate Data Science Certificate program.

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

The COMP 5960 listing, meant for non-CS graduate students and non-matriculated students, does not have pre-requisites.

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 as long as an attempt is made for each assignment.

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 10 minutes before the end of the day it is due, specifically 11:50pm. Once the 11:50pm 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. A strike occurs when you are reported for a major cheating (leading to failing a course), or two comparatively minor cheating instances. If you accumulate two strikes in any SoC courses, you will be unable to register for any future SoC courses.
https://www.cs.utah.edu/docs/misc/cheating_policy.pdf

If a student is caught cheating on a homework or quiz, they will at the least fail that homework or quiz and receive a minor sanction. For repeated, systematic, or ostentatious cheating the student will receive a failing grade and a major sanction. *Please do not cheat, it is not worth it*; the consequences are harsh and it is unfair to your peers. If feel you are falling behind on the material, reach out to the instructor, and he will work with you towards a solution.

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 Support and Inclusion

Students with Disabilities or with English as a non-First Language

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. CDS will work with you and the instructor to make arrangements for accommodations. All written information in this course can be made available in an alternative format with prior notification to the Center for Disability Services.

Extra support if also available for those for whom English is not their first language. Several resources on campus will support you with your language and writing development. These resources include: the Writing Center (http://writingcenter.utah.edu/); the Writing Program (http://writing-program.utah.edu/); the English Language Institute (http://continue.utah.edu/eli/). Contact the instructor if there is any additional support that would aid in this course.

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 https://safeu.utah.edu.

Student Names and Personal Pronouns

Class rosters are provided to the instructor with the students legal name as well as Preferred first name (if previously entered by you in the Student Profile section of your CIS account, which managed can be managed at any time). While CIS refers to this as merely a preference, I will honor you by referring to you with the name and pronoun that feels best for you in class or on assignments. Please advise me of any name or pronoun changes so I can help create a learning environment in which you, your name, and your pronoun are respected. If you need any assistance or support, please reach out to the LGBT Resource Center. https://lgbt.utah.edu/campus/faculty_resources.php

Diversity and Inclusivity

It is my intent that students from all diverse backgrounds and perspectives be well served by this course, that students learning needs be addressed both in and out of class, and that the diversity that students bring to this class be viewed as a resource, strength and benefit. It is my intent to present materials and activities that are respectful of diversity: gender, sexuality, disability, age, socioeconomic status, ethnicity, race, and culture. Your suggestions are encouraged and appreciated. Please let me know ways to improve the effectiveness of the course for you personally or for other students or student groups. In addition, if any of our class meetings conflict with your religious events, please let me know so that we can make arrangements for you.

Immigration is a complex phenomenon with broad impactthose who are directly affected by it, as well as those who are indirectly affected by their relationships with family members, friends, and loved ones. If your immigration status presents obstacles to engaging in specific activities or fulfilling specific course criteria, confidential arrangements may be requested from the Dream Center. Arrangements with the Dream Center will not jeopardize your student status, your financial aid, or any other part of your residence. The Dream Center offers a wide range of resources to support undocumented students (with and without DACA) as well as students from mixed-status families. To learn more, please contact the Dream Center at 801.213.3697 or `visitdream.utah.edu`.

Wellness

Personal concerns such as stress, anxiety, relationship difficulties, depression, cultural differences, etc., can interfere with a students ability to succeed cross and thrive at the University of Utah. For helpful resources contact the Center for Student Wellness at `www.wellness.utah.edu` or 801-581-7776.

If you are a student veteran, the U of Utah has a Veterans Support Center located in Room 161 in the Olpin Union Building. Hours: M-F 8-5pm. Please visit their website for more information about what support they offer, a list of ongoing events and links to outside resources: `http://veteranscenter.utah.edu/`.

Contact the instructor if there is any additional support that would aid in this course.

COVID, Masks, and Vaccines

As a health and safety measure, I recommend all students to get a COVID vaccine. For those concerned about your own infection or that of others, N95 (or similar) masks are recommended, but not required. If you are contagious with COVID (or anything else), it is recommended not to come to in-person aspects of class. Reach out to the instructor for other ways to stay up to date with class, if the answer is not already apparent. This course will follow university guidelines and recommendations as much as possible.

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.