Syllabus

Course: High-Dimensional Data Analysis (CS 5966 & 6966)

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

Class Meetings: Mondays and Wednesdays, 3pm – 4:20pm, AEB 310

Course Web Page: http://www.cs.utah.edu/~jeffp/teaching/HDDA.html

In modern data analysis, it is common to deal with data represented as high-dimensional vectors. These have become part of the default featurized or "learned" representation for large language models, images, audio, and just about any other structured data one wants to learn from. This course will cover the curse of dimensionality, as well as various methods to partially overcome it, including dimensionality reduction, margin-based approaches, locality sensitive hashing, and no-dimensional coresets. The focus will mainly be on methods with mathematically rigorous proofs, including analysis of correctness and efficiency; but keeping to methods that work in practice, and showing how to use these approaches on data.

The goal for the course is that students understand the real challenges of high-dimensional space, and methods that can be effectively applied in those settings. This understanding will come from both mathematical explanations and empirical observations. The focus will be on simple methods that work, and how to use those insights more generally.

Book

We will not follow a text book for this course. Resources (including open access) textbooks will be linked to from the course webpage.

Lecture Availability and Procedure

Lectures will be delivered live and in AEB 310 on the University of Utah campus.

Getting Help

Take advantage of the instructor and (TA) office hours. 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

For undergraduate students, prerequisites are CS 3190 and CS 4150.

For representing and working with high dimensions, some linear algebra is a necessity. The instructor will use geometric perspectives to explain phenomenons when possible. Some probabilistic arguments will be

necessary, especially concentration of measure bounds. While the ideas will conceptually require synthesis of advanced ideas, the aim will be explanations from first principals as much as possible.

Coding tasks will be agnostic to language, but convenient vector process (e.g., matlab or numpy in python) will be an essential component.

Grading

Quizzes will be worth 20% of the grade. There will be about 7. The lowest score dropped.
Assignments will be worth 30% of the grade. There will be 4.
A Project will be worth 30% of the grade.
The Final Exam worth 20% of the grade. It will be like the quizzes.

The homeworks will usually consist of an analytical problems set, and sometimes light programming exercises in a programming language of your choice.
The quizzes should take about 10 minutes each, and should be simple if you are attending and following the lectures.

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. Students caught cheating will get a 0 and will not be able to drop a low score on a quiz.

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/undergraduate/current-students/policy-statement-on-academi

If a student is caught cheating on a homework or quiz, they will at the least fail that homework or quiz (and not be able to drop) 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.

Welcoming Classroom Atmosphere

It is my intent that students from all backgrounds and perspectives be well served by this course, that students learning needs be addressed both in and out of class, and that the different views 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 these perspectives and experiences.

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.

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.