

SDS 385

Stat models for Big data

Fall 2021

TA : Rimli Sengupta

Professor
Office Hours
Office
Email

Purnamrita Sarkar
TBA via Zoom
Welch Hall 5.228B
`purna.sarkar@austin.utexas.edu`

Syllabus

Course Delivery

All classes will be held via Zoom until August 17th. Look at <https://zoom.its.utexas.edu/> for more information. Few things to remember, when you join you will be muted, and you will have to unmute yourself to talk. You can also use the raise hand feature to ask a question. In each class, there will be a preassigned moderator(s) who will keep an eye on who is raising hands/asking questions on chat, so that we have tons of communication.

Course Description

This course provides an introduction to large scale machine learning. The first part is focused on optimization and supervised learning, and the second part deals with algorithms for unsupervised learning.

- Optimization algorithms and analysis
 - Stochastic gradient descent
 - Coordinate ascent, and others
- Divide and conquer: distributed algorithms
- Sampling based algorithms/ randomized algorithms
 - Approximating matrix multiplication by row/column sampling
 - Johnson Lindenstrauss Lemma and related fast projection methods
 - Hashing and sketching (also extremely useful for nearest neighbor calculation)
- Sampling based methods for uncertainty estimation
 - Bootstrap
 - Subsampling
- Ranking methods
 - Random walk based methods like Pagerank

- Semisupervised learning
- Manifold regularization

Prerequisites Students are expected to have a good familiarity with Linear algebra, Calculus and undergraduate probability.

Textbook

This course is designed to be self-contained, and there is no required textbook. We will read a lot of papers.

Course website

<https://psarkar.github.io/sds385.html>

Evaluation

- 3-4 homeworks (30%),
 - Homeworks are due (every 2-3 weeks) before midnight on Sundays.
 - All homeworks need to be typed up in latex.
- weekly paper presentation(40%),
 - I will make a list of papers. You need to do the presentations in groups of 2.
 - Wednesdays we will have presentations from a group of 2 students for 40-50 minutes, where each should present a significant part of the paper. The presentation is worth 15% of the grade.
 - The details for how these classes will be carried out (online/ record before to ensure social distancing) will be announced later.
 - After the presentation, there will be a breakout session discussing the paper. This is 5% of the grade
 - By midnight each Wednesday, all of you should submit a one page review of the paper in question. The best way to do this would be to **read the paper before the class and start working on the review**. This will be 20% of the grade.
 - The review should be single-spaced and one page. Very similar to a NeurIPS review, I will look for three parts:
 - * Summarize the paper,
 - * Discuss the contribution (e.g., theoretical, methodological, algorithmic, empirical contributions). For each contribution, briefly state the level of significance (i.e., how much impact will this work have on researchers and practitioners in the future?). If you cannot think of three things, please explain why. Not all good papers will have three contributions),

* Discuss the strengths and weaknesses.

- Take home exam (30%)

Homework will be assigned biweekly and due via canvas..

Exam There will be one take home final exam.

Requests for Regrade: Clerical requests will be corrected without hassle. Other regrading requests must be submitted in writing within 2 days of the assignment/exam return. Be aware that the entire assignment/exam will be subject to regrading, and grades may go up or down.

Students with Disabilities

Students with disabilities may request appropriate academic accommodations from the Division of Diversity and Community Engagement, Services for Students with Disabilities, 512-471-6259, <http://www.utexas.edu/diversity/ddce/ssd/>.

Religious Holidays

By UT Austin policy, you must notify me of your pending absence at least fourteen days prior to the date of observance of a religious holy day. If you must miss a class, an examination, a work assignment, or a project in order to observe a religious holy day, you will be given an opportunity to complete the missed work within a reasonable time after the absence.

Scholastic Honesty

We expect students to behave with integrity. Students found Cheating on exams or homeworks will receive a score of zero for that exam or assignment, and may be subject to additional disciplinary action. For more information on the University of Texas scholastic dishonesty policy, see the 2006-2007 General Information Catalog, Appendix C.

Campus Safety

Please note the following recommendations regarding emergency evacuation from the Office of Campus Safety and Security, 512-471-5767, <http://www.utexas.edu/safety>:

- Occupants of buildings on The University of Texas at Austin campus are required to evacuate buildings when a fire alarm is activated. Alarm activation or announcement requires exiting and assembling outside.
- Familiarize yourself with all exit doors of each classroom and building you may occupy. Remember that the nearest exit door may not be the one you used when entering the building.
- Students requiring assistance in evacuation should inform the instructor in writing during the first week of class.
- In the event of an evacuation, follow the instruction of faculty or class instructors.
- Do not re-enter a building unless given instructions by the following: Austin Fire Department, The University of Texas at Austin Police Department, or Fire Prevention Services office.
- Behavior Concerns Advice Line (BCAL): 512-232-5050

- Further information regarding emergency evacuation routes and emergency procedures can be found at: <http://www.utexas.edu/emergency>.

SPECIAL NOTE: Please be aware that SB 11, concerning the concealed carrying of handguns on campus, does not take effect until August 2016. The possession of a firearm, illegal knife, or prohibited weapon on the grounds of an educational institution is currently a third-degree felony in Texas.

Class Recordings: Class recordings are reserved only for students in this class for educational purposes and are protected under FERPA. The recordings should not be shared outside the class in any form. Violation of this restriction by a student could lead to Student Misconduct proceedings.

Sharing of Course Materials is Prohibited: No materials used in this class, including, but not limited to, lecture hand-outs, videos, assessments (quizzes, exams, papers, projects, homework assignments), in-class materials, review sheets, and additional problem sets, may be shared online or with anyone outside of the class unless you have my explicit, written permission. Unauthorized sharing of materials promotes cheating. It is a violation of the University's Student Honor Code and an act of academic dishonesty. I am well aware of the sites used for sharing materials, and any materials found online that are associated with you, or any suspected unauthorized sharing of materials, will be reported to Student Conduct and Academic Integrity in the Office of the Dean of Students. These reports can result in sanctions, including failure in the course.