Second Semester Area B (Data Science)

Foundational Techniques in Machine Learning and Data Science

CSE 382M / M 393C / CS 395T

The University of Texas at Austin, Spring 2024

Times: MonWed 2-3:30PM CT
Location: Face-to-face in GDC 6.202
Dates: 28 classes, Wed Jan 17 – Mon Apr 29, with Mon Mar 11 and Wed Mar 13 off
Supporting Media: Canvas
Registrar Information: Unique 62230 / 54100 / 51410

Instructor: Joe Kileel, jkileel@math.utexas.edu
Office Hours: Tue 1-3PM CT in POB 3.434

Teaching Assistant: Yifan Zhang, yf.zhang@utexas.edu
Office Hours: Fri 2:30-4:30PM in EER 7.650

Description: This prelim course will be a mathematically rigorous introduction to topics from linear algebra, high-dimensional probability, optimization and statistics, which are foundational tools for data science, or the science of making predictions from structured data. A secondary aim of the course is for students to gain experience in exploring data science problems through computer programming.

Prerequisites: Undergraduate courses in probability and proof-based linear algebra. Basic computer programming for matrix computations and plotting data.

Textbook: The course has no official textbook. Material will be posted to Canvas. Parts of the course are based on *Foundations of Data Science* by Avrim Blum, John Hopcroft, Ravi Kannan (the preliminary version online or the published version).

Programming Language: You may use Python, Matlab, C, C++ or R. You may also use other languages if you receive the teaching assistant's permission.

Homework: Homework will be assigned roughly biweekly, for a total of ~5 assignments. These will test theory and basic programming. It is permitted for students to work together in groups. However each student must write up their own work individually. Homework will be posted and submitted via Gradescope on Canvas.

Midterm: One 75-minute midterm will be held in class, tentatively on March 6.

Final Project: Students will work in groups of two on final projects. Projects may apply different data science methods to real data sets and then interpret the results. Or they may explore theory beyond the class material and then provide numerical simulations. Students will submit a half-page proposal pitching their project idea to

the instructor. Later you will present your projects using slides, tentatively during the last four classes of the semester. You must also turn in a PDF copy of your slides.

Final Grade: In determining final letter grades on the plus/minus scale, homework will count for 40%, the midterm will count for 30%, and the final project will count for 30%. Letter grade cut-offs will be finalized at the end of the semester; however they will not be stricter than the following: A (100-93), A- (92-89), B+ (88-85), B (84-81), B- (80-77), C+ (76-73), C (72-69), C- (68-65), D (64-55), F (54-0). Throughout the semester, please check that your scores are recorded correctly on Canvas.

Tentative Course Plan:

1. High Dimensions (~3 classes)
Probability Review
Concentration of Measure
Curses and Blessings

2. Random Projections (~2 classes)
Johnson–Lindenstrauss lemma
Fast Pseudo-Random Projections

3. Linear Algebra (~3 classes)
Singular Value Decomposition
Principal Component Analysis
Power Method

4. Clustering (~2 classes)
k-means and Lloyd's Algorithm
Spectral Clustering

5. Classification (~2 classes)
Support Vector Machine and Kernel Trick
Neural Networks

6. Regression (~2 classes)
Linear, Polynomial, Kernel Ridge Regression
Neural Networks

7. Optimization (~3 classes)
Gradient Descent
Stochastic Gradient Descent

8. Sampling (~3 classes)
Markov Chains
Markov Chain Monte Carlo

9. Randomized Linear Algebra (~3 classes)
Subspace Random Projection Theorem
Randomized Sketching
Randomized Singular Value Decomposition

10. Miscellany (time permitting)
Compressed Sensing
Tensor Decompositions
Manifold Optimization

Accommodations: The University of Texas provides, upon request, academic accommodations for qualified students with disabilities. For more information, contact Disability and Access at 512-471-6259 or access@austin.utexas.edu

Student Honor Code: "I pledge, as a member of the University of Texas community, to do my work honestly, respectfully, and through the intentional pursuit of learning and scholarship."