

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 2026

Times: MonWed 2-3:30PM CT

Location: Face-to-face in GDC 6.202

Dates: 28 classes, Mon Jan 12 – Mon Apr 27, with Mon Jan 19, Mon Mar 16 and Wed Mar 18 off

Supporting Media: Canvas

Registrar Information: Unique 65820 / 57125 / 53649

Instructor: Joe Kileel, jkileel@math.utexas.edu

Office Hours: Tue 1-3PM CT in POB 3.434 (starting Jan 20)

Teaching Assistant: Gabriel Brown, ghbrown@utexas.edu

Office Hours: Thu 10AM-12PM CT in MBB 2.304 (starting Jan 15)

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 5-6 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 9.

Final Exam: There will be a cumulative final exam, held on April 30, May 1-2 or 4.

Late Homework or Missed Exam Policy: Per university rules, exceptions are permitted for the observation of religious holidays or military obligations. Other requests

will be handled by the instructor on a case-by-case basis.

Final Grade: In determining final letter grades on the plus/minus scale, homework will count for 40%, the midterm will count for 25%, and the final exam will count for 35%. Letter grade cut-offs will be finalized at the end of the semester; however they will not be stricter than the following: A (100-90), A- (89-85), B+ (84-80), B (79-75), B- (74-70), C+ (69-65), C (64-60), C- (59-55), D (54-50), F (49-0).

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 (~2 classes)

Singular Value Decomposition
Principal Component Analysis
Power Method

4. Clustering (~2 classes)

k -means and Lloyd's Algorithm
Spectral Clustering

5. Classification (~3 classes)

Support Vector Machine
Kernel Trick
Nearest Neighbors

6. Regression (~2 classes)

Linear and Polynomial Regression
Kernel Ridge Regression

7. Neural Networks (~4 classes)

Architectures
Universal Approximation
Backpropagation
Applications

8. Optimization (~3 classes)

Gradient Descent
Stochastic Gradient Descent

9. Sampling (~3 classes)

Markov Chains
Markov Chain Monte Carlo
Neural Network-Based Sampling

10. Randomized Linear Algebra (~2 classes)

Subspace Random Projection Theorem
Randomized Singular Value Decomposition

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

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."