CAS CS365 Foundations of Data Science
Spring 2026 Syllabus

Official course description:
Foundations of Data Science examines the core theoretical principles and techniques that underpin modern machine learning and data analysis. The course covers foundational topics in probability, statistics, linear algebra, and optimization, together with essential concepts in statistical learning. Emphasis is placed on developing mathematical intuition and practical problem-solving skills, and on applying computational tools to model, analyze, and interpret real-world datasets effectively.

Prerequisites:
students must have completed **CS112**, **CS131** (MA293) , **CS132** (MA242) and **CS237** (MA581). CS330 is not required but recommended.

Course staff:

Prof. Dora Erdos edori@bu.edu office: CDS 910
TF Patrick Lutz plutz@bu.edu

Class Tools (please sign up asap!):
Piazza: https://piazza.com/bu/spring2026/cascs365
Google Drive folder: https://tinyurl.com/4p5w88hn
Gradescope: https://www.gradescope.com/courses/1232203 entry code: **N2R8XR**

Lectures: Tue/Thu 3:30-4:45 pm SHA 110
Labs: Fridays B1: 1:25-2:15 pm BRB 122, B2: 2:30 - 3:20 PSY B35, B3: 3:35 - 4:25 MCS B33

Reading:
There is no assigned textbook for this course. Reading material will be provided in the course resources throughout the semester.

Course Communication:
Our primary form of communication will be through **Piazza**. Please post any class related question there - questions about course content, assignments, logistics, etc. If you know the answer to a fellow student's question, please don't hesitate to respond! It's much more beneficial to get an explanation from your peers - who have the same background and understanding as you - to the instructors. The course staff will be at hand too to respond to questions as needed. You are welcome to ask questions about the homework, but please don't post solutions. If you have a question specific to your solution, you are welcome to ask in a private post. If you have something private to discuss, you can also do so through private posts.

We are here to support you in your learning journey. Please don't hesitate to reach out if you are facing any kind of difficulty. We want you to succeed and will do everything we can to help you.

Grading:

- **30%** weekly homework assignments
- **30%** midterm exam (in class, **Thu, March 5, 2026**)
- **40%** final exam

Exams are closed book and no collaboration is permitted.

Homework:
There will be two types of homework assignments, paper-and-pencil style theoretical problems and data analysis.

The theoretical problems will consist of questions related to the concepts covered in lecture. You can expect to write mathematical proof-style solutions or perform computations. These problems are designed to be challenging. The material covered in labs will help you get started and/or make progress on them.

The data analysis part will consist of applying various algorithms to data or getting a deeper understanding of the theoretical concepts through observing its behavior when applied to data. In general you will not be expected to submit the code that you used. Rather, you will submit a pdf document answering questions about your findings on the data. Most likely your answer will include numerical outcomes and plots produced by your implementation, along with an explanation on what we can conclude based on the experimental results. You will get a more precise description of the expected format in class.

Assignments will be **due every week** on **Mondays**, roughly alternating between theory and implementation. You won't be able to make up missed assignments, however your **lowest theory assignment grade and your lowest implementation grade will be dropped**.

Late policy: You can submit up to 24 hours late for a late **penalty of 10%** of the grade. Assignments submitted more than 24 hours late will not be accepted.

Submission: Homework assignments have to be typed, hand-written assignments will not be accepted. You are allowed to use any text editor you like, but we highly recommend using LaTeX as the assignments will be heavy on mathematical formulas.

Collaboration and AI policy:

Collaboration: You are allowed - and encouraged! - to work together on assignments with your classmates. However, the answer you submit has to be **written by yourself using your own words**. This means that your text has to be different from your classmates' and should not be a product of AI. You can collaborate with up to three classmates, but you have to clearly list your collaborators on your assignment. If you use information from outside the material provided in the context of class (lecture and lab material, readings, Piazza, help during oh) you have to cite it.

AI policy: For the *theory assignments* you are **not allowed** to use AI to solve the problem. You can use AI to ask for explanation on concepts in the class material related to the assignment. But you cannot use any prompt that is derived from the text of the assignment.

For the *data analysis assignments* you are **free to implement the code any way you like**. This means you have permission to use off-the-shelf code packages or AI generated code. However, you have to actually **run the experiments yourself** and the **written analysis** of the outcome **has to be your own work** based on your ideas.

Note that, since most algorithms have a random component and/or often you will be asked to tweak your input data, it is highly unlikely that your classmate and you would end up with the exact same experimental results. This means that your outcome should look different. It's okay to compare your results and brainstorm on what conclusion you might draw from it, but then you have to write on your own how this applies to your results.

Academic Honesty:

Academic standards and the code of academic conduct are taken very seriously by the University, the College of Arts and Sciences, and the Department of Computer Science. Course participants must adhere to the CAS Academic Conduct Code. Please take the time to review this document if you are unfamiliar with its contents.
Any deviation from the Exam, Collaboration and AI policy - as laid out in this Syllabus - will be considered a violation of the Academic Conduct Policy. Any case of academic misconduct will be reported to the College and will also carry a grading penalty

Tentative schedule:

week	date	topic
1	Mon 1/19	
	Tue 1/20	intro, basic prob, conditional
	Thu 1/22	rnd vars, stats, pmf, pdf, continuous distributions, some common distributions
2	Mon 1/26	
	Tue 1/27	Gaussian and other mixture models (1D?), why focus so much on Gaussians? --> LLN, CLT, what does it mean to learn a distribution, fitting distributions - Likelihood, MLE,
	Thu 1/29	multi-dim Gaussian micture models, covariance mtx, MLE for multiple dim - mtx calculus (basic), EM intro
3	Mon 2/2	hw1 due (theory)
	Tue 2/3	soft clustering, EM (many examples? or more math)
	Thu 2/5	kNN, distances, Johnson-Lindenstrauss
4	Mon 2/9	hw2 due (theory)
	Tue 2/10	Lin algebra review, lin projections and implementing J-L
	Thu 2/12	SVD, PCA
5	Mon 2/16	hw3 due (analysis)
	Tue 2/17	Monday schedule
	Thu 2/19	computing SVD - mtx multiplication, master method
6 Last Day to Drop without W	Mon 2/23	hw 4 due (theory)
	Tue 2/24	low-rank approx using SVD and mtx decompositions (only touch on it)

	Thu 2/26	walks on graphs, Eigenvector centrality, PageRank - intro to stationary distributions
7	Mon 3/2	hw 5 due (analysis)
	Tue 3/3	Markov chains, Metropolis-Hastings sampling
	Thu 3/5	midterm
8	Wed 3/11	Spring recess
	Tue 3/10	
	Thu 3/12	
9	Mon 3/23	
	Tue 3/24	linear regression - least squares, technical computation (calc)
	Thu 3/26	fitting lines to data - general linear model. least squares fitting, measuring fit of regression
10	Mon 3/30	hw 6 due (analysis)
	Tue 3/31	regularization, penalty factors, approximate LS
Fri 4/3 last day to drop with W	Thu 4/2	logistic regression
11	Mon 4/6	hw 7 due (theory)
	Tue 4/7	Perceptron, SVM
	Thu 4/9	gradient descent, calc
12	Mon 4/13	hw 8 (theory)
	Tue 4/14	SGD
	Thu 4/16	more on GD, SGD
13	Mon 4/20	hw 9 (analysis)
	Tue 4/21	ML methods using our toolkit
	Thu 4/23	ML methods using our toolkit
14	Mon 4/27	hw 10 (analysis)
	Tue 4/28	ML methods using our toolkit
	Thu 4/30	voting