

CDS DS 120 B

Foundations of Data Science I

Spring 2023

Lecture Meeting Place: School of Hospitality (SHA) 110

Meeting Time: MWF 2:30 pm – 3:20 pm

Discussion Meeting Place: B2-B4: Cntr for Comp & Data Sciences (CDS) 164
B5: College of Arts and Sciences (CAS) B27

Meeting Times: B2: T 9:30 pm – 10:45 pm, B3: T 11:15 am – 12:05 pm,
B4: T 12:30 pm – 1:45 pm, B5: T 2:00 pm – 3:15 pm

Instructor: Prof. Lisa Wobbes

- **Office Hours:** See calendar on Piazza
- **Office Hours Location:** CCDS 1401
- **Email:** edwobbes@bu.edu

Teaching Assistants: Ellis Hamilton, Changxuan Fan

- **Office Hours:** See calendar on Piazza
- **Office Hours Location:** See calendar on Piazza
- **Email:** ellisham@bu.edu, fanchx@bu.edu

Course Assistants: Cooper Hassman, Ava Yip, Eline Leske

- **Office Hours:** See calendar on Piazza
- **Office Hours Location:** See calendar on Piazza
- **Email:** chassman@bu.edu, avayip@bu.edu, ejleske@bu.edu

Overview of the Course

CDS DS 120 is the first in a three-course sequence (including CDS DS 121 and CDS DS 122) that introduces students to theoretical foundations of data science. The three courses take an iterative (spiral) approach to a set of foundational material in calculus, linear algebra, and probability. DS 120 provides an introduction to key concepts from logic (mathematical induction, truth values, logical equivalence), probability and statistics (combinatorial analysis, random variables, central limit theorem), linear algebra (solving linear systems, inverses, determinants, vector spaces, linear regression), and calculus (higher-order derivatives, gradients,

Riemann sum, integrals). The course is aimed at building the foundations of mathematical thinking, specifically through the use of problem sets that require students to answer mathematically-posed questions.

Learning Objectives

CDS DS 120 consists of four modules, each with its own learning objectives.

- Module 1: Logic.
Ability to express yourself mathematically, providing basic mathematical proofs.
- Module 2: Probability and Statistics.
Understanding of fundamental principles and concepts of probability theory and statistics, ability to apply these concepts.
- Module 3: Linear Algebra.
Developing computational skills and intuition for linear algebra, ability to solve small-scale data science problems using linear algebra tools.
- Module 4: Calculus.
Developing computational skills for differential calculus and integral calculus required for data science.

Hub Learning Outcomes

Quantitative Reasoning I (QR1)

Getting Set Up

You will need to set up access to the following online materials. Instructions for how to do all of those setups are below.

Required: Piazza for discussion of assignments and all course material, except lecture handouts.

Required: Gradescope for assignment submission.

Required: Blackboard for lecture handouts.

Piazza

We will be using Piazza for class discussion. The system is well tuned to getting you help fast and efficiently from classmates, and instructors. Rather than emailing questions to the teaching staff, please post your questions on Piazza. We will also use Piazza for distributing materials such as homeworks and helpful resources.

When someone posts a question on Piazza, if you know the answer, please go ahead and post it. However, please *don't* provide answers to homework questions on Piazza. It's OK to tell people *where to look* to get answers, or to correct mistakes; just don't provide actual solutions to homeworks.

Setup: Our class Piazza page is at <https://piazza.com/bu/spring2024/ds120b/info>. If you registered before the semester started, you should have been automatically enrolled. If you are adding the class late, go to Piazza at that link and enroll yourself. If you have any problems, please contact a TA or a CA.

Gradescope

Assignments will be submitted via Gradescope (<https://www.gradescope.com/>). Graded assignments will be returned to you via Gradescope as well.

Setup: If you registered before the semester started, you should have been automatically enrolled. If you are adding the class late, go to Gradescope at the link above, and enroll yourself using the entry code ZW57X5. If you have any problems, please contact a TA or a CA.

Homeworks

1. There will be 10 homework assignments. The top 9 grades will be used to compute the total homework credit.
2. Homeworks will be assigned on Wednesdays.
3. Homeworks are due at 09:05 am on the following Wednesdays.
4. You can discuss homeworks in section meeting on Tuesdays. But don't expect that TAs will be going into detail – instead, they will answer specific questions!
5. You need to consistently work the problem sets each week. Plan to set aside a regular time each week to do them.
6. Most homeworks will contain bonus questions. You must earn full participation credit to activate the bonus points.
7. Homeworks will be submitted via *Gradescope*. See the next section.
8. **IMPORTANT:** You are responsible for submitting homework electronically on Gradescope by the stated due date and time. Assignments will be accepted on Gradescope up to 24 hours late for a 20% grade reduction. Later assignments will not be accepted.

Submitting Homework

There are three options for homework submission, in increasing order of quality:

1. You can scan handwritten notes into PDF. Note that these must be **clear** and **neat** because the grader will simply read them as best they can – if the grader cannot understand your handwriting easily, you may lose points on the assignment. If you use this option, you can scan from your mobile device if it comes out clearly enough. There are instructions on Piazza for how to scan and submit your homework via Gradescope.

2. You can write up your work in Word, using the built-in equation editor for the mathematics. Then save as PDF, and follow the same instructions for how to submit to Gradescope. Added benefit: no trees are destroyed.
3. You can learn and use \LaTeX . This is the tool that produces a professional, publishable PDF document. It is what hardcore computer scientists use. You can learn to use it quickly – I recommend starting with the cloud based system called Overleaf at

https://www.overleaf.com/learn/latex/Learn_LaTeX_in_30_minutes/.

If you want to install \LaTeX on your own computer (to use offline, for example) there are instructions at <http://www.latex-tutorial.com/>. Eventually you will find it useful for lots of your coursework (\LaTeX is required for DS 320!), so it makes sense to learn it now.

Course and Grading Administration

Final grades will be computed based on the following:

50% Homework assignments.

5% Participation.

10% Midterm 1.

10% Midterm 2.

25% Final (Cumulative).

Participation

To earn full participation credit, you need to attend at least 11 out of 14 discussion labs. However, up to 2 out of 11 discussion labs can be replaced by active participation on Piazza, in class, or by actively helping other students in office hours.

Final

I will determine grade cutoffs after all assignments and exams have been graded. Grade cutoffs will take into account my assessment of the difficulty level of the assignments and exams, and my assessment of what is expected for each letter grade.

Regrade policy

You have the right to request a regrade of any homework or exam question. All regrade requests must be submitted via Gradescope, and must describe a factual error in our assessment. If you request a regrade for one question, then we have the right to review the entire homework or exam. Beware that this may potentially result in a lower grade.

Important Dates

- Midterm 1: March 4.
- Midterm 2: April 17.

The final exam will take place between May 6 and May 10. Mark this week in your calendars!

Readings

You are expected to actively use the lecture handouts throughout this course. Each topic in the course schedule has a corresponding handout.

Office Hours

There are 8+ office hours each week. The schedule for office hours is on Piazza.

Accommodations for Students with Documented Disabilities

If you are a student with a disability or believe you might have a disability that requires accommodations, please contact the Office for Disability Services (ODS) at (617) 353-3658 or access@bu.edu to coordinate any reasonable accommodation requests. ODS is located at 25 Buick Street on the 3rd floor.

Course Feedback

Your feedback about any material you would like to cover or suggestions for how to improve the course is welcome at any time. All suggestions will be considered, though it cannot be guaranteed that the feedback will be incorporated. Please send your comments as a private Piazza note to Prof. Lisa Wobbes.

Academic Honesty

You may discuss homework assignments with classmates, but you are solely responsible for what you turn in. Collaboration in the form of discussion is allowed, but all forms of cheating (copying parts of a classmate's assignment, plagiarism from books or old posted solutions) are NOT allowed. We – both teaching staff and students – are expected to abide by the guidelines and rules of the Academic Code of Conduct (which is at <http://www.bu.edu/dos/policies/student-responsibilities/>).

You can probably, if you try hard enough, find solutions for homework problems online. Given the nature of the Internet, this is inevitable. Let me make a couple of comments about that:

1. If you are looking online for an answer because you don't know how to start thinking about a problem, talk to a TA, a CA, or myself, who may be able to give you pointers to get you started. Piazza is great for this – you can usually get an answer in an hour if not a few minutes.

2. If you are looking online for an answer because you want to see if your solution is correct, ask yourself if there is some way to verify the solution yourself. Usually, there is. You will understand what you have done *much* better if you do that.
3. If you are looking online for an answer because you don't have enough time and are getting close to the assignment deadline, think about this:
 - (a) what you are doing is intellectually dishonest,
 - (b) you are going to have to solve problems like this on the midterm and final,
 - (c) you can drop one homework without penalty,
 - (d) you can compensate for the lost points by submitting bonus questions.

So ... it would be better to simply submit what you have at the deadline (without going online to cheat) and plan to allocate more time for homeworks in the future. We care more about making an honest attempt on the homework than the final solution being exactly right.

Course Schedule

Date	Topic	Assigned	Due
01/19 (F)	Introduction		
Module 1			
01/22 (M)	Mathematical proofs		
01/24 (W)	Sets	HW 1	
01/26 (F)	Proofs by induction		
01/29 (M)	Mathematical statements		
01/31 (W)	Logical negation of quantified statements	HW 2	HW 1
02/02 (F)	Logical connectives		
02/05 (M)	Logical equivalence		
02/07 (W)	Logical negation of any mathematical statement	HW 3	HW 2
End Module 1			
Module 2			
02/09 (F)	Combinatorial analysis		
02/12 (M)	Axioms of probability		
02/14 (W)	Conditional probability	HW 4	HW 3
02/16 (F)	Random variables		
02/19 (M)	No class		
02/21 (W)	Distributions	HW 5	HW 4
02/23 (F)	Sampling and central limit theorem		
02/26 (M)	Confidence intervals		

02/28 (W)	Hypothesis testing		HW 5
End Module 2			
03/01 (F)	Review		
03/04 (M)	Midterm 1		
Module 3			
03/06 (W)	Introduction to linear algebra	HW 6	
03/08 (F)	Solving linear systems I		
03/18 (M)	Solving linear systems II		
03/20(W)	Vector manipulations	HW 7	HW 6
03/22 (F)	Matrix manipulations		
03/25 (M)	Inverses		
03/27 (W)	Determinants I	HW 8	HW 7
03/29 (F)	Determinants II		
04/01 (M)	Vector spaces I		
04/03 (W)	Vector spaces II	HW 9	HW 8
04/05 (F)	Projections		
04/08 (M)	Linear Regression I		
04/10 (W)	Linear Regression II		HW 9
End Module 3			
04/12 (F)	Review		
04/15 (M)	No class		
04/17 (W)	Midterm 2		

Module 4			
04/19 (F)	Functions and derivatives		
04/22 (M)	Higher-order and partial derivatives		
04/24 (W)	Introduction to gradient descent	HW 10	
04/26 (F)	Integrals I		
04/29 (M)	Integrals II		
End Module 4			
05/01 (W)	Review		HW 10