0 Syllabus

Overview

Instructors

- Professor: Allison McDonald (amcdon@bu.edu)
- TAs: Bhushan Suwal (bsuwal@bu.edu) and Akshat Kolekar (akolekar@bu.edu)

Meetings

- Lecture: Monday, Wednesday, Friday at 11:15am (CGS 505)
- Discussion labs:
  - Monday at 1:25-2:15pm (STH B20)
  - Monday at 2:30-3:20pm (STH B20)
  - Monday at 3:35-4:25pm (MCS B33)
  - Monday at 4:40-5:30pm (MCS B33)

Assignments

- Homework: Due on Fridays at 5pm unless otherwise posted
- In-class midterm exam dates: February 23 and April 5 (mark these dates on your calendar now!)
- Final exam: Yes, tentatively Tuesday, May 7 from 12:00pm – 2:00 pm

Websites

- Piazza: https://piazza.com/bu/spring2024/ds121, entry code ds121s24@bu
- Gradescope: https://www.gradescope.com/courses/711905, entry code B2YZ53

Course Information

Course description. DS 121 is the second in the three-course sequence (DS 120, 121, 122) that introduces students to theoretical foundations of Data Science. DS 121 covers an introduction to key concepts from Linear Algebra (vector space, independence, orthogonality and matrix factorizations). The DS theme running through the course is exploratory data analysis, enabling a better understanding of the data at hand. The course will link mathematical concepts with geometric and computational thinking, specifically through the use of problem sets that require students to answer mathematically-posed questions using computation and visualization.

Prerequisites. DS 120 (or equivalent) is a prerequisite, and DS 110 (or equivalent) is a corequisite.

To emphasize, it is expected that at least one of the following is true: you have taken DS 110, or you are currently taking DS 110. This course *will* contain programming examples in class and programming assignments in your homework.

Learning Outcomes

Quantitative Reasoning I.

- Learning Outcome #1: Students will demonstrate their understanding of core conceptual and theoretical tools used in quantitative reasoning, such as statistics, computing, and mathematics.
- Learning Outcome #2: Students will interpret quantitative models and understand a variety of methods of communicating them, such as graphs, tables, formulae, and schematics.
- Learning Outcome #3: Students will communicate quantitative information symbolically, visually, numerically, or verbally.
- Learning Outcome #4: Students will recognize and articulate the capacity and limitations of quantitative methods and the risks of using them improperly.

This class will introduce students to core tools used in quantitative reasoning, including vector spaces, independence, matrix decomposition, dimensionality reduction and linear systems. The homeworks will give students the chance to use computational tools to manipulate and make sense of data, communicate conclusions based on mathematical and computational representations, and identify misrepresentations based on incorrect use of the tools.

Digital/Multimedia Expression.

- Learning Outcome #1: Students will be able to craft and deliver responsible, considered, and well-structured arguments using media and modes of expression appropriate to the situation.
- Learning Outcome #2: Students will be able to demonstrate an understanding of the capabilities of various communication technologies and be able to use these technologies ethically and effectively.
- Learning Outcome #3: Students will be able to demonstrate an understanding of the fundamentals of visual communication, such as principles governing design, time-based and interactive media, and the audio-visual representation of qualitative and quantitative data.

Through the lectures and in homeworks, we will explore how to visually represent data in ways that support logical reasoning. Students will take complex multidimensional data and communicate arguments and interpretations using strategies such as projections, compression, clustering, and summarization.

Critical Thinking.

- Learning Outcome #1: Students will be able to identify key elements of critical thinking, such as habits of distinguishing deductive from inductive modes of inference, recognizing common logical fallacies and cognitive biases, translating ordinary language into formal argument, distinguishing empirical claims about matters of fact from normative or evaluative judgments, and recognizing the ways in which emotional responses can affect reasoning processes.
- Learning Outcome #2: Drawing on skills developed in class, students will be able to evaluate the validity of arguments, including their own.

Students will learn to think deeply about their data, to understand its properties, and based on the initial analyses translate their intuitive understanding of their data into formal problem definitions. Students will be able to evaluate whether models provide a good fit to data, and whether a given model is valid as a description of data.

Course Resources

Course websites. We will use two websites in this course, Piazza and Gradescope. Please sign up for both immediately!

- Piazza (https://piazza.com/bu/spring2024/ds121) contains the weekly course schedule, homework assignments and required reading; you should review the course schedule at least once per week. Also, course announcements will be made using Piazza; you are responsible for knowing these announcements, so please register for Piazza using an email address that you check regularly. We also encourage you to use Piazza as a resource to ask scientific questions to the instructors and your peers, and recommend that you mark posts as public unless they contain a personal question or reveal a partial solution to an assignment.
- Gradescope (https://www.gradescope.com/courses/711905, entry code B2YZ53) is the website that you must use to submit completed homework assignments and to respond to in-class questions.

Course textbooks. We will use (portions of) 5 textbooks during this class, all of which are available to download for free using the links in the Piazza resources tab: https://piazza.com/bu/spring2024/ds121/resources. (BU login may be required.)

Course meetings. The course instructor is Prof. Allison McDonald. I will lead lectures on Monday, Wednesday, Friday at 11:15-12:05pm in CGS 505. All lectures will be livestreamed over Zoom; if you cannot come to class in person (e.g., due to illness), you can join the Zoom livestream or review the lecture notes and video on Piazza after class.

Discussions. Two TFs (Bhushan and Akshat) will hold discussions on Mondays at 1:25-2:15pm (STH B20); 2:30-3:20pm (STH B20); 3:35-4:25pm (MCS B33); and 4:40-5:30pm (MCS B33). We actively encourage questions and interaction during all lectures and discussion labs.

Office hours. Students are welcome and encouraged to visit office hours. We will post the date, time, and location of all office hours on Piazza, and will also announce in class and on Piazza if the office hour times ever change. If you want to meet with me (the professor) but cannot make the scheduled times, then send me a private Piazza note with at least 3 suggestions for times that you are available to meet and we will find an alternate time.

Course Assessment

Your grade in this course will be based on four components:

- two in-class tests (15% each)
- a final exam (25%)
- homework (40%)
- participation (5%)

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.

Homework. There will be 10 homework assignments. I will drop the lowest score and the remaining 9 will count toward your homework grade (40%). Homework will typically be assigned on a Thursday and due on the following **Friday at 5pm on Gradescope**. Late homework will be accepted for 24 hours with a 20% penalty.

Homework assignments are assessed on the correctness and clarity of your solutions; the graders are looking for positive evidence that you understand the relevant concepts. Before submitting an assignment, you should crop down to the pertinent work, document all code, describe your thought process, display any relevant outputs, clearly mark your final answer, and explain your findings.

Submitting Homeworks. Homework assignments will include two types of questions: written mathematical questions that require you to calculate an answer or explain a concept, and computer-based data science questions that require you to write computer code in Python to analyze a dataset. All homeworks will be released as Jupyter Notebooks and submitted as PDFs.

Homework Resources

- Coding problems: I recommend using Google Colab (https://colab.research.google.com) for the programming questions, although it is not required. You may also wish to set up a local Jupyter server.
- Written problems: I highly encourage you to use markdown to format your answers. However, you may also take clear photos of written work.

Always review your file after uploading to check that it is legible and complete; we will only grade what we can see.

Exams. There will be three exams in this course: two in-class midterms on February 23 and April 5, and a final exam held on the university's Final Exam Schedule (most likely May 7, 12-2pm). Reserve these dates on your calendar now! All exams are cumulative: they may cover any material from lectures, discussion lab sections, homework, and required textbook reading.

Participation. To earn full participation credit, you must attend at least 11 out of the 14 discussion labs.

Course policies

Academic code of conduct. You must read and adhere to BU's Academic Code of Conduct, which is available here: https://www.bu.edu/academics/policies/academic-conduct-code/. Please familiarize yourself with this code, its definitions of misconduct and plagiarism, and its sanctions. Violations of this code will result in receiving a score of 0 on the homework or exam, and may be grounds for referral to BU's Academic Conduct Committee. If you have any questions about the policy, you must ask me in person or via private Piazza note before taking an action that might be a violation.

Plagiarism. All written work in this course must be original to you. If you consult outside texts, or other forms of assistance, cite these sources in the proper format—at a minimum, include the author, title, and website link for all external sources (books, journals, lectures, web sites, AI). We are required to report all suspected cases of plagiarism to the Academic Dean for review.

Academic integrity in computing coursework has some special aspects. Please review the examples of plagiarism as provided by the BU Computer Science department: https://www.bu.edu/cs/undergraduate/undergraduate-life/academic-integrity/.

Generative AI policy. All submitted work in this course must conform to the CDS Generative AI Assistance Policy, which you can read at https://www.bu.edu/cds-faculty/culture-community/gaia-policy/. Also, keep in mind that AI tools are often wrong!

Collaboration policy. The goal of homework and project assignments is to learn. Hence, I encourage you to use any and all resources that can help you to learn the material: computers/calculators, Piazza, lecture notes, textbooks, other websites, and your fellow classmates. That said, please always obey the following rules:

- You cannot copy solutions from anyone else, or give your solutions to a classmate to copy.
- You also cannot actively search for the solutions to the homework questions on the Internet or in any other source.
- Your submission must list (a) names of all classmates you worked with, (b) all websites you used besides the ones listed in the lecture notes or textbooks, and (c) all code that you used from other sources, including the exact prompts and responses from any AI tool. Taking ideas without attribution will be considered plagiarism.

By contrast, the goal of the exams is for you to show me what you have learned, so any form of collaboration is strictly prohibited. Computers and notes are also forbidden during exams unless I explicitly state otherwise. (That said, I encourage you to collaborate with classmates when studying lecture materials and preparing for the exams.)

Accommodations. BU strives to be accessible, inclusive, and diverse in our facilities, programming, and academic offerings. Your experience in this course is important to the teaching staff. If you have a disability or believe that you require a reasonable accommodation, please meet with BU Disability and Access Services as soon as possible at the beginning of the semester. Their office is at 25 Buick Street, Suite 300, and they can be contacted at 617-353-3658. Requests for accommodations are sent by that office to the Academic Dean who approves and returns them. Disability and Access Services then forwards them to the instructor. More information is at https://www.bu.edu/disability/

Learning environment. This course seeks to be inclusive of people of all genders, races, cultures, abilities, and sexual orientations. Please respect your fellow classmates and contribute toward a positive learning environment for everyone. While I actively encourage discussion and debate on ideas, I won't tolerate criticism of other people. Also, while you can use a computer for note-taking, do not use your laptop or phone in class for web surfing, sending messages, or anything else that can cause a distraction to your fellow classmates.

Absence policy. This course follows BU's policy on absences for religious observance. Otherwise, students should attend the lectures and discussion lab sections. Due to ongoing public health concerns: if you feel sick, please err on the side of caution and stay home. The lecture will be recorded and the video and lecture notes will be posted immediately.

Late work policy. 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, except in cases of long-term emergencies (e.g., family, medical); if this applies to you, inform us as soon as you are able. To accommodate the possibility of missed assignments, I will drop the lowest homework grade.

Make-up exam policy. If you have a valid conflict with a midterm exam, you must send me a written note as soon as you are aware, and with a minimum of 2 weeks notice (barring extenuating circumstances) so we can arrange a make-up exam. The final exam can only be rescheduled in accordance with the university policy: https://www.bu.edu/reg/calendars/final-exams/policy/.

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.

Course feedback. I welcome feedback from you at any time about any material you would like me to cover or suggestions for how I can improve the course. I promise to consider all suggestions, though I cannot promise that I will always be able to incorporate your feedback. Please send me any comments as a private Piazza note; you can send feedback anonymously if needed.

If you have any questions about the syllabus now or later, feel free to ask!