CDS DS 120: Foundations of Data Science

Instructor Name: Dr.Tanima Chatterjee

Office Location: CCDS 1402

Contact Information: tchatt2@bu.edu

Office Hours: Monday : 12:30pm-2:00pm, Wednesday : 12:30pm-2:00pm, Friday 12:30pm-2:00pm in CCDS 1402

TA1(Discussion Section A2 12:20pm-1:10pm, A3 1:25pm -2:15pm): **Abhaya Shukla**
OF: Friday: 3:00pm- 5pm, Wednesday: 2:00pm-3:00pm in yellow corner on 14th Floor near CCDS 1411.
Email: abhaya13@bu.edu

TA2(Discussion Section A4 2:30pm-3:20pm, A5 3:35-4:25pm): **Palak Kothari**
OF: Monday: 1:00pm- 2:30pm, Thursday: 1:00pm-2:30pm in yellow corner on 14th Floor near CCDS 1411.
Email: palak@bu.edu

CA1: **Max Schultz**
OF: Tuesday: 11:00am- 12:00pm, Saturday: 12:000pm-2:00pm in yellow corner on 14th Floor near CCDS 1411.
Email: maxjs@bu.edu

CA2: **Zihao Guo**
OF: Tuesday: 2:30pm- 4:00pm, Thursday: 2:30pm-4:00pm in yellow corner on 14th Floor near CCDS 1411.
Email: z2004@bu.edu

Peer tutors: Info coming soon!!!!

Course Description

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 covers an introduction to key concepts from Proofs, Logic, Linear Algebra (vector spaces, matrices, and linear systems), Calculus (Differentiation and Integration), and Probability (discrete and continuous random variables). The four areas find culminating applications in Algorithms, hypothesis testing, gradient descent for neural networks, and regression. A detailed topic list is provided at the end of this syllabus.

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.

Hub Learning Outcomes

Quantitative Reasoning I (QR1) Learning Outcomes:

Students will develop understanding of a large set of fundamental theoretical tools used in quantitative reasoning: derivatives, random variables, probability distributions, summary statistics, vector spaces, matrices, and linear systems. Through problem solving during discussion sections, students will develop skills in using computation to answer questions posed in mathematical terms.

Students will learn to interpret data in terms of probability distributions and also in terms of vectors. Students will develop skills presenting data visually in terms of plots, histograms, etc; symbolically in terms of matrix-vector equations and linear systems; and numerically in terms of summary statistics and fitted model parameters.

Students will develop an understanding of how regression relates to causality, the risks of misinterpreting correlations in data as indicating causal relationships, and techniques for establishing causal relationships.

Other Learning Outcomes (e.g., School, Department, and/or Program Outcomes)

On completion of DS 120, students will achieve proficiency in:

- Using probability to describe data and construct models, including models involving univariate distributions and discrete and continuous random variables.

- Using basic linear algebra (vectors and matrices) to describe and solve systems of linear equations, to represent data as points in vector spaces, to represent linear transformations in terms of matrices and matrix inverses, and to represent distances and lengths in terms of norms and inner products.

- Deriving derivatives of functions with multiple variables, and using differentiation to solve problems.

- Expressing solutions to mathematical and data analysis problems in computational terms and writing code to solve those problems.

This course contributes to the learning outcomes of the BS in Data Science with respect to:

- Foundations:

    - Mastery of the principal tools of data driven decision-making, including defining models, learning model parameters, and making predictions.

    - Understanding of how to use probabilistic models to make decisions.

- Integration and Application:
    - Ability to interpret and explain results .

Instructional Format, Course Pedagogy, and Approach to Learning

This course will be primarily a lecture heavy course where students will learn the mathematical foundations required for interpreting data during the lectures. The lectures don't require students to carry laptops. It will be pen and paper lectures with in class participation.

The discussion sessions will concentrate on applying the concepts learnt during the lectures to solve mathematical questions along with a review of the topics learnt during the lecture and on practicing further problems to bolster students' understanding of the concepts. The HomeWorks will serve as a method of applying

the concepts learnt during the lectures and discussions to applied problems with almost every lecture followed by a homework. HomeWorks are also meant to check the students understanding of the concepts taught in class.

Books and Other Course Materials

The students may find the following free references useful for the course material:

Hefferon, Linear Algebra. Available free online:
http://joshua.smcvt.edu/linearalgebra/

Lane, Online Statistics Education: A Multimedia Course of Study. Available free online: http://onlinestatbook.com/

Strang, Calculus. Available free online:
https://ocw.mit.edu/ans7870/resources/Strang/Edited/Calculus/Calculus.pdf

Courseware

We'll be using Blackboard and gradescope for distributing assignments, collecting assignments, grading, and the distribution of course notes. https://learn.bu.edu
We'll be using Piazza for Q&A.

Gradescope entry code : **DPRBDR**

Piazza Sign up link: **https://piazza.com/bu/spring2024/ds120**

Assignments and Grading

Students would be evaluated on their HomeWorks which would be given to them after every lecture. The HomeWork assignments are meant be short with occasional challenge questions to ensure the students' understanding of the concepts taught in class. HomeWorks will be due by the end of the week. HomeWork grading will involve grading one Random question from each homework so students are encouraged to solve and submit all the questions in the homeworks. **No HomeWorks sent via emails will be entertained**. **THERE WILL BE NO EXCEPTIONS TO THIS RULE.**

There will be occasional bonus questions which the students can use to make up for missed homeworks or lower homework grades.

Student participation is class (**especially discussions**) along with their involvement in Piazza will hold value as well.

If a student fails to attend at least 60% of the Discussions, they would not be graded for their Bonus Homeworks

In addition, students will be tested via two midterms and a final exam. All of these are written, in-class exams.

Grading:

40% Homework Assignments

25% Final Exam

15% Midterm1

15% Midterm2

5% Class Participation (Lecture, Piazza and **Discussion**)

The grade cutoffs will be determined after all the exams and assignments are graded and will depend on the difficulty of the exams/assignments and overall class performance. The grade boundaries will not be disclosed.

Resources/Support/How to Succeed in This Course:

1. To succeed in this course students should attend all lectures, come to discussion prepared with questions, complete all assignments on time, and discuss problems and material with fellow classmates.
2. Students are welcomed and encouraged to visit office hours.
3. The Education Resource Center offers free individual and group tutoring.
4. 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.

Community of Learning: Class and University Policies

1. **Courtesy expectations.** Students are responsible for supporting a courteous learning environment. Please show respect for other students' questions, and maintain an attentive attitude in class. If you have an advanced question to ask, try to make it accessible to the rest of the class. Explain things to other students with a polite tone.

2. Attendance & Absences. Attending class is optional but highly encouraged; behavior in class does carry some small participation weight. Regular absence without notice will, however, be penalized. Attending recitation carries some participation credit. However, either can be waived with a reasonable request for accommodation. We will in particular waive for religious holidays; see the University Policy on Religious Observance.

3. Assignment Completion & Late Work. Assignments will be submitted as PDF on Gradescope. After the deadline, **NO** late submissions will be allowed. ***NO HomeWorks submitted via email will be graded.***

4. Academic Conduct Statement
Students are expected to abide by the guidelines and rules of the Academic Code of Conduct. *https://www.bu.edu/academics/policies/academic-conduct-code/*

General Course Plan with Topics to be Covered: (Not written in stone!)

HomeWorks will be due by Saturday of every week.

Introduction

Module 1: Truths and Proofs

Introduction to Proof techniques and understanding a good proof

Introduction to proofs by Mathematical Induction

Introduction to Logic: Understanding mathematical statements and truth values, Existential vs. Universal Quantifiers, Specifying Quantification sets

Negation of Mathematical statements

Conditional statements, truth tables, Set Operations and Logical Connectives, Converse and Contrapositives, Biconditional Statements

Necessary and Sufficient Conditions, Proving logical equivalences

Module 2: Linear Algebra

Introduction to Linear Algebra, Gaussian elimination, RREF

Introduction to vectors and Matrices, Operations on Vectors and Matrices

Introduction to Vector spaces and Bases, Matrix Inverse, Linear combinations, Linear independence

Ranks, Isomorphisms and Kernels

Projections

Linear Regression: Line of best fit

**More linear fits, Introduction to Multiple Regression

Module 3 : Calculus and Gradient Descent

Introduction to Multivariable Calculus: review of Single Variable Calculus concepts, Derivatives w.r.t particular variables, second order derivatives, Maxima, Minima, Saddle Points

Introduction to Gradient: Second order Partial Derivative test, Directional Derivatives

Introduction to Gradient Descent and it's limitations

Lagrange Multipliers

**Introduction to Integral Calculus

Module 4: Probability and Statistics

Overview of Sets

Introduction to Probability, choosing with and without order, Random Variables, Binomial Distribution

Introduction to Conditional Probability, Joint Distribution, Introduction to Bayes Rule

Expectation

Distribution and variance

Gaussian Distribution and Central Limit Theorem

Sampling Errors and Confidence Intervals

Bias and Bessel's Corrections

t-Distributions

Hypothesis Testing : t-tests, z-tests

Important Dates (Tentative)

Midterm1: February 26st , 2024

Midterm2: March 27th , 2024

Final (Alternate): May 1st, 2024

Final: May 7th, 2024 (TBD by Registrar's Office)