Last Updated 08/21/2019

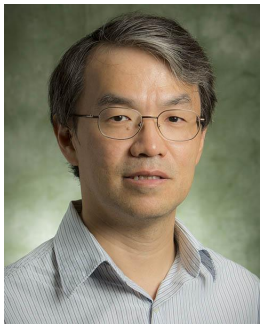
Binghamton University

Math 247/CS 207: Introduction To Data Science

Fall 2019 Syllabus

Instructor Information

Xingye Qiao
qiao@math.binghamton.edu
607-777-2593
Office: WH-134
Research Interests: Statistics, Machine Learning, High-dimensional Inference, Big Data Analytics, Precision Medicine.



Kenneth Chiu
kchiu@cs.binghamton.edu
607-777-7320
Office: EB Q-15
Research Interests: Cyberinfrastucture for Instruments and Sensors, Web Services Performance, Middleware for Scientific Computing.

Office Hours for XQ: Tuesday 9 to 10 am. Thursday 9 to 10:30 AM.
Office Hours for KC: Tuesday and Thursday, 7:30 PM in AA G007 (email if you will arrive after 7:30; no need to email if you will arrive at 7:30 PM)
The preferred mode of communication is posting questions on the discussion board, Piazza. Use the office hour only if it is difficult to communicate via the discussion board.

Teaching Assistants and Office Hours:

William T Huebner	whuebne1@binghamton.edu	Monday 5 pm
Zhen Lin	zlin89@binghamton.edu	Monday 4 pm
Derrick Lam	dlam15@binghamton.edu	Wednesday 3 pm

All TA's office hours are hold in EB N0.

Communication Policy

Preferably all communication will be done on Piazza. General questions about the course contents can be asked so that both the questions and the answers can be visible to and answered by all. Private message can also be sent on Piazza. Do not post code on Piazza unless it is part of the question.

Last Updated 08/21/2019

Course Information

Description

This course will provide a broad overview of data science's different areas, from statistics, machine learning to data engineering and many data science applications. The course teaches critical concepts and skills in computer programming and statistical inference, in conjunction with hands-on analysis of real-world datasets, including economic data, document collections, geographical data, and social networks. This course is designed to provide a non-technical introduction to the data science approach. It is intended both for students from non-quantitative fields and those from a quantitative field who are interested in data science. The prerequisite for this course includes working knowledge in high school math (Math 108 or equivalent), a course in introductory statistics (at the level of Math 147, Math 148 or AP statistics with grade 3+, or equivalent).

Course Objectives

- *Describe the life cycle of data science, the data science process and how its components interact.*
- *Identify different goals of data visualization and demonstrate proper use of visualization tools in different scenarios.*
- *Appreciate the significance of exploratory data analysis in data science.*
- *Use Python and/or R to collect, clean, wrangle, visualize and exploratorily analyze data.*
- *Explain in basic terms the meaning of statistical inference.*
- *Apply predictive models to data and make predictions using Python and/or R.*
- *Construct a learning pipeline including data collection, cleaning, analysis and inference.*
- *Communicate results of data analysis as well as the uncertainty in the learning process.*
- *Work effectively and synergically in teams on data science projects.*
- *Recognize common ethical and privacy issues in both data-driven research and applications of data science in the society. Criticize unethical use of statistical analysis and misinterpretations of data. Argue about ethical use of data science.*

General Education Information

It is expected that this course will have a General Education Designation for M - Mathematics/ Reasoning.

Prerequisite(s)

It is highly recommended that the student has taken an introductory statistics course before taking Math 247/CS 207. The curriculum and format is designed specifically for students who have not previously taken ***calculus-based*** *statistics or computer science courses. Students*

who have taken several statistics or computer science courses before should instead take a more advanced course.

Relationship to Other Courses

This course is the first course in the area of data science. It exposes students to aspects of data science and prepare them for more advanced courses in data management, machine learning, and statistics.

Format and Procedures

The course is divided into about 30 lectures over 15 weeks. There is lab assignment and homework every week. There are also three projects throughout the semester and one final exam.

Course Requirements

Required text

- *Computational and Inferential Thinking (CIT): https://www.inferentialthinking.com/*
- *Principles and Techniques of Data Science (PTDS): https://www.textbook.ds100.org/*

Both books are web-based and free to students. We will focus on CIT in the first half of the semester and PTDS in the second half of the semester.

Other Requirements

- *Internet, Blackboard (mycourse) access, Piazza access.*
- *Software: Python, R.*

Recommended Texts & Other Readings

- *Introduction to Statistical Learning (free online PDF available). This book is a great reference for machine learning and some of the statistics material in the class.*
- *Data Science from Scratch. This more applied book covers many of the topics in this class using Python but doesn't go into sufficient depth for some of the more mathematical material.*
- *Doing Data Science. This book provides a unique case-study view of data science which uses R and not Python.*
- *Python for Data Analysis. This book provides a good reference for the Pandas library.*

Credit Hours and Expectations

This course is a 4-credit course, which means that in addition to the scheduled lectures/ discussions, students are expected to do at least 9.5 hours of course-related work each week during the semester. This includes things like: completing assigned readings, participating in lab sessions, studying for tests and examinations, preparing written assignments, completing

internship or clinical placement requirements, and other tasks that must be completed to earn credit in the course.

Assignments

- *There will be 14 weekly lab assignments, 13 weekly homework and three projects.*
- *A final exam will be administered at the end of the semester.*

Python vs R Programming Languages

Both Python and R are popular programming languages. In this course, we will primarily use Python as the teaching vehicle. This choice is not necessarily an endorsement to Python or a disapproval of R. In fact, we conduct research using R all the time. We realize that for most students who have not been exposed to any programming languages before, teaching two new languages in one course will be too risky and ambitious. That said, we would like to stress that a good data scientist is not bound by the programming language he/she uses. To encourage students to explore the R programming language, we have included a "Bilingual Challenge". Students who take on the challenge will be provided some resources to learn the R programming language. They will be asked to complete a series of tasks to practice their bilingual skills. In particular, these may involve the "translation" of lecture demos, and lab or homework assignments from Python to R. Students who have successfully completed the challenge will receive 5% bonus point toward their final grade. Students could propose to use a different second language from R. A promising candidate can be the Julia Language (https://julialang.org/). Please contact either instructor if you would like to take upon this challenge and ask for details.

Grading

Activity	***Grade***
Lab	20%
Homework	25%
Projects	20%; 22.5% if opting out of attendance.
Final	30%; 32.5% if opting out of attendance.
Attendance	5%; 0% if opting out.
Bilingual Challenge	5% bonus point.

Grading Scheme

Grade	***Percent***
A	94 – 100%
A-	90 – <94%

B+	87 – <90%
B	83 – <87%
B-	80 – <83%
C+	77 – <80%
C	73 – <77%
C-	70 – <73%
D	64 – <70%
F	<64%

Accessing Grades

Students can assess their grades using the Blackboard.

Computer Setup

We assume that you have access to a working laptop or computer. If you have difficulty on that front, please let us know immediately. There might be ways to work around it.

Instructions for setting up the computer can be found on this Google Doc file. In particular, we ask you to install Anaconda as well as a few packages. Moreover, you will need to set up an okpy.org account to allow for auto-grading.

Course Policies

Penalties for Late Work

Lab assignments and homework that are late within 1 week will be deducted 50% of the grades. No late work is reviewed after 1 week or after the solution is posted. Late projects will have a 20% score penalty applied for each day after the deadline.

Final Exam Conflict

Students should not have to take three or more final examinations in one 24-hour period. If you find out that you are in such a situation, please let us know as soon as possible (before the withdrawal deadline.)

Absences Due to Religious Holidays

If you anticipate being absent because of any religious observance, please notify us in writing at least one week in advance. We will work together to accommodate.

Attendance & Participation

Class attendance and participation points are given to encourage your active class participation and discussion. You will be rewarded with a perfect score as long as you come to more than ⅔ of the class session (that is 20 sessions) and actively contribute to the class discussion during recitations and lectures. We will use a Google Form to record the attendance. You may use your

mobile devices or laptops to sign the attendance form. If you do not have the equipment to sign the attendance form, you must let the instructor know IN PERSON immediately after the class.

Use of Mobile Devices

Unexpected noises and movement automatically divert and capture people's attention, which means you are affecting everyone's learning experience if your cell phone, tablet, laptop, etc. makes noise or is visually distracting during class.
For this reason, I

- *ask you to restraint from checking your mobile devices during class except when signing the attendance form;*
- *allow you to take notes or program Python/R code on your laptop, but you must turn the sound off so that you do not disrupt other students' learning. If you are doing anything other than taking notes or conducting course related work on your laptop (including but not limited to checking social media, shopping, doing homework for another course, etc.), please sit in the back row so that other students are not distracted by your screen.*

Understand When You May Drop This Course

- *The deadline for dropping this course is: September 4, 2019.*
- *The deadline for withdrawing from this course is: October 28, 2019.*
- *It is possible to enter an incomplete contract with the instructors if you cannot complete the course in time. Consult either of the instructor.*

Academic Integrity

Binghamton University provides explicit guidelines in the Student Academic Honesty Code (see the University Bulletin - Academic Policies and Procedures for All Students). Unless specified otherwise in the syllabus, I expect the work you submit for grading to be yours and yours alone. Not acknowledging another's work with proper references, taking credit for someone else's work, letting your work appear in another student's paper, or fabricating "results" are grounds for failing the assignment and/or the course. If you have any questions about what constitutes plagiarism or cheating, please ask the instructors.

Disability-Related Equal Access Accommodations Statement

Students wishing to request academic accommodations to insure their equitable access and participation in this course should notify the instructor as soon as they are aware of their need for such arrangements. Authorizations from Services for Students with Disabilities (SSD) are generally required. We encourage you to contact SSD at (607) 777-2686 to schedule an appointment with the Director or Learning Disabilities Specialist. The SSD website (www.binghamton.edu/ssd/) includes information regarding their Disability Documentation Guidelines. The office is located in UU – 119.

Last Updated 08/21/2019

Campus Help for Students

University Tutoring Services

UTS offers free tutoring for undergraduate students at Binghamton University. All UTS tutoring appointments must be scheduled online through the my.binghamton.edu portal. Walk-in tutoring is also available for select courses. If you have any questions about UTS, call 607-777-9235, email uts@binghamton.edu, or visit the website: http://www.binghamton.edu/tutoring.

ITS Helpdesk/myCourses Support

Walk-in: Located in the Computer Center first floor lobby.
Call: 607-777-6420; E-mail: helpdesk@binghamton.edu.
https://www.binghamton.edu/its/

Libraries

The Libraries offer a wide variety and range of services including research assistance, instruction, user-friendly interfaces, digital preservation, digital scanners, and resource sharing.
Text: 607-205-8173; Call: 607-777-2345; Email: refquest@binghamton.edu
http://www.binghamton.edu/libraries

University Counseling Center

At some point during their college experience, students may encounter personal, social, or developmental issues that call for assistance beyond the advice provided by friends and family. That's where the University Counseling Center (UCC) can help. The UCC provides a variety of free and confidential counseling services delivered by professional counselors. All currently enrolled Binghamton University undergraduate students, graduate students and affiliated entities are eligible to receive these services free of charge. Services and programs available through the center include individual and group counseling, consultation, referral, and psychoeducational programs. For more information or to make an appointment, visit https://www.binghamton.edu/counseling.

When and Where to Seek Help

If you are experiencing undue personal or academic stress at any time during the semester or need to talk with someone about a personal problem or situation, I encourage you to seek support as soon as possible. I am available to talk with you about stresses related to your work in my class. Additionally, I can assist you in reaching out to any one of a wide range of campus resources, including:

1. *Dean of Students Office: 607-777-2804*
2. *Decker Student Health Services Center: 607-777-2221*
3. *University Police: On campus emergency, 911*
4. *Interpersonal Violence Prevention: 607-777-3062*

5. *Harpur Advising: 607-777-6305*
6. *Office of International Student & Scholar Services: 607-777-2510*