

ECE 20875: Python for Data Science

Spring 2020
Course Syllabus

Lectures: Mondays, Wednesdays, and Fridays, 12:30—1:20

Section 1: WALC 1055

Section 2: FRNY G124

Course web page: <http://www.cbrinton.net/ECE20875-2020-Spring.html>

Piazza discussion page: <https://piazza.com/purdue/spring2020/ece20875>

Instructors:

Section 1

David Inouye (dinouye@purdue.edu)

Office: EE 332

Office Hours: Mondays 1:30-2:30 and Thursdays 4:00-5:00

Section 2

Chris Brinton (cgb@purdue.edu)

Office: MSEE 342

Office Hours: Tuesdays 10:00–11:00 and Wednesdays 1:30–2:30

Graduate TAs:

Somosmita Mitra [Lead] (mitra26@purdue.edu)

Brandon Kozel (bkozel@purdue.edu)

Rajeev Sahay (sahayr@purdue.edu)

Undergraduate TAs:

Tyler Baumgartner [Lead] (baumgar6@purdue.edu)

Colin Harrison [Lead] (harri640@purdue.edu)

Aidan Abbott (abbott34@purdue.edu)

Sabriya Alam (alam10@purdue.edu)

Jason Chen (chen2156@purdue.edu)

Isha Ghodgaonkar (ighodgao@purdue.edu)

Chieh-En Li (li3261@purdue.edu)

Jerome Schweitzer (jschweit@purdue.edu)

Julia Taylor (taylo773@purdue.edu)

Chien-Hung Wang (wang4161@purdue.edu)

TA Lab Hours:

M–F 10:30–11:30 and 5:30–8:30 in EE 207

Course Outcomes: A student who successfully fulfills the course requirements will have demonstrated:

1. An understanding of regular expressions [1, 2]
2. An ability to use Python to write data analyses [1, 2, 6]
3. An ability to explain when particular data analyses are appropriate [1, 2, 3, 6]
4. An ability to explain the results of data analyses [2, 3, 5]
5. An ability to incorporate classes in their Python programs [1, 2, 6]
6. An ability to incorporate associative arrays in their programs [1, 2, 6]
7. An ability to work with a partner to choose appropriate analyses to solve a problem, perform those analyses, and interpret the results of those analyses [1, 2, 3, 5, 6]

These outcomes are extremely high level. In more detail, after taking this course you will be able to:

- Write programs in Python that incorporate: basic control structures; functions; data structures such as lists, tuples and associative arrays; classes and objects; iterators and generators; higher order functions; and regular expressions
- Perform basic scripting tasks in Bash and Python (e.g., redirecting the output of one program to the input of another, setting and reading environment variables)
- Understand the basics of sampling, estimation, and hypothesis testing
- Write data analyses that perform: visualization; textual analysis; regression; classification; clustering
- Perform more complex analyses using approaches like neural networks.

Course assessment: The achievement of the course objectives will be through 10-12 programming assignments (covering outcomes 2, 4, 5, 6, and 7), and three midterms (covering outcomes 1, 3, and 4).

Course grading: Grades will be assigned as follows:

- 45% — Programming assignments (all equally weighted)
- 45% — Three exams (15% each)
- 10% — Mini-project
- 5% bonus — Class participation

Class participation will be assessed by participation on the Piazza discussion board, as well as in class.

Programming assignments: Programming assignments will be due every 1-2 weeks, except for the weeks of the midterms and during the final two weeks (which is set aside for the mini-project). They will test the concepts covered in class, both programming and statistical. The following rules apply to *all* programming assignments:

1. All programs should run correctly in the versions of Python available on the Scholar cluster. You will have a Scholar account created for you at the beginning of the semester.
2. Some assignments will include a separate writeup as well as a code submission that will produce the required output when run.

3. Unless otherwise specified, assignments are due at 11:59 PM on the deadline.

Programming assignments will be submitted via GitHub Classroom (<https://classroom.github.com>). As such, *you are required to have a GitHub account*. These can be obtained for free at <https://github.com>.

Please fill out the following form to provide your GitHub account information:

https://docs.google.com/forms/d/e/1FAIpQLSeyoTyMh8t4EWR9JDo35gJUaQo-UTWc_zERxoJjG6gl2vTddQ/viewform

Assignment late submission policy: Except for medical and family emergencies (accompanied by verification), there will be *no individual extensions* granted for programming assignments or the mini-project. Late submissions will be scaled according to lateness, docking 10% from your score per day late, up to a maximum of 50%. Submissions more than 5 days late will be assigned a score of 0.

Exams: The first two exams will take place in the evening, roughly during the 6th and 11th weeks of the semester. Class periods will be canceled to compensate for these taking place outside of the lectures. The exact dates will be announced once they are determined. The third exam will take place during the final exam period.

Missing an exam will result in a grade of 0 for that exam. Exam reschedules will only be granted if you demonstrate that it is absolutely necessary (accompanied by verification) and obtain written approval from one of the instructors in advance of the exam.

Course discussion page: We will be using Piazza for class discussion. If you have questions about the course or the mini-project, we encourage you to post them on Piazza. It's a shared discussion forum, where your question can be answered by the instructors, the TAs or your fellow students!

Find our class's Piazza page at:

<https://piazza.com/purdue/spring2020/ece20875>.

Active participation on Piazza is one of the key ways of receiving points towards the bonus class participation.

Lecture materials: Lecture slides, coding examples, and tutorials will be released on the course website after the corresponding lecture. The course website can be found here: <http://www.cbrinton.net/ECE20875-2020-Spring.html>

Attendance: Students are expected to be present for every class. If a student misses a class, the student is responsible for catching up on any missed content (e.g., by talking with classmates and reading posted notes). The instructors will not recap lectures in office hours.

Lab/office hours: Lab hours and office hours are a venue for you to obtain in-person clarification to questions you have about the course material, assignments, mini-project, or exams. We will expect that you have been attending class regularly and reviewing the notes before asking questions in these venues, i.e., they must be used as a supplement, not an alternate, to the lectures. Additionally, we will expect that you have put some effort into understanding the material before coming to office hours. For instructor office hours, we will give preference to conceptual or clarification questions over debugging questions because debugging questions could be answered during the daily TA lab hours.

Email: Questions about course material or programming assignments *should be posted to Piazza* or raised during lecture or office hours. The professor and TAs will not answer programming questions via email. This is to allow other students who might have similar questions to benefit from our answers. Of course, if you have questions of a personal or confidential nature, we welcome your email.

Regrade requests: If you believe that we have made an error in grading, you may submit a regrade request with an explanation through the lead graduate TA (Somosmita). All such requests must be made within one week of when the grade was posted. Please note that during the regrading process we reserve the right to regrade any portion of your submission.

Course announcements: Course announcements, including changes in due dates, course topics, programming assignment details, etc., will be communicated in two ways:

1. Posts on Piazza
2. In lecture

Course Schedule: Below is a rough schedule of when and what will be covered in class:

Week	Topic(s) covered
1 1/13-1/19	Introduction Python basics (variables, control structures, functions) Git basics
2 1/20-1/26	1/20: MLK Day (No Class) Python basics (continued) Python data structures (lists, tuples, associative arrays)
3 1/27-2/2	Histograms Probability distributions Random variables
4 2/3-2/9	Filters Map/reduce Higher-order functions

Week	Topic(s) covered
5 2/10-2/16	Sampling and estimation Hypothesis testing Confidence intervals
6 2/17-2/23	Hypothesis testing (continued) Confidence intervals (continued) Exam 1
7 2/24-3/1	Regular expressions Basic text processing Bash scripting
8 3/2-3/8	Introduction to supervised learning Matrix algebra Regression and least squares
9 3/9-3/15	Regression (continued) Regularization Cross validation
3/16-3/22	Spring Break: No Class
10 3/23-3/29	Cross validation (continued) n-grams Natural language processing
11 3/30-4/5	Objects Inheritance Iterators and generators
12 4/6-4/12	Exam 2 Introduction to unsupervised learning Clustering: K-Means
13 4/13-4/19	Clustering: GMMs Classification: Naive Bayes Classification: kNN
14 4/20-4/26	Classification: Logistic regression Introduction to deep learning Perceptrons
15 4/27-5/3	Neural network architectures Backpropagation Conclusion

Academic honesty: Unless expressly allowed, you are expected to complete all assignments by yourself. However, you are allowed to discuss general issues with other students (programming techniques, clearing up confusion about requirements, etc.). You may discuss particular algorithmic issues on Piazza (but do not copy code!). *We will be using software designed to catch plagiarism in programming assignments, and all students found sharing solutions will be reported to the Dean of students.*

Punishments for academic dishonesty are severe, including receiving an F in the course or being expelled from the University. By departmental rules, all instances of cheating will be reported to the Dean. On the first instance of cheating, students will receive a 0 on the assignment; the second instance of cheating will result in a failure of the course.

Campus interruptions: In the event of a major campus emergency, course requirements, deadlines and grading percentages are subject to changes that may be necessitated by a revised semester calendar or other circumstances beyond the instructor's control. In such an event, information will be provided through the course website, Piazza, and email.