STATS 507: Data Science and Analytics using Python
Winter 2024

Description

Stats 507 surveys the software tools that are currently popular among data scientists in academia and industry. The course begins with an accelerated introduction to programming in Python. Next, we focus on Python's scientific computing stack: numpy, scipy, pandas, and scikit-learn. We also cover using git and the UNIX/Linux command line. The final part of the course is an introduction to deep learning using PyTorch.

Prerequisites

This course is designed to accommodate students entering with a wide variety of programming backgrounds. However:

1. All students must have some prior programming experience (though not necessarily in Python).
2. The amount of time required to do well in the course varies by background.
   - Students with the strongest programming backgrounds may complete each weekly homework assignment in as little as 5 hours.
   - Students with little prior programming experience will likely learn the most in this class, however, they may need to devote as much as 25 hours weekly to completing the homework assignments.
   - Most students complete the weekly homework in around 10 hours.

Instructors

Dr. Xian Zhang, *xianz@umich.edu*

Graduate student instructors

Julian Bernado: *bernado@umich.edu*

Roman Kouznetsov: *roko@umich.edu*

Office hours: see Canvas

Lectures

This is a lecture format class. We meet Monday & Wednesday, from 8:30- 10:00 am in AUD C AH. We will end by 9:50 am.

Textbook, Reading & Online Resources

The first part of the course is based on *Python for Everybody* by Charles Severance:

https://www.py4e.com/book.php

The last part of the course is based on *Understanding Deep Learning* by Simon J.D. Prince:

https://udlbook.github.io/udlbook/

No textbook is required for the remainder of the course. Lecture slides will be posted on Canvas. The following tutorials may also be useful:

- *NumPy quickstart*, https://numpy.org/devdocs/user/quickstart.html
- *SciPy tutorial*, https://docs.scipy.org/doc/scipy/tutorial/index.html
- *Pyplot tutorial*, https://matplotlib.org/stable/tutorials/pyplot.html
- *Pandas tutorial*, https://pandas.pydata.org/pandas-docs/stable/user_guide/
- *Seaborn tutorial*, https://seaborn.pydata.org/tutorial.html
- *Scikit-learn tutorial*, https://scikit-learn.org/stable/tutorial/
- *PyTorch* tutorial, https://pytorch.org/tutorials/

Course website

Homework assignments and grades will be posted on Canvas. Questions about homework may be asked through the Canvas discussion board. For questions about homework that cannot be asked without revealing a solution, please ask during GSI office hours rather than on Canvas (and rather than by email).

Grading

Final grades will be based on weekly homework assignments (70%) and a final project (30%). I expect the distribution of final grades will be similar to what it has been for previous offerings of this course: 45% A/A+, 70% A-/A/A+, 98% B- or above.

Homework & late days

Homework grades will be based on cumulative performance on approximately ten homework assignments. The exact number of homework assignments depends on factors such as lecture cancellations and how fast we cover material. Each homework assignment is worth a given number of points. Homework due dates are strict, and you may turn in work late only with the use of "late days", of which you have seven to use over the course of the semester. For each late day you spend, you extend the deadline of a homework by 24 hours. You may spend multiple late days per homework. Once you have turned in your homework you may not spend more late days to turn in your homework again. The purpose of this late day policy is to enable you to deal with unexpected circumstances (e.g., illness, family emergencies, job interviews) without having to come to me. Of course if dire circumstances arise (e.g., long-term illness that causes you to miss multiple weeks of lecture), please speak with me as promptly as possible.

Due to the university grading schedule, you may not use late days to extend any deadline of the last homework assignment: Monday, Dec. 9. 2024.

Final project

During the course, students are expected to learn Python well enough that they can efficiently solve basic programming problems. Students with those skills are well prepared for technical interviews.

The final project will test that students have acquired these skills. The specific format of the final project will be announced one month before the final class.

Ethics and class policies

From the LSA Community Standards of Academic Integrity: Academic dishonesty may be understood as any action or attempted action that may result in creating an unfair academic advantage for oneself or an unfair academic advantage or disadvantage for any other member or members of the academic community. Conduct, without regard to motive, that violates the academic integrity and ethical standards of the College community cannot be tolerated. See https://lsa.umich.edu/lsa/academics/academic-integrity.html for more information. You are welcome to discuss homework with your classmates, but the work that you turn in must be yours and yours alone, and you must disclose the names of those you spoke with in your homework, including both classmates and others outside the class. This disclosure applies whether a student has helped someone else or has received help. However, it is not necessary to disclose any discussion you have with the course instructor or the course GSIs. Violations of these or other university ethical standards surrounding academic honesty will be met with serious consequences and disciplinary action. Students who cheat on an assignment will received failing grades in the course. All incidents will be reported to the Rackham Graduate School Dean's Office, which typically imposes additional penalties.

Accommodations for students with disabilities

If you need an accommodation for a disability, please let me know as promptly as possible. Some aspects of this course may be modified to better suit you. As soon as you make me aware of your needs, we can work with the Services for Students with Disabilities (SSD) office to determine appropriate academic accommodations. SSD (734-763-3000; http://ssd.umich.edu) typically recommends accommodations through a Verified Individualized Services and Accommodations (VISA) form. Any information you provide SSD is confidential.