Syllabus
STATS507: Data Science in Python
Winter 2019, 3 Credits

Description

STATS507 surveys some of the tools and frameworks currently popular among data scientists and machine learning practitioners in academia and industry. The first half of the course will consist of an accelerated introduction to programming in Python. The second half of the course will survey tools for handling structured data (regular expressions, HTML/XML/JSON, databases), visualizing complex data, interacting with the UNIX/Linux command line, processing large data sets (Hadoop and Spark), and building models with Google TensorFlow.

Prerequisites There are no strict prerequisites for this course. All students should have some background in programming, preferably in Python.

Instructor

Keith Levin, `klevin@umich.edu`
Office: West Hall 272
Instructor office hours: WF 10:00 a.m. to 11:30 a.m., West Hall 438, or by appointment.

GSI

Roger Fan, `rogerfan@umich.edu`
GSI office hours: Mondays 1:00 p.m. to 4:00 p.m., SLC (2165 USB).

Meetings

Lecture: Wednesdays and Fridays, 8:30 a.m. to 10:00 a.m., DANA 1040.

Textbook, Readings & Online Resources

There is no physical textbook required for this course. In the first half of the course, we will make frequent reference to Allen B. Downey's *Think Python*, available at `http://greenteapress.com/wp/think-python-2e/` and to Charles Severance's *Python for Informatics*, available at `https://www.py4e.com/book`. Other required readings will be made available as we cover relevant material, and supplemental readings will be suggested the course for those who are interested in learning more.

All class resources will be made available on the course web page, `http://www.umich.edu/~klevin/teaching/Winter2019/STATS507/`. and to the course Canvas page. Please contact the instructor if any resources are missing from either of these websites. The instructor will make an effort to post slides a few days ahead of time so that they are available for printing before lecture. It is recommended, though not required, that students complete assigned readings before lecture.

Course Topics

- **Introduction to Python**. Data types. Common programming patterns. Classes and objects. Functional programming.
- **Visualization with** `matplotlib`. Basic plotting.
- **Processing Structured Data**. Regular expressions. Markup languages. Databases and SQL.
- **Basic UNIX/Linux**. Files and directories. `ssh` and basic commands. Text editors.

- **Big data and distributed processing**. Basics of parallel/cloud computing. The MapReduce framework. Hadoop and Spark.
- **Specifying and training models with** `TensorFlow`. Basics of Google TensorFlow. Function graphs. Symbolic differentiation.

Grading, Homeworks & Late Days

Grades will be based on cumulative performance on a set of eight to ten homeworks. There is no final exam for this course. The exact number of homework assignments is subject to change depending factors such as lecture cancellations and the speed with which we cover material. Each homework assignment is worth a given number of points, and grades will be based on a percentage out of the total possible points. Assignments later in the semester will be worth more points, on average, than those earlier in the semester. I reserve the right to curve these scores in the event of skewed class performance. Students may contest their grade on an assignment up to two (2) weeks from the day that an assignment's grades are released, after which grades may not be changed. In order to comply with the registrar's grading schedule, students may not contest any grade more than one (1) week past the grading of the final homework. Homework due dates are strict, and you may turn in work late only with the use of "late days", of which you have seven (7) to use over the course of the semester. By spending one late day, you may turn in your homework up to 24 hours after the deadline. 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 give you a way 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. **Note:** owing to the university grading schedule, you may not use late days to extend any deadline beyond the day of the final, Thursday, May 2.

Key Dates First lecture: Wednesday, January 9, 2019
Last lecture: Wednesday, April 22, 2019
Last homework due: Thursday, May 2 by 11:59 p.m. (this due date may not be changed using late days).

Ethics and class policies

The strength of our academic community, and indeed our society at large, depends on academic and personal integrity. As such, it is vital to the integrity of the university and the value of your degree that you do not commit academic misconduct during this course. Academic misconduct includes such actions as copying code from the web or from your fellow students, looking up solutions online, turning in assignments from other classes or previous iterations of this course, and hiring others to complete your work for you. You are welcome to discuss homeworks 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. Violations of these or other university ethical standards surrounding academic honesty will be met with serious consequences and disciplinary action. 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.

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 facilitate your participation and progress. As soon as you make me aware of your needs, we can work with the Services for Students with Disabilities (SSD) office to help us 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 is private and confidential and will be treated as such by the instructor.