Syllabus

STATS700-002 Topics in Statistics: Data Science and Analytics using Python
Fall 2017, 1.5 Credits

Description
STATS700-002 is a topics course surveying some of the tools and frameworks currently popular among data scientists and machine learning practitioners in academia and industry. The course will cover 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: Tuesdays and Thursdays 2:30 to 4:00 in West Hall 313 or by appointment.

GSI
Roger Fan, rogerfan@umich.edu
GSI office hours: Tuesdays 10:30 to 11:30.

Meetings
Lecture: Mondays and Wednesdays 2:30 p.m. to 4:00 p.m., Chem 1200

Textbook, Readings & Online Resources
There is no physical textbook required for this course. In the first few weeks of the course, we will make frequent reference to Charles Severance’s Python for Informatics, available at http://www.pythonlearn.com/book.php. A Python 3 version of the book is 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 throughout the course for those who are interested in learning more.

All class resources will be made available on Canvas and on the course webpage, at http://www-personal.umich.edu/~klevin/teaching/Fall2017/STATS700-002/. Please contact the instructor if any resources are missing from one or the other 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. Students should complete assigned readings before lecture.

Course Topics
- **Processing Structured Data.** Regular expressions. Markup languages. Databases and SQL.
- **Basic UNIX/Linux.** Files and directories. ssh and basic commands. Text editors.
- **Visualization with matplotlib.** Basic plotting.
- **Big data and distributed processing.** Basics of parallel/cloud computing. The MapReduce framework. Hadoop and Spark.

Homeworks & Late Days
Grades will be based on cumulative performance on a set of four homeworks, each worth approximately 25% of your grade, with the first and third homeworks being worth slightly more and the other two homeworks worth less. There is no final exam for this course.

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. Note that 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) 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: the due date of the last homework is fixed to the day of the final, December 19th. Owing to the university grading schedule I cannot allow you to use late days to delay any homeworks beyond this date, so you should use your late days on the first three homeworks!

Key Dates
First lecture: Monday, October 23, 2017
Last lecture: Monday, December 11, 2017
Last homework due: Tuesday, December 19 by 11:59 p.m. (note that this due date is strict and cannot 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 classmates, 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, the assignments, the in-class activities, and the way the course is usually taught 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 me.