Data Science II
STAT/CS 387

Time: Tu/Th 10:05-11:20
Place: Votey 207

Professor James Bagrow

Office: 206 Farrell Hall, Trinity Campus, 210 Colchester Ave.
Email: james.bagrow@uvm.edu

Office Hours: Wednesdays 8:30–11:30 AM, or by appointment
Course website: http://bagrow.com/ds2/
Textbook: None

Prerequisites: STAT/CS 287 Data Science I

Extracting meaning from data remains one of the most important tasks of science and industry. The Internet and modern computers have given us vast amounts of data, so it is more important than ever to understand how to collect, process, and analyze these data. A picture is worth a thousand words, so visualizations are crucial to summarize and communicate new discoveries.

Goals In this course students will continue their path towards data science mastery, learning:

1. statistical methods and problems from decision theory, causal inference, information theory, and more;
2. practical implementations of advanced statistical and data collection analyses;
3. how to deal with large-scale datasets, remote computing, and “big data”-ready pipelines;
4. scientific computing pipelines, software testing, “defensive” data analysis, and revision control;
5. to explore and synthesize the literature of cutting-edge data analytics;
6. to communicate data-driven results.

As with Data Science I, particular emphasis will be placed on nontraditional (non-numeric) data such as networks, text corpora, etc. and on developing good habits for rigorous and reproducible computational science.

Grades 45% for homework, 15% for reading reflections, 40% for final project and presentation.

• A homework template and a reading template are provided. Assignments that do not use these templates will not be accepted. Submissions should be clearly written, proofread, and typed; handwritten submissions will not be accepted. LaTeX is recommended for mathematical derivations in your assignments.

• Unexcused late assignments will be marked off 10% for every calendar day past due; assignments later than five days are automatically given a grade of zero. Late final projects and presentations will not be accepted. Check due dates carefully.

Programming This is a programming-intensive course taught using Python (version 3) and other open source tools. Python is free, easy to learn, and has many useful third-party packages. As with Data Science I, you should use the Anaconda Python working environment:

• Anaconda. A free, scientifically-focused “bundle” of Python and important Python libraries. It provides a text editor, enhanced interactive prompt called IPython, and a graphical package manager.
Unlike Data Science 1, at this level, students should already be well versed with programming and are expected to generate correct code on their own, and so instruction will focus on higher-level concepts (and not, for example, debugging error messages).

**Logistics**  Course materials will be distributed via Blackboard.

I assume you have a personal computer to work from. If this is not the case, please see me so we can make accommodations.

**Projects**  Students are free to choose a project topic they are interested in. All students should come to office hours or make an appointment with me to discuss their project topics before they get started, to make sure the project is acceptable for the course. A written report and code are to be handed in. When indicated, code will be graded on clarity and reproducibility, so you are expected to have simple, readable, and well commented code.

Final projects include a final presentation given during the scheduled final exam period for the course. Unless excused, attendance during the final exam is mandatory.

**Remarks:**

- The instructions for assignments and projects will contain criteria for organizing and formatting submissions, such as file name guidelines. Please follow these instructions carefully. Improperly formatted submissions may not be accepted. I reserve the right to deduct points for unorganized or illegible submissions.
- A 24-hour clock will be used for due dates, so noon is 12:00 and midnight is 00:00.
- Spreadsheets are not allowed!
- Please do not make course materials, assignments, your own work, etc. publicly available without my prior permission.
- Please do not copy someone else’s work. It does not help with the learning process and violates the UVM student Code of Ethics. This includes plagiarism of online and offline sources (see below for more).
- University attendance policy: The lectures will form the bulk of materials, so attendance is important. Please refer to the most recent UVM Catalogue: “Students are expected to attend all regularly scheduled classes. The instructor has the final authority to excuse absences.”
- I may convey important information to you via your UVM email account. If you do not use your uvm.edu email account, please have mail from this account forwarded to an account you check frequently. When emailing me, please include Data Science in the subject line.
- Offenses against academic integrity are any acts which would have the effect of unfairly promoting or enhancing one’s academic standing within the entire community of learners. Such acts are serious offenses, which insult the integrity of the entire academic community of the University. Any suspected violations of the policy will not be tolerated and all allegations will be forwarded to the Center for Student Ethics & Standards.
In keeping with University policy, any student with a documented disability interested in utilizing accommodations should contact SAS, the office of Disability Services on campus. SAS works with students and faculty in an interactive process to explore reasonable and appropriate accommodations, which are communicated to faculty in an accommodation letter. All students are strongly encouraged to meet with their faculty to discuss the accommodations they plan to use in each course. A student’s accommodation letter lists those accommodations that will not be implemented until the student meets with their faculty to create a plan. Contact SAS: A170 Living/Learning Center; 802-656-7753; access@uvm.edu; or https://www.uvm.edu/academicsuccess/student_accessibility_services

Religious Holidays: Students have the right to practice the religion of their choice. If you need to miss class to observe a religious holiday, please submit the dates of your absence to me in writing by the end of the second full week of classes. You will be permitted to make up work within a mutually agreed-upon time.