

CMSE 381: Fundamentals of Data Science Methods

Class: MWF 3:00 - 4:20, Engineering Building 2243

Credit hours: 4

Instructor: Elizabeth Munch, Ph.D. (She/her) muncheli@msu.edu
Assoc. Prof, Dept of CMSE, Dept of Math

Office: 428 South Shaw Lane
Engineering Bldg, Rm 1511

TA: Emily Bolger bolgerem@msu.edu

Slack: cmse-courses.slack.com (Click for invite link).
Join the [#cmse381-f22](#) channel.

Office Hours: Dr. Munch: Tues 9:30-10:30, Thurs 3:30-4:30
Zoom and EGR 1511

Emily Bolger: Wed 10:30-noon, Fri 1-2:30.
Zoom and EGR (EGR 1508A)

All office hours will be held online only at the below zoom link.

Office Hours Zoom Link: [Meeting ID: 978 4406 7031](#), Passcode: 621092

Course Description:

CMSE 381 is an introduction course to statistical and machine learning. It covers the fundamentals of data science methods, including unsupervised learning and supervised learning, feature extraction, dimension reduction, clustering, regression and classification

Course Objectives:

Understand how to find patterns and structure in data using statistical methods and machine learning algorithms, and be able to implement simple versions of such techniques. Understand the concepts of unsupervised and supervised learning broadly, in addition to having proficiency in common algorithms and ideas within each field

Pre-requisites:

(STT 180 and MTH 314 and CMSE 201 and STT 380)

or

(STT 180 and MTH 314 and CMSE 201 and STT 441 and STT 442)

Websites:

There are three places you need to access to get content for this course:

- Content for the class will be posted on the the github page, <https://github.com/msu-cmse-courses/cmse381-F22>.
Nearly every class will have a skeleton of slides, as well as a jupyter notebook for coding portions.
- The D2L website, <https://d2l.msu.edu/d2l/home/1579786>, has your grades kept as up to date as possible.
- Announcements and discussions will be facilitated on the department Slack. See the link at the top of this syllabus to join the [cmse-courses.slack.com](#) workspace (you might already be a member), and look for the [#cmse381-f22](#) channel once you are there. Students who post their favorite gif to the thread in the slack channel get a bonus point on their first homework.

Textbook:

- *An Introduction to Statistical Learning: With Applications in R*, 2nd edition. James, Witten, Hastie and Tibshirani. Available as a FREE download at <https://statlearning.com/>.
- Some additional reading for those interested, although we will not have any assignments from this book: *The Elements of Statistical Learning*. Hastie, Tibshirani, and Friedman. Available as a FREE download at <http://www-stat.stanford.edu/ElemStatLearn>.

Grading:

- **Homework:** There will be approximately weekly homework sets given which will be turned in for a grade. This homework will be typeset and turned in as a single pdf or no credit will be given. In previous years, I have seen students have the most success with generating their homeworks using a jupyter notebook, and then using the “print” feature in the web browser to create a pdf to upload. I highly encourage including drawings and figures with your homeworks to aid in your explanation; note that this could be simply be a photograph of a hand-drawn figure inside of your file. Similarly, code can be included using screenshots and/or by being included in the jupyter notebook.

Homeworks are due on the days marked in the schedule at midnight. Homeworks submitted before midnight the following day will still be graded with a 5% penalty. Homeworks submitted before midnight two days after will still be graded with a 15% penalty. After midnight 2 days after, no credit will be given, but see the dropped grades policy below. For example, if homework is due on Friday, those turned in before midnight on Saturday have a 5% penalty; those turned in before midnight Sunday will have a 15% penalty; and no homework will be accepted after that. While late homework will not be accepted, please see the *missed work policy* section below for informations on automatically dropped grades.

The material presented draws from many disparate backgrounds, and it is highly unlikely any one student will have experience with all of it. To that end, I encourage students to work together, discuss the problems, and teach each other while struggling together. My collaboration policy is as follows.

- I **do** assume you will talk to each other to work on things.
- I **do** assume you will google definitions while you are working on things.
- I **do** assume you will include an acknowledgement section in your homework mentioning the people and resources you used in the course of answering problems.

Example: I worked with Maryam Mirzakhani and Grigori Perelman while completing this assignment. I also used wikipedia to understand Galois cohomology and stack overflow to get me started on an algorithm for enumerating Hamiltonian cycles in a graph.

- I **do not** assume you will copy each other’s work or copy from the internet.

- **Quizzes:** Once a week, there will be a short in class quiz on basics from the previous week’s course. It will be at the beginning of class and will last for about 10 minutes. The sorts of things I will put on quizzes include checking understanding definitions and interpretations of the content to ensure that we are all at the same place with the material. Note that no makeups will be given, see the missed work policy below.
- **Exams:** There will be three non-cumulative miterms given. See the schedule below for dates. There is no final for the course.

- **Missed Work Policy:** At the end of the semester, I will drop the two lowest homework grades; and the two lowest quiz grades. For this reason, there is no late homework accepted and no makeups are available for quizzes.

Students with an issue (i.e. medical, bereavement) needing to miss either the midterm or the final exam need to get in touch with Dr. Munch as soon as possible, preferably before the exam is administered. A time to make up the exam will be organized with the student.

- **Points:** Your grade will be based on the total number of accumulated points from the semester; please note however that things get canceled or moved sometimes so all totals are approximate. At the end of the semester, I will add up the total number of points accumulated as the denominator for the semester. While the number of homeworks and quizzes is subject to change, the *estimated* number of points is below.

| <i>Estimated Points</i> | |
|-------------------------|-----------------------------------------------------------|
| Homeworks | (10 homeworks - 2 lowest grades) \times 20 points = 160 |
| Quizzes | (12 Quizzes - 2 lowest grades) \times 10 points = 100 |
| Midterm | (3 Midterms) \times 100 = 300 |
| TOTAL: | 560 |

Course activities:

- There will very likely be times when we can not meet in person. This could be due to university mandated shutdowns. This could be due to exposure of myself or someone else in the class to Covid-19. This could be due to my two young kids' school shutting down due to the same reasoning. In these cases, I ask for your flexibility and graciousness as we navigate this new world together.

I will post an announcement to Slack about the change in location as soon as I know that it needs to change. D2L is harder to access in general, but I will try to post any announcements there as well. Whenever possible, I will be sure to post the change in location by noon before class. In the case that we need to meet online, we will use the same zoom link as office hours, posted at the top of this syllabus.

Required materials:

You will need to ensure that you have the following:

- A laptop to bring to class. We will be coding examples of topics nearly every day, so be sure to bring it with you. In the case that you do not have access to a laptop to bring to class every day, please speak to me as soon as possible so we can figure out alternative accommodations.
- You need to have access to the github page before every class, as I will be posting skeleton slides and jupyter notebooks prior to class. These should be posted before noon the day of class. If you want the ability to take notes on the slides, you need to be sure to either have access to a tablet, or print out the slides in advance of class. I will not be providing print-outs of the slide decks.

Tentative Schedule:

Note that this schedule is only provided as a guide and is very likely subject to change. I will also maintain the schedule as it changes at this link: bit.ly/3SE0Phl.

| Lec # | Date | Topic | Reading | Homeworks |
|-------|----------|----------------------------------------------------|---------------|------------|
| 1 | W Aug 31 | Intro / First day stuff / Python Review Pt 1 | 1 | |
| 2 | F Sep 2 | What is statistical learning? / Python Review Pt 2 | 2.1 | |
| | M Sep 5 | No class - Labor day | | |
| 3 | W Sep 7 | Assessing Model Accuracy | 2.2 | HW #1 Due |
| 4 | F Sep 9 | Linear Regression | 3.1 | |
| 5 | M Sep 12 | More Linear Regression | 3.2 | |
| 6 | W Sep 14 | Even more linear regression | 3.3 | HW #2 Due |
| 7 | F Sep 16 | Probably more linear regression | | |
| 8 | M Sep 19 | Intro to classification, Logistic Regression | 4.1, 4.2, 4.3 | |
| 9 | W Sep 21 | More logistic regression | | HW #3 Due |
| 10 | F Sep 23 | Review | | |
| 11 | M Sep 26 | Midterm #1 | | |
| 12 | W Sep 28 | [No class, Dr Munch out of town] | | |
| 13 | F Sep 30 | [No class, Dr Munch out of town] | | |
| 14 | M Oct 3 | Leave one out CV | 5.1.1, 5.1.2 | |
| 15 | W Oct 5 | k-fold CV | 5.1.3 | |
| 16 | F Oct 7 | More k-fold CV | 5.1.4 | |
| 17 | M Oct 10 | CV for classification | 5.1.5 | HW #4 Due |
| 18 | W Oct 12 | Resampling methods: Bootstrap | 5.2 | |
| 19 | F Oct 14 | Subset selection | 6.1 | |
| 20 | M Oct 17 | Shrinkage: Ridge | 6.2.1 | HW #5 Due |
| 21 | W Oct 19 | Shrinkage: Lasso | 6.2.2 | |
| 22 | F Oct 21 | Dimension Reduction | 6.3 | |
| | M Oct 24 | No class - Fall break | | |
| 21 | W Oct 26 | More dimension reduction; High dimensions | 6.4 | |
| 22 | F Oct 28 | Polynomial & Step Functions. | 7.1, 7.2 | HW #6 Due |
| 23 | M Oct 31 | Review | | |
| 24 | W Nov 2 | Midterm #2 | | |
| 25 | F Nov 4 | Basis functions, Regression Splines | 7.3, 7.4 | |
| 26 | M Nov 7 | Smoothing Splines; Local regression; GAMs | 7.5-7.7 | |
| 27 | W Nov 9 | Decision Trees | 8.1 | |
| 28 | F Nov 11 | Ensemble methods | 8.2 | HW #7 Due |
| 29 | M Nov 14 | Maximal Margin Classifier | 9.1 | |
| 30 | W Nov 16 | SVC | 9.2 | |
| 31 | F Nov 18 | SVM | 9.3, 9.4, 9.5 | HW #8 Due |
| 32 | M Nov 21 | More SVM | | |
| 33 | W Nov 23 | Single layer NN | 10.1 | |
| | F Nov 25 | No class - Thanksgiving | | |
| 35 | M Nov 28 | Multi Layer NN | 10.2 | HW #9 Due |
| 36 | W Nov 30 | CNN | 10.3 | |
| 37 | F Dec 2 | Unsupervised Learning & Clustering | 12.1, 12.4 | |
| 38 | M Dec 5 | More Clustering | 12.4 | HW #10 Due |
| 39 | W Dec 7 | Review | | |
| 40 | F Dec 9 | Midterm #3 | | |

Inclusive classroom behavior:

Respectful and responsible behavior is expected at all times, which includes not interrupting other students, refraining from non-course-related use of electronic devices or additional software during class sessions, and not using offensive or demeaning language in our discussions. Flagrant or repeated violations of this expectation may result in ejection from the classroom, grade-related penalties, and/or involvement of the university Ombudsperson. In particular, behaviors that could be considered discriminatory or harassing, or unwanted sexual attention, will not be tolerated and will be immediately reported to the appropriate MSU office (which may include the MSU Police Department).

In addition, MSU welcomes a full spectrum of experiences, viewpoints, and intellectual approaches because they enrich the conversation, even as they challenge us to think differently and grow. However, we believe that expressions and actions that demean individuals or groups comprise the environment for intellectual growth and undermine the social fabric on which the community is based. These demeaning behaviors are not welcome in this classroom.

Accommodations for Students with Disabilities:

(from the Resource Center for Persons with Disabilities (RCPD): Michigan State University is committed to providing equal opportunity for participation in all programs, services and activities. Requests for accommodations by persons with disabilities may be made by contacting the Resource Center for Persons with Disabilities at 517-884-RCPD or on the web at <http://rcpd.msu.edu>. Once your eligibility for an accommodation has been determined, you will be issued a Verified Individual Services Accommodation ("VISA") form. Please present this form to the instructor at the start of the term and/or two weeks prior to the accommodation date (test, project, etc.). Requests received after this date may not be honored.

Academic honesty:

Intellectual integrity is the foundation of the scientific enterprise. In all instances, you must do your own work and give proper credit to all sources that you use in your papers and oral presentations – any instance of submitting another person's work, ideas, or wording as your own counts as plagiarism. This includes failing to cite any direct quotations in your essays, research paper, class debate, or written presentation. The MSU College of Engineering adheres to the policies of academic honesty as specified in the General Student Regulations 1.0, Protection of Scholarship and Grades, and in the all-University statement on Integrity of Scholarship and Grades, which are included in Spartan Life: Student Handbook and Resource Guide. Students who plagiarize will receive a 0.0 in the course. In addition, University policy requires that any cheating offense, regardless of the magnitude of the infraction or punishment decided upon by the professor, be reported immediately to the dean of the student's college. (See also the Academic Integrity webpage.)

It is important to note that plagiarism in the context of this course includes, but is not limited to, directly copying another student's solutions to assignments; copying materials from online sources, textbooks, or other reference materials without citing those references in your source code or documentation, or having somebody else do your in-class work or homework on your behalf. Any work that is done in collaboration with other students should state this explicitly, and have their names as well as yours listed clearly.

More broadly, we ask that students adhere to the Spartan Code of Honor academic pledge, as written by the Associated Students of Michigan State University (ASMSU):

"As a Spartan, I will strive to uphold values of the highest ethical standard. I will practice honesty in my work, foster honesty in my peers, and take pride in knowing that honor is worth more than grades. I will carry these values beyond my time as a student at Michigan State University, continuing the endeavor to build personal integrity in all that I do."

Limits to confidentiality:

Essays, journals, and other materials submitted for this class are generally considered confidential pursuant to the University's student record policies. However, students should be aware that University employees, including instructors, may not be able to maintain confidentiality when it conflicts with their responsibility to report certain issues to protect the health and safety of MSU community members and others. As the instructor, I must report the following information to other University offices (including the Department of Police and Public Safety) if you share it with me:

- suspected child abuse/neglect, even if this maltreatment happened when you were a child,
- allegations of sexual assault or sexual harassment when they involve MSU students, faculty, or staff, and
- credible threats of harm to oneself or to others.

These reports may trigger contact from a campus official who will want to talk with you about the incident that you have shared. In almost all cases, it will be your decision whether you wish to speak with that individual. If you would like to talk about these events in a more confidential setting you are encouraged to make an appointment with the MSU Counseling Center.

Drops and Adds:

The last day to add this course is **September 7, 2022**. The last day to drop this course with a 100 percent refund and no grade reported is **Sept 23, 2022**. The last day to drop this course with no refund and no grade reported is **Oct 18, 2022**. See the registrar's page for up-to-date information regarding dates. You should immediately make a copy of your amended schedule to verify you have added or dropped this course.

Changes to Syllabus:

The syllabus may also be adjusted if needed. These types of changes will be announced during class, by email and/or in the course's slack channel.

Disruptive Behavior:

Article 2.III.B.4 of the Student Rights and Responsibilities (SRR) for students at Michigan State University states: "The student's behavior in the classroom shall be conducive to the teaching and learning process for all concerned." Article 2.III.B.10 of the SRR states that "The student and the faculty share the responsibility for maintaining professional relationships based on mutual trust and civility." General Student Regulation 5.02 states: "No student shall . . . interfere with the functions and services of the University (for example, but not limited to, classes . . .) such that the function or service is obstructed or disrupted. Students whose conduct adversely affects the learning environment in this classroom may be subject to disciplinary action through the Student Judicial Affairs office.

Grief Absence Policy:

Michigan State University is committed to ensuring that the bereavement process of a student who loses a family member during a semester does not put the student at an academic disadvantage in their classes. If you require a grief absence, you should complete the "Grief Absence Request" web form no later than one week after knowledge of the circumstance. I will work with you to make appropriate accommodations so that you are not penalized due to a verified grief absence.