CMSE 381: Fundamentals of Data Science Methods
Credit hours: 4

Class Location: (Section 001)
Olds Hall - Room 109
(Section 002)
Holden Hall - Room C135

Class Time: (Section 001)
MWF 3:30 - 4:40 PM
(Section 002)
MWF 12:30 - 1:40 PM

Instructor: (Section 001)
Guanqun Cao, Ph.D. (she/they) `caoguanq@msu.edu`
Assoc. Prof, Dept of CMSE, Dept of Statistics & Probability
(Section 002)
Mengsen Zhang, Ph.D. (she/they) `mengsen@msu.edu`
Assist. Prof, Dept of CMSE

Office: Dr. Cao:
C426 Wells Hall
Dr. Zhang:
428 South Shaw Lane
Engineering Bldg, Rm 1514

TA: (Section 001) Christy Lu (`luqiao@msu.edu`)
(Section 002) Siyu Guo (`guosiyu1@msu.edu`)

Slack: cmse-courses.slack.com (Click for invite link).
Join the `#cmse381-f25-sec1` or `#cmse381-f25-sec2` channel.

Office Hours: Dr. Cao: MW 2:15 - 3:15 pm
Zoom
Dr. Zhang: W 10 am - 12 pm
Zoom
TA: Siyu Guo: To be announced
Christy Lu: To be announced
Details on the calendar posted on the course webpage

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 four places you need to access to get content for this course:

- Content for the class will be posted on the course webpage:

  msu-cmse-courses.github.io/CMSE381-F25/

  which can also be accessed from cmse.msu.edu/CMSE381. Nearly every class will have a skeleton of slides, as well as a jupyter notebook for coding portions.

- The D2L website 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-f25-sec1` or `#cmse381-f25-sec1` 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.

- We use Crowdmark (`https://crowdmark.com/`) for grading. You will upload your homeworks here, and you will also have access to graded quizzes and tests from this website. Once we have the first homework assignment, you will get an invitation from the website to be able to upload your stuff.

Textbook:

- *An Introduction to Statistical Learning: With Applications in Python.* James, Witten, Hastie, Tibshirani, and Taylor. This is our primary textbook. Available as a FREE download at `https://statlearning.com/`.

- Some additional reading for those interested, although we will not have any assigments from this book: *The Elements of Statistical Learning.* Hastie, Tibshirani, and Friedman. Available as a FREE download at `http://www-stat.stanford.edu/ElemStatLearn`.

Basic Expectations:

- We expect you to attend each class for the full 70 min duration. During this time, we expect you to take detailed notes on, or beside, the skeleton slides provided before each class. We also expect you to complete the jupyter notebook.

- We expect you to read the assigned chapters in the text book as listed in the course schedule (see the most up-to-date version on course website).

- We expect you to actively participate in group work and interactive Q&A sessions.

- We expect you to complete all homework assignments, quizzes, exams, and a semester project.

Grading:

- **Homework:** There will be approximately weekly homework sets given which will be turned in for a grade. Your answer to each homework question will be submitted on Crowdmark. Some answers can be directly typed in. For some questions, you will need to upload your code and jupyter notebook outputs either as a pdf (using the "print" feature in the web browser) or as screenshot(s). We 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, we encourage students to work together, discuss the problems, and teach each other while struggling together. Our collaboration policy is as follows.

- We **do** assume you will talk to each other to work on things.
- We **do** assume you will google definitions while you are working on things.
- We **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.
- We **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 we 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 section for dates. There is no final for the course.

- **Project:**

  There will be a final project where you will take a publicly available data set, use tools learned in class to analyze the data, and submit a written report of your results. This project will be due at the end of the semester, but we will have several checkpoints throughout the semester.

- **Missed Work Policy:** At the end of the semester, we 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 your section instructor 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, we will add up the total number of points accumlated as the denomiantor for the semester. While the number of homeworks and quizzes is subject to change, the **estimated** number of points is below.

	Estimated Points
Homeworks	(9 homeworks - 2 lowest grades) $\times$ 20 points = 140
Quizzes	(10 Quizzes - 2 lowest grades) $\times$ 10 points = 80
Midterm	(3 Midterms) $\times$ 100 = 300
Final Project	100 points
TOTAL:	620

  Your final grade will calculated using the following scale:

Grade	% Points
4.0	≥ 90
3.5	≥ 85
3.0	≥ 80
2.5	≥ 75
2.0	≥ 70
1.5	≥ 65
1.0	≥ 60
0.0	< 60

Course activities:

- There may be times when we can not meet in person, e.g., when the instructors are sick. We will post an announcement to Slack about the change in location as soon as we know that it needs to change. D2L is harder to access in general, but we will try to post any announcements there as well. Whenever possible, we will be sure to post the change in location by noon before class. In the case that we need to meet online, the Zoom link will be added to the course calendar on the course website.

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 accomodations.

- You need to have access to the course website before every class, as we 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. we will not be providing print-outs of the slide decks.

Tentative Schedule:
Note that this schedule is only provided as a guide. What content is covered on which day is **very likely subject to change**. However, the dates of the midterms **will not change**. I maintain the schedule as it changes at this link:
msu-cmse-courses.github.io/CMSE381-F25/Course_Info/Schedule.html.

Lec #	Date		Topic	Reading	HW	Pop Quizzes	Notes
1	M	8/25	Intro / Python Review	1			
2	W	8/27	What is statistical learning	2.1		Q1	
3	F	8.29	Assessing Model Accuracy	2.2.1, 2.2.2			
	M	9/1	Labor Day - No Class				
4	W	9/3	Linear Regression	3.1		Q2	
5	F	9/5	More Linear Regression	3.1	HW #1 Due Sun 9/7		
6	M	9/8	Multi-linear Regression	3.2			
7	W	9/10	Probably More Linear Regression	3.3		Q3	
8	F	9/12	Last of the Linear Regression		HW #2 Due Sun 9/14		
9	M	9/15	Intro to classification, Bayes classifier, KNN classifier	2.2.3			
10	W	9/17	Logistic Regression	4.1, 4.2, 4.3.1-3		Q4	
11	F	9/19	Multiple Logistic Regression / Multinomial Logistic Regression	4.3.4-5	HW #3 Due Sun 9/21		
	M	9/22	***Project Day & Review***				
	W	9/24	**Midterm #1**				
12	F	9/26	Leave one out CV	5.1.1, 5.1.2			
13	M	9/29	k-fold CV	5.1.3			
14	W	10/1	More k-fold CV	5.1.4-5		Q5	
15	F	10/3	k-fold CV for classification	5.1.5			
16	M	10/6	Subset selection	6.1			
17	W	10/8	Shrinkage: Ridge	6.2.1			
18	F	10/10	Shrinkage: Lasso	6.2.2	HW #4 Due Sun 10/12		
19	M	10/13	PCA	6.3			
20	W	10/15	PCR	6.3		Q6	
	F	10/17	***Review***				
	M	10/20	Fall Break				
	W	10/22	**Midterm #2**				
21	F	10/24	Polynomial & Step Functions	7.1-7.2	HW #5 Due Sun 10/28		
22	M	10/27	Step Functions; Basis functions; Start Splines	7.2-7.4		Q7	
23	W	10/29	Regression Splines	7.4			
24	F	10/31	Decision Trees	8.1	HW #6 Due Sun 11/2		
25	M	11/3	Random Forests	8.2.1, 8.2.2			
26	W	11/5	Maximal Margin Classifier	9.1		Q8	
27	F	11/7	SVC	9.2	HW #7 Due Sun 11/9		
28	M	11/10	SVM	9.3, 9.4			
29	W	11/12	Single Layer NN	10.1		Q9	
30	F	11/13	Multi Layer NN	10.2	HW #8 Due Sun 11/16		
31	M	11/17	CNN	10.3			
32	W	11/19	Unsupervised learning / clustering	12.1, 12.4		Q10	
33	F	11/21	Virtual: Project Office Hours		HW #9 Due Sun 11/23		
	M	11/24	***Review***				
	W	11/26	**Midterm #3**				
	F	11/28	Thanksgiving				
	M	12/1	Virtual: Project Office Hours				
	W	12/3	Virtual: Project Office Hours				
	F	12/5			**Project Due**		
	M	12/8					
	W	12/10					
	F	12/12	**No final exam**		**Honors Project Due**		

Acceptable and Unacceptable Use of Generative AI:
This course is all about the use of machine learning, and so understanding the potential for use of generative AI tools such as ChatGPT, Dall-e, etc., is directly applicable to our learning objectives. That being said, the extremely fast emergence of this tool means that we are all still learning about how it can be used, and working to understand the use of this tool in an ethical manner, particularly in our classroom context. In class, we will draft a list of examples of how we as a class think that that these tools can and cannot be used for the purposes of this course.
Your use of AI tools must be properly documented and cited in order to stay within university policies on academic integrity and the Spartan Code of Honor Academic Pledge. **Clearly indicate where generative AI has been used in your assignments and provide proper citations for the AI tools and models you utilized, along with the prompts you used to generate content.** This can be done through comments in your code, annotations in your documentation, or a dedicated section in your homework or project reports. Properly citing both the tools and the specific prompts for each piece of generated content helps ensure transparency, reproducibility, and gives credit to the creators of the AI tools. Any assignment that is found to have used generative AI tools in unauthorized ways is subject to being given a 0. When in doubt about permitted usage, please ask for clarification.

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 August 29, 2025. The last day to drop this course with a 100 percent refund and no grade reported is September 18, 2025. The last day to drop this course with no refund and no grade reported is October 13, 2025. 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.