San José State University
Department of Mathematics & Statistics
Math 185, Learning from Large Data, Section 2, Spring 2022

Course and Contact Information

Instructor:	Guangliang Chen
Office Location:	MQH 413
Telephone:	(408) 924-5131
Email:	guangliang.chen@sjsu.edu
Office Hours:	TR 1-2:15pm (on Zoom, meeting ID: 422 306 1605) and by appointment.
Class Days/Time:	TR 10:30-11:45am
Classroom:	Sweeney Hall 347 (Before 2/14 on Zoom, meeting ID: 838 9188 0012, registration required)
Prerequisites:	Math 32, Math 39, and one of CS 22A, CS 46A, CMPE 30, or MATH 50.

Course Description

Linear algebra essentials, matrix computing in software, benchmark data sets in machine learning, data processing, plotting and visualization, dimension reduction (PCA, LDA), spectral clustering, linear regression, support vector machines, neural networks, and scalability of machine learning algorithms.

Faculty Web Page and MYSJSU Messaging

Course materials such as syllabus, handouts, notes, assignment instructions, etc. can be found on Canvas Learning Management System course login website at http://sjsu.instructure.com.

Piazza

The course uses Piazza at piazza.com/sjsu/spring2022/math185 as a venue for communication and discussions outside class. Please post all course-related questions on piazza for fastest response and broadest impact.

Course Learning Outcomes (CLO)

Upon successful completion of this course, students will be able to:

1. Load data sets from the internet to software and process them into matrices
2. Plot data sets in software to gain insights and create high-quality figures
3. Explain the objective and procedure of each of PCA and LDA, and perform dimension reduction on high dimensional data to visualize them in low dimensions
4. Describe the formulation of each of the machine learning algorithms for clustering, regression and classification, and apply them to given data sets in software
5. Analyze the scalability of machine learning algorithms such as spectral clustering and linear regression

Required Texts/Readings

There is no required textbook but learning material such as the instructor's lecture slides and relevant online resources will be provided from time to time, e.g, *Learning with Large Datasets*, Andrew Ng (Stanford). URL: https://www.coursera.org/lecture/machine-learning/learning-with-large-datasets-CipHf

Recommended readings

"An introduction to Statistical Learning with Applications in R", by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. Second Edition, Springer. Freely available at https://www.statlearning.com.

"Neural Networks and Deep Learning", by Michael Nielson, Determination Press, December 2019. Freely available at http://neuralnetworksanddeeplearning.com.

Other technology requirements / equipment / material

This course will make intensive use of specialized software (MATLAB) to carry out various computing tasks on large data sets. Familiarity with MATLAB is very helpful but not required.

Students are required to have an electronic device (laptop or desktop) with a camera and built-in microphone in order to attend the Zoom-based classes. Additionally, they should have access to a scanner (physical or cell phone app) in order to scan and submit their work.

Students are responsible for ensuring that they have access to reliable Wi-Fi during tests. If students are unable to have reliable Wi-Fi, they must inform the instructor, as soon as possible or at the latest one week before the test date to determine an alternative. See the Learn Anywhere website at https://www.sjsu.edu/learnanywhere/equipment/index.php for current Wi-Fi options on campus.

Any student that needs accommodations or assistive technology due to a disability should work with the Accessible Education Center (AEC), and notify the instructor at the earliest possible time.

Zoom Classroom Etiquette

Please arrive at each Zoom meeting on time. If you have to miss a class occasionally, please find out from the instructor or your classmates regarding what's said/done in that class and act accordingly.

Students are encouraged to turn on their cameras (when without privacy concerns) during each Zoom meeting. If using a virtual background, it should be appropriate and professional and should NOT suggest or include content that is objectively offensive or demeaning.

Please raise your hand to get the instructor's permission before you speak up in class. Alternatively, you can type your question or answer (when responding to the instructor's question) in the chat window.

To help keep background noise to a minimum, make sure you mute your microphone when not speaking.

Recording Policy

All lectures will be recorded and shared with the whole class; however, you should still make every effort to attend all classes. The recordings will be deleted at the end of the semester. If you would prefer to remain anonymous during these recordings, then please speak with the instructor about possible alternatives.

Students are not allowed to record without instructor permission: Students are prohibited from recording class activities (including class lectures, office hours, advising sessions, etc.), distributing class recordings, or posting class recordings. Materials created by the instructor for the course (syllabi, lectures and lecture notes, presentations, etc.) are copyrighted by the instructor. The university policy (S12-7) is in place to protect the privacy of students in the course, as well as to maintain academic integrity through reducing the instances of cheating. Students who record, distribute, or post these materials will be referred to the Student Conduct and Ethical Development office. Unauthorized recording may violate university and state law.

CoS Covid-19 Safety Policy

"All students registered for a College of Science (CoS) class with an in-person component must view the CoS COVID-19 Training slides and the SJSU Phased Adapt Plan website and acknowledge reading them according to their instructor's directions. By working together to follow these county and SJSU safety practices, we can keep our college safer. Students who do not follow COVID-19 Safety practice(s) outlined in the training, the SJSU Phased Adapt Plan, or instructions from their instructors, TAs or CoS Safety Staff may be dismissed from CoS buildings, facilities or field sites. Please review this training as needed throughout the semester, as updates will be implemented as changes occur (and posted to the same links)."

Course Requirements and Assignments

Course requirements include regular homework assignments, two midterm exams, and a final project.

Students are expected to attend all classes and actively participate in classroom discussions, as they often lead to a deeper understanding of the concepts and better course grades.

The homework assignments will typically contain both math and coding questions. For the mathematical questions, you must show steps to earn full credit, while for the programming questions, you need to present your results in an organized way, interpret them carefully, and attach the code you used to obtain the results.

The midterms will be closed-book and closed-notes; more information will be given in class.

Success in this course is based on the expectation that students will spend, for each unit of credit, a minimum of 45 hours over the length of the course (normally three hours per unit per week) for instruction, preparation/studying, or course related activities, including but not limited to internships, labs, and clinical practica. Other course structures will have equivalent workload expectations as described in the syllabus.

Final Examination or Evaluation

This course ends with a data analysis project that aims to provide students with an opportunity to practice and apply the methods learned in class to data sets from the internet.

The class will be divided into groups of size two to work on the projects. In special circumstances, a single-person group can be approved by the instructor.

The data sets used by different groups must be distinct. Each data set must have at least 5000 instances and 10 features, and requires advanced approval of the instructor. It is advised that you select a data set as early as possible, because data sets will be available on a first-propose, first-get basis and you also need enough time to complete your project.

You will be asked to report your results through a short oral presentation in class and meanwhile submit a report that contains all the details:

- Your presentation needs to present a high-level summary of your work but you should still give some necessary specifics, such as data information and parameter values for certain algorithms. It should be clear, organized, logical, and self-sustained. We will reserve the final exam day for your presentations.
- Your report must be written using your own language (copying from other places is strictly prohibited and will be given zero points). In addition, it needs to contain a clear structure with the following components: Title, Author, Abstract, Introduction, Experiments, Conclusions (or Discussions), and References. Your report will also be due on the scheduled final exam day.

Your presentation and report will be graded based on clarity, completeness, correctness and originality.

Grading Information

Students may collaborate on homework but must write their own codes and solutions independently. Copying and other forms of cheating will not be tolerated and will be reported to the SJSU Office of Student Conduct.

Students must submit homework on time to receive full credit. Late submissions within 24 hours of the due time can still be accepted but will receive a penalty of 10% of the total number of points. Submissions that are late for more than one day (24 hours) will not be accepted for any reason.

No make-up exam will be given if any student misses a midterm exam, unless the student has a legitimate excuse (such as illness or other personal emergencies) and can provide documented proof.

For both homework and tests, it is the student's work (in terms of correctness, completeness, and clarity), not just the answer, that is graded. Thus, correct answers with no or poorly written supporting steps may receive very little credit.

The grading scale and weights of the assignments (for determining the semester average) are:

- Homework: 25%
- Midterm 1: 25%
- Midterm 2: 30%
- Final project: 20%

The following cutoffs will be used for assigning students' course grades (however, the instructor reserves the right to slightly adjust these percentages in order to better reflect the actual distribution of the class in the end):

A+: 98% to 100%	B+: 86% to 89%	C+: 73% to 75%	D+: 63% to 64%	F: 0% to 55%
A: 93% to 97%	B: 80% to 85%	C: 68% to 72%	D: 58% to 62%	
A-: 90% to 92%	B-: 76% to 79%	C-: 65% to 67%	D-: 56% to 57%	

Classroom Protocol

- The class starts on time, so please do not be late.
- If you miss a class, you are responsible for finding out what's said/done in that class (such as new announcement, deadline change, etc.) and responding accordingly.
- Please make sure to turn off or mute your cell phone during class.
- Please do not perform irrelevant or distracting activities in class.
- Academic dishonesty at any level is not tolerated and will be surely reported to the Office of Student Conduct and Ethical Developments.

University Policies

Per University Policy S16-9 *(http://www.sjsu.edu/senate/docs/S16-9.pdf)*, relevant university policy concerning all courses, such as student responsibilities, academic integrity, accommodations, dropping and adding, consent for recording of class, etc. and available student services (e.g. learning assistance, counseling, and other resources) are listed on Syllabus Information web page (https://www.sjsu.edu/curriculum/courses/syllabus-info.php). Make sure to visit this page to review and be aware of these university policies and resources.

Disclaimer: *The instructor reserves the final right to interpret, and make changes to, all the policies that are stated in this course syllabus.*

Math 185 Learning from Large Data, Tentative Course Schedule

Week	Date	Topics, Readings, Assignments, Deadlines
1	Jan 27	Introduction and overview
2	Feb 1	Review of symmetric matrices and positive definiteness
2	3	Rayleigh quotient
3	8	Singular value decomposition
3	10	Matrix computing in software
4	15	Benchmark data sets in machine learning
4	17	Data plotting and visualization
5	22	Principal component analysis
5	24	Principal component analysis
6	Mar 1	Linear discriminant analysis
6	3	Linear discriminant analysis
7	8	Backup / Review
7	10	Midterm 1
8	15	Spectral clustering
8	17	Spectral clustering
9	22	Scalable spectral clustering for large data sets
9	24	Applications of spectral clustering
March 28 – April 1 Spring break (no class)		
10	Apr 5	Linear regression
10	7	Linear regression
11	12	Linear regression via (stochastic) gradient descent
11	14	Support vector machines
12	19	Support vector machines
12	21	Support vector machines on high dimensional data
13	26	Backup / Review
13	28	Midterm 2 (comprehensive)
14	May 3	Neural networks
14	5	Neural networks
15	10	Neural networks
15	12	Transfer learning
	20	Project presentations (Friday, 9:45am-12pm) Reports due 11:59pm