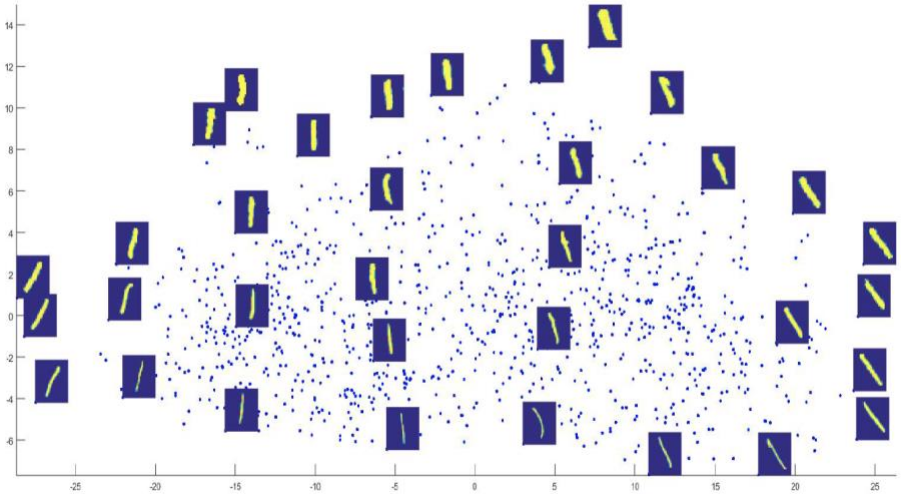
San José State University
Department of Mathematics & Statistics

Math 250 Mathematical Data Visualization, Spring 2023



Course and Contact Information

Instructor:	Dr. Guangliang Chen
Office:	MH 417
Email:	guangliang.chen@sjsu.edu
Class Days/Time:	MW 12 - 1:15pm
Classroom:	MQH 235
Office Hours:	M 10:30 - 11:30am, W 1:30 - 3pm, and by appointment
Prerequisites:	Math 32 and Math 39 (each with a grade of B or better), Math 163 (C or better)

Faculty Web Page and MYSJSU Messaging

Course materials such as syllabus, lecture slides, and reading material can be found on the course page at http://www.sjsu.edu/faculty/guangliang.chen/Math250.html. Assignments and grades will be posted on Canvas Leaning Management System course login website at http://sjsu.instructure.com.

Piazza

The course uses Piazza at https://piazza.com/sjsu/spring2023/math250 as a venue for communication outside class. Please post all course-related questions on piazza for fastest response and broadest impact.

Catalog Description

Matrix computing in software, data plotting in 3 dimensions, advanced linear algebra, dimensionality reduction, visualization of high dimensional data, and spectral clustering. 3 units.

Course Learning Outcomes (CLO)

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

- Use software to carry out various linear algebra operations
- Plot and visualize data of various types in software and create publication-quality graphics
- Perform matrix singular value decomposition and other advanced linear algebra operations
- Apply dimensionality reduction to high dimensional data and visualize them in low dimensions
- Develop a basic understanding of the machine learning task of clustering

Required Texts/Readings

None, but instructor's notes for all listed topics will be provided.

Recommended Readings

1. *Probabilistic Machine Learning: An Introduction*, by Kevin Patrick Murphy. MIT Press, March 2022. A draft copy is freely available at https://probml.github.io/pml-book/book1.html.
2. *Foundations of Data Science*, Avrim Blum, John Hopcroft, and Ravindran Kannan. Cambridge University Press; 1st edition (January 1, 2020). An unofficial January 2018 version of the book is publicly available from the authors' website at https://www.cs.cornell.edu/jeh/book.pdf.

Technology and Equipment Requirements

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

Additionally, they should have access to a scanner (physical or cell phone app) in order to submit their work to Canvas electronically.

Course Requirements and Assignments

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

Students are expected to attend all classes and actively participate in class discussions, as they often lead to a deeper understanding of the concepts and are also strongly associated with course grade.

The homework assignments will typically contain both theory and coding questions. For the theory questions, you must show steps to earn full credit, while for the programming questions, you need to present your results in an organized, meaningful 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 provided in class later.

Final Examination or Evaluation

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

The class will be divided into groups of size two to work on the projects.

The data sets used by different groups must be distinct. Each data set must have at least 5000 instances and 10 features (with at least a categorical feature), 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 depth. The grading rubric will be provided later in the course.

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.

You 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.

The instructor aims to give 12 homework assignments, and your lowest homework score will be dropped.

No make-up exam will be given if you miss a midterm exam, unless you have a legitimate excuse (such as illness or other personal emergencies) and can provide documented proof.

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

The weights in determining the semester average are:
- Homework: 20%
- Midterm 1: 30%
- Midterm 2: 35%
- Final project: 15%

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%	

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 250 Tentative Course Schedule, Spring 2023

Week	Date		Topics	Assignments
1	W	January 25	Course introduction and overview	Hw0, Matlab Onramp
2	M	30	Matrix algebra	
2	W	February 1	Matrix algebra	Hw1
3	M	6	Matrix computing in MATLAB	Hw2
3	W	8	Benchmark data sets in machine learning	
4	M	13	Data plotting and visualization in 3D	Hw3
4	W	15	Rayleigh quotient	
5	M	20	Rayleigh quotient	Hw4
5	W	22	Singular value decomposition (SVD)	
6	M	27	Singular value decomposition (SVD)	Hw5
6	W	March 1	Generalized inverse and pseudoinverse	
7	M	6	Generalized inverse and pseudoinverse	Hw6
7	W	8	Matrix norm and low-rank approximation	
8	M	13	Matrix norm and low-rank approximation	Hw7
8	W	15	Backup / Review	
9	M	20	Midterm 1	
9	W	22	Principal component analysis (PCA)	
March 27 – 31: Spring Recess (no classes)				
10	M	April 3	Principal component analysis (PCA)	
10	W	5	Principal component analysis (PCA)	Hw8
11	M	10	Linear discriminant analysis (LDA)	
11	W	12	Linear discriminant analysis (LDA)	Hw9
12	M	17	Multidimensional scaling (MDS)	
12	W	19	ISOmap	Hw10
13	M	24	Backup / Review	
13	W	26	Midterm 2 (comprehensive)	
14	M	May 1	Laplacian Eigenmaps	
14	W	3	Laplacian Eigenmaps	Hw11
15	M	8	Spectral clustering	
15	W	10	Scalable spectral clustering	
16	M	15	Last class	
	W	17	Project presentations (9:45am – 12pm) Reports due 11:59pm	