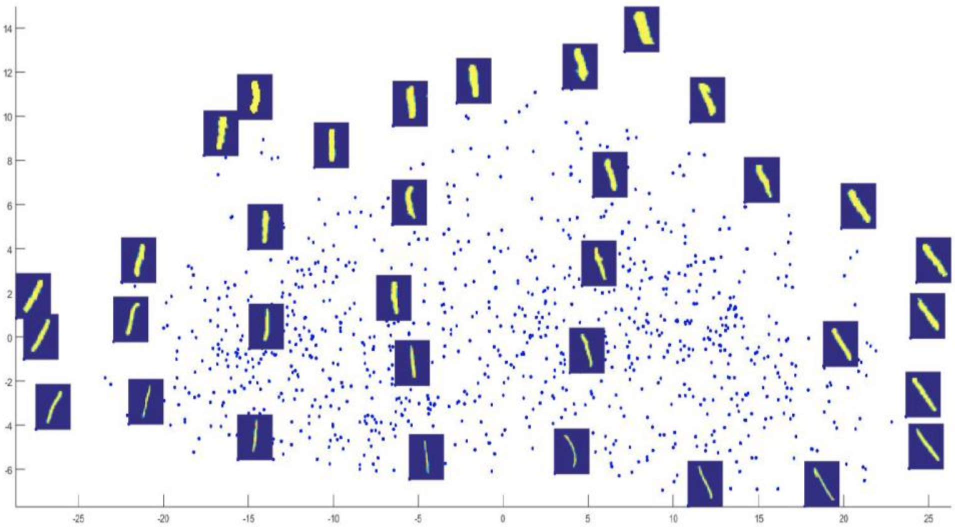
San José State University
Department of Mathematics & Statistics

Math 253: Mathematical Methods for Data Visualization, Section 1, Spring 2020



Course and Contact Information

Instructor:	Guangliang Chen
Office Location:	MQH 417
Telephone:	(408) 924-5131
Email:	guangliang.chen@sjsu.edu
Office Hours:	TR 10:20-11:50am
Class Days/Time:	TR 12:00-1:15pm
Classroom:	MH 223
Prerequisites:	Math 32 and Math 129A (each with a grade of B or better), and Math 163

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/Math253S20.html. Assignments and grades will be posted on Canvas Leaning Management System course login website at http://sjsu.instructure.com.

Piazza

The course will use Piazza as a venue for communication and discussions outside of the class meetings. Please post all course-related questions on piazza for fastest response and broadest benefit.

Course Description

Programming basics; data plotting and graphing in 3D or less; advanced linear algebra; dimensionality reduction; visualization of high dimensional data; and applications to clustering and classification. 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 and statistical computing tasks
- Create publication-quality pictures and graphs
- Perform matrix singular value decomposition and other advanced linear algebra operations
- Apply various dimensionality reduction techniques to high dimensional data and visualize them in low dimensions
- Acquire first-hand experience with large data sets
- Develop a basic understanding of machine learning tasks such as clustering and classification

Required Texts/Readings

Required Textbook

Foundations of Data Science, by Avrim Blum, John Hopcroft, and Ravindran Kannan. New book (upcoming).

An unofficial January 2018 version of the book is publicly available from the authors' website at https://www.cs.cornell.edu/jeh/book.pdf, which we will use for this course. Other learning material such as lecture slides, papers, and websites will be provided from time to time in class.

Other technology requirements / equipment / material

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

Students taking this course will be asked to use the following data for learning and practice:

- MNIST Handwritten Digits (available at http://yann.lecun.com/exdb/mnist/), which consists of 70,000 digital images of size 28x28 of handwritten digits 0…9 collected from about 250 people
- Fashion-MNIST (available at https://github.com/zalandoresearch/fashion-mnist), which resembles the MNIST data set in all ways except for different contents
- USPS Zip Code Data (available at http://statweb.stanford.edu/~tibs/ElemStatLearn/data.html), which consists of 9,300 size 16x16 grayscale images of handwritten digits scanned from envelops
- 20 Newsgroups (available at http://qwone.com/~jason/20Newsgroups/), consisting of about 19,000 text documents that are divided into 20 groups (according to their topics)

Smaller data sets such as those from the UCI Machine Learning Repository (at http://archive.ics.uci.edu/ml/) will also be used for teaching demonstration and homework assignments.

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 classroom 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 involve both theory and programming. 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.

The midterms will be closed-book; but cheat sheets of certain size will be allowed. More information will be given by the instructor in class.

Final Examination or Evaluation

This course ends with an individual project that aims to provide students with the culminating experience and additional learning opportunities.

Examples of a good project for this course are the following (not an exhaustive list):

- Presenting a dimensionality reduction algorithm that is not covered in class. You must describe the new method clearly and with sufficient detail, and demonstrate it on both toy and real data.
- Nontrivial improvement of a dimensionality reduction algorithm learned in this course. You must demonstrate the performance of your implementation on data sets, and compare with the old implementation.
- A nontrivial application of a method learned in this course to a large, interesting data set with explorations of different options and parameter settings.
- An empirical study of several algorithms using a few data sets to study their strength and weakness and compare their performance.
- Proving a nontrivial theoretical result that is not done in class.

The topic and content of the final project is to be determined between each of you and the instructor. You will need to submit a 1-page proposal about one month before the end of semester that describes what you intend to do in your project, and you must get the approval from the instructor for the project you work on.

To maximize the chance of your proposal getting approved, make sure that it is as clear as possible and provides all necessary information for evaluation, such as:

- The title of your project;
- Description of the problem
- Your proposed work with clearly stated goals
- Data sets that you plan to use (with urls)
- Reference papers that are relevant to your project;
- Significance of your proposed work, e.g., potential applications.

It is advised that you select a project as early as possible, because projects will be available on a first-propose, first-get basis and you also need enough time to complete your project. Your proposal will be referred to when grading your project presentation and report.

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 by focusing on the main ideas but you should still give all necessary specifics, like parameter values, etc. 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, Your proposed method or study, 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 must submit homework on time to receive full credit (late homework will receive a 20% penalty for each extra day). Your lowest homework score will be dropped.

No make-up exam will be given, unless there is a legitimate excuse such as illness or personal emergencies (with documented proof).

Student must show all their work for both homework and the midterm. Note that it is their 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 weights in determining the semester average are:

- Homework: 20%
- Midterm 1: 25%
- Midterm 2: 33%
- Final project: 22%

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+: 97% to 100%	B+: 85% to 89%	C+: 72% to 75%	D+: 62% to 65%	F: 0% to 55%
A: 93% to 97%	B: 80% to 85%	C: 68% to 72%	D: 58% to 62%	
A-: 90% to 93%	B-: 75% to 80%	C-: 65% to 68%	D-: 55% to 58%	

University Policies

Per University Policy S16-9 *(http://www.sjsu.edu/senate/docs/S16-9.pdf)*, relevant information to all courses, such as academic integrity, accommodations, dropping and adding, consent for recording of class, etc. is available on Office of Graduate and Undergraduate Programs' Syllabus Information web page at http://www.sjsu.edu/gup/syllabusinfo/".

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

Math 253 Mathematical Methods for Data Visualization, Course Schedule

This schedule is subject to change with fair notice which will be made in class if there is a delay in progress.

Week	Date		Topics
1	January	23	Course introduction and overview
2		28	Review of linear algebra
2		30	Matrix computing in MATLAB
3	February	*4	High quality data plotting in 3D
3		6	High quality data plotting in 3D
4		11	High quality data plotting in 3D

Week	Date		Topics
4		13	Spectral decomposition of symmetric matrices
5		18	Singular value decomposition (SVD)
5		20	Singular value decomposition (SVD)
6		25	Moore-Penrose pseudoinverse and Rayleigh quotient
6		27	Matrix norm and low-rank approximation
7	March	3	Applications
7		5	Review
8		10	Midterm 1
8		12	Principal component analysis (PCA)
9		17	Principal component analysis (PCA)
9		19	Linear discriminant analysis (LDA)
10		24	Linear discriminant analysis (LDA)
10		26	Multidimensional scaling (MDS)
	March 30 – April 3:		*Spring Recess*
11	April	7	ISOmap
11		9	Locally linear embedding (LLE)
12		14	Laplacian Eigenmaps
12		16	Review
13		21	Midterm 2 (cumulative)
13		**23	Data clustering
14		28	Data clustering
14		30	Data classification
15	May	5	Data classification
15		7	Last class
		13	Project presentations (9:45am-12pm)
	Tests & project		**Midterms: Weeks 8 and 13** **Project proposal: due Week 11** **Project presentations: due Final Exam Day** **Project report: due Final Exam Day**

*Last day to drop without a W grade
**Semester withdrawal deadline