MATH 250: Mathematical Data Visualization

Course overview: data as representation

Peter A. Gao

January 27, 2025

San José State University

What is the focus of this course?

- **How do we use vectors and matrices to represent images/networks/other forms of information as data?**
  - *data as representation*

- **How can we summarize and visualize high-dimensional datasets?**
  - *linear algebra, dimensionality reduction, data visualization*

- **How do we extract insights from large, often unstructured, datasets?**
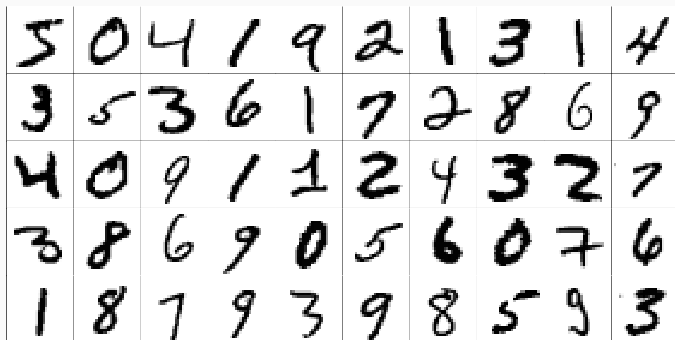 - *pattern recognition, data mining, classification, clustering*

Figure 1: Handwritten digits from MNIST dataset [1]

Example: handwritten digits

```
       [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24] [,25] [,26] [,27] [,28]
 [1,]    0    0    0    0    0    0    0    0    0    0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
 [2,]    0    0    0    0    0    0    0    0    0    0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
 [3,]    0    0    0    0    0    0    0    0    0    0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
 [4,]    0    0    0    0    0    0    0    0    0    0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
 [5,]    0    0    0    0    0    0    0    0    0    0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
 [6,]    0    0    0    0    0    0    0    0    0    0     0     0     3    18    18    18   126   136   175    26   166   255   247   127     0     0     0     0
 [7,]    0    0    0    0    0    0    0   30   36   94   154   170   253   253   253   253   253   225   172   253   242   195    64     0     0     0     0     0
 [8,]    0    0    0    0    0    0    0   49  238  253   253   253   253   253   253   253   253   251    93    82    82    56    39     0     0     0     0     0
 [9,]    0    0    0    0    0    0    0   18  219  253   253   253   253   253   198   182   247   241     0     0     0     0     0     0     0     0     0     0
[10,]    0    0    0    0    0    0    0    0   80  156   107   253   253   205    11     0    43   154     0     0     0     0     0     0     0     0     0     0
[11,]    0    0    0    0    0    0    0    0    0   14     1   154   253    90     0     0     0     0     0     0     0     0     0     0     0     0     0     0
[12,]    0    0    0    0    0    0    0    0    0    0     0   139   253   190     2     0     0     0     0     0     0     0     0     0     0     0     0     0
[13,]    0    0    0    0    0    0    0    0    0    0     0    11   190   253    70     0     0     0     0     0     0     0     0     0     0     0     0     0
[14,]    0    0    0    0    0    0    0    0    0    0     0     0    35   241   225   160   108     1     0     0     0     0     0     0     0     0     0     0
[15,]    0    0    0    0    0    0    0    0    0    0     0     0     0    81   240   253   253   119    25     0     0     0     0     0     0     0     0     0
[16,]    0    0    0    0    0    0    0    0    0    0     0     0     0     0    45   186   253   253   150    27     0     0     0     0     0     0     0     0
[17,]    0    0    0    0    0    0    0    0    0    0     0     0     0     0     0    16    93   252   253   187     0     0     0     0     0     0     0     0
[18,]    0    0    0    0    0    0    0    0    0    0     0     0     0     0     0     0     0   249   253   249    64     0     0     0     0     0     0     0
[19,]    0    0    0    0    0    0    0    0    0    0     0     0     0     0     0    46   130   183   253   253   207     2     0     0     0     0     0     0
[20,]    0    0    0    0    0    0    0    0    0    0     0     0     0    39   148   229   253   253   253   250   182     0     0     0     0     0     0     0
[21,]    0    0    0    0    0    0    0    0    0    0     0     0    24   114   221   253   253   253   253   201    78     0     0     0     0     0     0     0
[22,]    0    0    0    0    0    0    0    0   23   66   213   253   253   253   253   198    81     2     0     0     0     0     0     0     0     0     0     0
[23,]    0    0    0    0    0    0   18  171  219  253   253   253   253   195    80     9     0     0     0     0     0     0     0     0     0     0     0     0
[24,]    0    0    0    0   55  172  226  253  253  253   253   244   133    11     0     0     0     0     0     0     0     0     0     0     0     0     0     0
[25,]    0    0    0    0  136  253  253  253  253  212   135   132    16     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
[26,]    0    0    0    0    0    0    0    0    0    0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
[27,]    0    0    0    0    0    0    0    0    0    0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
[28,]    0    0    0    0    0    0    0    0    0    0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
```

Figure 2: Handwritten digit from MNIST dataset, as matrix

Each handwritten digit is digitized and represented using a 28 x 28 pixel grid of numbers between 0 (white) and 255 (black).

We can represent each image using a 28 x 28 matrix, or a 1 x 784 row vector. In other words, each image is a point in the 784-dimensional space $\mathbb{R}^{784}$.

This course focuses on the task of **dimensionality reduction**. Many modern datasets have hundreds/thousands/millions of dimensions/features/variables.

Reducing dimension **can**:

- save run time and storage space
- reduce noise and multi-collinearity, facilitating statistical analysis
- facilitate data visualization

What if we wish to develop an algorithm for classifying our digits?

Working in the original 784-dimensional space may not be necessary (for example, the corners of each image are almost always 0 and not very informative).

Feature selection approaches examine a subset of the original dimensions/features.

- ex. subset selection, ridge regression, LASSO

This course will focus on **feature extraction**: constructing a low-dimensional representation of the original high-dimensional data

- ex. principal component analysis, manifold learning

Instead of working in $\mathbb{R}^{784}$, what if we work in a lower-dimensional space?

In this class, we'll learn what it means to construct a lower-dimensional representation of a high-dimensional object and then visualize our representations.

Note that when we digitize handwritten digits, we must choose a resolution (28 x 28 in this case). The higher the resolution, the more information is contained in our digital representation.

In this sense, our data are typically just representations of some real world objects. Lowering the "resolution" of our representation loses information but typically saves space and can simplify analysis.

Figure 3: Coastline paradox, illustrated by Alexandre Van de Sande.

Dimensionality reduction facilitates many common statistical and machine learning tasks:

- **visualization**
- regression (MATH 261A)
- classification (MATH 251)
- clustering (MATH 252)

We will touch on all of these, but our focus will be on linear algebra and visualization tools that are foundational for machine learning.

Course overview

This course covers **matrix methods and linear algebra** for machine learning and high-dimensional data analysis.

Our focus will be on **dimension reduction**, as applied to **data visualization**.

We will also cover computation and tools for visualization in R.

This course is used in the following ways:

- Prerequisite for **MATH 251 Statistical and Machine Learning Classification**
- Elective for the regular **MS Statistics** program
- Required for the **MS Statistics Machine Learning Specialization**
- Required for the **MS Data Science** program (joint with CS)

Prerequisites:

- **Math 32**: Multivariate calculus
- **Math 39**: Linear algebra
- **Math 163**: Probability theory

In addition, though there is no formal programming prerequisite, you are expected to solve problems and make visualizations using R.

Coursework

- **Homework (25%)**: Roughly two weeks per assignment, involving both problem sets (proofs) and R labs.
- **Exams**: Two midterm exams (**25% each**)
- **Final project (25%)**: Final report and presentation during last week of class

Late policy

In general, the late policy is as follows:

- Any assignment that is received late but less than 24 hours late will receive a penalty of 25%.
- Any assignment that is received 24–48 hours late will receive a penalty of 50%.
- Assignments will not be accepted more than 48 hours late.

Extensions for academic purposes (ex. conference presentation or job interview) or extreme circumstances (ex. illness, emergency) will generally be granted. Email me at least 24 hours before an assignment is due to request an extension, along with the reason.

- You may discuss problems, approaches, and solutions with your classmates.
- You must credit anyone with whom you worked on each assignment.
- All submitted work must be your own; you should not submit code or answers copied from any resource including your classmates.
- If you have any questions, ask!

Students may use external online resources including discussion forums (ex. StackOverflow) large language model-based chatbots (ex. ChatGPT) as aids for learning and understanding course material.

However, submitting code or answers for course assignments obtained using resources like StackOverflow or ChatGPT is **not permitted**.

When external resources are consulted for an assignment, they must be cited clearly at the top of your submission.

Key advances in machine learning and LLMs rely on applications of linear algebra and optimization–building a strong foundation in these topics may be important for your personal and career goals.

Even if LLMs can be a useful tool for your learning, remember: there is no substitute for practicing your problem-solving and developing your mathematical intuition.

Exams

There will be two in-class midterm exams, both closed note and closed book.

The second midterm will be cumulative (essentially an early final).

These exams will test your understanding of the **linear algebra theory and technical details of the dimension reduction methods** covered in class.

Practice problems will be provided before the exams.

Final project

This course culminates with a project in which you will explore a dataset or dimension reduction technique of your choice.

You will prepare a brief oral presentation and write a final report.

Examples:

- Present a dimension reduction method not covered in class.
- Apply a dimension reduction method covered in this course to a real-world dataset and interpret the results.
- Compare two dimension reduction methods using a real-world dataset or simulated data.
- Extend an existing dimension reduction method or review a theoretical result/paper not discussed in class.

If you ever have questions about materials, please contact me.
If you need any kind of accommodations, please let me know as soon as possible.

- **Course website**: Course slides, assignment instructions.
- **Canvas**: Official syllabus, receiving grades, data.
- **Gradescope**: Submitting assignments.

Required Texts:

- Strang, G. (2019). Linear Algebra and Learning From Data. Available for free via Library Course Reserves.

- Murphy, K. P. (2022). Probabilistic Machine Learning: An Introduction. Available for free via Library Course Reserves.

- Murphy, K. P. (2023). Probabilistic Machine Learning: Advanced Topics. Available for free via Library Course Reserves.

Course overview

Suggested texts:

- Fessler, J.A. and Nadakuditi, R. R. (2024). Linear Algebra for Data Science, Machine Learning, and Signal Processing.
- Baker, J. D. (2024). Applied Multivariate Statistics in R. Available for free here.
- Healy, K. (2019). Data Visualization: A Practical Introduction. Draft version available here.
- Strang, G. (2023). Introduction to Linear Algebra, Sixth Edition.

Texts:

- **Math 32**: Multivariate calculus
- **Math 39**: Linear algebra
- **Math 163**: Probability theory

In addition, though there is no formal programming prerequisite, you are expected to solve problems and make visualizations using R.

Your responsibilities

This course may be challenging and demanding. We are combining difficult theory (linear algebra) and applications (dimension reduction and visualization).

However, this material is crucial for establishing the theoretical and applied foundations of machine learning and high-dimensional statistics.

As such, you should strive to:

- Attend all classes
- Participate in class discussion
- Read the textbooks before and after class
- Think carefully through homework problems
- **Ask questions early and often**

Office hours: MW 10:30-11:30 or by appointment in MH426 (or Zoom, by appointment)

Canvas: Discussion board; up to 2% extra credit for participation on the discussion board.

Email: peter.gao [at sjsu]. Feel free to send me a reminder after 48 hours have passed. Please include [MATH 250] in your subject line.

Examples

- Image compression and deblurring
- Regression for high-dimensional data
- Clustering and classification of high-dimensional data
- Dimension reduction for visualization
- Topic modeling and LLM embeddings for text data

Original (2419 kb) Rank-50 approximation (276 kb)
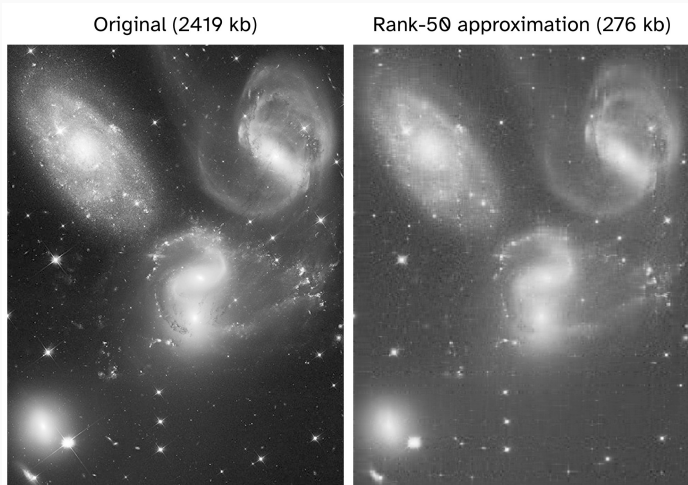
Figure 4: Stephan's quintet
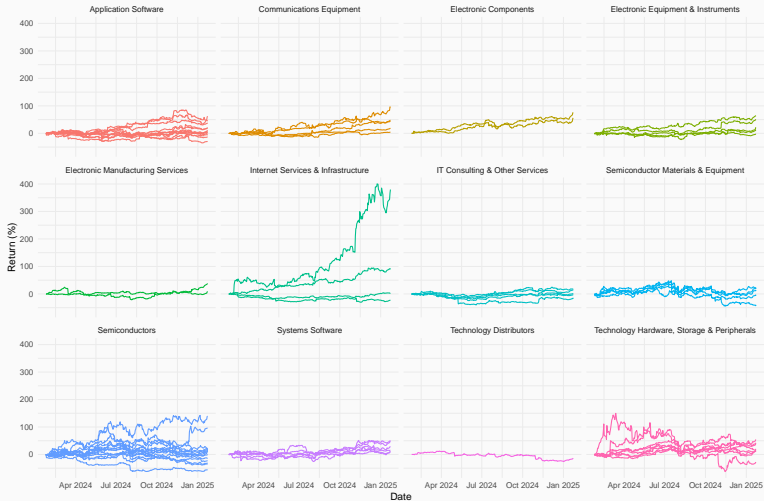
Figure 5: Performance of IT stocks, 2024-2025

	AAPL	ACN	ADBE	ADI	ADSK	AKAM	AMAT
1	0.00000000	0.00000000	0.0000000	0.00000000	0.000000000	0.0000000	0.0000000
2	−0.90712209	−0.23390738	−1.3893836	−1.70806403	0.524332601	−0.4852063	−3.3192439
3	−1.25663206	0.37909438	1.2287518	−0.05068733	2.176767025	0.7035380	−2.4039906
4	−3.15702989	−0.17207470	0.8641467	−1.29751521	1.874876354	0.8086689	−3.7015578
5	−5.03167552	−2.16701551	−0.7709833	−2.50380252	0.818271126	−0.3477279	−4.8253499
6	−3.76474101	−0.27692606	0.8561078	−2.68120044	1.318771529	0.3072920	−3.2786906
7	−4.28490093	0.27155180	1.9563739	−2.63557875	1.866936329	0.6065017	−2.5777745
8	−3.34243475	−0.43017855	1.2721229	−2.40243559	0.421051662	0.7682330	−0.8920863
9	−2.50810896	−1.94654788	−2.4800029	−3.84693169	0.885797707	0.6631083	−2.2765497
10	−2.45145725	−1.42227470	−1.0809922	−2.70147073	2.653423109	1.5445544	−1.0021496
11	−3.01281913	−0.93025526	−1.0793844	−1.64217204	3.876865845	2.5230430	0.7298815
12	−2.73986307	−0.07258940	0.7436803	−1.15559997	5.930483685	3.7684001	7.6521990
13	−3.61539081	−1.21256378	−1.7250779	−1.64723768	4.087391676	3.6309155	7.4784151
14	−4.70206776	−2.45469829	−3.3232680	−6.39128252	1.990073393	1.1240493	4.4488168
15	−5.16042868	−0.25810385	−2.8783519	−5.56512699	4.139032146	−7.1648070	7.8549483
16	−5.30977888	−0.11560807	−5.1623910	−4.94678642	3.773584906	−9.7768103	8.7064811
17	−6.10805002	−0.66139470	−12.1944235	−4.59199058	2.601782639	−11.7014401	15.6056315
18	−6.49430949	−2.41705387	−12.9573776	−4.00405932	1.394238136	−12.3726367	9.5638035
19	−6.10289486	−2.96553173	−13.5018785	−1.81449659	0.305860684	−13.3430370	10.2531405
20	−5.04712527	−0.16938347	−13.6544713	−2.73188003	2.077462159	−13.5128629	15.6983086

Figure 6: Performance of IT stocks, 2024-2025, as matrix

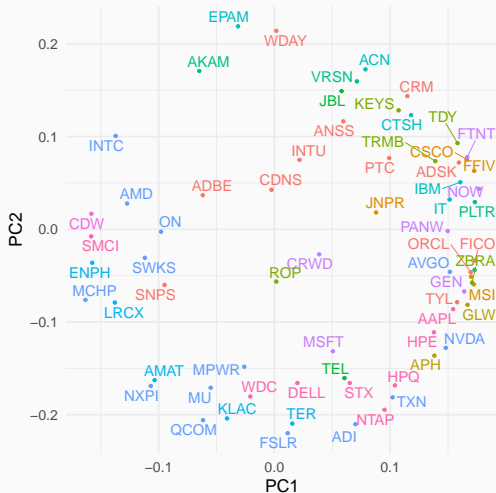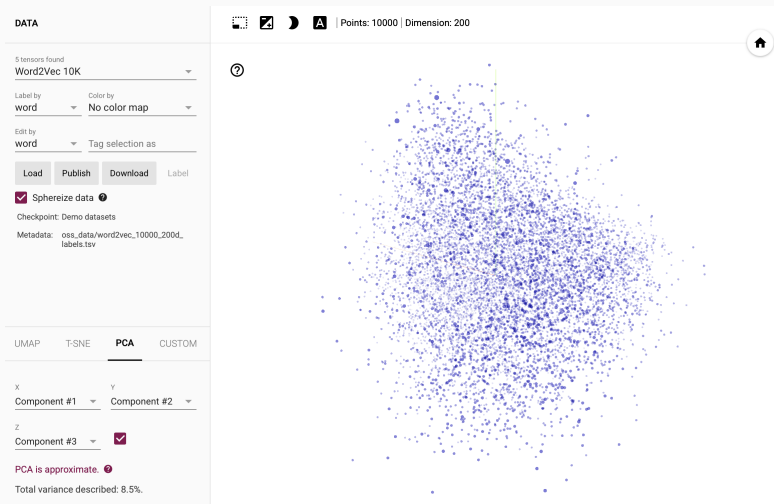**Figure 7:** Performance of IT stocks, 2024-2025, projected to 2D space

Figure 8: Embedding Projector

Getting to know each other

Form groups of 4-5 and discuss the following:

- Introduce yourself (names, major/program)
- What are you excited/nervous/confused about with regards to this course? What questions do you have?
- Have you ever used R? Programmed?
- Why did you choose to study statistics/data science? What is one area of interest you would like to use statistics/data science to study?

📄 Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner,
"Gradient-based learning applied to document recognition,"
Proceedings of the IEEE, vol. 86, pp. 2278–2324, Nov. 1998.