Review and a Look Ahead

Dennis Sun
Stanford University
DATASCI / STATS 112

March 17, 2023

1. Review
2. A Look Ahead
3. Data Science Classes in Spring 2023
4. Final Project

1. Review
2. A Look Ahead
3. Data Science Classes in Spring 2023
4. Final Project

Goal of this Class

The goal of this class was to equip you with data science tools to solve real-world problems.

This class focused mostly on the tools, but I hope

- the guest lecture on Wednesday
- your final project

give you a glimpse of how the tools fit into the larger data science process.

Themes of this Class

1. the different shapes that data can take
   - tabular
   - textual
   - hierarchical (JSON, XML)
   - image
   - geospatial
2. the unifying role of distance metrics
   - document similarity (cosine distance)
   - k-nearest neighbors
   - k-means clustering
   - hierarchical clustering
   - distances on the surface of the Earth (Haversine distance)

We've Covered A Lot!

You would need to take a lot of courses to get all the content of this course:

- machine learning (CS 129, EE 104)
- natural language processing (CS 124)
- databases (CS 145)
- geographic information systems (EARTHSYS 144)
- descriptive statistics (STATS 60/110/141)

All of these courses go deeper into their respective areas, but you know much of the content needed to solve real problems.

DATASCI 112 vs. Typical Machine Learning Class

In a typical machine learning class, you'd also learn:

- more machine learning algorithm (decision trees, neural nets, etc.)
- the math behind them (loss functions, gradient descent)
- how to implement them from scratch

In practice, it would just be another Scikit-Learn import:

```
from sklearn.... import ...
```

To use machine learning in practice, it is much more crucial that you understand general concepts like:

- training vs. test error
- cross-validation
- ensemble methods

which was the focus of this class.

Let's Analyze the Exam Scores

Click on the icon below to be taken to the Colab.

Course Evaluations

Please remember to fill out your course evaluations!

Your feedback is especially important to me, since this was the first time the course was offered.

If there's any topic/lesson that was memorable, let me know so that I make sure to keep it the next time I teach the course!

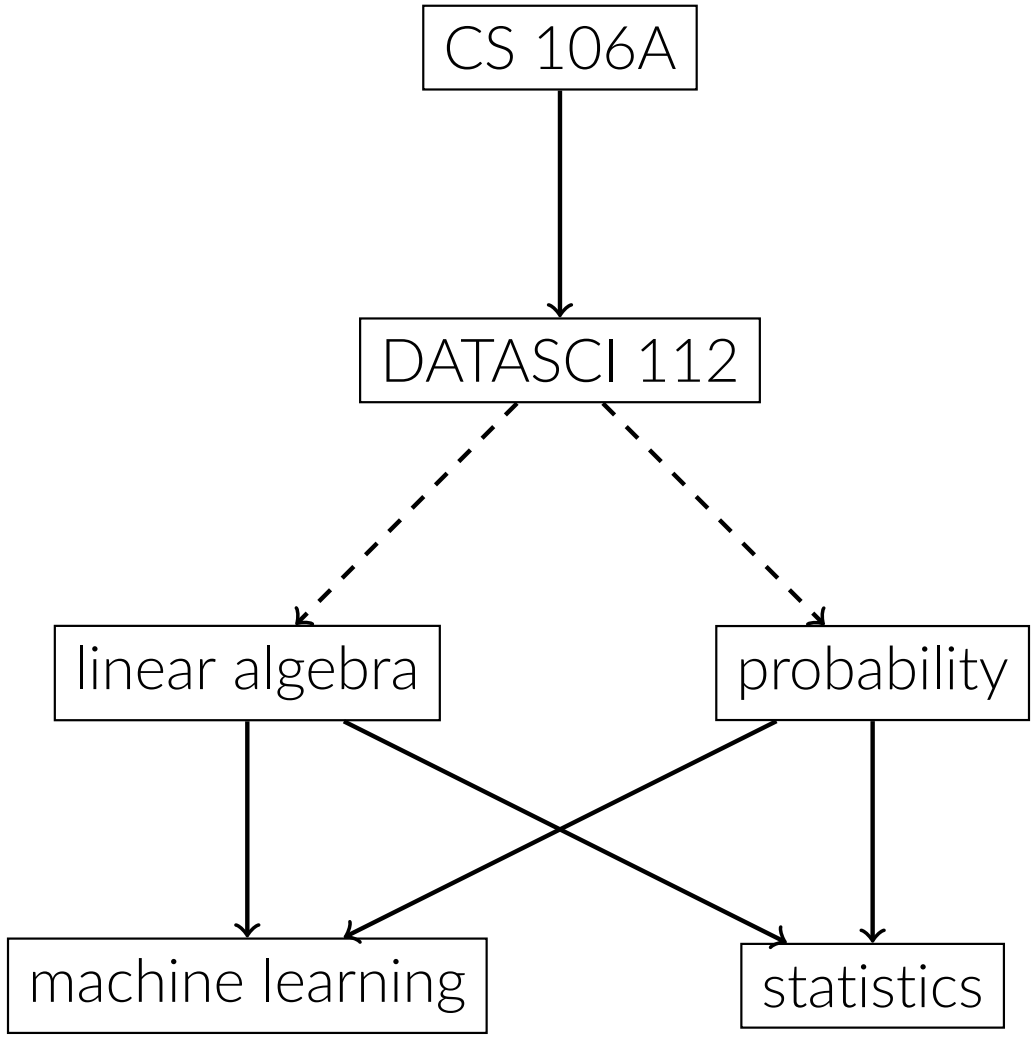
1 Review

2 A Look Ahead

3 Data Science Classes in Spring 2023

4 Final Project

Path to Data Science



B.S./B.A. Data Science

- Consider majoring in Data Science!
  - The B.S. is more mathematical, with the goal of equipping you to develop new data science methods. It used to be called Mathematical and Computational Science (MCS).
  - The B.A. is more focused on applying data science to policy issues.
- I am one of the faculty advisors of the B.S. in Data Science!
- There will likely be tweaks to the curriculum in the coming years. Come talk to me if the major doesn't suit your needs!

Any questions about the major?

Computer Science vs. Data Science

There is a lot of overlap between CS (especially if you do the AI track) and Data Science.

In my opinion, the big distinction is systems vs. math.

- Do Computer Science if you like systems programming (e.g., CS 107) more than math.
- Do Data Science if you like math (e.g., MATH 51, 113) more than systems programming.

You can always major in one and co-term in the other.

Please don’t hesitate to reach out to me in the future if you have any questions about Data Science.

I would be happy to chat with you!

1 Review

2 A Look Ahead

3 Data Science Classes in Spring 2023

4 Final Project

DATASCI 120: Data Narratives

By taking DATASCI120, you will grow in your ability to communicate ideas and insights with data. During the quarter, we will explore multiple components of a well-crafted narrative based on data—from a discussion of data sources to visualization, and from pattern detection to generalizable conclusions. The class does not introduce advanced data analysis techniques. It rather focuses on the essential elements of an inquiry conducted with data and places a special emphasis on how to record and communicate them.

Each student needs to identify a dataset and a question that they are going to explore. As we examine the different components of a data inquiry, the students will carry out a corresponding weekly analysis/writing assignment on the data they have identified, gradually building material for the narrative that will constitute their final paper. Students will support each other, working in pairs, as first readers and reviewers.

The class fulfills the WIM requirement for the majors in Data Science and MCS. It prepares students for summer research experiences in Statistics and Data Science. It is also a great class to take while you are writing up results from your independent data intensive research. Prerequisites: Stats191 or comparable class. Attendance required.

Spr 2023 | Units: 3 | MW 3:00-4:20pm
Instructors: Chiara Sabatti (PI), Isaac Gibbs (TA)

Other Classes to Consider

Linear Algebra (after MATH 51)

- ENGR 108: Introduction to Matrix Methods
  (TuTh 10:30 - 11:50)
- MATH 104: Applied Matrix Theory
  (MWF 1:30 - 2:20)

Probability

- STATS 116: Theory of Probability
  (MTuWTh 10:30 - 11:20)
- CS 109: Introduction to Probability for Computer Scientists
  (MWF 10:30 - 11:50)

1 Review

2 A Look Ahead

3 Data Science Classes in Spring 2023

4 Final Project

Final Project Presentations

- Check the Final Project Presentation Schedule. (There should be no surprises.) The order is random.
- On the day of your presentation, bring your laptops, come on time. (Note: the presentations are in a different room!)
- I'll supply candy and coffee.
- We have many presentations to get through; it is important that you stick to the 5-7 minute limit. Rehearse your talk to make sure you do not go over. You will be cut off at 7 minutes, which may hurt your "Presentation" score.
- You will have to fill out peer evaluations for presentations. Take these seriously; they count toward your "Participation" score.

Remaining Tasks

You only have three remaining tasks for this class:

- Assignment 8 (optional, only if you want to replace your lowest assignment score)
- Final Project
- Course Evaluations

Thanks for a great quarter!

I'm excited to see the projects you produce!