INFO 1998: Introduction to Machine Learning

CDS Education

We explore, learn, and educate big minds.

Lecture 1: Introduction

INFO 1998: Introduction to Machine Learning

CDS Education

We explore, learn, and educate big minds.

Agenda

1. **Meet the Team**: Who are we?
2. **Introduction**: What is Data Science / Machine Learning?
3. **Course Syllabus**: What will we learn?
4. **Course Logistics**: How will we learn?
5. **Getting Started**: Software & Demo

Who are we?

Cornell Data Science

Project Team

- Intelligent Systems
- Insights
- Data Engineering
- Algorithmic Trading

Community Outreach

- Education
  - INFO 1998
  - Workshops
  - Online Tutorials

Course Manager

Who you'll have to bear with

Tanmay Bansal

Information Science '21

tb444@cornell.edu

Took INFO 1998 in Fall 2017

***Still** haven't taken the swim test*

Course Instructors

Backbone of INFO 1998

Dylan Tsai

CS ‘21

Is probably taking 33 credits

Chris Elliott

IS ‘20

Veteran Web-Scraper

Camilo Cedeno-Tobon

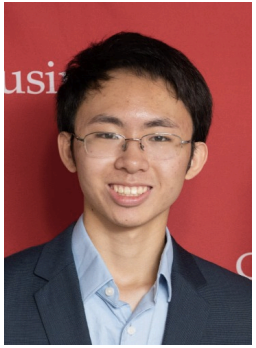
ORIE ‘21

Break-dancer

Emily Weed

Stats ‘22

Professional Chef



Jerry Sun

CS ‘23

Will debate you

Raye Liu

ORIE ‘22

aka Whitney Houston

What is Data Science?

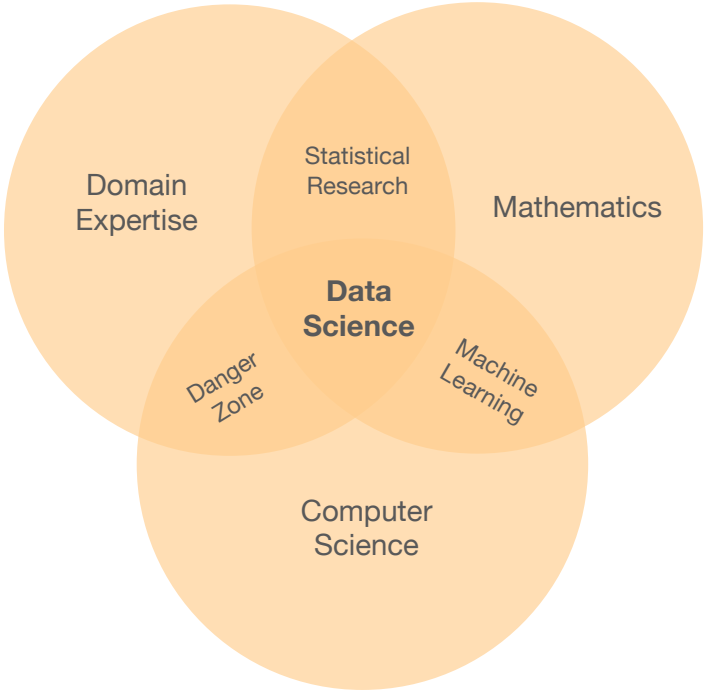
What is Data Science?

“By "Data Science", we mean almost everything that has something to do with data: Collecting, analyzing, modeling..... yet the most important part is its applications --- all sorts of applications.”

Journal of Data Science

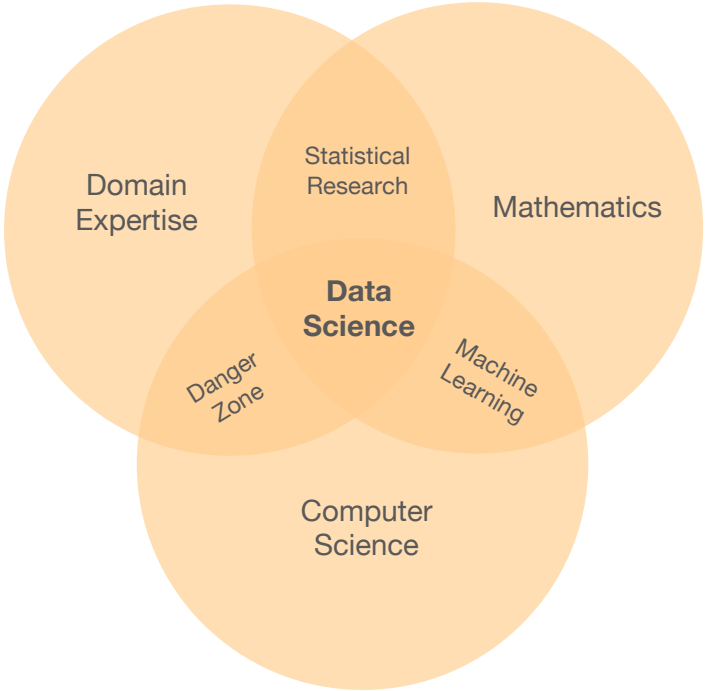
What is Data Science?

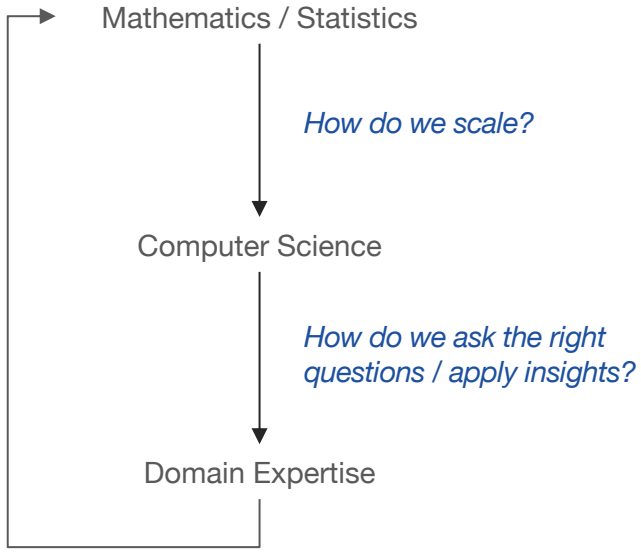
Conway's Data Science Venn Diagram



What is Data Science?

Conway's Data Science Venn Diagram





Data Science ≠ Machine Learning

Applications of Data Science

We'll be back to this slide!

Predictive	Stock Prices
	Netflix Recommendations
Preventive	Medical Diagnosis
	Social Impact Analytics
Real-Time	Digital Advertising
	Autonomous Vehicles

Course Objectives and Syllabus

What you should aim to understand by the end of the course

OBJECTIVES	SYLLABUS
Manipulating Data	Data Manipulation / Visualization *Lectures 1-3*
Communicating Data	
Understanding of ML as a concept	Fundamentals of Machine Learning *Lectures 4-5*
Intuitive understanding of ML models	
Implementation of ML models	Supervised Learning *Lectures 6-8*
Comfort Using Python	
Applications in Industry	Unsupervised Learning *Lecture 9*
Project Experience	

Sample Final Projects

“0 – 100, Real Quick” - Drake

(1) Determining indicators for a candidate’s success in Canadian Elections

Kevin Zhou, Jerry Sun

(2) Predicting Player Performance/Value using NFL Data

Brian Bobby, Max Brody, Teddy Klausner

FAQs

Is this class a good fit for you?

1) Will I become a Data Scientist / Machine Learning Engineer?

No, you will not. The course covers a breadth of concepts, helps build intuitive understanding of some models, but does not dive too deep into the mathematical complexities (since this is a 1000-level course). However, feel free to come to office hours if you're interested in learning more.

2) How much time commitment is this course?

Depends. If you want to have a strong command over the material so that you can get a head start in this field, you will have to read a little more and be prepared to spend time with our TAs to go over concepts in more depth. If you want to acquire just street-fighting machine learning skills, that's fine too – it'll require little more than 1 hour per week.

3) I have no background in CS / Stats – am I underprepared?

Not at all! We'll teach you everything you need to know, but you may have to spend a little more time getting comfortable with Python. A number of non-STEM graduate students have taken this class in the past to understand basics that they could apply to their research. A large number of freshmen also take the course because they're excited to learn more about the field. TL;DR: If you're interested, give it a shot!

Enrollment

Let's get this credit

1

Fill out
tiny.cc/info1998_sp20

Enrollment

Let's get this credit

1

Fill out
tiny.cc/info1998_sp20

2

Enroll in Piazza
(INFO 1998)
You will be added to CMS over the weekend

For fun, download
'Pizza for Piazza'

Enrollment

Let's get this credit

1

Fill out
tiny.cc/info1998_sp20

2

Enroll in Piazza
(INFO 1998)
You will be added to CMS over the weekend

For fun, download
'Pizza for Piazza'

3

Enroll through Student Center after obtaining pins

Course Logistics

How is the class structured (and graded)?

10 assignments (1 assignment per lecture) *Drop lowest score!*	**50%**
Mid-semester Group Project *2-3 students*	**15%**
Final Group Project *2-3 students*	**35%**

Passing Grade: 70%

Getting Started

Get the required software

Jupyter Notebook

- Documentation (Code + Visuals)
- Supports Python, Julia, etc.
- Easy to share

Python (3)

- Easy to learn, Readable
- Industry Standard
- Great documentation, online resources

Installation

Get the required software

1.

Anaconda https://www.anaconda.com/distribution/

2.

Open **Terminal** (MacOS) / **Command Prompt** (Windows),
Type and enter: `jupyter notebook`

Demo

Next Steps

- **Installation**: Seek help at Office Hours!
- **Assignment 1**: Due at 5:30pm on Wednesday, Feb 19, 2020 on CMS (Will be enrolled soon!)
- **Next Lecture**: Data Manipulation

CDS Education

We explore, learn, and educate big minds.