

# Statistical Methods in AI (CS7.403)

## Lecture-1: Intro and Administrivia

Ravi Kiran ([ravi.kiran@iiit.ac.in](mailto:ravi.kiran@iiit.ac.in))

<https://ravika.github.io>



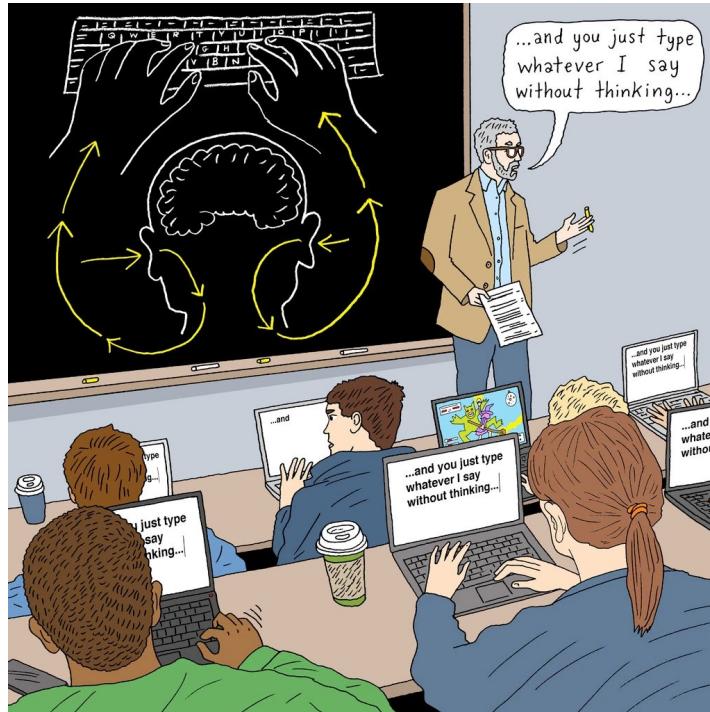
Center for Visual Information Technology (CVIT)  
IIIT Hyderabad

# No laptops, phones





In a series of experiments at Princeton University and the University of California, Los Angeles, students were randomly assigned either laptops or pen and paper for note-taking at a lecture. Those who had used laptops had substantially worse understanding of the lecture, as measured by a standardized test, than those who did not.



Do not come in with ONLY your phone.  
You will be asked to leave.

# SMAI (Statistical Methods in AI)

- SMAI ~ Introduction to Machine Learning



**Yi Ma**

@YiMaTweets

...

Mathematics is the art of giving the same name to different things(Henri Poincaré). Machine learning is the art of giving different names to the same thing.

# Machine Learning



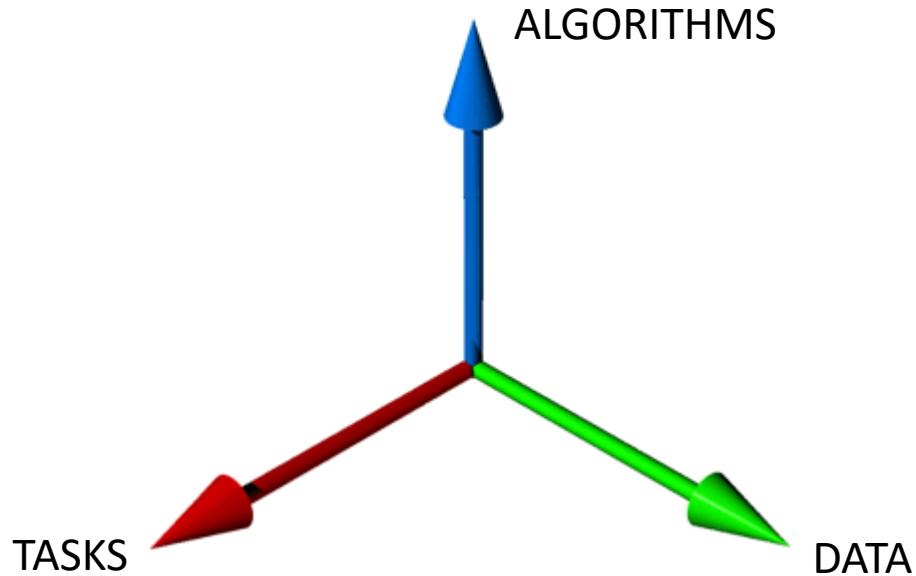
Study of **Algorithmic methods** that use **data** to improve their **knowledge** of a **task**

# What is ML ? (alternate definition)

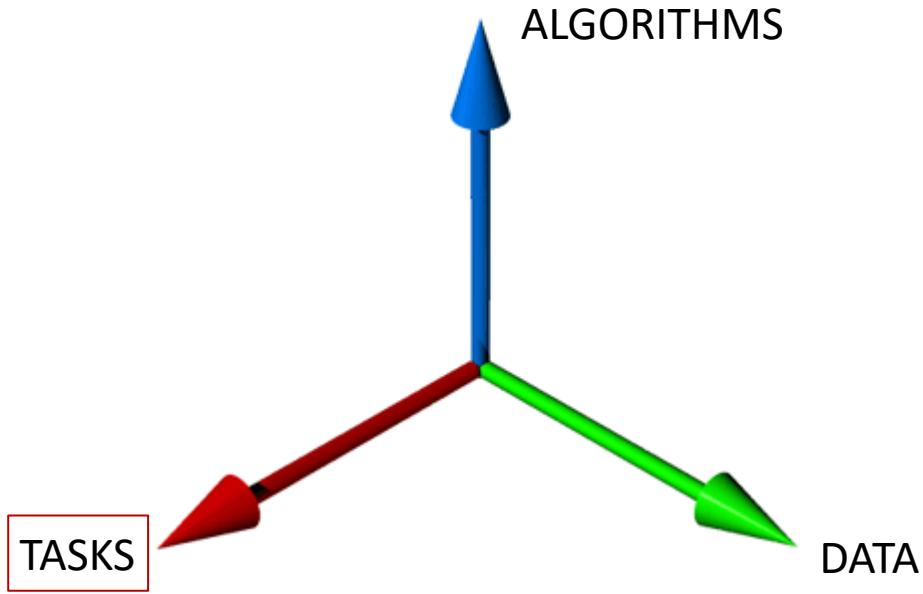


- Computer program that learns from **experience E** in order to improve its **performance P** on a **task T** (Tom Mitchell)

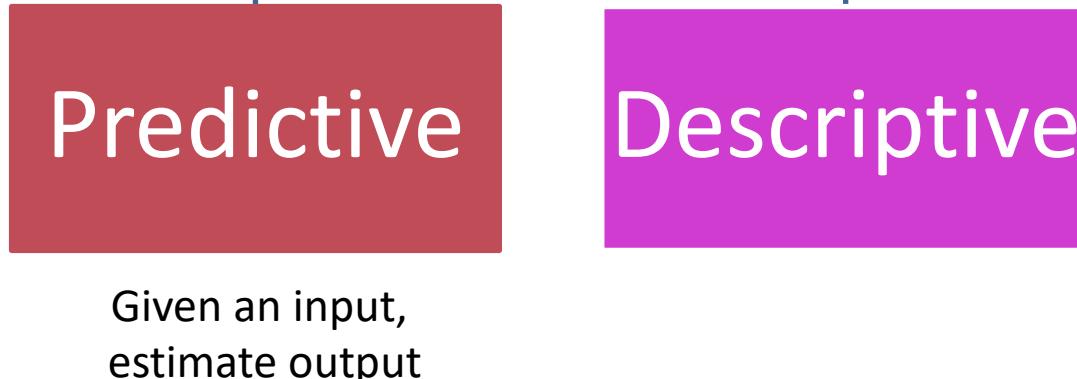
# 3 axes of ML



# 3 axes of ML



# ML Tasks



# ML::Tasks → Predictive

Feature Space  $\mathcal{X}$

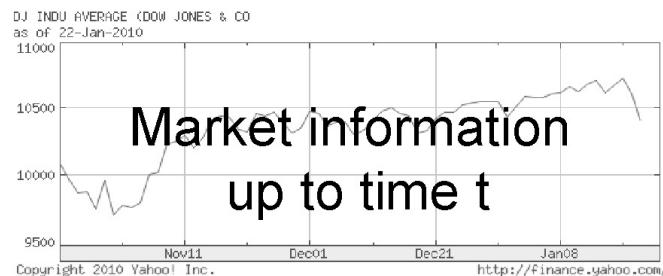


Words in a document

Label Space  $\mathcal{Y}$

“Sports”  
“News”  
“Science”

...



Share Price  
“\$ 24.50”



**Task:** Given  $X \in \mathcal{X}$ , predict  $Y \in \mathcal{Y}$ .

# ML::Tasks → Predictive → Classification

Feature Space  $\mathcal{X}$

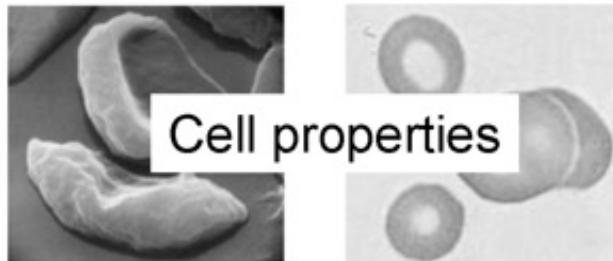


Label Space  $\mathcal{Y}$

“Sports”  
“News”  
“Science”  
...



Cell properties



“Anemic cell”  
“Healthy cell”



**Task:** Given  $X \in \mathcal{X}$ , predict  $Y \in \mathcal{Y}$ .

**Discrete Labels**

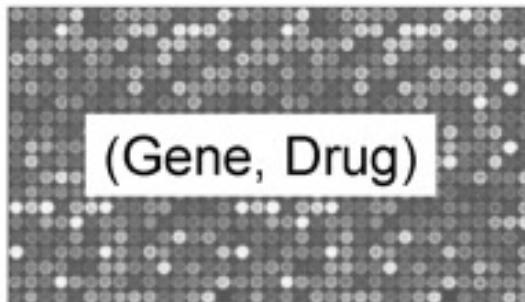
# ML::Tasks → Predictive → Regression

Feature Space  $\mathcal{X}$



Label Space  $\mathcal{Y}$

Share Price  
"\$ 24.577"



Expression level  
"6.88"

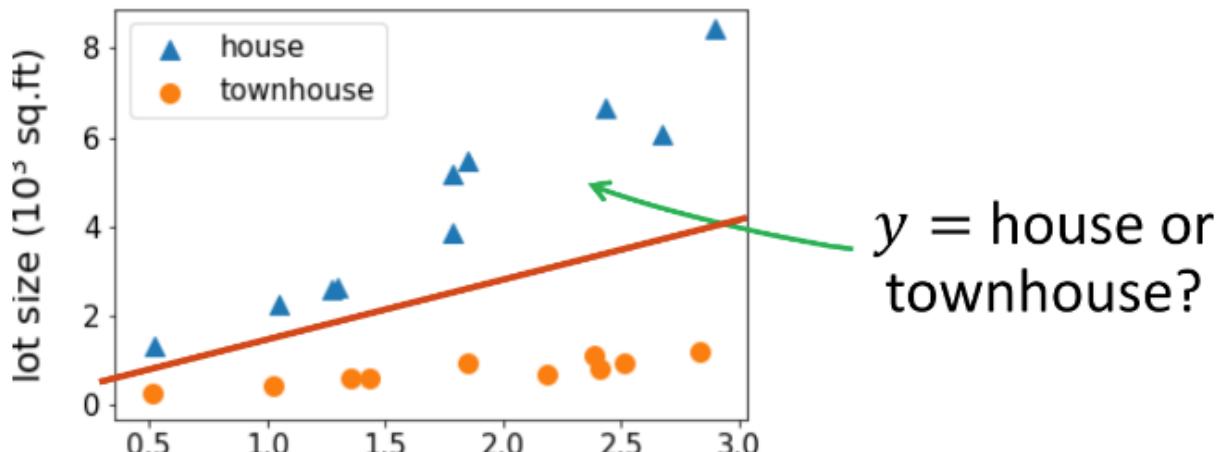
**Task:** Given  $X \in \mathcal{X}$ , predict  $Y \in \mathcal{Y}$ .

**Continuous Labels**

# Regression vs Classification

- regression: if  $y \in \mathbb{R}$  is a continuous variable
  - e.g., price prediction
- classification: the label is a discrete variable
  - e.g., the task of predicting the types of residence

(size, lot size) → house or townhouse?



# Supervised Learning in Computer Vision

## ➤ Image Classification

➤  $x$  = raw pixels of the image,  $y$  = the main object

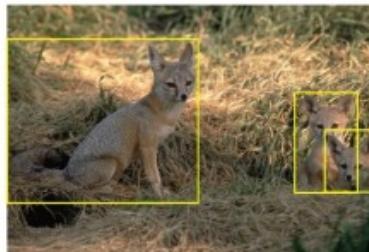


- There is too much info in raw data
- Relevant info is hidden
- Feature Extraction:  
Extract useful info (X) from raw data

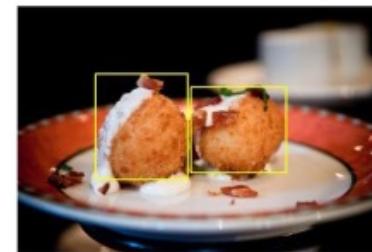
# Supervised Learning in Computer Vision

- Object localization and detection

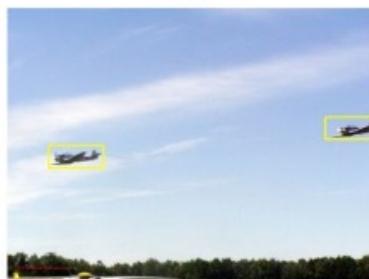
- $x$  = raw pixels of the image,  $y$  = the bounding boxes



kit fox



croquette



airplane

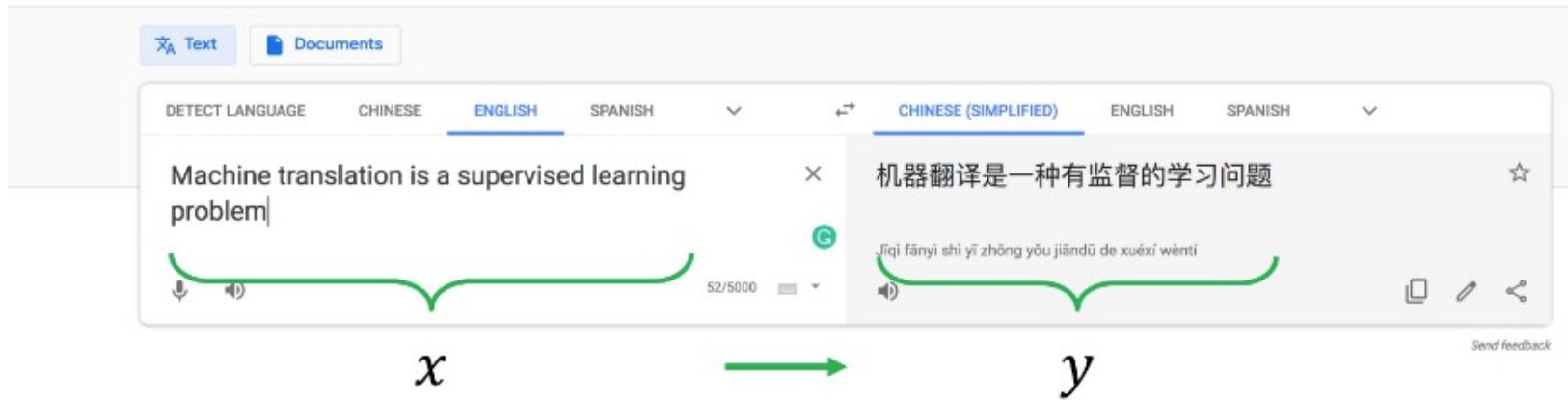


frog

# Supervised Learning in Natural Language Processing

## ➤ Machine translation

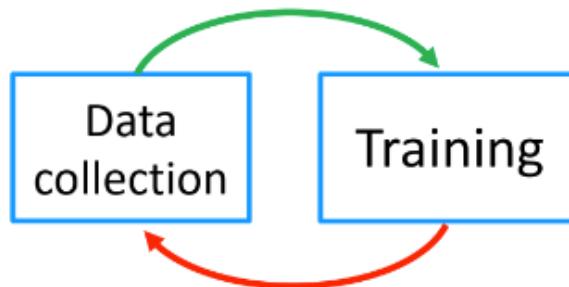
Google Translate



# Reinforcement Learning

- The algorithm can collect data interactively

Try the strategy and collect feedbacks

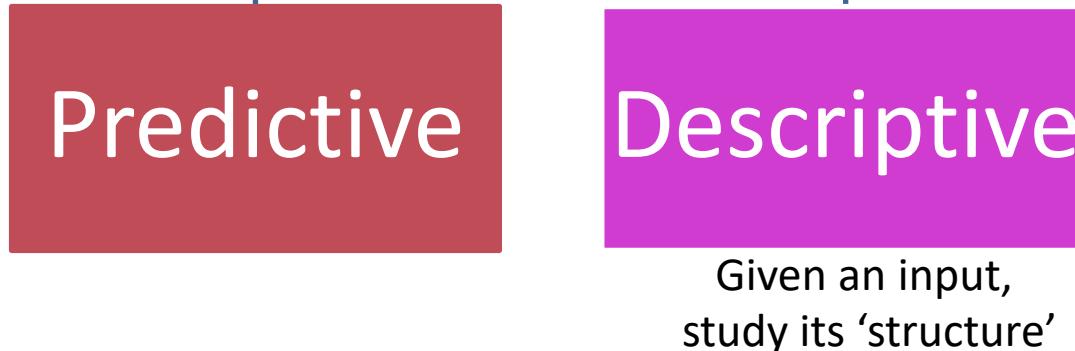


Improve the strategy based on the feedbacks

# ML Tasks



# ML Tasks

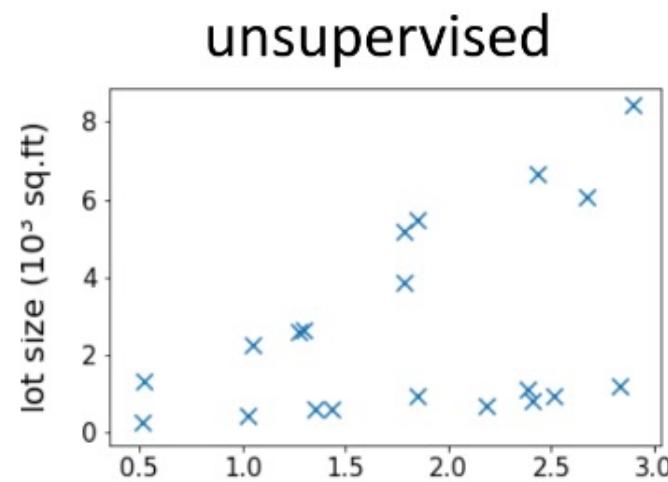
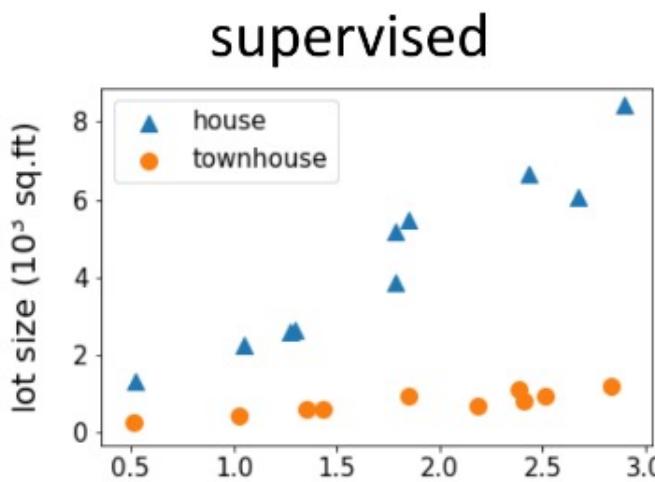


# ML::Tasks → Descriptive

- Study/Exploit the ‘structure’ of data
  - Density Estimation
  - Clustering
  - Dimensionality Reduction
- Also studied as ‘Unsupervised Learning’
  - ‘Input’ data without paired ‘Output’

# Unsupervised Learning

- Dataset contains **no labels**:  $x^{(1)}, \dots x^{(n)}$
- **Goal** (vaguely-posed): to find interesting structures in the data



# Unsupervised Learning → Density Estimation

Aka “learning without a teacher”

Feature Space  $\mathcal{X}$



Words in a document



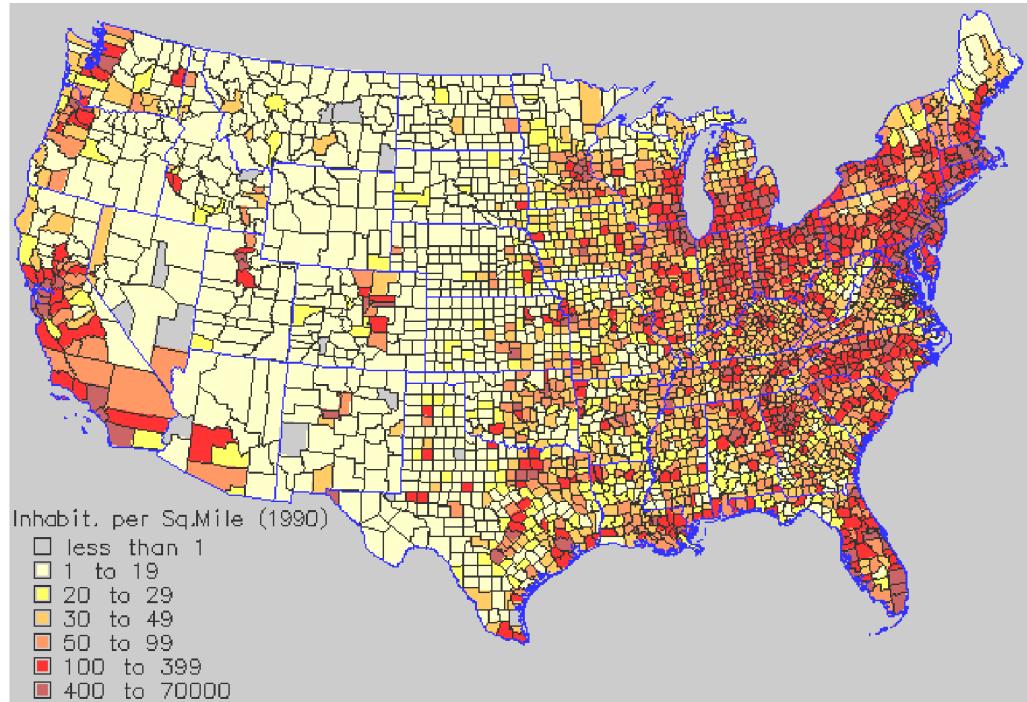
Word distribution  
(Probability of a word)

**Task:** Given  $X \in \mathcal{X}$ , learn  $f(X)$ .

# Unsupervised Learning → Density Estimation

Population density

**Task:** Given  $X \in \mathcal{X}$ , learn  $f(X)$ .



# Unsupervised Learning → Clustering

**Task:** Given  $X \in \mathcal{X}$ , learn  $f(X)$ .

Group similar things e.g. images

[Goldberger et al.]



# Unsupervised Learning → Web Search

Google search results for "alphabet".

Search filters: All, Images, News, Videos, Maps, More, Settings, Tools.

Image filters: printable, font, calligraphy, phonetic, fancy, cursive, handwriting, spanish, a to z, arabic, military, lettering, sign language.

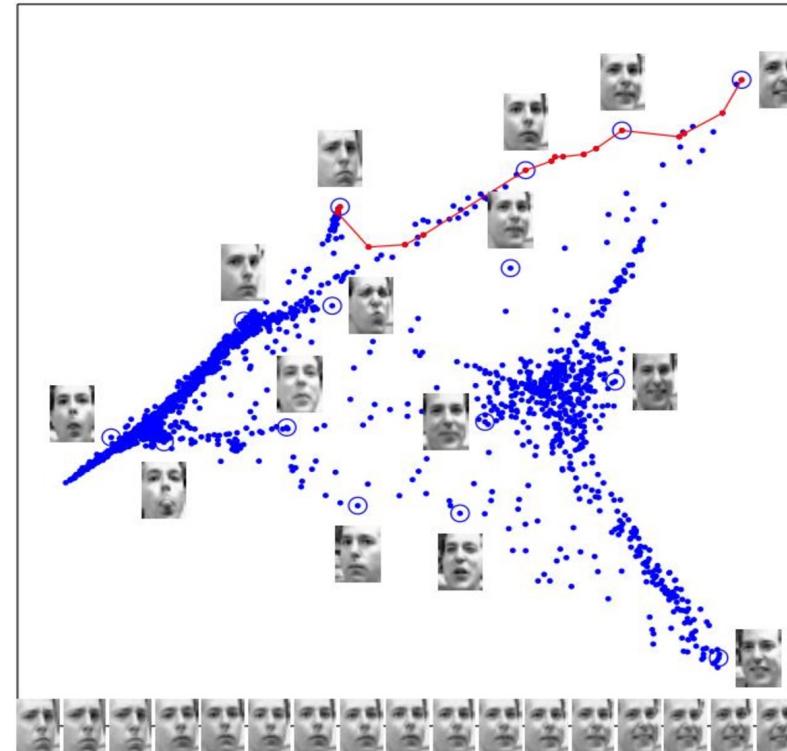
Results:

- Talk to Me Alphabet | ABCya!** (vectorstock.com) - A colorful, hand-drawn style alphabet from A to Z.
- Morse code alphabet Roy...** (vectorstock.com) - A grid showing Morse code equivalents for the English alphabet.
- Patient Provider Communication ...** (patientprovidercommunication.org) - A grid of letters with small icons representing common sounds.
- Alphabet Vectors, Photos and PSD files ...** (freepik.com) - A collection of colorful, stylized vector letters.
- Colorful Capital Letters Alphabet...** (123rf.com) - A collection of colorful, stylized capital letters.
- Why are the letters of the alphabet in ...** (theguardian.com) - A collection of colorful, stylized lowercase letters.
- Cursive Alphabet Modern ...** (amazon.com) - A grid of cursive letters.
- MFT Stitched Alphabet Die** (sevenhillscrafts.co.uk) - A grid of cursive letters.
- Fun english alphabet one ...** (vectorstock.com) - A colorful, hand-drawn style alphabet from A to Z.
- Moldtsov alphabet - Wikip...** (en.wikipedia.org) - A grid of letters with Cyrillic equivalents.
- The Alphabet Chart Grade ...** (carsondellosa.com) - A colorful chart showing the English alphabet with corresponding numbers and symbols.
- Turkish language, alphabets and ...** (omniglot.com) - A grid of letters with Turkish equivalents.
- FolkArt Alphabet Heavy Ty...** (homedepot.com) - A grid of stylized, heavy-set letters.
- Patient Provider Communication ...** (patientprovidercommunication.org) - A grid of letters with small icons representing common sounds.
- Definition of Alphabet by M...** (merriam-webster.com) - A grid of letters with definitions.
- We are the Alphabet - YouTube** (youtube.com) - A video thumbnail showing two large, friendly red letters A and a with arms and legs.

# Unsupervised Learning → Dimensionality Reduction + Visualization

Images have thousands or millions of pixels.

Can we give each image a coordinate,  
such that similar images  
are near each other?

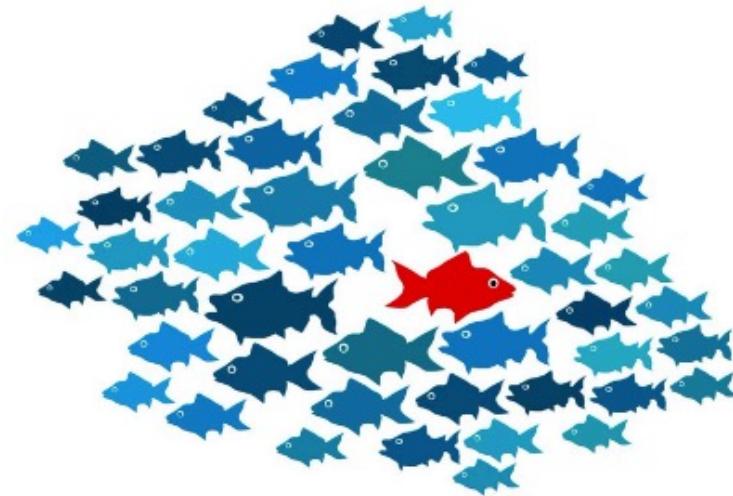


## Unsupervised Learning: Anomaly/Outlier detection

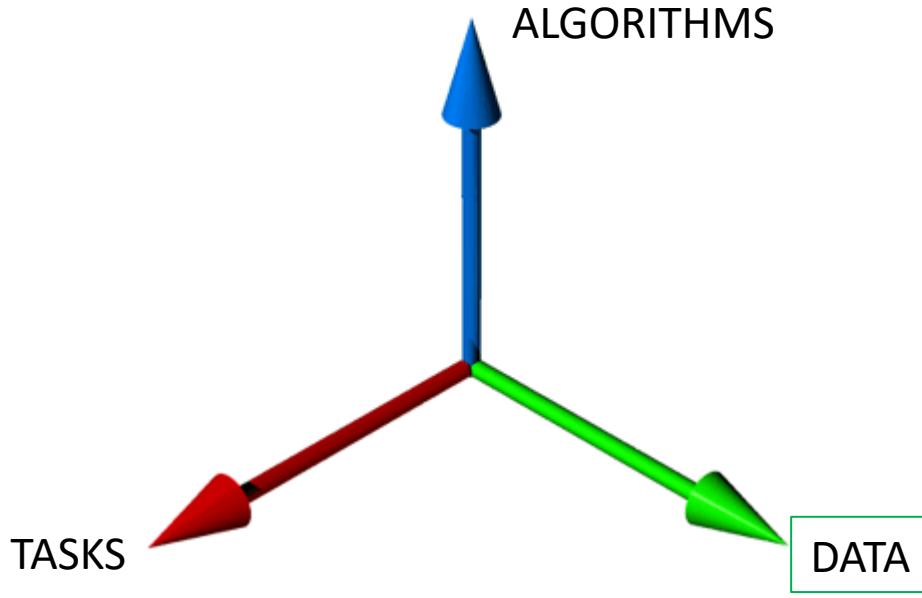
**Anomaly Detection:** The task of finding samples in a dataset that raise suspicion.

**Problem:** Usually, what exactly you are looking for is unknown.

**Solution:** Use statistics and characteristics of dataset to find outliers.

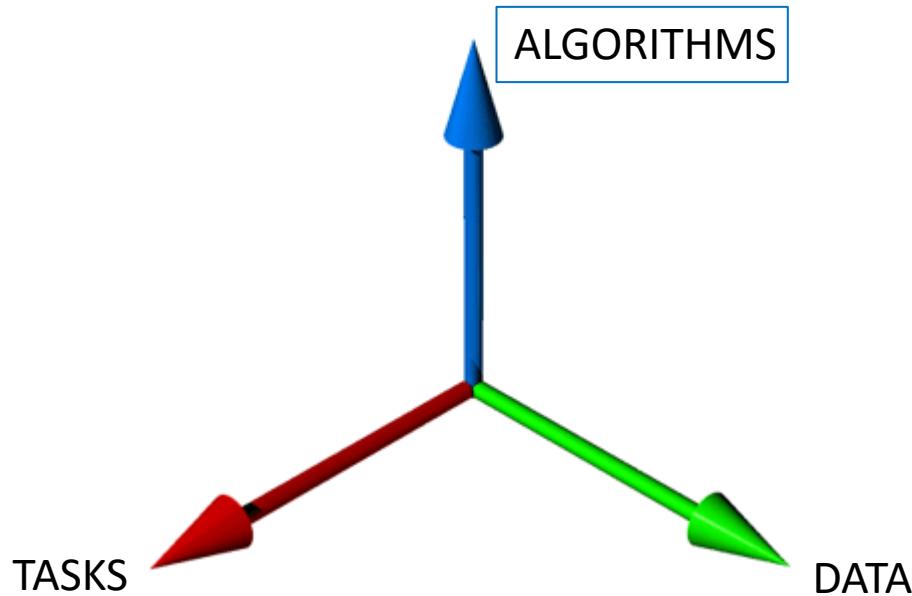


# 3 axes of ML



- Fully Observed
- Partially Observed
  - Some variables systematically not observed (e.g. 'topic' of a document)
  - Some variables missing some of the time (e.g. 'faulty sensor' readings)

# 3 axes of ML



# Approaches

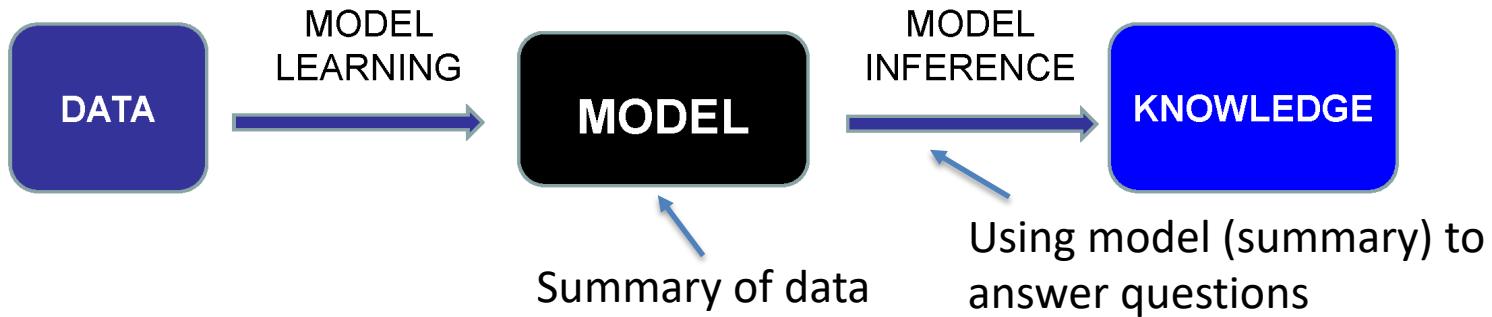
Model-based

Model-free

# Model-based ML



Algorithmic methods that use **data** to improve their **knowledge** of a **task**



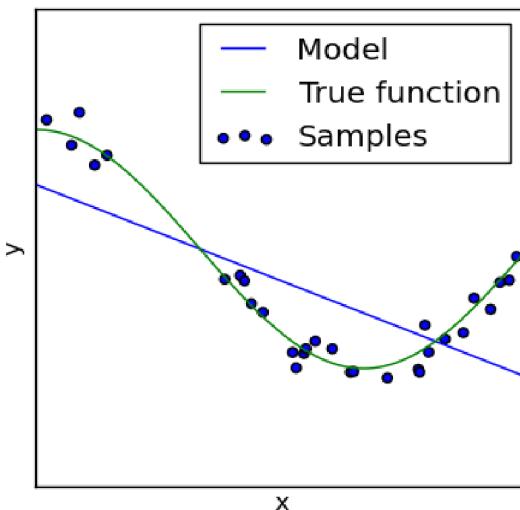
Model-based  
ML

Parametric

Non-parametric

# Parametric Models

- “Fixed-size” models that do not “grow” with the data
- More data just means you learn/fit the model better

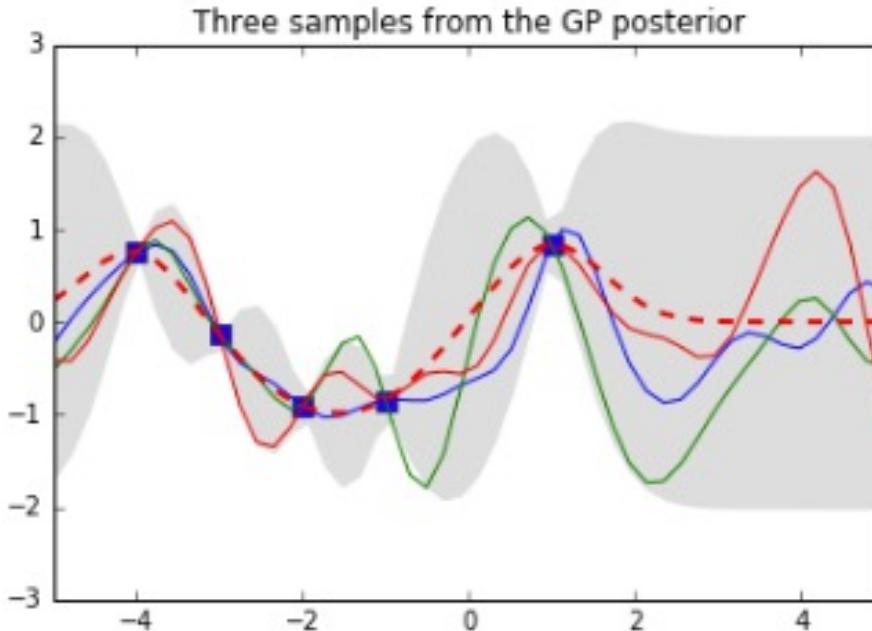


Fitting a simple line (2 params)  
to a bunch of one-dim. samples

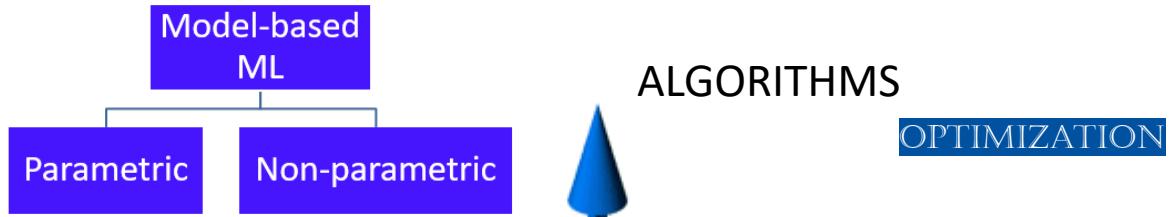
Model: data = point on line + noise

# Nonparametric Models

- Models that grow with the data
- More data means a more complex model

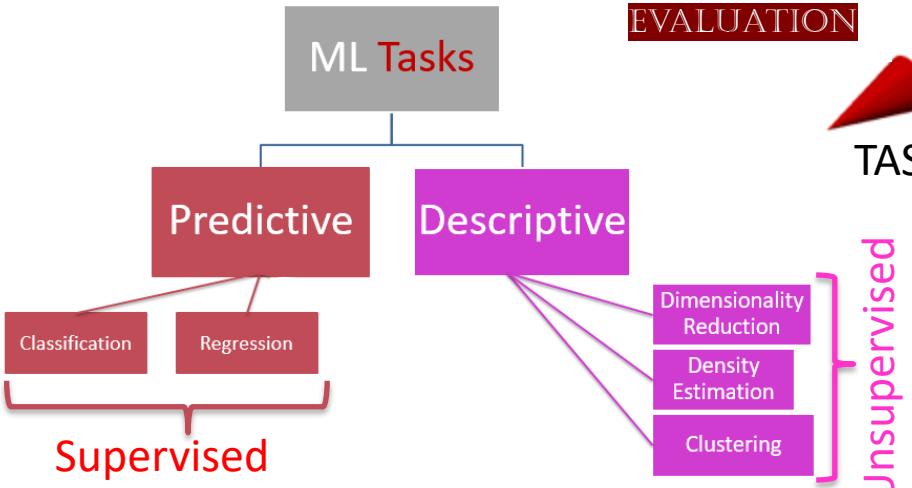


Gaussian Process



ALGORITHMS

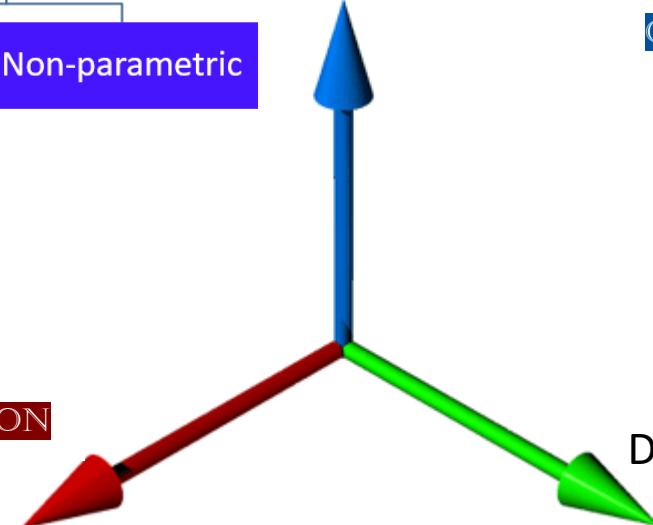
OPTIMIZATION



EVALUATION

TASKS

Unsupervised



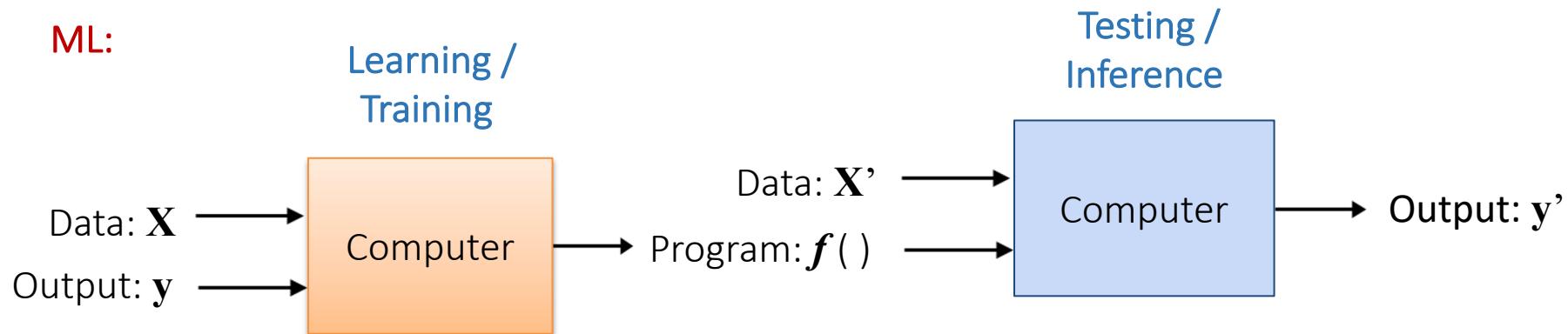
- Fully Observed
- Partially Observed
  - Some variables systematically not observed (e.g. ‘topic’ of a document)
  - Some variables missing some of the time (e.g. ‘faulty sensor’ readings)

# Programming vs. Machine Learning

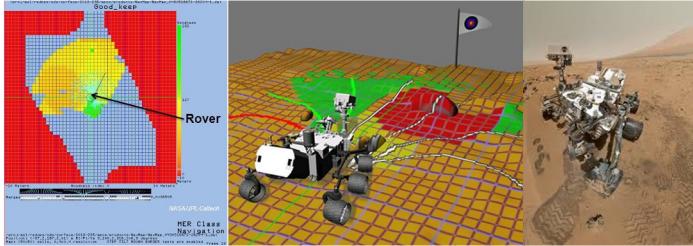
Programming:



ML:



# When to “Learn”

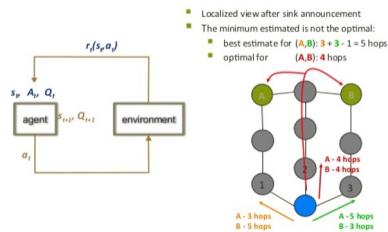


Human expertise does not exist  
(‘learning’ to navigate on Mars)

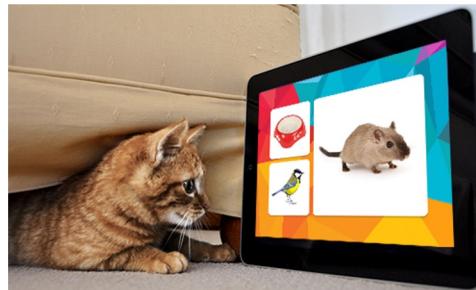


Humans unable to explain their expertise  
(‘learning’ to understand speech)

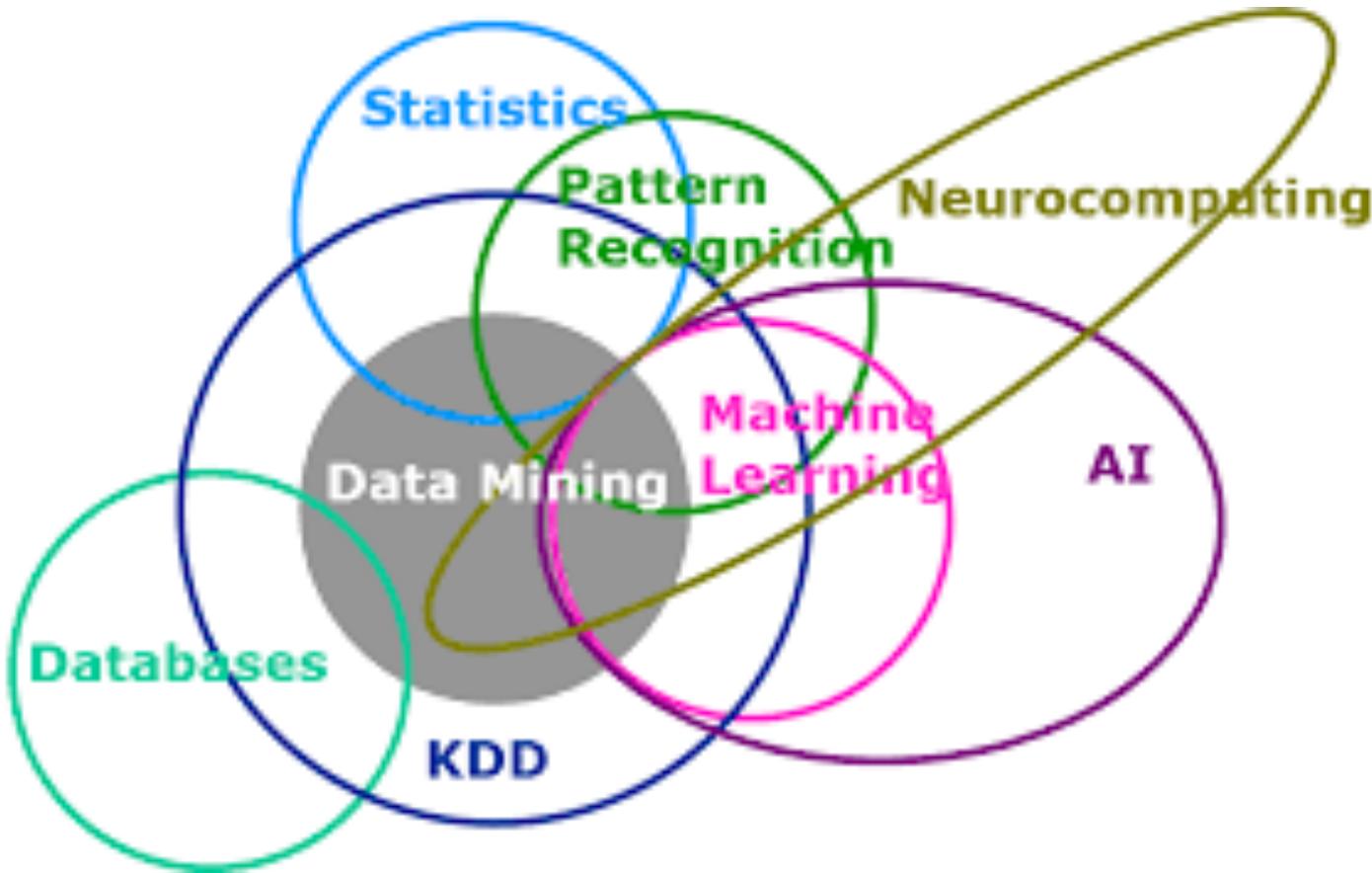
FROMS: Multicast routing with Q-Learning



Solution changes over time  
(‘learning’ to route network packet traffic)



Solution needs to be adapted to particular cases  
(user-specific ‘learning’)



# ML v/s Statistics

- Statistics:
  - Common assumption: Data is generated by a model
  - Cares about: How well does data fit the model ?
- ML
  - (Generally) Cares about: How well does model fit the data ?

# About the course (CS7.403)

- Timings: Tue, Fri (Himalaya 205, 5.10p – 6.40p)
- Tutorial: Sat, Himalaya 205, 11.40a – 12.40p

# Course Overview

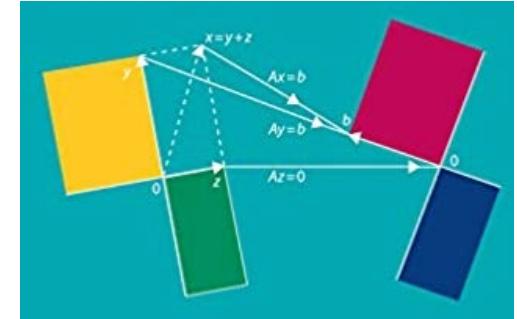
- Part-1 : Supervised Learning
- Part-2 : Unsupervised Learning
- Part-3 : Neural Networks
- Part-4 : Bias-Variance, Nonparametric estimation
- Part-5 : ML for Sequential Data
- Part-6 : Misc. topics

# Pre-requisites

- Programming
- Python + Numpy
  - [https://cs229.stanford.edu/lectures-spring2022/cs229-python review slides.pdf](https://cs229.stanford.edu/lectures-spring2022/cs229-python_review_slides.pdf)
  - [https://cs229.stanford.edu/lectures-spring2022/cs229-python review materials.zip](https://cs229.stanford.edu/lectures-spring2022/cs229-python_review_materials.zip)

# Pre-requisites

- Mathematics
  - Linear Algebra
    - Matrix, Vector operations
    - Systems of equations, Matrix Form ( $Ax = b$ ), Conditions for existence of solution
    - Rank
    - Invertibility of matrix
    - Eigenvectors, Eigenvalues,
    - Semi-definiteness of matrix
    - Decompositions (Singular Value Decomposition, Eigendecomposition)
    - Properties of symmetric matrices



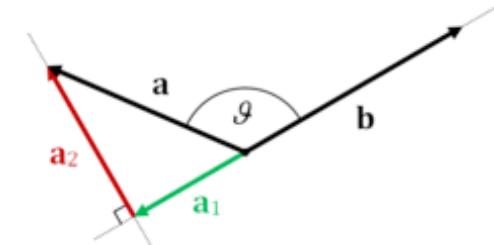
[https://courses.engr.illinois.edu/ece498rc3/fa2016/material/linearAlgebra\\_4pgs.pdf](https://courses.engr.illinois.edu/ece498rc3/fa2016/material/linearAlgebra_4pgs.pdf)

<https://github.com/kenjihiranabe/The-Art-of-Linear-Algebra>

[https://cs229.stanford.edu/lectures-spring2022/cs229-linear\\_algebra\\_review.pdf](https://cs229.stanford.edu/lectures-spring2022/cs229-linear_algebra_review.pdf)

# Pre-requisites

- Mathematics
  - Coordinate Geometry
    - Distance of point from a line
    - Distance between two parallel lines
  - Vector Calculus
    - Dot product , Projections



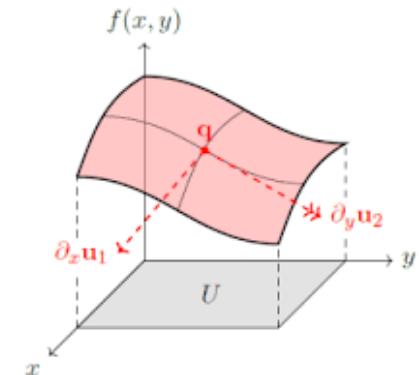
<http://studyphysicswithme.com/blog/2016/11/07/vectors-vector-spaces/>

# Pre-requisites

## – Calculus

- Derivative of single variable,  $y = f(x)$
- Partial derivative
- Chain Rule
- Gradient

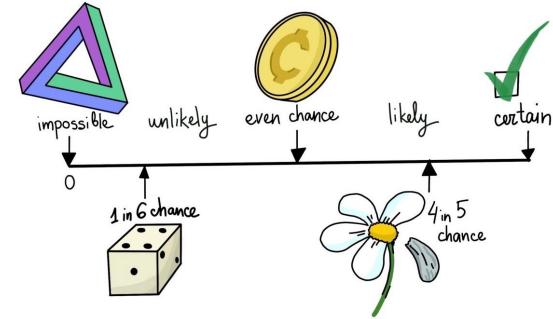
<http://tutorial.math.lamar.edu/getfile.aspx?file=B,41,N>



# Pre-requisites

## — Probability

- Axioms of probability
- Sample Space, Event
- Discrete, Continuous distributions
  - Uniform, Bernoulli, Geometric
  - Gaussian
- Expectation of a random variable



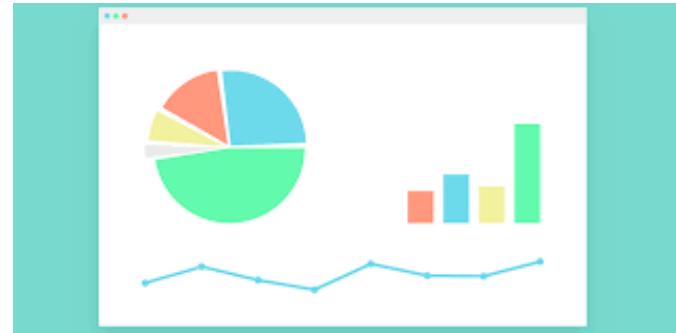
Cheat-sheet: <https://stanford.edu/~shervine/teaching/cme-106/>

[http://www.wzchen.com/s/probability\\_cheatsheet.pdf](http://www.wzchen.com/s/probability_cheatsheet.pdf)

# Pre-requisites

- Statistics

- Mean, Median, Mode
- Standard Deviation



Cheat-sheet: <https://stanford.edu/~shervine/teaching/cme-106/>

# Course Objectives

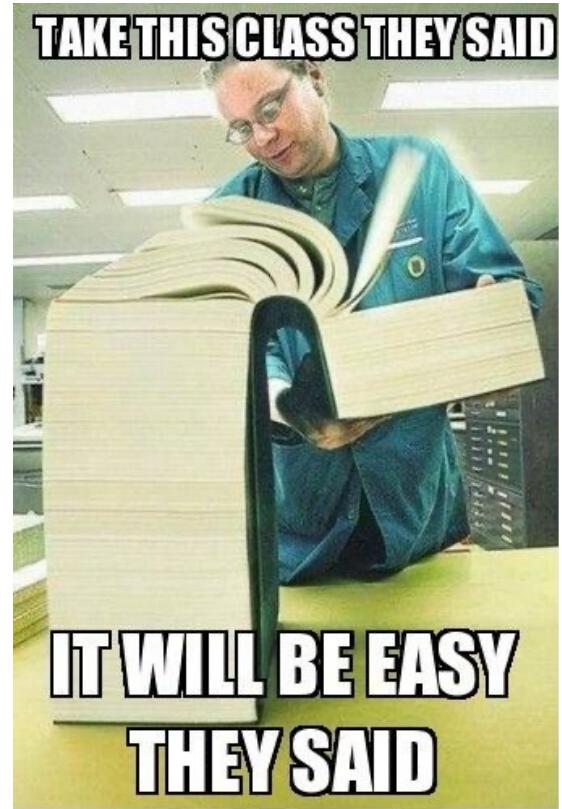
- Determine whether ML is suitable for a problem
- Formulate a problem as a ML problem (data ,representations, tasks, algorithms)
- Understand and apply ML method(s)
- Be aware of ML pitfalls, follow best practices
- Be ready to dive deeper (into ML theory or applied areas)

# About the course - TAs

- TBA

# About the course – Grading Scheme 1

- Assessment
  - Final Exam (35 %)
  - Mid semester Exam (25 %)
  - 2 Quizzes (15 %)
  - 5 Assignments (20 %)
  - Scribe Class Notes (5 %)

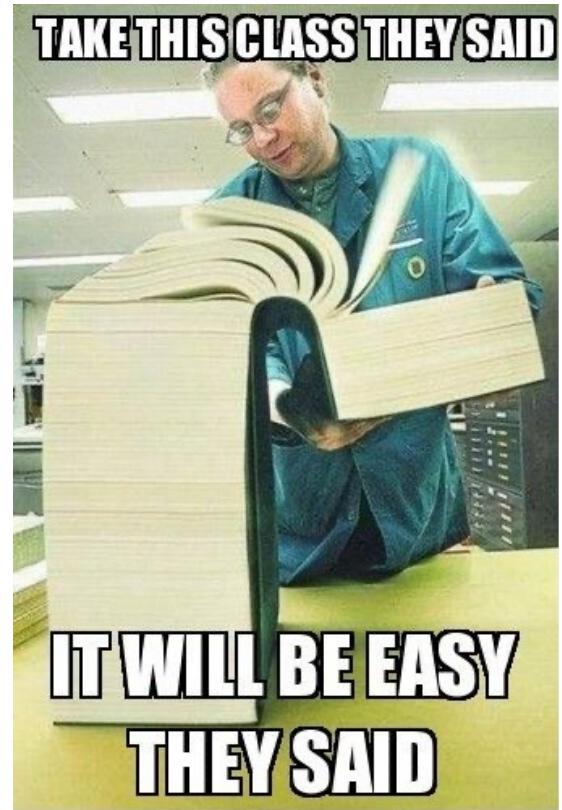


# No projects (by default)



# About the course – Grading Scheme 2

- Assessment
  - Mid semester Exam (25 %)
  - 2 Quizzes (15 %)
  - **4 Assignments (20 %) [best 4 of 5]**
  - Scribe Class Notes (5 %)
  - **Project (max 25 projects, max 2 members per team) → 35%**
- Decide grading scheme after mid-sem



# About the course

- Make travel plans only for Dec 5<sup>th</sup> night and beyond.
- Exam Absence policy: Approval from Academic Office necessary. No re-exam. Marks will be calibrated based on your relative class performance in other exams.

# About the course - assignments

- Code
  - Python (scikit-learn, Google Colab, Jupyter notebook)
  - Neural Networks: Pytorch, Keras

## TOP 10 DEEP LEARNING FRAMEWORKS

10. You can't
9. Rank them
8. Because each has
7. Their own merits
6. That make them better
5. Tools for certain tasks
4. Than the others
3. Just appreciate that they're
2. All used to build great stuff
1. Pytorch



Andrej Karpathy @karpathy

Following

I've been using PyTorch a few months now and I've never felt better. I have more energy. My skin is clearer. My eye sight has improved.

11:56 AM - 26 May 2017

384 Retweets 1,519 Likes



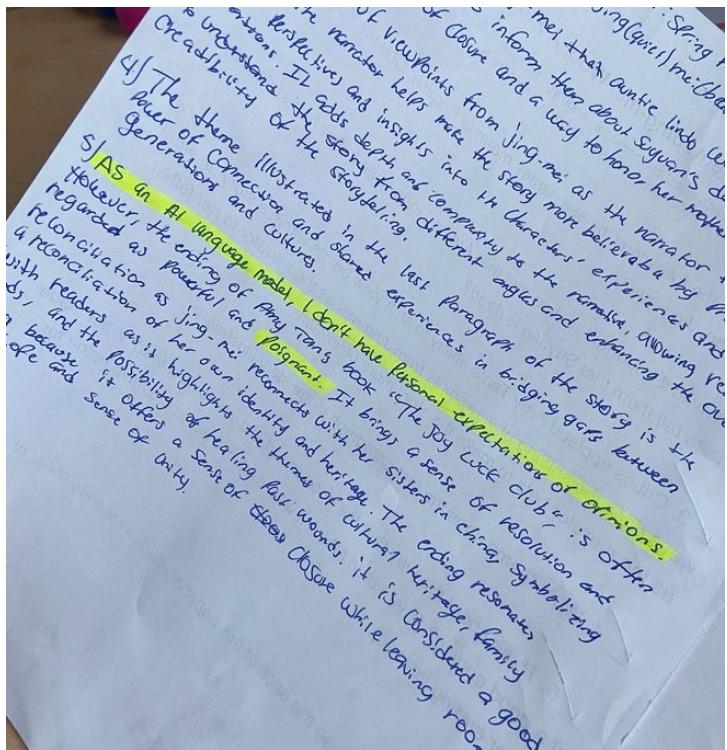
33 384 1.5K

# About the course - scribing

- Scribing = Course notes
- 8 students chosen randomly without repetition
- Student names will be called out at end of lecture
- To be submitted within a week of the lecture (0/1 – no late submissions).
- Very lightweight (and easy marks) if group coordinates well
- Add resources/diagrams you feel will help the class
- Scribe notes will be shared with class

# About the course – collaboration policy

- OK to discuss assignment questions and approaches
  - But work must be your own (no copying – partially or fully)
  - We will be checking for copying/plagiarism
  - Better to own up than be caught !



# About the course – collaboration policy

- If you use Github Copilot, ChatGPT and variants, mention how you used them.
- Evidence of mass copying will lead to assignments becoming optional and higher weightage for exams.



# About the course – Grading Policy

- **Assignment Late Policy:** 50% if one day late; zero percent if more than one day late
- **A one-time late submission bonus:** With maximum of three days delay. You must adhere to standard late submission policy after using your bonus. No exceptions will be made. You'll need to inform TAs before assignment deadline if you wish to use the late submission bonus.

# About the course – Grading Policy

- **Grades Policy:**
  - No revisiting grade components after assignment/paper checking

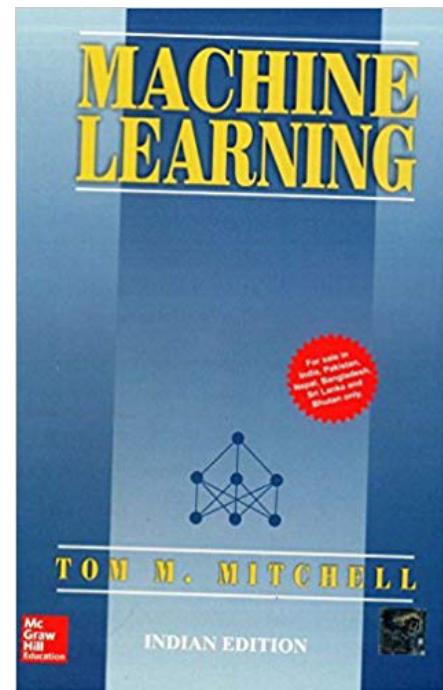
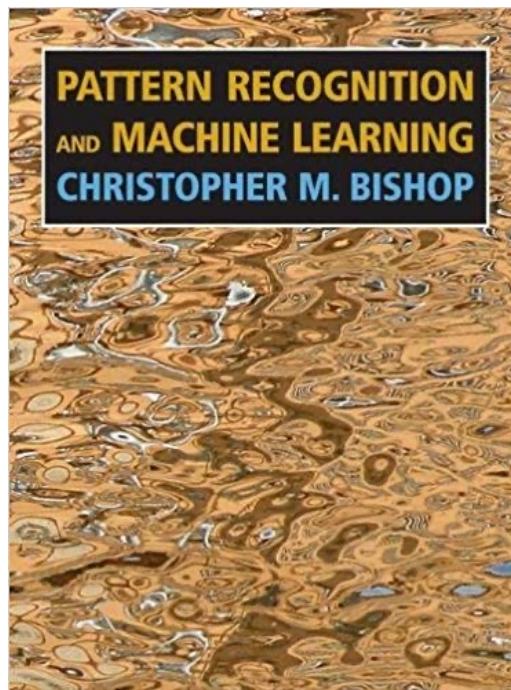
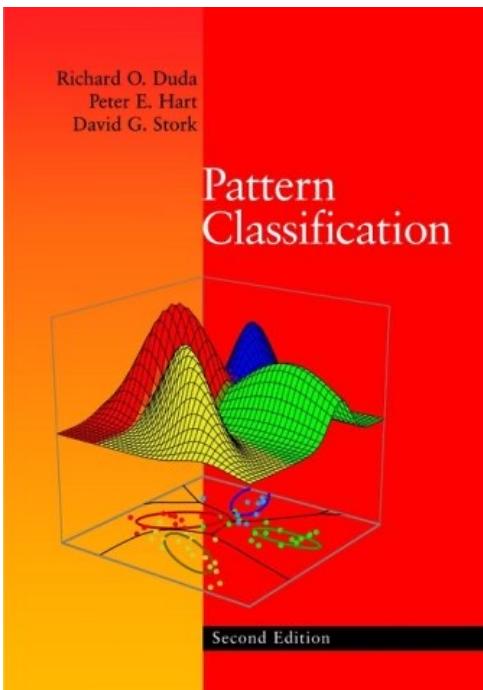
# About the course – Grading Policy

- **Grades Policy:**
  - Largest gaps in sorted order typically form the grade boundaries (A,A-,B,B- etc.).
  - Alternate gap locations may be used (subjective!)

# About the course – Attendance Policy

- IIIT's policy: < 85% attendance → 1 grade drop

# About the course - Textbooks

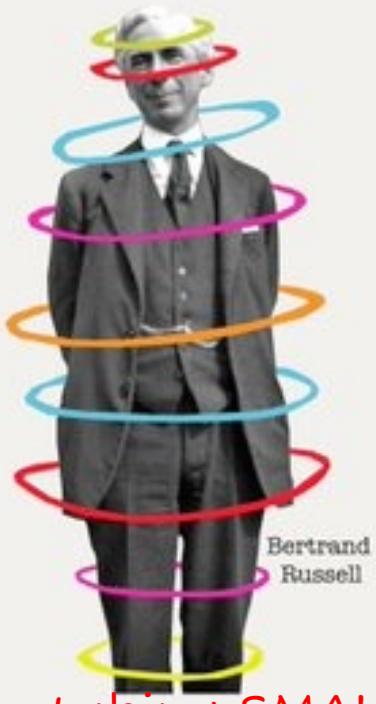


# About the course - Material

- Will be provided on a per lecture basis
- Scattered Resources across Internet

# Survey

- For those **seriously** planning to take the course ...
- Take the anonymous survey:  
<https://forms.gle/maNnF9eipqcMZcccA>
- Deadline to submit survey: **3rd August 2023**
- ... To understand your background
- ... Will help tailor the course content



taking SMAI

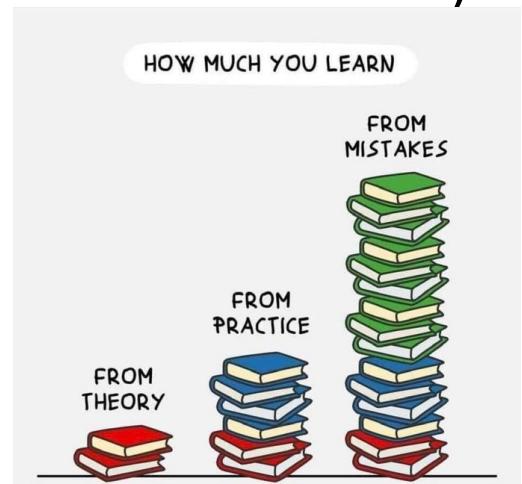
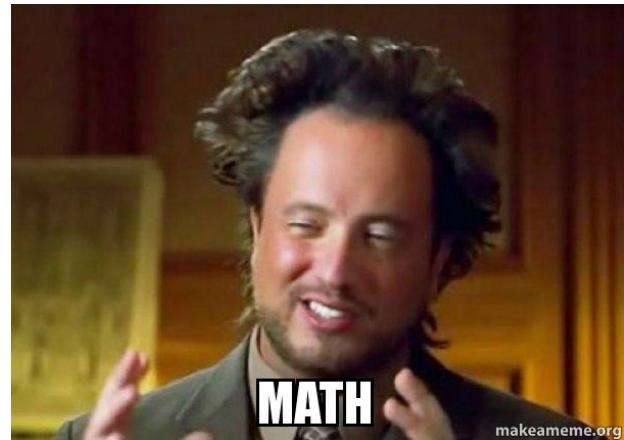
There are **two motives** for reading a book; one, that you enjoy it;  
the other, that you can ~~boast about~~ mention **in your CV**

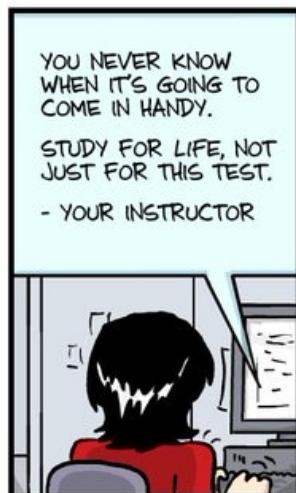
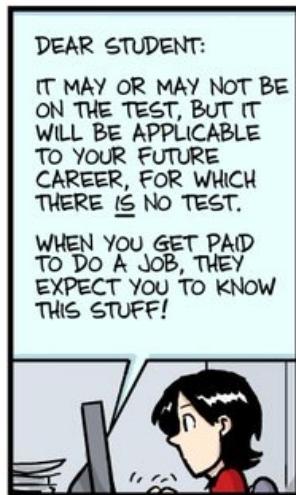
# Feedback and Office Hours

- Do not wait until end of course!
- Office Hours:
  - Sat 10AM – 11AM
  - Or by appointment
- Anonymous emails (e.g. [worriedSMAI@gmail.com](mailto:worriedSMAI@gmail.com)) are fine, but do not send gossip/spam ☺

# Additionally ...

- **Understand**, don't just memorize
- Love the math, not just the toolbox !
- Capture the broad ideas and insights (useful years down the line)
- Implement ! No substitute for experience.







Teresa Kubacka  
@paniterka\_ch

...

What I'm angry about is that this is the exact opposite of what the Explainable AI movement has been trying to achieve. Code Interpreter is the ultimate black box AI, which you cannot debug, cannot tweak, cannot rely on, but which very confidently delivers "interesting insights".

2:22 AM · Jul 11, 2023 · 238 Views



Hassan Hayat 🔥 @TheSeaMouse · Jul 9

"Future of programming is english... why do I need to learn any of this when I can prompt ChatGPT..."

...



Ask ChatGPT for a CNN on the MNIST dataset and then swap it with your own images. Teach the model whatever you want. ChatGPT is here to help you not replace you.



1



8



1,187



Hassan Hayat 🔥 @TheSeaMouse · Jul 9

"But can't I just call a model from HuggingFace?"

...



Sure, you will probably do that, but what if it breaks or is slightly off, how would you fix it? How would you modify it? Learn the fundamentals, you will be armed with the confidence to move forward when things inevitably break



1



5



985



Hassan Hayat 🔥 @TheSeaMouse · Jul 9

"Don't I need super advanced calculus math to do AI"

...



Do you know recursive functions? If so, you are better at math than 99% of all humans ever. Don't rush towards the math, pick a problem and solve it. You will naturally pick up the skills you need on the way.



1



8



864



# A tale of two airplanes



"The Gimli Glider – 30 years later"

[https://www.youtube.com/watch?v=sbVG\\_6yKLM8](https://www.youtube.com/watch?v=sbVG_6yKLM8)



"Fatal Flight 447: Chaos in the Cockpit"

<https://youtu.be/jM3CwBYX-ms>



OML

@OffensiveML

Just deal with it!  
#MachineLearning

```
1
2 import pandas as np
3 import torch as tf
4 import numpy as pd
5 import matplotlib.pyplot as nn
6 import sklearn as plt
```



A black and white photograph of Marvin Minsky, a man with a shaved head and glasses, wearing a light-colored sweater. He is holding a large astronomical telescope. The background is dark.

**"No computer has  
ever been designed  
that is ever aware  
of what it's doing.**

**But most of the  
time, we aren't  
either."**

***Marvin Minsky*  
1927-2016**