CS4780/5780 - Machine Learning

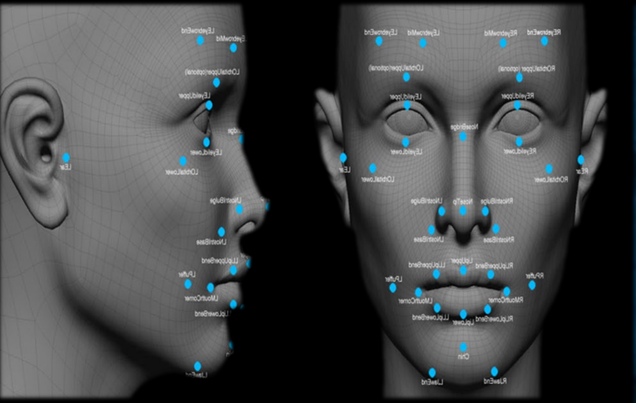
Fall 2019

Nika Haghtalab & Thorsten Joachims
Cornell University
Department of Computer Science

Reading: UML Chapter 1

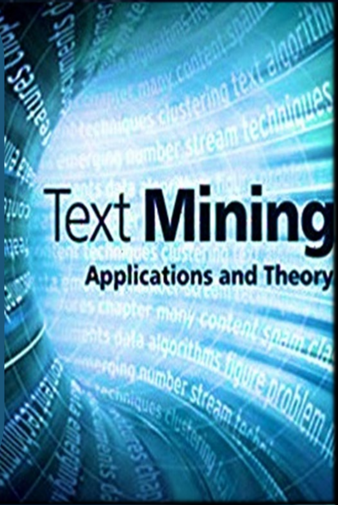
Outline of Today

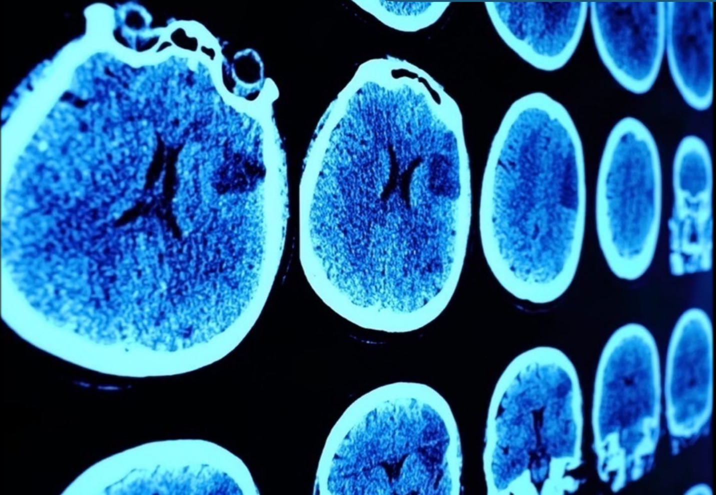
- Who we are?
  - Prof: Nika Haghtalab & Thorsten Joachims
  - TAs: Aman Agarwal, Himank Yadav, Jerry Chee, Lucas Chen, DB Lee, Nandini Nayar
  - Consultants: Many – you will get to know them.
- What is machine learning?
- Syllabus
- Administrivia

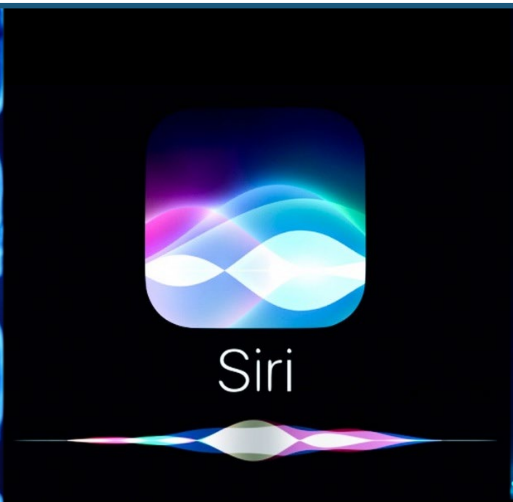


Machine Learning (ML)

Programs that improve with experience.







Revolutionizing Science and Technology

“A breakthrough in machine learning would be worth ten Microsofts.” (Bill Gates, Microsoft)

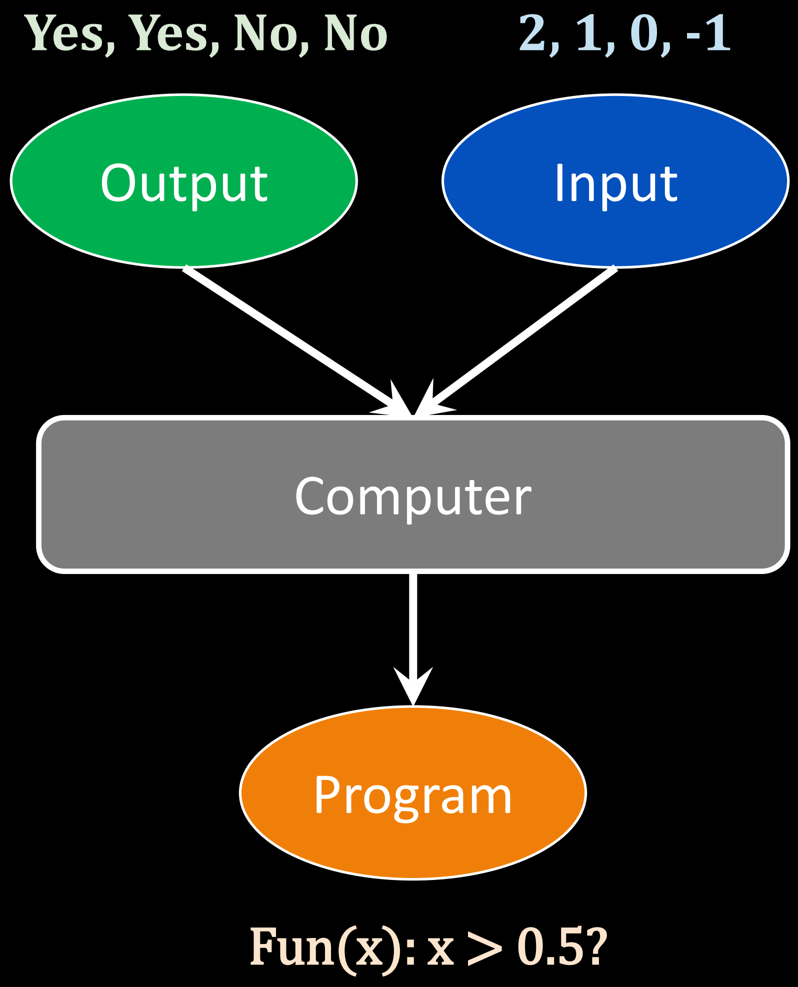
“It will be the basis and fundamentals of every successful huge IPO win in 5 years.” (Eric Schmidt, Google / Alphabet

“AI and machine learning are going to change the world and we really have not begun to scratch the surface.” (Jennifer Chayes, Microsoft / Berkeley)

“ML is transforming sector after sector of the economy, and the rate of progress only seems to be accelerating.” (Daphne Koller, Stanford / Coursera/ Insitro)

“Machine learning is the next Internet” (Tony Tether, DARPA)

What is Machine Learning?



Yes, Yes, No, No

2, 1, 0, -1

Output

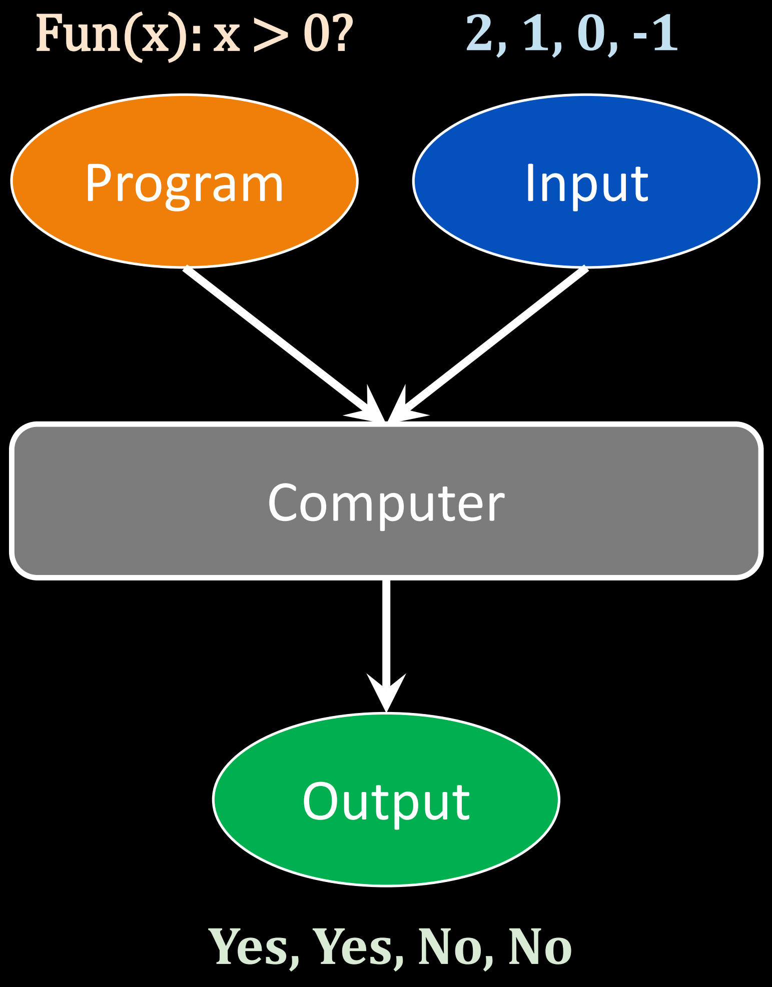
Input

Computer

Program

Fun(x): x > 0.5?

Machine Learning



Fun(x): x > 0?

2, 1, 0, -1

Program

Input

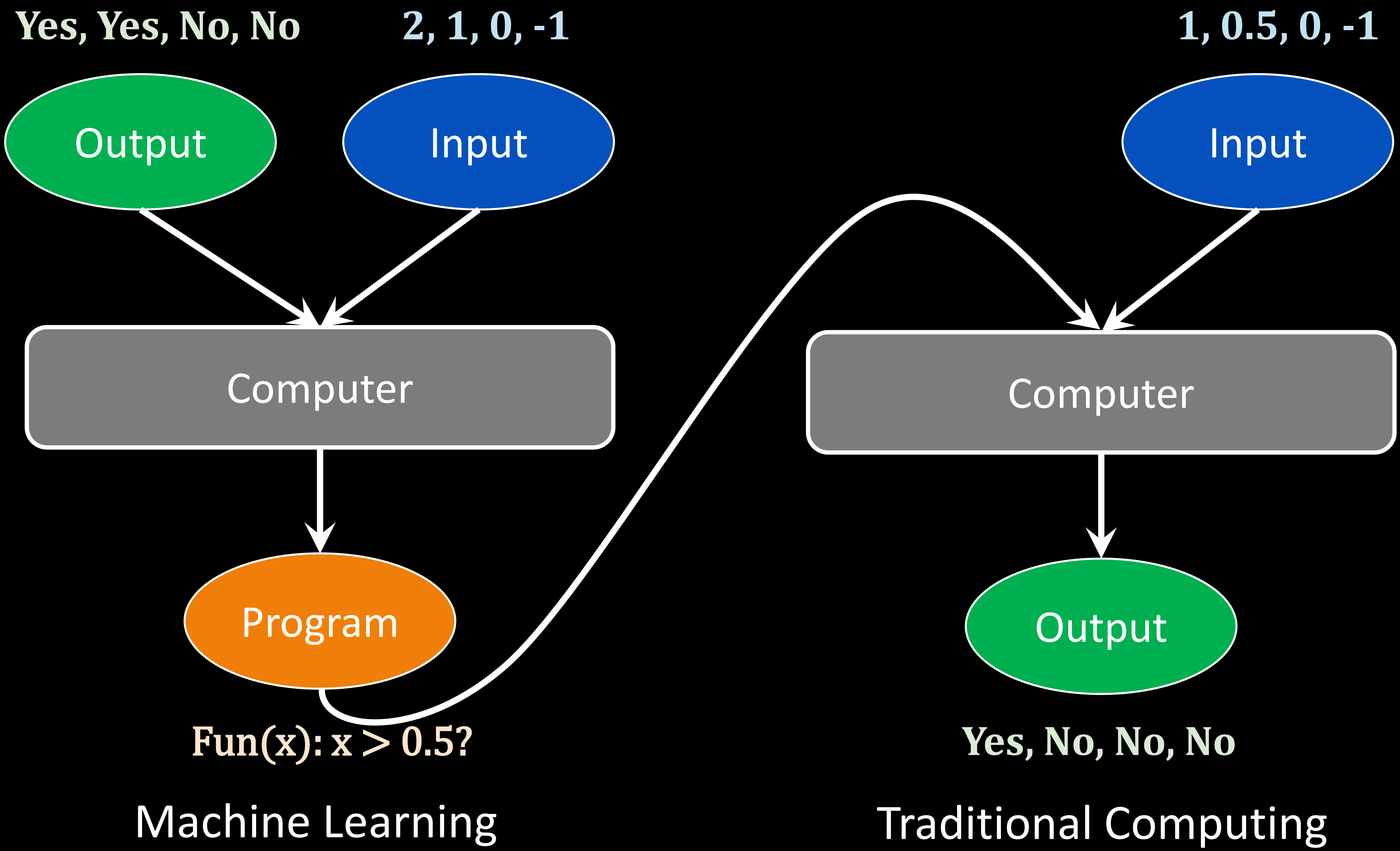
Computer

Output

Yes, Yes, No, No

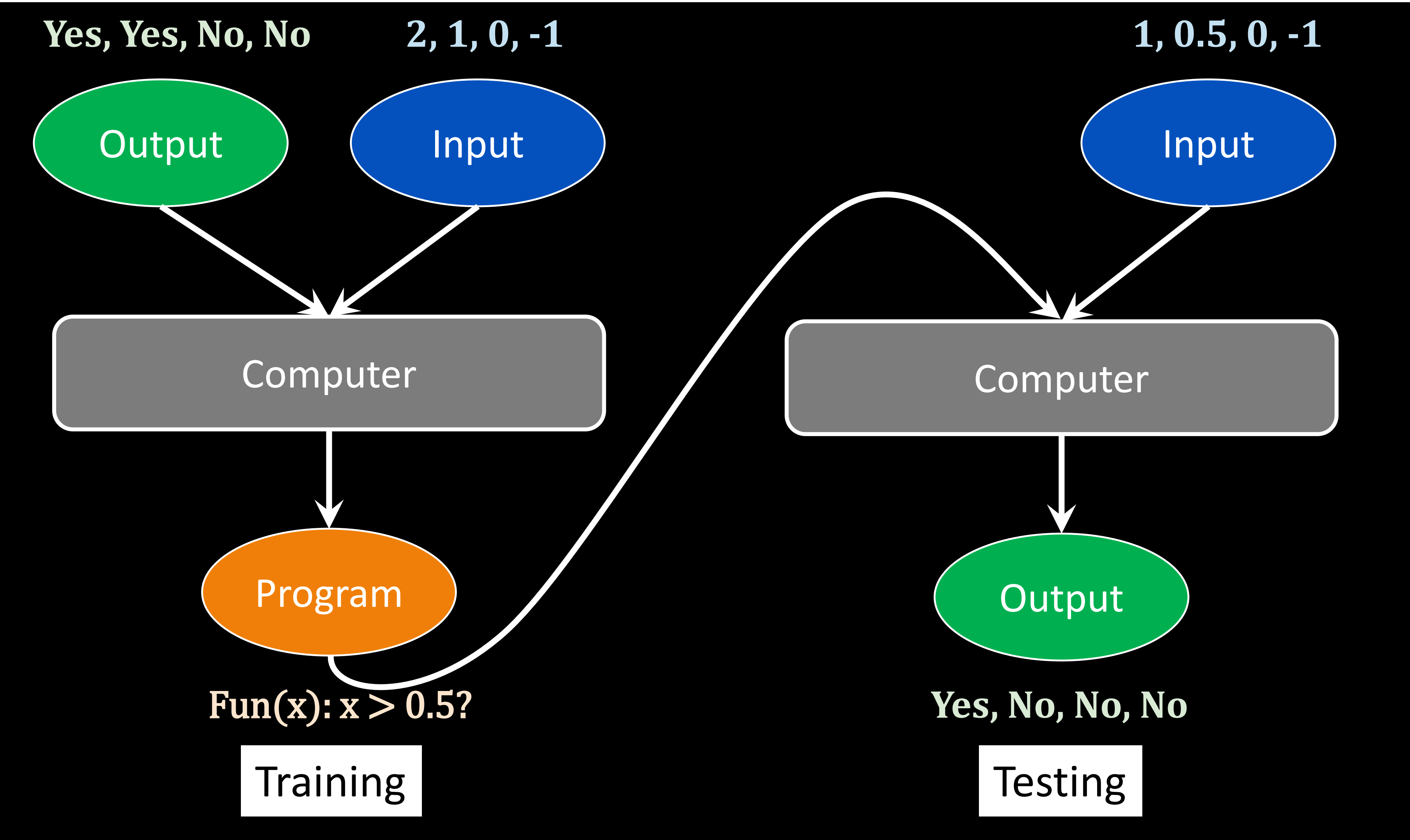
Traditional Computing

What is Machine Learning?



Tom Mitchell, 1997:

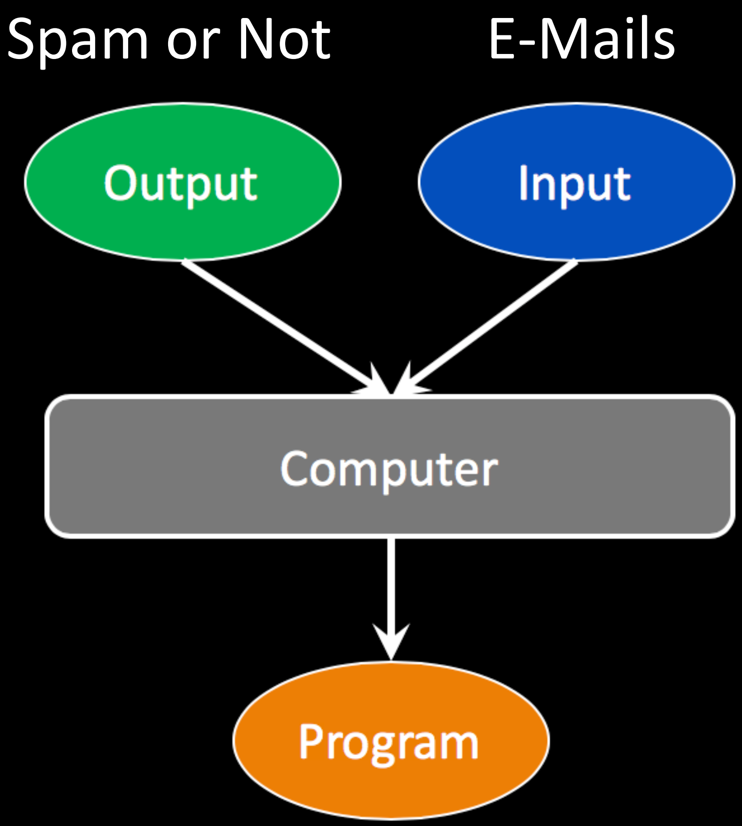
A computer program **A** is said to learn from experience **E** with respect to some class of tasks **T** and performance measure **P**, if its performance at tasks in **T**, as measured by **P**, improves with experience **E**.



Learning to Detect Spam

Use past emails and whether or not they were flagged as spam.

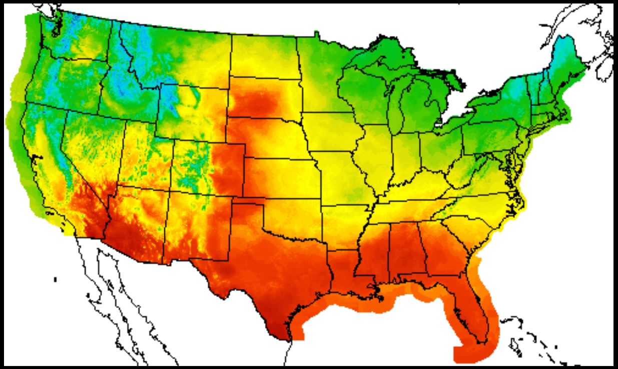
Learn a program that takes a future email and decides whether it is a spam:
E.g. If the email is from an unknown sender, has a misspelling, and has "Million Dollars" in it flag as spam.



Applications of ML

Use past data to …

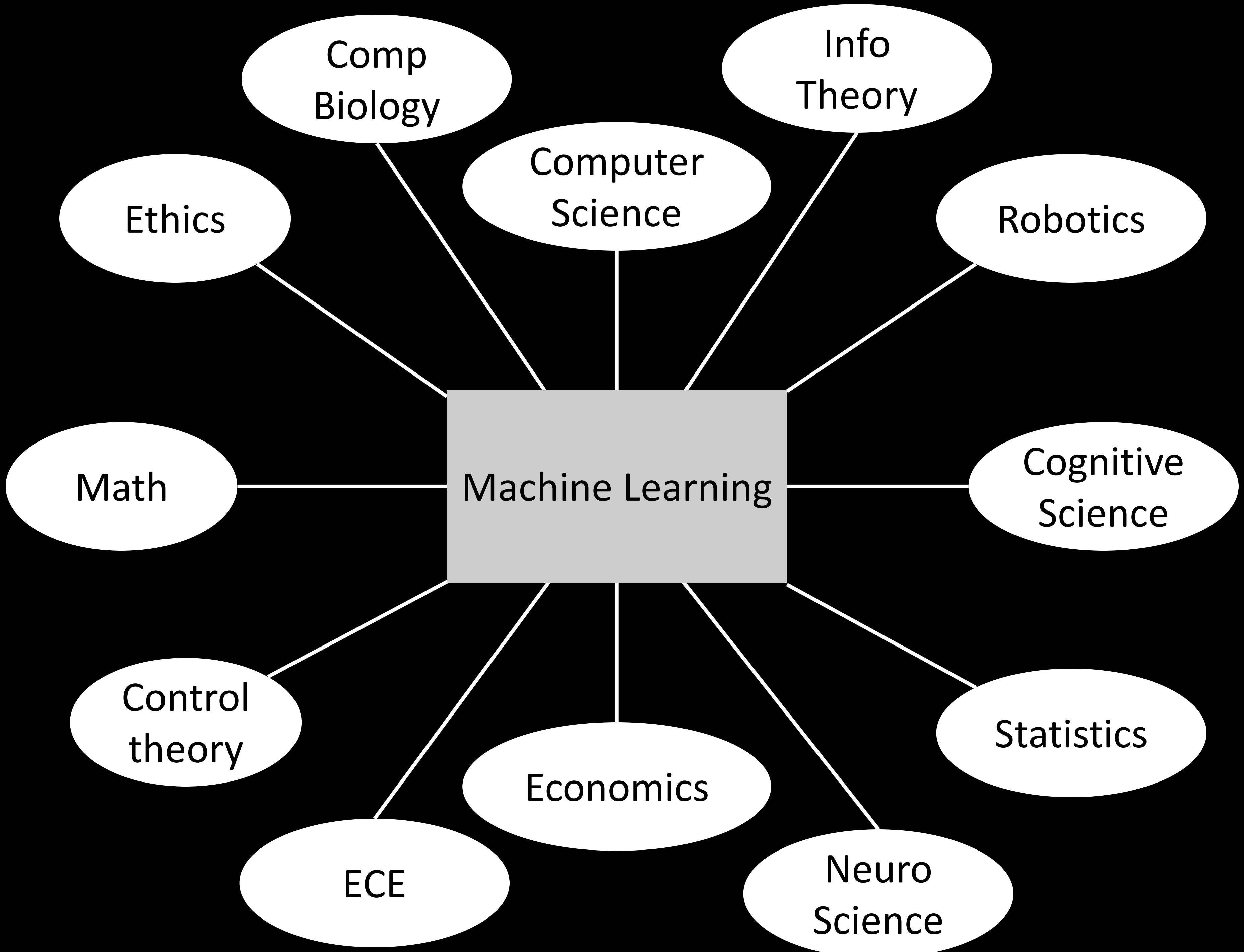
Detect spam



Predict weather

Classify images

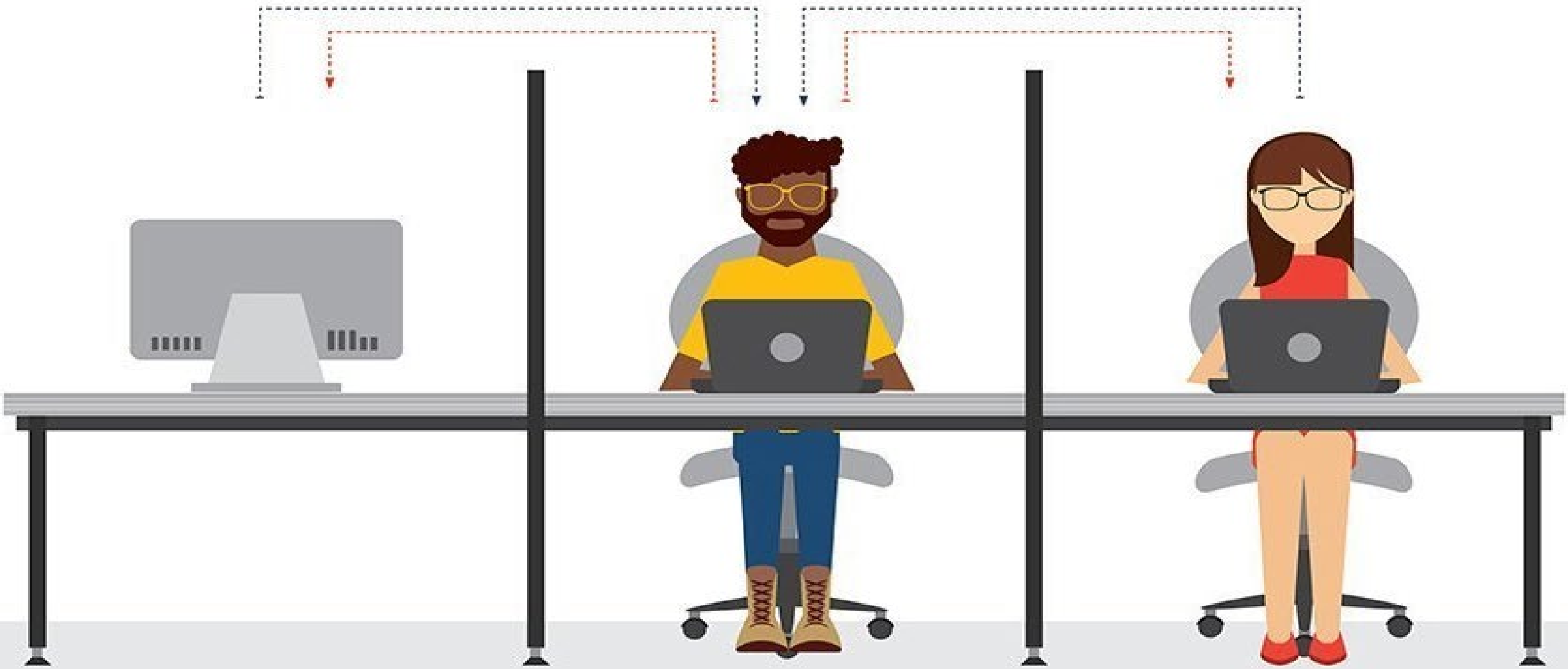
<u>Other Examples:</u>
Fraud Detection, Flagging inappropriate social media posts, Natural Language Processing, Document Classification, Designing Economic Mechanisms, Computational Advertising, …


Comp Biology
Info Theory
Computer Science
Ethics
Robotics
Math
Machine Learning
Cognitive Science
Control theory
Statistics
Economics
ECE
Neuro Science

Alan Turing

The Turing Test, 1950

A machine is intelligent if its answers are indistinguishable from a human's.



Checkers Program, 1952

Arthur Samuel

Created a Checkers-playing program that got better overtime.

Also introduced the term “Machine Learning”.

Frank Rosenblatt

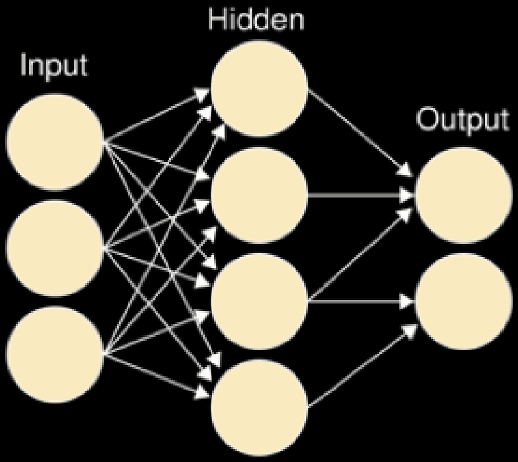
@ Cornell!

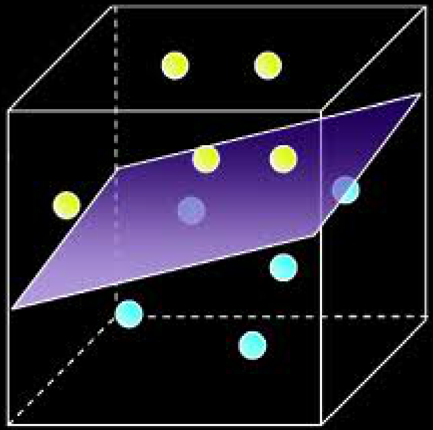
Perceptron, 1957

Predecessor of deep networks.

Separating two classes of objects using a linear threshold classifier.

Provable learning and convergence guarantees.





1960s: Lots of hope for AI to solve everything!

AI didn’t live up to the hype!

- 1966: Machine Translation failed.
- 1970: Minsky and Papert argued against Perceptron.
- 1971: Speech Understanding failed.
- 1973: Lighthill report torn apart AI.

> “In no part of the field have the discoveries made so far produced the major impact that was then promised”

- 1974: The UK and US stopped funding AI research.

The AI Winter, 1974-1980

Rebirth as Machine Learning

Machine Learning:

- Originally, a bit of a name game to get funding.
- Fundamentally a different approach to intelligence:

<u>Machine Learning</u>	<u>Artificial Intelligence</u>
Data-driven	Knowledge-based
Bottom-up approach	Heavy use of logic
	Top-down approach

Foundations of ML, 1980s-present

Formal notions of learnability from Data.

- When data-driven learning is possible?
  → Probably Approximately Correct Learning (PAC) by Valiant.
  → How much data is required?
- What’s the difference between great and mediocre learners?
  → Improving the performance of a learning algorithm.
  → Boosting algorithm of Freund and Schapire.
- How to deal with difficult and noisy learning problems?
  → (Soft Margin) Support Vector Machines by Cortes and Vapnik
- What to do when the learning task evolves over time?
  → Online learning framework.

TD-Gammon, 1992

Gerald Tesauro at IBM thought a neural network to play Backgammon.

The net played 100K+ games **against itself** and beat the world champion.

Algorithm found new techniques that people had erroneously ruled out.

Deep Blue, 1997

IBM's Deep Blue won against Kasparov in chess.

The crucial winning move was made due to machine learning methods developed by Gerald Tesauro.

Expanding the reach, 2000s

Learning to rank

→ Powering search engines: Google, Bing, …

Topic Modeling:

→ Detecting and organizing documents by subject matter.

→ Making sense of the unstructured data on the web.

Online economy:

→ Ad placement and pricing.

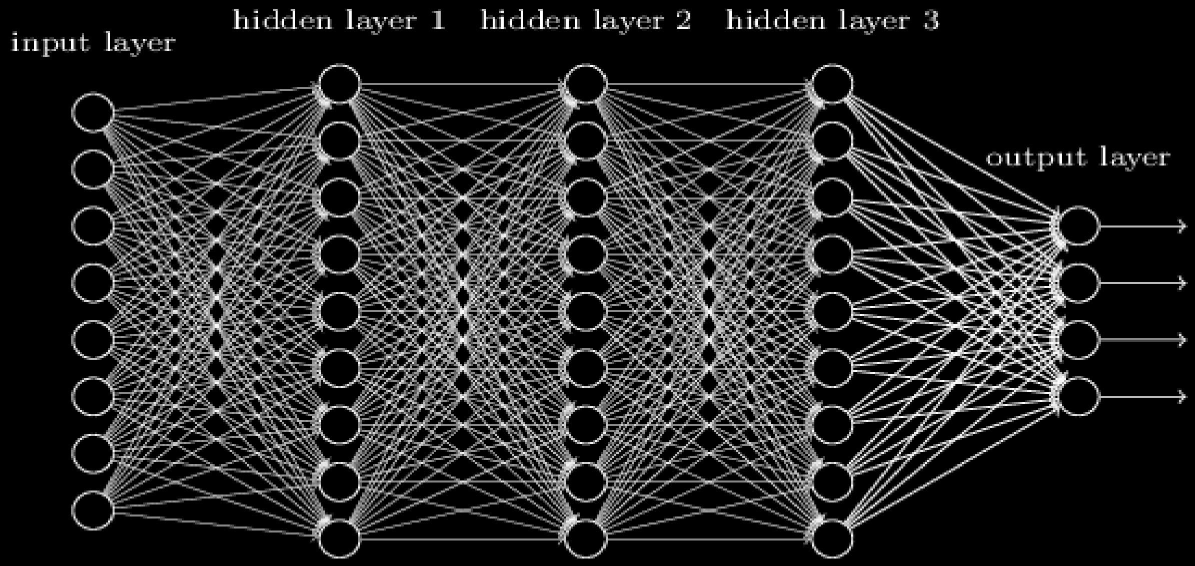
→ Product recommendation.

Machine learning became profitable!

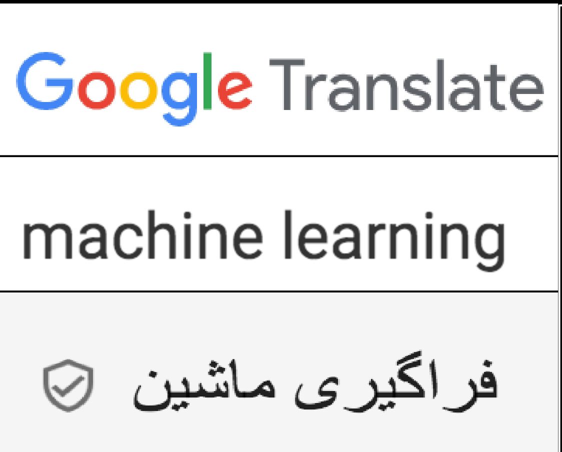
Return of Neural Networks, 2010s

Neural networks return and excel at image recognition, speech recognition, …

The 2018 Turing award was given to Yoshua Bengio, Geoff Hinton, and Yann LeCun.

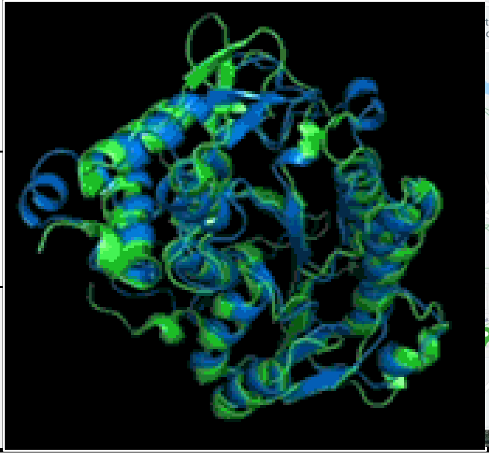


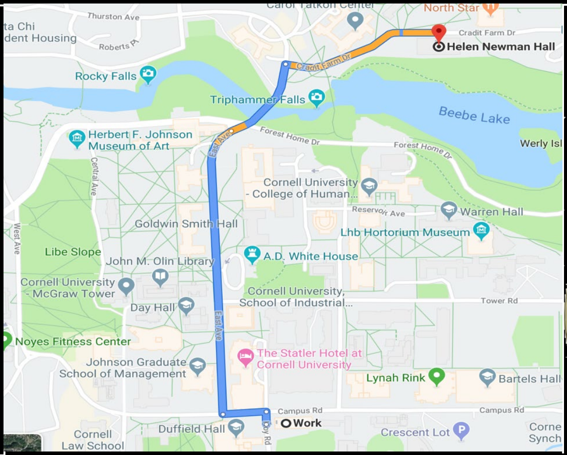
Surrounded by Machine Learning

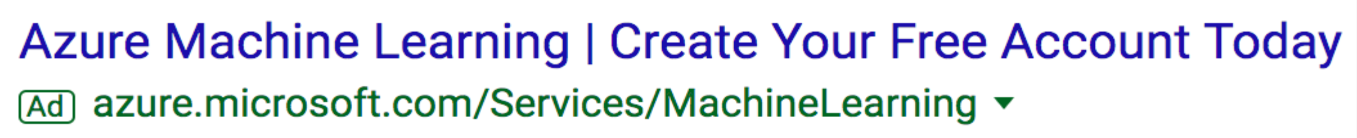

Google Translate

machine learning

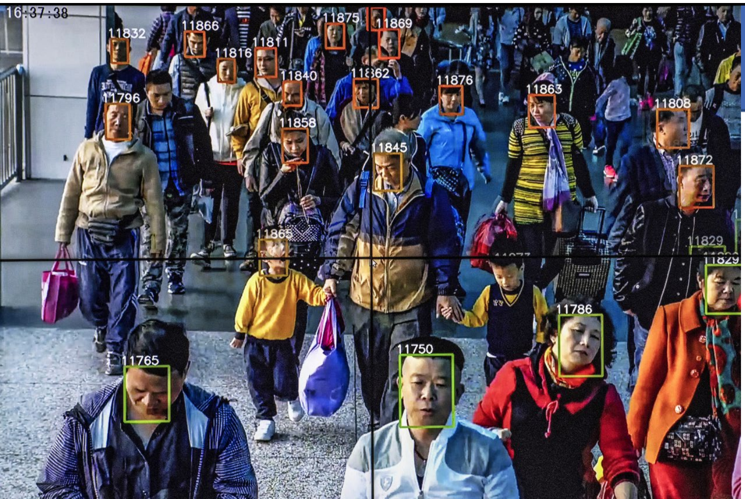
فراگیری ماشین

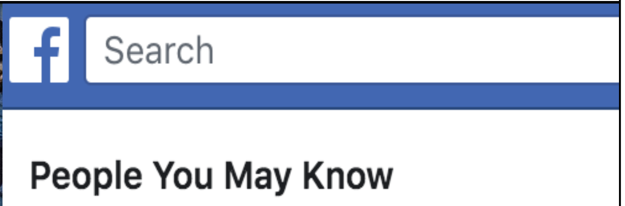





Azure Machine Learning | Create Your Free Account Today

Ad azure.microsoft.com/Services/MachineLearning ▾




Search

People You May Know

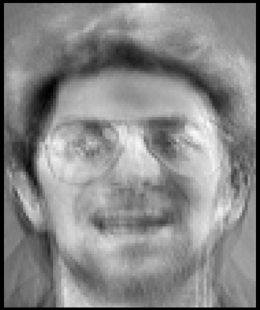
Nika Haghtalab

Your instructor?

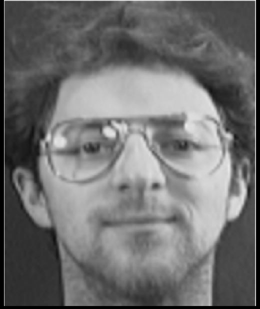
“With great power, there must also come
– great responsibility!”

Data Privacy

Learning models leak training data (Fredrickson et al. ‘15)



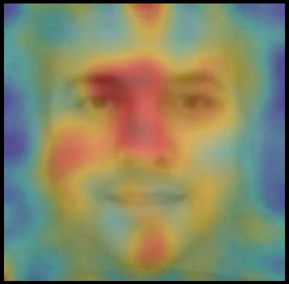
Leaked data

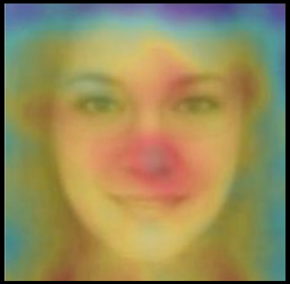


Real image

Learning algorithms detect sexual orientation better than people

(Wang & Kosinski’17)





Formal definitions of data privacy:

- K- anonymity (Sweeney)
- Differential Privacy (Dwork, McSherry, Nissim, Smith).

Latanya Sweeney

Cynthia Dwork

Frank McSherry

Kobbi Nissim

Adam Smith

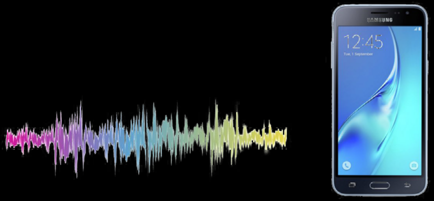
Robust and Secure ML

Image Recognition

Misreading traffic signs (Eykholt et al)

Speech recognition

Hide commands in noise (Carlini & Wagner)



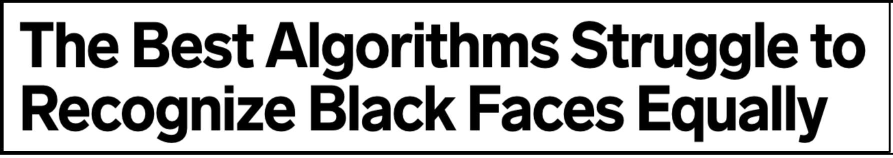
Poisoning Attacks

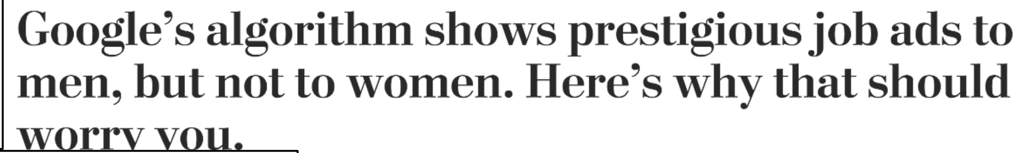
Tay (chat bot) became inflammatory in 16 hr.

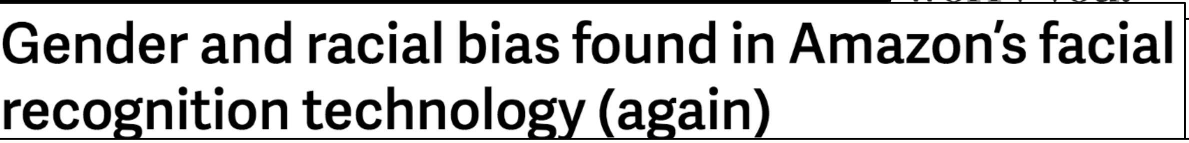
How to create robust and secure machine learning algorithms?

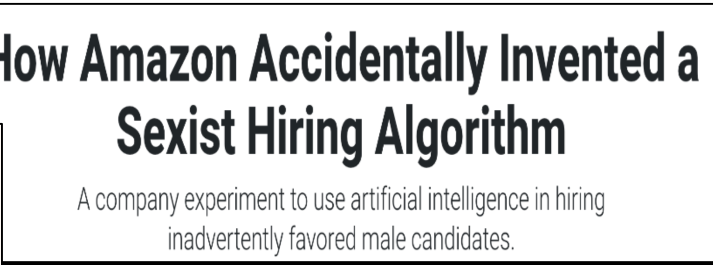
Learning and the Society

- Bad dynamics, perpetuating and worsening stereotypes and biases.
- Who carries the burden of bad prediction?
- How to design good dynamics?

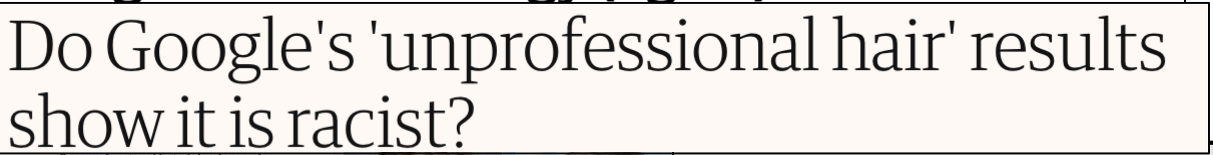

The Best Algorithms Struggle to Recognize Black Faces Equally

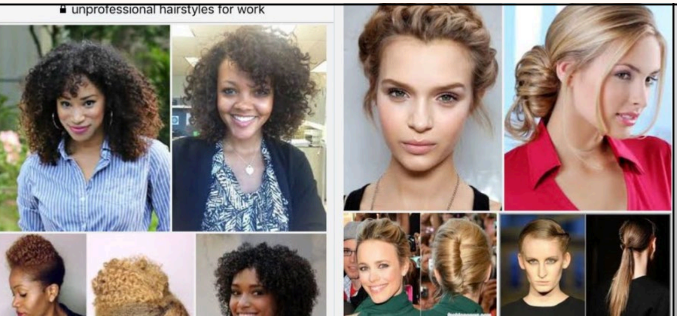

Google's algorithm shows prestigious job ads to men, but not to women. Here's why that should worry you.

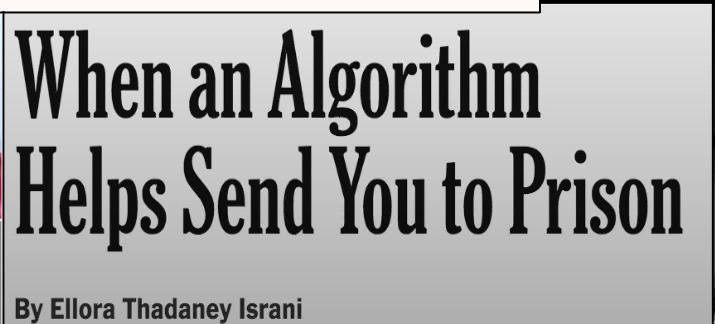

Gender and racial bias found in Amazon's facial recognition technology (again)


How Amazon Accidentally Invented a Sexist Hiring Algorithm

A company experiment to use artificial intelligence in hiring inadvertently favored male candidates.


Do Google's 'unprofessional hair' results show it is racist?




When an Algorithm Helps Send You to Prison

By Ellora Thadaney Israni

Challenging Questions

Machine learning and Artificial Intelligence will shape the future, what kind of a future do we want?

What is the role of machine learning?

→ ML for good versus ML for profit.

How do automation and learning change the quality of life?

→ Job loss and displacement, life satisfaction, safety and security?

How do we approach machine learning and (inter-)national security? Weaponization of machine learning and AI?

This Course

Syllabus

- **Supervised Batch Learning**: decision theoretic foundation, model selection, model assessment, empirical risk minimization
- **Instance-based Learning**: K-Nearest Neighbors, collaborative filtering
- **Decision Trees**: TDIDT, attribute selection, pruning and overfitting
- **Linear Rules**: Perceptron, logistic regression, linear regression
- **Support Vector Machines**: Optimal hyperplane, margin, duality, kernels, stability
- **Deep Learning**: multi-layer perceptrons, deep networks, stochastic gradient
- **Generative Models**: generative vs. discriminative, naive Bayes, linear discriminant analysis
- **Structured Output Prediction**: predicting sequences, hidden markov model, rankings
- **Statistical Learning Theory**: generalization error bounds, VC dimension
- **Online Learning**: experts, bandits, online mistake bounds

→ Understand ML beyond the individual algorithms (theory, design, use)

Related Courses

- Follow-up Courses
  - CS4786: Machine Learning for Data Science
  - CS4787: Principles of Large-Scale Machine Learning Systems
- Related Courses
  - CS4700: Foundations of Artificial Intelligence
  - CS4850: Mathematical Foundations for the Information Age
  - CS4300: Language and Information
  - CS4740: Natural Language Processing
  - CS6780: Advanced Machine Learning
  - CS6784: Advanced Topics in Machine Learning
  - CS6740: Advanced Language Technologies
  - More courses in Robotics, Computer Vision, etc.

Pre-Requisites

- Pre-Requisites
  - Programming skills (e.g. CS 2110)
  - Basic linear algebra (e.g. MATH 2940)
  - Basic multi-variable calculus
  - Basic probability theory (e.g. CS 2800)
- Pre-Requisite Assessment
  - Multiple choice, 1% of final course grade
  - Get real about whether you are ready for this class
  - Available on Gradescope (via course homepage)
  - Due on Tuesday, Sep 3, at noon.
    - Everybody gets a 2 day extension (Thursday, Sep 5, at noon)

Textbook and Course Material

- Main Textbooks
  - Shai Shalev-Shwartz, Shai Ben-David, "Understanding Machine Learning - From Theory to Algorithms", Cambridge University Press, 2014. (online PDF)
- Additional References (optional)
  - Kevin Murphy, “Machine Learning – a Probabilistic Perspective”, MIT Press, 2012. (online via CU Library)
  - See other references on course web page
- Course Notes
  - Slides available on course homepage
  - Writing on whiteboard
  - Video of lectures available on course homepage

Homework Assignments

- Assignments
  - Max 5 homework assignments – typically 1 week.
  - Problem sets without programming.
- Policies
  - Assignments are due via Gradescope (see course homepage).
  - Assignments turned in late will be charged a 1 percentage point reduction of the cumulated final homework grade for each period of 24 hours for which the assignment is late.
  - Everybody has 5 “free” late days. Use them wisely.
  - No assignments will be accepted after the solutions have been made available (typically 3-4 days after deadline).
  - Typically collaboration of 2-3 students (see each assignment for detailed collaboration policy).
  - We run automatic cheating detection. Must state all sources of material used. Please review Cornell Academic Integrity Policy!

Programming Projects

- Assignments
  - Max 8 programming projects – typically 1 week each.
  - Typically programming with (mostly) auto-grading for mastery.
- Policies
  - Assignments are distributed and submitted via Vocareum.
  - You will get an invite to Vocareum on Monday/Tuesday ($30 fee) (if not, fill in registration form on course homepage).
  - Assignments turned in late will be charged a 1 percentage point reduction of the cumulated final homework grade for each period of 24 hours for which the assignment is late.
  - Everybody has 5 “free” late days (separate from assignments).
  - No assignments will be accepted after the solutions have been made available (typically 3-4 days after deadline).
  - Typically collaboration of 2-3 students (see each assignment for detailed collaboration policy).
  - We run automatic cheating detection. Must state all sources of material used. Please review Cornell Academic Integrity Policy!

Forming Groups

- Groups really help with learning!
  - Groups are required for both homeworks and projects.
  - Groups can be different for each homework and/or project.
- Several ways of finding group partners
  - Self-selected groups
  - WICC Event Tuesday, Sep 3, 6:00-7:00pm, Gates 3rd floor lounge
  - Automatic matching
    - Sign up via web form if you are looking for partners (with some questions about preferences)
    - Deadline typically the day after homework/project came out
    - We will send you your partners

Exams

- Midterm Exam
  - October 24, 7:30pm
- Final Exam
  - December 15, 7:00pm

If you cannot make these dates, let us know within the next 7 days.

Grading

- Deliverables

– Midterm Exam	(25% of Grade)
– Final Exam	(25% of Grade)
– Homeworks	(30% of Grade)
– Projects	(18% of Grade)
– PreReq Assessment	(1% of Grade)
– Participation	(1% of Grade)

- Outlier elimination
  - For homeworks and projects, the lowest grade is replaced by the second lowest grade.

Enrollment

- If you are not yet enrolled, sign up for the waitlist.
- If you are enrolled, but want to drop, please do so ASAP.
- Students are added from the waitlist via the policy at

  https://www.cs.cornell.edu/courseinfo/enrollment

How to Get in Touch

- Online
  - Course Homepage (slides, video, references, policies, office hours)
    - http://www.cs.cornell.edu/Courses/cs4780/2019fa/
  - Piazza forum (questions and comments, self-sign up)
  - Gradescope (homeworks and prereq assessment, self sign-up with code)
  - Vocareum (programming projects, invitations forthcoming)
- Office Hours
  - Thorsten Joachims:
    - Fridays 11:00am – 12:00pm, 418 Gates Hall
  - Nika Haghtalab
    - Wednesdays 1:30pm – 2:30pm, 315 Gates Hall
  - Other office hours:
    - See course homepage
- Reading: UML Chapter 1