CS109: Probability for Computer Scientists

Chris Piech

Teaching at Stanford

8,000+ students over 10 years

CS106A	CS106B	CS109	CS221
Programming Methodologies	Programming Abstractions	Probability for Computer Scientists	Intro to Artificial Intelligence
CURRENT	LAST: FALL 2016	LAST: FALL 2018	LAST: SUM 2013

Created a research lab in:

AI for Social Good (esp Education)

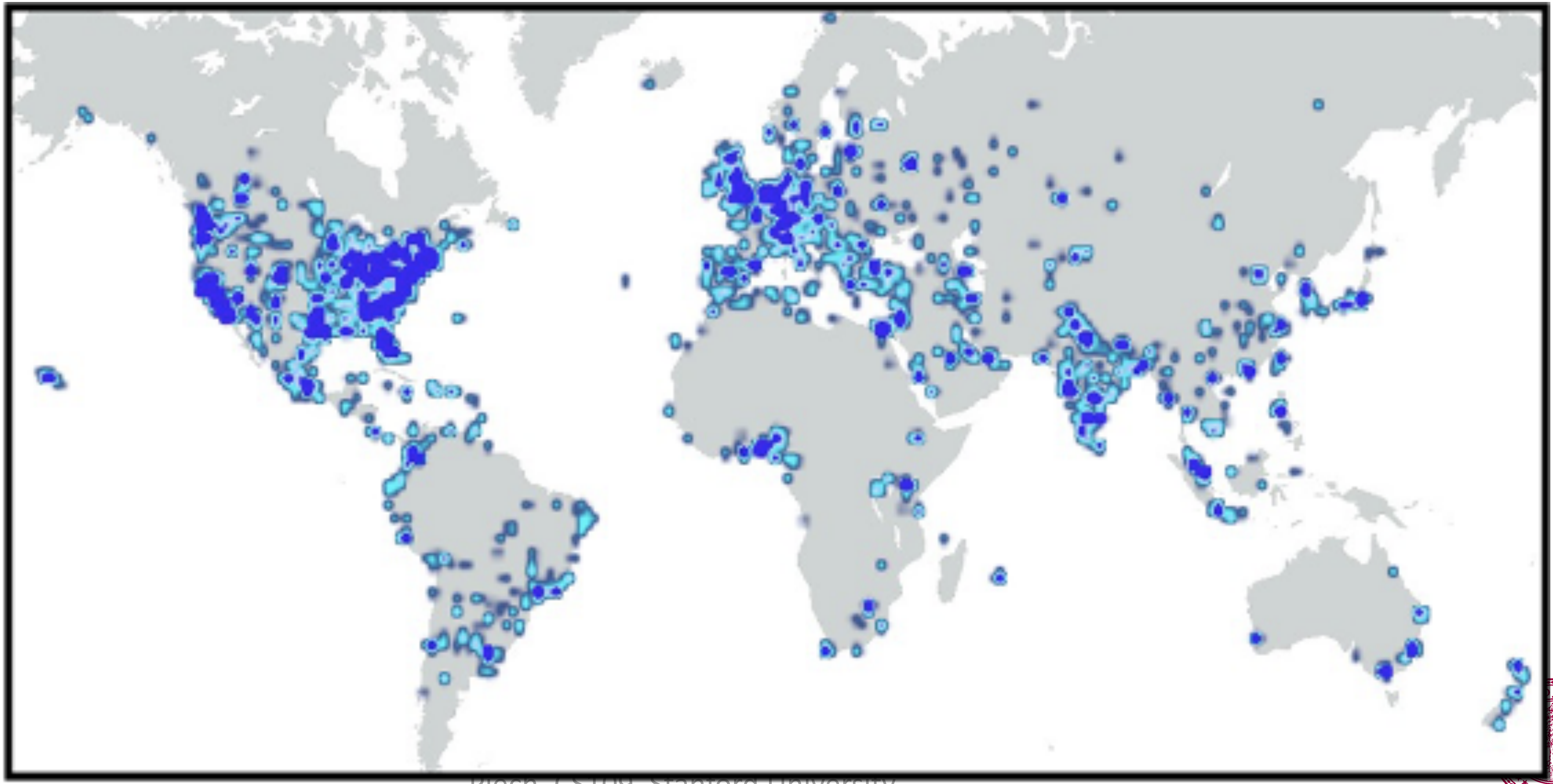
Grew up in Nairobi, Kuala Lumpur before Stanford!

Long History in CS109

I took the first CS109 back when I looked like this

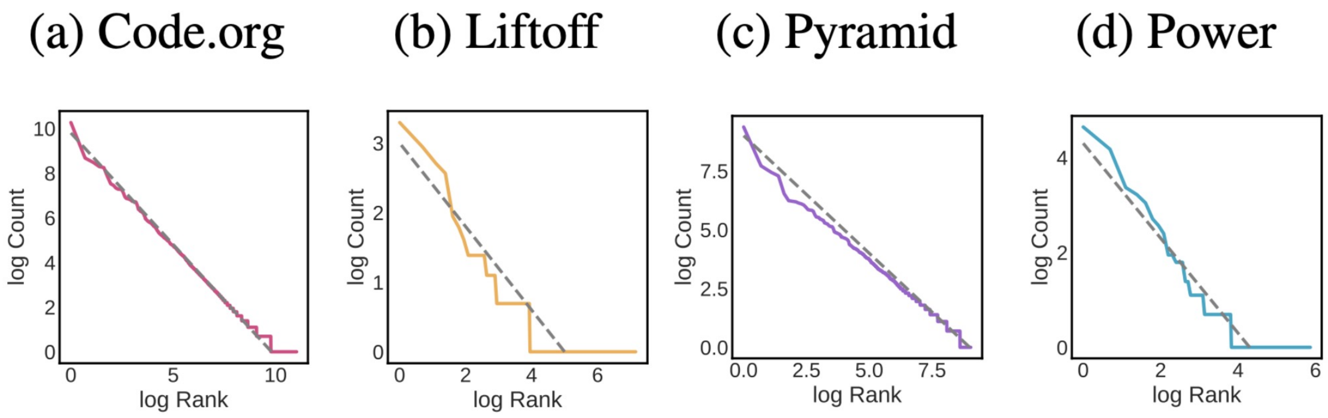
Been teaching it since 2014

What Sort of Thing Does Chris Research?

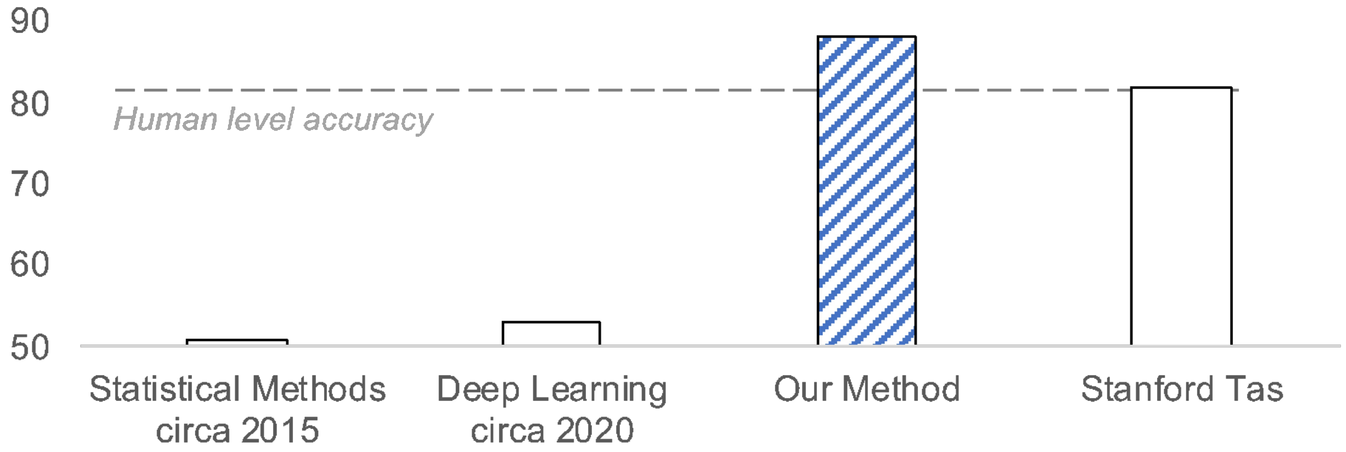


Can you “rubic grade” free response work?

Math question: Can you learn to generate contrasts that help people learn?



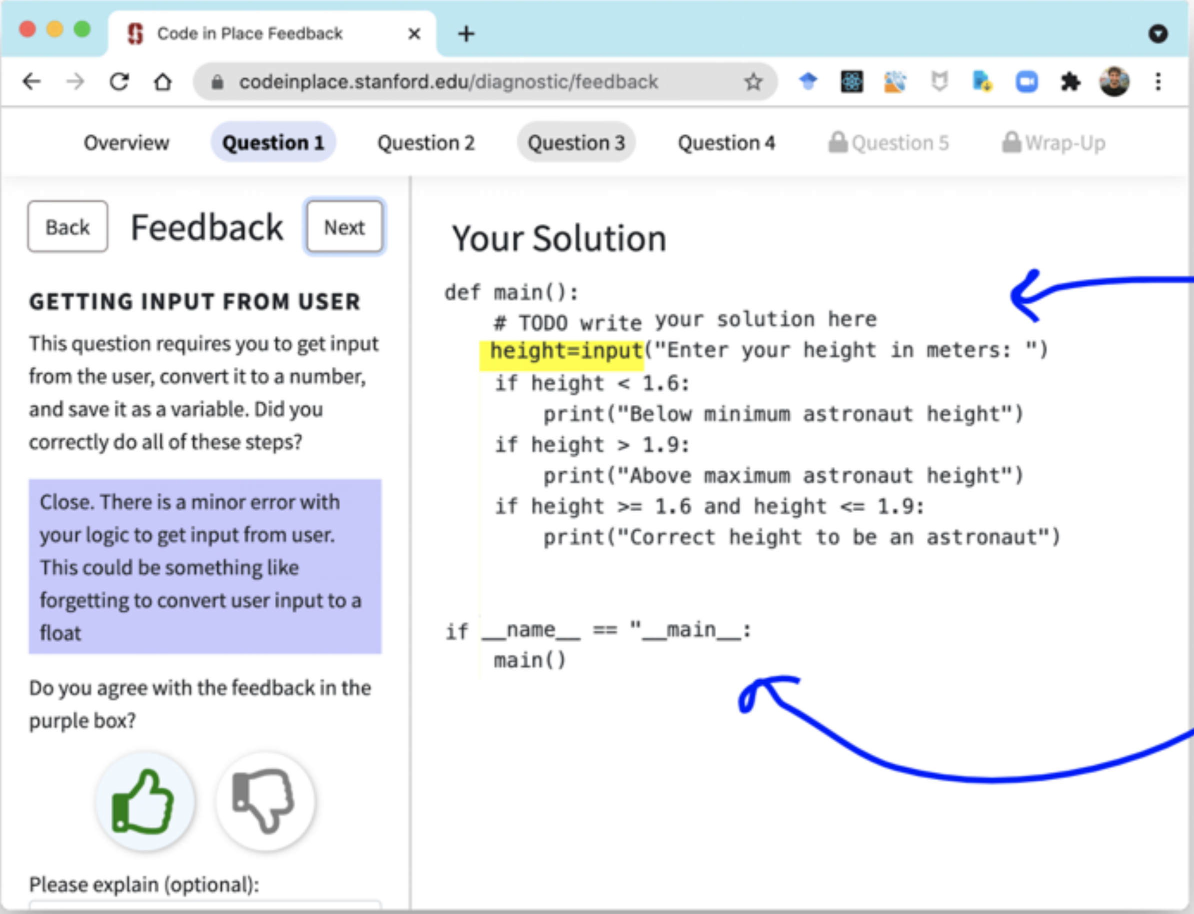
Rubric Level Accuracy on Few-Shot Novel Question Grading (% Rubric Items)



Deployed to grade 16,000 submissions in Code.org

AI generated feedback

Students evaluate the feedback

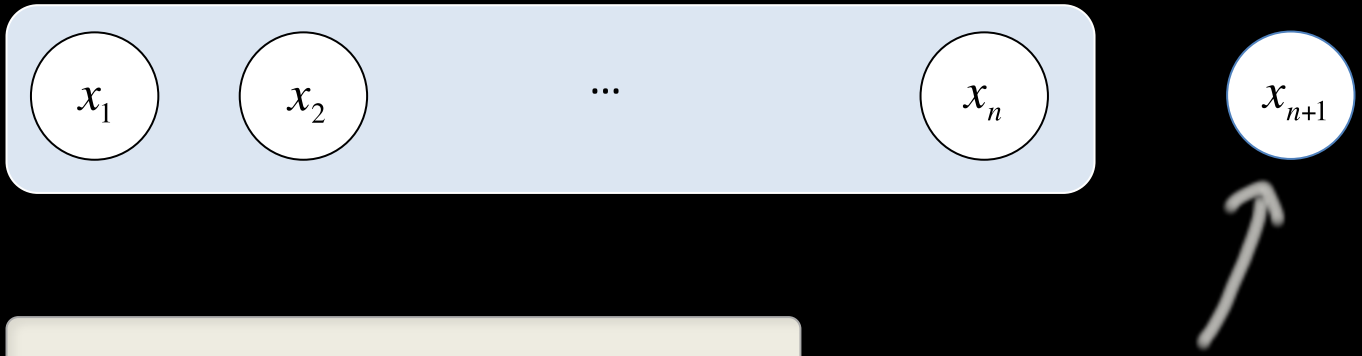


Algorithm uses attention to highlight where in the code the error comes from

Syntax error (missing ") here would prevent auto graders from being useful.

Knowledge Tracing for Feedback

Given *n* historical answers:



Predict the next one

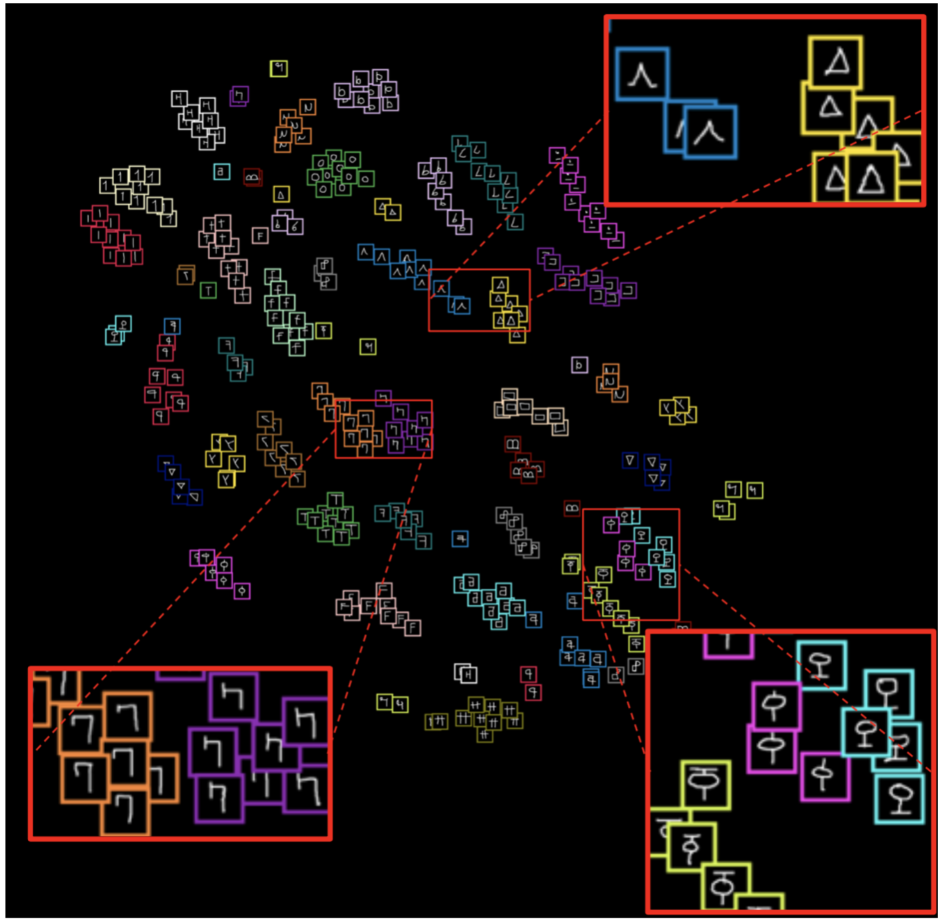
Answer is a tuple:

$$x_i = \{q_i, a_i\}$$

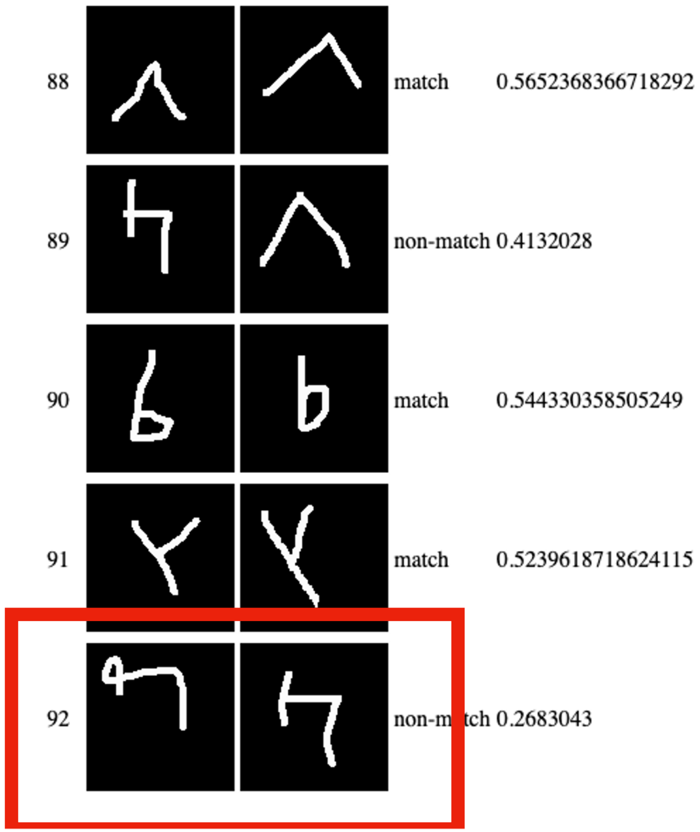
Question

Student response

More fun with probability for teaching



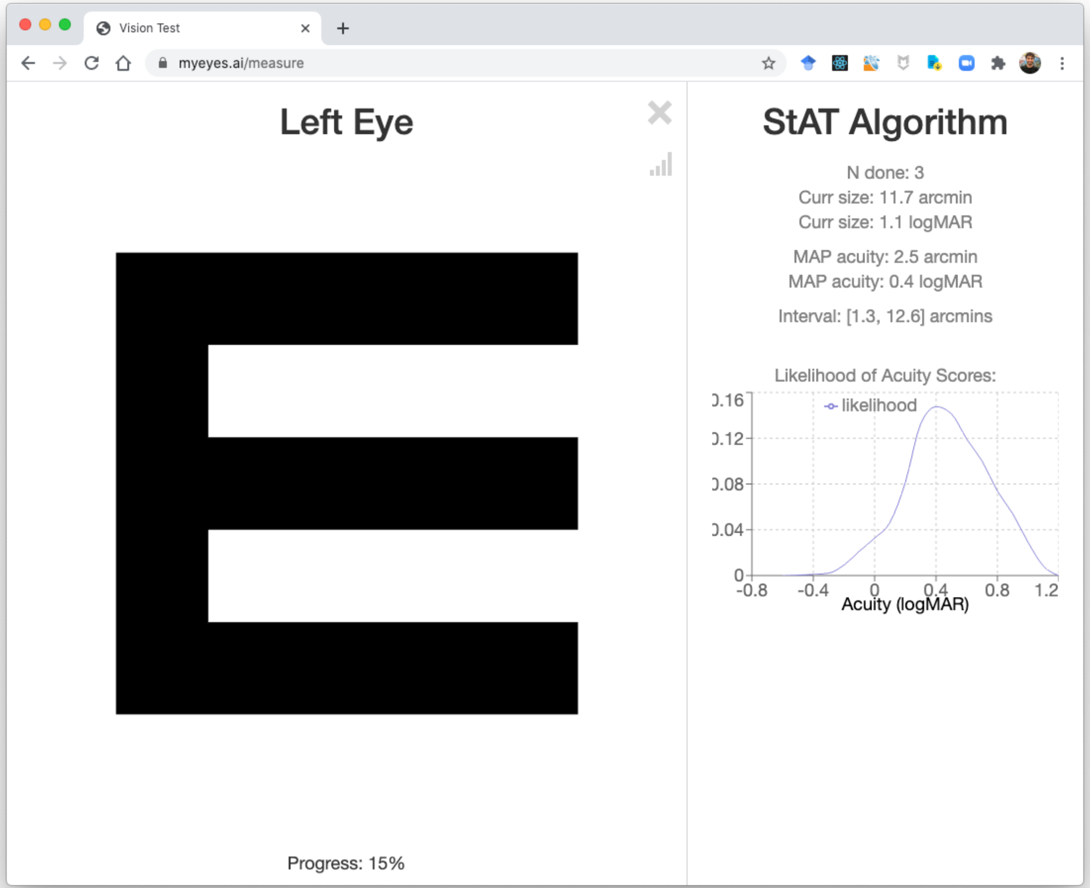
Teaching different nko characters

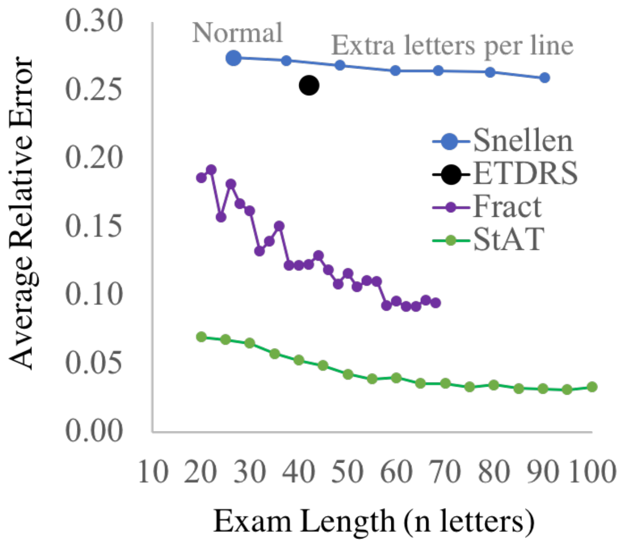


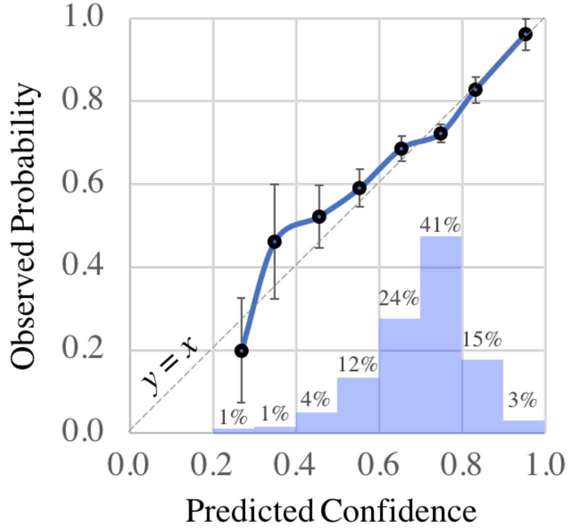
Math question: Can you learn to generate contrasts that help people learn?

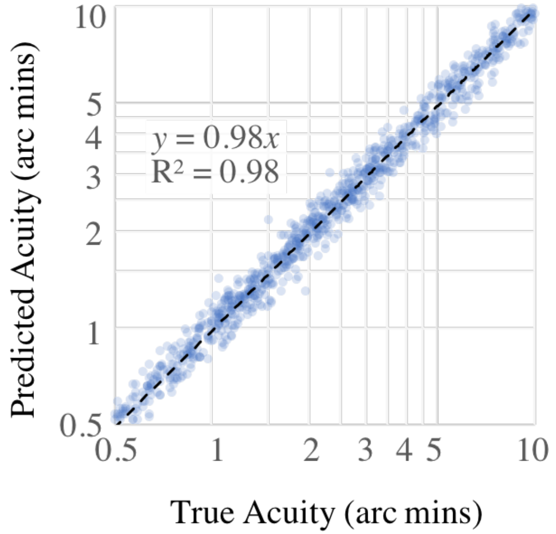
Probability in my Research: Better Eye Exam

Jan 2020, With a former CS109 student, Ali Malik





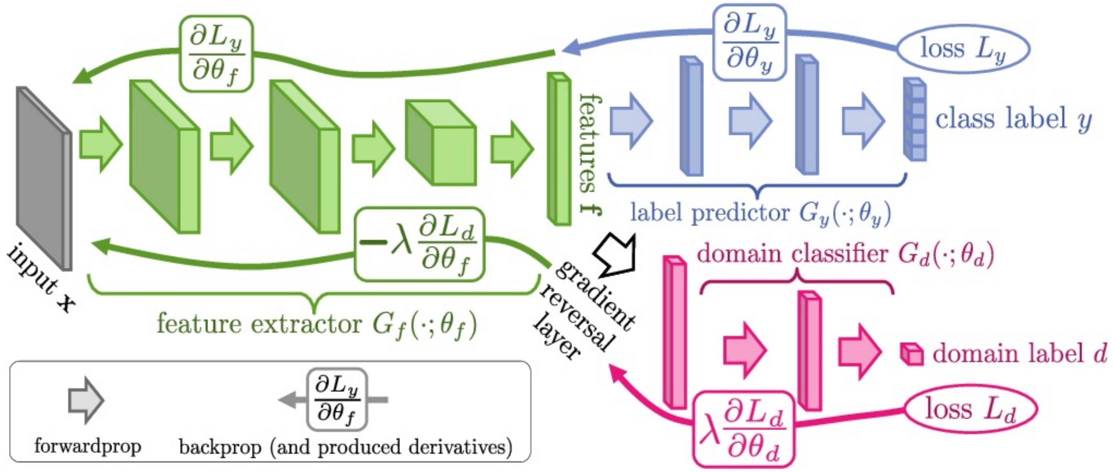




Math question: Estimate a continuous valued number.
Get to run noisy experiments of your choosing.

Fair AI with Adversarial Network

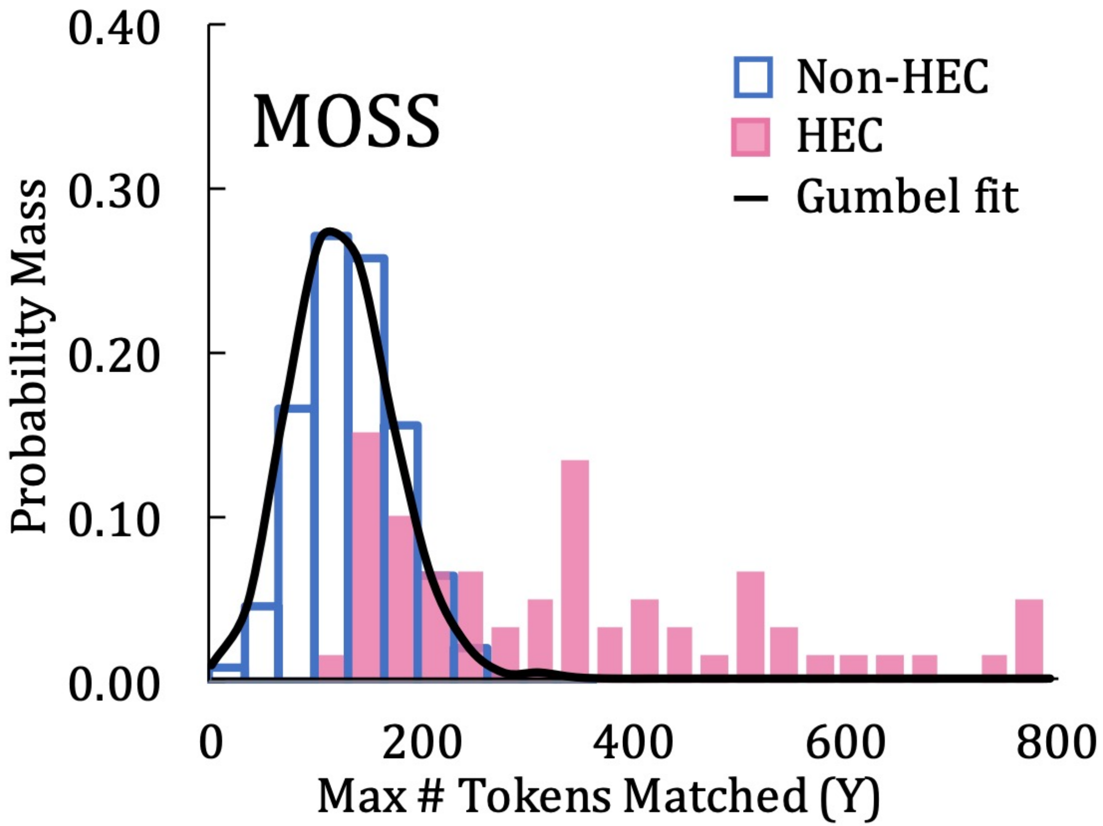
2018, with undergrads Christina Wadsworth and Francesca Vera

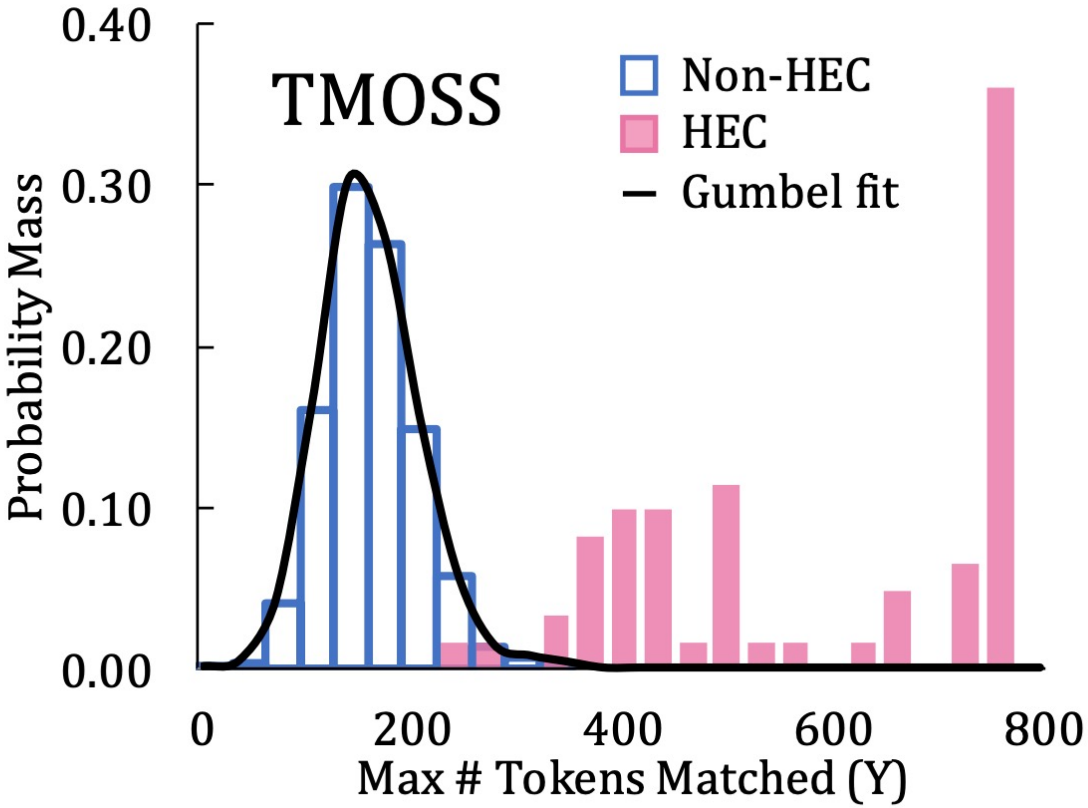


Model	Accuracy	FP Gap	FN Gap
COMPAS Scores (Our Test Set)	0.68	0.17	0.22
Our Recidivism Model	0.70	0.15	0.27
Our Chosen Adversarial Model	0.70	**0.01**	**0.02**
Bechavod et al. AVD Penalizers (2017)	0.65	**0.02**	**0.04**
Bechavod et al. SD Penalizers (2017)	0.66	**0.02**	**0.03**
Bechavod et al. Vanilla Regularized (2017)	0.67	0.20	0.30
Zafar et al. (2017)	0.66	**0.03**	0.11
Zafar et al. Baseline (2017)	0.66	**0.01**	0.09
Hardt et al. (2016)	0.65	**0.01**	**0.01**

Math question: Can you remove racism from a deep learning predictor?

Excessive Collaboration





So many things to love in this world

I am part of a trifecta with this new member

Amazing Teaching Team





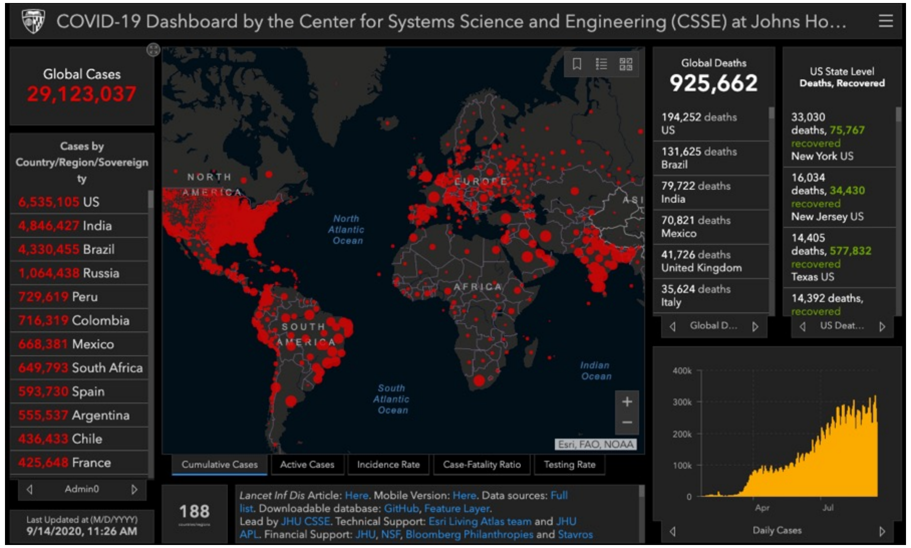



Course mechanics

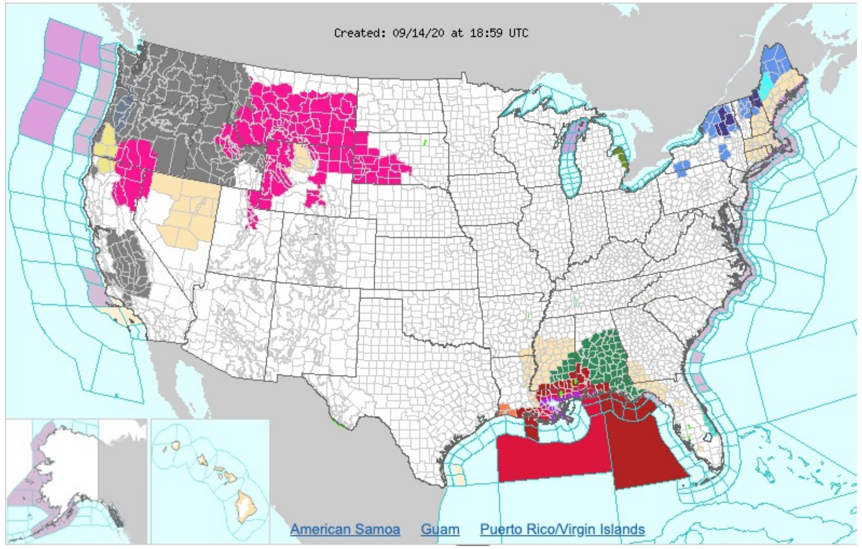
(this is a light version. Please read the handout for details).

What makes this quarter important

We are seeing a huge surge in **statistics, predictions, and probabilistic models** shared through global news, governing bodies, and social media.



Global cases of COVID-19
as of September 14th (JHU)
https://coronavirus.jhu.edu/map.html



National Weather Service Alerts
https://www.weather.gov/

FiveThirtyEight 2020

The New York Times 2020

US Presidential Election 2020 prediction forecasts
https://fivethirtyeight.com/
https://www.nytimes.com/

Essential Information

cs109.stanford.edu

Are you in the right place?

Prerequisites

What you really need:

CS106B/X (important):

- Recursion
- Hash Tables
- Binary Trees
- Programming

CS103 (ok as a corequisite):

- Proof techniques (induction)
- Set theory
- Math maturity

Math 51 or CME 100 (important)

- Multivariate differentiation
- Multivariate integration
- Basic facility with linear algebra (vectors)

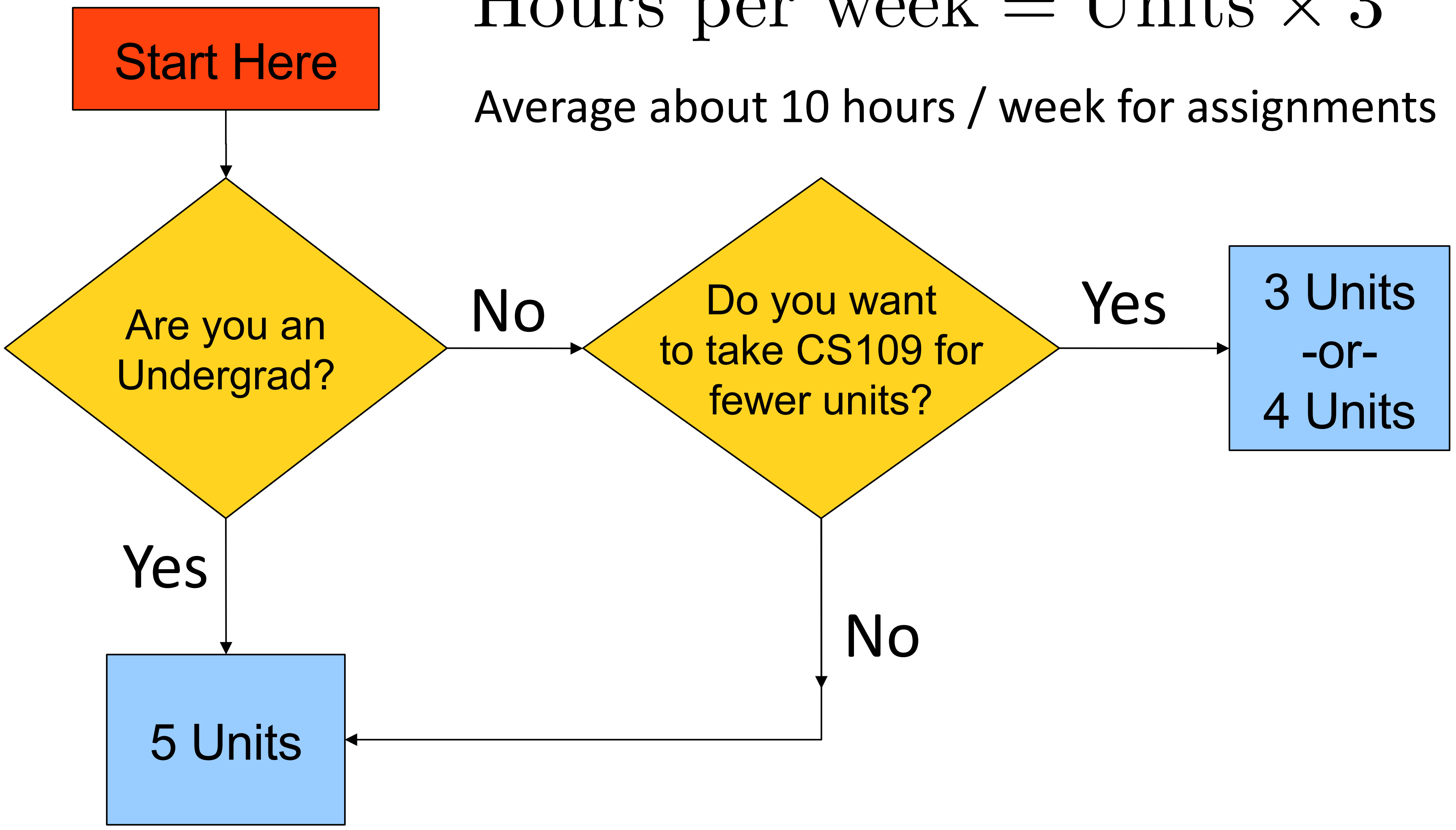
Coding in CS109

Review session on Friday

CS109 Units

$$\text{Hours per week} = \text{Units} \times 3$$

Average about 10 hours / week for assignments



Class Breakdown

40% **6 Assignments**

20% **Midterm**
2 hour exam, Oct 26th, 7pm

28% **Final**
3 hour exam, Dec 9th, 12:15pm

5% **Concept Checks**
Daily after lecture

7% **Section Participation**

Is Class Online?

Not officially recorded, though I will try to screen capture. We will try to provide online options for parts of the course

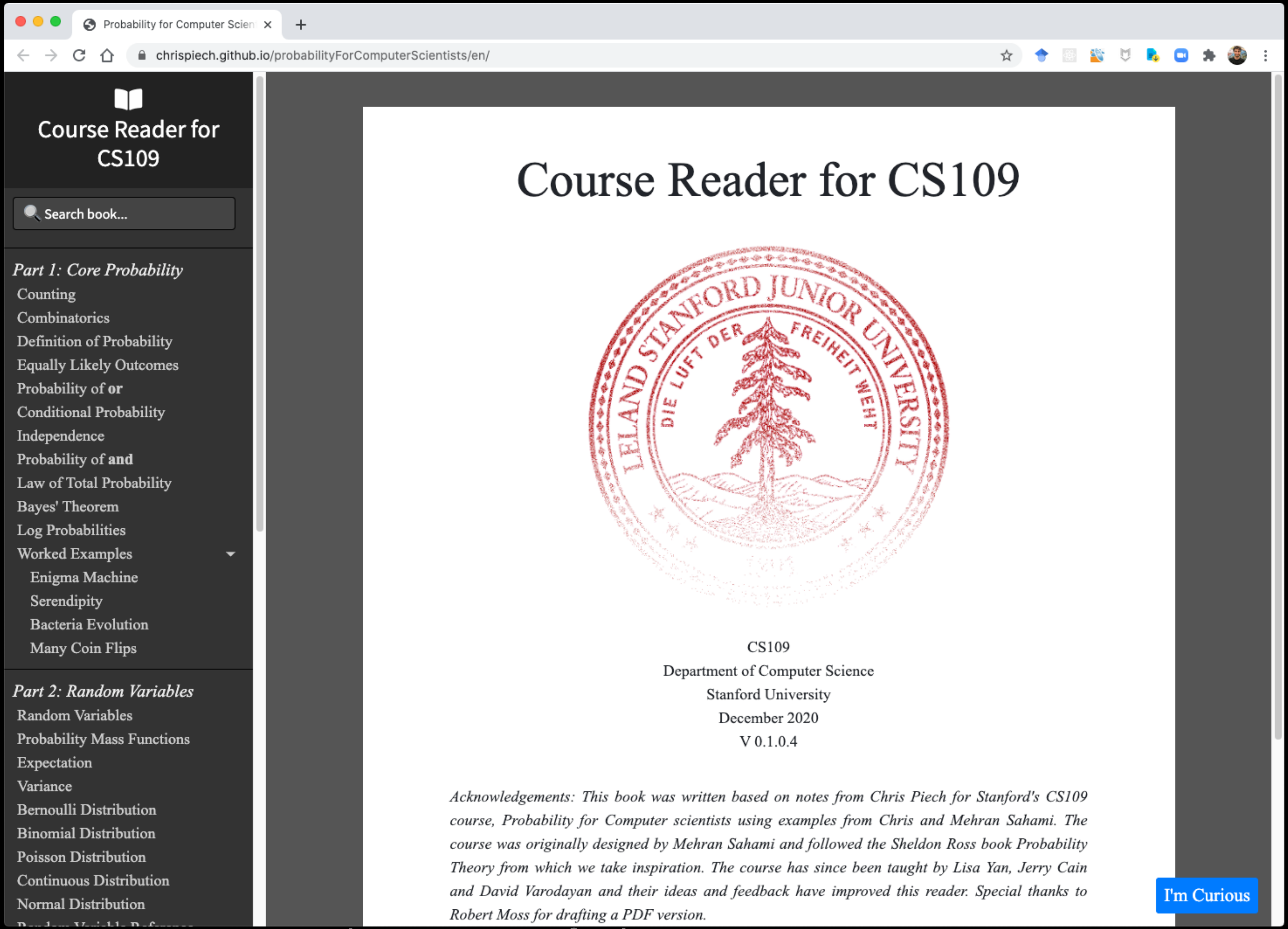
Ask questions

Q&A forum
All announcements

"Working" office hours
start on Thursday

Email cs109@cs.stanford.edu

Brand new this quarter: Course Reader!



Story of Modern AI

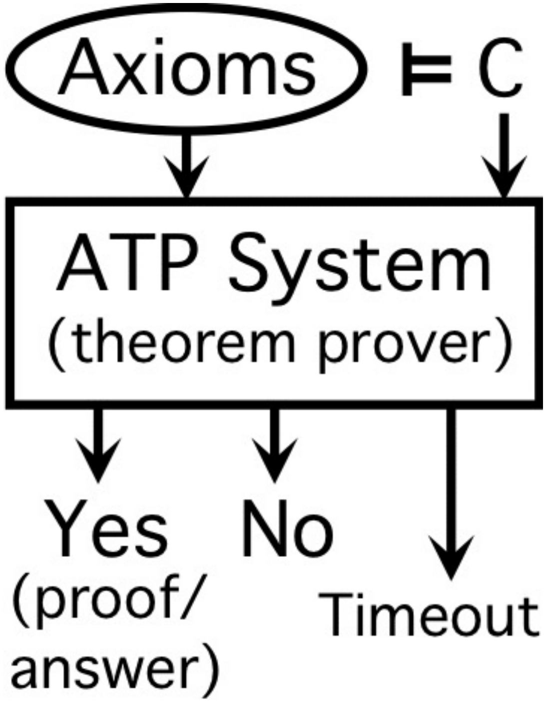
Modern AI
or, How we learned to combine probability and programming

Brief History

Early Optimism 1950s

1952

1955



Early Optimism 1950s

"Machines will be capable, within twenty years, of doing any work a man can do."
–Herbert Simon, 1952

Underwhelming Results 1950s to 1980s

The spirit is willing but the flesh is weak.

↓

(Russian)

↓

The vodka is good but the meat is rotten.

The world is too complex

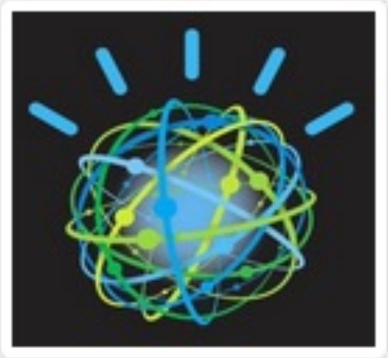

BRACE YOURSELVES
WINTER IS COMING

Something is going on in the world of AI

Big Milestones Part 1

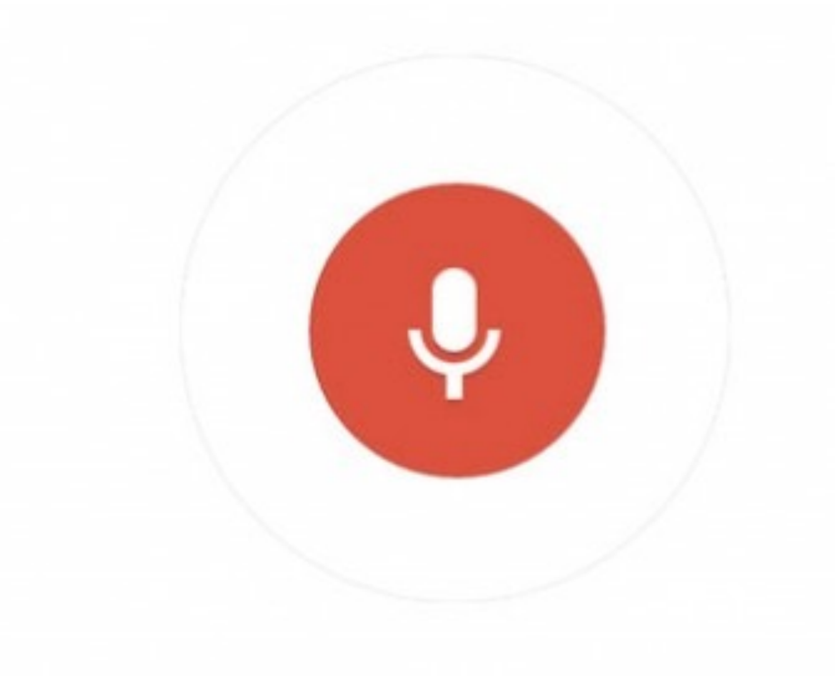
1997 Deep Blue

2005 Stanley



2011 Watson

I was told speech was 30 years out



Almost perfect…

The last remaining board game

Computers Making Art

Self Driving Cars

What is going on?

[suspense]

Focus on one problem

Computer Vision

Chihuahua or muffin?

Can you do it?

Chihuahua or Muffin?

Chihuahua or Muffin?

How about now?

What a computer sees

0	0	1	0	1	0	1	0	0	0	1	1	1	0	1
1	0	0	1	0	1	1	1	0	1	0	0	0	0	0
1	1	1	0	1	0	0	1	1	0	0	1	0	1	0
1	1	1	1	1	0	0	0	0	0	1	1	0	1	1
0	0	0	1	1	0	0	1	0	0	0	1	1	1	0
1	0	0	1	1	0	0	0	1	0					
1	1	0	1	1	0	0	1	1	0					
1	0	1	0	0	1	0	0	1	0					
0	0	0	0	1	0	1	0	1	1					
0	1	1	0	0	0	0	0	1	1					
0	0	1	0	1	1	1	0	0	0					
0	1	1	1	0	1	0	0	1	0					
1	1	0	0	0	0	0	0	0	0					
0	0	0	0	0	0	0	0	1	1					
0	0	1	1	1	0	1	0	1	1					

What a human sees

Very hard to code

```
public class Chihuahua extends ConsoleProgram {

   public void run() {
      println("Todo: Write program");
   }

}
```

Two Great Ideas

1. Probability from Examples

2. Artificial Neurons

Two Great Ideas

1. Probability from Examples

2. Artificial Neurons

Probability from Examples

When does the magic happen?

Lots of Data + Sound Probability

Machine Learning

Basically just a rebranding of statistics and probability.

Computer Vision is Still Hard

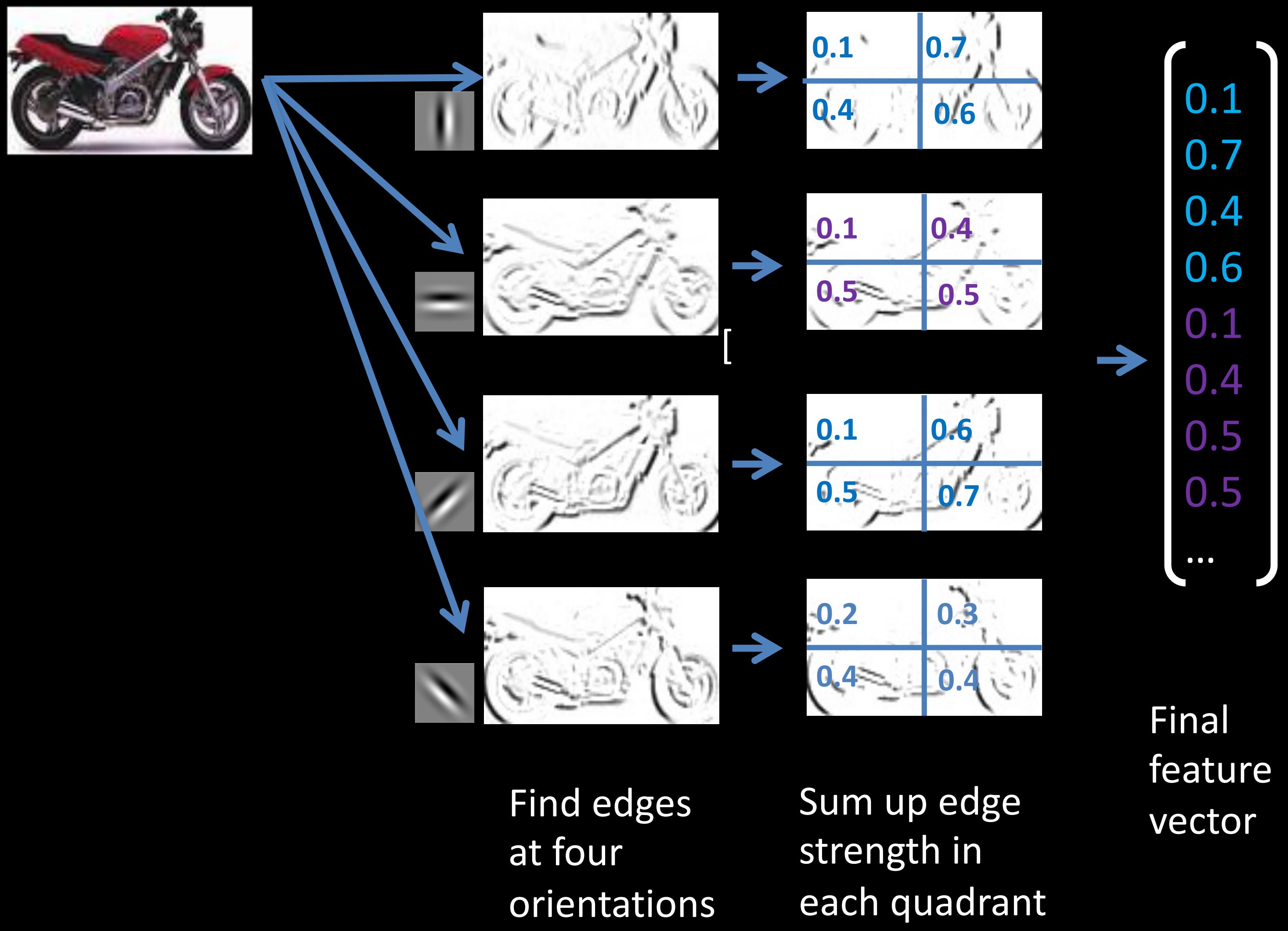
You see this:

But the camera sees this:

194	210	201	212	199	213	215	195	178	158	182	209
180	189	190	221	209	205	191	167	147	115	129	163
114	126	140	188	176	165	152	140	170	106	78	88
87	103	115	154	143	142	149	153	173	101	57	57
102	112	106	131	122	138	152	147	128	84	58	66
94	95	79	104	105	124	129	113	107	87	69	67
68	71	69	98	89	92	98	95	89	88	76	67
41	56	68	99	63	45	60	82	58	76	75	65
20	43	69	75	56	41	51	73	55	70	63	44
50	50	57	69	75	75	73	74	53	68	59	37
72	59	53	66	84	92	84	74	57	72	63	42
67	61	58	65	75	78	76	73	59	75	69	50

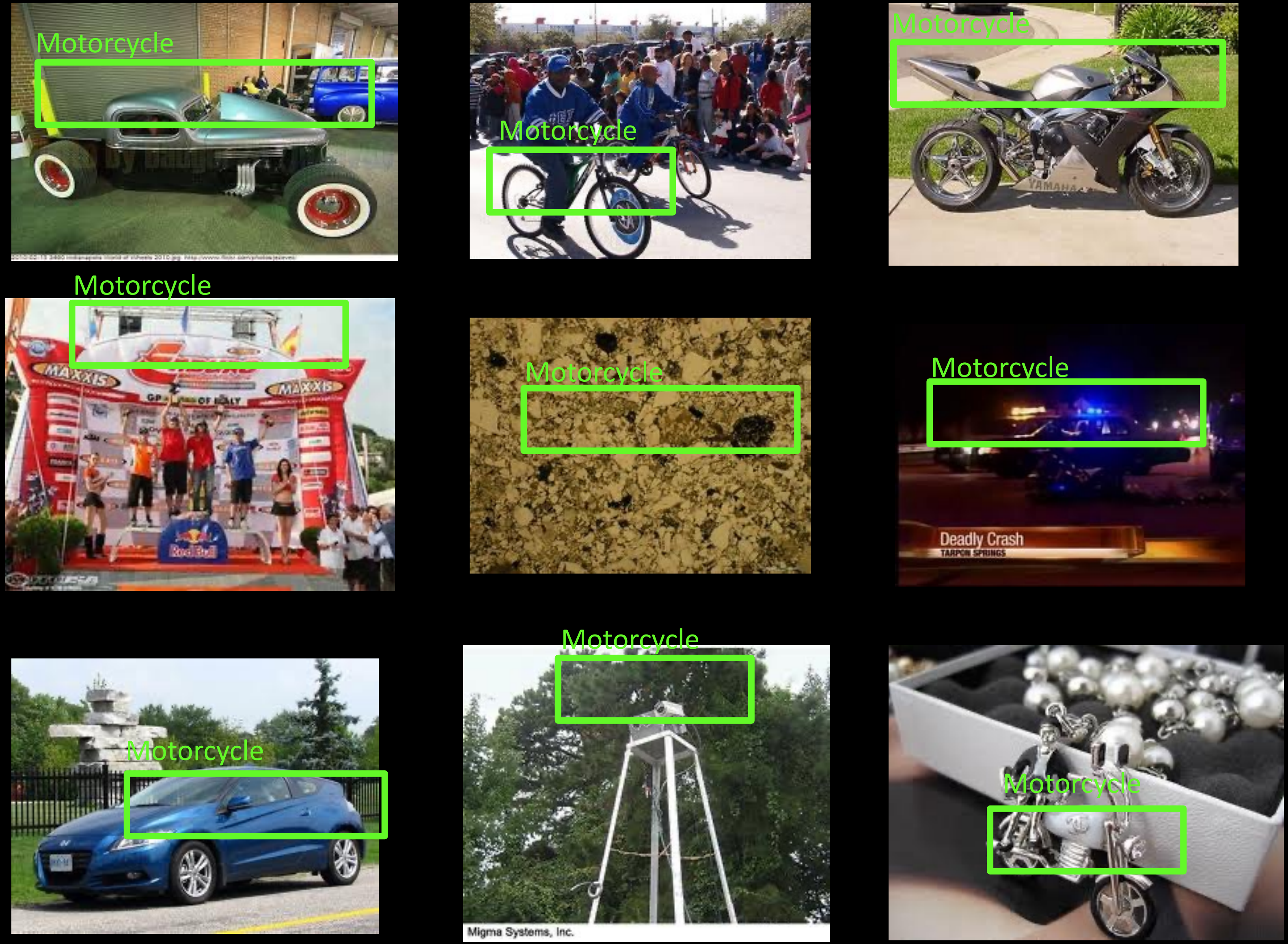
[Andrew Ng]

Computer Vision is Still Hard



[Andrew Ng]

Straight ML not perfect



Some Great Thinkers

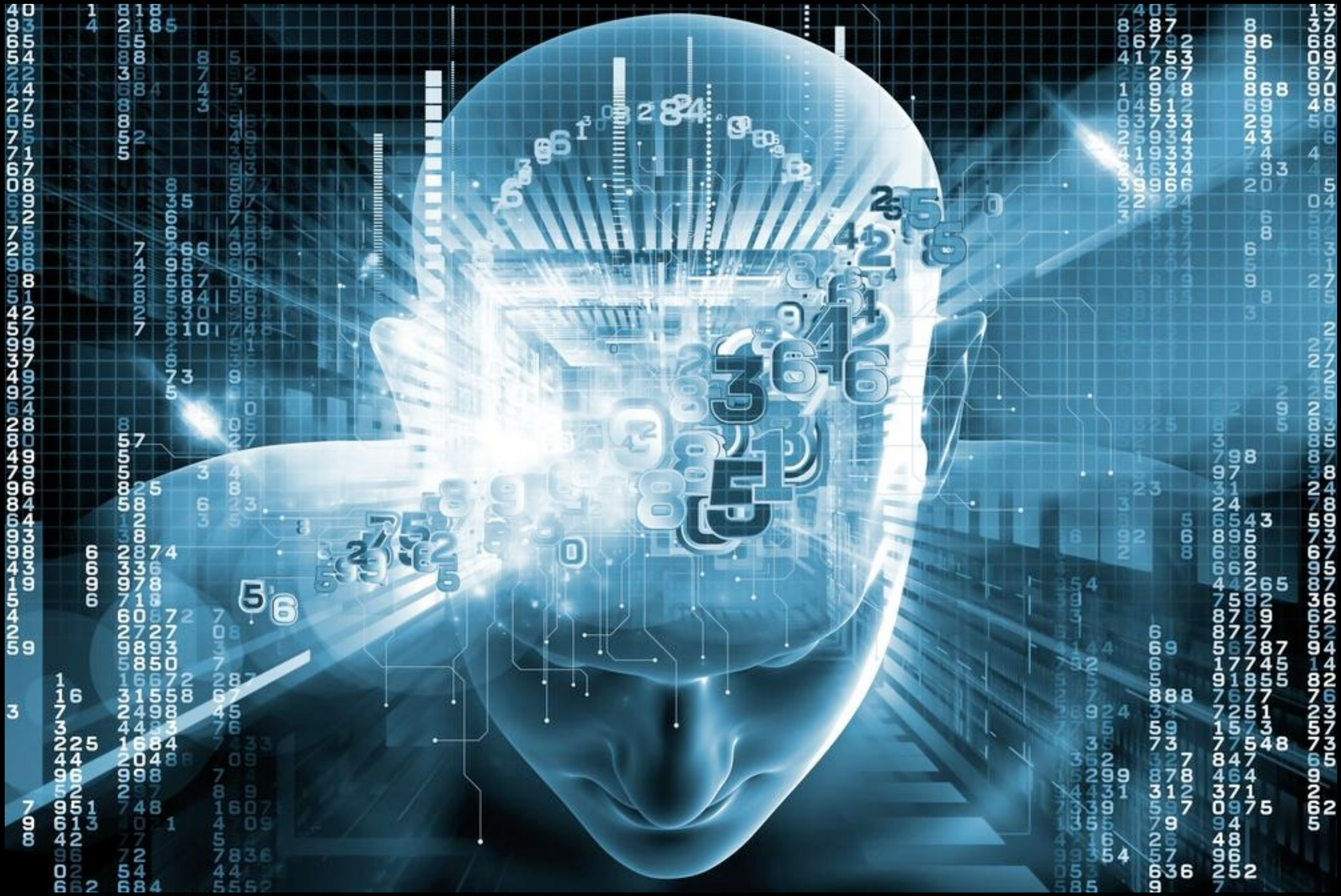
Daphne Koller

Two Great Ideas

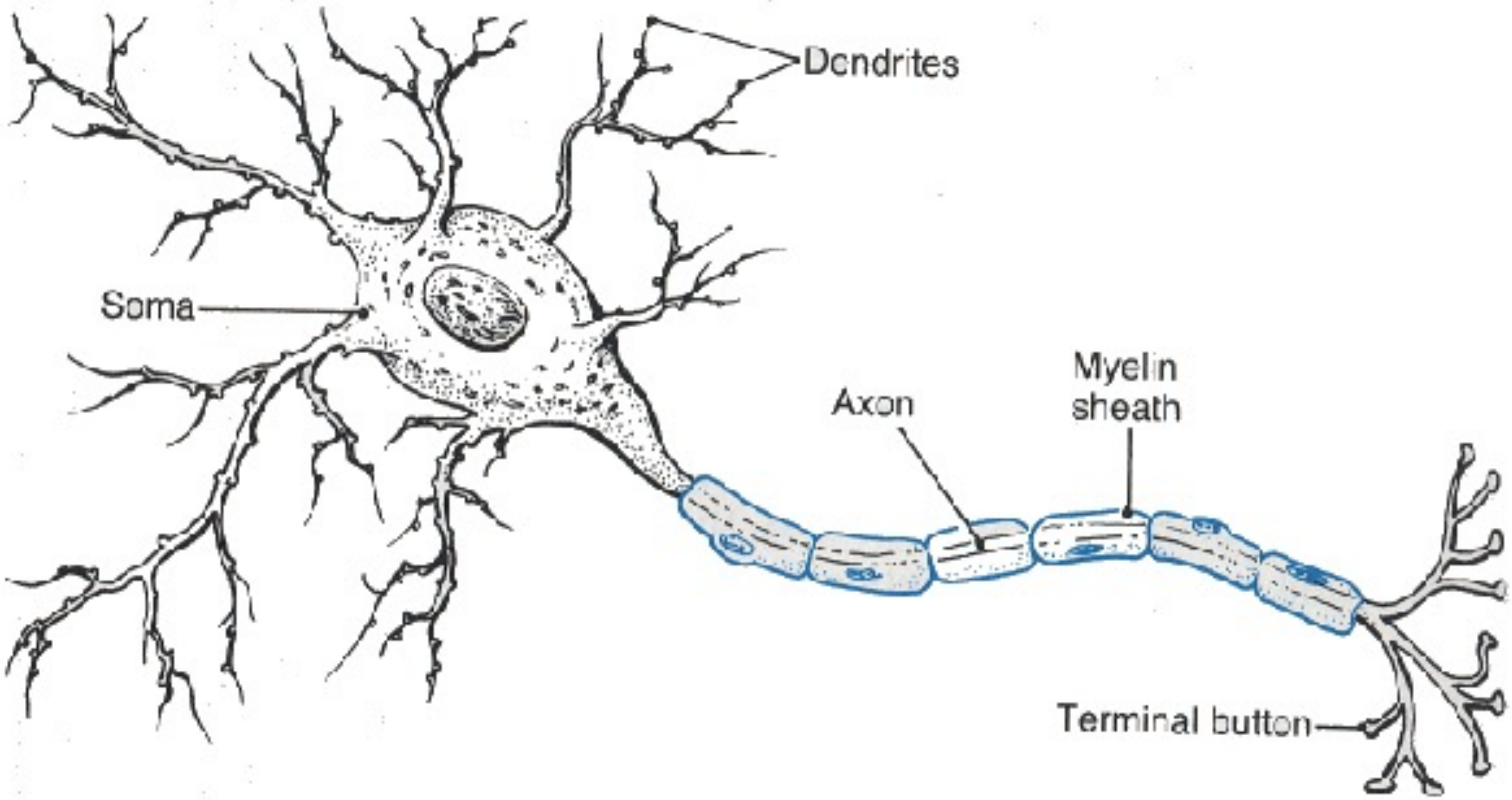
1. Probability from Examples

2. Artificial Neurons

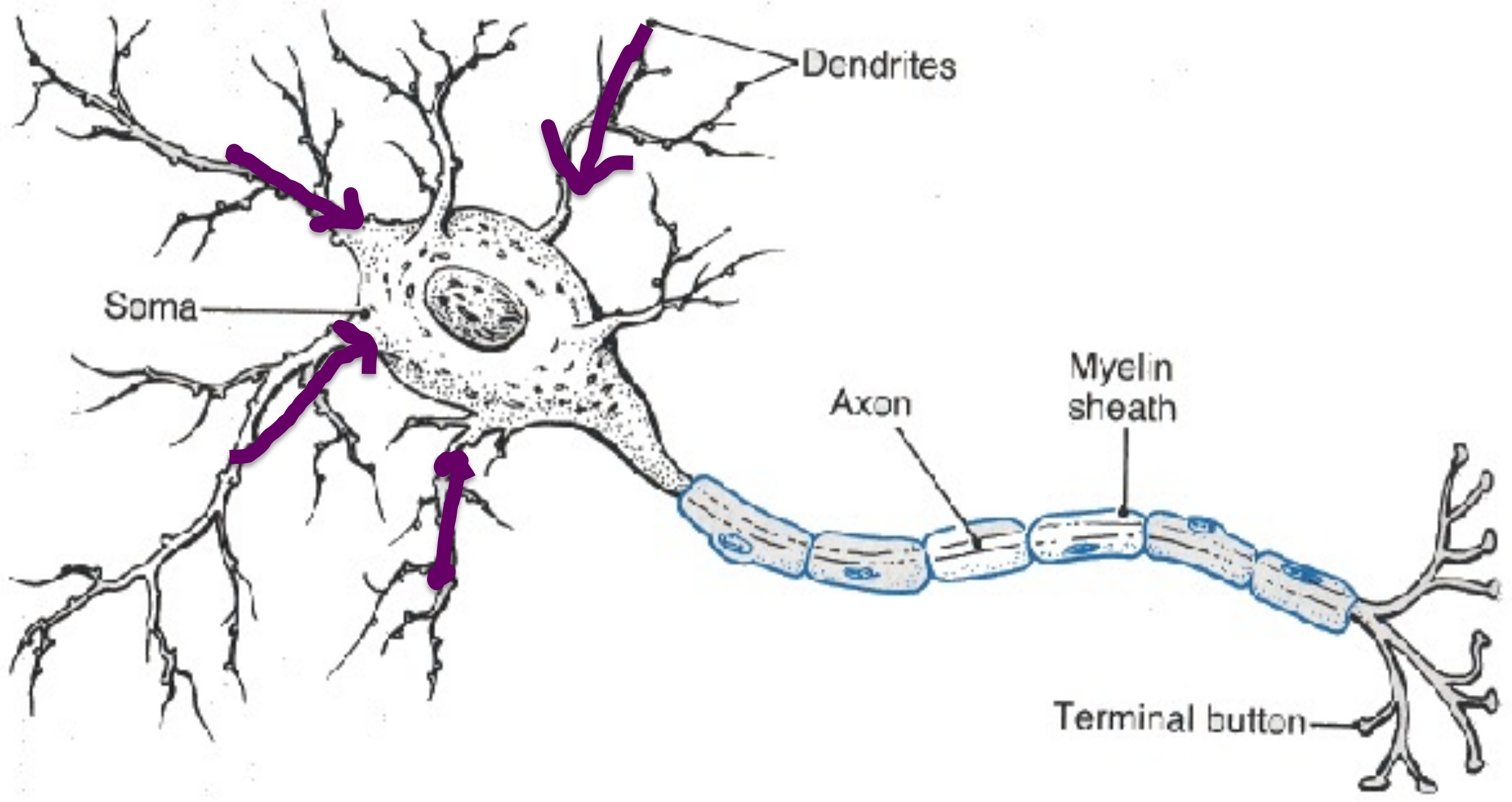
Artificial Neurons



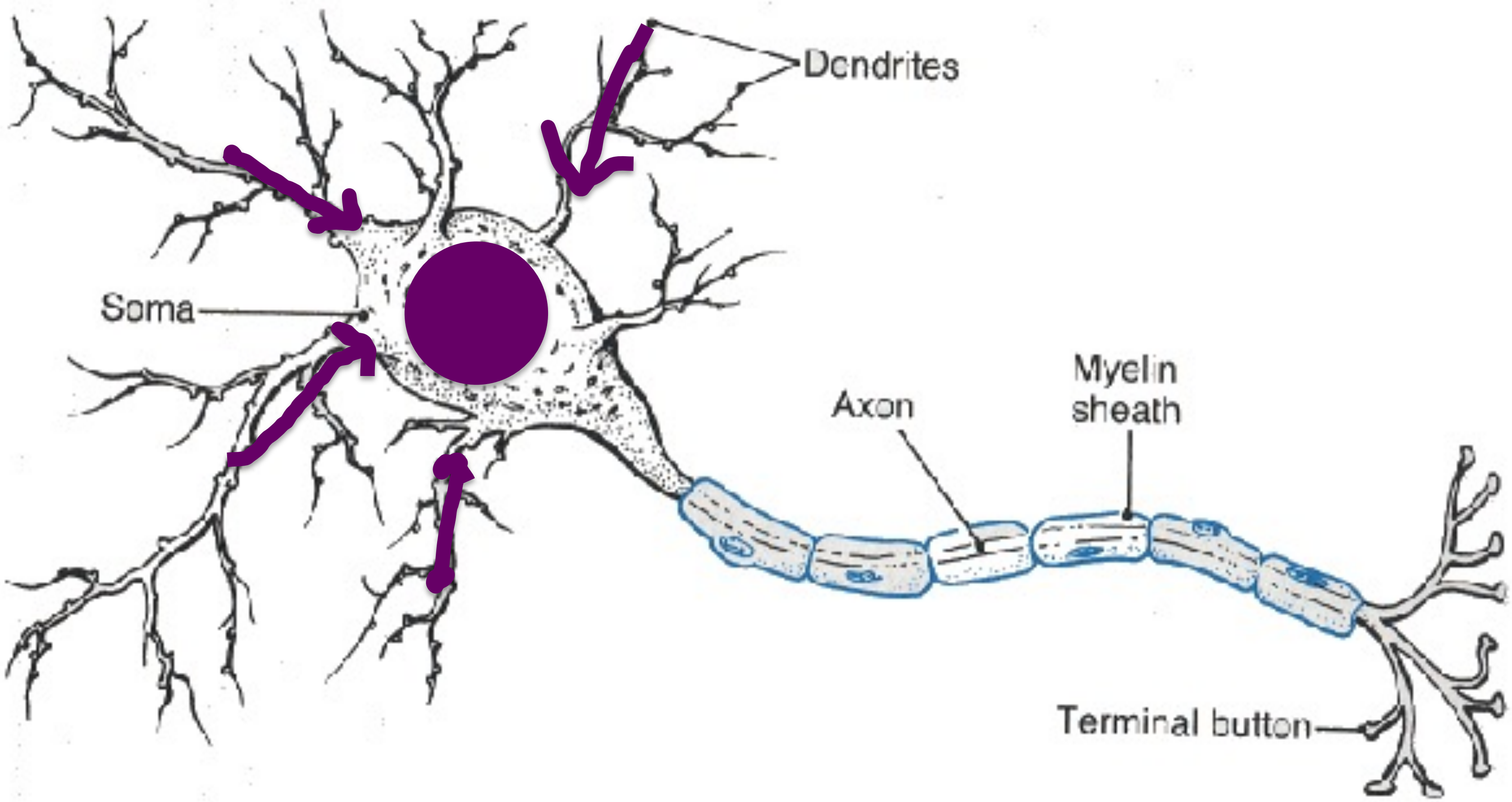
Human Neuron



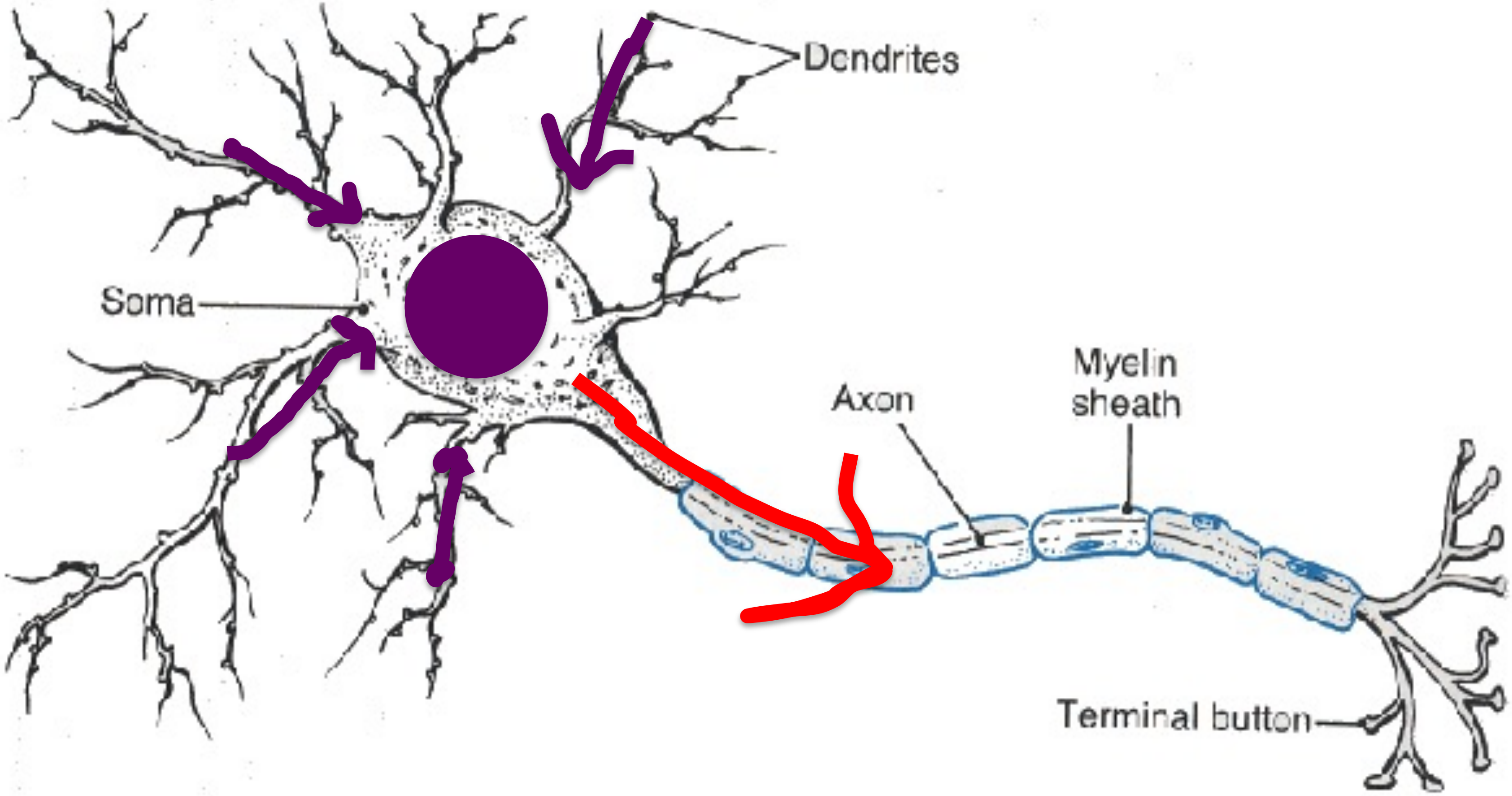
Human Neuron



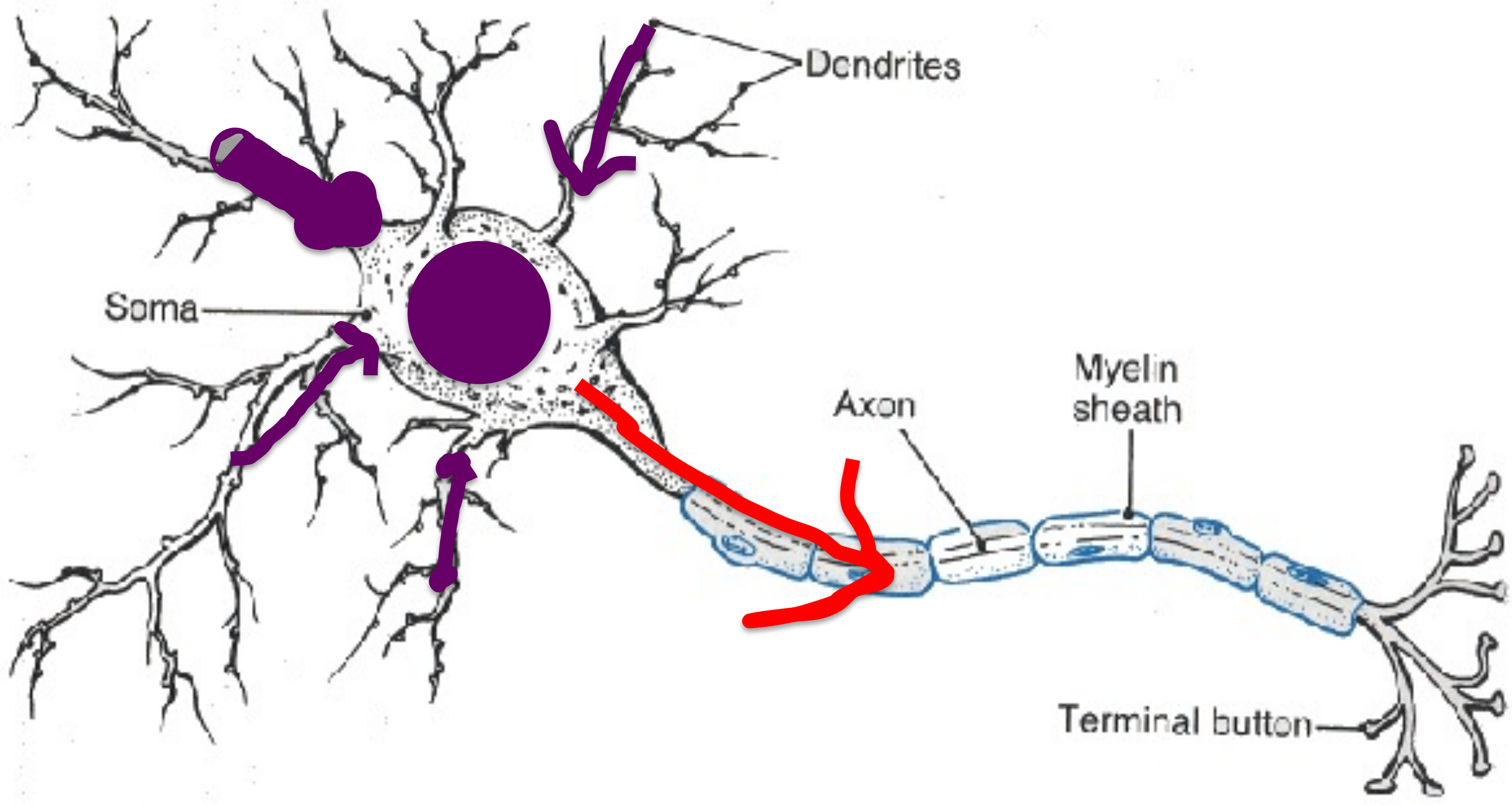
Human Neuron



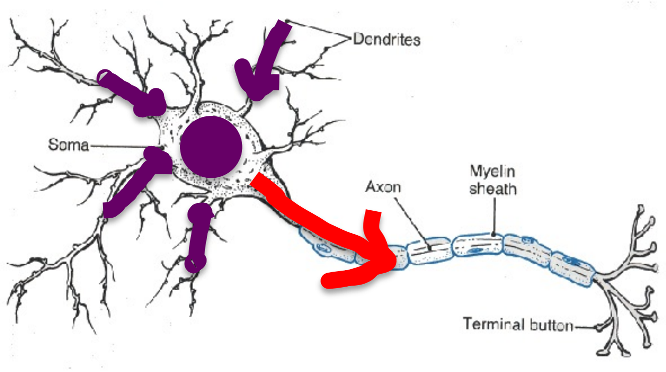
Human Neuron

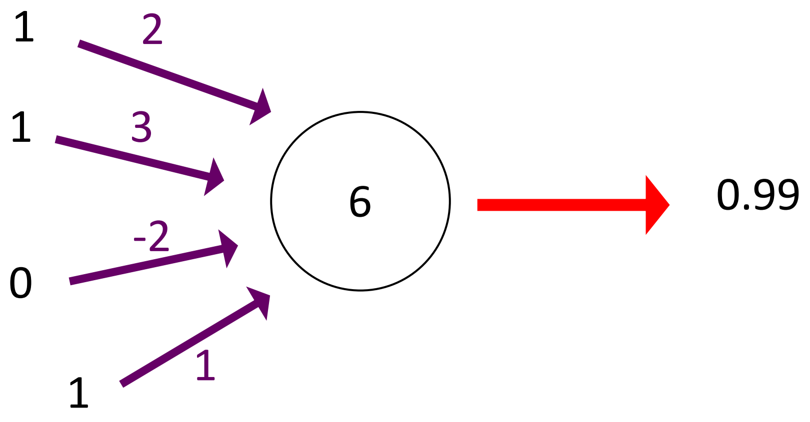


Human Neuron



Artificial Neuron

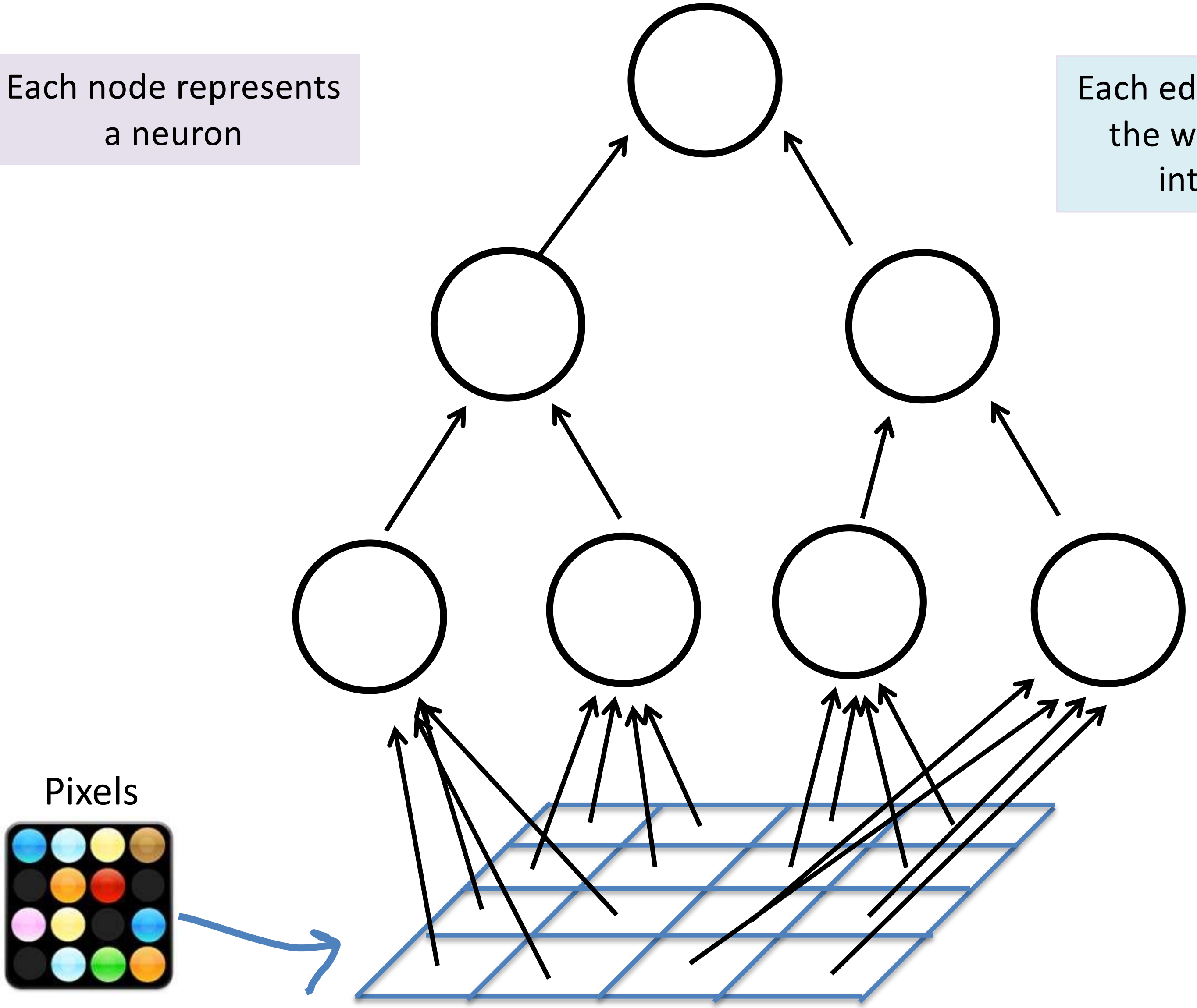




Artificial Neural Network

Each node represents a neuron

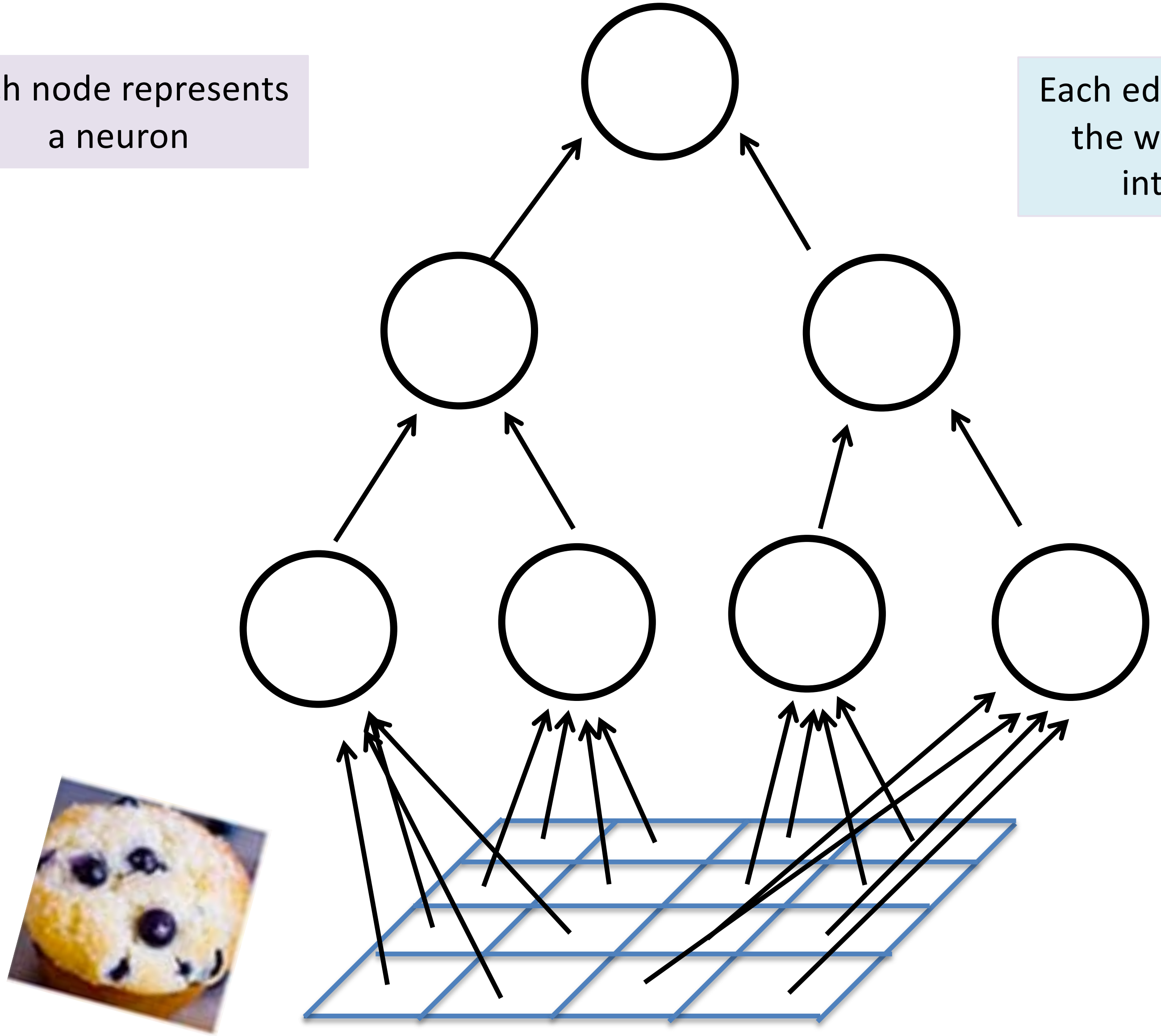
Each edge represents the weight of the interaction



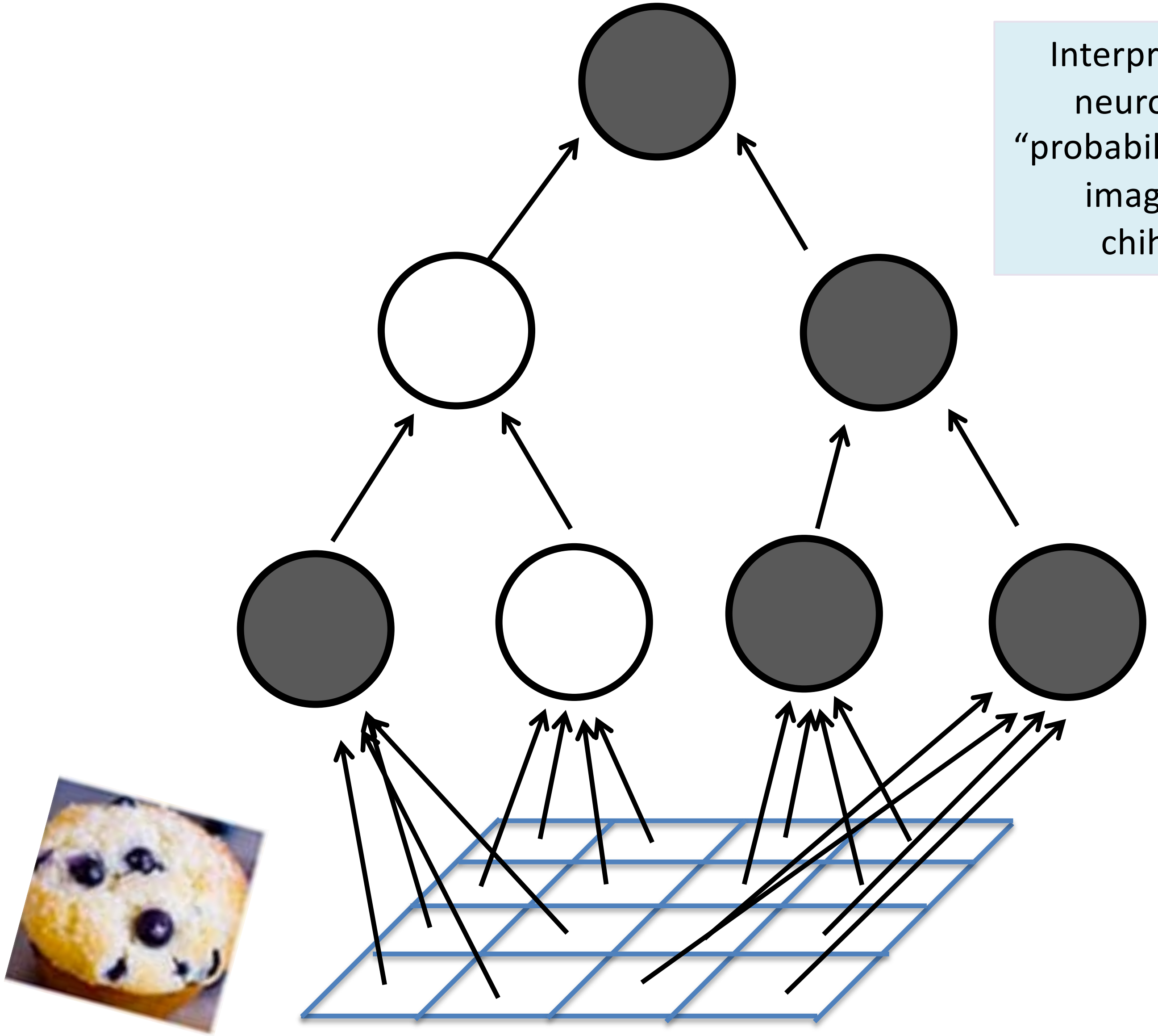
Artificial Neural Network

Each node represents a neuron

Each edge represents the weight of the interaction



Artificial Neural Network



Interpret the last neuron as the "probability" that the image is of a chihuahua

Update Neural Network

$$P(Y = 1 | X = \mathbf{x}) = \hat{\mathbf{y}}$$

$$\hat{y} = \sigma\left(\sum_{j=0}^{m_h} \mathbf{h}_j \theta_j^{(\hat{y})}\right)$$

For one datum

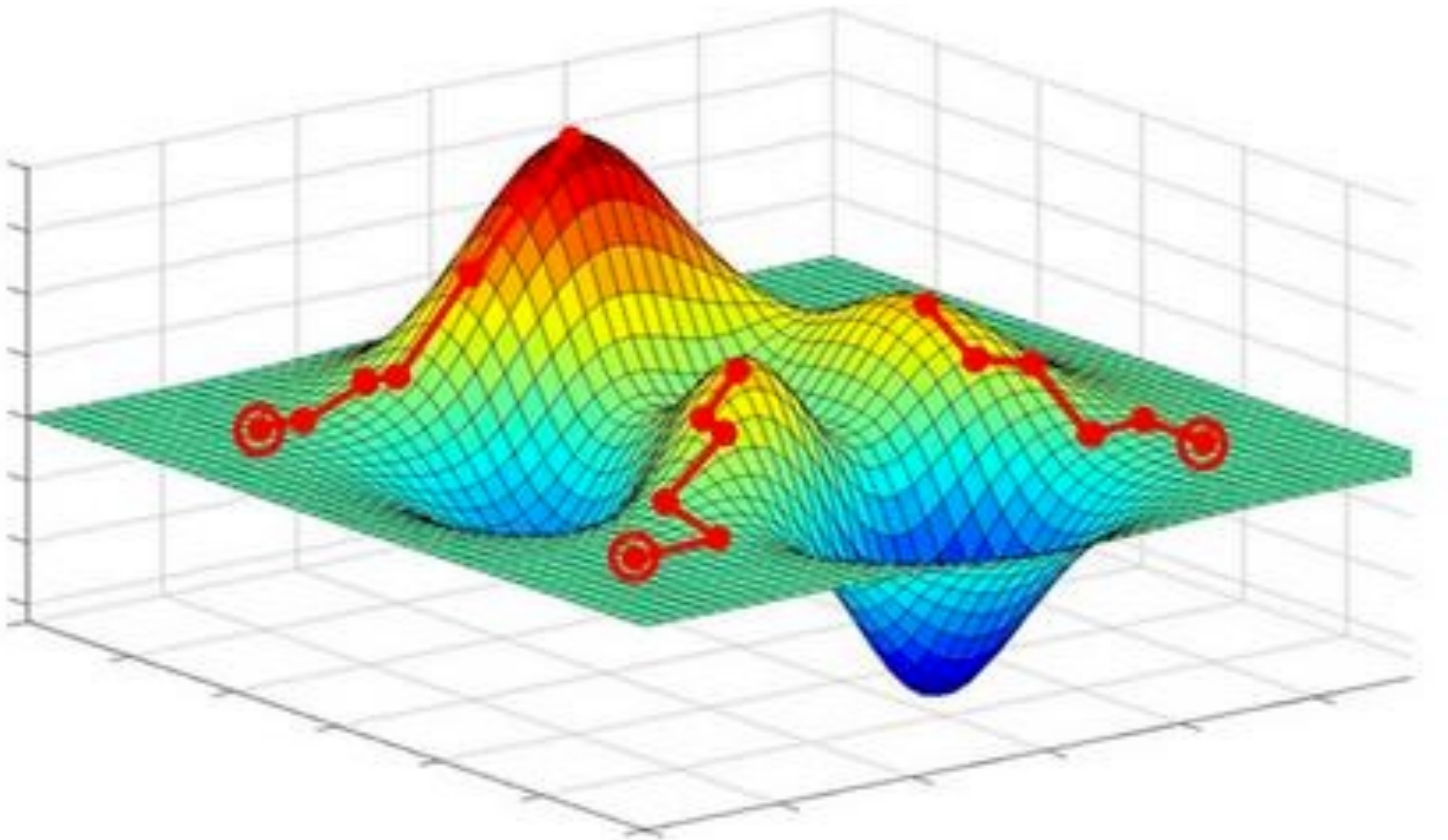
$$P(Y = y | X = \mathbf{X}) = (\hat{y})^y (1 - \hat{y})^{1-y}$$

For IID data

$$L(\theta) = \prod_{i=1}^{n} P(Y = y^{(i)} | X = \mathbf{x}^{(i)})$$

$$= \prod_{i=1}^{n} (\hat{y}^{(i)})^{y^{(i)}} \cdot \left[1 - (\hat{y}^{(i)})\right]^{(1-y^{(i)})}$$

Gradient Descent



Walk uphill and you will find a local maxima
(if your step size is small enough)

Gradient of Probability

$$\frac{\partial L}{\partial \theta_i^{(\hat{y})}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial \theta_i^{(\hat{y})}}$$

$$\hat{y} = \sigma\left(\sum_{j=0}^{m_h} \mathbf{h}_j \theta_j^{(\hat{y})}\right)$$

$$\frac{\partial \hat{y}}{\partial \theta_i^{(\hat{y})}} = \sigma\left(\sum_{j=0}^{m_h} \mathbf{h}_j \theta_j^{(\hat{y})}\right)\left[1 - \sigma\left(\sum_{j=0}^{m_h} \mathbf{h}_j \theta_j^{(\hat{y})}\right)\right] \cdot \frac{\partial}{\partial \theta_i^{(\hat{y})}} \sum_{j=0}^{m_h} \mathbf{h}_j \theta_j^{(\hat{y})}$$

$$= \hat{y}[1 - \hat{y}] \cdot \frac{\partial}{\partial \theta_i^{(\hat{y})}} \sum_{j=0}^{m_h} \mathbf{h}_j \theta_j^{(\hat{y})}$$

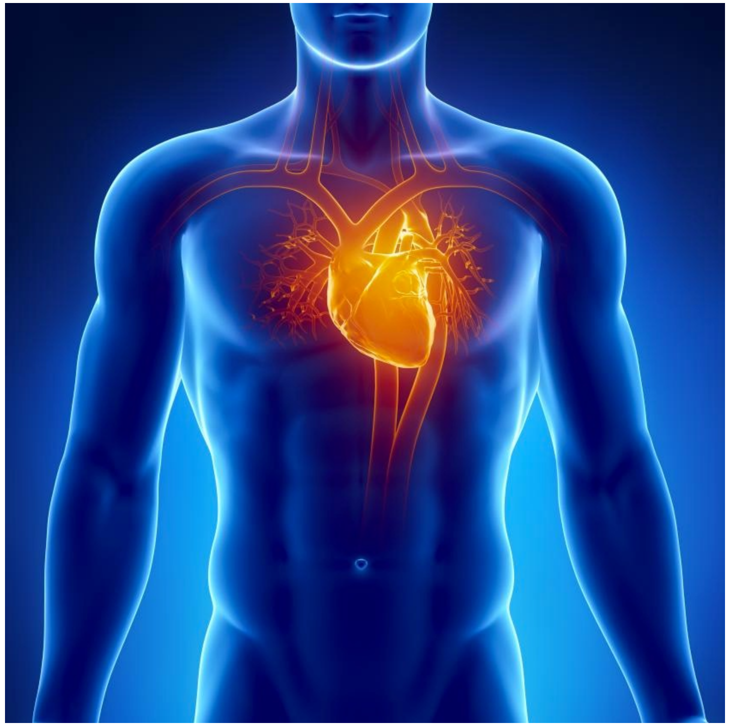
$$= \hat{y}[1 - \hat{y}] \cdot h_i$$

You will be able to do this.

Where you will be by the end of class

CS109: Theory Class focused on Applications

Heart



Ancestry

23andMe

Netflix

NETFLIX

Where is this Useful?

A machine learning algorithm performs **better than** the best dermatologists.

Developed recently, at Stanford.

Esteva, Andre, et al. "Dermatologist-level classification of skin cancer with deep neural networks." *Nature* 542.7639 (2017): 115-118.

Not once, but twice, AI was revolutionized by people who understood probability theory.

End of Story

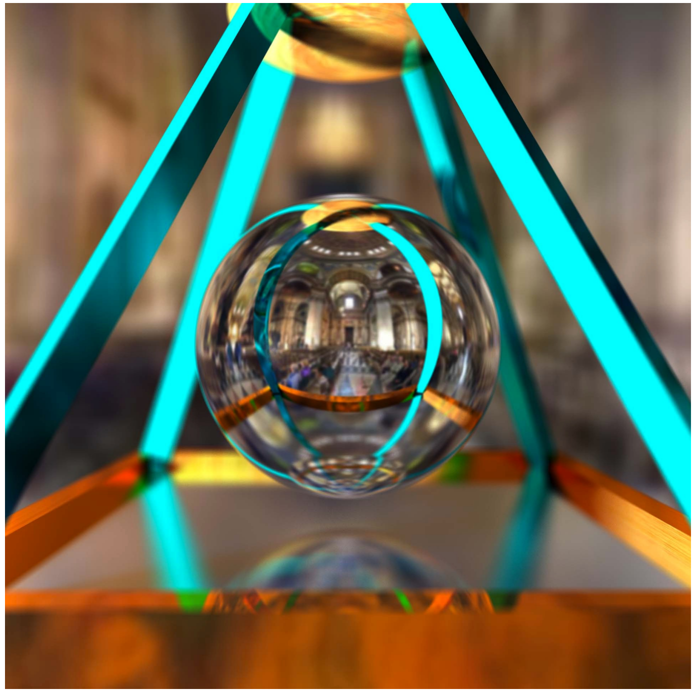
Except it isn't the end of the story…

Probability is more than just machine learning

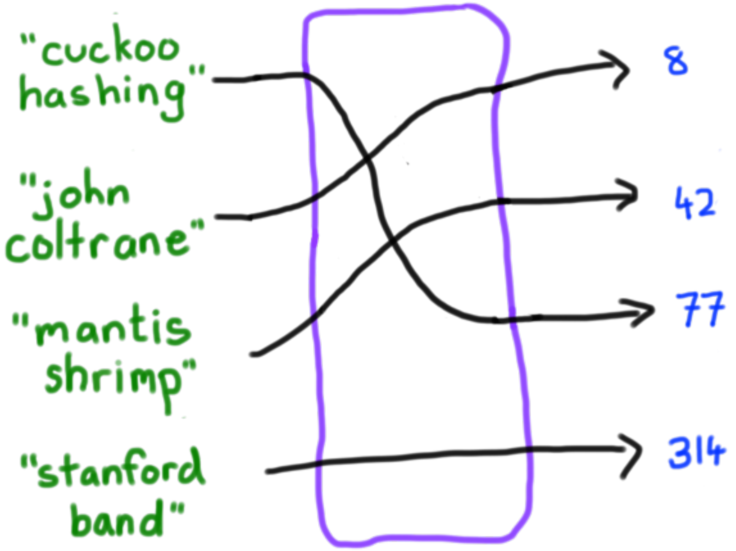
Abundance of important problems

Algorithms and Probability

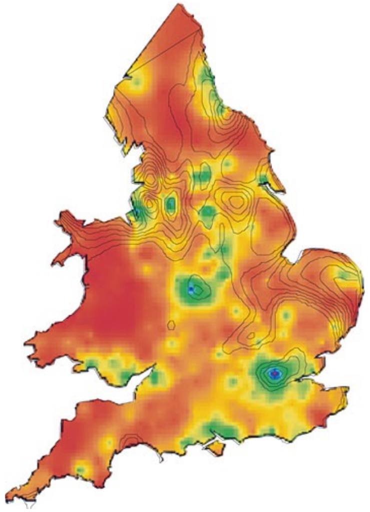
Eg Raytracing



Eg HashMaps



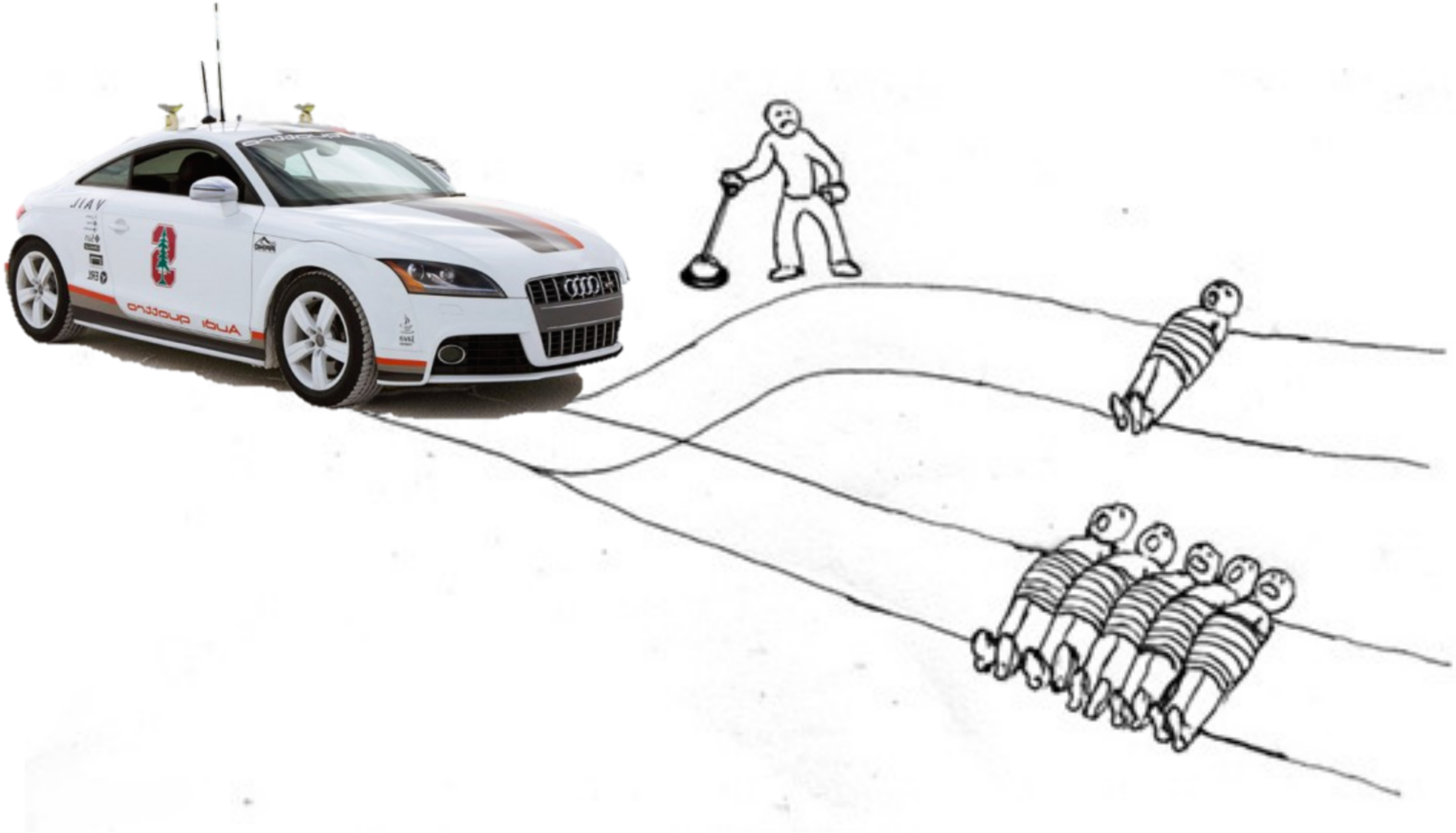
The next medical revolution?



Autocomplete

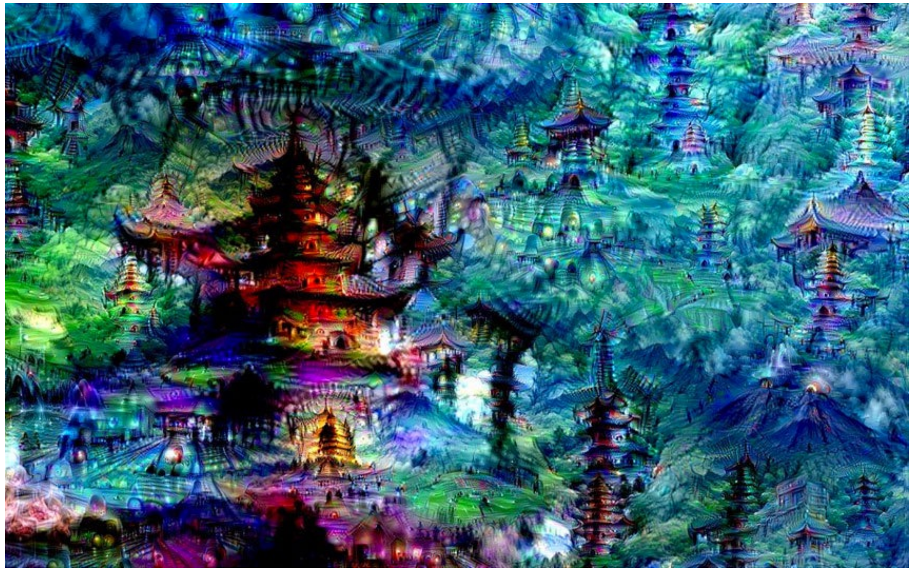
Recommender Systems

Philosophy and Ethics

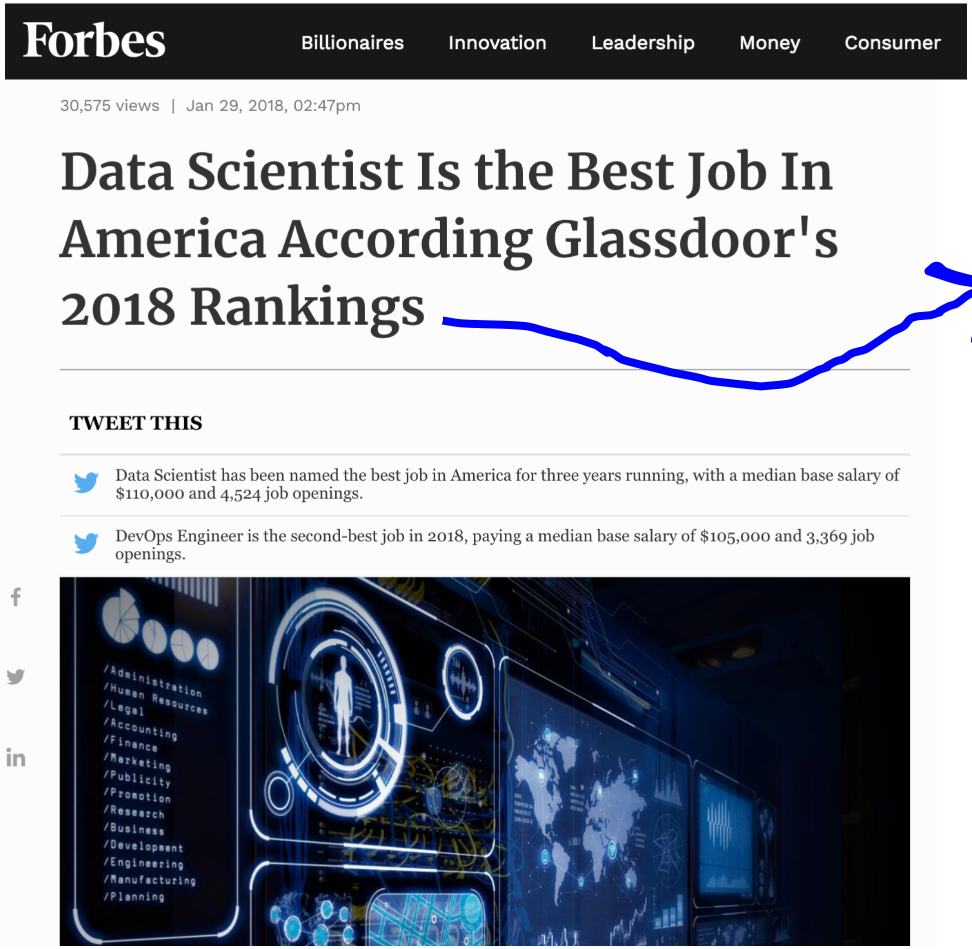


Art and Probability






Most Desired Skill in Industry



Forbes | Billionaires | Innovation | Leadership | Money | Consumer

30,575 views | Jan 29, 2018, 02:47pm

Data Scientist Is the Best Job In America According Glassdoor's 2018 Rankings

TWEET THIS

Data Scientist has been named the best job in America for three years running, with a median base salary of $110,000 and 4,524 job openings.

DevOps Engineer is the second-best job in 2018, paying a median base salary of $105,000 and 3,369 job openings.

Job Score is based on:

- Earning potential
- Number of jobs
- Job satisfaction rating

"Data science and machine learning are generating more jobs than candidates right now, making these two areas the *fastest growing employment areas*."

9.8 times more jobs than five years ago.

[LinkedIn's 2017 U.S. Emerging Jobs Report]

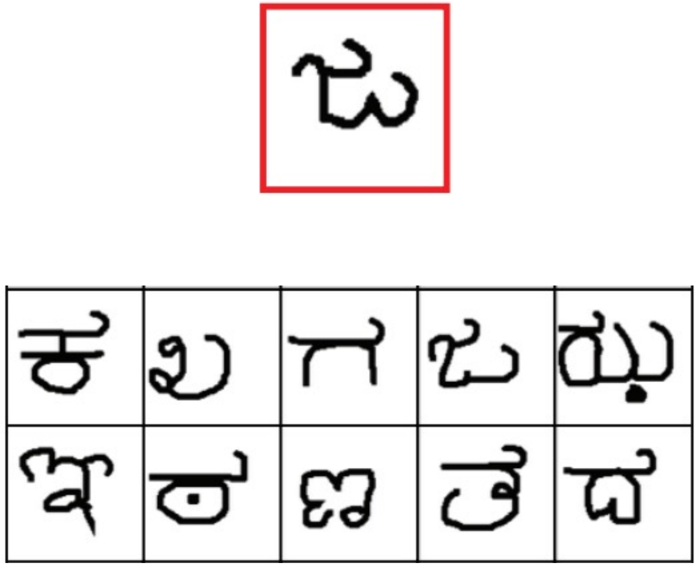
Most Desired Skill in Academia

Most CS PhD students list their highest desiderata upon graduation as:

"Better understanding of probability"

Open Problem: One Shot Learning

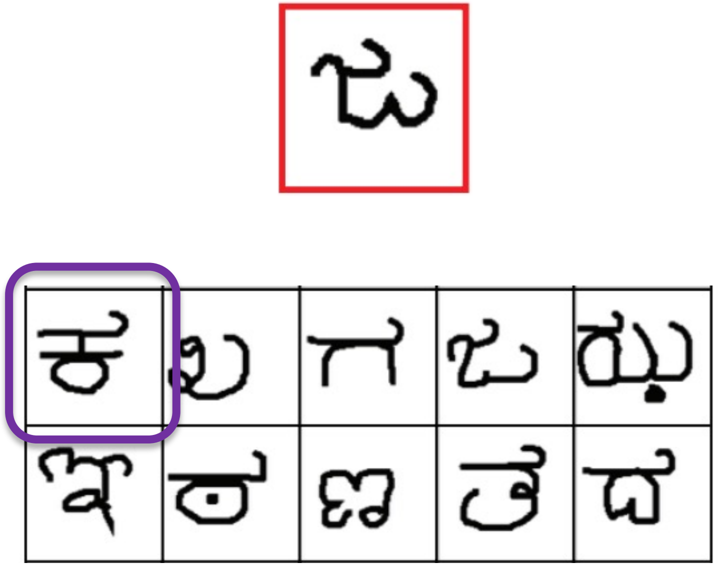
B Lake, R Salakhutdinov, J Tenenbaum. Science 2015.
Human-level concept learning through probabilistic program induction.



Current deep learning methods are not enough to move the needle as far as we want, **especially on socially relevant problems** that often do not have the benefit of massive public datasets. The best new ideas are coming from probability theory

Open Problem: One Shot Learning

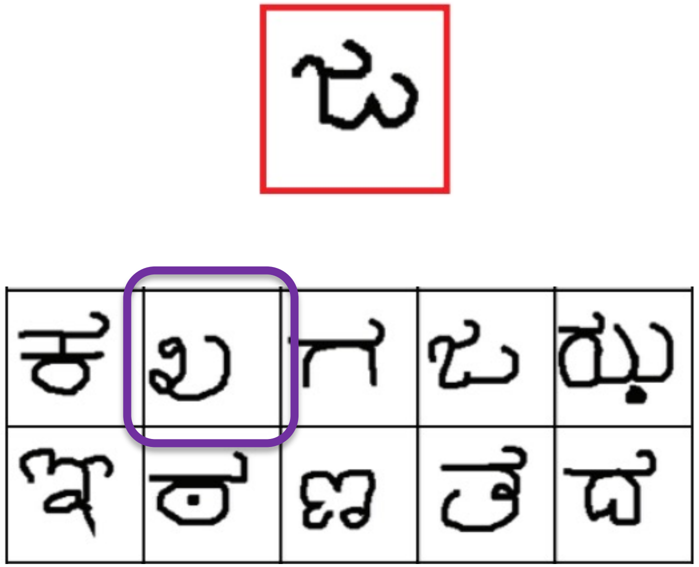
B Lake, R Salakhutdinov, J Tenenbaum. Science 2015.
Human-level concept learning through probabilistic program induction.



Current deep learning methods are not enough to move the needle as far as we want, **especially on socially relevant problems** that often do not have the benefit of massive public datasets. The best new ideas are coming from probability theory

Open Problem: One Shot Learning

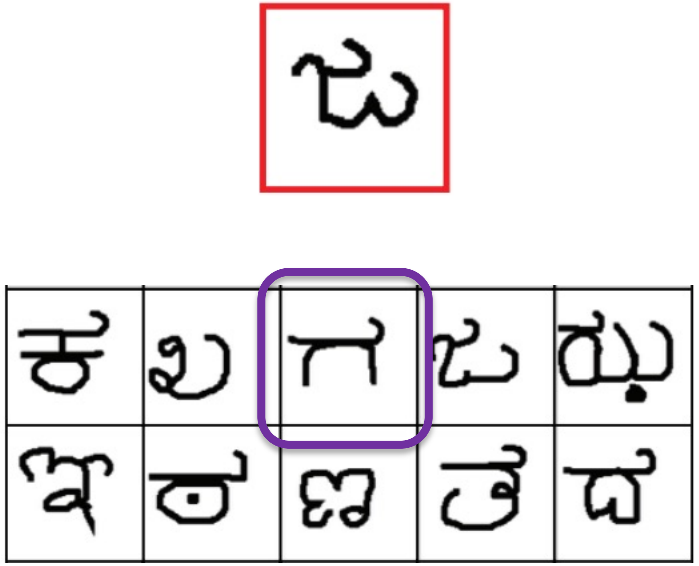
B Lake, R Salakhutdinov, J Tenenbaum. Science 2015.
Human-level concept learning through probabilistic program induction.



Current deep learning methods are not enough to move the needle as far as we want, **especially on socially relevant problems** that often do not have the benefit of massive public datasets. The best new ideas are coming from probability theory

Open Problem: One Shot Learning

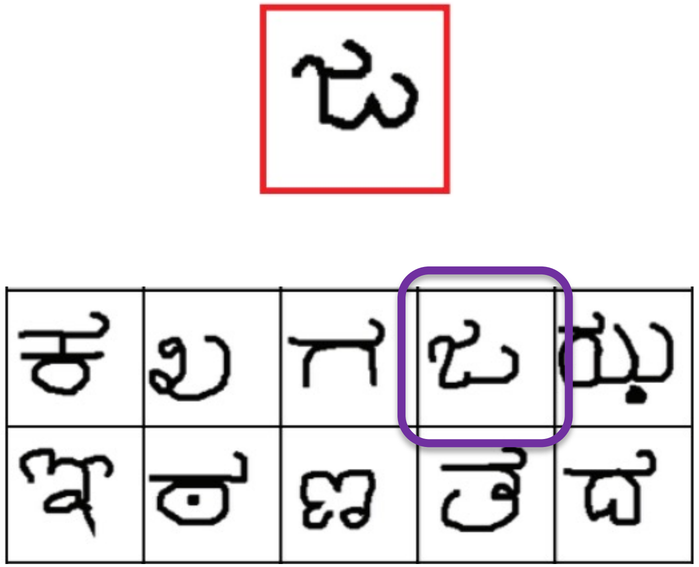
B Lake, R Salakhutdinov, J Tenenbaum. Science 2015.
Human-level concept learning through probabilistic program induction.



Current deep learning methods are not enough to move the needle as far as we want, **especially on socially relevant problems** that often do not have the benefit of massive public datasets. The best new ideas are coming from probability theory

Open Problem: One Shot Learning

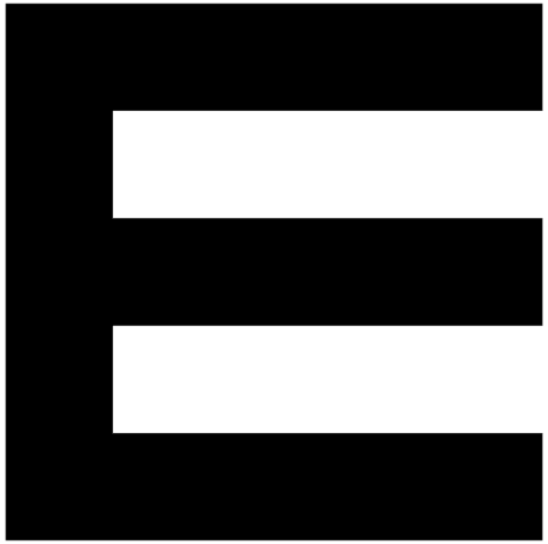
B Lake, R Salakhutdinov, J Tenenbaum. Science 2015.
Human-level concept learning through probabilistic program induction.



Current deep learning methods are not enough to move the needle as far as we want, **especially on socially relevant problems** that often do not have the benefit of massive public datasets. The best new ideas are coming from probability theory

Learn Real Skills in CS109

Spring 2017



Patient sees a series of letters of different font size, and for each, answers correct or incorrect

You decide that the vision tests given by eye doctors could have more precise results if we used an approach inspired by logistic regression. In a vision test a user looks at a letter with a particular font size and either correctly guesses the letter or incorrectly guesses the letter.

You assume that the probability that a particular patient is able to guess a letter correctly is:

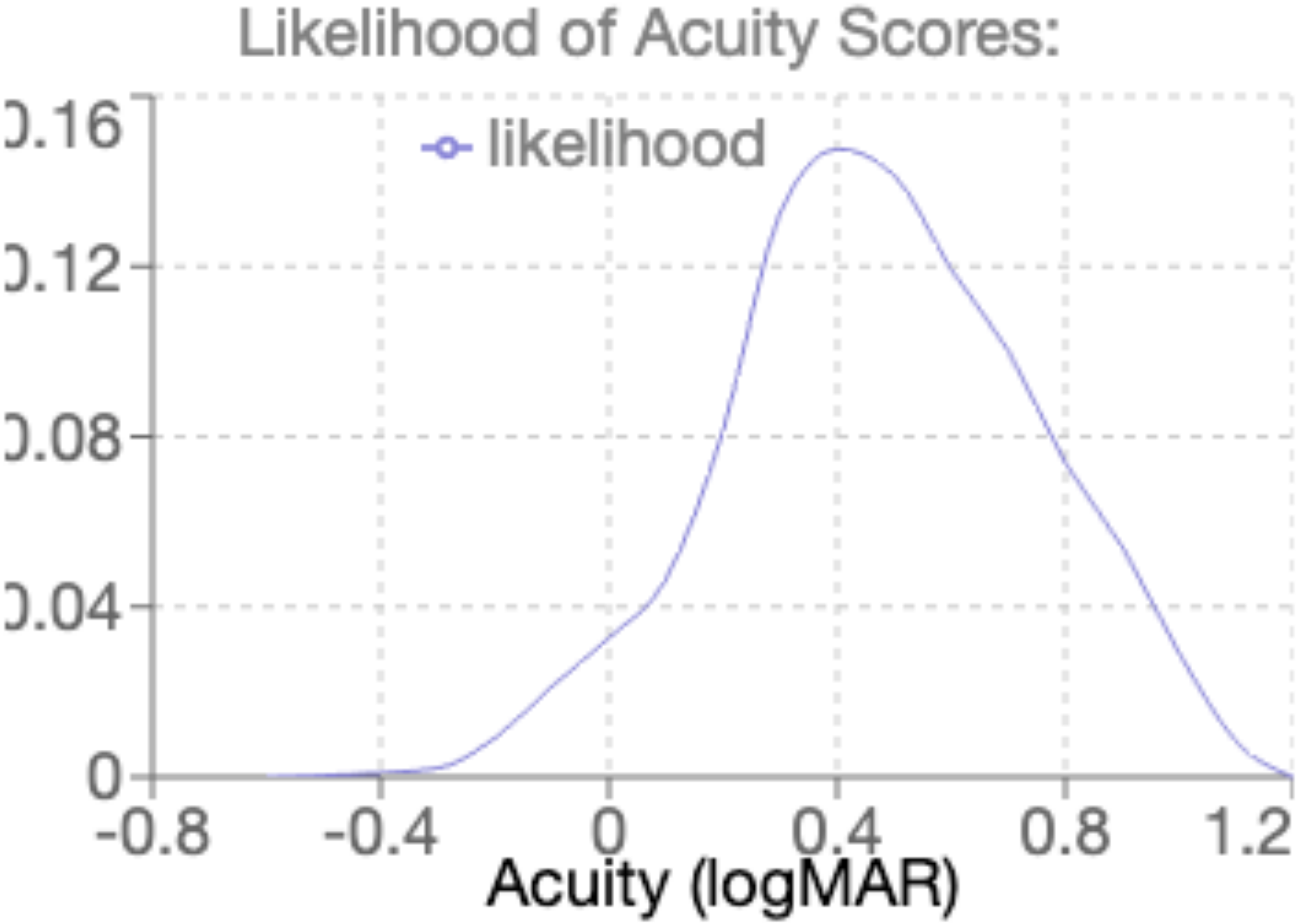
$$p = \sigma(\theta - f)$$

Where θ is the user's vision score and f is the font size of the letter.

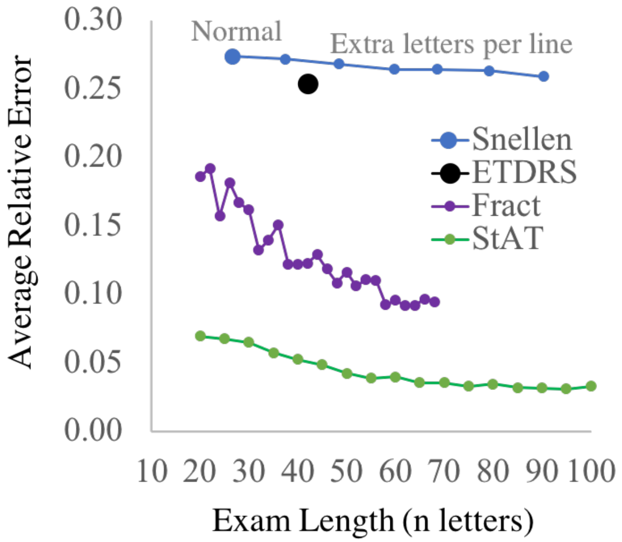
Explain how you could estimate a user's vision score (θ) based on their 20 responses $(f^{(1)}, y^{(1)}) \dots (f^{(20)}, y^{(20)})$, where $y^{(i)}$ is an indicator variable for whether the user correctly identified the ith letter and $f^{(i)}$ is the font size of the ith letter. Solve for any and all partial derivatives required by your answer.

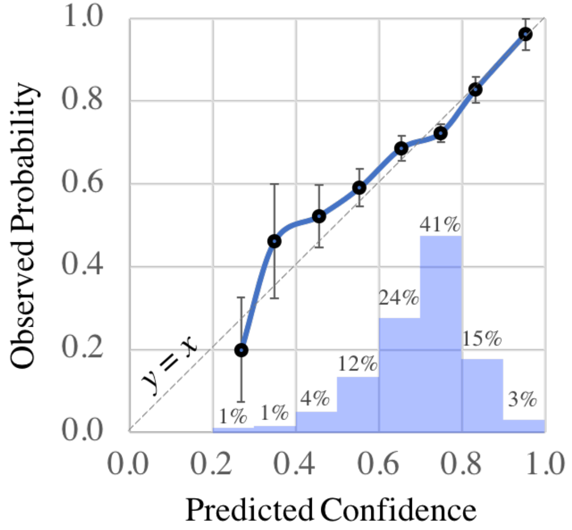
Learn Real Skills in CS109

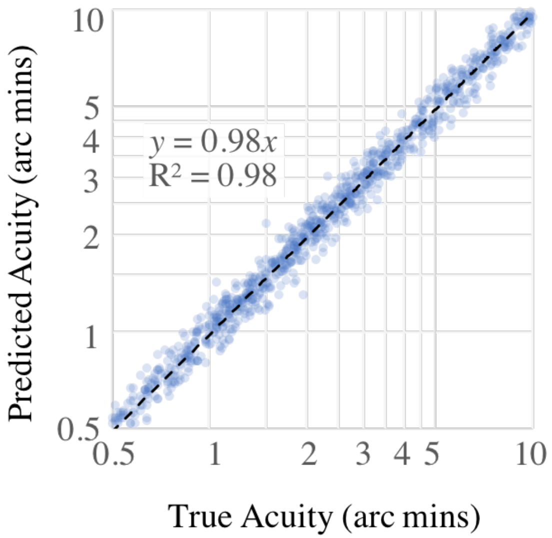
A patient has answered 20 “letter sizes” and got a few correct. What is your belief in how well they can see?



Now state of the art for eye exam theory







Learn Real Skills in CS109

The Stanford Acuity Test: A Precise Vision Test Using Bayesian Techniques and a Discovery in Human Visual Response

Chris Piech,*[1] Ali Malik,*[1] Laura M Scott, Robert T Chang,[2] Charles Lin[2]
[1]Department of Computer Science, Stanford University
[2] Department of Ophthalmology, Stanford University
{piech, malikali}@cs.stanford.edu, {rchang3, lincc}@stanford.edu

Abstract

Chart-based visual acuity measurements are used by billions of people to diagnose and guide treatment of vision impairment. However, the ubiquitous eye exam has no mechanism for reasoning about uncertainty and as such, suffers from a well-documented reproducibility problem. In this paper we make two core contributions. First, we uncover a new parametric probabilistic model of visual acuity response based on detailed measurements of patients with eye disease. Then, we present an adaptive, digital eye exam using modern artificial intelligence techniques which substantially reduces acuity exam error over existing approaches, while also introducing the novel ability to model its own uncertainty and incorporate prior beliefs. Using standard evaluation metrics, we estimate a 74% reduction in prediction error compared to the ubiquitous chart-based eye exam and up to 67% reduction compared to the previous best digital exam. For patients with eye disease, the novel ability to finely measure acuity from home could be a crucial part in early diagnosis. We provide a web implementation of our algorithm for anyone in the world to use. The insights in this paper also provide interesting implications for the field of psychometric Item Response Theory.

1 Introduction

Reliably measuring a person's visual ability is an essential component in the detection and treatment of eye diseases around the world. However, quantifying how well an individual can distinguish visual information is a surprisingly difficult task—without invasive techniques, physicians rely on chart-based eye exams where patients are asked visual questions and their responses observed.

Historically, vision has been evaluated by measuring a patient's *visual acuity*: a measure of the font size at which a patient can correctly identify letters shown a fixed distance away. Snellen, the traditional eye exam chart, determines this statistic by marching down a set of discrete letter sizes, asking the patient a small number of questions per size to indentify the size where the patient gets less than half the letters correct. This approach is simple and is used daily in the treatment of patients; yet, it suffers from some notable shortcomings. Acuity exams such as these exhibit high variance in their results due to the large role that chance plays in the final diagnosis, and the approximation error incurred by the need to discretise letter sizes on a chart. On the other hand, digital exams can show letter of any size and can *adaptively* make decisions based on intelligent probabilistic models. As such they have potential to address the shortcomings of analog charts.

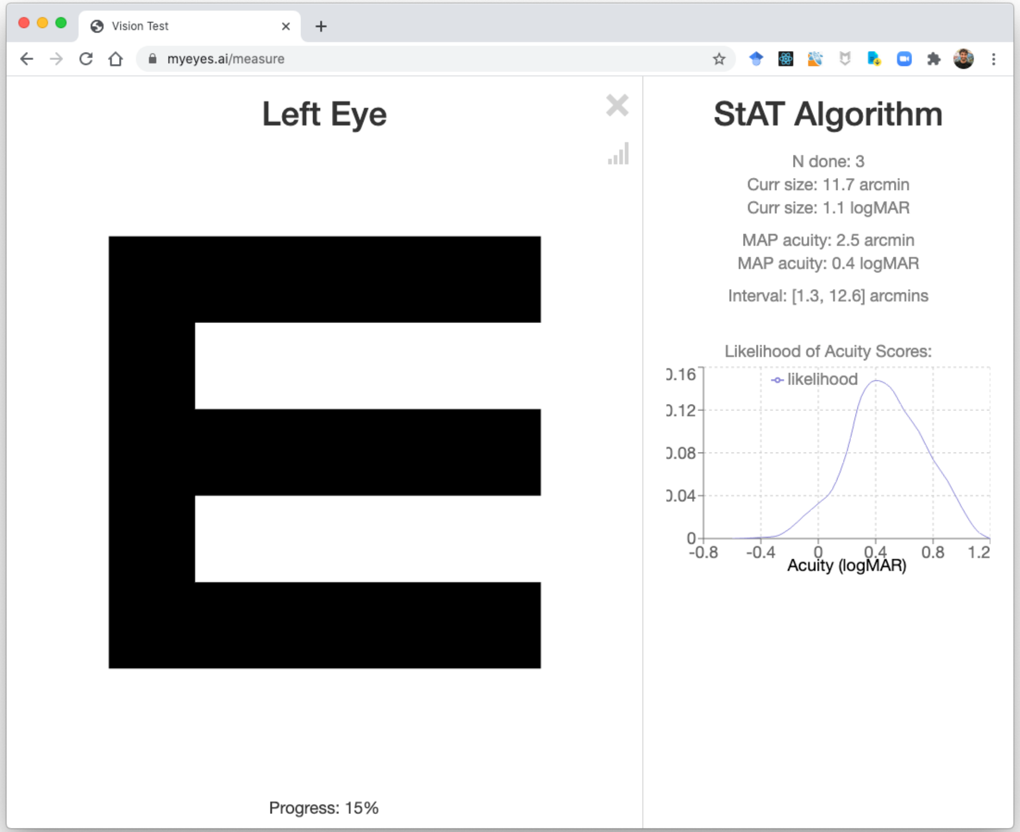
While promising, contemporary digital exams have yet to dramatically improve accuracy over traditional chart-based approaches. The current best digital exam uses a psychometric Item Response Theory (IRT) algorithm for both selecting the next letter size to query and for making a final prediction of acuity. Under simulation analysis, this digital exam results in a 19% reduction in error over traditional chart-based approaches. The separate fields of reinforcement learning and psychometric IRT have independently explored how to effectively make decisions under uncertainty. By merging the good ideas from both disciplines we can develop a much better visual acuity test.

In this paper we make two main contributions. First, we revisit the human Visual Response Function—a function relating the size of a letter to the probability of a person identifying it correctly—and discover that it follows an interpretable parametric form that fits real patient data. Second, we present an algorithm to measure a person's acuity which uses several Bayesian techniques common in modern artificial intelligence. The algorithm, called the Stanford Acuity Test (StACT)[1], has the following novel features:

1. Uses the new parametric form of the human Visual Response Function.
2. Returns a soft inference prediction of the patient's acuity, enabling us to express a calibrated confidence in the final result.
3. Uses a posterior probability matching algorithm to adaptively select the next letter size shown to a user. This effectively balances exploration of the acuity belief.

*Equal contribution.

[1]The previous state-of-the-art, FrACT, was named after Freiburg, the city in which it was developed. We continue in this tradition.



Foundation for your future

But its not always intuitive

But Its not Always Intuitive

A patient has a positive Zika test.

What is the probability they have zika?

- *0.8% of people have zika*
- *Test has 90% positive rate for people with zika*
- *Test has 7% positive rate for people without zika*

The right answer is 9%

Probability = Important + Needs Study

Delayed gratification

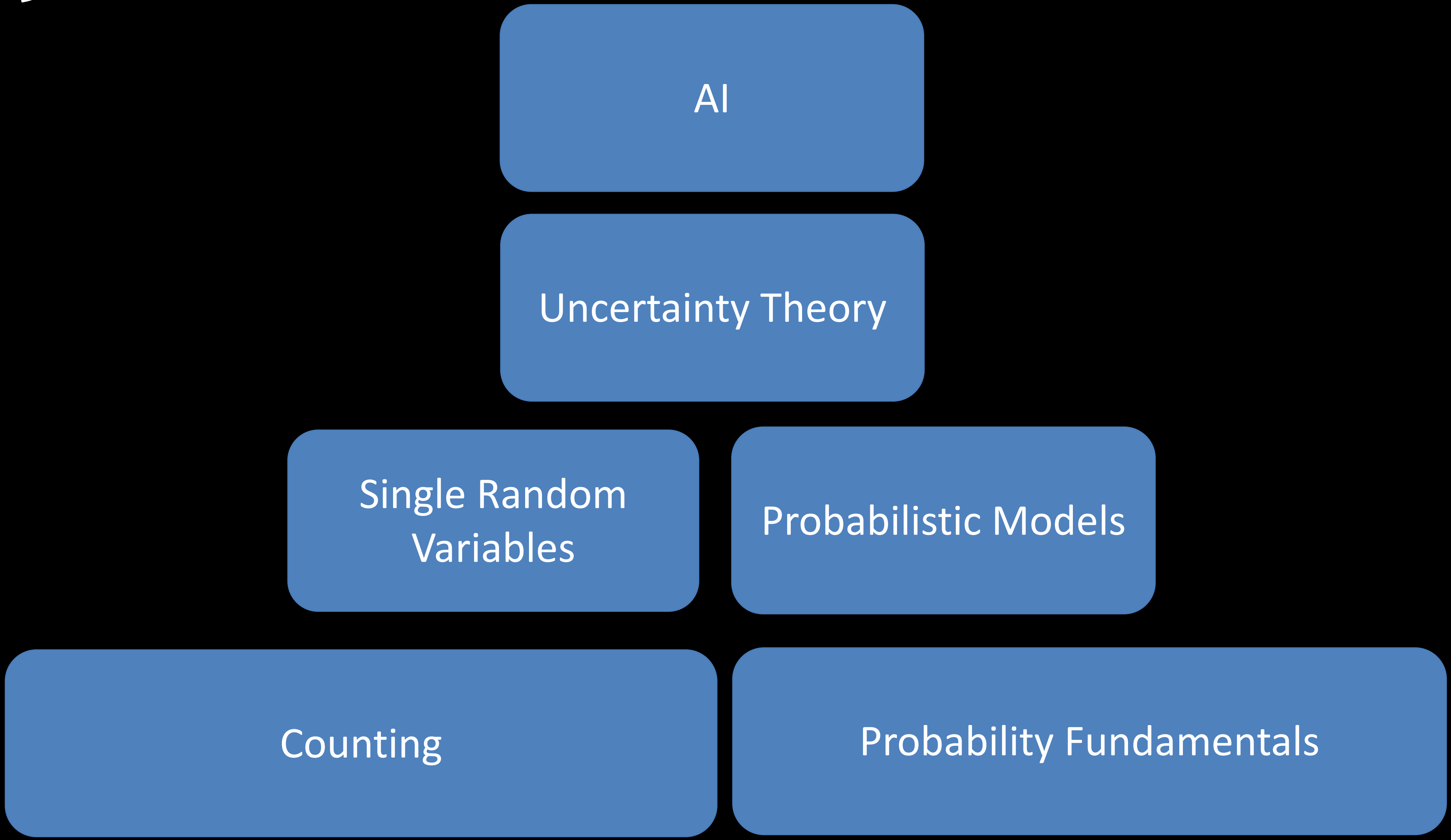
CS109 View of Probability

Teach you how to write programs
that most people are not able to write.

CS109 View of Probability

Teach you the theory you need to do the math that most people are not able to do.

CS109


AI
Uncertainty Theory
Single Random Variables
Probabilistic Models
Counting
Probability Fundamentals

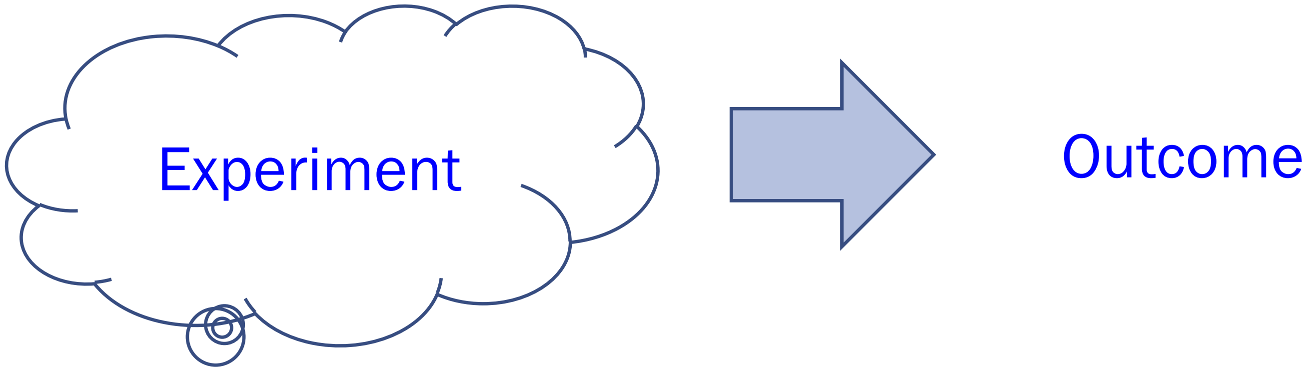
Lets dive in…

2 min pedagogic pause.

Counting I

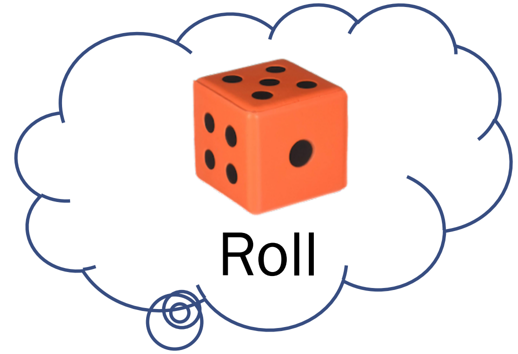
What is Counting?

An experiment in probability:



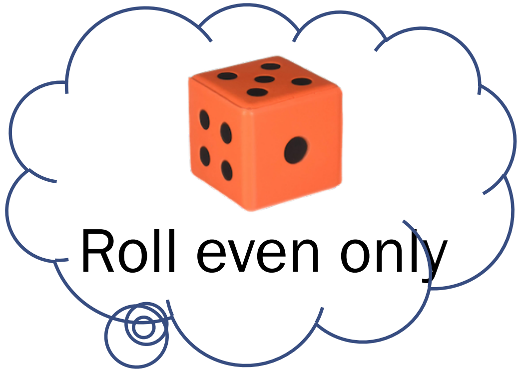
Counting: How many possible outcomes can occur from performing this experiment?

What is Counting?

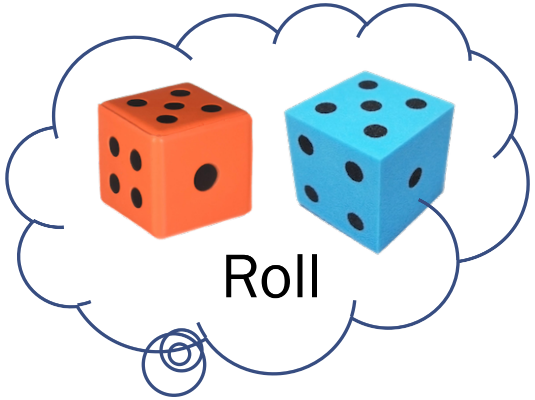


6

{1, 2, 3,
4, 5, 6}



3 {2, 4, 6}



36

{(1, 1) , (1, 2), (1, 3), (1, 4), (1, 5), (1, 6),
(2, 1) , (2, 2), (2, 3), (2, 4), (2, 5), (2, 6),
(3, 1) , (3, 2), (3, 3), (3, 4), (3, 5), (3, 6),
(4, 1) , (4, 2), (4, 3), (4, 4), (4, 5), (4, 6),
(5, 1) , (5, 2), (5, 3), (5, 4), (5, 5), (5, 6),
(6, 1) , (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)}

Step Rule of Counting (aka Product Rule of Counting)

If an experiment has two steps, where

> The first step's outcomes are from Set A, where $|A| = m$,
> <u>and</u> the second step's outcomes are from Set B, where $|B| = n$,
>
> and |B| is unaffected by outcome of first step.

Then the number of outcomes of the experiment is

$$|A||B| = mn.$$

Two-step experiment

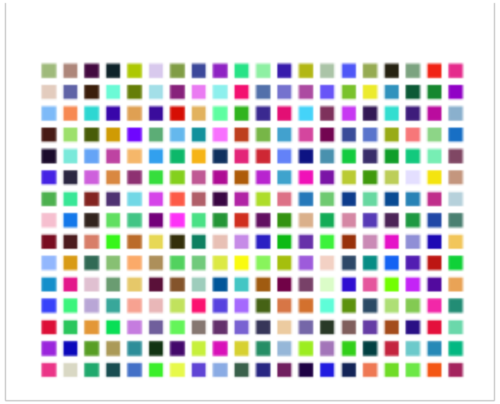
→ A → B

[On board]

How Many Unique Images?

Each pixel can be one of 17 million distinct colors

(a) 12 million pixels



(b) 300 pixels

(c) 12 pixels

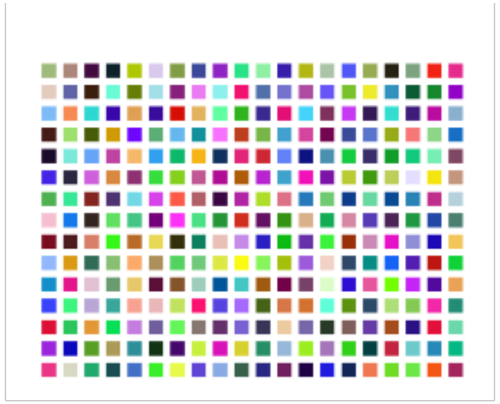
$$(17 \text{ million})^n$$

How Many Unique Images?

Each pixel can be one of 17 million distinct colors

(a) 12 million pixels

$\approx 10^{86696638}$



(b) 300 pixels

$\approx 10^{2167}$

(c) 12 pixels

$\approx 10^{86}$

$(17 \text{ million})^n$

Sum Rule of Counting

If the outcome of an experiment can be either from

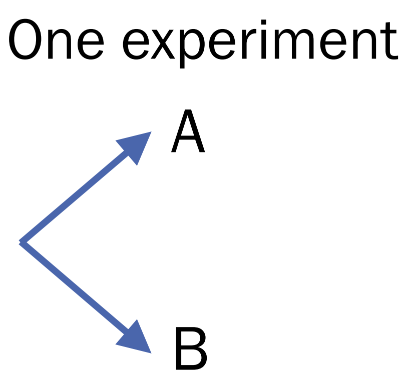
Set A, where $|A| = m$,

<u>or</u> Set B, where $|B| = n$,

where $A \cap B = \emptyset$,

Then the number of outcomes of the experiment is

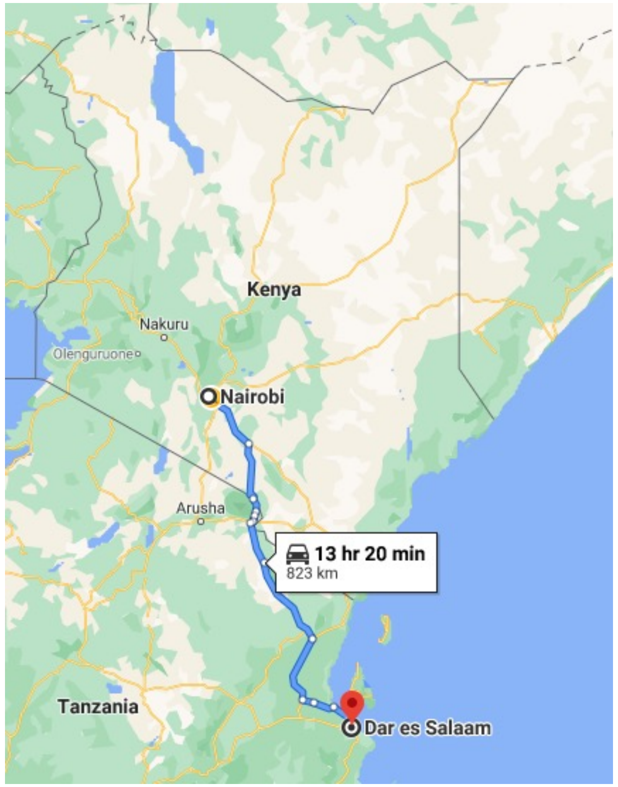
$|A| + |B| = m + n$.



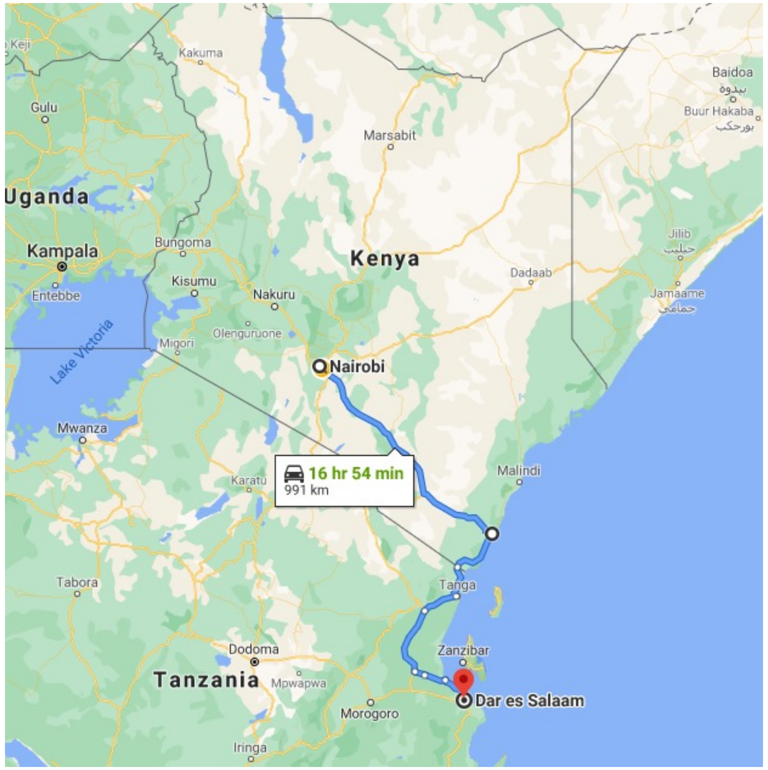
How many routes

Question: All routes considered by google maps from Nairobi to Dar es Salaam go through either Mt Kilimanjaro **or** Mombasa. How many total routes are considered?

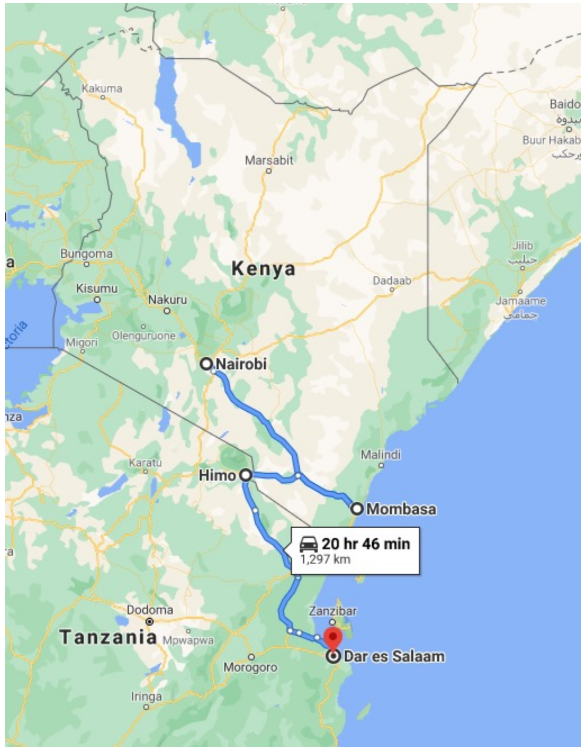
20 routes go through Mt Kili



10 routes go through Mombasa



0 go through both



Answer: 20 + 10

How Many Bit Stings?

Problem: A 6-bit string is sent over a network. The valid set of strings recognized by the receiver must either start with "01" or end with "10". How many such strings are there?

Answer

2^4 start with 01	2^4 end with 10
010000	000010
010001	000110
010010	001010
010011	001110
010100	010010
010101	010110
010110	011010
010111	011110
011000	100010
011001	100110
011010	101010
011011	101110
011100	110010
011101	110110
011110	111010
011111	111110
Set A	Set B

How Many Bit Stings?

Problem: A 6-bit string is sent over a network. The valid set of strings recognized by the receiver must either start with "01" or end with "10". How many such strings are there?

Answer

2^4 start with 01	2^4 end with 10
010000	000010
010001	000110
010010	001010
010011	001110
010100	**010010**
010101	**010110**
010110	**011010**
010111	**011110**
011000	100010
011001	100110
011010	101010
011011	101110
011100	110010
011101	110110
011110	111010
011111	111110
Set A	Set B

How Many Bit Stings?

Problem: A 6-bit string is sent over a network. The valid set of strings recognized by the receiver must either start with "01" or end with "10". How many such strings are there?

Answer

2^4 start with 01	2^4 end with 10
010000	000010
010001	000110
010010	001010
010011	001110
010100	010010
010101	010110
010110	011010
010111	011110
011000	100010
011001	100110
011010	101010
011011	101110
011100	110010
011101	110110
011110	111010
011111	111110
Set A	Set B

How Many Bit Stings?

Problem: A 6-bit string is sent over a network. The valid set of strings recognized by the receiver must either start with "01" or end with "10". How many such strings are there?

Answer

$$N = |A| + |B| - |A \text{ and } B|$$
$$= 16 + 16 - 4$$
$$= 28$$

2^4 start with 01	2^4 end with 10
010000	000010
010001	000110
010010	001010
010011	001110
010100	**010010**
010101	**010110**
010110	**011010**
010111	**011110**
011000	100010
011001	100110
011010	101010
011011	101110
011100	110010
011101	110110
011110	111010
011111	111110
Set A	Set B

Or Rule of Counting (aka Inclusion/ Exclusion)

If the outcome of an experiment can be either from

Set A, where $|A| = m$,

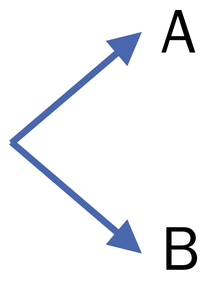
or Set B, where $|B| = n$,

where $A \cap B$ $may\ not\ be\ empty$,

Then the number of outcomes of the experiment is

N = $|A| + |B| - |A \cap B|$.

One experiment



Challenge Problem

1. Strings
   - How many *different* orderings of letters are possible for the string BAYES?
   - How about BOBA?

BOBA, ABOB, OBBA…

🤔

Concept Check!

Incredible time and school at which to study probability! Exciting.