CS109: Probability for Computer Scientists

Instructors

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!

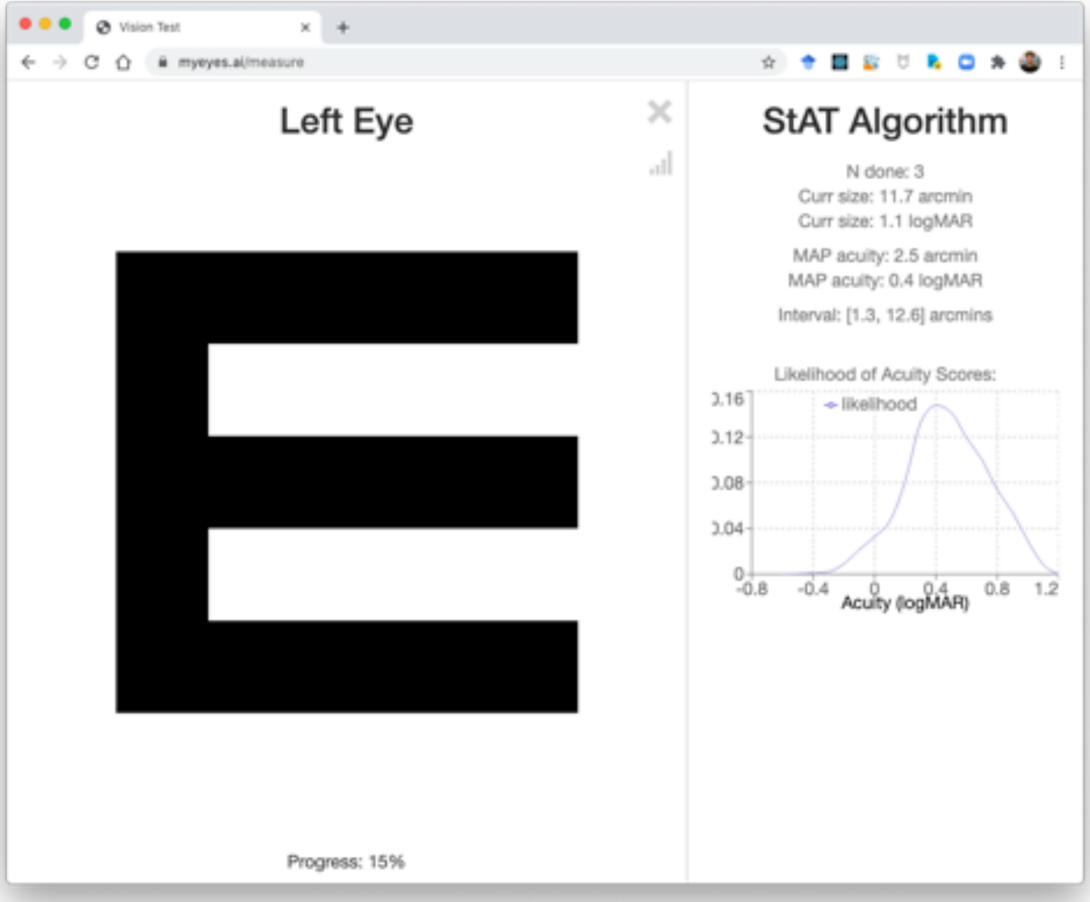
Long History in CS109

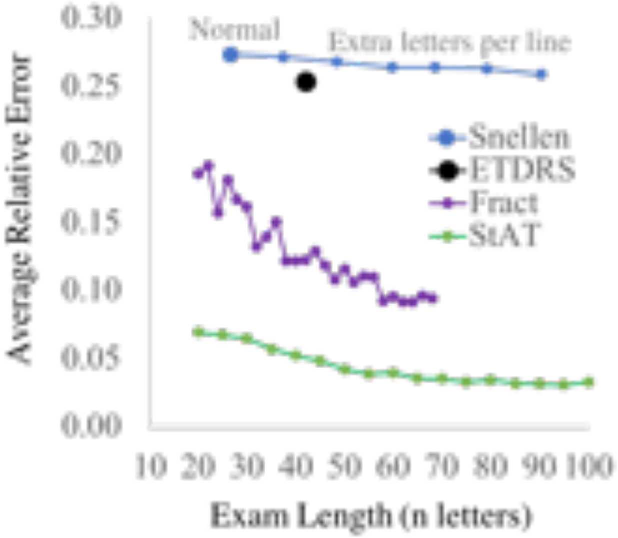
I took the first CS109 back when I looked like this

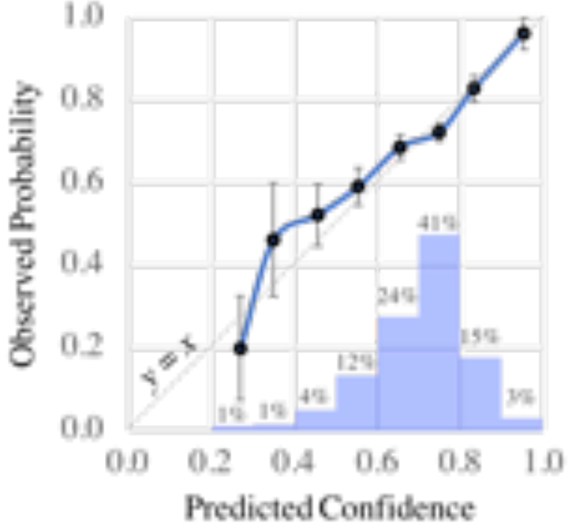
Been teaching it since 2014

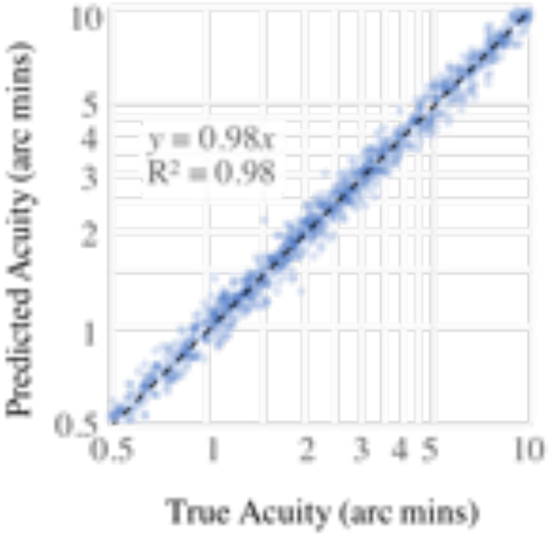
Probability in my Research: Better Eye Exam

Jan 2020, With a former CS109 student, Ali Malik





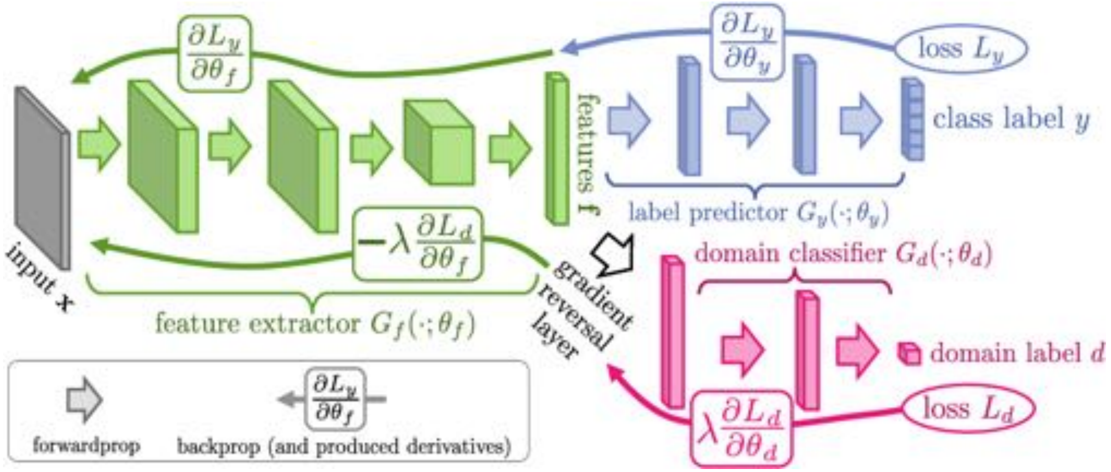




Math question: Estimate a continuous valued number.
Get to run noisy experiments of your choosing.

Fair AI with Adversarial Network

2018, with undergrads Christina Wadsworth and Francesca Vera



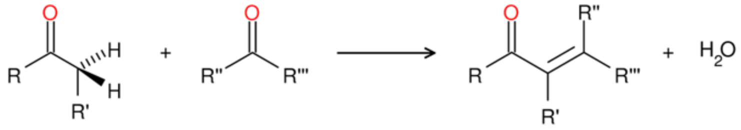
Model	Accuracy	FP Gap	FN Gap
COMPAS Scores (Our Test Set)	0.68	0.17	0.22
Our Recidivism Model	0.70	0.15	0.27
Our Chosen Adversarial Model	0.70	**0.01**	**0.02**
Bechavod et al. AVD Penalizers (2017)	0.65	**0.02**	**0.04**
Bechavod et al. SD Penalizers (2017)	0.66	**0.02**	**0.03**
Bechavod et al. Vanilla 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?

So many things to love in this world

Jerry Cain

I went here from 1987 through 1991 and majored in chemistry.



Then I came here for a PhD in chem, switched to CS

Received MSCS 1998
Lecturer: not so new
Speculation: Taught more classes than anyone else here.

<u>My interests over time</u>

Chemistry and Physics	
STEM Education	
Make Sense Need Nods	

Why Jerry likes probability

- I majored in chemistry, and my undergraduate research was rooted in surface science and **statistical** mechanics.
- When I switched to CS as a grad student here, I focused on CS theory and all the beautiful mathematics that comes with it. Mathematics feeds the soul.
- Probability has revived parts of AI and information theory that were thought to be borderline dead when I was getting my MSCS degree here.
- PV=nRT?

1974

1996

$$PV = \frac{1}{3}Nmv_{\text{rms}}^2. \quad f(v) = 4\pi\left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}} v^2 e^{-\frac{mv^2}{2kT}} \quad v_{\text{rms}}^2 = \int_0^\infty v^2 f(v)\,dv = 4\pi\left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}} \int_0^\infty v^4 e^{-\frac{mv^2}{2kT}}\,dv$$

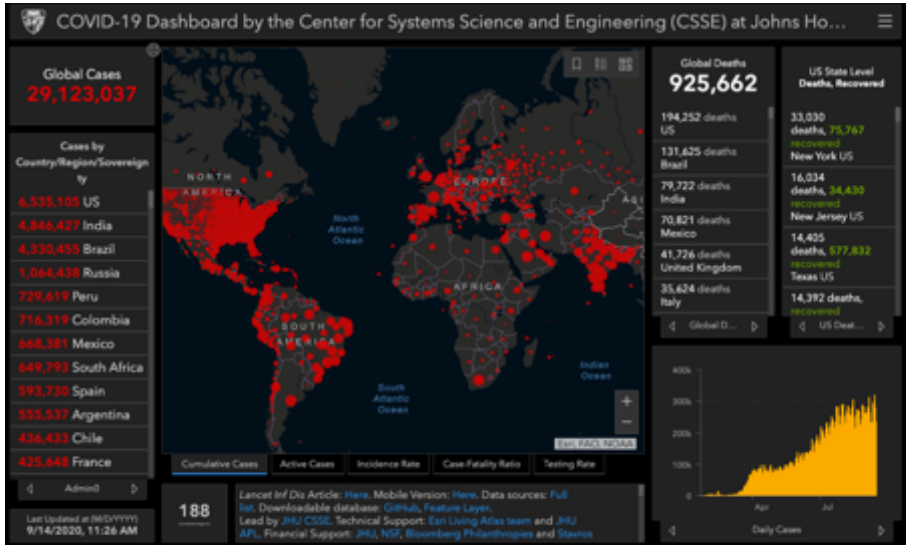
Amazing Teaching Team

Course mechanics

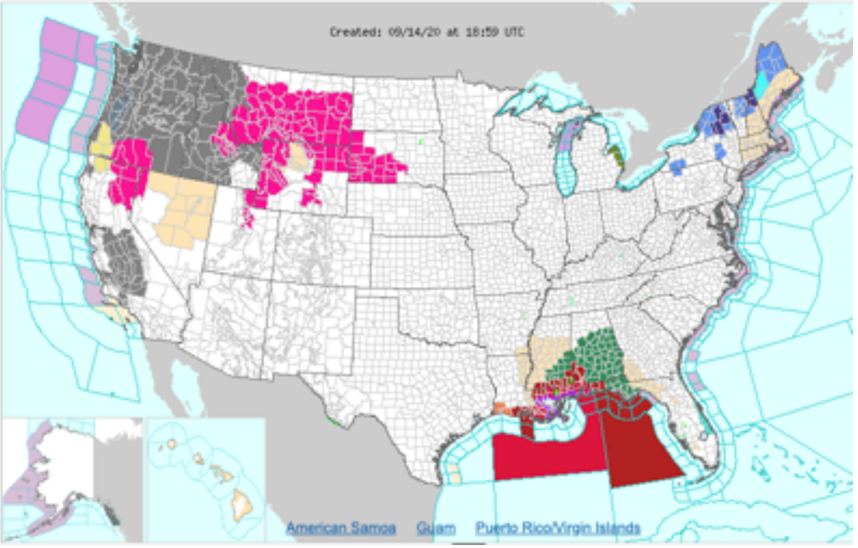
(this is a light version. Please read the handout for details).

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/

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.

The challenge of delivering Stanford-class education online reflects our university's commitment to fostering a **diverse body of students**.

CS109 Spring 2020

Essential Information

cs109.stanford.edu

Zoom Zoom

- **Camera on** during breakouts.
- **Camera off** during lecture is just fine!
- Come to lecture, it’s a great time.
- **Ask questions** in the chat.
- Use Ed for questions after lecture.

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)

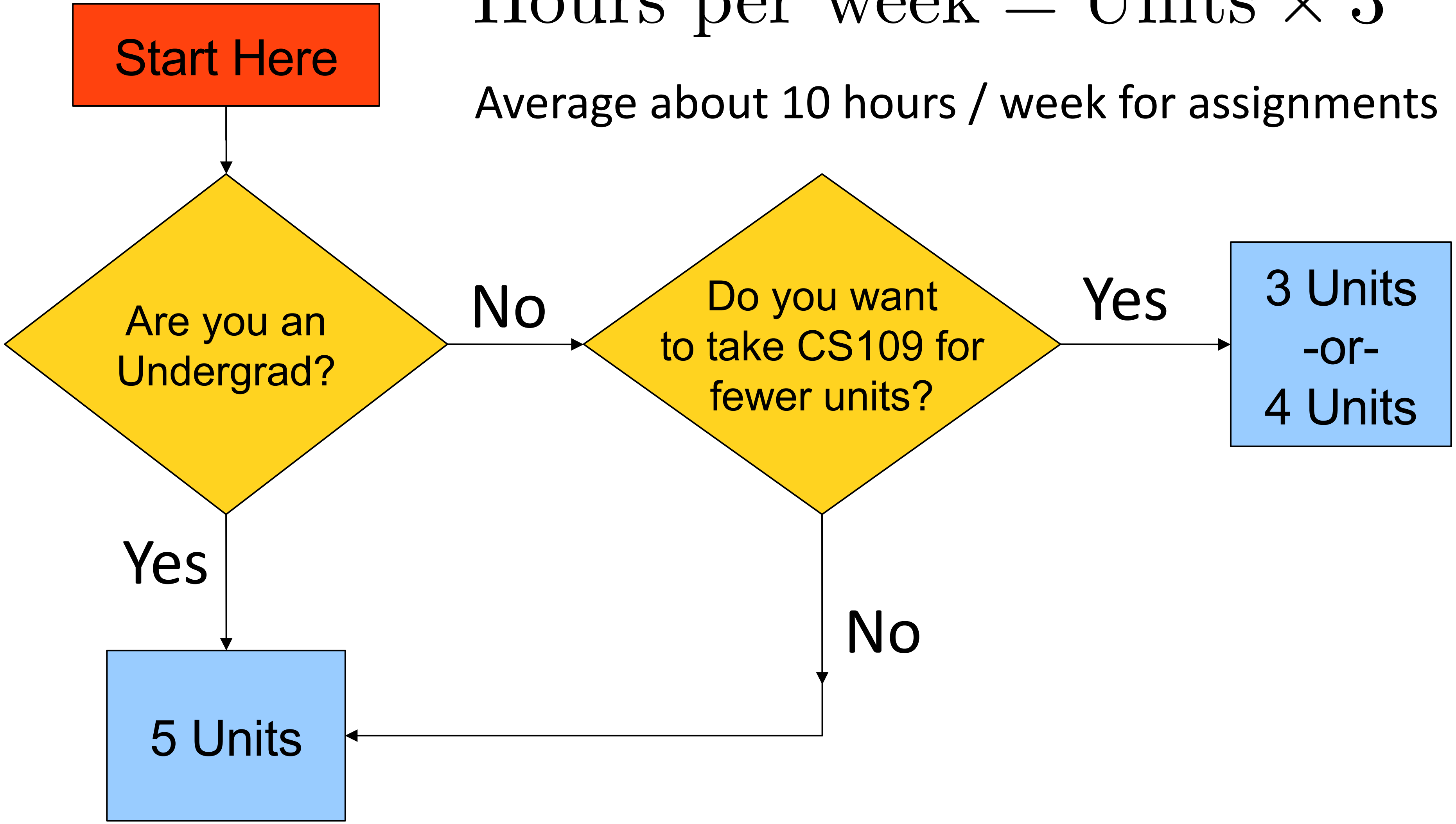
Coding in CS109

Review session on Friday

CS109 Units

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

Average about 10 hours / week for assignments



Class Breakdown

42%	**6 Assignments**
42%	**3 Quizzes** 48 hour take home tests
10%	**Concept Checks** Daily after lecture
6%	**Section Participation**

Ask questions

Q&A forum
All announcements

"Working" office hours start after Wednesday class

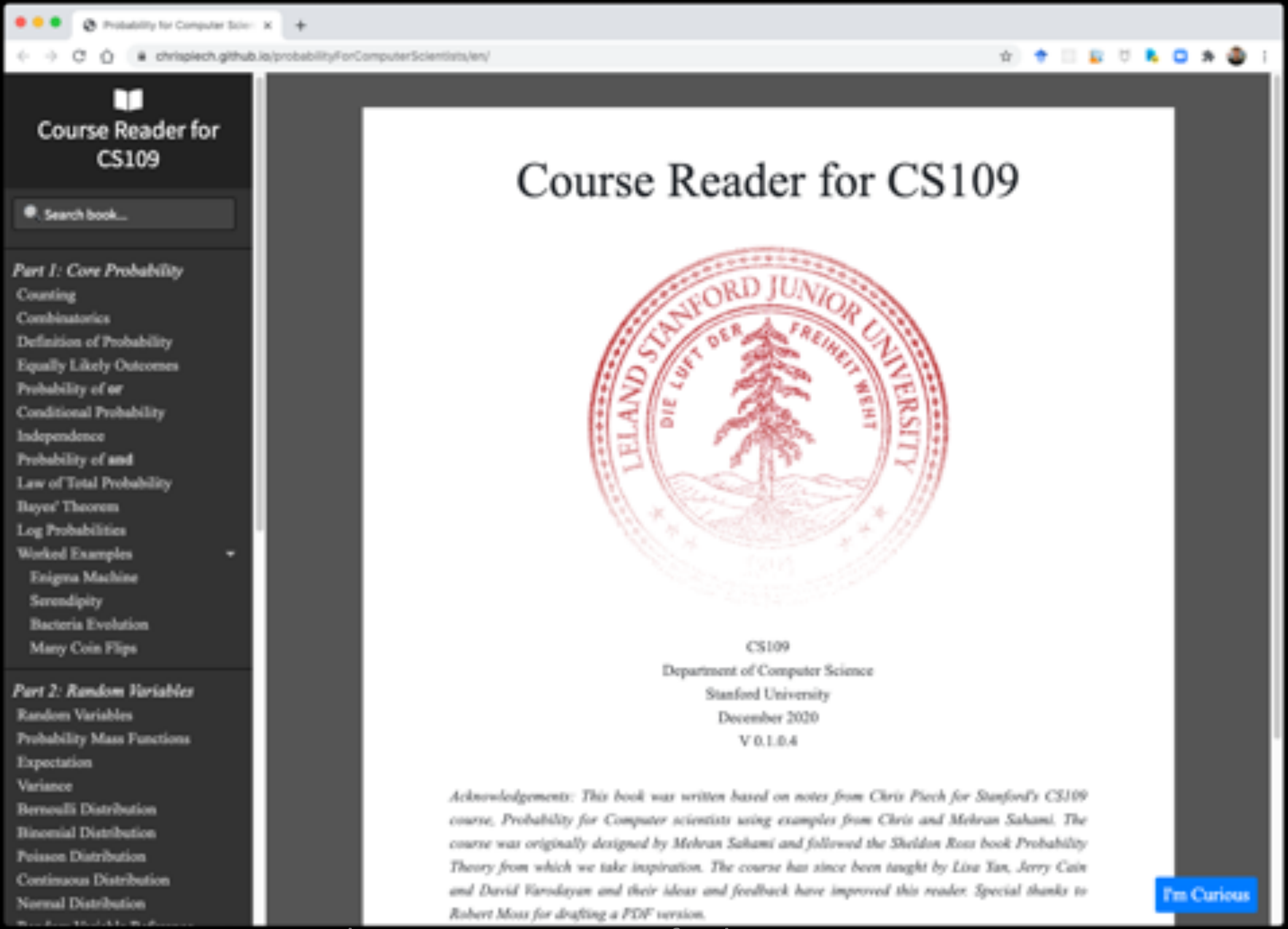
Email cs109@cs.stanford.edu

Companion class: CS109A

- CS109A is an extra 1-unit "Pathfinders" or "ACE" section with additional support, practice, and instruction
- Meets for an additional weekly section and has additional review session
- Entry by application — see course website for details

Georgia Sampaio

Brand new this quarter: Course Reader!



Story of Modern AI

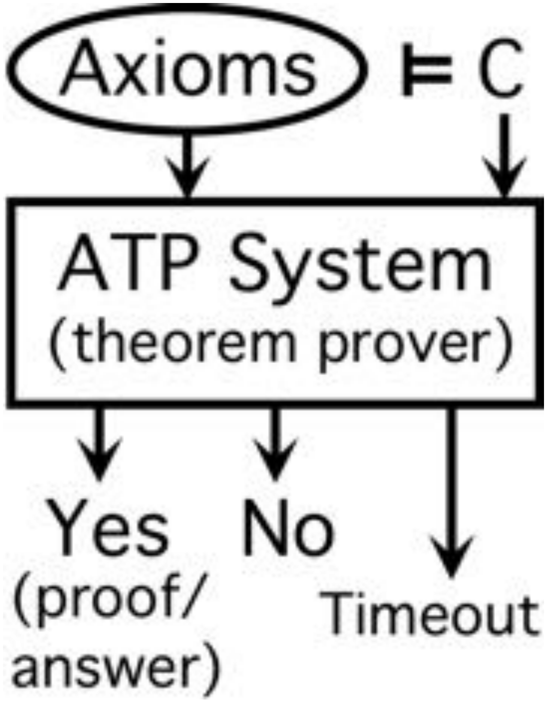
Modern AI
or, How we learned to combine probability and programming

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

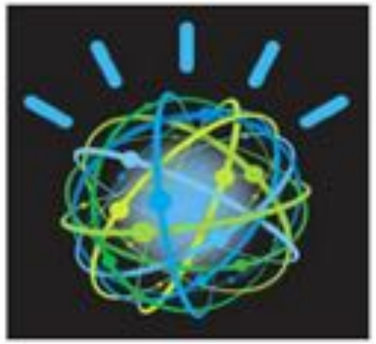

BRACE YOURSELVES
WINTER IS COMING

Something is going on in the world of AI

Big Milestones Part 1

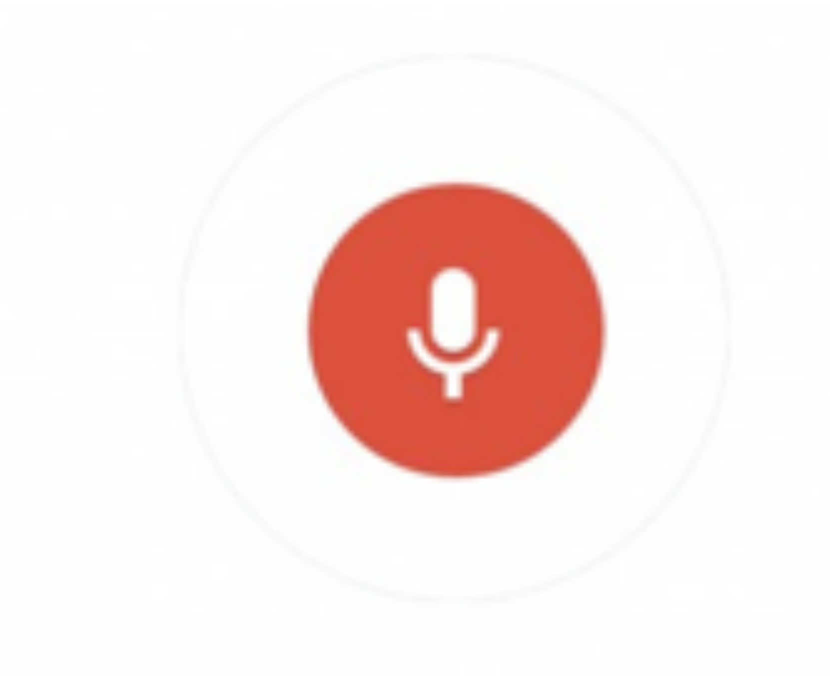
1997 Deep Blue

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

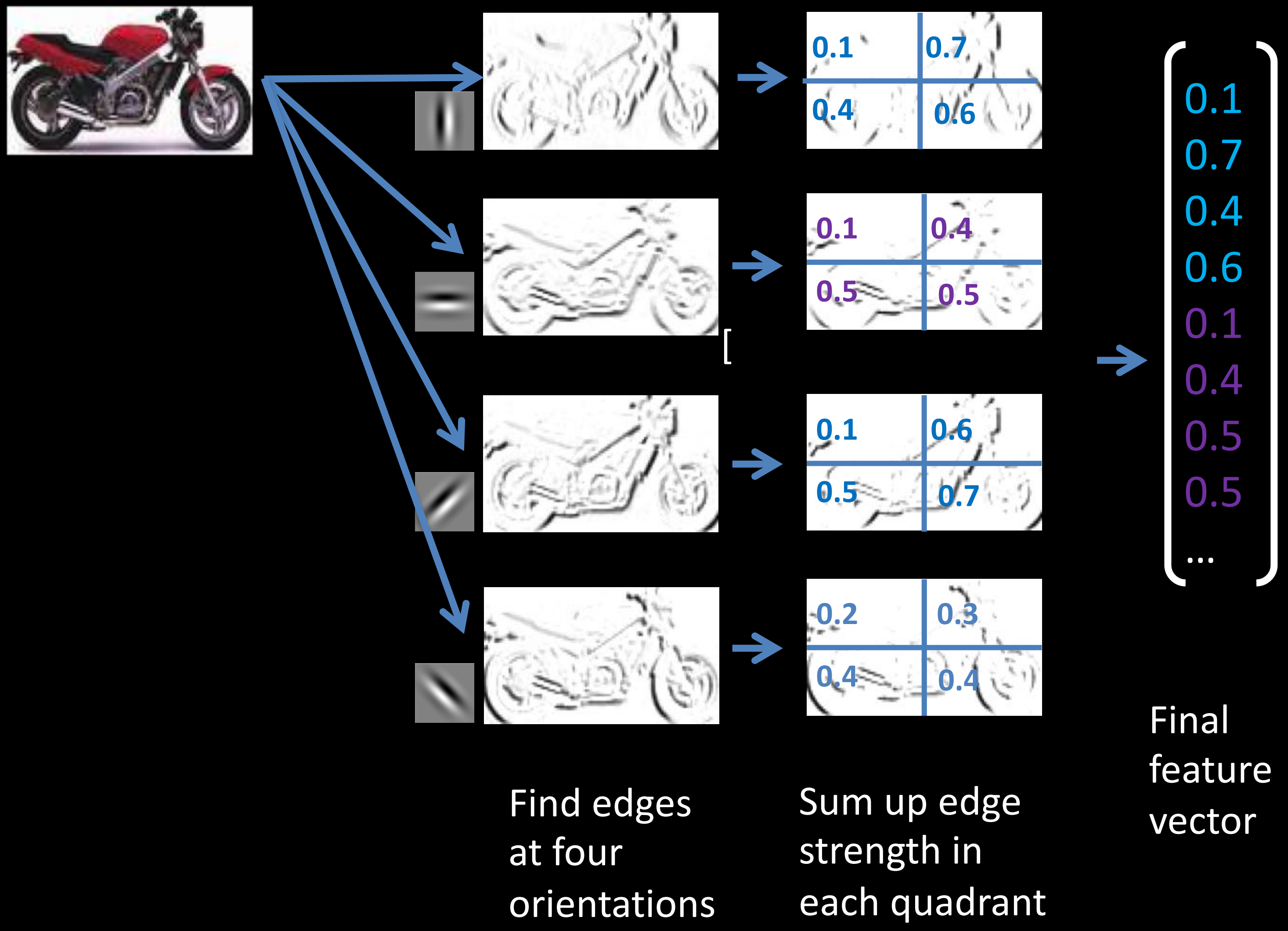
You see this:

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

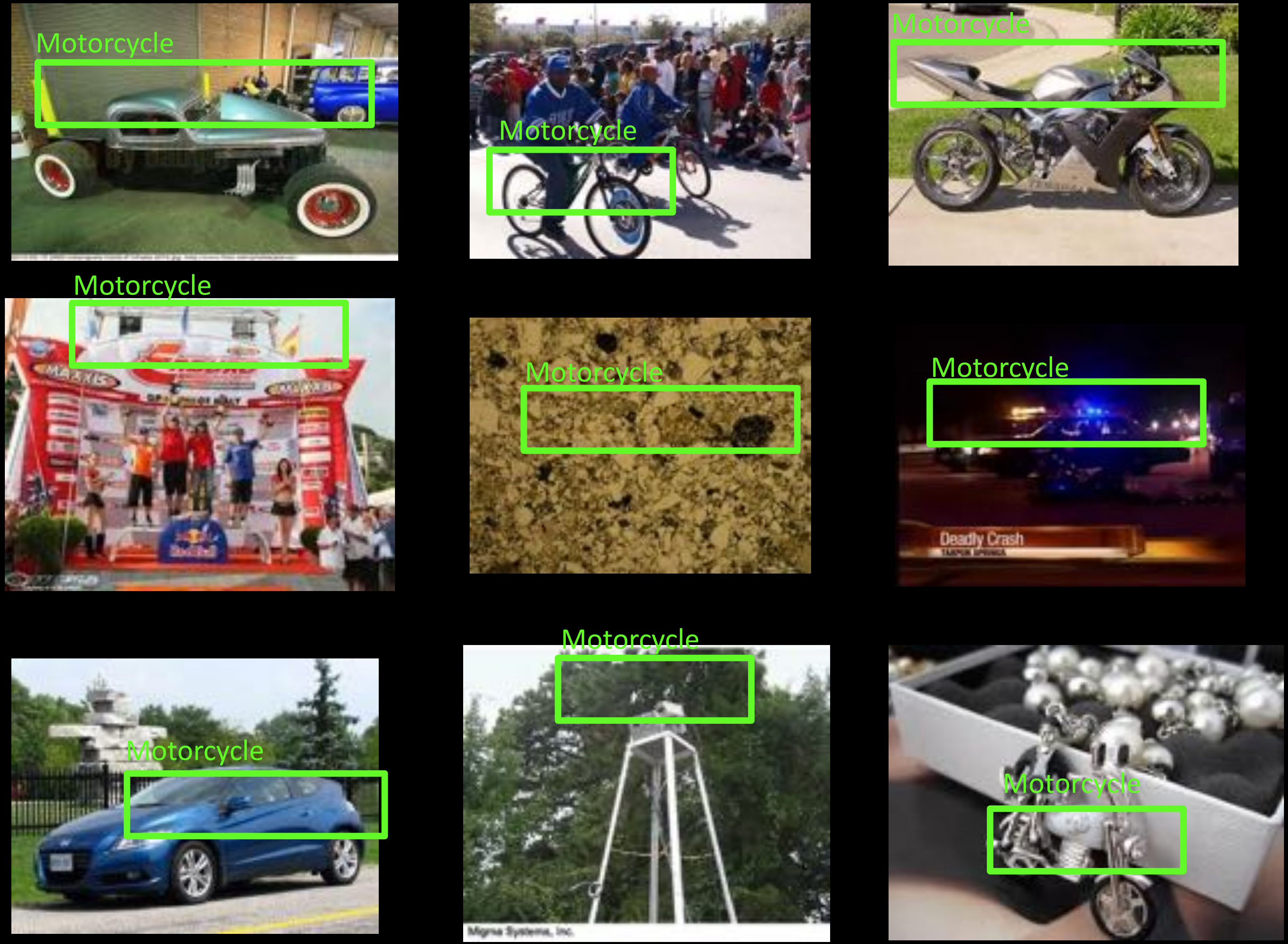
[Andrew Ng]

Computer Vision is Still Hard



[Andrew Ng]

Straight ML not perfect



Some Great Thinkers

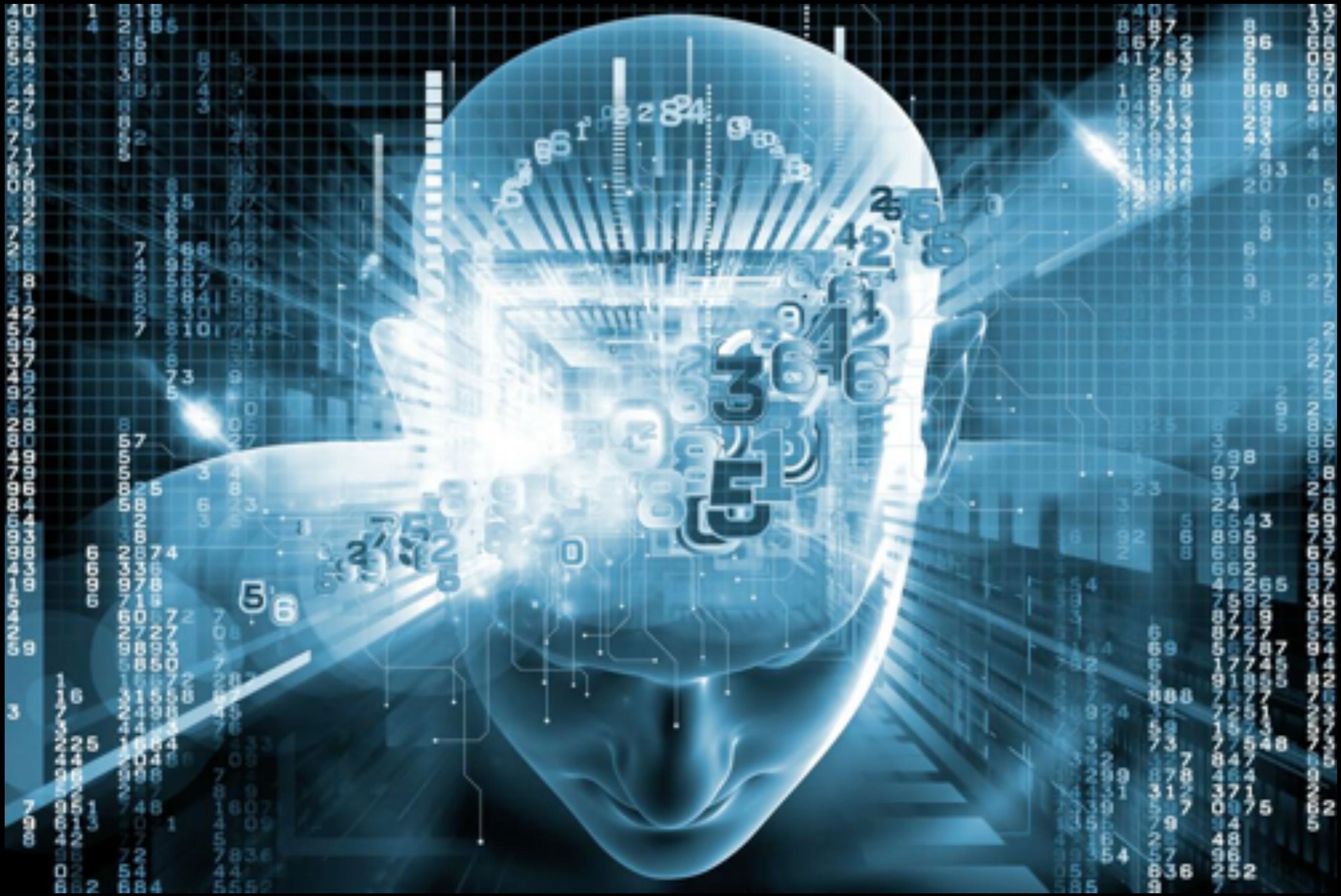
Daphne Koller

Two Great Ideas

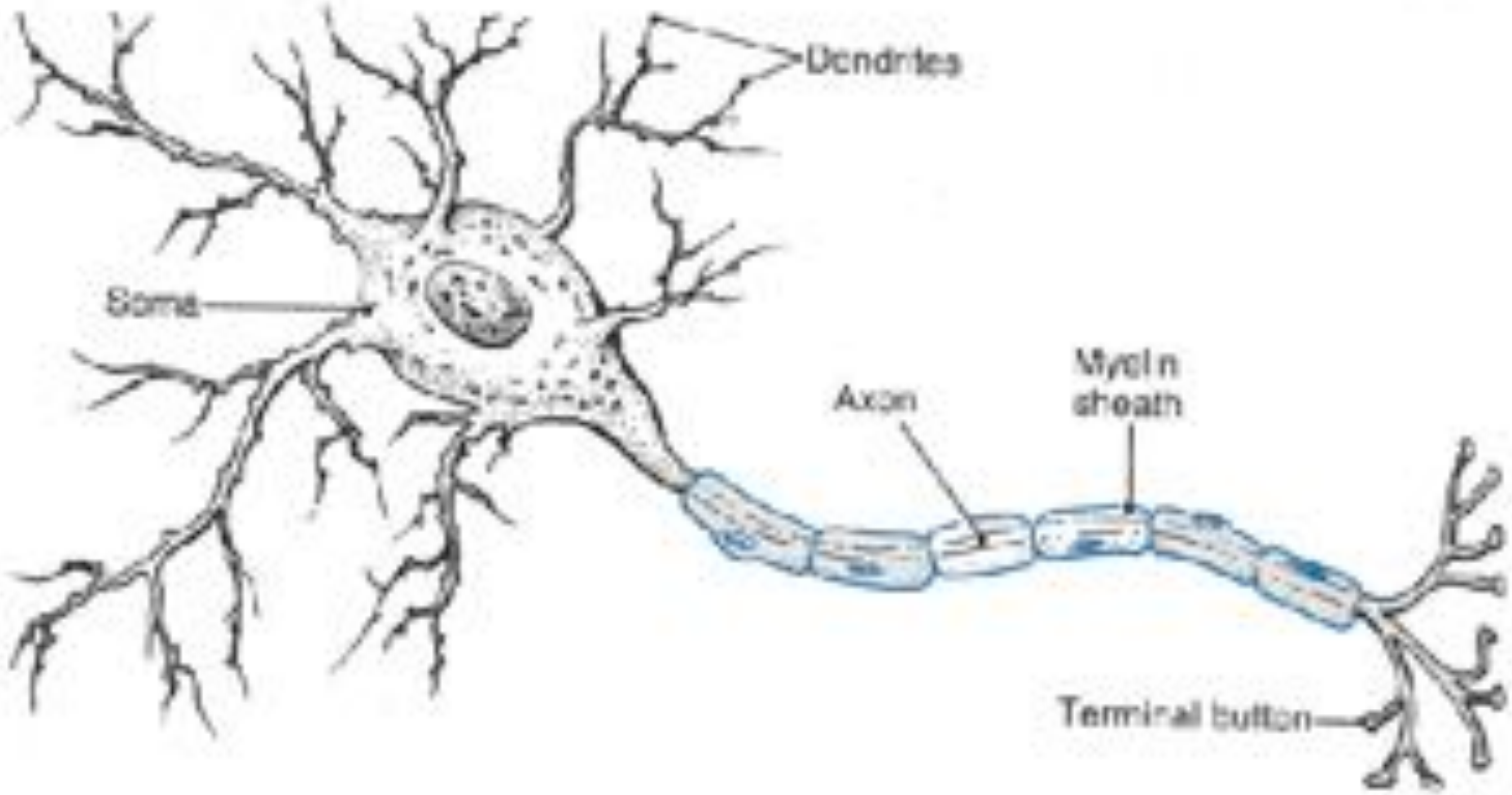
1. Probability from Examples

2. Artificial Neurons

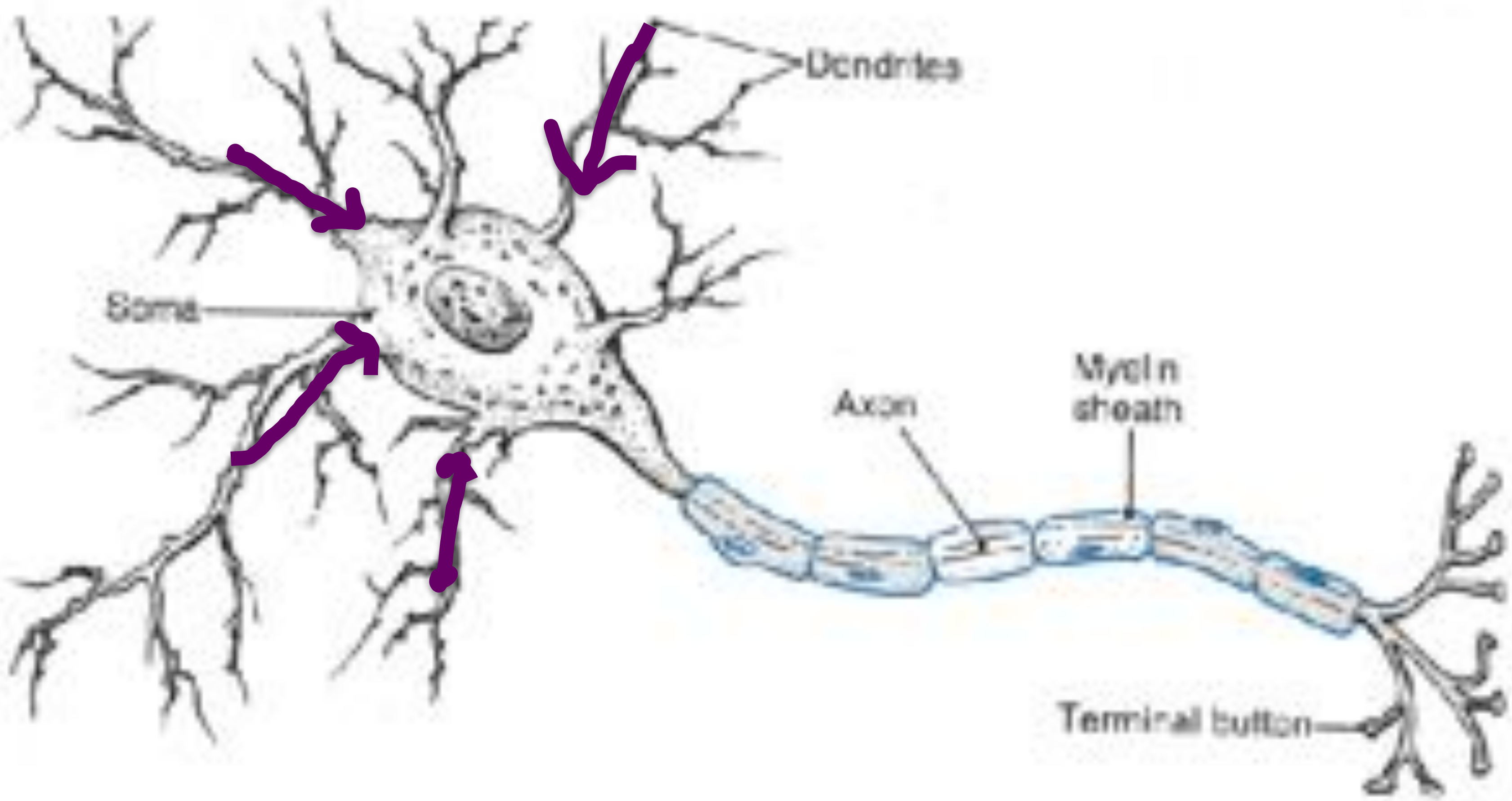
Artificial Neurons



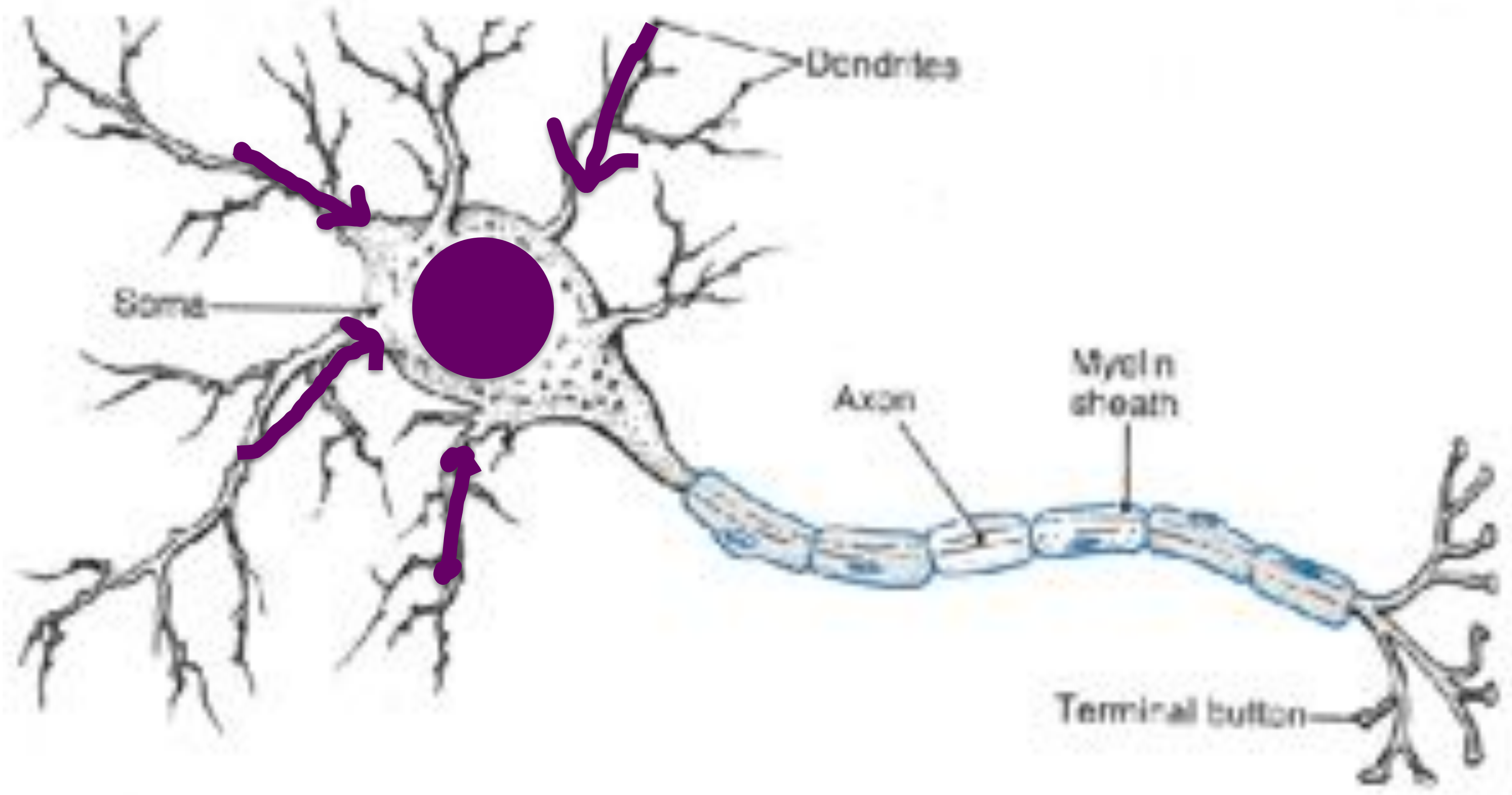
Human Neuron



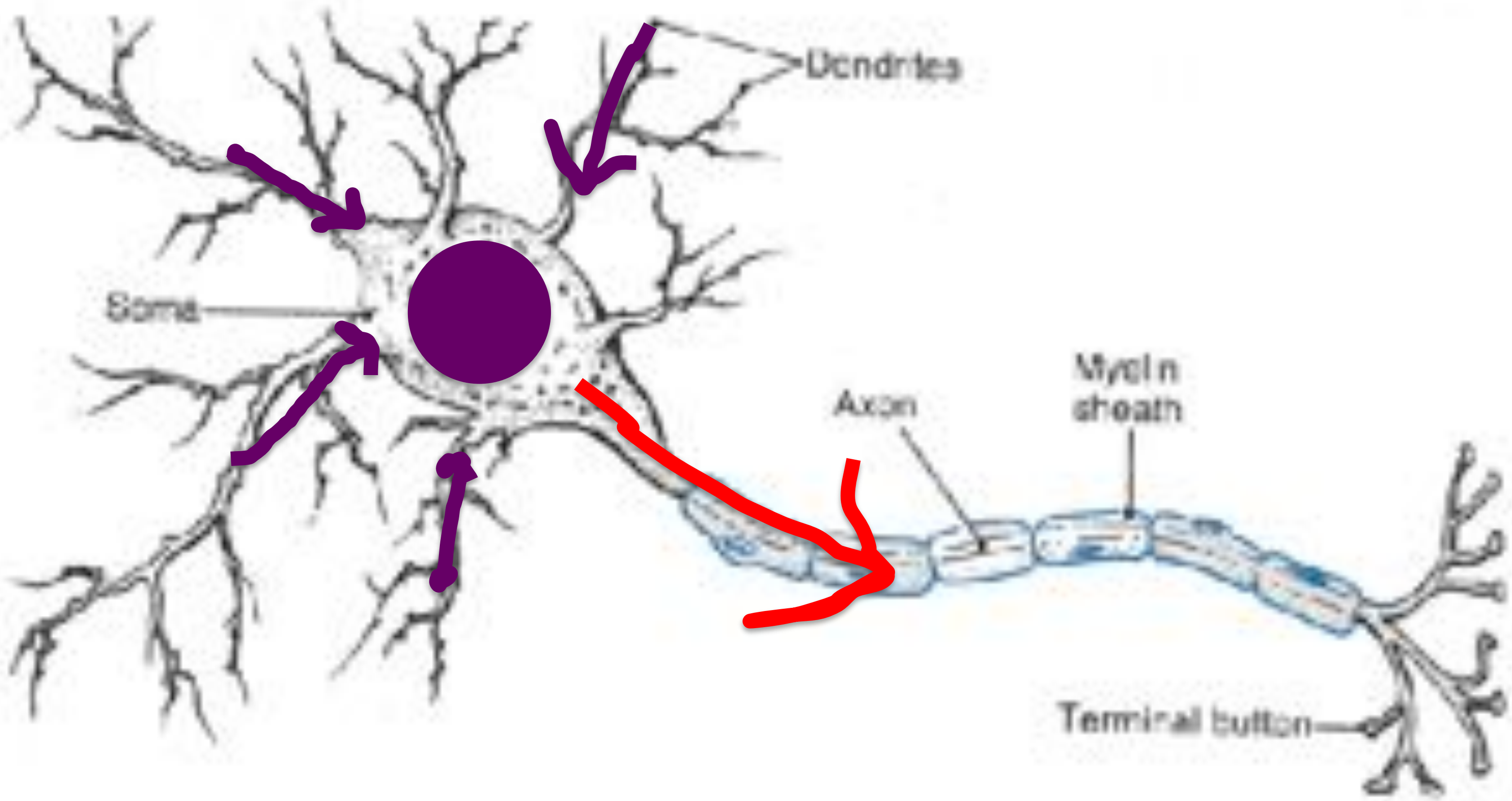
Human Neuron



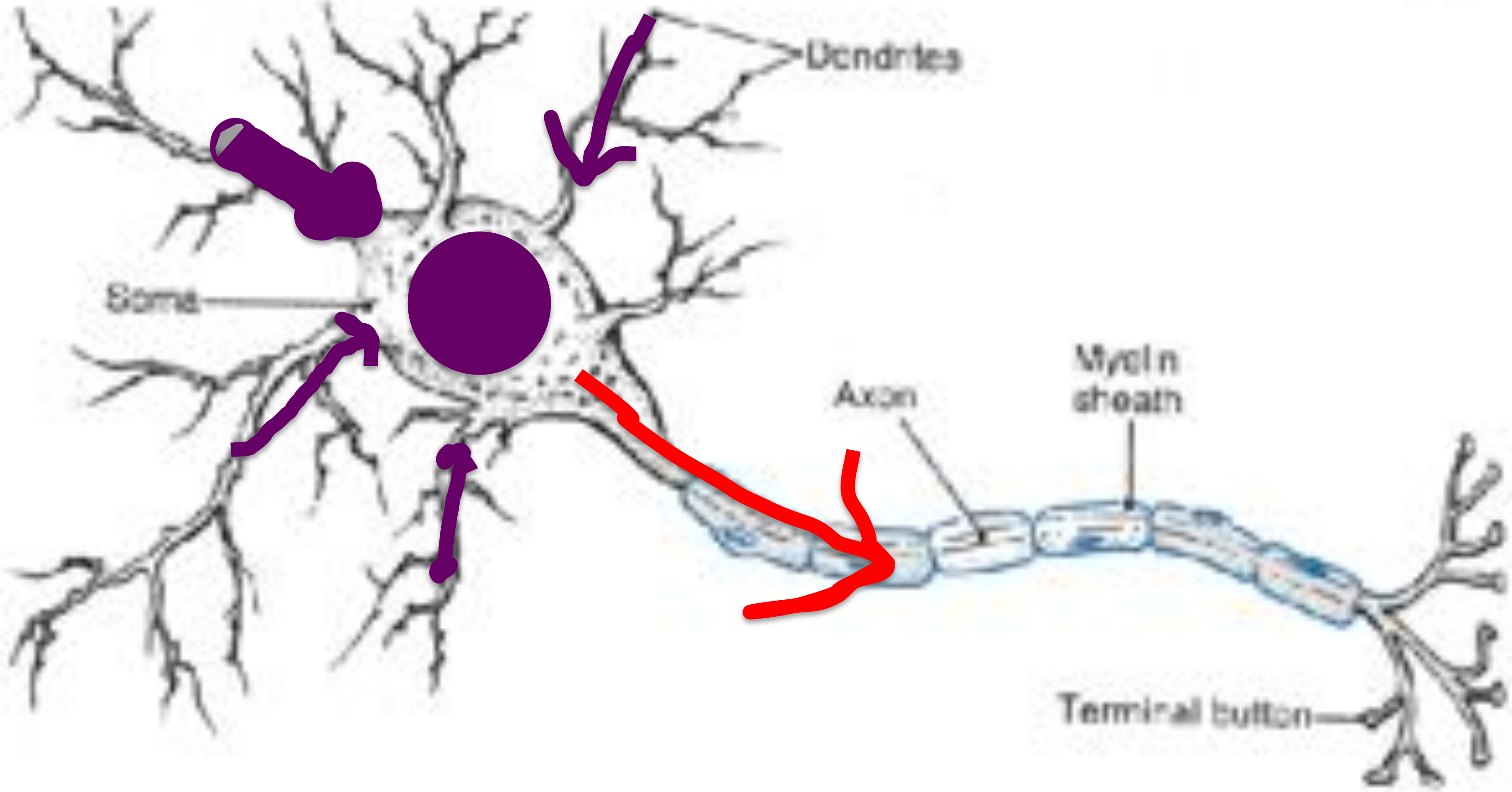
Human Neuron



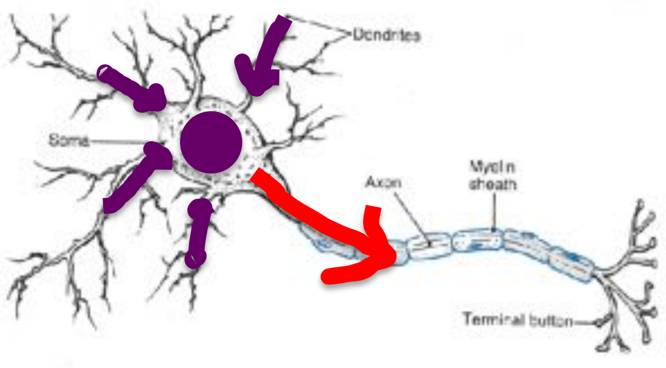
Human Neuron

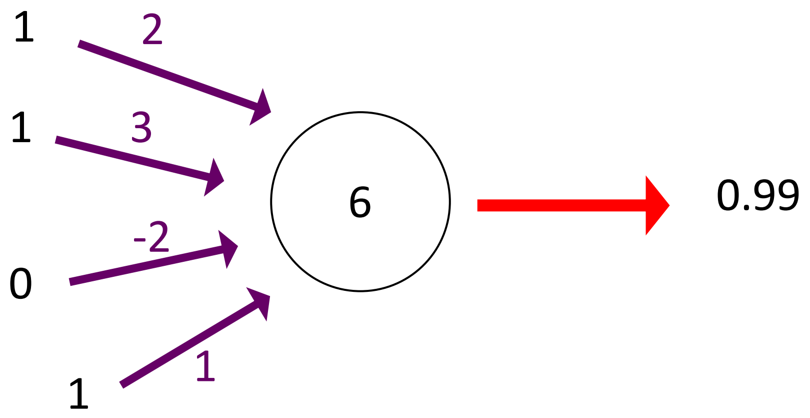


Human Neuron



Artificial Neuron

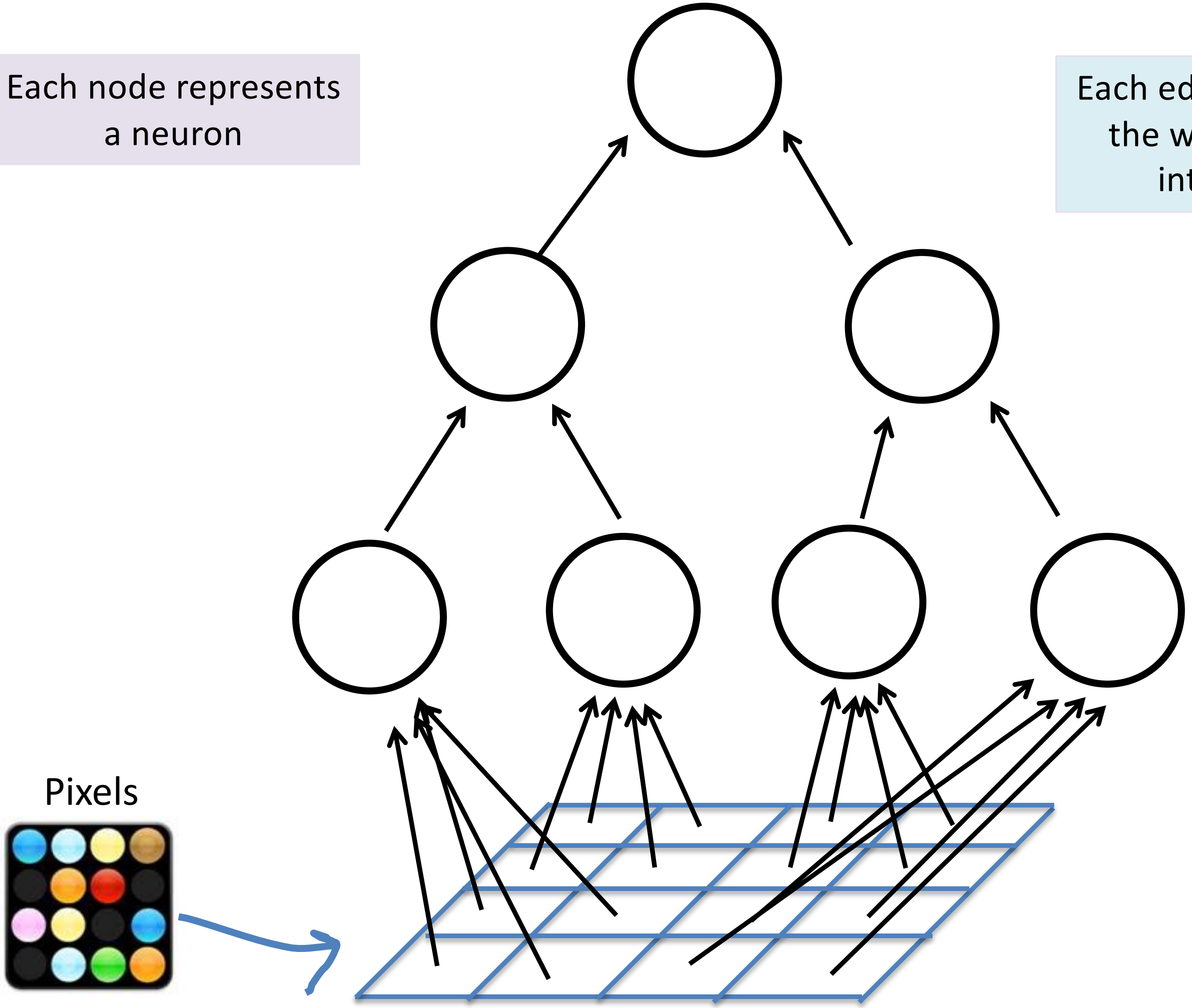




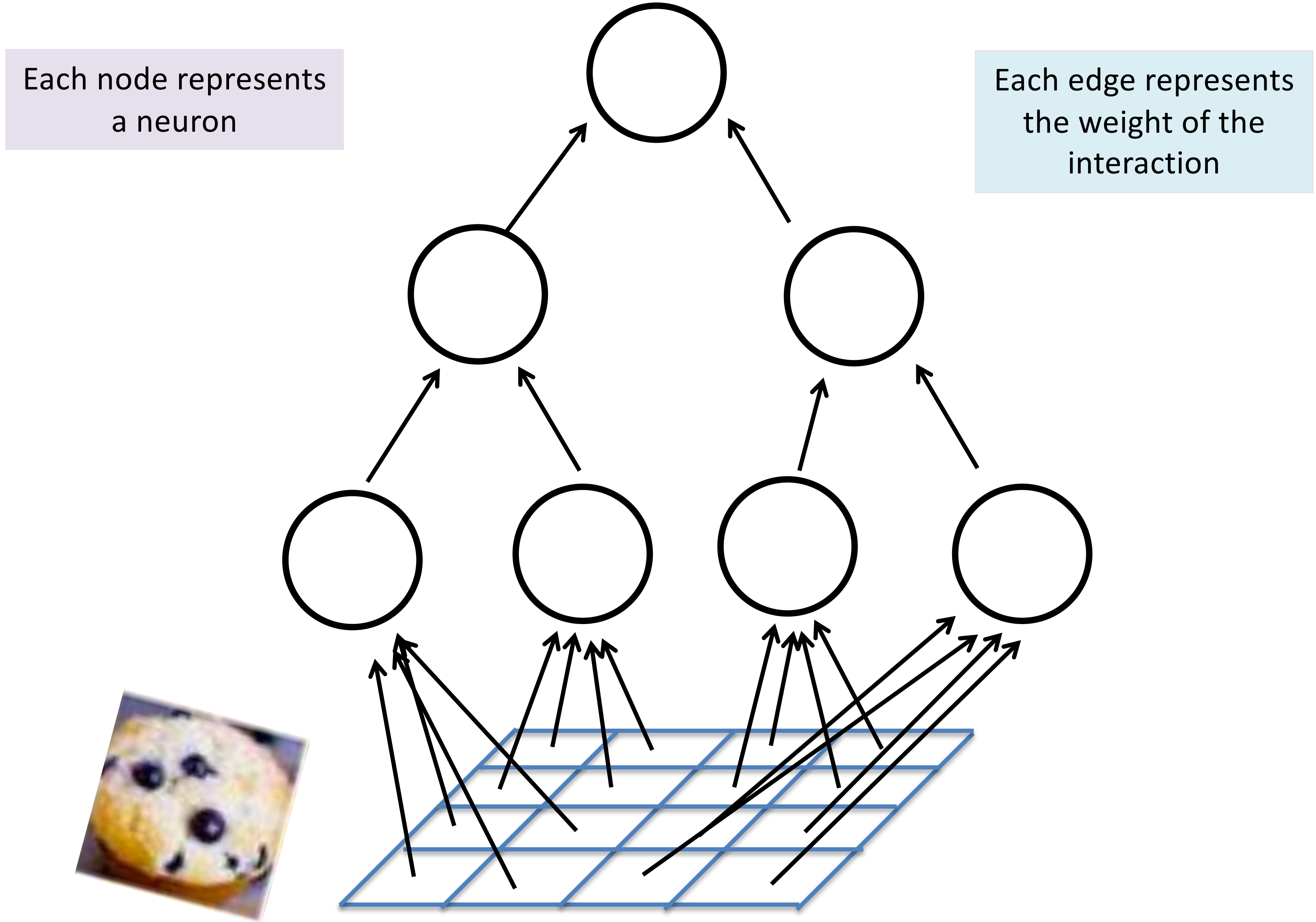
Artificial Neural Network

Each node represents a neuron

Each edge represents the weight of the interaction



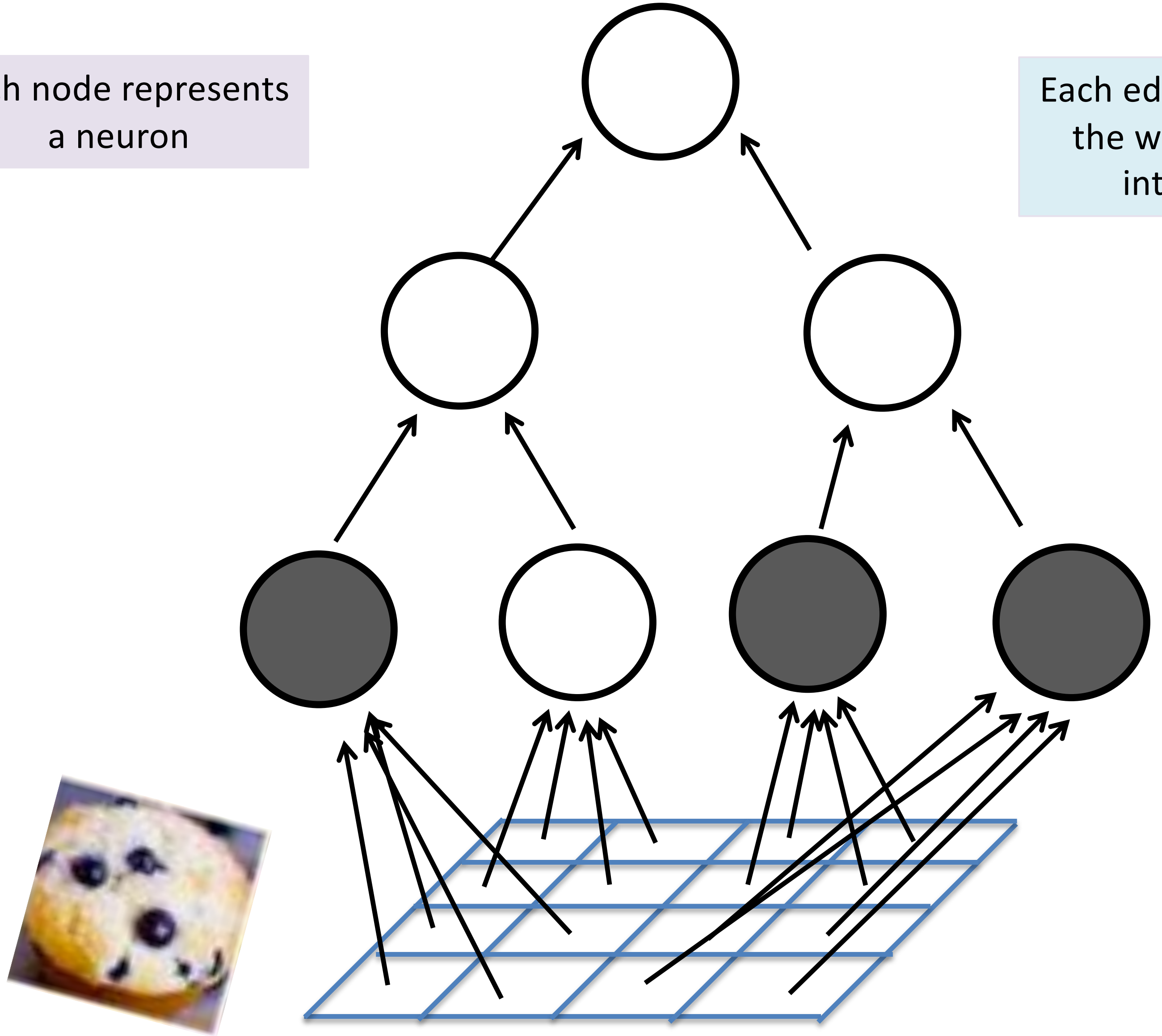
Artificial Neural Network



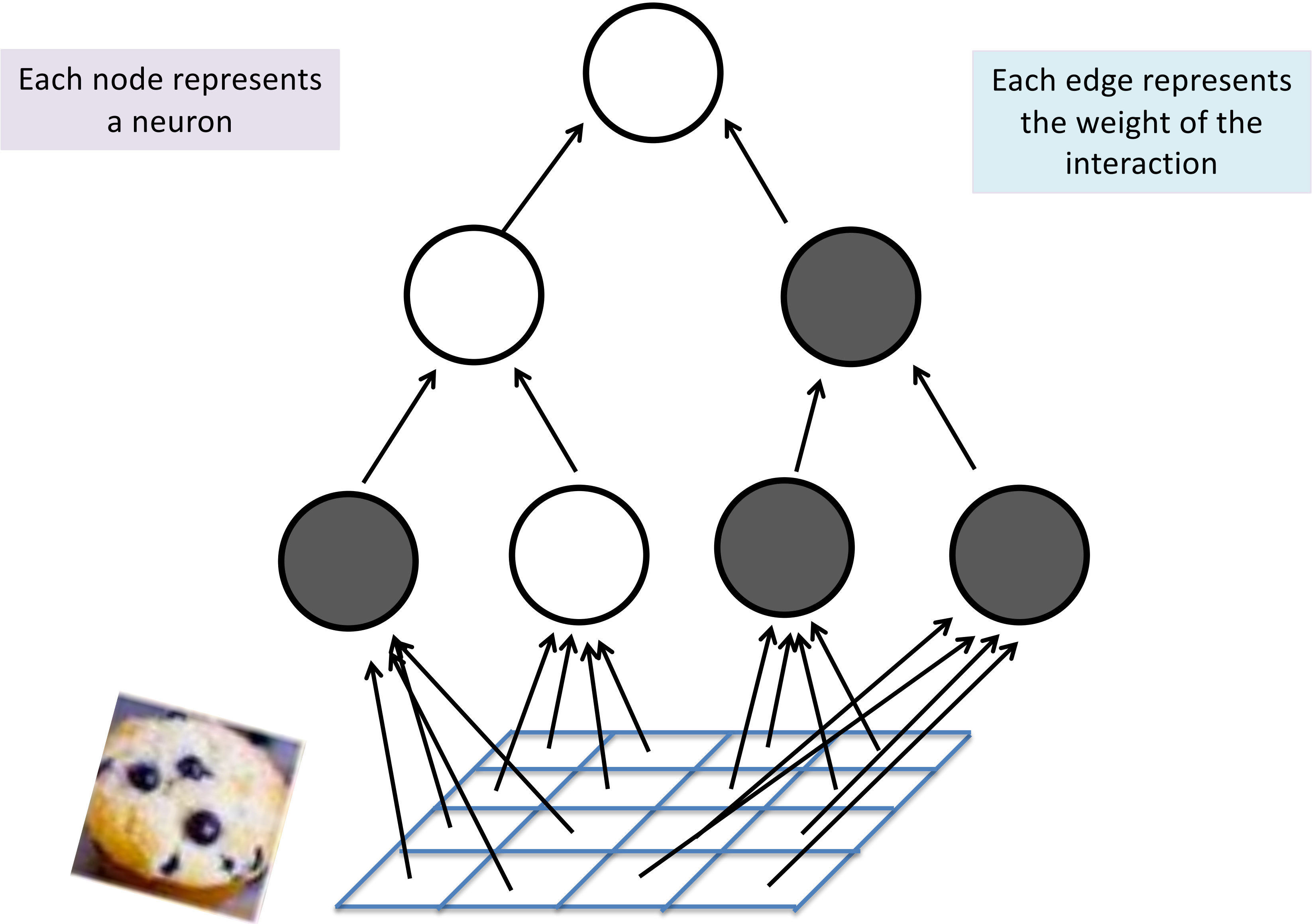
Artificial Neural Network

Each node represents a neuron

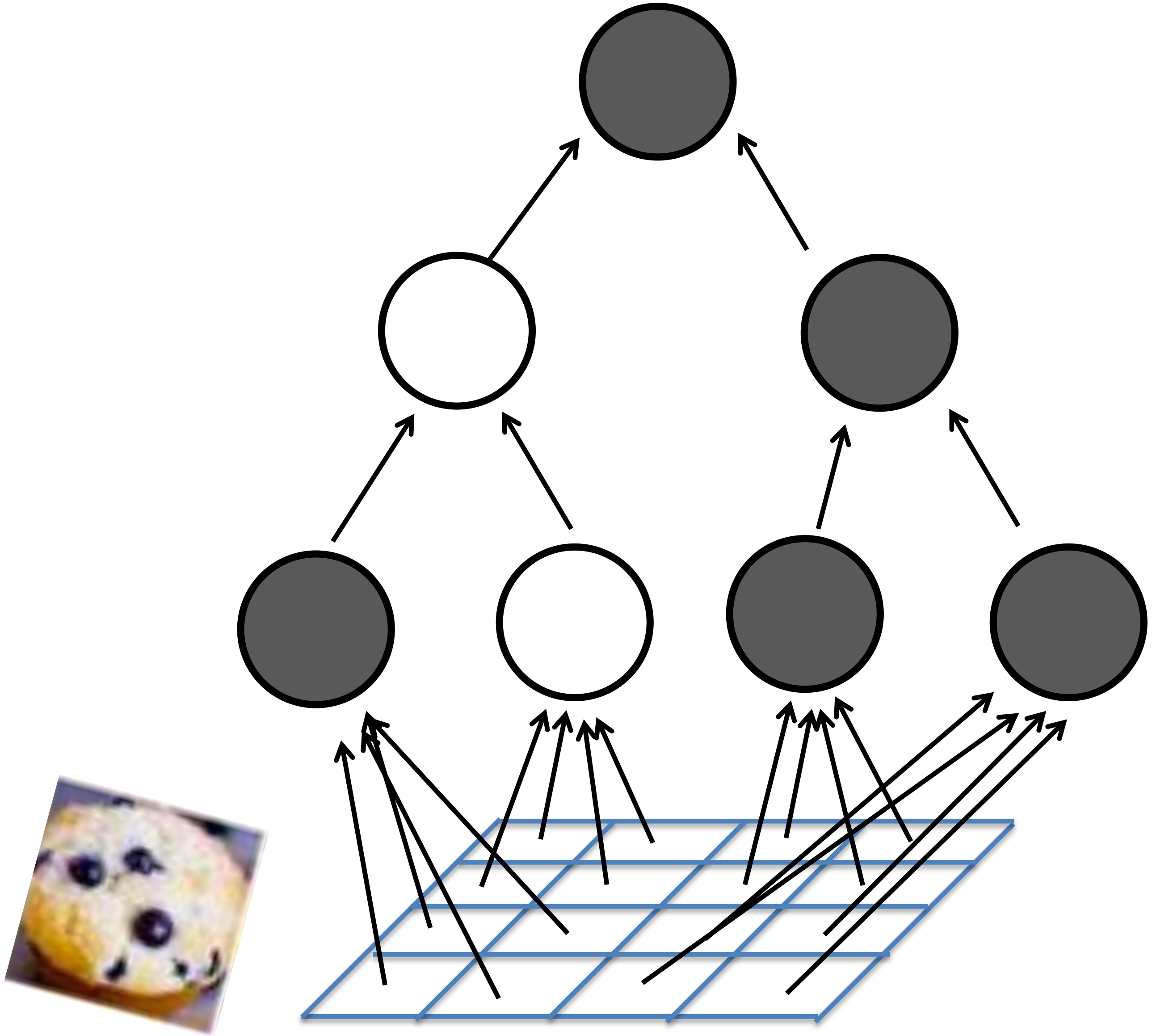
Each edge represents the weight of the interaction



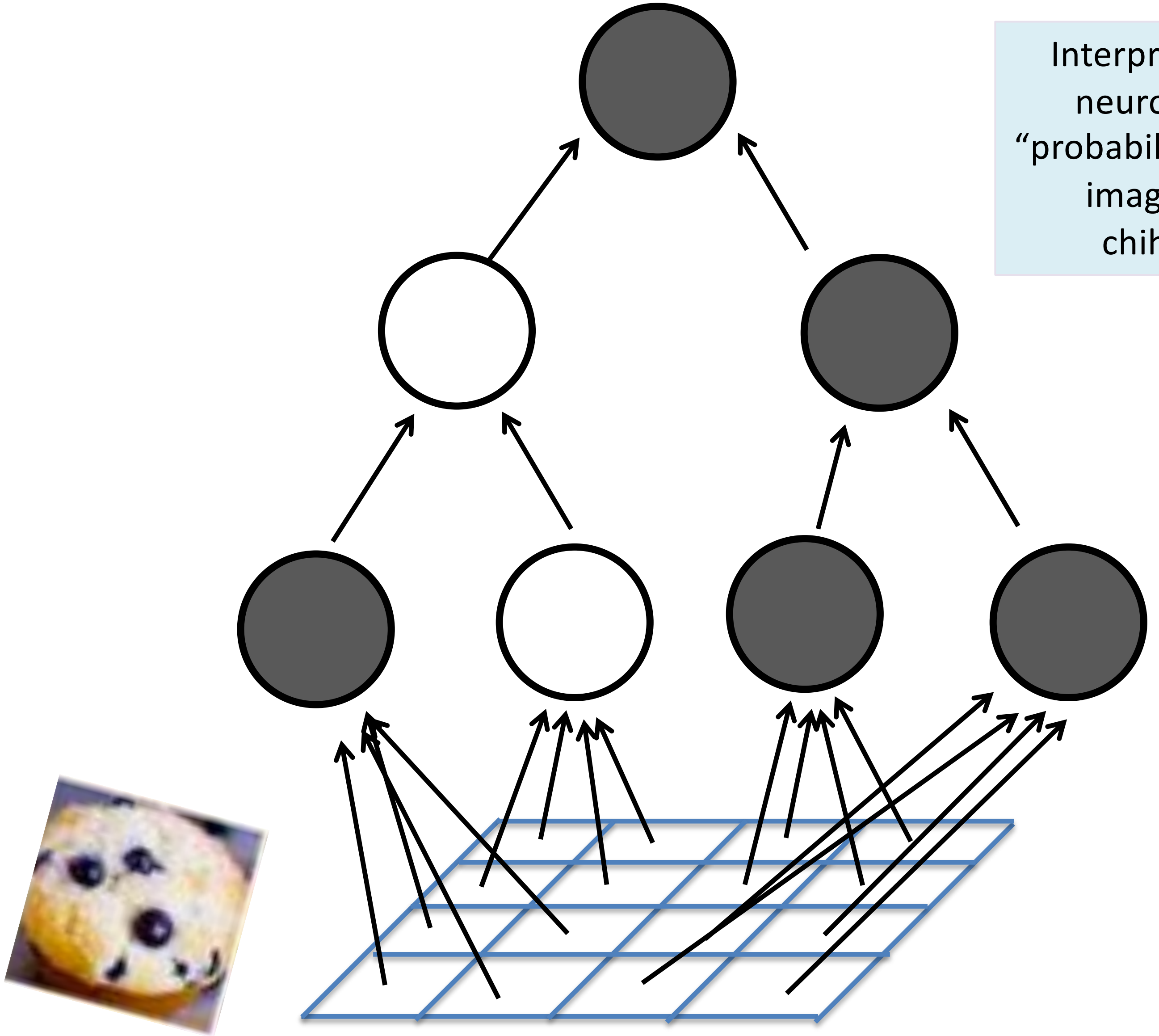
Artificial Neural Network



Artificial Neural Network



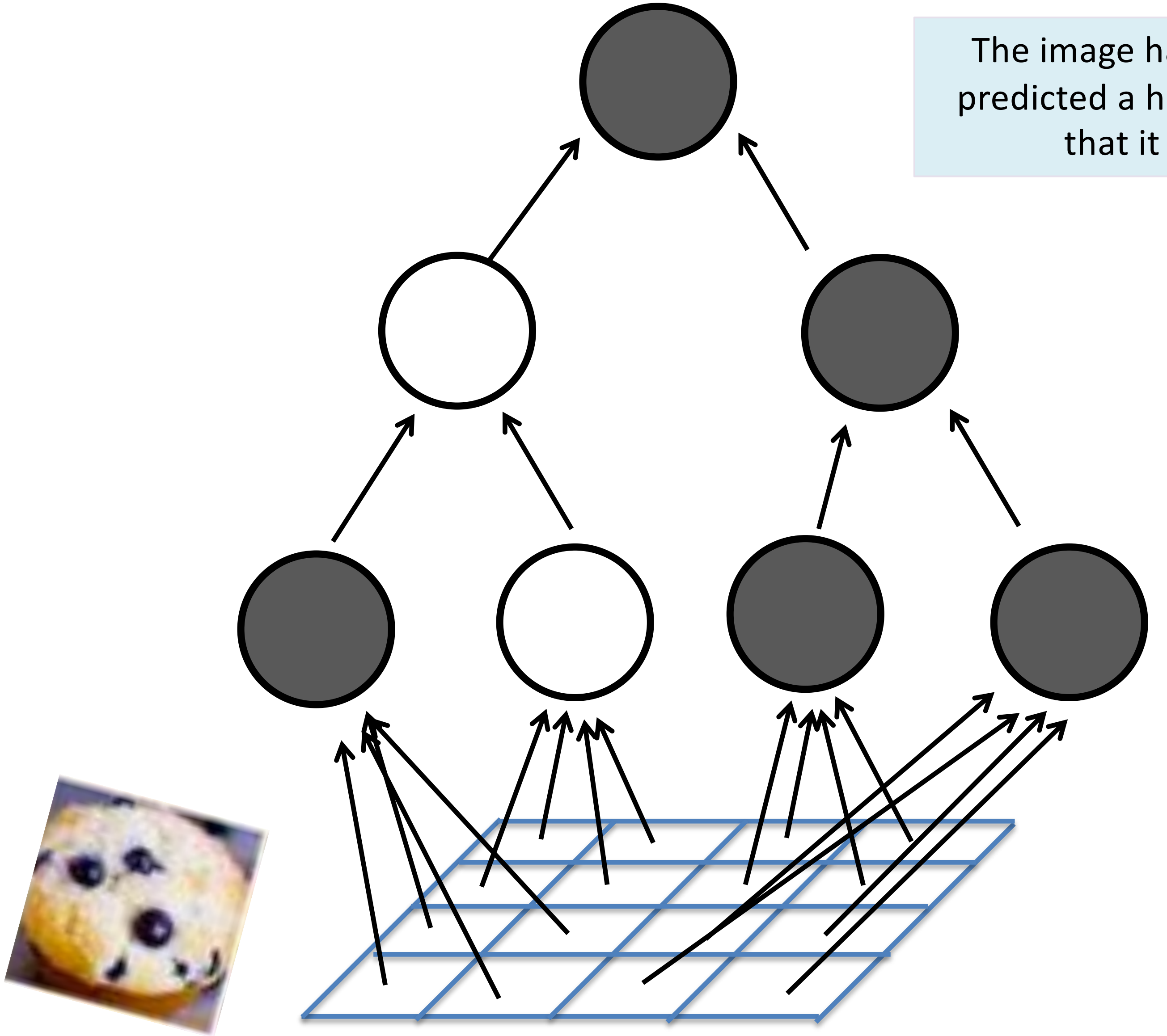
Artificial Neural Network



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

Artificial Neural Network

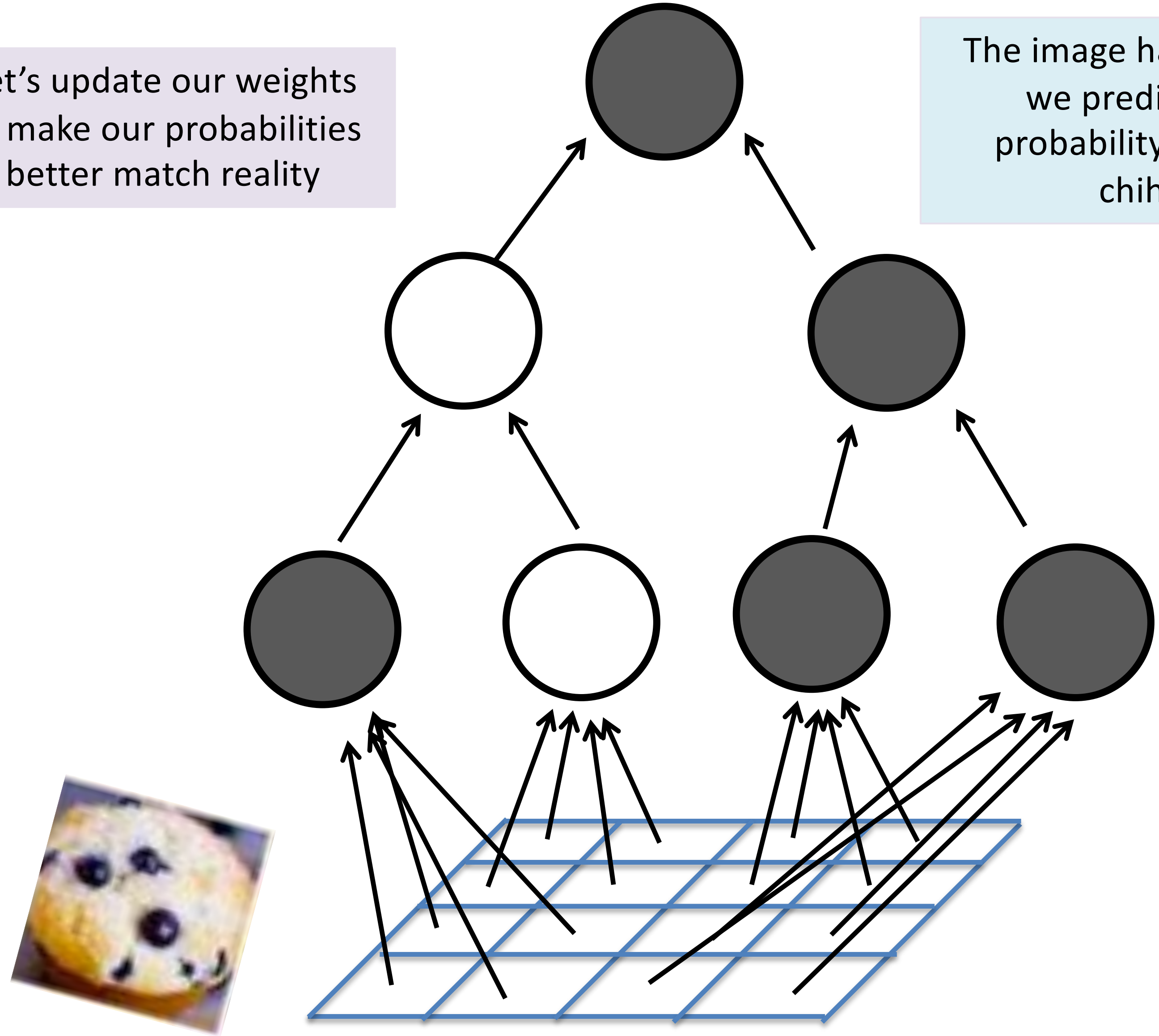
The image had a 0 but we predicted a high probability that it was a 1



Artificial Neural Network

Let's update our weights to make our probabilities better match reality

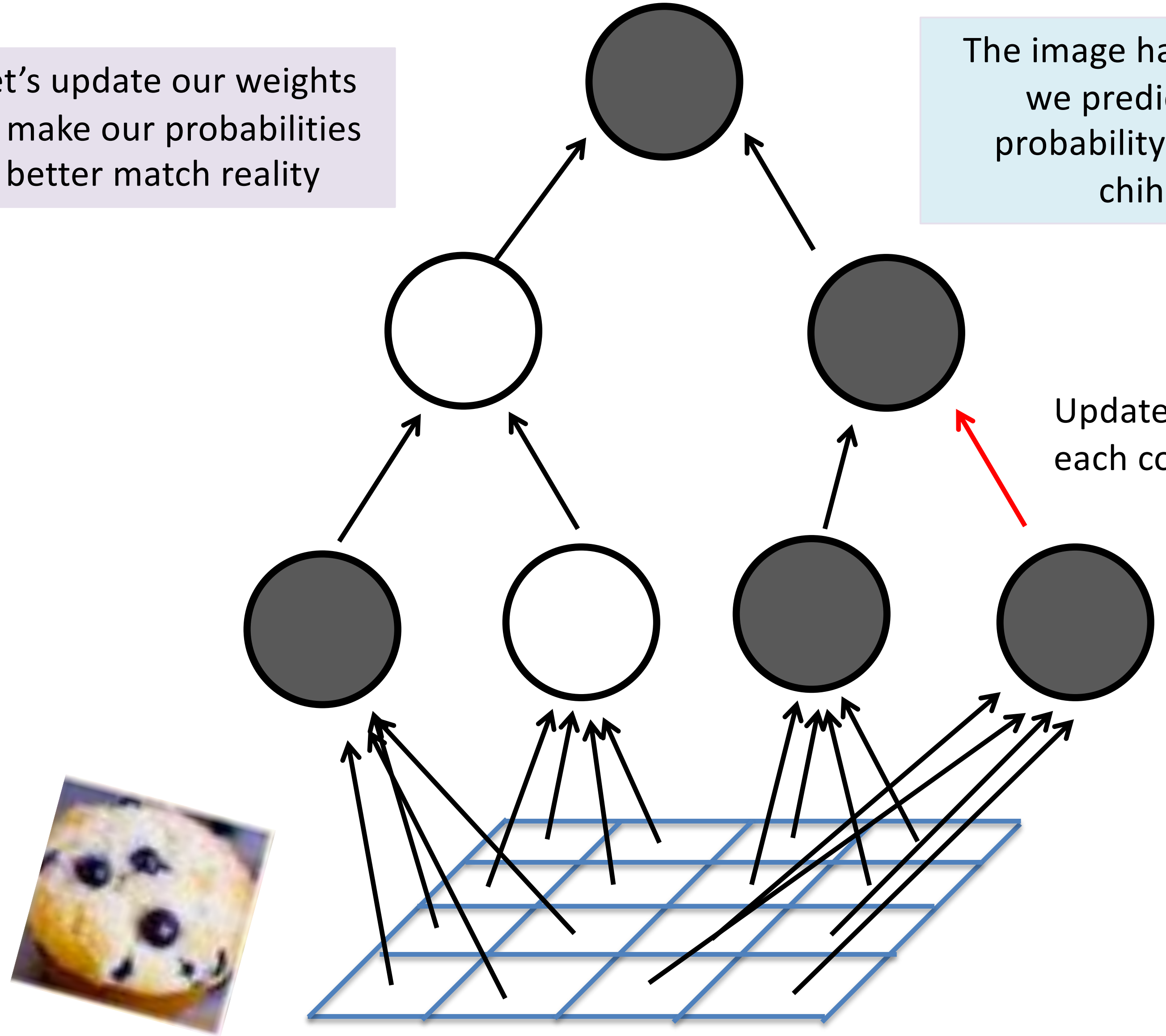
The image had a muffin but we predicted a high probability that it was a chihuahua



Artificial Neural Network

Let's update our weights to make our probabilities better match reality

The image had a muffin but we predicted a high probability that it was a chihuahua



Update the parameter at each connection

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

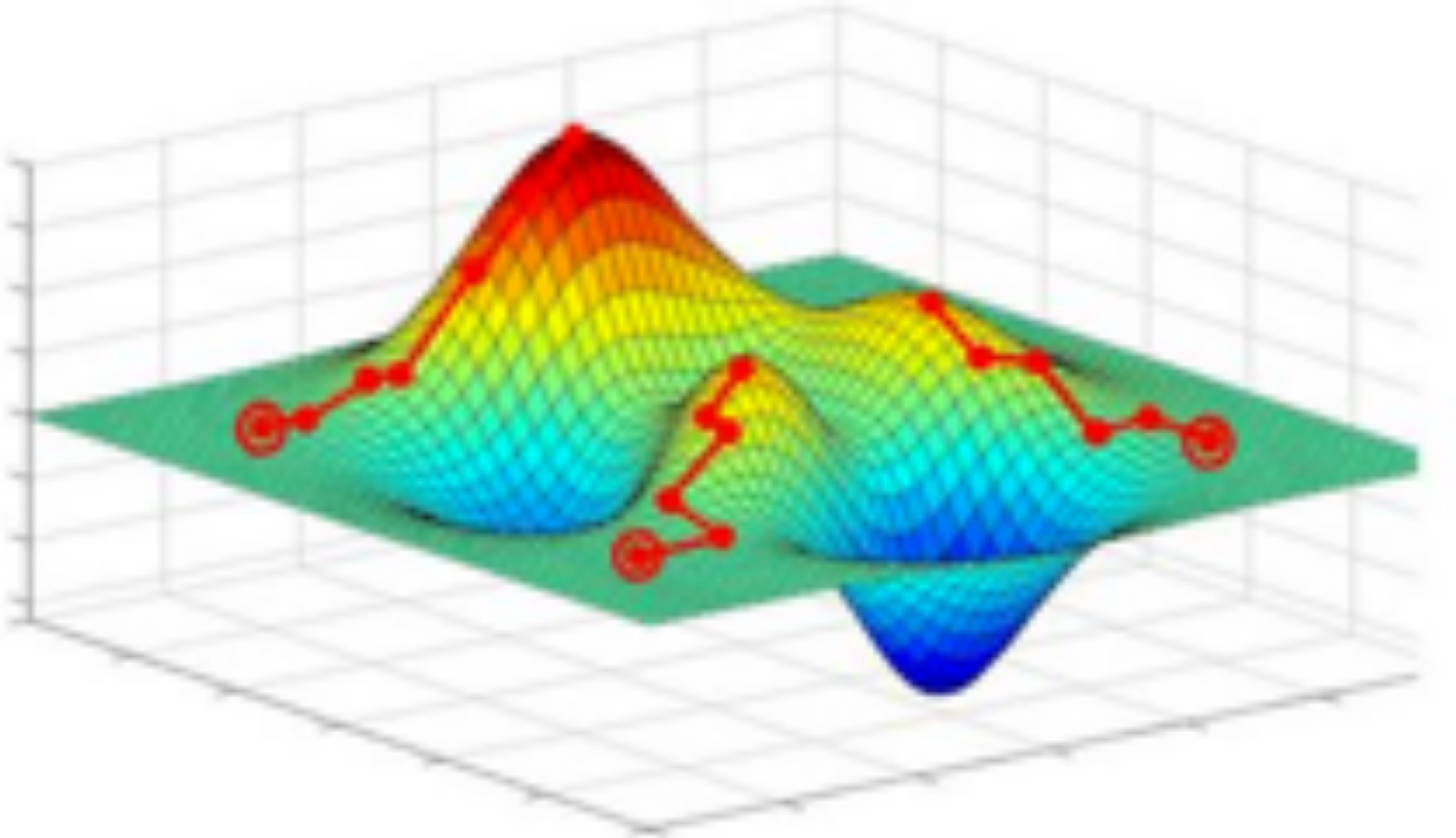
$$P(Y = y | X = \mathbf{X}) = (\hat{y})^y (1 - \hat{y})^{1-y}$$

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

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

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

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

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

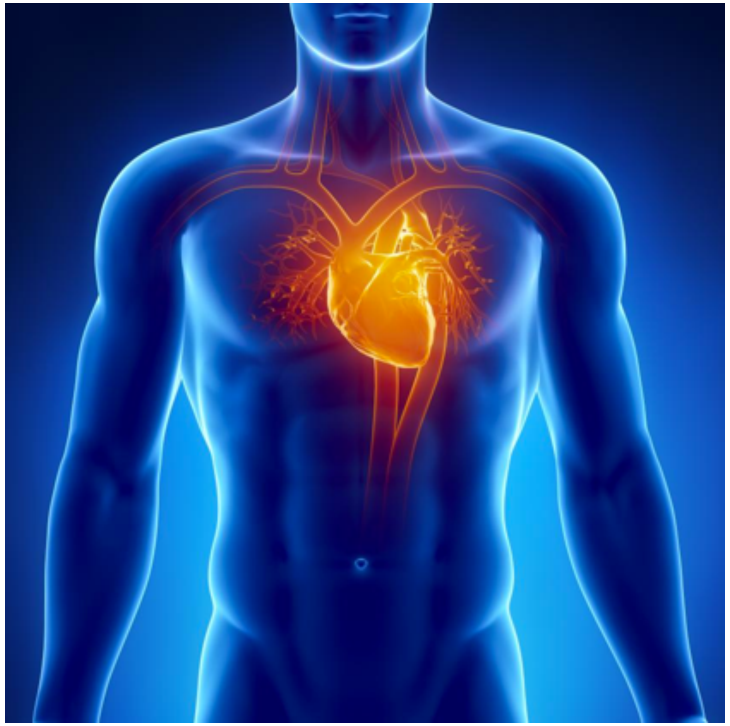
$$= \hat{y}[1-\hat{y}] \cdot h_i$$

You will be able to do this.

Where you will be by the end of class

CS109: Theory Class focused on Applications

Heart



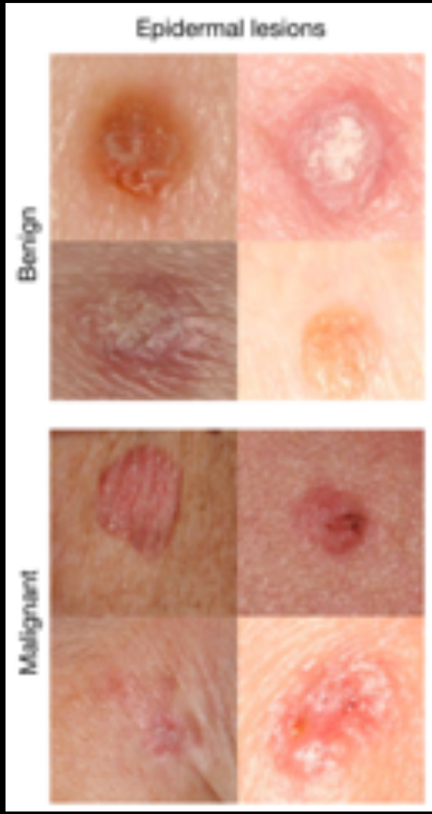
Ancestry

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.

AI Augmented Education?

50 thousand students

Knowledge Tracing for Feedback

Given *n* historical answers:

x_1 x_2 … x_n

x_{n+1}

Predict the next one

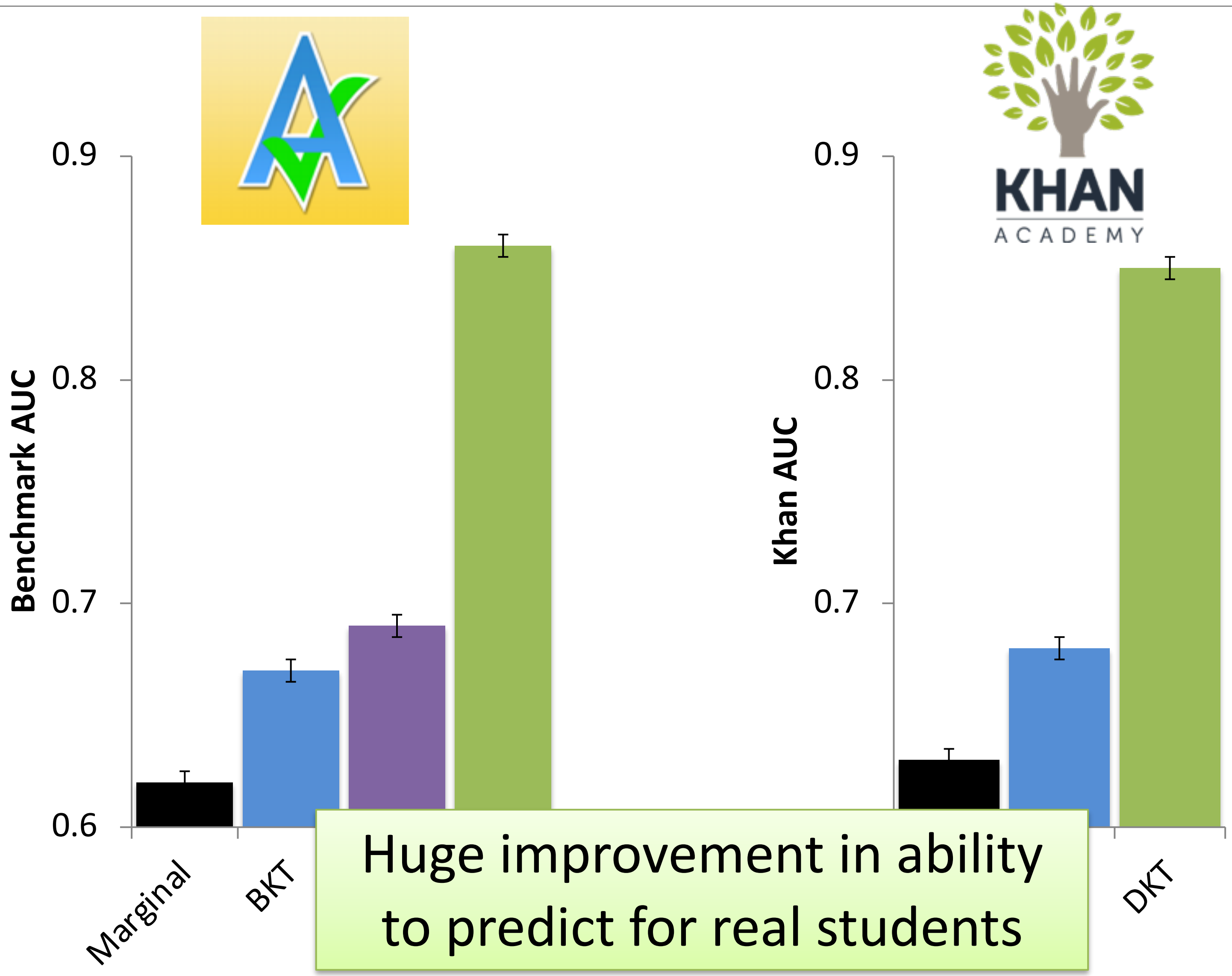
Answer is a tuple:

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

Question

Student response

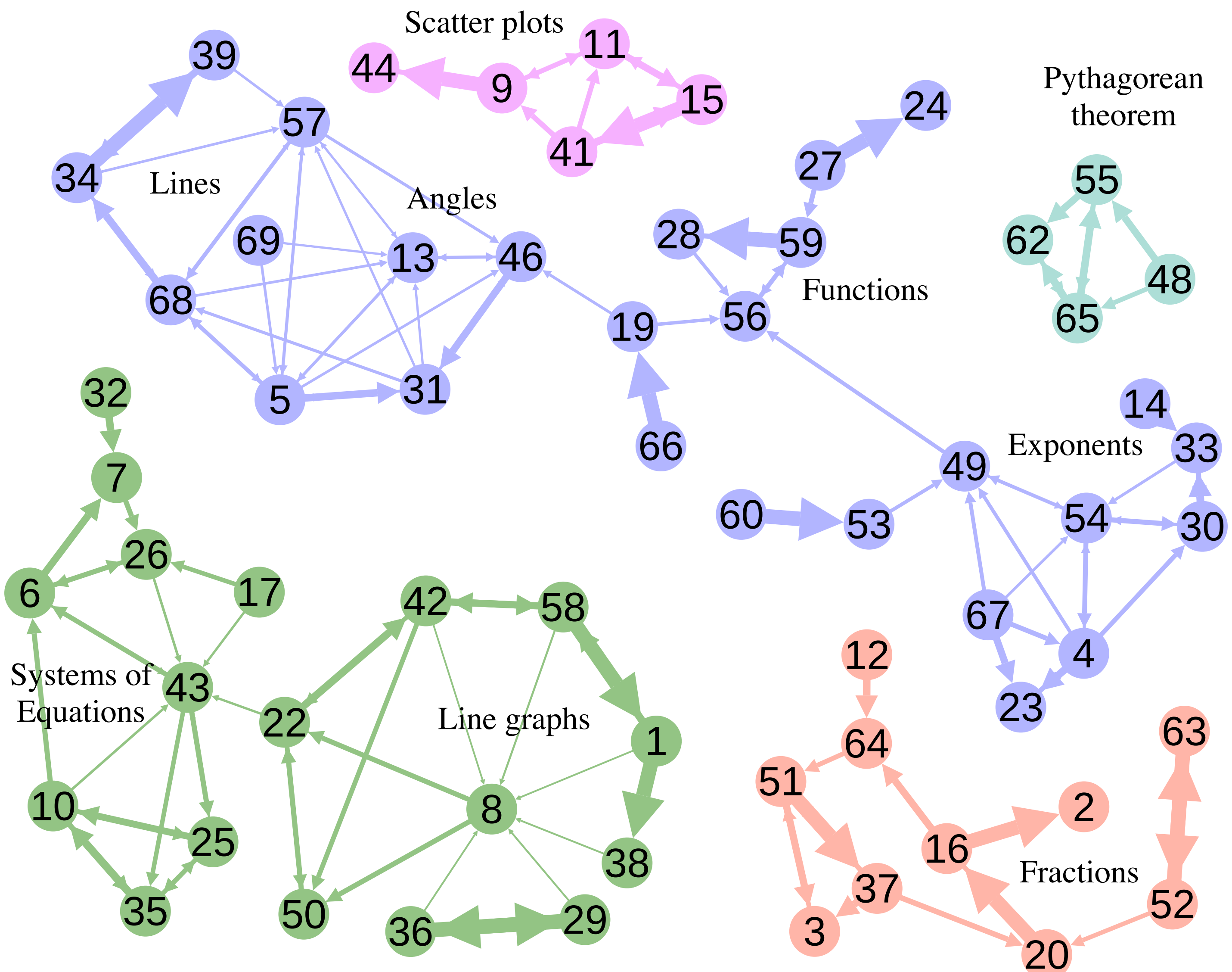
Prediction Results



Huge improvement in ability to predict for real students

Piech et al, 2015

Learns Concept Relations



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

End of Story

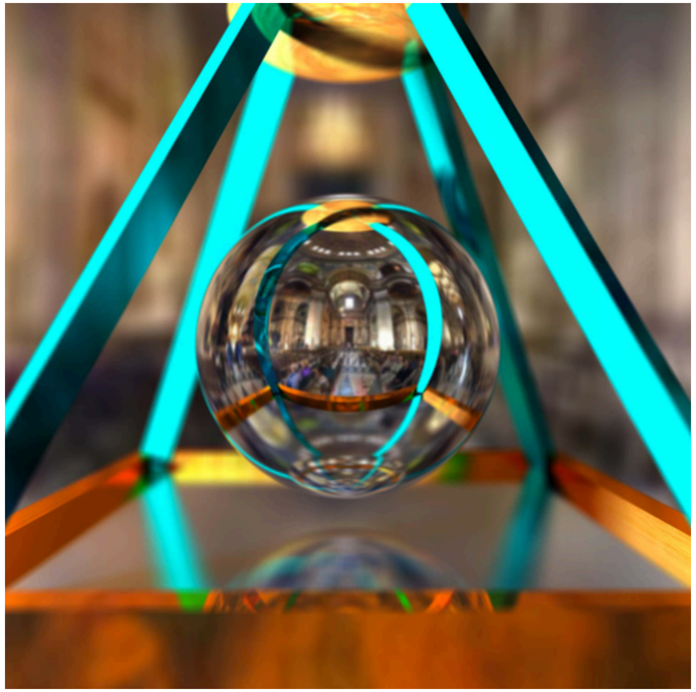
Except it isn't the end of the story…

Probability is more than just machine learning

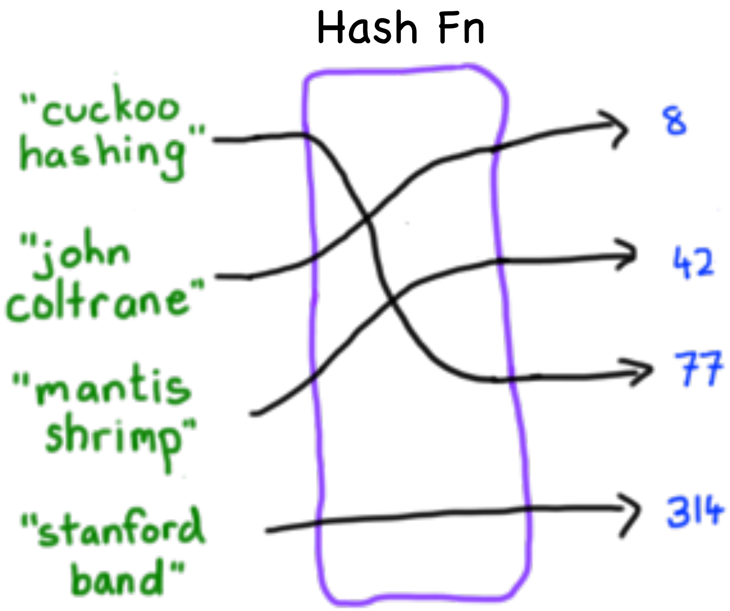
Abundance of important problems

Algorithms and Probability

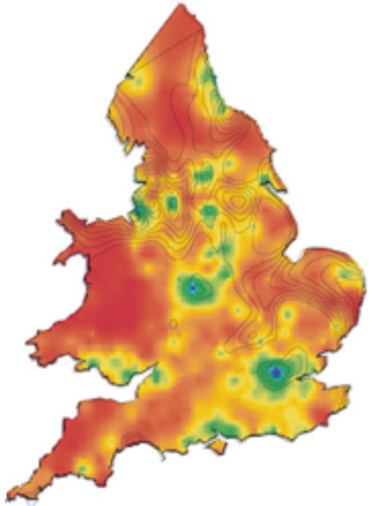
Eg Raytracing



Eg HashMaps



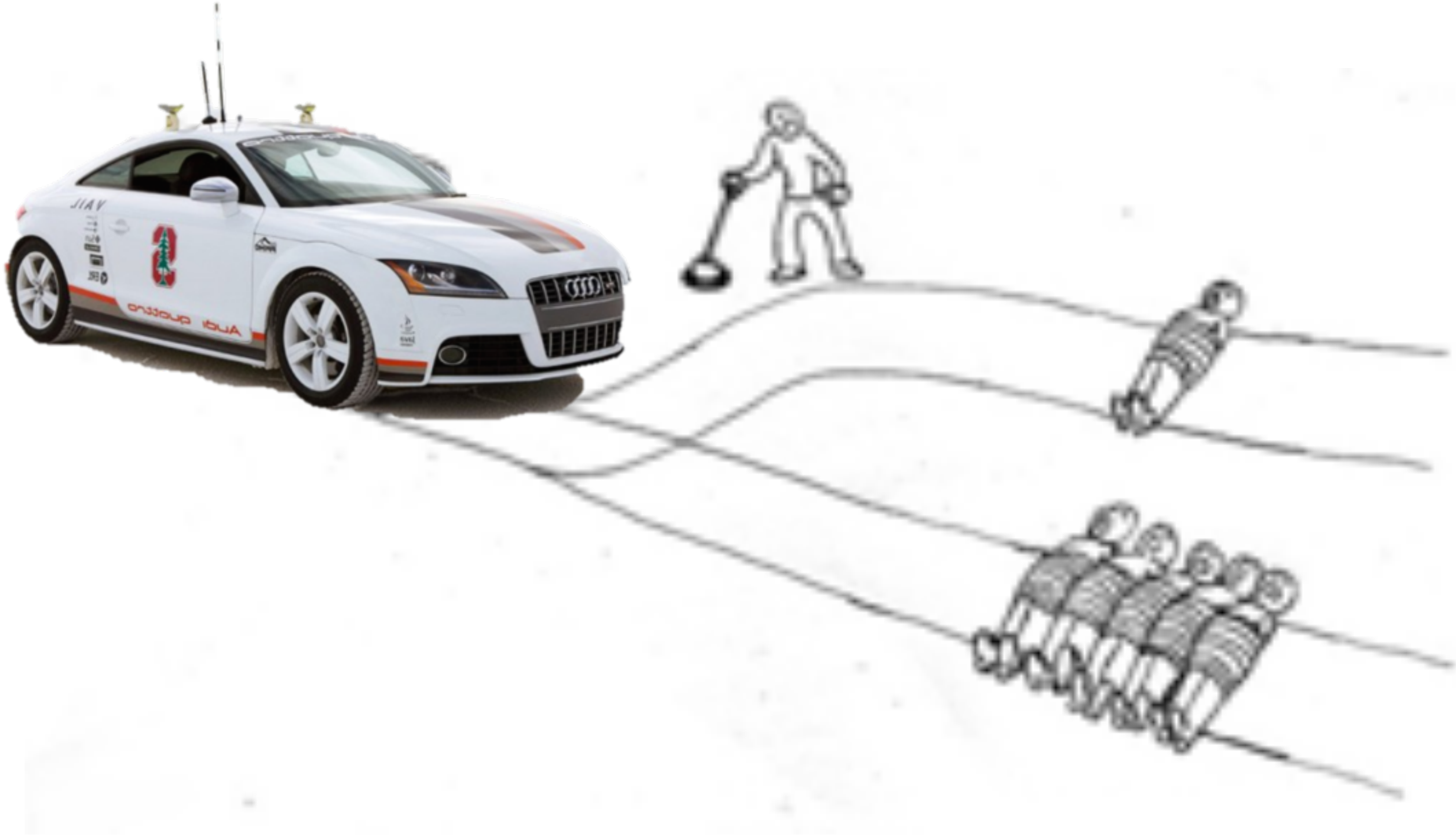
The next medical revolution?



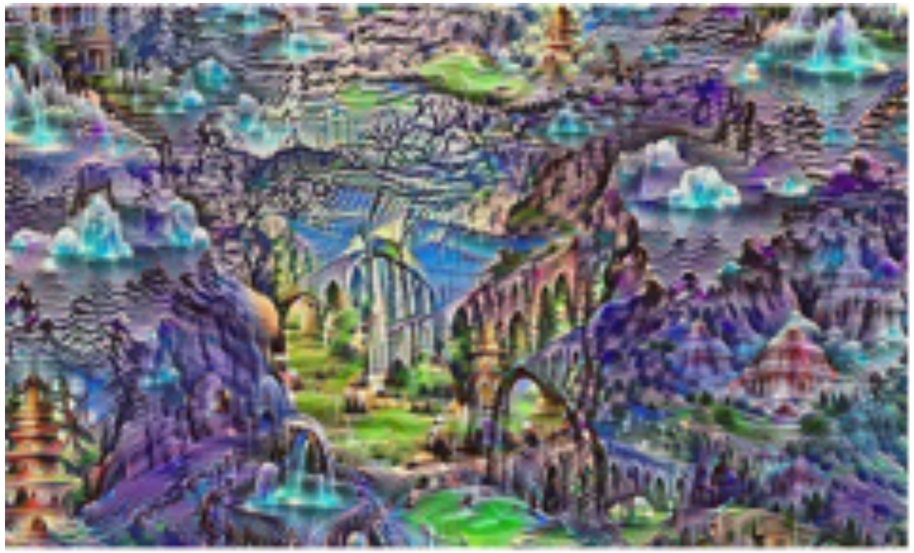
Autocomplete

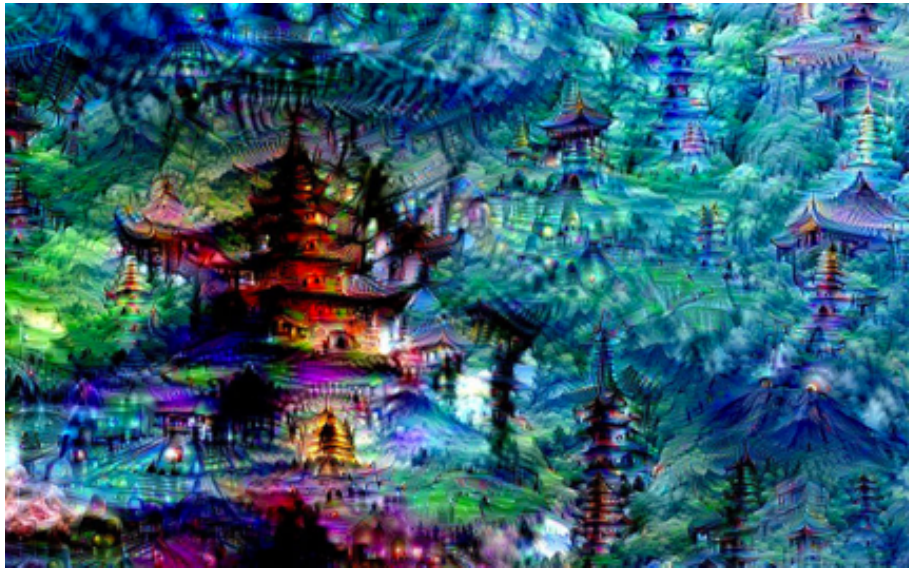
Recommender Systems

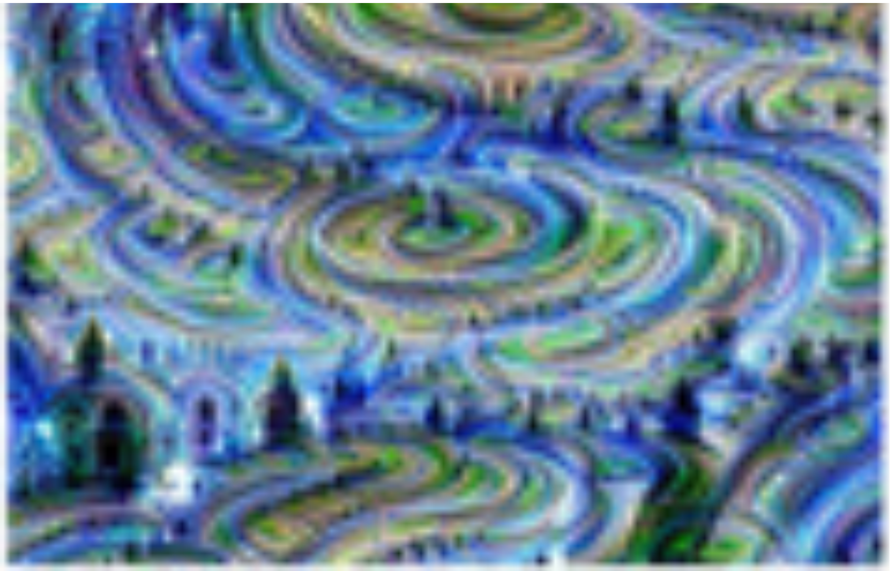
Philosophy and Ethics



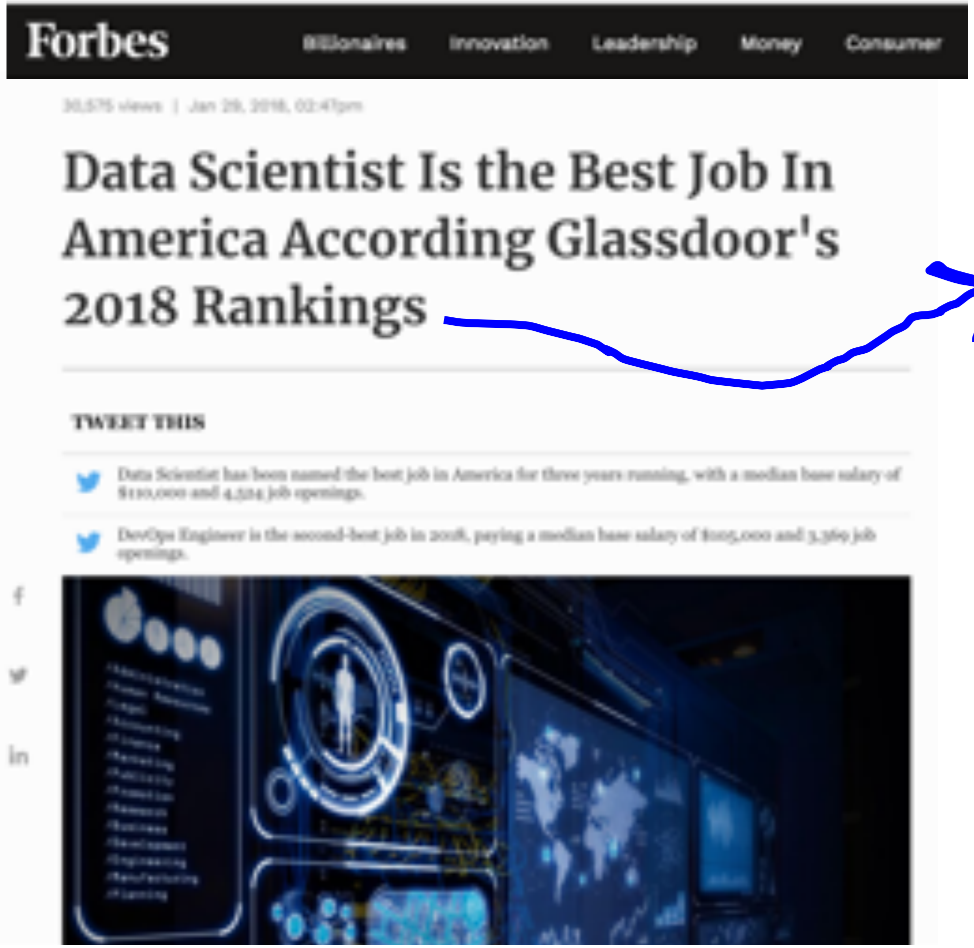
Art and Probability







Most Desired Skill in Industry



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"

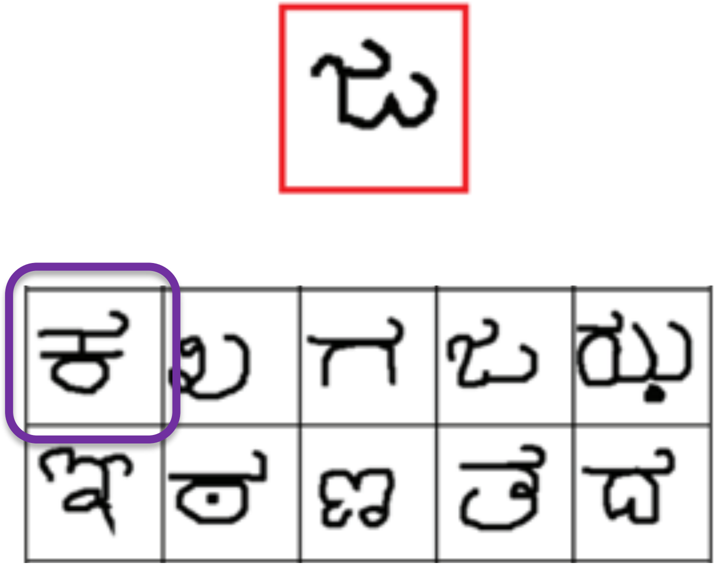
Open Problem: One Shot Learning

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

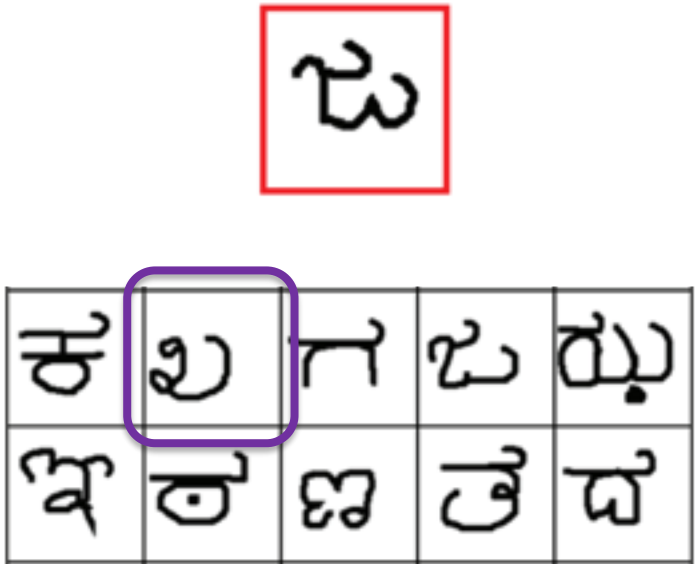
B Lake, R Salakhutdinov, J Tenenbaum. Science 2015.
Human-level concept learning through probabilistic program induction.



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

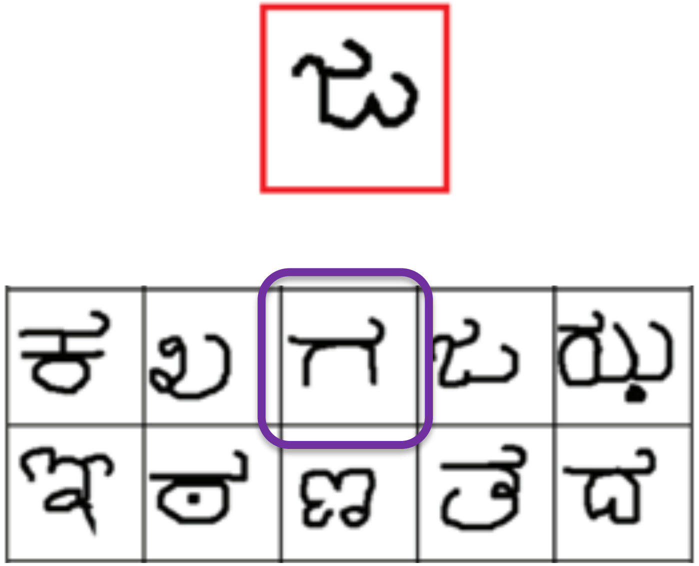
B Lake, R Salakhutdinov, J Tenenbaum. Science 2015.
Human-level concept learning through probabilistic program induction.



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

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

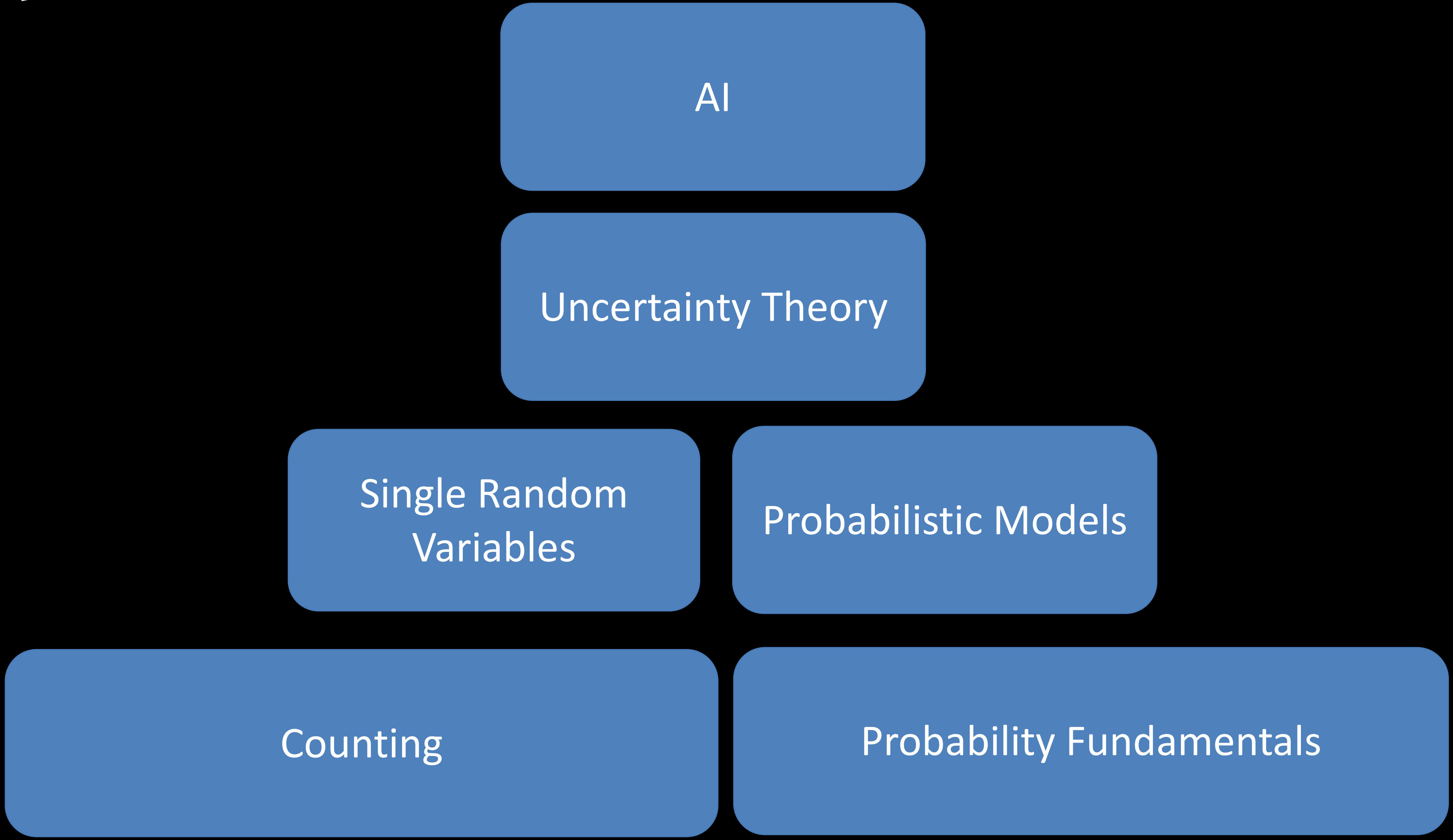
Delayed gratification

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.


AI
Uncertainty Theory
Single Random Variables
Probabilistic Models
Counting
Probability Fundamentals

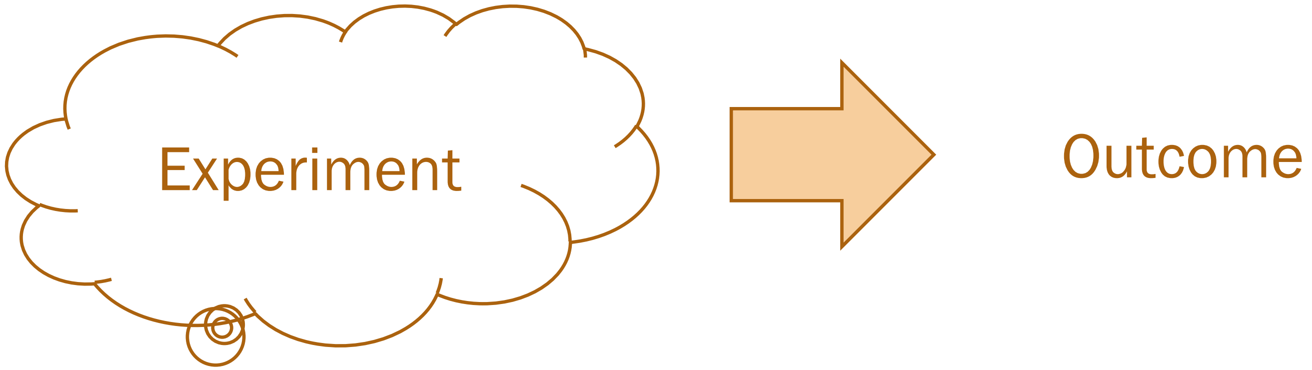
Lets dive in…

2 min pedagogic pause.

Counting I

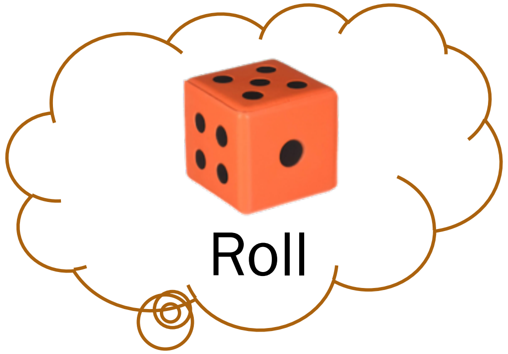
What is Counting?

An experiment in probability:



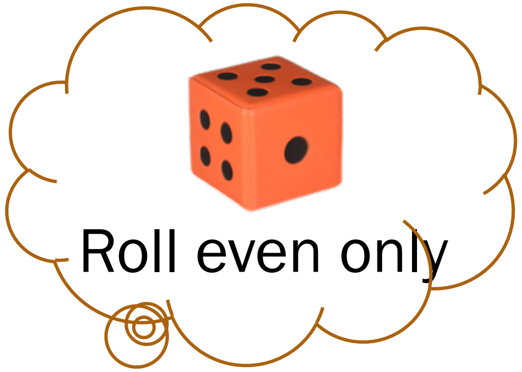
Counting: How many possible outcomes can occur from performing this experiment?

What is Counting?



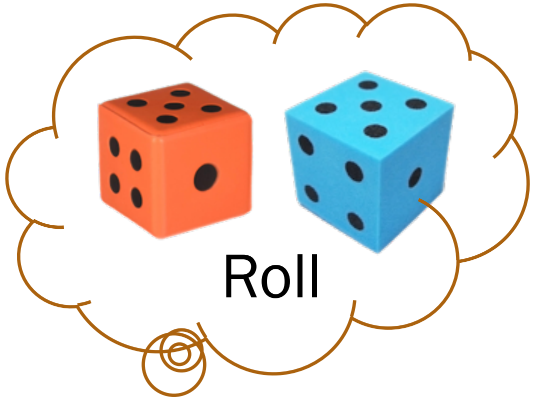
6

{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 $|A| = m$,
and 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

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

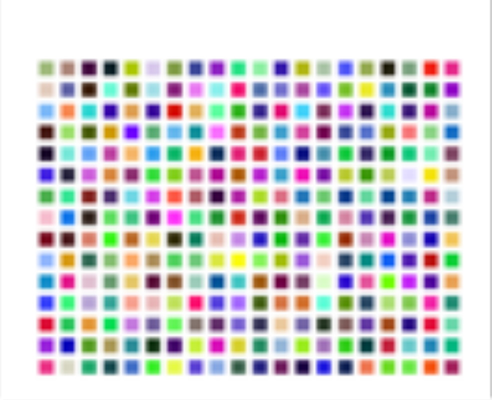
Two-step experiment

$\rightarrow A \rightarrow B$

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 $|A| = m$,

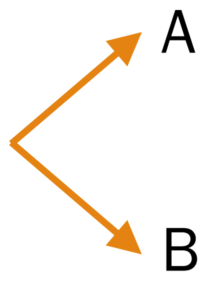
or Set B, where $|B| = n$,

where $A \cap B = \emptyset$,

Then the number of outcomes of the experiment is

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

One experiment

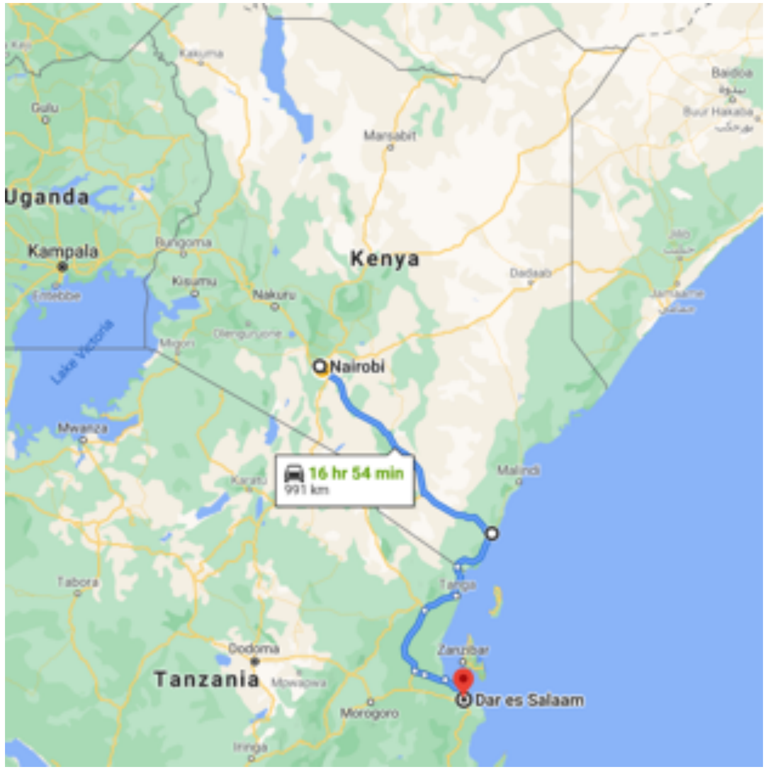


How many routes

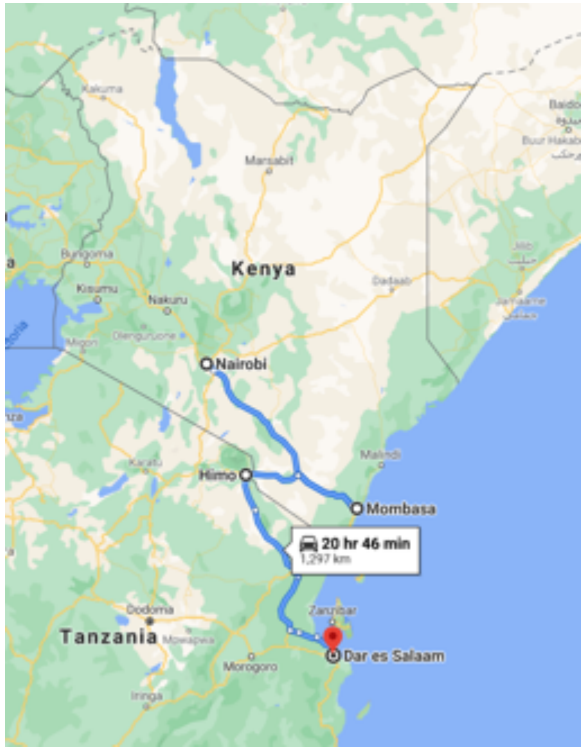
Question: All routes considered by google maps from Nairobi to Dar es Salaam go through either Mt Kilimanjaro **or** Mombasa. How many total routes are considered?

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

```
010000
010001
010010
010011
010100
010101
010110
010111
011000
011001
011010
011011
011100
011101
011110
011111
```

Set A

2^4 end with 10

```
000010
000110
001010
001110
010010
010110
011010
011110
100010
100110
101010
101110
110010
110110
111010
111110
```

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 $|A| = m$,

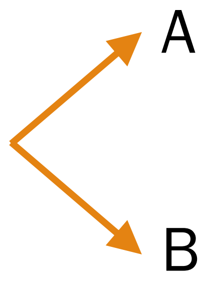
<u>or</u> Set B, where $|B| = n$,

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.