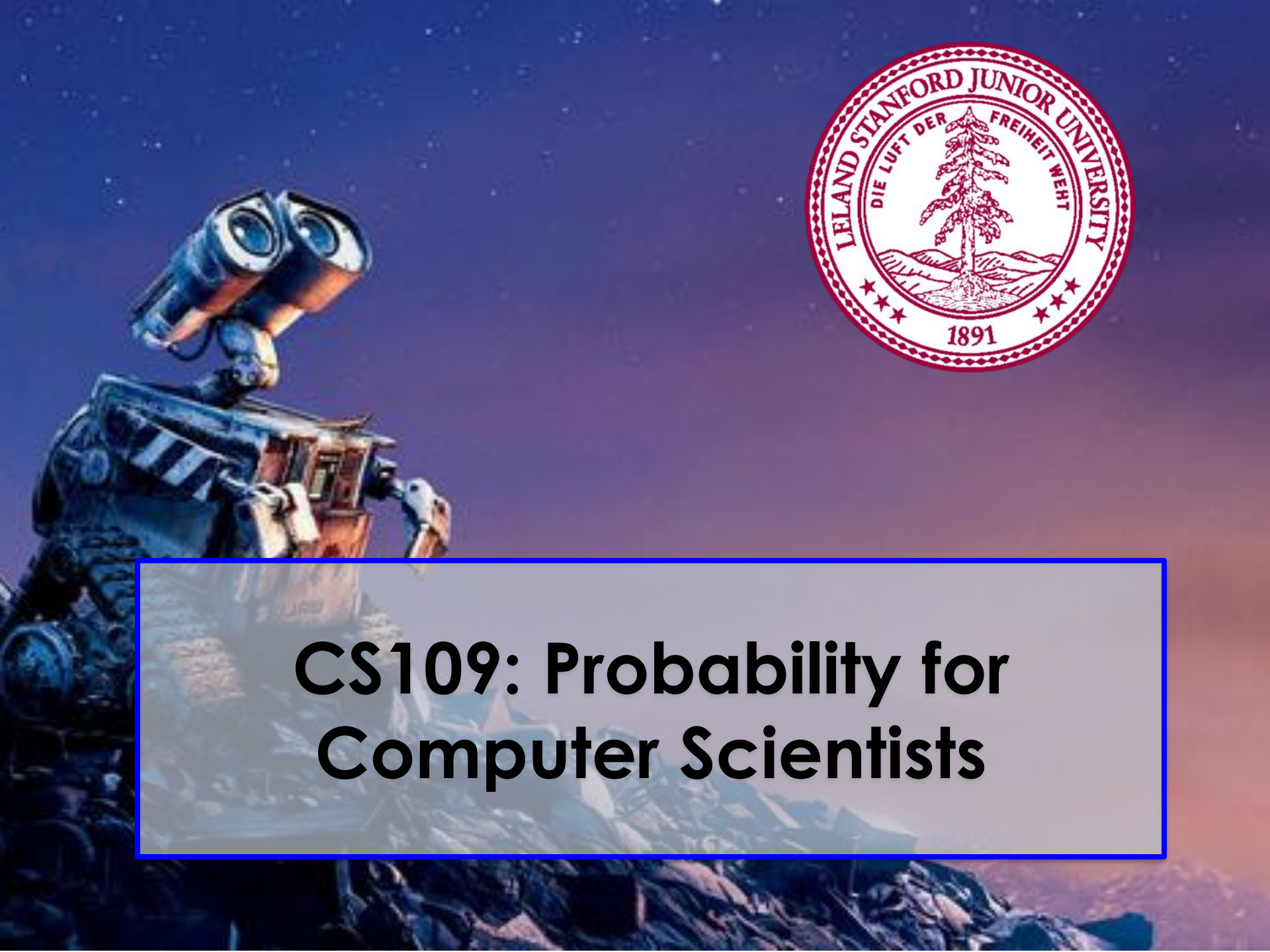


CS109: Probability for Computer Scientists

Chris Piech

My parents are interesting folks

I originally concentrated in graphics and worked at Pixar

- Childhood: Nairobi, Kenya
- High School: Kuala Lumpur, Malaysia
- Stanford University Ph.D. in Deep Learning
- Research lab on AI for Social Good

The problem I really want to solve is to make high quality more education accessible

I Took the First CS109 Class

Back when I looked like this ☺

Teaching Team

Professor: Chris Piech
piech@cs.stanford.edu
Thurs 1-3pm
Gates 202

TA: Julia Daniel
jdaniel7@stanford.edu
TBA
TBA

TA: Ana-Maria Istrate
aistrate@stanford.edu
TBD

TA: Annie Hu
anniehu@stanford.edu
TBD
TBD

TA: Conner Smith
csmith95@stanford.edu
TBD
TBD

TA: Cooper Raterink
crat@stanford.edu
TBD
TBD

TA: David Liang
dwliang@stanford.edu
TBD
TBD

TA: Nathan Dass
ndass@stanford.edu
TBD
TBD

TA: Nicholas Bowman
nbowman@stanford.edu
TBD
TBD

TA: Noah Arthurs
narthurs@stanford.edu
TBD
TBD

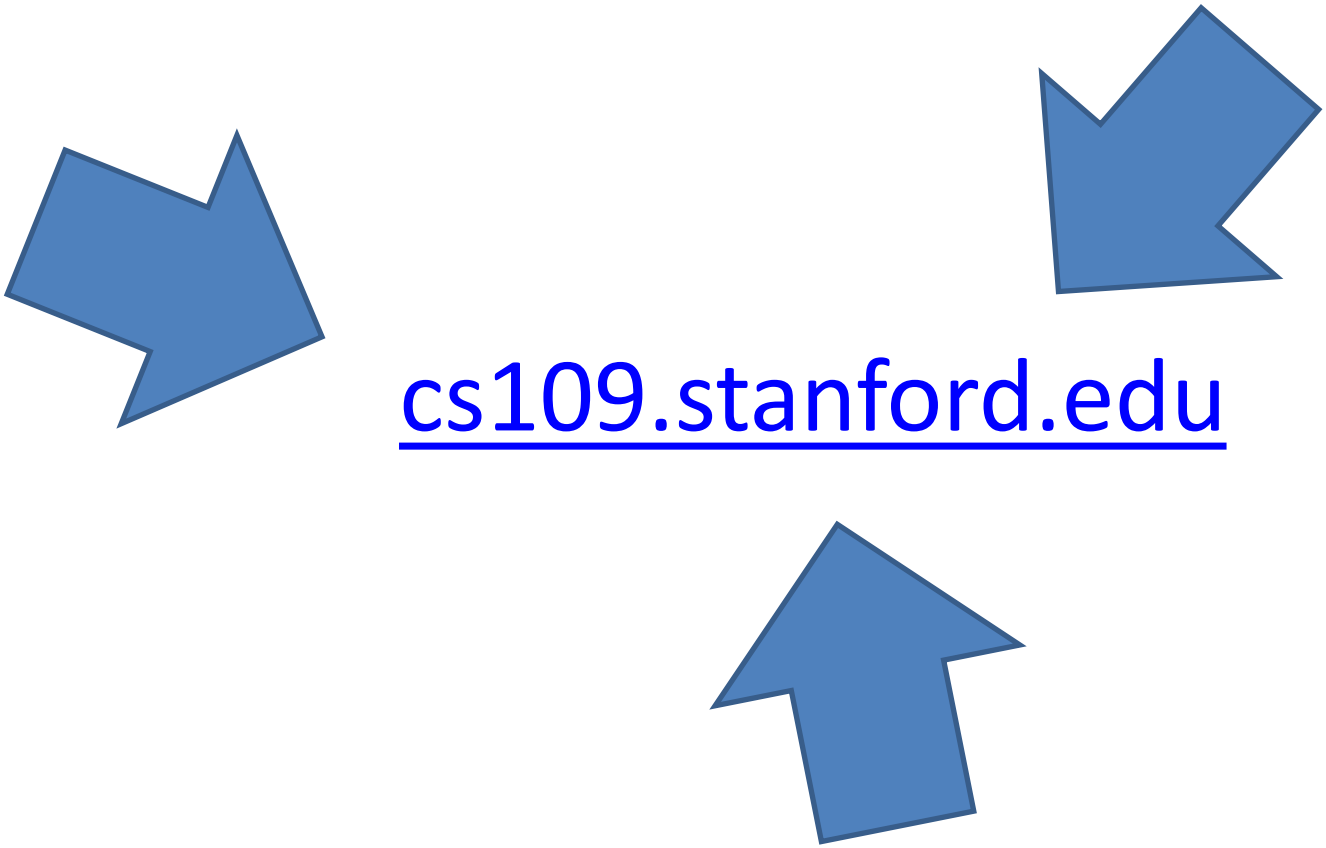
TA: Oishi Banerjee
oishib@stanford.edu
TBD
TBD

TA: Sam Schwager
sams95@stanford.edu
TBD
TBD

Course mechanics

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

Essential Information



cs109.stanford.edu

Are you in the right place?

Prereqs

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

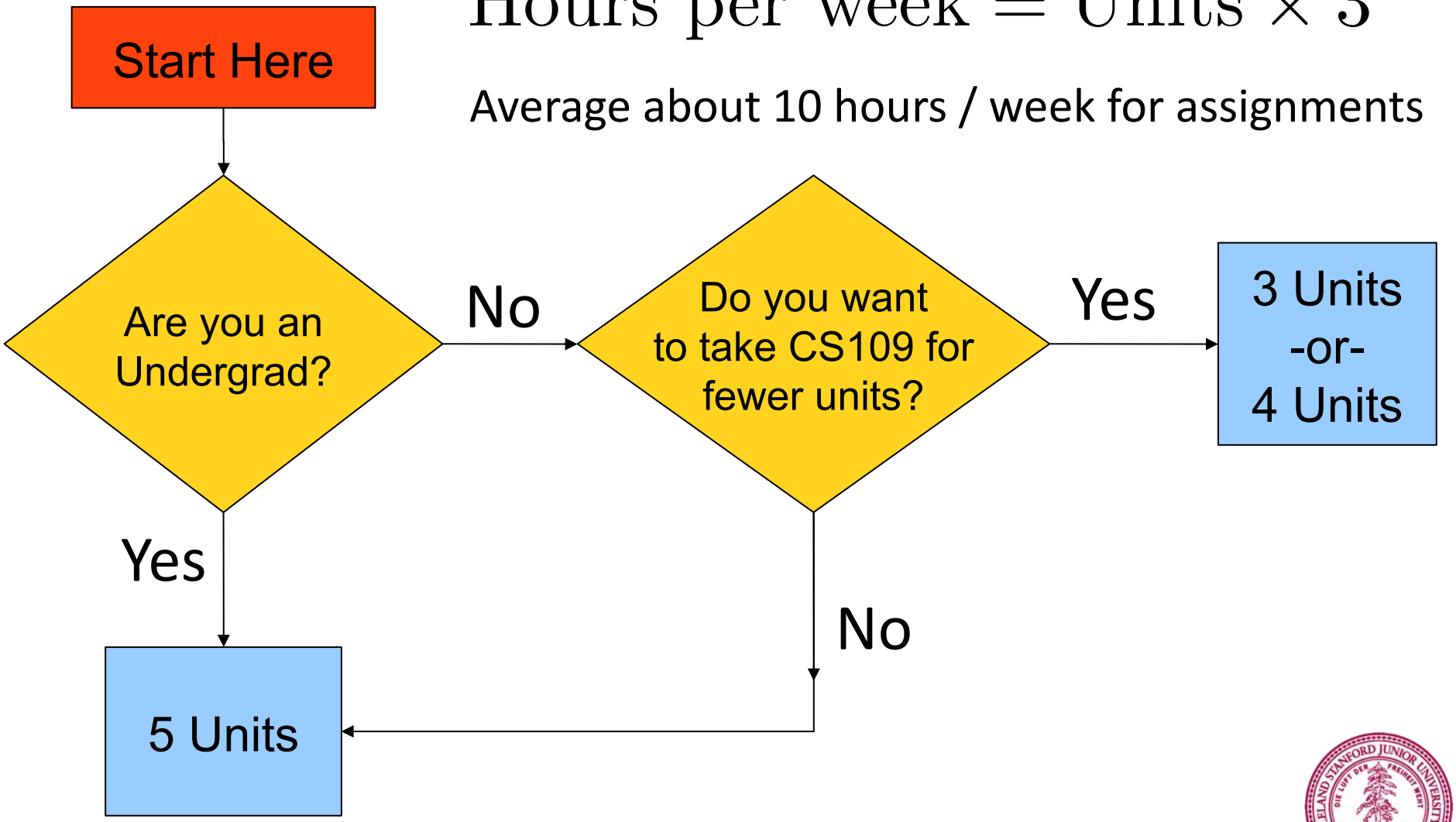
Staff Contact

- Post to Piazza for clarification
- Go to Working Office Hours
- Email cs109@cs.stanford.edu
- Email Chris or go to his office for course level issues.

CS109 Units

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

Average about 10 hours / week for assignments



Not Videotaped

* And you should expect to learn more

Class Breakdown

45% **6 Assignments**

20% **Midterm**
Tuesday Oct 30th, 7-9pm

30% **Final**
Wed Dec 12th, 3:30-6:30pm

5% **Section Participation**

Late Days

2


LELAND STANFORD JUNIOR UNIVERSITY
DIE LUFT DER FREIHEIT WEHT
1891

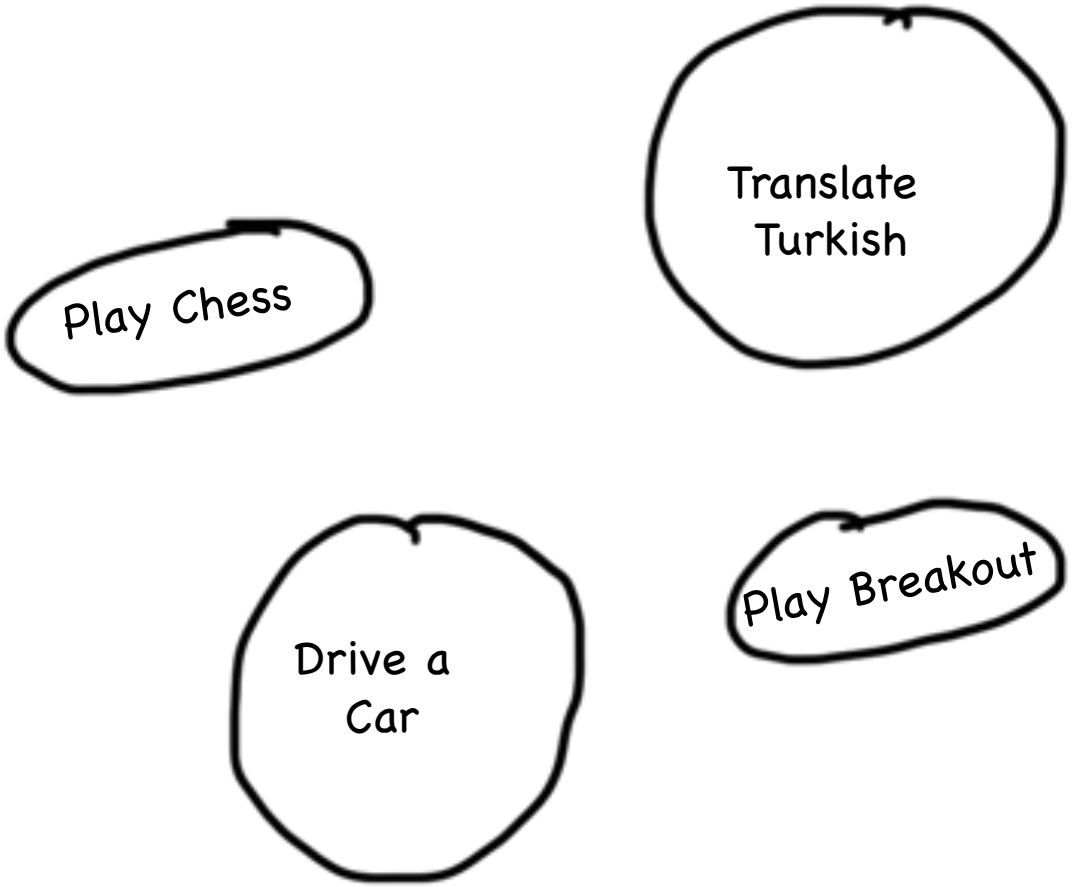
The Student Honor Code

Story of Modern AI

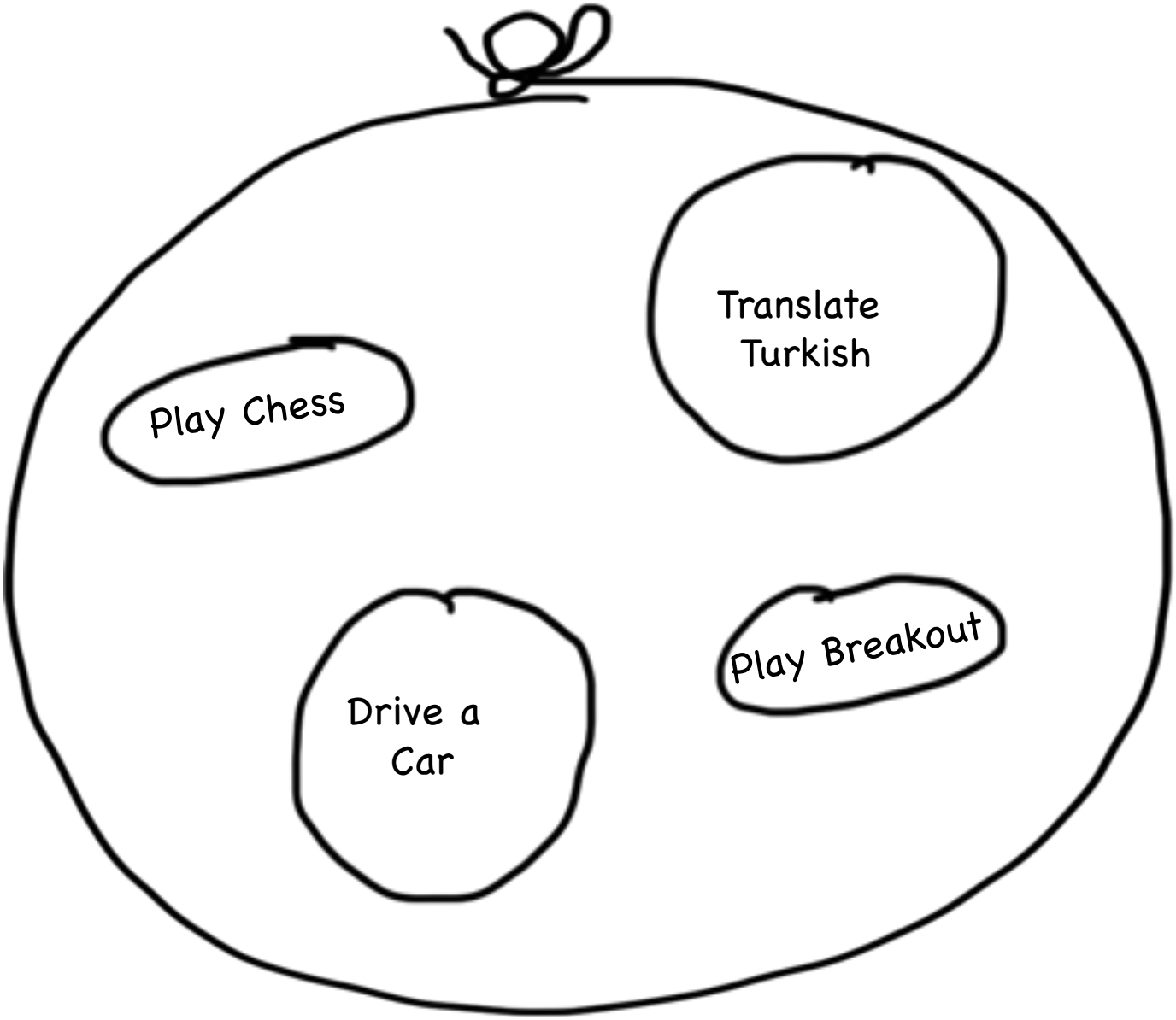
Modern AI
or, How we learned to combine probability and programming

Brief History

Narrow Intelligence



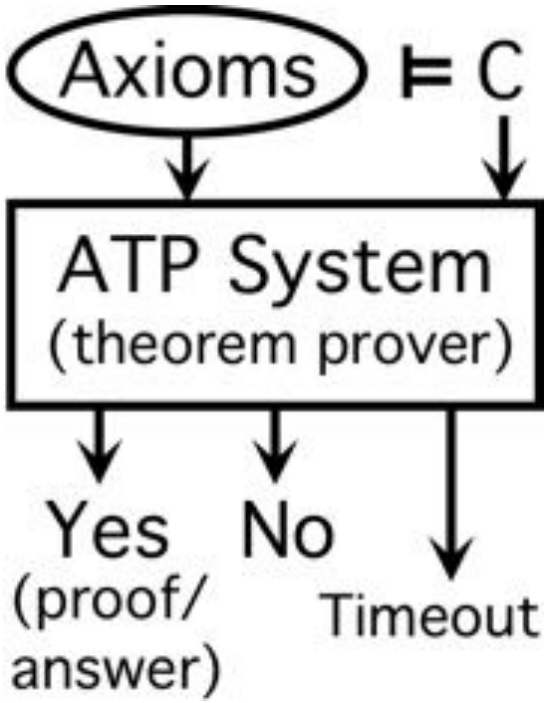
General Intelligence



Early Optimism 1950

1952

1955



Early Optimism 1950

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

Underwhelming Results 1950s to 1980s

The world is too complex


BRACE YOURSELVES
WINTER IS COMING

Something is going on in the world of AI

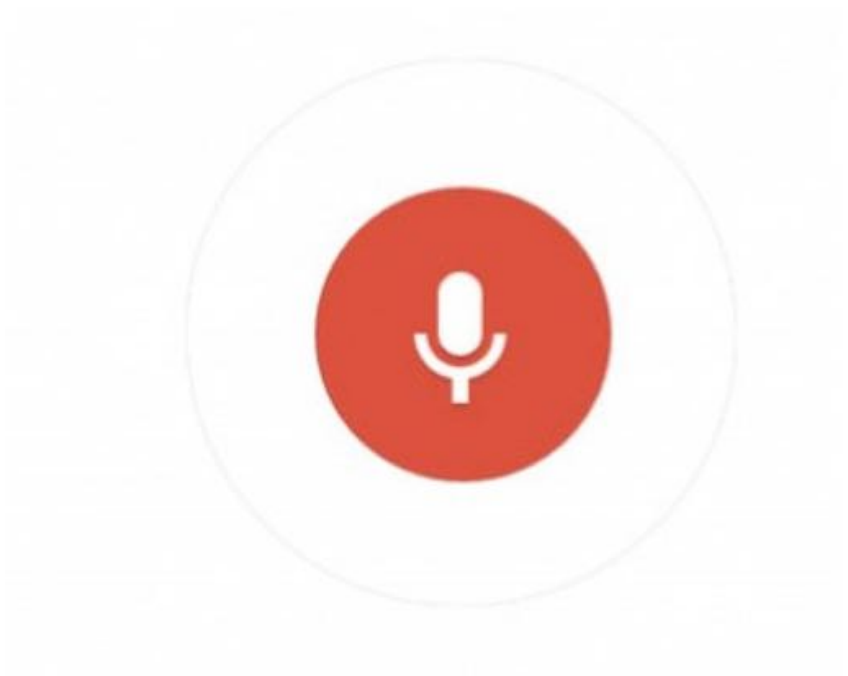
Big Milestones Pt 1

1997 Deep Blue

2005 Stanley

2011 Watson

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

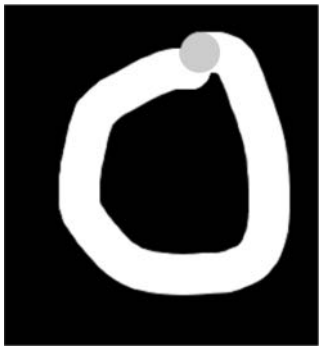
Computer Vision

Classification

That is a picture of a **one**

Classification

That is a picture of a **zero**



Classification

That is a picture of an **zero**

* It doesn't have to be correct all of the time

Can you do it?

What number is this?

What number is this?

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					
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 Program


```
public class HarryHat 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

1. Probability From Examples

When Does the Magic Happen?

Lots of Data + Sound Probability

Machine Learning

Basically just a rebranding of statistics and probability.

Vision is Hard

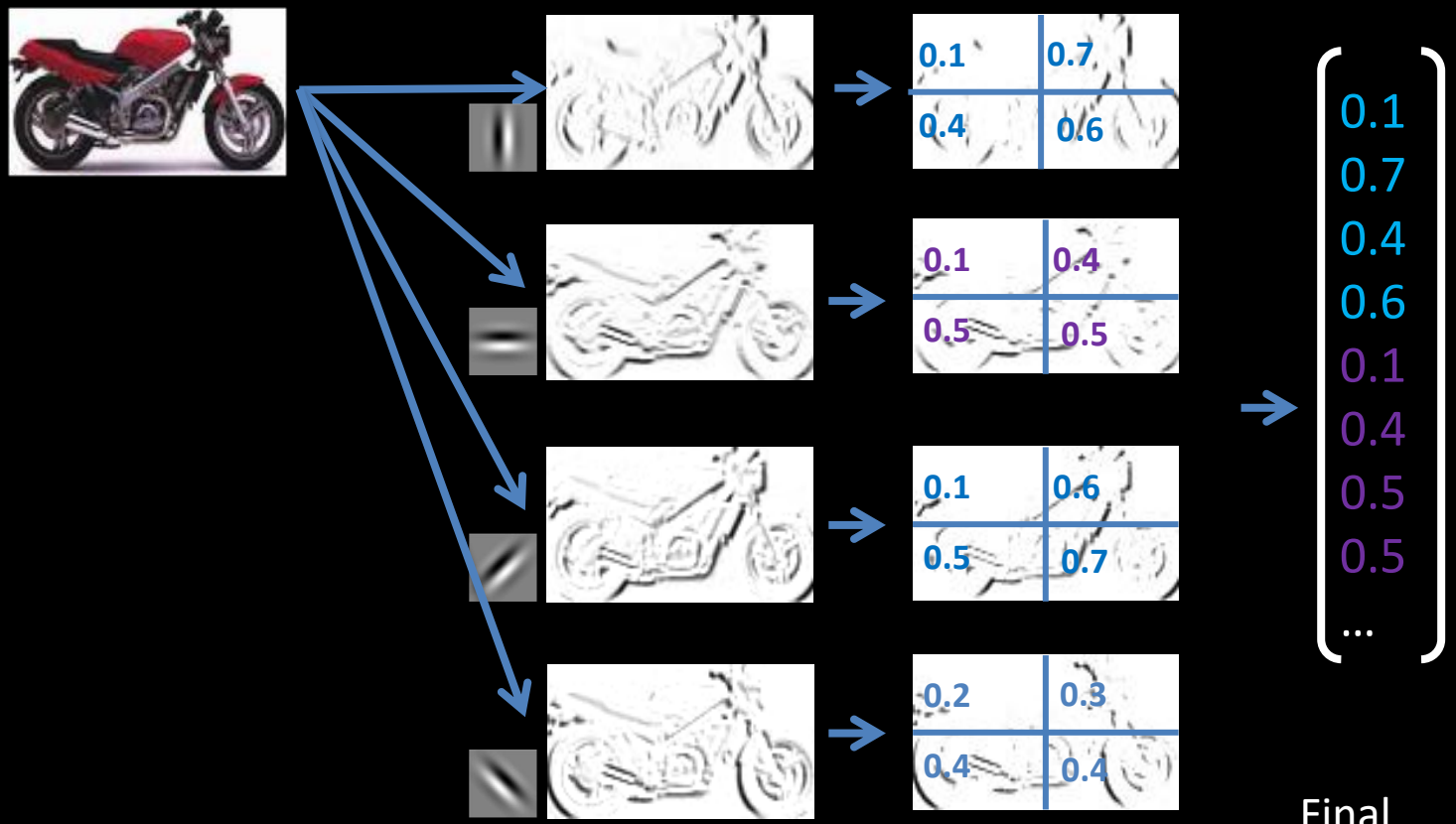
You see this:

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]

Human Designed Features



[Andrew Ng]

Some Great Thinkers

Daphne Koller

Straight ML Not Perfect…

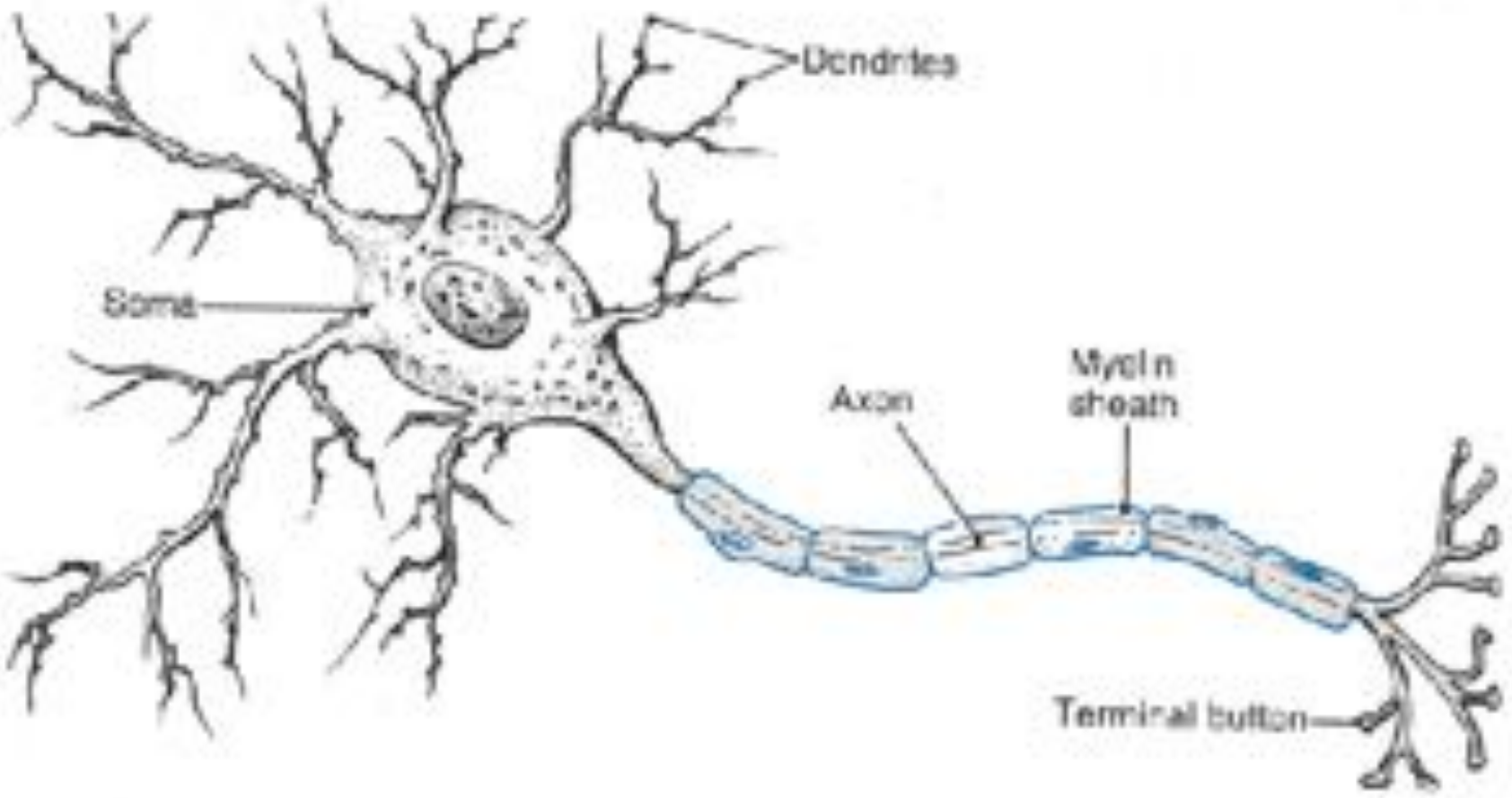
Two Great Ideas

1. Probability from Examples

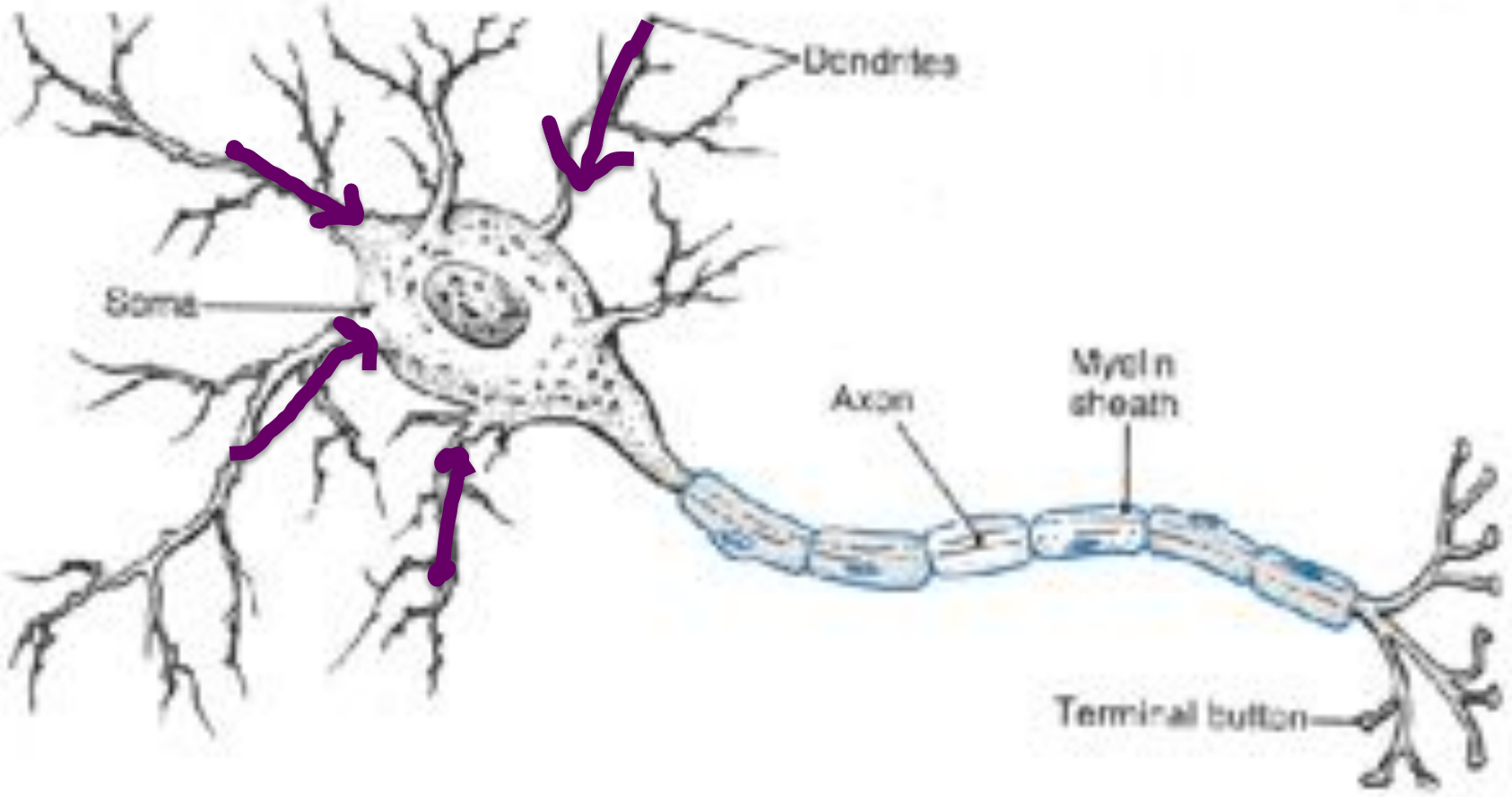
2. Artificial Neurons

2. Artificial Neurons

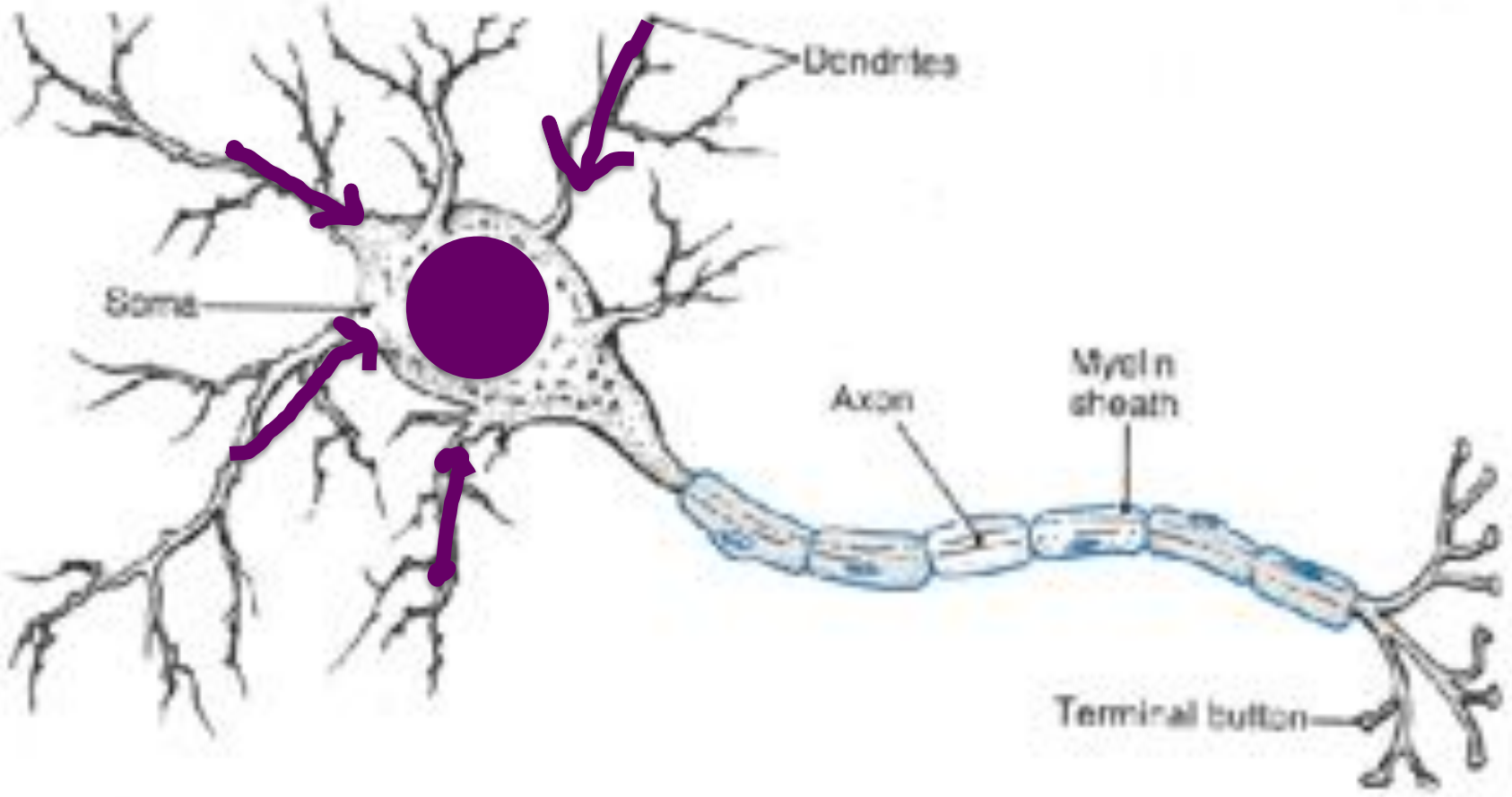
Neuron



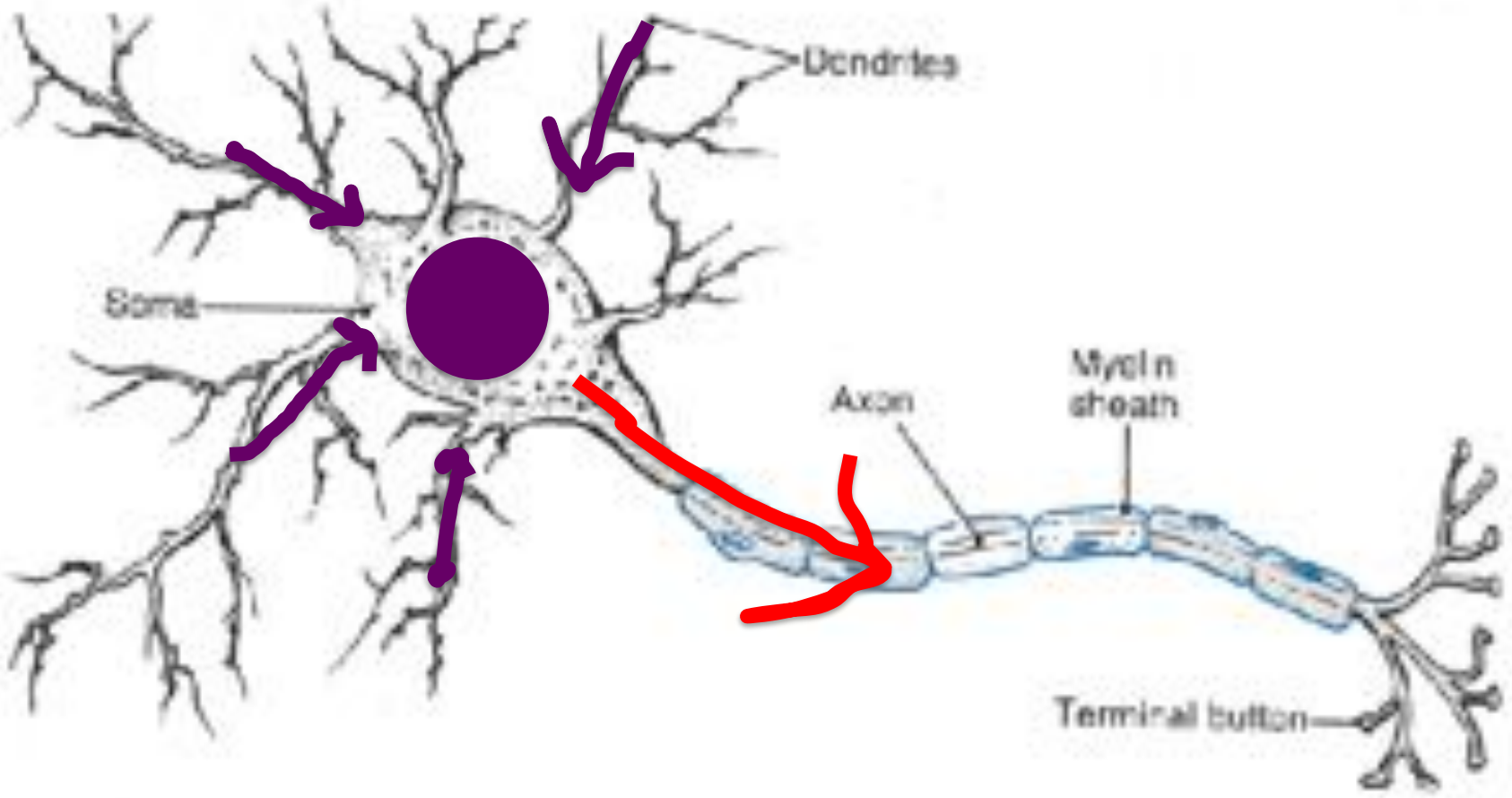
Neuron



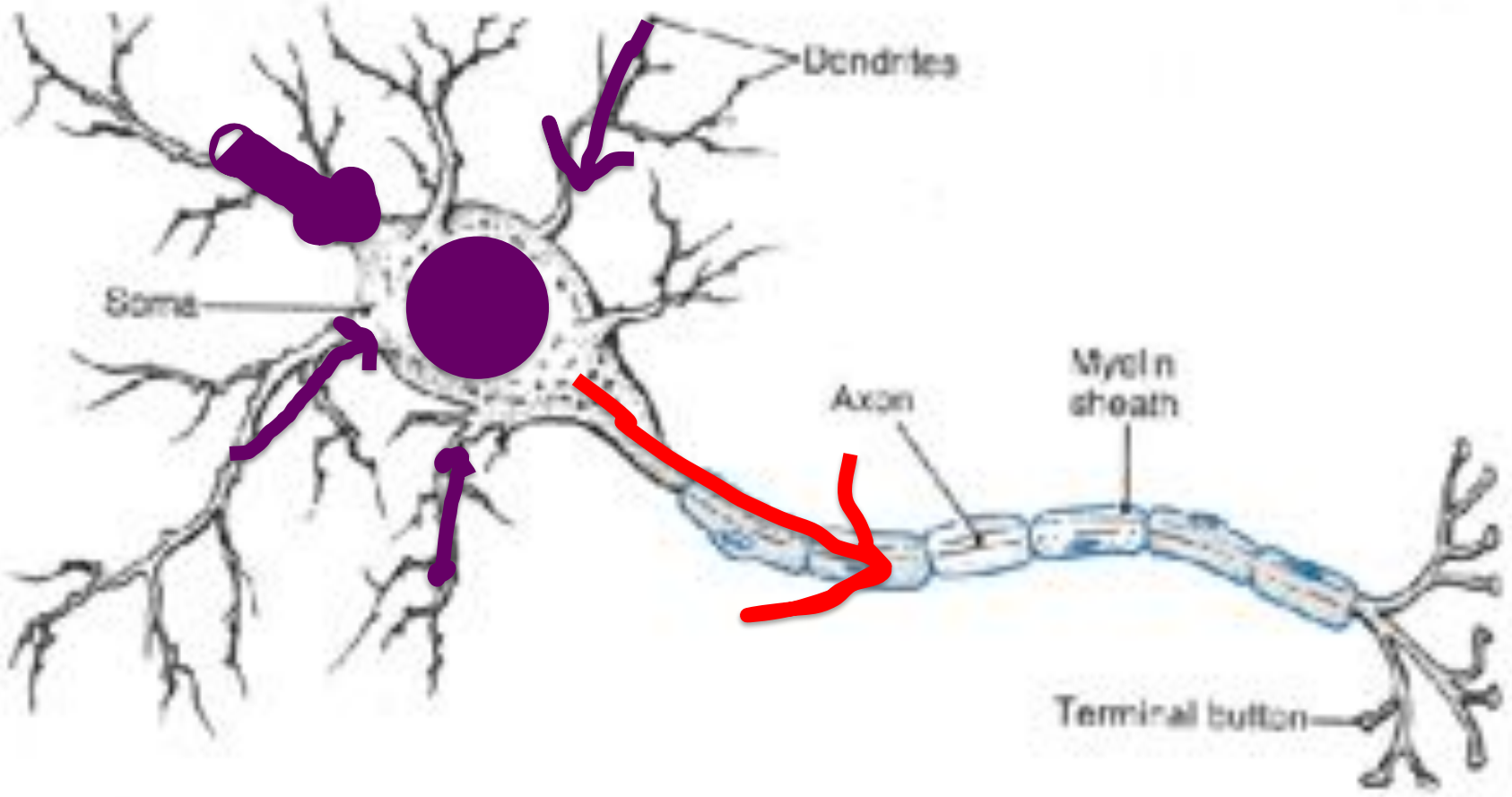
Neuron



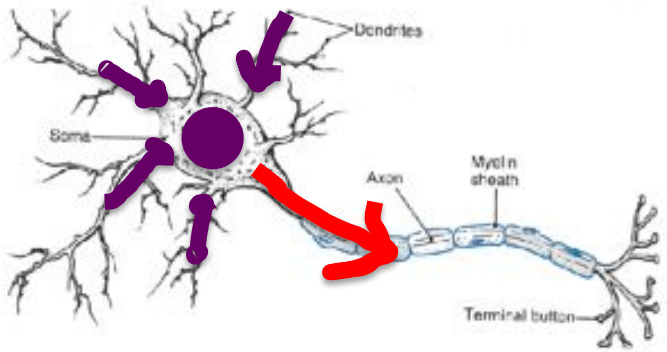
Neuron

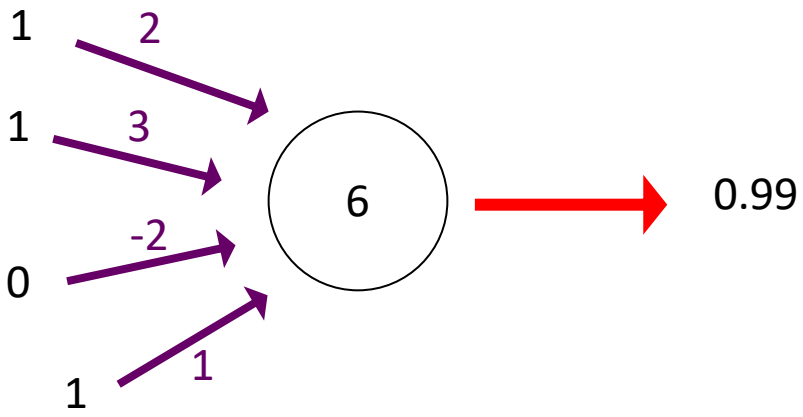


Some Inputs are More Important

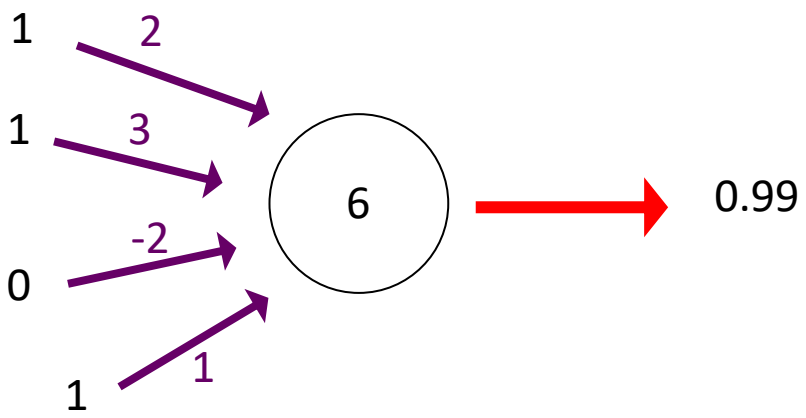


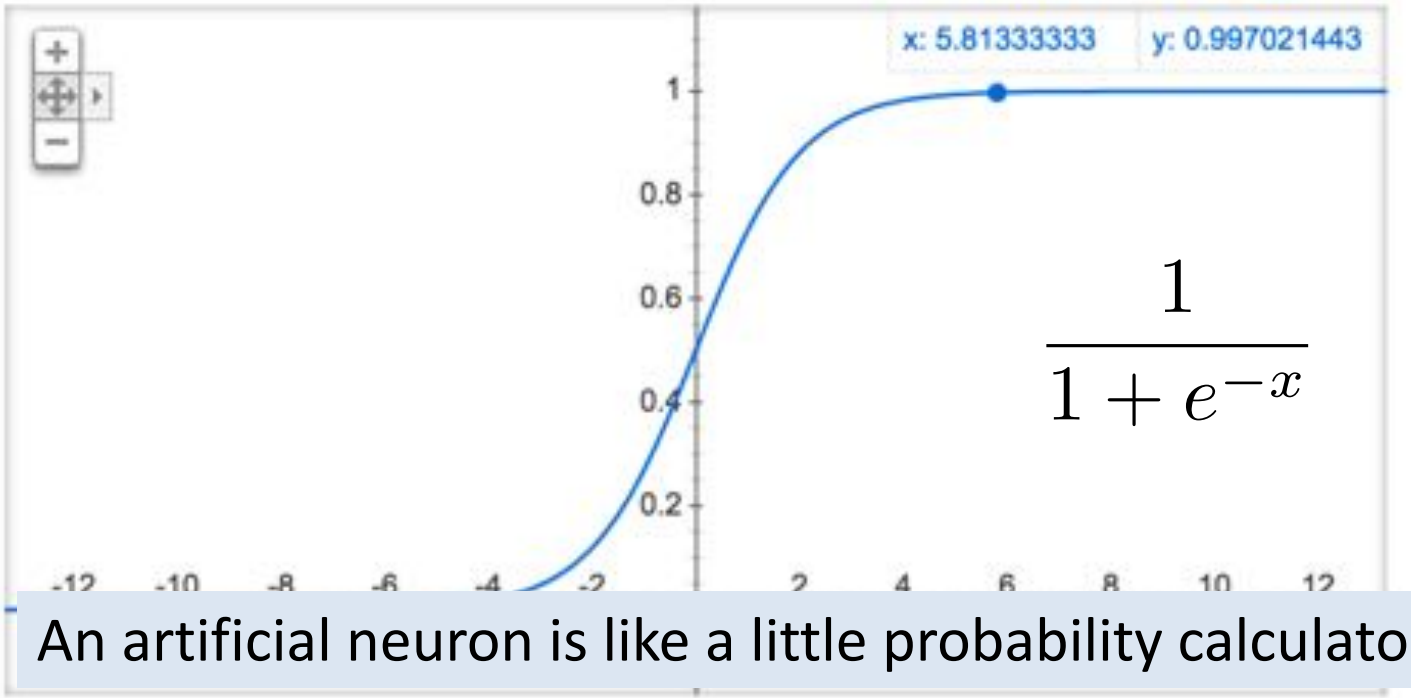
Artificial Neuron





Sigmoid Function



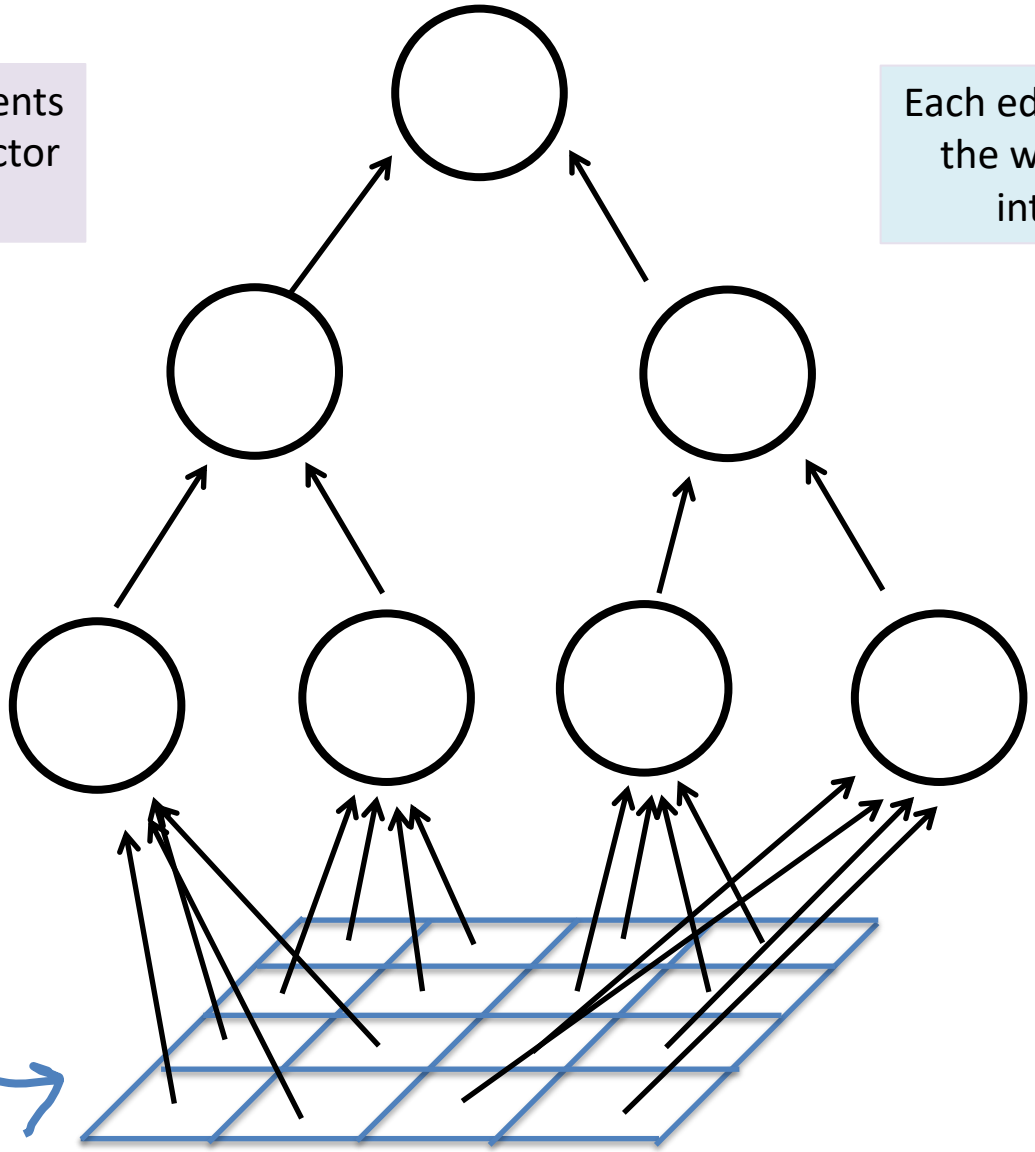


An artificial neuron is like a little probability calculator

Neural Network

Each node represents a neuron (or a vector of neurons)

Each edge represents the weight of the interaction



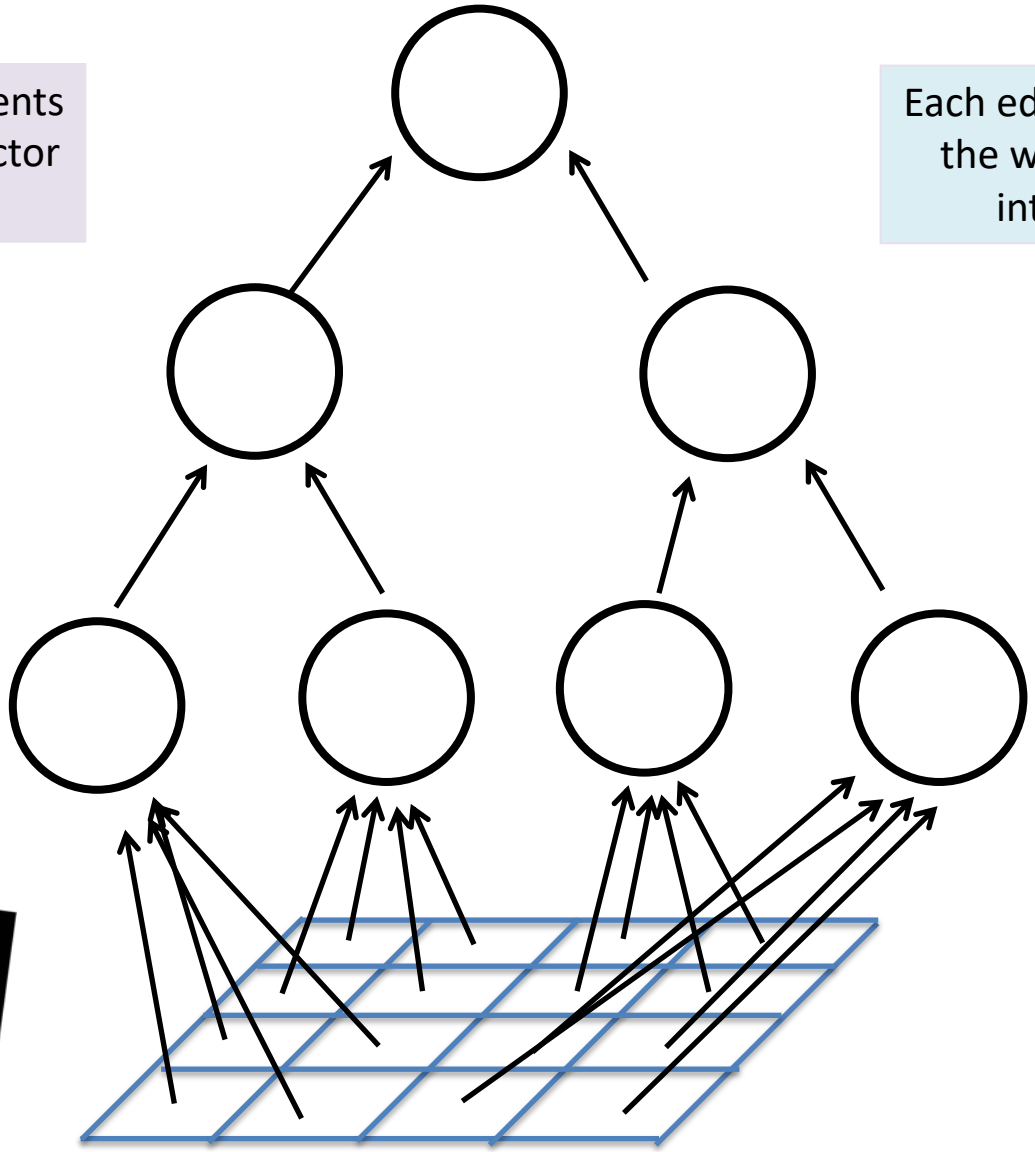
Pixels

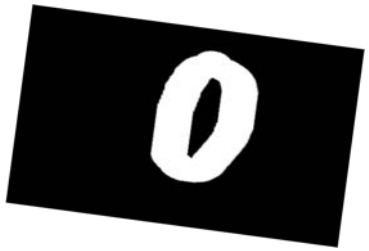
Forward Pass…

Forward Pass

Each node represents a neuron (or a vector of neurons)

Each edge represents the weight of the interaction

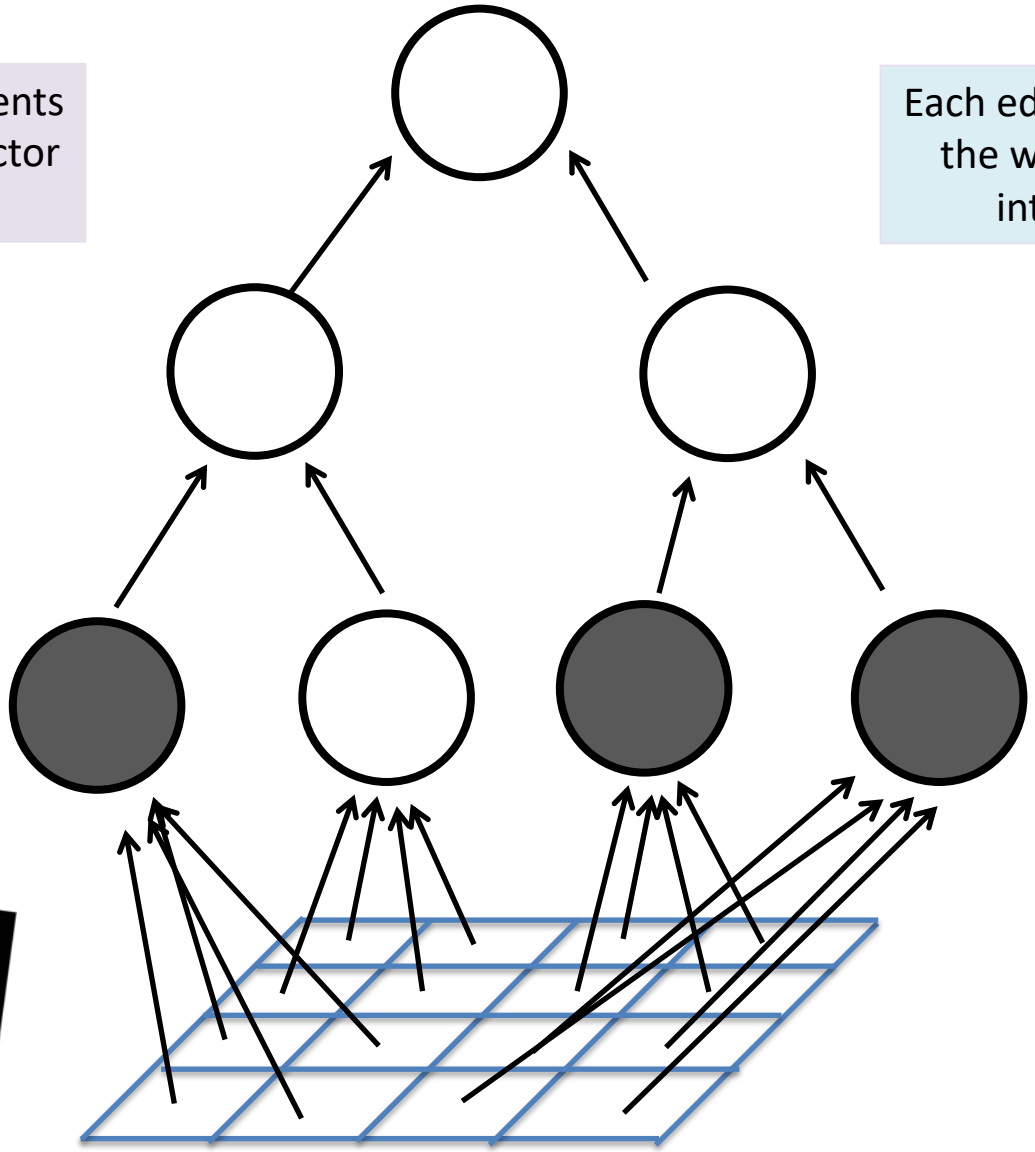


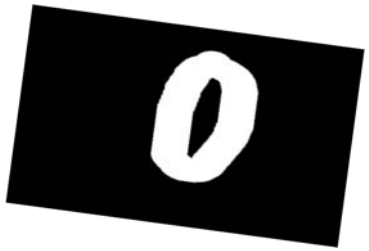


Forward Pass

Each node represents a neuron (or a vector of neurons)

Each edge represents the weight of the interaction

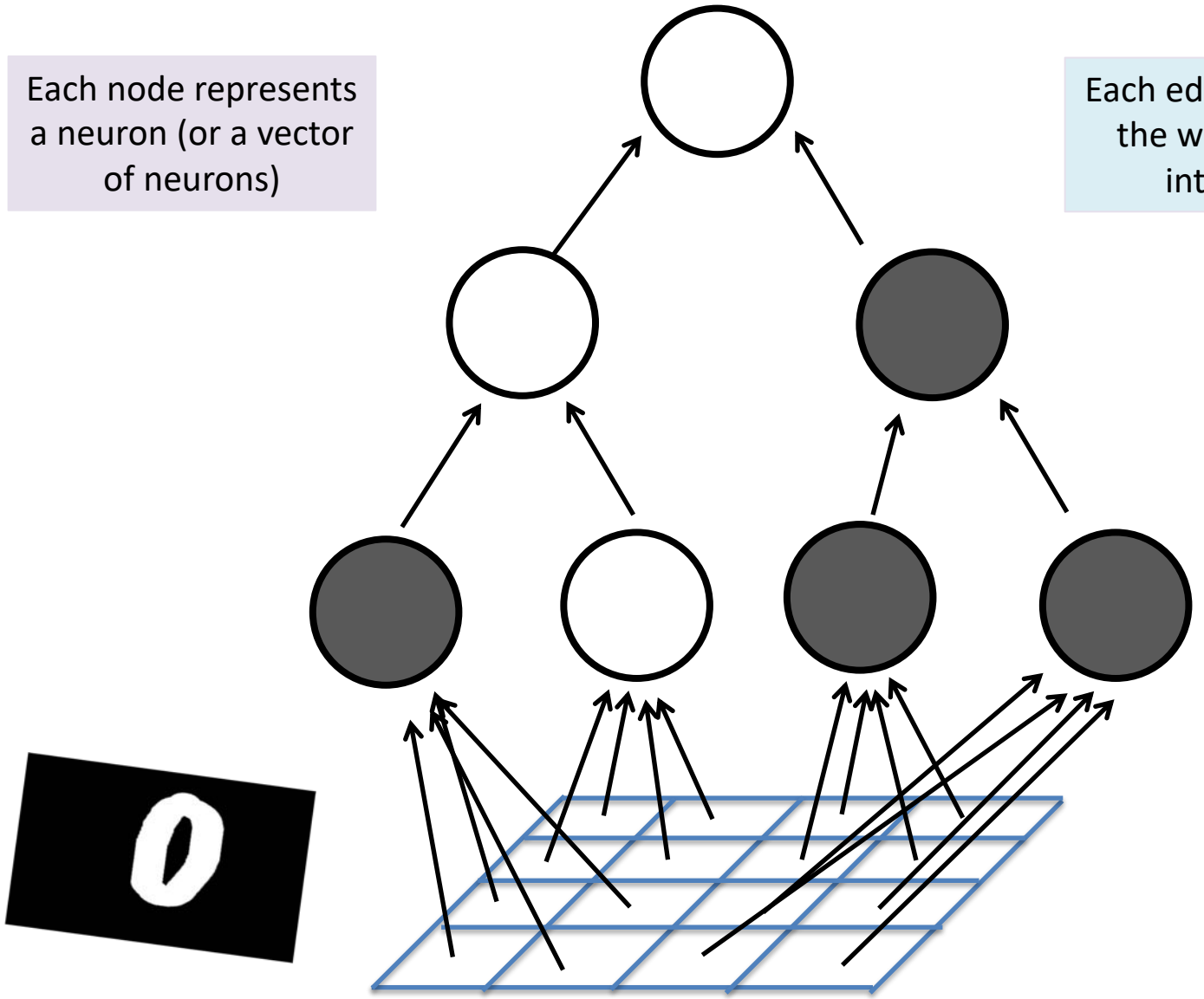




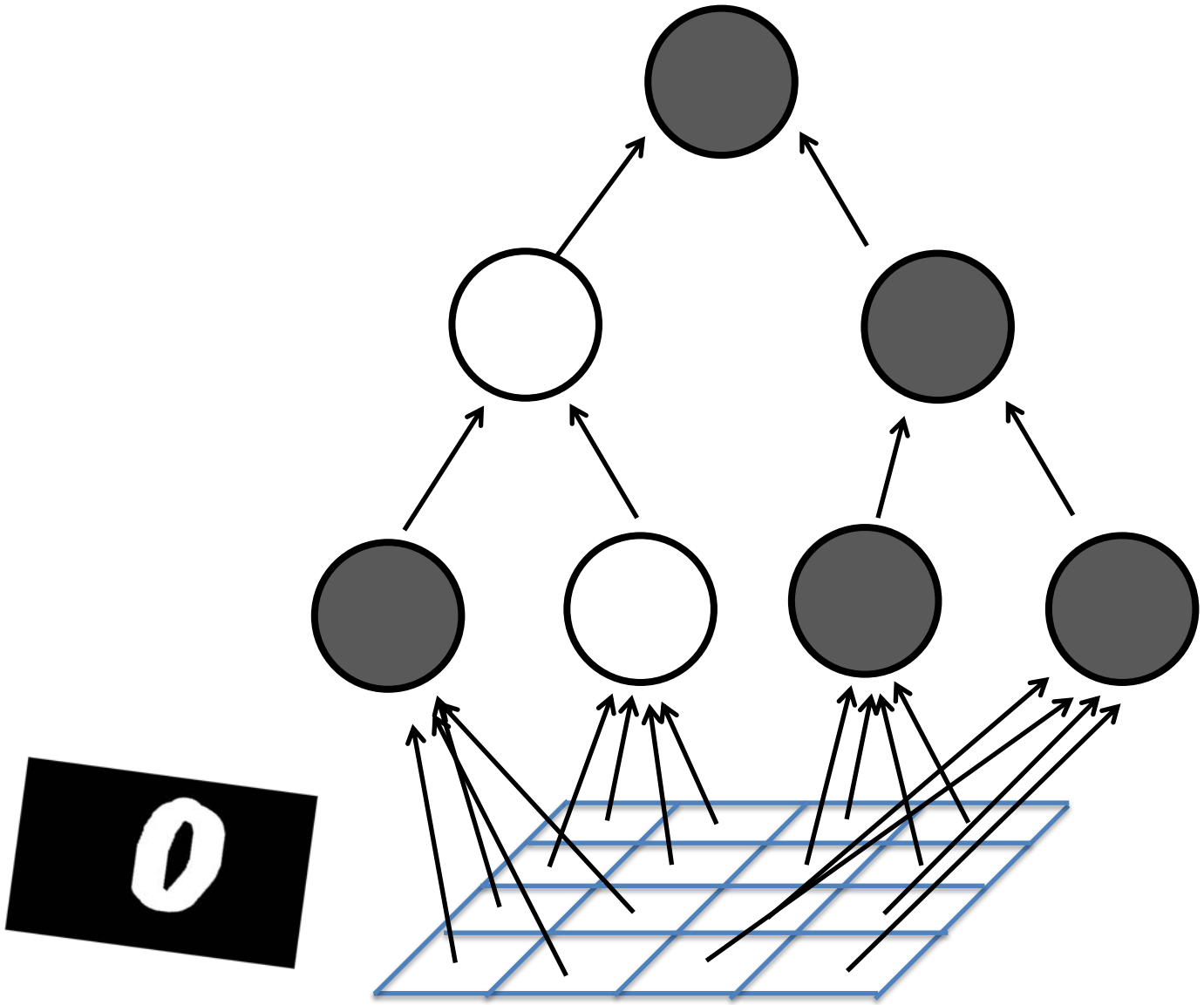
Forward Pass

Each node represents a neuron (or a vector of neurons)

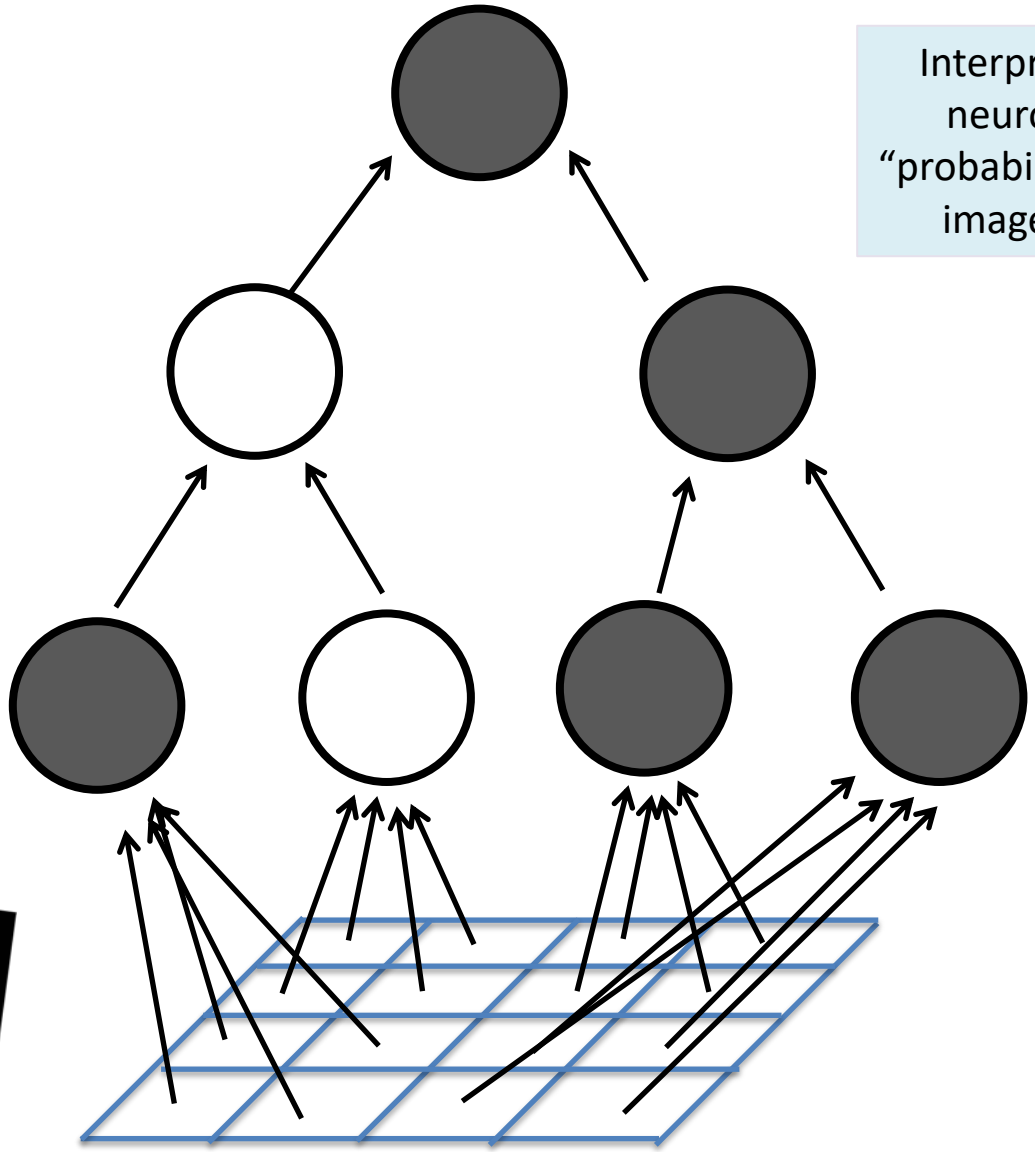
Each edge represents the weight of the interaction



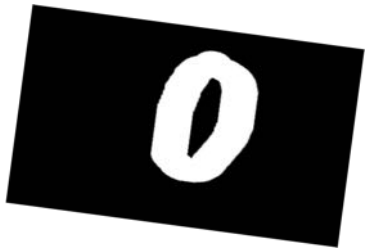
Forward Pass



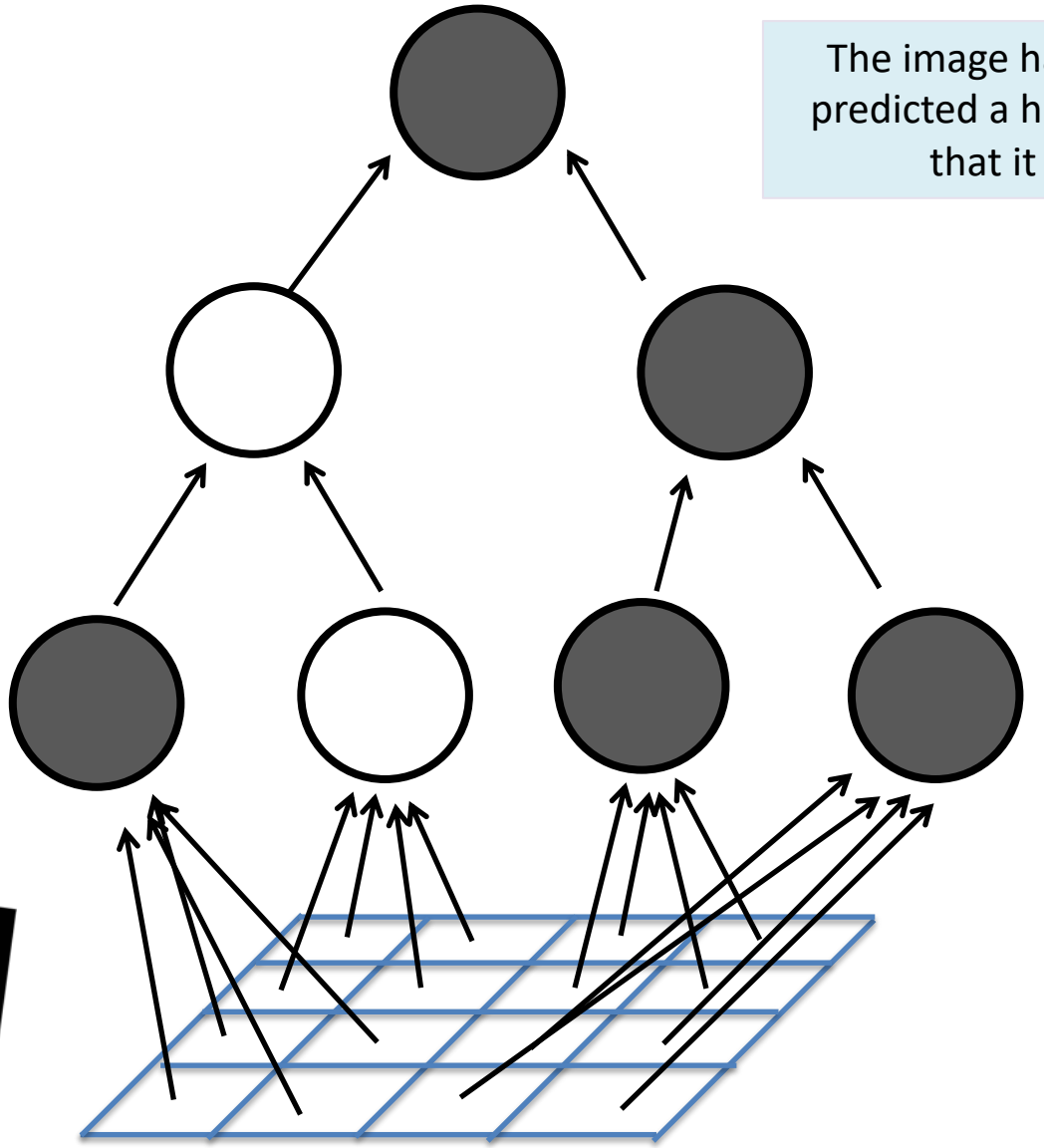
Forward Pass



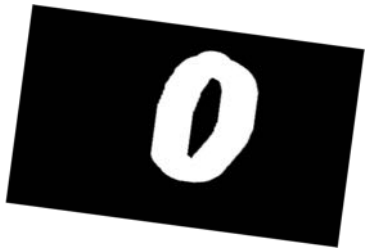
Interpret the last neuron as the "probability" that the image is of a 1



Backward Pass



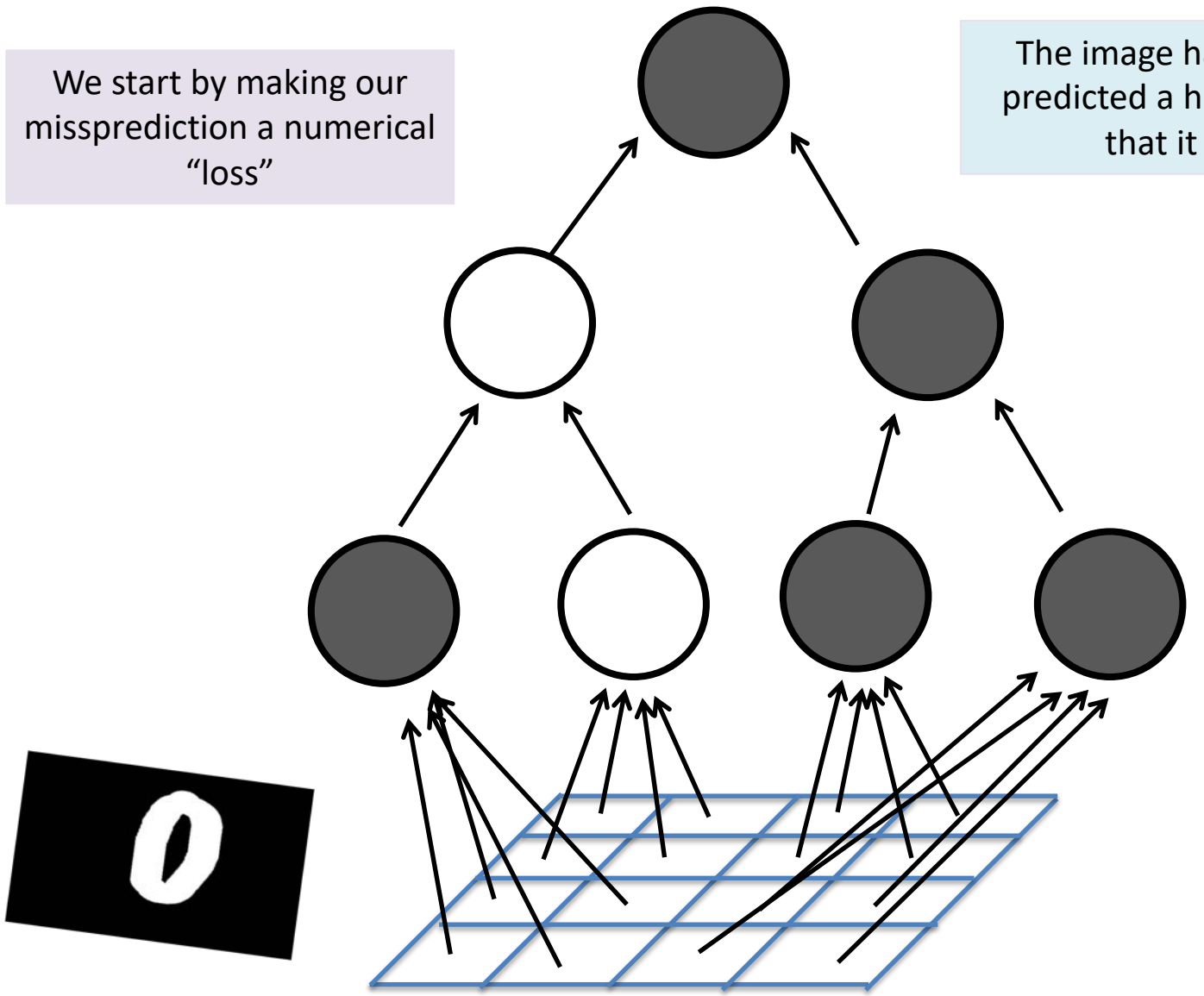
The image had a 0 but we predicted a high probability that it was a 1



Backward Pass

We start by making our misspredicton a numerical "loss"

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

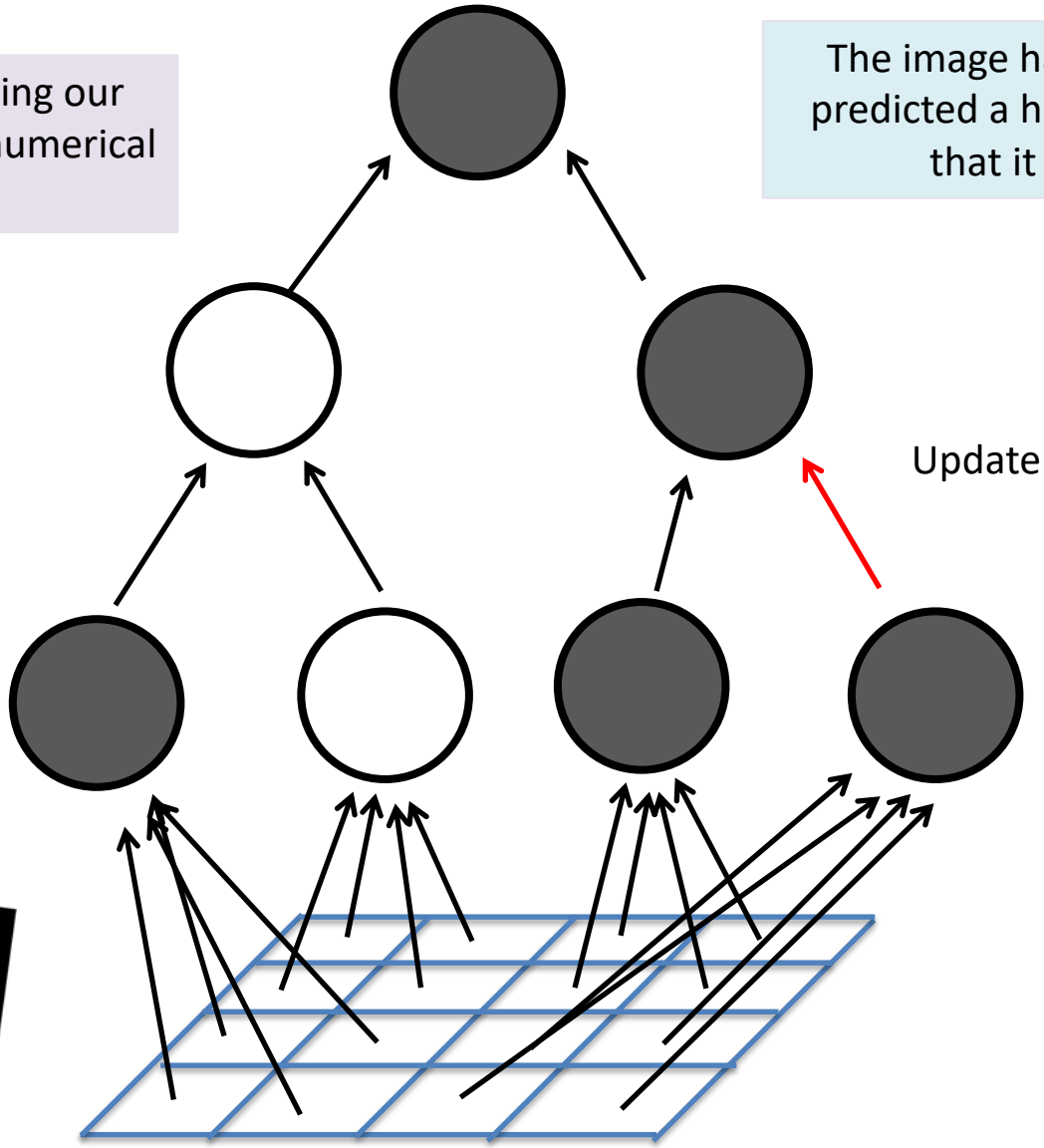


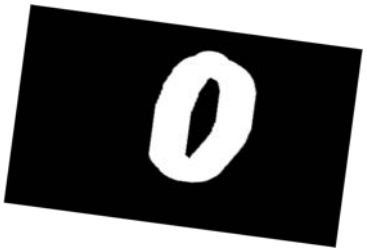
Backward Pass

We start by making our misspredication a numerical "loss"

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

Update each connection





Chose weights that maximize the probability of the right answers

$$P(Y = 1|X = \mathbf{x}) = \hat{\mathbf{y}} \qquad \hat{y} = \sigma\left(\sum_{j=0}^{m_h} \mathbf{h}_j \theta_j^{(\hat{y})}\right)$$

For one datum

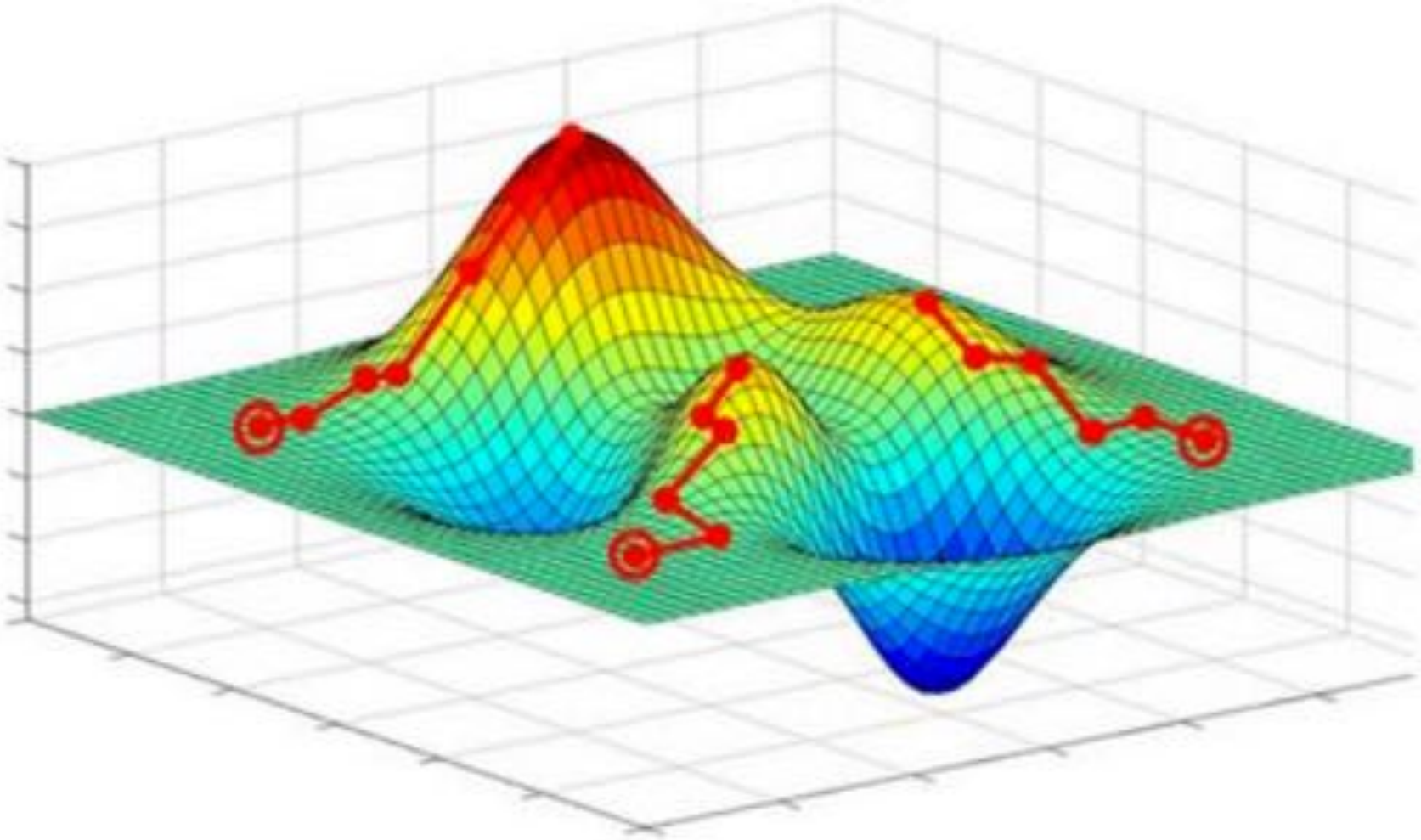
$$P(Y = y|X = \mathbf{X}) = (\hat{y})^y (1 - \hat{y})^{1-y}$$

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 Ascent



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

Gradient of output layer params

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

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

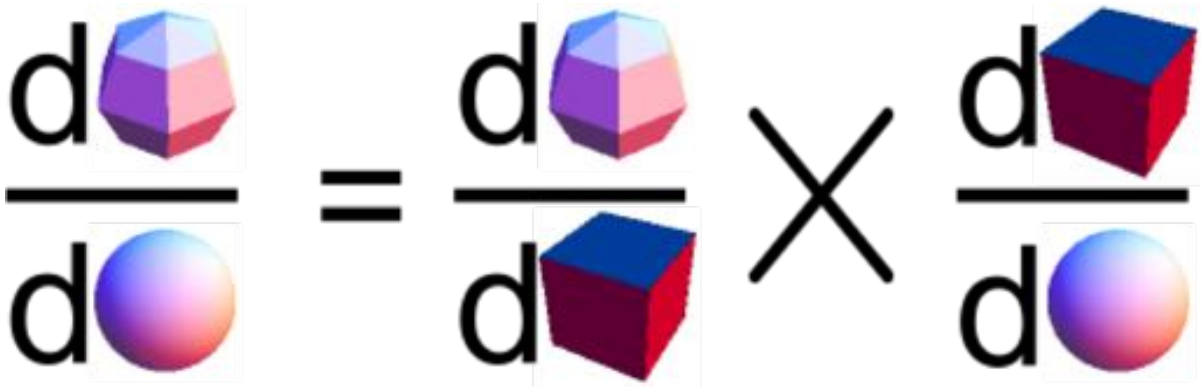
$$\frac{\partial \hat{y}}{\partial \theta_i^{(\hat{y})}} = \sigma\left(\sum_{j=0}^{m_h} \mathbf{h}_j \theta_j^{(\hat{y})}\right)\left[1 - \sigma\left(\sum_{j=0}^{m_h} \mathbf{h}_j \theta_j^{(\hat{y})}\right)\right] \cdot \frac{\partial}{\partial \theta_i^{(\hat{y})}} \sum_{j=0}^{m_h} \mathbf{h}_j \theta_j^{(\hat{y})}$$

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

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

That looks scarier than it is

Chain Rule Down the Network



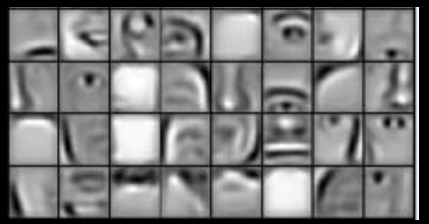
Where you will be by the end of class

When you train,
something really neat happens

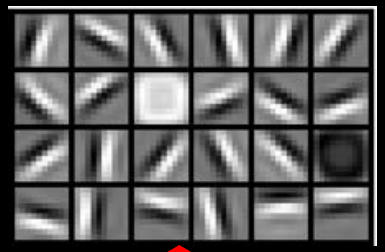
Visualize the Weights

Training set: Aligned images of faces.

object models



object parts (combination of edges)

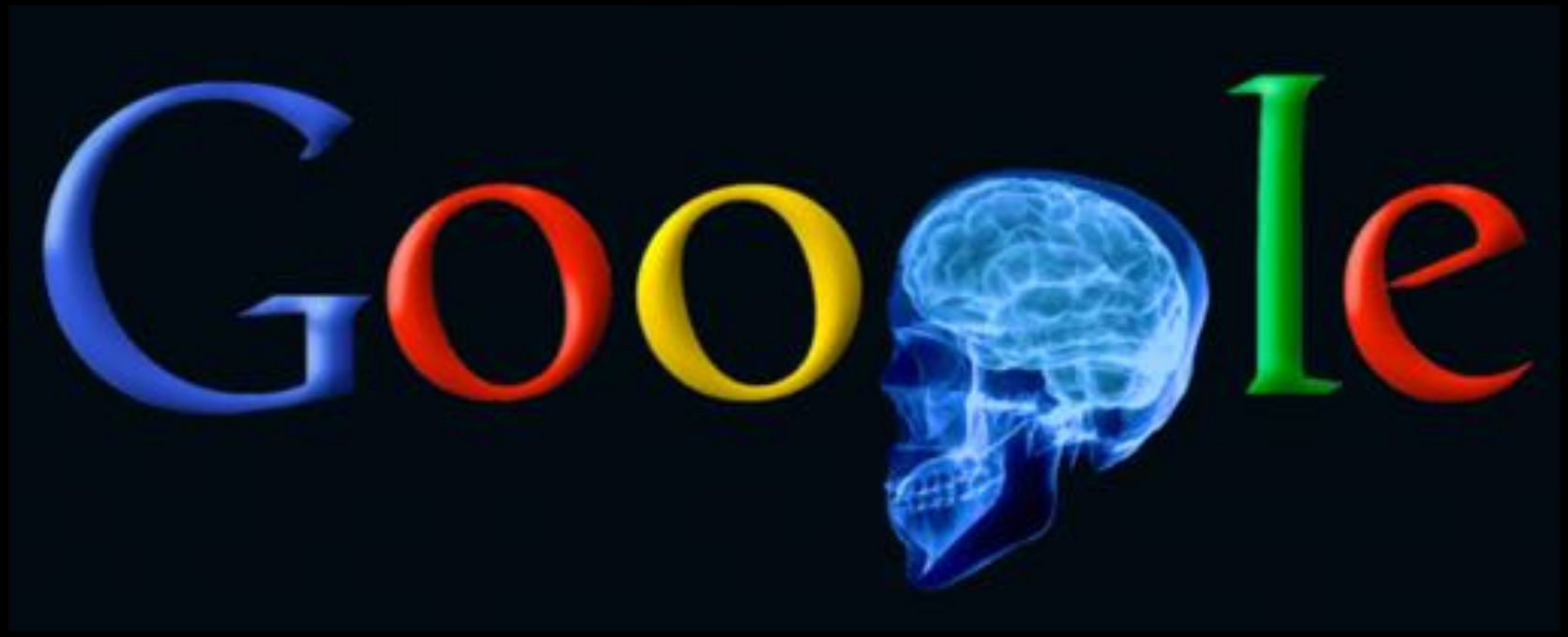


edges

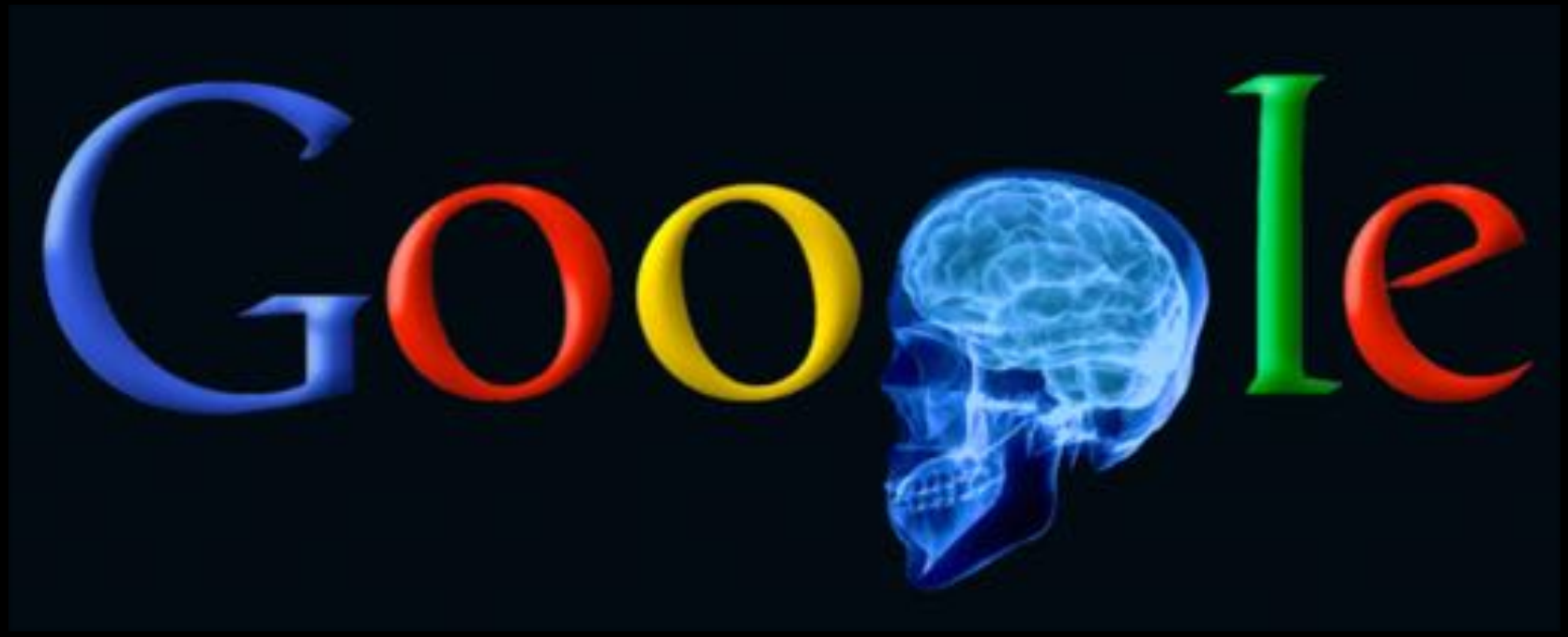
pixels

[Honglak Lee]

Google Brain



Google Brain



1 Trillion Artificial Neurons

A Neuron That Fires When It Sees Cats

Top stimuli from the test set

Optimal stimulus
by numerical optimization

Le, et al., *Building high-level features using large-scale unsupervised learning*. ICML 2012

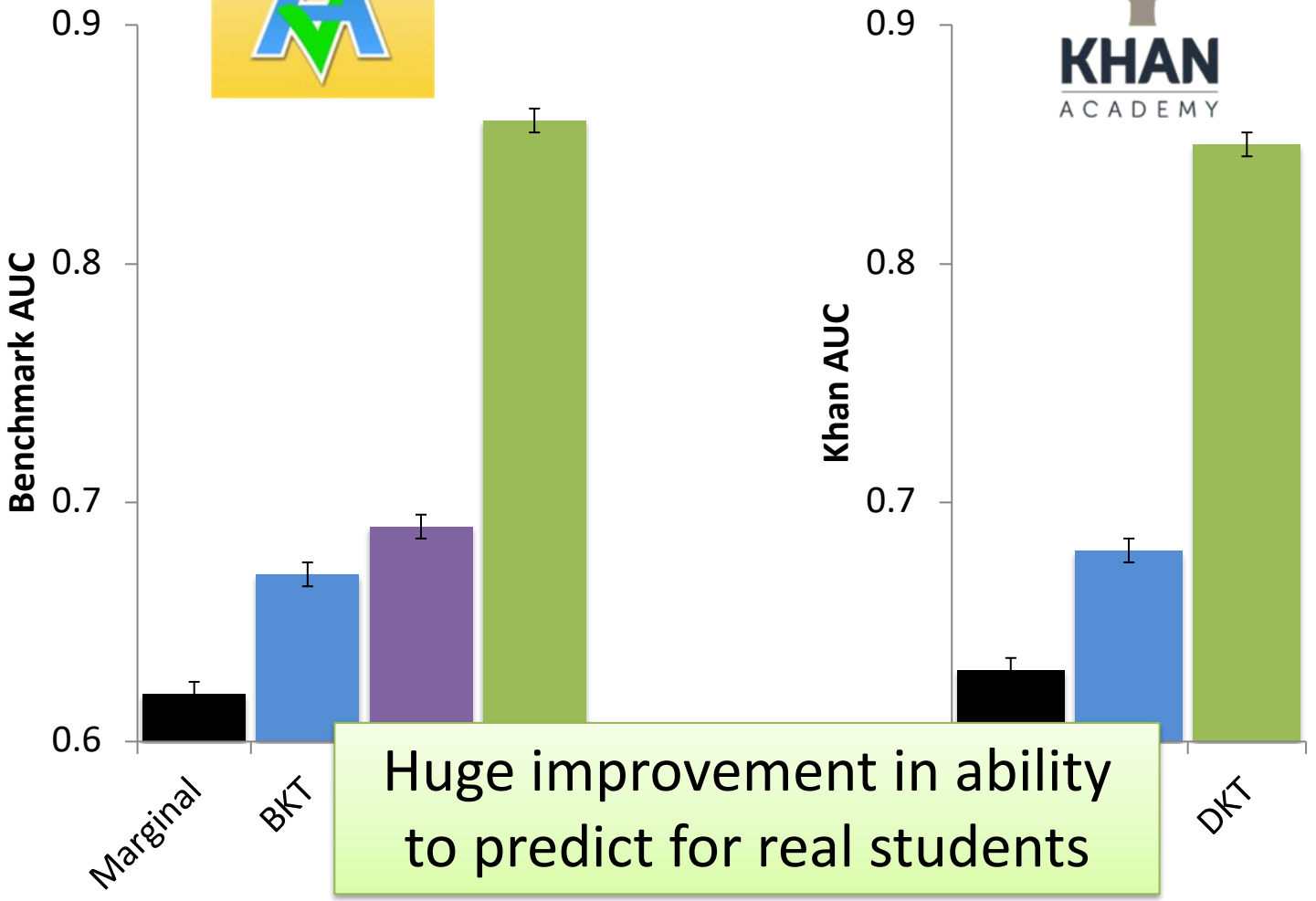

We're essentially
a cat detection company

Other Neurons

Le, et al., *Building high-level features using large-scale unsupervised learning*. ICML 2012

Autonomous Tutor

Prediction Results



Huge improvement in ability to predict for real students

Piech et al, 2015

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

End of Story

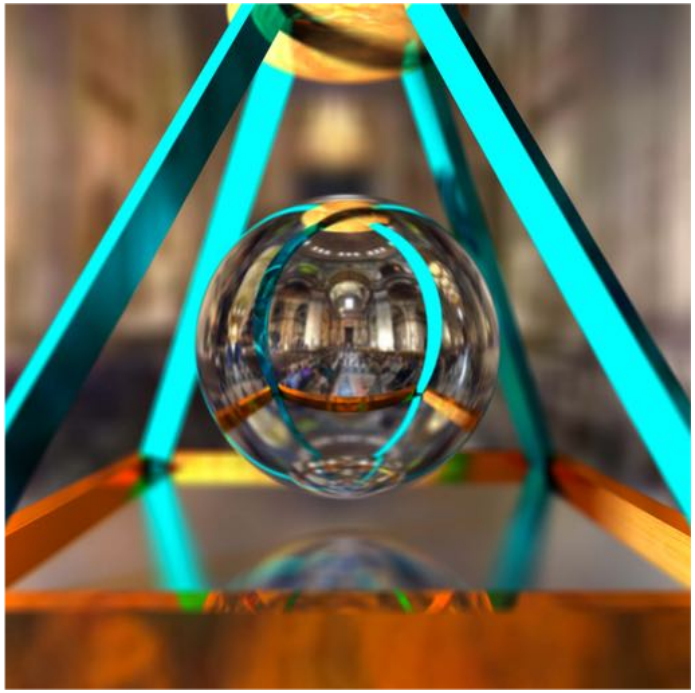
Except it isn’t the end of the story…

Probability is more than just machine learning

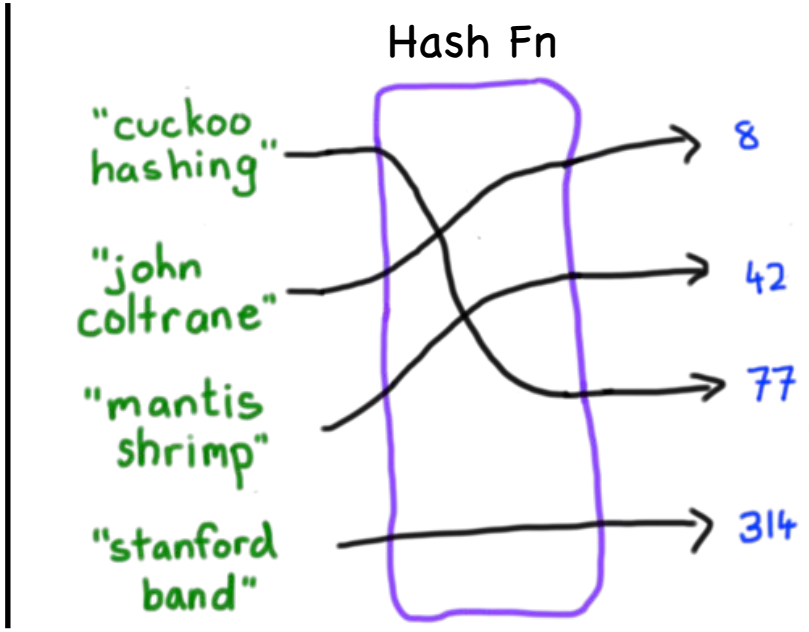
Abundance of Important Problems

Algorithms and Probability

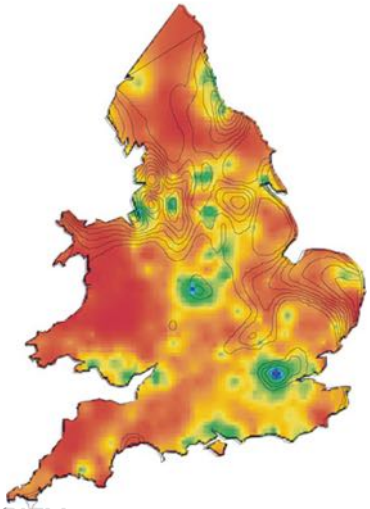
Eg Raytracing



Eg HashMaps



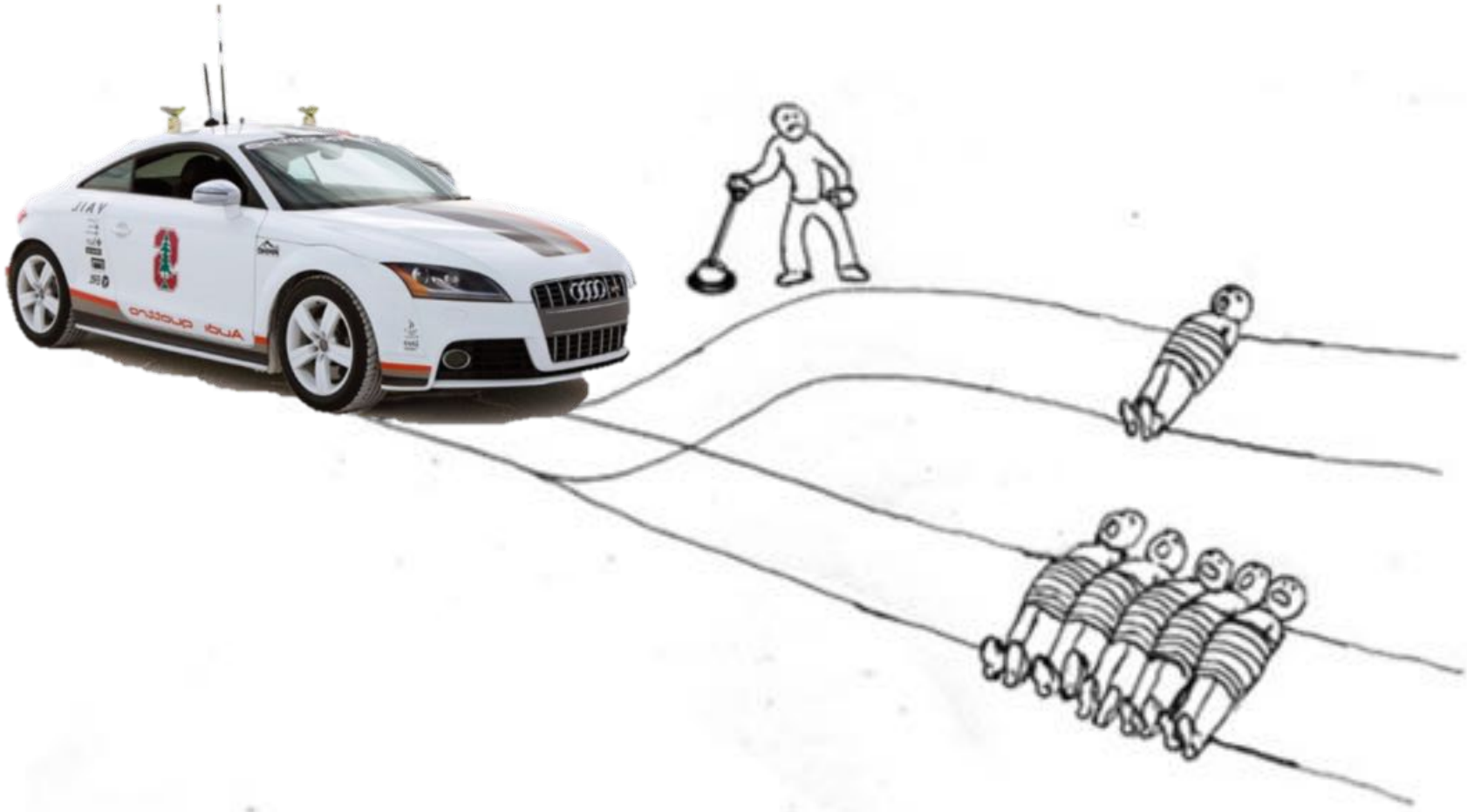
Medicine and Probability



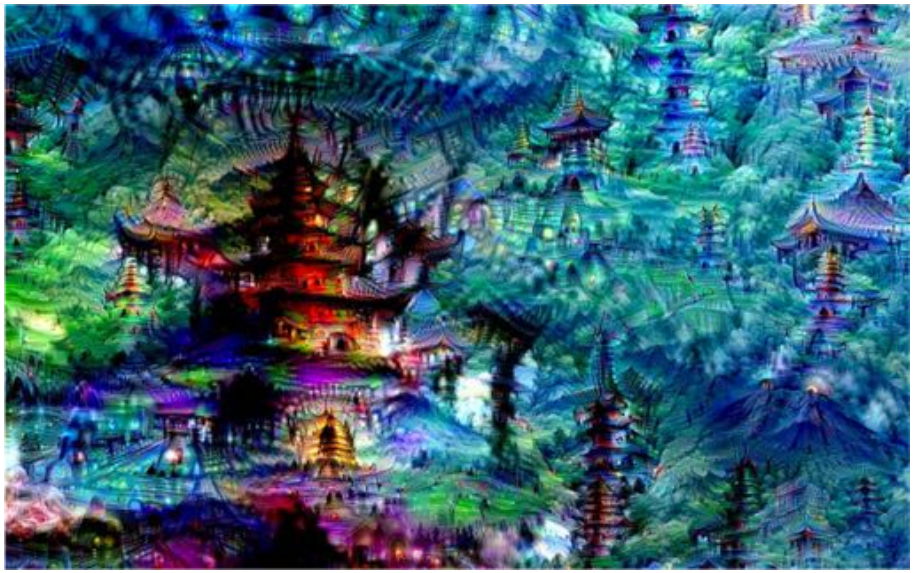
Autocomplete

Probability in Practice

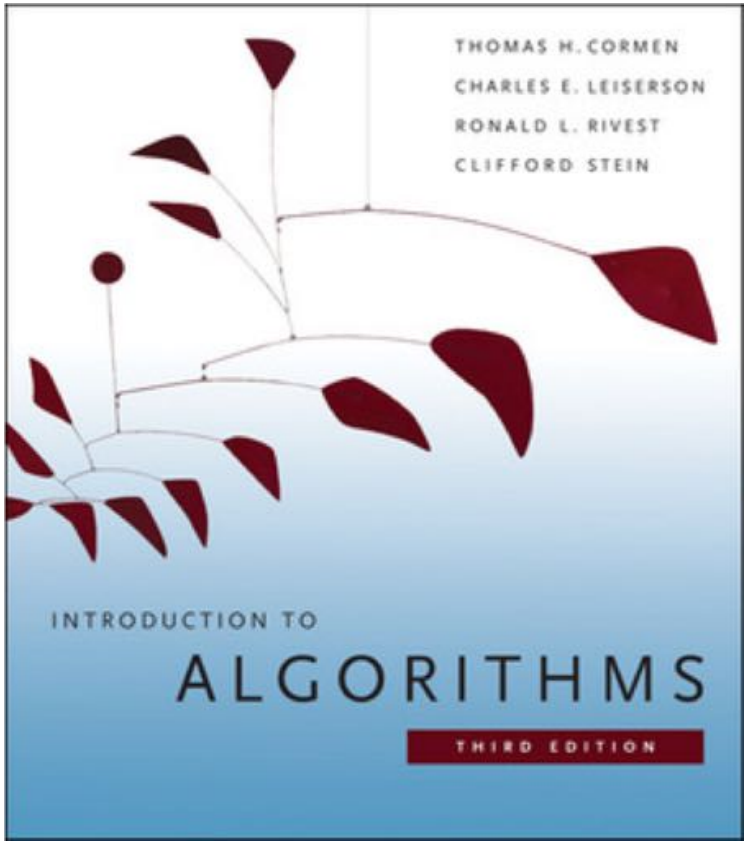
Philosophy and Probability

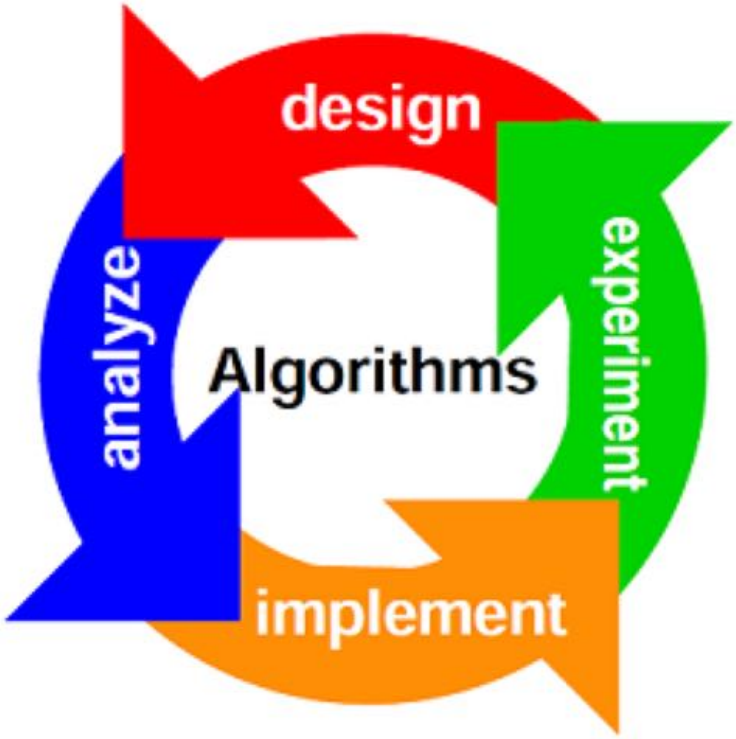


Art and Probability



Probabilistic Analysis of Algorithms





#1 Most Desired Skill in Industry

Microsoft's competitive advantage, [Bill Gates] responded, was its expertise in "Bayesian [probabilistic] networks.“

(from Los Angeles Times, Oct. 28, 1996)

“The sexy job in the next 10 years will be statisticians.”

-Hal Varian, Chief Economist at Google

(from New York Times, August 6, 2009)

#1 Most Desired Skill in Industry

"I believe over the next decade computing will become even more ubiquitous and intelligence will become ambient. The coevolution of software and new hardware form factors will intermediate and digitize — many of the things we do and experience in business, life and our world. This will be made possible by an ever-growing network of connected devices, incredible computing capacity from the cloud, insights from big data, and intelligence from machine learning."

-- Satya Nadella (CEO, Microsoft)

Email to all employees on first day as CEO (Feb. 04, 2014)

#1 Most Desired Skill in Academia

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

"Better understanding of probability"

Foundation for your future

But its not always intuitive

Zika Test

Positive Zika.

What is the probability of zika?

- *0.08% of people have zika*
- *90% positive rate for people with zika*
- *7% positive rate for people without zika*

The right answer is 1%

Probability = Important + Needs Study

Delayed gratification

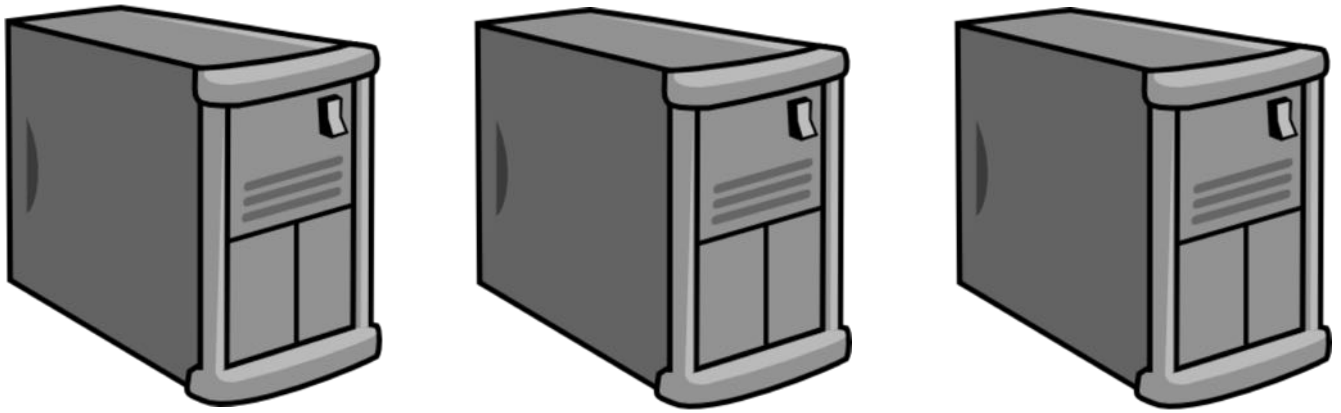
What is CS109?

Traditional View of Probability

CS View of Probability

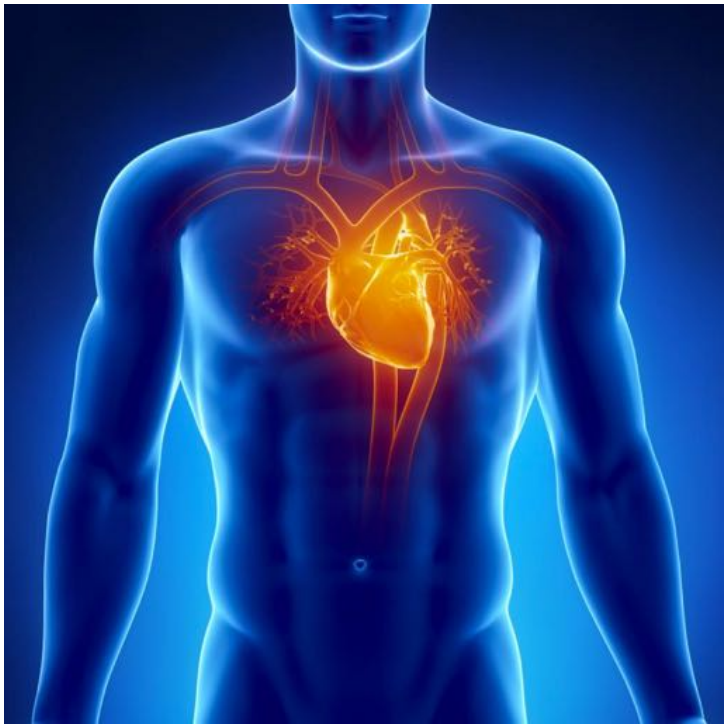
`http://www.site.com`

Give you the tools necessary to build and understand probabilistic CS algorithms.



CS View of Probability

Heart



Ancestry

Netflix

CS View of Probability

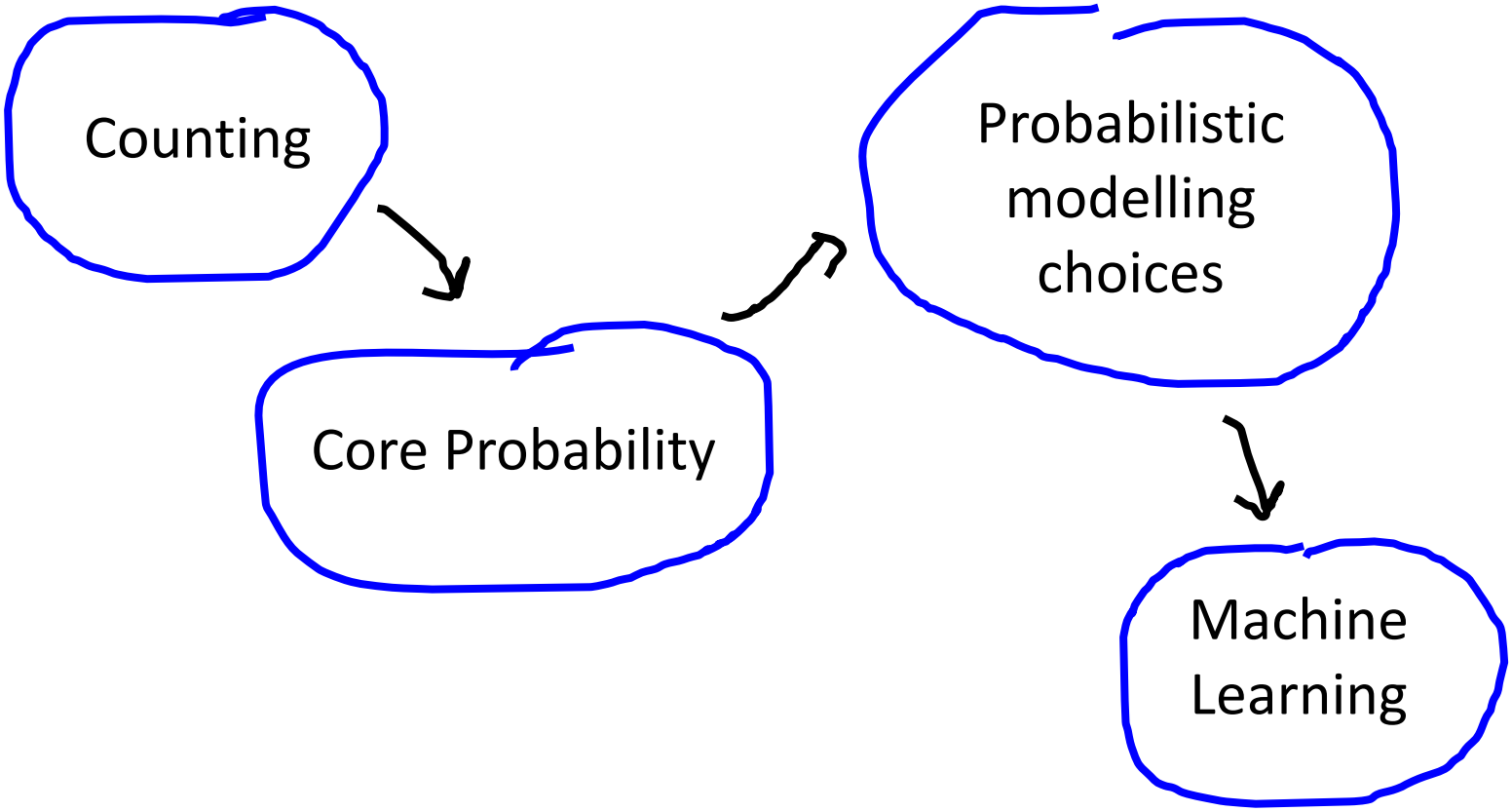
CS View of Probability

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

Lets dive in…

Counting

Our Route


Counting
Core Probability
Probabilistic modelling choices
Machine Learning