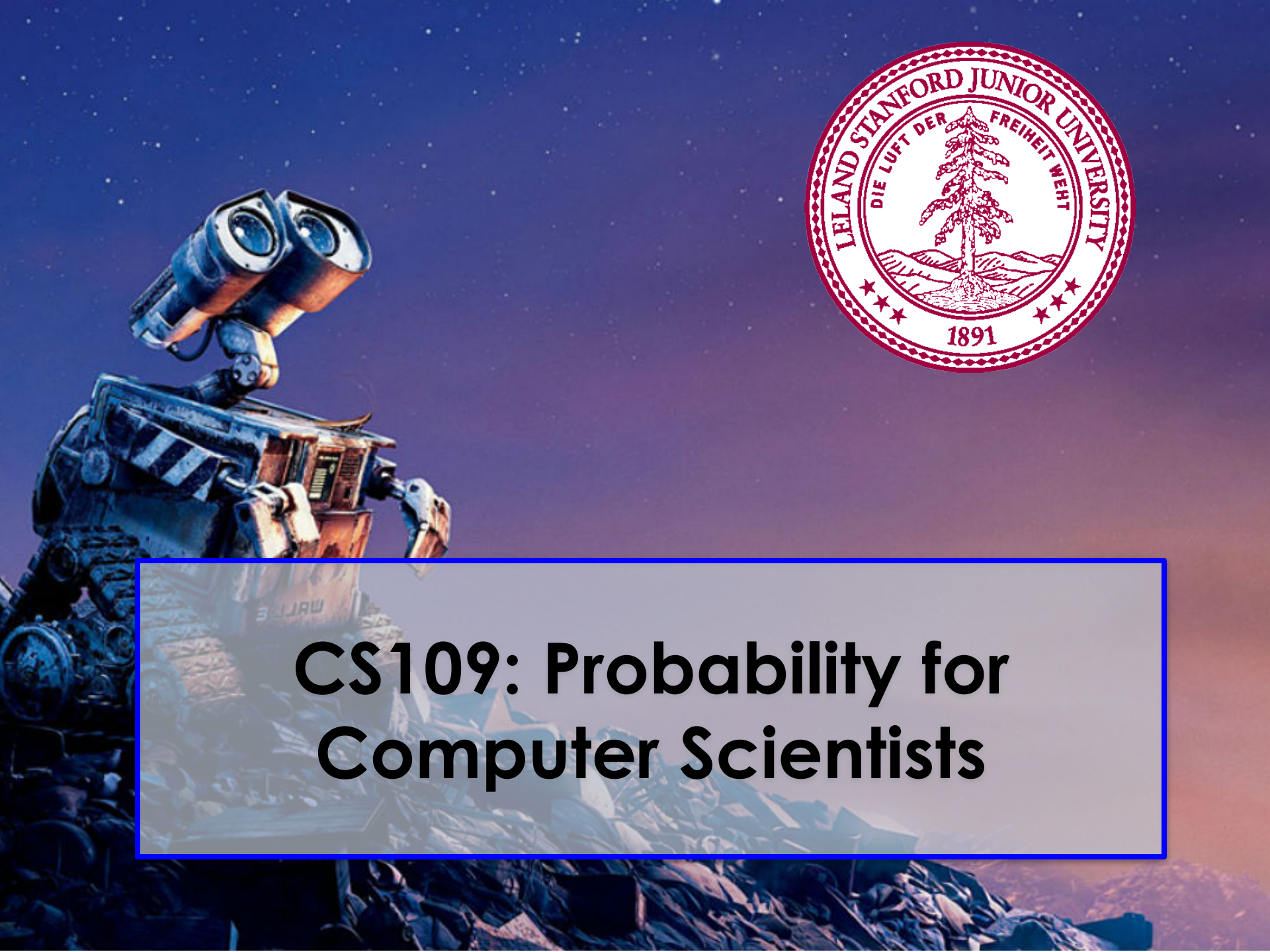


A background image of the WALL-E robot from the Pixar movie "WALL-E". The robot is a small, boxy, blue and orange machine with large, expressive eyes. It is standing on a rocky, desolate landscape under a dark, starry sky. The robot's body is covered in various mechanical details, including gears, pipes, and a small antenna. It has a small, rectangular body with a large, circular head and two large, round eyes. The robot is looking towards the right side of the frame.

# CS109: Probability for Computer Scientists

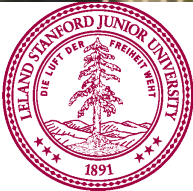
# Chris Piech

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

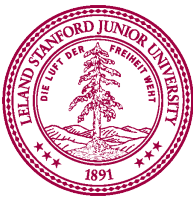




# I Took the First CS109 Class



Back when I looked like this 😊



# Teaching Assistants

Several outstanding TAs:

Ben Ulmer (Head TA)

Elliott Chartock,

Brendan Corcoran,

Julia Daniel,

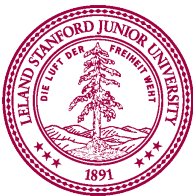
Gobi Dasu,

Ana-Maria Istrate,

Luke Johnston

Yuling Liu,

Eric Redondo



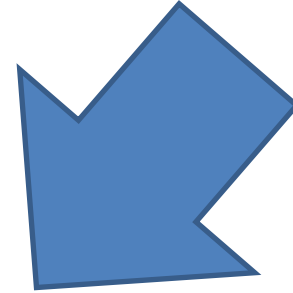
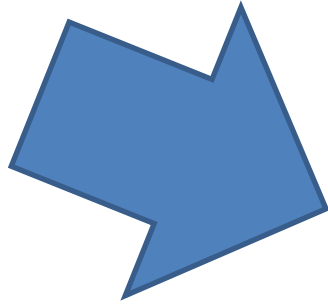
Who are you?



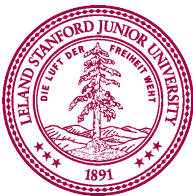
# Course mechanics

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

# Essential Information



[cs109.stanford.edu](https://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)



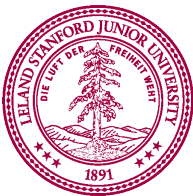


# Coding in CS109



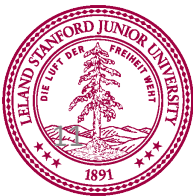
Review session on Friday

Piech, CS106A, Stanford University



# Staff Contact

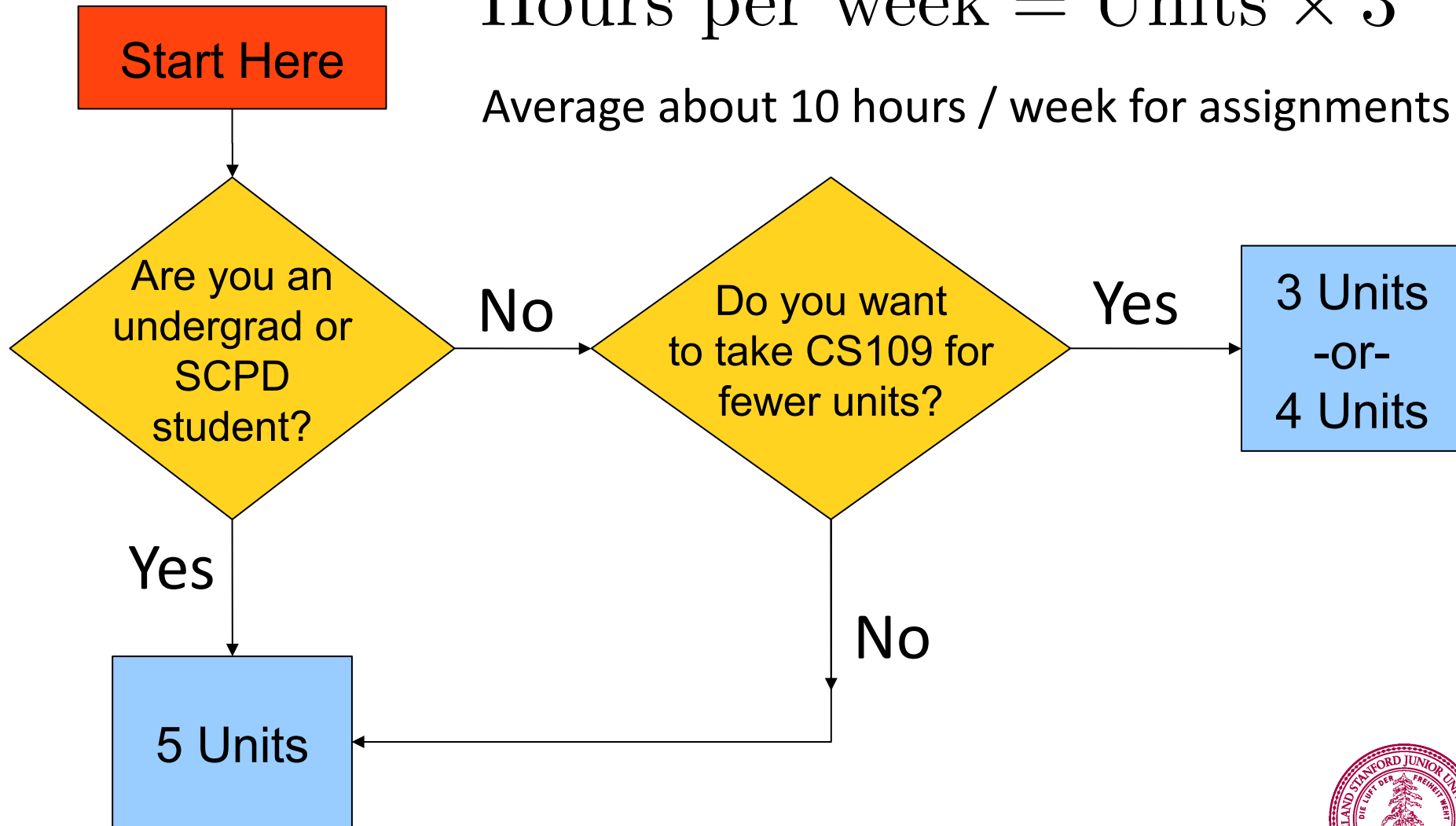
- Post to Piazza for clarification
- Go to Working Office Hours
- Email [cs109@cs.stanford.edu](mailto:cs109@cs.stanford.edu)
- Email Chris or go to his office for course level issues.



# CS109 Units

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

Average about 10 hours / week for assignments



# Not Videotaped



- \* But we are going to look into it
- \*\* Capturing lecture participation

# Class Breakdown

**45%**

**6 Assignments**

**20%**

**Midterm**

Tuesday May 9<sup>th</sup>, 7-9pm

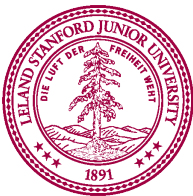
**30%**

**Final**

Tuesday June 13<sup>th</sup>, 8:30-11:30am

**5%**

**Section Participation**



# Late Days

2



# The Student Honor Code





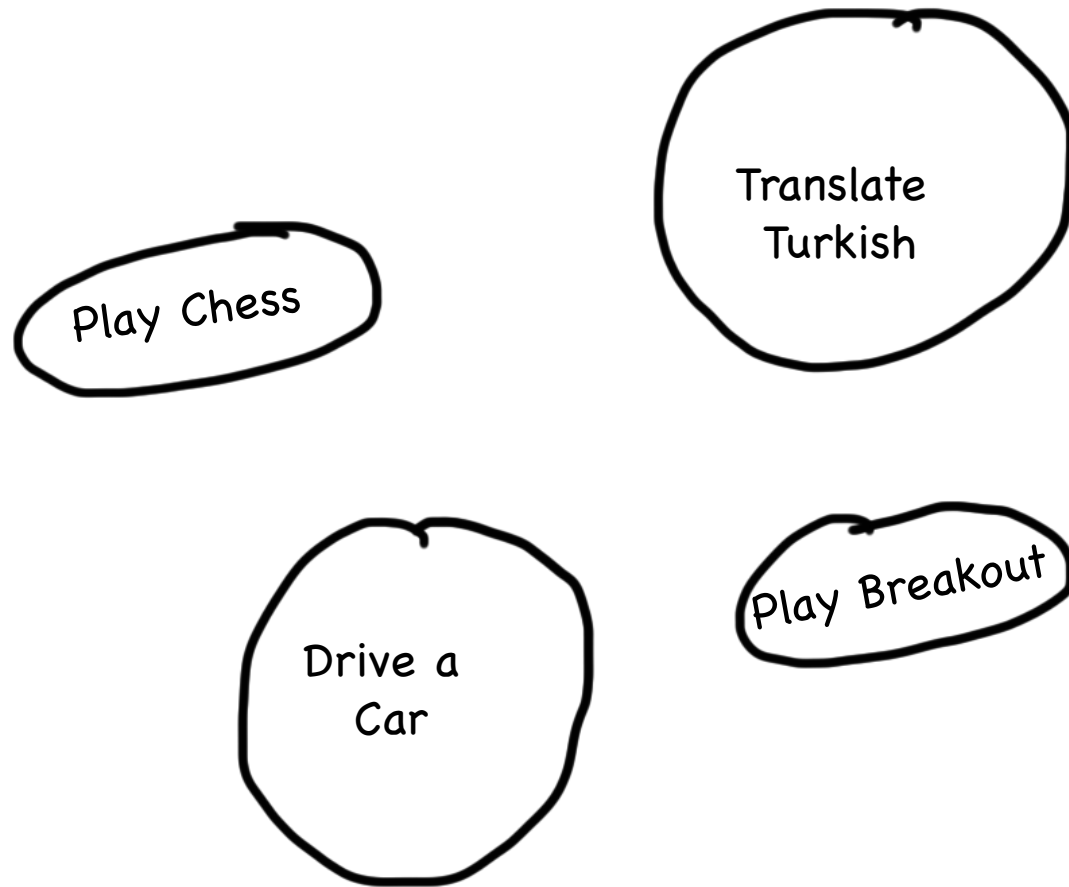
# Story of Modern AI

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

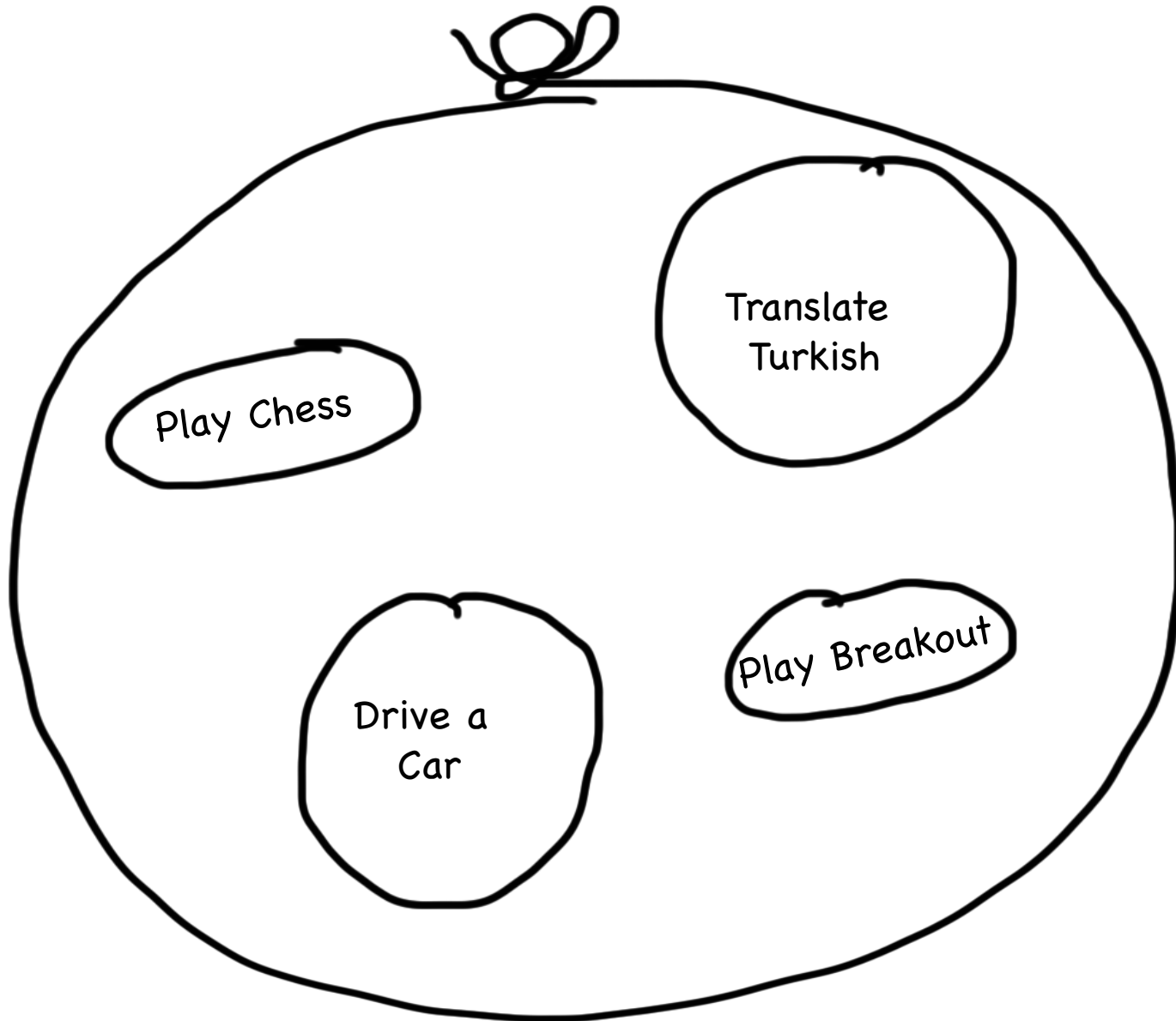
# Brief History



# Narrow Intelligence



# General Intelligence

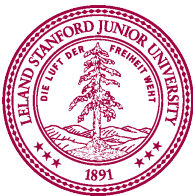
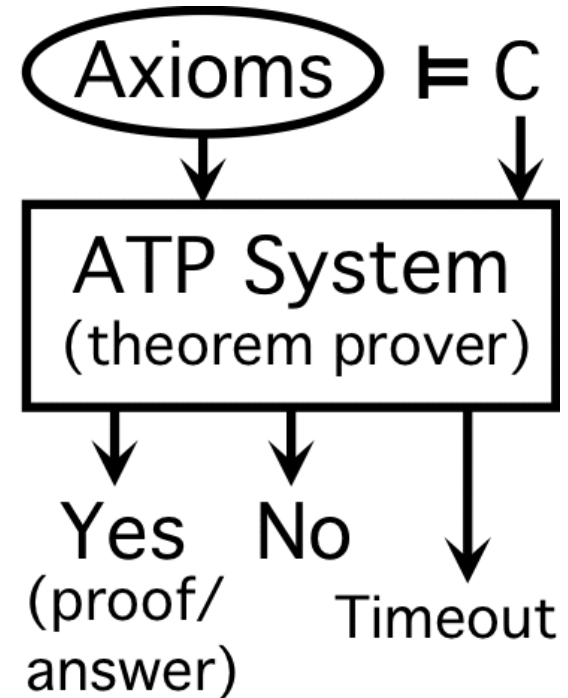


# Early Optimism 1950

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 spirit is willing but the flesh is weak.*

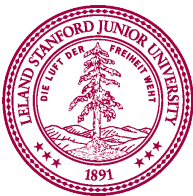


(Russian)



*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

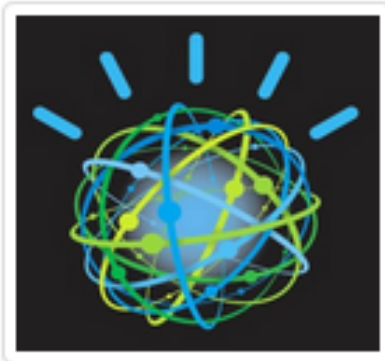
# Big Milestones Pt 1



1997 Deep Blue

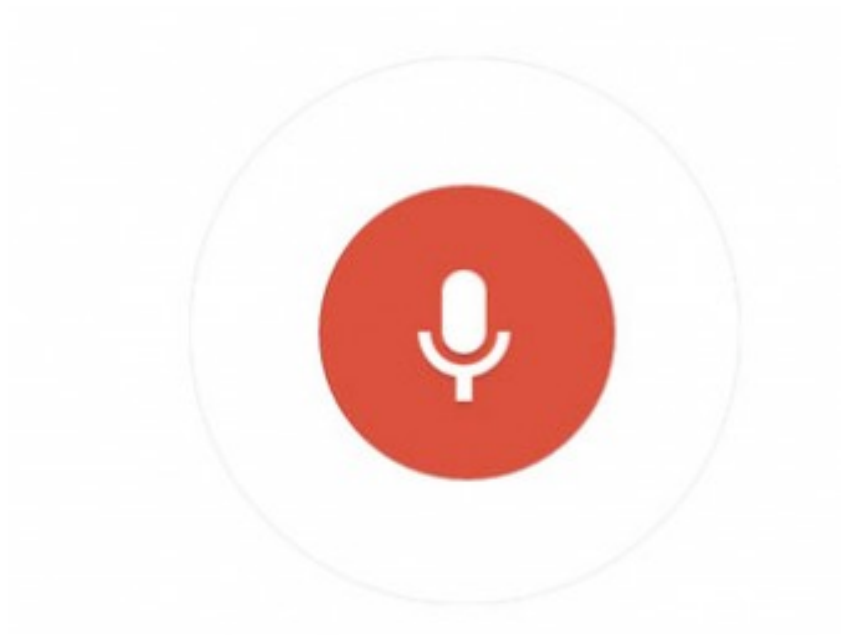


2005 Stanley

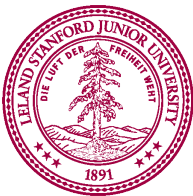


2011 Watson

# 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



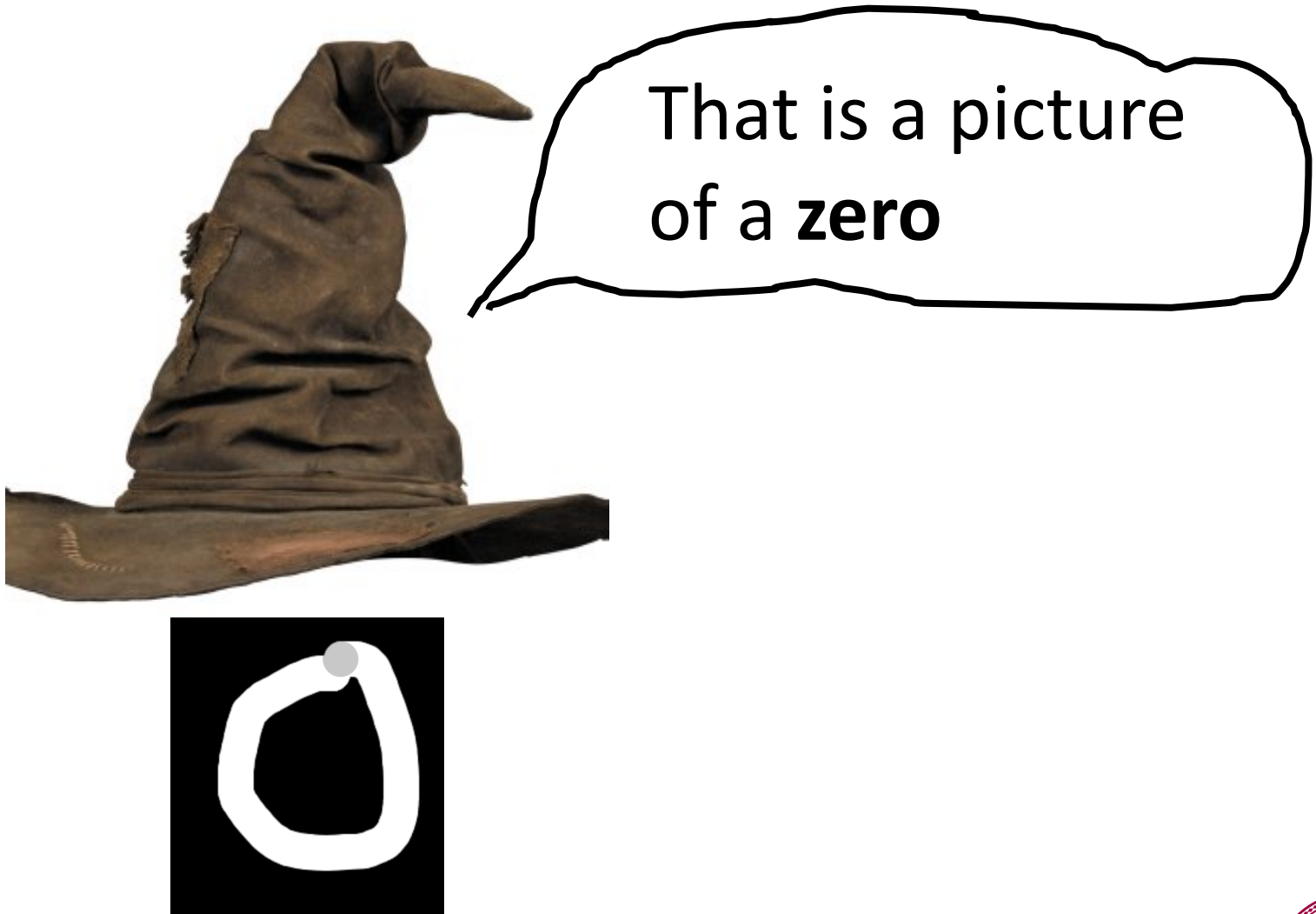
# Classification



That is a picture  
of a **one**



# Classification



# Classification



That is a picture  
of an **zero**

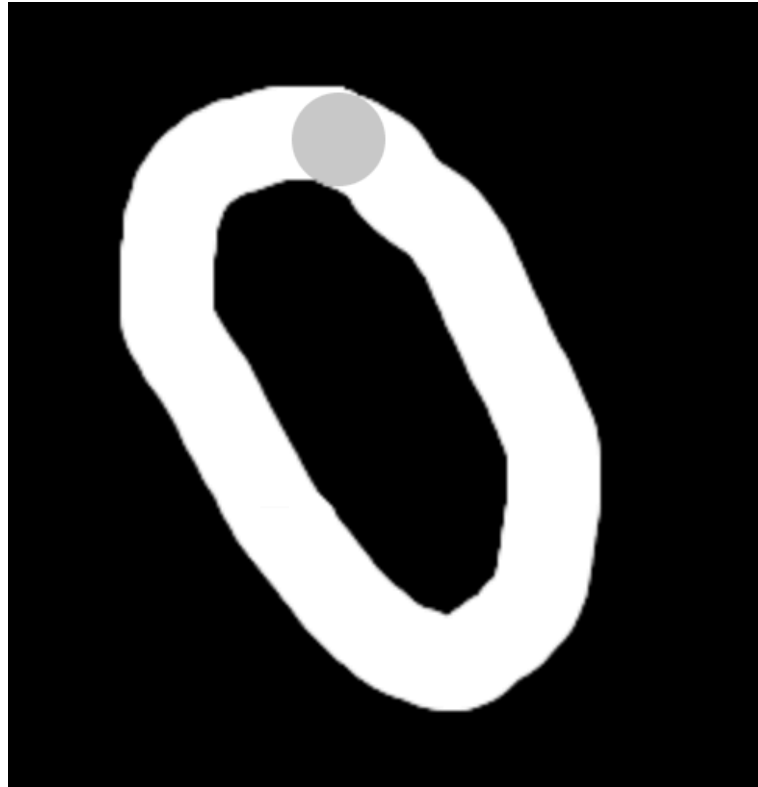


\* It doesn't have to be  
correct all of the time

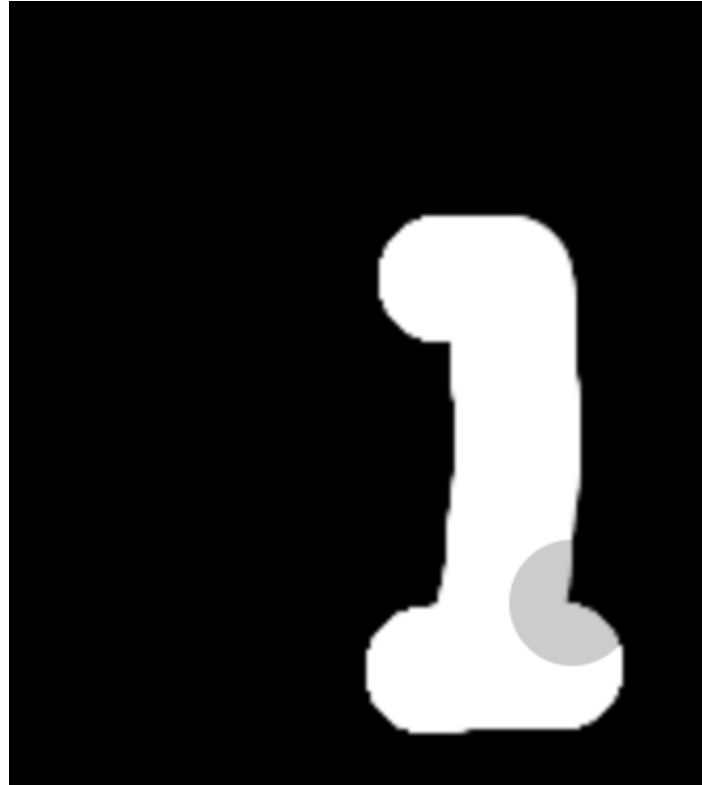


Can you do it?

# 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**

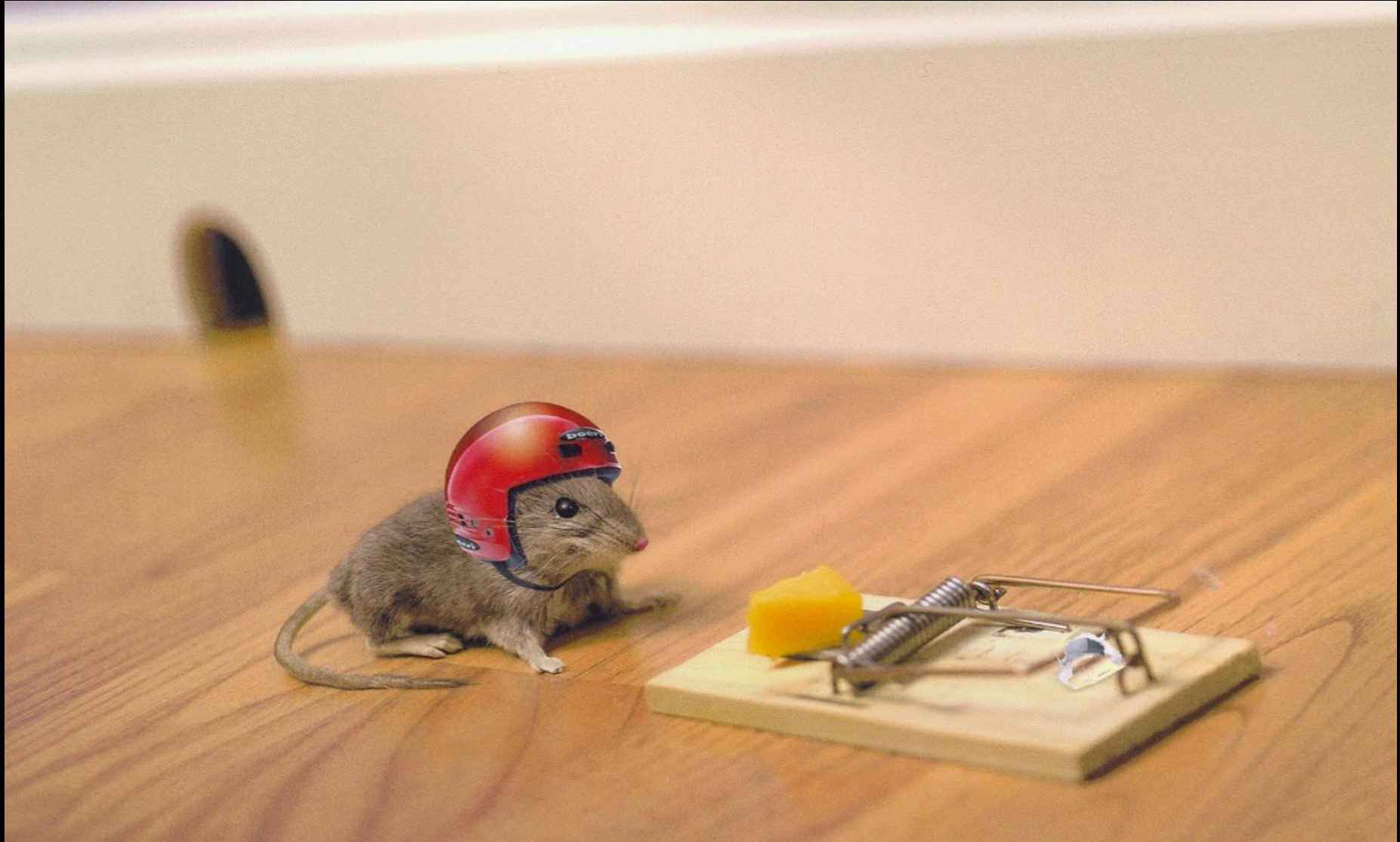
**2. Artificial Neurons**

# Two Great Ideas

**1. Probability from Examples**

**2. Artificial Neurons**

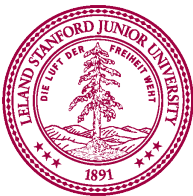
# 1. Probability From Examples





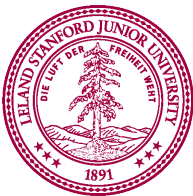
# When Does the Magic Happen?

Lots of  
Data + Sound  
Probability



# Machine Learning

Basically just a rebranding of statistics  
and probability.



# Vision is Hard

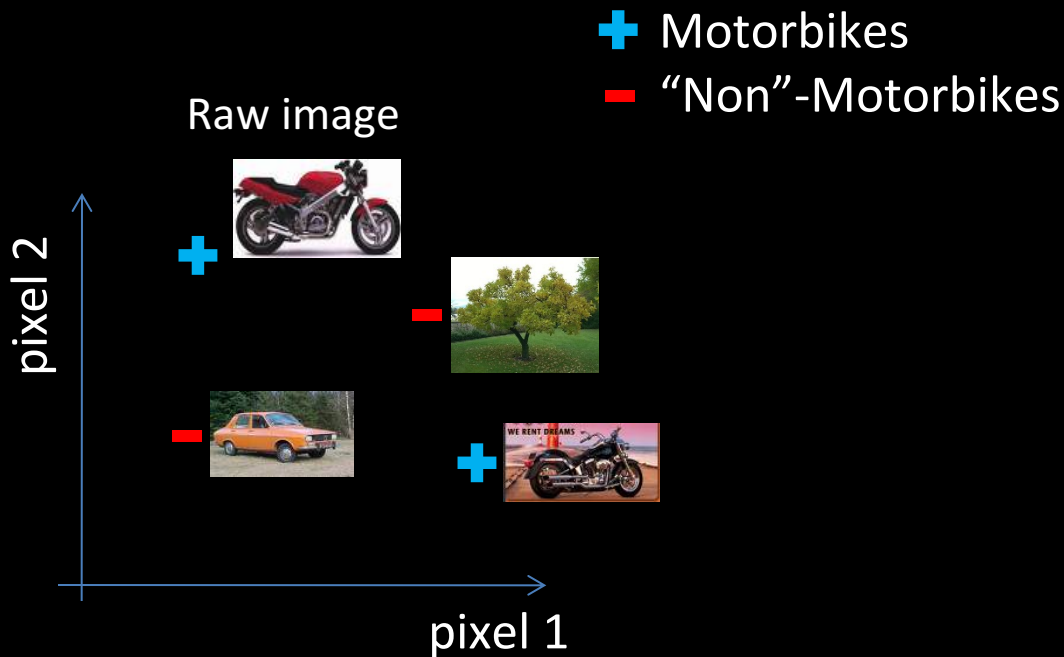
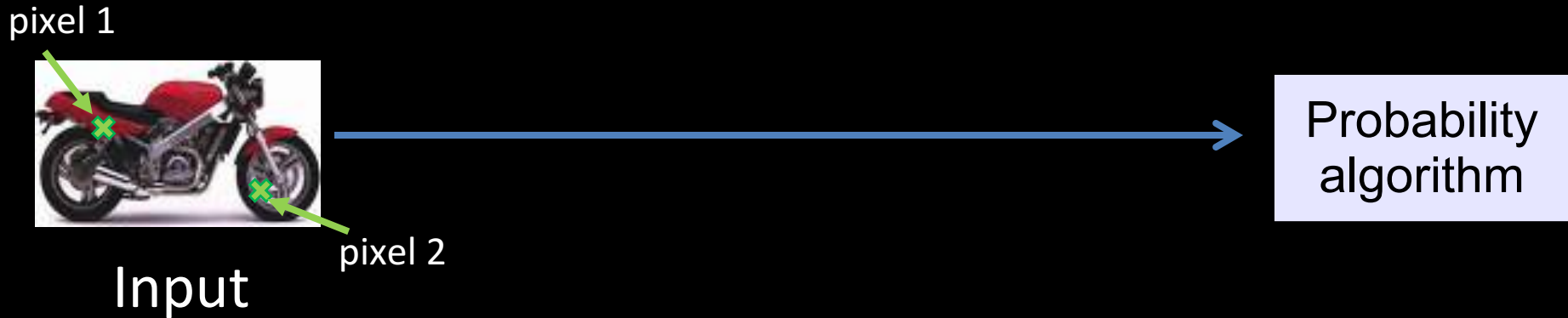
You see this:



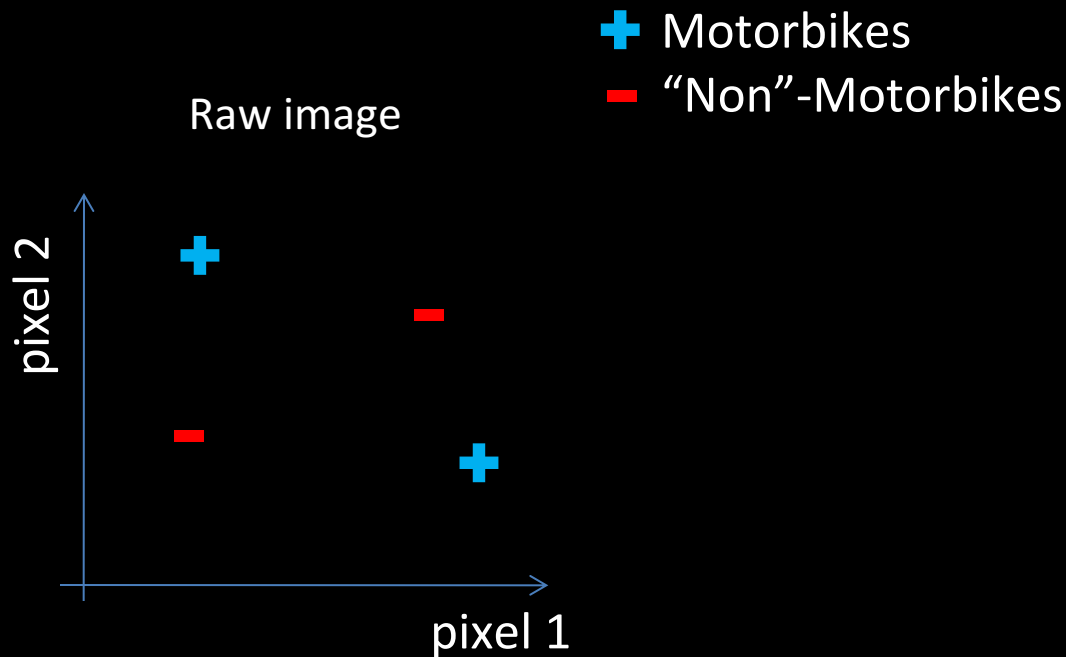
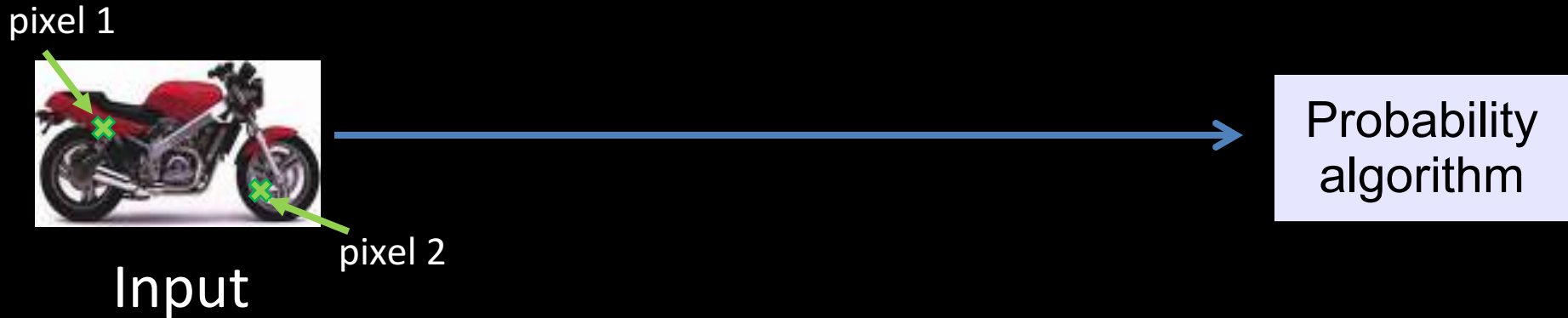
But the camera sees this:

|     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 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  |

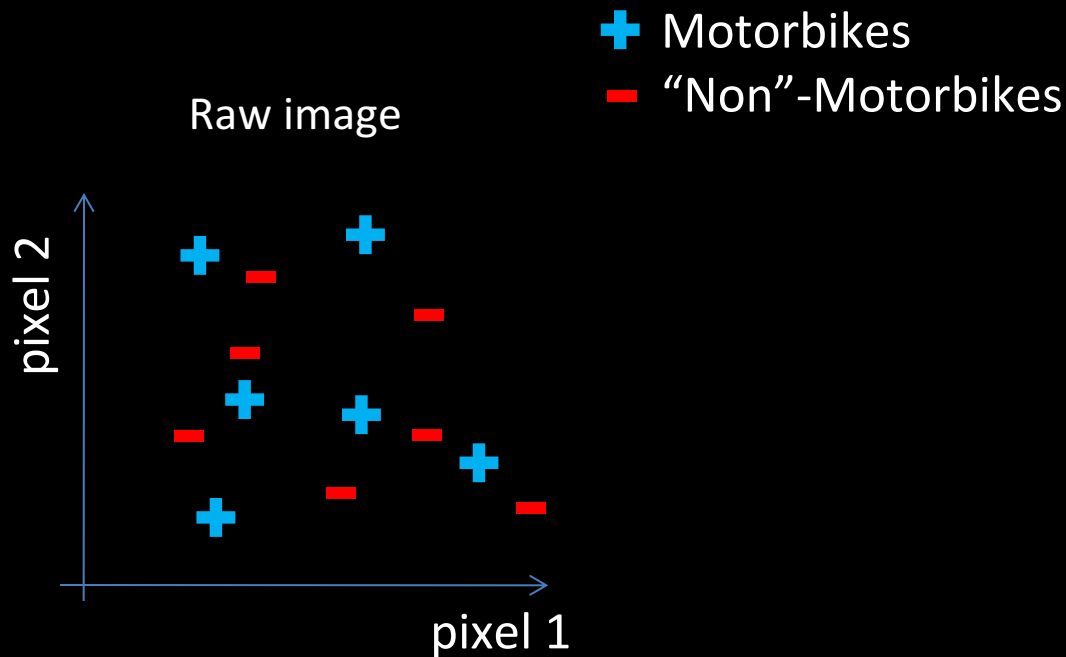
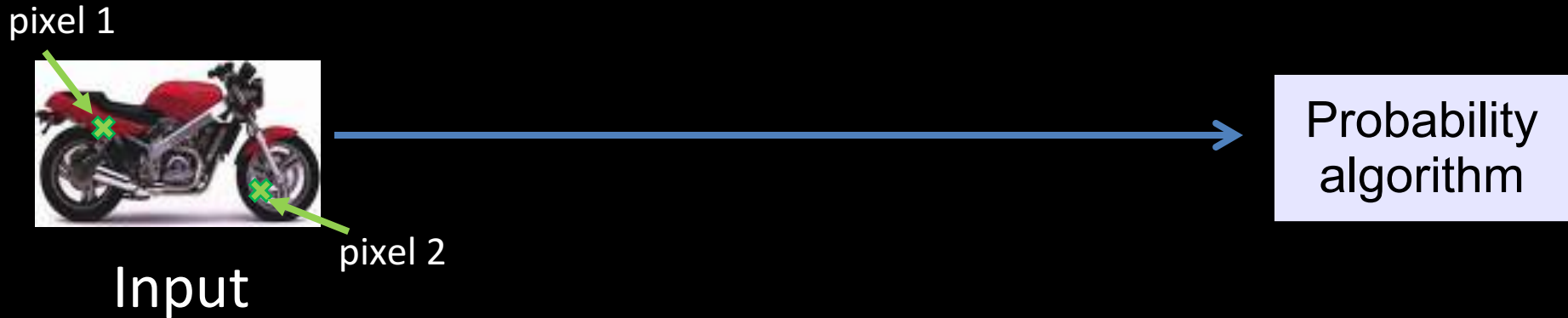
# Machine Learning



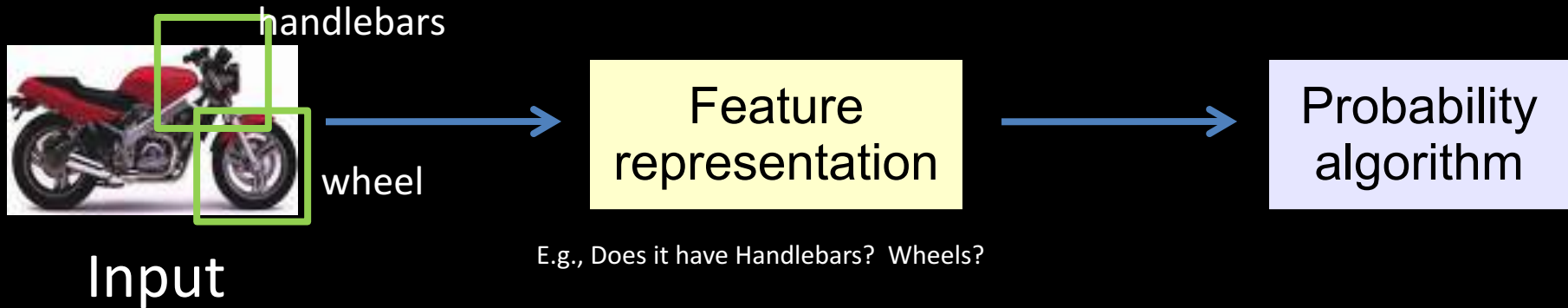
# Machine Learning



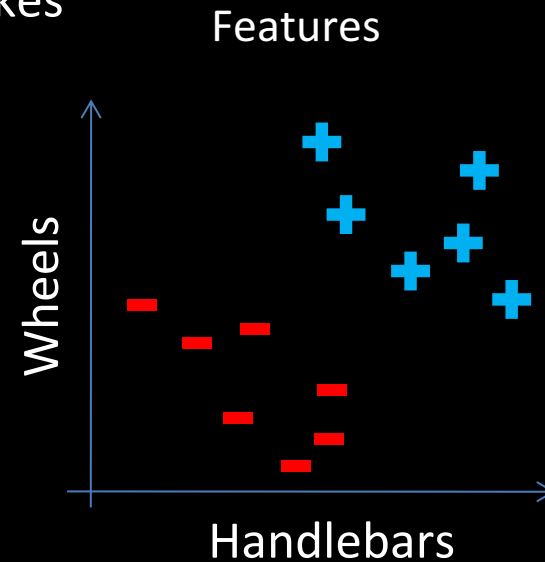
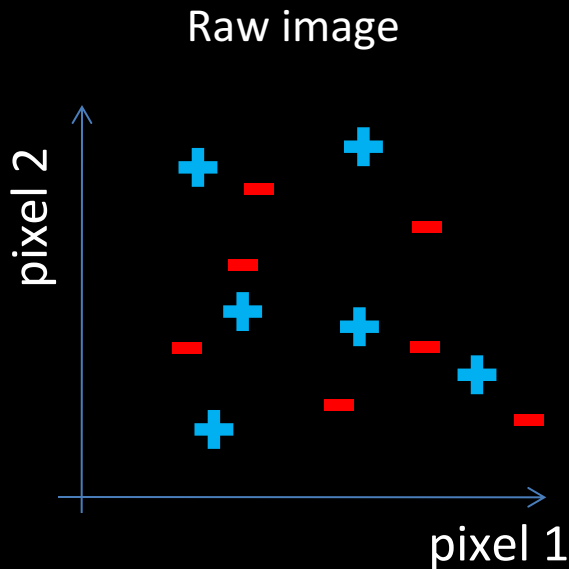
# Machine Learning



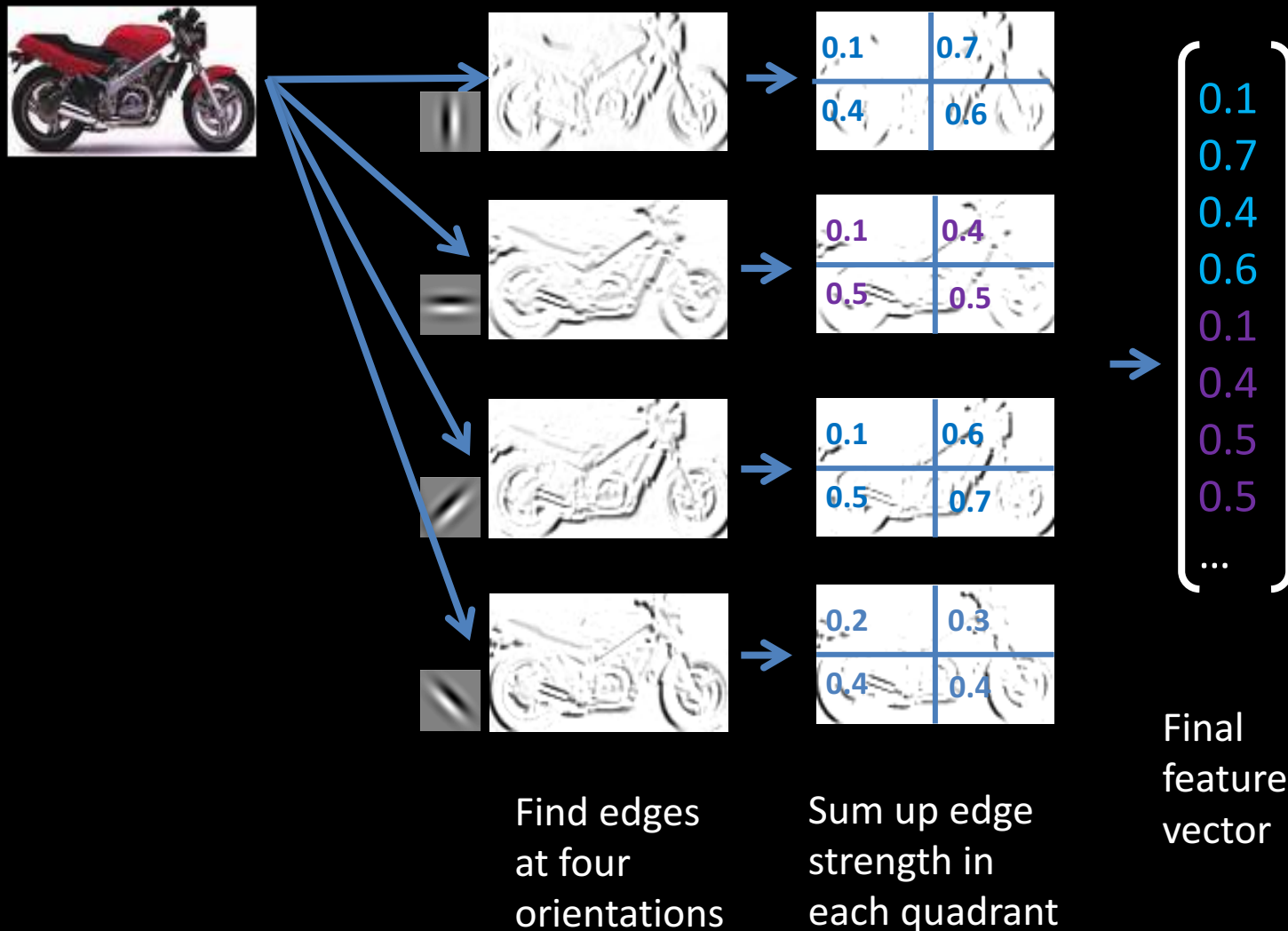
# Machine Learning



+ Motorbikes  
- "Non"-Motorbikes



# Human Designed Features



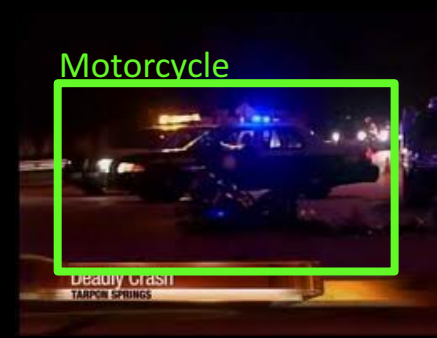


# Some Great Thinkers



Daphne Koller

# Straight ML Not Perfect...



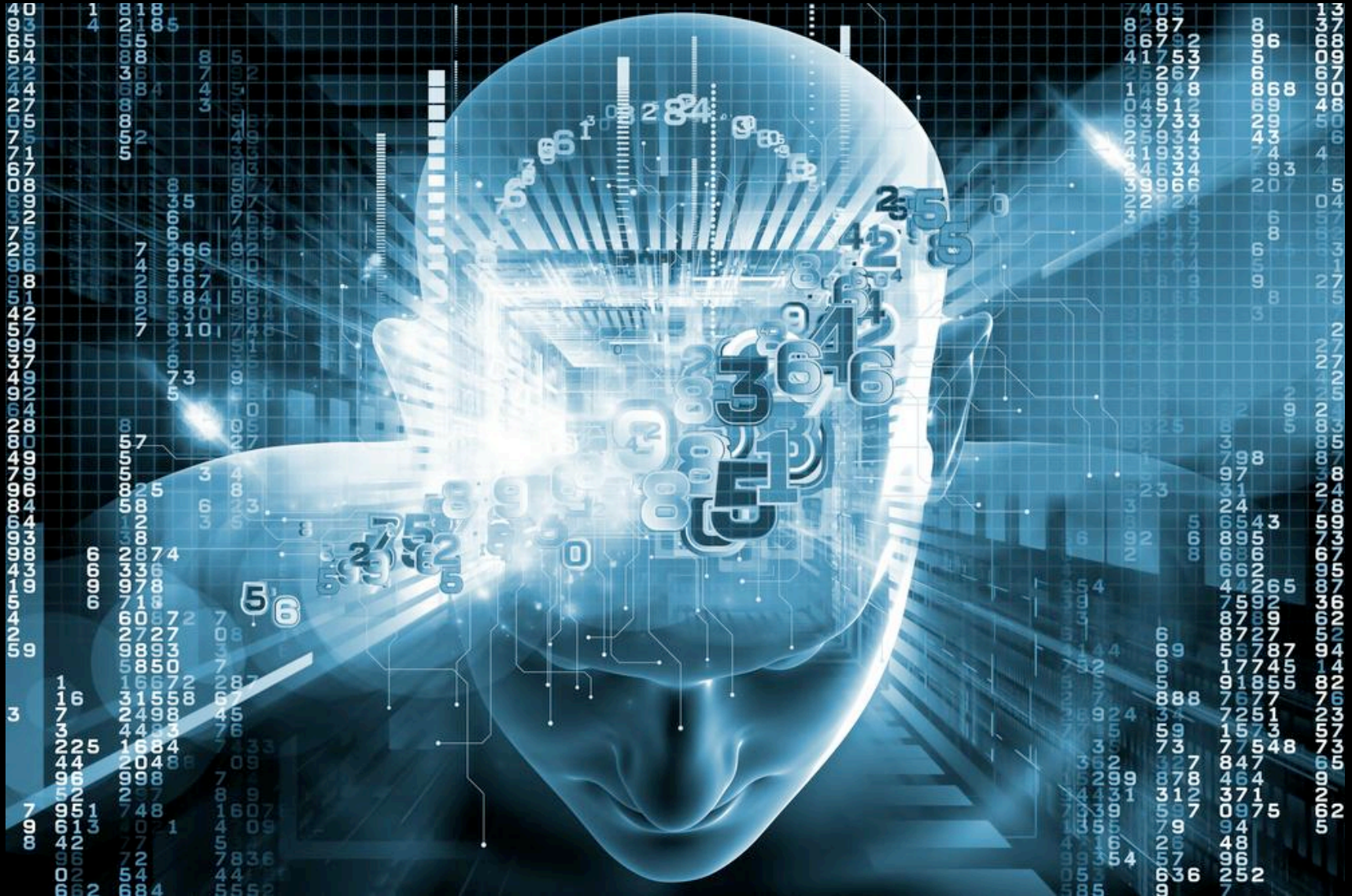
# Two Great Ideas

**1. Probability from Examples**

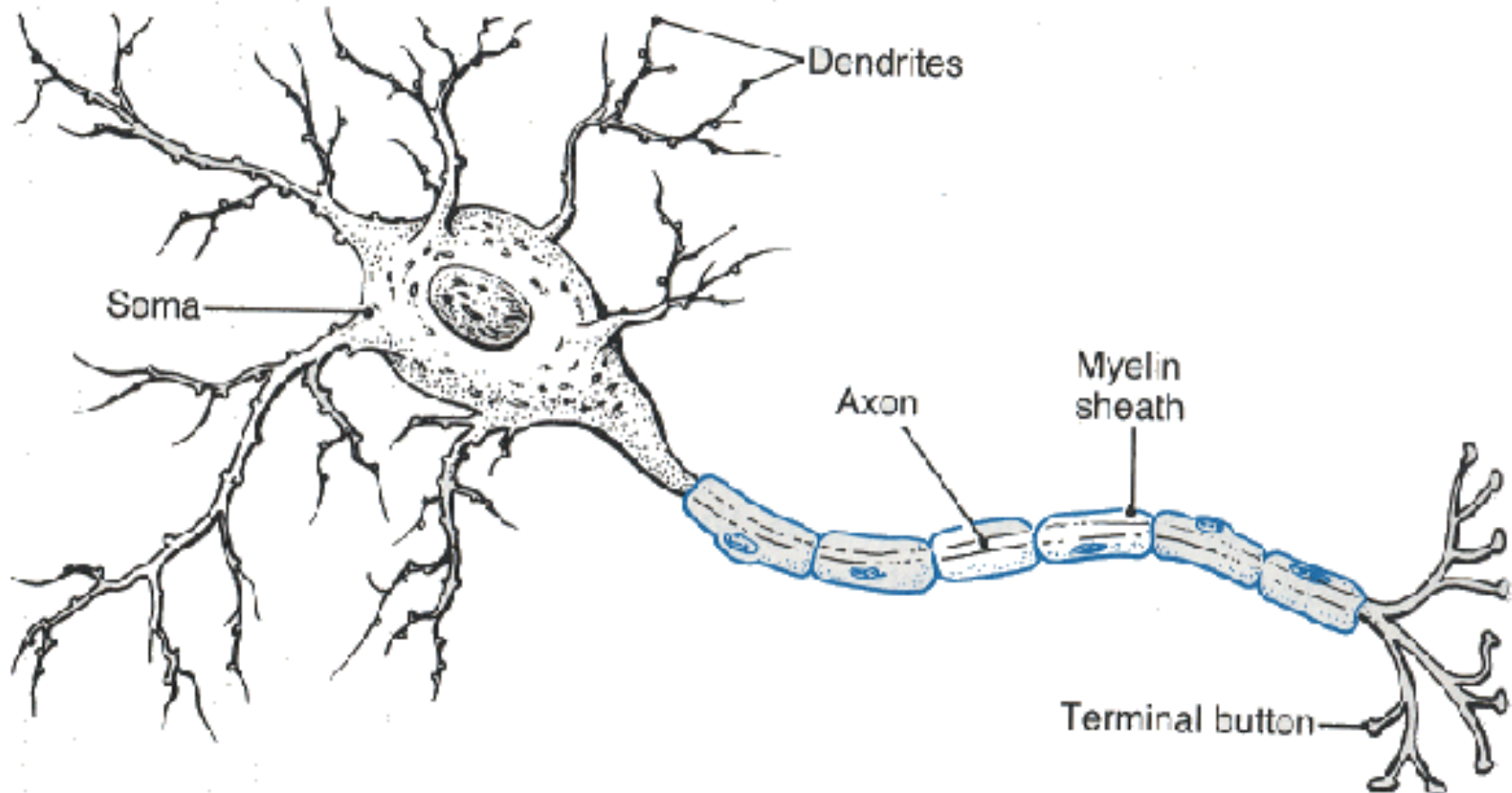
**2. Artificial Neurons**



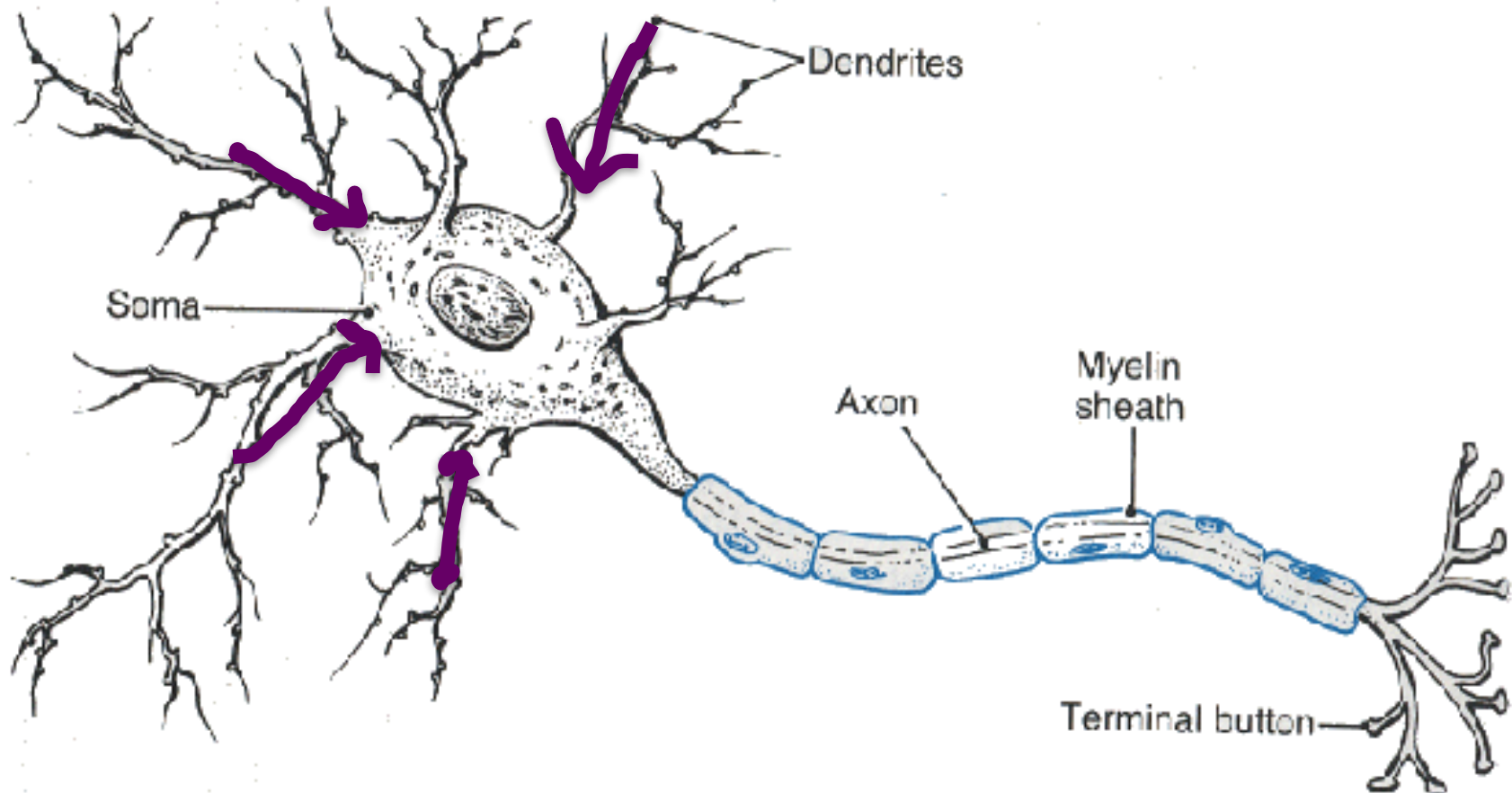
## 2. Artificial Neurons



# Neuron

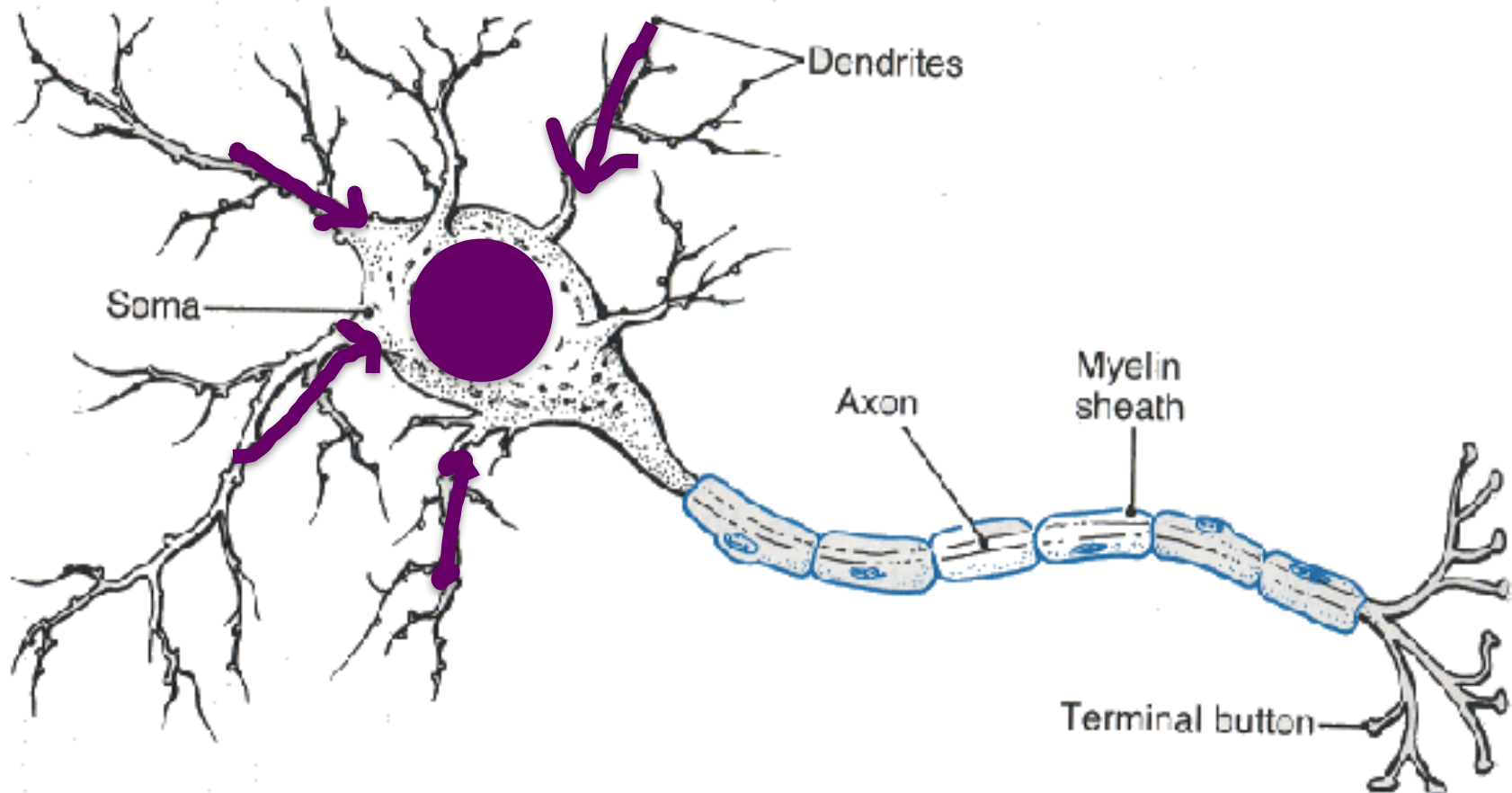


# Neuron

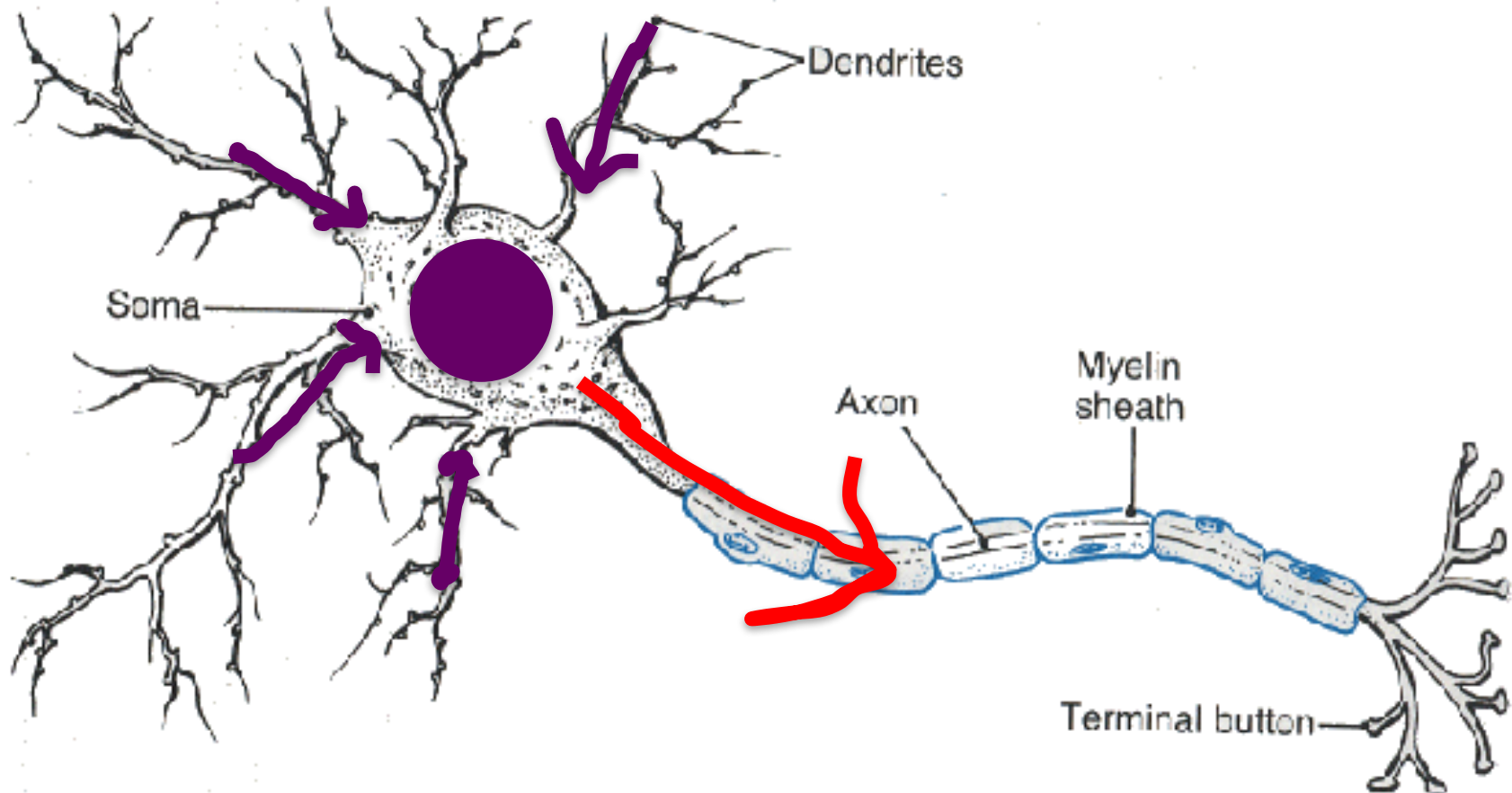




# Neuron

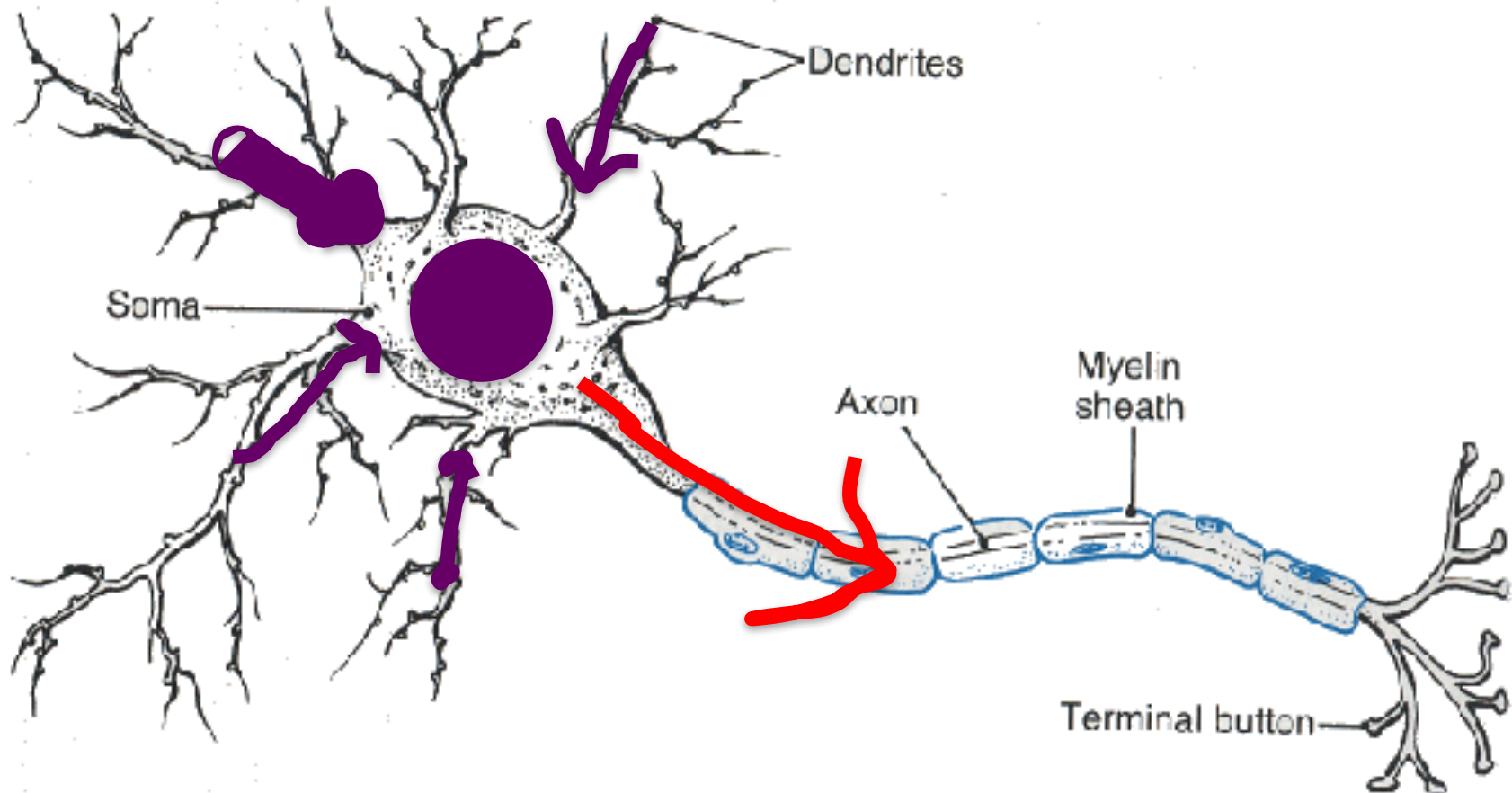


# Neuron

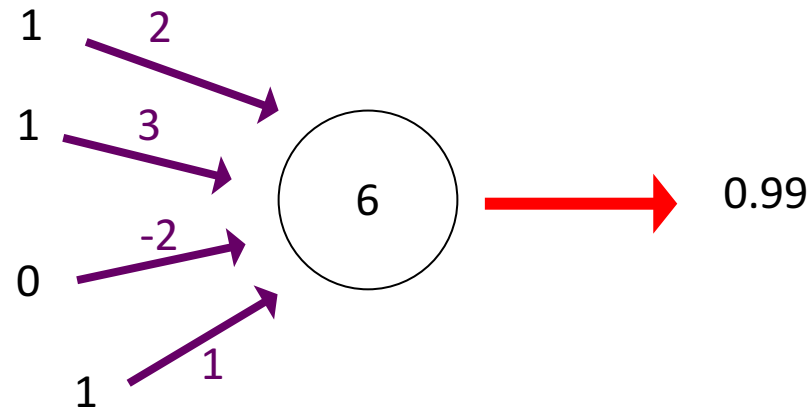
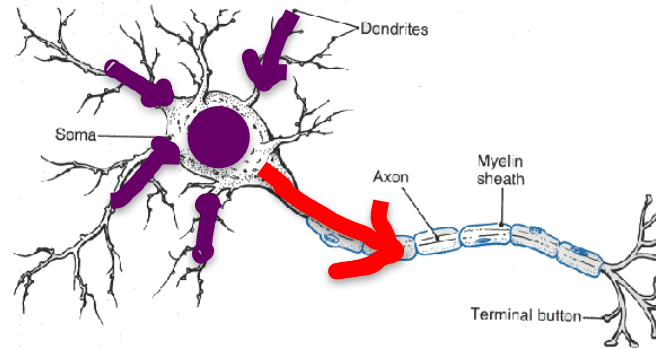




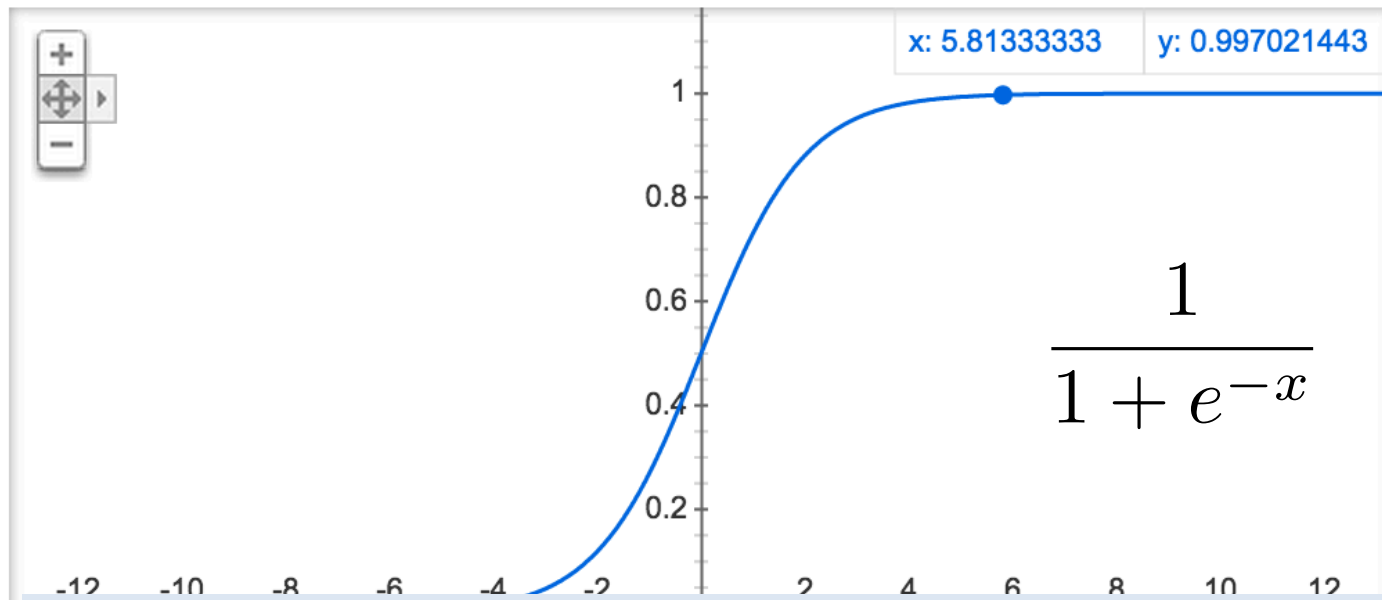
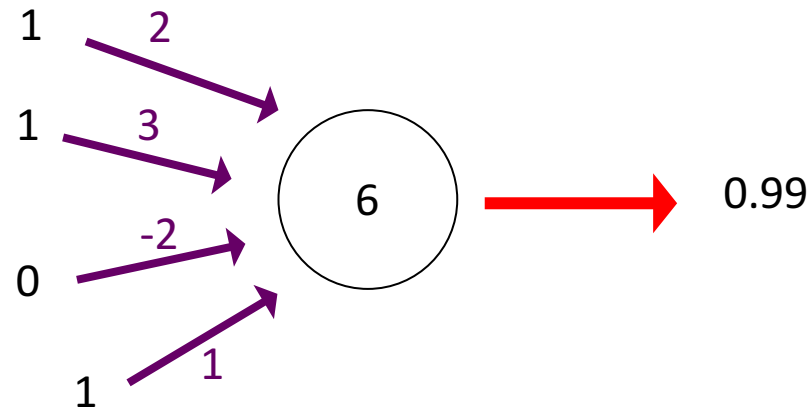
# Some Inputs are More Important



# Artificial Neuron



# Sigmoid Function

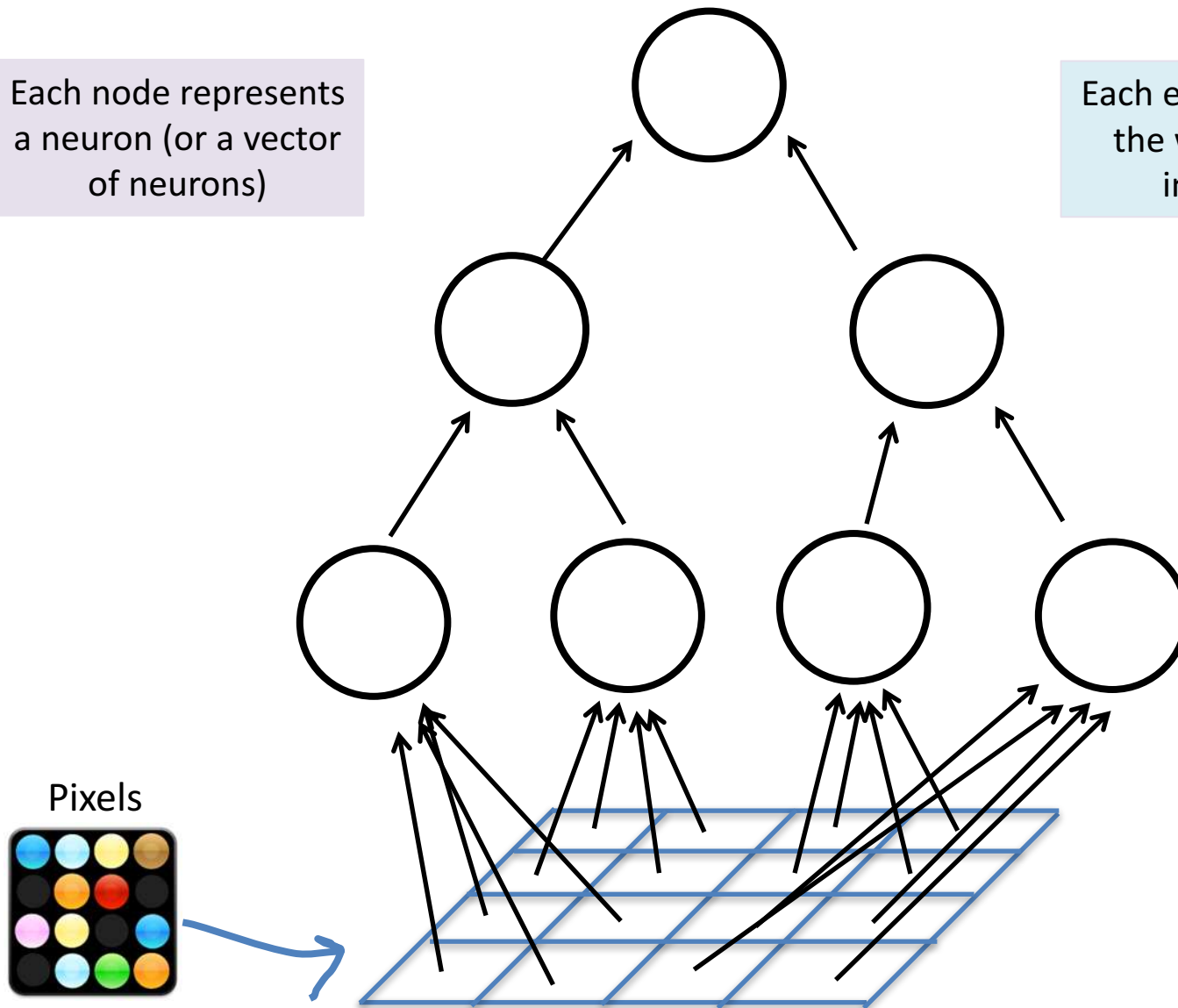


An artificial neuron is like a little probability calculator

# Neural Network

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

Each edge represents  
the weight of the  
interaction

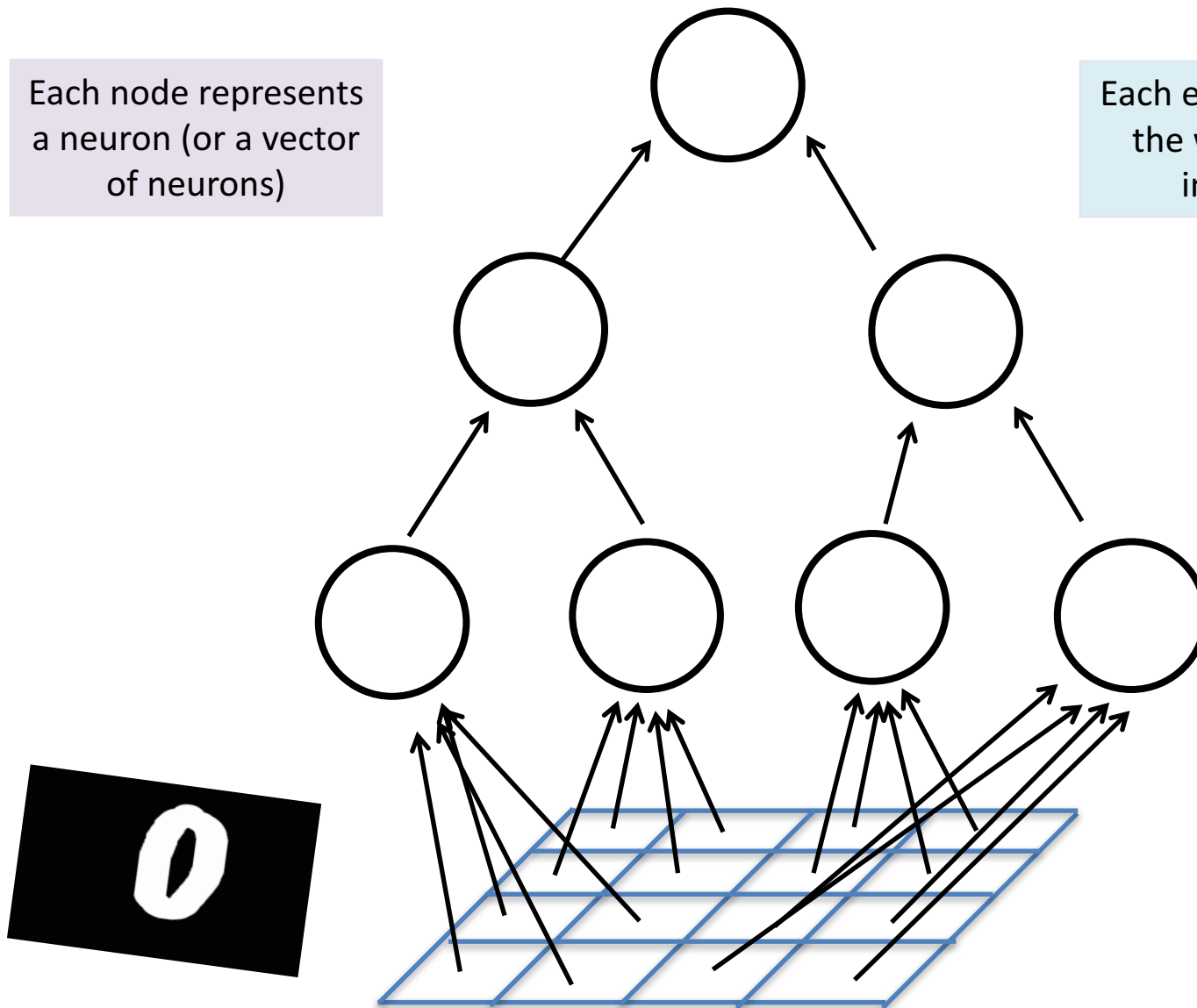


Forward Pass...

# Forward Pass

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

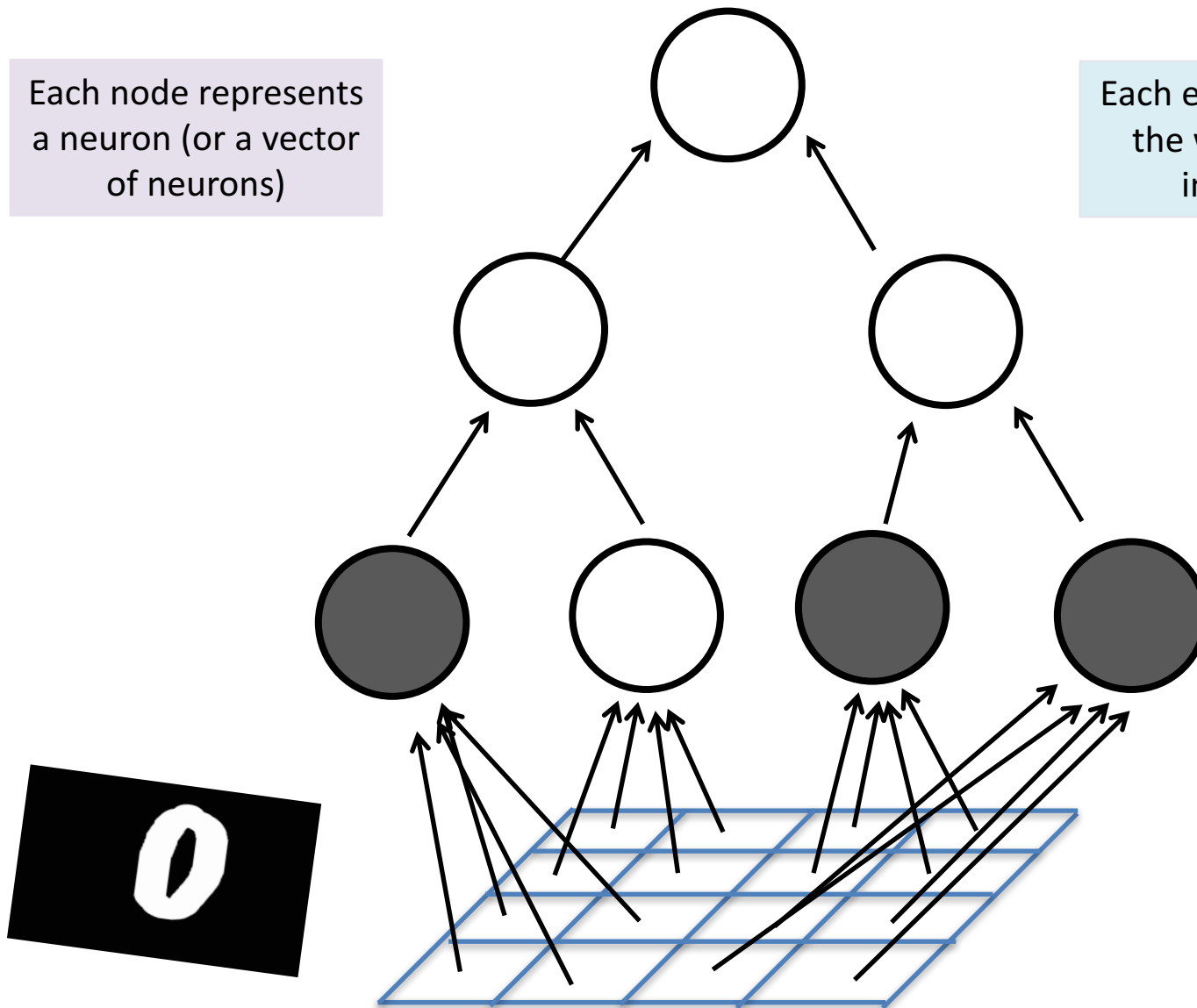
Each edge represents  
the weight of the  
interaction



# Forward Pass

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

Each edge represents  
the weight of the  
interaction

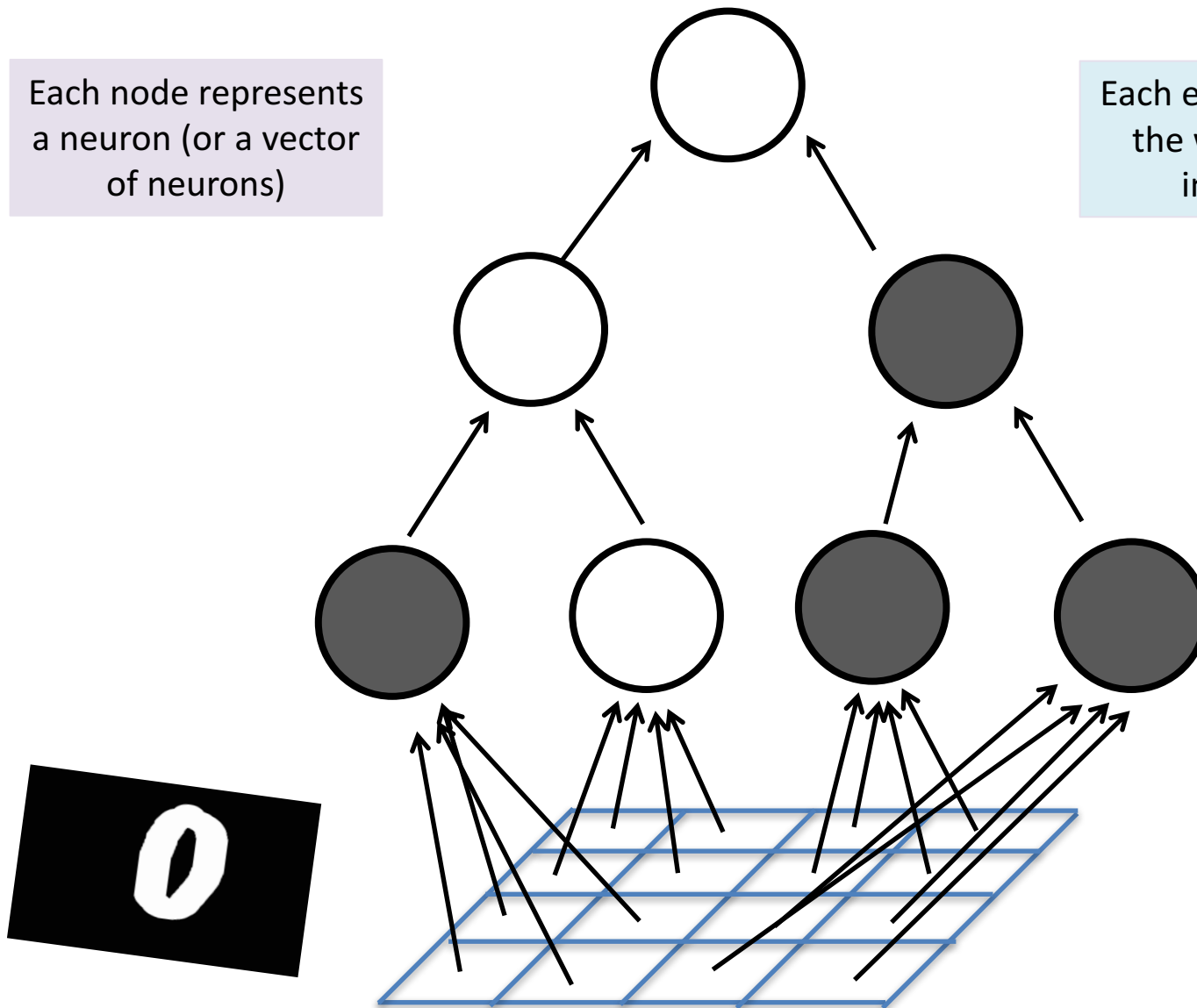




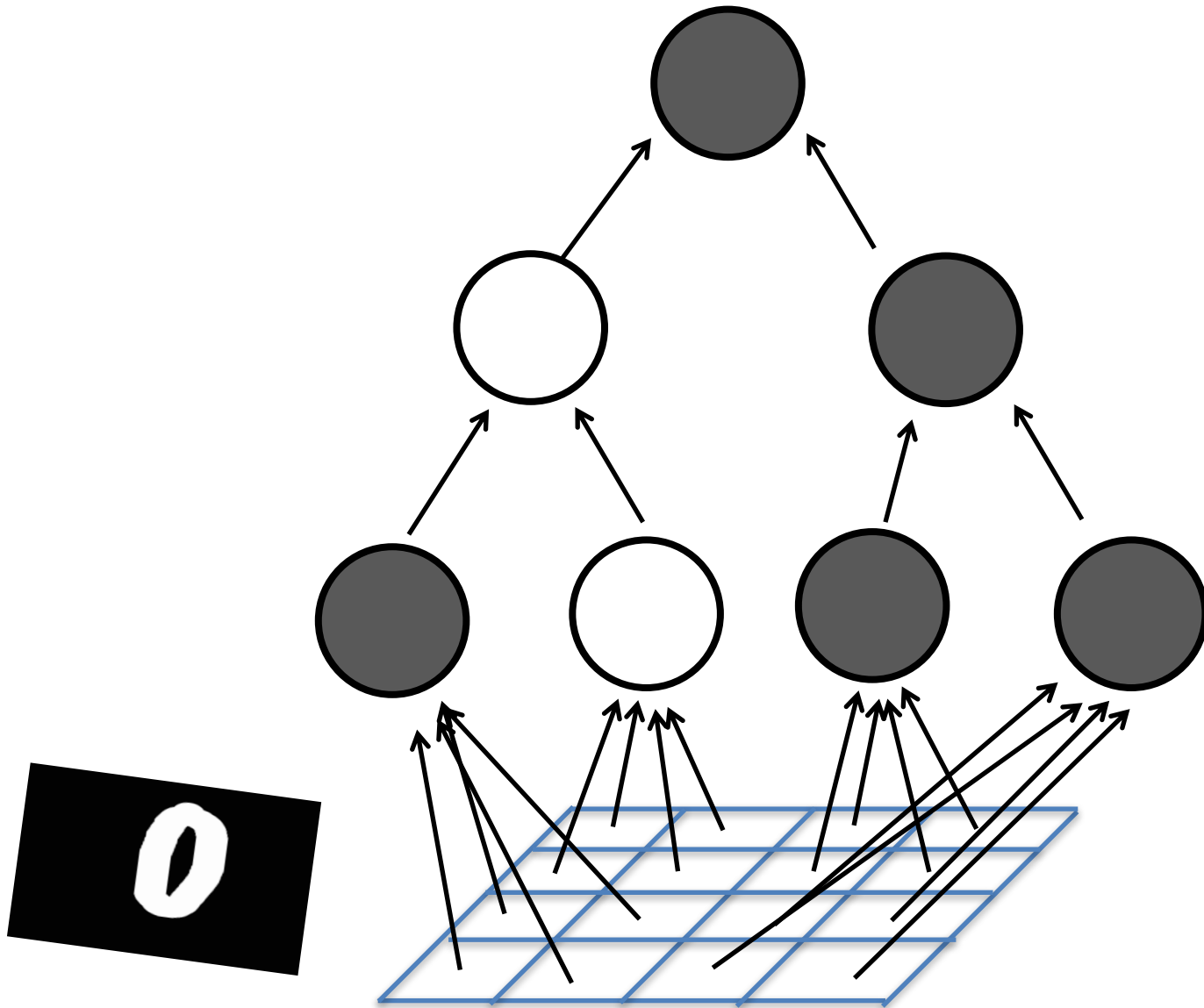
# Forward Pass

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

Each edge represents  
the weight of the  
interaction

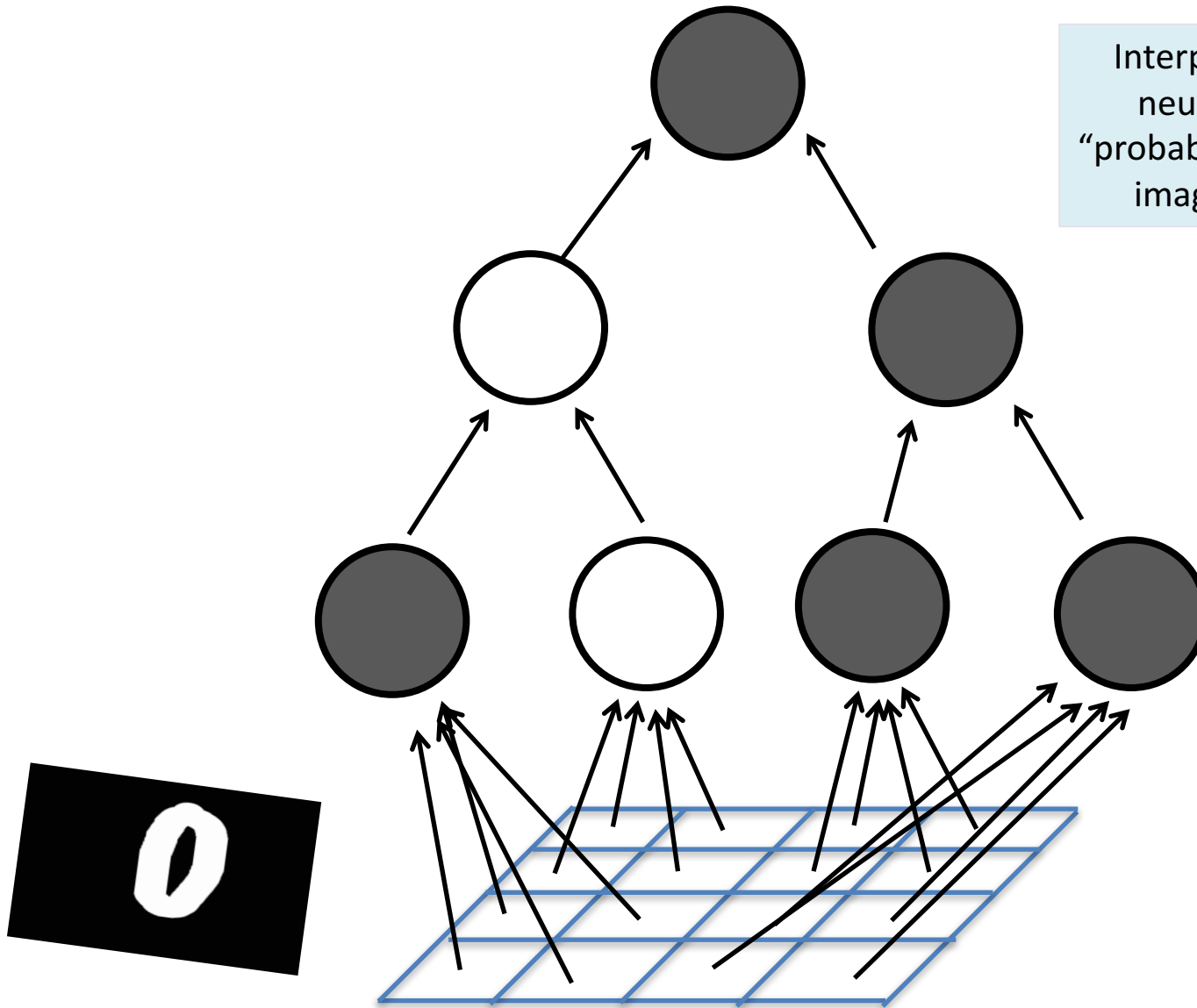


# Forward Pass

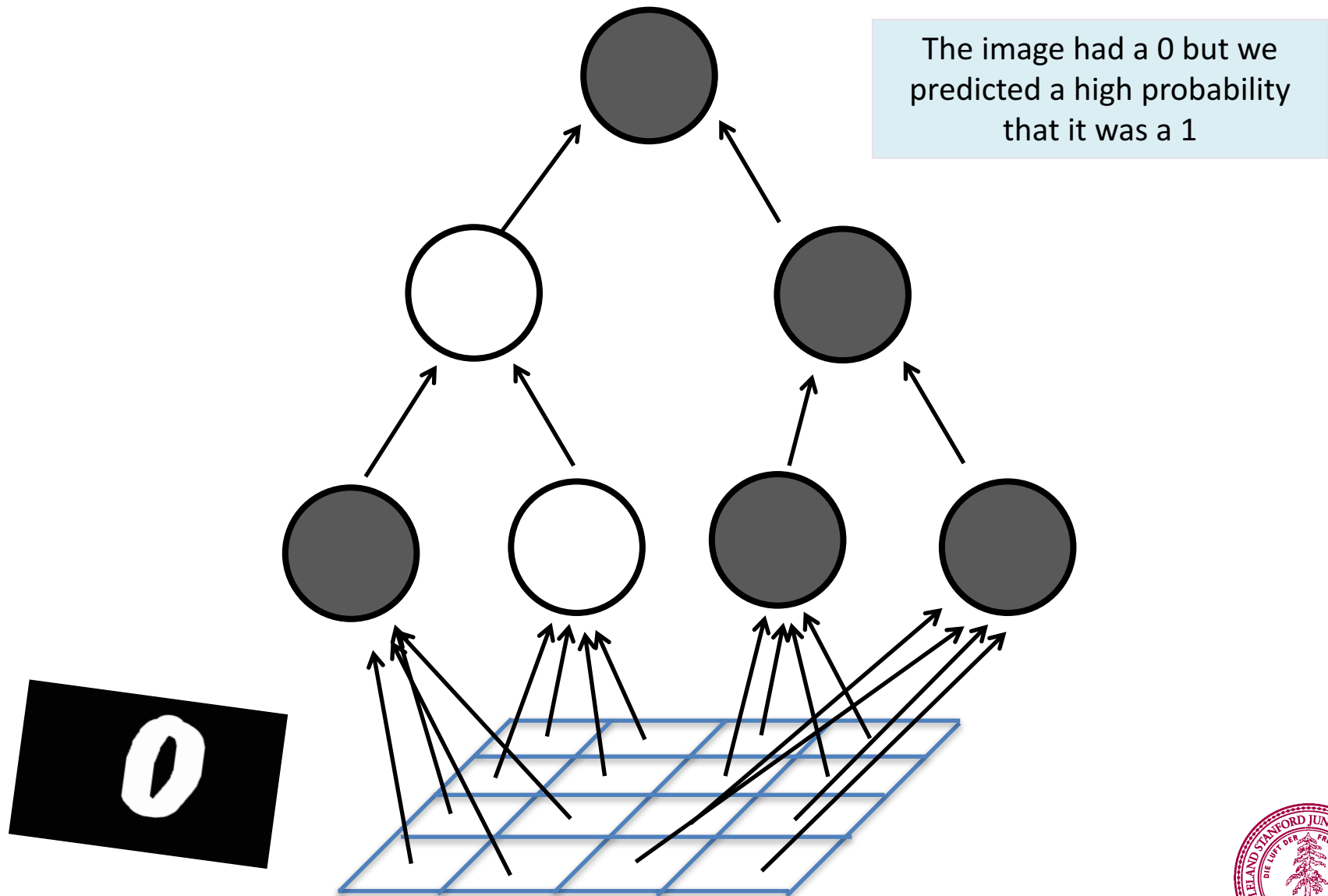


# Forward Pass

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



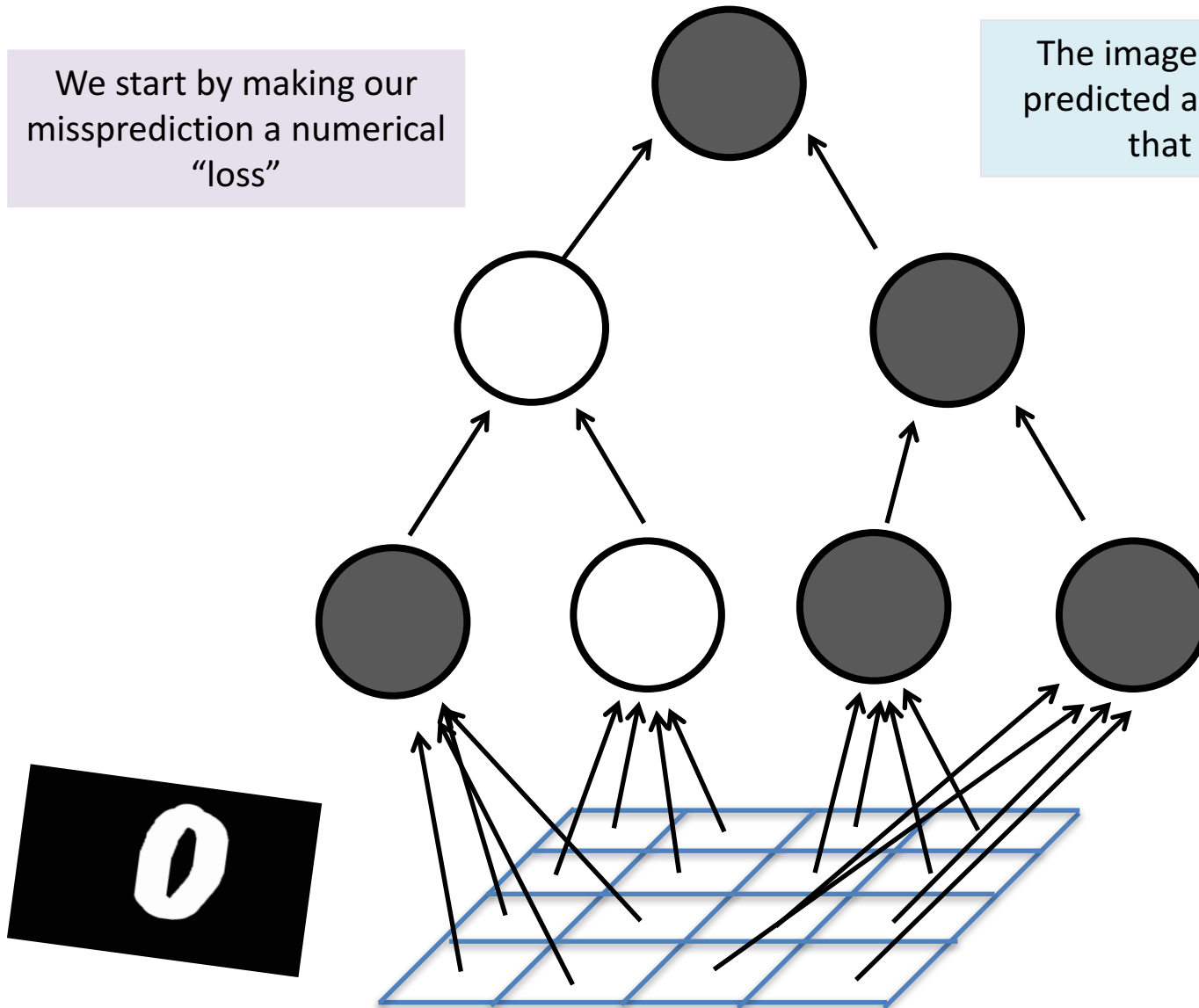
# Backward Pass



# Backward Pass

We start by making our missprediction a numerical “loss”

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

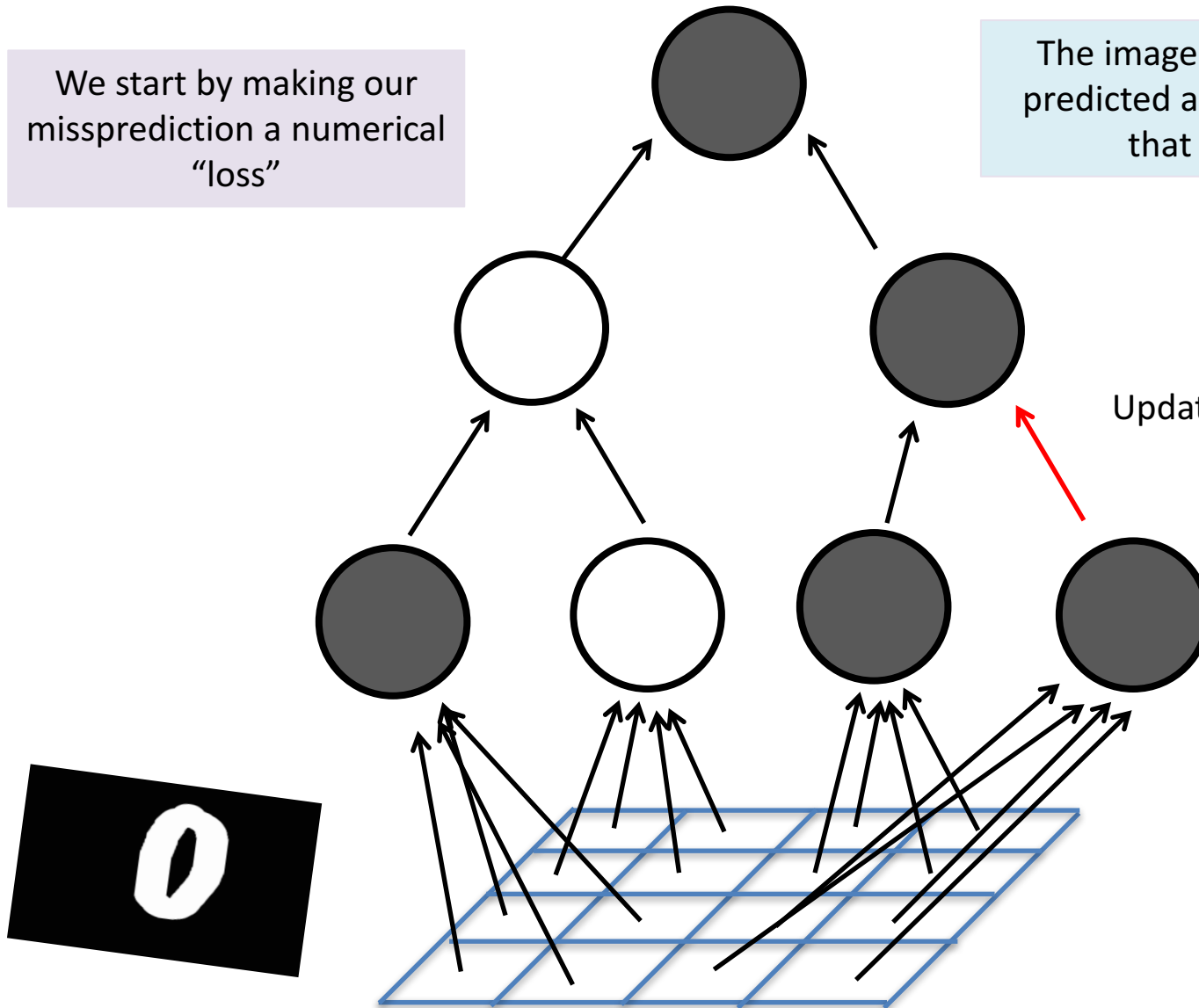


# Backward Pass

We start by making our missprediction 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

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

---

For one datum

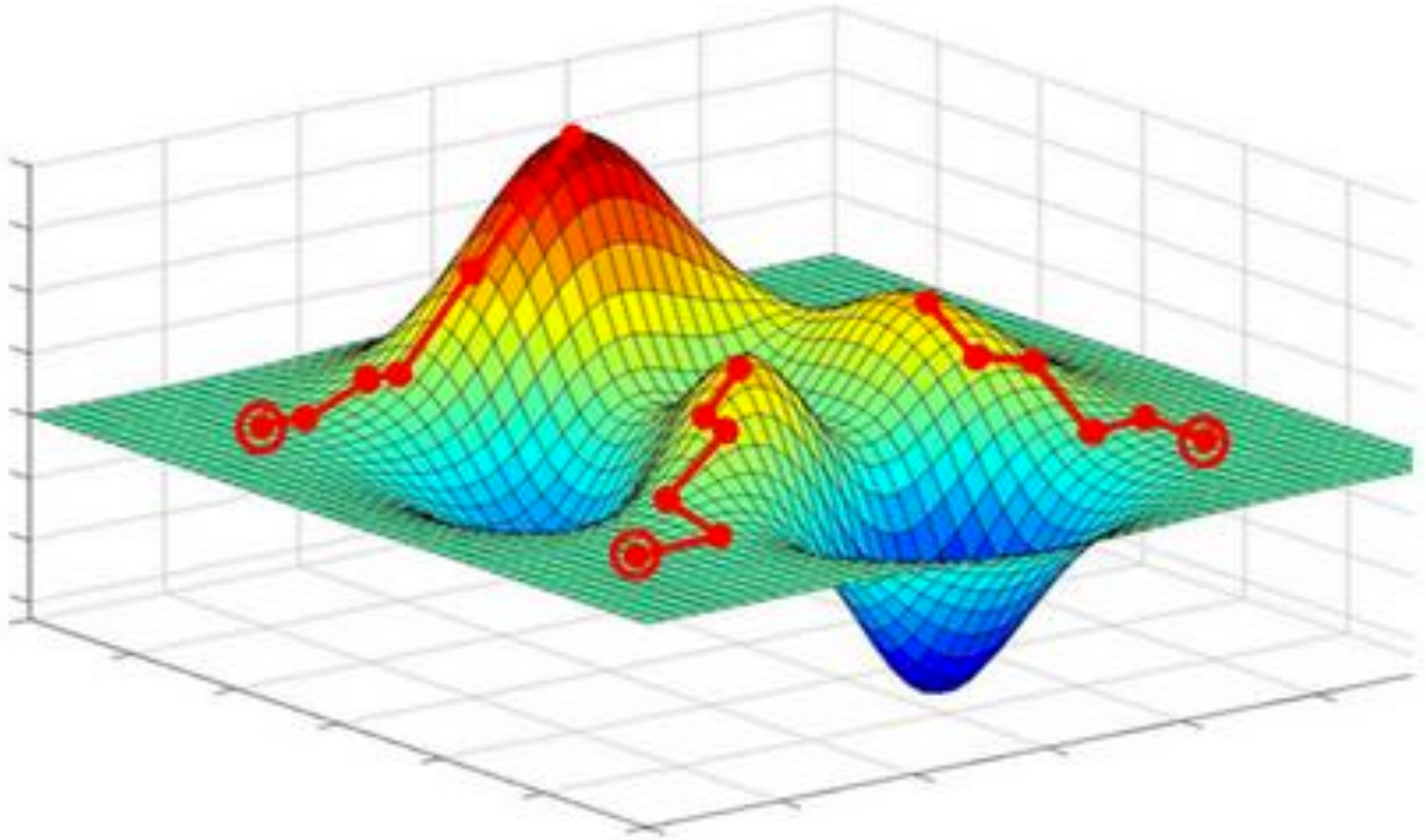
$$P(Y = y|X = \mathbf{X}) = (\hat{y})^y (1 - \hat{y})^{1-y}$$

For IID data

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n P(Y = y^{(i)}|X = \mathbf{x}^{(i)}) \\ &= \prod_{i=1}^n (\hat{y}^{(i)})^{y^{(i)}} \cdot \left[ 1 - (\hat{y}^{(i)}) \right]^{(1-y^{(i)})} \end{aligned}$$

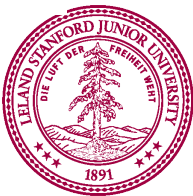


# Gradient Ascent



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

Piech, CS106A, Stanford University





# 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

$$\frac{d \text{ [dodecahedron]}}{d \text{ [sphere]}} = \frac{d \text{ [dodecahedron]}}{d \text{ [cube]}} \times \frac{d \text{ [cube]}}{d \text{ [sphere]}}$$



Where you will be by the end of class

When you train,  
something really neat happens

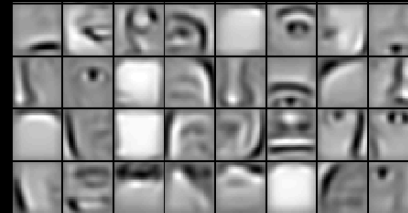
# Visualize the Weights



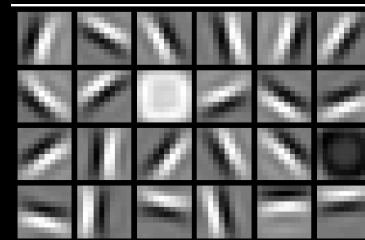
Training set: Aligned images of faces.



object models



object parts  
(combination  
of edges)

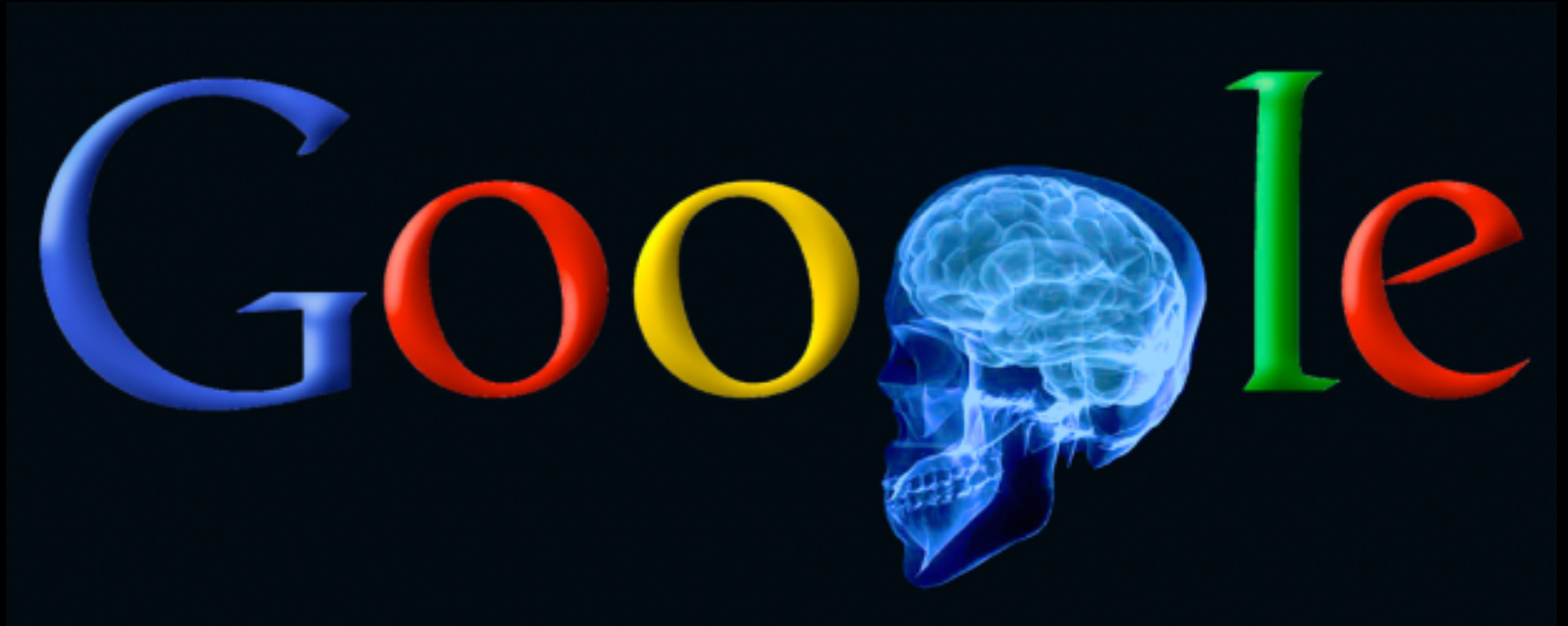


edges

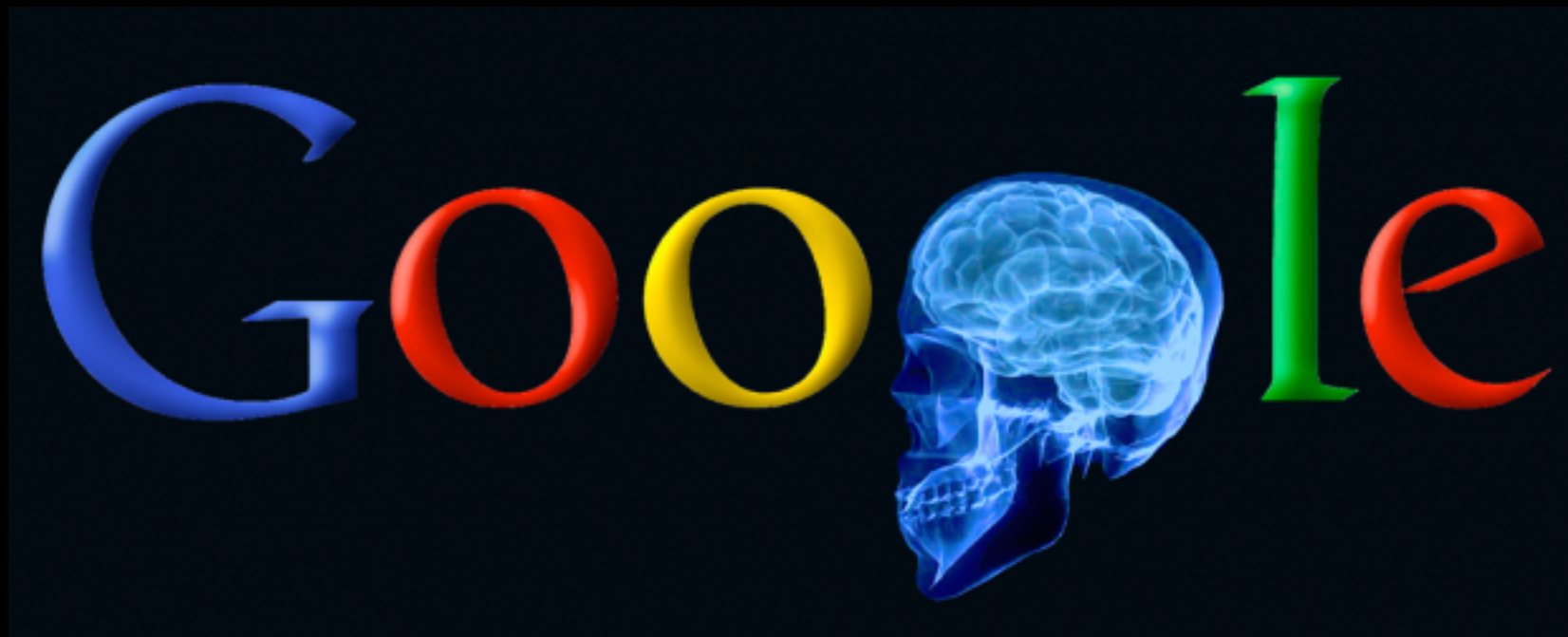


pixels

# 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



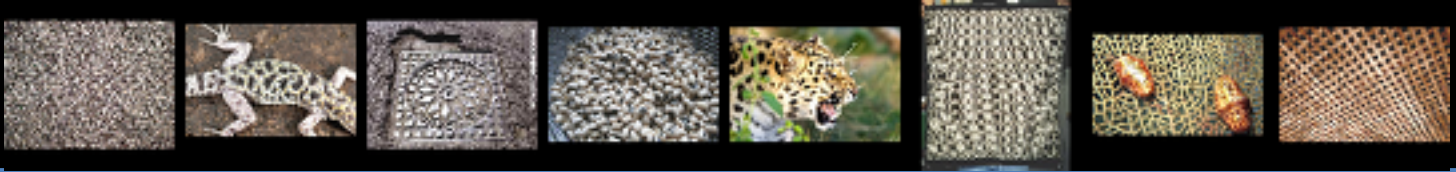
**We're essentially**

**a cat detection company**



# Other Neurons

Neuron 1



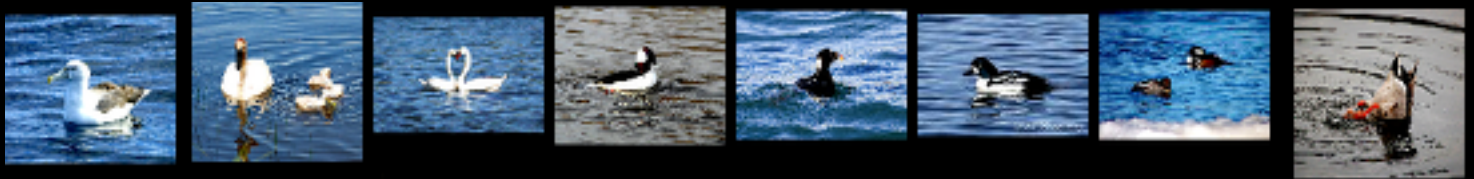
Neuron 2



Neuron 3



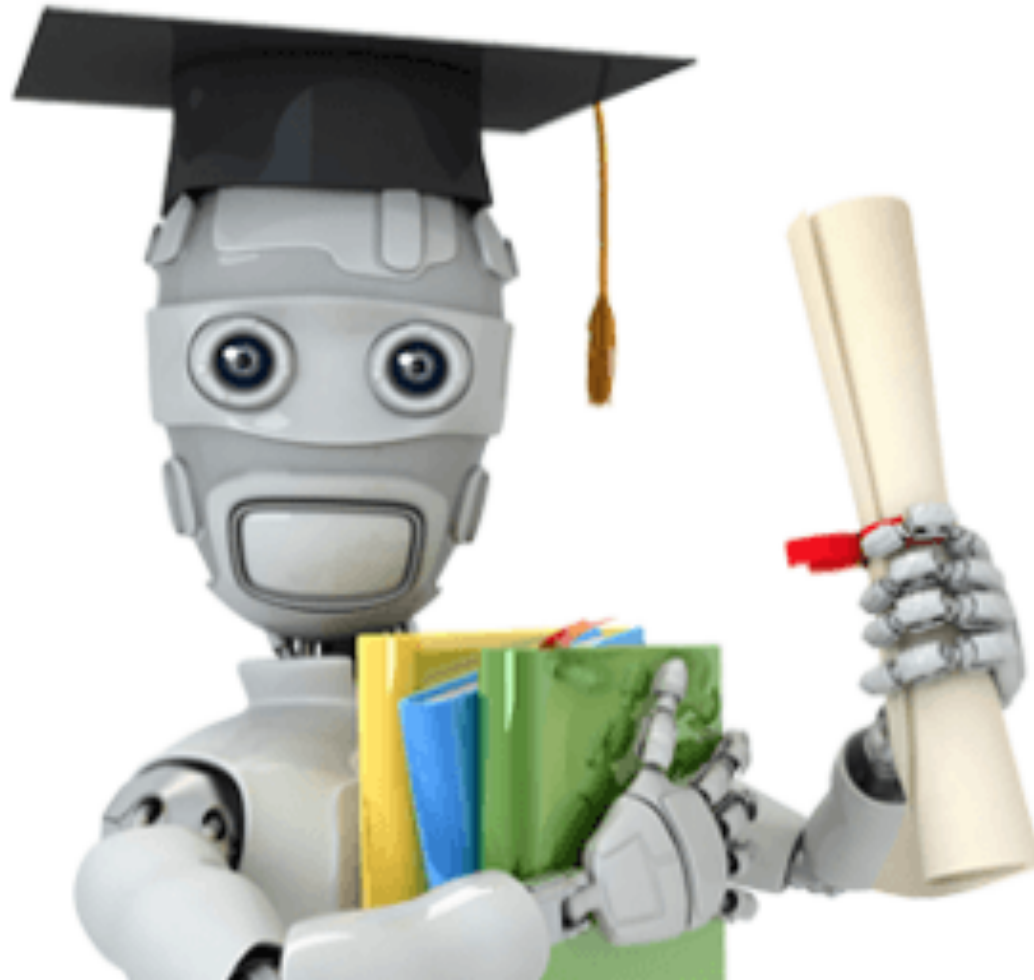
Neuron 4



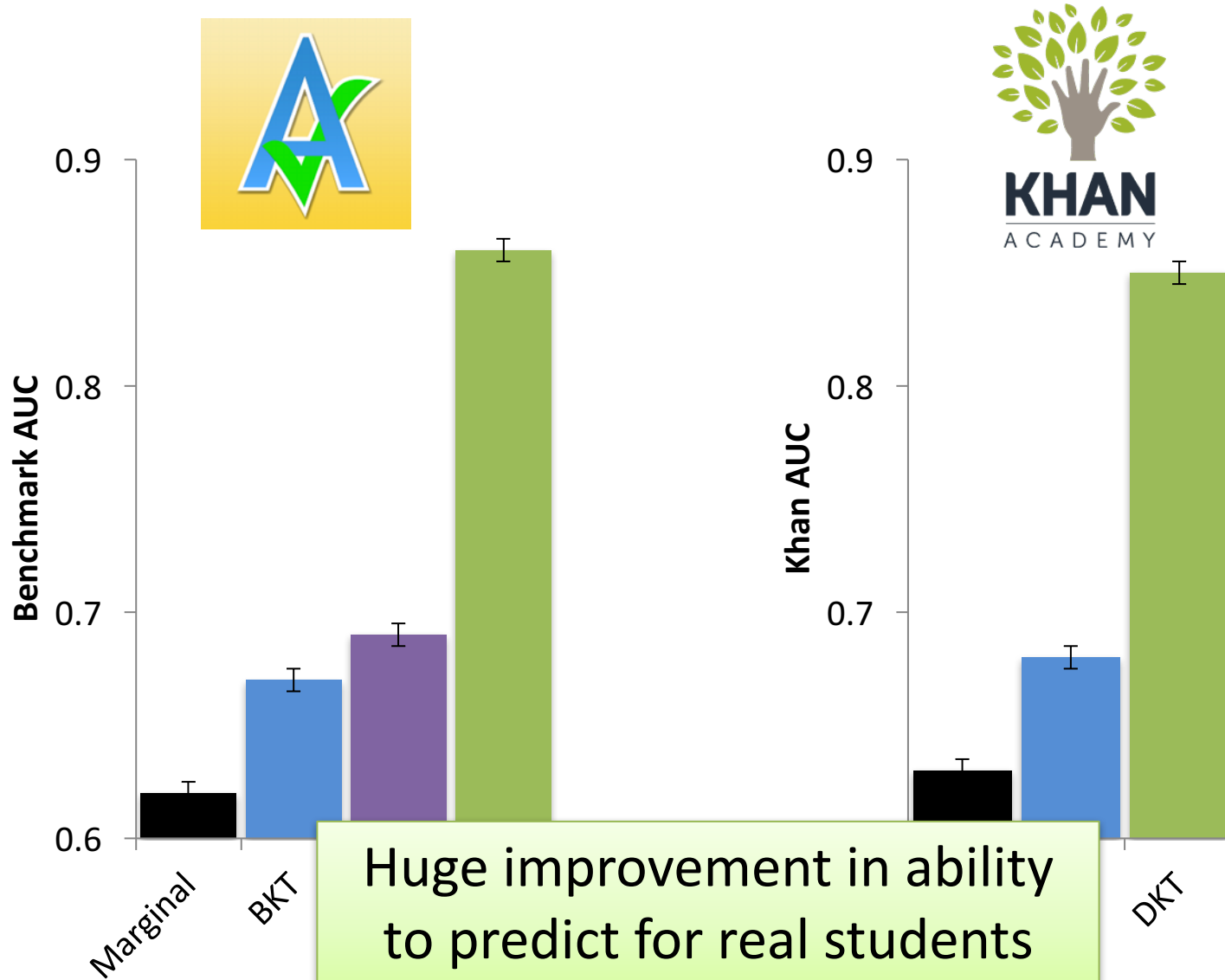
Neuron 5



# Autonomous Tutor



# Prediction Results



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

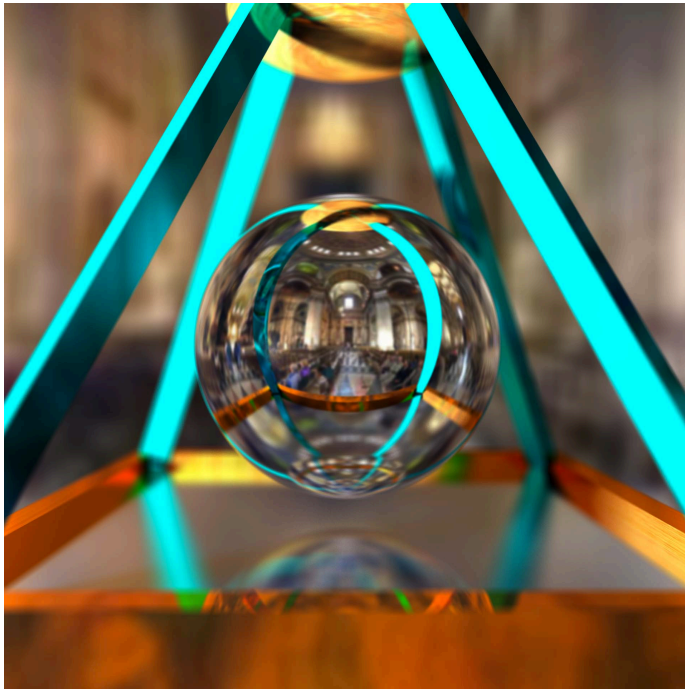


# Abundance of Important Problems

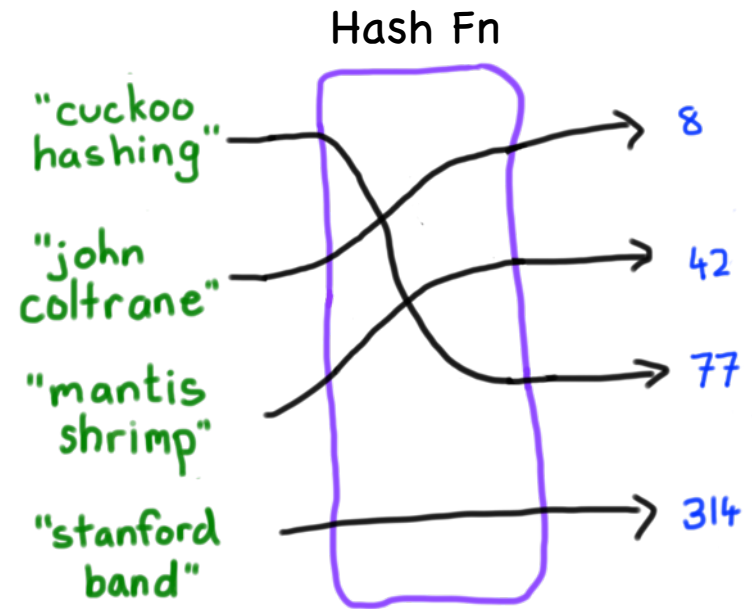


# Algorithms and Probability

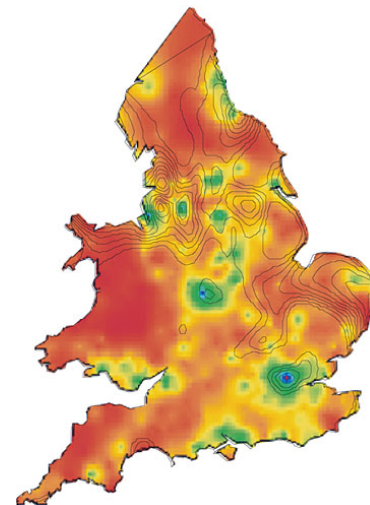
Eg Raytracing



Eg HashMaps



# Medicine and Probability



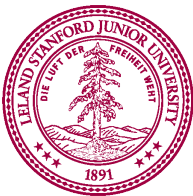
# Autocomplete



dinosaurs we

[Advanced Search](#)  
[Language Tools](#)

- dinosaurs websites for kids
- dinosaurs we're back
- dinosaurs webcomic
- dinosaurs webquest
- dinosaurs were made up by the cia to discourage time travel
- dinosaurs website
- dinosaurs went extinct
- dinosaurs weight
- dinosaurs we are scientists
- dinosaurs weed episode





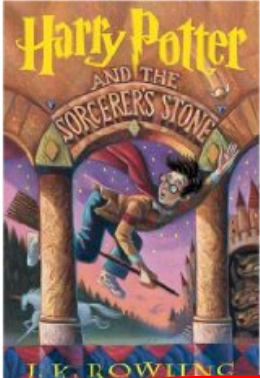
# Probability in Practice

amazon.com Hello. Sign in to get [personalized recommendations](#). New customer? [Start here](#). FREE 2-Day Shipping, No Minimum Purchase

Your Amazon.com Today's Deals Gifts & Wish Lists Gift Cards Your Account | Help

Shop All Departments Search Books GO Cart Your Lists

Books Advanced Search Browse Subjects Hot New Releases Bestsellers The New York Times® Best Sellers Libros En Español Bargain Books Textbooks



**Harry Potter and the Sorcerer's Stone (Book 1) (Hardcover)**  
by [J.K. Rowling](#) (Author), [Mary GrandPré](#) (Illustrator)  
★★★★☆ (5,471 customer reviews)

List Price: ~~\$24.99~~  
Price: **\$15.92** & eligible for **FREE Super Saver Shipping** on orders over \$25.  
[Details](#)  
You Save: **\$9.07 (36%)**

**In Stock.**  
Ships from and sold by **Amazon.com**. Gift-wrap available.

Quantity: 1  
[Add to Shopping Cart](#)  
or  
[Sign in](#) to turn on 1-Click ordering.  
or  
[Add to Cart with FREE Two-Day Shipping](#)  
Amazon Prime Free Trial required. Sign up when you check out. [Learn More](#)

**Customers Who Bought This Item Also Bought**

Page 1 of 20

  
**Harry Potter and the Prisoner of Azkaban (Book 3)** by J.K. Rowling  
★★★★☆ (2,599) **\$16.49**

  
**Harry Potter and the Goblet of Fire (Book 4)** by J.K. Rowling  
★★★★☆ (5,186) **\$19.79**

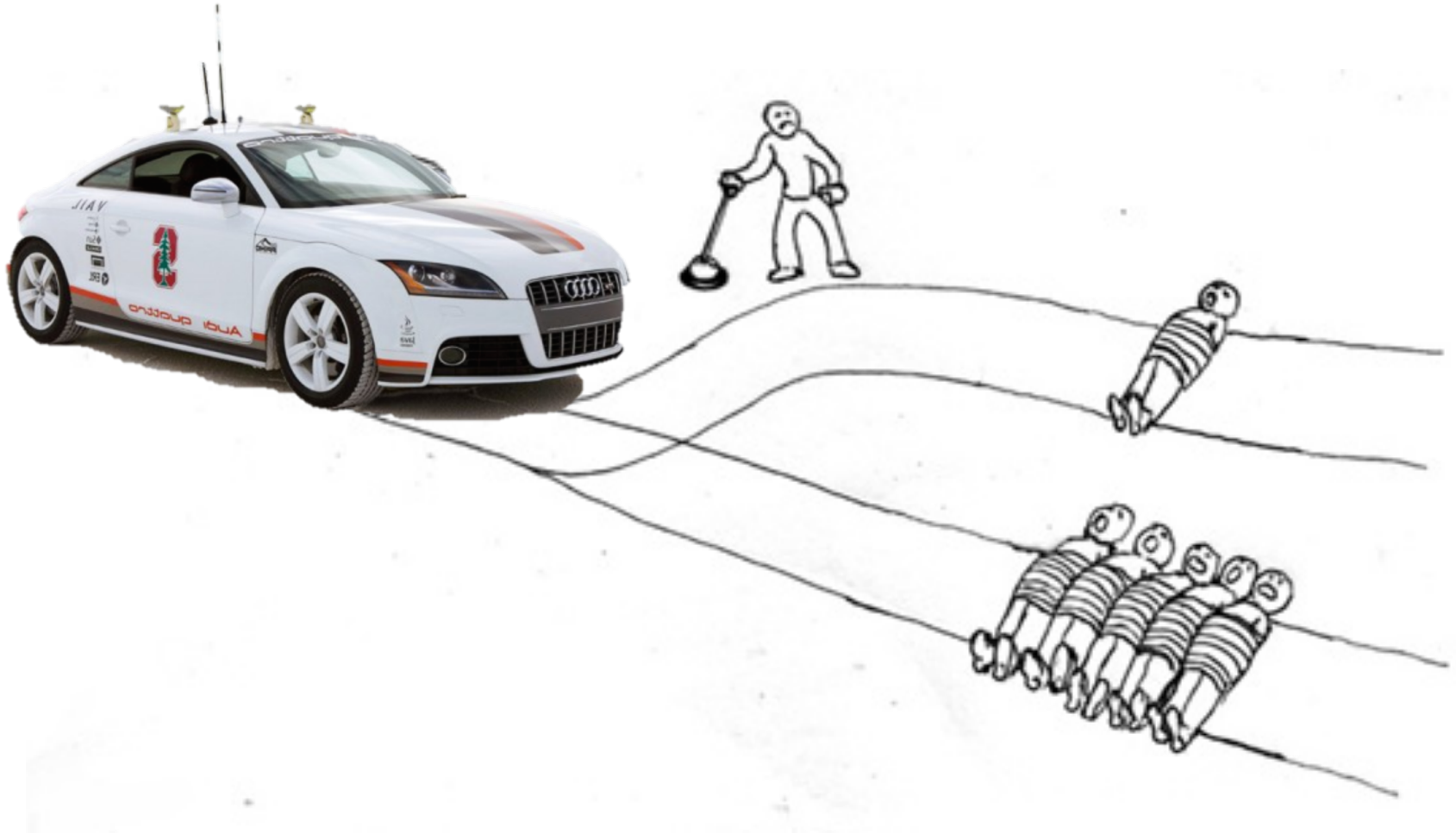
  
**Harry Potter and the Order of the Phoenix (Book 5)** by J. K. Rowling  
★★★★☆ (5,876) **\$10.18**

  
**Harry Potter and the Half-Blood Prince (Book 6)** by J.K. Rowling  
★★★★☆ (3,597) **\$10.18**

  
**The Tales of Beedle the Bard, Collector's Ed...** by J. K. Rowling  
★★★★☆ (176)

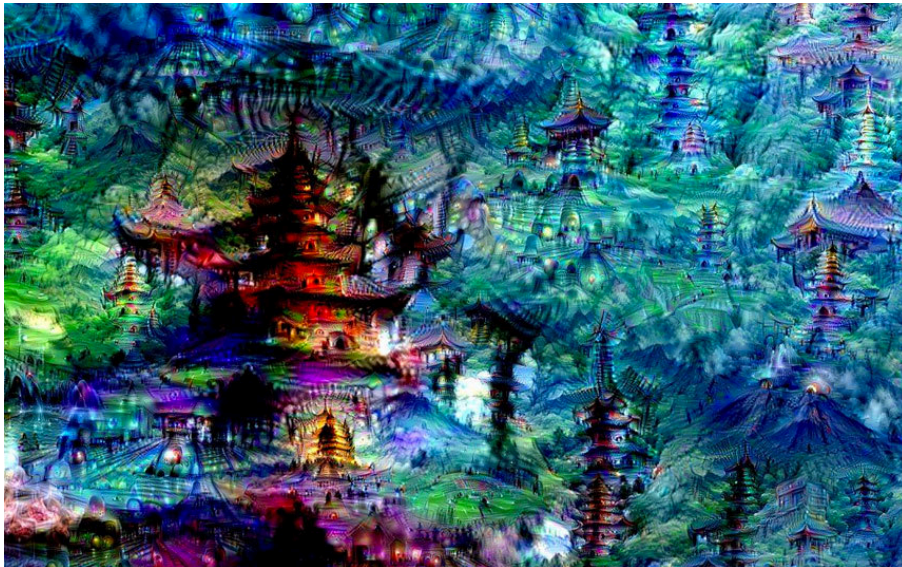
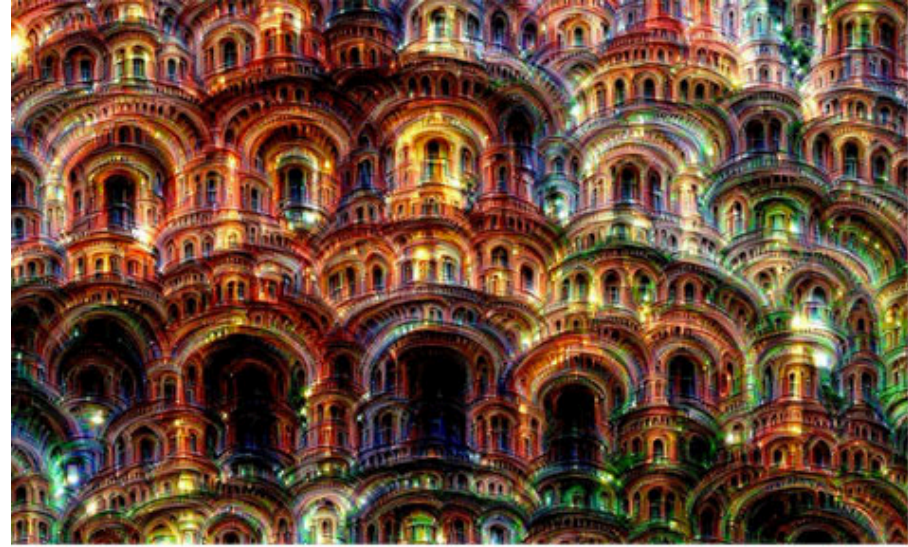
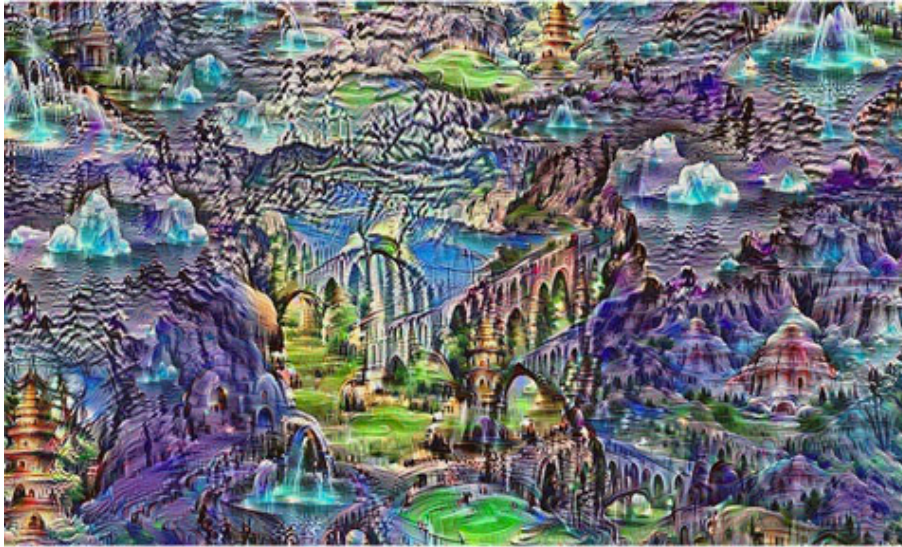


# Philosophy and Probability



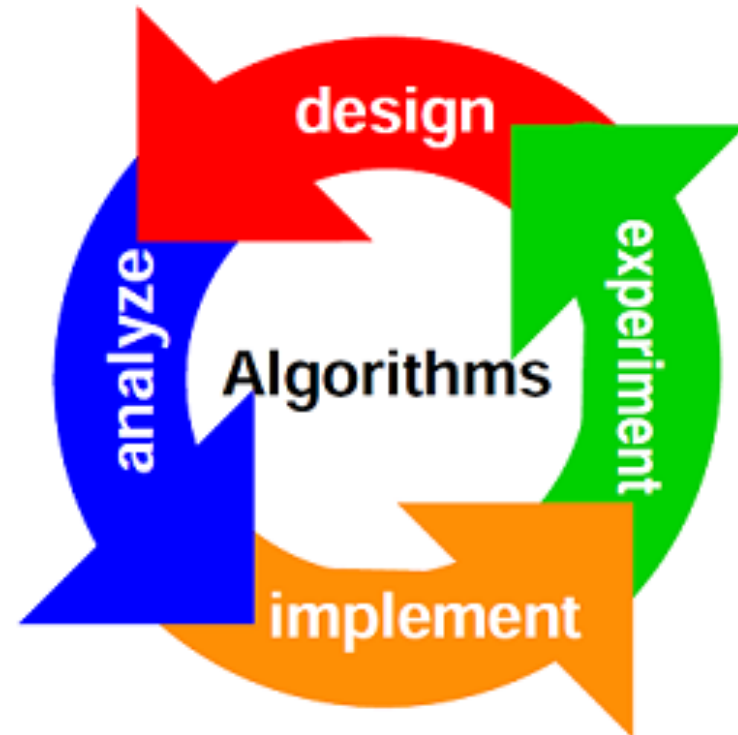
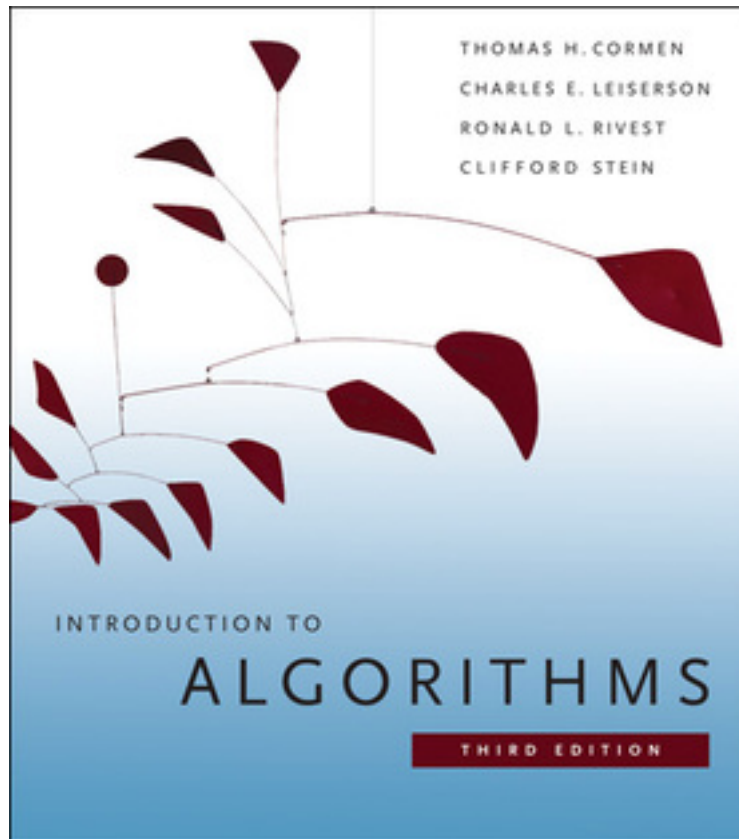


# Art and Probability





# 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

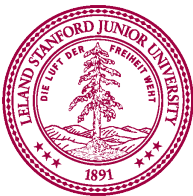


Positive Zika.

*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?

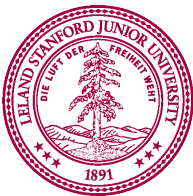
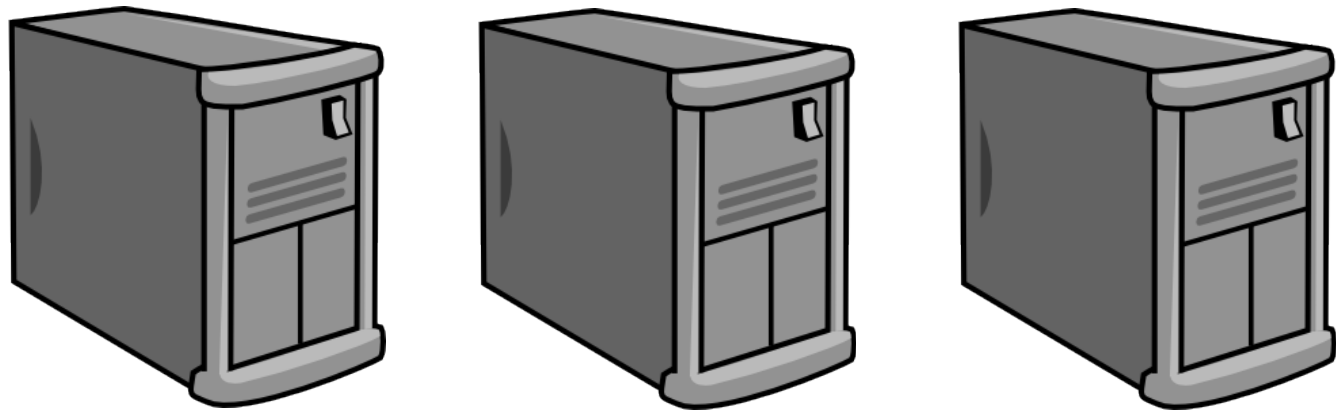
# Traditional View of Probability



# CS View of Probability

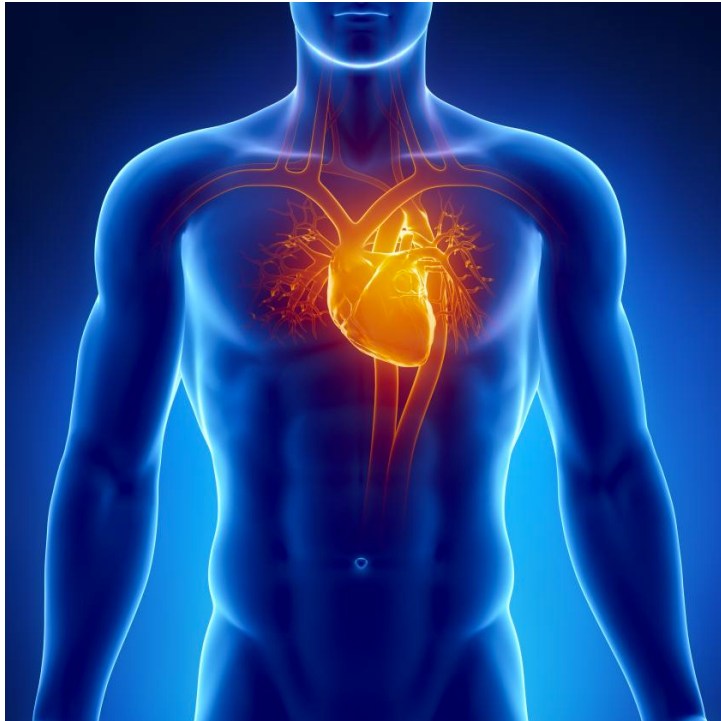
<http://www.site.com>

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



# CS View of Probability

Heart



Ancestry



23andMe

Netflix

The Netflix logo, which consists of the word "NETFLIX" in a bold, white, sans-serif font. The letters have a slight 3D effect with a dark shadow. The logo is centered on a solid red rectangular background.

NETFLIX

# CS View of Probability



Piech, CS106A, Stanford University



# CS View of Probability

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

Lets dive in...

# Counting





# Our Route

