Reinforcement

Chris Piech
CS109, Stanford University

Announcements

Home Stretch for CS109 (Nov 17th) #541

y, November 17, 2025 at 11:40:08 AM PST

3 hours ago in General

UNPIN STAR WATCHING 249 VIEWS

23

Hi lovely CS109 class!

I hope you had a great weekend and are feeling ready for Thanksgiving break, coming up.

We have made it to a big milestone in class. In today's lecture I will introduce the final piece of new content. Congrats! So, what happens next in CS109?

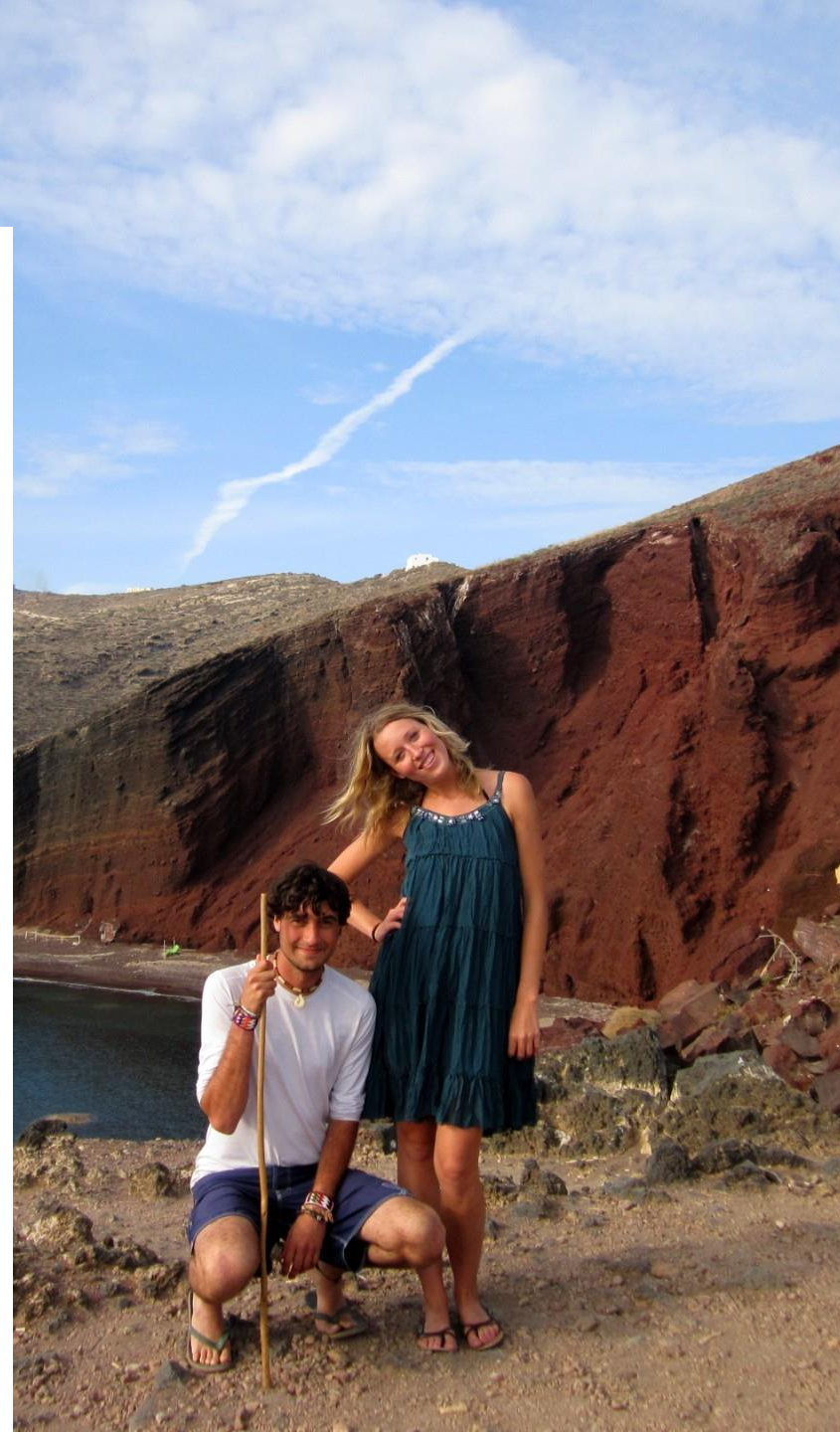
Lots of Practice!

This quarter for the final two weeks of class we are going to focus intentionally on giving you a ton of applied problems that are designed to give you **practice** (for the final and for life beyond CS109). You have learned a ton of great content and we want every single one of you to feel like you know this core front to back :-). To that end, the final pset (that goes out tomorrow morning) is going to be all review problems. This is new in CS109, and its in response to listening to lots of students and asking: "what can we do to best set you all up for success?" I am committed to making this practice fun and rewarding. I will mix in some basics as well as some more complicated problems. Its made for everyone -- see you in lecture!

Sign up for Final PEP

As you know, PEP is a key part of class (and required for your participation grade). PEP will be the Monday and Tuesday after thanksgiving break. We just turned signups on: Final PEP Signup.

Extra Late Days



Today

Regression

Night Sight

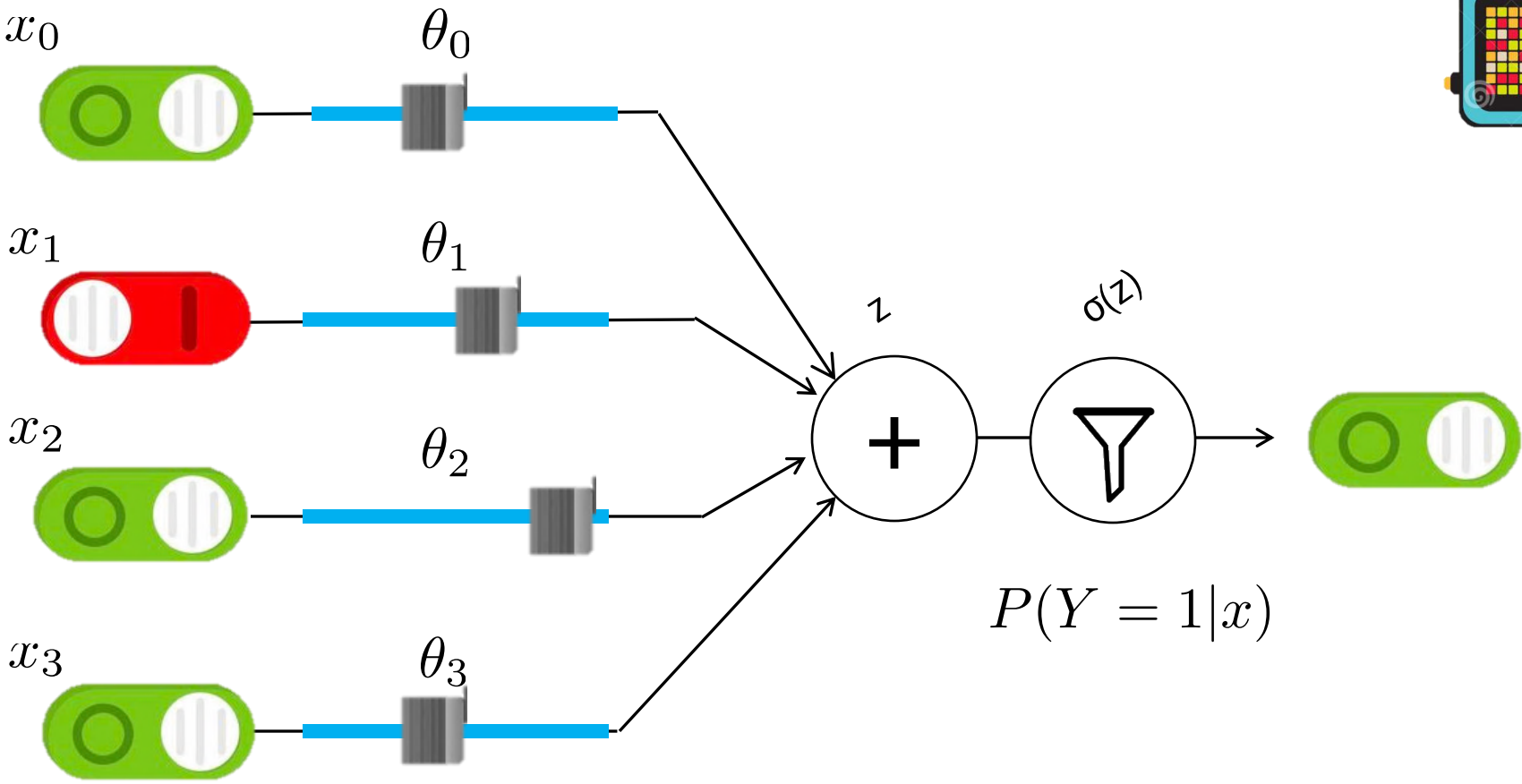
Birthday Paradox

Wisdom of the Crowds

Ratings out of 5: What to Buy?

Review

Logistic Regression

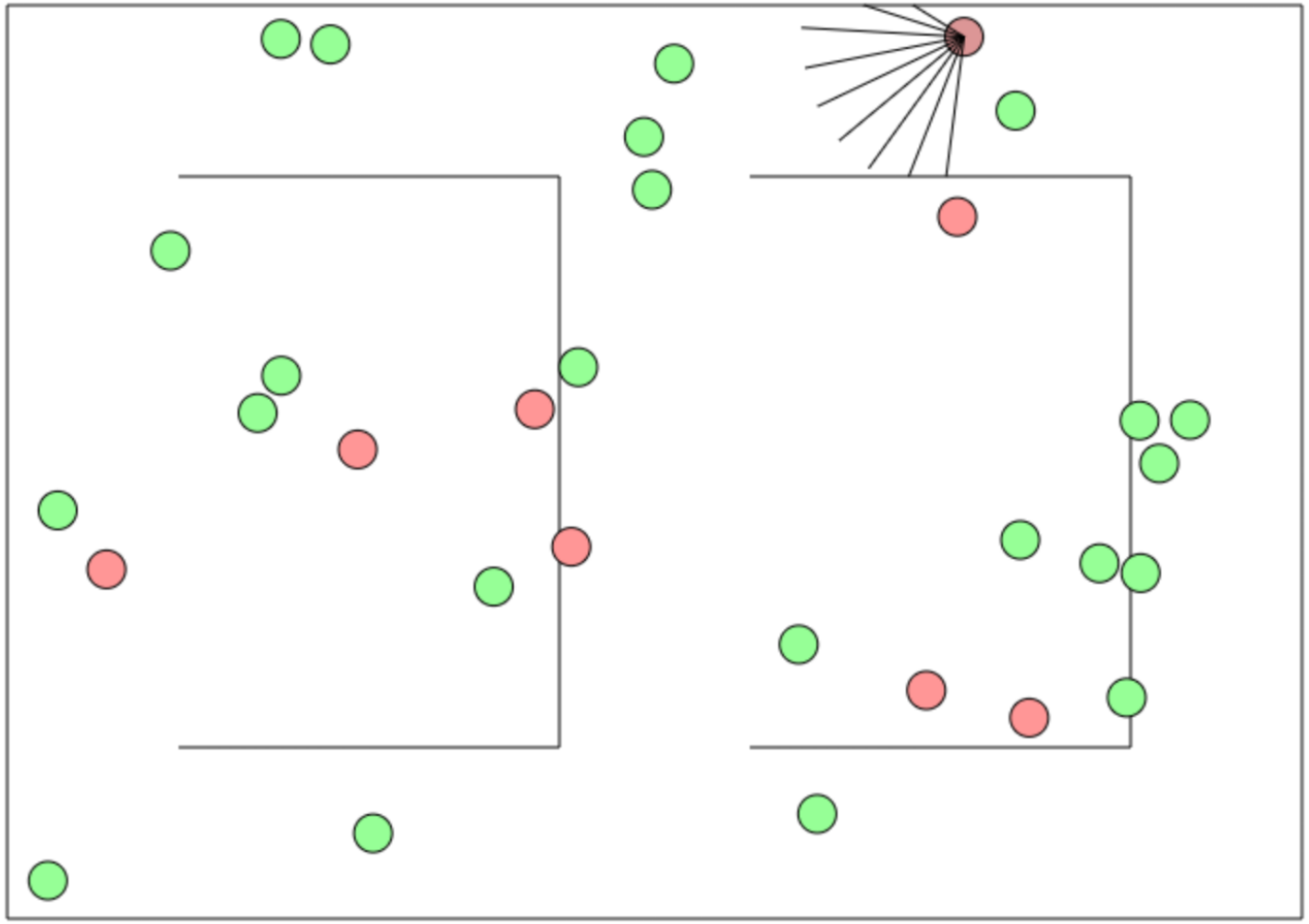


$$P(Y = 1|\mathbf{X} = \mathbf{x}) = \sigma\Big(\sum_i \theta_i x_i\Big)$$

But wait….

Is all of ML Classification?

Lets start training a Critter



http://cs.stanford.edu/people/karpathy/convnetjs/demo/rldemo.html

Types of Machine Learning Tasks

Muli-Class Classification

Regression

Reinforcement Learning

Generation

Types of Machine Learning Tasks

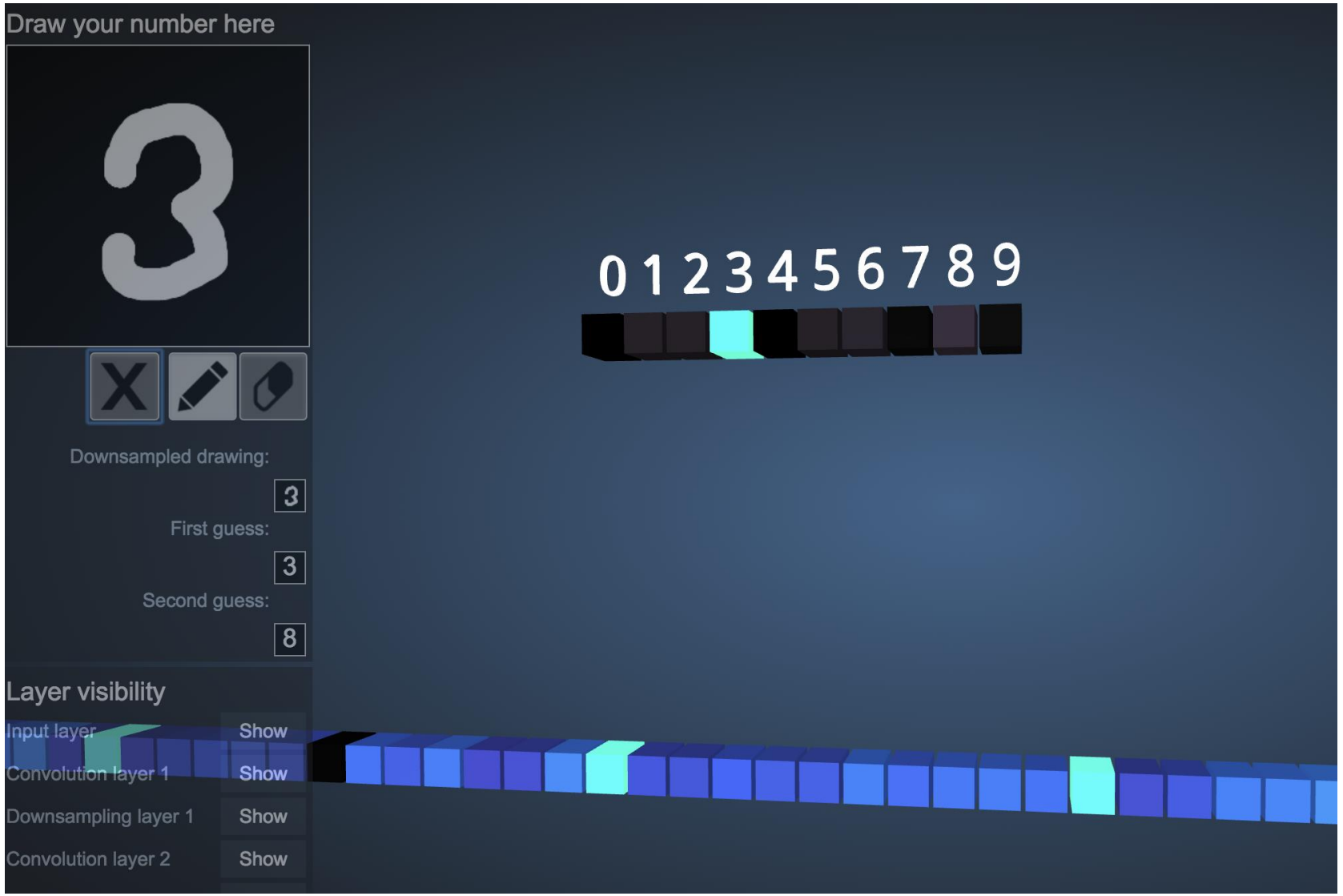
Muli-Class Classification

Regression

Reinforcement Learning

Generation

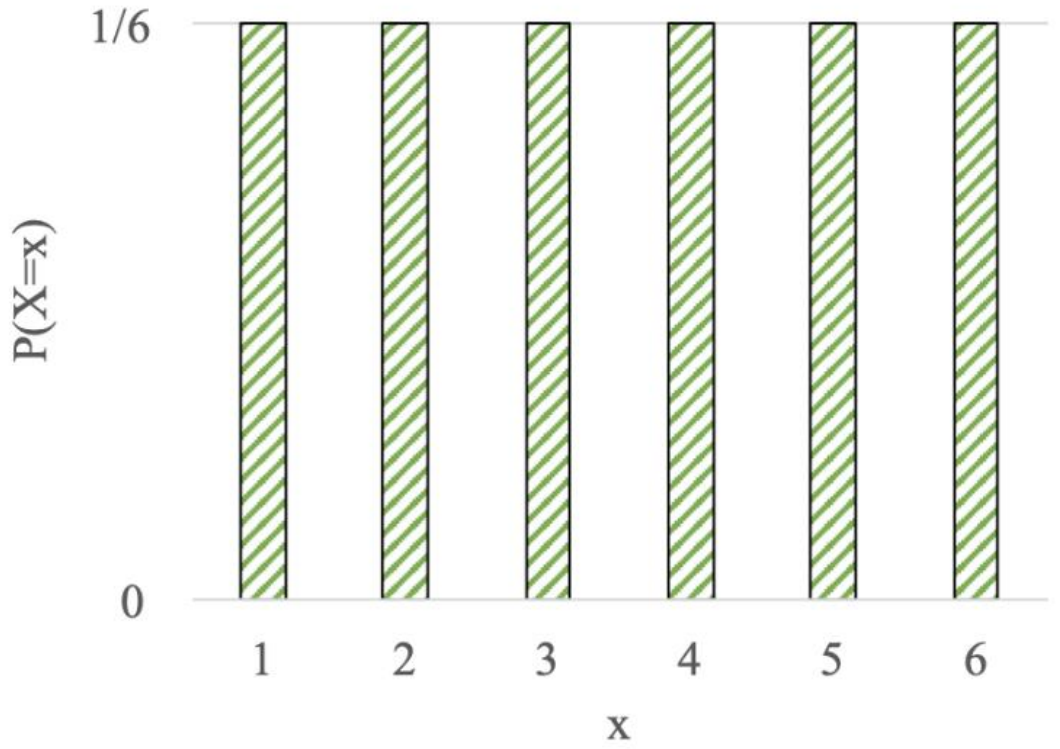
Multiple Outputs



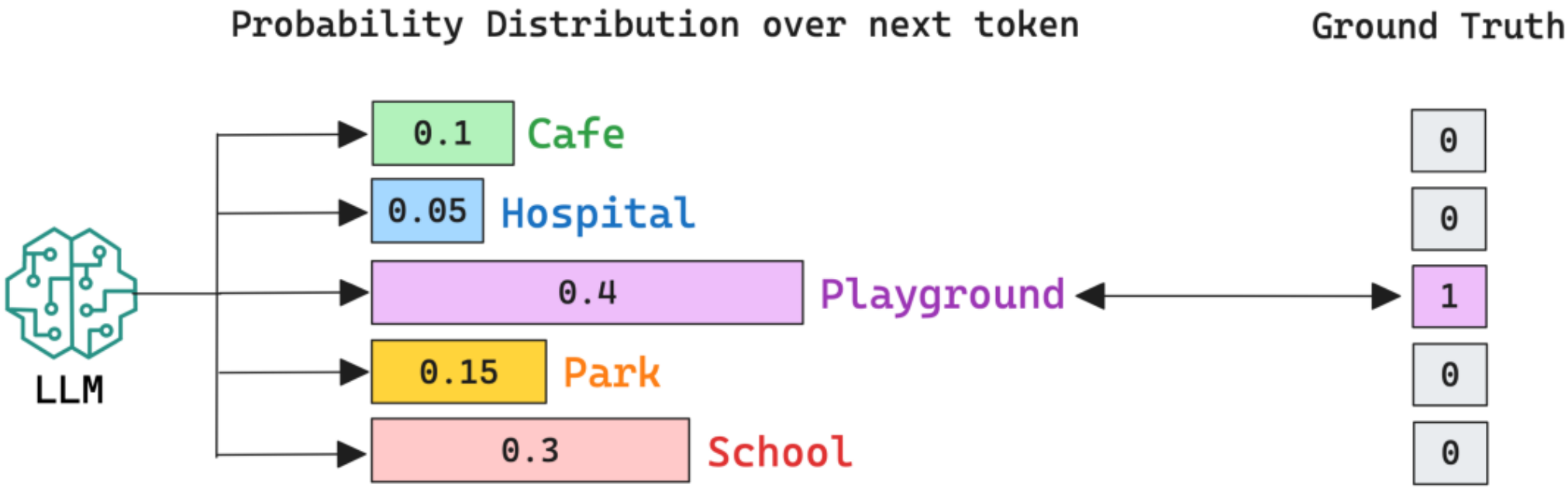
Categorial Example from Class: Thinking of an Animal

Notice that these are not numbers

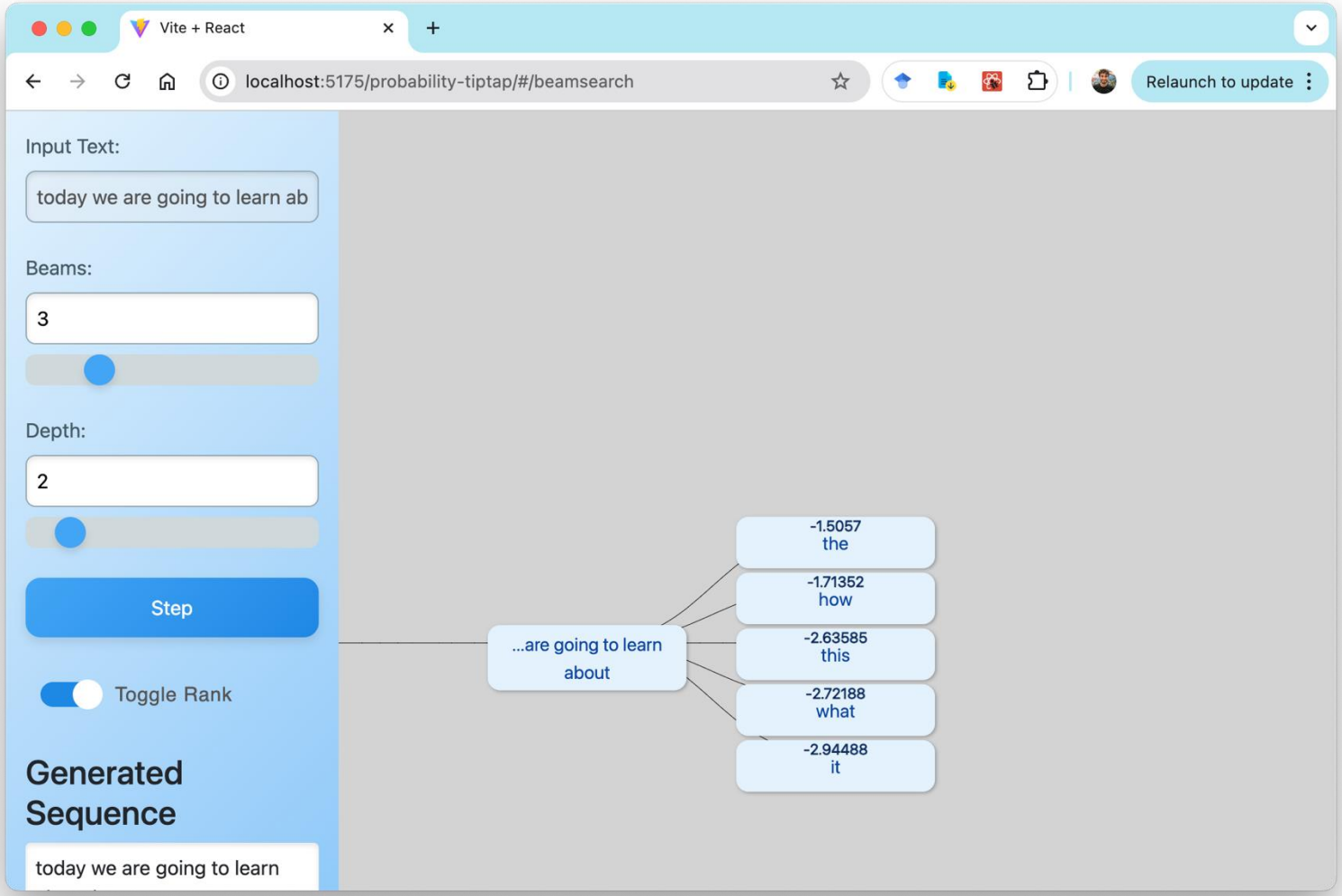
Categorial Example from Class: Single Dice Roll



Output of an LLM is a Categorical for Next Token



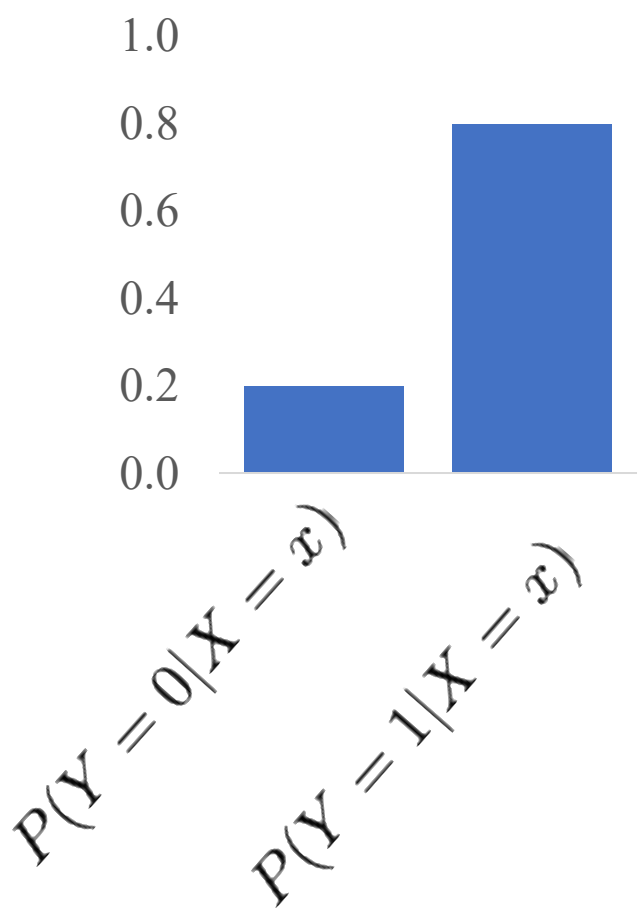
Output of an LLM is a Categorical for Next Token



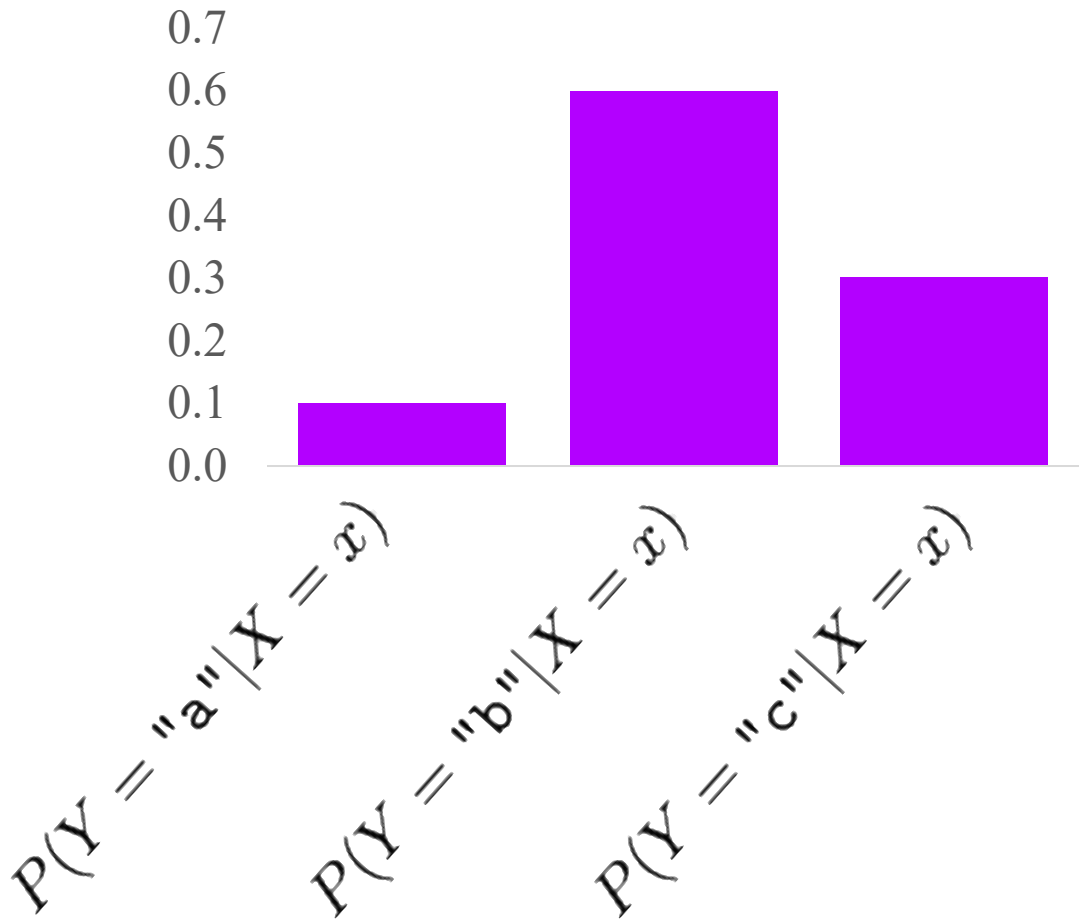
Thanks to Justin Blumencranz and Adam Boswell

Logistic Regression to Predict a Categorical?

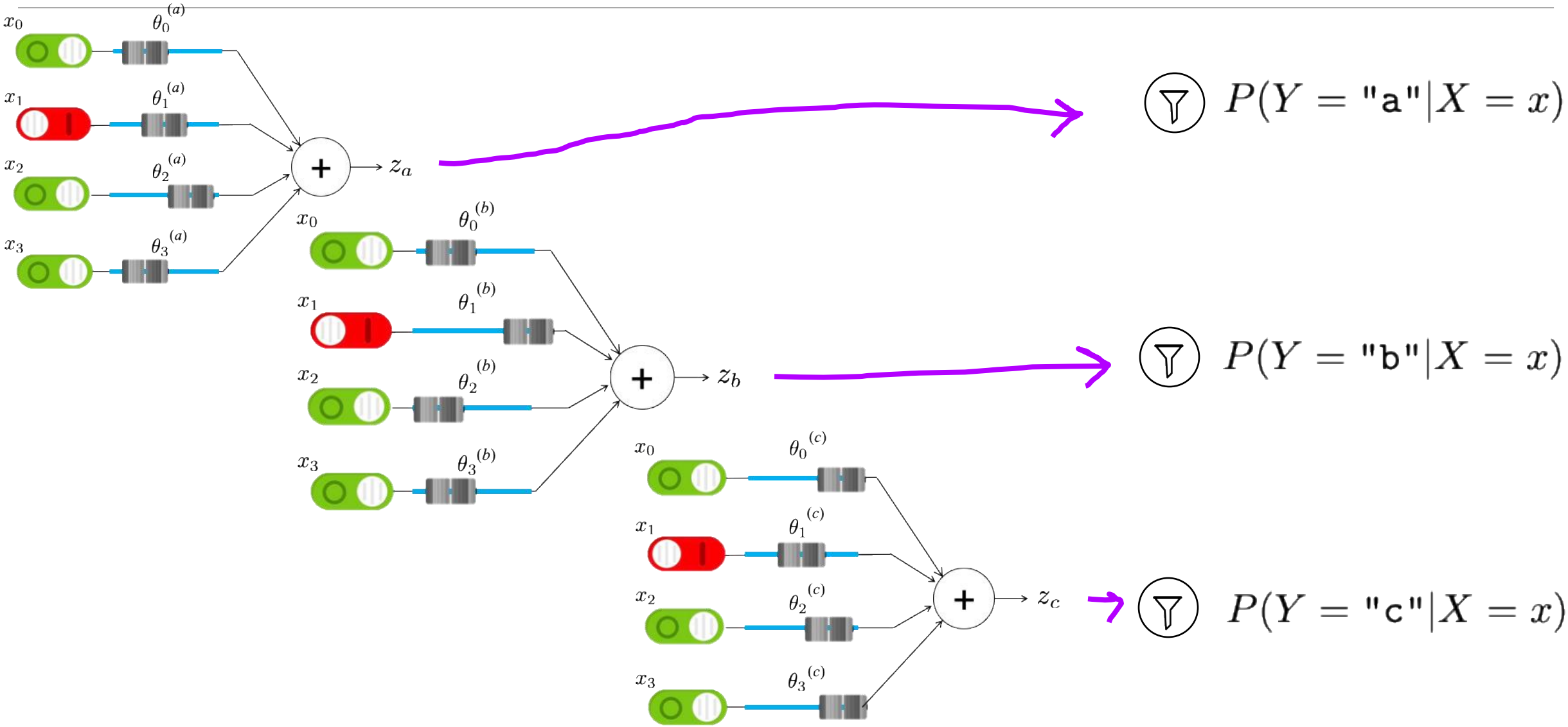
Standard Logistic Regression



Multi-Class Logistic Regression



Logistic Regression to Predict a Categorical



End Review

Types of Machine Learning Tasks

Muli-Class Classification

Regression

Reinforcement Learning

Generation

Regression: Predicting Real Numbers

	Opposing team ELO	Points in last game	...	At Home?	Output
					# Points
Game 1	84	105		1	120
Game 2	90	102		0	95
		⋮			⋮
Game *n*	74	120		0	115

Same Notation for Training Data

Training Data: assignments all random variables **X** and Y

Assume IID data:

n training datapoints

$$(\mathbf{x}^{(1)}, y^{(1)}), (\mathbf{x}^{(2)}, y^{(2)}), \ldots (\mathbf{x}^{(n)}, y^{(n)})$$

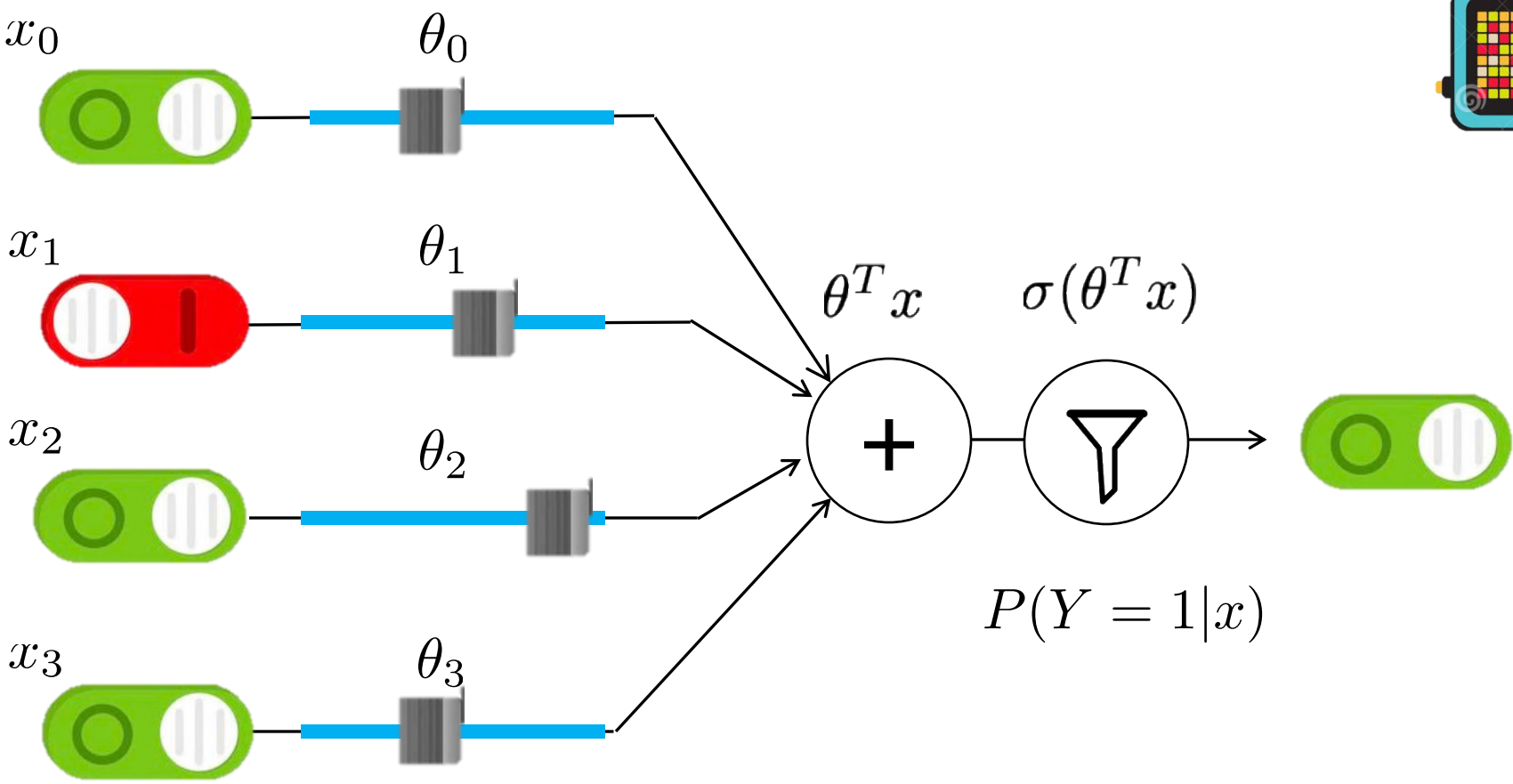
$$m = |\mathbf{x}^{(i)}|$$

Each datapoint has m features and a single output

Regression: Predicting Real Numbers

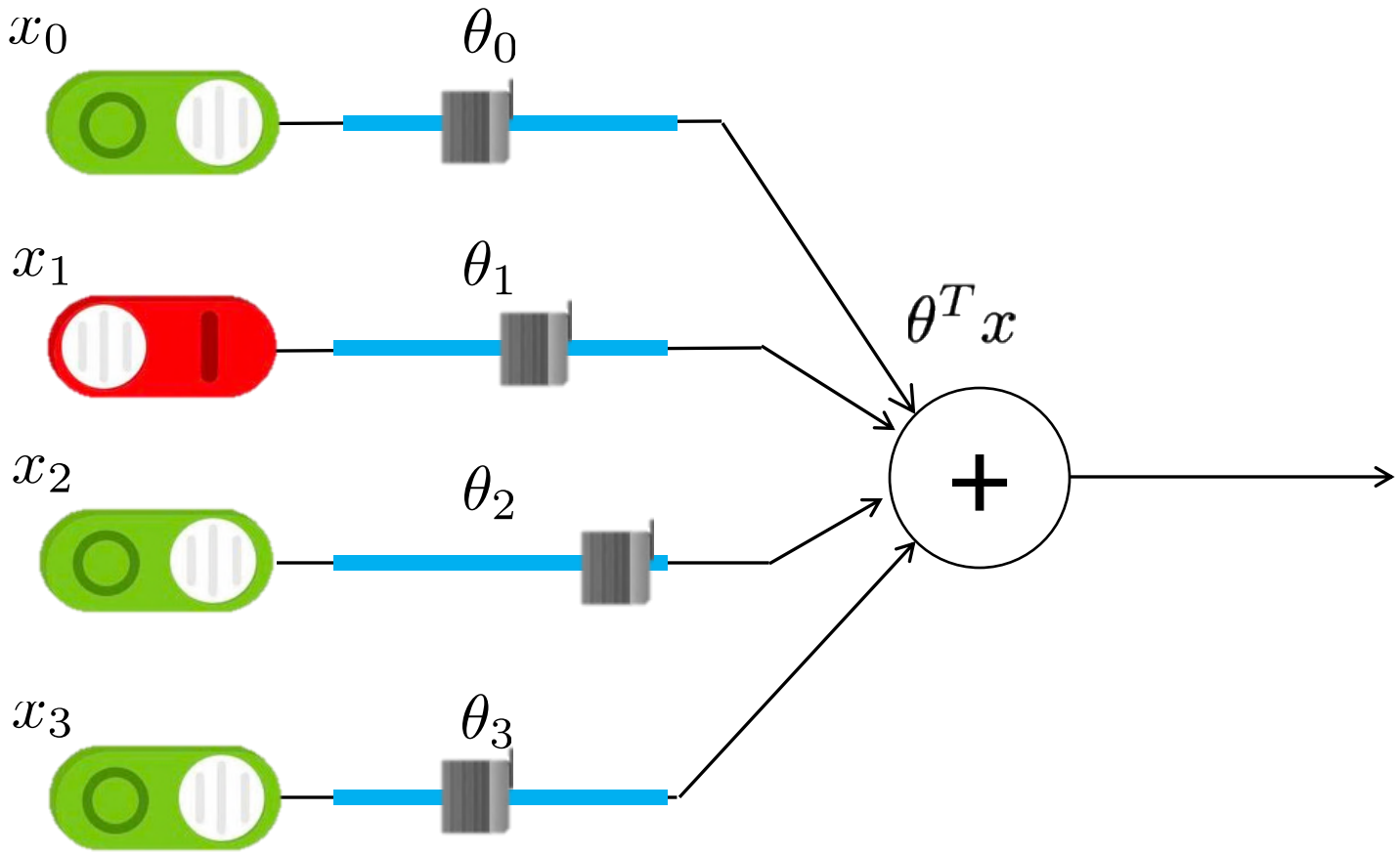
	Opposing team ELO	Points in last game	...	At Home?	Output
					# Points
Game 1	84	105		1	120
Game 2	90	102		0	95
		⋮			⋮
Game *n*	74	120		0	115

Logistic Regression

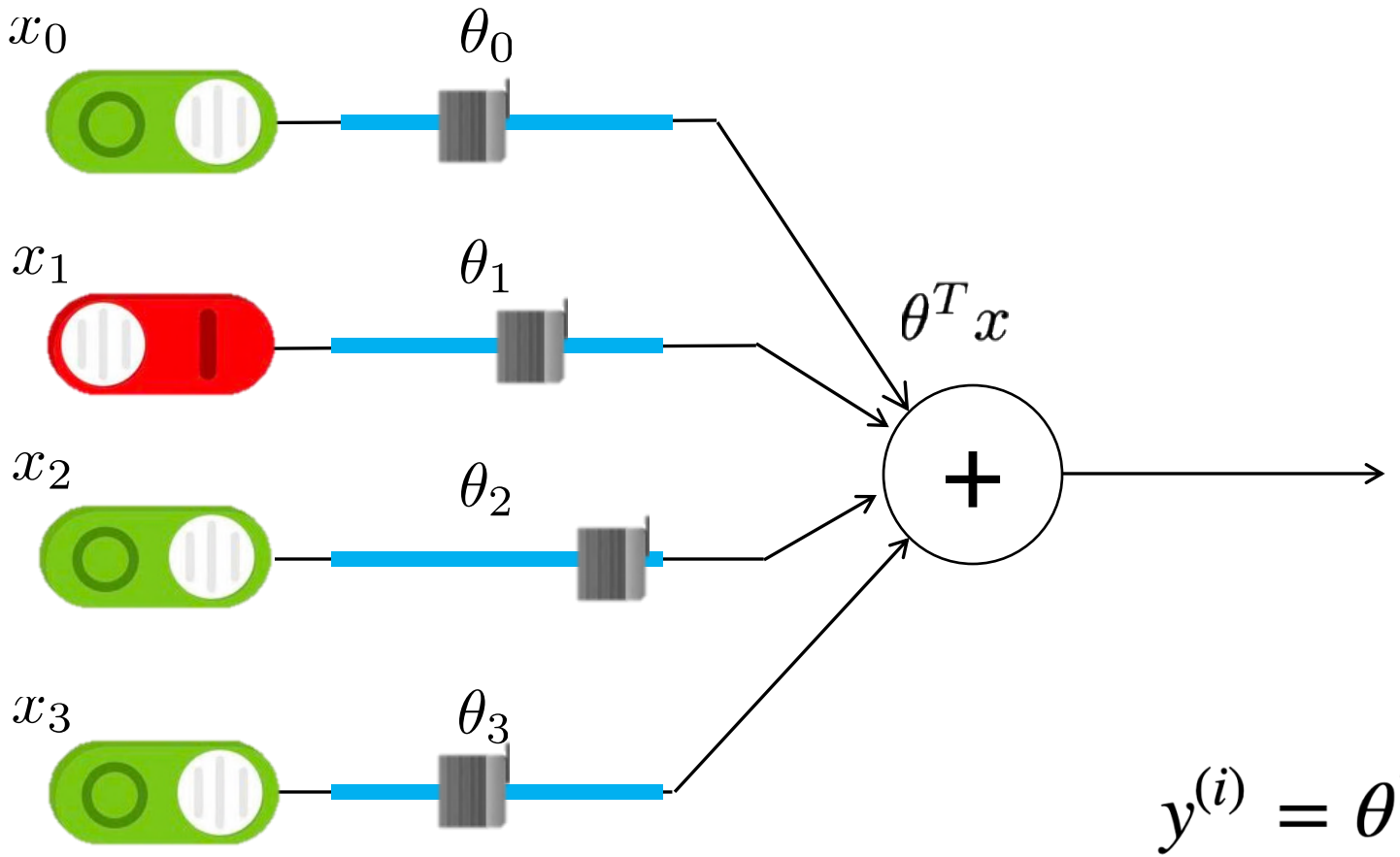


$$P(Y = 1|\mathbf{X} = \mathbf{x}) = \sigma\Big(\sum_i \theta_i x_i\Big)$$

Linear Regression



Linear Regression



$$y^{(i)} = \theta^T x^{(i)} + Z$$

$$Z \sim N(0, \sigma^2)$$

Linear Regression Model

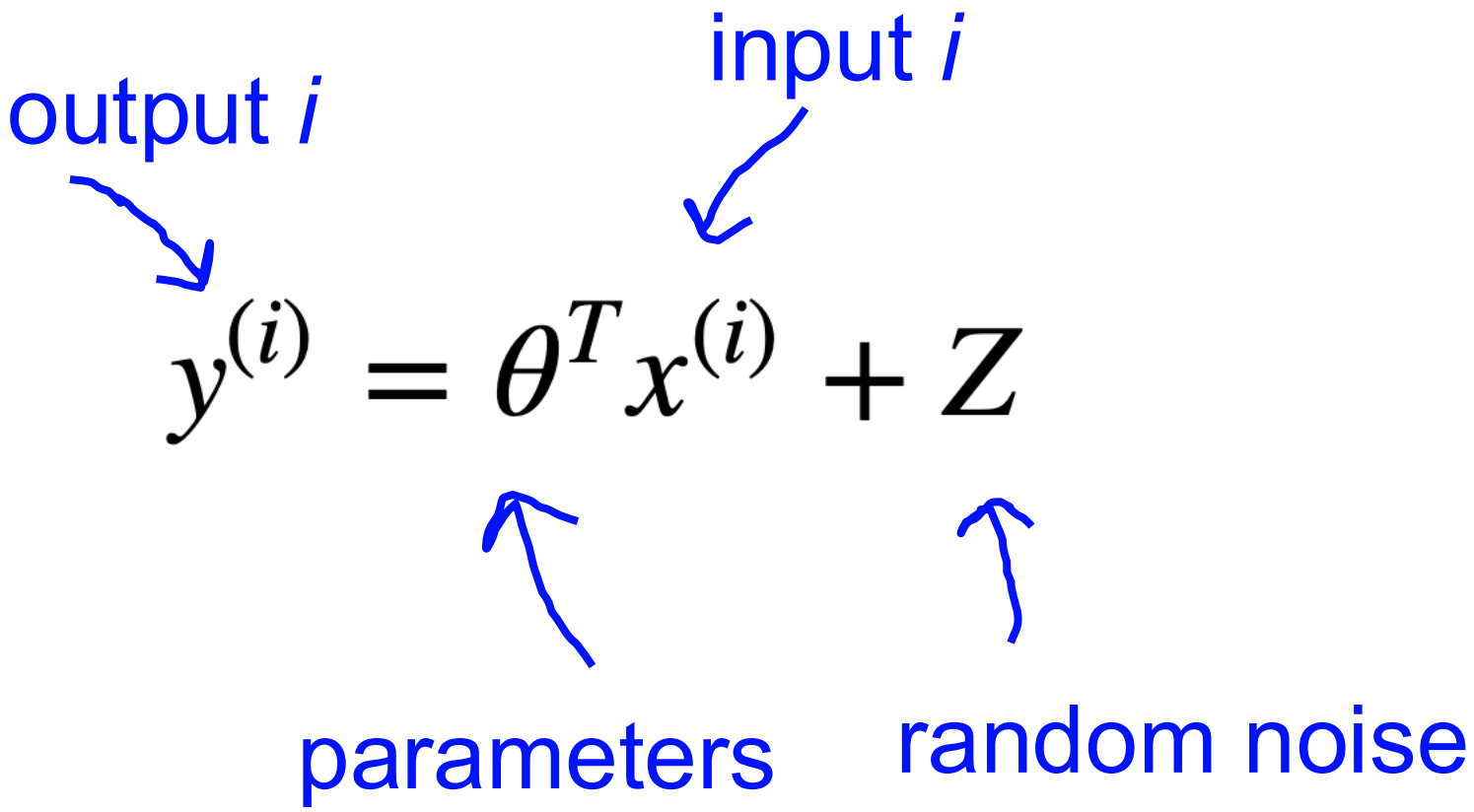
output *i*

input *i*

$$y^{(i)} = \theta^T x^{(i)}$$

parameters

Linear Regression Model



Linear Regression Model

output *i*

input *i*

$$y^{(i)} = \theta^T x^{(i)} + Z$$

parameters

random noise

Noise is *N* with mean 0

$$Z \sim N(0, \sigma^2)$$

Linear Regression Model

output *i*

noise

$$y^{(i)} = \theta^T x^{(i)} + Z$$

noise is *N* with mean 0

$$Z \sim N(0, \sigma^2)$$

1. Linear Transform

If X is a Normal such that $X \sim N(\mu, \sigma^2)$ and Y is a linear transform of X such that $Y = aX + b$ then Y is also a Normal where:

$$Y \sim N(a\mu + b, a^2\sigma^2)$$

Linear Regression Model

output *i*

noise

$$y^{(i)} = \theta^T x^{(i)} + Z$$

noise is *N* with mean 0

$$Z \sim N(0, \sigma^2)$$

Output is normal too:

$$y^{(i)} \sim N(\theta^T x^{(i)}, \sigma^2)$$

Log Likelihood

Assume: $y^{(i)} \sim N(\theta^T x^{(i)}, \sigma^2)$

Data: $(\mathbf{x}^{(1)}, y^{(1)}), (\mathbf{x}^{(2)}, y^{(2)}), \ldots (\mathbf{x}^{(n)}, y^{(n)})$

$$LL(\theta) = \sum_{i=1}^{n} \log \left[f(y^{(i)}) \right]$$

$$= \sum_{i=1}^{n} \log \left[\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y^{(i)}-\mu}{\sigma}\right)^2} \right]$$

Normal distribution PDF

$$= \sum_{i=1}^{n} \log \left[\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y^{(i)}-\theta^T x^{(i)}}{\sigma}\right)^2} \right]$$

Substitute in the mean

$$= \sum_{i=1}^{n} \log \left[\frac{1}{\sigma\sqrt{2\pi}} \right] - \frac{1}{2} \left(\frac{y^{(i)} - \theta^T x^{(i)}}{\sigma} \right)^2$$

Apply the log

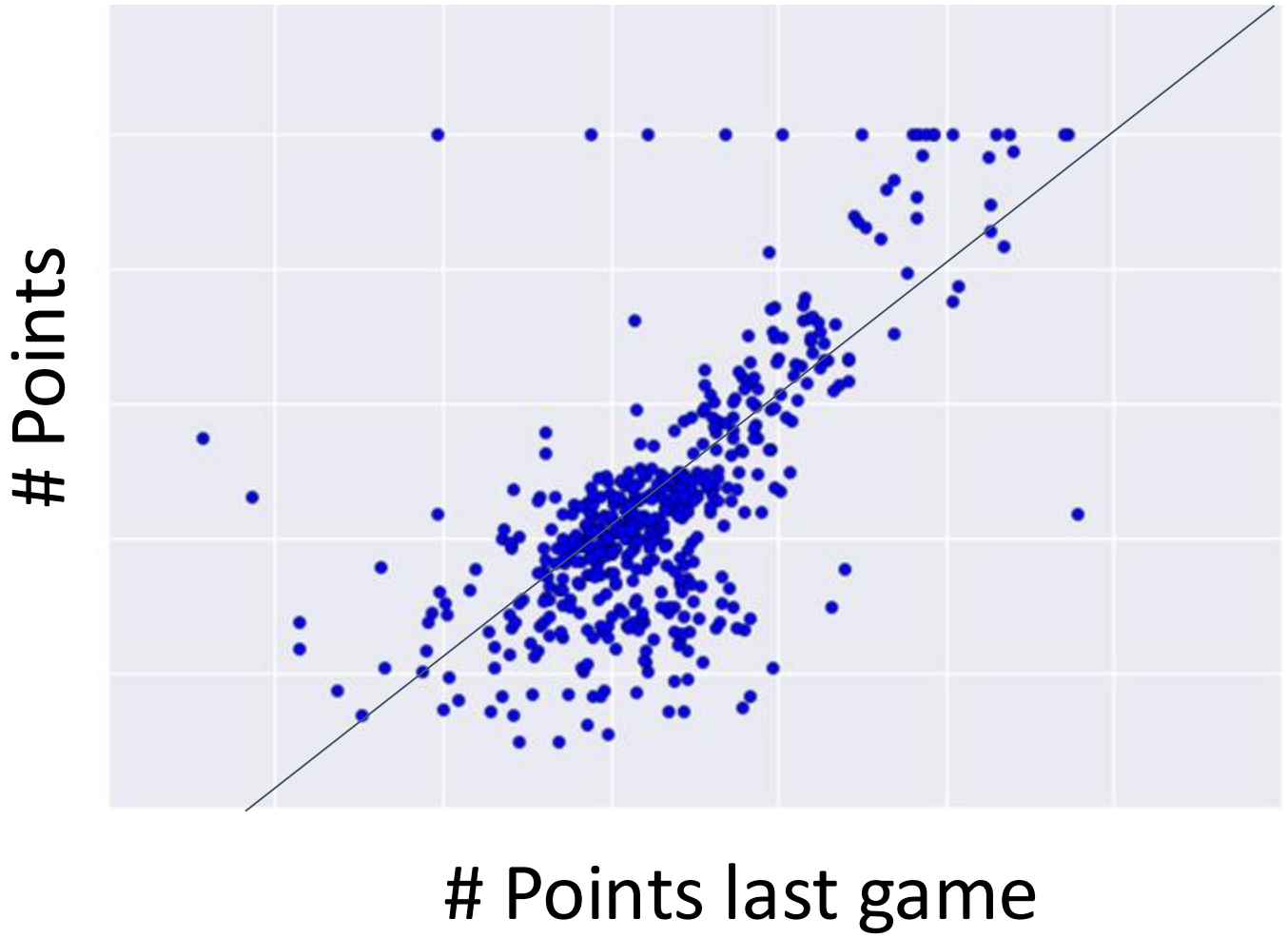
Optimization

Log likelihood: $$LL(\theta) = \sum_{i=1}^{n} \log\left[\frac{1}{\sigma\sqrt{2\pi}}\right] - \frac{1}{2}\left(\frac{y^{(i)} - \theta^T x^{(i)}}{\sigma}\right)^2$$

$$\begin{aligned}
\underset{\theta}{\text{argmax}}\, LL(\theta) &= \underset{\theta}{\text{argmax}} \sum_{i=1}^{n} \log\left[\frac{1}{\sigma\sqrt{2\pi}}\right] - \frac{1}{2}\left(\frac{y^{(i)} - \theta^T x^{(i)}}{\sigma}\right)^2 \\
&= \underset{\theta}{\text{argmax}} - \sum_{i=1}^{n} \frac{1}{2}\left(\frac{y^{(i)} - \theta^T x^{(i)}}{\sigma}\right)^2 \qquad \text{Simplify} \\
&= \underset{\theta}{\text{argmax}} - \sum_{i=1}^{n} \left(y^{(i)} - \theta^T x^{(i)}\right)^2 \qquad \text{Simplify}
\end{aligned}$$

Hey it's the sum of squared errors!

Linear Regression with one Input



Derivative is Necessary for Gradient Ascent

$$\frac{\partial LL(\theta)}{\partial \theta_j} = -\sum_{i=1}^{n} \frac{\partial}{\partial \theta_j} \left(y^{(i)} - \theta^T x^{(i)} \right)^2$$

Derivative of a sum

$$= -\sum_{i=1}^{n} 2\left(y^{(i)} - \theta^T x^{(i)} \right)(-x_j^{(i)})$$

Chain rule

$$= \sum_{i=1}^{n} 2\left(y^{(i)} - \theta^T x^{(i)} \right) \cdot x_j^{(i)}$$

Simplify

Types of Machine Learning Tasks

Muli-Class Classification

Regression

Reinforcement Learning

Generation

Types of Machine Learning Tasks

Muli-Class Classification

Regression

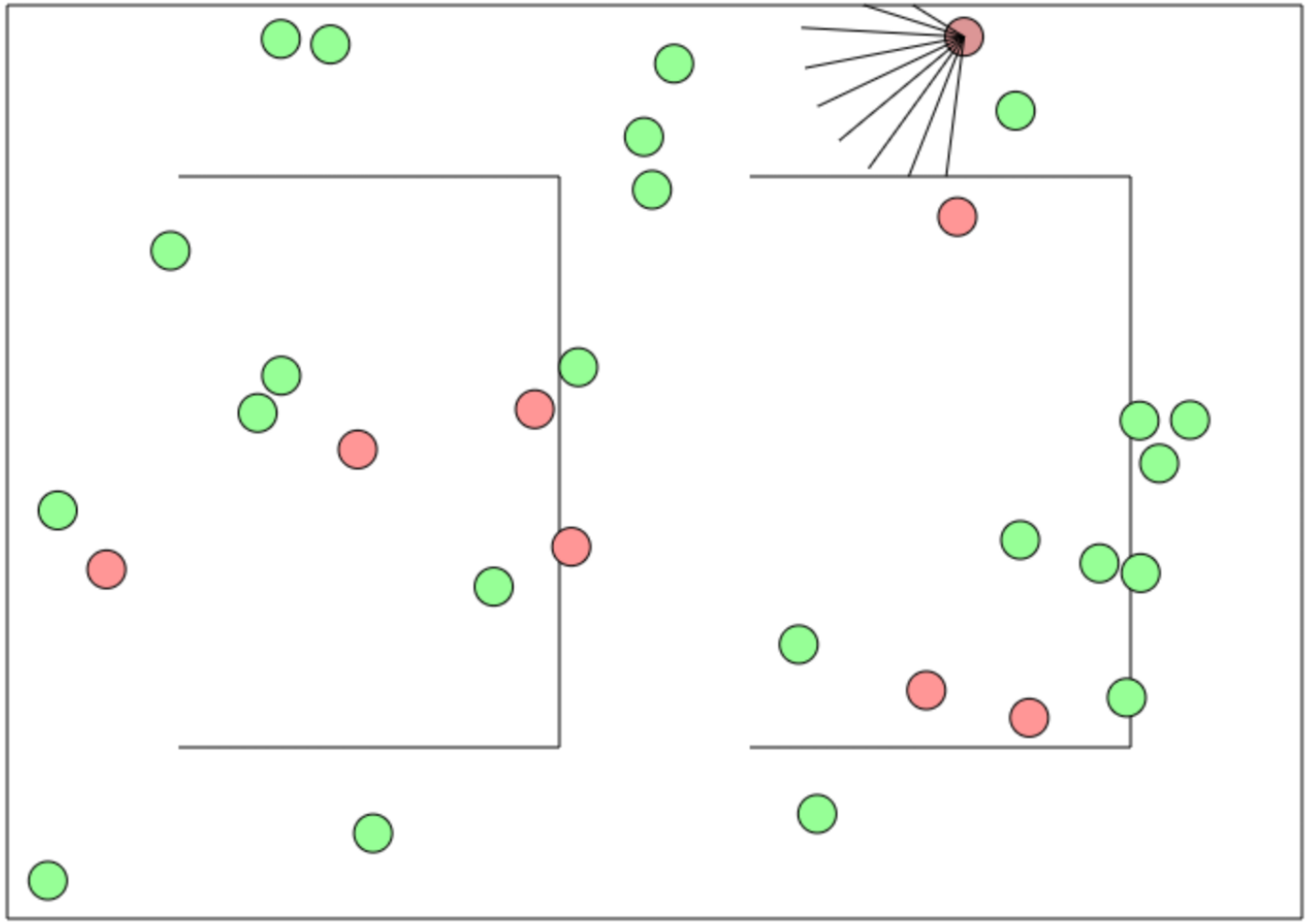
Reinforcement Learning

Generation

Making Decisions?

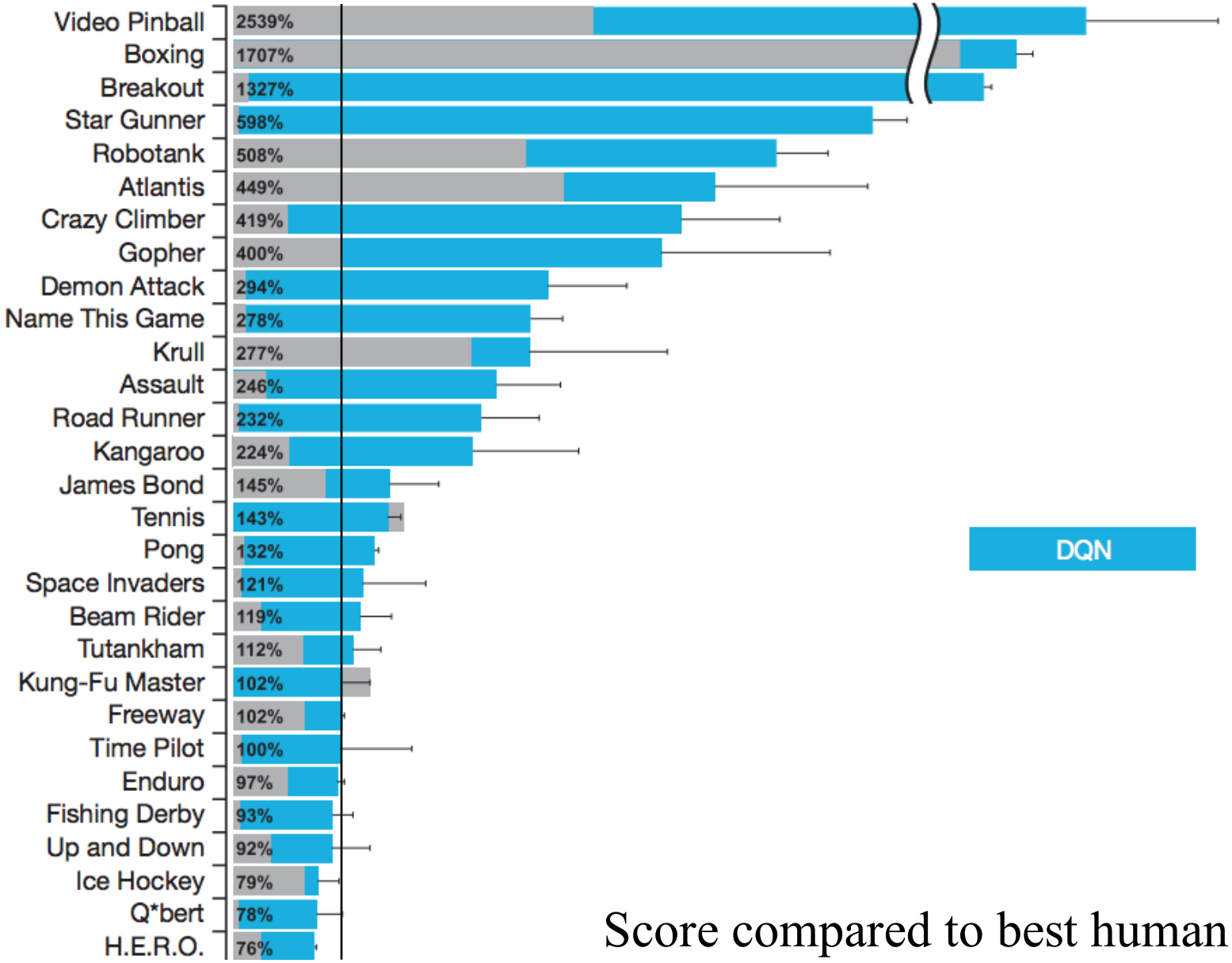
Deep Reinforcement Learning
Instead of having the output of a model be a probability you can make it an expectation.

Deep Reinforcement Learning



http://cs.stanford.edu/people/karpathy/convnetjs/demo/rldemo.html

Deep Mind Atari Games

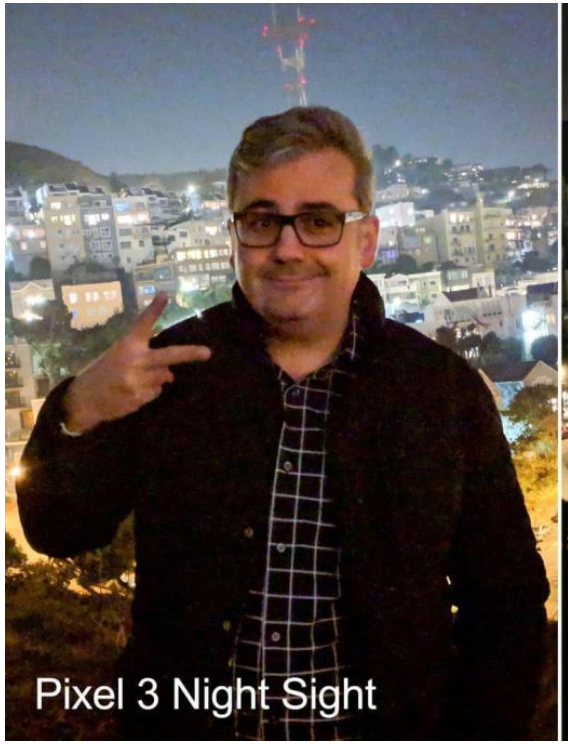


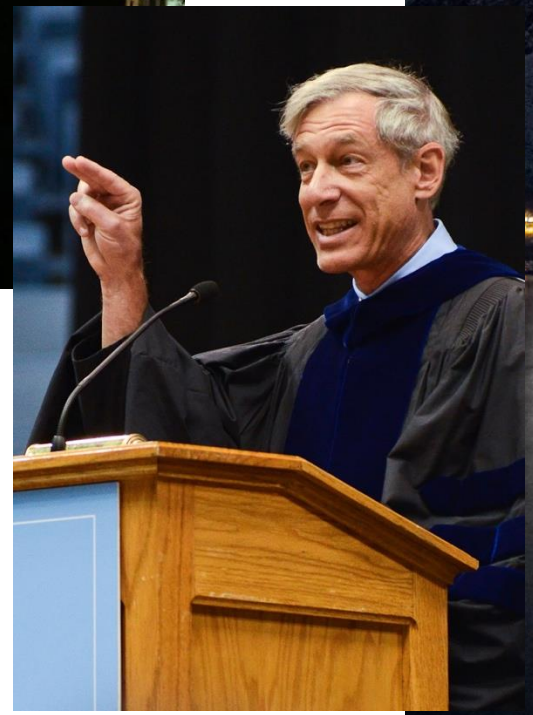
Score compared to best human

Review

Night Sight

Night Sight

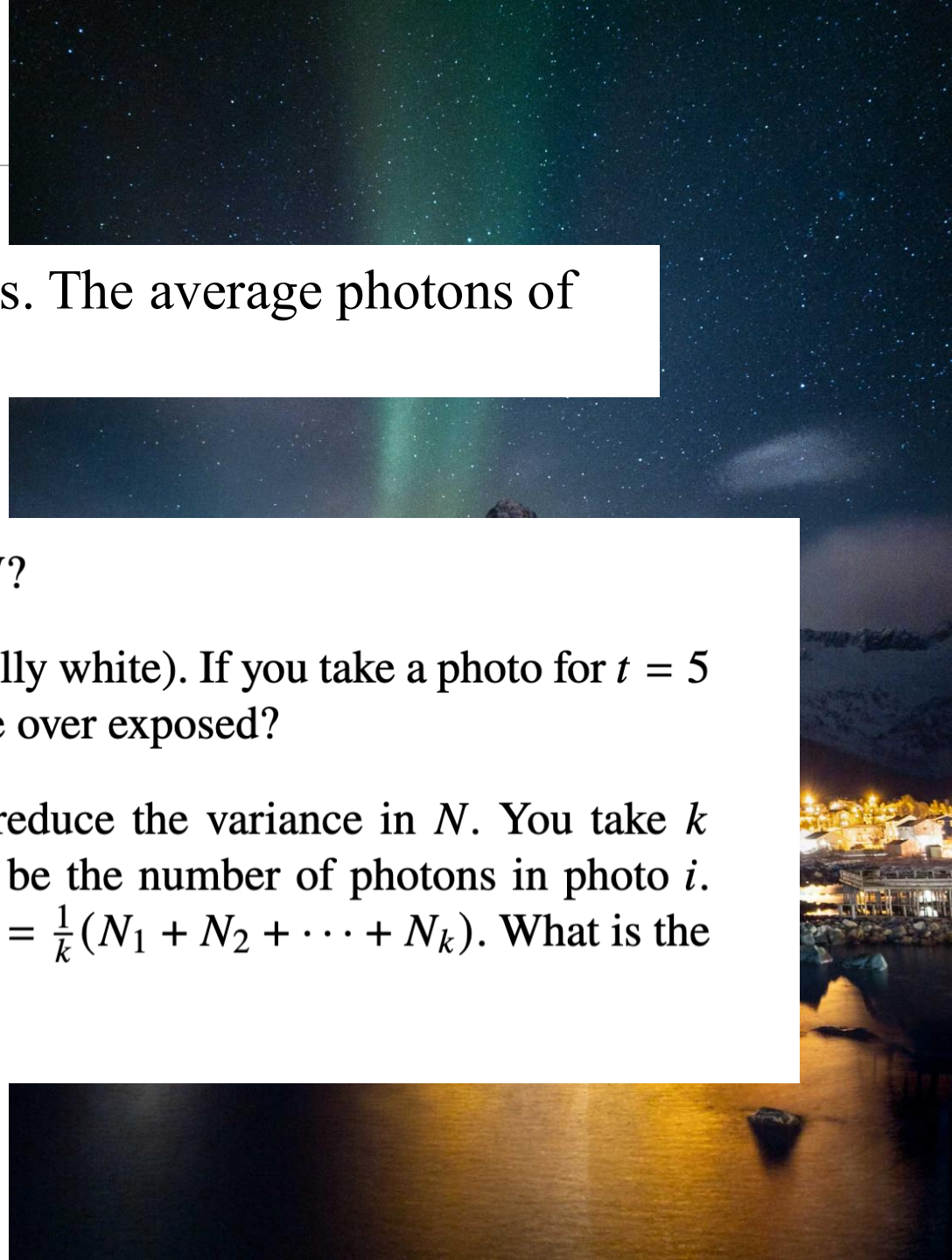




Mark Levoy, Stanford Emeritus Professor

https://static.googleusercontent.com/media/hdrplusdata.org/en//hdrplus.pdf

Night Sight



Light hits sensors on a camera as a Poisson process. The average photons of light that hit a given pixel in one second is r.

a. A photo is taken for t seconds. What is the variance of N?

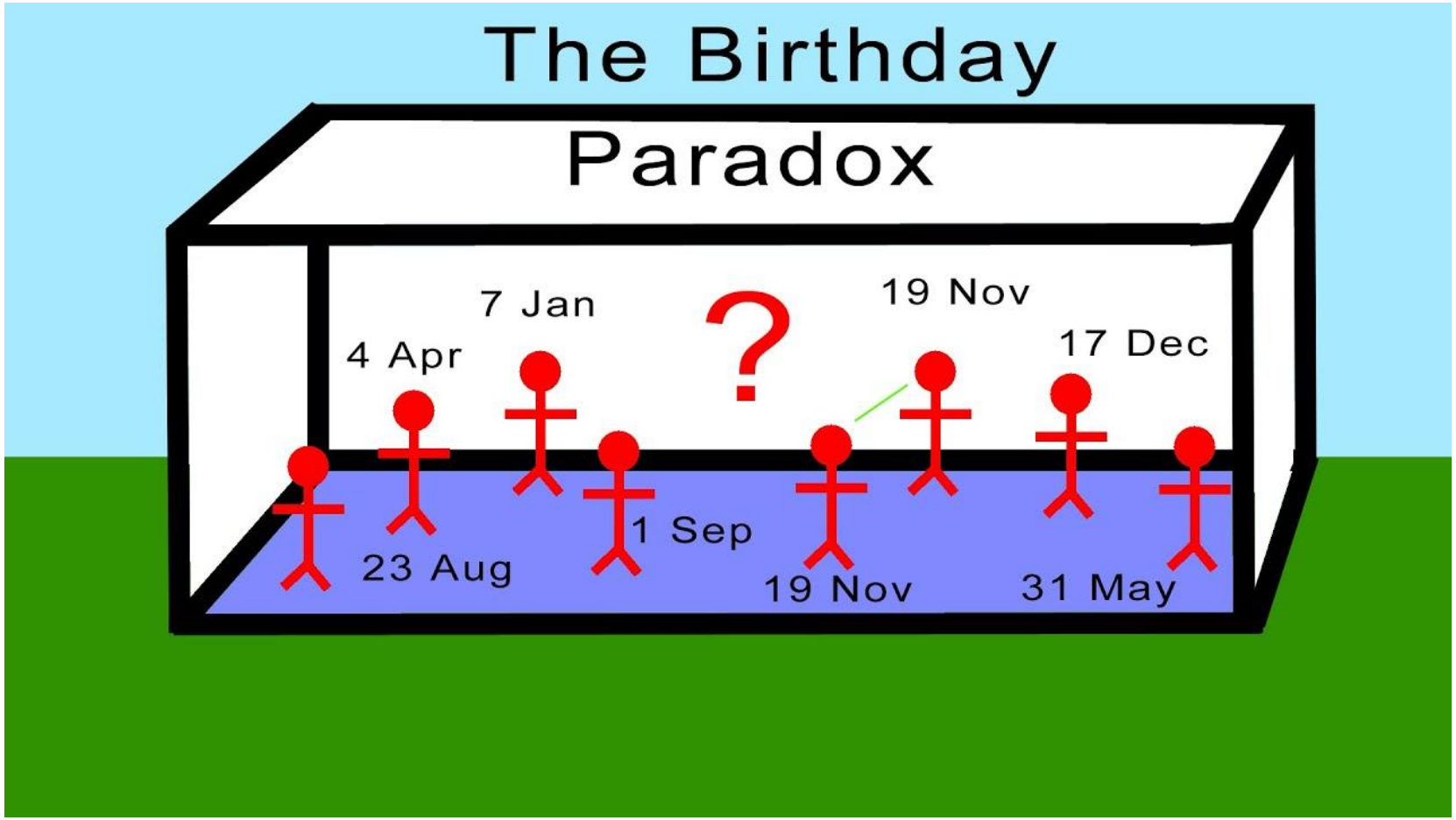
b. If $N > 100$, the pixel will be over exposed (and will be fully white). If you take a photo for $t = 5$ seconds, what is the probability that a given pixel will be over exposed?

c. If you take several photos and average them, you can reduce the variance in N. You take k photos, each with exposure length of t seconds. Let N_i be the number of photons in photo i. The average number of photons across the k photos is $N = \frac{1}{k}(N_1 + N_2 + \cdots + N_k)$. What is the variance of N?

Birthday Pradox

Birthday Pradox

There are *n* people in a room. What is the probability that **at least one** pair of people share a birthday?



Bunnies and Foxes

- 4 bunnies and 3 foxes in a toy box. 3 drawn.
  - What is P(1 bunny and 2 fox drawn)?

Equally likely sample space? Thought experiment

3 foxes

4 bunnies

The Choice of Sample Space is Yours!

	Distinct	Indistinct
Unordered	$\{F_1, B_2, B_3\}$ $\{F_1, F_2, F_3\}$	{2 foxes, 1 bunny} {3 foxes} {3 bunnies}
Ordered	$[F_1, B_2, B_3]$ $[F_1, F_2, F_3]$	[fox, bunny, fox] [fox, fox, fox] [bunny, bunny, bunny]

Which choice will lead to equally likely outcomes?

Bunnies and Foxes

- 4 bunnies and 3 foxes in a Bag. 3 drawn.
  - What is P(1 bunny and 2 foxes drawn)?
- **Ordered and Distinct**:
  - Pick 3 ordered items: $|S| = 7 * 6 * 5 = 210$
  - Pick bunny as either 1st, 2nd, or 3rd item:
    $|E| = (4 * 3 * 2) + (3 * 4 * 2) + (3 * 2 * 4) = 72$
  - P(1 bunny, 2 foxes) = 72/210 = 12/35
- **Unordered and Distinct**:
  - $|S| = \binom{7}{3} = 35$
  - $|E| = \binom{4}{1}\binom{3}{2} = 12$
  - P(1 bunny, 2 foxes) = 12/35

Make indistinct items **distinct** to get equally likely sample space outcomes

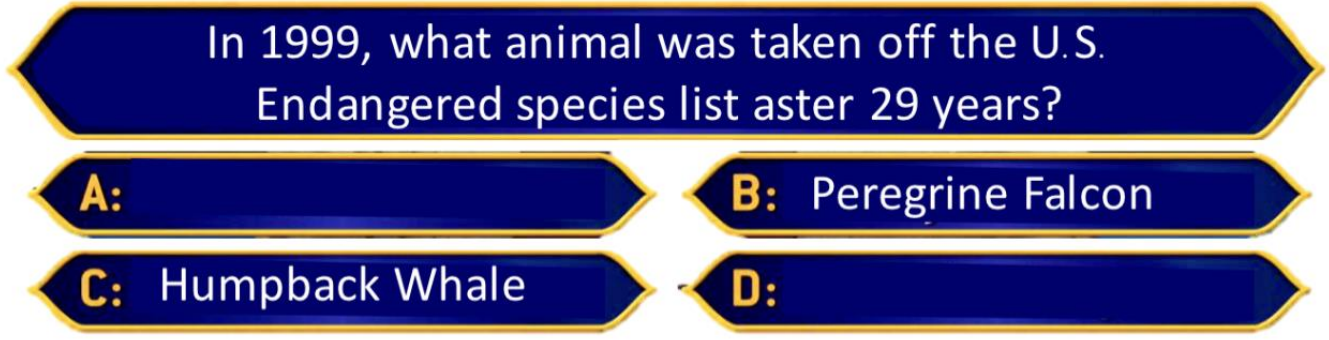
*You will need to use this “trick” with high probability

Wisdom of the Crowds

Wisdom of the Crowds

There are two answers for each audience member to chose from, a Correct answer and an Incorrect answer.

- 10% of the audience are knowledgeable about the problem (call them experts). An expert votes for the Correct answer with a probability of 0.7, otherwise they vote for the Incorrect answer.
- 90% of the audience are not knowledgeable (call them non-experts). A non-expert votes randomly with equal likelihood between the Correct answer and the Incorrect answer.



Wisdom of the Crowds

There are two answers for each audience member to chose from, a Correct answer and an Incorrect answer.

- 10% of the audience are knowledgeable about the problem (call them experts). An expert votes for the Correct answer with a probability of 0.7, otherwise they vote for the Incorrect answer.
- 90% of the audience are not knowledgeable (call them non-experts). A non-expert votes randomly with equal likelihood between the Correct answer and the Incorrect answer.

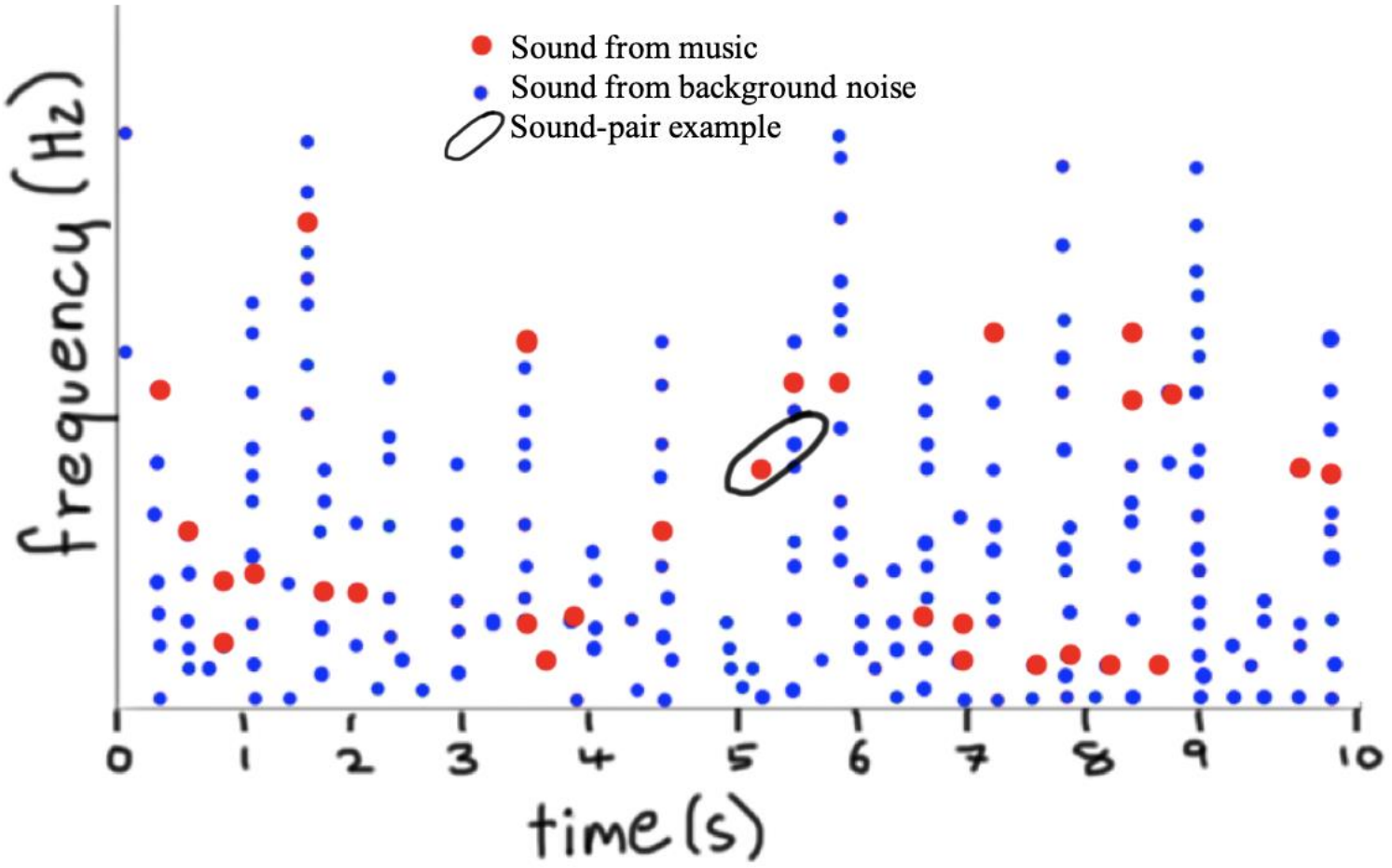
a. What is the probability that exactly k of the experts vote for the Correct answer? You may assume that k is a number between 0 and 20 inclusive.

b. If exactly k of the experts vote for the Correct answer, what is the probability that the Correct answer will get at least 101 votes? (hint: the Correct answer needs at least 101 - k more votes from the non-experts).

c. Write an expression for the exact probability that the Correct answer will get at least 101 votes.

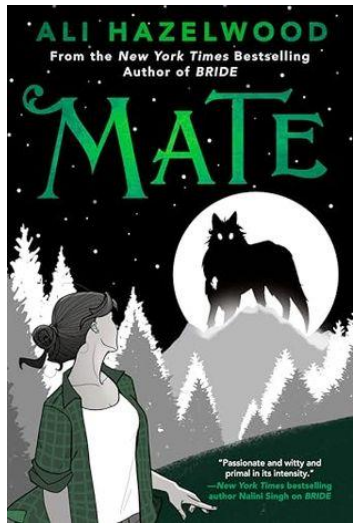
d. Use an approximation to estimate the probability that the Correct answer gets at least 101 votes. You may leave your answer in terms of roots and/or values that could be looked up from the ϕ table. For full credit your approximation calculation should *not* include a summation or integral.

Shazam

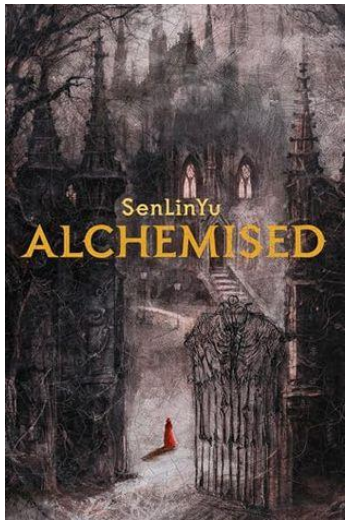

Sound from music
Sound from background noise
Sound-pair example
frequency (Hz)
0
1
2
3
4
5
6
7
8
9
10
time (s)

Rating Preference

Which Book Would you Prefer?



Rating	1	2	3	4	5
Count	0	0	0	1	4



Rating	1	2	3	4	5
Count	0	1	2	20	200

Dirchlet is the multivariate generalization of the Beta

Assume: Rating probabilities are distributed as a Dirchlet with the following joint PDF:

$$f(x_1, \ldots, x_5) = K \cdot \prod_{i=1}^{5} x_i^{c_i}$$

How many times you saw rating i

Probability that you will get rating 1

Probability that you will get rating k

You can assume you have access to

```
# take a sample from the joint
x_list = sample_dirchlet()

# evaluate the joint pdf
likelihood = pdf(x_list)
```