CS 232: AI

Fall 2023

Prof. Carolyn Anderson
Wellesley College

New policy: earn extra late days

You can earn bonus late days by attending a CS research talk. To be eligible:

- The talk must be about CS research or AI research in a related field
- The talk must be in-person (so that you have the ability to ask questions)
- You must write a paragraph about the talk and what you learned and email it to me.

Research

- *Manipulation-robust citizens' assembly selection* studies how to reduce incentives for people to try to increase their chances of being chosen for the assembly by misreporting their features.

- *The distortion of public-spirited participatory budgeting* studies the welfare of participatory budgeting outcomes in a beyond-worst-case model of voter behavior: instead of considering only their own interests, voters also weigh the interests of others. This model is motivated by the potential for this behavior to be cultivated in practice, via democratic deliberation.

Bailey Flanagan

October 25th

Computer Science Colloquium Series | Fall 2023

Supporting Responsible AI Practices in Public Sector Contexts

Anna is a third year PhD student at Carnegie Mellon's Human-Computer Interaction Institute. Her research focuses on improving the design, evaluation, and governance of AI technologies used to inform complex, consequential decisions in real-world organizations. In addition to her research, she will share prior experiences forming collaborations with public sector agencies, doing research internships with industry groups, travelling to conferences, and mentoring undergraduate students. The session will end with an open Q/A discussion on applying to and doing a PhD in Computer Science / Human-Computer Interaction and other topics.

WELLESLEY

Accessibility and Disability Resources:
accessibility@wellesley.edu

Anna Kawakami '21

Nov 2nd, 12:45-2:00 | SCI H401
Lunch will be served

Questions??? eni.mustafara

November 2nd

YLLATAILY Discussion

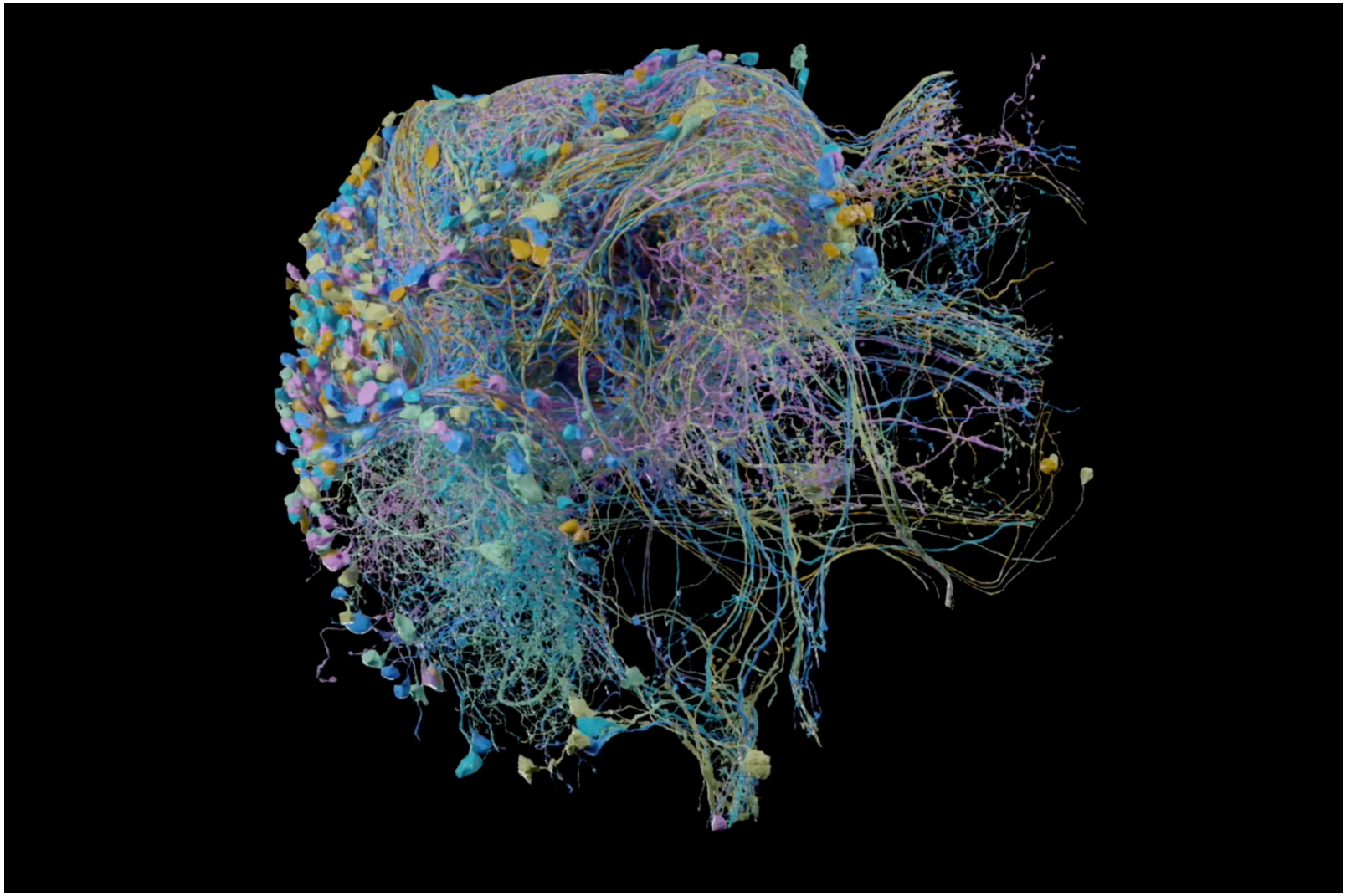
Drawing on the last two chapters of YLLATAILY, come up with some rules of thumb for identifying misleading AI headlines

Neural Networks

This is someone's brain

estimated to have 100 trillion synapses connecting 100 billion neurons

This is not your brain

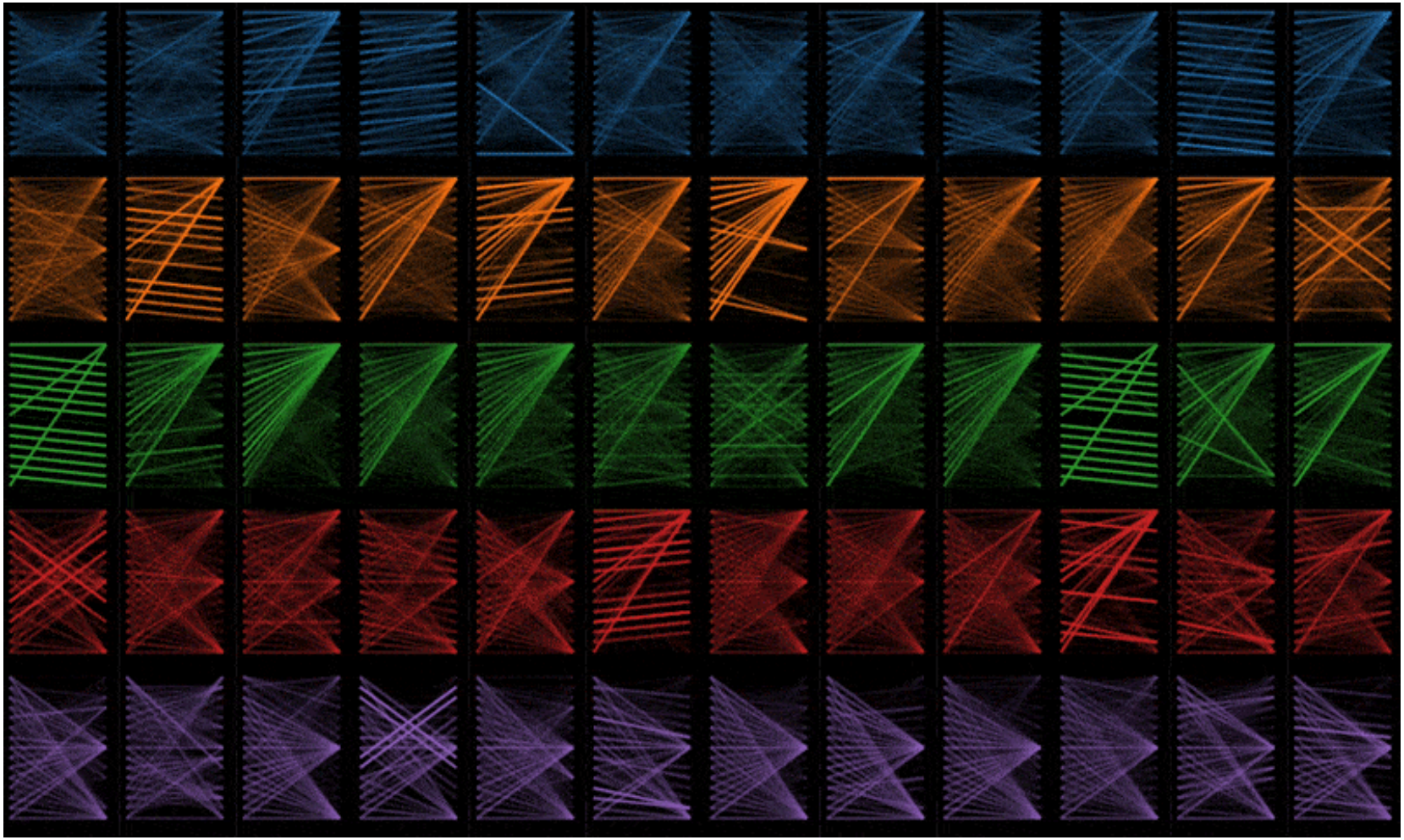


https://www.janelia.org/project-team/flyem/hemibrain

It's a fruit fly brain

20 million synapses connecting ~25,000 neurons

This is also not your brain

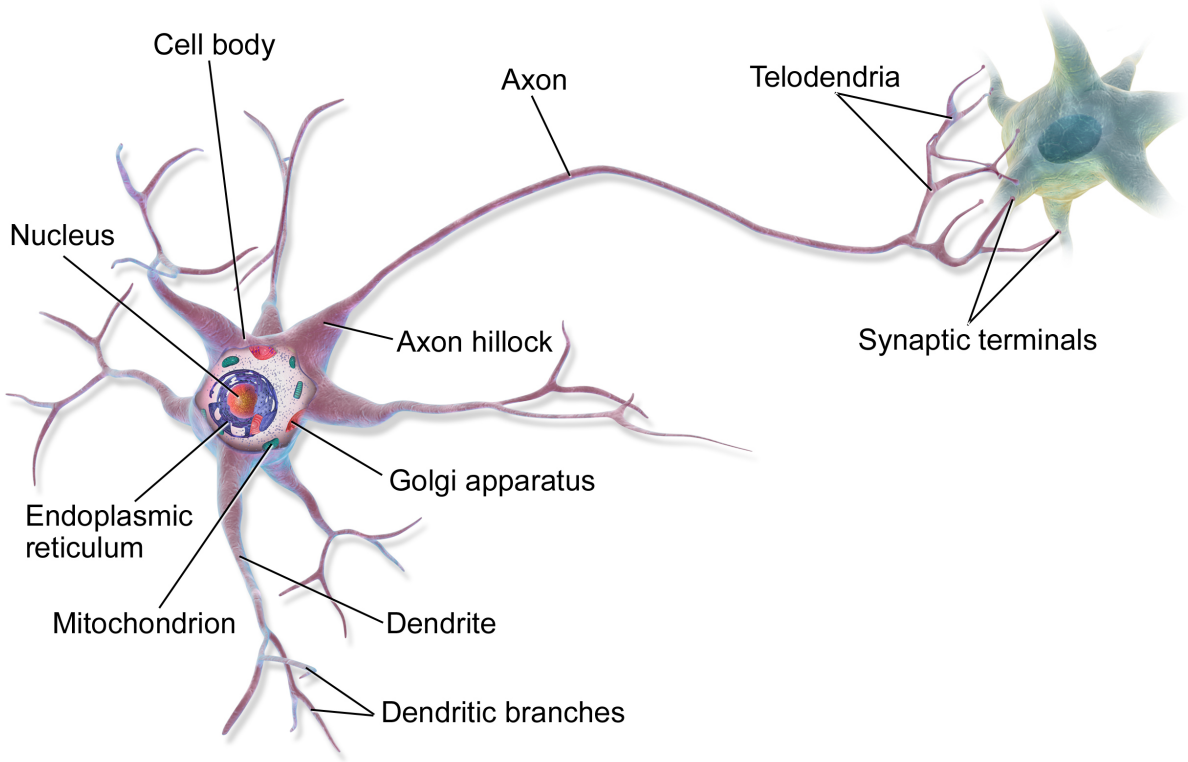


https://github.com/jessevig/bertviz

It's BERT, a big neural network

110 million parameters

This is in your brain

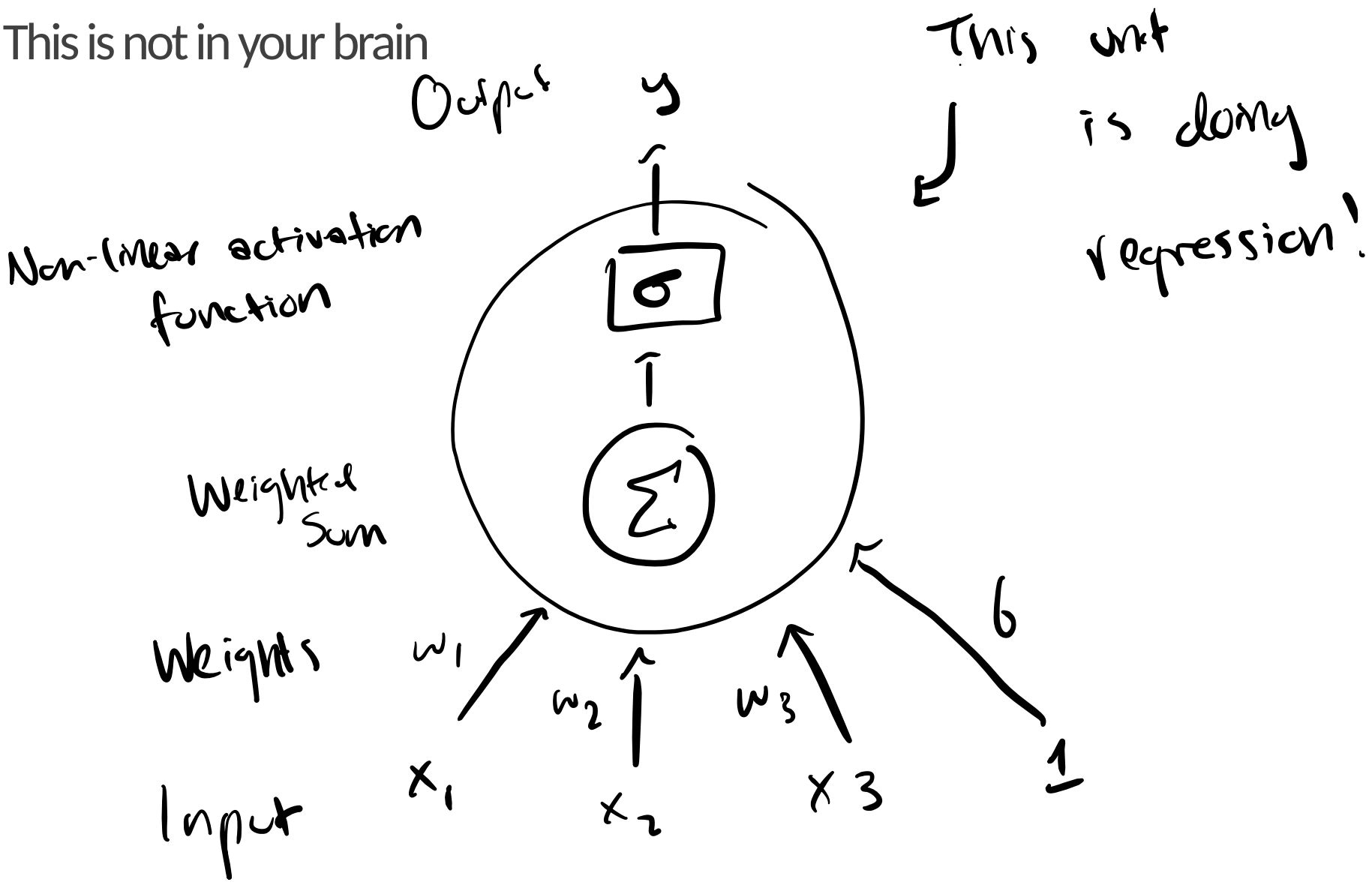


By BruceBlaus - Own work, CC BY 3.0, https://commons.wikimedia.org/w/index.php?curid=28761830

Slide borrowed from Jurafsky & Martin Edition 3

Neural Network Unit

This is not in your brain



Units in Neural Networks

Slides borrowed from Jurafsky & Martin Edition 3

Neural unit

$$z = b + \sum_i w_i x_i$$

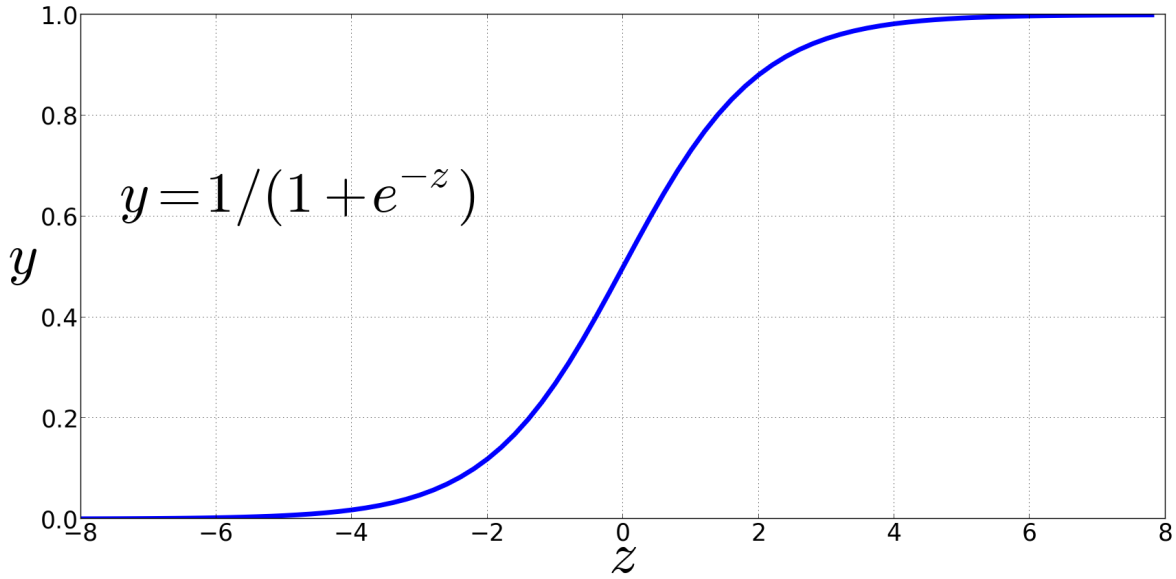
$$z = wx + b$$

$$y = \sigma(z)$$

Non-Linear Activation Functions

We're already seen the sigmoid for logistic regression:

Sigmoid

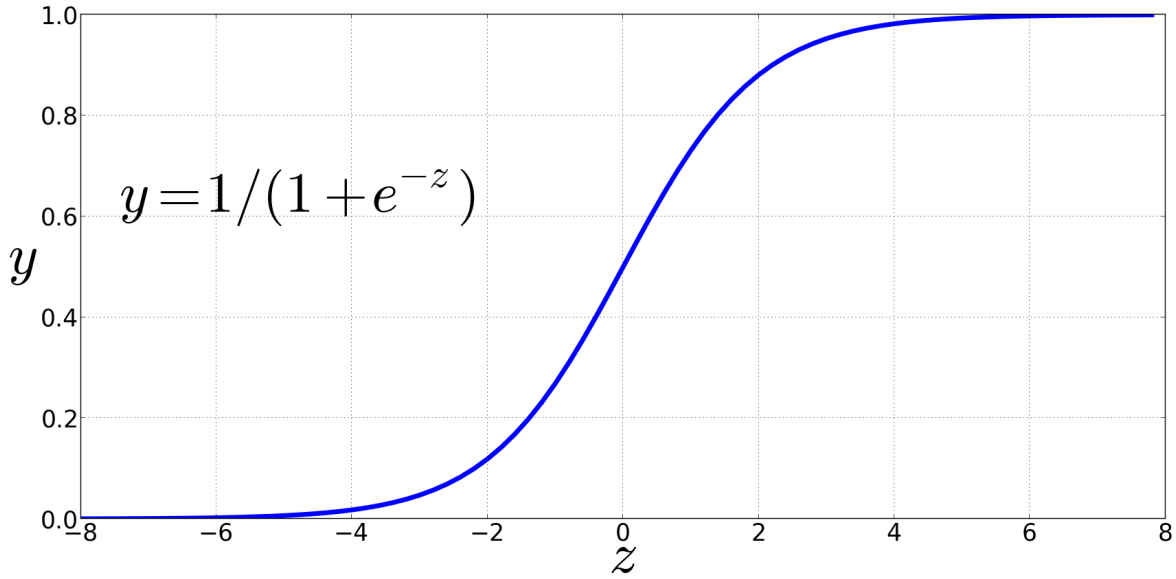


Non-Linear Activation Functions

We're already seen the sigmoid for logistic regression:

Sigmoid

$$y = s(z) = \frac{1}{1 + e^{-z}}$$



Final function of a unit:

$$y = \sigma(wx + b) = \frac{1}{1 + \exp(-(wx + b))}$$

Spot the differences

Neural Network Unit

$$z = b + \sum w_i x_i$$

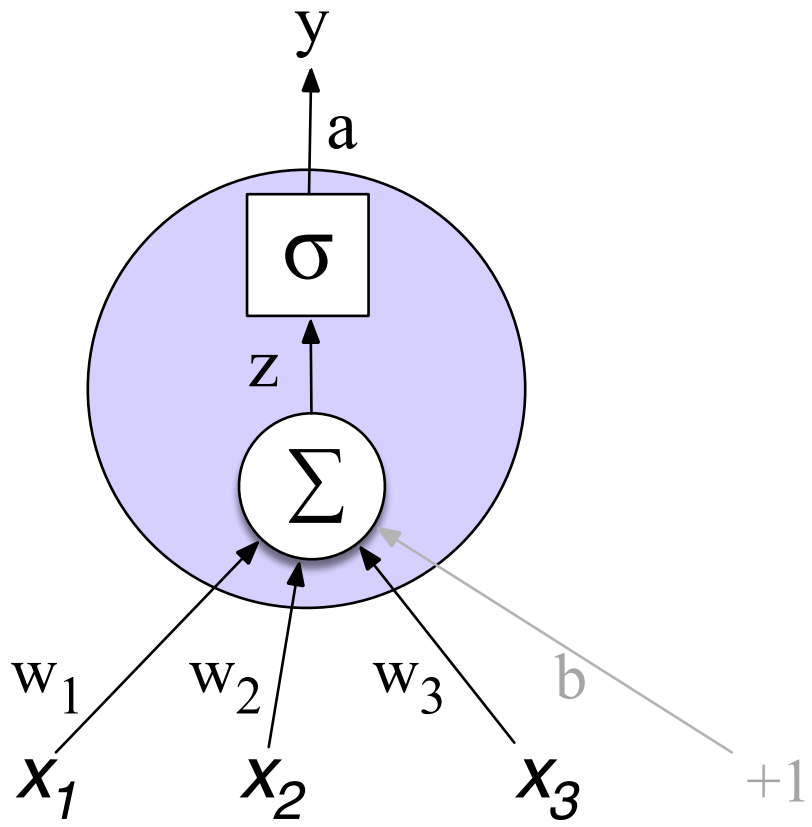
$$y = \sigma(wx + b)$$

Logistic Regression

$$z = \left(\sum w_i x_i\right) + b$$

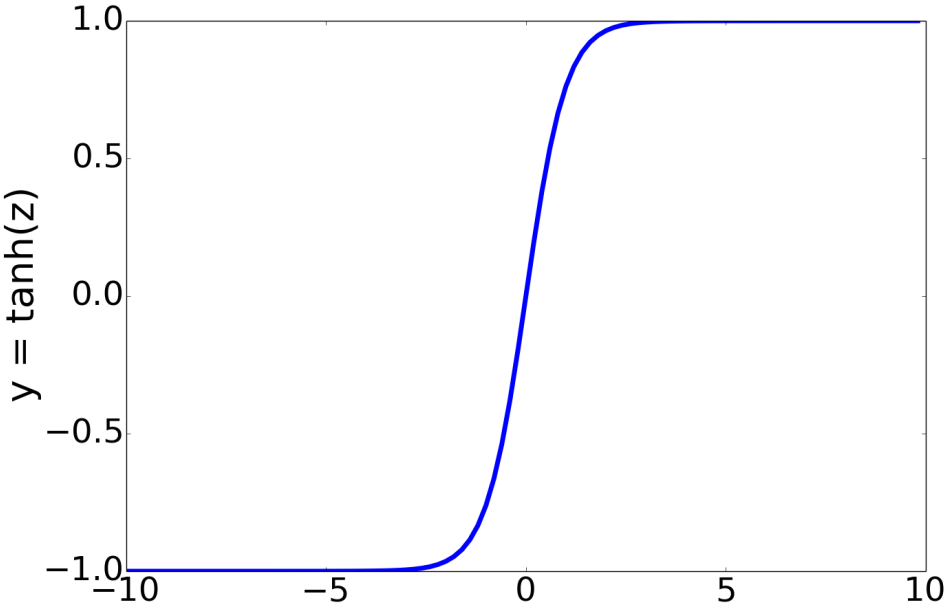
$$y = \sigma(wx + b)$$

Final unit again



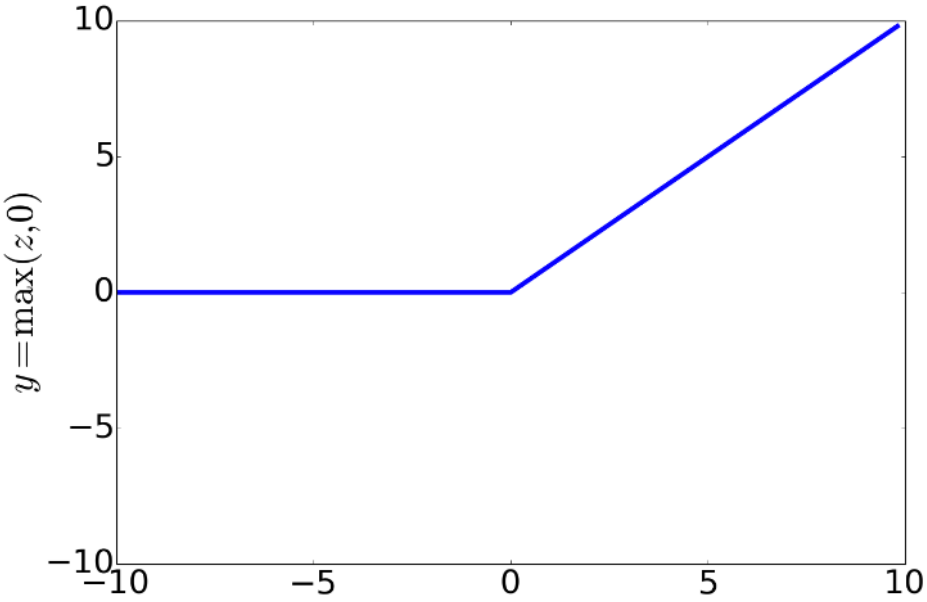
Non-Linear Activation Functions besides sigmoid

Most Common:



tanh

$$y = \frac{e^z - e^{-z}}{e^z + e^{-z}}$$



ReLU

Rectified Linear Unit

$$y = max(z, 0)$$

Example: XOR

The XOR problem

Minsky and Papert (1969)

Can neural units compute simple functions of input?

AND			OR			XOR		
x1	x2	y	x1	x2	y	x1	x2	y
0	0	0	0	0	0	0	0	0
0	1	0	0	1	1	0	1	1
1	0	0	1	0	1	1	0	1
1	1	1	1	1	1	1	1	0

Perceptrons

A very simple neural unit

- Binary output (0 or 1)
- No non-linear activation function

$$y = \begin{cases} 0, & \text{if } \mathbf{w} \cdot \mathbf{x} + b \leq 0 \\ 1, & \text{if } \mathbf{w} \cdot \mathbf{x} + b > 0 \end{cases}$$

Solving AND

Deriving AND

$$y = \begin{cases} 0, & \text{if } \mathbf{w} \cdot \mathbf{x} + b \leq 0 \\ 1, & \text{if } \mathbf{w} \cdot \mathbf{x} + b > 0 \end{cases}$$

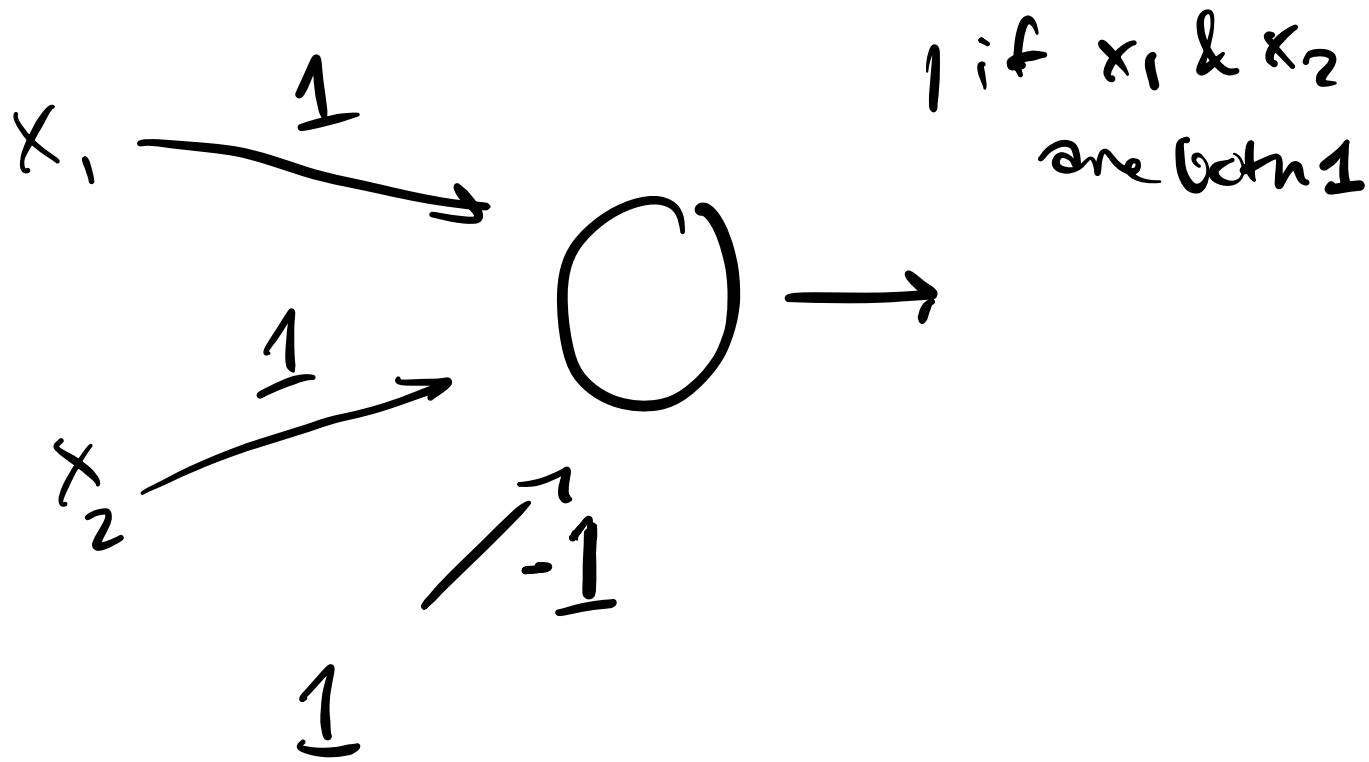
Goal: return 1 if x1 and x2 are 1

AND

x1	x2	y
0	0	0
0	1	0
1	0	0
1	1	1

Deriving AND

Goal: return 1 if x1 and x2 are 1



Exercise: solving OR

OR

x1	x2	y
0	0	0
0	1	1
1	0	1
1	1	1

Deriving OR

$$y = \begin{cases} 0, & \text{if } \mathbf{w} \cdot \mathbf{x} + b \leq 0 \\ 1, & \text{if } \mathbf{w} \cdot \mathbf{x} + b > 0 \end{cases}$$

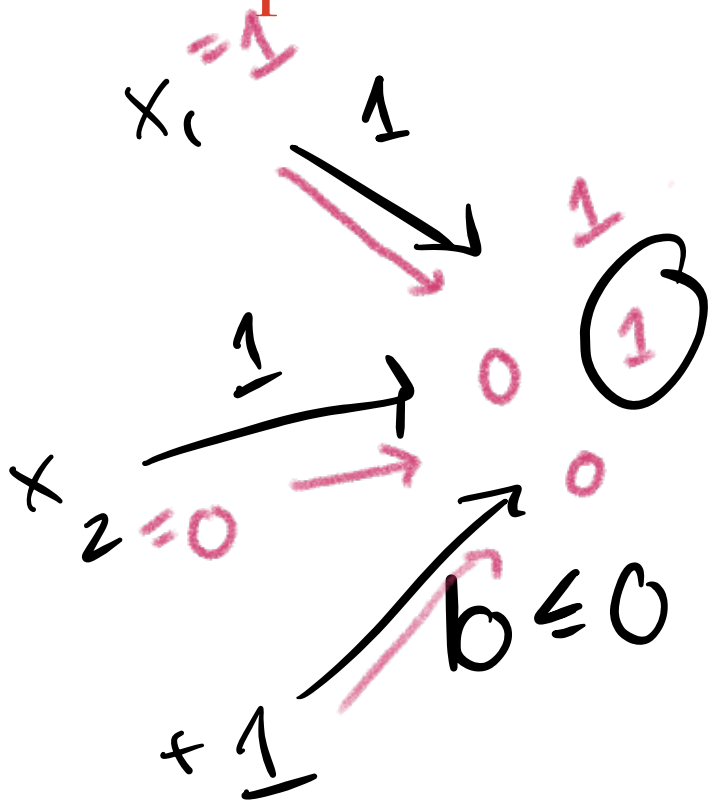
Goal: return 1 if either input is 1

OR		
x1	x2	y
0	0	0
0	1	1
1	0	1
1	1	1

Deriving OR

$w_1 0 + w_2 0 + b = 0$

Goal: return 1 if either input is 1



$w_1 x_1 + w_2 x_2 + b \geq 0$

$= 1$ if x_1 or x_2 is 1

$w_1 1 + w_2 1 + 0 \geq 1$

$w_1 + w_2 \geq 1$