Bootcamp 2023

Basics of Neural Networks

UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE & ENGINEERING
Centre for Analytics and Artificial Intelligence Engineering

A bit about myself

- Bachelor's in AI from the University of Edinburgh
- Master's from UofT in applications of AI
- Joined CARTE in 2020, where I work with faculty, students and staff to bring AI and ML to a wide range of problems
- Recently: consulting with various levels of the university on LLM strategies in education

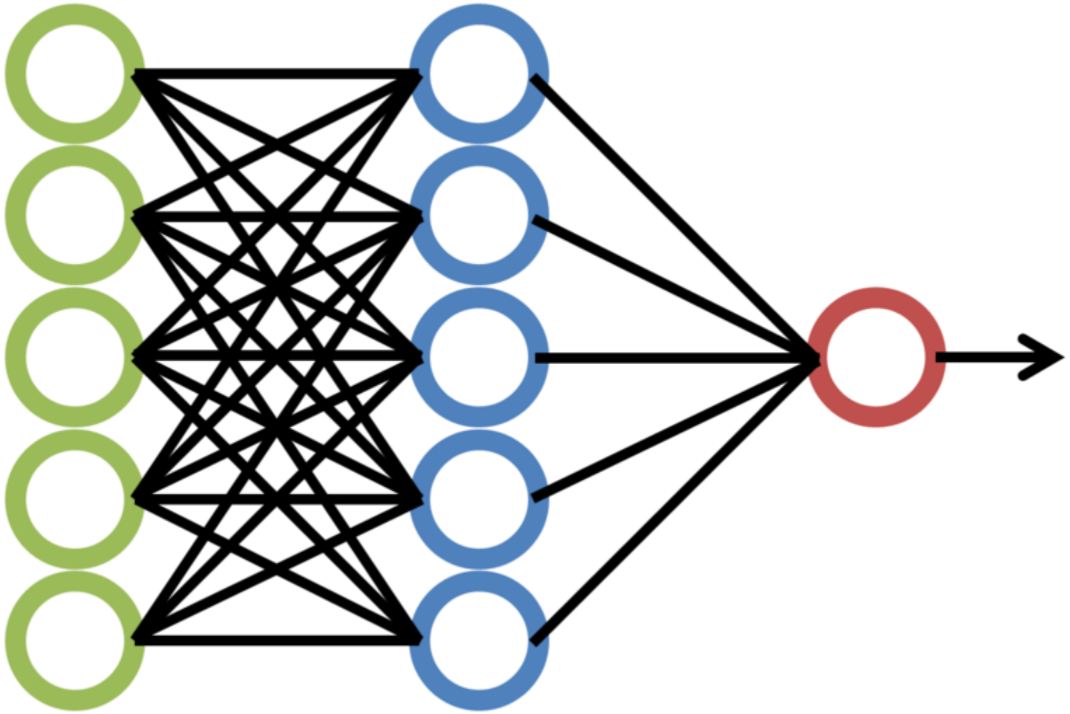
Teaching assistants

- Shehnaz Islam
- Focused on researching the automation of Patient Safety Incident Report classification through NLP, ML, and Deep Learning techniques.

- Nakul Upadhya
- Researching Interpretable Sequence Models with a focus on Forecasting

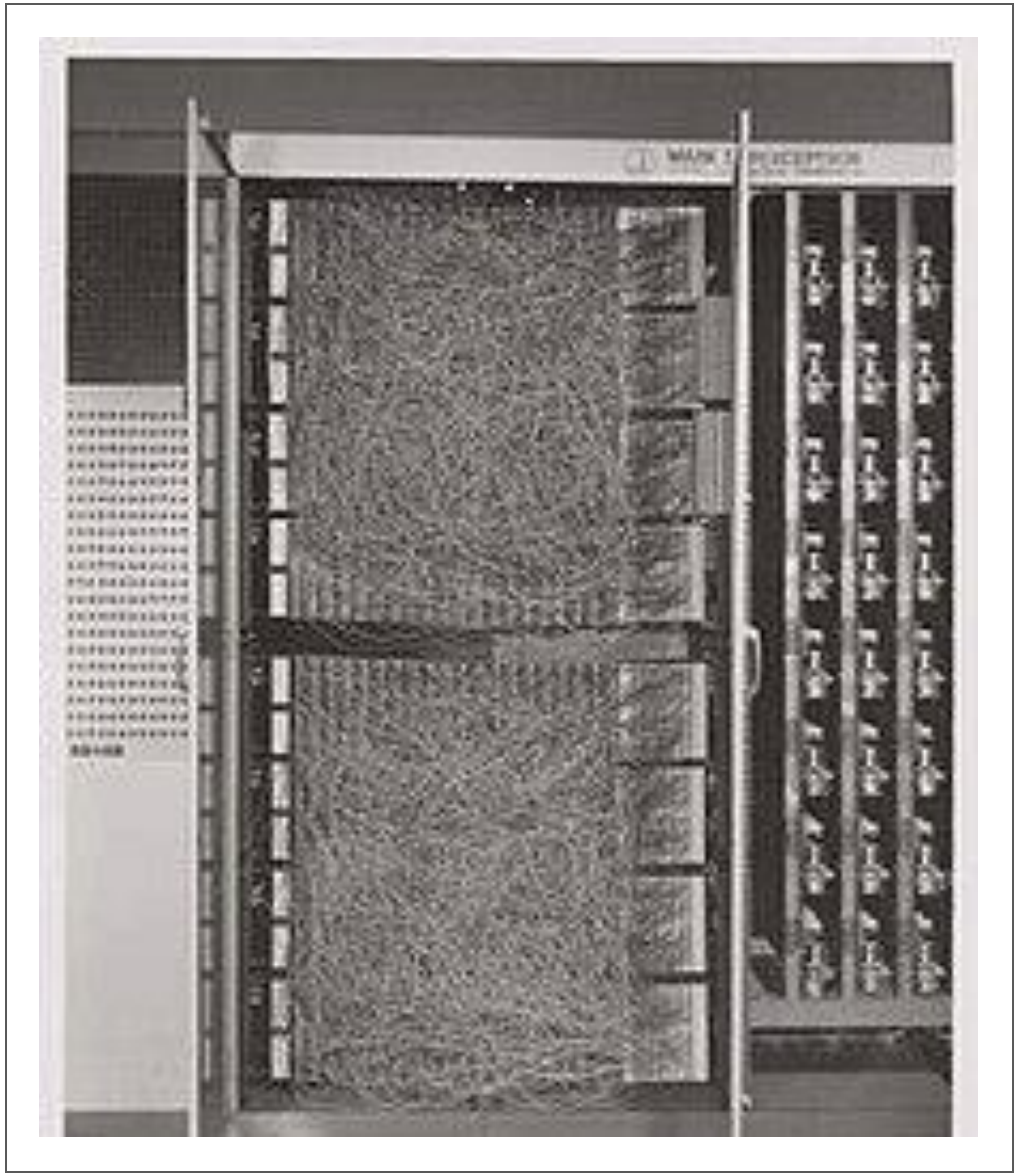
What is a neural network?

- Complex structure of interconnected computing nodes (neurons)
- Can identify patterns and trends in complex data
- NNs operate on the principle of “learning” from data, using a process that mimics how biological brains learn



History of NNs

- 1940s – Early Beginnings
  - Concept of a neural network is first proposed: "A Logical Calculus of Ideas Immanent in Nervous Activity"
- 1950s – The Perceptron
  - With funding from the US Navy, Cornell builds the Mark 1 Perceptron, a physical neural network
  - The New York Times reported the perceptron to be "the embryo of an electronic computer that [the Navy] expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence."



History of NNs

- 1960s – The First AI Winter
  - Despite the excitement of the 50s, NN research stalled
  - A highly influential book – *Perceptrons (1969)* – showed that these early neural networks were severely limited
- 1980s – Backpropagation
  - The discovery of backpropagation allowed for the first time the creation of multi-layer neural networks that could efficiently learn from examples
- 1990s – Support Vector Machines and the Second AI Winter
  - NN research stalled again due to the rising popularity of SVMs, which provided a better theoretical framework and outperformed the NNs of the day

History of NNs

- 2000s – Dawn of the Deep Learning Era
  - The term “deep learning” began to circulate, reflecting a new focus on deeper, multi-layered neural networks
  - Advances in hardware, datasets, and training techniques allowed the development of much more sophisticated networks
- 2010s – Breakthroughs and Wide Adoption
  - With the success of AlexNet, Convolutional Neural Networks gained prominence and became a go-to method for image tasks
  - Recurrent Neural Networks show impressive results in natural langauge understanding
  - Tech giants begin to heavily invest in deep learning technology

History of NNs

- 2020s – Transformers and the Era of Large Language Models
  - The Transformer model, introduced in the paper "Attention is All You Need", starts demonstrating state-of-the-art performance in language tasks
  - An increasing focus on large-scale models with billions, or even trillions, of parameters begins, leading to unprecedented performance…
  - …but also raising questions about computational efficiency, environmental impact, and accessibility.

Neural Networks: From Linear to Non-Linear

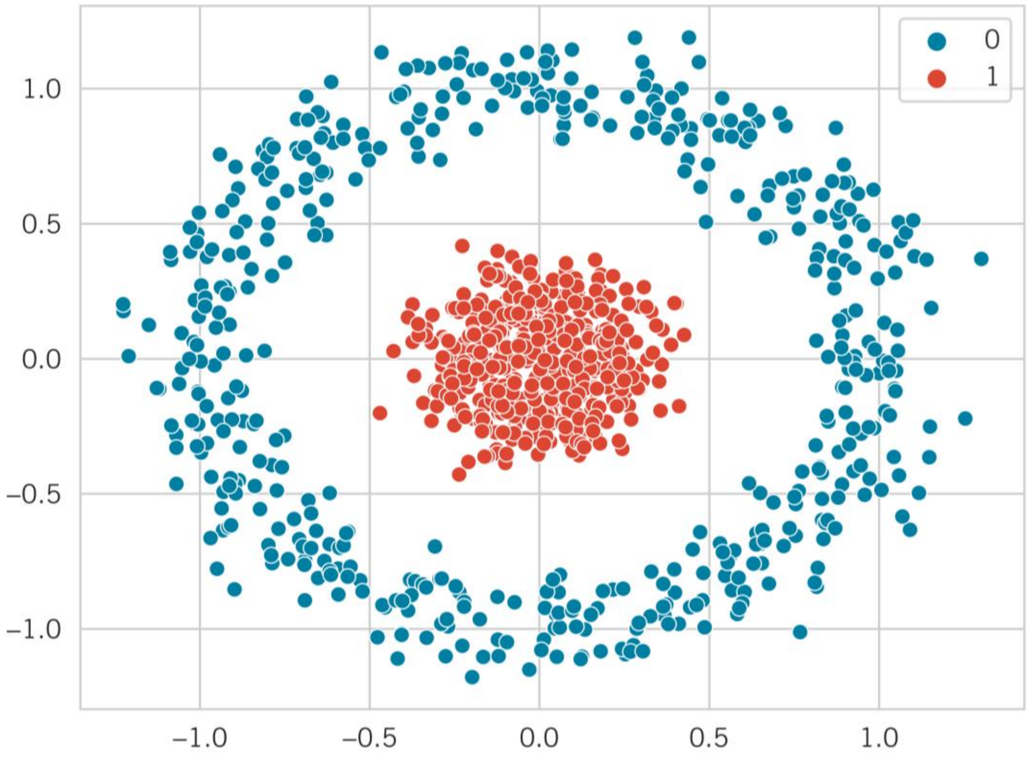
- Linear score function: $f = Wx$

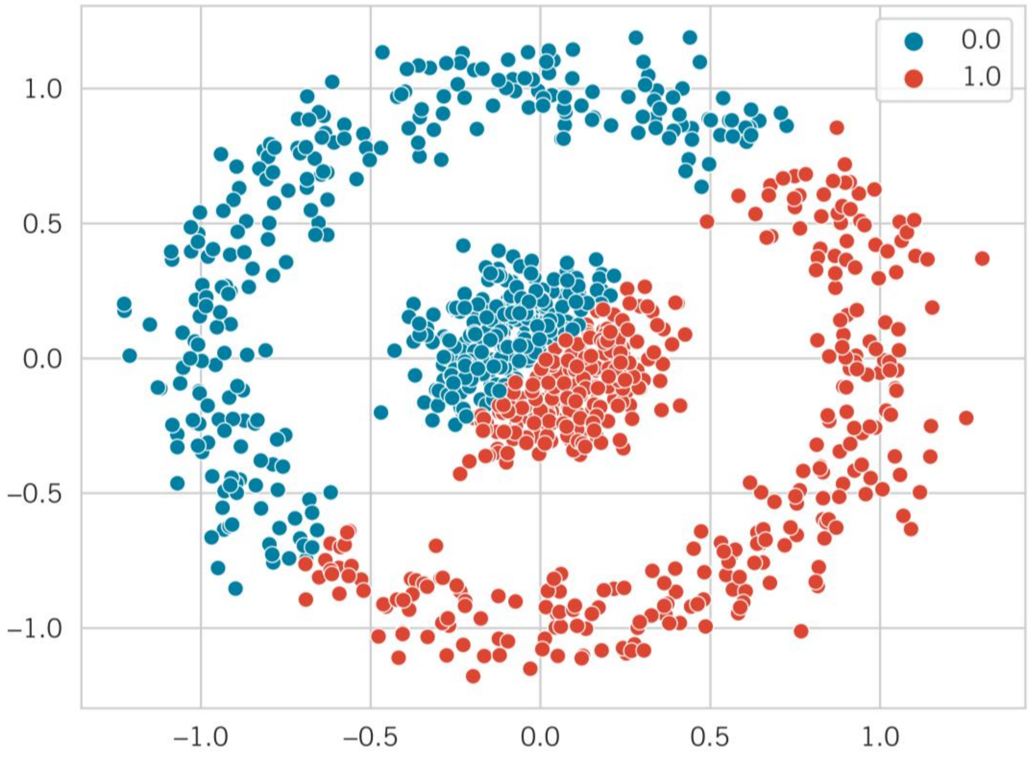
$$x \in \mathbb{R}^D, W \in \mathbb{R}^{C \times D}$$

- 2-layer Neural Network: $f = W_2 \max(0, W_1 x)$

$$x \in \mathbb{R}^D, W_1 \in \mathbb{R}^{H \times D}, W_2 \in \mathbb{R}^{C \times H}$$

Why do we want non-linearity?

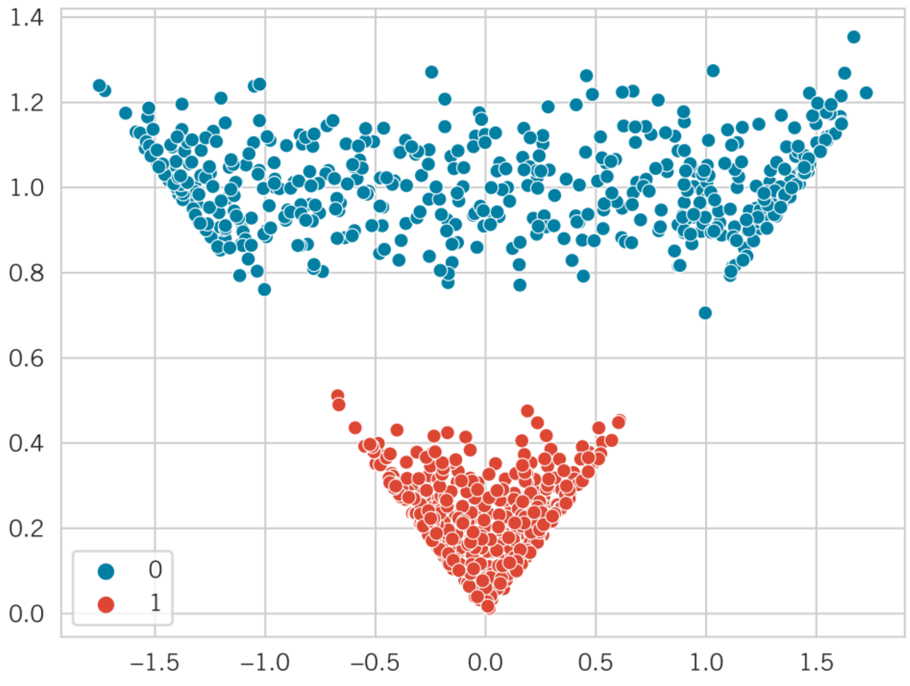




Cannot apply a linear classifier!

Why do we want non-linearity?

- After applying feature transformation, points become linearly separable



Neural Networks: Also called fully-connected

- Linear score function: $f = Wx$

$$x \in \mathbb{R}^D, W \in \mathbb{R}^{C\times D}$$

- 2-layer Neural Network: $f = W_2 \max(0, W_1 x)$

$$x \in \mathbb{R}^D, W_1 \in \mathbb{R}^{H\times D}, W_2 \in \mathbb{R}^{C\times H}$$

"Neural Network" is a very broad term; these are more accurately called "fully-connected networks" or sometimes "multi-layer perceptrons" (MLP)

Neural Networks: 3 layers

- Linear score function: $f = Wx$

  $x \in \mathbb{R}^D, W \in \mathbb{R}^{C \times D}$

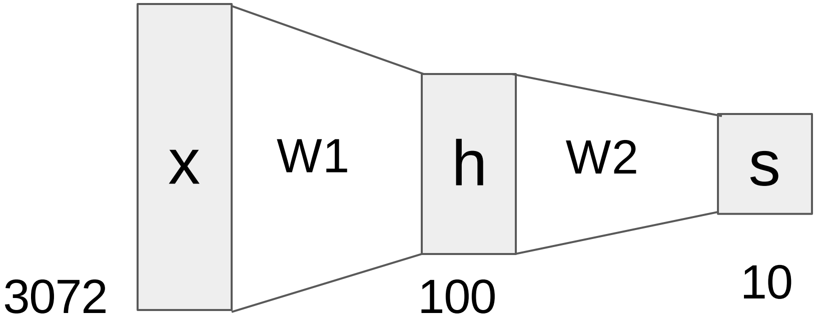
- 2-layer Neural Network: $f = W_2 \max(0, W_1 x)$
  or 3-layer: $f = W_3 \max(0, W_2 \max(0, W_1 x))$

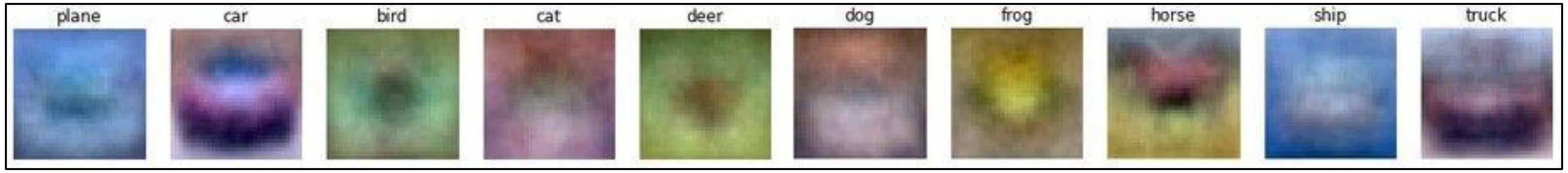
  $x \in \mathbb{R}^D, W_1 \in \mathbb{R}^{H_1 \times D}, W_2 \in \mathbb{R}^{H_2 \times H_1}, W_3 \in \mathbb{R}^{C \times H_1}$

Neural Networks: Hierarchical computation

- 2-layer Neural Network: $f = W_2 \max(0, W_1 x)$

  $x \in \mathbb{R}^{D}, W_1 \in \mathbb{R}^{H \times D}, W_2 \in \mathbb{R}^{C \times H}$





Learn 100 templates instead of 10.

Share templates between classes

Neural Networks: Why is max so important?

- 2-layer Neural Network: $f = W_2 \max(0, W_1 x)$

$$x \in \mathbb{R}^D, W_1 \in \mathbb{R}^{H \times D}, W_2 \in \mathbb{R}^{C \times H}$$

- We call the function $\max(0, z)$ the <u>activation function</u>.

What if we try to build a neural network without one?

$$f = W_2 W_1 x \quad W_1 \in \mathbb{R}^{H \times D}, W_2 \in \mathbb{R}^{C \times H}$$

Neural Networks: Why is max so important?

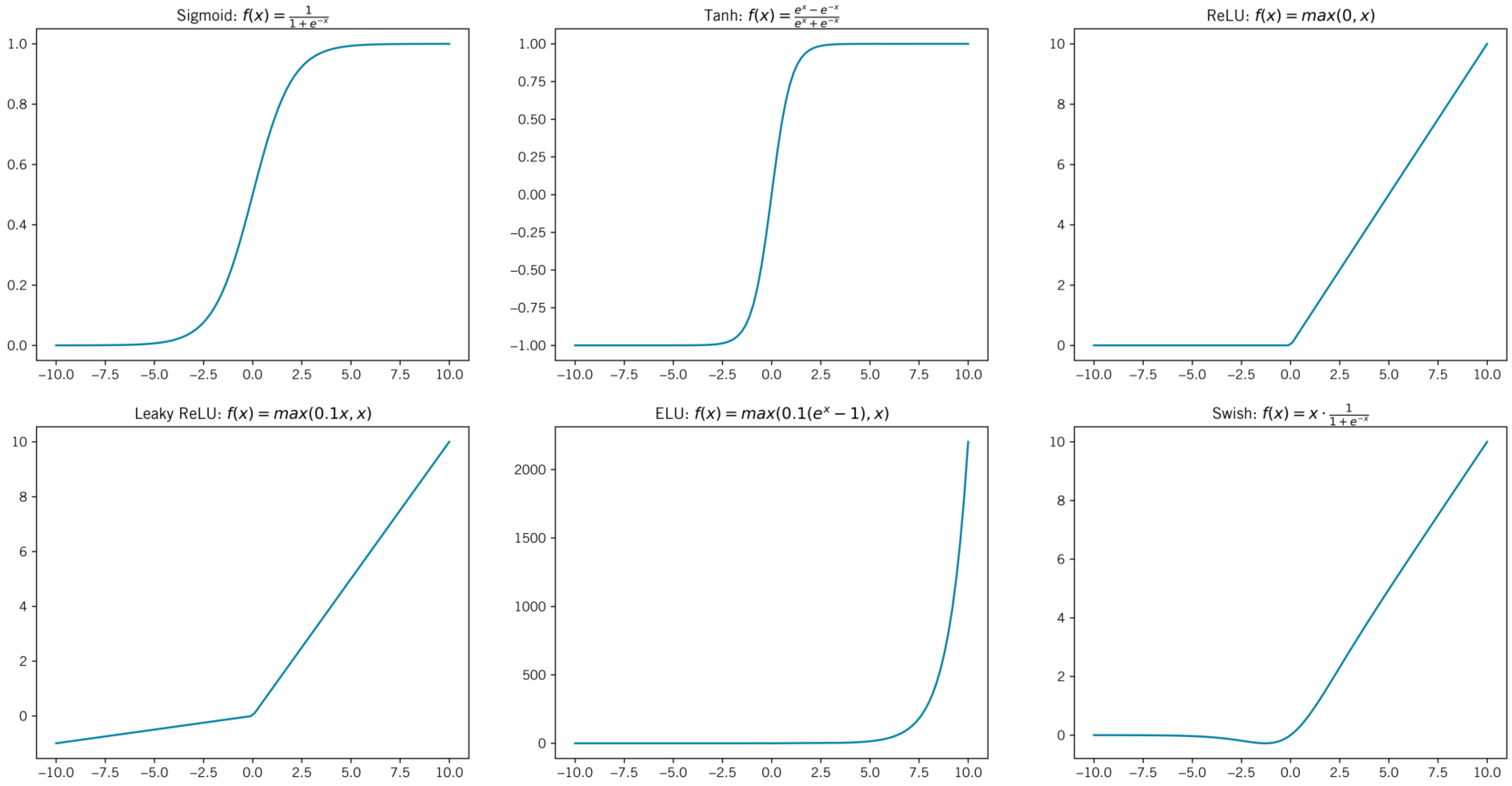
- $\max(0, z)$ the activation function.

What if we try to build a neural network without one?

$$
\begin{aligned}
f &= W_2 W_1 x \quad W_1 \in \mathbb{R}^{H \times D}, W_2 \in \mathbb{R}^{C \times H} \\
W_3 &= W_2 W_1 \quad W_3 \in \mathbb{R}^{C \times D} \\
\therefore f &= W_3 x
\end{aligned}
$$

We end up with a linear classifier again!

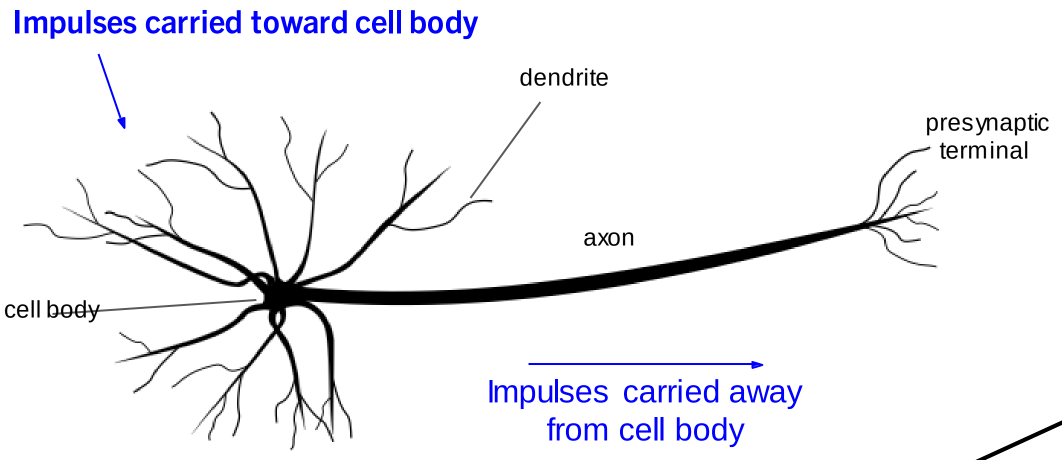
Activation functions

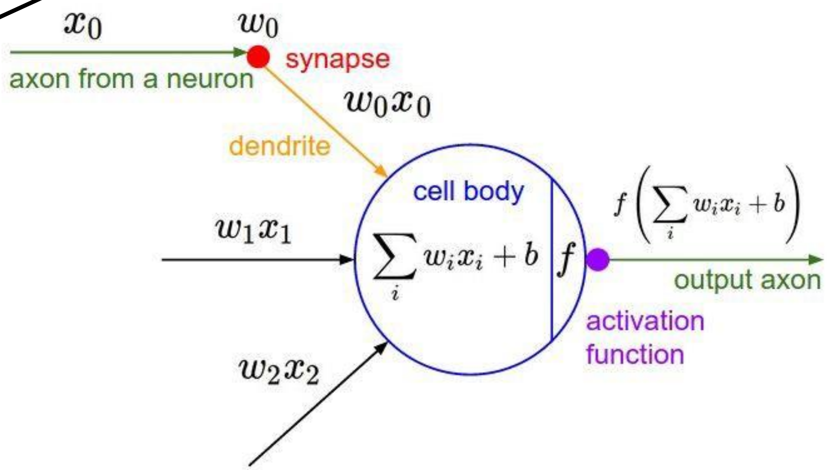


The Neuron Metaphor

- Neural networks were inspired by our understanding of the brain and how neurons interact.
- An artificial neuron in a neural network takes in multiple inputs, applies a function to them, and generates an output – mirroring the basic functionality of a biological neuron.
- This analogy has been extremely useful for explaining and visualizing how these artificial structures work.

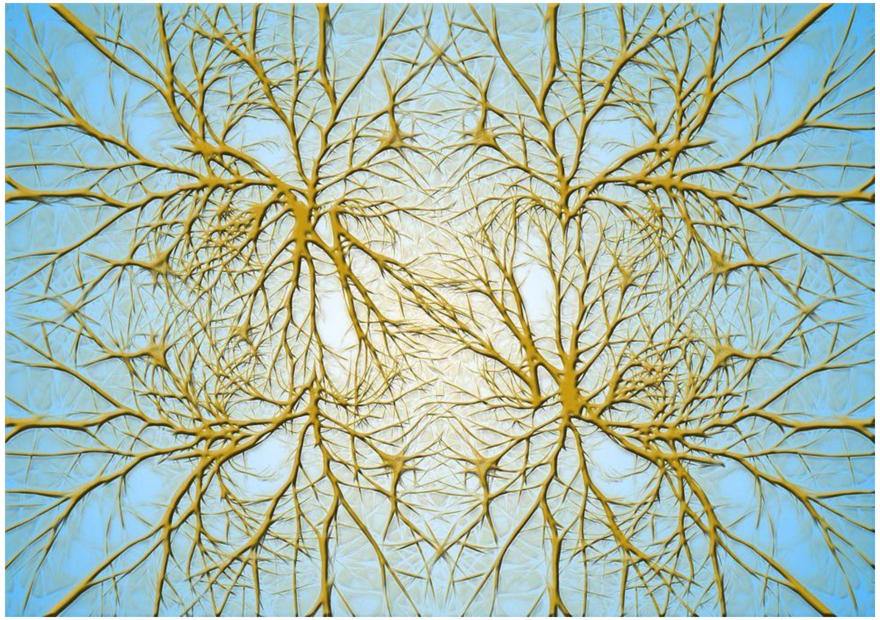
The Neuron Metaphor



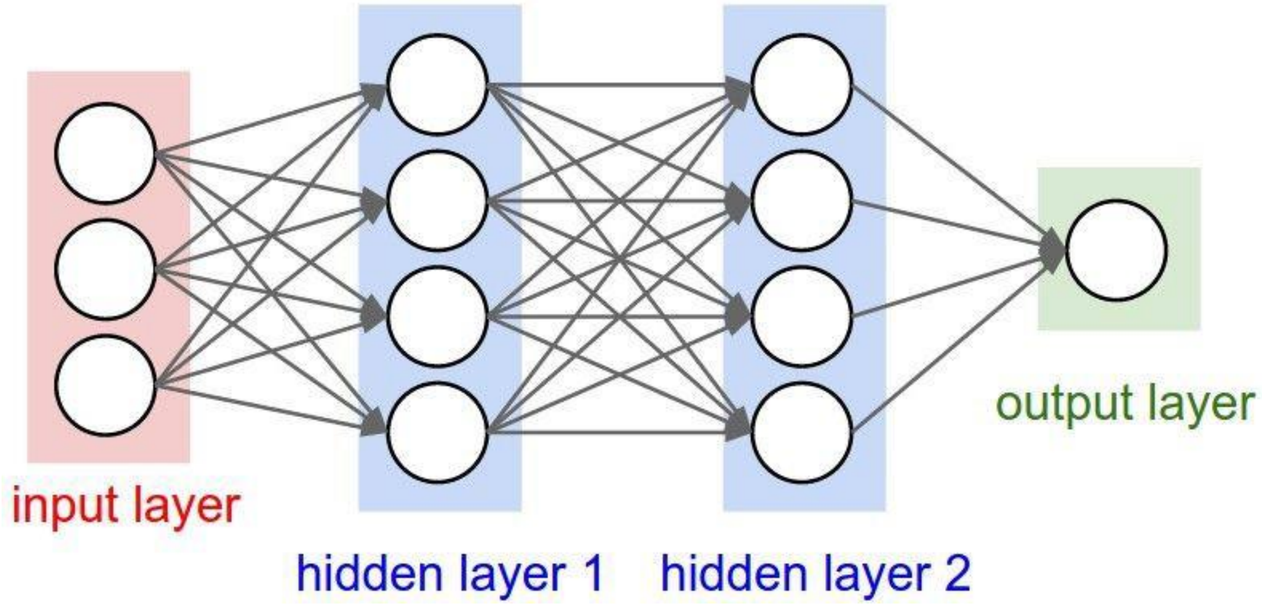


The Metaphor Breaks Down

Biological Neurons:
Complex connectivity patterns



Neurons in a neural network:
Organized into regular layers for computational efficiency



The Metaphor Breaks Down

- Biological neurons are vastly more complex: they use a mixture of electrical and chemical signals, have complex temporal dynamics, and can restructure their own connections.
- The brain is not just a feed-forward network: it has many complex feedback loops, which are not typically found in artificial neural networks.
- The brain isn't easily divided into distinct layers, as we do in artificial neural networks.

The Metaphor Breaks Down

- Over-reliance on the analogy can lead to misunderstandings about how neural networks function and their capabilities.
- This can lead to unrealistic expectations about what neural networks can do, or to overgeneralizations about their functioning.
- For instance, claiming a neural network "thinks" or "understands" like a human brain is misleading.
- To further progress, it's important to view artificial neural networks as mathematical/statistical tools, and not overstate the comparison to the human brain.

Neural Network Playground

https://playground.tensorflow.org