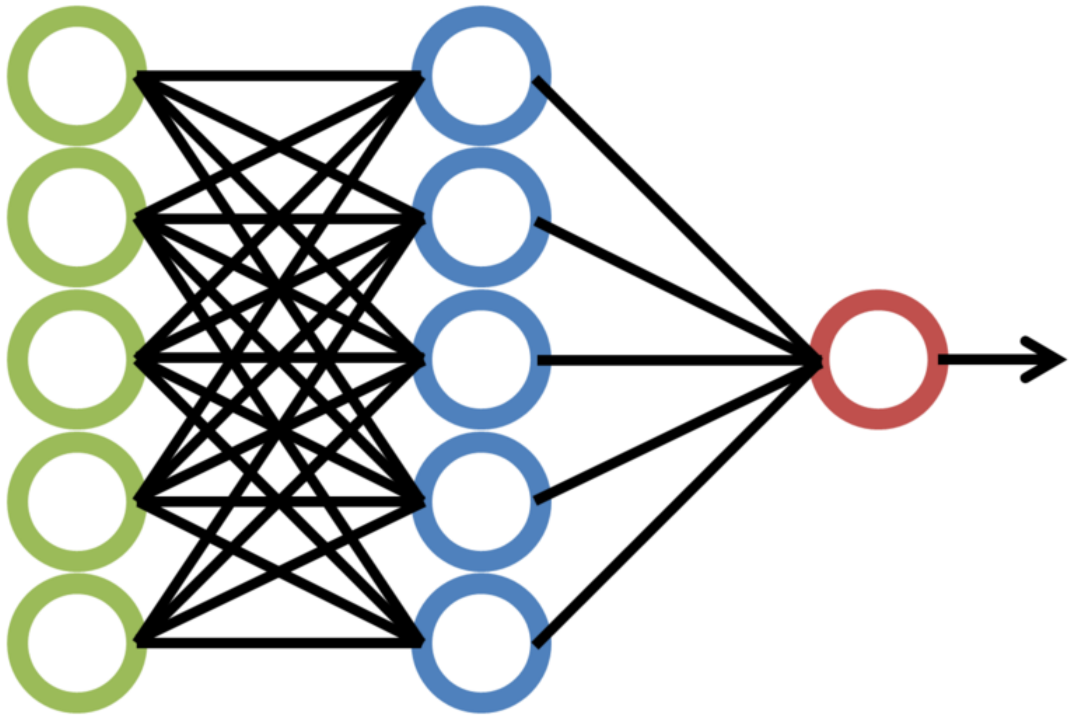
Neural Networks

Alex Olson

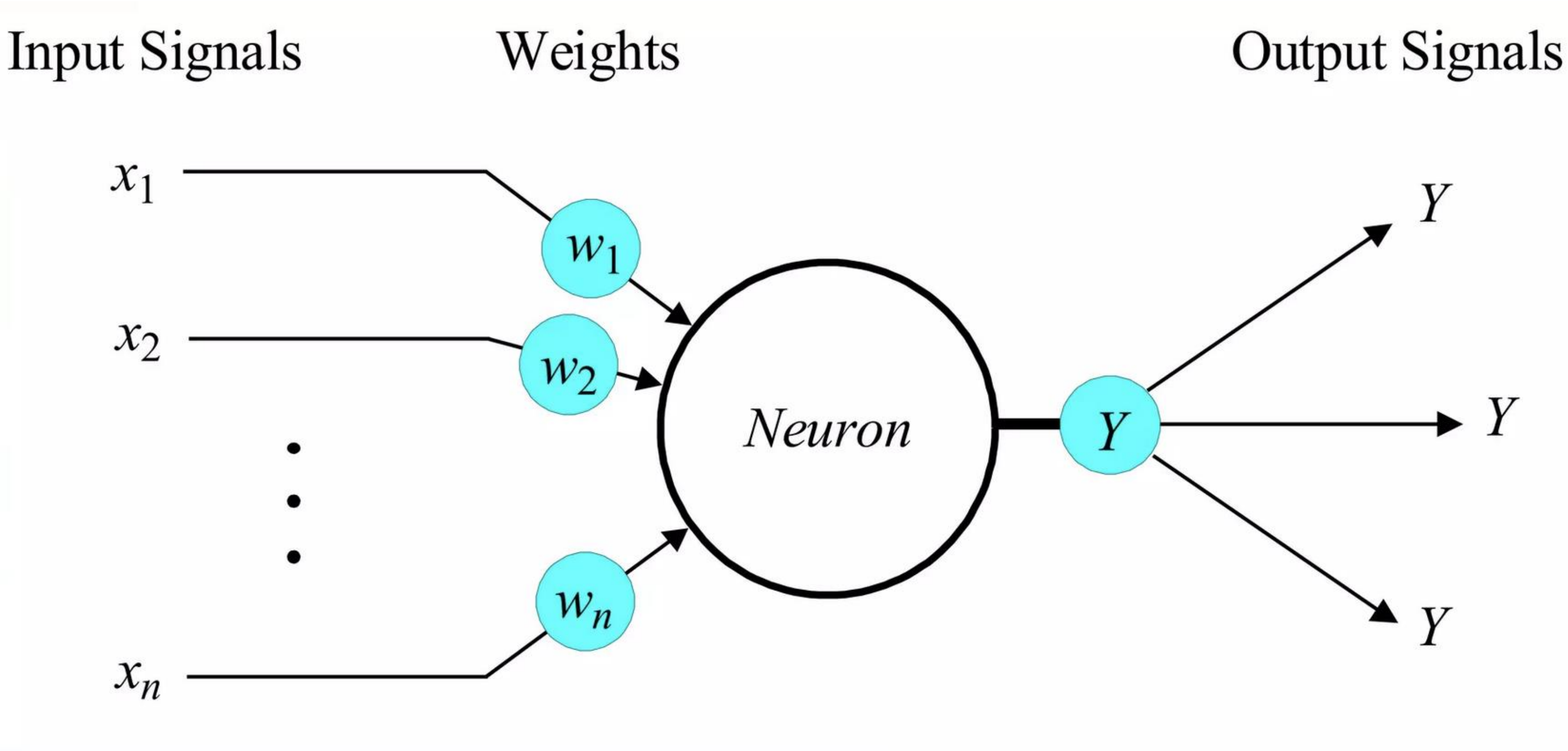
UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE & ENGINEERING
Centre for Analytics and Artificial Intelligence Engineering

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

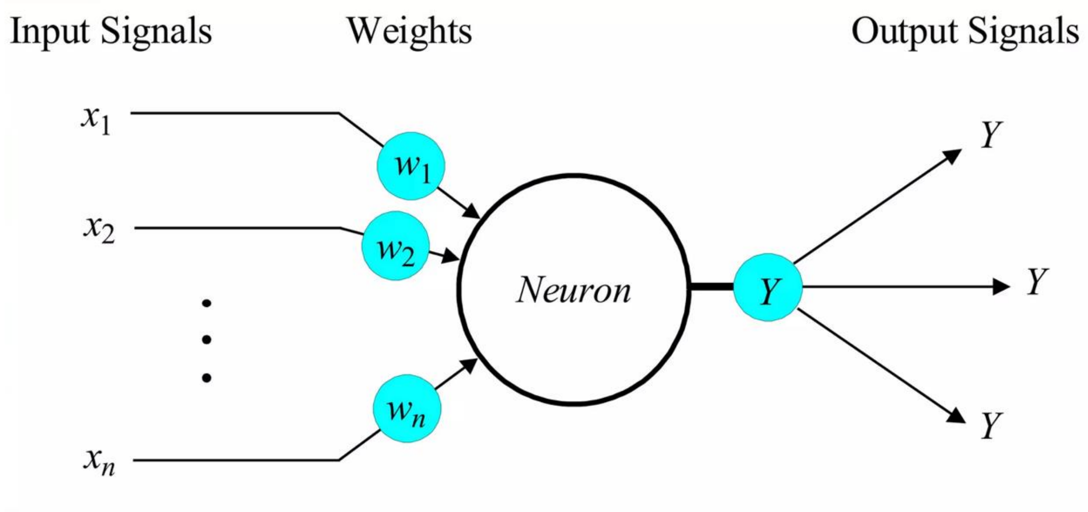


View inside an artificial neuron



View inside an artificial neuron

- Behaves like a linear regression model:
- $w_1x_1 + w_2x_2 + \ \dots + w_nx_n$
- Weights correspond to how much the neuron “cares” about each input

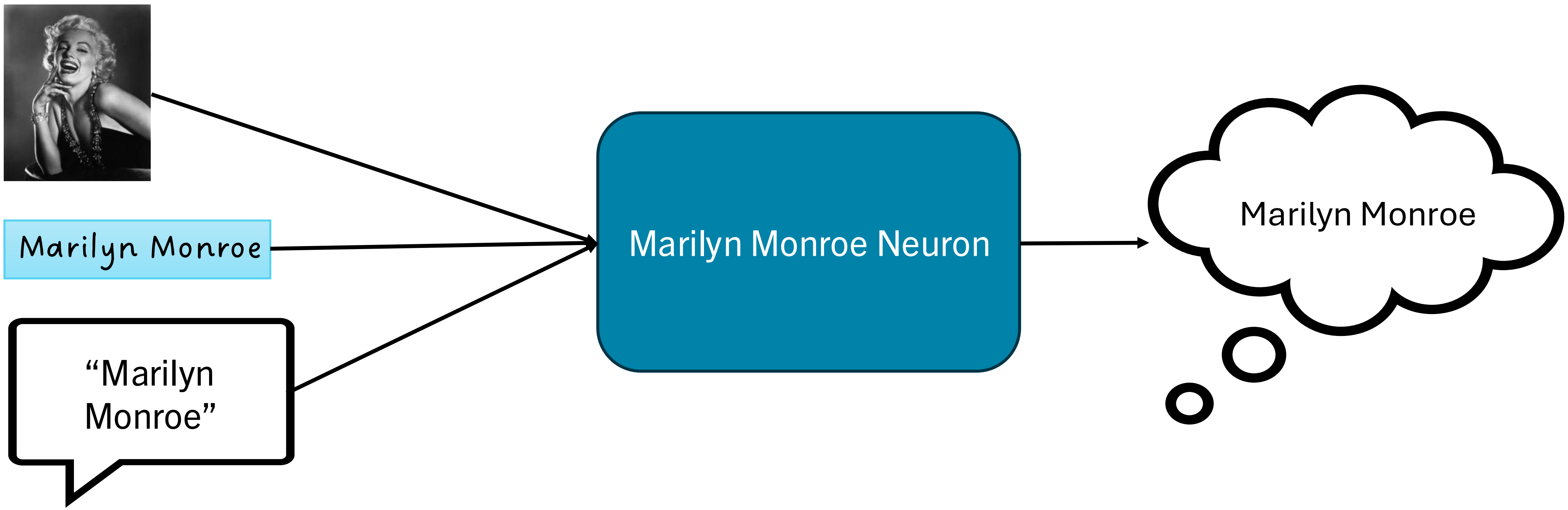


Back to the brain: the Marilyn Monroe Neuron

- Study conducted on patients with epilepsy
- Researchers use specialized equipment to measure the "excitement" of individual neurons in a patient's brain
- Measuring a neuron, the researchers showed patients a series of images
- In each patient, they found around five neurons that fired when the patient looked at a specific person

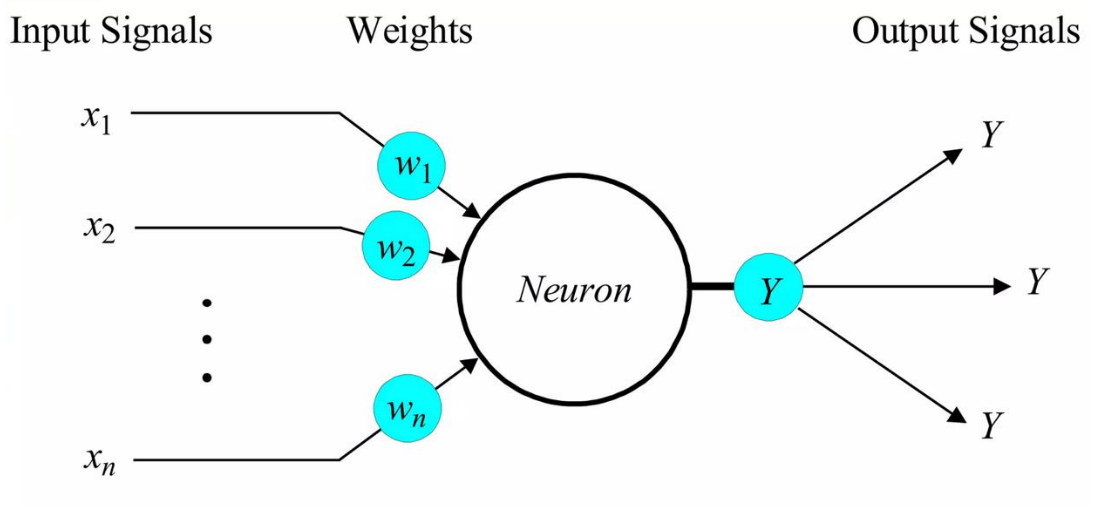
Back to the brain: the Marilyn Monroe Neuron

- Once a "celebrity" neuron was identified, the researchers wanted to know if it would still fire for representations other than images



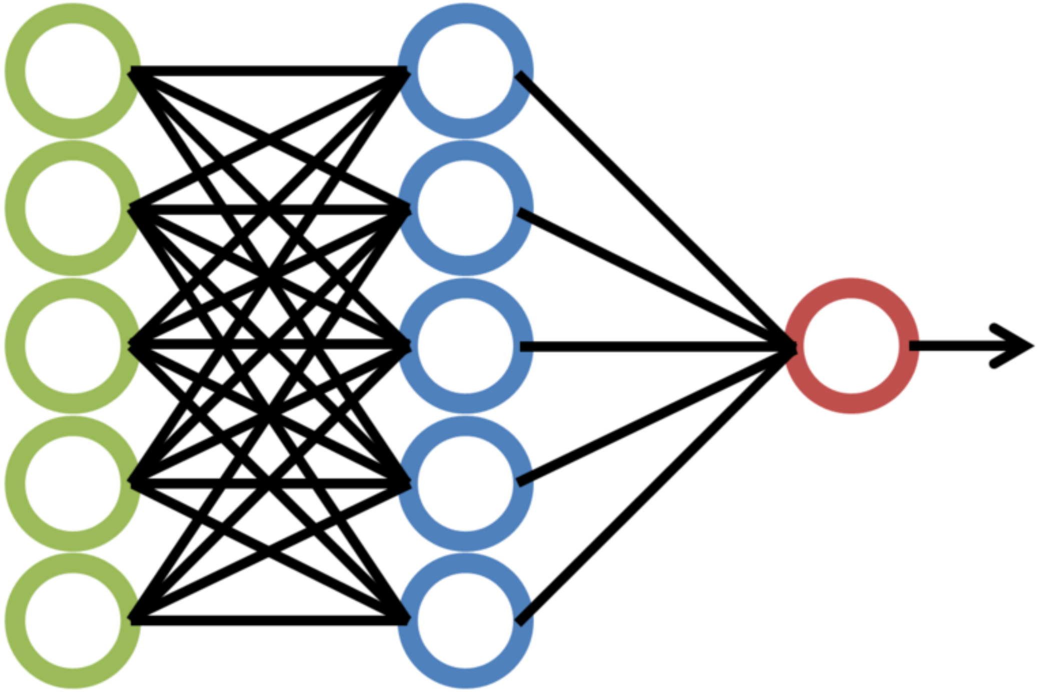
Marilyn Monroe ANN

- Weights would be high from neurons that react to different representations of Marilyn Monroe
- Weights would be low for neurons that react to other people, or concepts



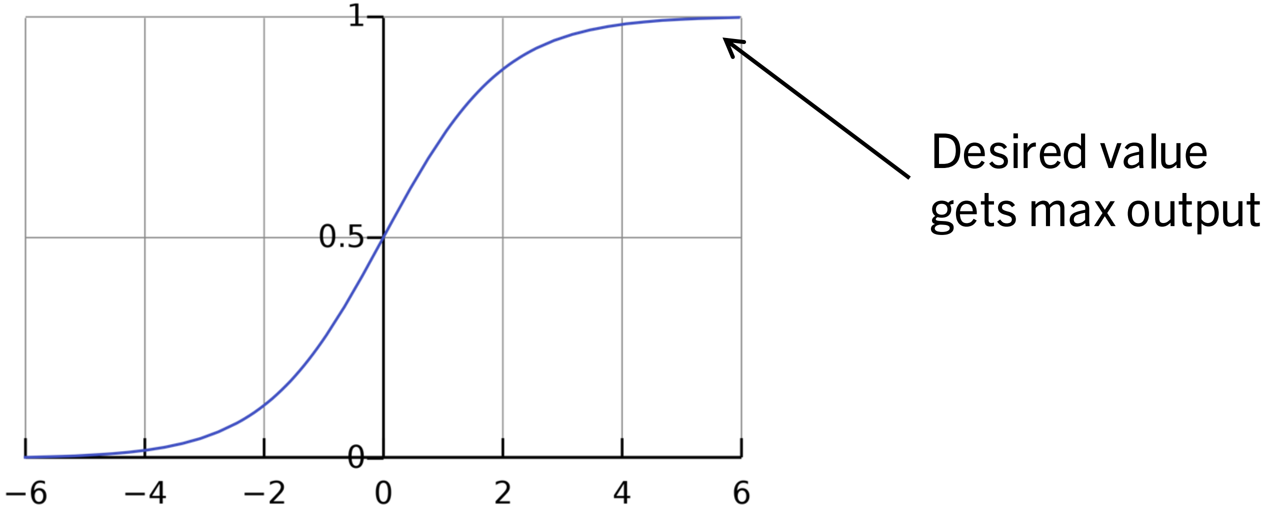
ANNs

- Each neuron considers the responses of the neurons in the previous layer
- It learns to pay attention to the neurons that are excited about what it's excited about
- Ignores the neurons that are excited about other things

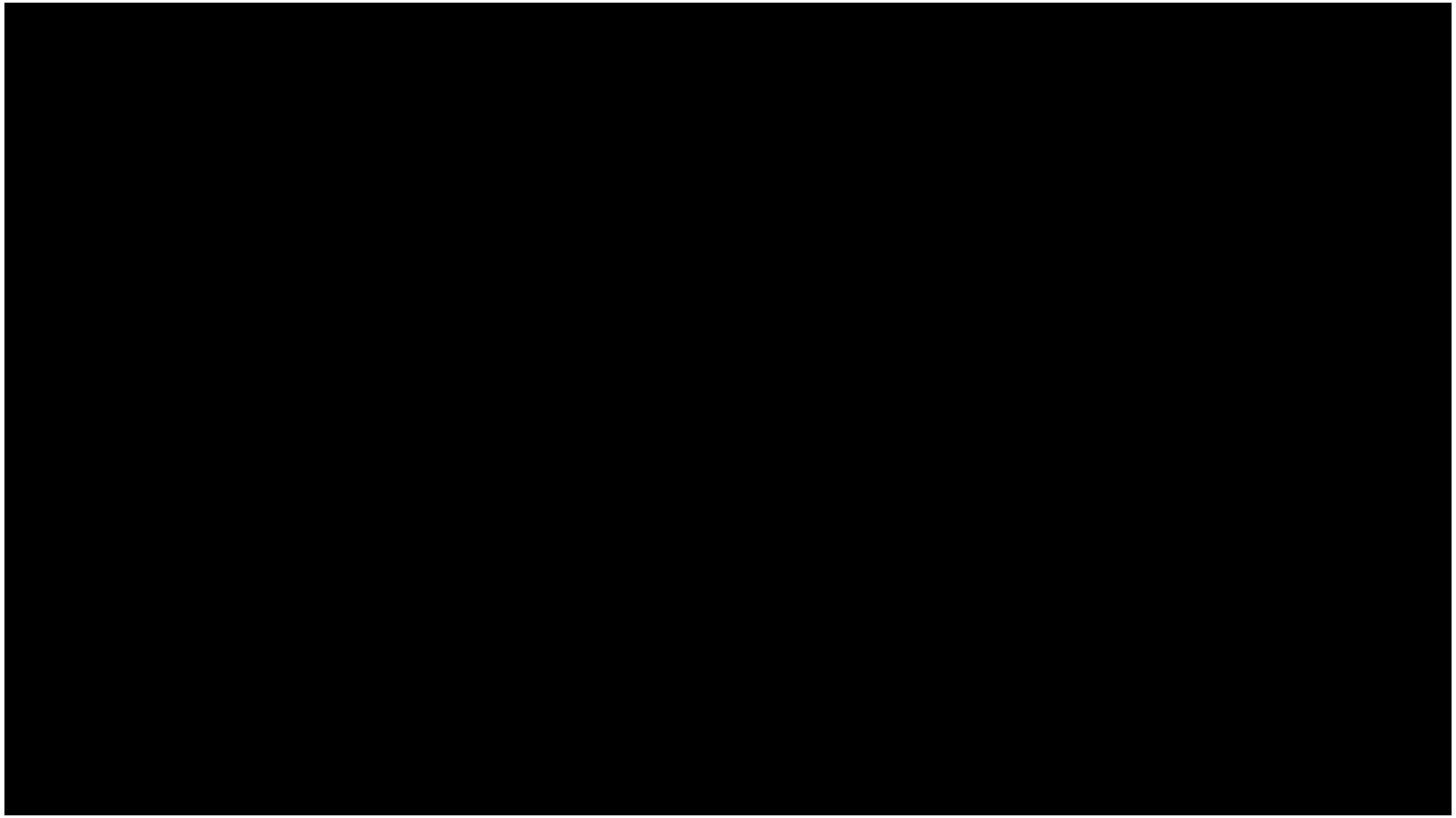


Activation Functions

- Basic approach: when I see enough activity, I get excited
- Below threshold: 0
- Above threshold: 1
- More useful: gradually increase excitement as we see more activity
- In practice: many different activation functions!

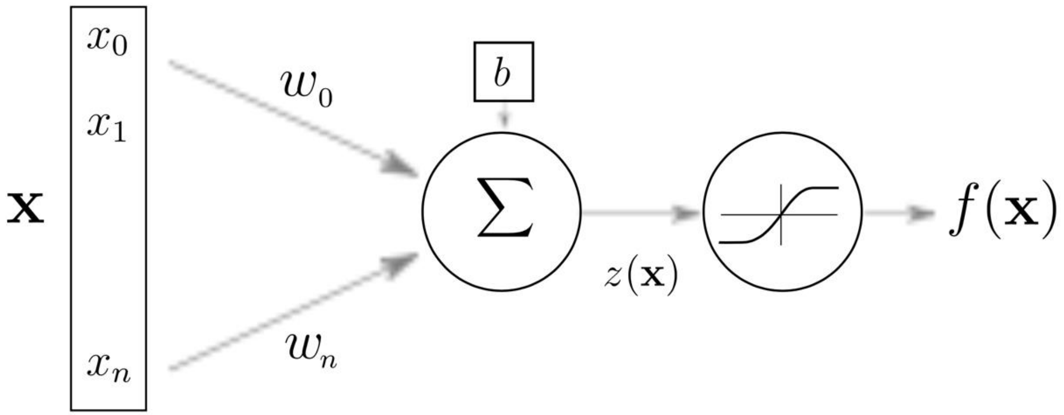


How do Neural Networks actually work?



Artificial Neuron

- $z(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x} + b$
- $f(\boldsymbol{x}) = g(\boldsymbol{w}^T \boldsymbol{x} + b)$
- $\boldsymbol{x}$ input
- $f(\boldsymbol{x})$ output
- $\boldsymbol{w}$, b weights and bias
- g activation function

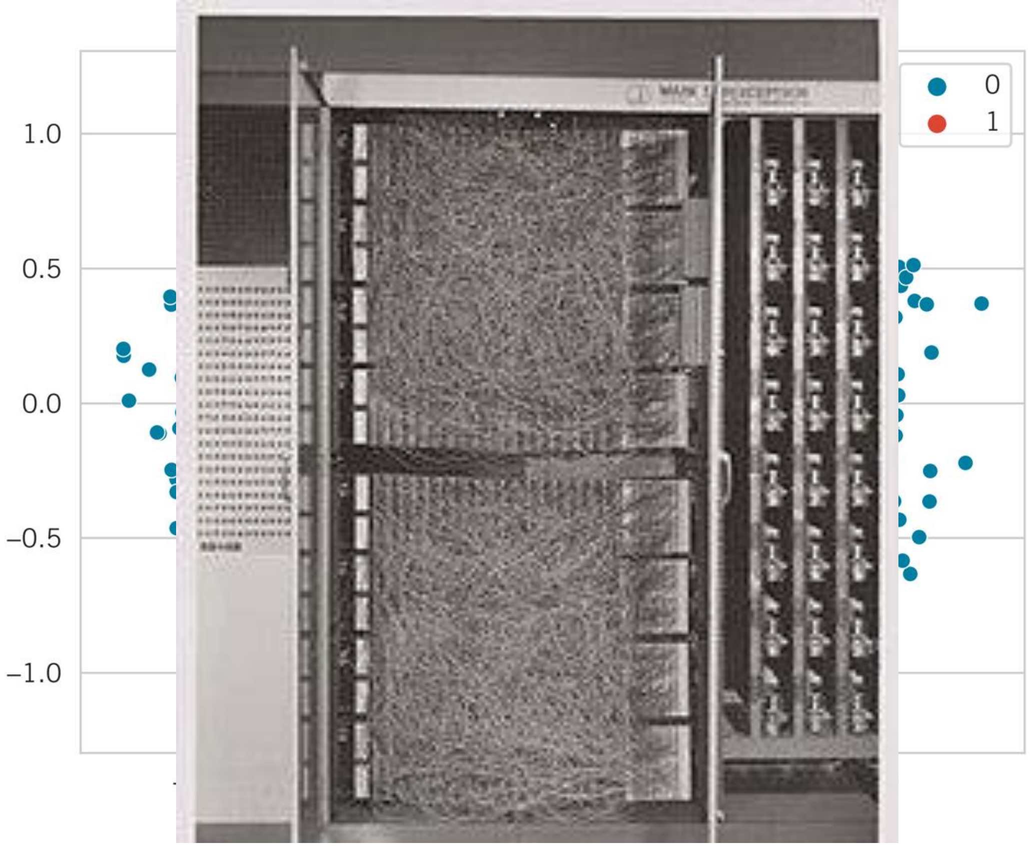


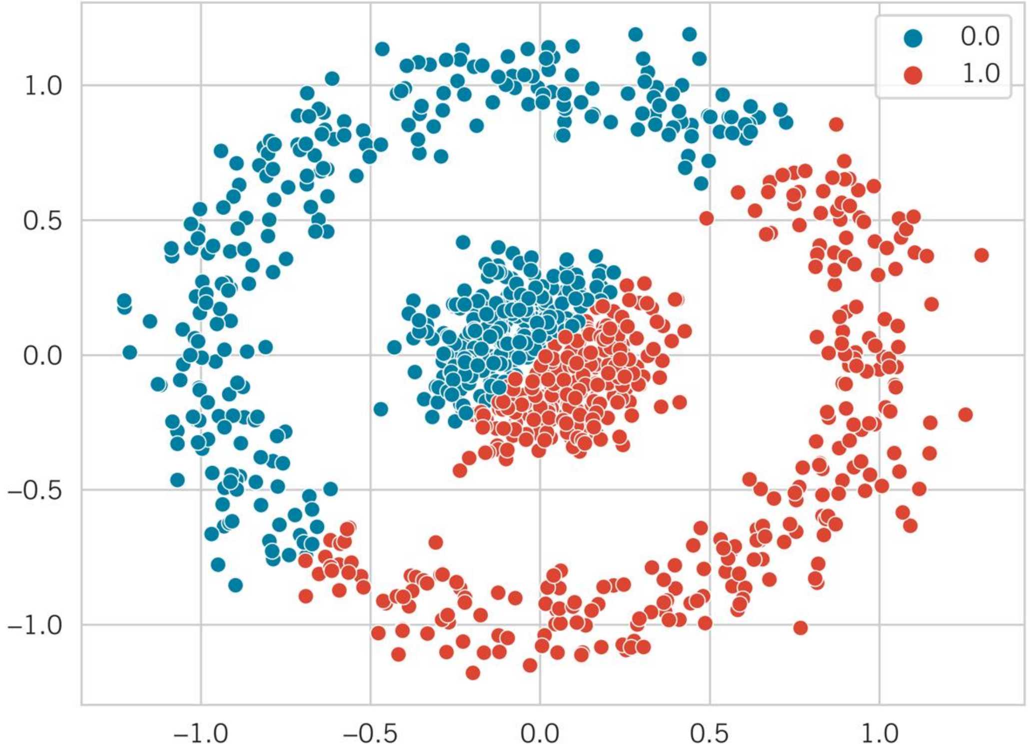
UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE & ENGINEERING
Centre for Analytics and Artificial Intelligence Engineering

Concrete Example

- Say we have two input values x_1, x_2 and one output value $f(x)$
- Our weights could be $\boldsymbol{w} = [3, -2]$ and bias $b = 1$
- Our non—linearity could be $g(z) = \max(0, z)$ (aka. ReLU)
- Now $z(\boldsymbol{x}) = 3x_1 - 2x_2 + 1$
- $f(\boldsymbol{x}) = \max(z(\boldsymbol{x}), 0)$ or $f(\boldsymbol{x}) = \max(3x_1 - 2x_2 + 1,0)$
- Every neuron in a neural network is a function just like this one!

Why do we want non-linearity?

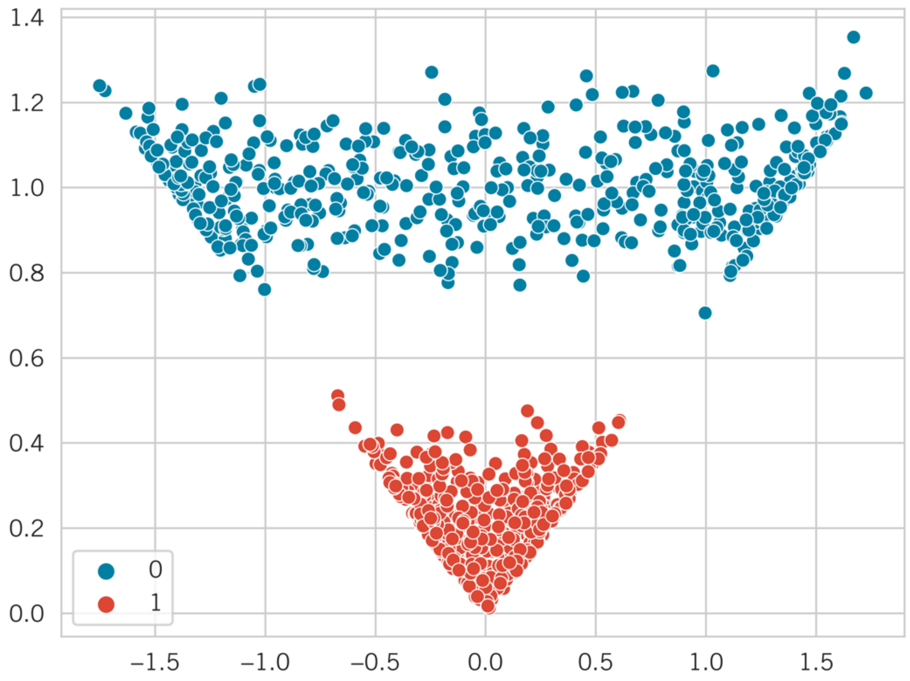




Cannot apply a linear classifier!

Why do we want non-linearity?

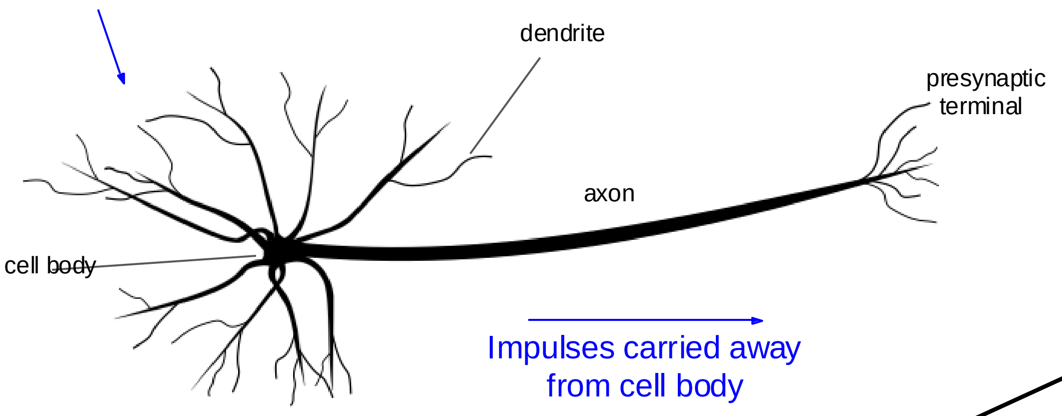
- After applying feature transformation, points become linearly separable

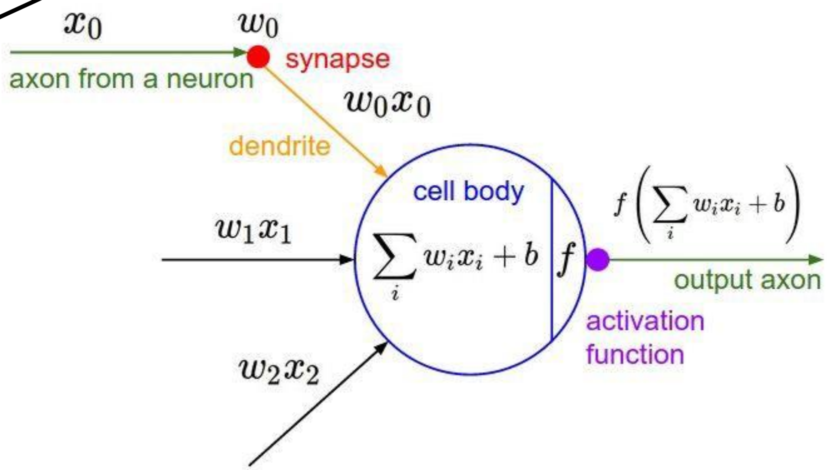


The Neuron Metaphor

- 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.

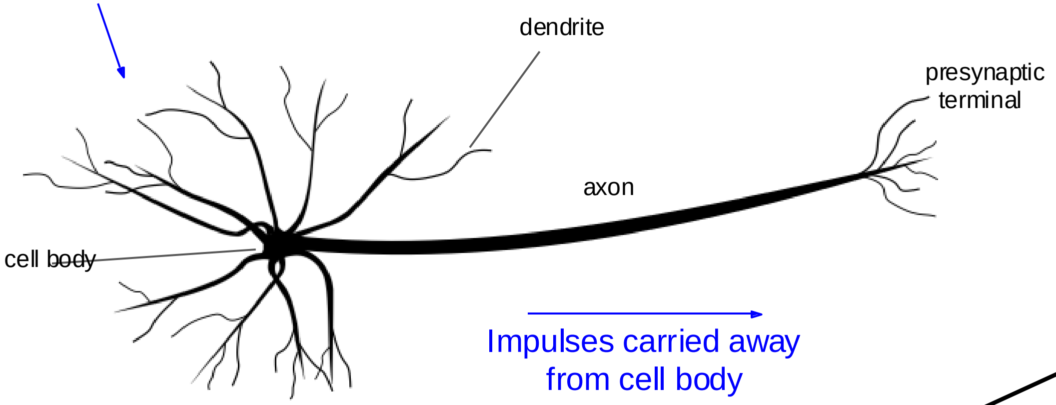
The Neuron Metaphor



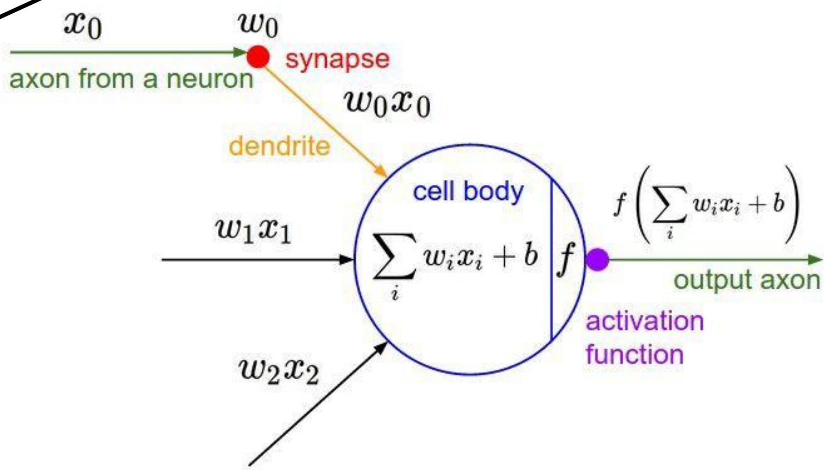


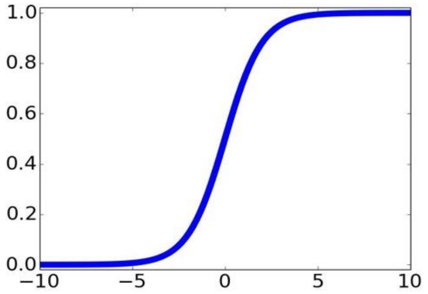
The Neuron Metaphor

Impulses carried toward cell body



Impulses carried away from cell body





sigmoid activation function

$$\frac{1}{1+e^{-x}}$$

Training the Network

- Find parameters that **minimize the total error**
- **Loss** for a given sample is the total error in predictions made
- Going through the network, the predictions are dependent on the settings of the parameters
- We have a mathematical function representing the network
- A way of measuring how ”good” it is
- How do we find the *parameters* that minimize the *loss*?

Gradient Descent

- Let's imagine we have a single parameter, p
- We can compute the relationship between p and our prediction: the $derivative\ of\ the\ loss\ with\ respect\ to\ p$
- The derivative tells us whether increasing p will increase the error, or decrease it
- To minimize loss, we make a set of predictions, compute the derivative using the total error, and adjust p *away* from the error

Gradient Descent

- We can use gradient descent to play "guess what number I'm thinking of"
- If your guess is too high, you decrease it
- If your guess is too low, you increase it
- The error function is a parabola
- By finding the lowest point on the parabola, you find the best guess

Visualizing Gradient Descent

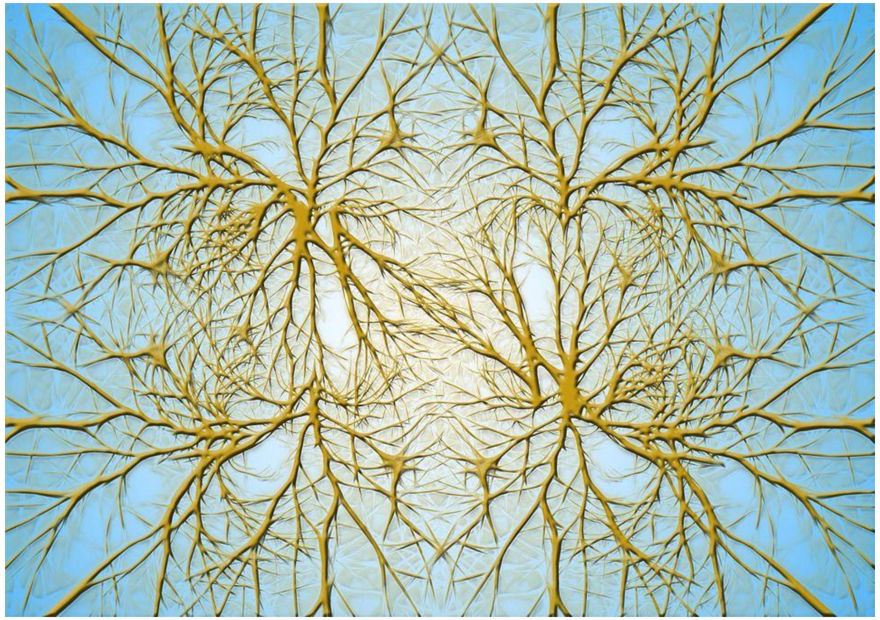
- https://uclaacm.github.io/gradient-descent-visualiser/#playground

Stochastic Gradient Descent

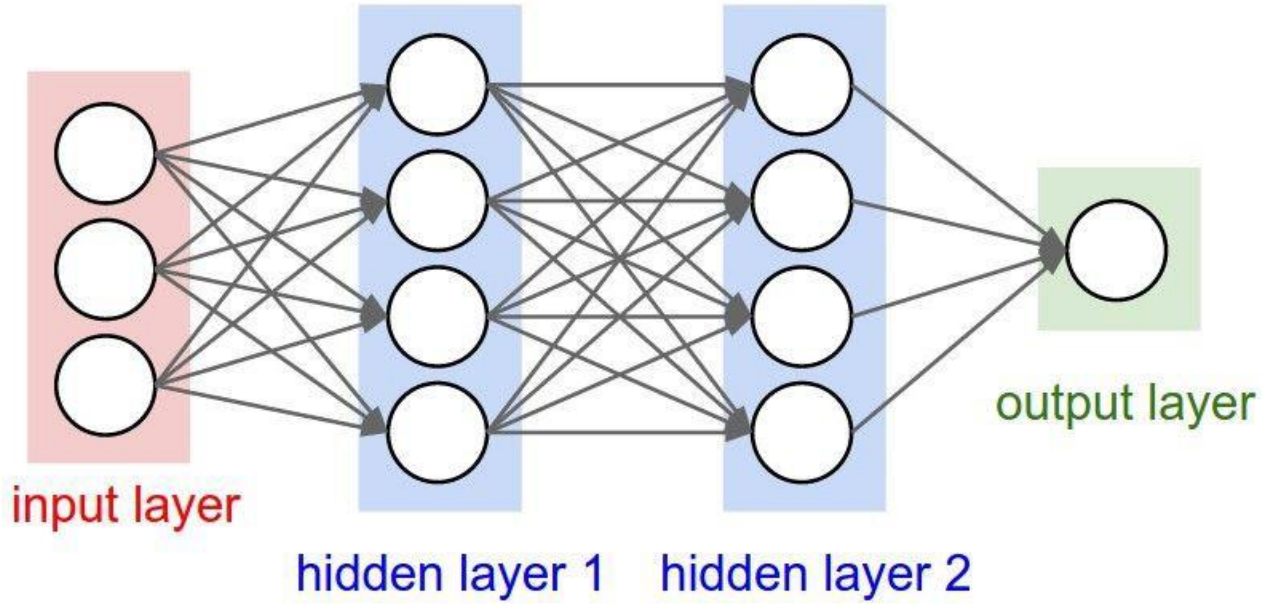
- Traditional Gradient Descent uses the entire dataset to compute the gradient, which can be computationally expensive
- Stochastic Gradient Descent (SGD) updates the parameters using only a single data point (or a small batch)
- In SGD, for each iteration, a data point (or batch) is randomly selected to compute the gradient
- Since only a subset of data is used, the gradient estimation can be noisy, leading to a less smooth path towards the minimum
- However, SGD is much faster than traditional gradient descent

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

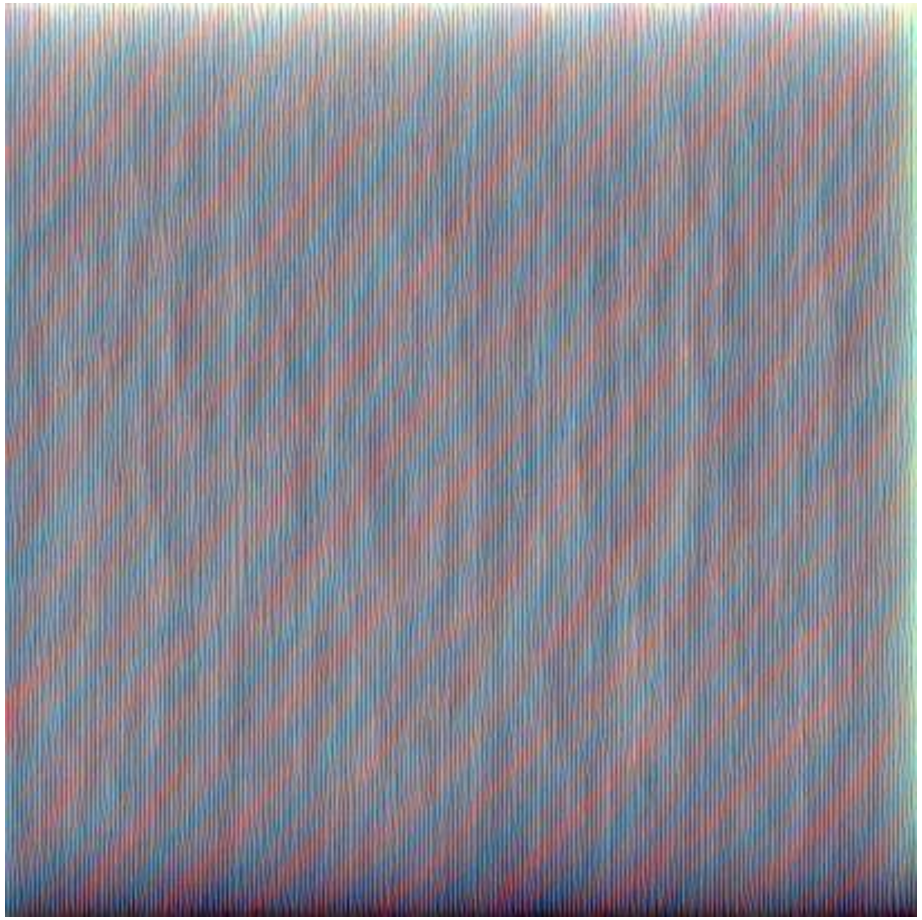
https://playground.tensorflow.org

Going past the fully connected network

- In many image tasks, we want to be able to recognize something regardless of where it is in the image
- For fully-connected networks, the order of the inputs is fixed
- No “shift invariance”

Going past the fully connected network

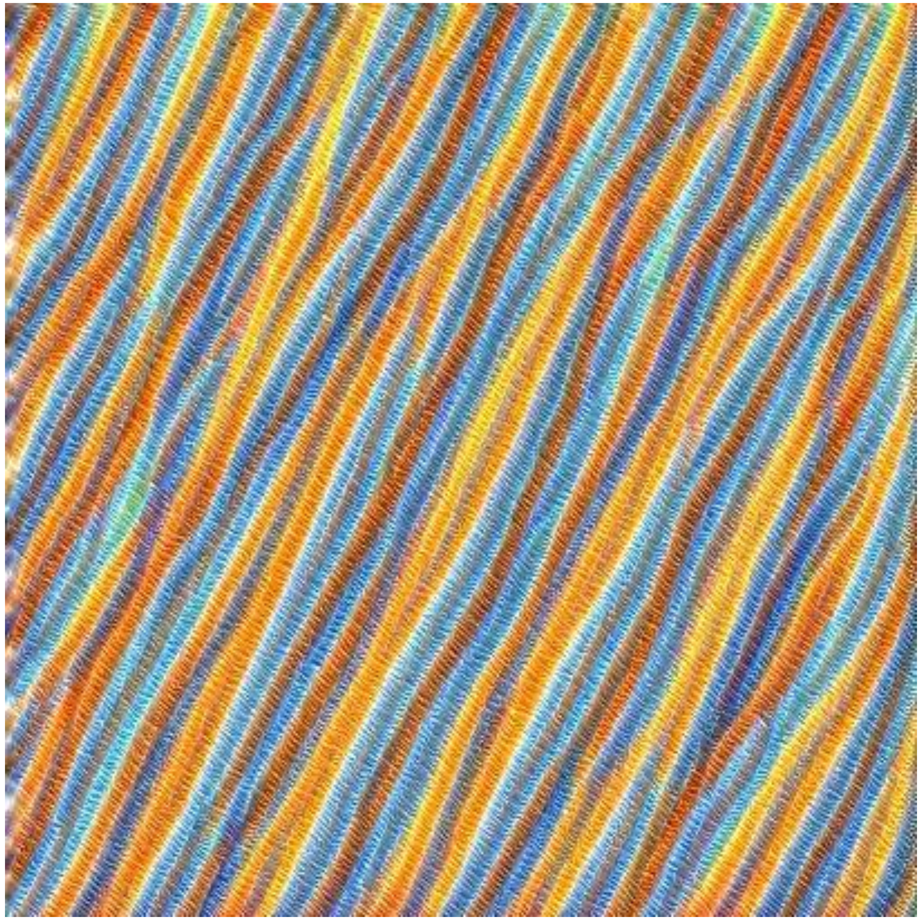
- In the 1950s and 60s, researchers showed that the brain contains neurons which respond to specific patterns, regardless of where they appear
- Combinations of very basic patterns can then be recognized as a more complicated one!



VGG-16, neuron in layer 7

Going past the fully connected network

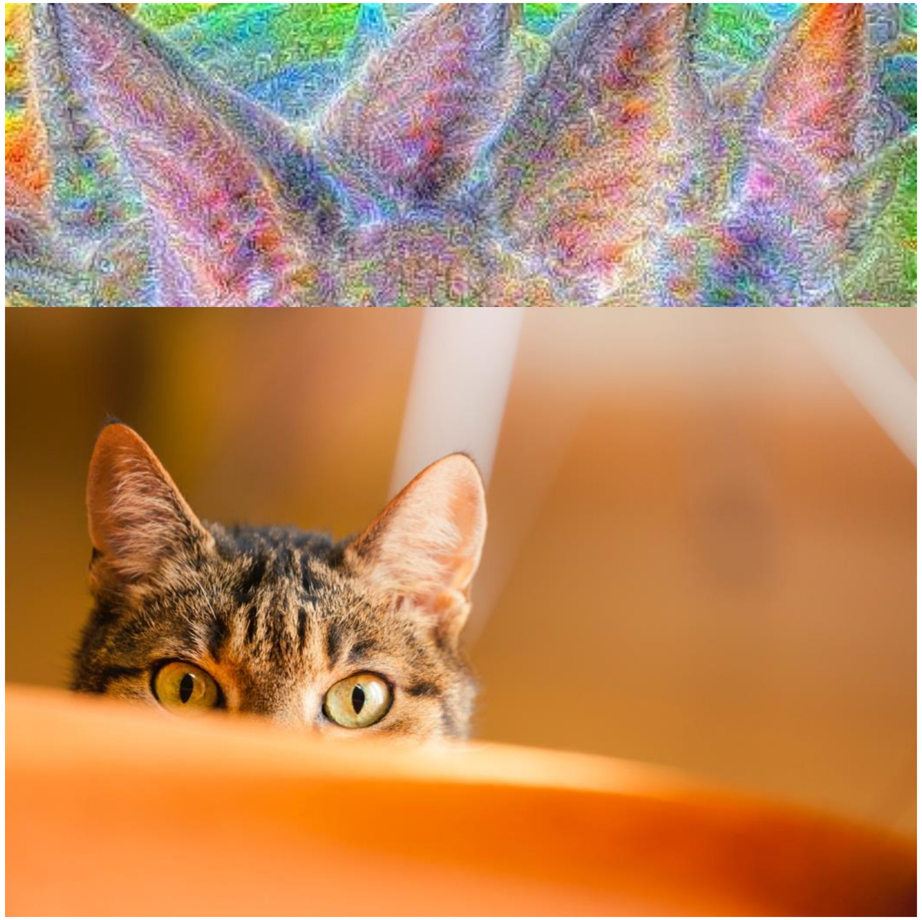
- In the 1950s and 60s, researchers showed that the brain contains neurons which respond to specific patterns, regardless of where they appear
- Combinations of very basic patterns can then be recognized as a more complicated one!



VGG-16, neuron in layer 14

Going past the fully connected network

- In the 1950s and 60s, researchers showed that the brain contains neurons which respond to specific patterns, regardless of where they appear
- Combinations of very basic patterns can then be recognized as a more complicated one!



VGG-16, neuron in layer 40

Interactive CNNs

https://adamharley.com/nn_vis/cnn/2d.html