Neural Systems

THE UNIVERSITY OF
TENNESSEE
KNOXVILLE

Additional Lab 2 Information

“Iterations”

- Every time you run lab2.py (even with the same parameter values), you should get different results
- This is because LEAP uses a different random seed to generate the population each time
- To understand how a set of parameters behaves in general, we are running lab2.py 20 times (iterations) for each parameter set

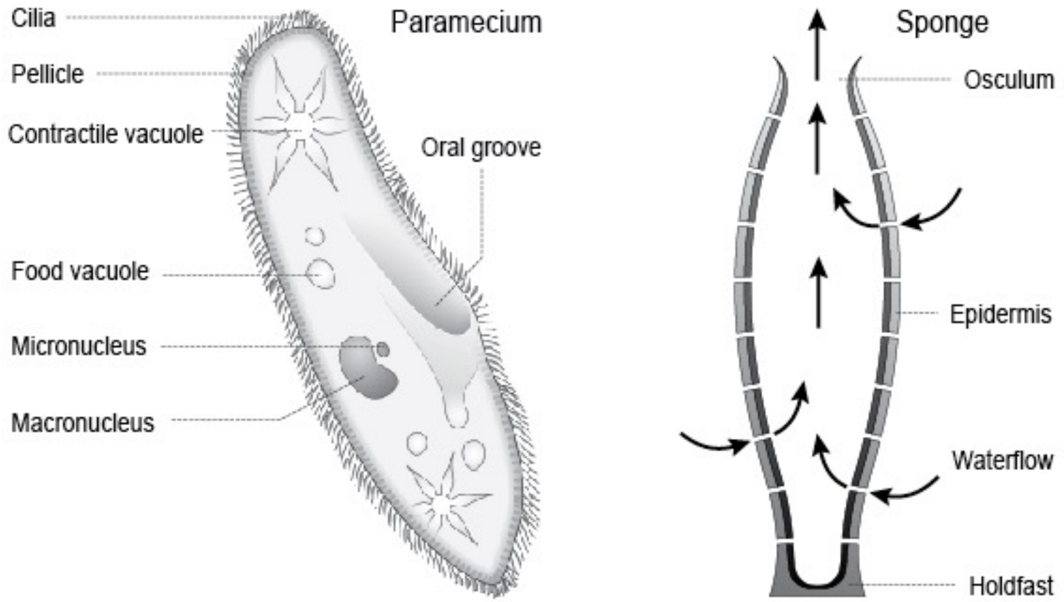
CSV File

- In the CSV file you create, you will have a row for each generation from each of these files, for a total of 5120*30 =153,600 rows.
- You should write a program that parses the 5120 CSV files that are created.
- For each generation g (step = g) in each file, you will calculate:
  - The average fitness in that generation
  - The maximum fitness in that generation
  - The best genome in that generation
  - Whether a solution (fitness = 1) was found: This will be 0 or 1 in the CSV file
  - How many solutions were found in that generation
  - CS 527: The diversity metric for that generation

Biological Neural Systems

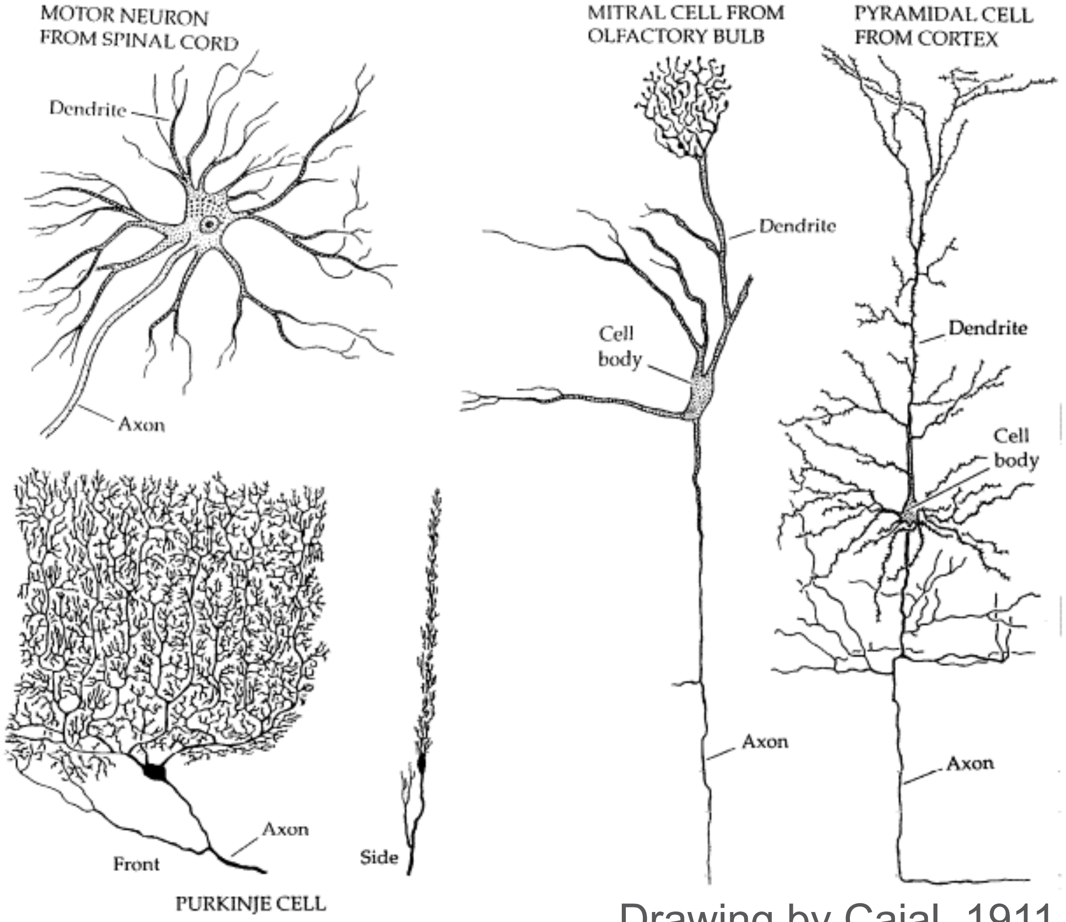
Why Nervous Systems?

- Not all animals have nervous systems; some use only chemical reactions
  - Paramecium and sponge move, eat, escape, display habituation
- Nervous systems give advantages:
  - Selective transmission of signals across distant areas: more complex bodies
  - Complex adaptation: survival in changing environments



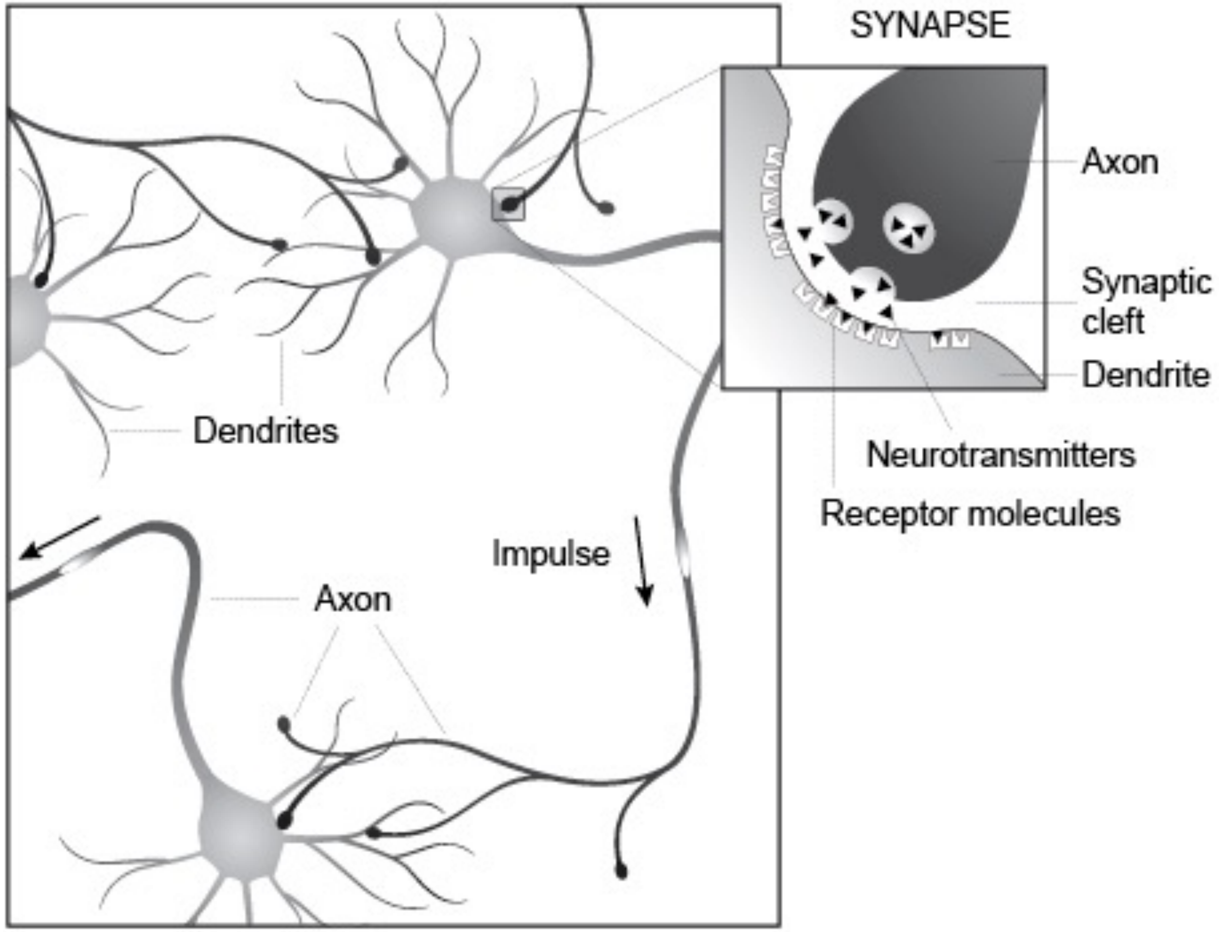
What Makes Brains Different?

- Components and behavior of individual neurons are very similar across animal species and, presumably, over evolutionary history (Parker, 1919)
- Evolution of the brain seems to occur mainly in the architecture, that is how neurons are interconnected.
- First classification of neurons by Cajal in 1911 was made according to their connectivity patterns



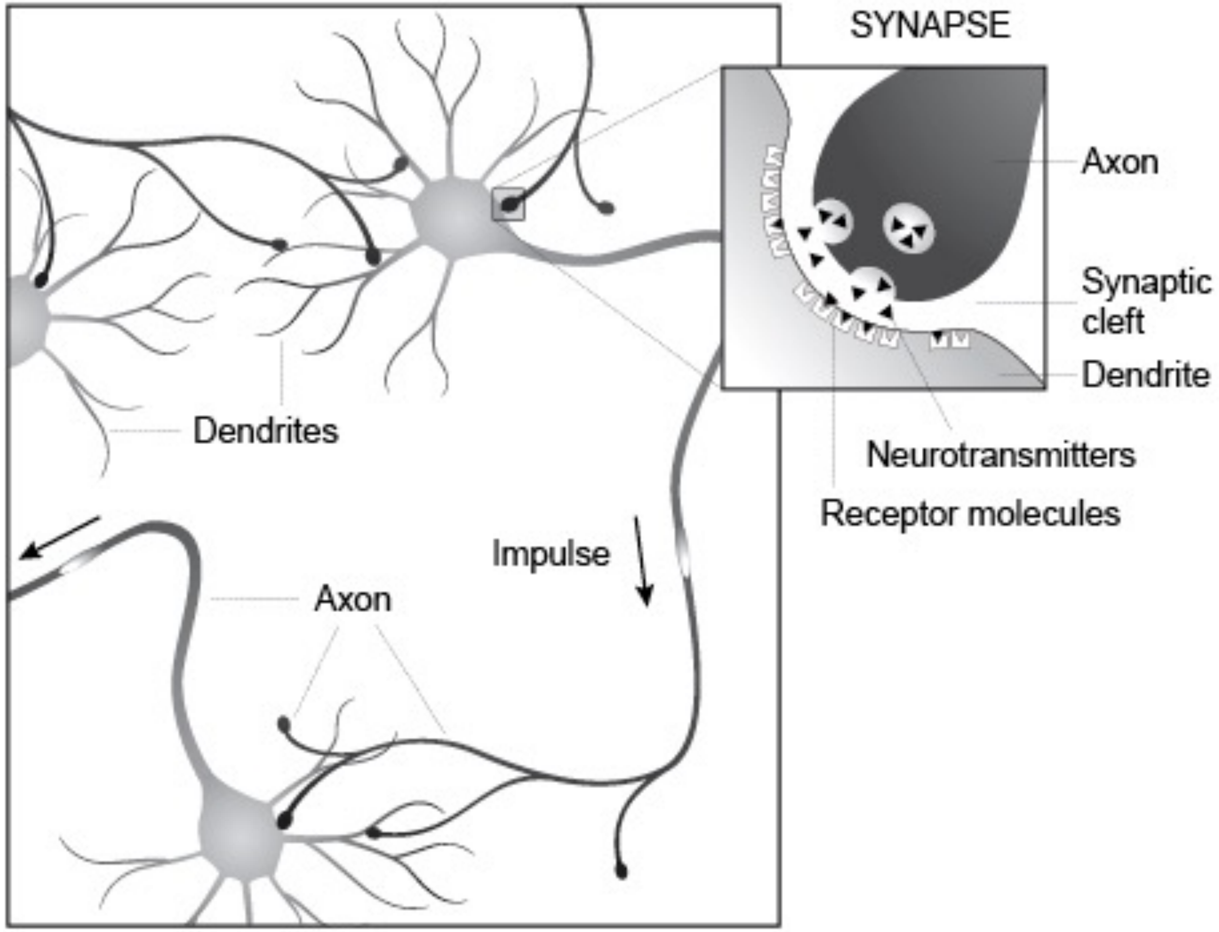
Drawing by Cajal, 1911

Biological Neurons



- Dendrites receive signals from other neurons
- Axons carry outgoing electrical signals emitted by the neuron
- Axons branch out to connect to other neurons

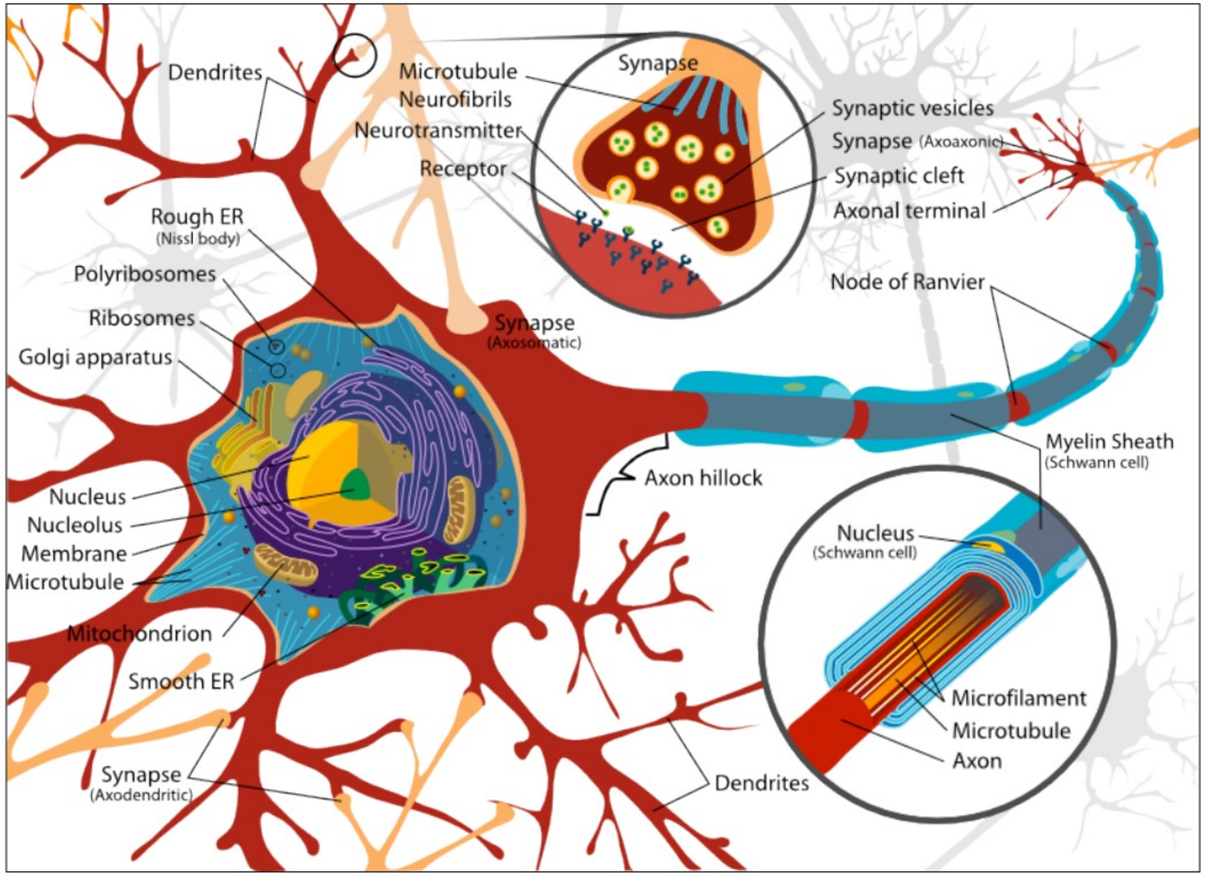
Biological Synapses



- Transmission of signals between neurons are mediated by electrochemical devices called synapses
- Incoming electrical signals trigger the release of neurotransmitters, which open the molecular gates on dendrites to allow the flow of electrically charged particles (ions)

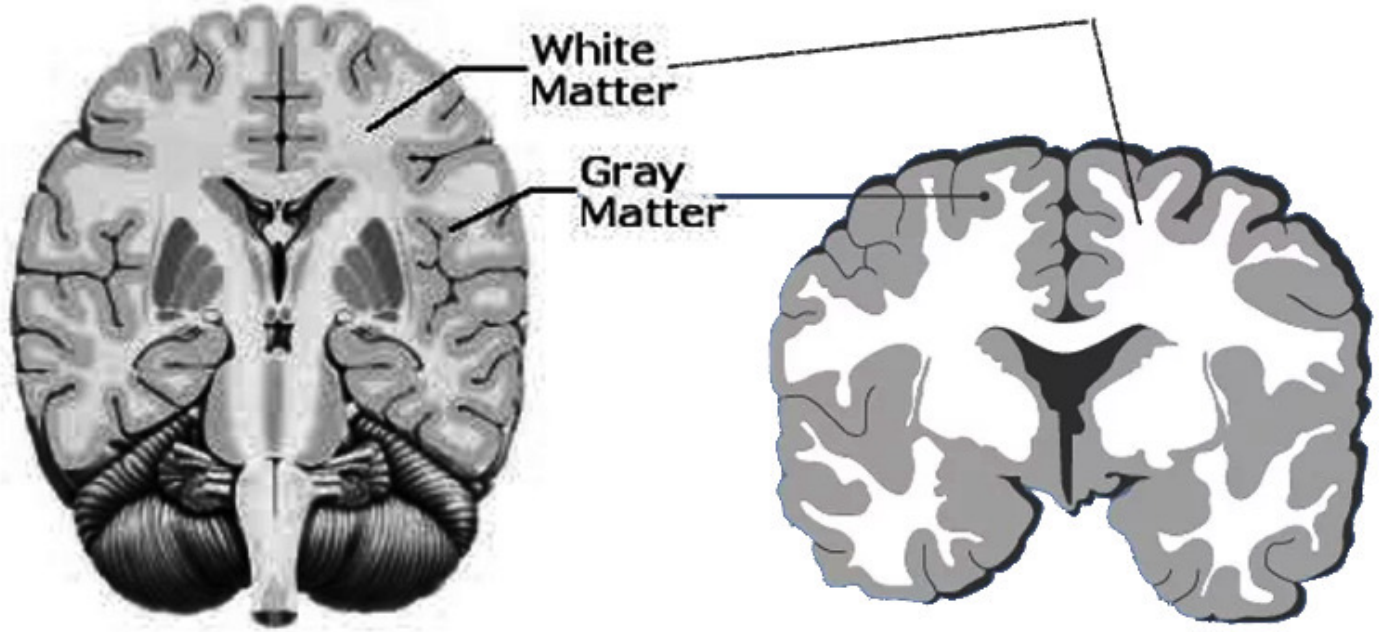
Biological Neurons/Synapses

- In the central nervous system, multiple types of cells provide myelin sheaths along axons.
- Myelin is a fat that provides an insulating layer for the axon. The thickness of the myelin sheath controls the propagation delay of signals along the axon.
- Myelin sheaths are separated along the axon by nodes of Ranvier.
  - The action potential traveling along the axon is regenerated at each of the nodes of Ranvier.



Gray Matter vs. White Matter

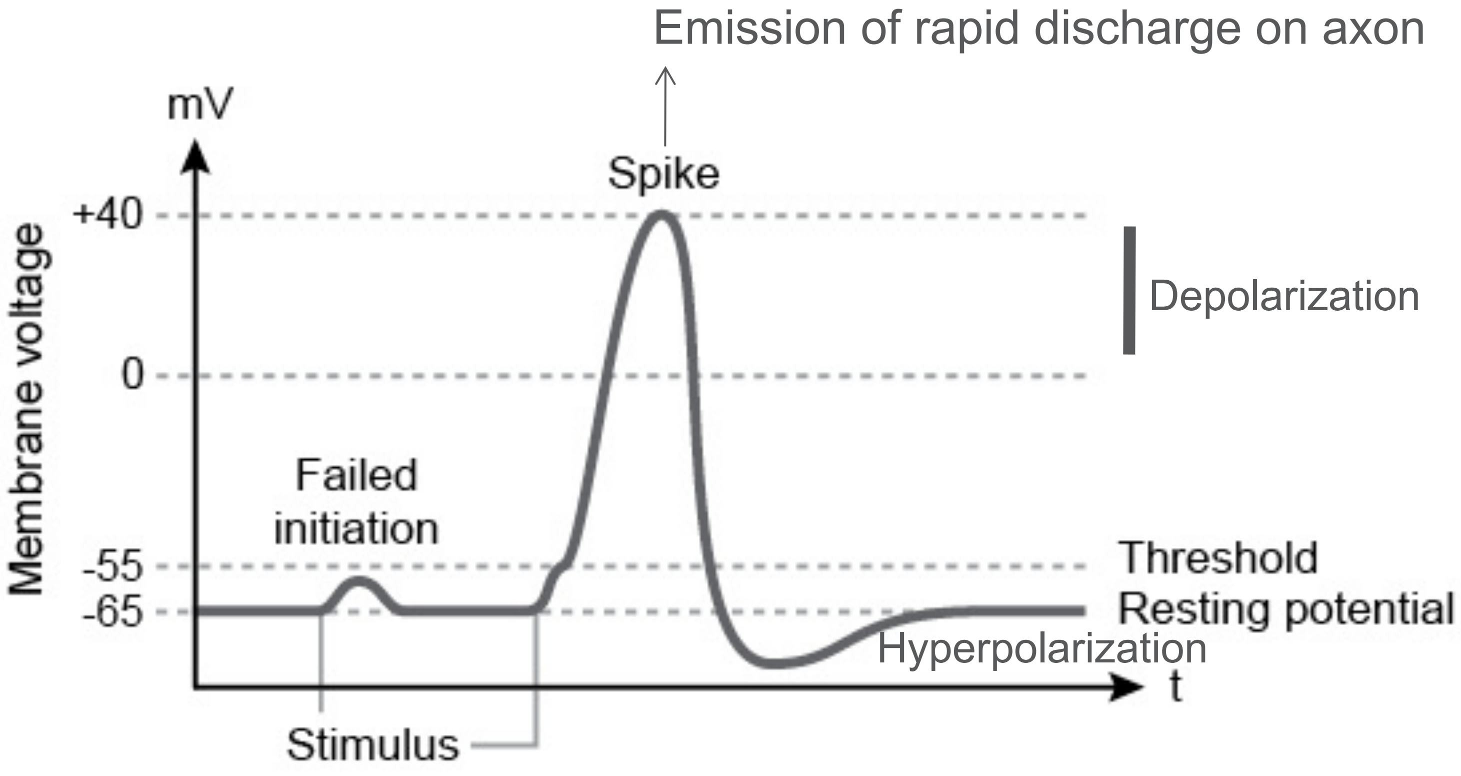
- Nervous tissue in the brain is categorized into two types:
  - Gray matter consists mainly of the neuron cell bodies (which gives gray matter its color) and the dendrites and non-myelinated axons
  - White matter is predominantly myelinated axons. The color of white matter is due to the whitish color of myelin



Mercadante AA, Tadi P. Neuroanatomy, Gray Matter. [Updated 2021 Jul 31]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. [Figure, Gray and White matter. Image courtesy S Bhimji MD] Available from: https://www.ncbi.nlm.nih.gov/books/NBK553239/figure/article-36416.image.f1/

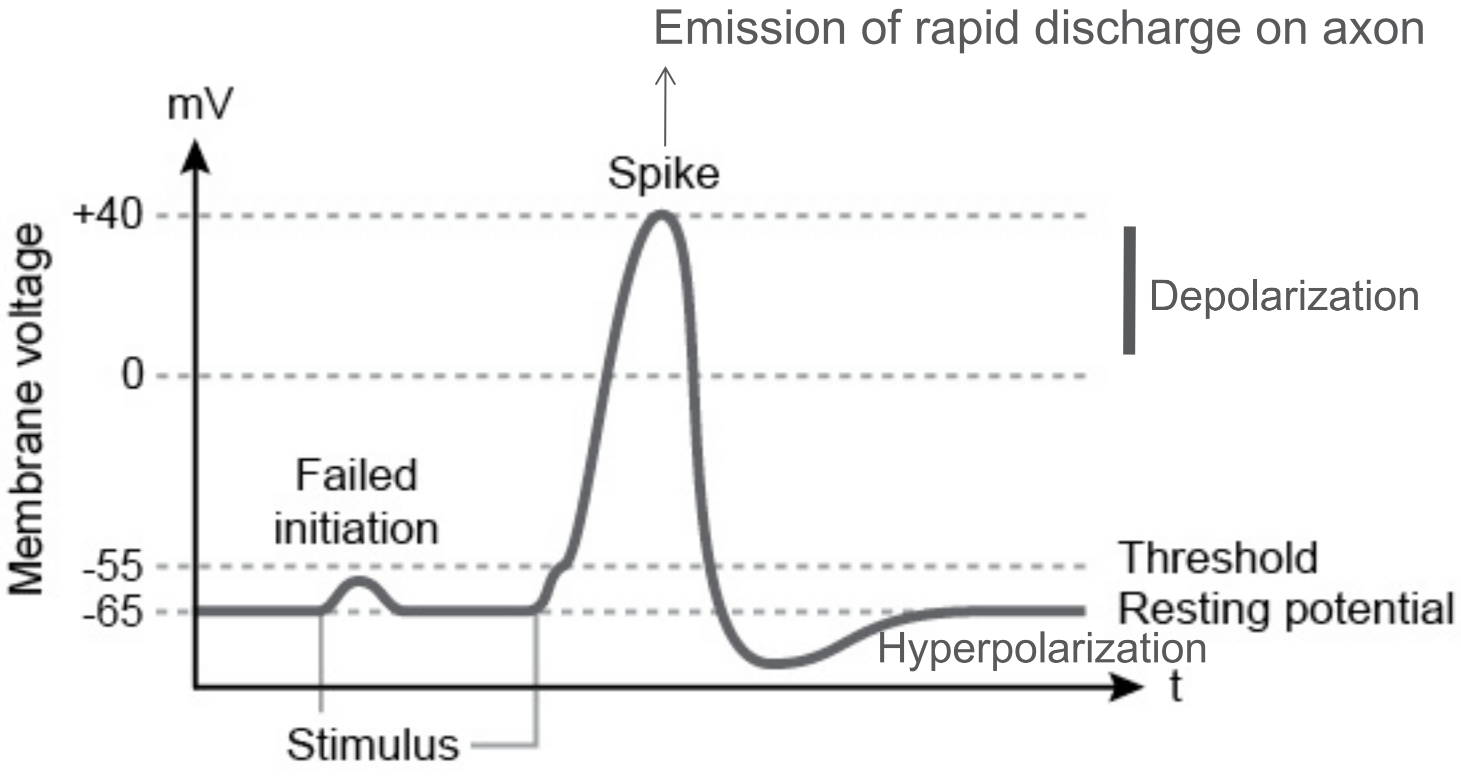
Membrane Dynamics

- Ions generate a voltage difference across the membrane that travels from the dendrite to the body
- Creates a voltage difference between the interior and environment
- Voltage difference is the activation level or potential



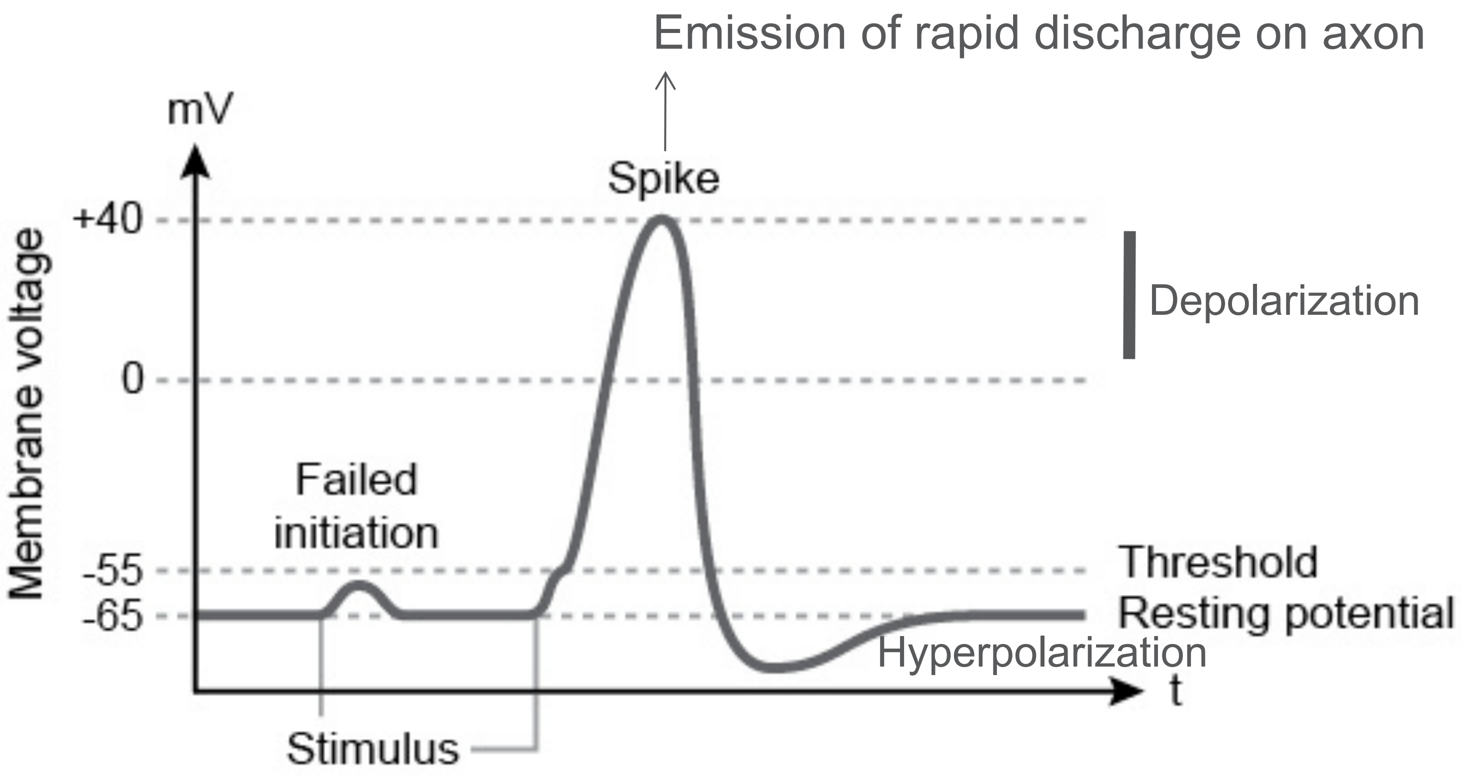
Membrane Dynamics

- A neuron propagates an electrical signal along its axon when the activation level is larger than the threshold
- This creates an action potential (an electrical discharge)
- Action potential is followed by a hyperpolarization period



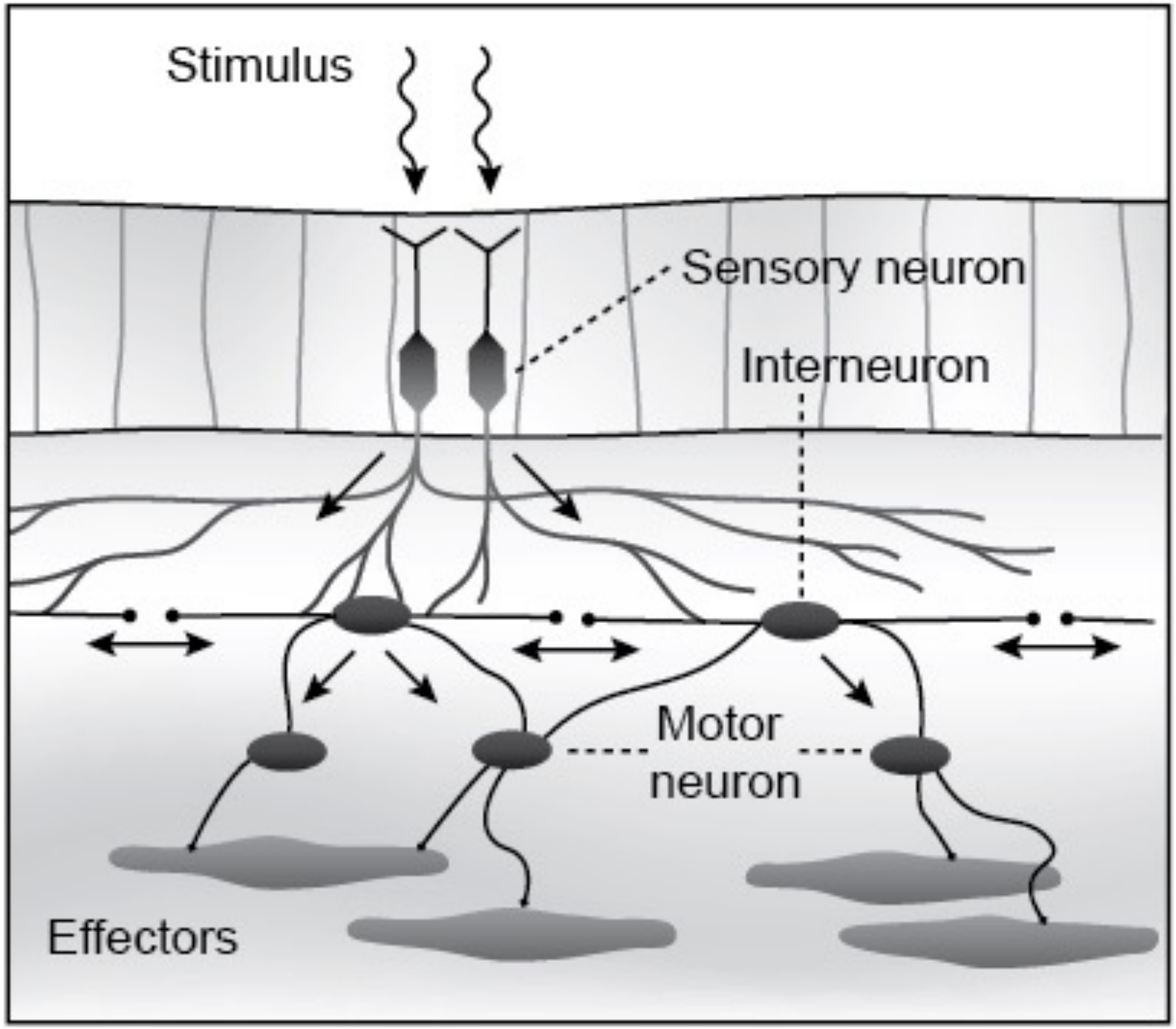
Membrane Dynamics

- This cycle lasts approximately 3-50 ms, depending on the type of ion channels involved (Hodgkin and Huxley, 1952)
- Output discharge is often called a spike and it is common to say that the neuron has “fired”



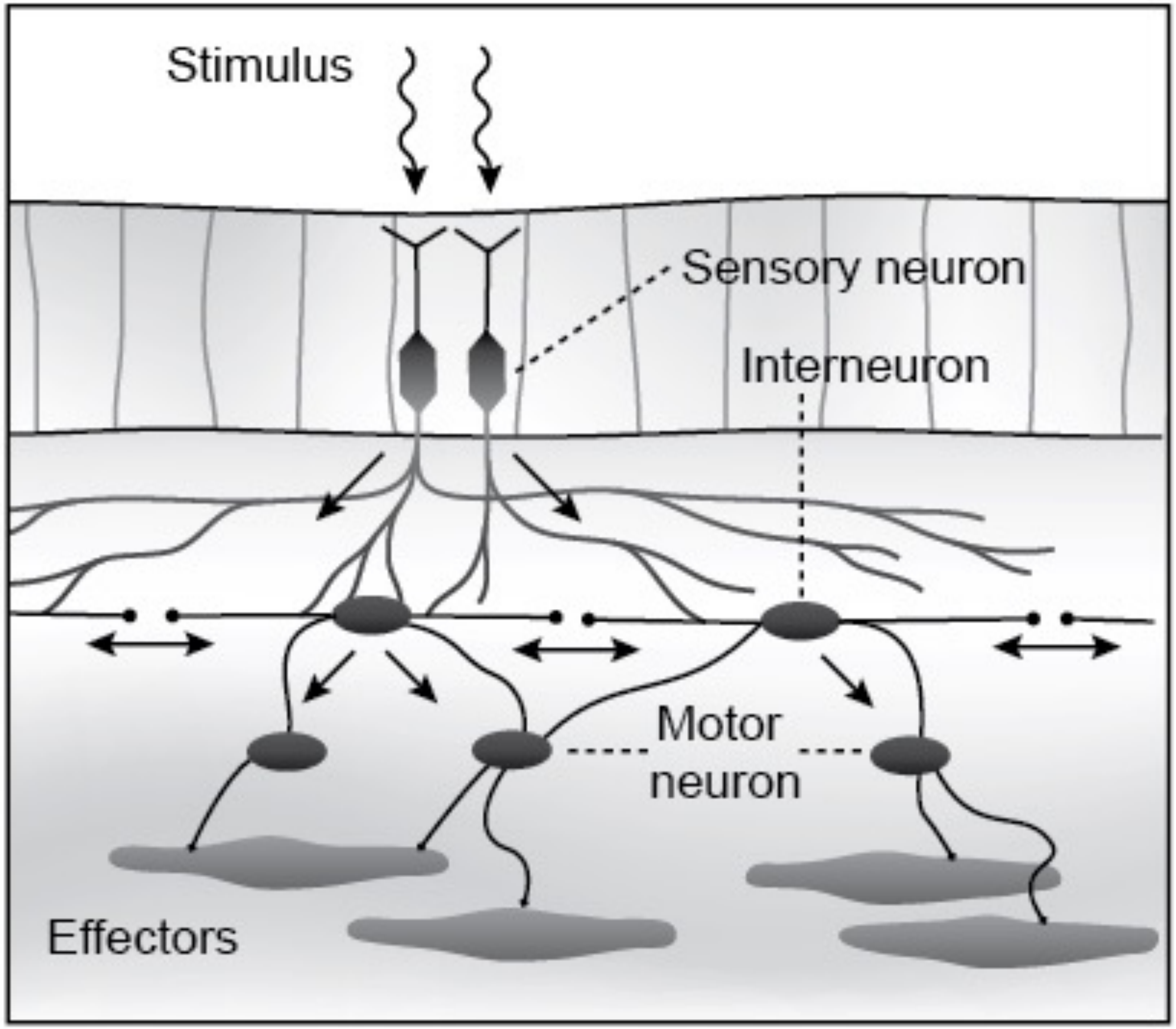
Types of Neurons

- Sensory neurons:
  - Peripheral cells that have an input detector exposed to the environment and an output connection that can diverge to make contact with neurons or effectors (muscle cells)
  - Massively distributed and parallel connections to be combined
  - Are used to synchronize when connected with effectors



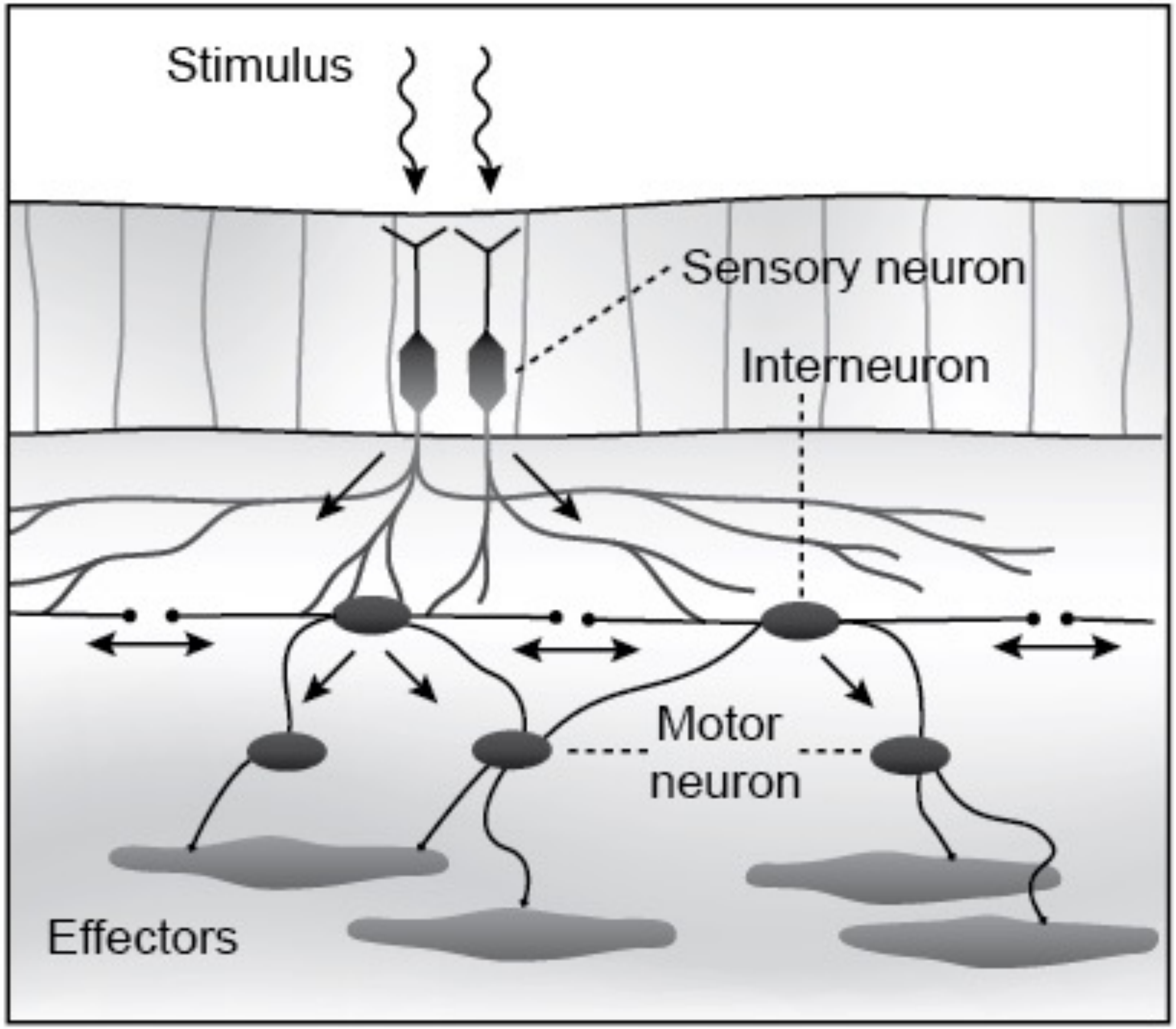
Types of Neurons

- Motor neurons:
  - Peripheral cells that send signals directly to effectors or other motor neurons
  - Combines signals from various sources for more complex effects



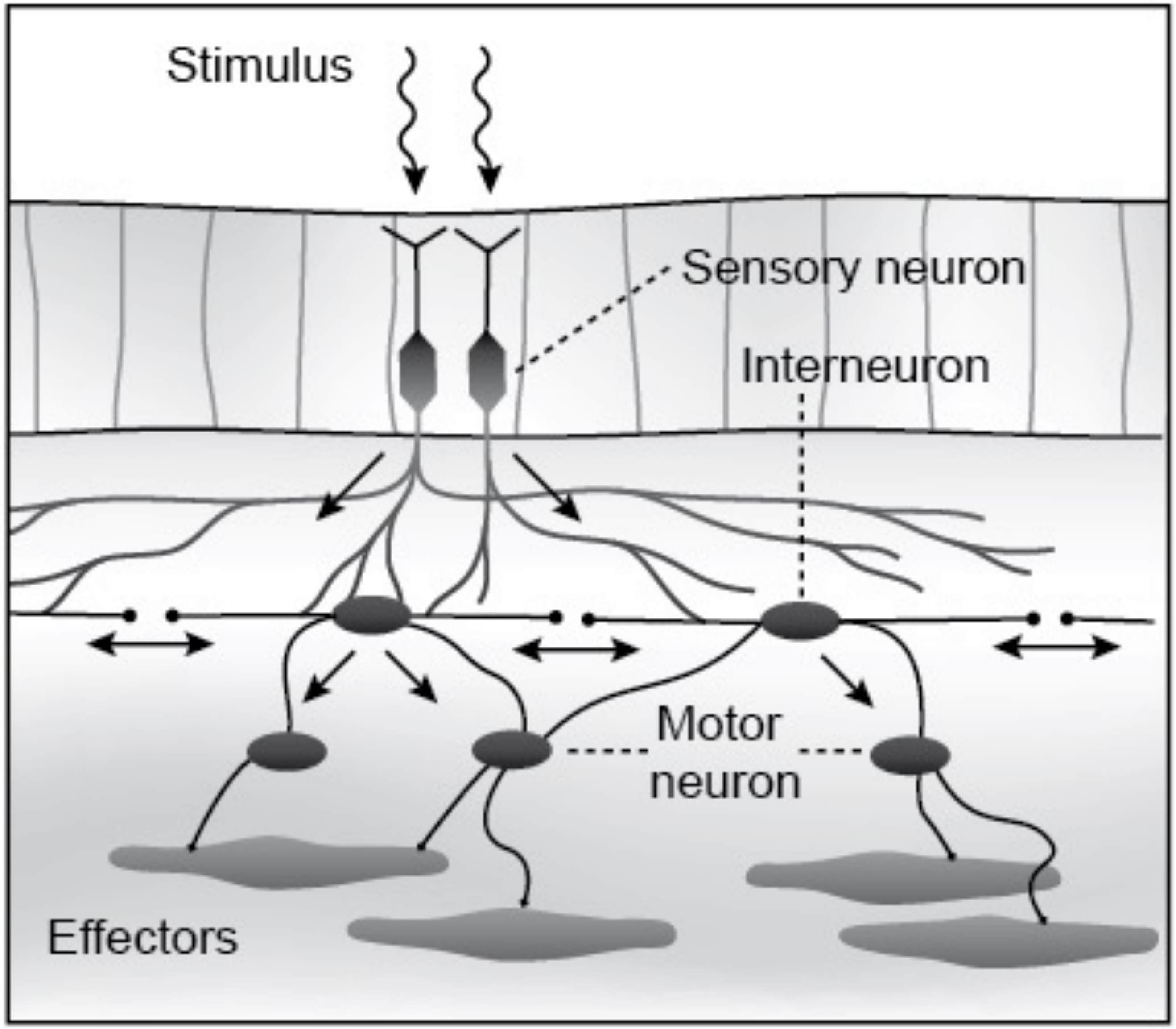
Types of Neurons

- Interneurons:
  - Not directly connected to the environment
  - Allow for more complex computations and signal transformation
  - Increases the complexity of the topologies that can be realized

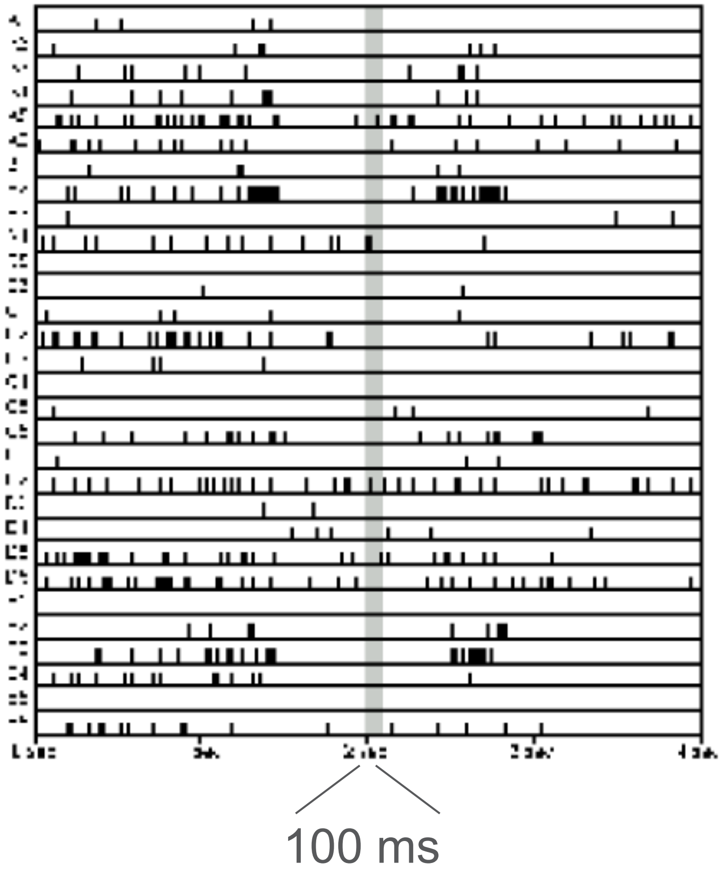


Types of Neurons

- Interneurons can be:
  - Excitatory: tend to increase activation of post-synaptic neurons
  - Inhibitory: tend to establish synaptic connections that decrease or block activity of post-synaptic neurons



How do neurons communicate?



Firing rate → McCulloch-Pitts

Firing time → Spiking neurons

Ongoing Argument in the Community!

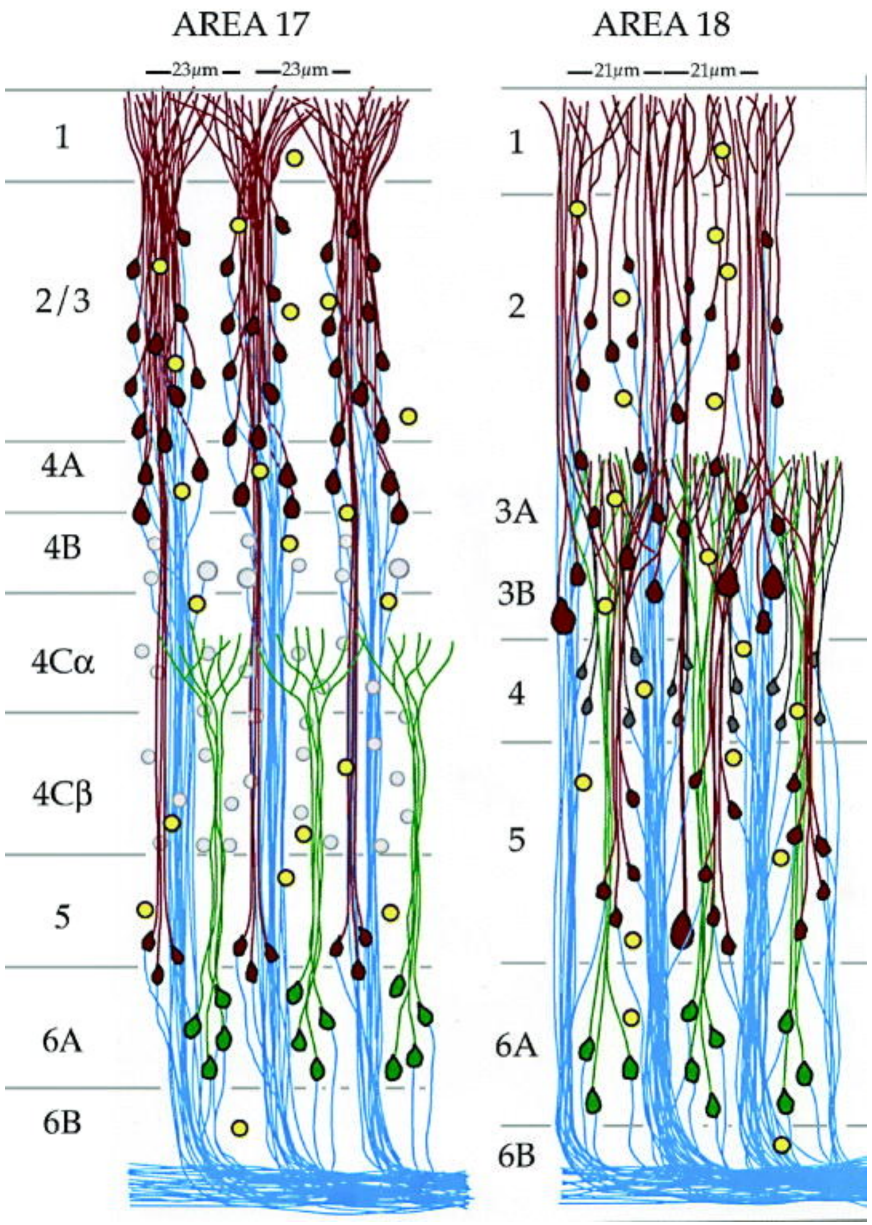
Adapted from companion slides for the book *Bio-Inspired Artificial Intelligence: Theories, Methods, and Technologies* by Dario Floreano and Claudio Mattiussi, MIT Press

Neural Topology

- Most neurons receive connections from and project to neighboring neurons.
  - Layout of neurons close to sensory areas tends to preserve topological relations of the receptors
  - Neighboring neurons tend to respond to similar patterns of stimulation
  - Nervous systems are organized in local circuits characterized by specific patterns of connectivity

Minicolumn

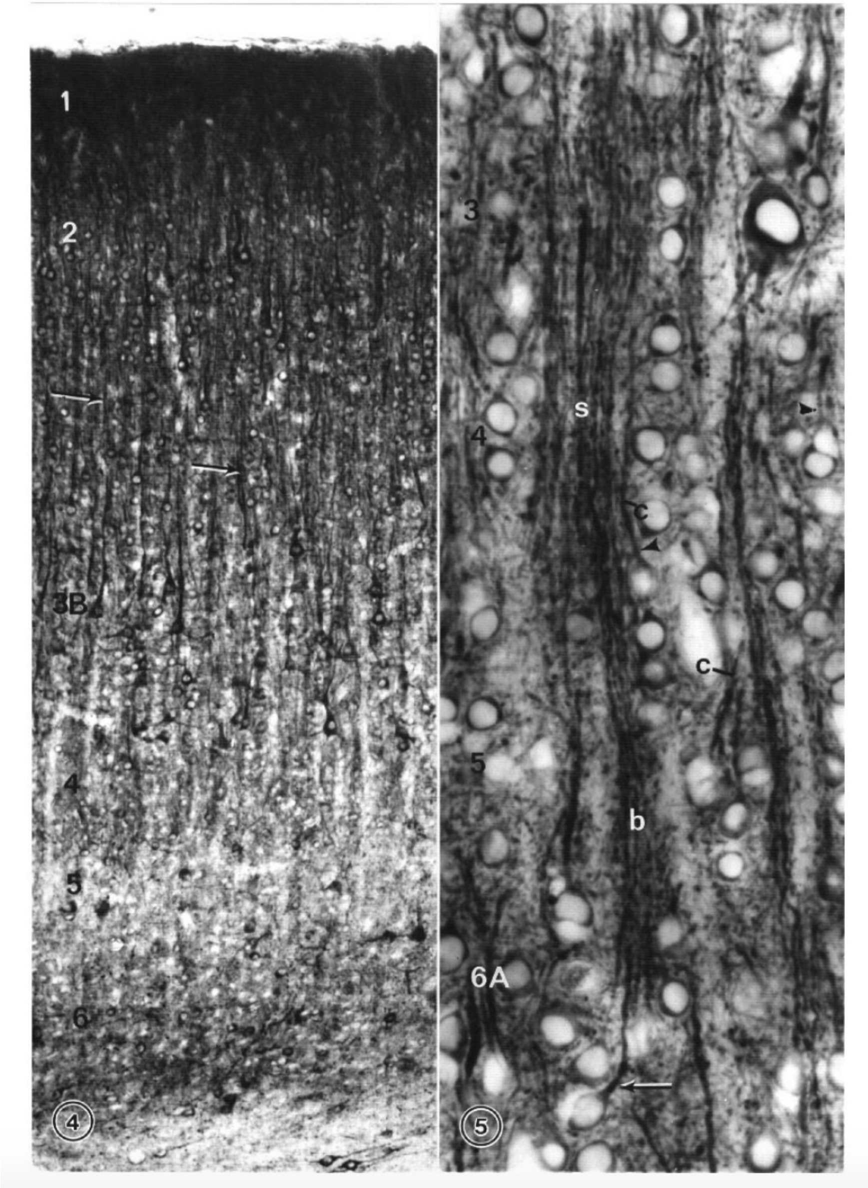
- Up to ~100 neurons
- 20–50μ diameter
- Length: 0.8 (mouse) to 3mm (human)
- ~ $6{\times}10^5$ synapses
- 75–90% synapses outside minicolumn
- Interacts with $1.2{\times}10^5$ other minicolumns
- Mutually excitable
- Also called microcolumn



Peters, Alan. "The morphology of minicolumns." *The neurochemical basis of Autism*. Springer, Boston, MA, 2010. 45-68.
Peters, A., J. M. Cifuentes, and C. Sethares. "The organization of pyramidal cells in area 18 of the rhesus monkey." Cerebral cortex (New York, NY: 1991) 7.5 (1997): 405-421.

Minicolumn

- Up to ~100 neurons
- 20–50μ diameter
- Length: 0.8 (mouse) to 3mm (human)
- ~ 6×105 synapses
- 75–90% synapses outside minicolumn
- Interacts with 1.2×105 other minicolumns
- Mutually excitable
- Also called microcolumn

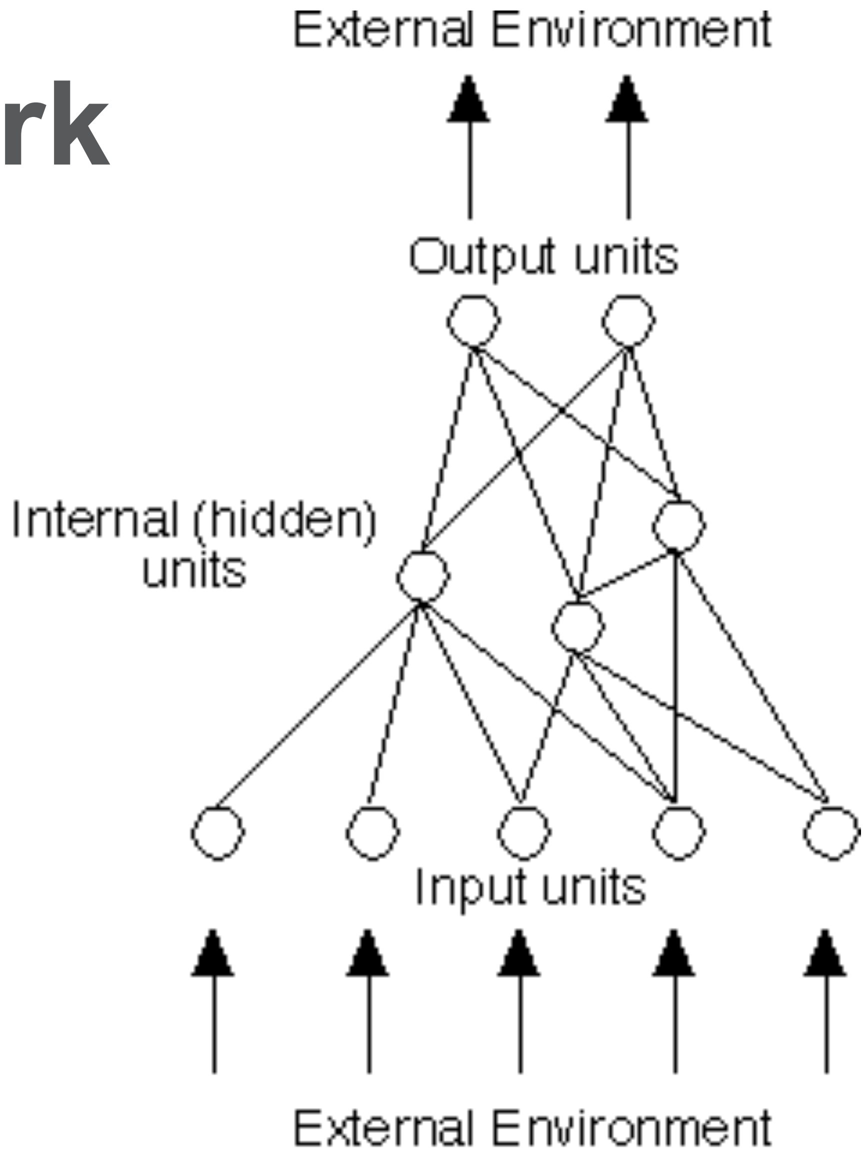


Peters, Alan. "The morphology of minicolumns." *The neurochemical basis of Autism*. Springer, Boston, MA, 2010. 45-68.
Peters, A., J. M. Cifuentes, and C. Sethares. "The organization of pyramidal cells in area 18 of the rhesus monkey." Cerebral cortex (New York, NY: 1991) 7.5 (1997): 405-421.

Artificial Neural Systems

An Artificial Neural Network

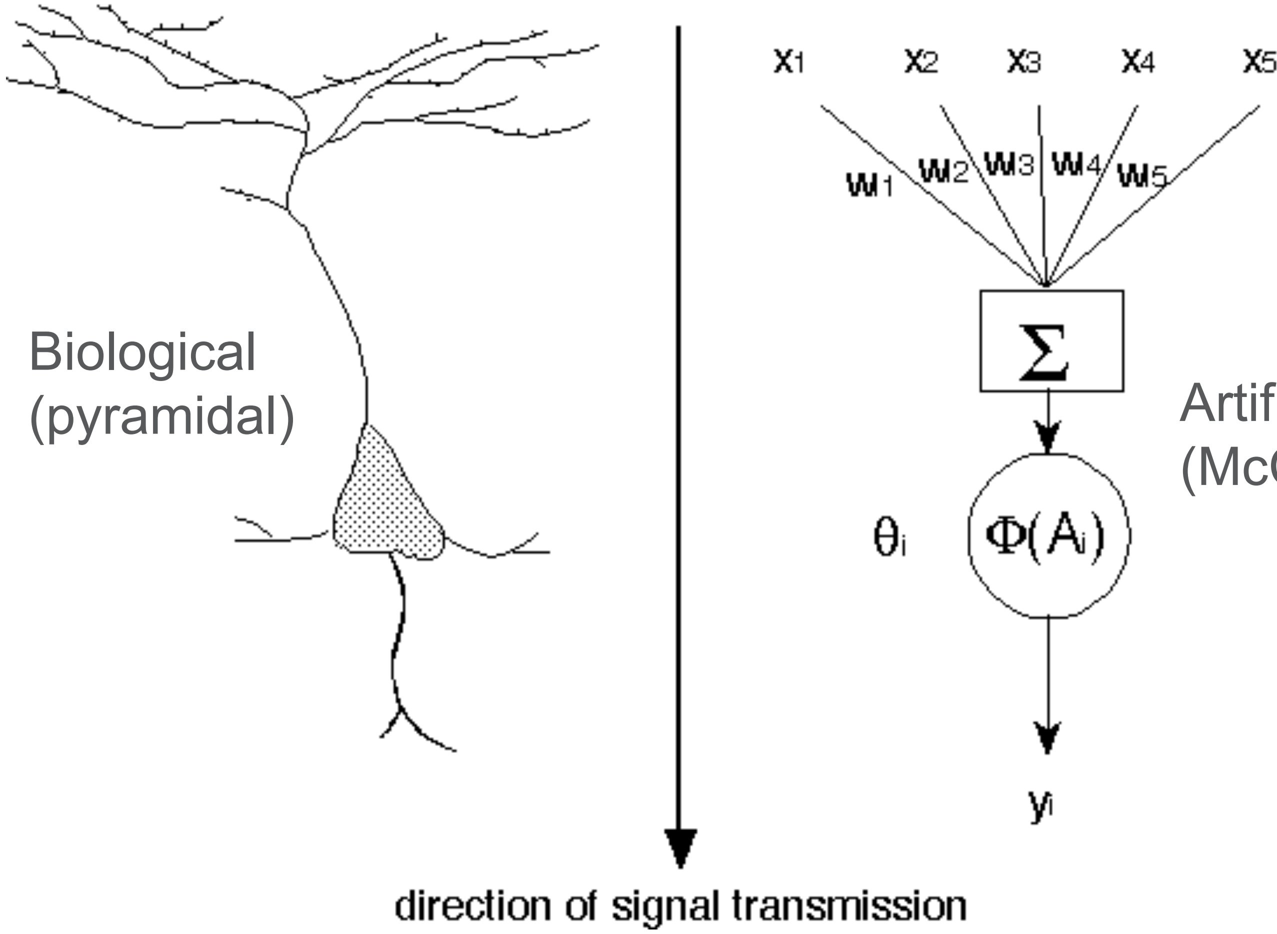
- A neural network communicates with the environments through input units and output units. All other elements are called internal or hidden units.
- Units are linked by uni-directional connections.
- A connection is characterized by a weight and a sign that transforms the signal.



Why Artificial Neural Networks?

- **Robustness**: Can handle signal degradation gracefully
- **Flexibility**: Can be applied to a wide variety of different problem types, including those where an analytical solution is not known
- **Generalization**: Trained on a limited number of samples, but can be operate on input patterns that are similar but never seen before
- **Content-based retrieval**: Some network models can retrieve memories when input patterns are missing information or corrupted by noise

Biological vs. Artificial Neurons



Biological (pyramidal)

Artificial (McCulloch-Pitts)

$$A_i = \sum_{j=1}^{N} w_{i,j} x_j$$

$$y_i = \Phi(A_i - \theta_i)$$

Adapted from companion slides for the book *Bio-Inspired Artificial Intelligence: Theories, Methods, and Technologies* by Dario Floreano and Claudio Mattiussi, MIT Press

McCulloch-Pitts Neuron

- Proposed by McCulloch and Pitts in 1943
- ***Not*** meant to be bio-realistic, but to capture some of the behavior of neurophysiological behaviors
- Started with binary inputs and outputs, some restrictions on the possible weights, and a more flexible threshold value.

$$A_i = \sum_{j=1}^{N} w_{i,j} x_j$$

$$y_i = \Phi(A_i - \theta_i)$$

Threshold to Bias

$$A_i = \sum_{j=1}^{N} w_{i,j} x_j$$

$$y_i = \Phi(A_i - \theta_i)$$

$$x_0 = -1$$

$$w_{i,0} = \theta_i$$

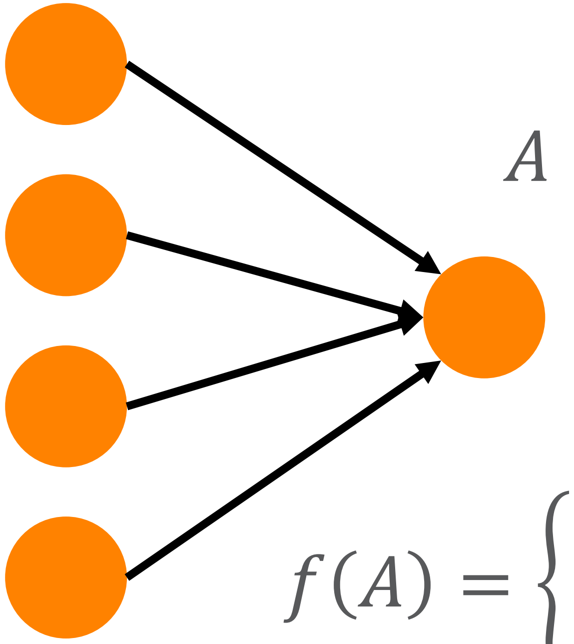
$$A_i = \sum_{j=0}^{N} w_{i,j} x_j$$

$$y_i = \Phi(A_i)$$

ADALINE (Adaptive Linear Neuron), introduced by Widrow and Hoff in 1960

Perceptron

- Linear classifier developed by Rosenblatt in 1958
- Single layer perceptrons were shown to only be able to deal with linearly separable data
  - Minsky and Papert showed it was impossible for this class of networks to solve the XOR problem

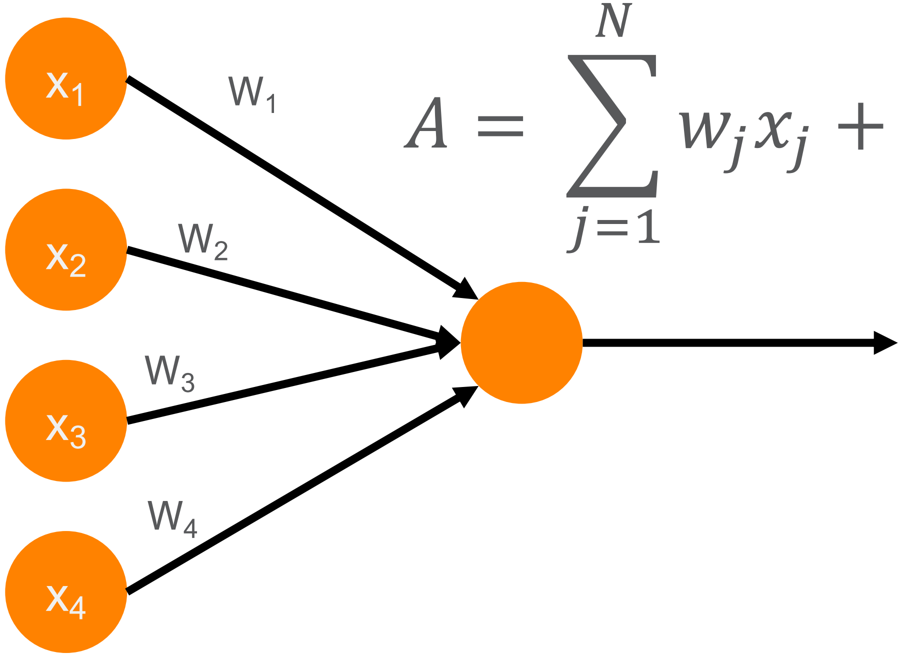


$$A = \sum_{j=1}^{N} w_j x_j$$

$$f(A) = \begin{cases} 0 \ if \ A < 0 \\ 1 \ if \ A \geq 0 \end{cases}$$

Adapted from companion slides for the book *Bio-Inspired Artificial Intelligence: Theories, Methods, and Technologies* by Dario Floreano and Claudio Mattiussi, MIT Press

How do you train a perceptron?

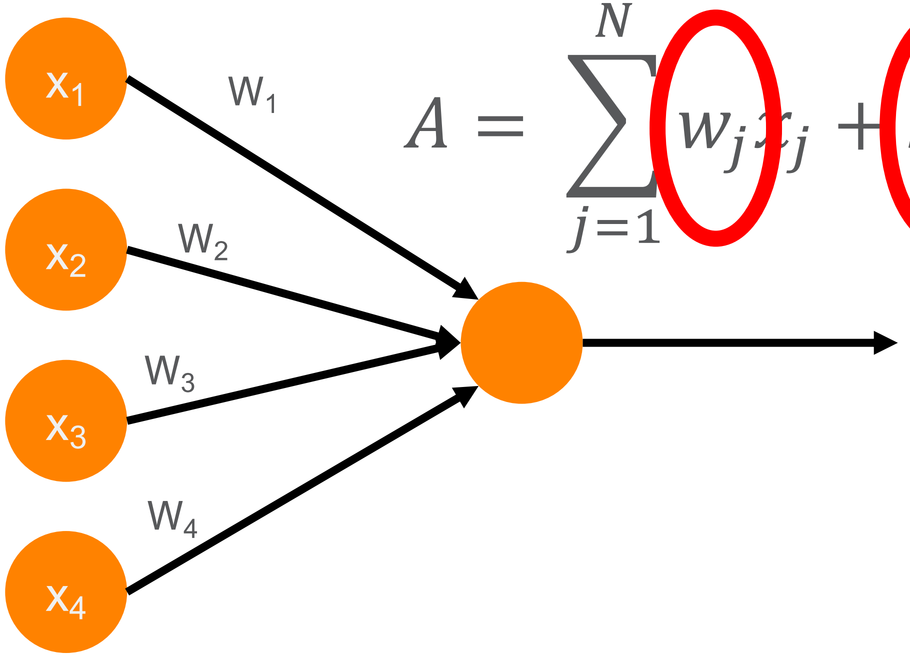


$$A = \sum_{j=1}^{N} w_j x_j + b$$

$$f(A) = \begin{cases} 0 \; if \; A \leq 0 \\ 1 \; if \; A > 0 \end{cases}$$

How do you train a perceptron?

$$A = \sum_{j=1}^{N} w_j x_j + b$$



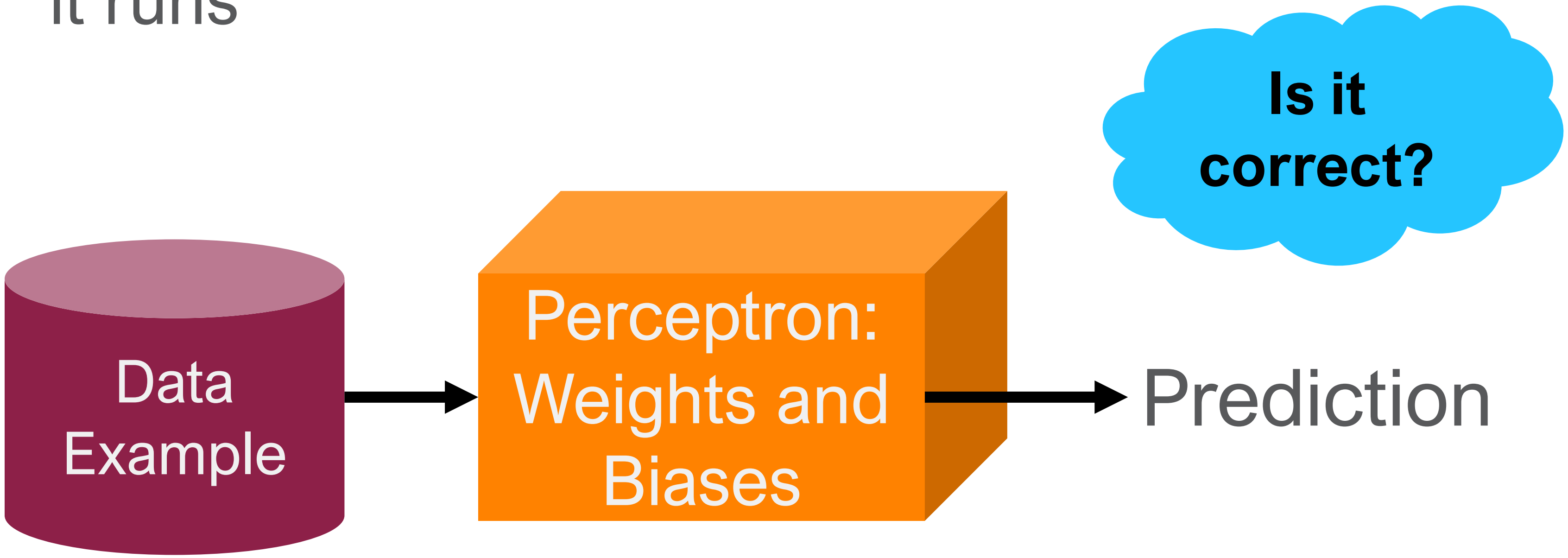
$$f(A) = \begin{cases} 0 \ if \ A \leq 0 \\ 1 \ if \ A > 0 \end{cases}$$

Error-Driving Update: Perceptron Algorithm

- The perceptron maintains a guess at parameters (weights and bias) as it runs

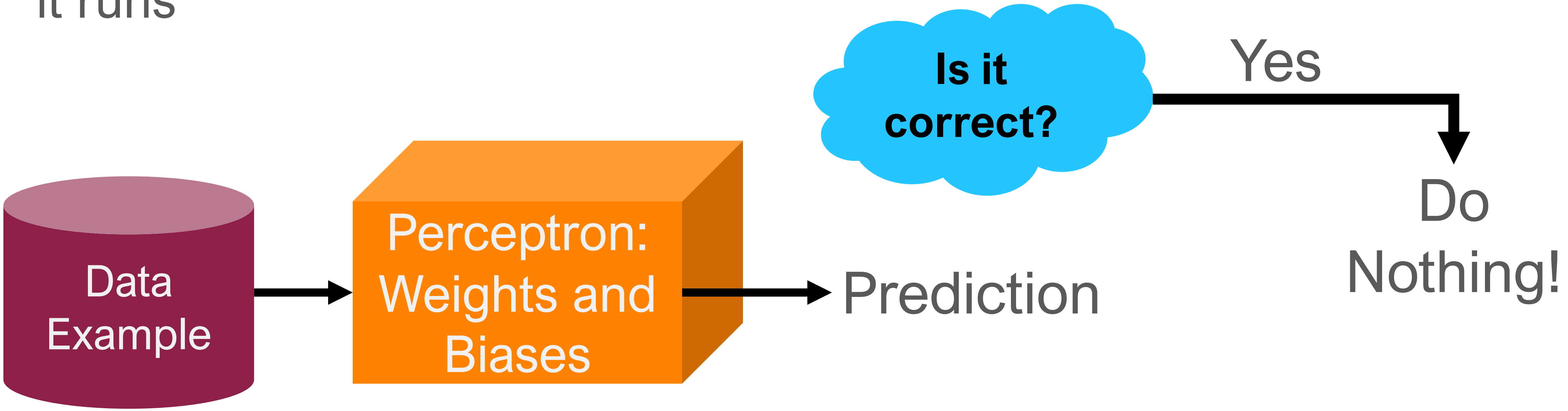
Error-Driving Update: Perceptron Algorithm

- The perceptron maintains a guess at parameters (weights and bias) as it runs



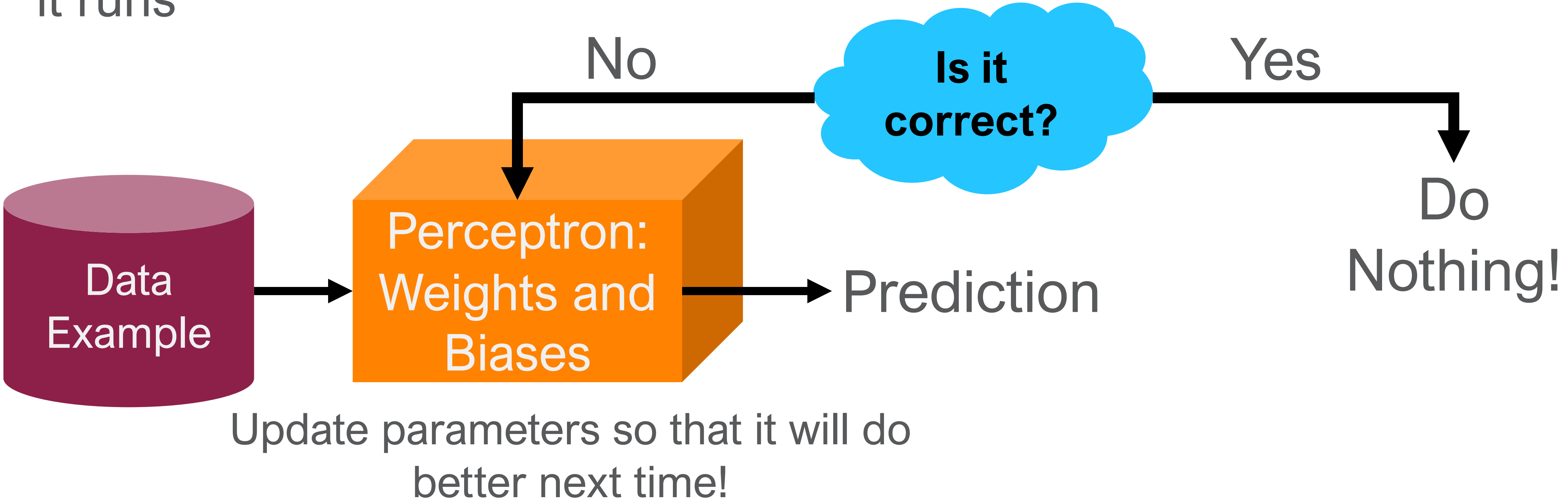
Error-Driving Update: Perceptron Algorithm

- The perceptron maintains a guess at parameters (weights and bias) as it runs



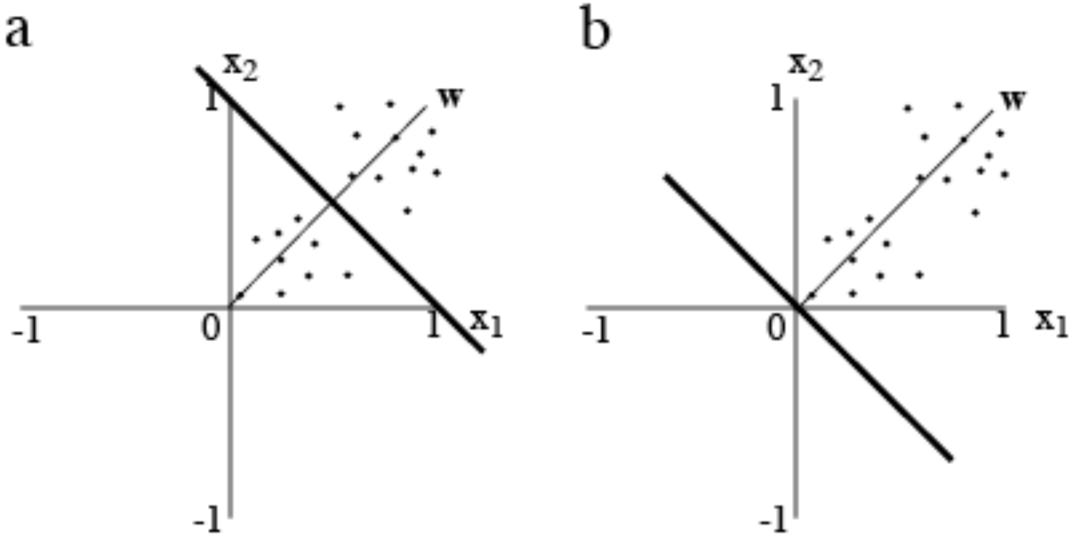
Error-Driving Update: Perceptron Algorithm

- The perceptron maintains a guess at parameters (weights and bias) as it runs



What is an individual neuron doing?

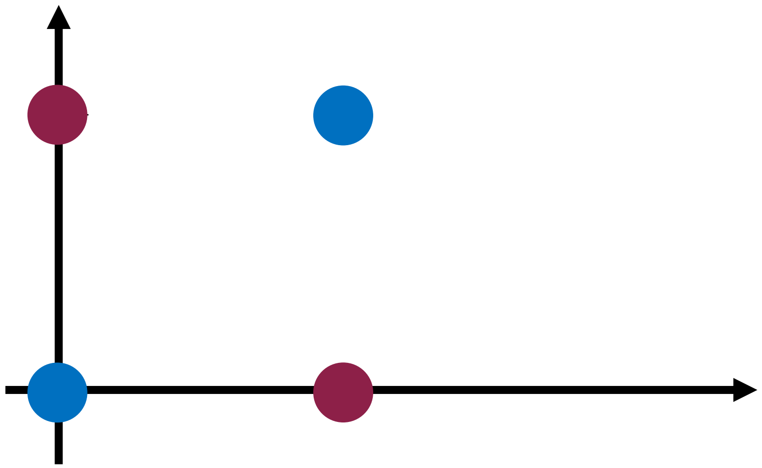
- You can think of a neuron as tracing a separation line in the input space between the classes of input patterns that produce different responses



$$\vartheta > 0 \qquad\qquad \vartheta = 0$$

Limitations of the Perceptron

- Although the perceptron is useful, it is FUNDAMENTALLY limited
- It can only deal with data that is linearly separable
- The XOR problem is not linearly separable!



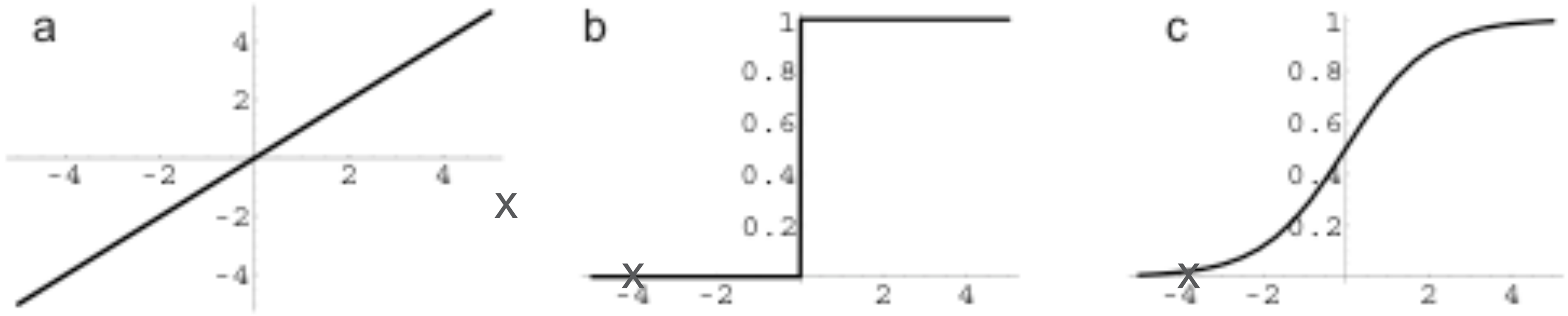
Limitations of the Perceptron

- The XOR problem (and all related problems) was so significant that it killed research in classifiers with linear decision boundaries for a decade or two
  - Minsky and Papert published on the limitations of perceptrons in 1969
  - It contributed to the first AI Winter!

Making the Perceptron more complex…

Activation Functions

Identity $\Phi(x)$ | Step $\Phi(x)$ | Sigmoid $\Phi(x)$



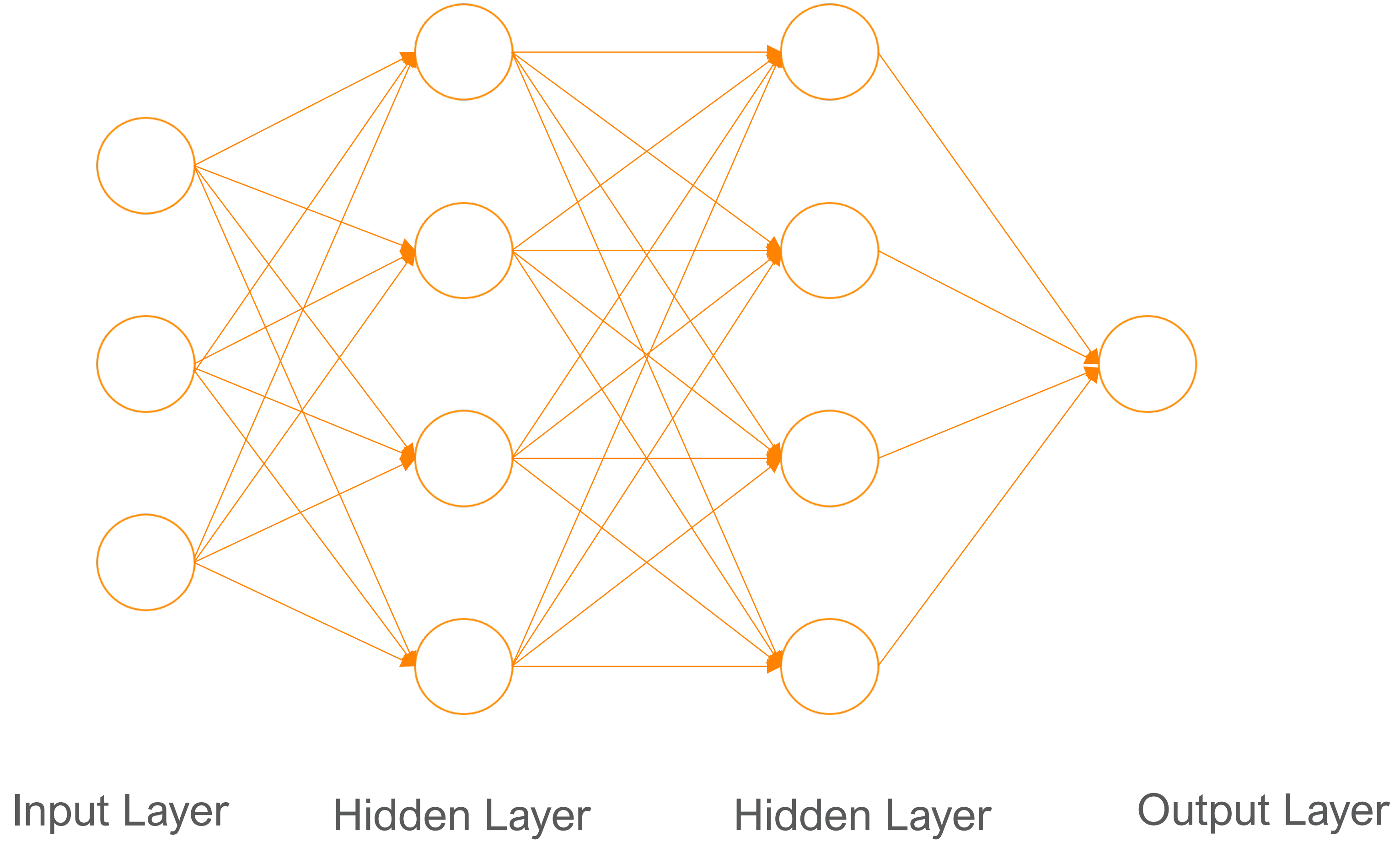
Sigmoid function:

- continuous
- non-linear
- monotonic
- bounded
- asymptotic

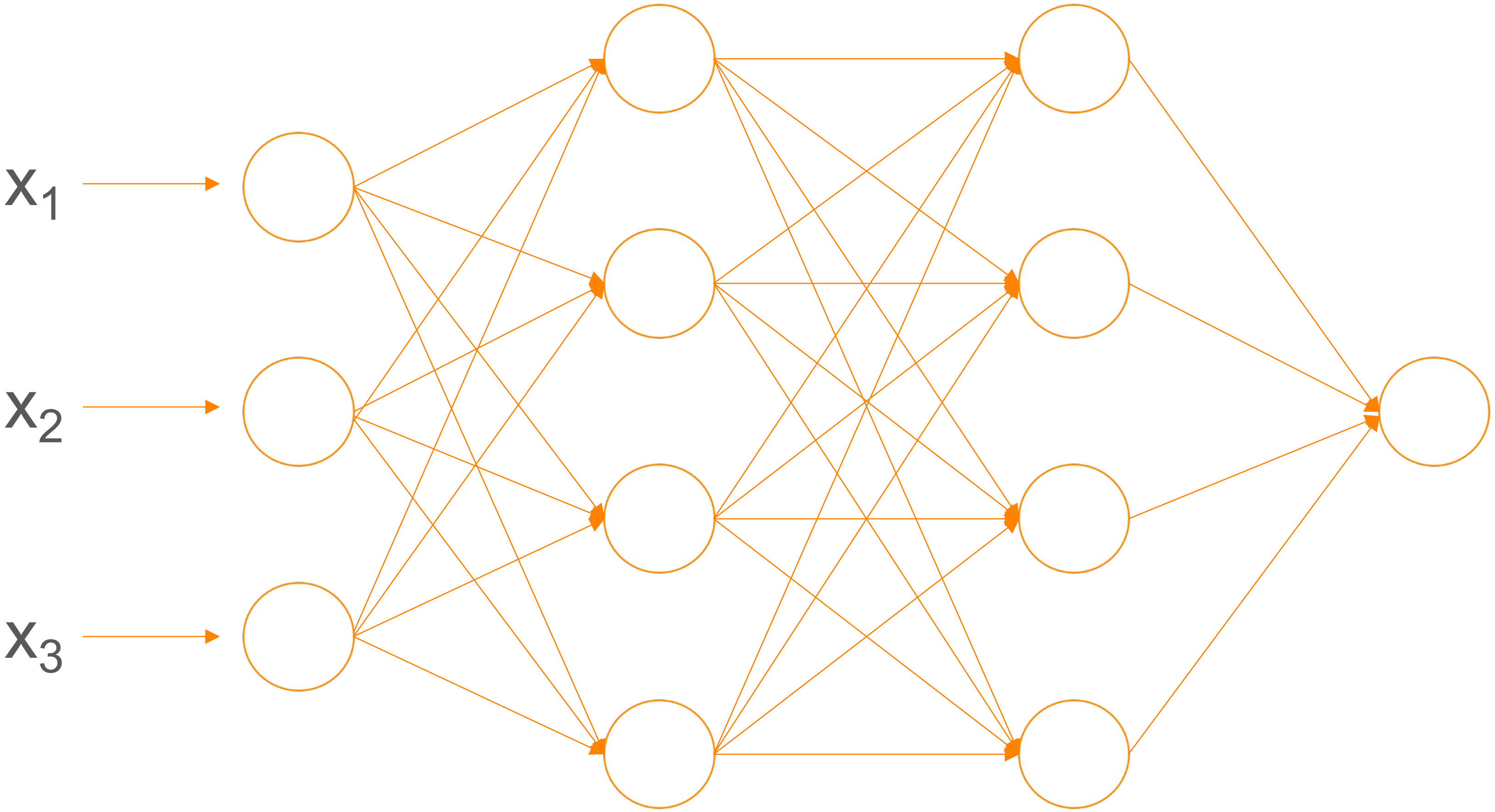
$$\Phi(x) = \frac{1}{1 + e^{-kx}}$$

Adapted from companion slides for the book *Bio-Inspired Artificial Intelligence: Theories, Methods, and Technologies* by Dario Floreano and Claudio Mattiussi, MIT Press

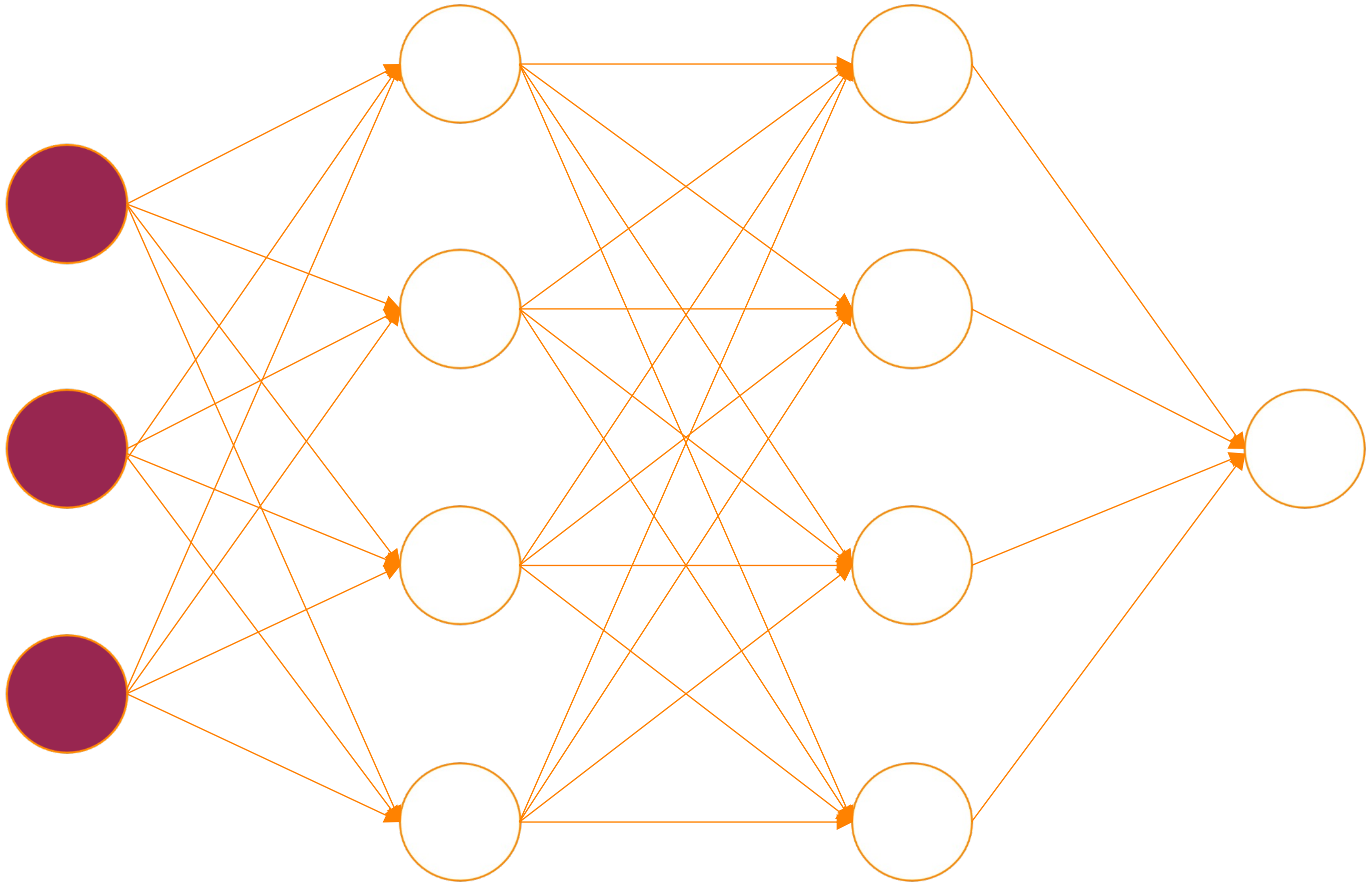
Feed-Forward Neural Networks



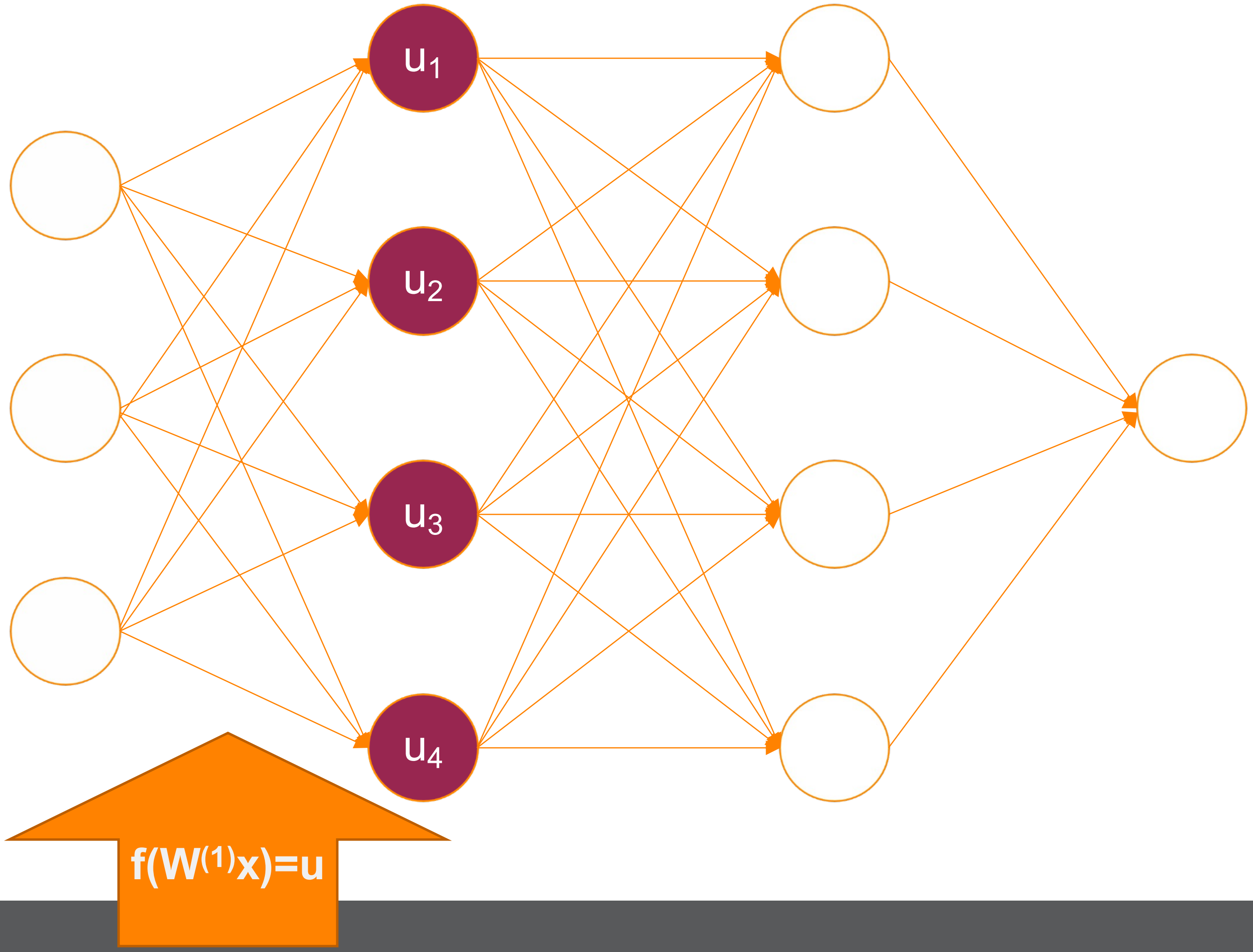
Feed-Forward Neural Networks



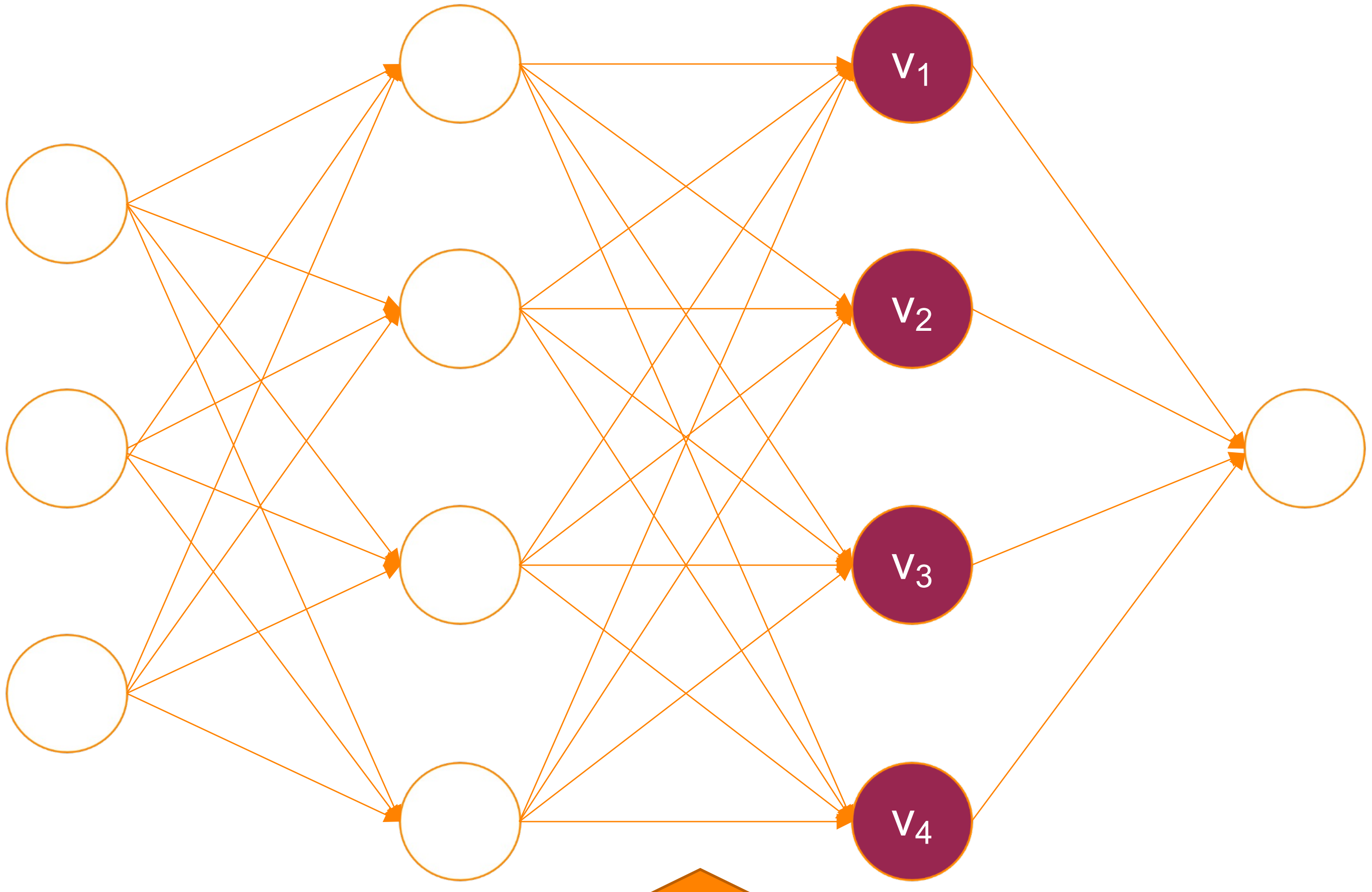
Feed-Forward Neural Networks



Feed-Forward Neural Networks

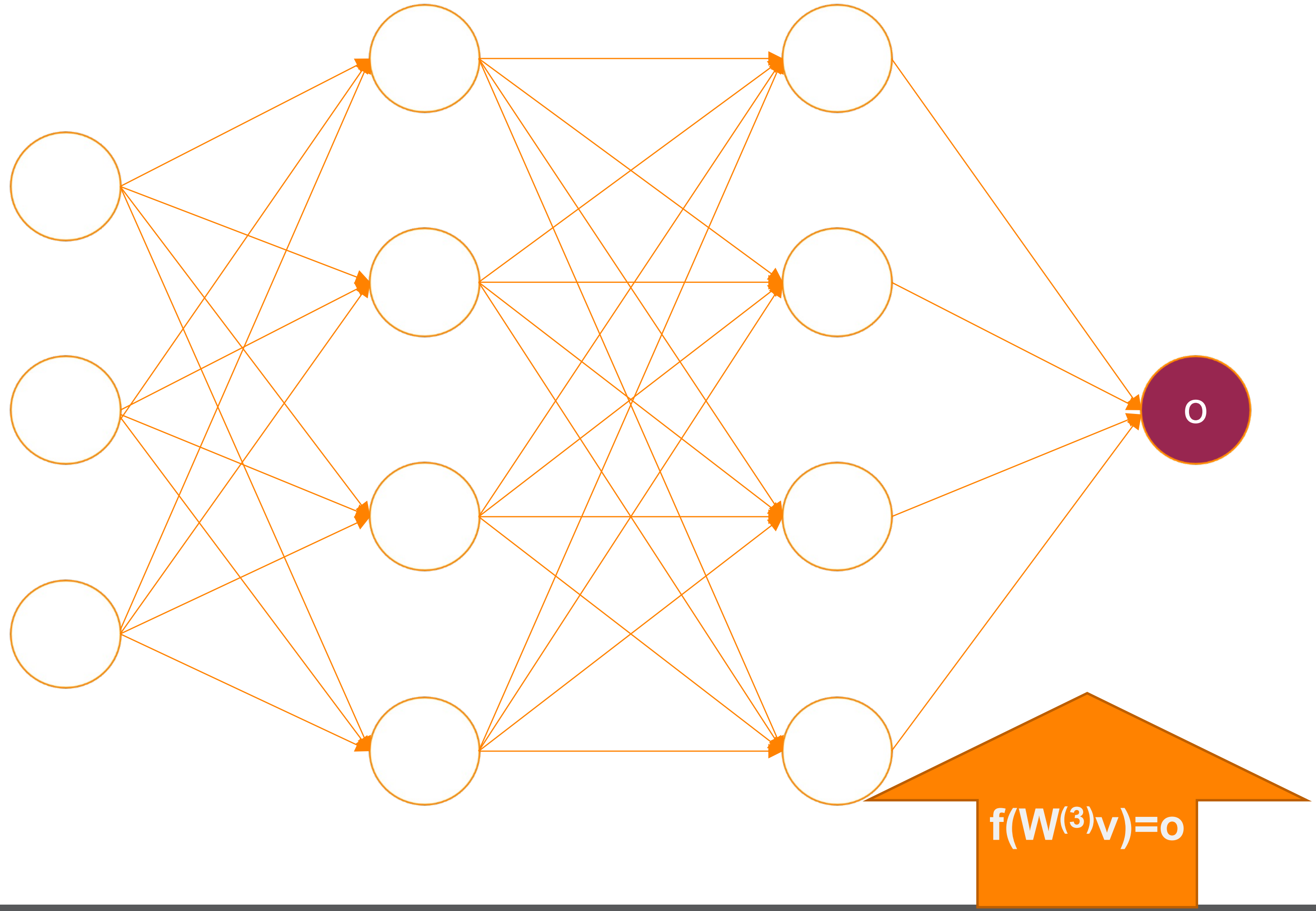


Feed-Forward Neural Networks



$f(W^{(2)}u)=v$

Feed-Forward Neural Networks



More Layers!

- Perceptrons have simple learning mechanisms, but they could not be applied for networks with hidden layers
- Backpropagation was named and popularized by Rumelhart, Hinton & Williams in 1986
  - Though it definitely had predecessors prior to that.
- Back-propagation is an approach that leverages the chain rule and gradient descent to update weights in a multi-layer network
- It allowed for training for hidden layers

What do hidden layers give us?

Universal Function Approximator

- Theorem: Let F be a continuous function on a bounded subset of D-dimensional space. Then, there is a two-layer neural network $\hat{F}$ with a finite number of hidden units that approximate F arbitrarily well. Specifically, for all x in the domain of F, $\left|F(x) - \hat{F}(x)\right| < \epsilon$.

- What does this mean?
  - Two-layer networks can approximate *any* function!

Practicalities

- If you have a function F and some error tolerance parameter ϵ, you can construct a two-layer network that will compute F

- Going from one layer to two completely changes the representational power of the network!

- However, it doesn't tell us what the network needs to look like…

More Layers?!

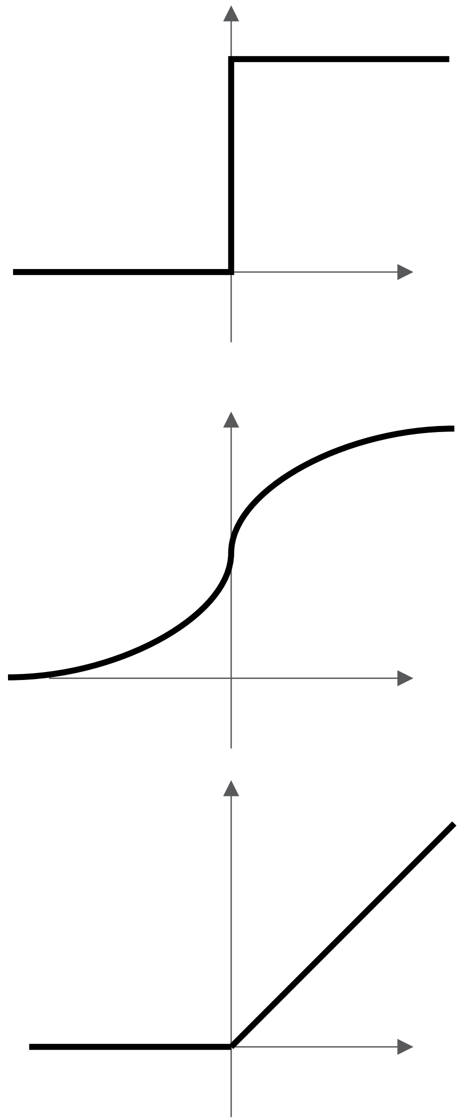
- As multi-layer perceptrons were popularized, and more layers were added, the vanishing gradients problem with back-propagation was found
  - The "deeper" the network, the less likely it was that back-propagation would work effectively
- The sigmoid function was part of the problem, which has led to the popular usage of the rectified linear unit (ReLU) activation function:

$$f(x) = \max(x, 0)$$

More Layers?!

- As multi-layer perceptrons were popularized, and more layers were added, the vanishing gradients problem with back-propagation was found
  - The ”deeper” the network, the less likely it was that back-propagation would work effectively
- The sigmoid function was part of the problem, which has led to the popular usage of the rectified linear unit (ReLU) activation function:

$$f(x) = \max(x, 0)$$



Tensorflow Playground

playground.tensorflow.org

Announcements

- In the next lecture, we'll go over recurrent neural networks and Hopfield networks
  - This lecture will be asynchronous as well
- There is an associated quiz on Canvas for this lecture that is due before by midnight on Tuesday, February 21
  - This is graded for participation, not correctness