Artificial Neural Networks (ANN)

Dr. Kisor K. Sahu

SMMME, IITBBS

Brain

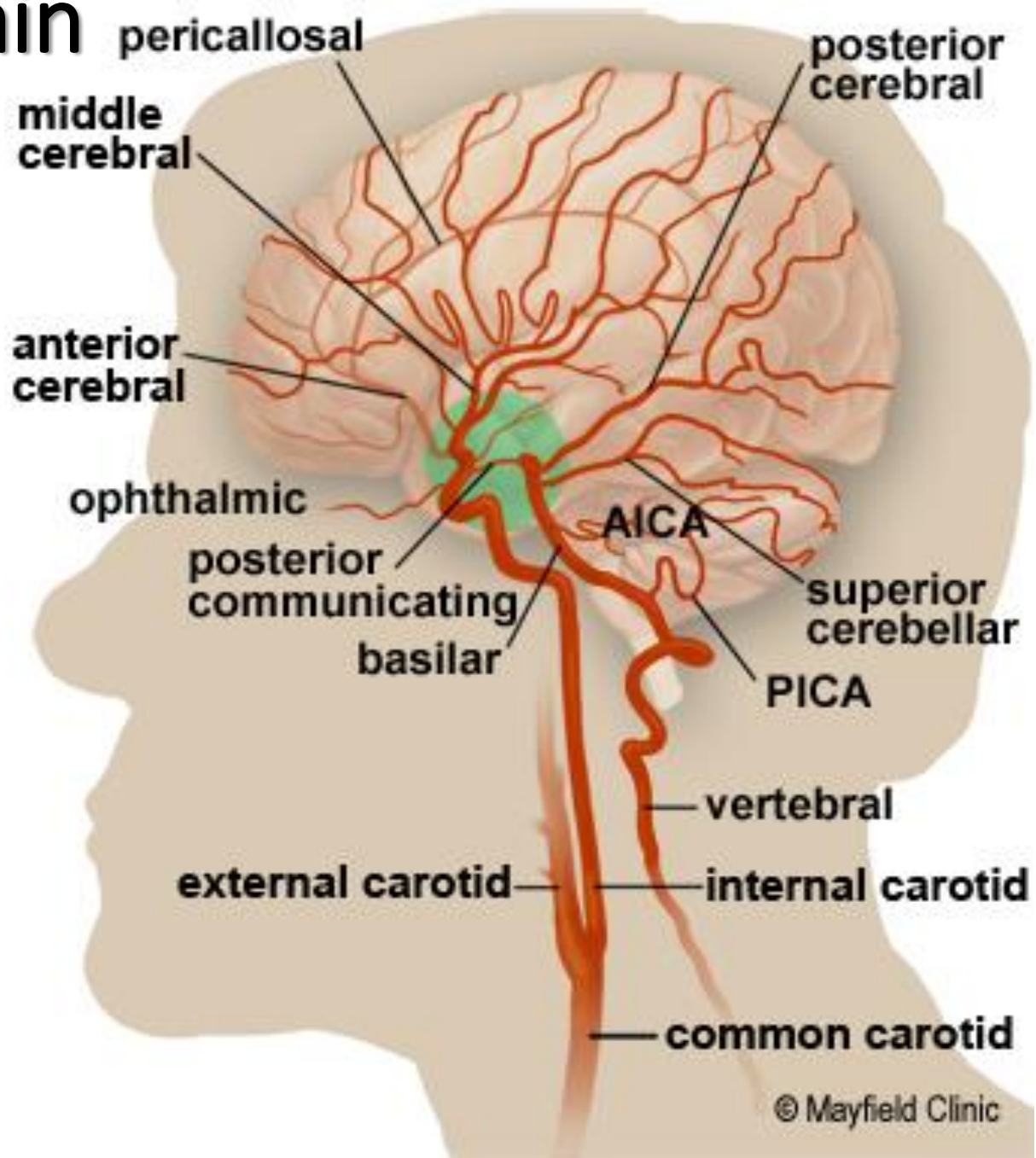


Table 1

Properties	Computer	Human Brain
Number of Basic Units	Up to 10 billion transistors	~100 billion neurons; ~100 trillion synapses
Speed of Basic Operation	10 billion/sec.	< 1,000/sec.
Precision	1 in ~4.2 billion (for a 32-bit processor)	~1 in 100
Power Consumption	~100 watts	~10 watts
Information Processing Mode	Mostly serial	Serial and massively parallel
Input/Output for Each Unit	1–3	~1,000
Signaling Mode	Digital	Digital and analog

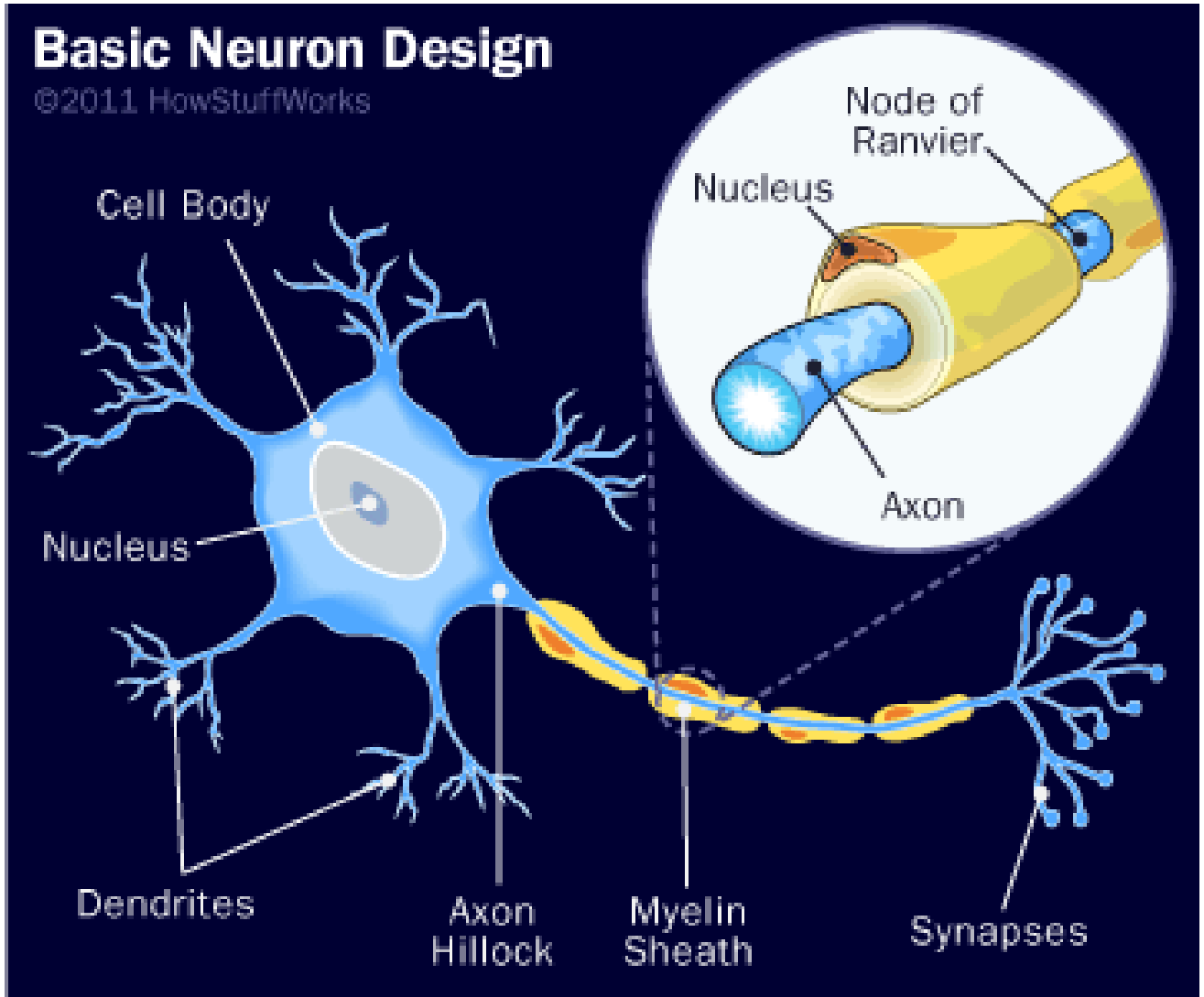
a) Based on personal computers in 2008.

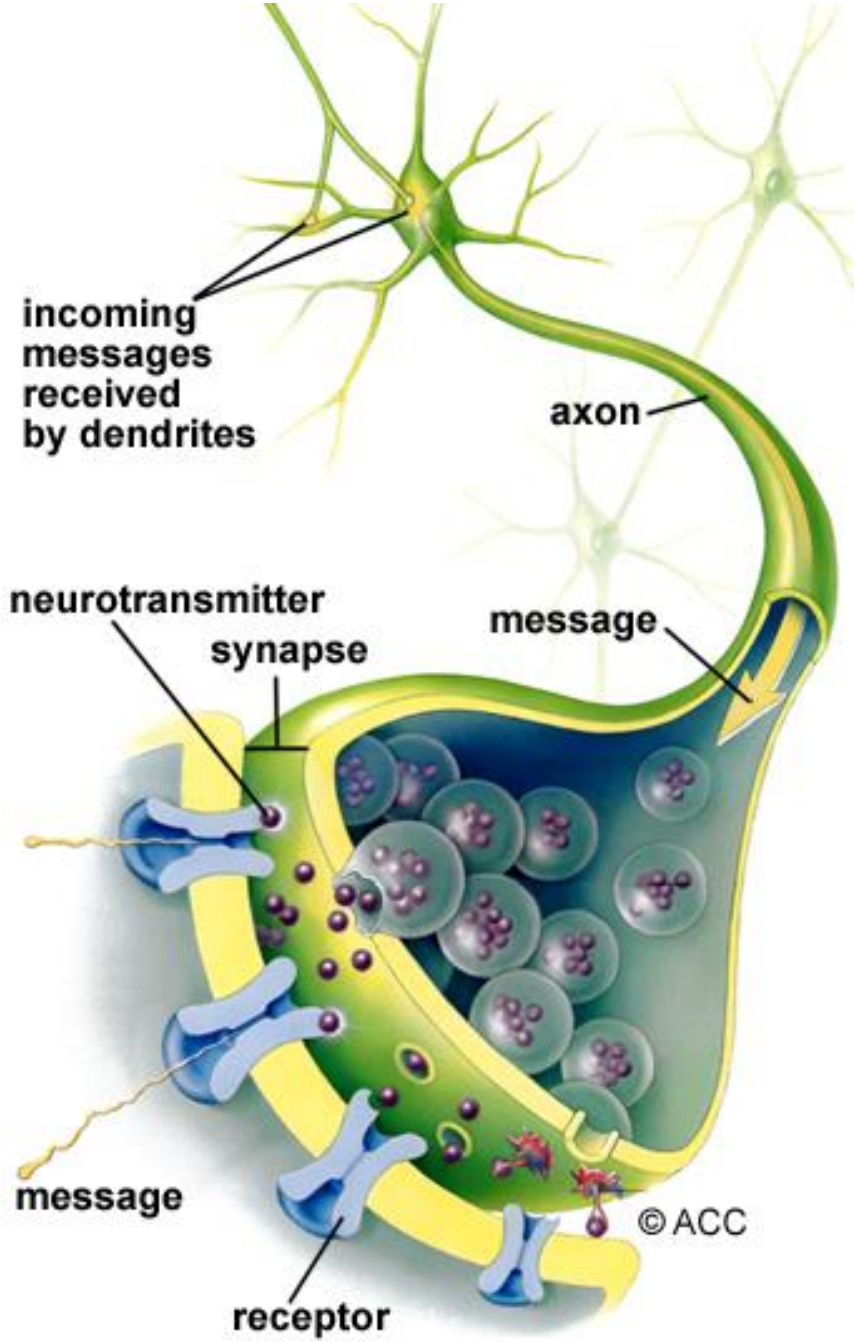
b) The number of transistors per integrative circuit has doubled every 18–24 months in the past few decades; in recent years the performance gains from this transistor growth have slowed, limited by energy consumption and heat dissipation.
References: John von Neumann, The Computer and the Brain (New Haven: Yale University Press, 2012); D. A. Patterson and J. L. Hennessy, Computer Organization and Design (Amsterdam: Elsevier, 2012).

Brain

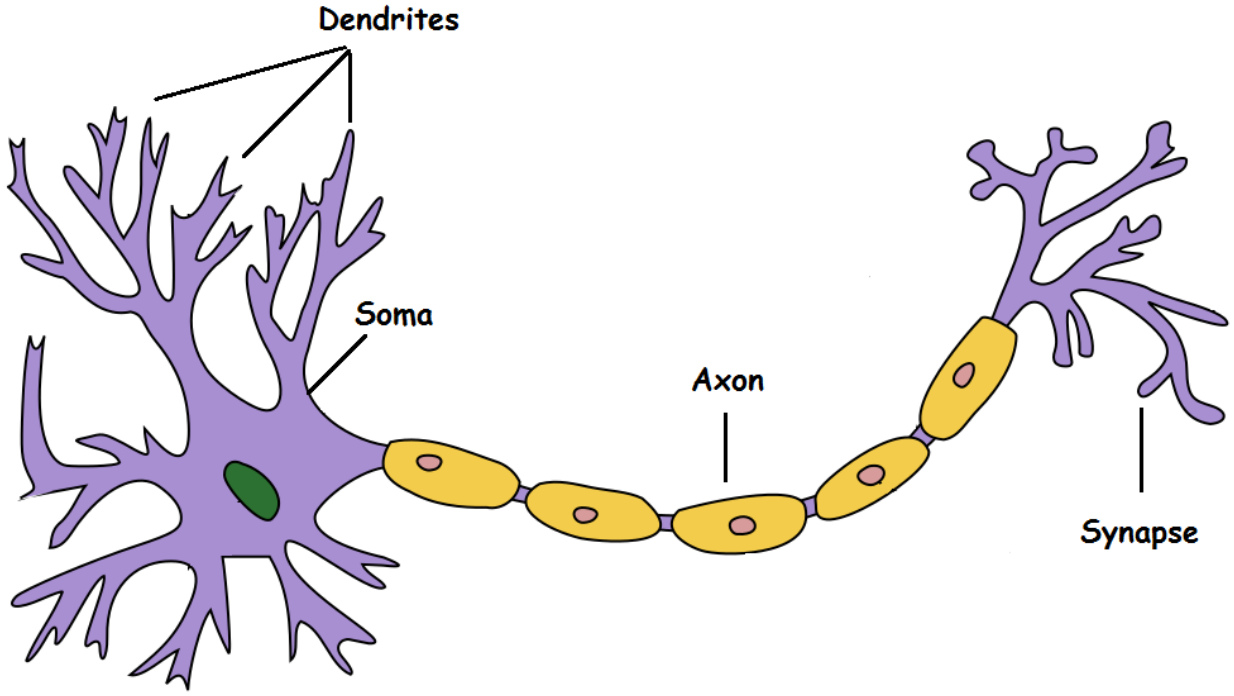
Properties	Computer	Human Brain
Number of Basic Units	Up to 10 billion transistors	~100 billion neurons; ~100 trillion synapses
Speed of Basic Operation	10 billion/sec.	< 1,000/sec.
Precision	1 in ~4.2 billion (for a 32-bit processor)	~1 in 100
Power Consumption	~100 watts	~10 watts
Information Processing Mode	Mostly serial	Serial and massively parallel
Input/Output for Each Unit	1–3	~1,000
Signaling Mode	Digital	Digital and analog

Neurones


Basic Neuron Design
©2011 HowStuffWorks
Cell Body
Nucleus
Dendrites
Axon Hillock
Myelin Sheath
Synapses
Node of Ranvier
Nucleus
Axon


incoming messages received by dendrites
axon
neurotransmitter
synapse
message
message
receptor
© ACC

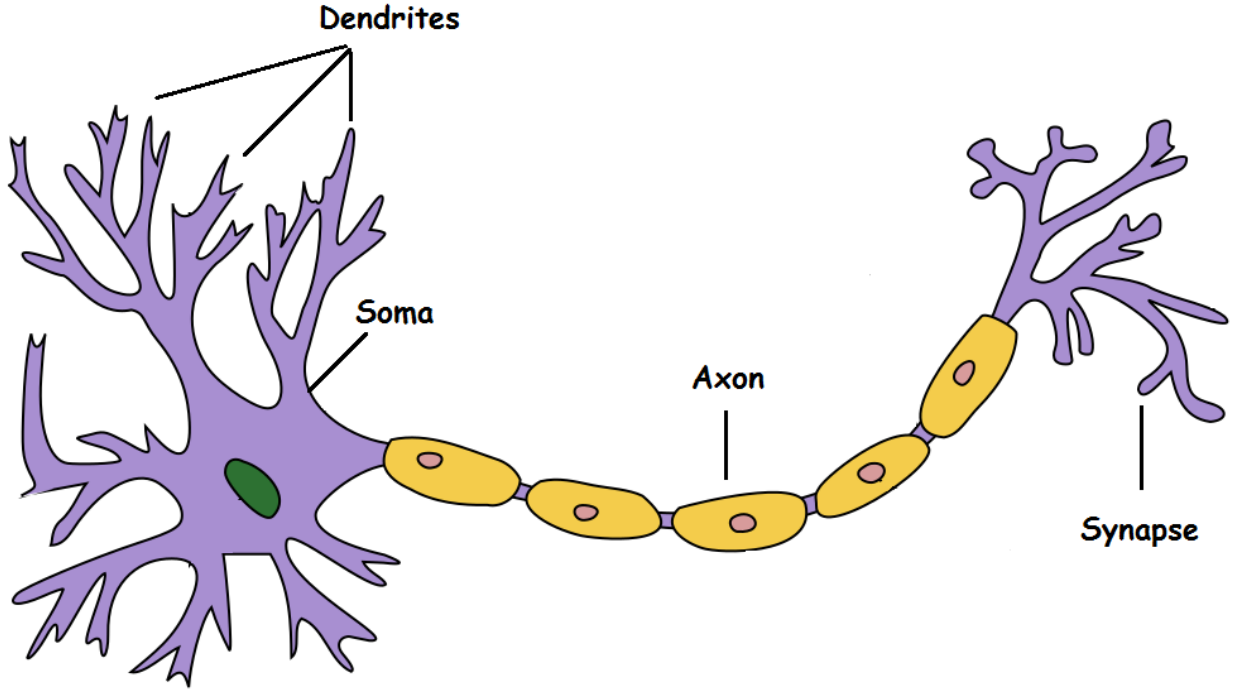
Working of a Biological Neuron



- **Dendrites** – They are tree-like branches, responsible for receiving the information from other neurons it is connected to. In other sense, we can say that they are like the ears of neuron.

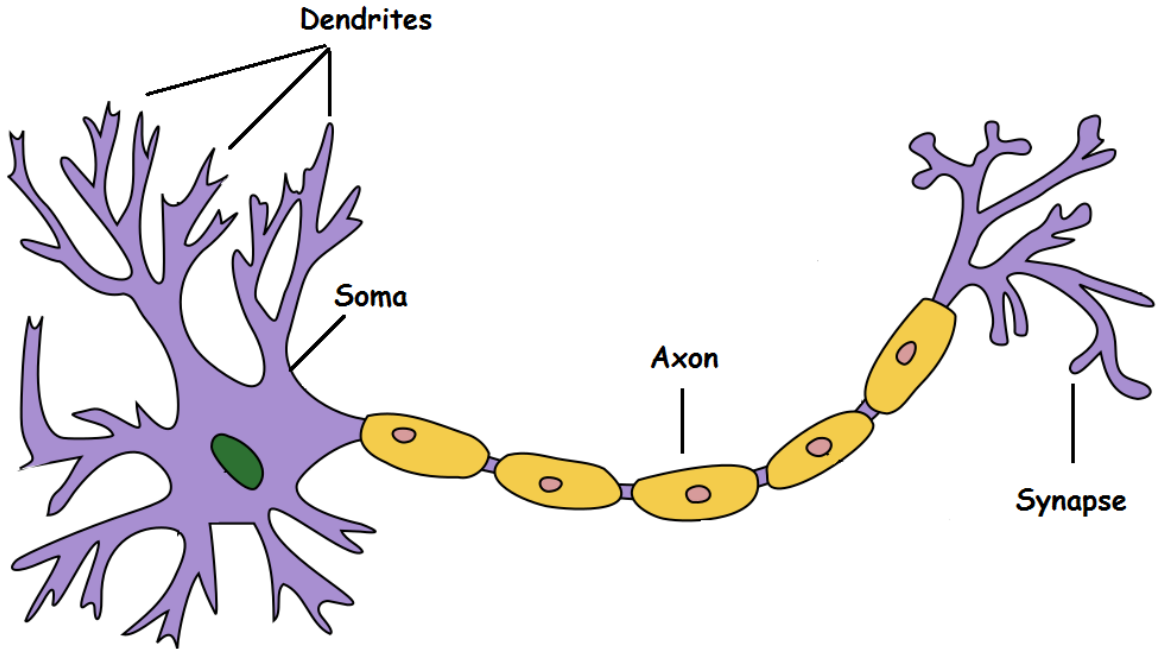
- **Soma** – It is the cell body of the neuron and is responsible for processing of information, they have received from dendrites.
- **Axon** – It is just like a cable through which neurons send the information.
- **Synapses** – It is the connection between the axon and other neuron dendrites.

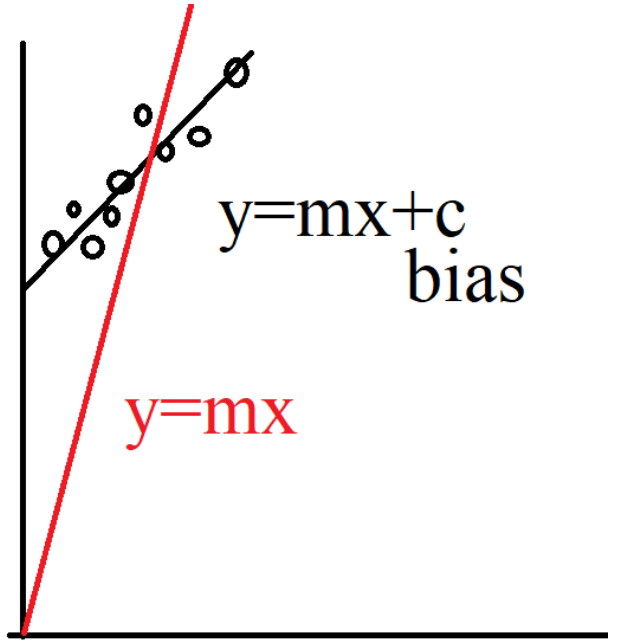
ANN vs. BNN

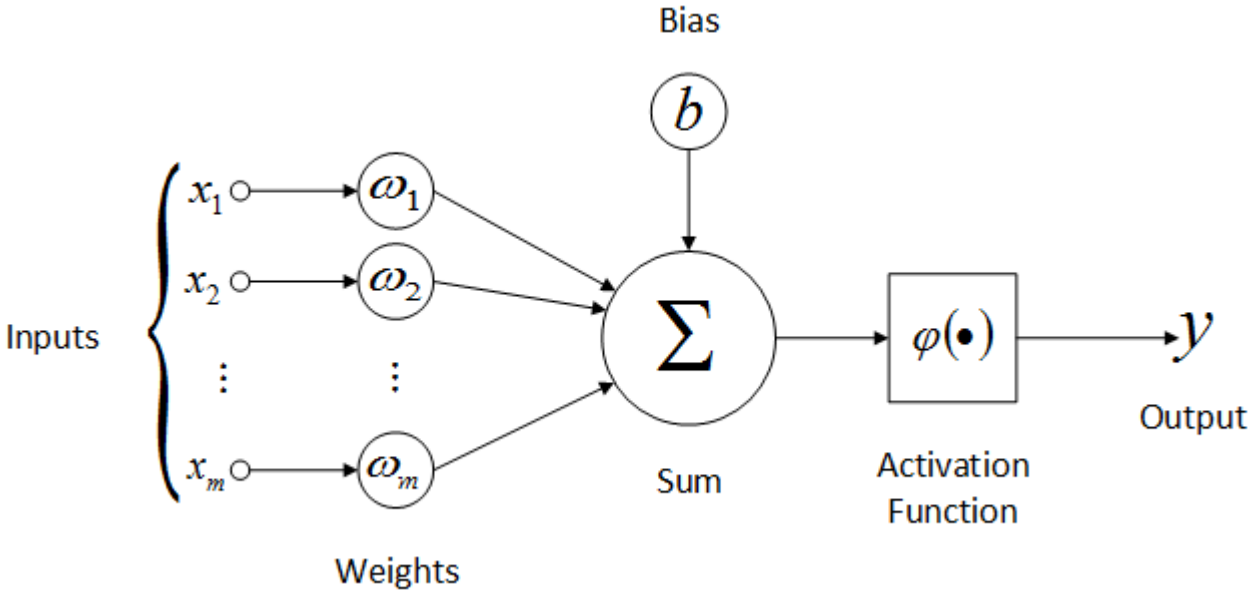


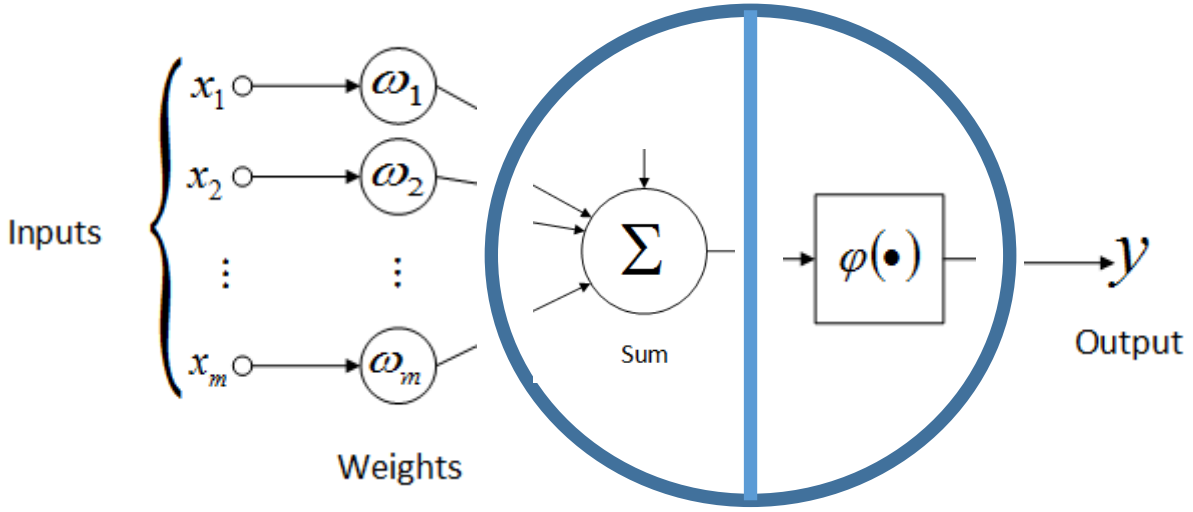
Biological Neural Network (BNN)	Artificial Neural Network (ANN)
Soma	Node
Dendrites	Input
Synapse	Weights or Interconnections
Axon	Output

Mathematical model of a neuron

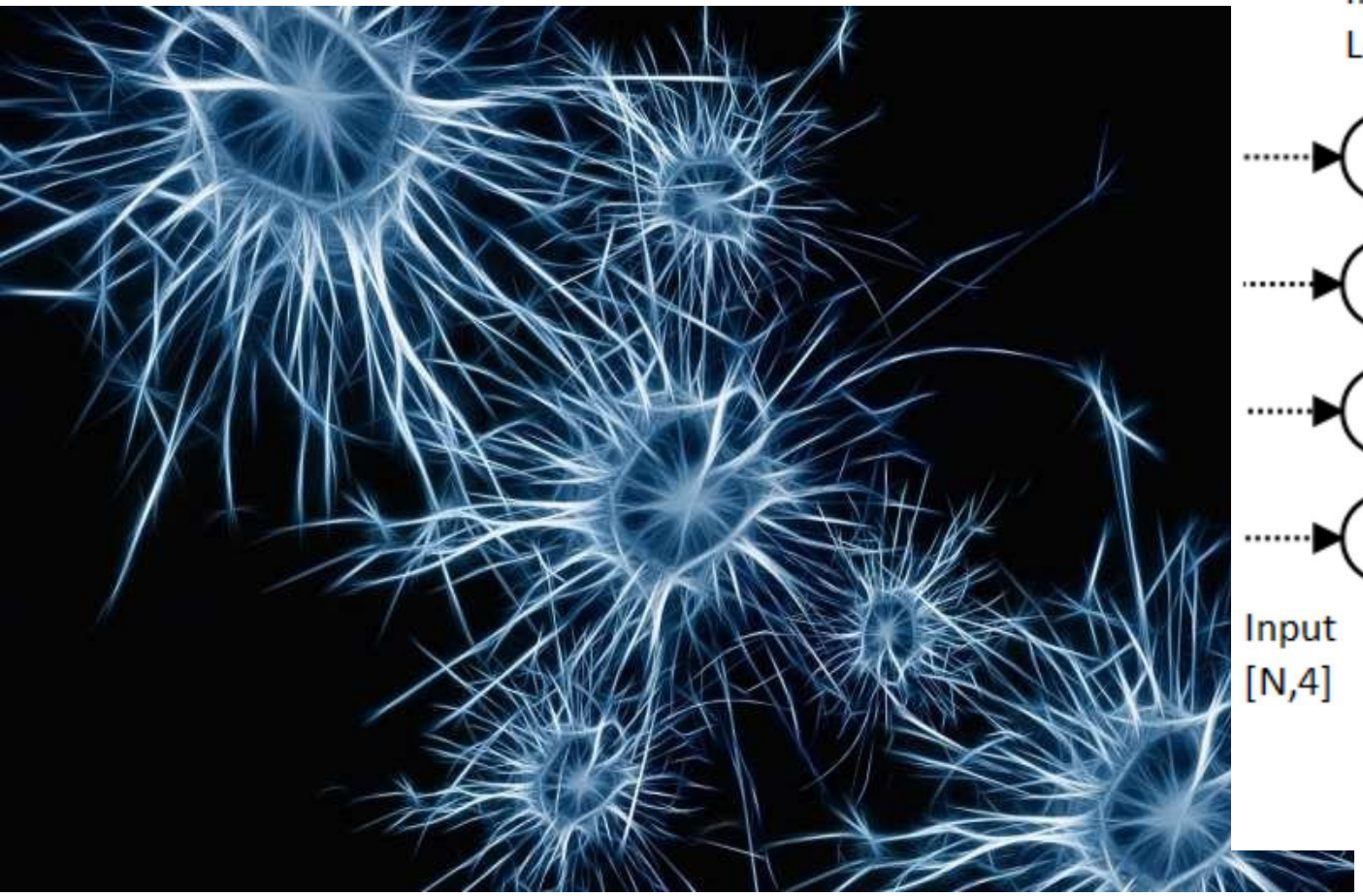


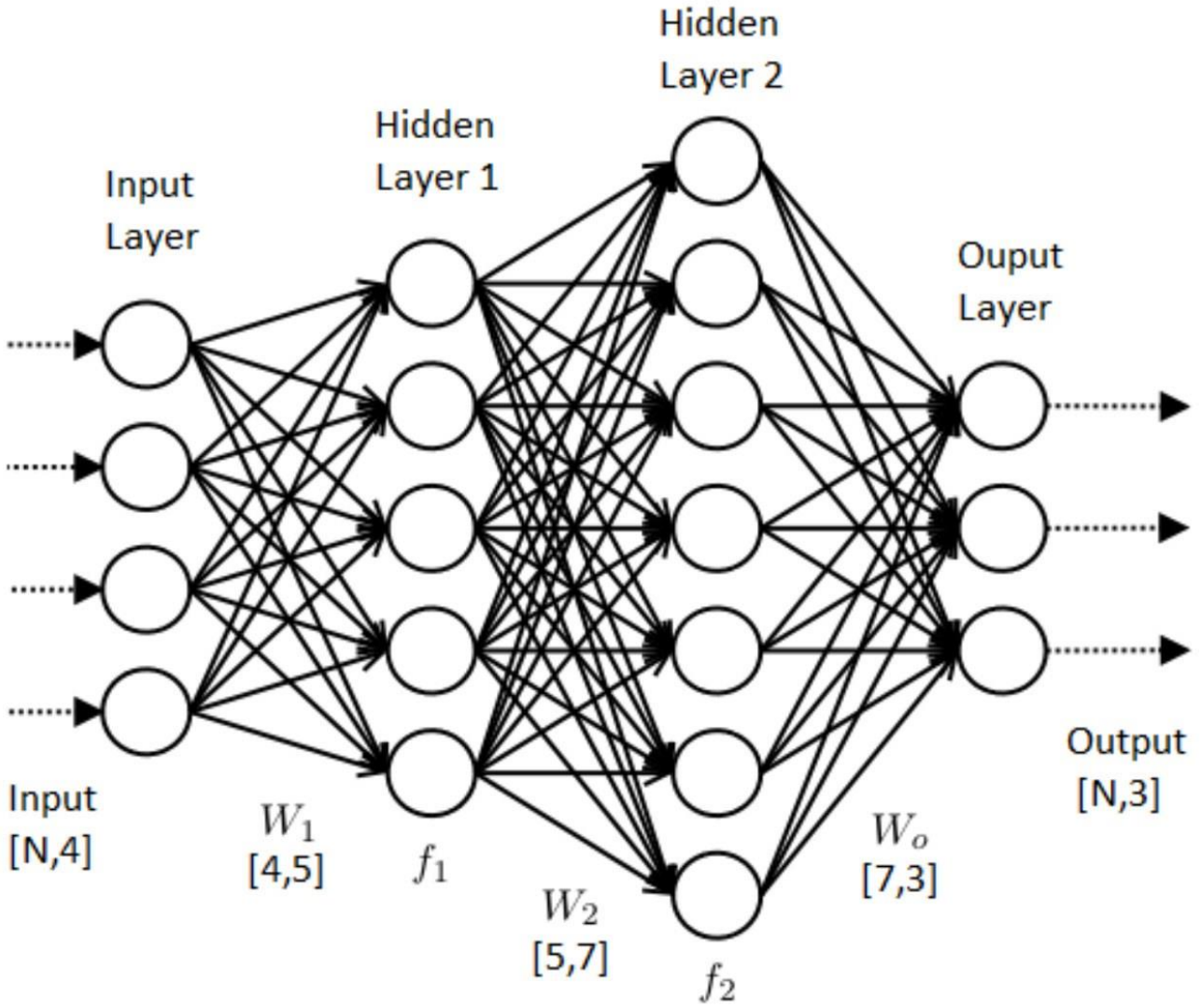




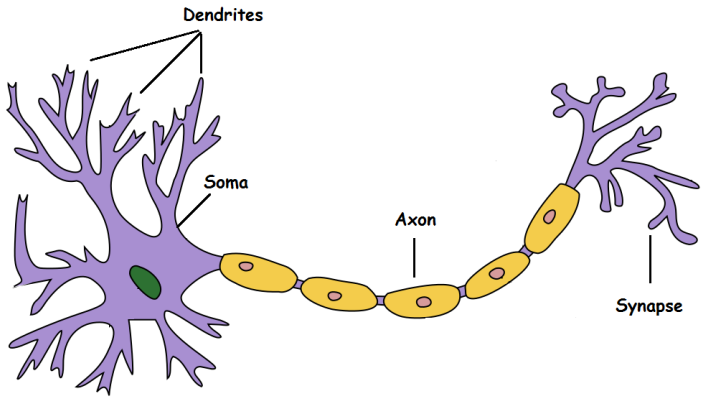


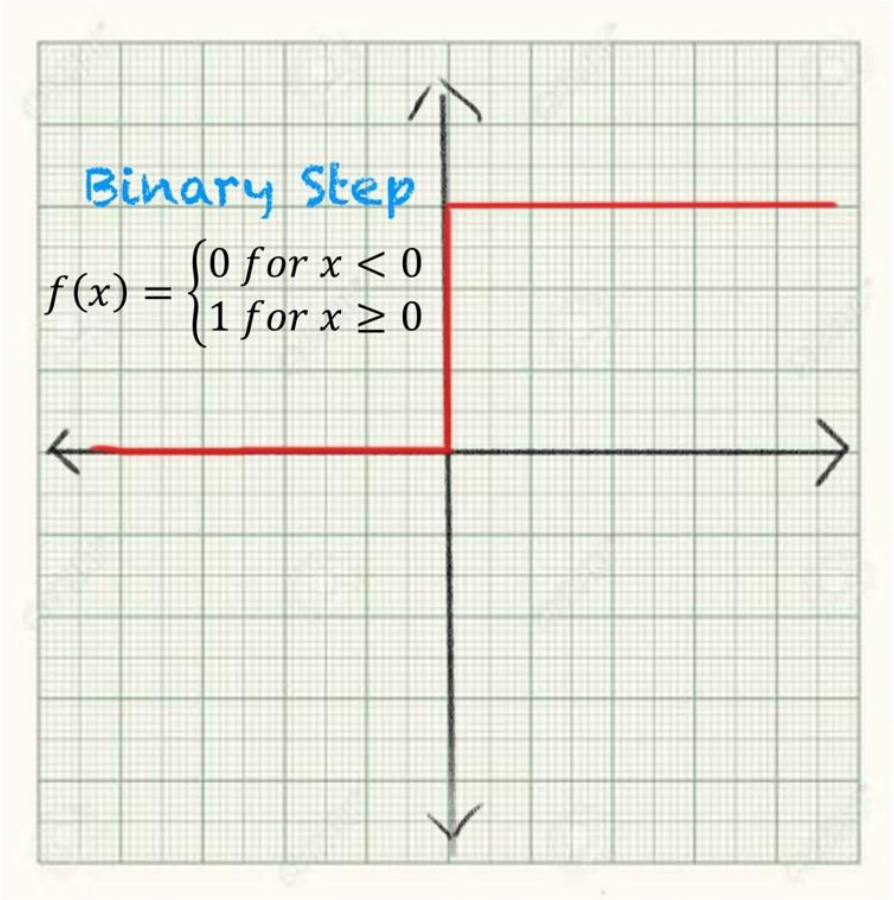
Neural networks

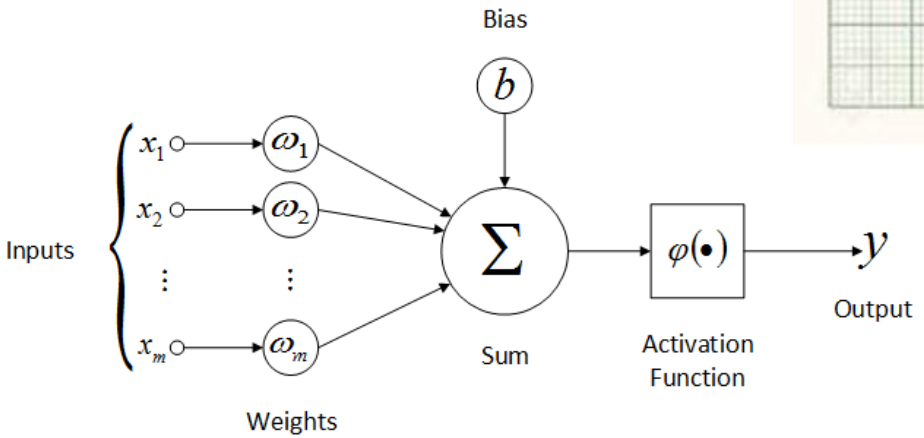




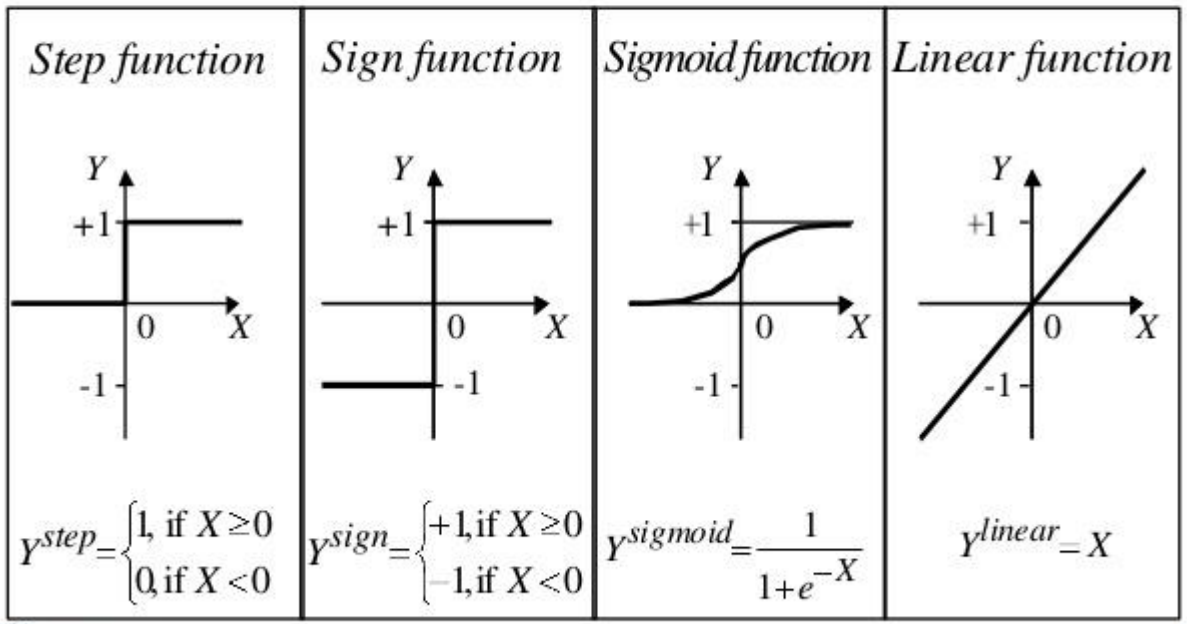
Activation functions: First idea





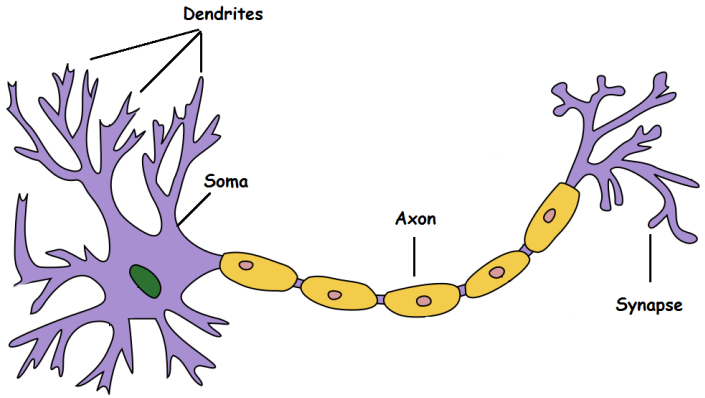


Some Activation functions of a neuron



Activation Functions	
Name	**Formula**
Identity	$A(x) = x$
Sigmoid	$A(x) = \frac{1}{1+e^{-x}}$
Tanh	$A(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$
Step	$A(x) = \begin{matrix} -1 & if\ x < 0 \\ 1 & if\ x \geq 0 \end{matrix}$

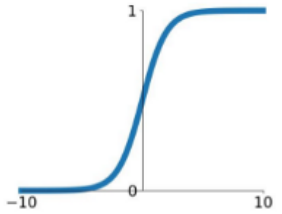
Different activation functions



Activation Functions

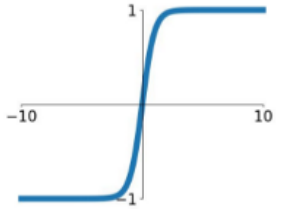
Sigmoid

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



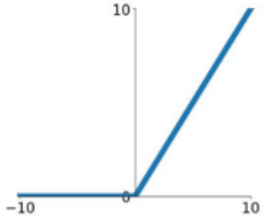
tanh

$$\tanh(x)$$



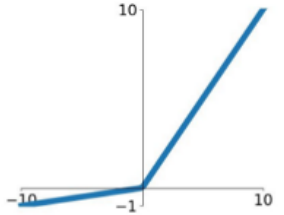
ReLU

$$\max(0, x)$$



Leaky ReLU

$$\max(0.1x, x)$$

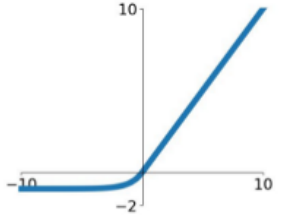


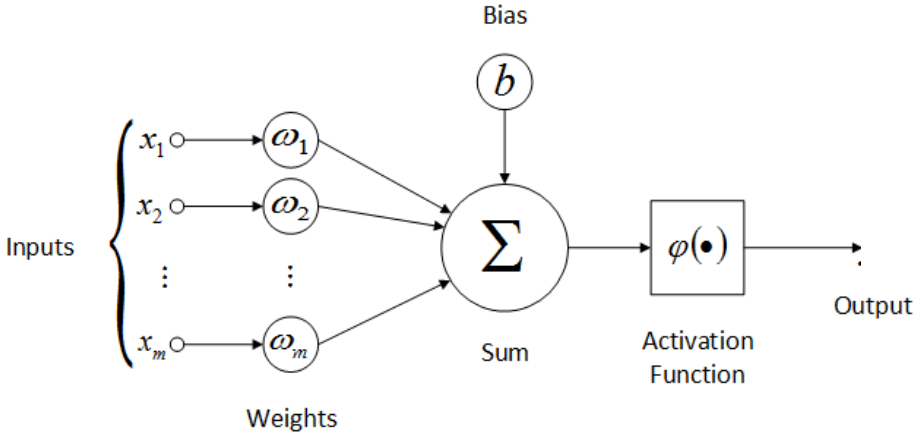
Maxout

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

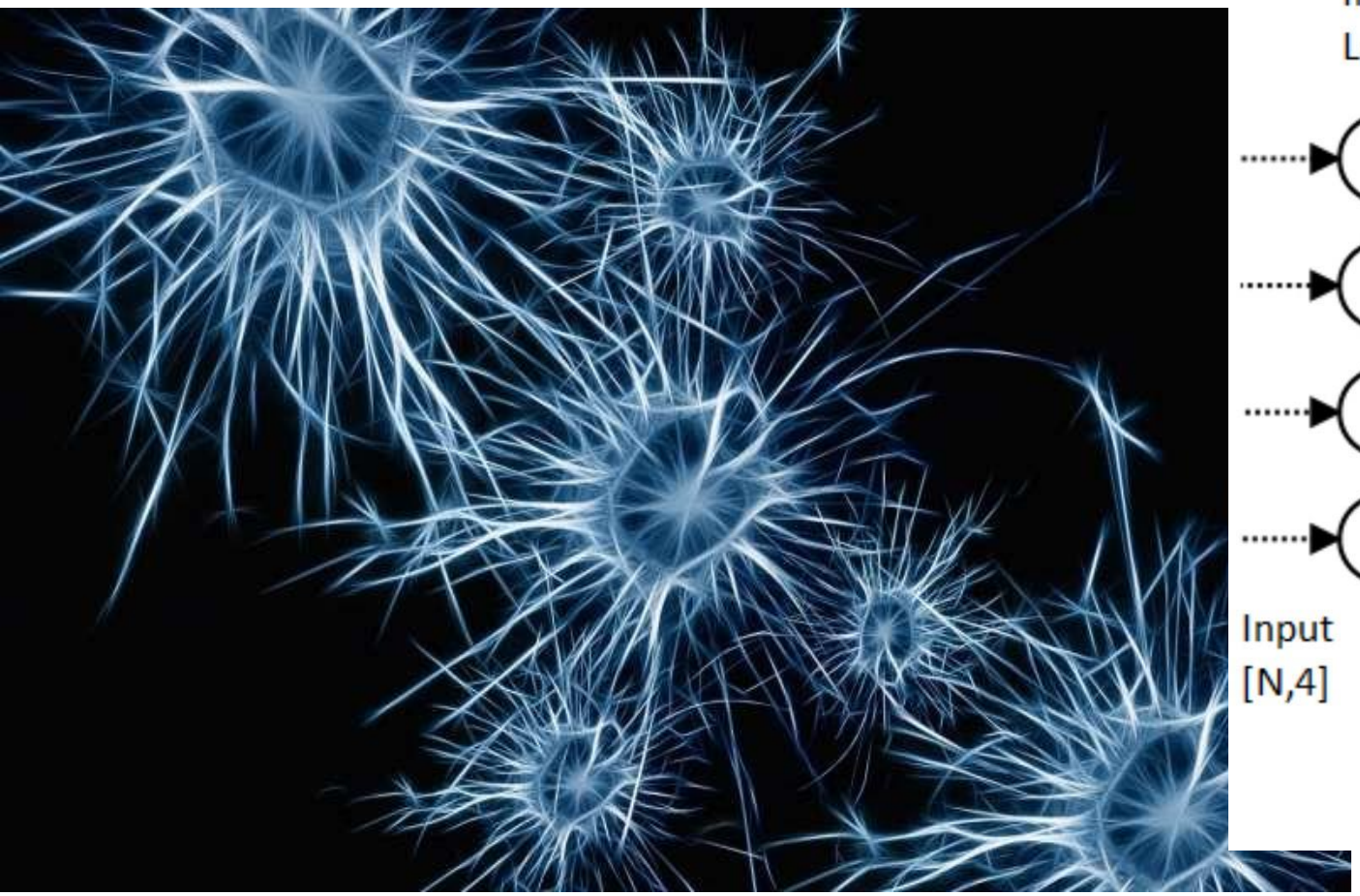
ELU

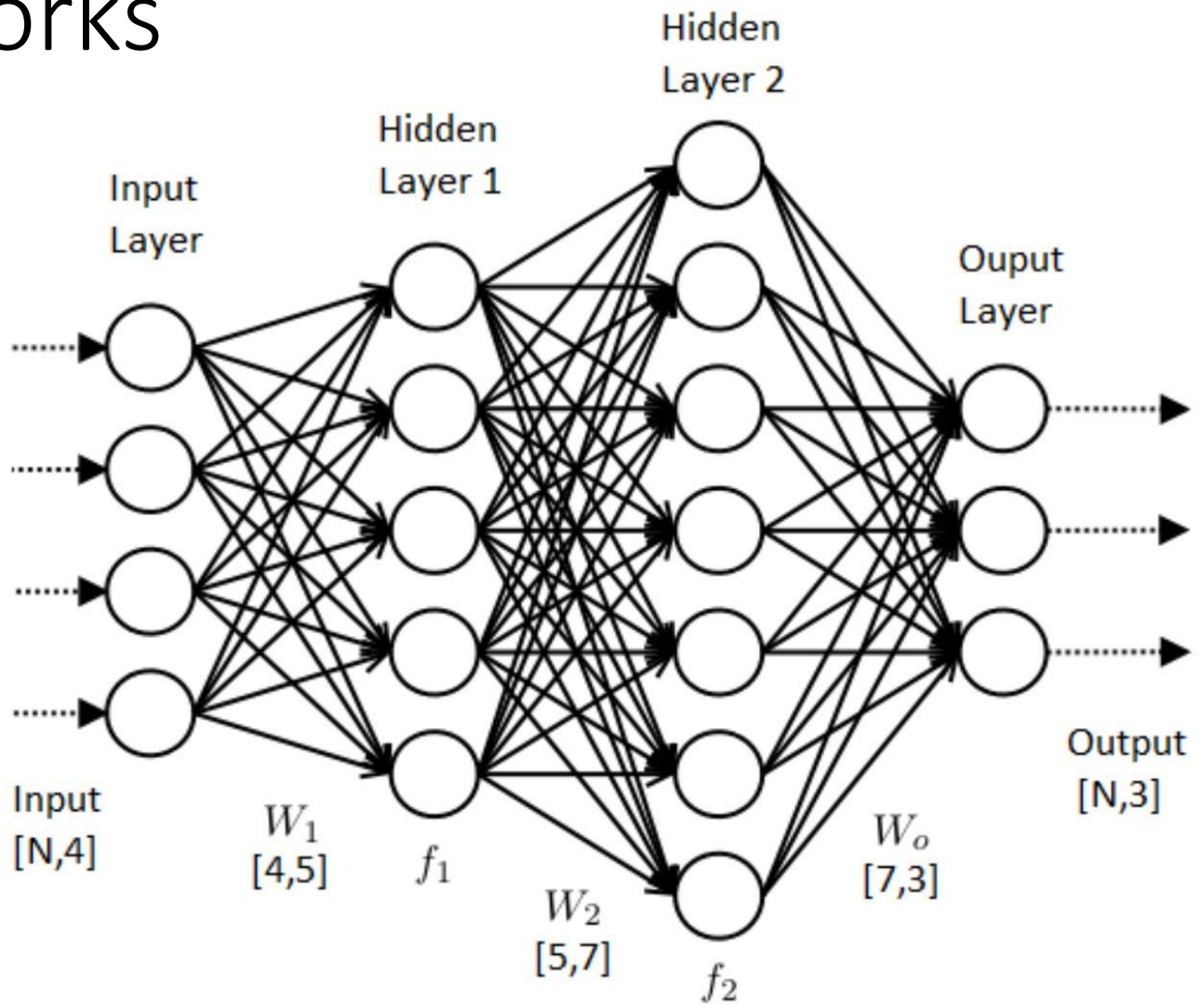
$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$

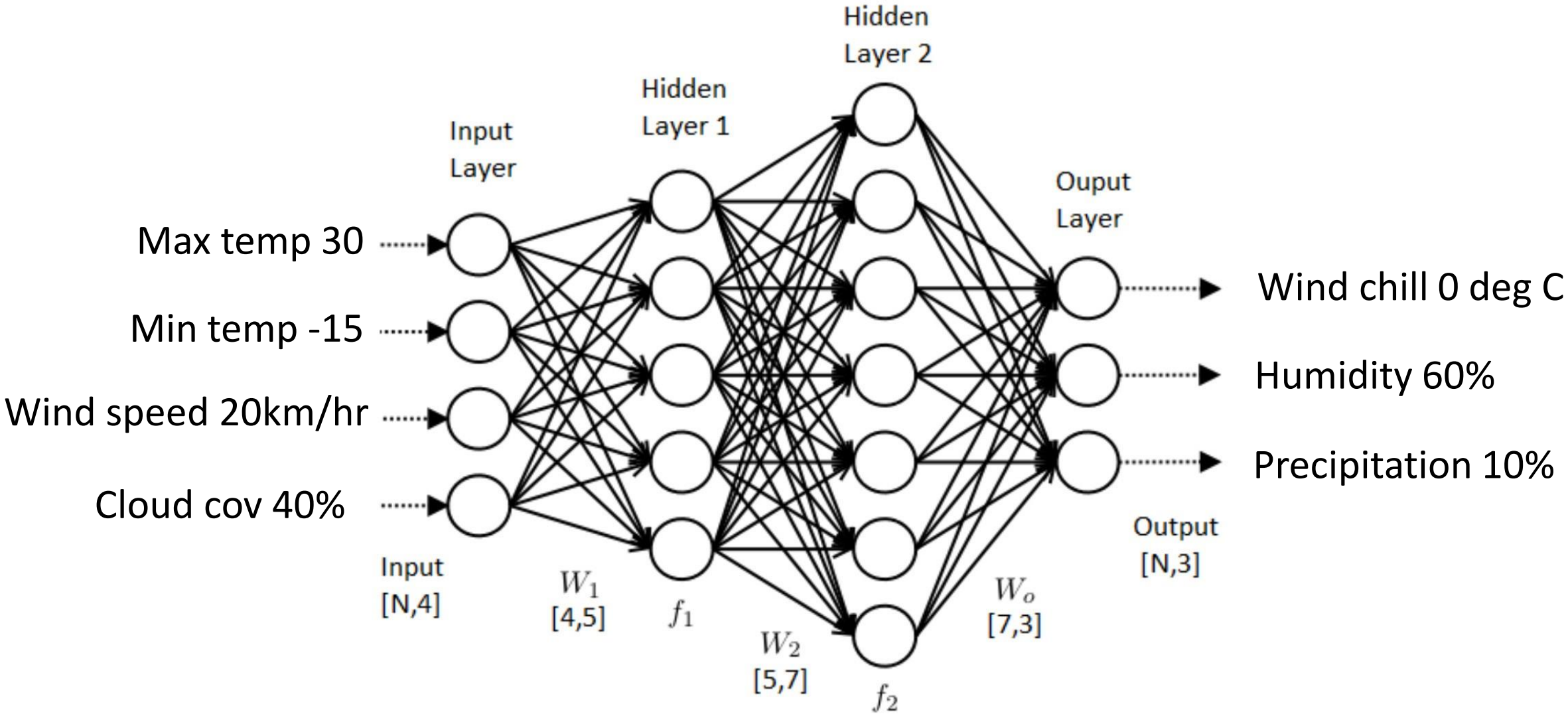




Revisiting Neural networks



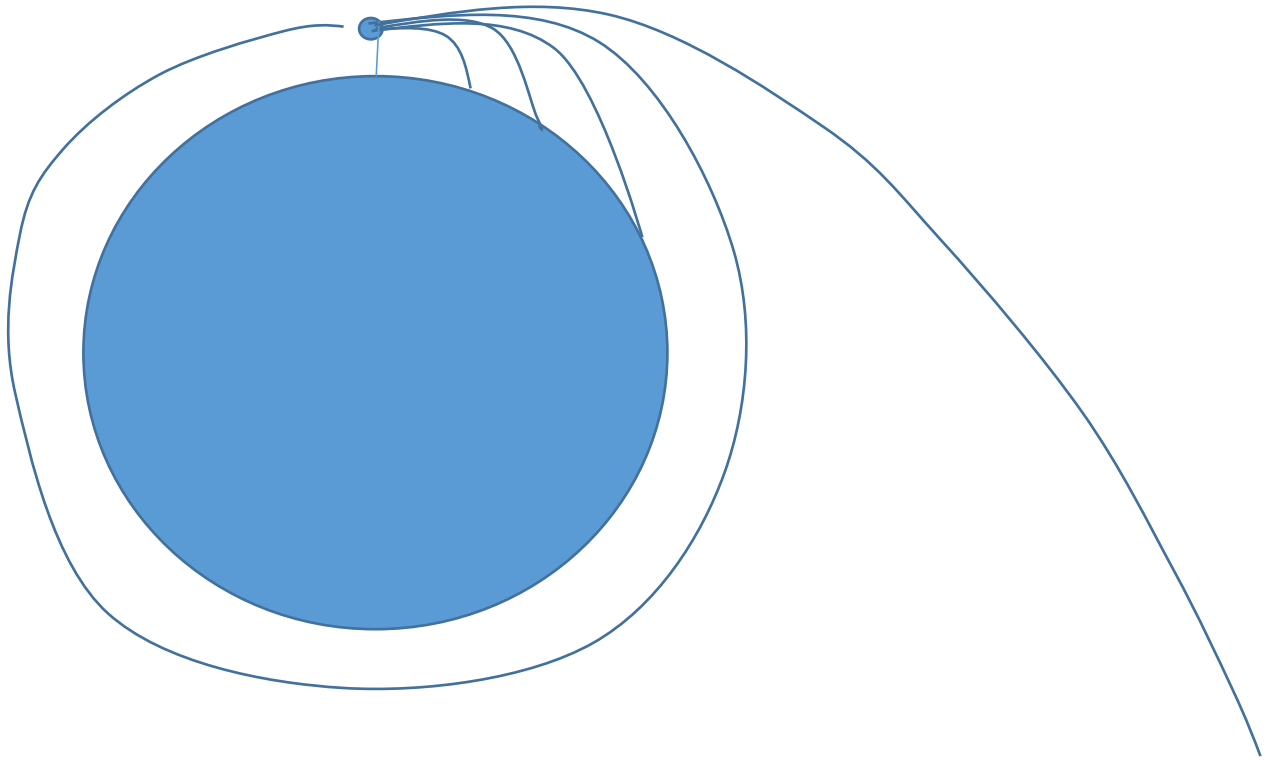




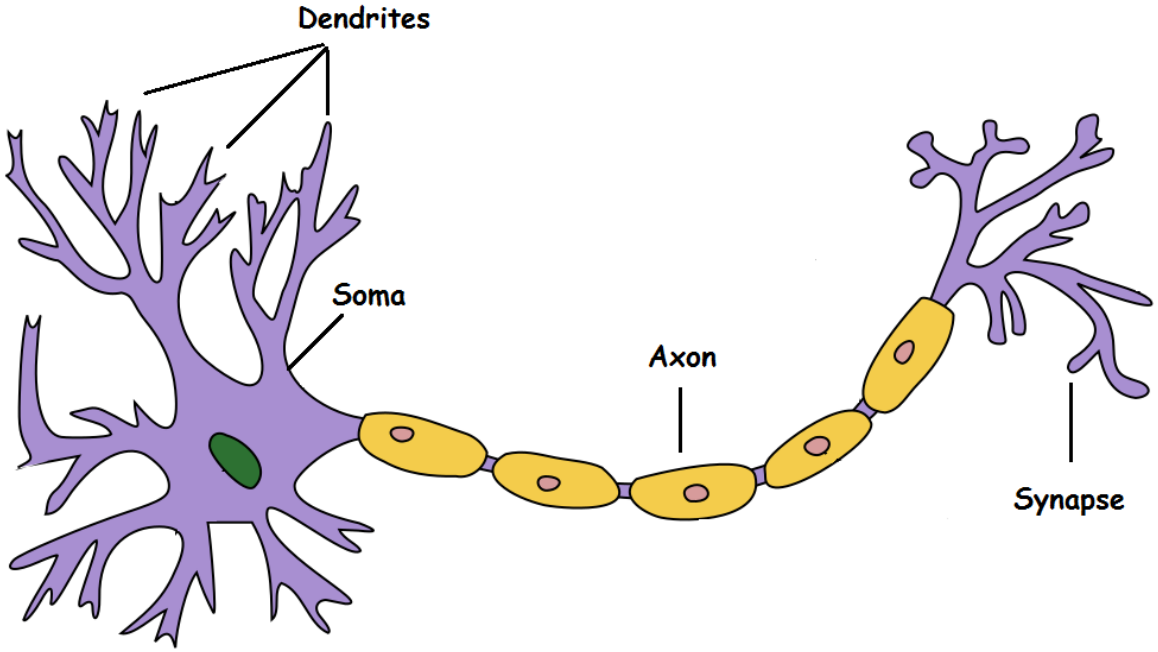
main() sin(x) f(x)

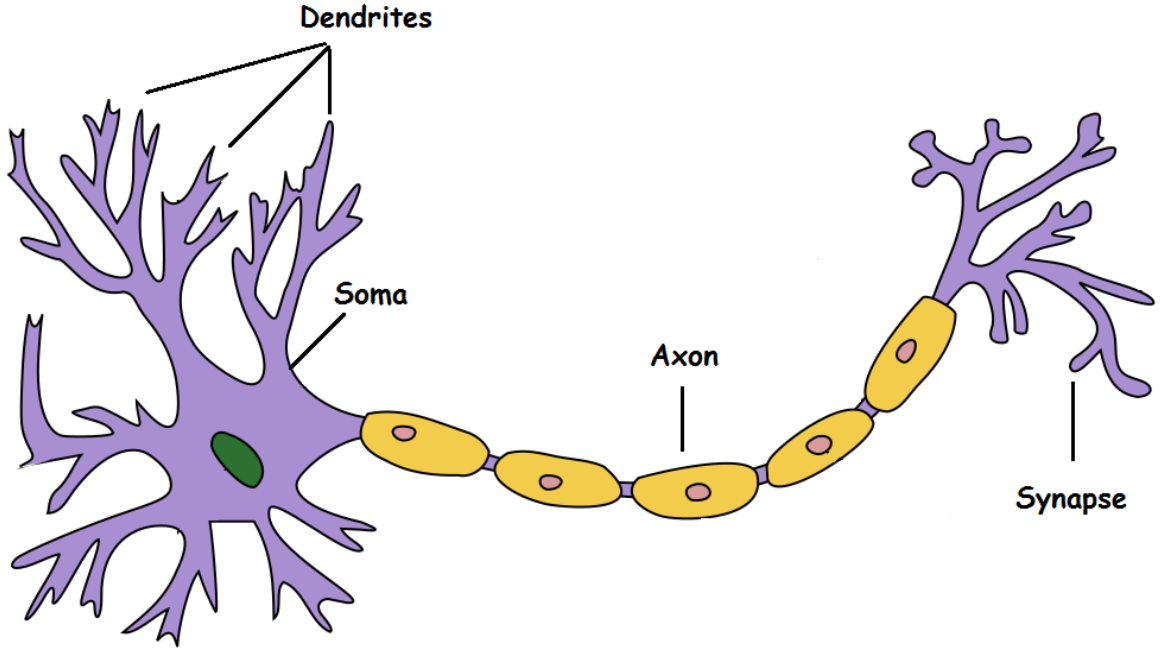
marks(h, w, hf,wf)

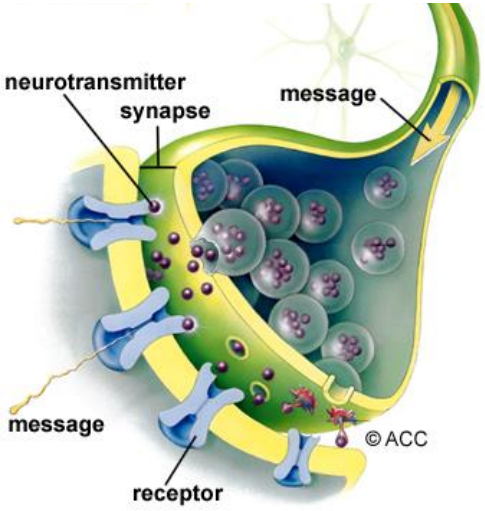
$z=f(x,w) \approx f^*(x) \approx d$

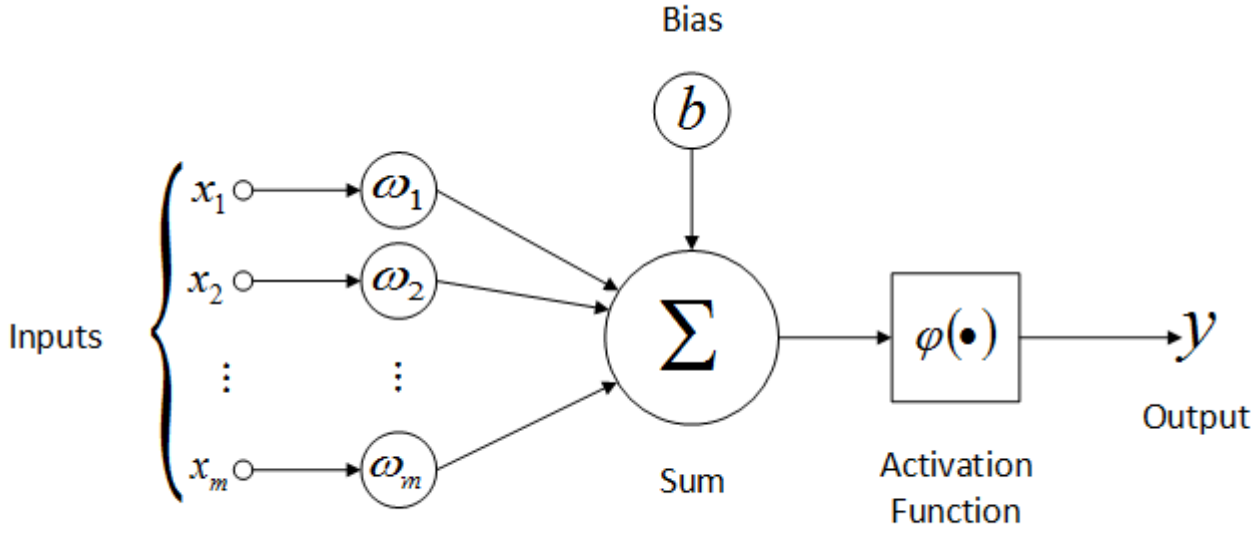


Mathematical model of a neuron

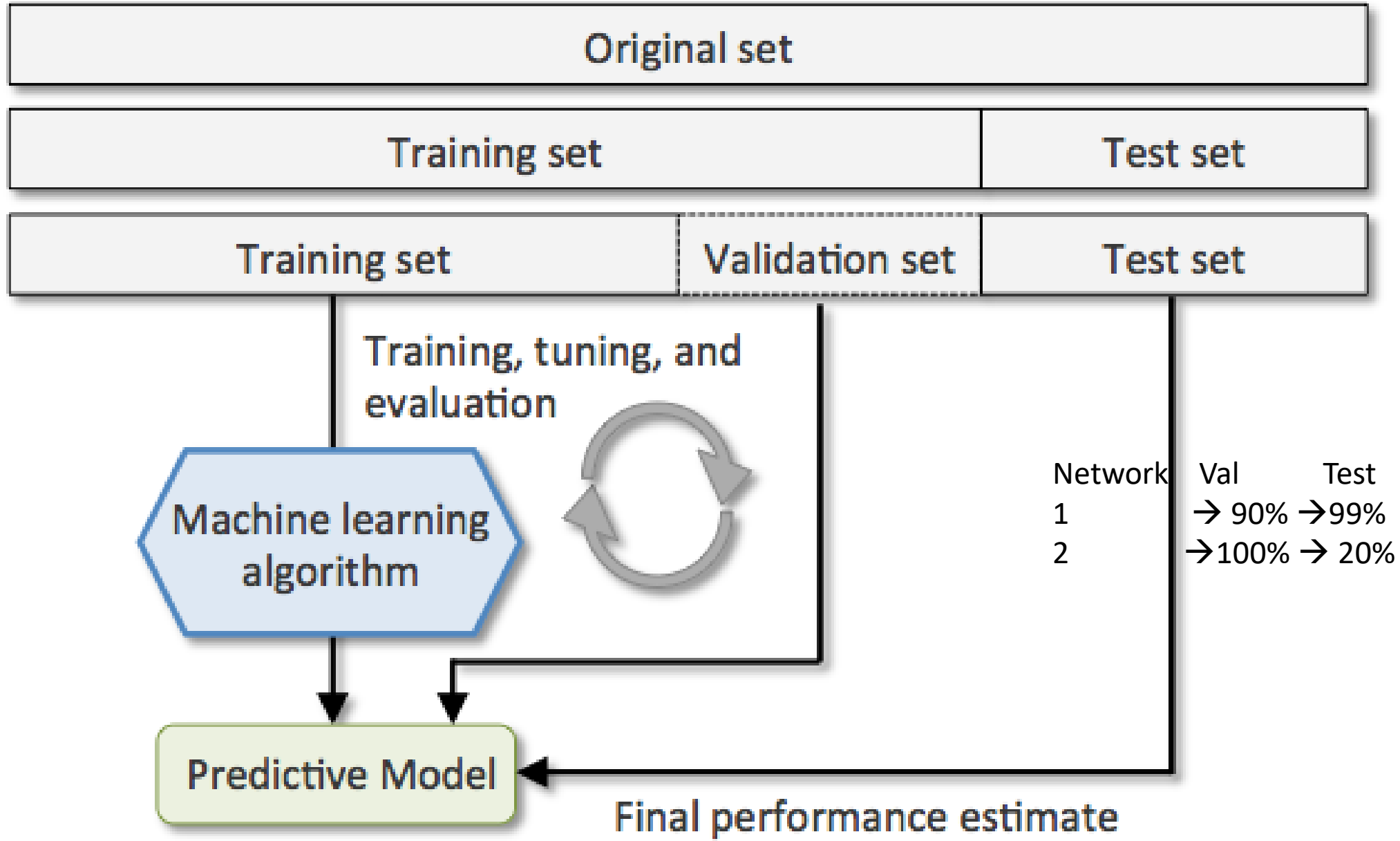








Massaging the data



Network	Val	Test
1	→ 90%	→99%
2	→100%	→ 20%

Key terminologies

Epoch

An Epoch represents one iteration over the entire dataset.

Batch

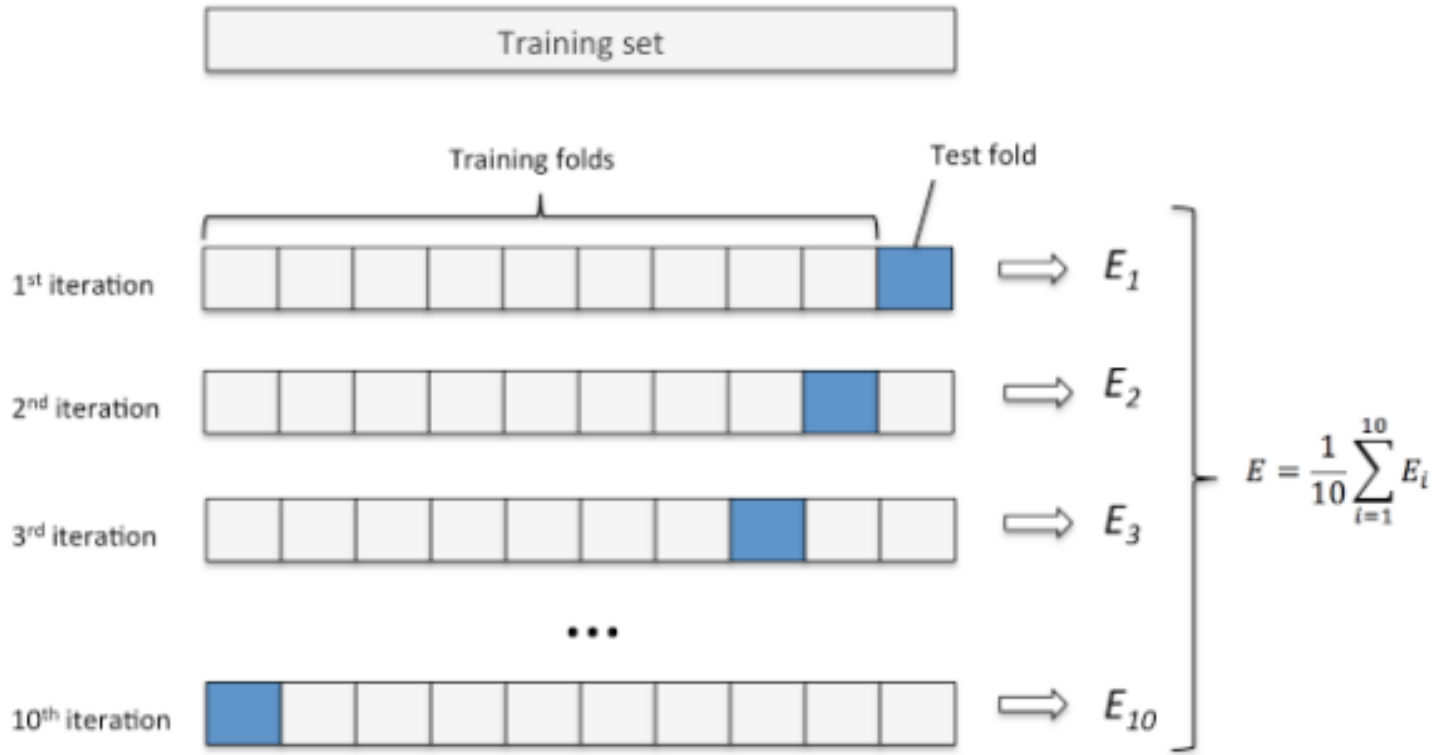
We cannot pass the entire dataset into the neural network at once. So, we divide the dataset into number of batches.

Iteration

If we have 10,000 images as data and a batch size of 200, then an epoch should contain 10,000/200 = 50 iterations.

K-Fold Cross Validation

In K-fold cross-validation (CV), we randomly split the training dataset into K folds without replacement, where $K-1$ folds are used for the model training and the remaining 1 fold is for testing. This procedure is repeated K times so that we obtain K models and K performance estimates. Then we take their **average** as the final performance estimate. The following figure illustrate the 10-fold CV:



We can apply K-fold CV to either the hyperparameter tuning, performance reporting, or both. The advantage of this approach is that the performance is less sensitive to unfortunate splits of data. In addition, it utilize data better since each example can be used for both training and validation/testing.

Let's use K-Fold CV to select the hyperparamter `n_neighbors` of the `KNeighborsClassifier`:

In matrix notation, 100 images at a time