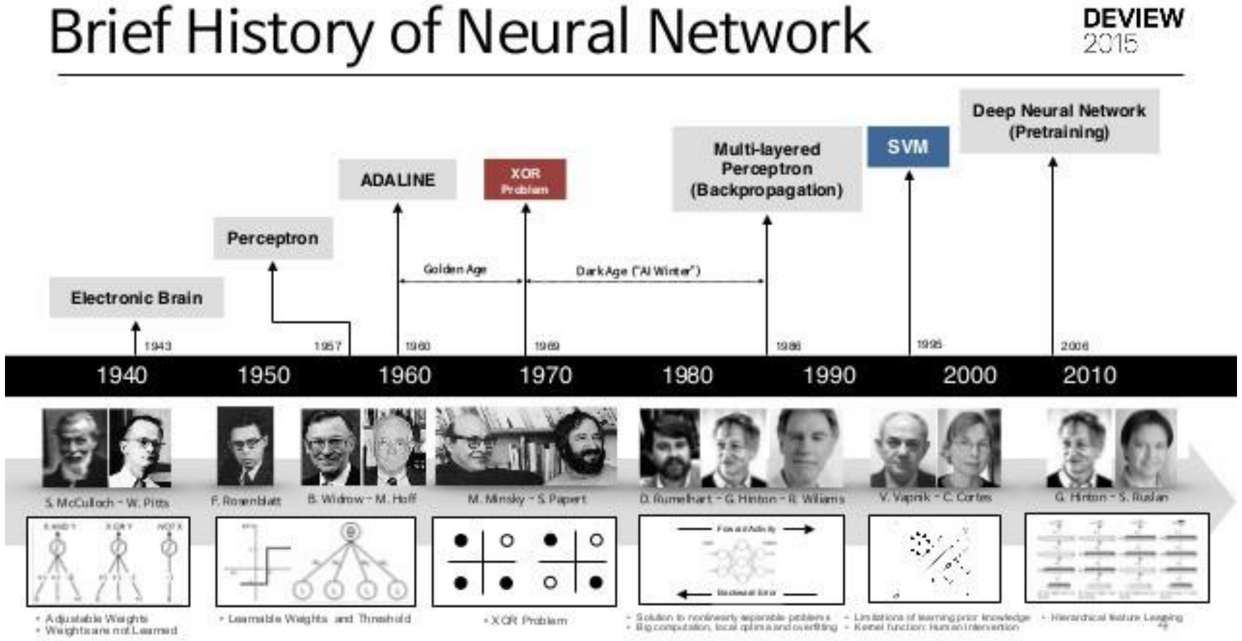
- Artificial Neural Networks, especially Convolutional Neural Networks currently provide the best solutions to many problems in image processing, computer vision and natural language processing.

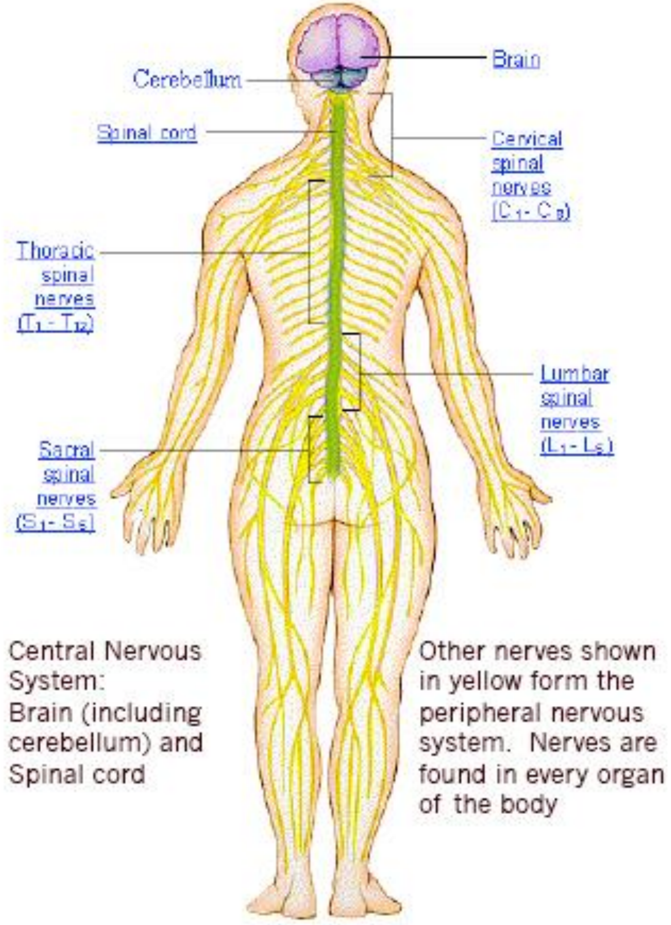
- Artificial neural networks, usually simply called neural networks, are computing systems inspired by the biological neural networks that constitute animal brains.

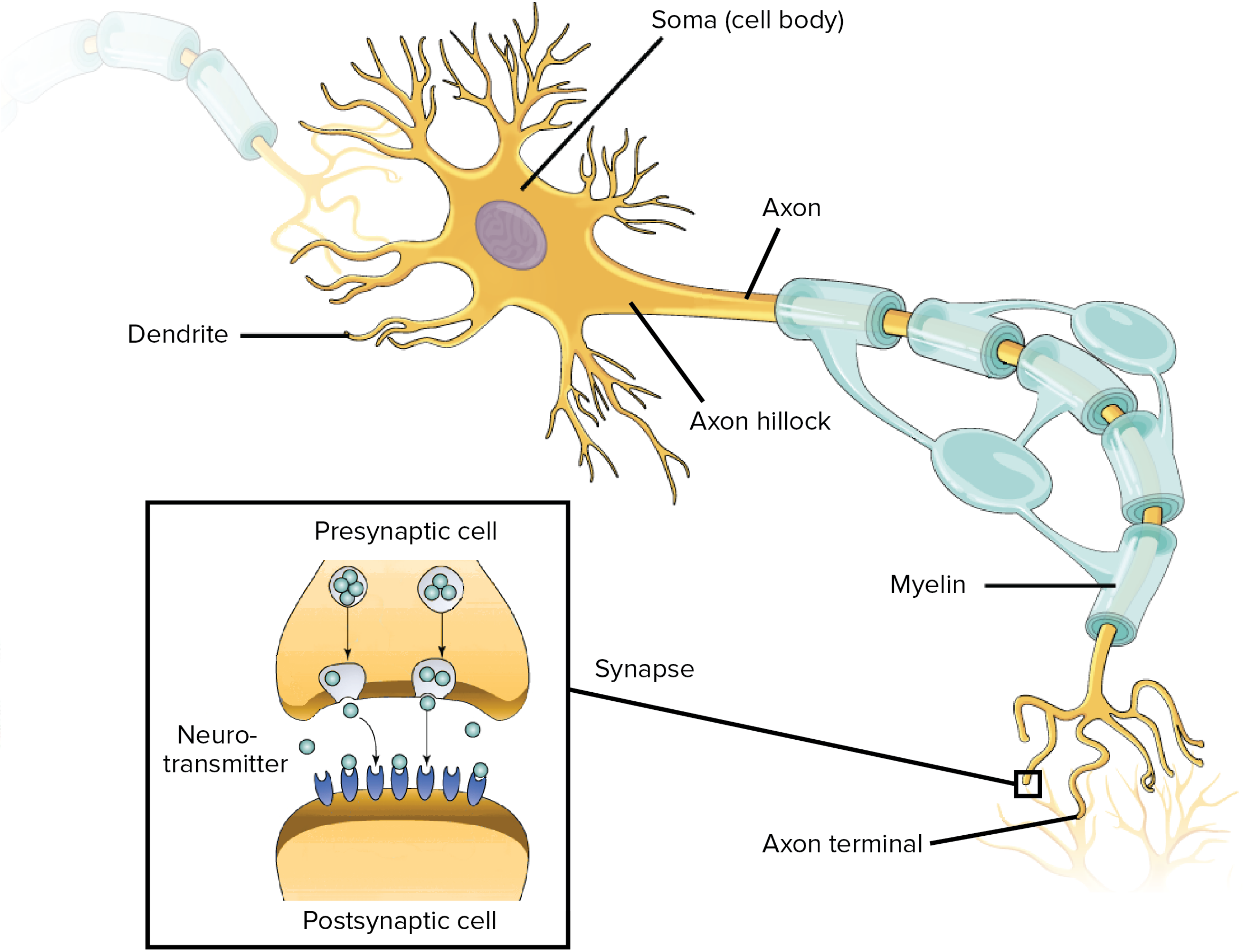
- Universal Approximation Theorem states that that with enough hidden neurons, there exists some set of connection weights that can approximate any function.

- The history of artificial neural networks (ANN) began with Warren McCulloch and Walter Pitts (1943) who created a computational model for neural networks based on algorithms called threshold logic.

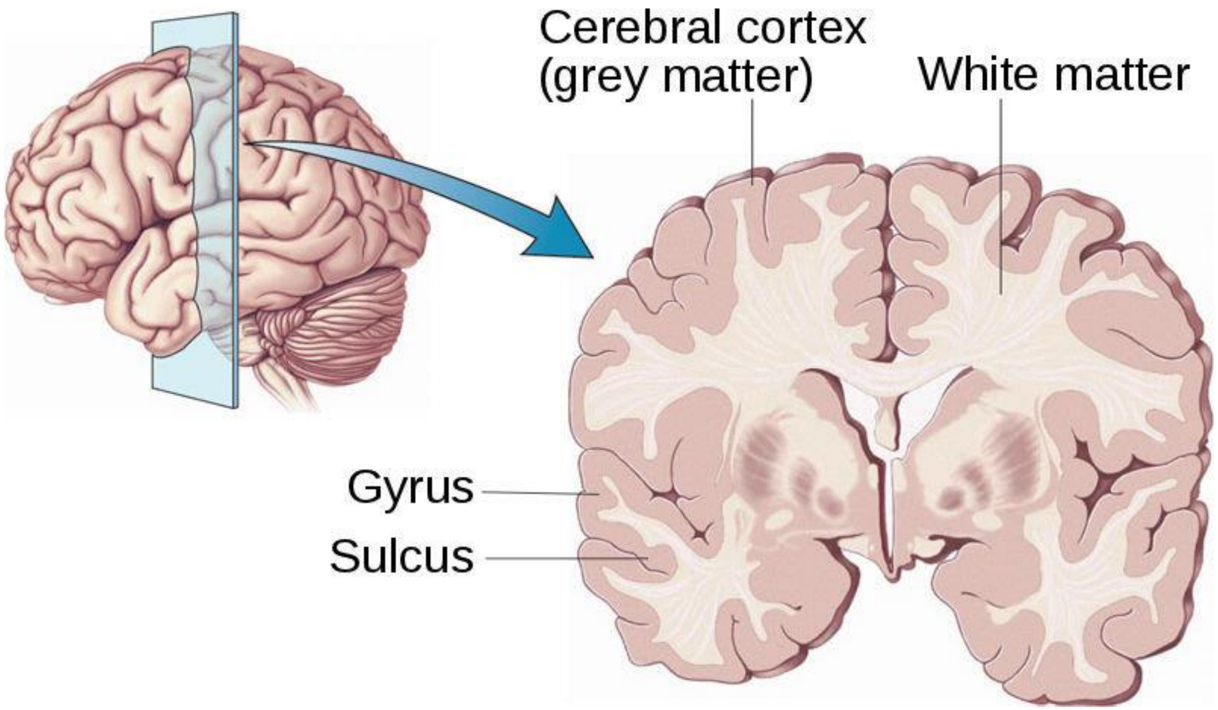


https://medium.com/coinmonks/neural-networks-bb11fb9a8266



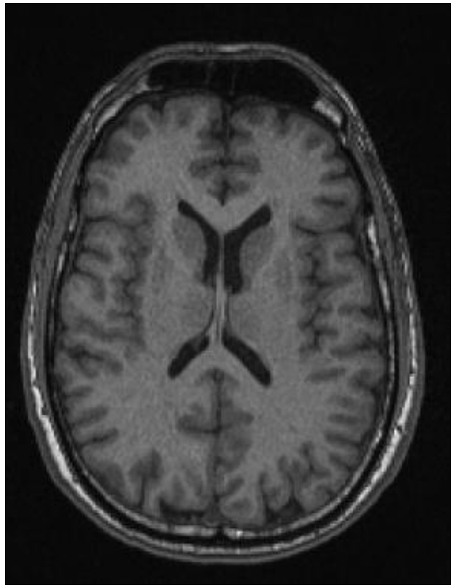


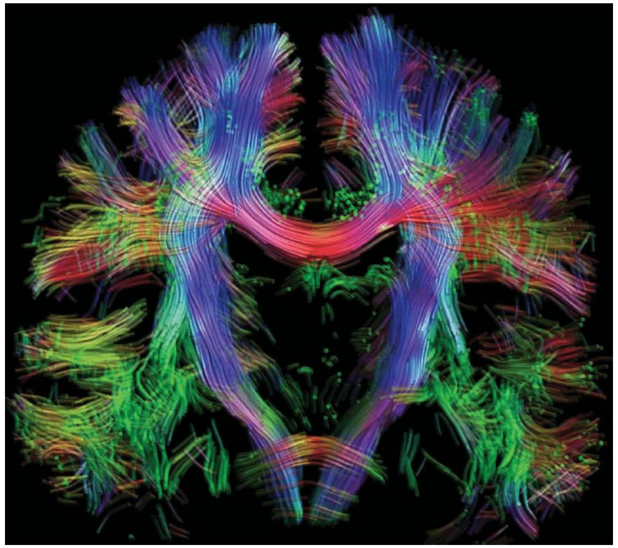
https://www.ninds.nih.gov/Disorders/Patient-Caregiver-Education/Life-and-Death-Neuron

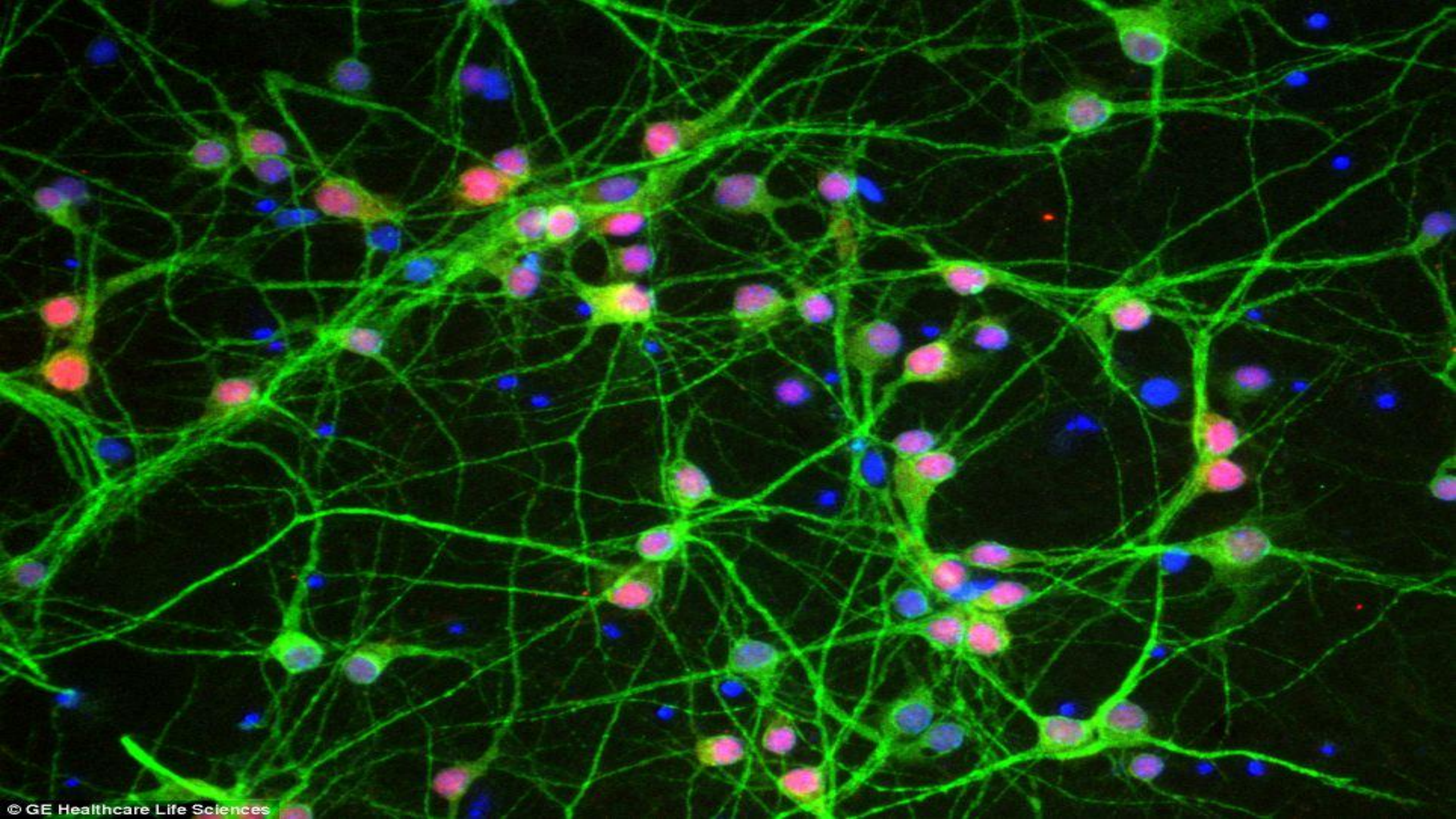


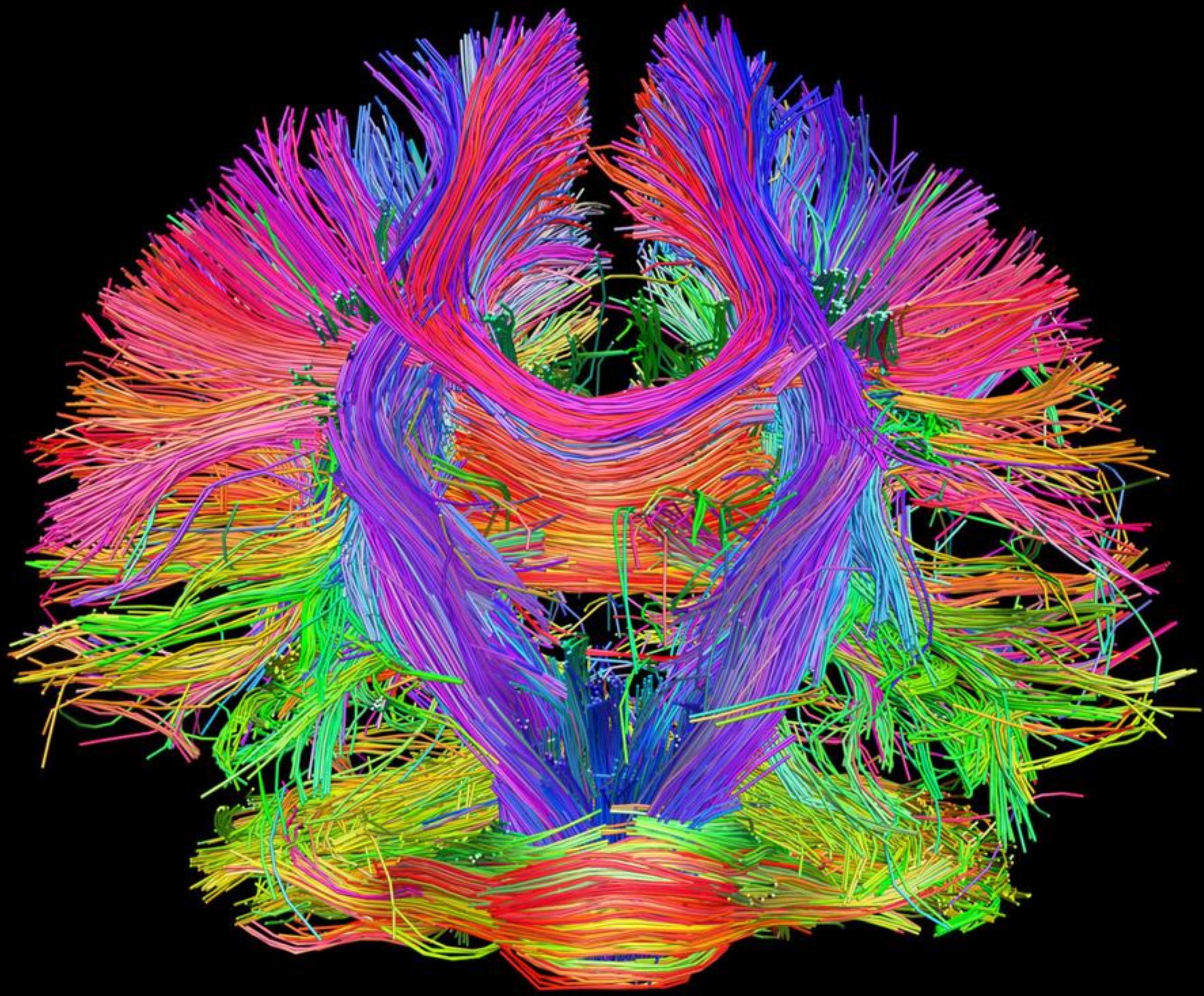
smooth cerebral cortex

highly folded cerebral cortex

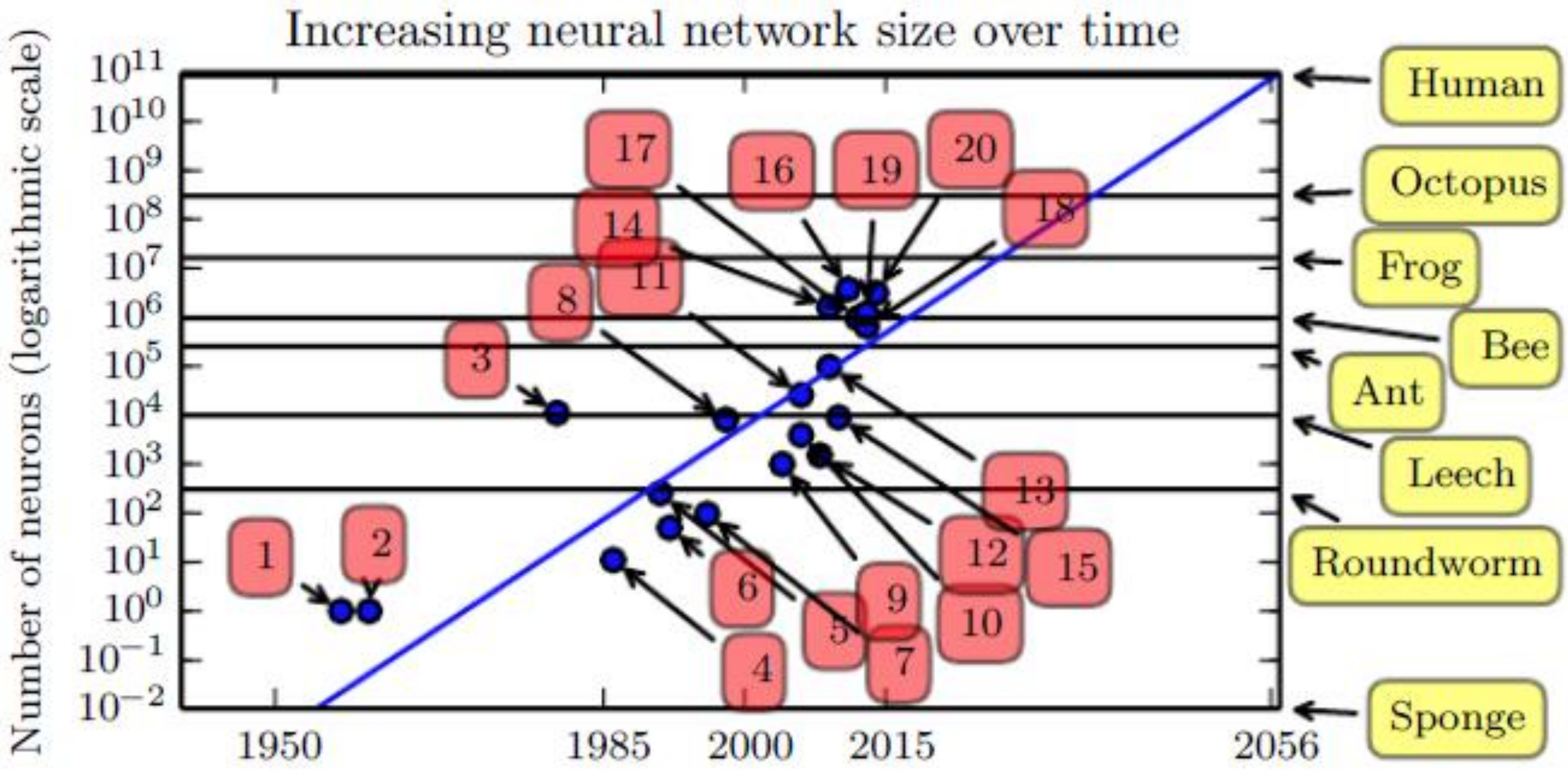


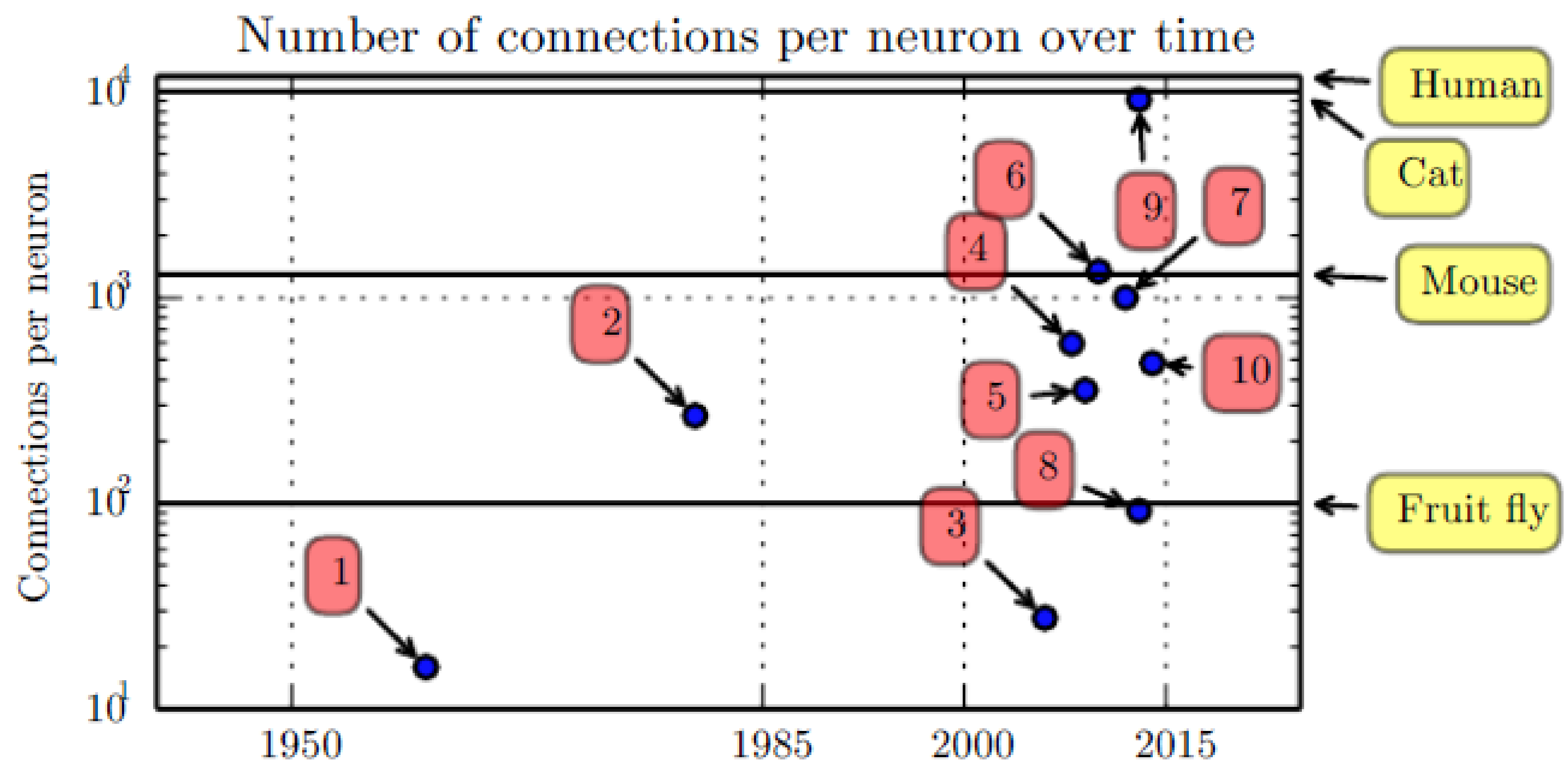






1. Perceptron (Rosenblatt, 1958, 1962)
2. Adaptive linear element (Widrow and Hoff, 1960)
3. Neocognitron (Fukushima, 1980)
4. Early back-propagation network (Rumelhart *et al.*, 1986b)
5. Recurrent neural network for speech recognitio
6. Multilayer perceptron for speech recognition (
7. Mean field sigmoid belief network (Saul *et al.*,
8. LeNet-5 (LeCun *et al.*, 1998b)
9. Echo state network (Jaeger and Haas, 2004)
10. Deep belief network (Hinton *et al.*, 2006)
11. GPU-accelerated convolutional network (Chella
12. Deep Boltzmann machine (Salakhutdinov and
13. GPU-accelerated deep belief network (Raina et
14. Unsupervised convolutional network (Jarrett et
15. GPU-accelerated multilayer perceptron (Ciresa
16. OMP-1 network (Coates and Ng, 2011)
17. Distributed autoencoder (Le *et al.*, 2012)
18. Multi-GPU convolutional network (Krizhevsky
19. COTS HPC unsupervised convolutional networ
20. GoogLeNet (Szegedy *et al.*, 2014a)





1. Adaptive linear element (Wi
2. Neocognitron (Fukushima, 1
3. GPU-accelerated convolutio
4. Deep Boltzmann machine (Salakhutdinov and Hinton, 2009a)
5. Unsupervised convolutional network (Jarrett *et al.*, 2009)
6. GPU-accelerated multilayer perceptron (Ciresan *et al.*, 2010)
7. Distributed autoencoder (Le *et al.*, 2012)
8. Multi-GPU convolutional network (Krizhevsky *et al.*, 2012)
9. COTS HPC unsupervised convolutional network (Coates *et al.*, 2013)
10. GoogLeNet (Szegedy *et al.*, 2014a)

BULLETIN OF
MATHEMATICAL BIOPHYSICS
VOLUME 5, 1943

A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY

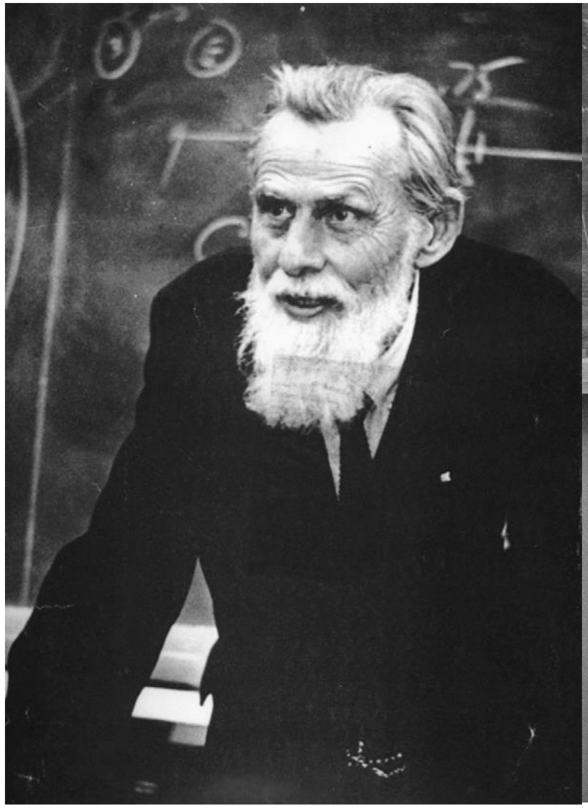
WARREN S. MCCULLOCH AND WALTER PITTS

FROM THE UNIVERSITY OF ILLINOIS, COLLEGE OF MEDICINE,
DEPARTMENT OF PSYCHIATRY AT THE ILLINOIS NEUROPSYCHIATRIC INSTITUTE,
AND THE UNIVERSITY OF CHICAGO

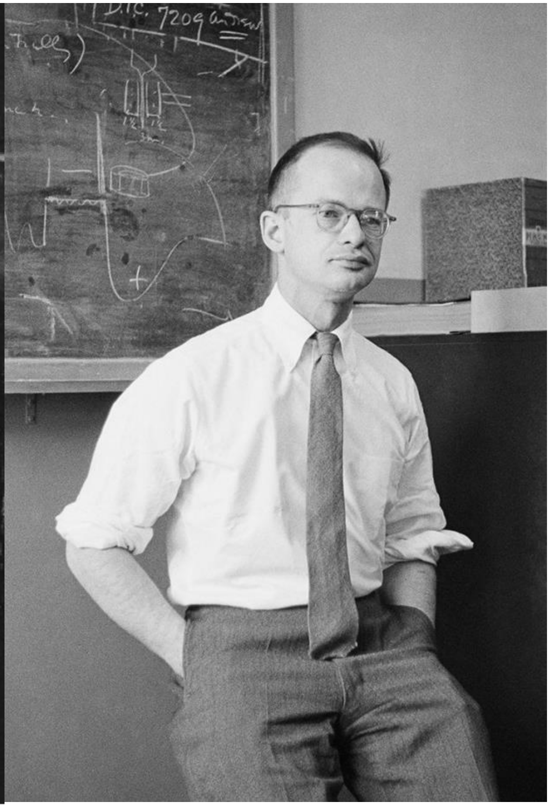
Because of the "all-or-none" character of nervous activity, neural events and the relations among them can be treated by means of propositional logic. It is found that the behavior of every net can be described in these terms, with the addition of more complicated logical means for nets containing circles; and that for any logical expression satisfying certain conditions, one can find a net behaving in the fashion it describes. It is shown that many particular choices among possible neurophysiological assumptions are equivalent, in the sense that for every net behaving under one assumption, there exists another net which behaves under the other and gives the same results, although perhaps not in the same time. Various applications of the calculus are discussed.

I. *Introduction*

Theoretical neurophysiology rests on certain cardinal assumptions. The nervous system is a net of neurons, each having a soma and an axon. Their adjunctions, or synapses, are always between the axon of one neuron and the soma of another. At any instant a neuron has some threshold, which excitation must exceed to initiate an impulse. This, except for the fact and the time of its occurrence, is determined by the neuron, not by the excitation. From the point of excitation the impulse is propagated to all parts of the neuron. The velocity along the axon varies directly with its diameter, from less than one meter per second in thin axons, which are usually short, to more than 150 meters per second in thick axons, which are usually long. The time for axonal conduction is consequently of little importance in determining the time of arrival of impulses at points unequally remote from the same source. Excitation across synapses occurs predominantly from axonal terminations to somata. It is still a moot point whether this depends upon irreciprocity of individual synapses or merely upon prevalent anatomical configurations. To suppose the latter requires no hypothesis *ad hoc* and explains known exceptions, but any assumption as to cause is compatible with the calculus to come. No case is known in which excitation through a single synapse has elicited a nervous impulse in any neuron, whereas any



Warren McCulloch

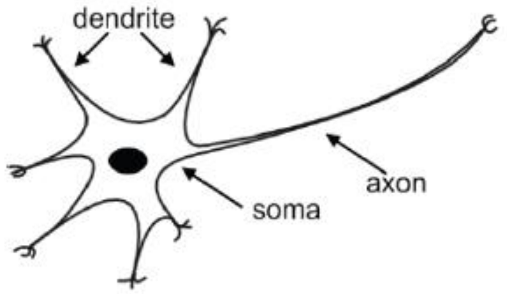


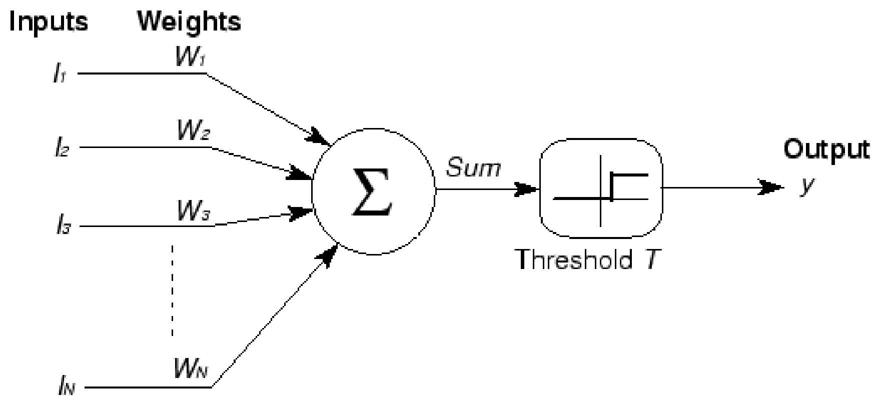
Walter Pitts

1943, "A Logical Calculus of the Ideas Immanent in Nervous Activity". With Walter Pitts. In: *Bulletin of Mathematical Biophysics* Vol 5, pp 115–133.

The McCulloch-Pitts Model of Neuron (1942 model)

- The early model of an artificial neuron is introduced by Warren McCulloch and Walter Pitts in 1943.
- This is the first mathematical model of a neural network.



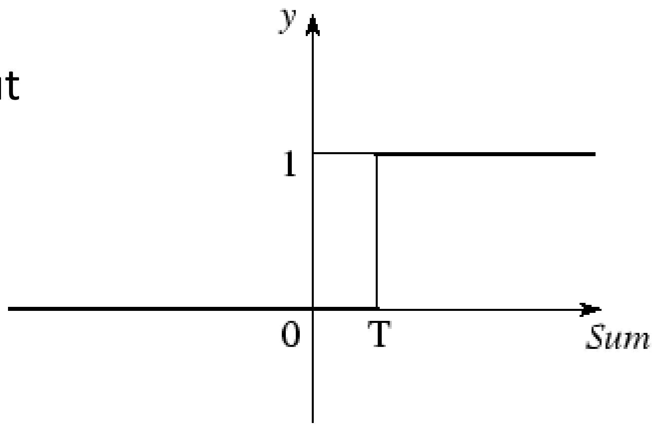


$$Sum = \sum_{i=1}^{N} I_i W_i$$

$$Output = \begin{cases} 0 & \text{if Sum} \leq T \\ 1 & \text{if Sum} > T \end{cases}$$

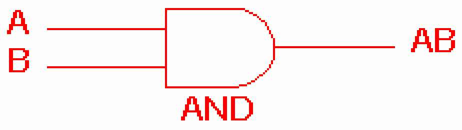
$$y = f(Sum)$$

- The main feature of their neuron model is that a weighted sum of input signals is compared to a threshold to determine the neuron output.
- When the sum is greater than or equal to the threshold, the output is 1.
- When the sum is less than the threshold, the output is 0.

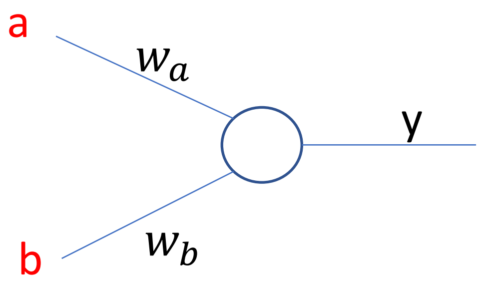


Step activation Function

- They demonstrated that networks of these neurons could, in principle, compute any arithmetic or logical function.
- Unlike biological networks, the parameters of their networks had to be designed, as no training method was available.
- However, the perceived connection between biology and digital computers generated a great deal of interest.



2 Input AND gate		
A	B	A.B
0	0	0
0	1	0
1	0	0
1	1	1



$$y = \begin{cases} 0 & \text{if Sum} \leq T \\ 1 & \text{if Sum} > T \end{cases}$$

$$0 \times w_a + 0 \times w_b \leq T, \qquad T \geq 0$$

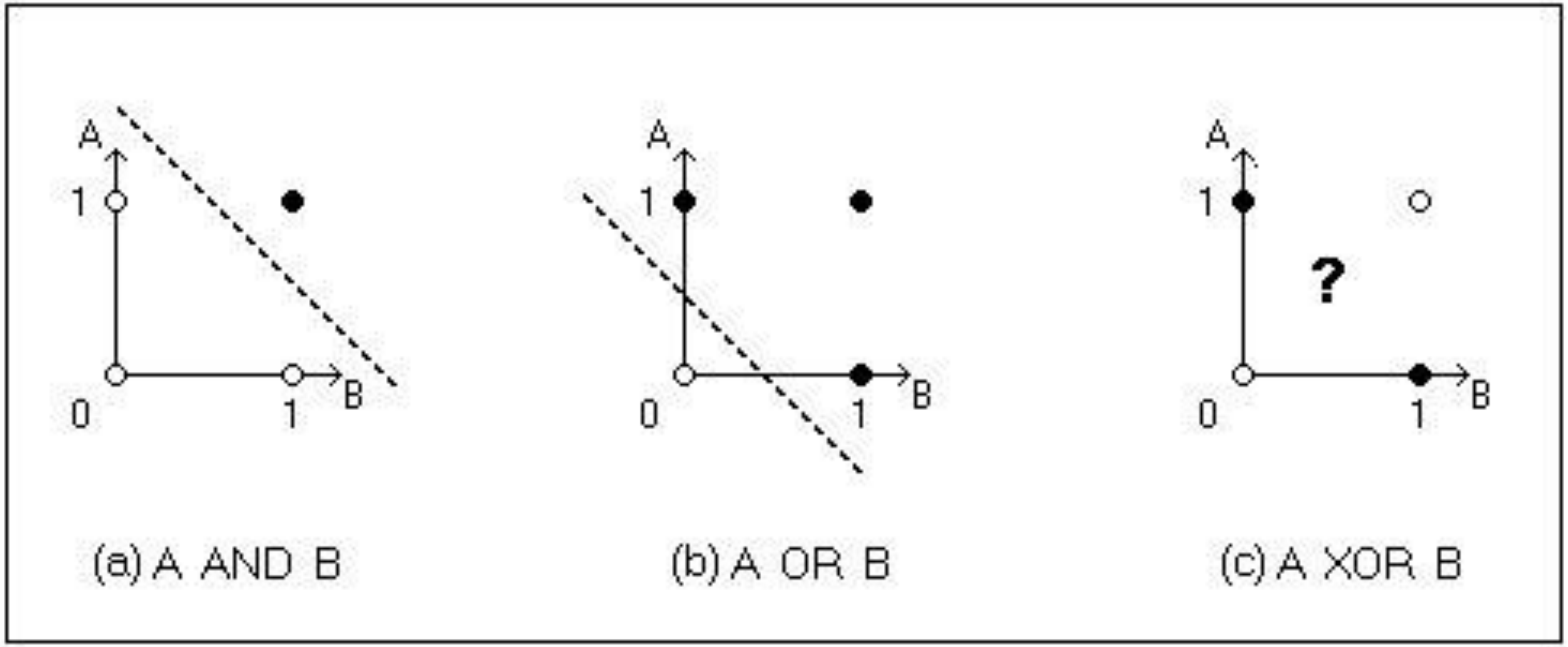
$$1 \times w_a + 0 \times w_b \leq T, \qquad w_a \leq T$$

$$0 \times w_a + 1 \times w_b \leq T, \qquad w_b \leq T$$

$$1 \times w_a + 1 \times w_b > T, \qquad w_a + w_b > T$$

$$T = 1.5, w_a = 1, w_b = 1$$

Geometric Interpretation Of M-P Neuron



Psychological Review
Vol. 65, No. 6, 1958

THE PERCEPTRON: A PROBABILISTIC MODEL FOR INFORMATION STORAGE AND ORGANIZATION IN THE BRAIN[1]

F. ROSENBLATT

Cornell Aeronautical Laboratory

If we are eventually to understand the capability of higher organisms for perceptual recognition, generalization, recall, and thinking, we must first have answers to three fundamental questions:

1. How is information about the physical world sensed, or detected, by the biological system?
2. In what form is information stored, or remembered?
3. How does information contained in storage, or in memory, influence recognition and behavior?

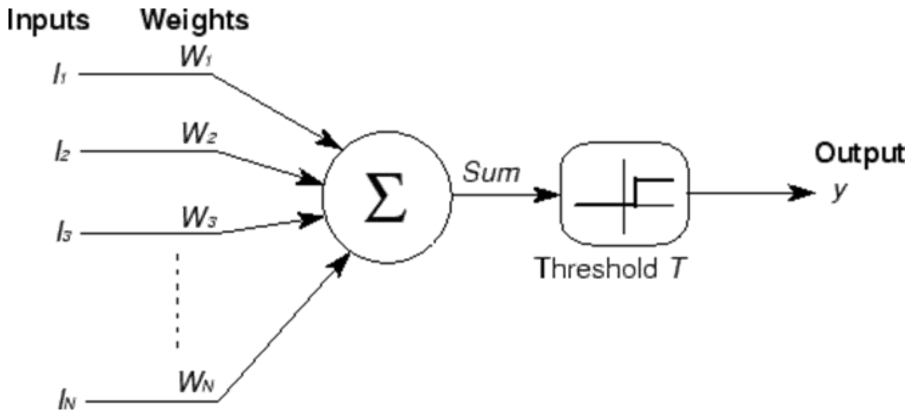
The first of these questions is in the province of sensory physiology, and is the only one for which appreciable understanding has been achieved. This article will be concerned primarily with the second and third questions, which are still subject to a vast amount of speculation, and where the few relevant facts currently supplied by neurophysiology have not yet been integrated into an acceptable theory.

With regard to the second question, two alternative positions have been maintained. The first suggests that storage of sensory information is in the form of coded representations or images, with some sort of one-to-one mapping between the sensory stimulus and the stored pattern. According to this hypothesis, if one understood the code or "wiring diagram" of the nervous system, one should, in principle, be able to discover exactly what an organism remembers by reconstructing the original sensory patterns from the "memory traces" which they have left, much as we might develop a photographic negative, or translate the pattern of electrical charges in the "memory" of a digital computer. This hypothesis is appealing in its simplicity and ready intelligibility, and a large family of theoretical brain models has been developed around the idea of a coded, representational memory (**2**, **3**, **9**, **14**). The alternative approach, which stems from the tradition of British empiricism, hazards the guess that the images of stimuli may never really be recorded at all, and that the central nervous system simply acts as an intricate switching network, where retention takes the form of new connections, or pathways, between centers of activity. In many of the more recent developments of this position (Hebb's "cell assembly," and Hull's "cortical anticipatory goal response," for example) the "responses" which are associated to stimuli may be entirely contained within the CNS itself. In this case the response represents an "idea" rather than an action. The important feature of this approach is that there is never any simple mapping of the stimulus into memory, according to some code which would permit its later reconstruction. Whatever in-

[1] The development of this theory has been carried out at the Cornell Aeronautical Laboratory, Inc., under the sponsorship of the Office of Naval Research, Contract Nonr-2381(00). This article is primarily an adaptation of material reported in Ref. 15, which constitutes the first full report on the program.

Frank Rosenblatt

Rosenblatt, Frank (1958), The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain, Cornell Aeronautical Laboratory, Psychological Review, v65, No. 6, pp. 386–408.



- The neurons in these networks were similar to those of McCulloch and Pitts.
- Rosenblatt's key contribution was the **introduction of a learning rule** for training perceptron networks to solve pattern recognition problems
- In addition to the variable weight values, the perceptron model added an extra input that represents *bias*. Thus, the modified equation from is now as follows:

$$Sum = \sum_{i=1}^{N} I_i W_i + b$$

$$\text{bias}, b = -t$$

$$Output = \begin{cases} 0 & \text{if Sum} \leq 0 \\ 1 & \text{if Sum} > 0 \end{cases}$$

M-P model

$$Sum = \sum_{i=1}^{N} I_i W_i$$

$$Output = \begin{cases} 0 & \text{if Sum} \leq T \\ 1 & \text{if Sum} > T \end{cases}$$

Bias is a measure of how easy it is to get the perceptron to output 1.

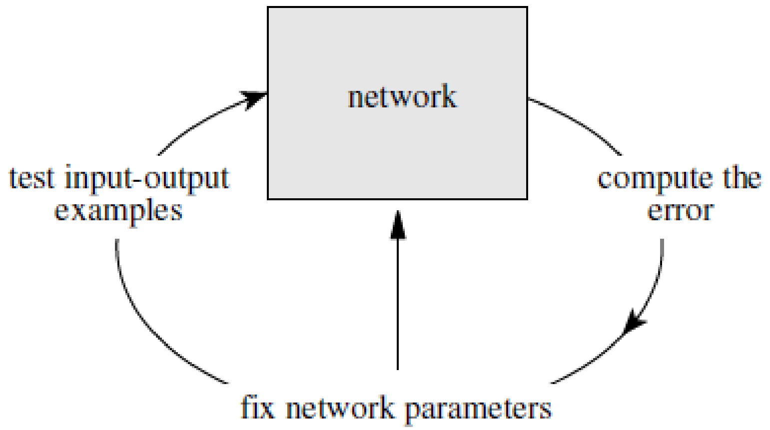
Perceptron learning rule

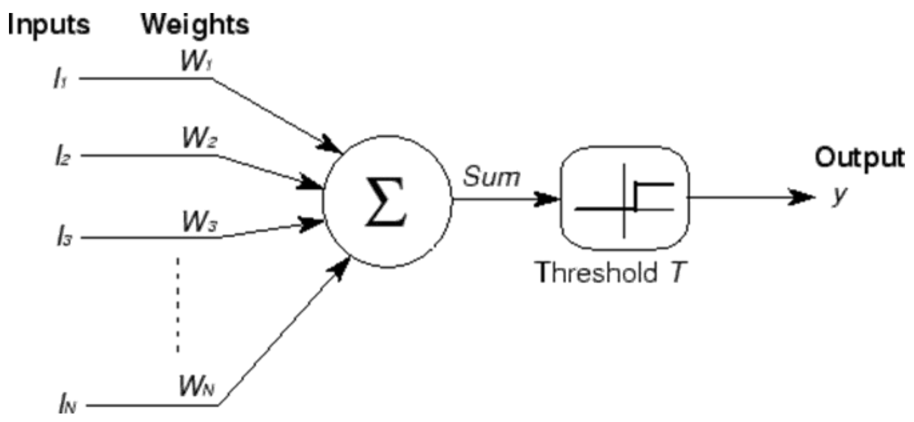
$$w(i+1) = w(i) + \Delta w$$

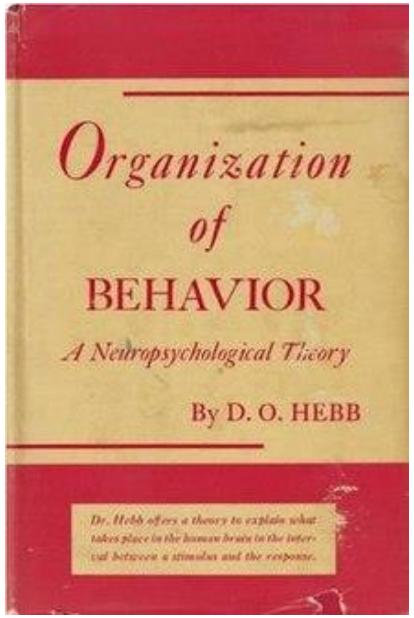
$$\Delta w = \eta . Error . x$$

$$Error = t - y$$

$$b(i+1) = b + \eta E$$







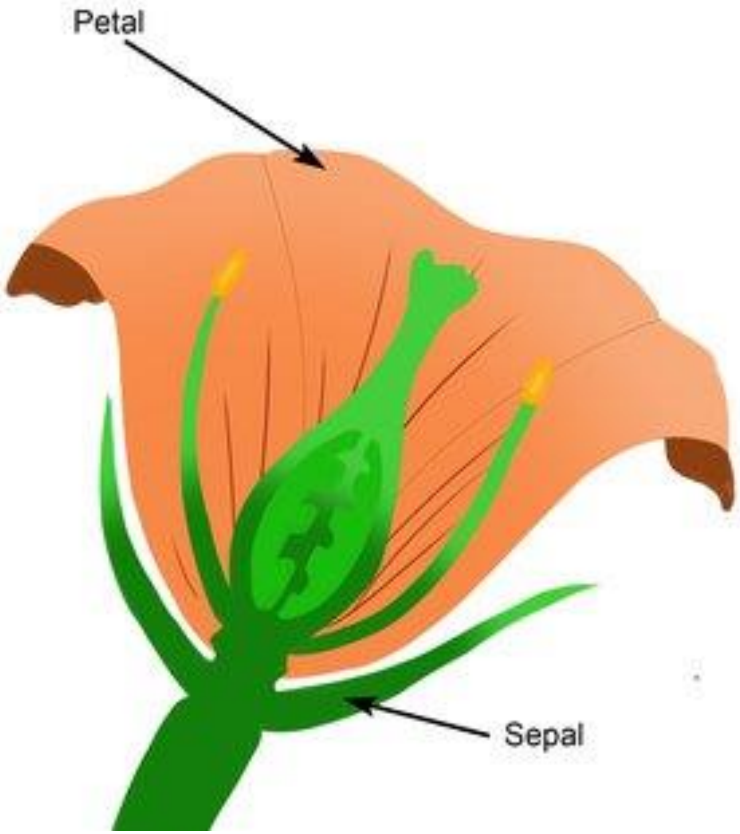
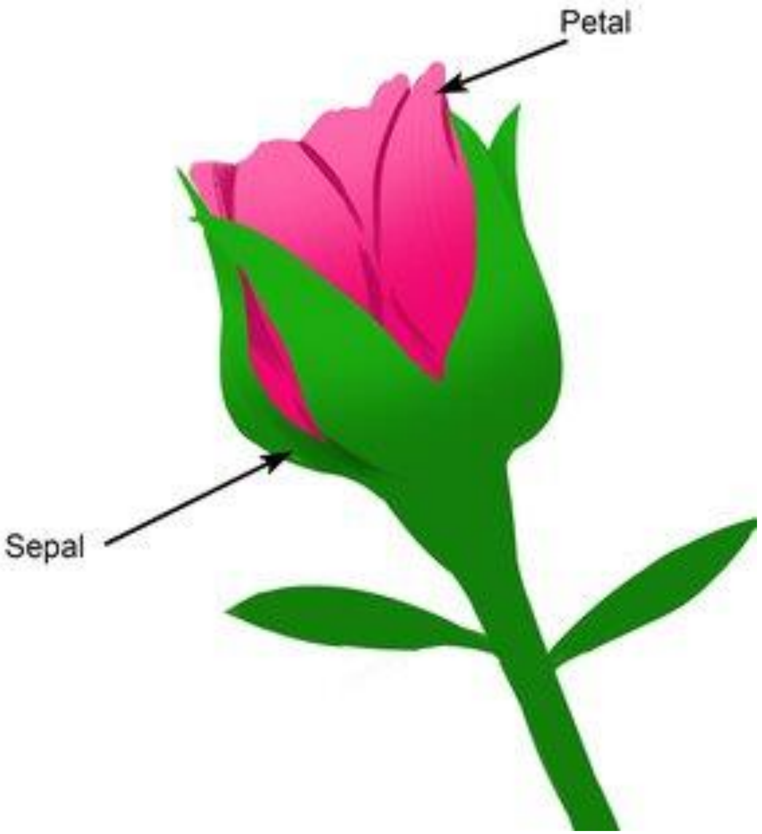
x = input
η is the learning rate
y = expected output
w = weights
b = bias

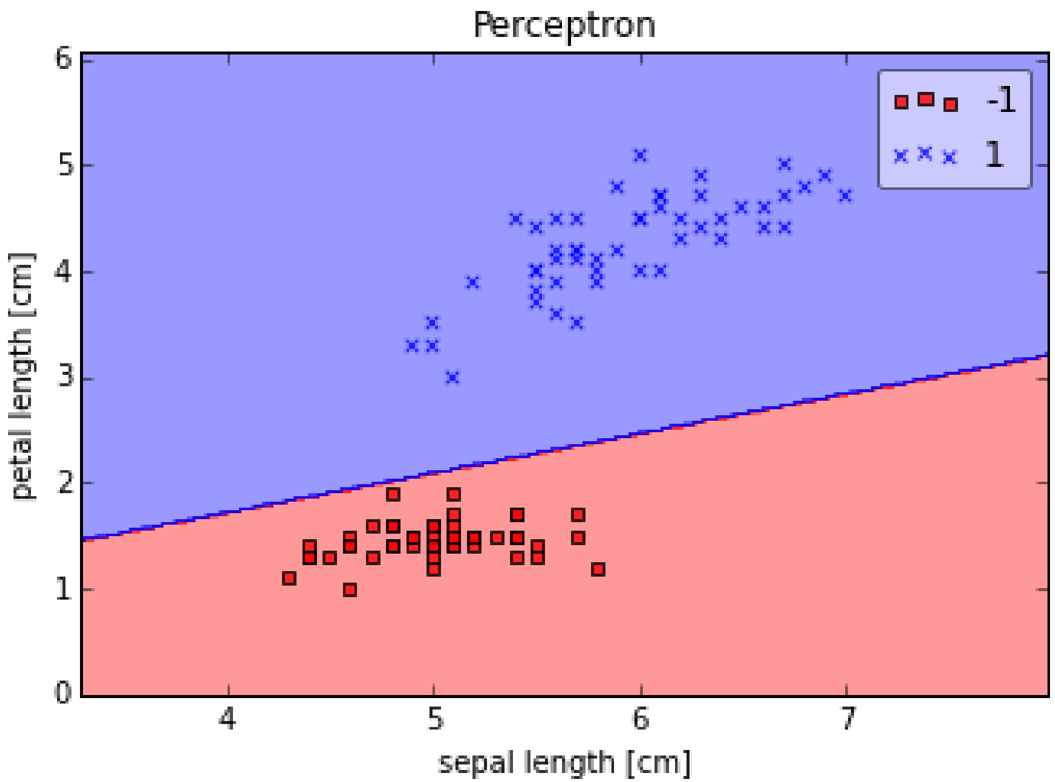
The Perceptron convergence theorem states that for any data set which is linearly separable the Perceptron learning rule is guaranteed to find a solution in a finite number of steps.

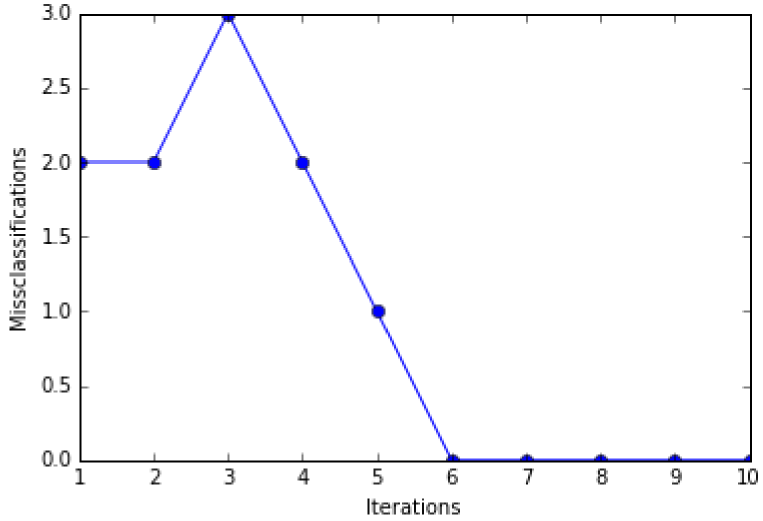
Hebb, D.O. (1949). The Organization of Behavior. New York: Wiley & Sons.

Iris versicolor

Iris setosa

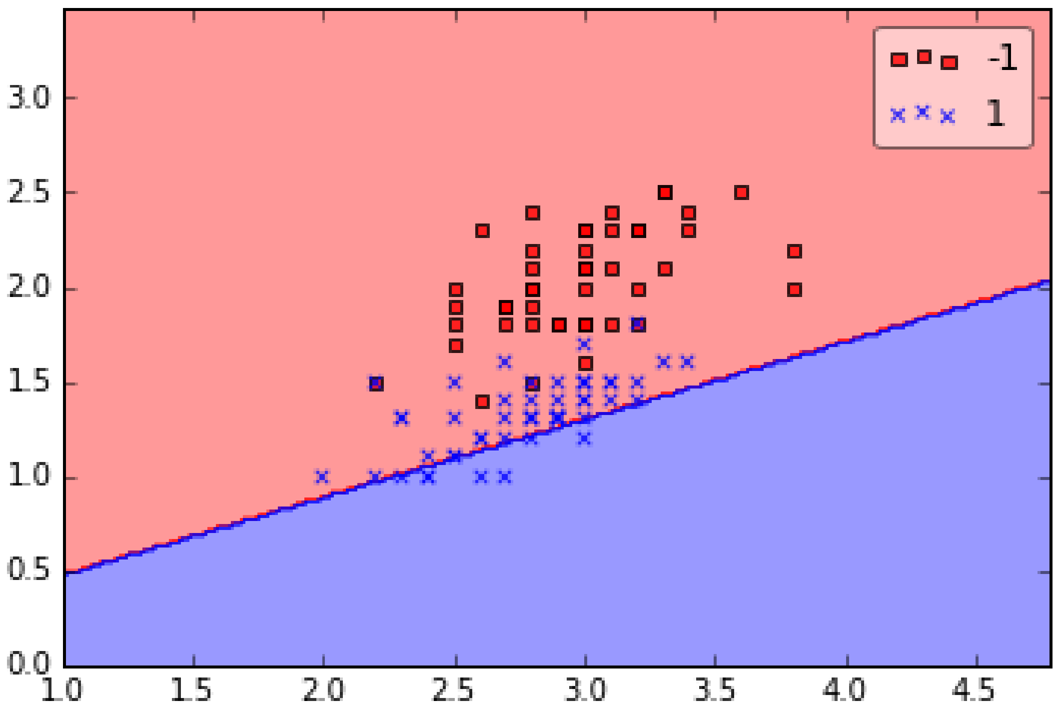

Petal
Sepal

Petal
Sepal

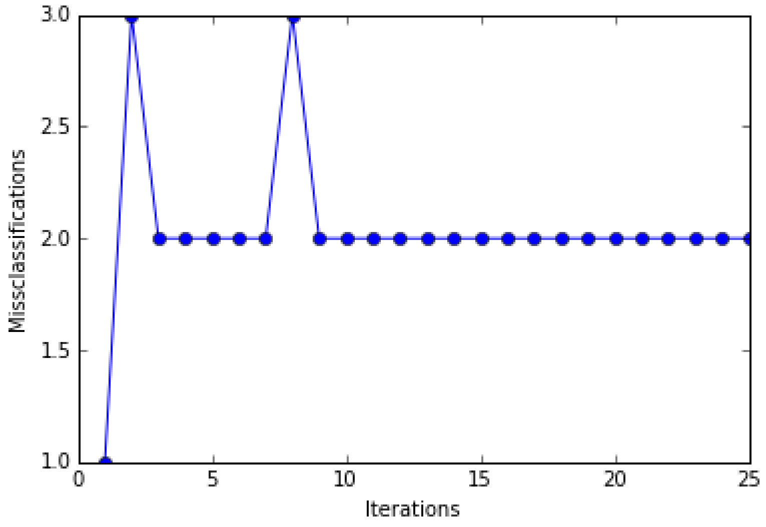

Perceptron
-1
1
petal length [cm]
sepal length [cm]
0
1
2
3
4
5
6
4
5
6
7

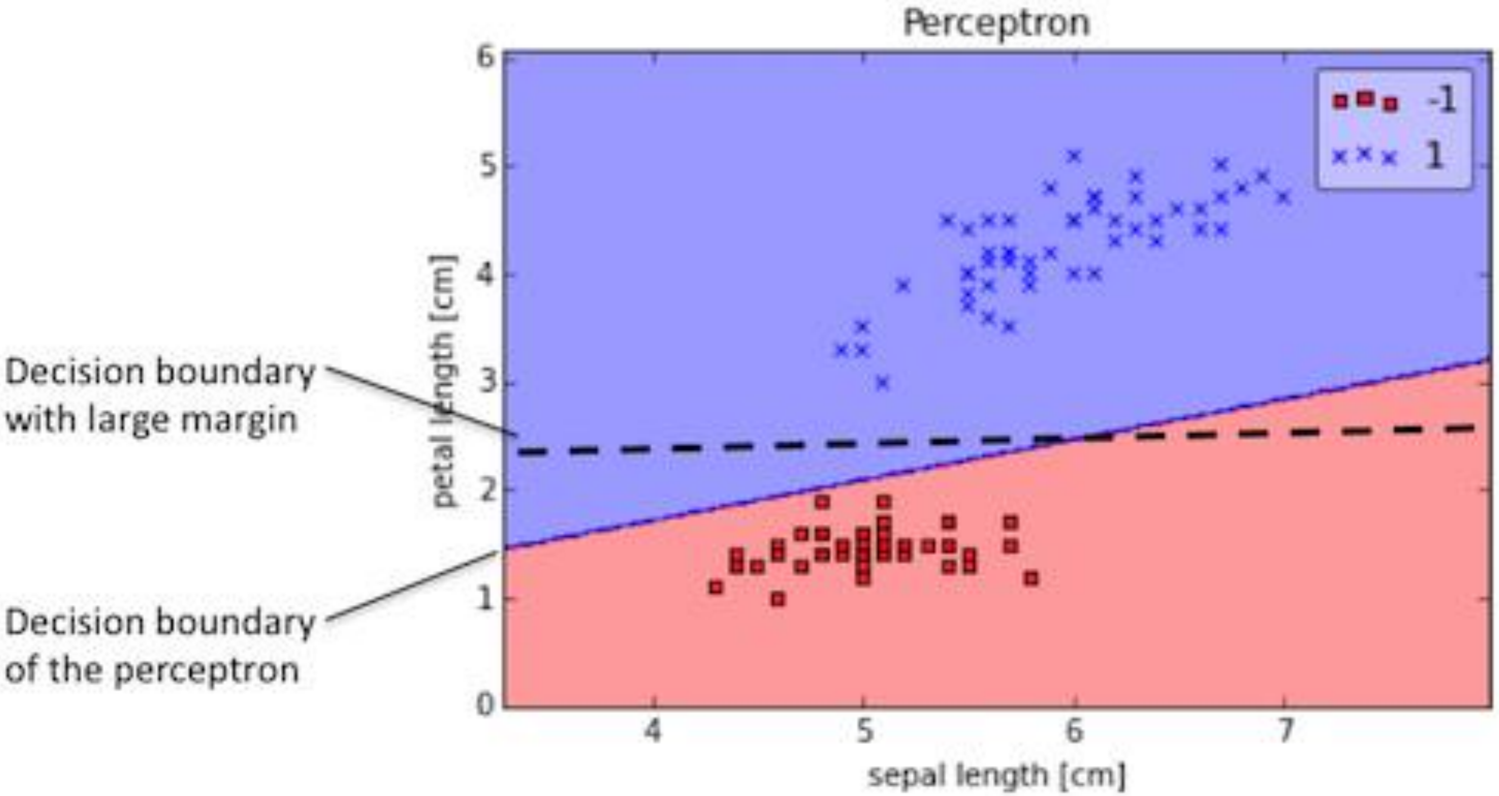

Missclassifications
Iterations
0.0
0.5
1.0
1.5
2.0
2.5
3.0
1
2
3
4
5
6
7
8
9
10

Iris virginica

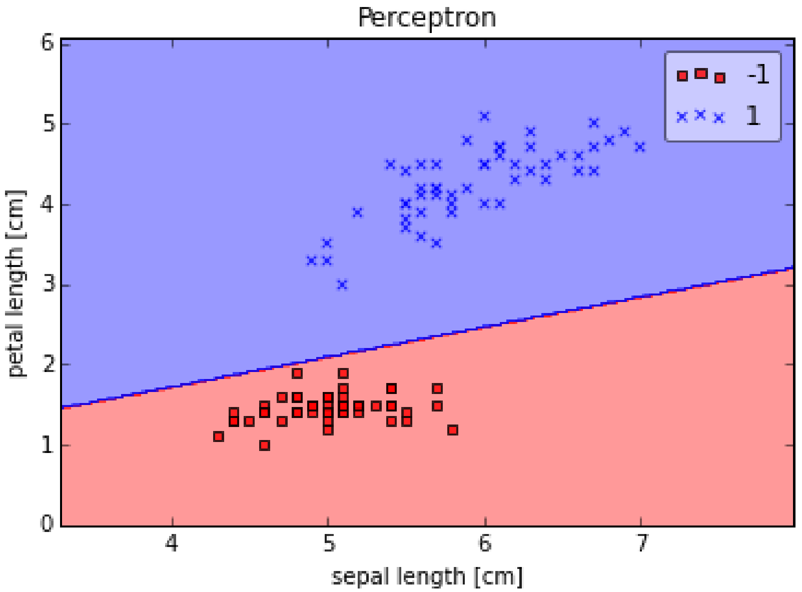
Iris versicolor

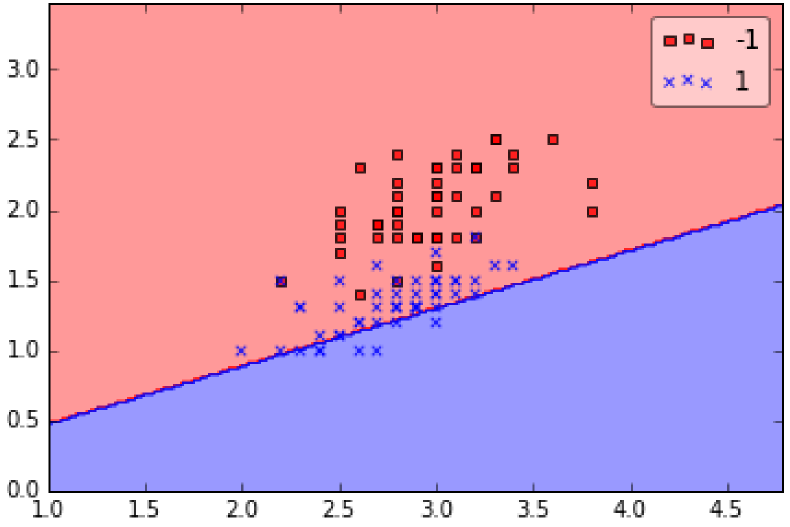

-1
1
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
0.0
0.5

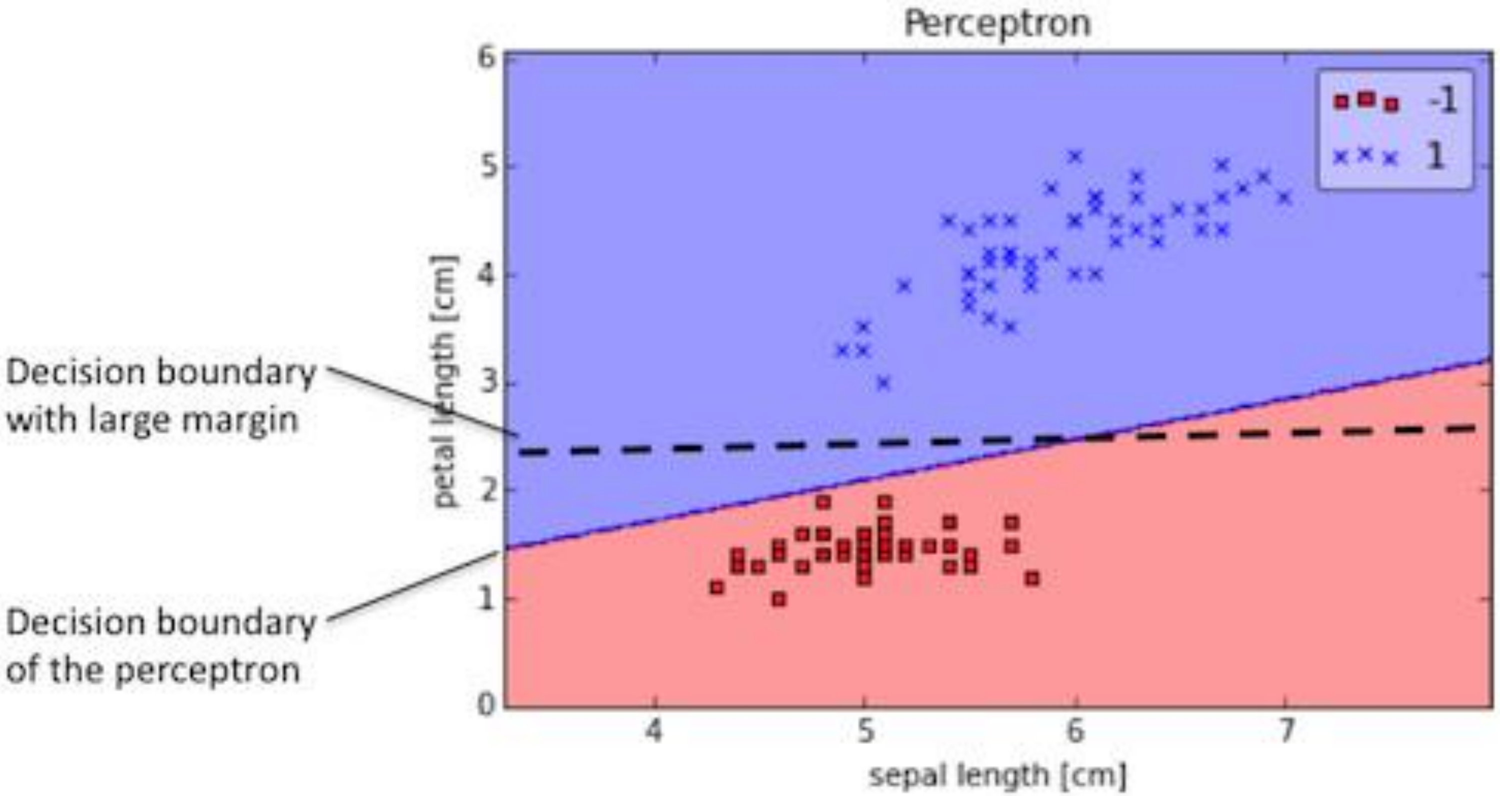

Missclassifications
Iterations
0
5
10
15
20
25
1.0
1.5
2.0
2.5
3.0



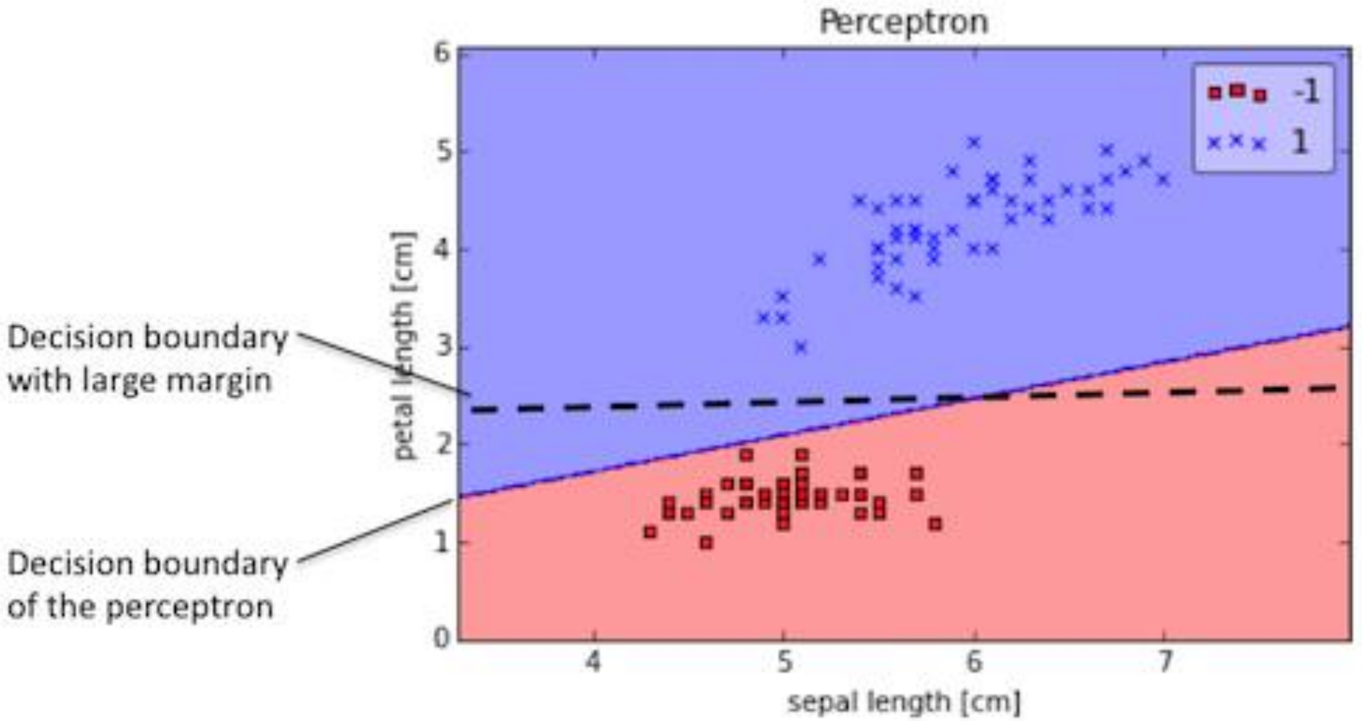
Example of a large-margin decision boundary.



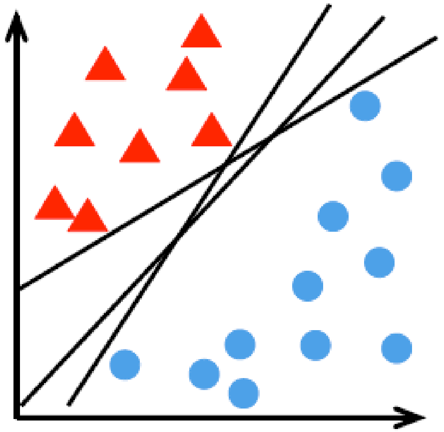




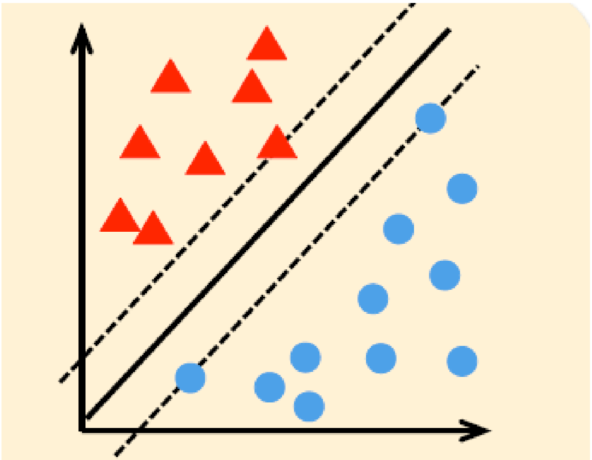
Example of a large-margin decision boundary.



Example of a large-margin decision boundary.



These decision boundaries are very close to some items in the training data.
They have small margins.



This decision boundary is as far away from any training items as possible.
It has a large margin.

- **Margin** = the distance of the decision boundary to the closest items in the training data.
- We want to find a classifier whose decision boundary is furthest away from the nearest data points. (This classifier has the largest margin).