Neural networks: a brief history

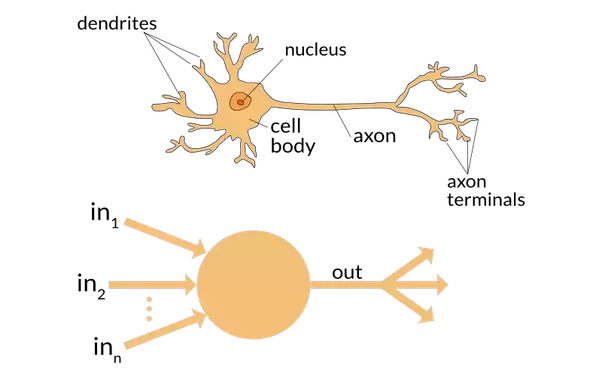
Source: translation from the [Fastai book](https://github.com/fastai/fastbook) [Your deep learning journey](https://github.com/fastai/fastbook/blob/master/01_intro.ipynb)

# Mathematical model of an artificial neuron

In 1943 **Warren McCulloch**, a neurophysiologist, and **Walter Pitts**, a logician, teamed up to develop **a mathematical model of an artificial neuron**. They declared that:

*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*. (Pitts and McCulloch; A Logical Calculus of the Ideas Immanent in Nervous Activity)

They realised that **a simplified model of a real neuron could be represented using simple addition and thresholding** as shown here:



# The Perceptron: the ability to learn

**Frank Rosenblatt** further (1956) developed the artificial neuron to give it the **ability to learn**. Even more importantly, he worked on building the first device that actually used these principles: The Mark I **Perceptron**.

Rosenblatt wrote about this work: "**we are about to witness the birth of such a machine – a machine capable of perceiving, recognizing and identifying its surroundings without any human training or control**". The perceptron was built, and was able to successfully recognize simple shapes.

# Perceptrons: start of the first AI Winter

An MIT professor named **Marvin Minsky** (who was a grade behind Rosenblatt at the same high school!) along with **Seymour Papert** wrote a **book, called "Perceptrons"**, about Rosenblatt's invention. They showed that a single layer of these devices was unable to learn some simple, critical mathematical functions (such as XOR).

In the same book, they also showed that using multiple layers of the devices would allow these limitations to be addressed. Unfortunately, only the first of these insights was widely recognized, as a result of which the global academic community nearly entirely gave up on neural networks for the next two decades.

# Neural Networks

Perhaps the most pivotal work in neural networks in the last 50 years is the multi-volume *Parallel Distributed Processing* (PDP), released in 1986 by MIT Press. Chapter 1 lays out a similar hope to that shown by Rosenblatt:

: *…people are smarter than today's computers because the brain employs a basic computational architecture that is more suited to deal with a central aspect of the natural information processing tasks that people are so good at. …we will introduce a computational framework for modeling cognitive processes that seems… closer than other frameworks to the style of computation as it might be done by the brain.* (PDP, chapter 1)

The premise that PDP is using here is that **traditional computer programs work very differently to brains**, and that might be why computer programs had (at that point) been so bad at doing things that brains find easy (such as recognizing objects in pictures). The authors claim that the PDP approach is "closer than other frameworks" to how the brain works, and therefore it might be better able to handle these kinds of tasks.

In fact, the approach laid out in PDP is very similar to the approach used in today's neural networks. The book defined "Parallel Distributed Processing" as requiring:

1. A set of *processing units*
2. A *state of activation*
3. An *output function* for each unit
4. A *pattern of connectivity* among units
5. A *propagation rule* for propagating patterns of activities through the network of connectivities
6. An *activation rule* for combining the inputs impinging on a unit with the current state of that unit to produce a new level of activation for the unit
7. A *learning rule* whereby patterns of connectivity are modified by experience
8. An *environment* within which the system must operate

We will see in this book that modern neural networks handle each of these requirements.

# 1980’s: first applications of NN

In the 1980's most models were built with a second layer of neurons, thus avoiding the problem that had been identified by Minsky (this was their "pattern of connectivity among units", to use the framework above). And indeed, neural networks were widely used during the 80s and 90s for real, practical projects. However, again a misunderstanding of the theoretical issues held back the field. In theory, adding just one extra layer of neurons was enough to allow any mathematical model to be approximated with these neural networks, but in practice such networks were often too big and slow to be useful.

# More layers!

Although researchers showed 30 years ago that to get practical good performance you need to use even more layers of neurons, it is only in the last decade that this has been more widely appreciated. Neural networks are now finally living up to their potential, thanks to the understanding to use more layers as well as improved ability to do so thanks to improvements in computer hardware, increases in data availability, and algorithmic tweaks that allow neural networks to be trained faster and more easily.

We now have what Rosenblatt had promised: "a machine capable of perceiving, recognizing and identifying its surroundings without any human training or control".