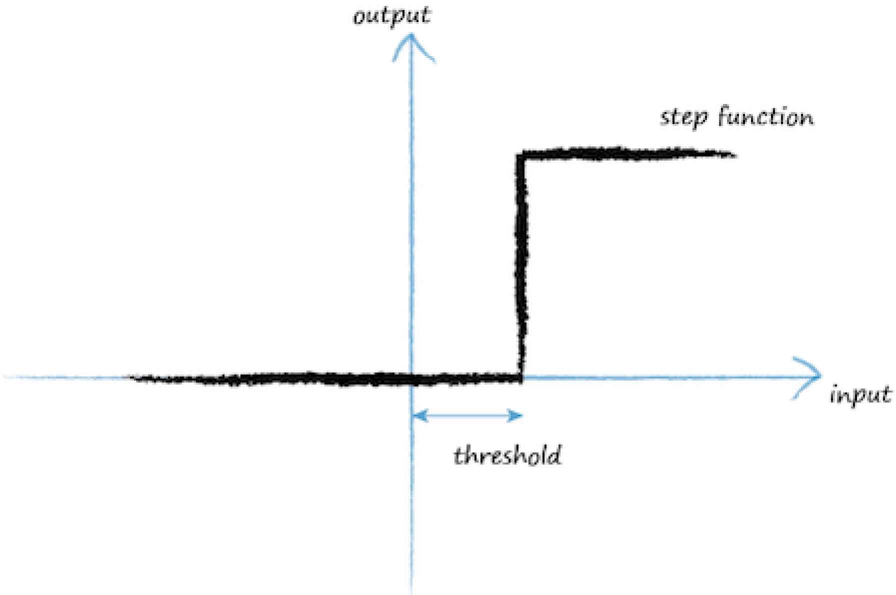
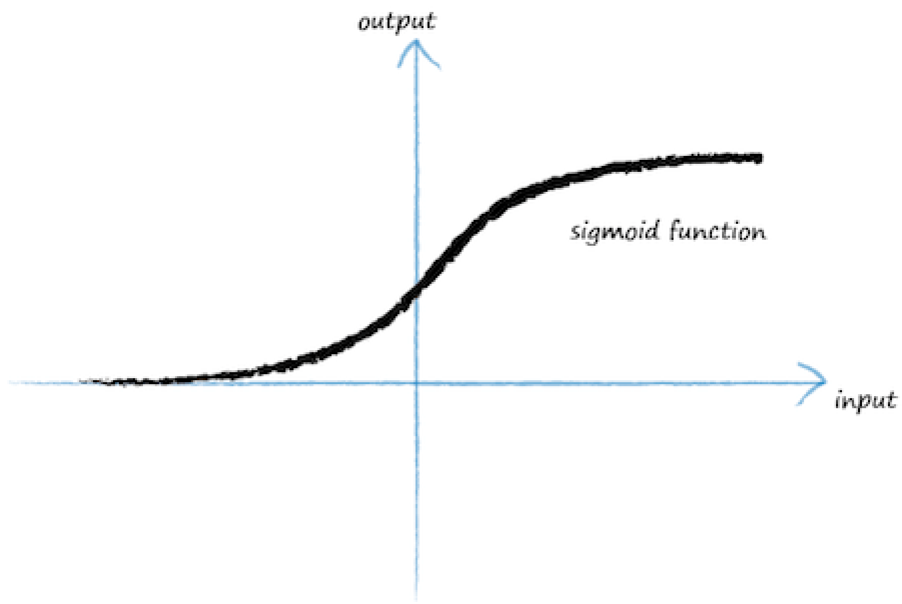
What is an Activation Function?

An introduction to Activation function and its use in calculating the output of a neural network.

A function that takes the input signal and generates an output signal, but takes into account some kind of threshold is called an *activation function*. Mathematically, there are many such activation functions that could achieve this effect. A simple *step function* could do this:



You can see for low input values; the output is zero. However, once the threshold input is reached, output jumps up. An artificial neuron behaving like this would be like a real biological neuron. The term used by scientists actually describes this well, they say that neurons fire when the input reaches the threshold. We can improve on the step function. The S-shaped function shown below is called the *Sigmoid function*. It is smoother than the cold hard step function, and this makes it more natural and realistic. Nature rarely has cold hard edges!



This smooth S-shaped sigmoid function is what we will continue to use for making our own neural network. Artificial intelligence researchers will also use other, similar looking functions, but the sigmoid is simple and actually very common too, so we're in good company. The sigmoid function, sometimes also called the *logistic function*, is

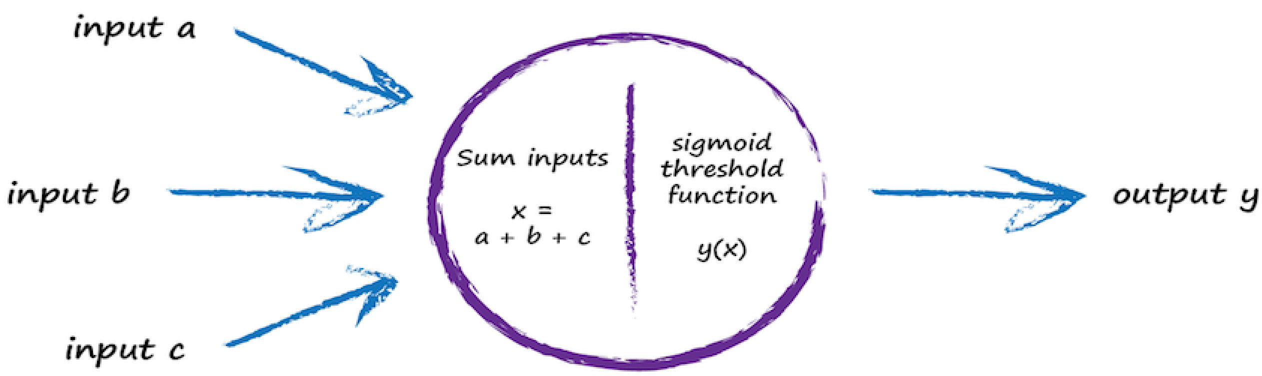
$$y = \frac{1}{1 + e^{-x}}$$

That expression isn't as scary as it first looks. The letter *e* is a mathematical constant 2.71828... It's a very interesting number that pops up in all sorts of areas of mathematics and physics, and the reason I've used the dots ... is because the decimal digits keep going on forever. Numbers like that have a fancy name, transcendental numbers. That's all very well and fun, but for our purposes, you can just think of it as 2.71828. The input *x* is negated and *e* is raised to the power of that -*x*. The result is added to 1, so we have $1 + e^{-x}$. Finally, the inverse is taken of the whole thing, that is 1 is divided by $1 + e^{-x}$. That is what the mildly scary looking function above does to the input *x* to give us the output value y. So not so scary after all! Just out of interest, when *x* is zero, e^{-x} is one because anything raised to a power of zero is 1. So y becomes $1/(1+1)$ or simply $1/2$, a half. So the basic sigmoid cuts the y-axis at $y = 1/2$.

There is another, very powerful reason for choosing this sigmoid function over the many many other S-shaped functions we could have used for a neuron's output. The reason is that this sigmoid function is much easier to do

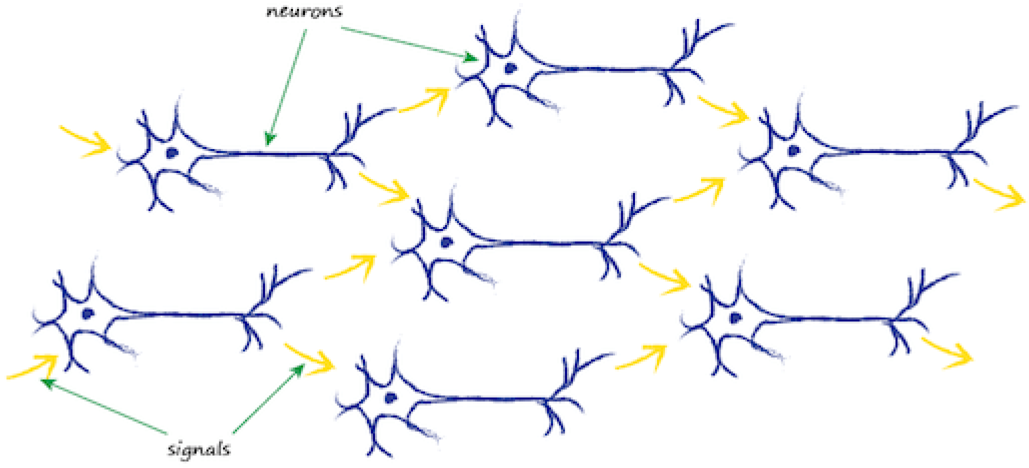
calculations with than other S-shaped functions, and we'll see why in practice later. Let's get back to neurons, and consider how we might model an artificial neuron. The first thing to realize is that real biological neurons take many inputs, not just one. We saw this when we had two inputs to the Boolean logic machine, so the idea of having more than one input is not new or unusual.

What do we do with all these inputs? We simply combine them by adding them up, and the resultant sum is the input to the sigmoid function which controls the output. This reflects how real neurons work. The following diagram illustrates this idea of combining inputs and then applying the threshold to the combined sum:



If the combined signal is not large enough then the effect of the sigmoid threshold function is to suppress the output signal. If the sum *x* if large enough the effect of the sigmoid is to fire the neuron. Interestingly, if only one of the several inputs is large and the rest small, this may be enough to fire the neuron. What's more, the neuron can fire if some of the inputs are individually almost, but not quite, large enough because when combined the signal is large enough to overcome the threshold. Intuitively, this gives you a sense of the more sophisticated, and in a sense fuzzy, calculations that such neurons can do.

The electrical signals are collected by the dendrites, and these combine to form a stronger electrical signal. If the signal is strong enough to pass the threshold, the neuron fires a signal down the axon towards the terminals to pass onto the next neuron's dendrites. The following diagram shows several neurons connected in this way:



The thing to notice is that each neuron takes input from many before it, and also provides signals to much more if it happens to be firing. Now, in the next lesson, we will try to sum up our whole discussion until this point and try to come up with a scientific replica of a biological brain.