In biologically inspired neural networks, the activation function is usually an abstraction representing the rate of [action potential](http://en.wikipedia.org/wiki/Action_potential) firing in the cell. In its simplest form, this function is [binary](http://en.wikipedia.org/wiki/Binary_function)—that is, either the [neuron](http://en.wikipedia.org/wiki/Neuron) is firing or not. The function looks like \phi(v_i)=U(v_i), where U is the [Heaviside step function](http://en.wikipedia.org/wiki/Heaviside_step_function). In this case a large number of neurons must be used in computation beyond linear separation of categories.

A line of positive [slope](http://en.wikipedia.org/wiki/Slope) may also be used to reflect the increase in firing rate that occurs as input current increases. The function would then be of the form \phi(v_i)=\mu v_i, where \mu is the slope. This activation function is linear, and therefore has the same problems as the binary function. In addition, networks constructed using this model have [unstable convergence](http://en.wikipedia.org/wiki/BIBO_stability) because neuron inputs along favored paths tend to increase without bound, as this function is not [normalizable](http://en.wikipedia.org/wiki/Normalizing_constant).

All problems mentioned above can be handled by using a normalizable [sigmoid](http://en.wikipedia.org/wiki/Sigmoid_function) activation function. One realistic model stays at zero until input current is received, at which point the firing frequency increases quickly at first, but gradually approaches an [asymptote](http://en.wikipedia.org/wiki/Asymptote) at 100% firing rate. Mathematically, this looks like \phi(v_i)=U(v_i)\tanh(v_i), where the [hyperbolic tangent](http://en.wikipedia.org/wiki/Hyperbolic_tangent) function can also be replaced by any [sigmoid function](http://en.wikipedia.org/wiki/Sigmoid_function). This behavior is realistically reflected in the neuron, as neurons cannot physically fire faster than a certain rate. This model runs into problems, however, in computational networks as it is not [differentiable](http://en.wiktionary.org/wiki/differentiation), a requirement in order to calculate [backpropagation](http://en.wikipedia.org/wiki/Backpropagation).

The final model, then, that is used in [multilayer perceptrons](http://en.wikipedia.org/wiki/Multilayer_perceptron) is a sigmoidal activation function in the form of a hyperbolic tangent. Two forms of this function are commonly used: \phi(v_i)=\tanh(v_i) whose range is normalized from -1 to 1, and \phi(v_i) = (1+\exp(-v_i))^{-1} is vertically translated to normalize from 0 to 1. The latter model is often considered more biologically realistic, but it runs into theoretical and experimental difficulties with certain types of computational problems.

<http://ufldl.stanford.edu/wiki/index.php/Neural_Networks>