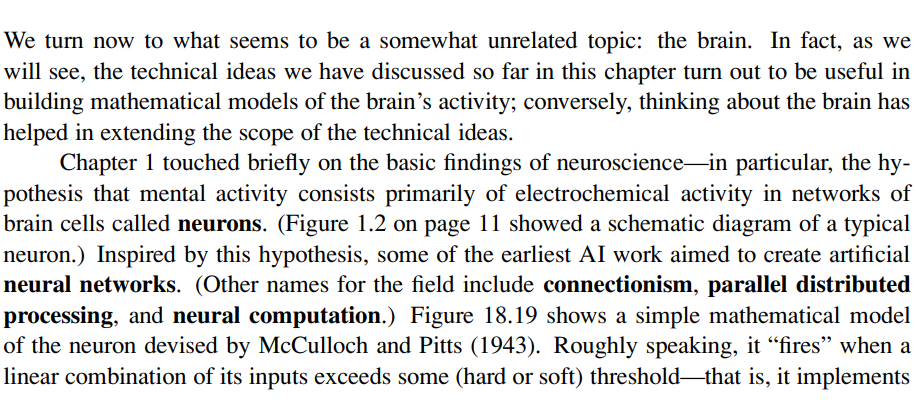
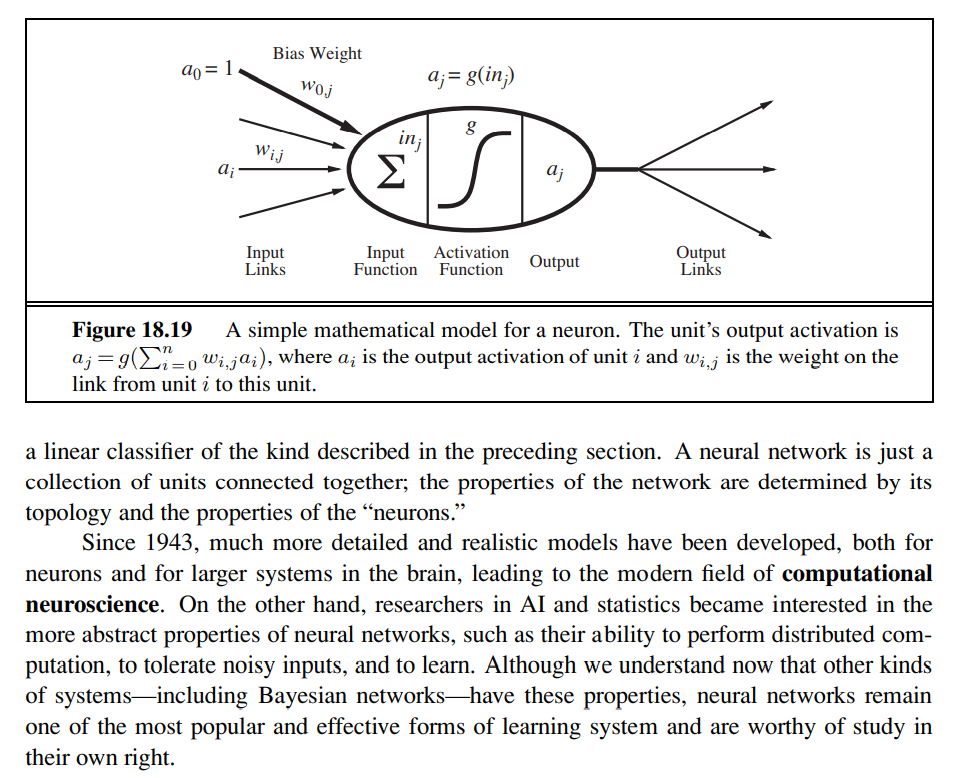
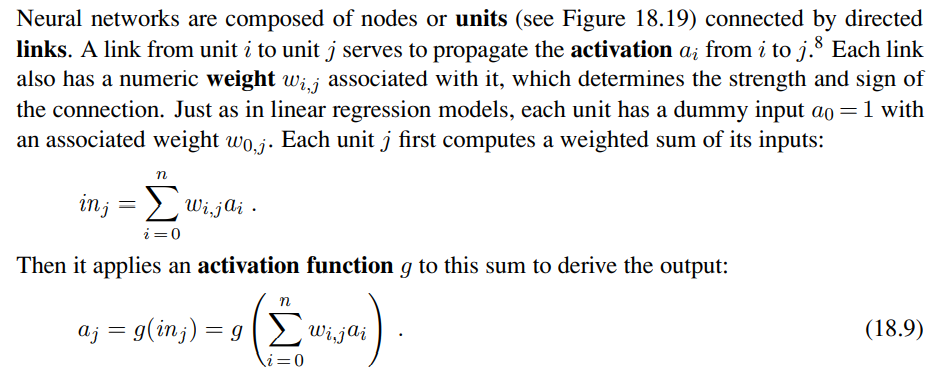
# Artificial Neural Networks



Therefore, artificial neural networks are inspired by the neuroscience hypothesis that brain activity is based on the electrochemical activity in the network of brain cells which are called neurons. In other words, according to neuroscience our mental processes are carried out thanks to the electrochemical activities that occur inside our network of neurons in the brain.





Therefore, given that artificial neural networks are a mathematical model of the network of neurons inside our brain then they consist of many **units** (also called nodes or neurons) which are connected to each other through directed links. In other words, artificial neural networks are directed graphs of nodes where the nodes are called **units** or **neurons.**

Each unit i outputs a single value which will be carried towards all nodes j such that there exists a link from node I to node j.

The links in the artificial neural networks are weighted links. The weight in the link from unit i to unit j is called wij.

Each neuron j receives an input called inj which is equal to the weighted sum of all outputs from nodes i, called ai , such that they have a directed link to node j. Each output ai is weighted by the weight of the directed link from node i to node j.

It is important to notice that each unit j has a dummy input a0 = 1 associated with a weight w0j as in linear regression.

Once a unit j receives inj then it applies an **ACTIVATION FUNCTION** g such that its output is the following:

aj = g (inj)

Normally, such an activation function is either the hard-threshold or the logistic function (soft-threshold).

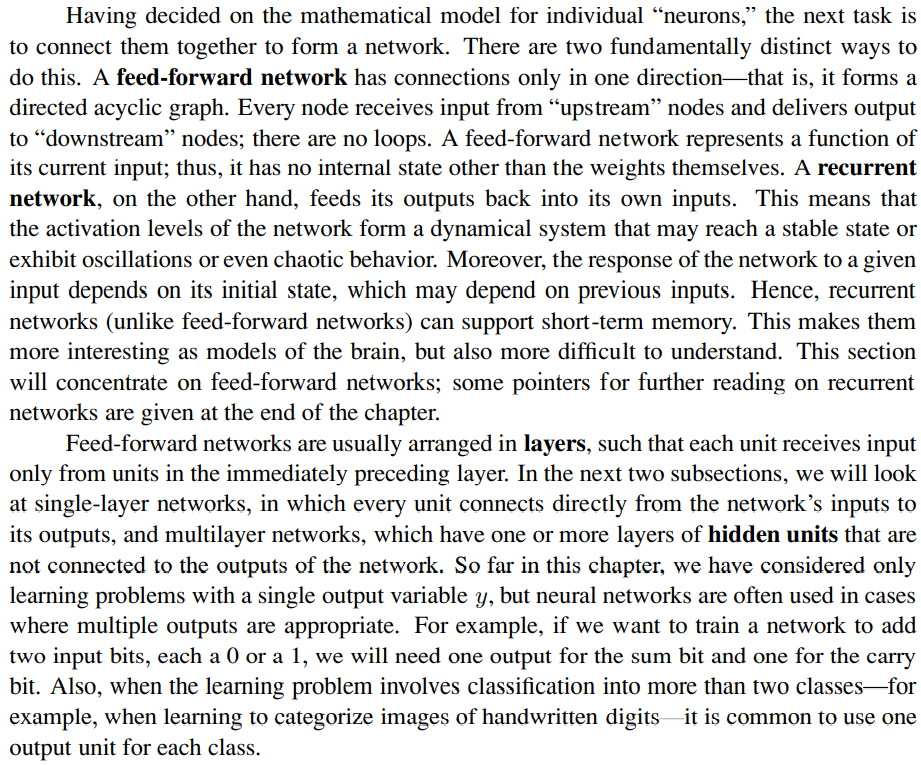
Therefore, we can see that a single unit in a artificial neural network is just a linear classifier as it receives a linear combination of inputs which are fed into the activation function which will output either 0 or 1 if the activation function is the hard-threshold or a continuous range of values in [0,1] (0 and 1 excluded) if the activation function is the logistic function.

Given that we use an artificial neural network with a single neuron to do a binary classification of data points in n-dimensions then this neuron will have n+1 directed weighted links to it, they are n+1 links because we need to add the dummy link for the weight wo, and in order to train this neuron we already know what to do as this problem is just a simple linear classification. In other words, we just need to perform gradient descent in order to find the best weights that minimise the loss.

Units in an artificial neural network that use as activation function the hard-threshold are called **PERCEPTRONS.**

Units in an artificial neural network that use as activation function the logistic-function are called **SIGMOID PERCEPTRONS.**

Both of these nonlinear activation functions ensure the important property that the entire network of units can represent a nonlinear function. I need to find out why this is true.

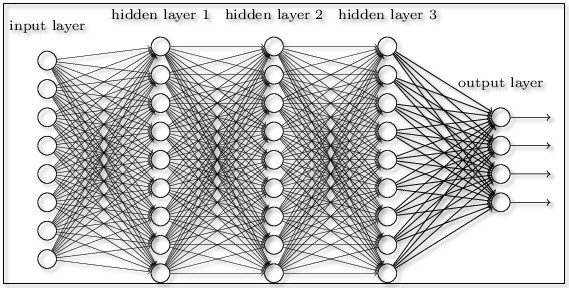


Therefore, we can categorise artificial neural networks into two categories:

1. Feed-forward network
2. Recurrent network

A feed-forward network is an artificial neural network which forms a directed acyclic graph. In other words, we can imagine that a feed-forward network comprises some layers of units where each neuron of the same unit is not linked to the other neurons of the same layer, each neuron of layer l receives inputs only from neurons in layer l-1, each neuron of layer l outputs only to neurons in layer l+1.

A representation of a feed-forward network is the following:

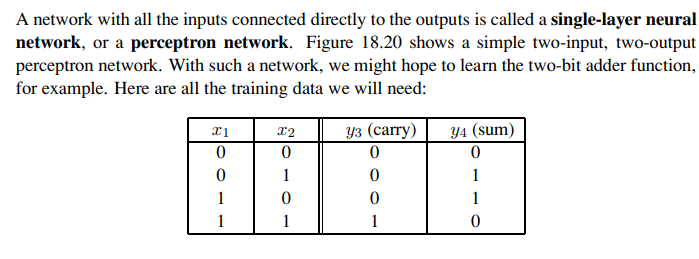


As we can see we can distinguish the layers of a feed forward neural network into:

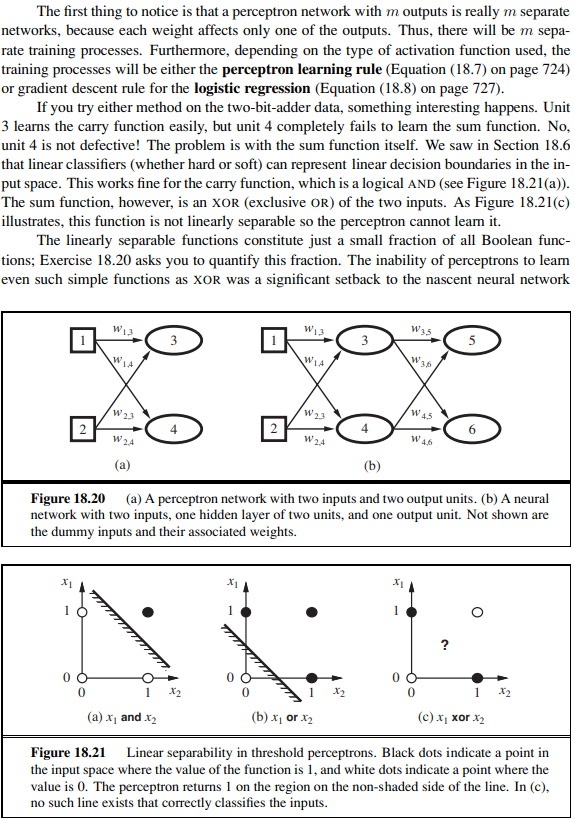
1. Input layer => comprises a set of units that just output the input values of the problem
2. Hidden layer => comprises a set of units that are neither output units or input units and they are between the input layer and the output layer
3. Output layer => comprises a set of units that output the final result of the computation

A single layer feed-forward neural network comprises only the input layer and the output layer. Multiple layers feed-forward neural networks comprise an input layer, an output layer and a finite number of layers between the input layer and the output layer which are called hidden-layers.

It is common to have multiple units in the output layer when solving classification problems involving more than 2 categories.



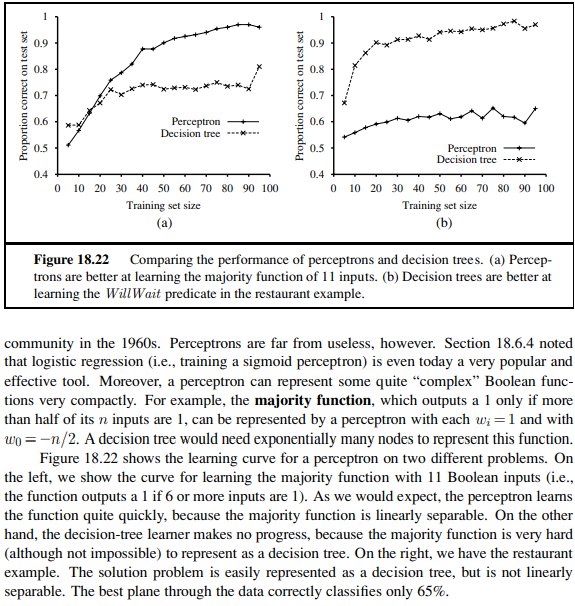
Therefore, a single-layer feed forward neural network is called **PERCEPTRON NETWORK.**



Therefore, in the presence of a **PERCEPTRON NETWORK** with m units in the output layer, we actually have m different perceptron networks with only one unit in the output layer. Thus, we already know how to train a **PERCEPTRON NETWORK** with m output units as it just consists of training the m output units separately. We know how to train the units as it consists of just solving a linear classification problem.

In the case of a perceptron unit then we are going to use the perceptron rule, otherwise in the case of a sigmoid perceptron we are going to use logistic regression through gradient descent.

As we can see, while a unit can easily learn to perform the “and” and “or” Boolean functions, it can’t learn the “xor” Boolean function. This occurs because while the “and” and “or” input values are linearly classifiable as shown in 18.21, the “xor” input values are not linearly classifiable as there is no straight line that can separate the two categories 0 and 1.



Therefore, what we have learnt is that the capabilities of a perceptron network are limited as it possible to use only in those binary classification problems which are linearly classifiable. However, nonetheless perceptron networks are not useless as can represent complex Boolean functions very compactly like the Majority function (which is linearly classifiable). As an example, a decision tree would need exponential many nodes to represent a majority function.