Artificial Neural Network : Introduction

Credits:

Soft Computing Applications, Dr. Debasis Samanta, IIT Kharagpur

RC Chakraborty, www.myreaders.info

"Principles of Soft Computing" by S.N. Sivanandam & SN Deepa

Neural Networks, Fuzzy Logic and Genetic Algorithms by S. Rajasekaran and GAV Pai

Neural networks- a comprehensive foundation by Simon S Haykin

Biological nervous system

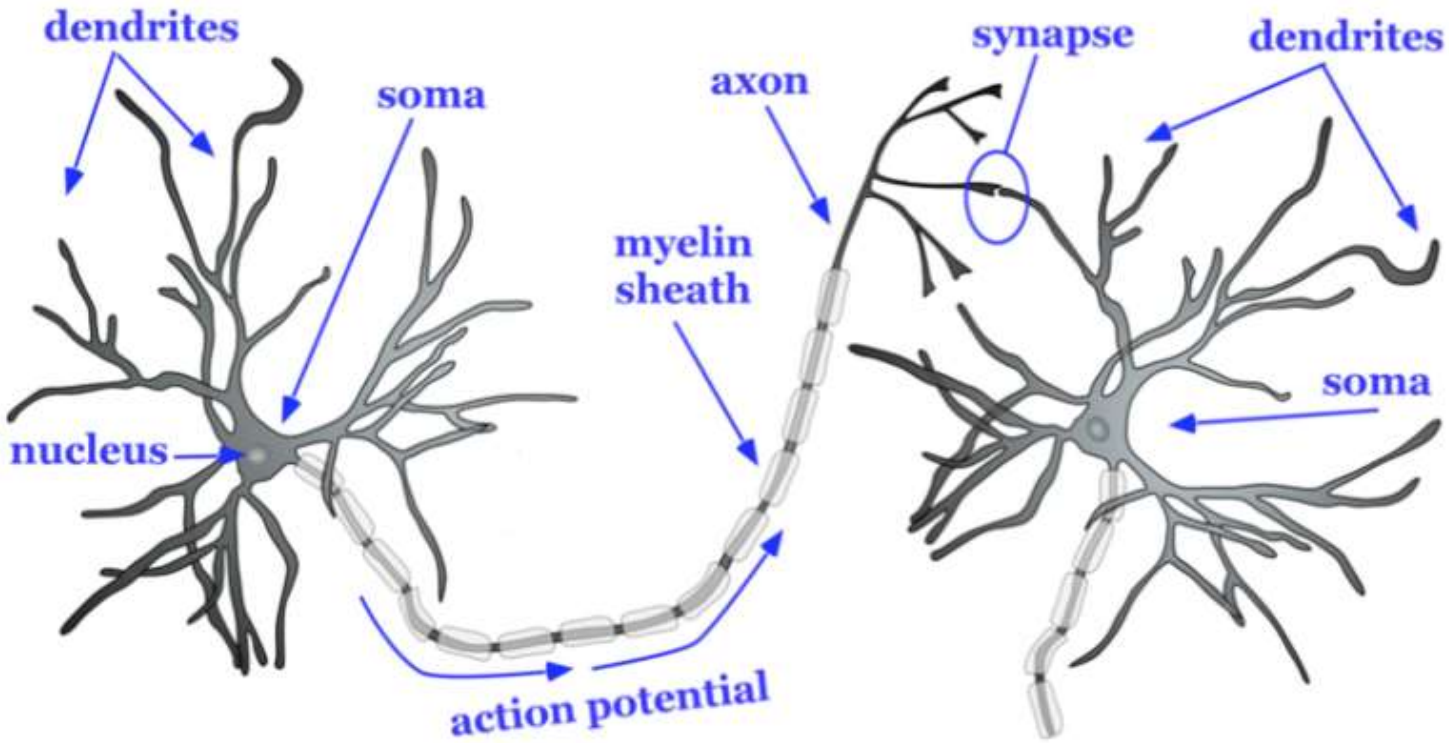
- Biological nervous system is the most important part of many living things, in particular, human beings.
- **brain** is the center of human nervous system.
- In fact, any biological nervous system consists of a large number of interconnected processing units called **neurons**.
- Each neuron is approximately $10\mu m$ long and they can operate in parallel.
- Typically, a human brain consists of approximately 10^{11} neurons communicating with each other with the help of **electrical impulses.**

Brain: Center of the nervous system

Computation

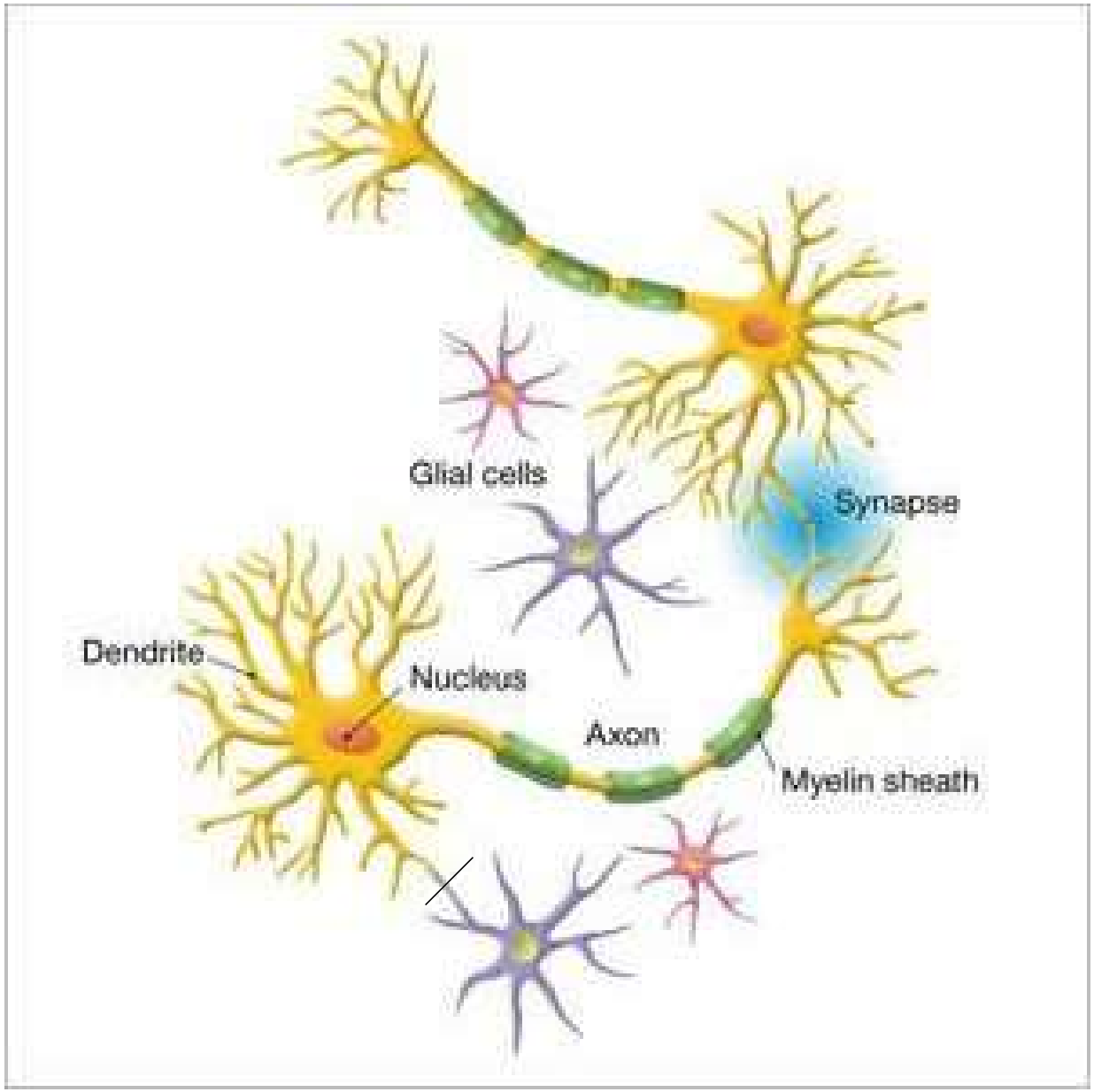
	processing elements	element size	energy use	processing speed	style of computation	fault tolerant	learns	intelligent, conscious
	10^{14} synapses	10^{-6} m	30 W	100 Hz	parallel, distributed	yes	yes	usually
	10^{8} transistors	10^{-6} m	30 W (CPU)	10^{9} Hz	serial, centralized	no	a little	not (yet)

Neuron: Basic unit of nervous system



Basic structure of biological neuron

Neuron and its working



Neuron and its working

Figure shows a schematic of a biological neuron. There are different parts in it : dendrite, soma, axon and synapse.

Dendrite : Extends away from the cell body and is the main input to the neuron. It receives information from thousands of other neurons. Receive information

Soma : It is also called a cell body, and just like as a nucleus of cell. Process information

Axon : A long cylindrical fibre. Carries information to other neuron

Synapse : It is a junction where axon makes contact with the dendrites of neighboring dendrites.

Nucleus: is located within the soma, contains genetic information, directs protein synthesis, and supplies the energy and the resources the neuron needs to function.

Neuron and its working

- There is a chemical in each neuron called neurotransmitter.
- A signal (also called sense) is transmitted across neurons by this chemical.
- That is, all inputs from other neuron arrive to a neurons through dendrites.
- These signals are accumulated at the synapse of the neuron and then serve as the output to be transmitted through the neuron.
- An action may produce an electrical impulse, which usually lasts for about a millisecond.
- Note that this pulse generated due to an incoming signal and all signal may not produce pulses in axon unless it crosses a **threshold value**.
- Also, note that an action signal in axon of a neuron is commutative signals arrive at dendrites which summed up at soma.

Definition of Neural Networks (NN)

According to the DARPA Neural Network Study (1988, AFCEA International Press, p. 60):

- ... a neural network is a system composed of many simple processing elements operating in parallel whose function is determined by network structure, connection strengths, and the processing performed at computing elements or nodes.

According to Haykin (1994), p. 2:

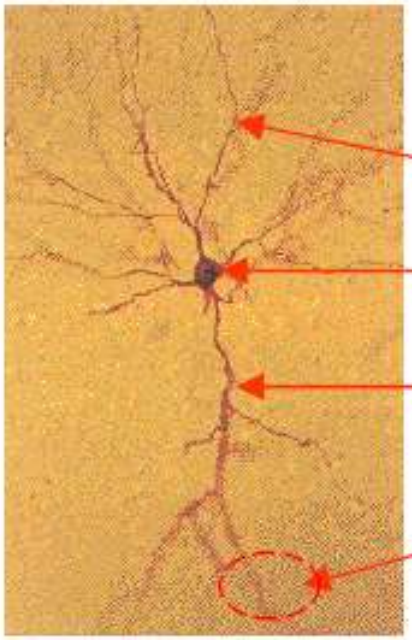
A neural network is a massively parallel distributed processor that has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects:

- Knowledge is acquired by the network through a learning process.
- Interneuron connection strengths known as synaptic weights are used to store the knowledge.

Artificial neural network

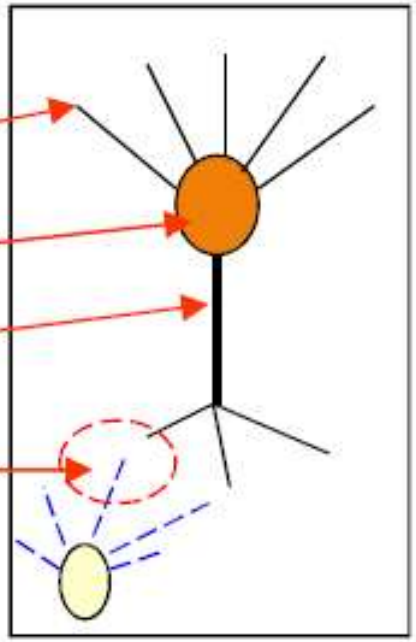
- In fact, the human brain is a highly complex structure viewed as a massive, highly interconnected network of simple processing elements called **neurons**.
- Artificial neural networks (ANNs) or simply we refer it as neural network (NNs), which are simplified models (i.e. imitations) of the biological nervous system, and obviously, therefore, have been motivated by the kind of computing performed by the human brain.
- The behavior of a biolgical neural network can be captured by a simple model called artificial neural network.

Analogy between BNN and ANN



Hippocampal Neurons
Source: heart.cbl.utoronto.ca/ ~berj/projects.html

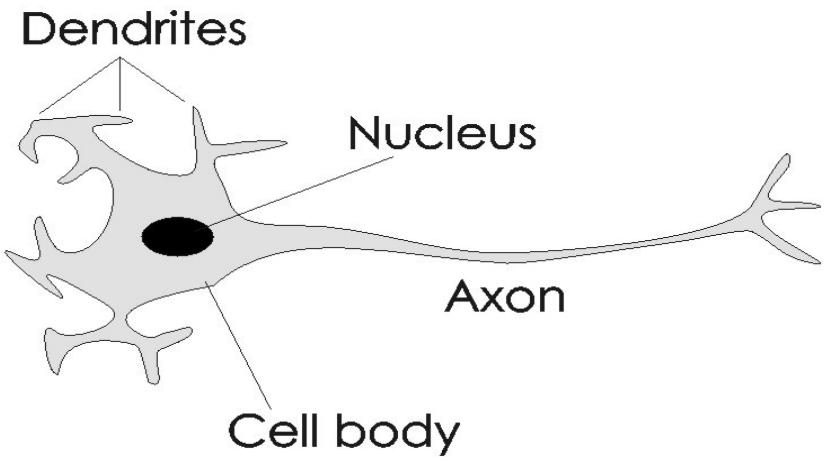
Dendrite
Soma
Axon
Synapse

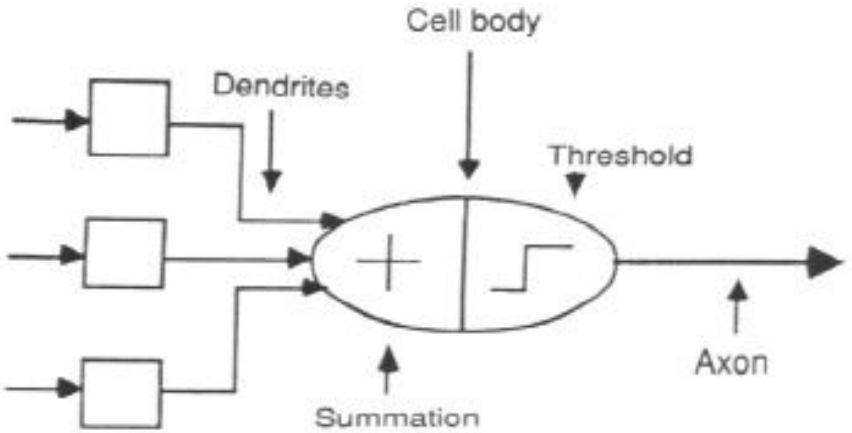


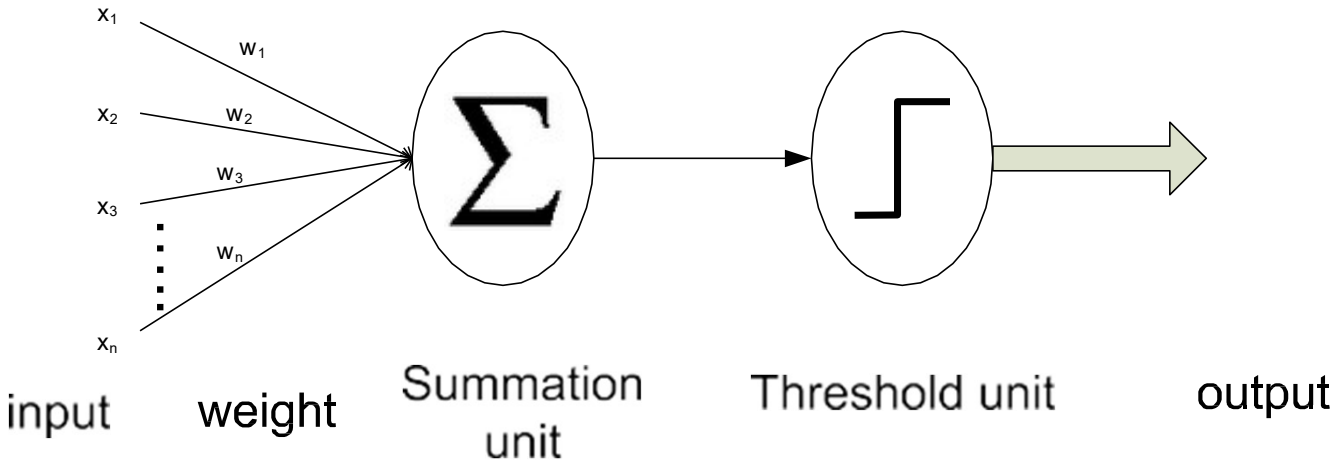
Schematic

Hippocampus: region of the brain that is associated primarily with memory.

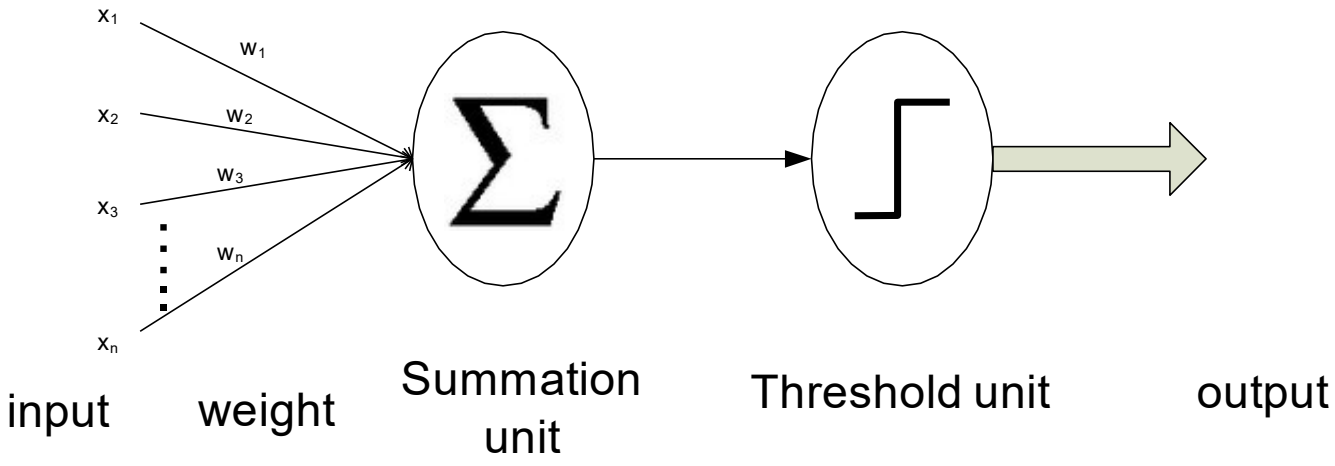
Analogy between BNN and ANN







Artificial neural network



- Here, $x_1, x_2, \cdots, x_n$ are the n inputs to the artificial neuron.
- $w_1, w_2, \cdots, w_n$ are weights attached to the input links.

Artificial neural network

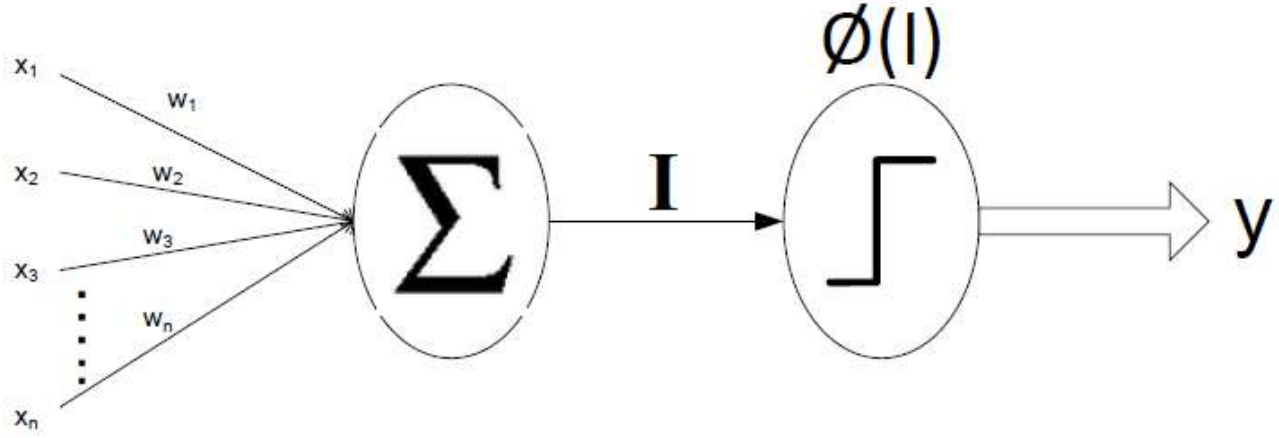
- Note that, a biological neuron receives all inputs through the dendrites, sums them and produces an output if the sum is greater than a threshold value.
- The input signals are passed on to the cell body through the synapse, which may accelerate or retard an arriving signal.
- It is this acceleration or retardation of the input signals that is modeled by the **weights**.
- An effective synapse, which transmits a stronger signal will have a correspondingly larger weights while a weak synapse will have smaller weights.
- Thus, weights here are multiplicative factors of the inputs to account for the strength of the synapse.

Artificial neural network

- Hence, the total input say *I* received by the soma of the artificial neuron is

$$I = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n = \sum_{i=1}^{n} w_i x_i$$

- To generate the final output *y* , the sum is passed to a filter Φ Called transfer function, which releases the output.
- That is, $y = \Phi(I)$



Artificial neural network

- A very commonly known transfer function is the **thresholding function**.
- In this thresholding function, sum (i.e. I) is compared with a threshold value θ.
- If the value of I is greater than θ, then the output is 1 else it is 0 (this is just like a simple linear filter).
- In other words,

$$y = \phi\left(\sum_{i=1}^{n} w_i x_i\right)$$

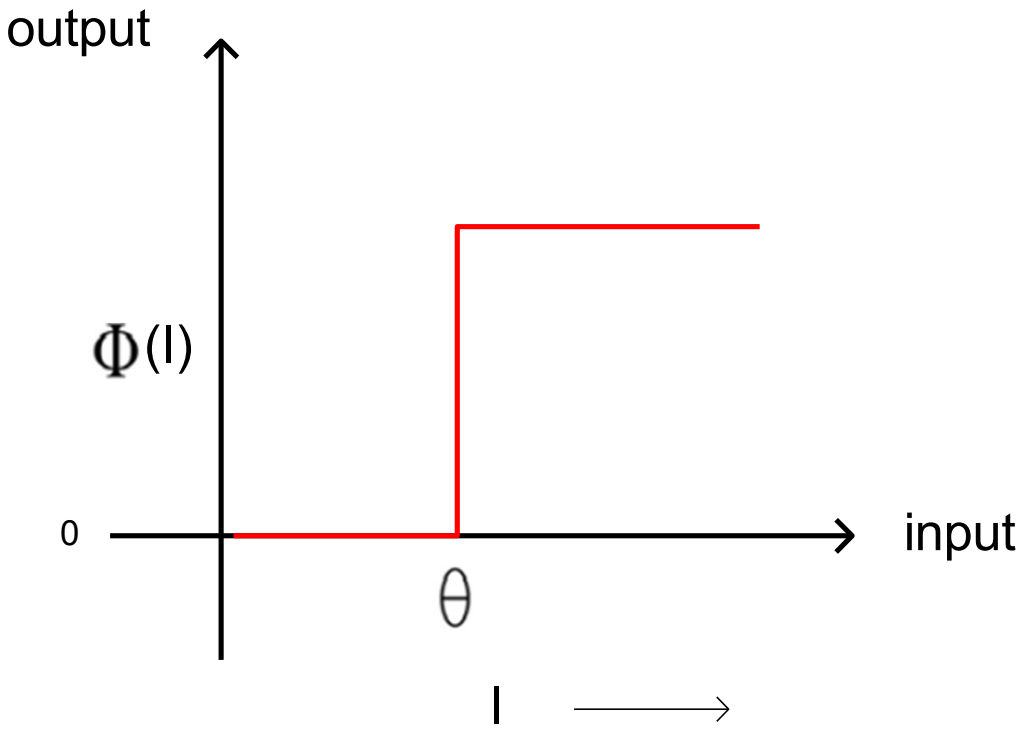
where

$$\phi(I) = \begin{cases} 1 & , \text{if } I > \theta \\ 0 & , \text{if } I \leq \theta \end{cases}$$

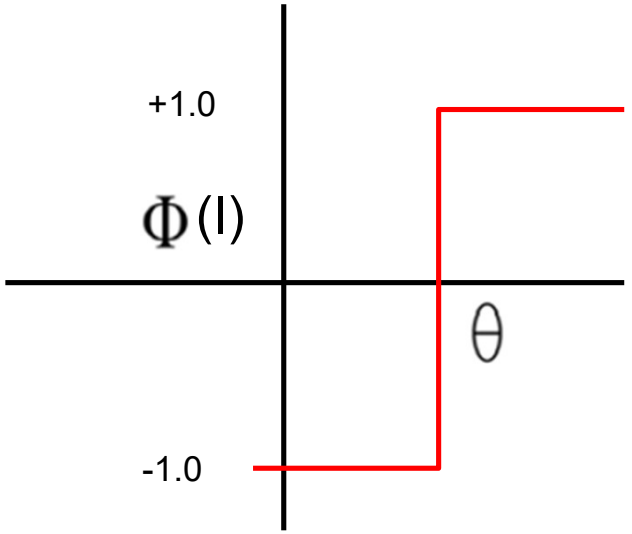
Such a Φ is called step function(also known as Heaviside function).

Artificial neural network

Thresholding functions can be of two types: Binary (1, 0) or Bipolar [-1, 1].



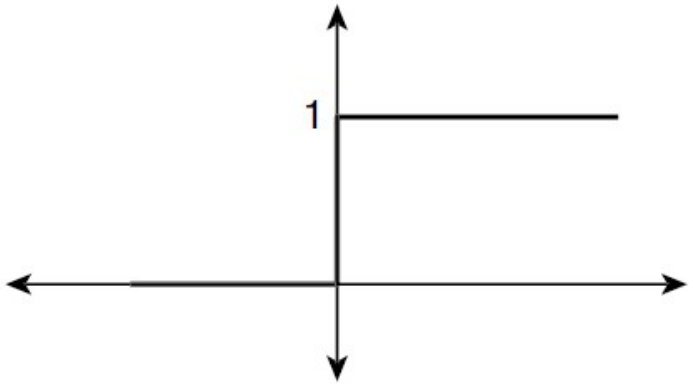
(a) Binary Threshold function



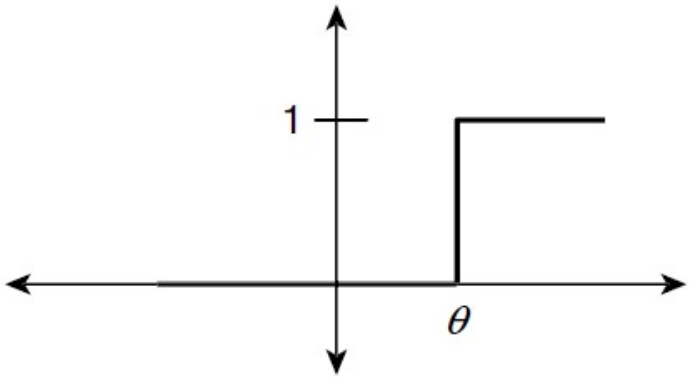
(b) Bipolar Threshold function

Artificial neural network

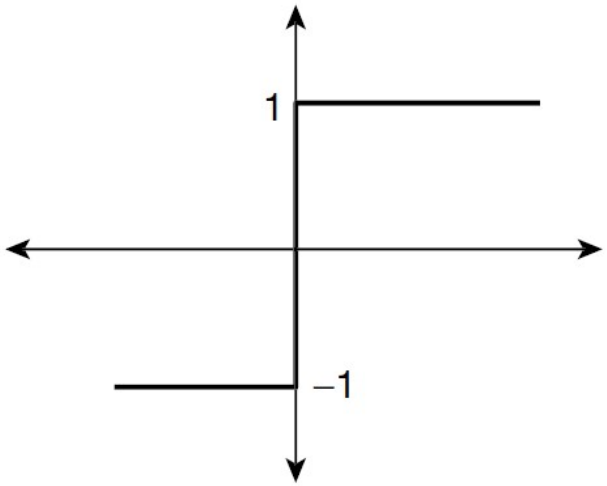
Step vs Thresholding functions



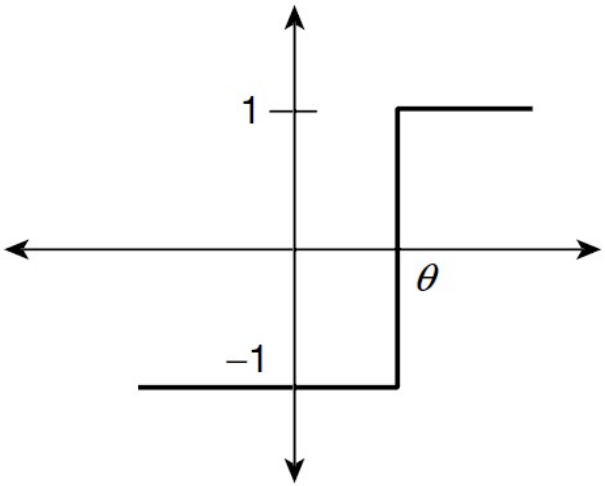
Binary Step



Binary Threshold



Bipolar Step



Bipolar Threshold

Transformation functions

- **Binary transfer function** : The output of the transfer function is lies in the range of *0* to 1. It is generally used in perceptron neuron.

  In other words,

$$\phi(I) = \begin{cases} 1 & , \text{if } I > \theta \\ 0 & , \text{if } I \leq \theta \end{cases}$$

Binary Threshold

- **Bipolar transfer function :** The output of the transfer function is lies in the range of −1.0 to +1.0. It is also known Quantizer function and it defined as

$$\phi(I) = \begin{cases} +1 & , \text{if } I > \theta \\ -1 & , \text{if } I \leq \theta \end{cases}$$

Bipolar Threshold

Transformation functions

Signum transfer function

In Bipolar Threshold function When $\theta = 0$, it is same as $sgn(I)$.

1 if weighted sum (I) is positive.

0 if weighted sum (I) is zero.

-1 if weighted sum (I) is negative.

First use of the Activation functions

The first mathematical model of a biological neuron was presented by McCulloch and Pitts. (1943).

Binary thresholding function were used in the model.

McCulloch-Pitts model does not exhibit any learning but just serves as a basic building block for significant work in NN research.

Question?

How to use Neural Network to learn pattern?

Simple Related Example

Let us predict the student X will pass the exam or not?

Output (y): Pass (1) Fail (0)
Inputs (x): features that influence or don't influence the prediction

study hour (x1)
Name of the student (x2)
Previous Knowledge(x3)
Smartness (x4)

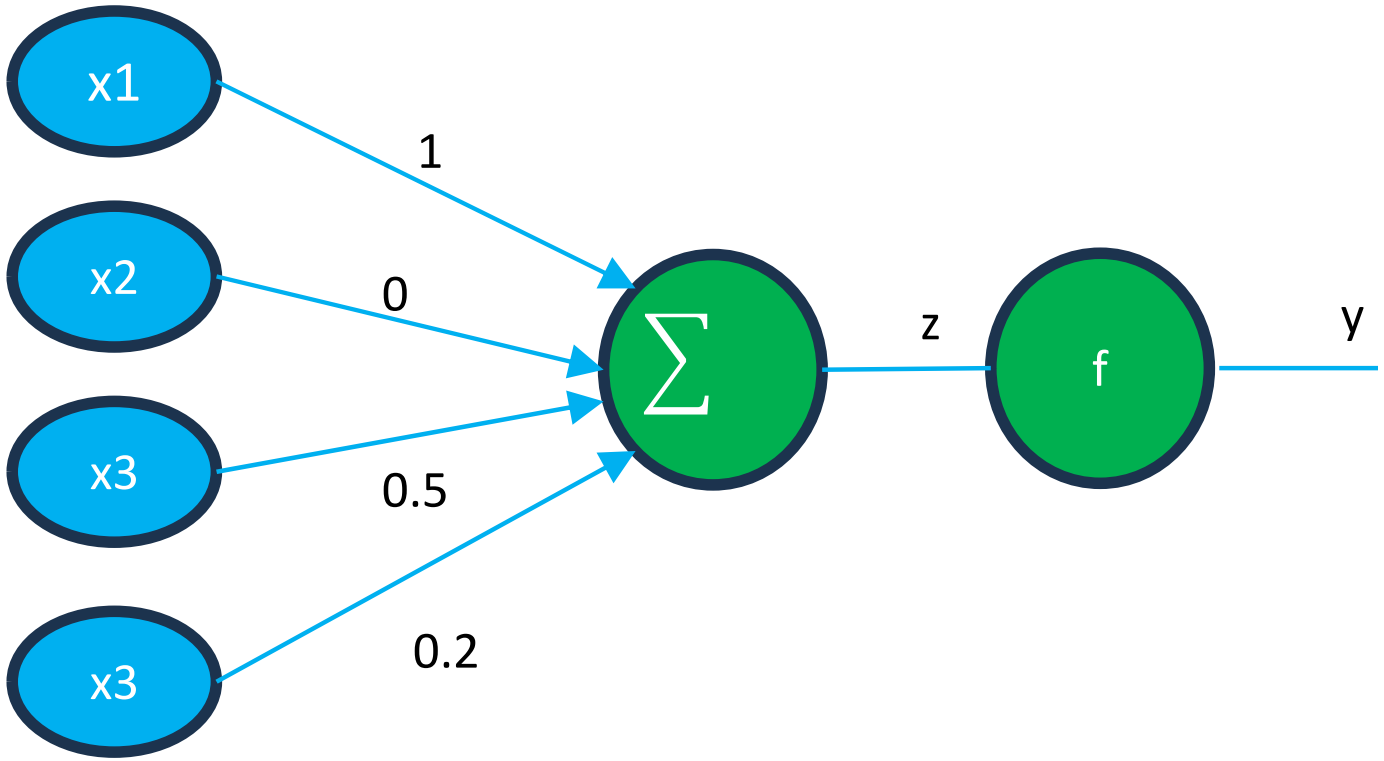
$$y = f(x) = \mathbf{(1)}\, x1 + \mathbf{(0)}\, x2 + \mathbf{(0.5)}\, x3 + \mathbf{(0.2)} x4$$

Given x1, x2, x3, x4 as input, what value shall we multiply, so that we can get our desired output?

Simple Related Example

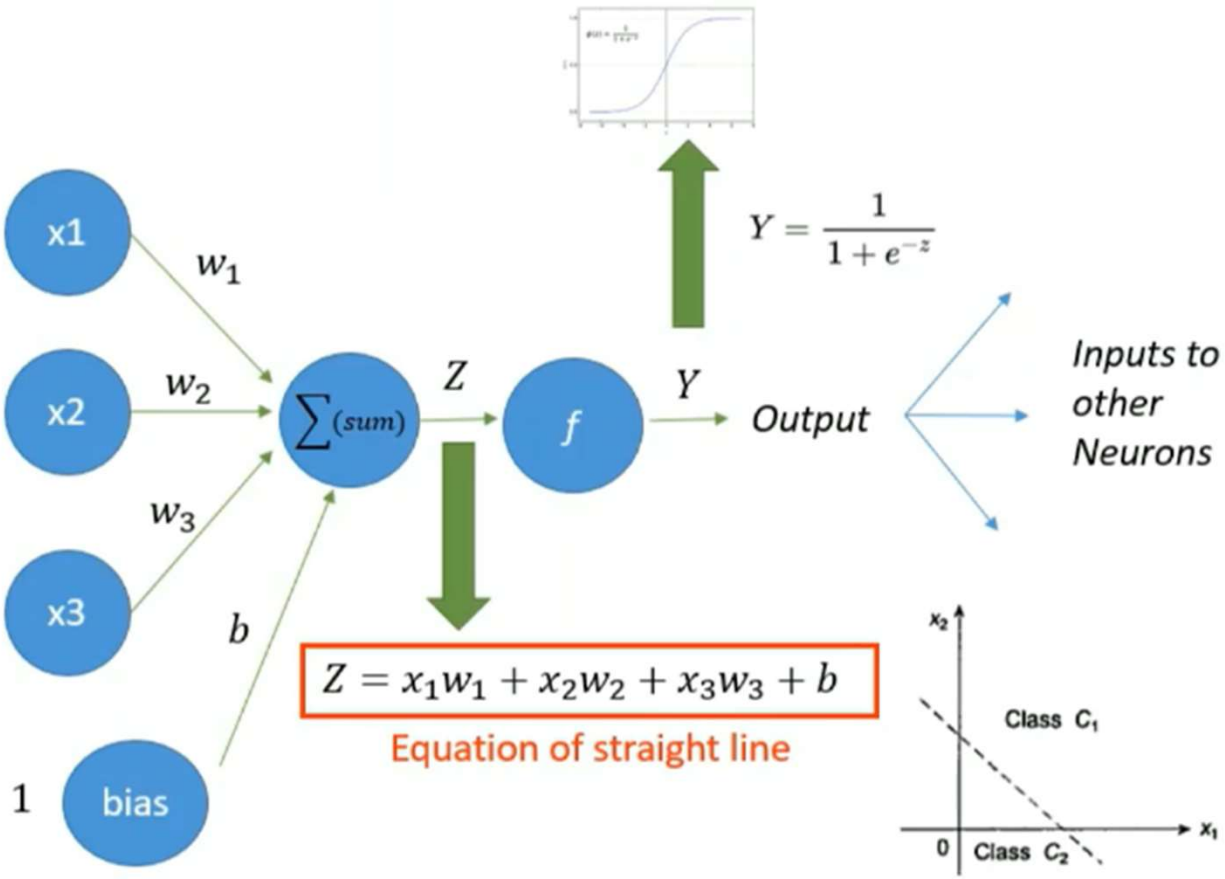
$$y = f(x) = \mathbf{(1)}\ x1 + \mathbf{(0)}\ x2 + \mathbf{(0.5)}\ x3 + \mathbf{(0.2)}\ x4$$

Weights



Note# Weights give importance to the inputs.

Modelling the Brain



$y = mx + c$

A perceptron may be able to separate the inputs by a straight line

Example

Input vector x = [1;2;5;8]

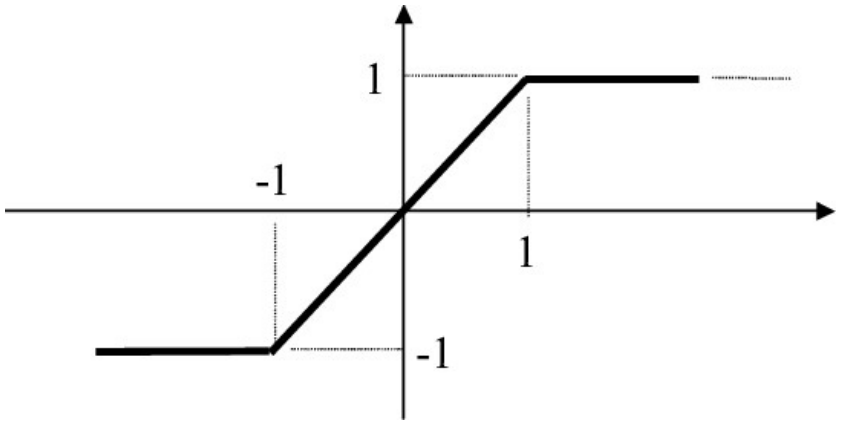
Weight vector w = [1; 1; -1; 2]

Find the output of the neuron if binary step function is used as activation.

Ans: y = 1

Other transformation functions

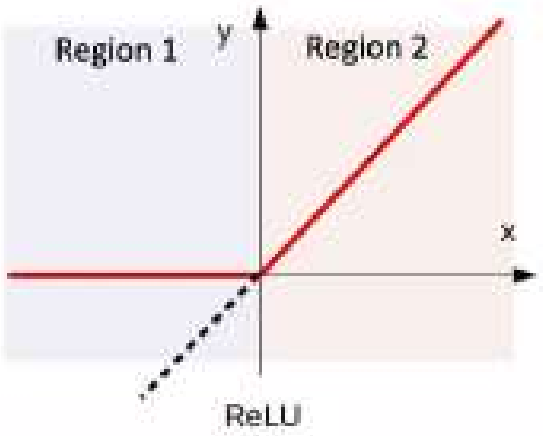
- **Piecewise Linear**: a connected sequence of line segments. This is also called saturating linear function and can have either binary or bipolar.

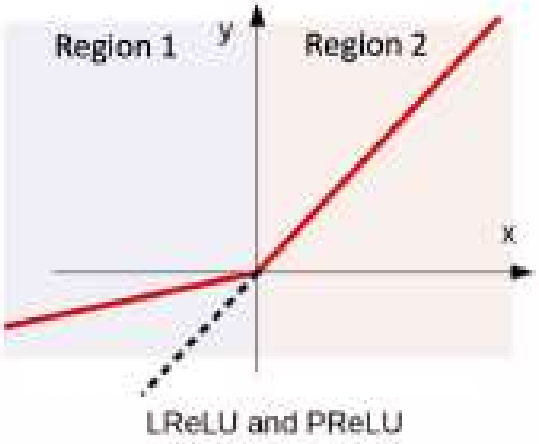


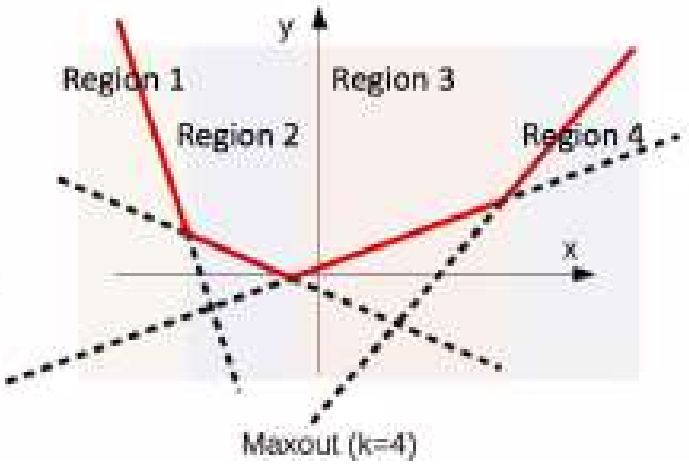
$$\emptyset(I) = \begin{cases} 1 & if\ I > 1 \\ I & if\ -1 \leq I \leq 1 \\ -1 & if\ I < -1 \end{cases}$$

Other transformation functions

- **Piecewise Linear** can have other forms.







Other transformation functions

- **Sigmoid transfer function** : This function is a continuous function that varies gradually between the asymptotic values 0 and 1 (called log-sigmoid) or -1 and +1 (called Tan-sigmoid) function and is given by

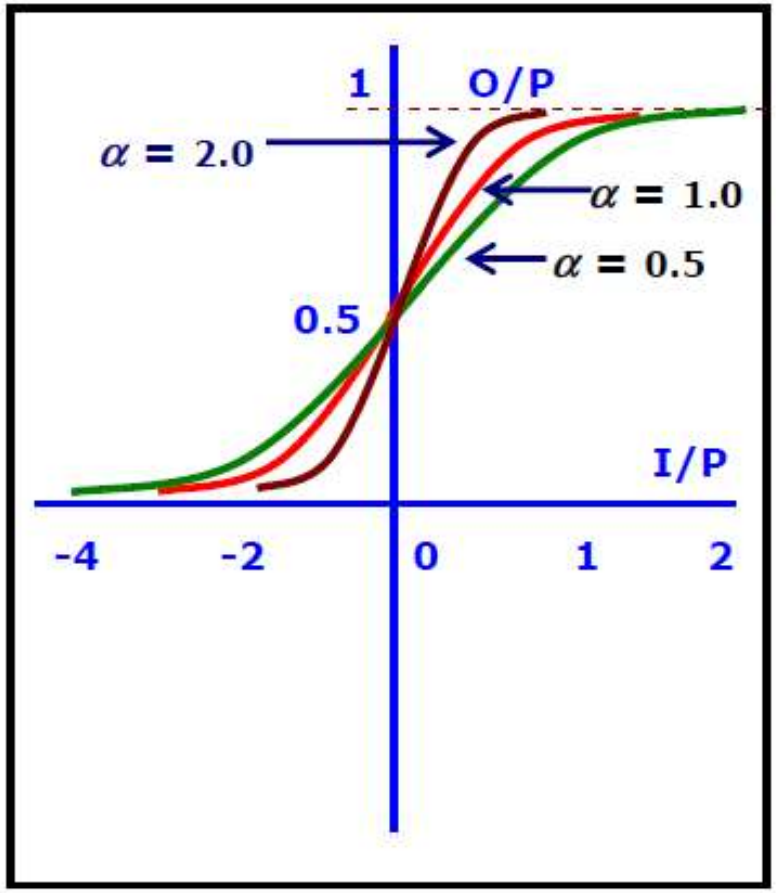
$$\phi(I) = \frac{1}{1+e^{-\alpha I}} \text{ [log-Sigmoid]}$$

$$\phi(I) = tanh(I) = \frac{e^{\alpha I}-e^{-\alpha I}}{e^{\alpha I}+e^{-\alpha I}} \text{ [tan-Sigmoid]}$$

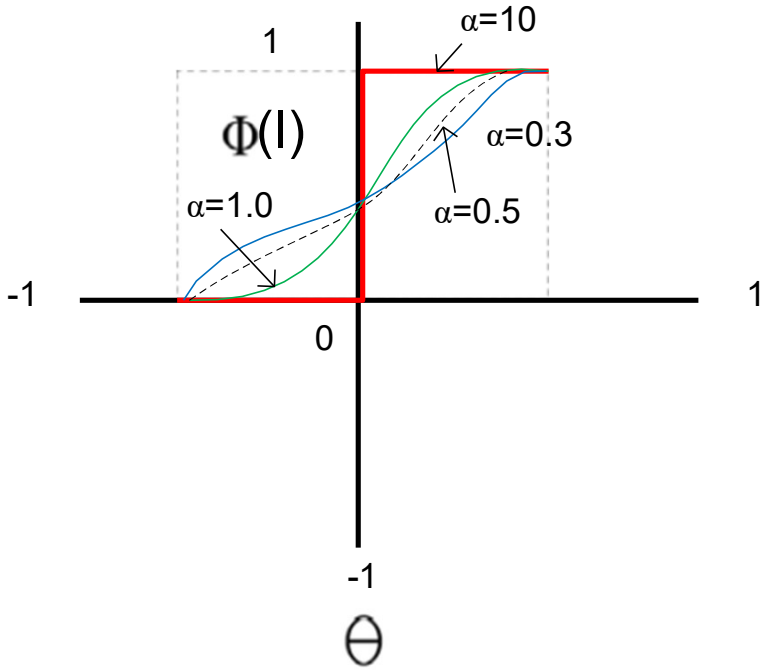
Here, α is the coefficient of transfer function/ Slope parameter.

Other transformation functions

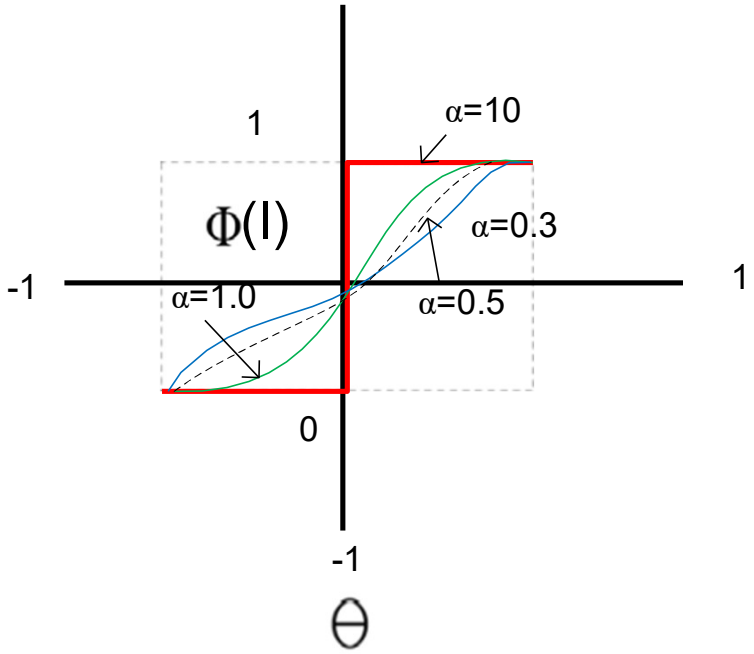
Unipolar Sigmoidal



Transfer functions in ANN

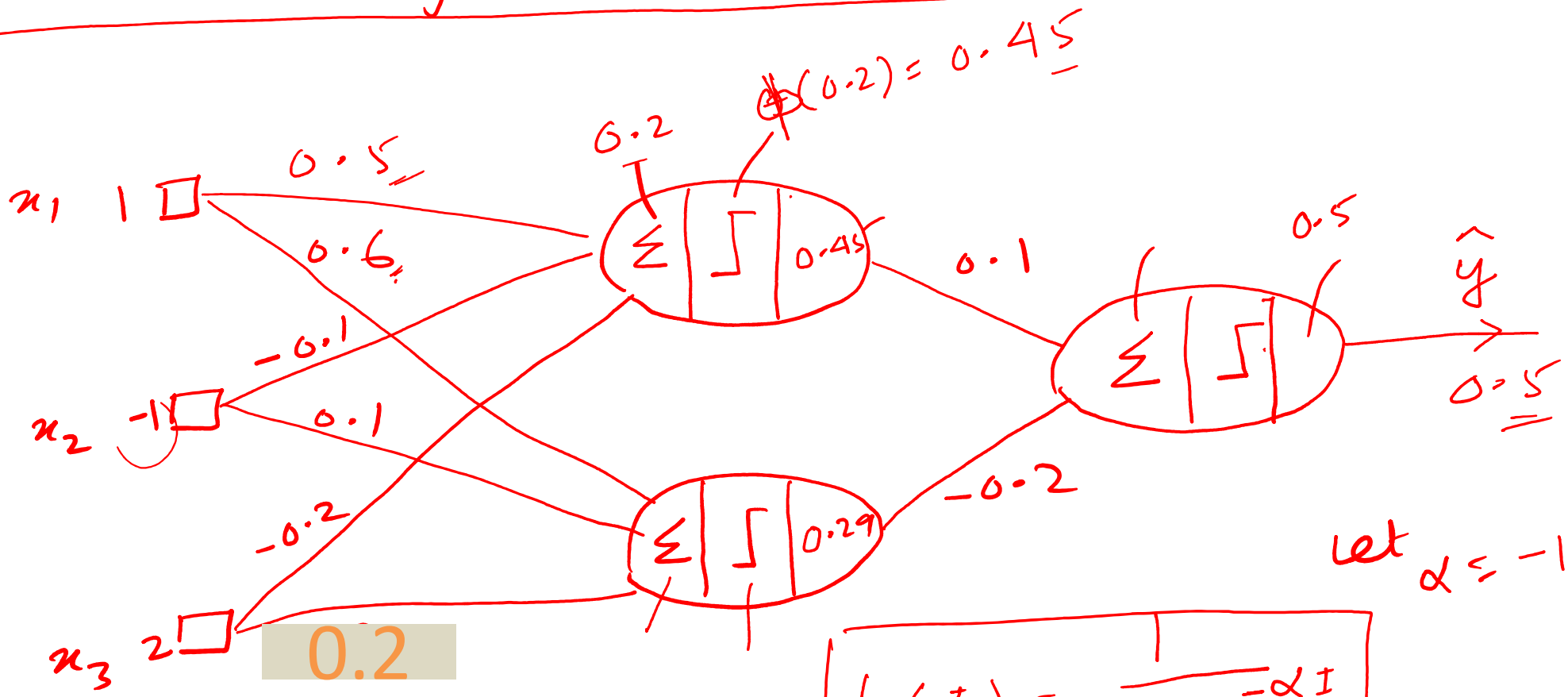


(a) Log-Sigmoid transfer function
(Unipolar Sigmoidal)



(b) Tan-Sigmoid transfer function
(Bipolar Sigmoidal)

Prediction using a particular ANN model



let $\alpha = -1$

$$\phi(I) = \frac{1}{1 + e^{-\alpha I}}$$

predicted $\hat{y} = 0.5$

Actual $y = 1$. Error $= 1 - 0.5 = 0.5$

Advantages of ANN

- ANNs exhibits mapping capabilities, that is, they can map input patterns to their associated output pattern.
- The ANNs learn by examples. Thus, an ANN architecture can be trained with known example of a problem before they are tested for their inference capabilities on unknown instance of the problem. In other words, they can identify new objects previously untrained.
- The ANNs posses the capability to generalize. This is the power to apply in application where exact mathematical model to problem are not possible.

Advantages of ANN

- The ANNs are robust system and fault tolerant. They can therefore, recall full patterns from incomplete, partial or noisy patterns.
- The ANNS can process information in parallel, at high speed and in a distributed manner. Thus a massively parallel distributed processing system made up of highly interconnected (artificial) neural computing elements having ability to learn and acquire knowledge is possible.