Sigmoid Functions and Their Usage in Artificial Neural Networks

Taskin Kocak
School of Electrical Engineering and Computer Science

Applications of Calculus II:

Inverse Functions

Example problem

- Calculus Topic: Inverse functions
- Section 7.6 #19: Prove identity

$$\frac{1+\tanh x}{1-\tanh x} = e^{2x}$$

Example problem (cont.)

$$\frac{1+\tanh x}{1-\tanh x} = \frac{1+?}{1-?}$$

$$\tanh x = ?$$

$a)\ \dfrac{1}{\sinh x}$ $b)\ \dfrac{\cosh x}{\sinh x}$ $c)\ \dfrac{\sinh x}{\cosh x}$ $d)\ \sinh x \cosh x$

Example problem (cont.)

$$\frac{1+\tanh x}{1-\tanh x}=\frac{1+\dfrac{\sinh x}{\cosh x}}{1-\dfrac{\sinh x}{\cosh x}}=\frac{1+\dfrac{\dfrac{e^{x}-e^{-x}}{2}}{\dfrac{e^{x}+e^{-x}}{2}}}{1-\dfrac{\dfrac{e^{x}-e^{-x}}{2}}{\dfrac{e^{x}+e^{-x}}{2}}}$$

Example problem (cont.)

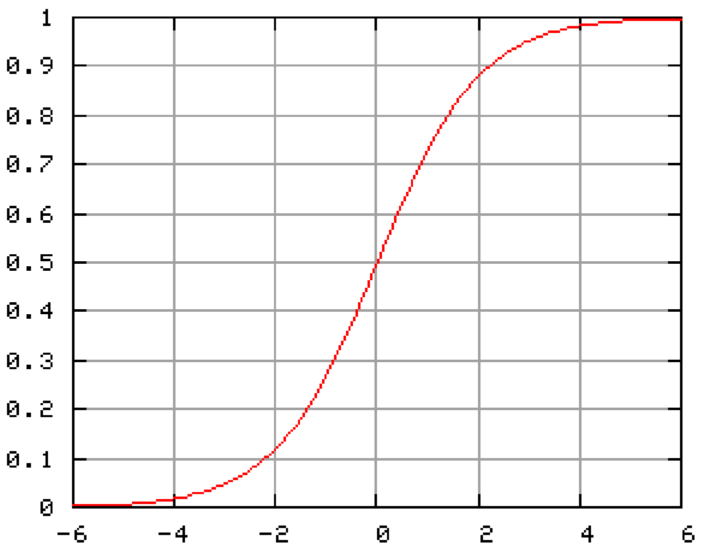
$$\frac{1+\tanh x}{1-\tanh x}=\frac{1+\dfrac{\sinh x}{\cosh x}}{1-\dfrac{\sinh x}{\cosh x}}=\frac{1+\dfrac{\frac{e^{x}-e^{-x}}{2}}{\frac{e^{x}+e^{-x}}{2}}}{1-\dfrac{\frac{e^{x}-e^{-x}}{2}}{\frac{e^{x}+e^{-x}}{2}}}$$

$$=\frac{1+\dfrac{e^{x}-e^{-x}}{e^{x}+e^{-x}}}{1-\dfrac{e^{x}-e^{-x}}{e^{x}+e^{-x}}}=\frac{\dfrac{e^{x}+e^{-x}+e^{x}-e^{-x}}{e^{x}+e^{-x}}}{\dfrac{e^{x}+e^{-x}-e^{x}+e^{-x}}{e^{x}+e^{-x}}}=\frac{2e^{x}}{2e^{-x}}=e^{2x}$$

Learning Objectives

1. Determine the relationships between the biological and artificial neural networks
2. Usage of artificial neural networks – The OR example
3. The classic XOR problem

Sigmoid functions



$$sig(t) = \frac{1}{1 + e^{-t}}$$

A sigmoid function produces a curve with an "S" shape.

The example sigmoid function shown on the left is a special case of the logistic function, which models the growth of some set.

Sigmoid, hyperbolic functions, and neural networks

- In general, a sigmoid function is real-valued and differentiable, having a non-negative or non-positive first derivative, one local minimum, and one local maximum. The logistic sigmoid function is related to the hyperbolic tangent as follows

$$1 - 2sig(x) = 1 - 2\frac{1}{1+e^{-x}} = -\tanh\frac{x}{2}$$

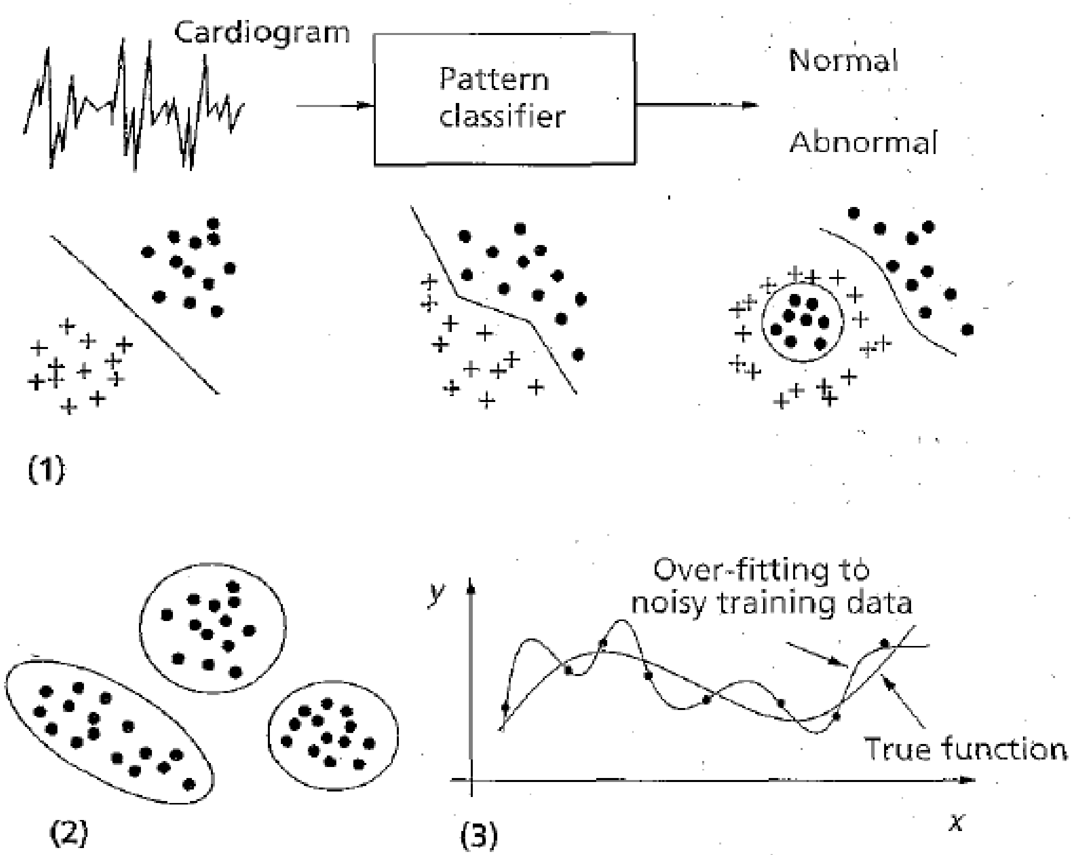
Sigmoid, hyperbolic functions, and neural networks

- Sigmoid functions are often used in artificial neural networks to introduce nonlinearity in the model.
- A neural network element computes a linear combination of its input signals, and applies a sigmoid function to the result. A reason for its popularity in neural networks is because the sigmoid function satisfies a property between the derivative and itself such that it is computationally easy to perform.

$$\frac{d}{dt} sig(t) = sig(t)(1 - sig(t))$$

- Derivatives of the sigmoid function are usually employed in learning algorithms.

Artificial neural networks: Motivation

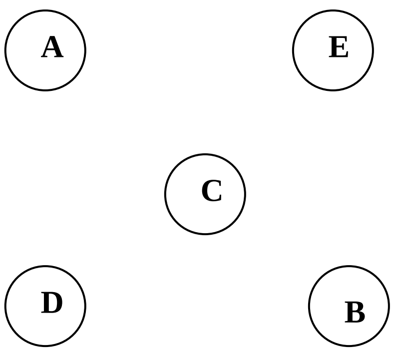


1. Pattern classification
2. Clustering
3. Function approximation

Artificial neural networks: Motivation

Optimization – Traveling salesman problem

Start from city A and visit all the cities. What's the shortest path?



a) A-C-B-D-E-A

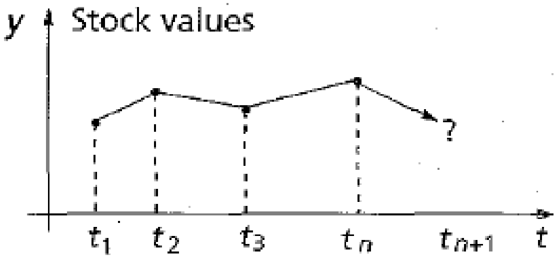
b) A-E-B-D-C-A

c) A-B-C-D-E-A

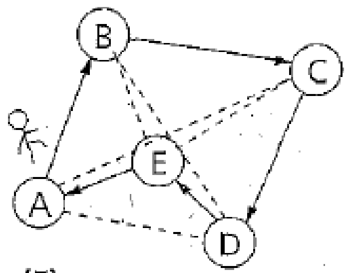
d) A-D-C-E-B-A

e) A-C-E-D-B-A

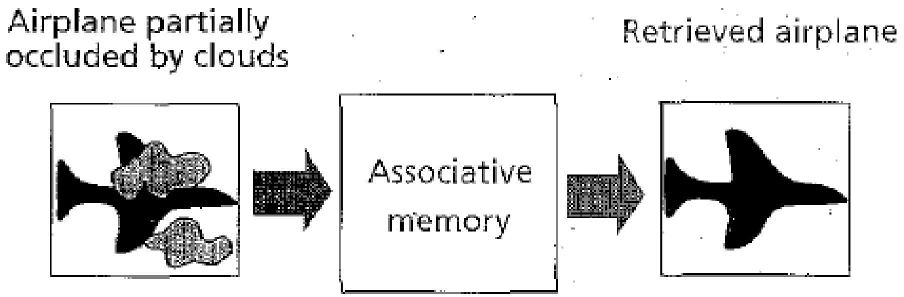
Artificial neural networks: Motivation



(4)



(5)



(6)

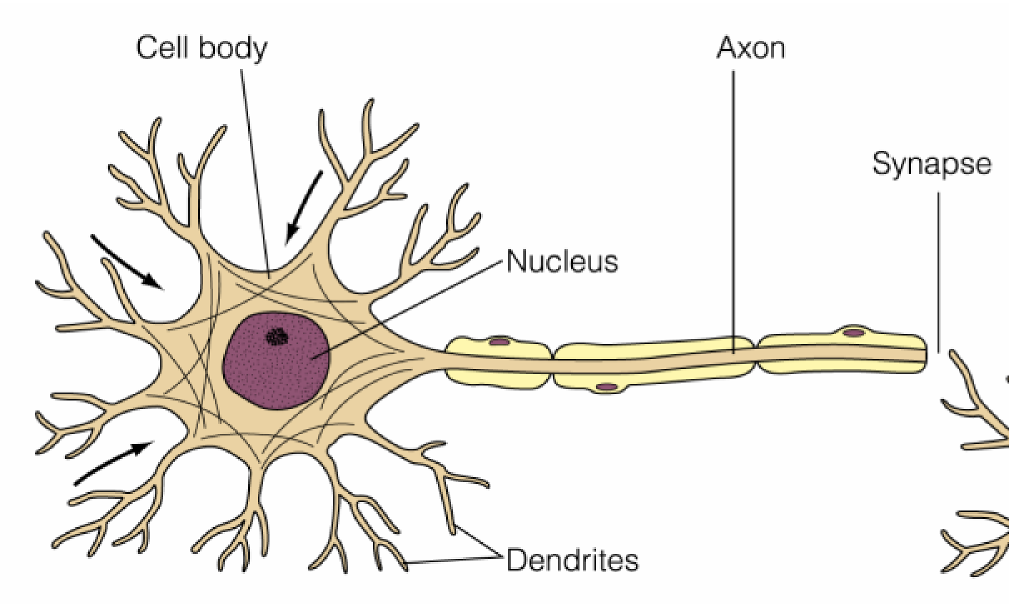
4. Prediction / forecasting

5. Optimization

6. Content addressable memory

Biological neuron

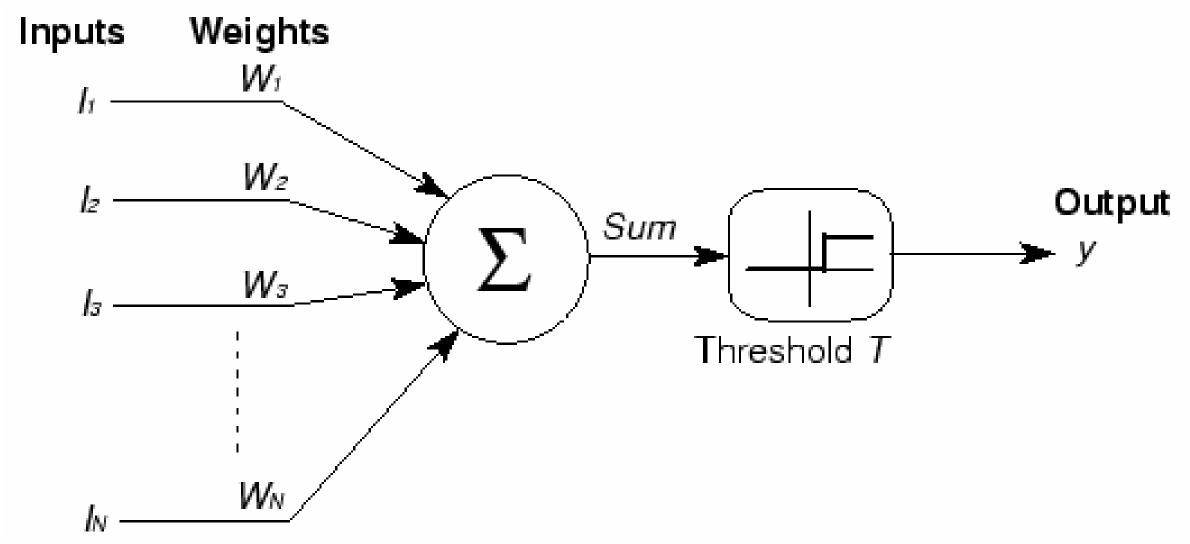
A *neuron* (or nerve cell) is a special biological cell that processes information. It is composed of a cell body, or *soma*, and two types of out-reaching tree-like branches: the *axon* and the *dendrites*.



The cell body has a nucleus that contains information. A neuron receives signals (impulses) from other neurons through its dendrites (receivers) and transmits signals generated by its cell body along the axon (transmitter).

Artificial neural networks

Inspired by biological neural networks, artificial neural networks are massive parallel computing systems consisting of an extremely large number of simple processors with many interconnections. McCulloch and Pitts proposed a binary threshold unit as a computational model for an artificial neuron.

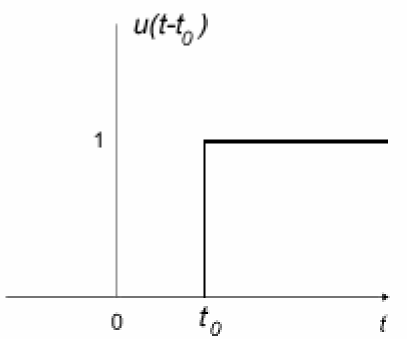


Activation function - Unit step function

- Activation function: A function used to transform the activation level of neuron (weighted sum of inputs) to an output signal.

$$y = \Theta\left(\sum_{j=1}^{N} w_j I_j - T\right)$$

- Unit step function is one of the activation functions



$$u(t-t_0) = \begin{cases} 0 & if\ t < t_0 \\ 1 & if\ t > t_0 \end{cases}$$

Assessment of Learning Objective #1

Similarities between the biological and artificial neural networks

Connection weights in ANN represent which biological structure?

a) Cell body b) Nucleus c) Axon d) Dendrites e) Synapse

Usage of artificial neural networks – The OR example

- we will utilize the McCulloch-Pitt model to train a neural network to learn the logic OR function. The OR function we will use is a two-input binary OR function given in Table 1.

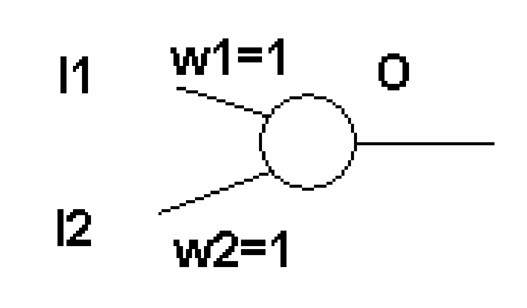
Table 1: OR function

I_1	I_2	Output
0	0	0
0	1	?
1	0	?
1	1	1

a) 0   b) 1

The OR example

First, we will use one neuron with two inputs. Note that the inputs are given equal weights by assigning the weights (w's) to '1'. The threshold, *T*, is set to 0 in this example.



We calculate the output as follows:

1) Compute the total weighted inputs

$$X = \sum_{i=1}^{2} I_i w_i$$

$$X = I_1 w_1 + I_2 w_2 = I_1 1 + I_2 1 = I_1 + I_2$$

The OR example

2) Calculate the output using the logistic sigmoid activation function

$$O = sig(X - T) = sig(X) = \frac{1}{1 + e^{-X}}$$

Now, let's try it for the inputs given in Table 1.

For $I_1=0$ and $I_2=0$; $X=0$,

$$O = \frac{1}{1 + e^{0}} = \frac{1}{1+1} = 0.5$$

The OR example (cont.)

For I_1=0 and I_2=1, and I_1=1 and I_2=0; X=1,

$$O = \frac{1}{1+e^{-1}} = \frac{1}{1+0.37} \cong 0.73$$

For I_1=1 and I_2=1; X=2,

$$O = \frac{1}{1+e^{-2}} = \frac{1}{1+0.14} \cong 0.88$$

- For all cases the results match with Table 1 assuming that ‘0.5’ and below are considered as ‘0’ and above as ‘1’.

Assessment of Learning Objective #2

1. (Two minute discussion) If the weights were 0.5 rather than 1, will the network still function like OR?

a) Yes b) No

Assessment of Learning Objective #2

2. (5-minute paper) In groups of two students, discuss whether the same network can be used to learn the AND function? (Hint: You may change the threshold(=0.5) if necessary)

Table 2: AND function

I_1	I_2	Output
0	0	0
0	1	?
1	0	?
1	1	1

a) 0 b) 1

The classic XOR problem

Table 3: XOR function

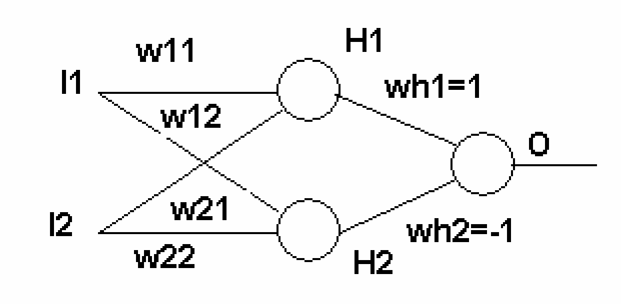
I_1	I_2	Output
0	0	0
0	1	1
1	0	1
1	1	0

If we use the same one-neuron model to learn the XOR (exclusive or) function, the model will fail.

The first three cases will produce correct results; however, the last case will produce ‘1’, which is not correct.

The classic XOR problem (cont.)

The solution is to add a middle (hidden in ANN terminology) layer between the inputs and the output neuron



Choose the weights w11=w12=w21=w22=1. Use a different sigmoid function, which is given with a certain threshold for each neuron:

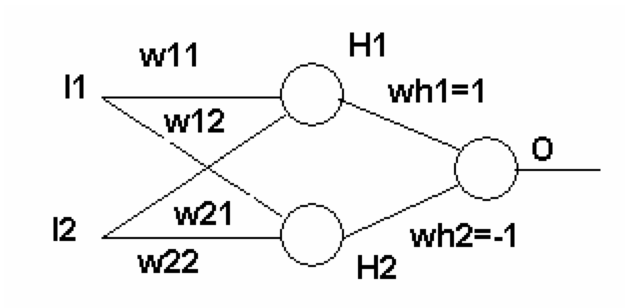
$$sig_{H1}(x) = \frac{1}{1+e^{-(x-0.5)}}$$

$$sig_{H2}(x) = \frac{1}{1+e^{-(x-1.5)}}$$

$$sig_{O}(x) = \frac{1}{1+e^{-(x-0.2)}}$$

Confirm by calculating the neuron outputs for each possible input combinations that this neural network is indeed functioning like an XOR. (Hint: The output equal or below 0.5 is considered ‘0’, otherwise ‘1’)

Neuron calculation



I1	I2	XOR	X	H1	H2	O	Out
0	0	0	0				
0	1	1	1				
1	0	1	1				
1	1	0	2				

w11=w12=w21=w22=1

$$sig_{H1}(x) = \frac{1}{1+e^{-(x-0.5)}}$$

$$sig_{H2}(x) = \frac{1}{1+e^{-(x-1.5)}}$$

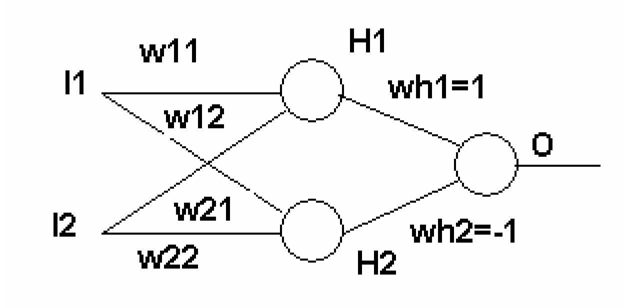
$$sig_{O}(x) = \frac{1}{1+e^{-(x-0.2)}}$$

$$sig_{H1}(0) = \frac{1}{1+e^{-(0-0.5)}} = 0.3775$$

$$sig_{H1}(1) = \frac{1}{1+e^{-(1-0.5)}} = 0.6225$$

$$sig_{H1}(2) = \frac{1}{1+e^{-(2-0.5)}} = 0.8176$$

Neuron calculation (2)



I1	I2	XOR	X	H1	H2	O	Out
0	0	0	0	0.3775			
0	1	1	1	0.6225			
1	0	1	1	0.6225			
1	1	0	2	0.8176			

w11=w12=w21=w22=1

$$sig_{H1}(x) = \frac{1}{1+e^{-(x-0.5)}}$$

$$sig_{H2}(x) = \frac{1}{1+e^{-(x-1.5)}}$$

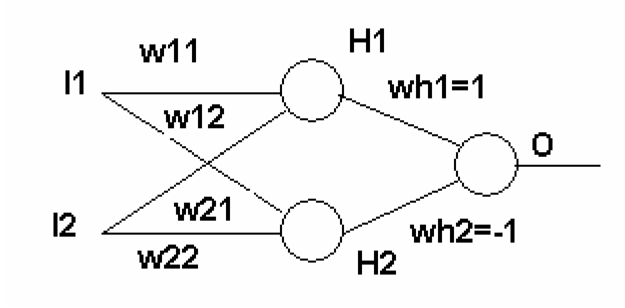
$$sig_{O}(x) = \frac{1}{1+e^{-(x-0.2)}}$$

$$sig_{H2}(0) = \frac{1}{1+e^{-(0-1.5)}} = 0.1824$$

$$sig_{H2}(1) = \frac{1}{1+e^{-(1-1.5)}} = 0.3775$$

$$sig_{H2}(2) = \frac{1}{1+e^{-(2-1.5)}} = 0.6225$$

Neuron calculation (3)



I1	I2	XOR	X	H1	H2	O	Out
0	0	0	0	0.3775	0.1824		
0	1	1	1	0.6225	0.3775		
1	0	1	1	0.6225	0.3775		
1	1	0	2	0.8176	0.6225		

w11=w12=w21=w22=1

$$sig_{H1}(x) = \frac{1}{1+e^{-(x-0.5)}}$$

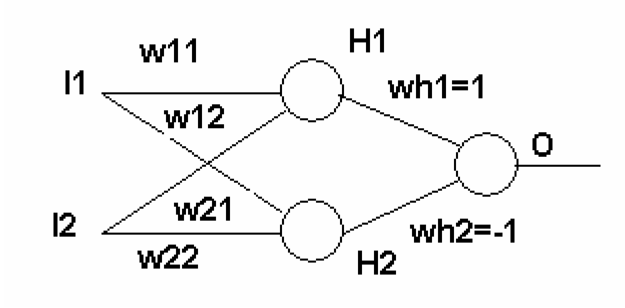
$$sig_{H2}(x) = \frac{1}{1+e^{-(x-1.5)}}$$

$$sig_{O}(x) = \frac{1}{1+e^{-(x-0.2)}}$$

$$sig_O(H2-H1) = sig_O(0.1951) = \frac{1}{1+e^{-(0.1951-0.2)}} = 0.4988$$

$$sig_O(H2-H1) = sig_O(0.2450) = \frac{1}{1+e^{-(0.2450-0.2)}} = 0.5112$$

Neuron calculation (4)



I1	I2	XOR	X	H1	H2	O	Out
0	0	0	0	0.3775	0.1824	0.4988	**0**
0	1	1	1	0.6225	0.3775	0.5112	**1**
1	0	1	1	0.6225	0.3775	0.5112	**1**
1	1	0	2	0.8176	0.6225	0.4988	**0**

w11=w12=w21=w22=1

$$sig_{H1}(x) = \frac{1}{1+e^{-(x-0.5)}}$$

$$sig_{H2}(x) = \frac{1}{1+e^{-(x-1.5)}}$$

$$sig_{O}(x) = \frac{1}{1+e^{-(x-0.2)}}$$

Assuming that ‘0.5’ and below are considered as ‘0’ and above as ‘1’.

Example applications

http://www.heatonresearch.com/articles/42/page1.html

http://www.williewheeler.com/software/bnn.html

http://staff.aist.go.jp/utsugi-a/Lab/BSOM1/index.html