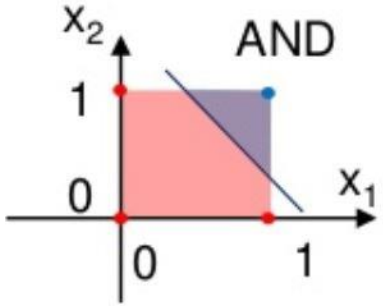
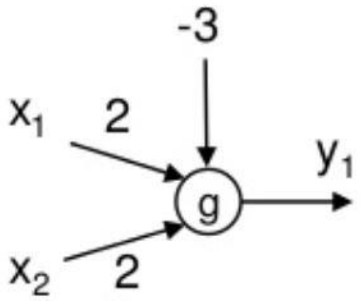
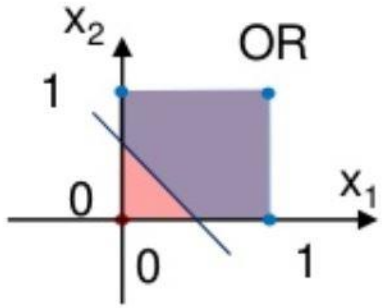

language
evaluation
NLP
learning
systems
text
words
rules
data
tasks
translation
recognition
often
languages
information
system
speech
statistical
machine
algorithms
task
input
based
segmentation
processing
human
Given
example
edit
natural
research
e.g
models
sentence
set
possible
also
many
standard
one
especially
corpora
different
problem
determine
chunk
difficult
separate
word
sentences
typically
named
English
considered
real-world
make
Note
written
entity
developed
generation
criteria
ambiguity
section
However
sources
mainly
extraction
specific
people
tagger
generally
morphology
disambiguation
resolution
etc
gold
parsers
compared
meaning
Automatic
type
refer
semantics
vs
project
result
conducted
automatic
semantic
intelligence
general
large
names
common
door
may
procedures
including
compare
French
number
quality
within
machine-learning
referred
output
grammar
automatically
however
corresponding
Many
understanding
related
use
results
complexity
world
i.e
although
hand-written
complex
using
much
identifying
POS
devoted
termed
reliable
given
probabilistic
usually
decisions
entities
problems
campaign
focused
concerned
1980s
serve
sometimes
multiple
made
identify
knowledge
larger
Natural
analysis
process
IR

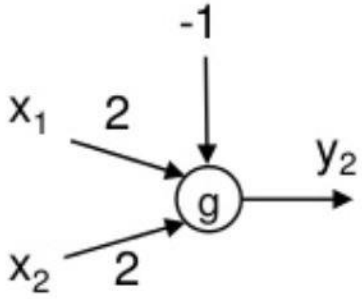
Neural Network





Input vector (x1,x2)	Class AND
(0,0)	0
(0,1)	0
(1,0)	0
(1,1)	1

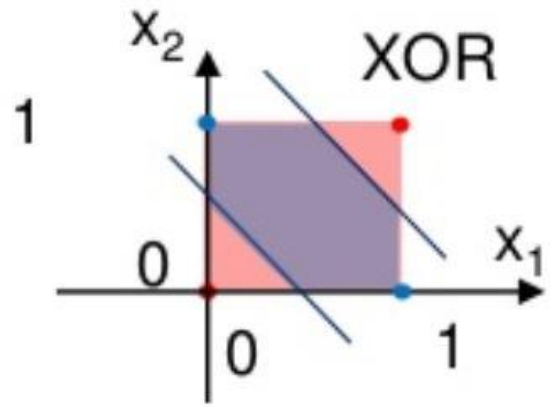




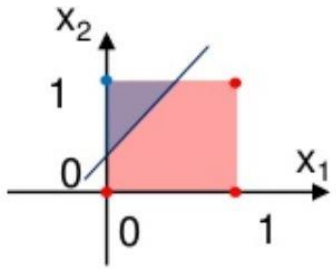
Input vector (x1,x2)	Class OR
(0,0)	0
(0,1)	1
(1,0)	1
(1,1)	1

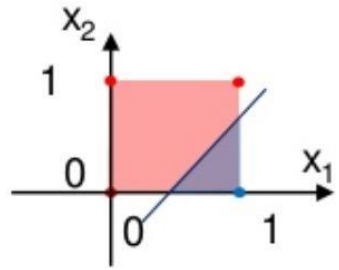
$$y_1 = g(\mathbf{w}^T\mathbf{x} + b) = u((2 \quad 2) \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - 3)$$

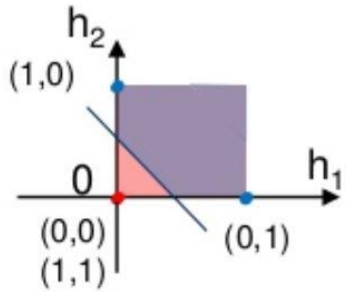
$$y_2 = g(\mathbf{w}^T\mathbf{x} + b) = u((2 \quad 2) \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - 1)$$

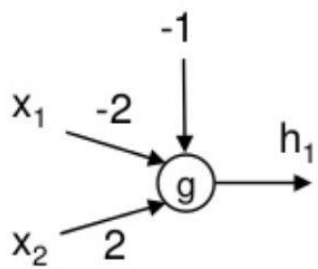


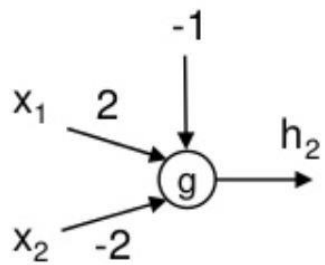
Input vector (x1,x2)	Class XOR
(0,0)	0
(0,1)	1
(1,0)	1
(1,1)	0

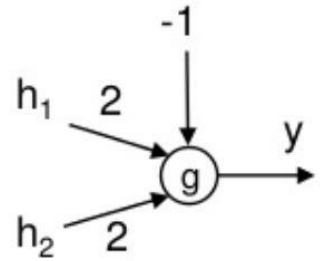








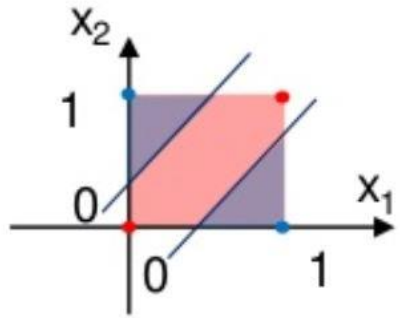


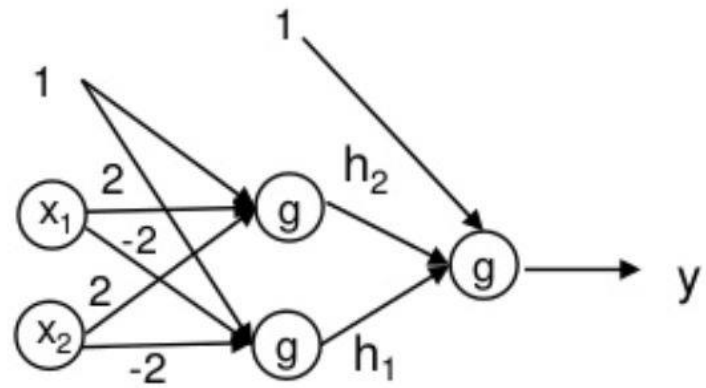


$$h_1 = g\left(\mathbf{w_{11}^T}\,\mathbf{x} + b_{11}\right) = u\left((-2 \quad 2) \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - 1\right)$$

$$h_2 = g\left(\mathbf{w_{12}^T} \quad \mathbf{x} + b_{12}\right) = u\left((2 \quad -2) \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + 1\right)$$

$$y = g\left(\mathbf{w_2^T} \quad \mathbf{h} + b_2\right) = u\left((2 \quad -2) \cdot \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} + 1\right)$$





Input layer | Hidden layer | Output Layer

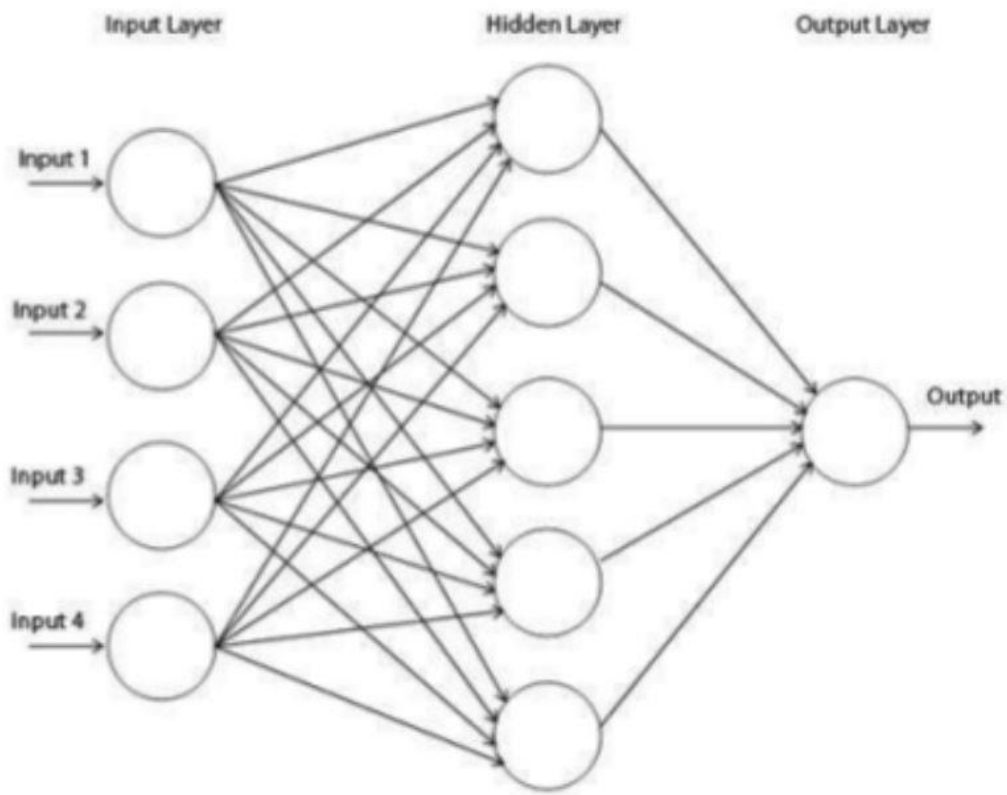
$$h_1 = g\left(\mathbf{w_{11}^T}\,\mathbf{x} + b_{11}\right) = u\left((-2 \quad 2) \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - 1\right)$$

$$h_2 = g\left(\mathbf{w_{12}^T} \quad \mathbf{x} + b_{12}\right) = u\left((2 \quad -2) \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + 1\right)$$

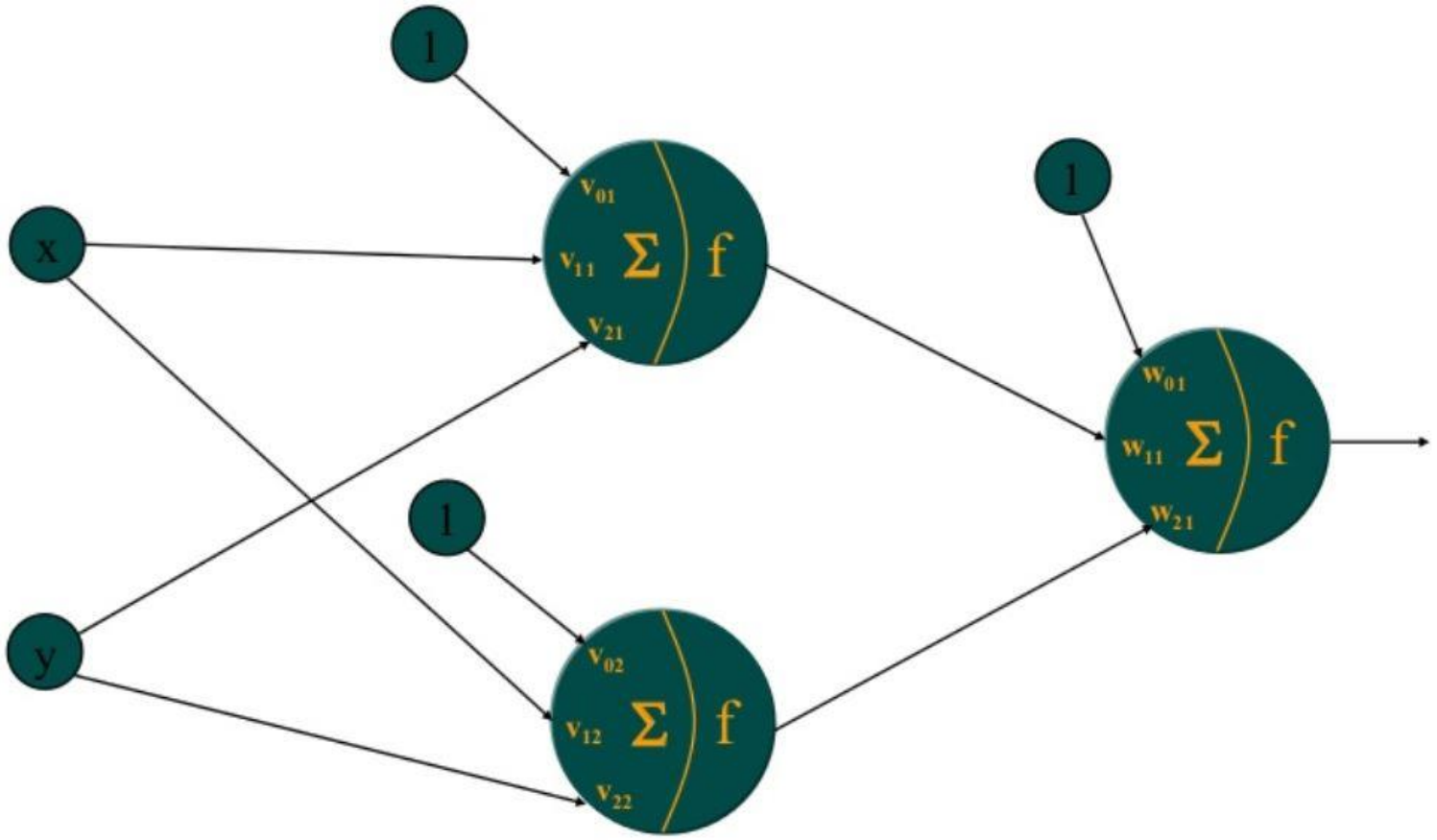
$$y = g\left(\mathbf{w_2^T} \quad \mathbf{h} + b_2\right) = u\left((2 \quad -2) \cdot \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} + 1\right)$$

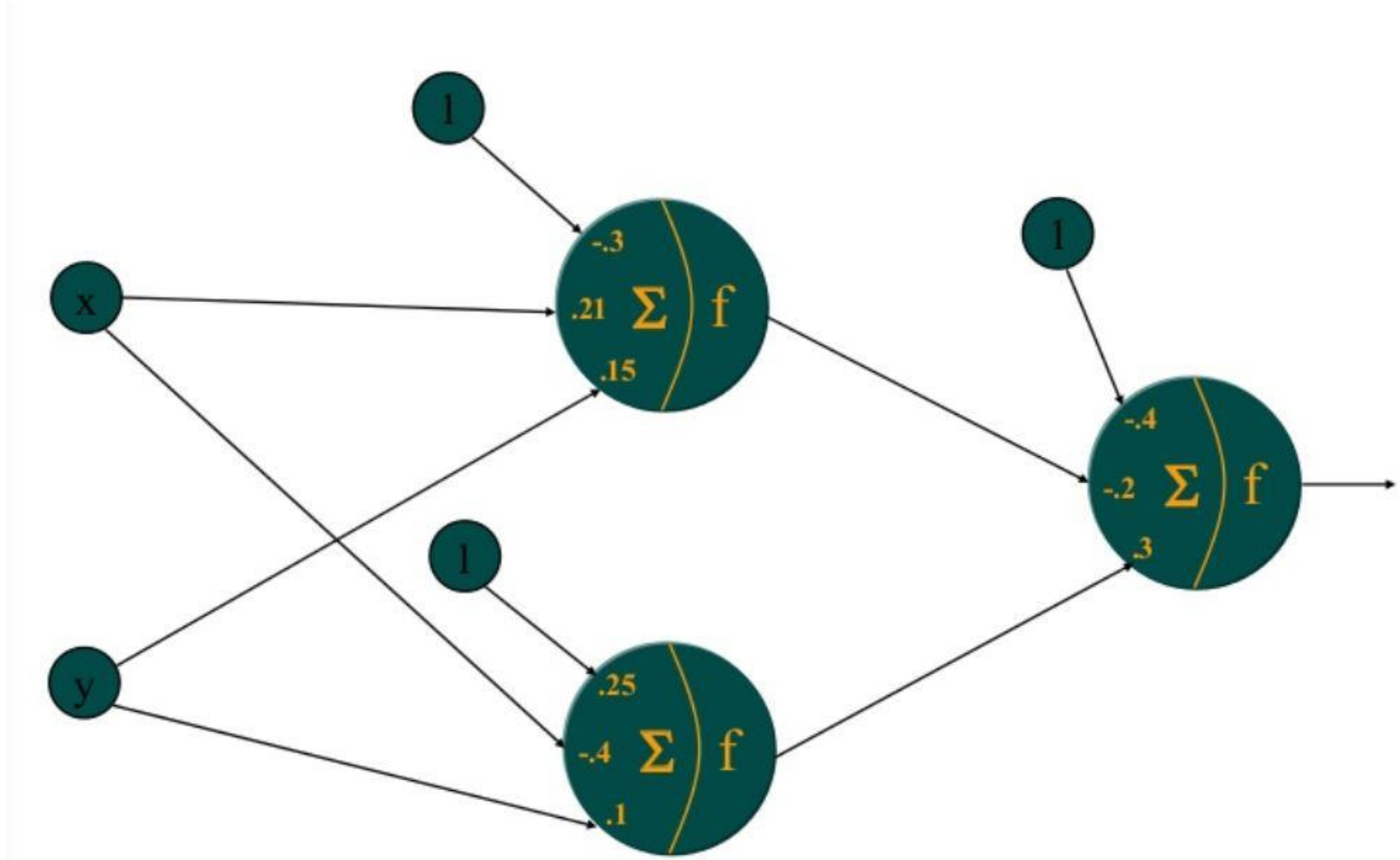
Neural Networks

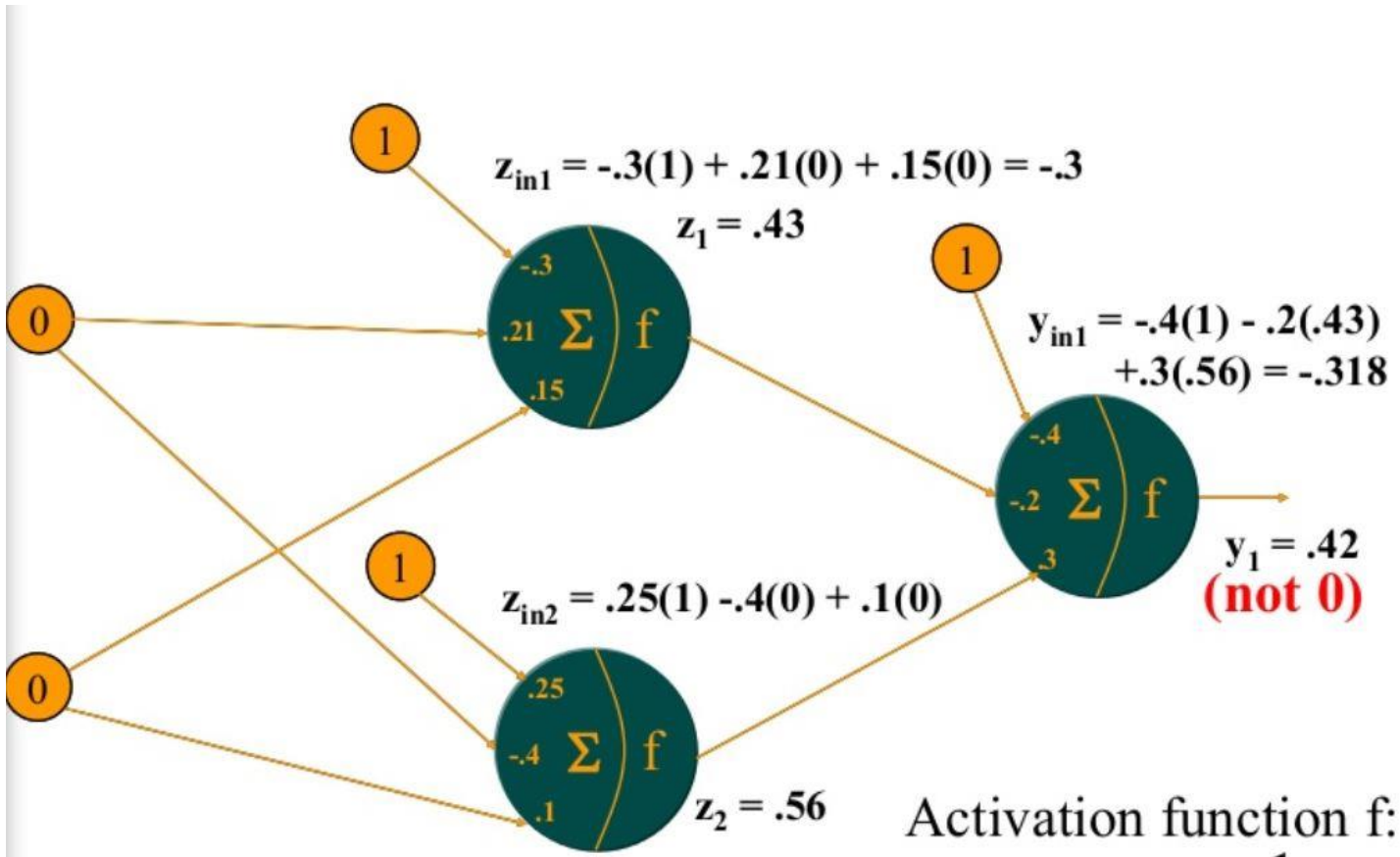
- A neural network is simply a composition of simple neurons into several layers.
- Each neuron simply computes a linear combination of its inputs, adds a bias and passes the result through an activation function g(x).
- The network can contain one or more hidden layers. The outputs of these hidden layers can be thoght of as a new representation of the data (new features).
- The final output is the target variable (y=f(x))


Input Layer
Hidden Layer
Output Layer
Input 1
Input 2
Input 3
Input 4
Output

Example:



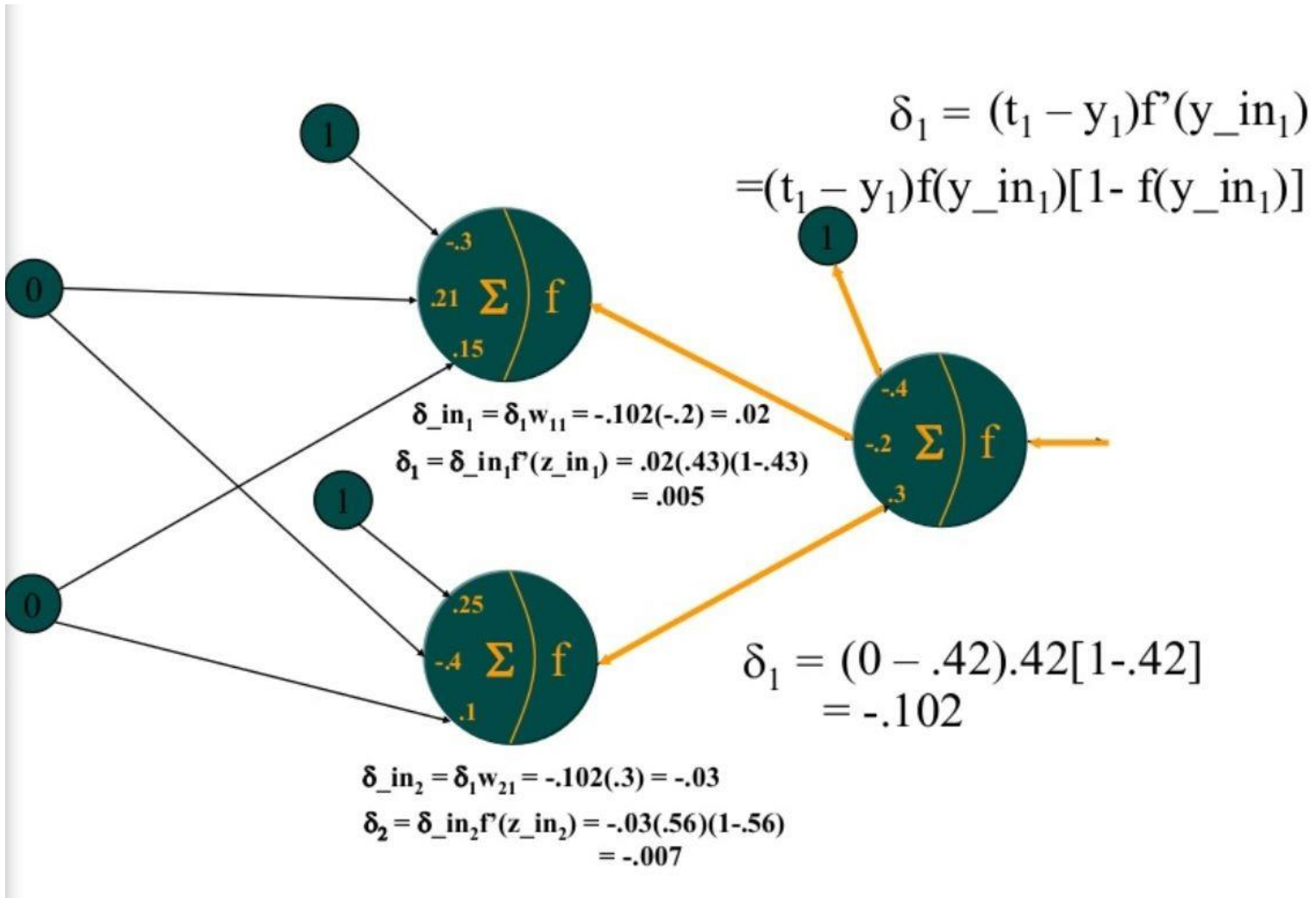

1
x
-.3
.21
Σ
f
.15
1
-.4
-.2
Σ
f
.3
1
y
.25
-.4
Σ
f
.1

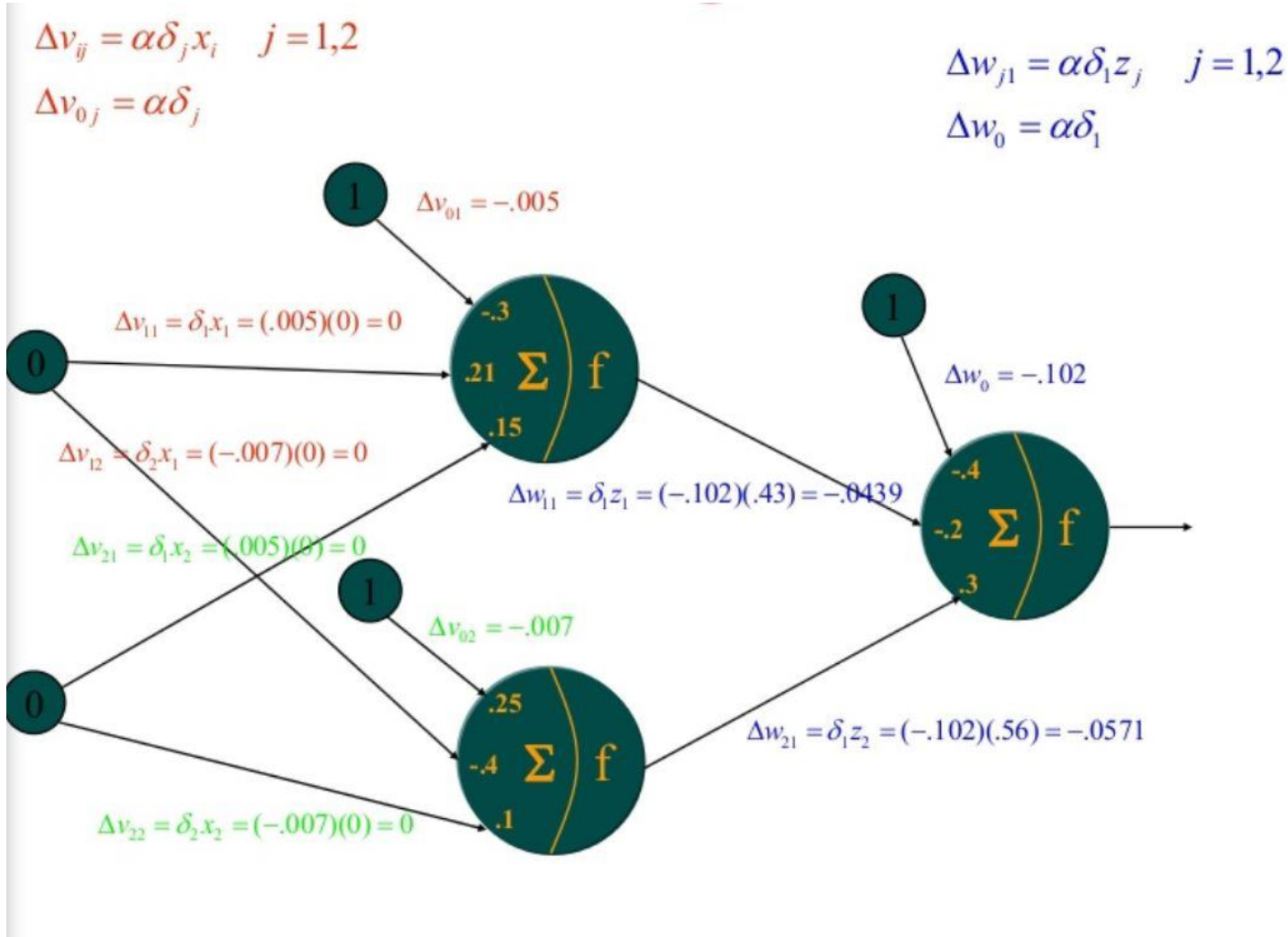


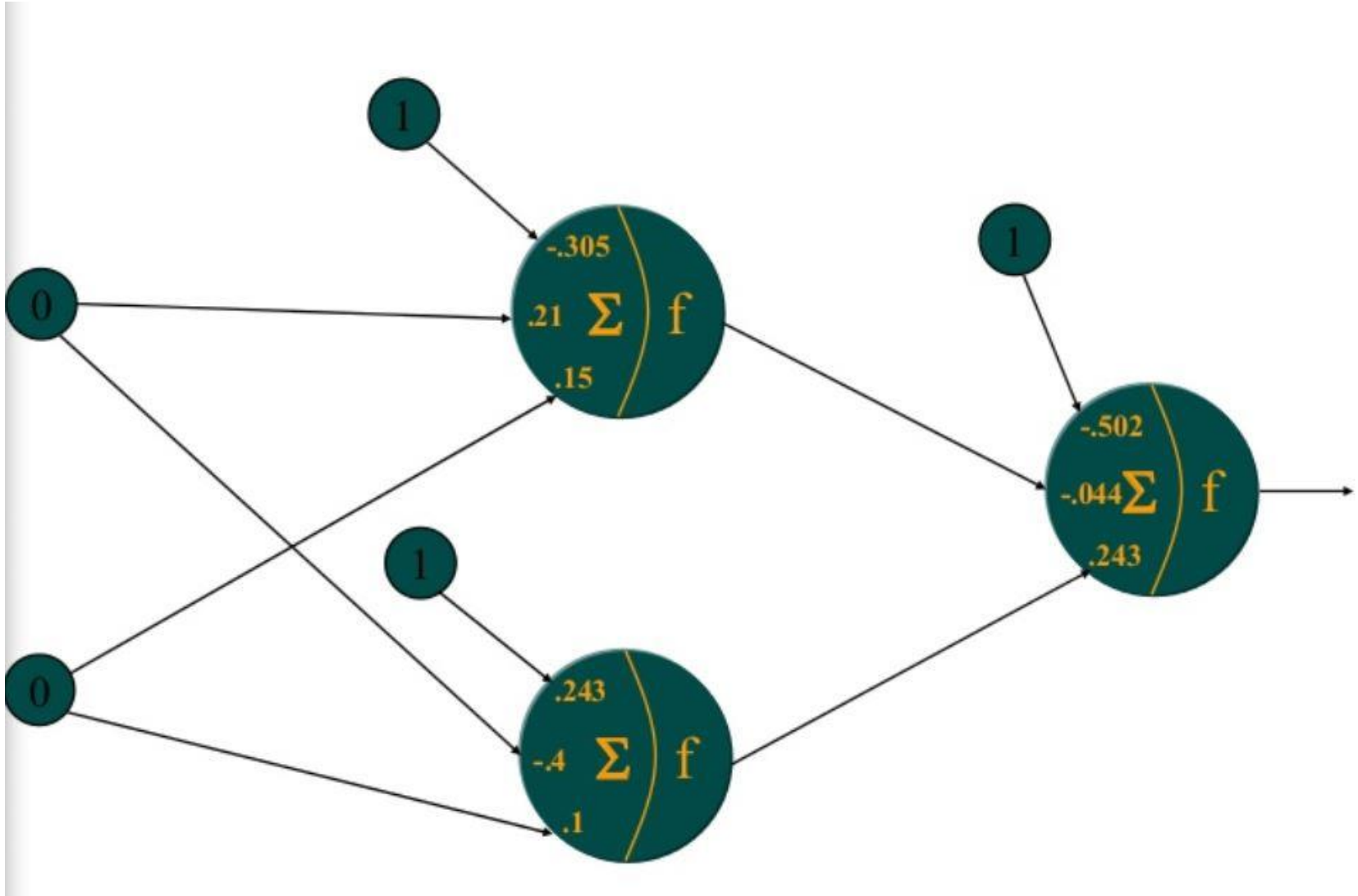
Training Case: (0 0 0)

Activation function f:

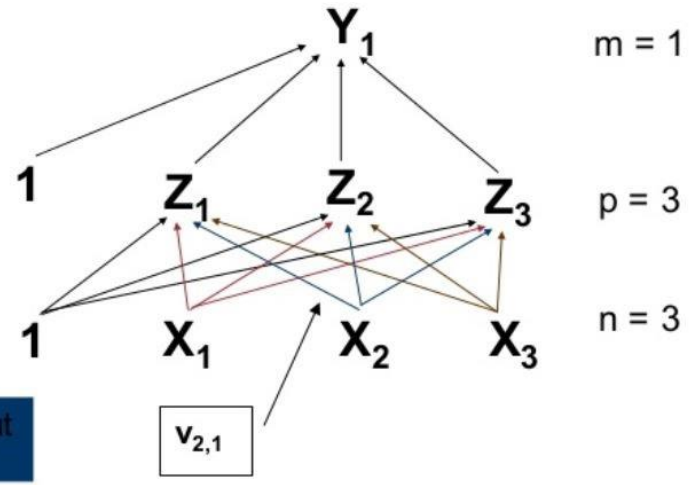
$$f(x) = \frac{1}{1+e^{-x}}$$


$\delta_1 = (t_1 - y_1)f'(y\_in_1)$
$=(t_1 - y_1)f(y\_in_1)[1- f(y\_in_1)]$
1
0
0
1
1
-.3
.21
.15
Σ
f
-.4
-.2
.3
.25
-.4
.1
$\delta\_in_1 = \delta_1 w_{11} = -.102(-.2) = .02$
$\delta_1 = \delta\_in_1 f'(z\_in_1) = .02(.43)(1-.43)$
$= .005$
$\delta_1 = (0 - .42).42[1-.42]$
$= -.102$
$\delta\_in_2 = \delta_1 w_{21} = -.102(.3) = -.03$
$\delta_2 = \delta\_in_2 f'(z\_in_2) = -.03(.56)(1-.56)$
$= -.007$


$\Delta v_{ij} = \alpha \delta_j x_i \quad j = 1,2$
$\Delta v_{0j} = \alpha \delta_j$
$\Delta w_{j1} = \alpha \delta_1 z_j \quad j = 1,2$
$\Delta w_0 = \alpha \delta_1$
1
$\Delta v_{01} = -.005$
0
$\Delta v_{11} = \delta_1 x_1 = (.005)(0) = 0$
$\Delta v_{12} = \delta_2 x_1 = (-.007)(0) = 0$
$\Delta v_{21} = \delta_1 x_2 = (.005)(0) = 0$
0
1
$\Delta v_{02} = -.007$
$\Delta v_{22} = \delta_2 x_2 = (-.007)(0) = 0$
-.3
.21
.15
Σ
f
.25
-.4
.1
Σ
f
1
$\Delta w_0 = -.102$
$\Delta w_{11} = \delta_1 z_1 = (-.102)(.43) = -.0439$
$\Delta w_{21} = \delta_1 z_2 = (-.102)(.56) = -.0571$
-.4
-.2
.3
Σ
f


1
-.305
.21
Σ
f
.15
1
0
-.502
-.044
Σ
f
.243
1
0
.243
-.4
Σ
f
.1

Example



$X = [0.6 \quad 0.8 \quad 0]$

$t = 0.9$

Desired output for X input

$v_0 = [0 \quad 0 \quad -1]$

$\alpha = 0.3$

$w_0 = [-1]$

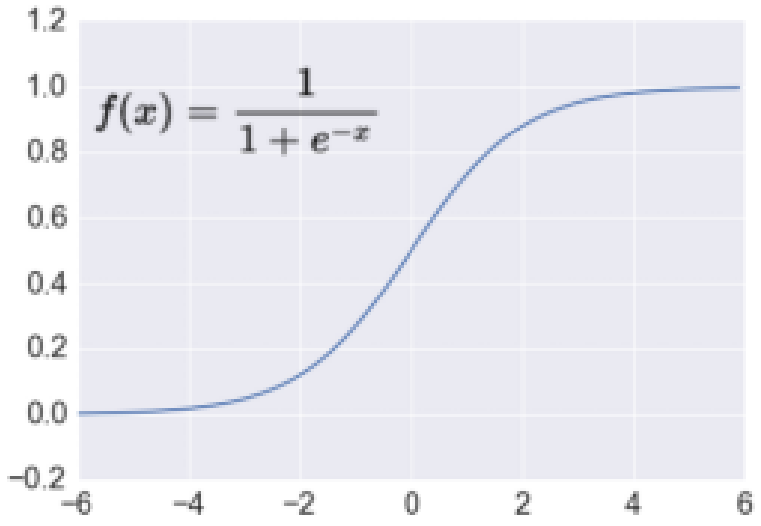
$$v = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 2 \\ 0 & 3 & 1 \end{bmatrix}$$

$$f(x) = \frac{1}{(1+e^{-x})}$$

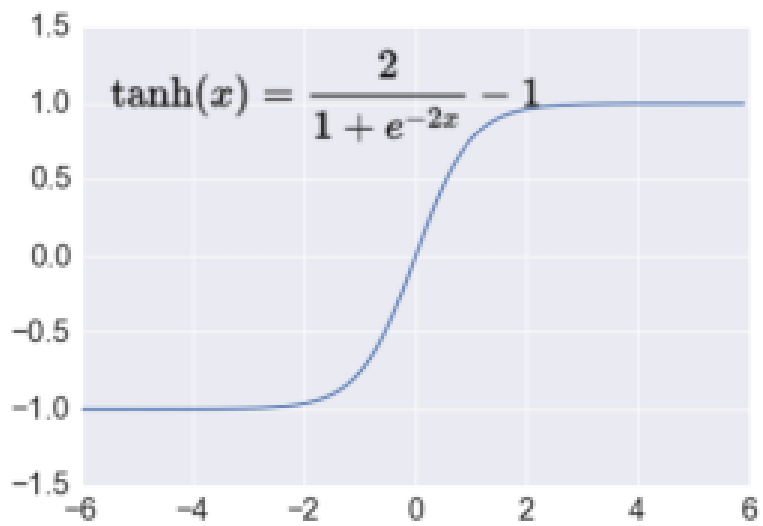
$$w = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}$$

Activation Functions

Sigmoid



TanH



ReLU