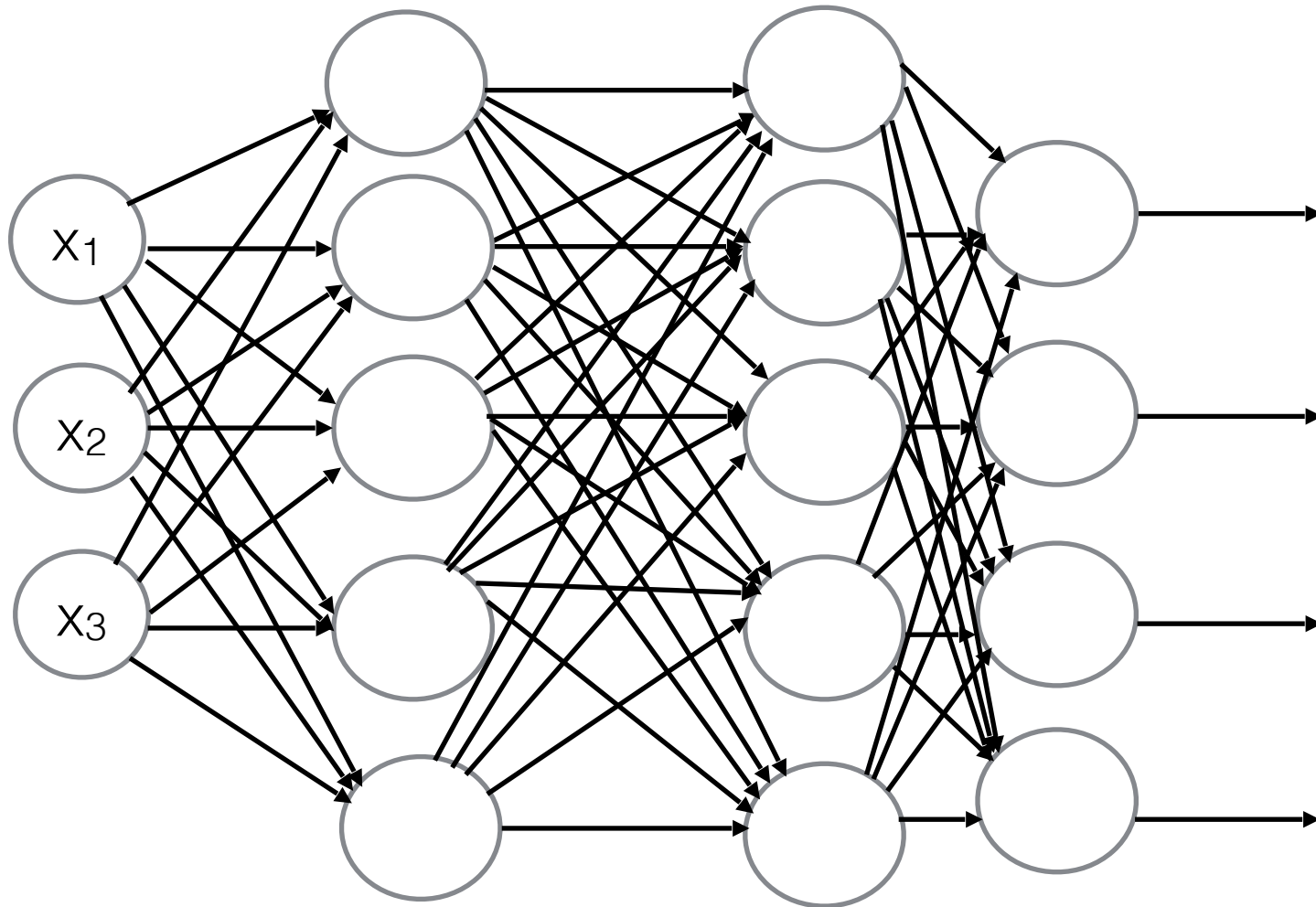


# Artificial Neural Networks

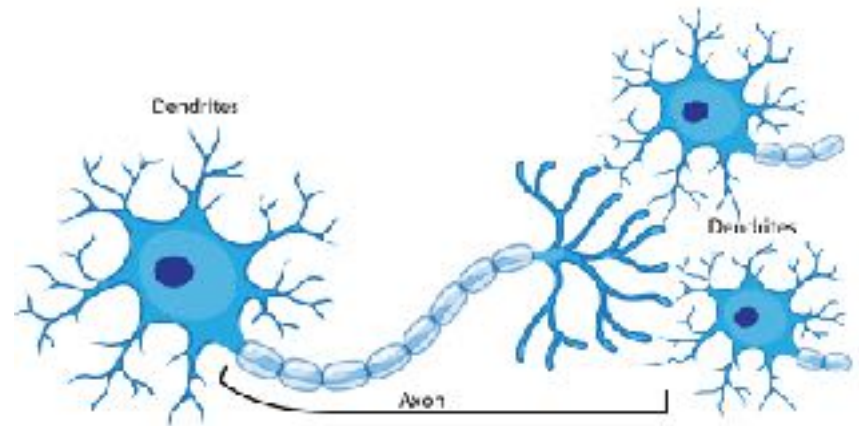


# What are they?

Inspired by the Human Brain.

The human brain has about 86 Billion neurons and requires 20% of your body's energy to function.

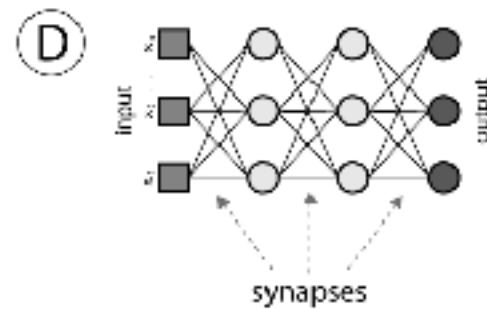
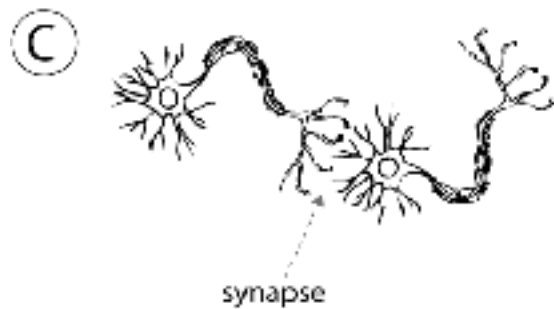
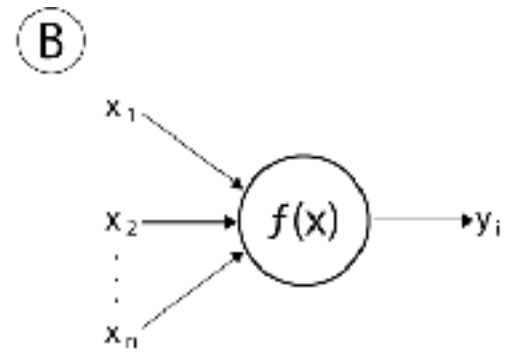
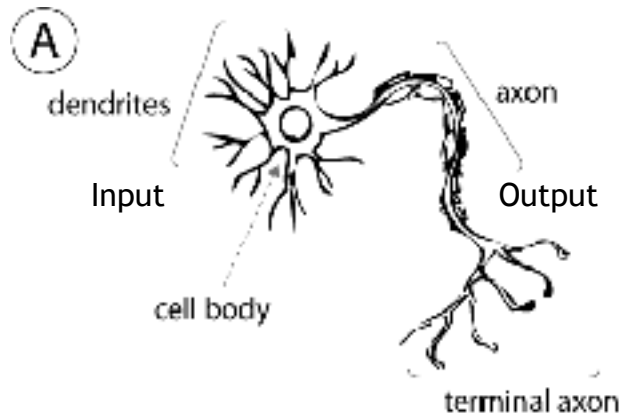
These neurons are connected to between 100 Trillion to 1 Quadrillion synapses!



# What are they?

Human

Machine



# What are they?

1. Originally developed by Warren McCullough and Walter Pitts (1944)
2. First trainable network, perceptron, proposed by Frank Rosenblatt (1957)
3. Started off as an unsupervised learning tool.
  1. Had problems with computing time and could not compute XOR
  2. Was abandoned in favor of other algorithms
4. [Werbos's](#) (1975) [backpropagation](#) algorithm
  1. Incorporated supervision and solved XOR
  2. But were still too slow vs. other algorithms e.g., Support Vector Machines
5. Backpropagation was accelerated by GPUs in 2010 and shown to be more efficient and cost effective

# GPUS

GPUS handle parallel operations much better (thousands of threads per core) but are not as quick as CPUs. However, the matrix multiplication steps in ANNs can be run in parallel resulting in considerable time + cost savings. The best CPUs handle about 50GB/s while the best GPUs handle 750GB/s memory bandwidth.

	CPU i9 Xseries	GeForce GTX 1080
Cores	18 (36 threads)	2560
Clock Speed (GHz)	4.4	1.6G
Memory	Shared	8GB
Price (\$)	1799	549

# Deep neural networks uncover what the brain likes to see

November 4, 2019

---

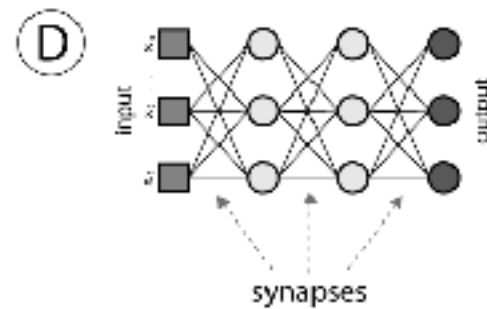
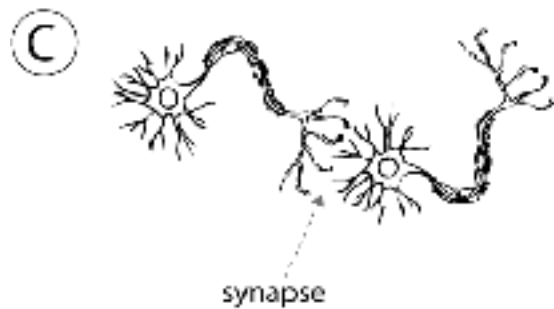
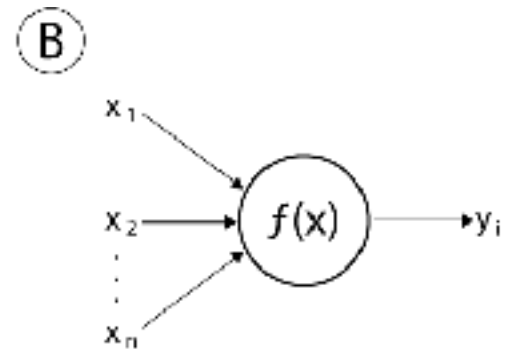
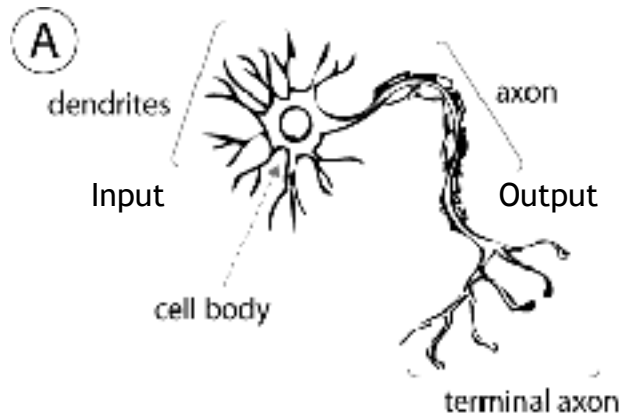


Researchers at Baylor College of Medicine and the University of Tübingen in Germany have developed a novel computational approach to finding stimuli that neurons in the brain 'like.' They built deep artificial neural networks that can accurately predict the neural responses produced by a biological brain to arbitrary visual stimuli.

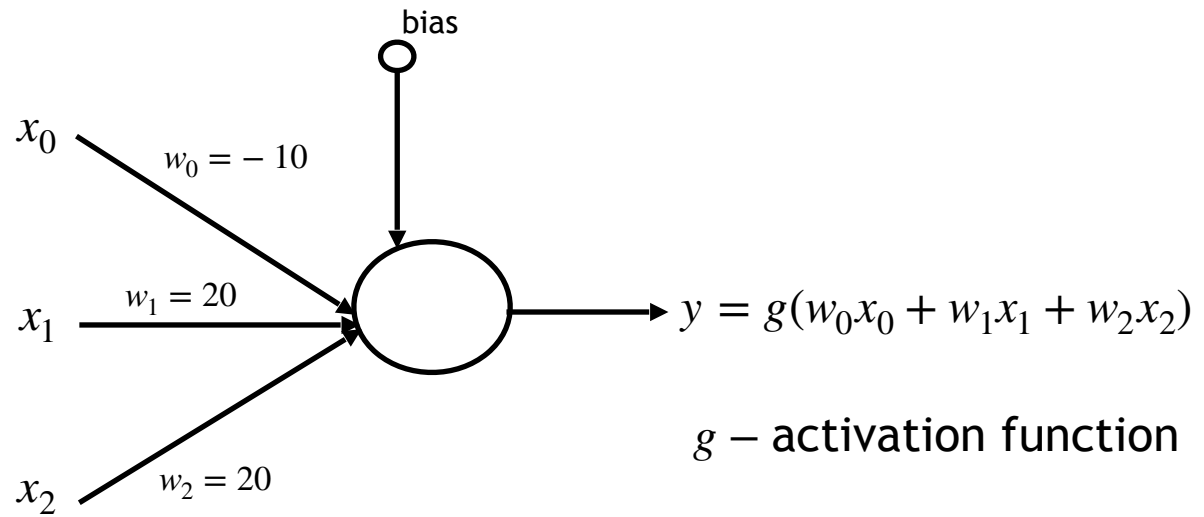
# What are they?

Human

Machine



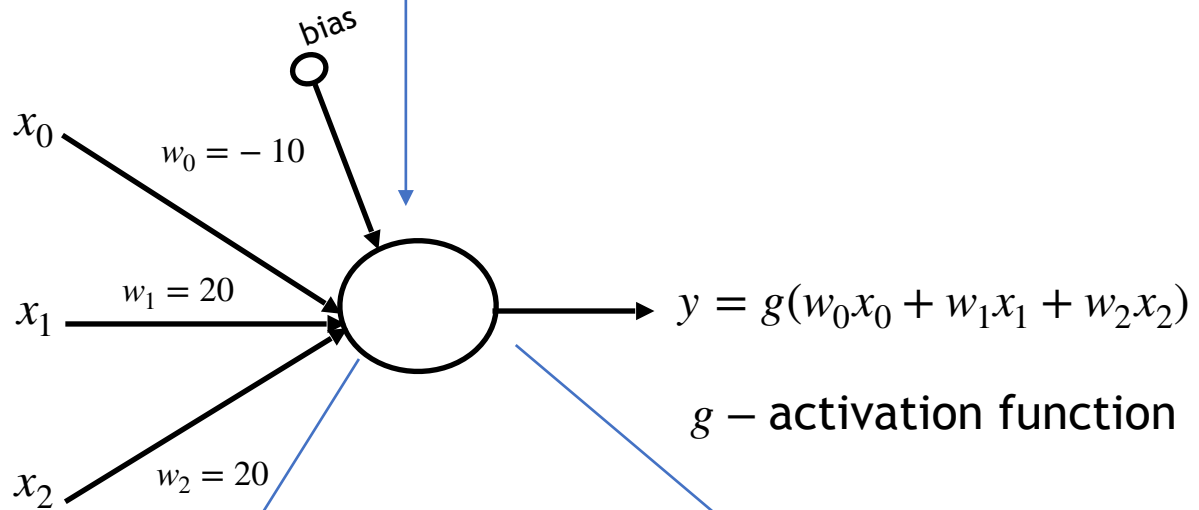
# Neuron





# Neuron

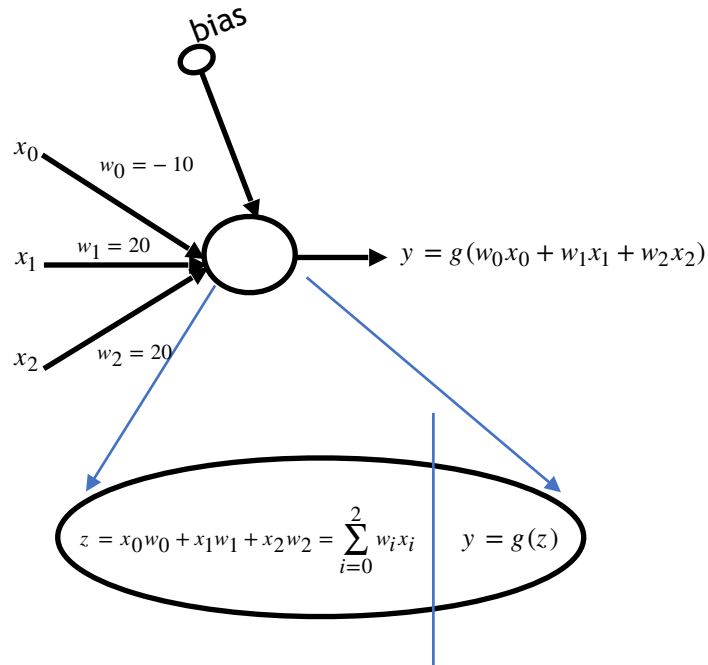
Inputs



$g$  – activation function

$$z = x_0w_0 + x_1w_1 + x_2w_2 = \sum_{i=0}^2 w_ix_i \quad y = g(z)$$

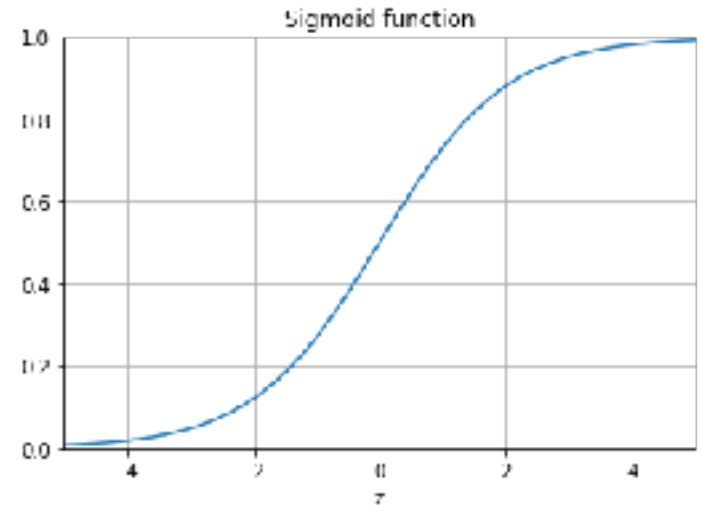
# Activation function



$g$  – activation function

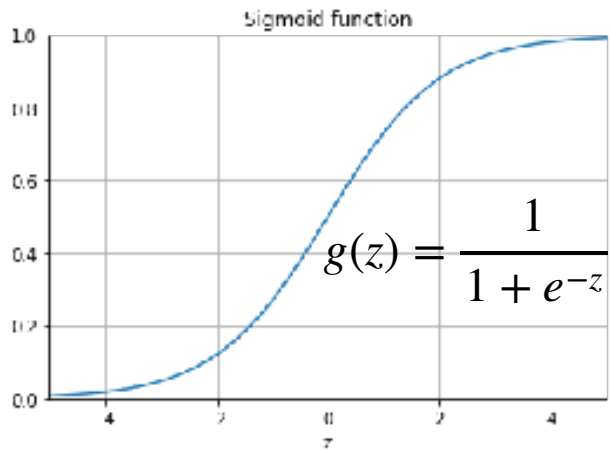
$$g(z) = \frac{1}{1 + e^{-z}}$$

$$z = w_0x_0 + w_1x_1 + w_2x_2$$

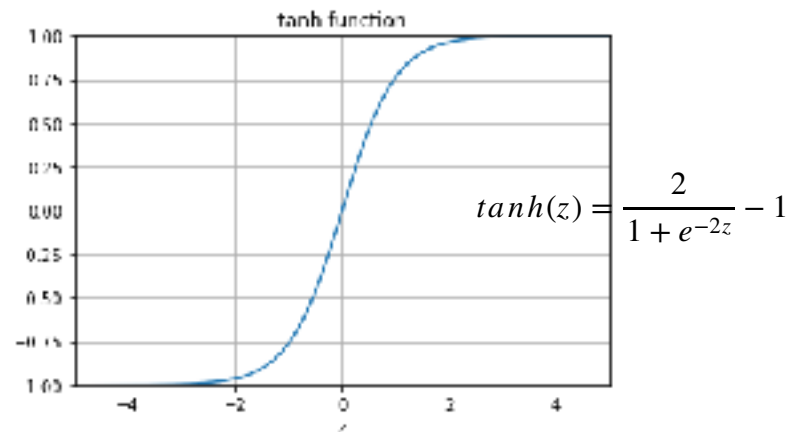


# Activation Functions

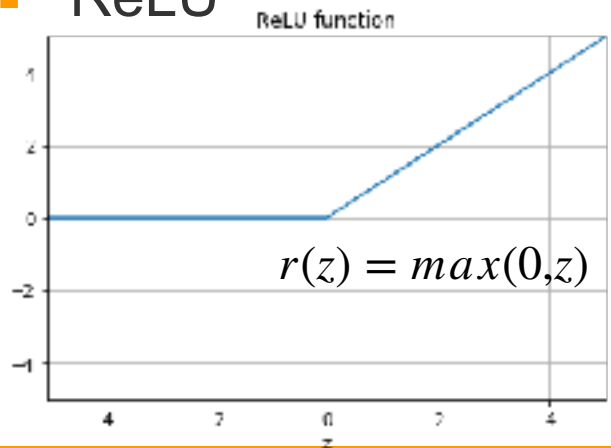
## ■ Sigmoid



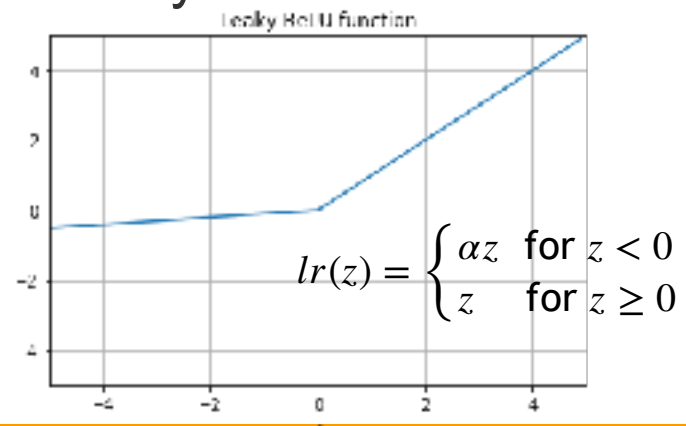
## ■ Hyperbolic tangent



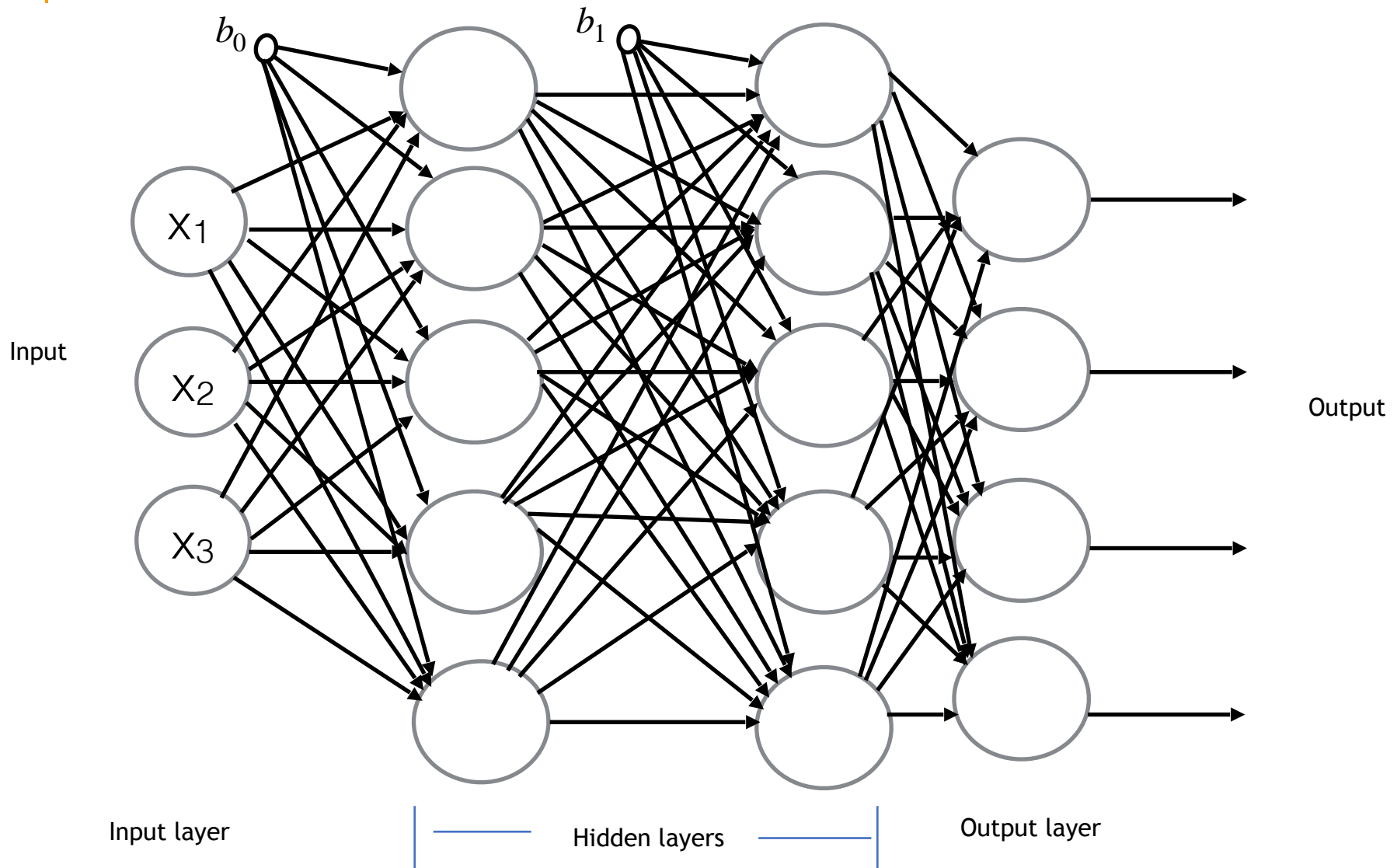
## ■ ReLU



## ■ Leaky ReLU

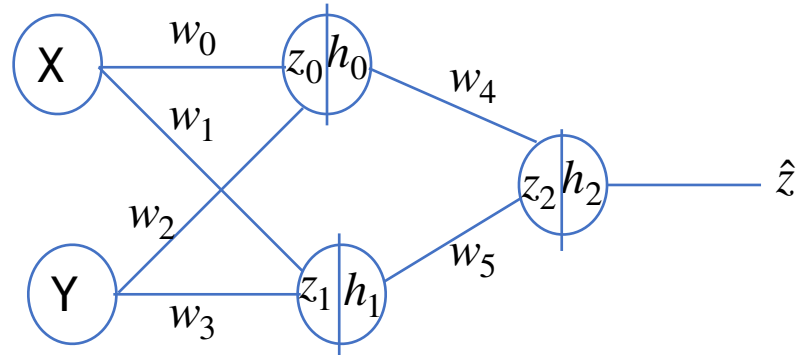


# Neural Network



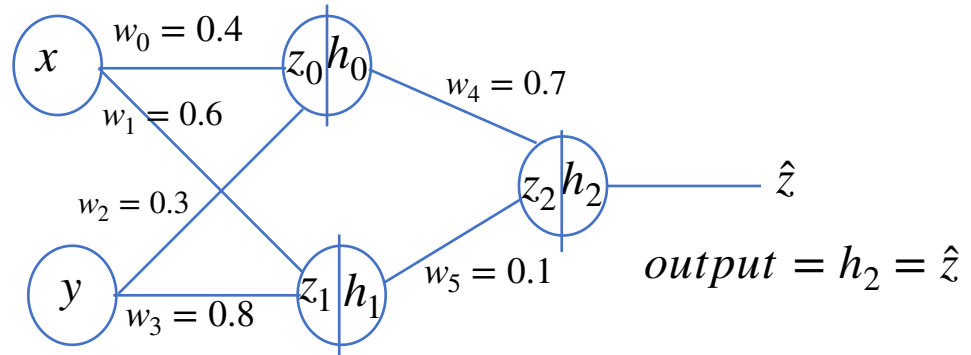
# XOR Gate

X	Y	$z_a$
0	0	0
0	1	1
1	0	1
1	1	0



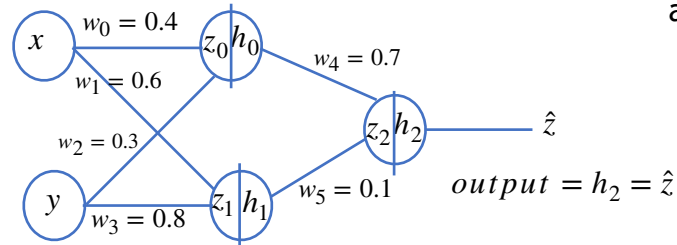
# XOR Gate

$x$	$y$	$z_a$
0	0	0
0	1	1
1	0	1
1	1	0



# Feedforward propagation

$x$	$y$	$z_a$
0	0	0
0	1	1
1	0	1
1	1	0



activation function,  $g(z) = \frac{1}{1 + e^{-z}}$

$$\begin{aligned}
 z_0 &= w_0x + w_2y \\
 x = 1 &= 0.4 \times 1 + 0.3 \times 1 \\
 &= 0.7
 \end{aligned}$$

$$\begin{aligned}
 h_0 &= g(z_0) = \frac{1}{1 + e^{-z_0}} \\
 &= \frac{1}{1 + e^{-0.7}} \\
 h_0 &= 0.67
 \end{aligned}$$

$$\begin{aligned}
 z_1 &= w_1x + w_3y \\
 y = 1 &= 0.6 \times 1 + 0.8 \times 1 \\
 &= 1.4
 \end{aligned}$$

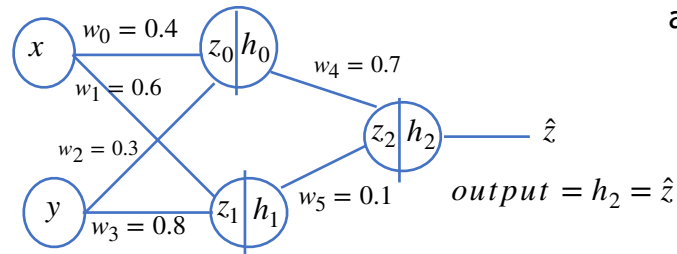
$$\begin{aligned}
 h_1 &= g(z_1) = \frac{1}{1 + e^{-z_1}} \\
 &= \frac{1}{1 + e^{-1.4}} \\
 h_1 &= 0.8
 \end{aligned}$$

$$\begin{aligned}
 z_2 &= w_4h_0 + w_5h_1 \\
 &= 0.7 \times 0.67 + 0.1 \times 0.8 \\
 &= 0.63
 \end{aligned}$$

$$\begin{aligned}
 h_2 &= g(z_2) = \frac{1}{1 + e^{-z_2}} \\
 &= \frac{1}{1 + e^{-0.63}} \\
 h_2 &= 0.65
 \end{aligned}
 \quad h_2 = \hat{z} = 0.65$$

# Feedforward propagation

$x$	$y$	$z_a$
0	0	0
0	1	1
1	0	1
1	1	0



activation function,  $g(z) = \frac{1}{1 + e^{-z}}$

$$\begin{aligned}
 z_0 &= w_0x + w_2y \\
 x = 1 \quad &= 0.4 \times 1 + 0.3 \times 1 \\
 &= 0.7
 \end{aligned}$$

$$\begin{aligned}
 h_0 &= g(z_0) = \frac{1}{1 + e^{-z_0}} \\
 &= \frac{1}{1 + e^{-0.7}} \\
 h_0 &= 0.67
 \end{aligned}$$

$$\begin{aligned}
 z_2 &= w_4h_0 + w_5h_1 \\
 &= 0.7 \times 0.67 + 0.1 \times 0.8 \\
 &= 0.63
 \end{aligned}$$

$$\begin{aligned}
 h_2 &= g(z_2) = \frac{1}{1 + e^{-z_2}} \\
 &= \frac{1}{1 + e^{-0.63}} ; h_2 = \hat{z} = 0.65 \\
 h_2 &= 0.65
 \end{aligned}$$

$$\begin{aligned}
 z_1 &= w_1x + w_3y \\
 y = 1 \quad &= 0.6 \times 1 + 0.8 \times 1 \\
 &= 1.4
 \end{aligned}$$

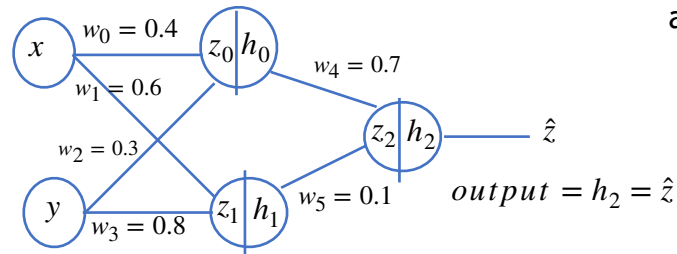
$$\begin{aligned}
 h_1 &= g(z_1) = \frac{1}{1 + e^{-z_1}} \\
 &= \frac{1}{1 + e^{-1.4}} \\
 h_1 &= 0.8
 \end{aligned}$$

$$\text{Mean Squared Error, } E = \frac{1}{2}(\hat{z} - z_a)^2 = \frac{1}{2}(0.65 - 0)^2 = 0.21$$



# Backpropagation

$x$	$y$	$z_a$
0	0	0
0	1	1
1	0	1
1	1	0



activation function,  $g(z) = \frac{1}{1 + e^{-z}}$

$$E = \frac{1}{2}(\hat{z} - z_a)^2$$

$$g'(z) = \frac{\partial}{\partial z} g(z) = \frac{\partial}{\partial z} \left( \frac{1}{1 + e^{-z}} \right) = g(z)[1 - g(z)]$$

Mean Squared Error,  $E = \frac{1}{2}(\hat{z} - z_a)^2 = \frac{1}{2}(h_2 - z_a)^2 = \frac{1}{2}(0.65 - 0)^2 = 0.21$

$$\frac{\partial E}{\partial w_4} = \frac{\partial E}{\partial h_2} \cdot \frac{\partial h_2}{\partial z_2} \cdot \frac{\partial z_2}{\partial w_4} = 0.65 \times 0.228 \times 0.67 = 0.09$$

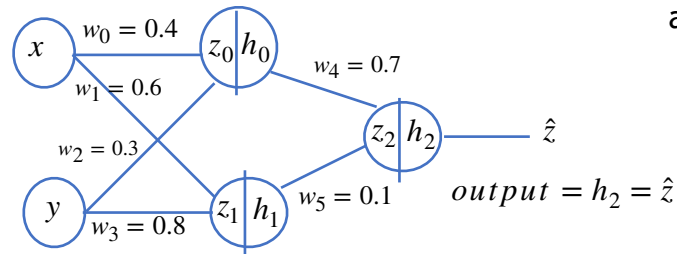
$$\frac{\partial E}{\partial h_2} = \frac{\partial}{\partial h_2} \left[ \frac{1}{2}(h_2 - z_a)^2 \right] = h_2 - z_a = 0.65$$

$$\frac{\partial h_2}{\partial z_2} = \frac{\partial}{\partial z_2} g(z_2) = \frac{\partial}{\partial z_2} \left( \frac{1}{1 + e^{-z_2}} \right) = g(z_2)(1 - g(z_2)) = h_2(1 - h_2) = 0.65 \times 0.35 = 0.228$$

$$\frac{\partial z_2}{\partial w_4} = \frac{\partial}{\partial w_4} [w_4 h_0 + w_5 h_1] = h_0 = 0.67$$

# Backpropagation

$x$	$y$	$z_a$
0	0	0
0	1	1
1	0	1
1	1	0



activation function,  $g(z) = \frac{1}{1 + e^{-z}}$

$$E = \frac{1}{2}(\hat{z} - z_a)^2$$

$$g'(z) = \frac{\partial}{\partial z} g(z) = \frac{\partial}{\partial z} \left( \frac{1}{1 + e^{-z}} \right) = g(z)[1 - g(z)]$$

Mean Squared Error,  $E = \frac{1}{2}(\hat{z} - z_a)^2 = \frac{1}{2}(h_2 - z_a)^2 = \frac{1}{2}(0.65 - 0)^2 = 0.21$

$$\frac{\partial E}{\partial w_5} = \frac{\partial E}{\partial h_2} \cdot \frac{\partial h_2}{\partial z_2} \cdot \frac{\partial z_2}{\partial w_5} = 0.65 \times 0.228 \times 0.8 = 0.118$$

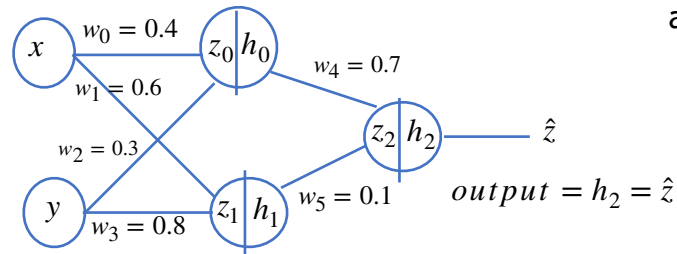
$$\frac{\partial E}{\partial h_2} = \frac{\partial}{\partial h_2} \left[ \frac{1}{2}(h_2 - z_a)^2 \right] = h_2 - z_a = 0.65$$

$$\frac{\partial h_2}{\partial z_2} = \frac{\partial}{\partial z_2} g(z_2) = \frac{\partial}{\partial z_2} \left( \frac{1}{1 + e^{-z_2}} \right) = g(z_2)(1 - g(z_2)) = h_2(1 - h_2) = 0.65 \times 0.35 = 0.228$$

$$\frac{\partial z_2}{\partial w_5} = \frac{\partial}{\partial w_5} [w_4 h_0 + w_5 h_1] = h_1 = 0.8$$

# Backpropagation

$x$	$y$	$z_a$
0	0	0
0	1	1
1	0	1
1	1	0



activation function,  $g(z) = \frac{1}{1 + e^{-z}}$

$$E = \frac{1}{2}(\hat{z} - z_a)^2$$

$$g'(z) = \frac{\partial}{\partial z} g(z) = \frac{\partial}{\partial z} \left( \frac{1}{1 + e^{-z}} \right) = g(z)[1 - g(z)]$$

Mean Squared Error,  $E = \frac{1}{2}(\hat{z} - z_a)^2 = \frac{1}{2}(h_2 - z_a)^2 = \frac{1}{2}(0.65 - 0)^2 = 0.21$

$$\frac{\partial E}{\partial w_0} = \frac{\partial E}{\partial h_0} \cdot \frac{\partial h_0}{\partial z_0} \cdot \frac{\partial z_0}{\partial w_0} = 0.103 \times 0.221 \times 1 = 0.227$$

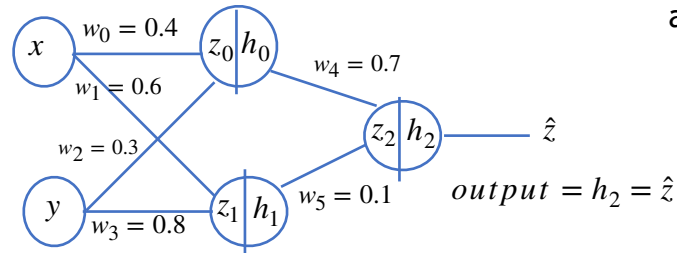
$$\frac{\partial E}{\partial h_0} = \frac{\partial E}{\partial h_2} \cdot \frac{\partial h_2}{\partial z_2} \cdot \frac{\partial z_2}{\partial h_0} = h_2 \times h_2(1 - h_2) \times w_4 = 0.65 \times 0.65(1 - 0.65) \times 0.7 = 0.103$$

$$\frac{\partial h_0}{\partial z_0} = \frac{\partial}{\partial z_0} g(z_0) = \frac{\partial}{\partial z_0} \left( \frac{1}{1 + e^{-z_0}} \right) = g(z_0)(1 - g(z_0)) = h_0(1 - h_0) = 0.67 \times 0.33 = 0.221$$

$$\frac{\partial z_0}{\partial w_0} = \frac{\partial}{\partial w_0} [w_0 x + w_2 y] = x = 1$$

# Backpropagation

$x$	$y$	$z_a$
0	0	0
0	1	1
1	0	1
1	1	0



activation function,  $g(z) = \frac{1}{1 + e^{-z}}$

$$E = \frac{1}{2}(\hat{z} - z_a)^2$$

$$g'(z) = \frac{\partial}{\partial z} g(z) = \frac{\partial}{\partial z} \left( \frac{1}{1 + e^{-z}} \right) = g(z)[1 - g(z)]$$

Mean Squared Error,  $E = \frac{1}{2}(\hat{z} - z_a)^2 = \frac{1}{2}(h_2 - z_a)^2 = \frac{1}{2}(0.65 - 0)^2 = 0.21$

$$\frac{\partial E}{\partial w_2} = \frac{\partial E}{\partial h_0} \cdot \frac{\partial h_0}{\partial z_0} \cdot \frac{\partial z_0}{\partial w_2} = 0.103 \times 0.221 \times 1 = 0.227$$

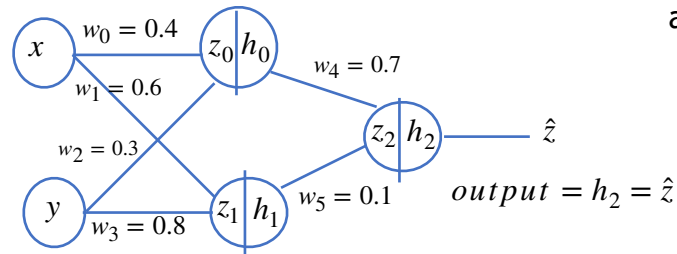
$$\frac{\partial E}{\partial h_0} = \frac{\partial E}{\partial h_2} \cdot \frac{\partial h_2}{\partial z_2} \cdot \frac{\partial z_2}{\partial h_0} = h_2 \times h_2(1 - h_2) \times w_4 = 0.65 \times 0.65(1 - 0.65) \times 0.7 = 0.103$$

$$\frac{\partial h_0}{\partial z_0} = \frac{\partial}{\partial z_0} g(z_0) = \frac{\partial}{\partial z_0} \left( \frac{1}{1 + e^{-z_0}} \right) = g(z_0)(1 - g(z_0)) = h_0(1 - h_0) = 0.67 \times 0.33 = 0.221$$

$$\frac{\partial z_0}{\partial w_2} = \frac{\partial}{\partial w_2} [w_0 x + w_2 y] = y = 1$$

# Backpropagation

$x$	$y$	$z_a$
0	0	0
0	1	1
1	0	1
1	1	0



activation function,  $g(z) = \frac{1}{1 + e^{-z}}$

$$E = \frac{1}{2}(\hat{z} - z_a)^2$$

$$g'(z) = \frac{\partial}{\partial z} g(z) = \frac{\partial}{\partial z} \left( \frac{1}{1 + e^{-z}} \right) = g(z)[1 - g(z)]$$

Mean Squared Error,  $E = \frac{1}{2}(\hat{z} - z_a)^2 = \frac{1}{2}(h_2 - z_a)^2 = \frac{1}{2}(0.65 - 0)^2 = 0.21$

$$\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial h_1} \cdot \frac{\partial h_1}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_1} = 0.014 \times 0.16 \times 1 = 0.0022$$

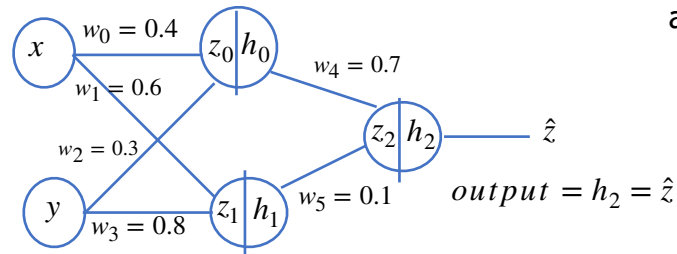
$$\frac{\partial E}{\partial h_1} = \frac{\partial E}{\partial h_2} \cdot \frac{\partial h_2}{\partial z_2} \cdot \frac{\partial z_2}{\partial h_1} = h_2 \times h_2(1 - h_2) \times w_5 = 0.65 \times 0.65(1 - 0.65) \times 0.1 = 0.014$$

$$\frac{\partial h_1}{\partial z_1} = \frac{\partial}{\partial z_1} g(z_1) = \frac{\partial}{\partial z_1} \left( \frac{1}{1 + e^{-z_1}} \right) = g(z_1)(1 - g(z_1)) = h_1(1 - h_1) = 0.8 \times 0.2 = 0.16$$

$$\frac{\partial z_0}{\partial w_1} = \frac{\partial}{\partial w_1} [w_1 x + w_3 y] = x = 1$$

# Backpropagation

$x$	$y$	$z_a$
0	0	0
0	1	1
1	0	1
1	1	0



activation function,  $g(z) = \frac{1}{1 + e^{-z}}$

$$E = \frac{1}{2}(\hat{z} - z_a)^2$$

$$g'(z) = \frac{\partial}{\partial z} g(z) = \frac{\partial}{\partial z} \left( \frac{1}{1 + e^{-z}} \right) = g(z)[1 - g(z)]$$

Mean Squared Error,  $E = \frac{1}{2}(\hat{z} - z_a)^2 = \frac{1}{2}(h_2 - z_a)^2 = \frac{1}{2}(0.65 - 0)^2 = 0.21$

$$\frac{\partial E}{\partial w_3} = \frac{\partial E}{\partial h_1} \cdot \frac{\partial h_1}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_3} = 0.014 \times 0.16 \times 1 = 0.0022$$

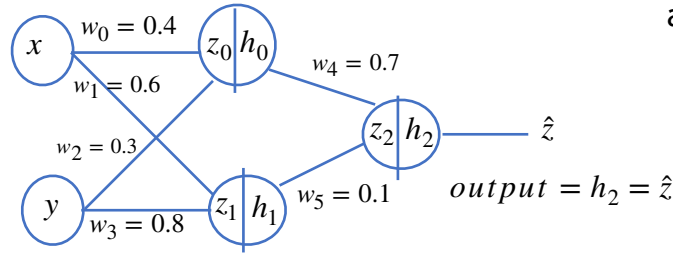
$$\frac{\partial E}{\partial h_1} = \frac{\partial E}{\partial h_2} \cdot \frac{\partial h_2}{\partial z_2} \cdot \frac{\partial z_2}{\partial h_1} = h_2 \times h_2(1 - h_2) \times w_5 = 0.65 \times 0.65(1 - 0.65) \times 0.1 = 0.014$$

$$\frac{\partial h_1}{\partial z_1} = \frac{\partial}{\partial z_1} g(z_1) = \frac{\partial}{\partial z_1} \left( \frac{1}{1 + e^{-z_1}} \right) = g(z_1)(1 - g(z_1)) = h_1(1 - h_1) = 0.8 \times 0.2 = 0.16$$

$$\frac{\partial z_0}{\partial w_1} = \frac{\partial}{\partial w_1} [w_1 x + w_3 y] = y = 1$$

# Gradient descent weight updates

x	y	$z_a$
0	0	0
0	1	1
1	0	1
1	1	0



activation function,  $g(z) = \frac{1}{1 + e^{-z}}$

$$E = \frac{1}{2}(\hat{z} - z_a)^2$$

$$g'(z) = \frac{\partial}{\partial z} g(z) = \frac{\partial}{\partial z} \left( \frac{1}{1 + e^{-z}} \right) = g(z)[1 - g(z)]$$

Mean Squared Error,  $E = \frac{1}{2}(\hat{z} - z_a)^2 = \frac{1}{2}(h_2 - z_a)^2 = \frac{1}{2}(0.65 - 0)^2 = 0.21$

*feedforward*

$$\begin{aligned}
 z_0 &= w_0 x + w_2 y & h_0 &= g(z_0) = \frac{1}{1 + e^{-z_0}} \\
 x = 1 & & &= \frac{1}{1 + e^{-0.7}} \\
 &= 0.4 \times 1 + 0.3 \times 1 & &= 0.67 \\
 &= 0.7 & & \\
 z_1 &= w_1 x + w_3 y & h_1 &= g(z_1) = \frac{1}{1 + e^{-z_1}} \\
 y = 1 & & &= \frac{1}{1 + e^{-1.4}} \\
 &= 0.6 \times 1 + 0.8 \times 1 & &= 0.8 \\
 &= 1.4 & & \\
 z_2 &= w_4 h_0 + w_5 h_1 & h_2 &= g(z_2) = \frac{1}{1 + e^{-z_2}} \\
 &= 0.7 \times 0.67 + 0.1 \times 0.8 & &= \frac{1}{1 + e^{-0.63}} \\
 &= 0.63 & &h_2 = \hat{z} = 0.65 \\
 & & &h_2 = 0.65
 \end{aligned}$$

*backprop*

$$\begin{aligned}
 \frac{\partial E}{\partial w_4} &= \frac{\partial E}{\partial h_2} \cdot \frac{\partial h_2}{\partial z_2} \cdot \frac{\partial z_2}{\partial w_4} = 0.65 \times 0.228 \times 0.67 = 0.09 & \frac{\partial E}{\partial w_0} &= \frac{\partial E}{\partial h_0} \cdot \frac{\partial h_0}{\partial z_0} \cdot \frac{\partial z_0}{\partial w_0} = 0.103 \times 0.221 \times 1 = 0.227 & \frac{\partial E}{\partial w_1} &= \frac{\partial E}{\partial h_1} \cdot \frac{\partial h_1}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_1} = 0.014 \times 0.16 \times 1 = 0.0022 \\
 \frac{\partial E}{\partial w_5} &= \frac{\partial E}{\partial h_2} \cdot \frac{\partial h_2}{\partial z_2} \cdot \frac{\partial z_2}{\partial w_5} = 0.65 \times 0.228 \times 0.8 = 0.118 & \frac{\partial E}{\partial w_2} &= \frac{\partial E}{\partial h_0} \cdot \frac{\partial h_0}{\partial z_0} \cdot \frac{\partial z_0}{\partial w_2} = 0.103 \times 0.221 \times 1 = 0.227 & \frac{\partial E}{\partial w_3} &= \frac{\partial E}{\partial h_1} \cdot \frac{\partial h_1}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_3} = 0.014 \times 0.16 \times 1 = 0.0022
 \end{aligned}$$

Learning rate,  $\alpha = 0.01$

*grad descent weight update*

$$w_4 = w_4 - \alpha \frac{\partial E}{\partial w_4}; w_5 = w_5 - \alpha \frac{\partial E}{\partial w_5}; w_0 = w_0 - \alpha \frac{\partial E}{\partial w_0}; w_1 = w_1 - \alpha \frac{\partial E}{\partial w_1}; w_2 = w_2 - \alpha \frac{\partial E}{\partial w_2}; w_3 = w_3 - \alpha \frac{\partial E}{\partial w_3};$$