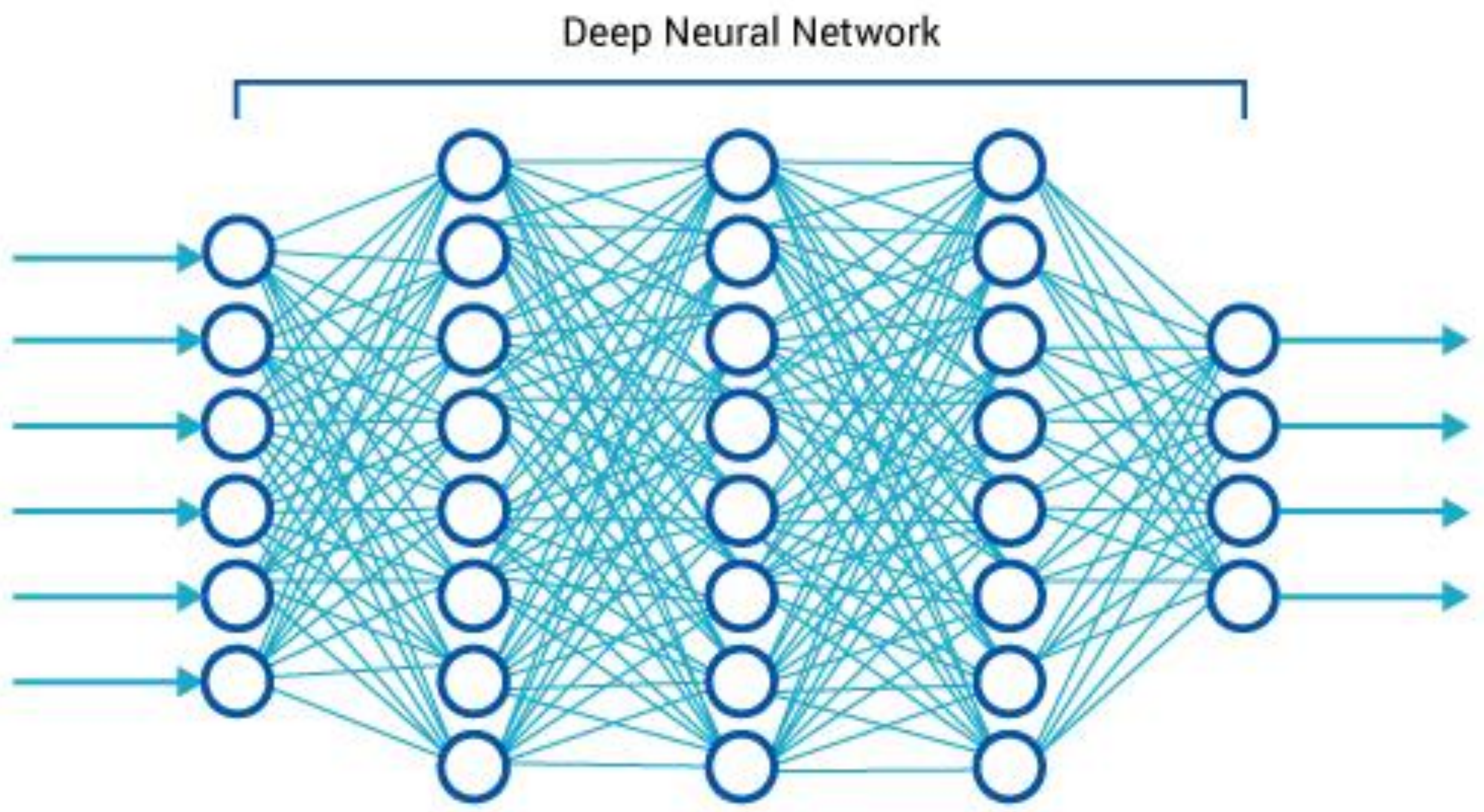
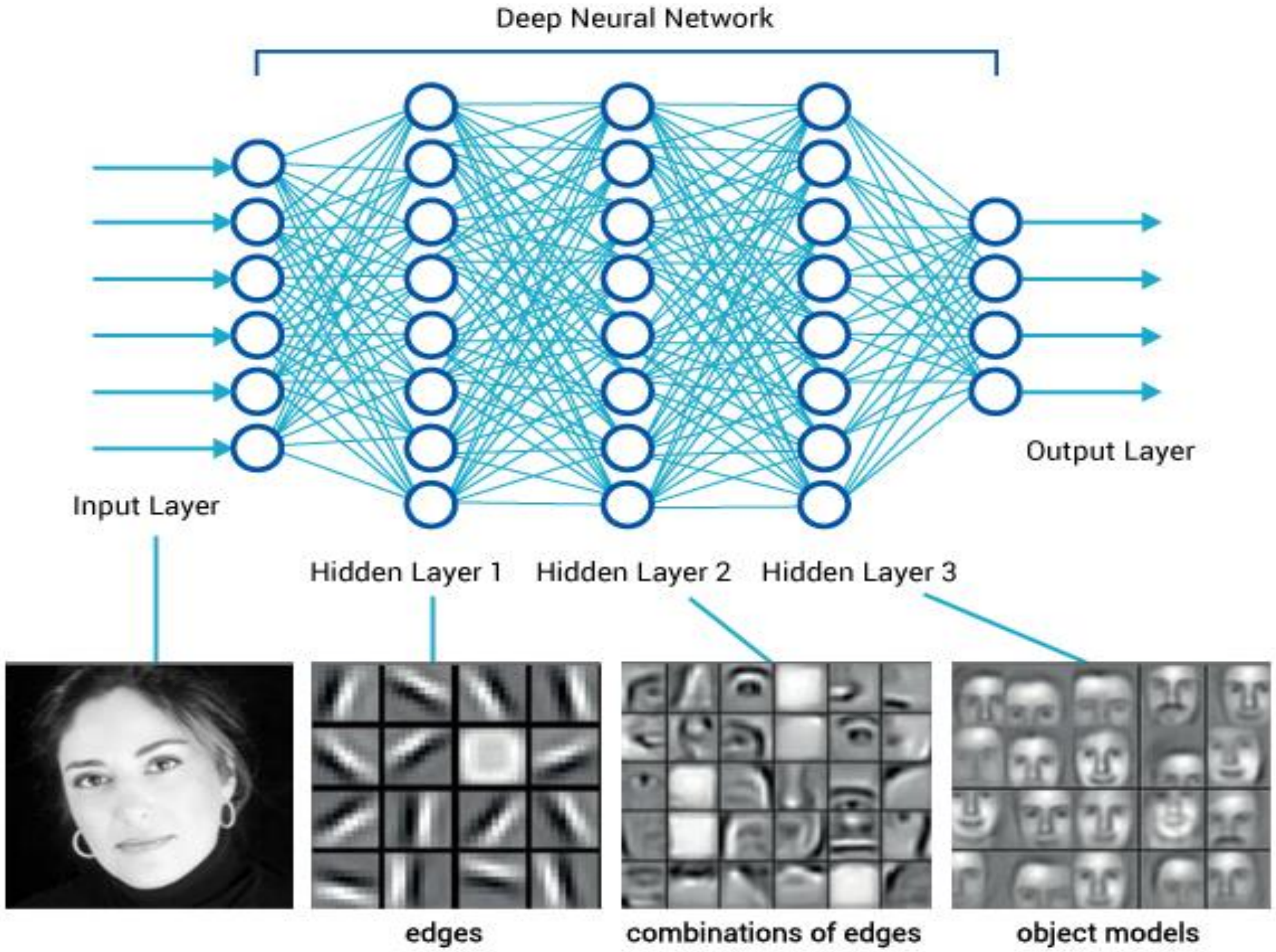
Deep Learning



Deep Learning

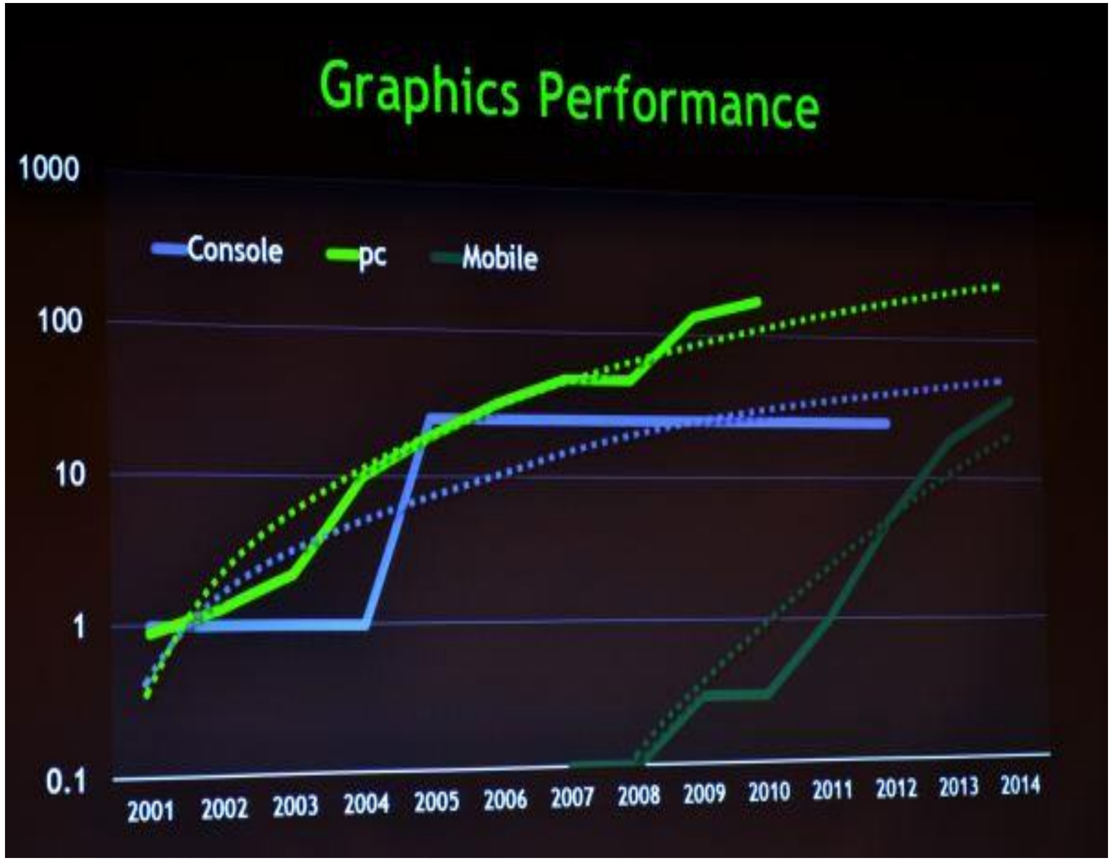


Why is deep learning a growing trend?

- few feature engineering

- state-of-the-art performance

Deep Learning



Deep Learning Heroes

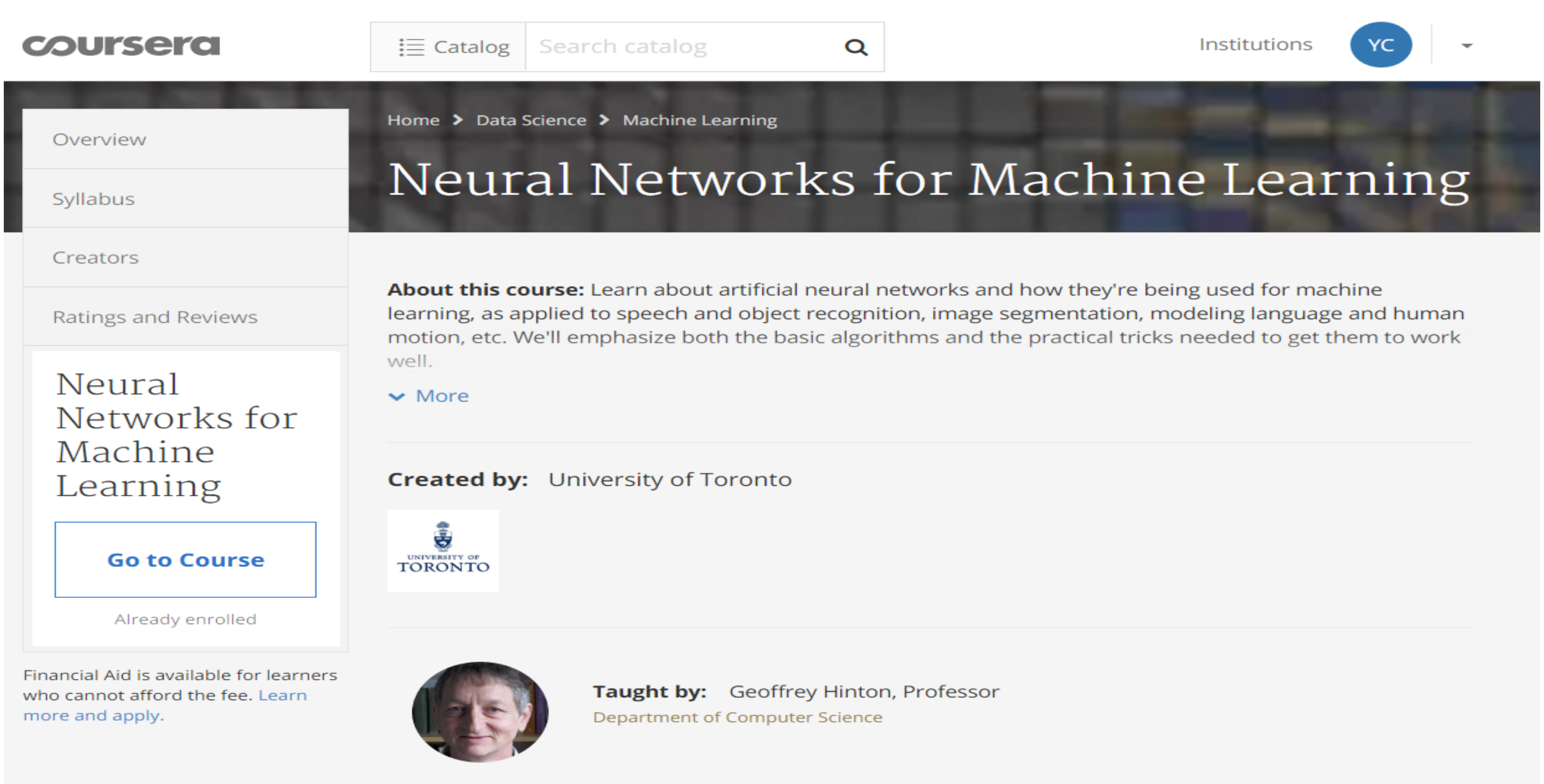
[Geoffrey Hinton]

[Yann LeCun]

[Andrew Ng]

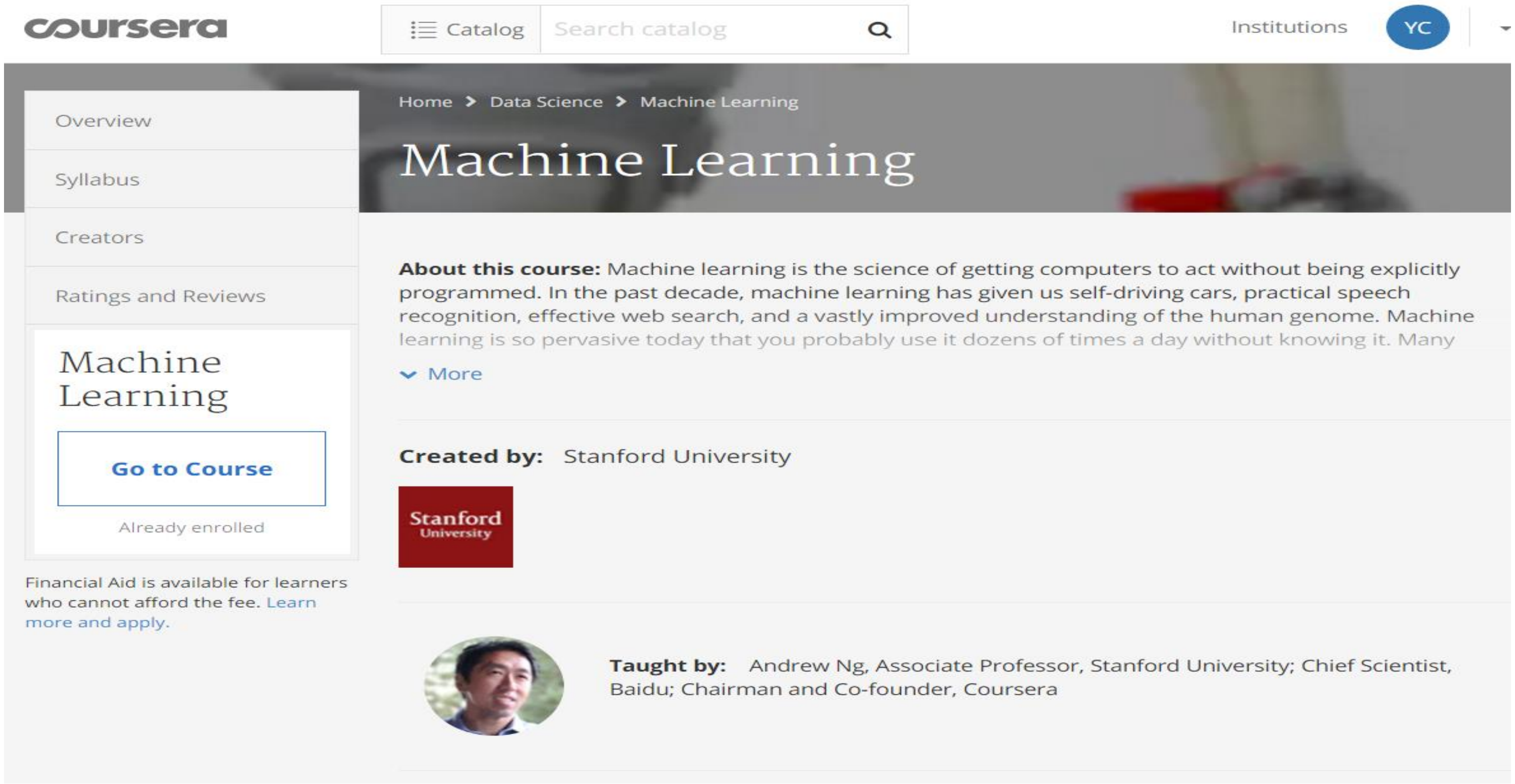
[Yoshua Bengio]

Lectures for Neural Network



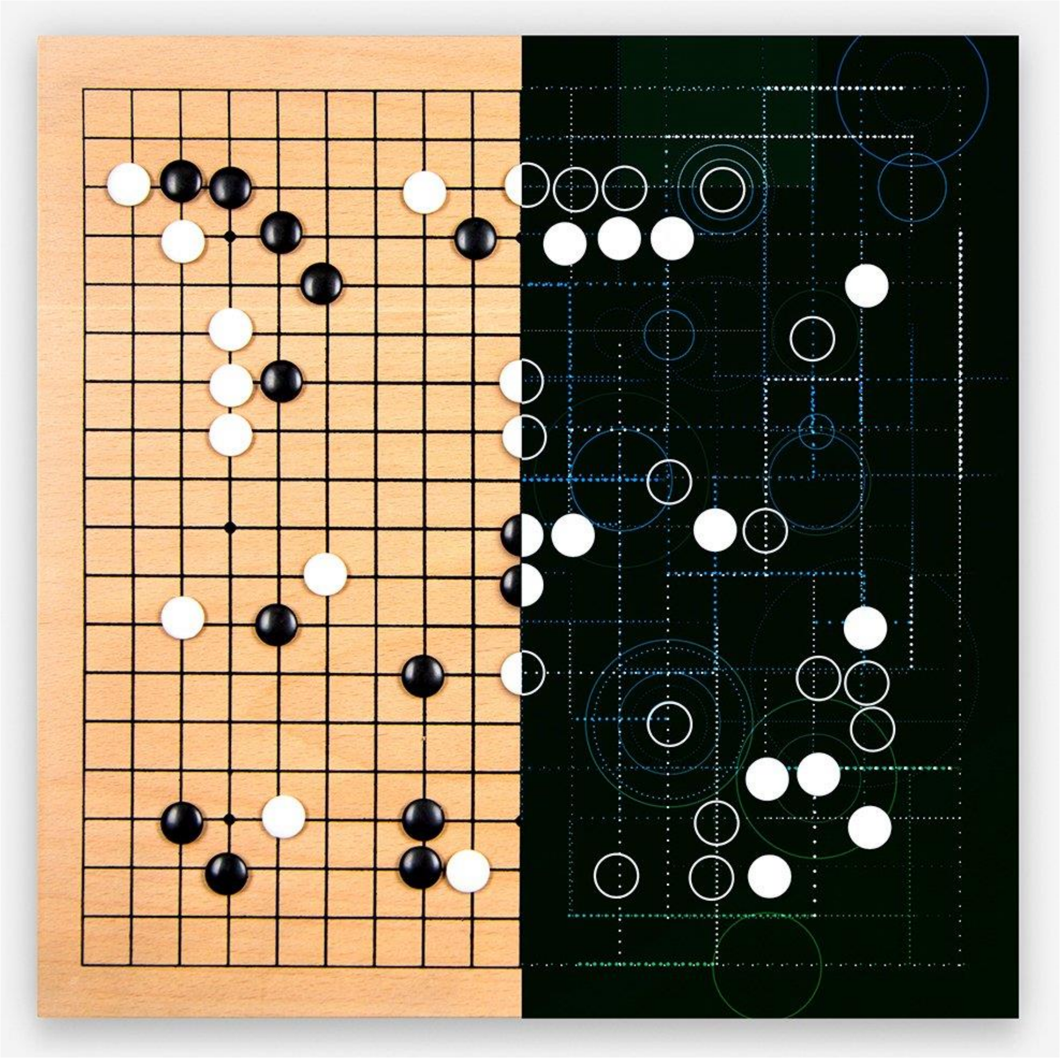
[Neural Networks for Machine Learning – University of Toronto]

Lectures for Neural Network

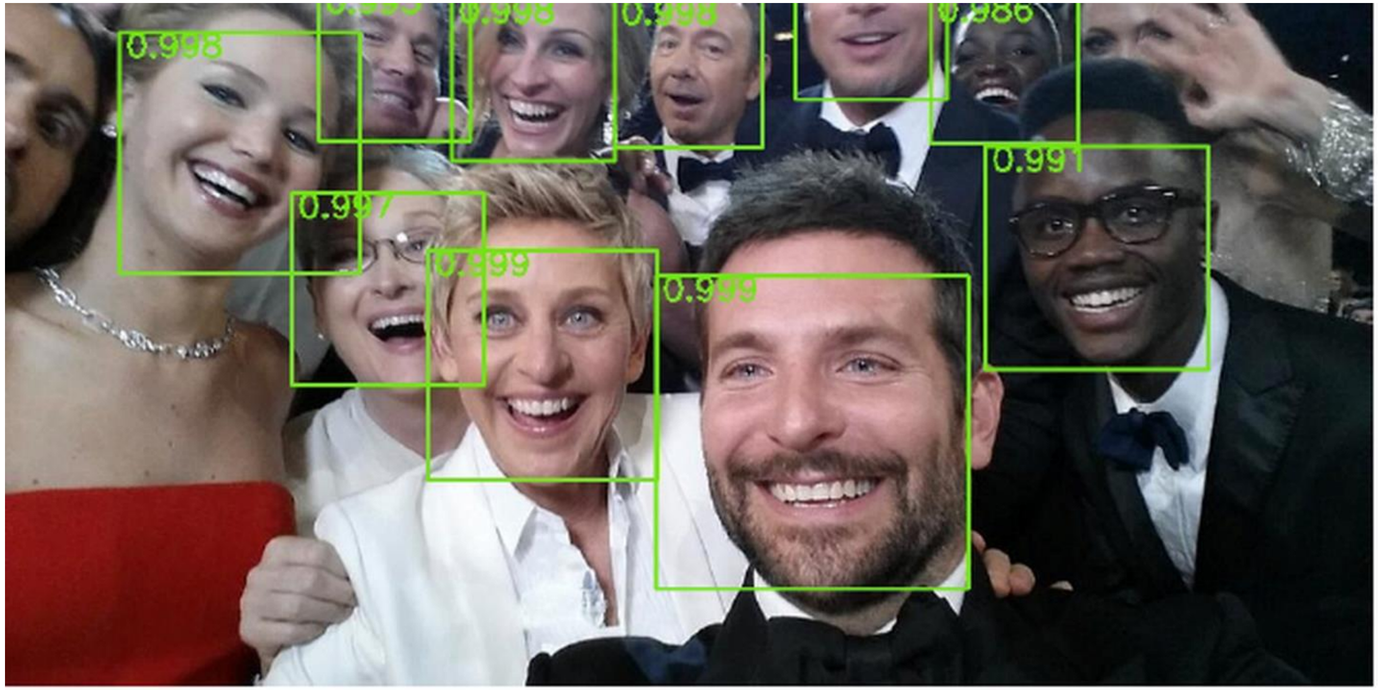


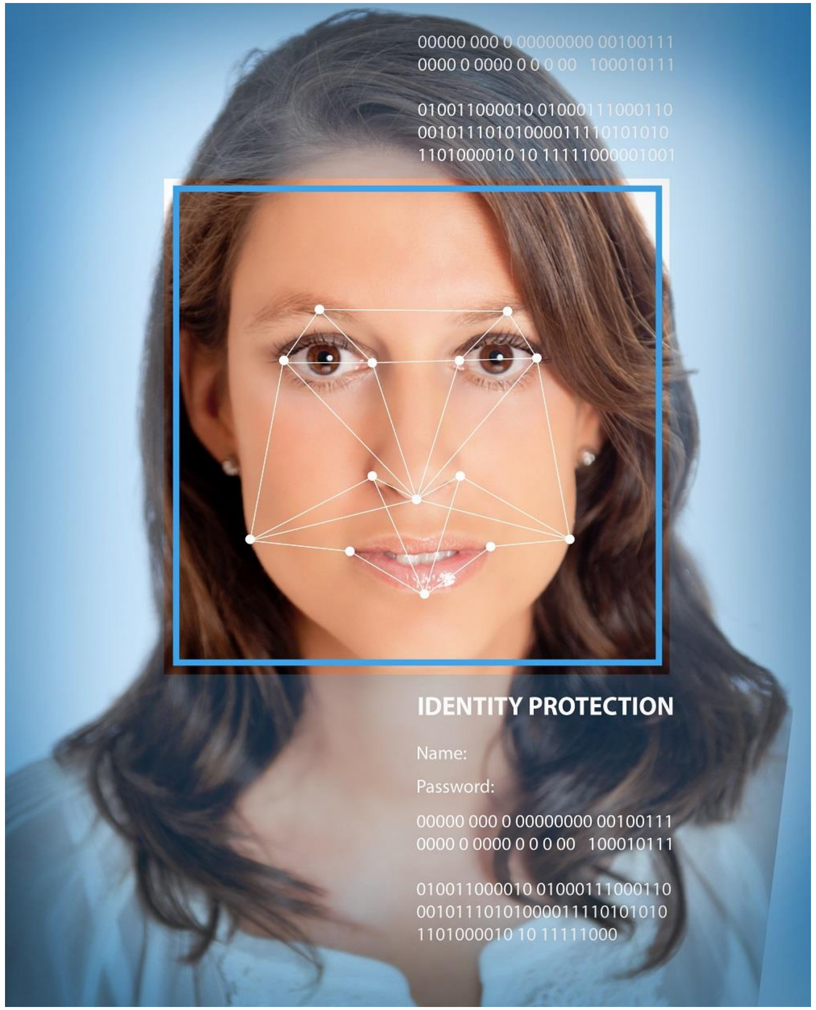
[Machine Learning – Stanford University]

Game AI

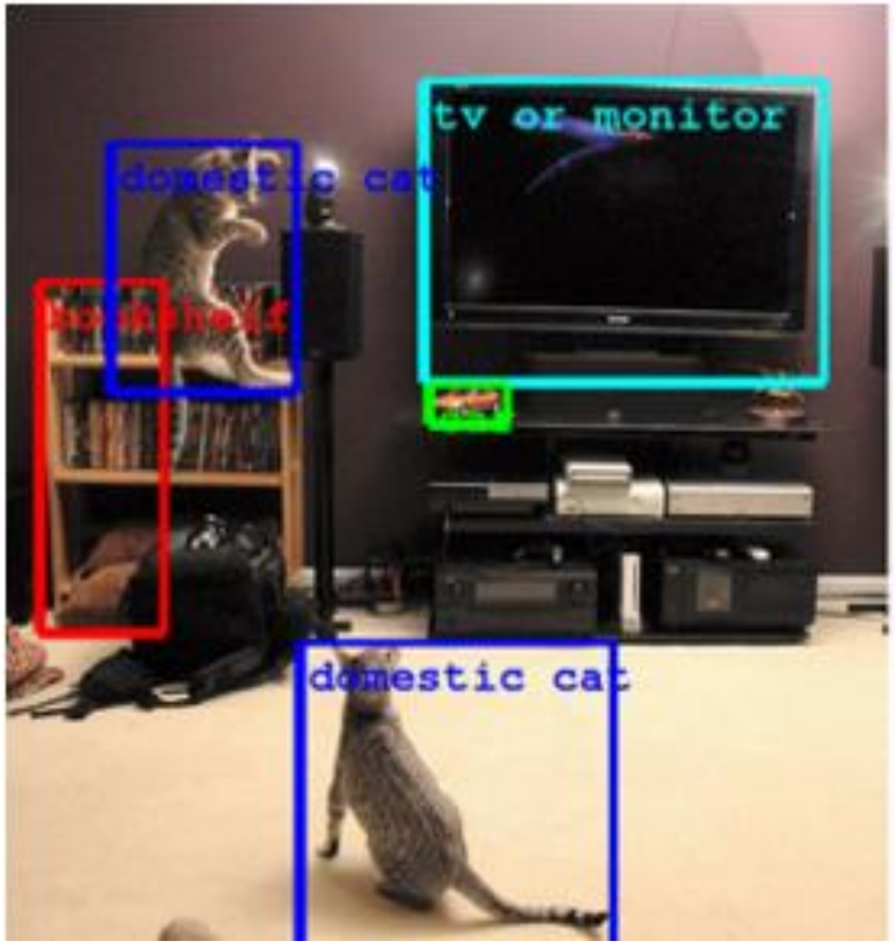


Face Detection & Recognition





Object Detection & Recognition



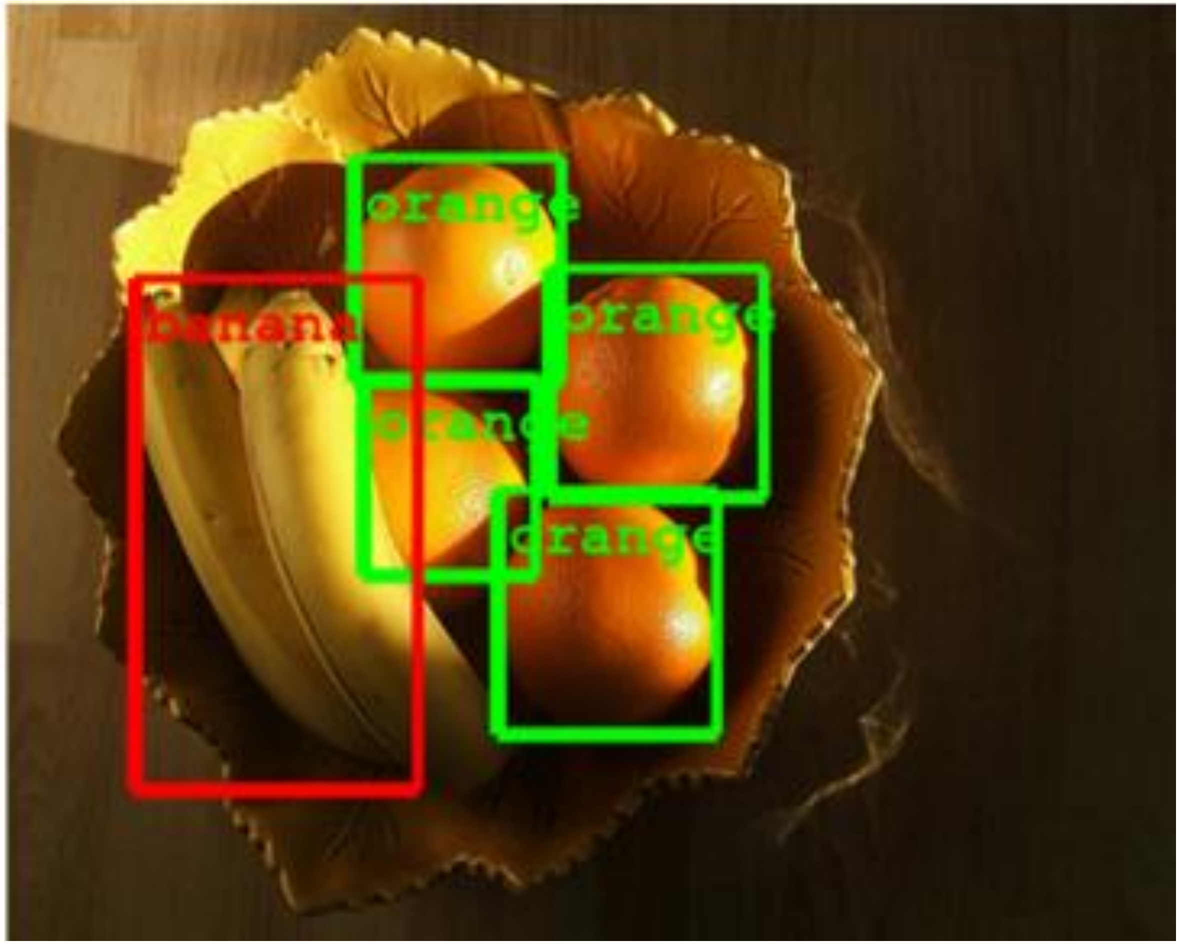
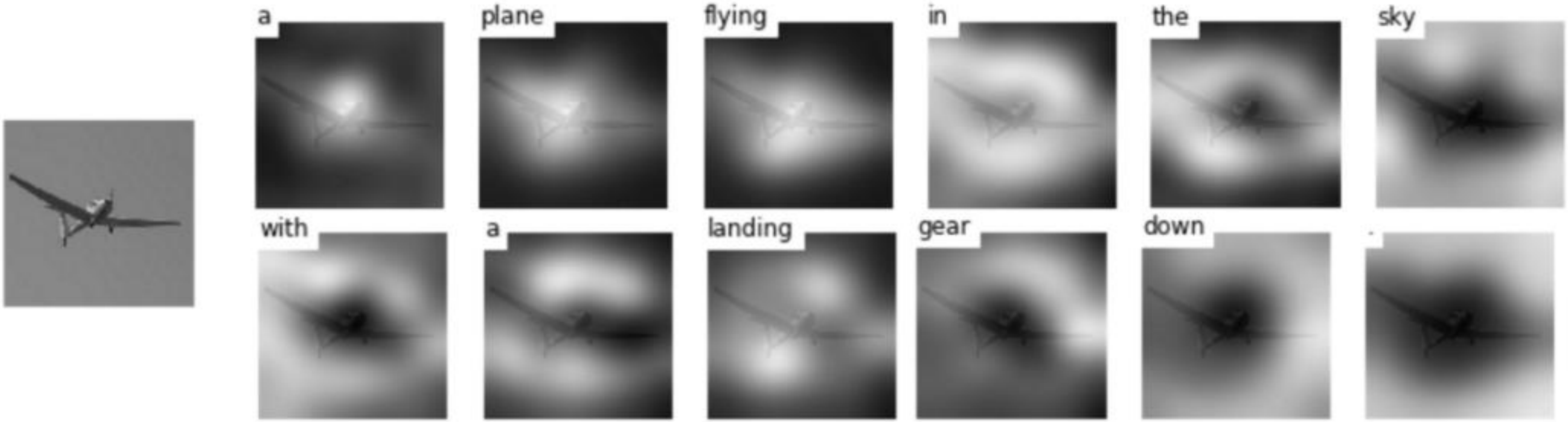
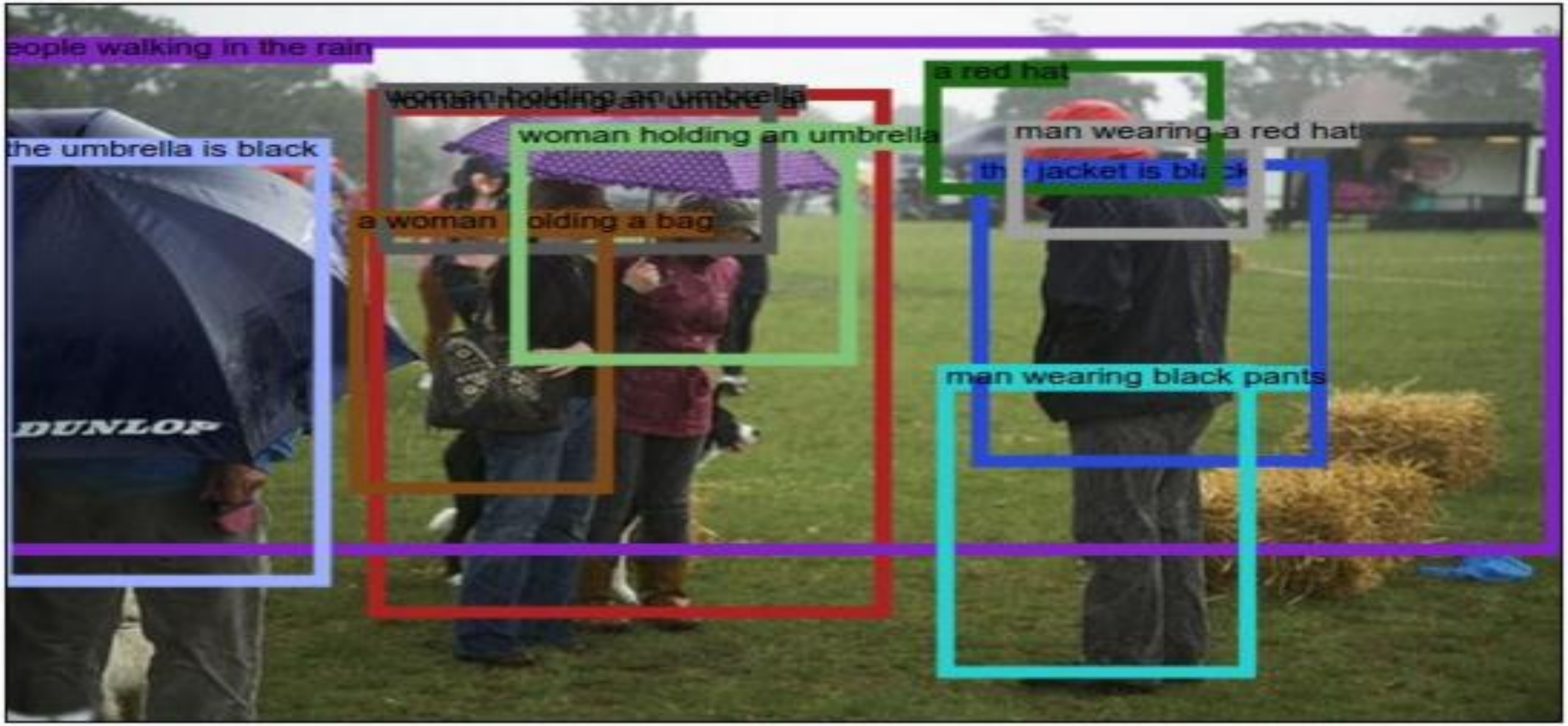


Image Captioning



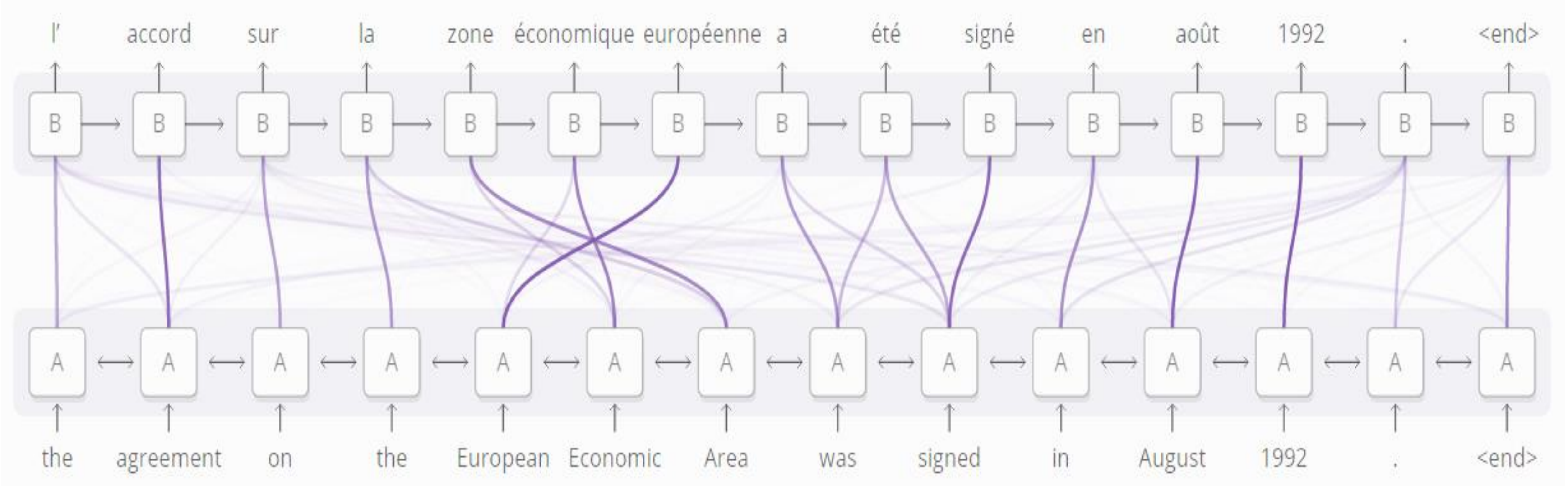
https://github.com/yunjey/show-attend-and-tell

Image Captioning

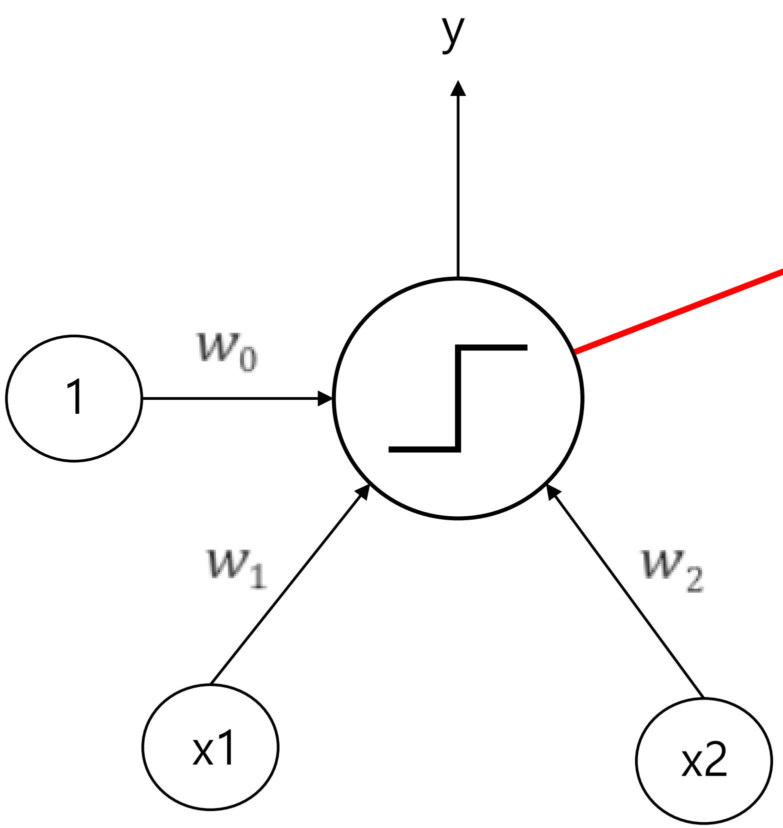


[DenseCap: Fully Convolutional Localization Networks for Dense Captioning]

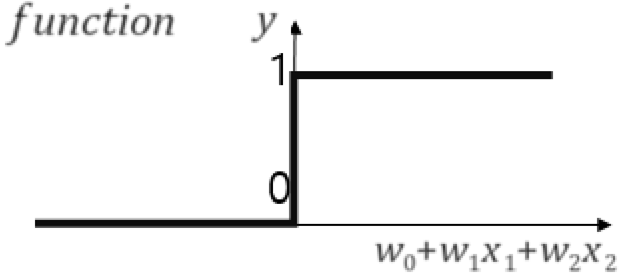
Machine Translation



Single Layer Perceptron



hard thresholding function



$$y = \begin{cases} 1 & w_0+w_1x_1+w_2x_2 \geq 0 \\ 0 & otherwise \end{cases}$$

Inputs: x1, x2

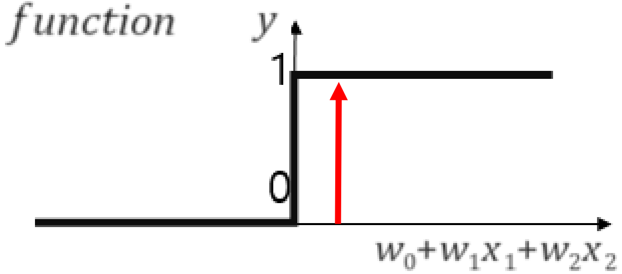
Output: y

Weights: w0, w1, w2

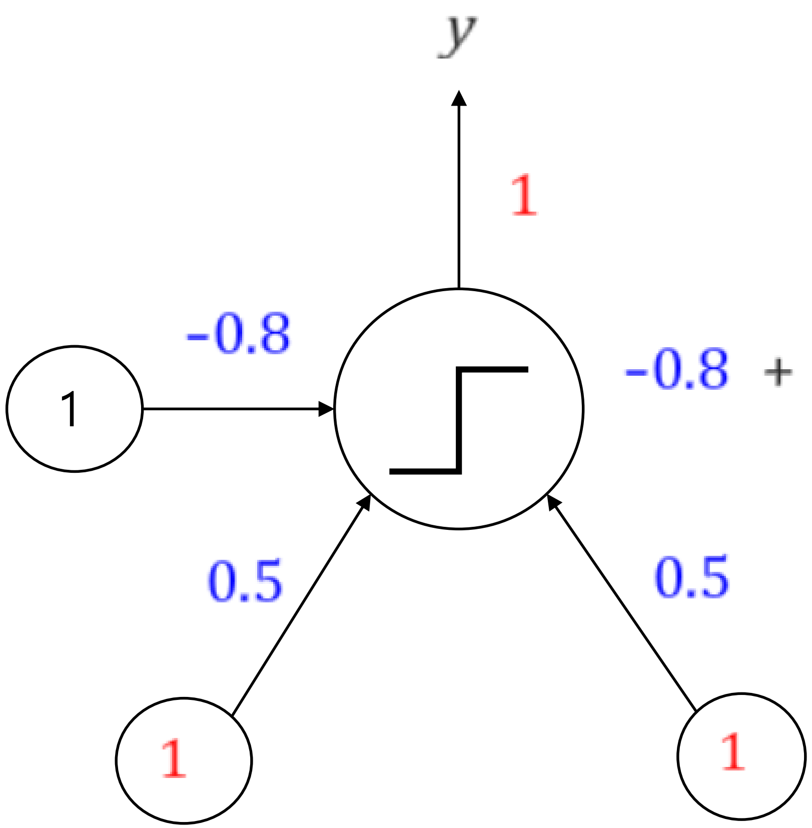
Bias: 1

Single Layer Perceptron

hard thresholding function



$$y = \begin{cases} 1 & w_0+w_1x_1+w_2x_2 \geq 0 \\ 0 & otherwise \end{cases}$$

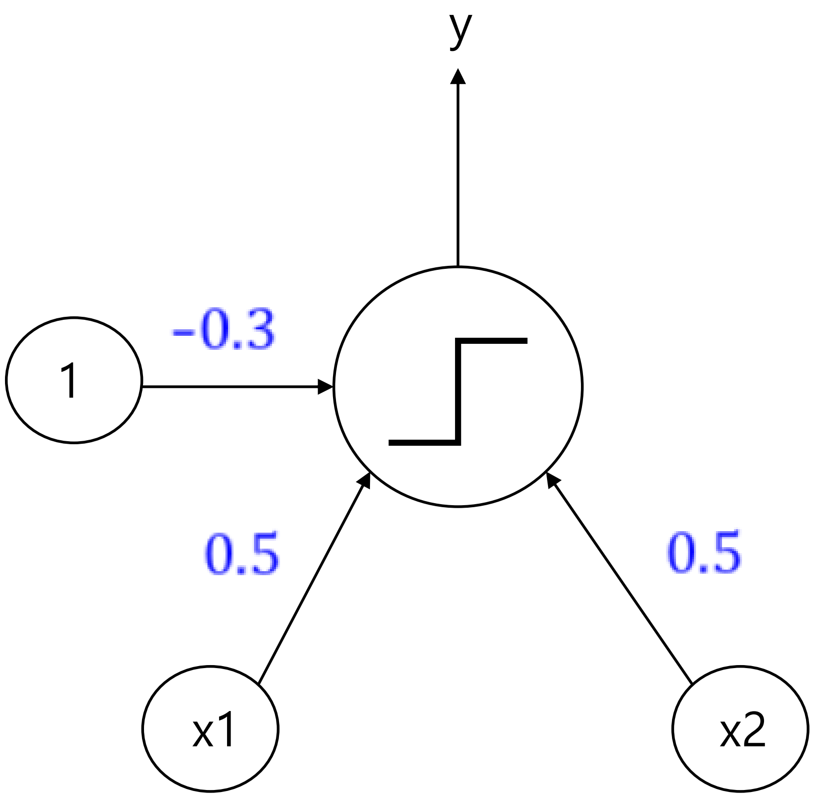


$$-0.8 + 0.5\times 1 + 0.5\times 1 = 0.2$$

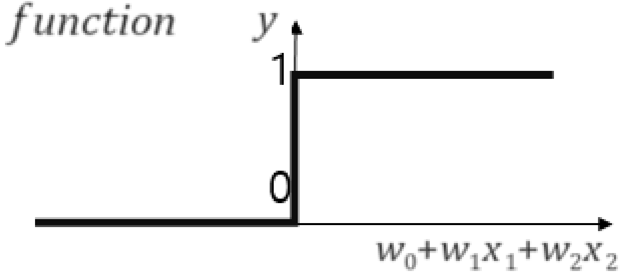
AND Gate

x_1	x_2	y
0	0	0
0	1	0
1	0	0
1	1	1

Single Layer Perceptron



hard thresholding function



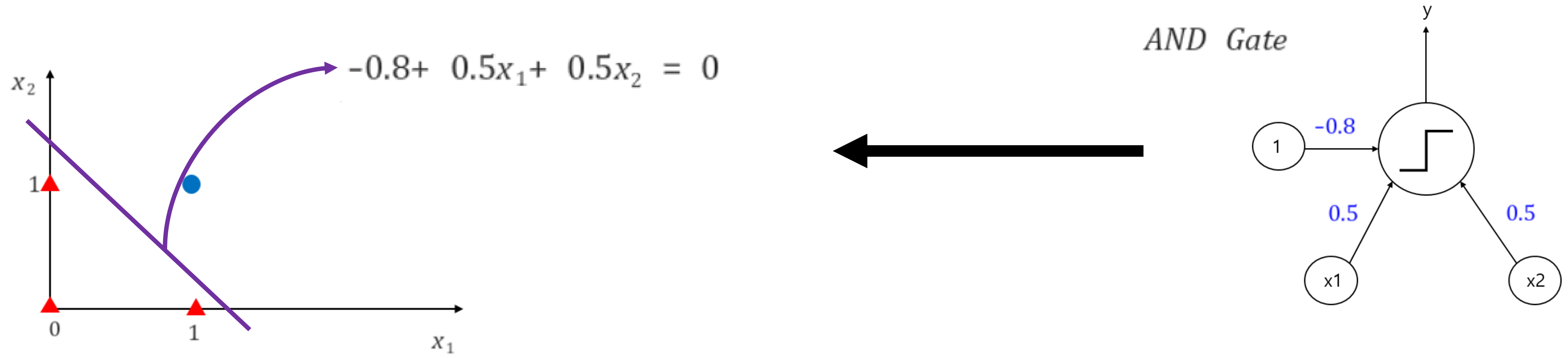
$$y = \begin{cases} 1 & w_0+w_1x_1+w_2x_2 \geq 0 \\ 0 & otherwise \end{cases}$$

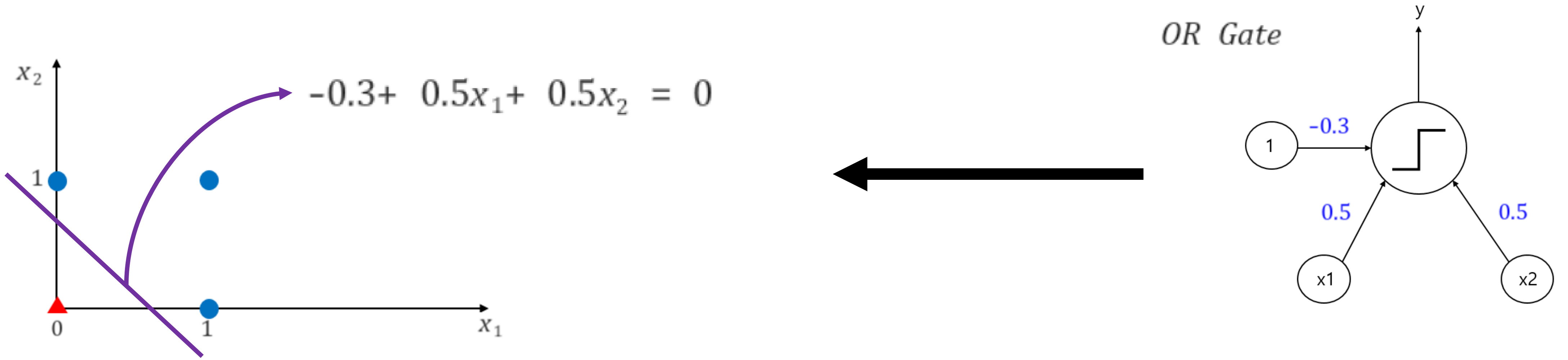
OR Gate

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	1

Single Layer Perceptron

● : 1 ▲ : 0

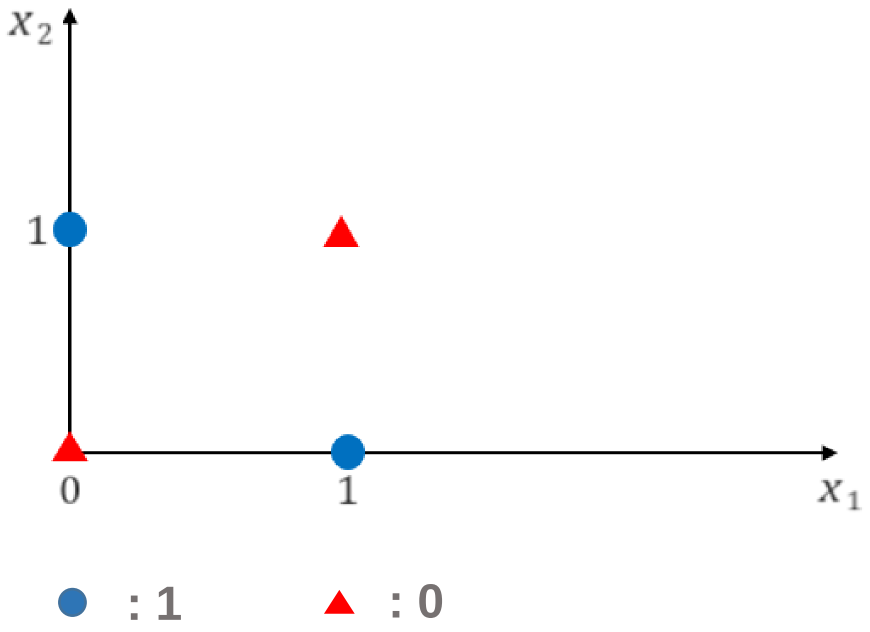




XOR Gate

Is it possible to solve XOR problem using a single layer perceptron?

No. Single layer perceptron can only solve linear problem. XOR problem is non-linear.

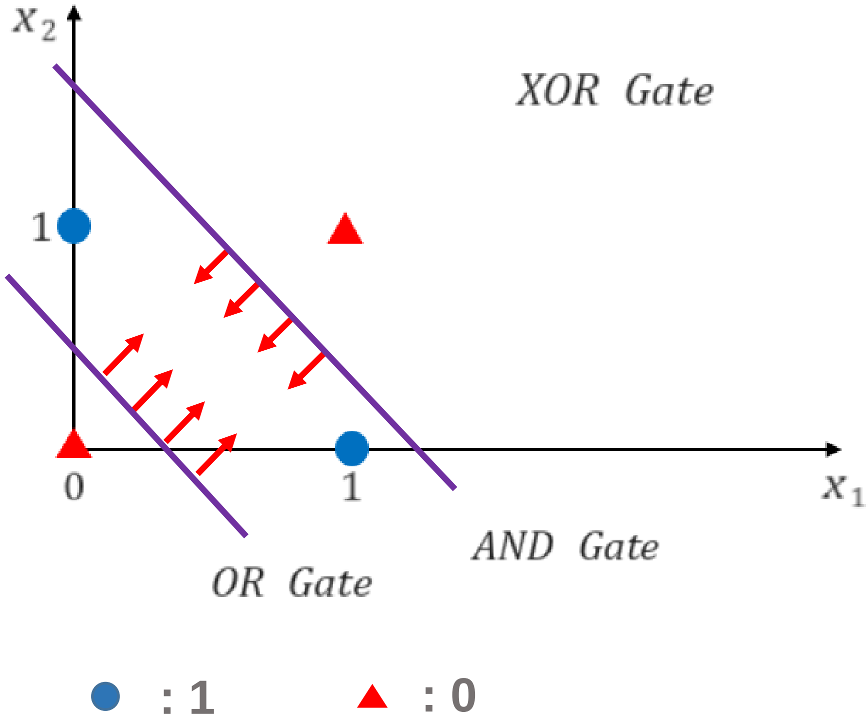


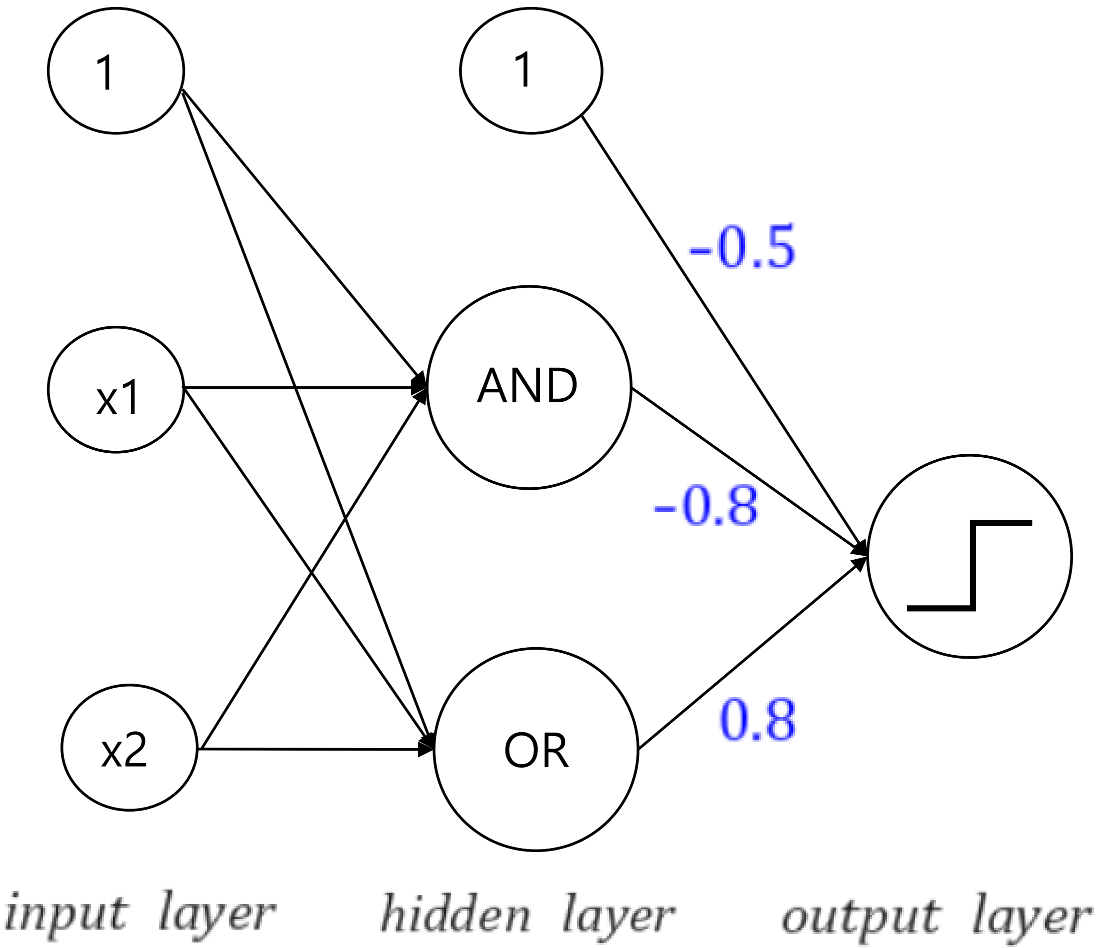
$XOR\ \ Gate$

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

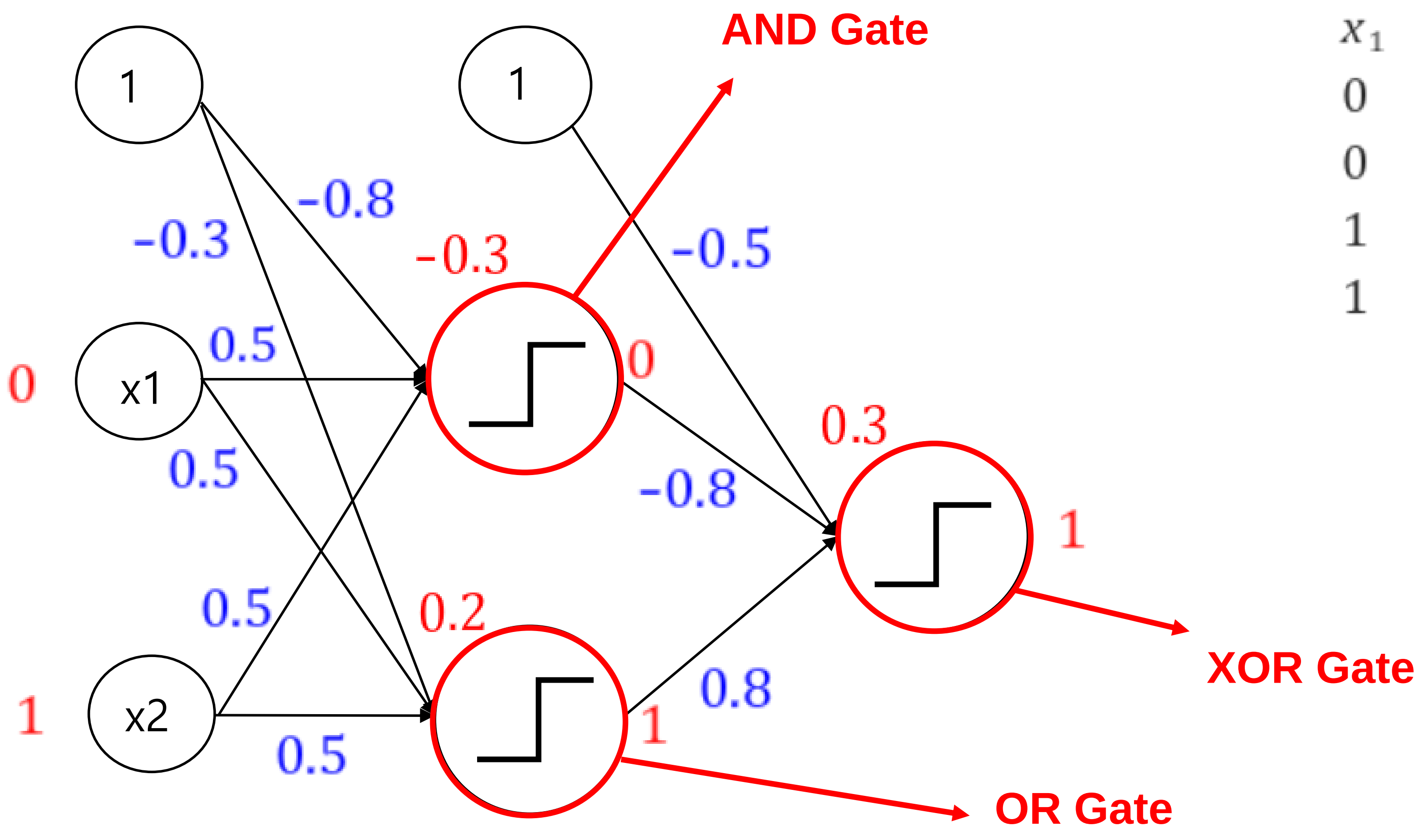
Multi Layer Perceptron

But if we use 2 single layer perceptron, we can solve XOR problem. This model is called multi layer perceptron.





Multi Layer Perceptron



OR Gate

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	1

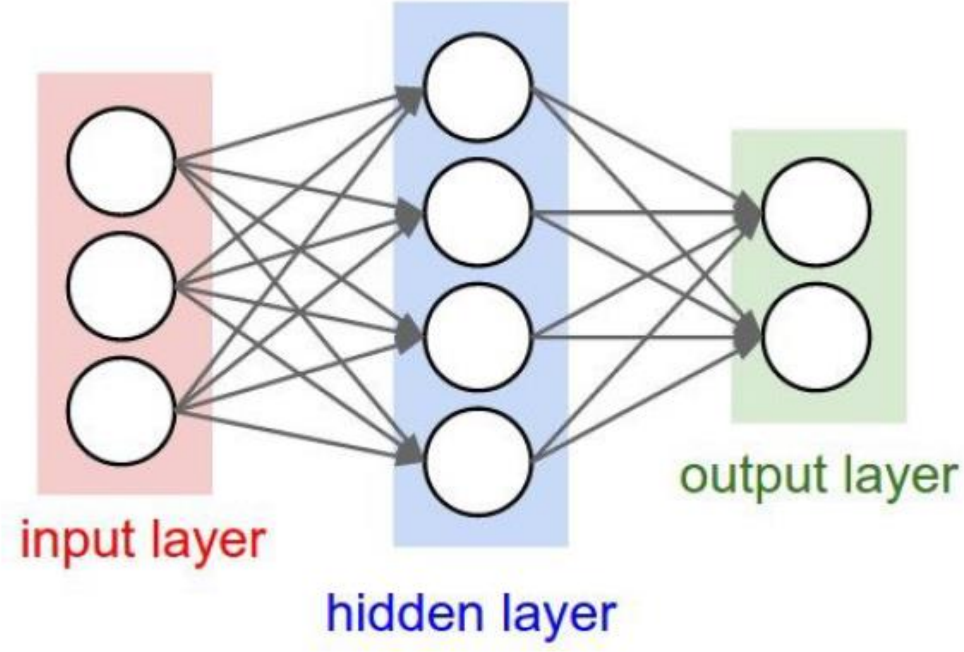
AND Gate

x_1	x_2	y
0	0	0
0	1	0
1	0	0
1	1	1

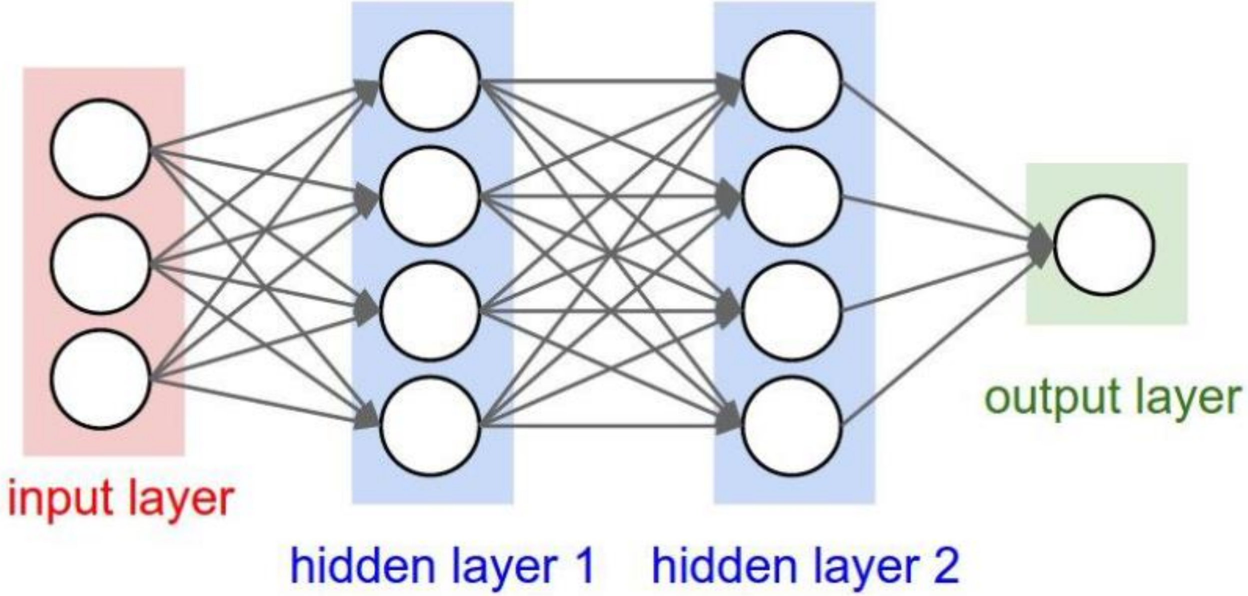
XOR Gate

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

Neural Network Architecture

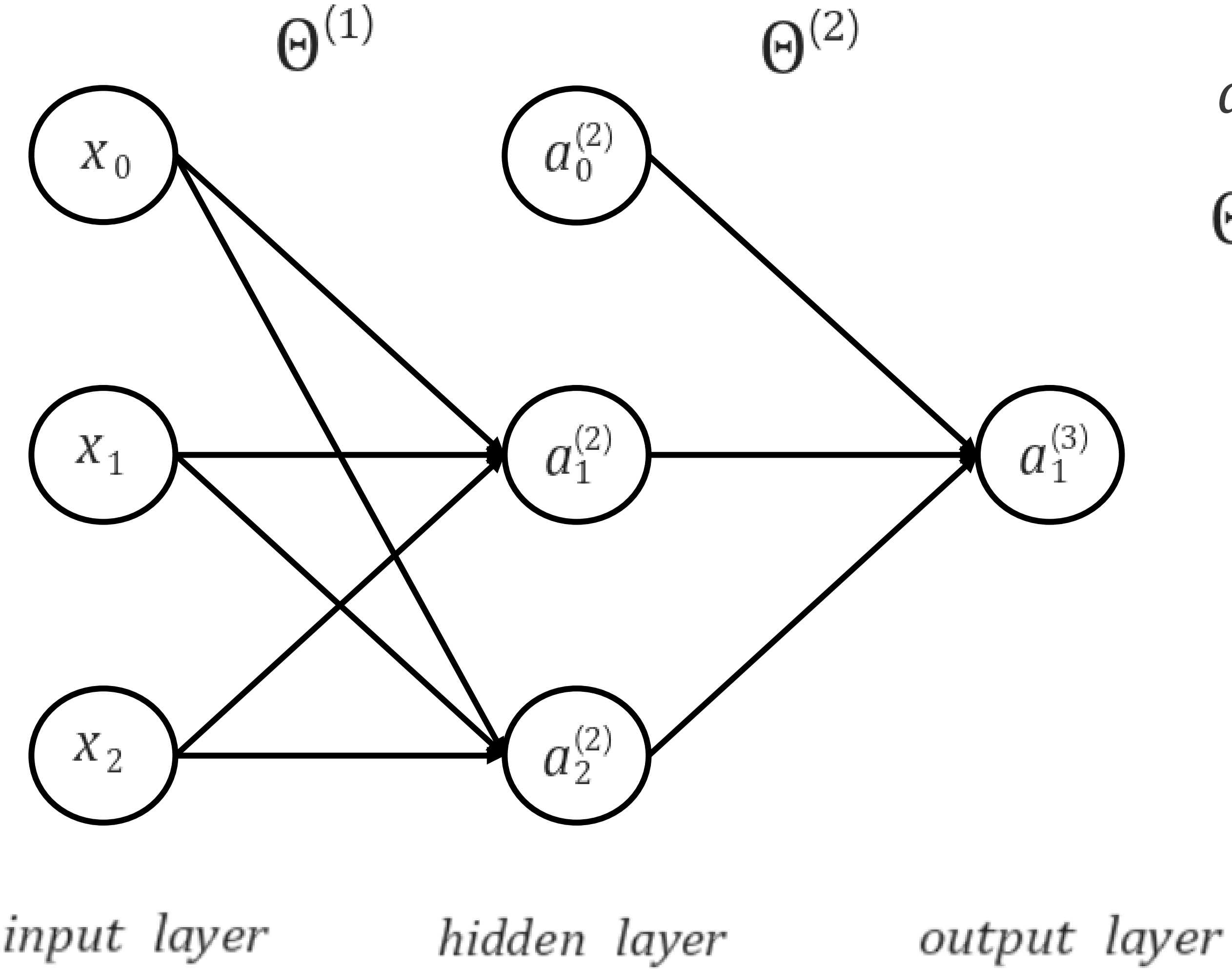


“2-layer Neural Network”
or
“1-hidden-layer Neural Network”



“3-layer Neural Network”
or
“2-hidden-layer Neural Network”

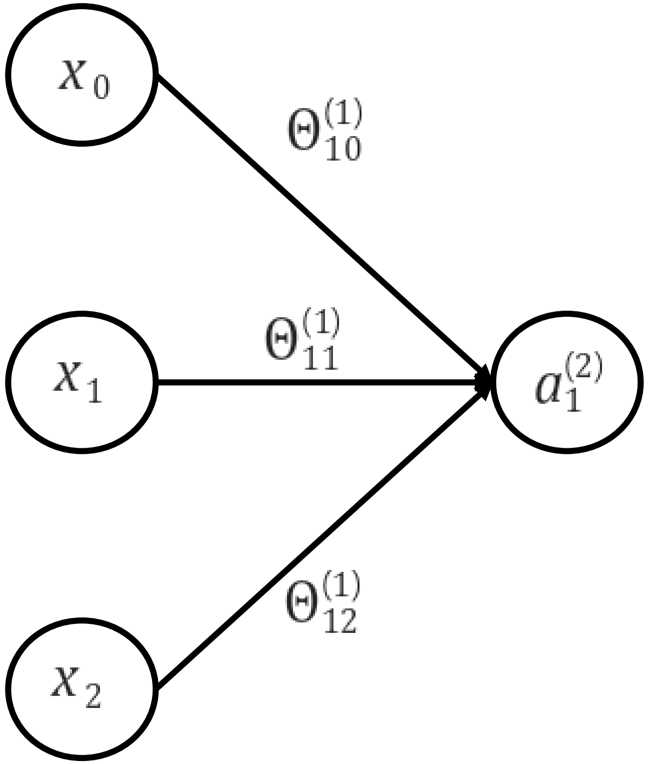
Forward Propagation



a_i^j : "activation" of i-th unit in j-th layer

$\Theta^{(j)}$: matrix of weights controlling function mapping form j-th layer to (j+1)-th layer

Forward Propagation

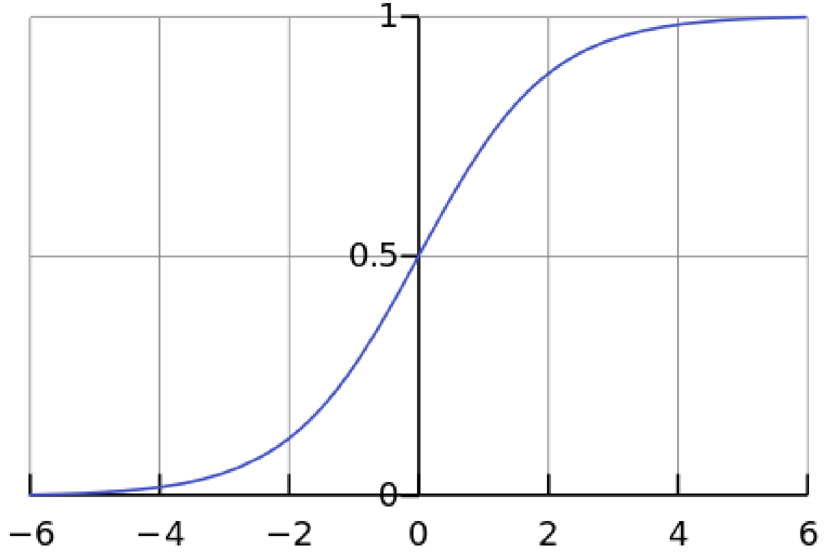


$$z_1^{(2)} = \Theta_{10}^{(1)} x_0 + \Theta_{11}^{(1)} x_1 + \Theta_{12}^{(1)} x_2$$

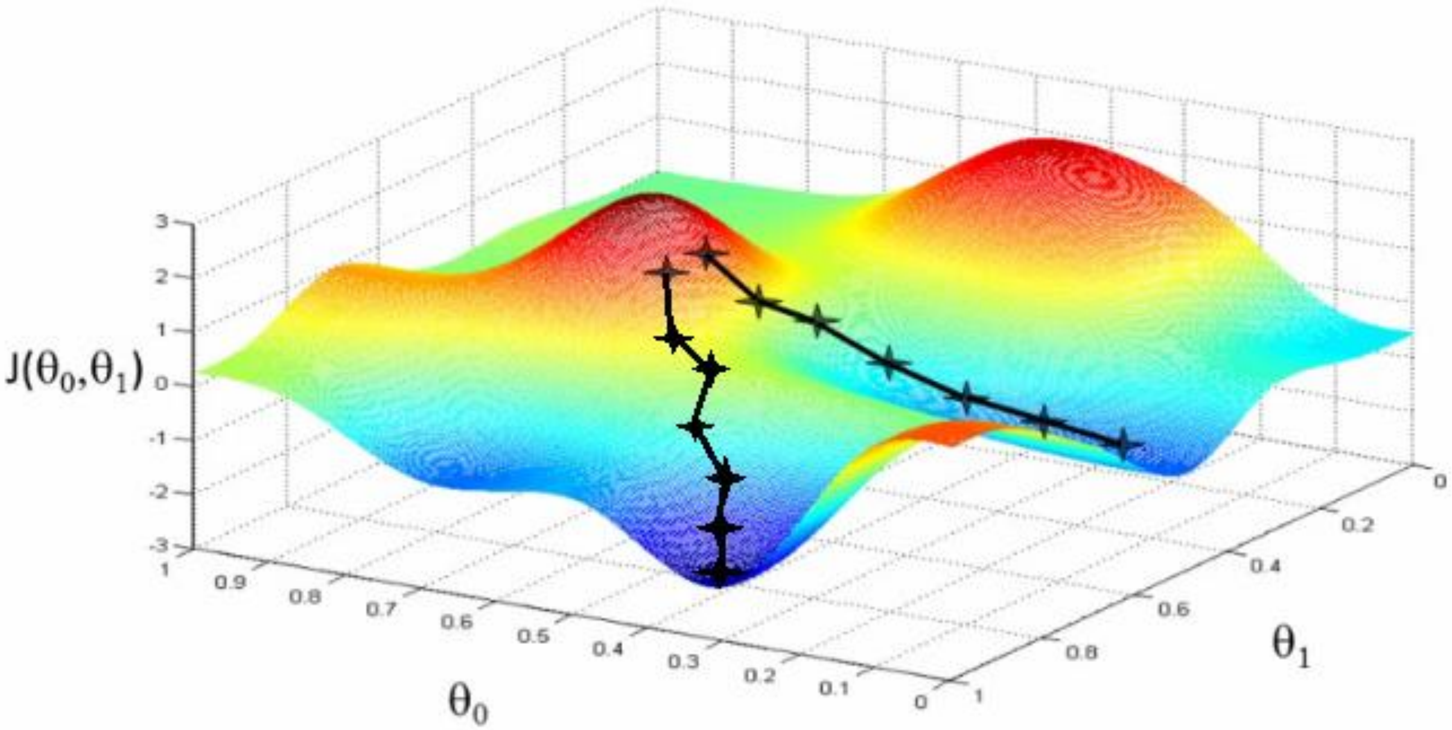
$$a_1^{(2)} = g(z_1^{(2)})$$

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

logistic sigmoid



Gradient Descent



$$compute \quad \frac{\partial J}{\partial \theta_0} \quad and \quad \frac{\partial J}{\partial \theta_1}$$

$$update \;\; weights \;\; with$$

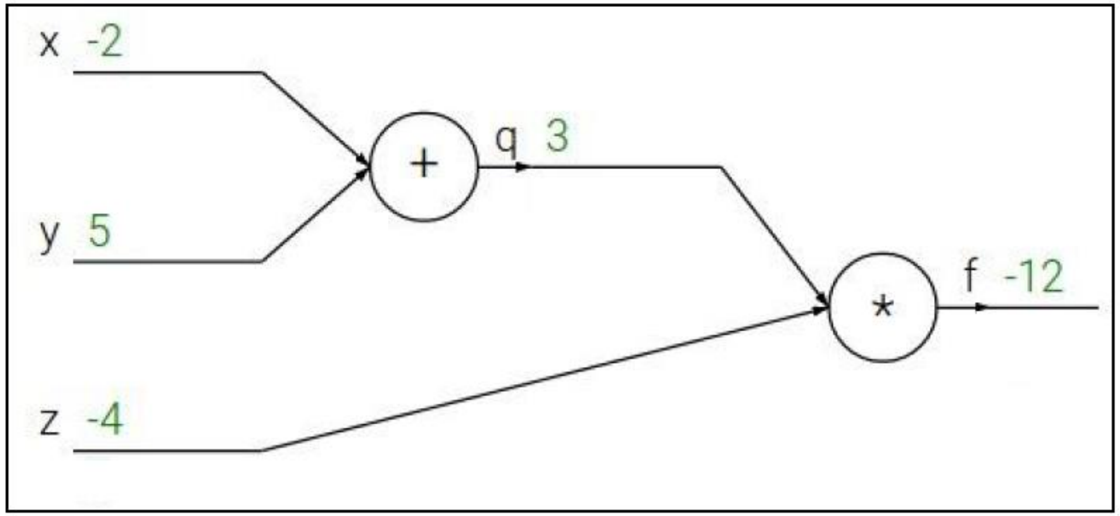
$$\boxed{\begin{aligned} \theta_0 &:= \theta_0 - \alpha \cdot \frac{\partial J}{\partial \theta_0} \\ \theta_1 &:= \theta_1 - \alpha \cdot \frac{\partial J}{\partial \theta_1} \end{aligned}}$$

Gradient “descent” optimization with learning rate ‘alpha’ (e.g. 0.01)

Computational Graph

$$f(x, y, z) = (x + y)z$$

e.g. x = -2, y = 5, z = -4



Computational Graph

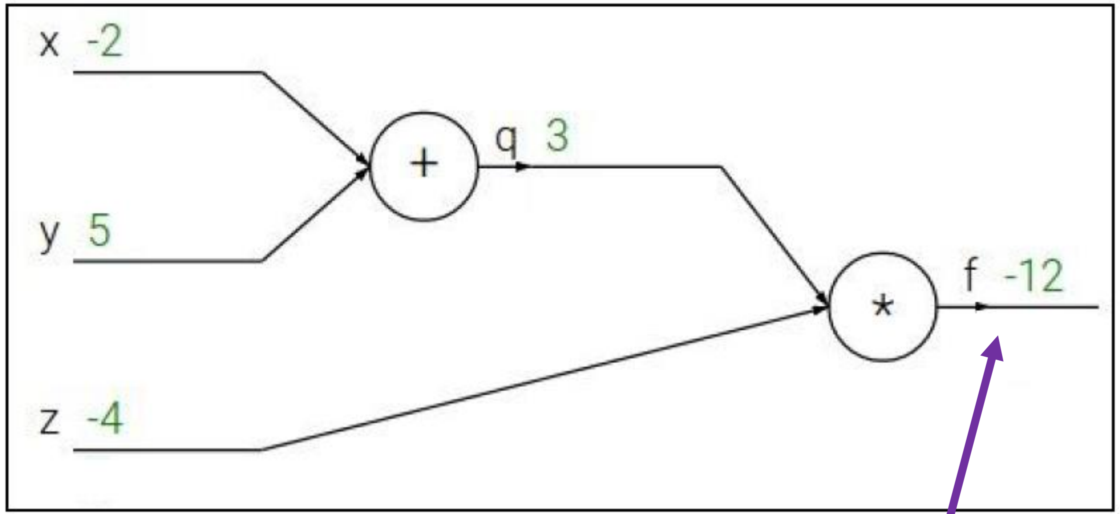
$$f(x, y, z) = (x + y)z$$

e.g. x = -2, y = 5, z = -4

$$q = x + y \qquad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \qquad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



$$\frac{\partial f}{\partial f}$$

Computational Graph

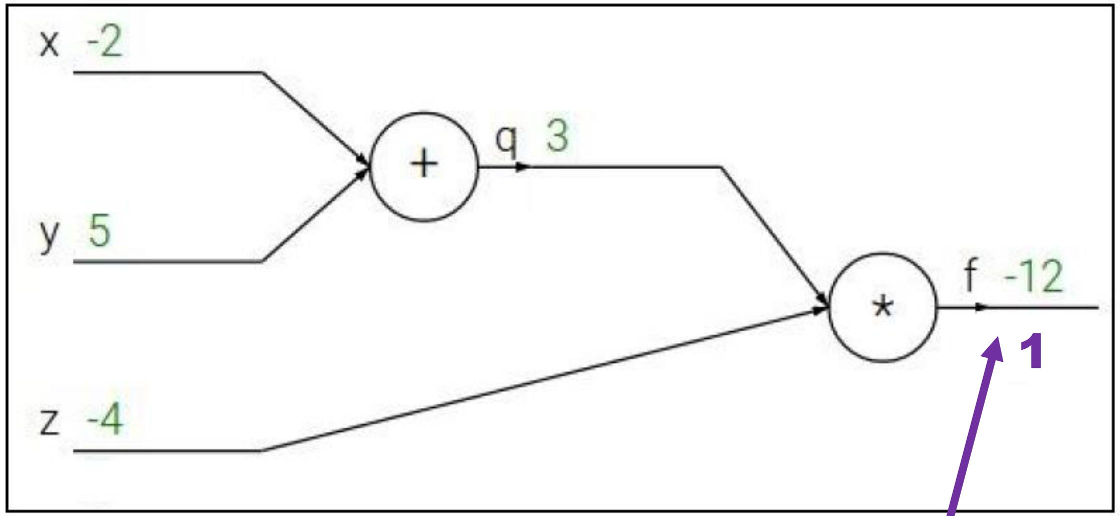
$$f(x, y, z) = (x + y)z$$

e.g. x = -2, y = 5, z = -4

$$q = x + y \qquad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \qquad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



$$\frac{\partial f}{\partial f}$$

Backpropagation using Chain Rule

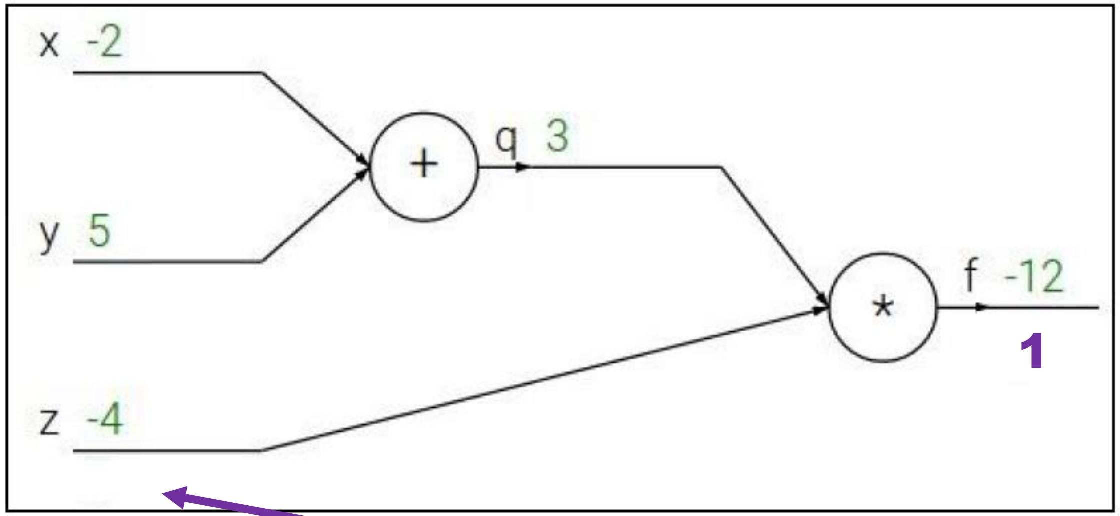
$$f(x, y, z) = (x + y)z$$

e.g. x = -2, y = 5, z = -4

$$q = x + y \qquad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \qquad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



$$\boxed{\frac{\partial f}{\partial z}}$$

Backpropagation using Chain Rule

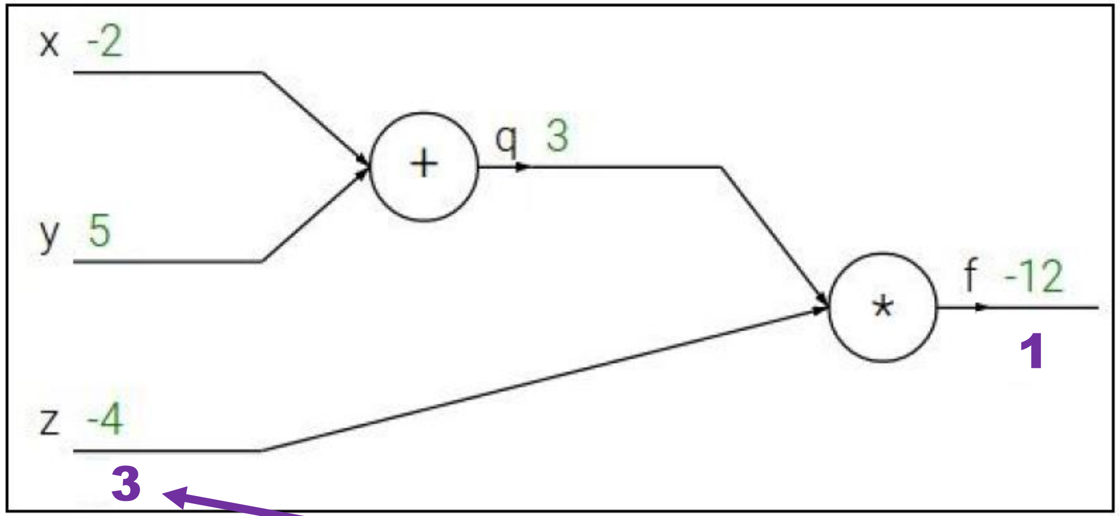
$$f(x, y, z) = (x + y)z$$

e.g. x = -2, y = 5, z = -4

$$q = x + y \qquad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \qquad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



$$\boxed{\frac{\partial f}{\partial z}}$$

Computational Graph

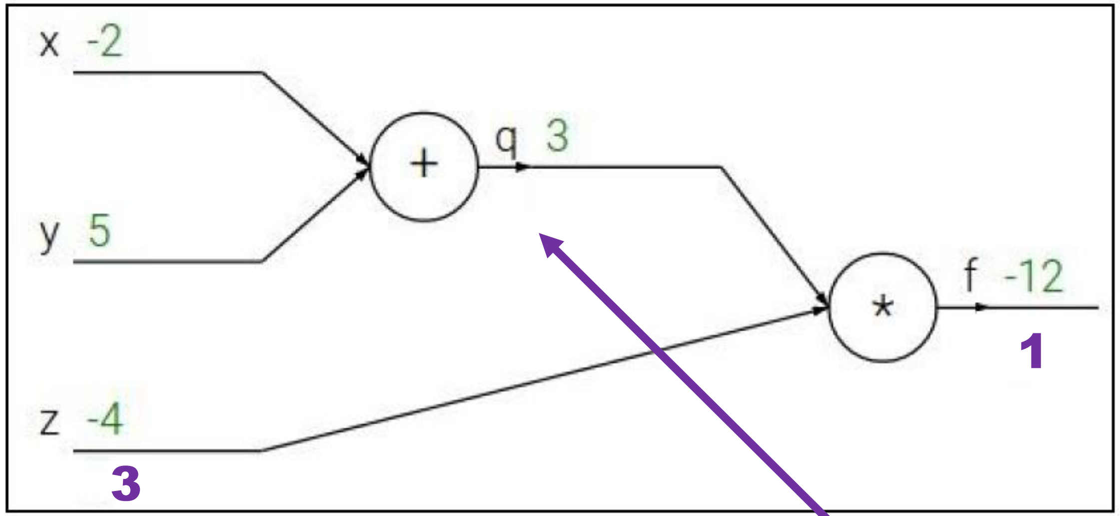
$$f(x, y, z) = (x + y)z$$

e.g. x = -2, y = 5, z = -4

$$q = x + y \qquad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \qquad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



$$\boxed{\frac{\partial f}{\partial q}}$$

Backpropagation using Chain Rule

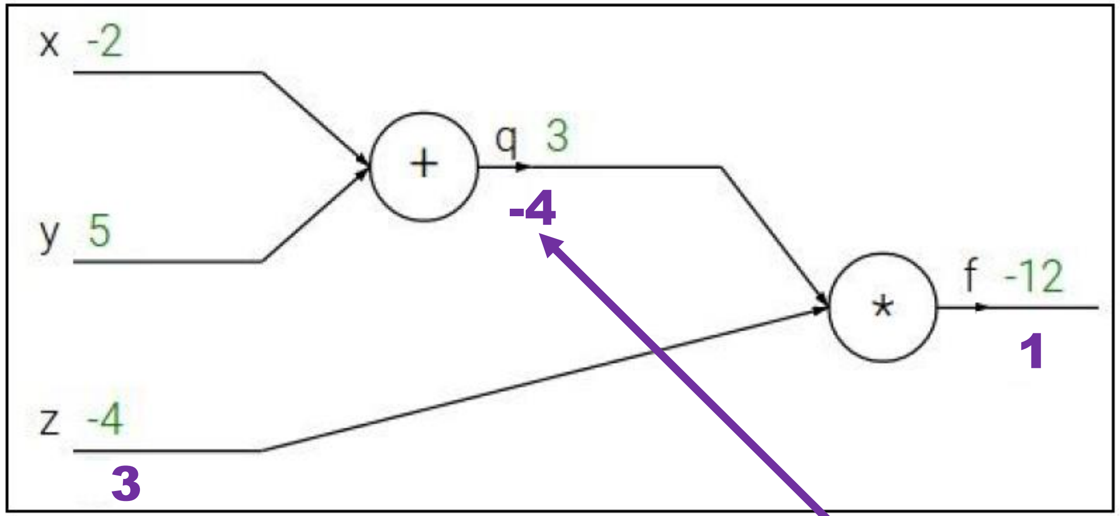
$$f(x, y, z) = (x + y)z$$

e.g. x = -2, y = 5, z = -4

$$q = x + y \qquad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \qquad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Want: $\dfrac{\partial f}{\partial x}, \dfrac{\partial f}{\partial y}, \dfrac{\partial f}{\partial z}$



$$\boxed{\frac{\partial f}{\partial q}}$$

Backpropagation using Chain Rule

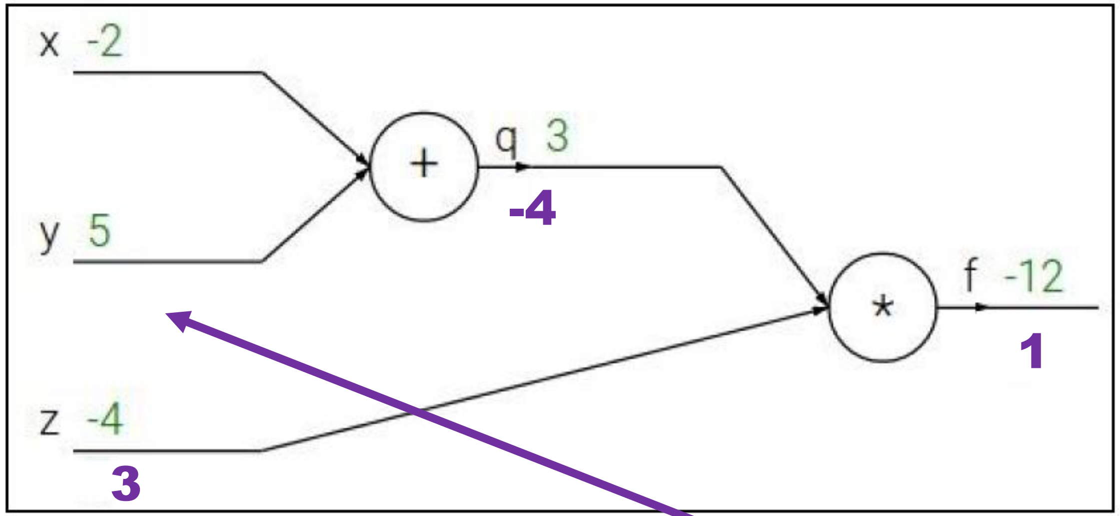
$$f(x, y, z) = (x + y)z$$

e.g. x = -2, y = 5, z = -4

$$q = x + y \qquad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \qquad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



$$\boxed{\frac{\partial f}{\partial y}}$$

Backpropagation using Chain Rule

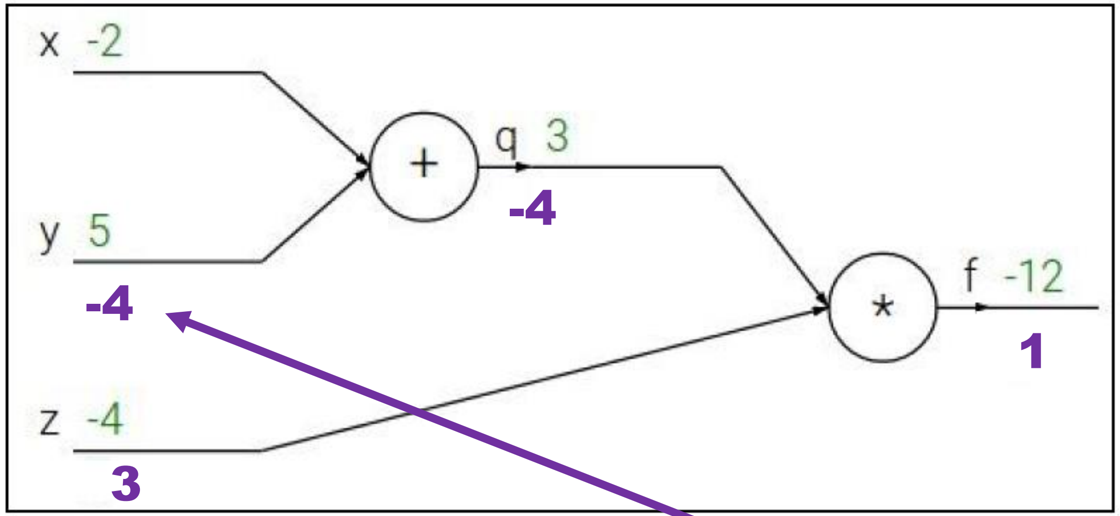
$$f(x, y, z) = (x + y)z$$

e.g. x = -2, y = 5, z = -4

$$q = x + y \qquad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \qquad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



$$\boxed{\frac{\partial f}{\partial y}}$$

Chain rule:

$$\boxed{\frac{\partial f}{\partial y} = \frac{\partial f}{\partial q} \frac{\partial q}{\partial y}}$$

Backpropagation using Chain Rule

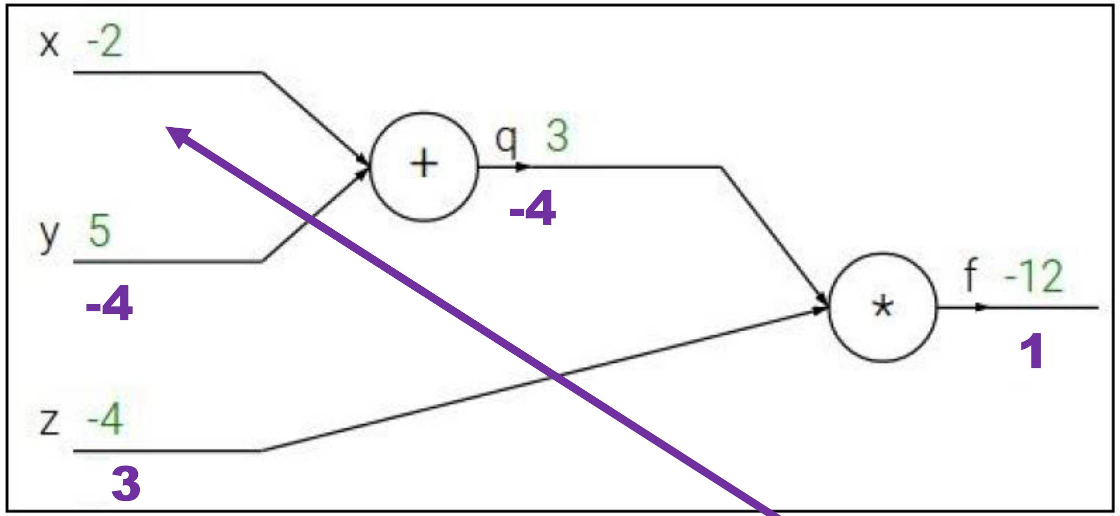
$$f(x, y, z) = (x + y)z$$

e.g. x = -2, y = 5, z = -4

$$q = x + y \qquad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \qquad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$

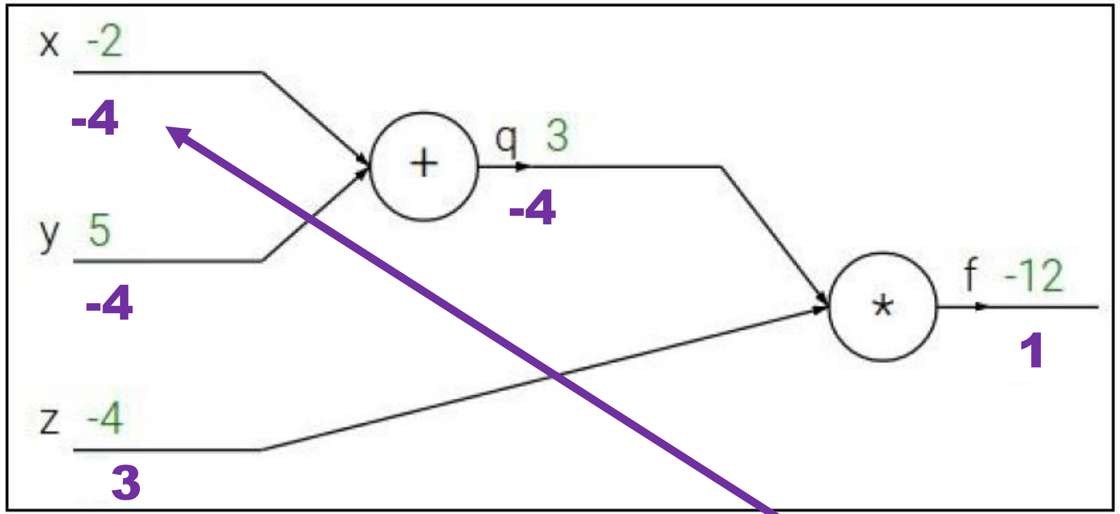


$$\frac{\partial f}{\partial x}$$

Backpropagation using Chain Rule

$$f(x, y, z) = (x + y)z$$

e.g. x = -2, y = 5, z = -4



$$\boxed{\frac{\partial f}{\partial x}}$$

$$q = x + y \qquad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \qquad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$

Chain rule:

$$\boxed{\frac{\partial f}{\partial x} = \frac{\partial f}{\partial q}\frac{\partial q}{\partial x}}$$