# Backpropagation on the loss function

**Kietikul Jearanaitanakij**

Department of Computer Engineering, KMITL

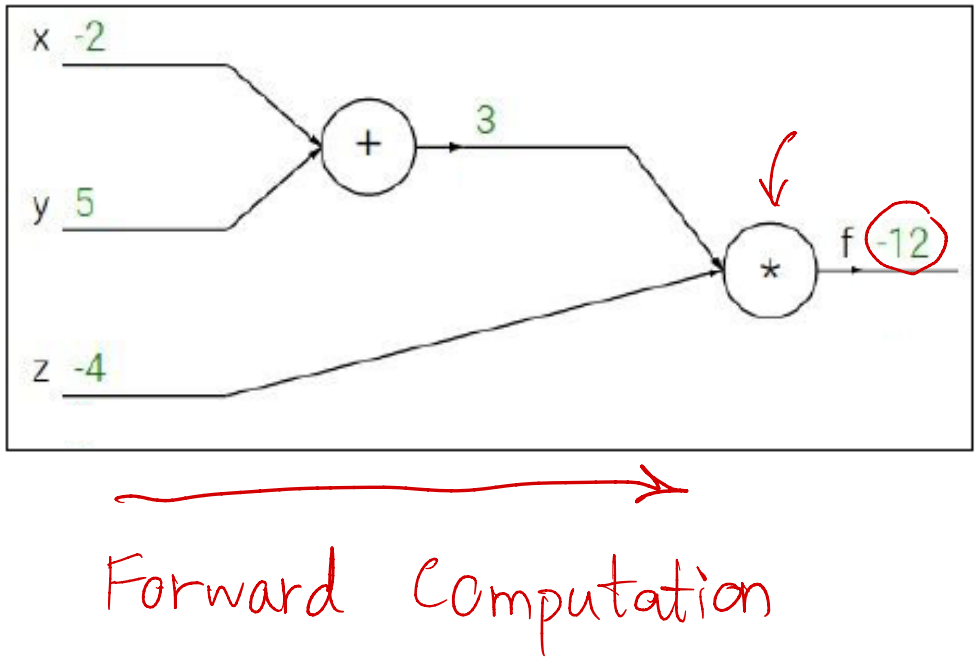
(Slides are adapted from cs231n @Stanford University)

# Backpropagation(Revisited)

Backpropagation: a simple example

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

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



Backpropagation: a simple example

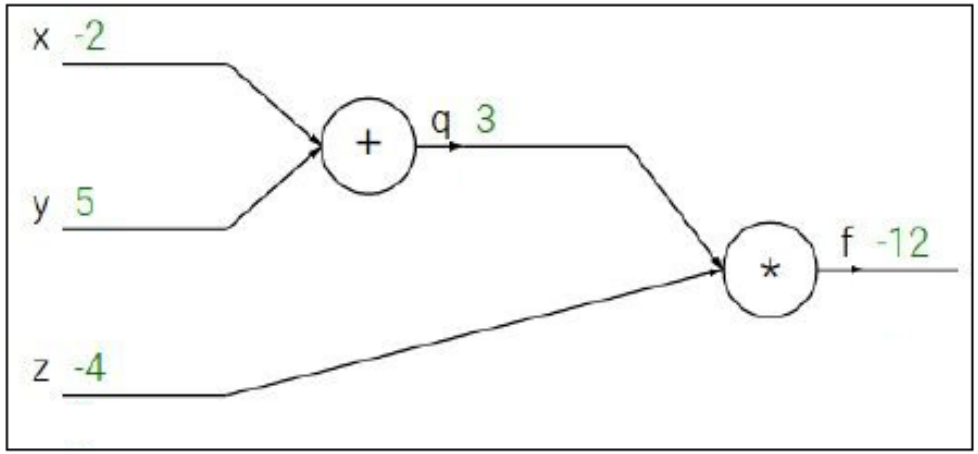
$$f(x, y, z) = (x + y)z$$

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



## Backpropagation: a simple example

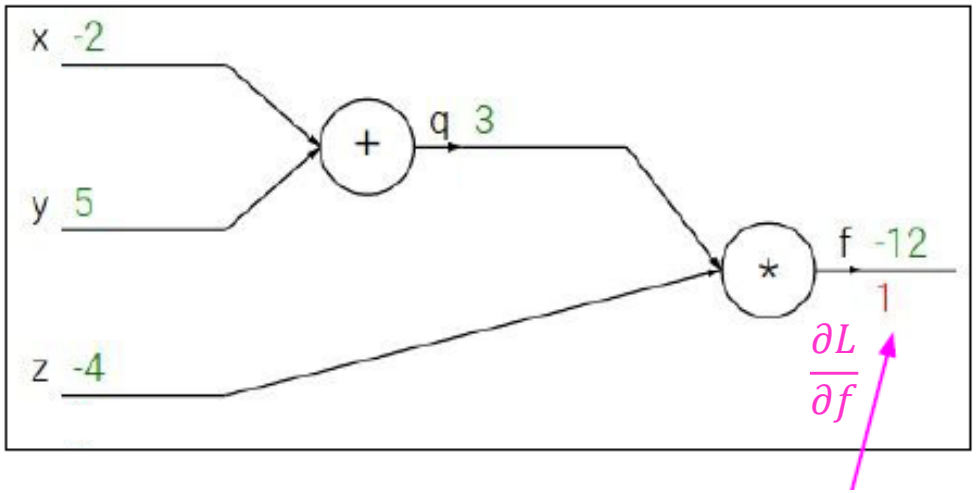
$$f(x, y, z) = (x + y)z$$

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




Suppose gradient of Loss wrt f equals to 1

Backpropagation: a simple example

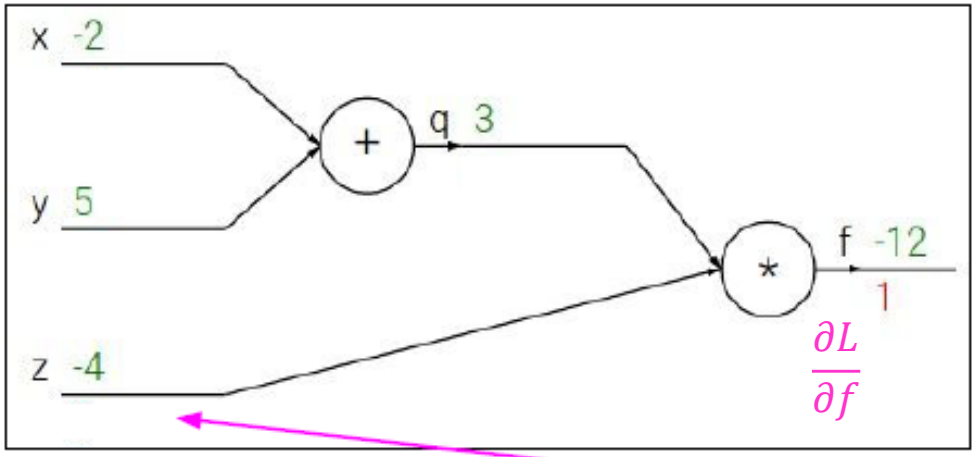
$$f(x, y, z) = (x + y)z$$

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




$$\frac{\partial L}{\partial z} = \frac{\partial L}{\partial f} \cdot \frac{\partial f}{\partial z}$$

Chain rule:

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

Backpropagation: a simple example

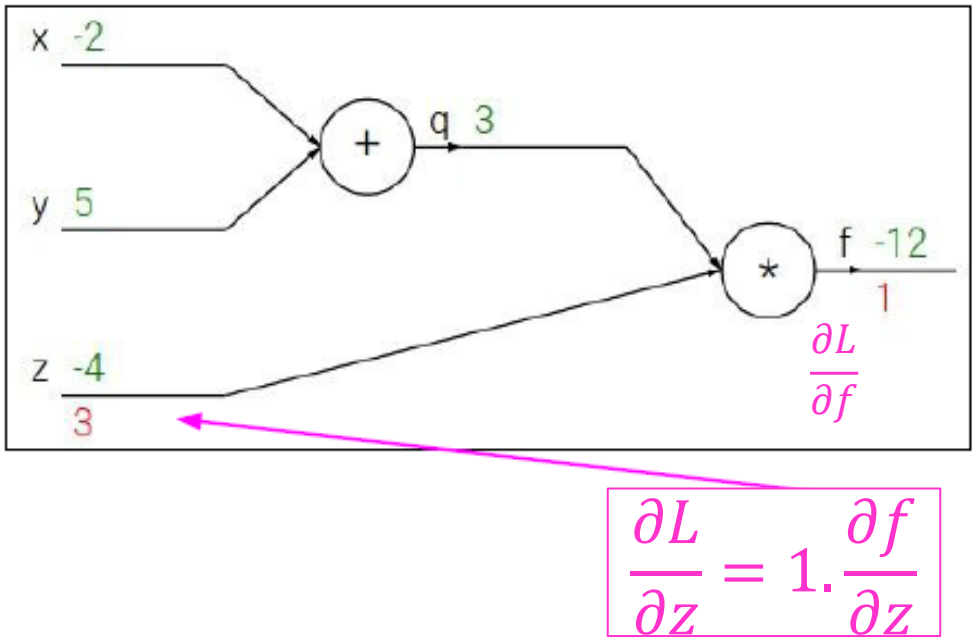
$$f(x, y, z) = (x + y)z$$

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



Backpropagation: a simple example

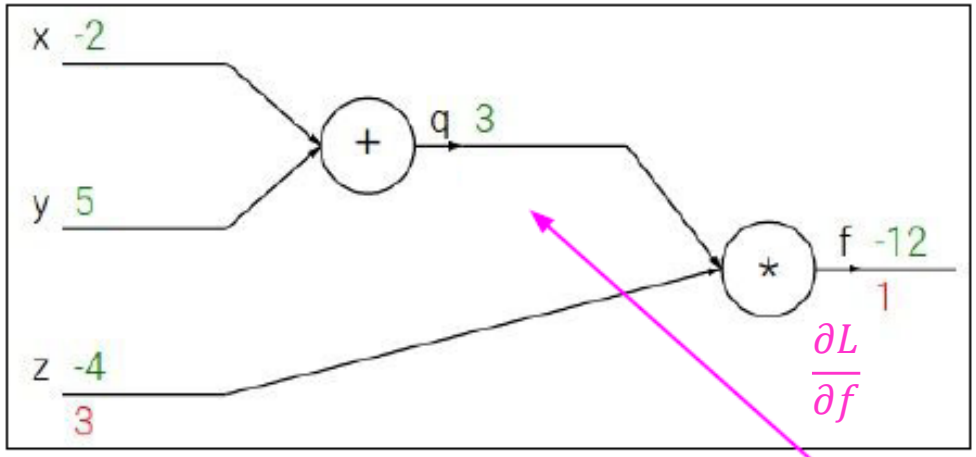
$$f(x, y, z) = (x + y)z$$

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




$$\frac{\partial L}{\partial q} = \frac{\partial L}{\partial f} \cdot \frac{\partial f}{\partial q}$$

Backpropagation: a simple example

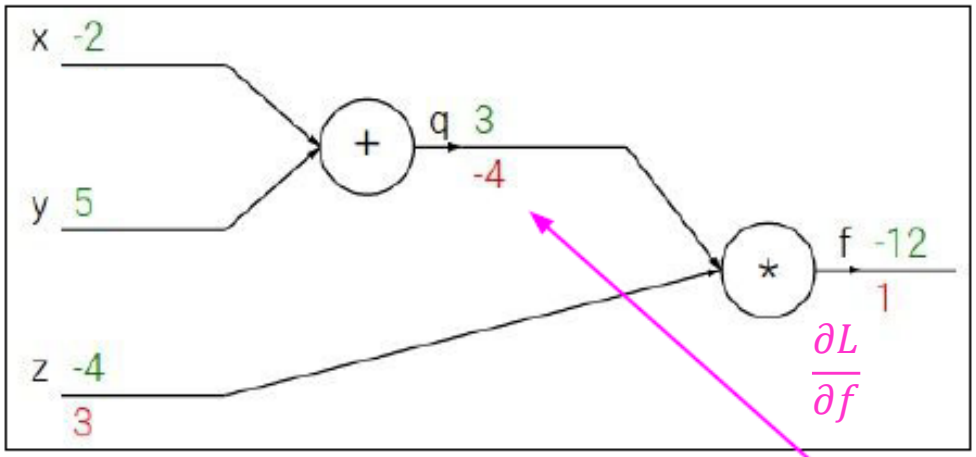
$$f(x, y, z) = (x + y)z$$

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




$$\frac{\partial L}{\partial q} = 1.\frac{\partial f}{\partial q}$$

Backpropagation: a simple example

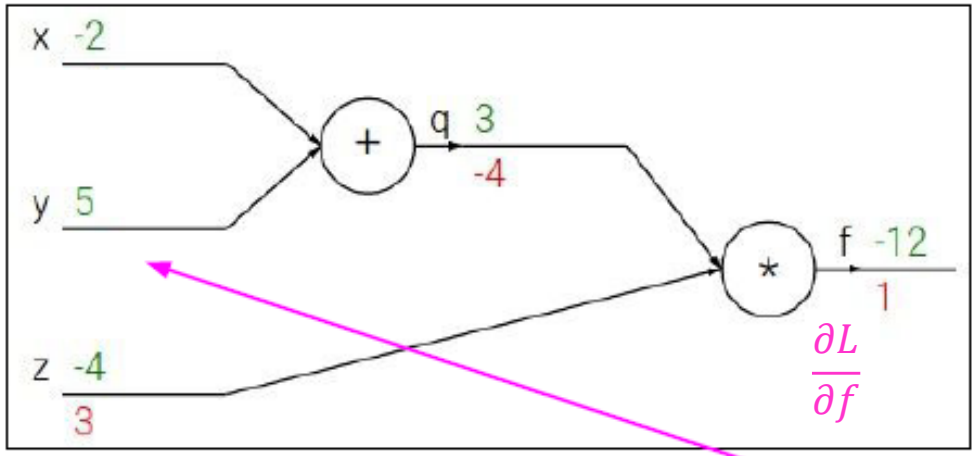
$$f(x, y, z) = (x + y)z$$

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




$$\frac{\partial L}{\partial y} = \frac{\partial L}{\partial q} \cdot \frac{\partial q}{\partial y}$$

Backpropagation: a simple example

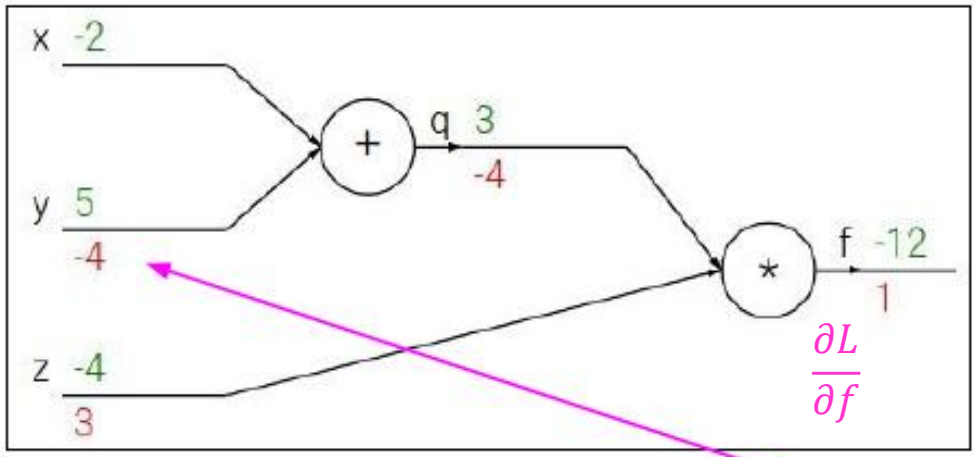
$$f(x, y, z) = (x + y)z$$

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




$$\frac{\partial L}{\partial y} = -4.\frac{\partial q}{\partial y}$$

Backpropagation: a simple example

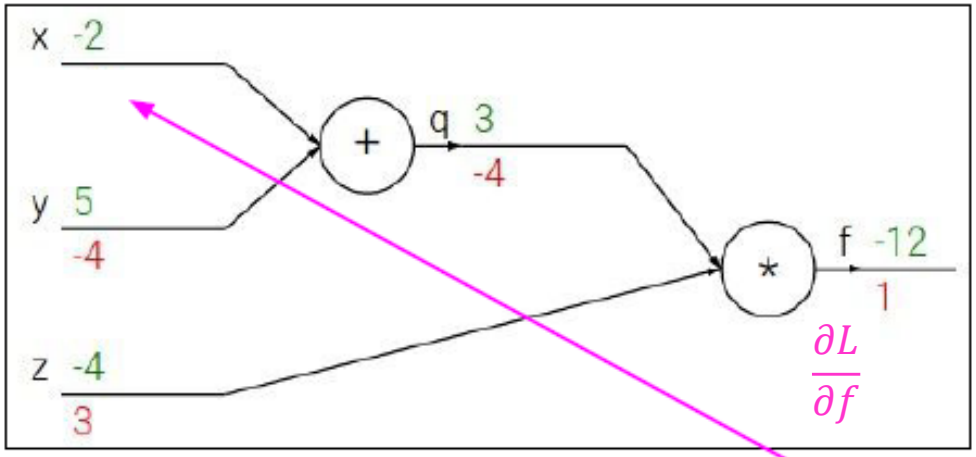
$$f(x, y, z) = (x + y)z$$

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




$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial q} \cdot \frac{\partial q}{\partial x}$$

## Backpropagation: a simple example

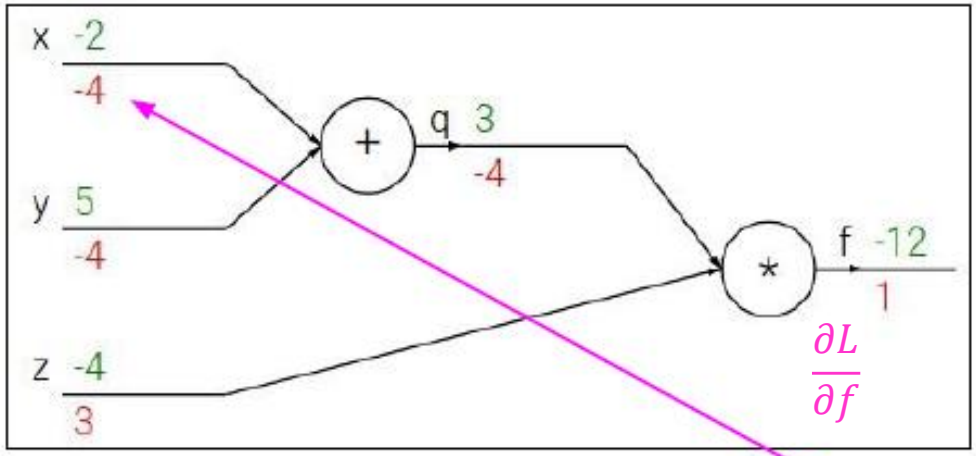
$$f(x, y, z) = (x + y)z$$

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

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

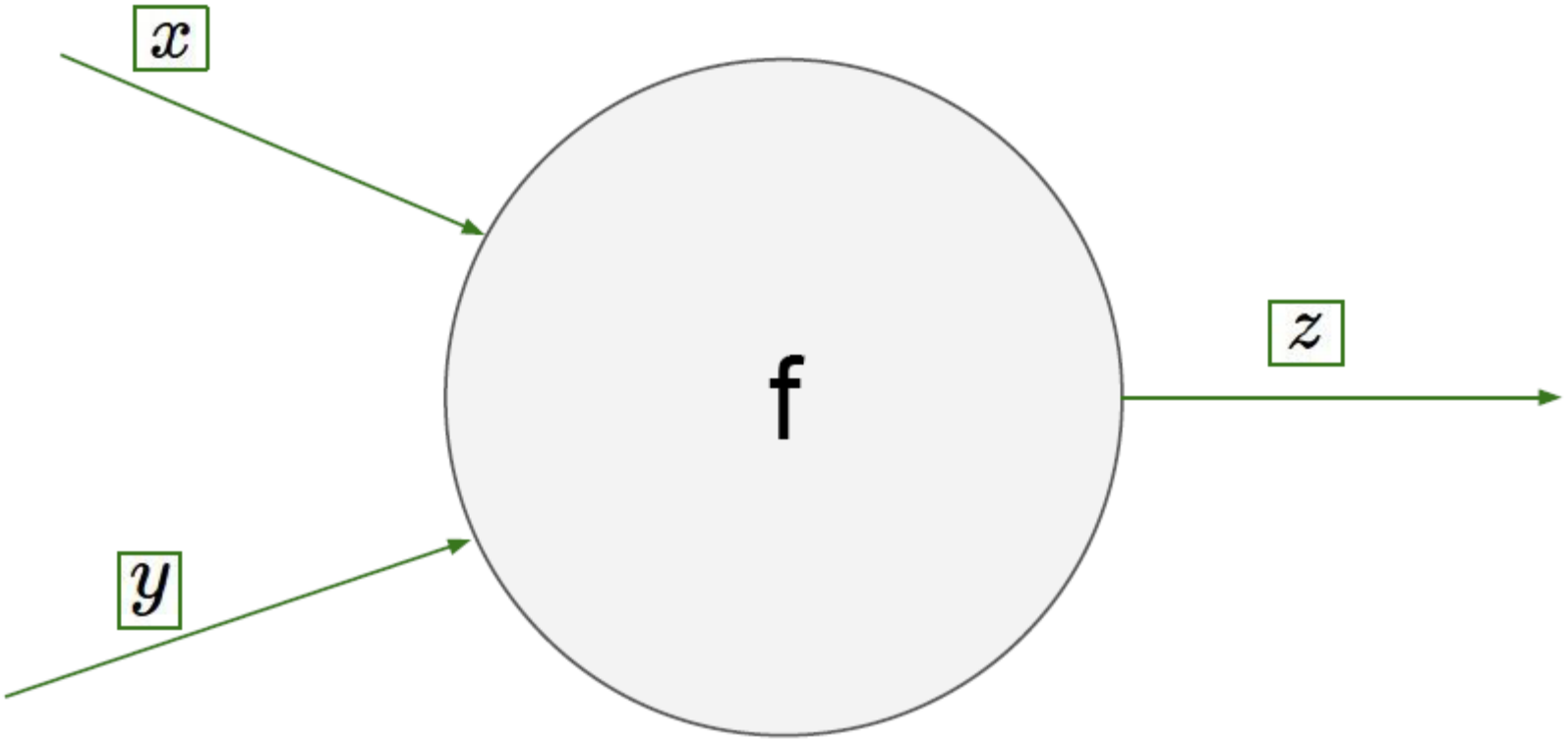
$$f = qz \qquad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

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

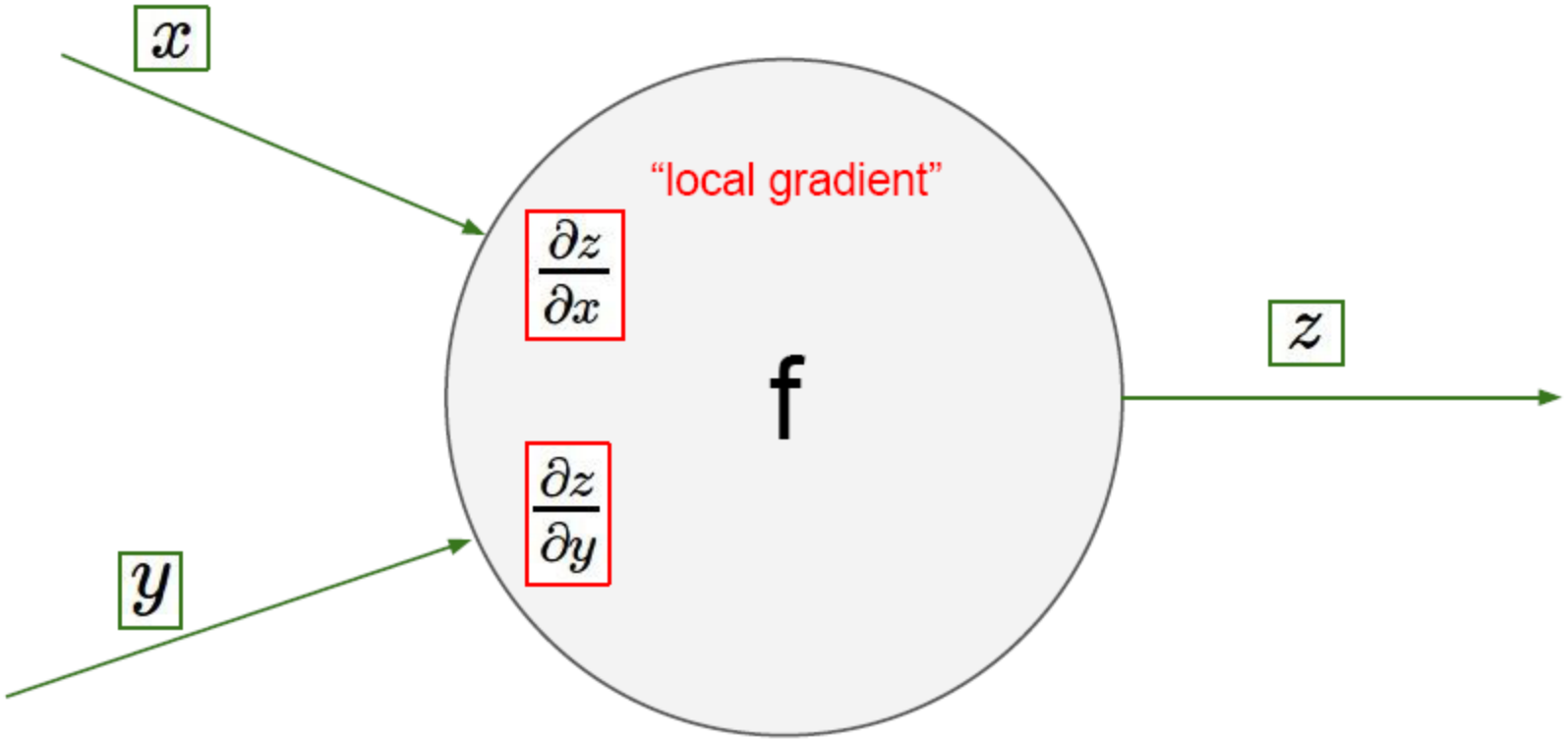



$$\frac{\partial L}{\partial x} = -4.\frac{\partial q}{\partial x}$$

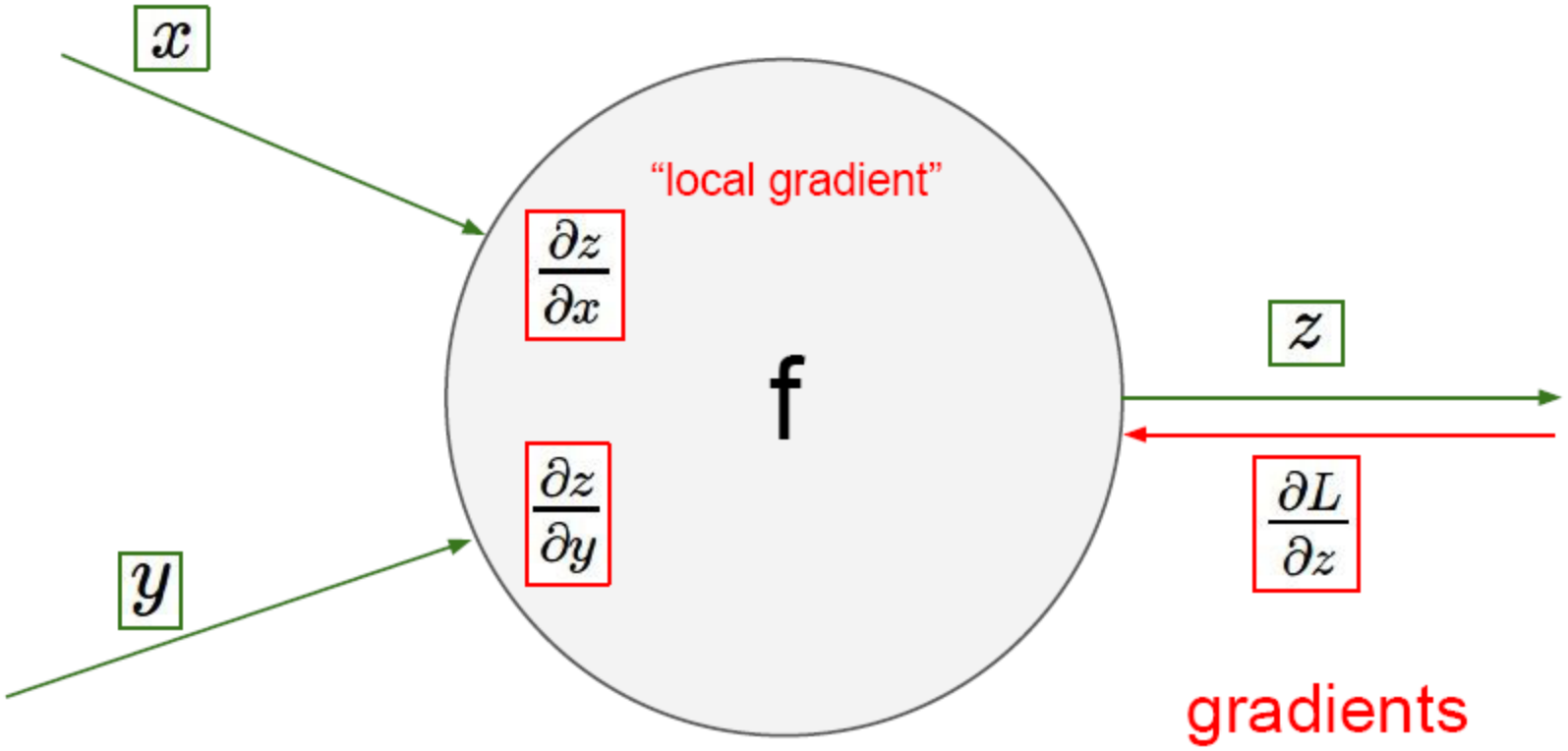
# Backpropagation of gradient in general



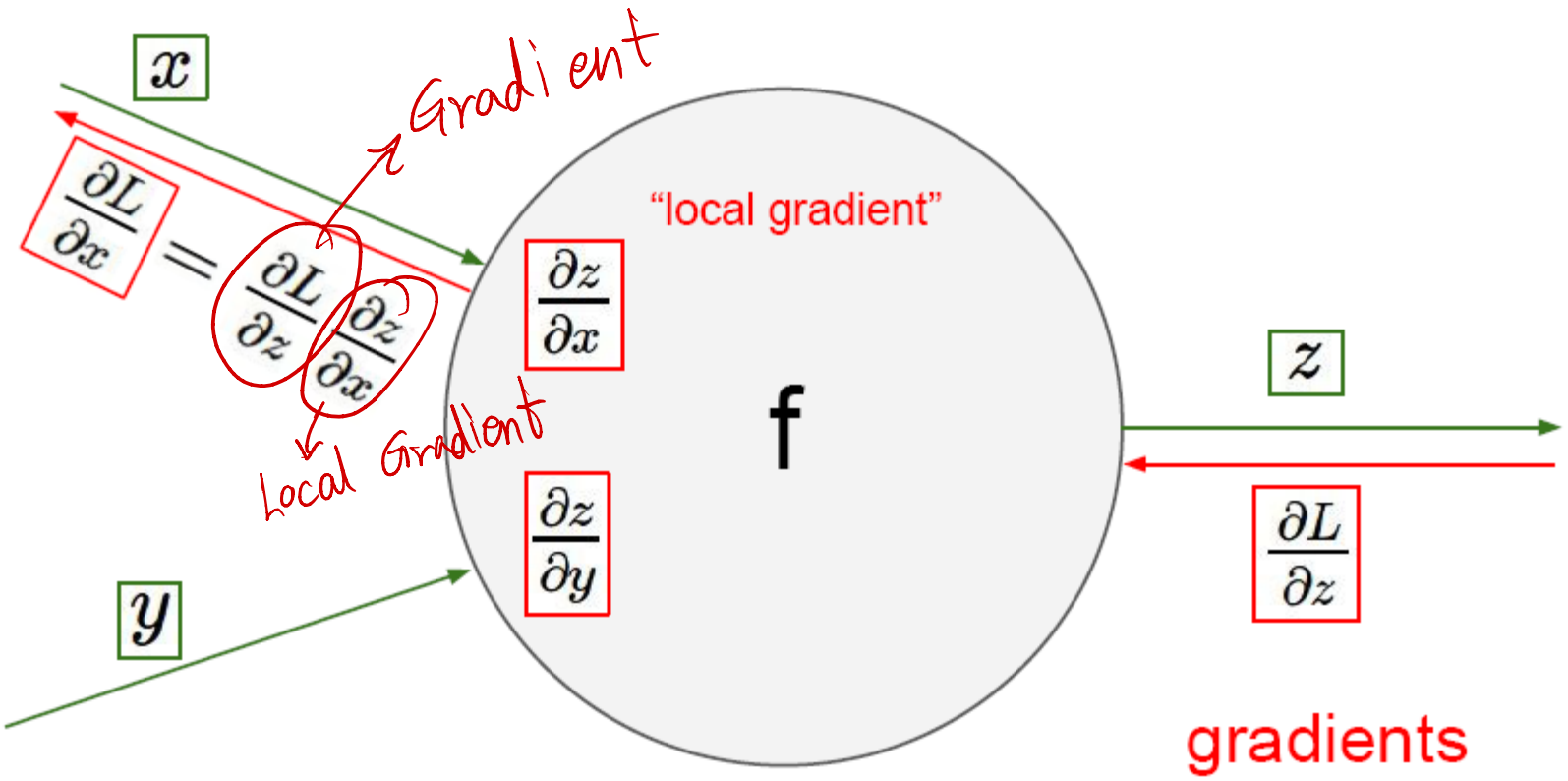
# Backpropagation of gradient in general



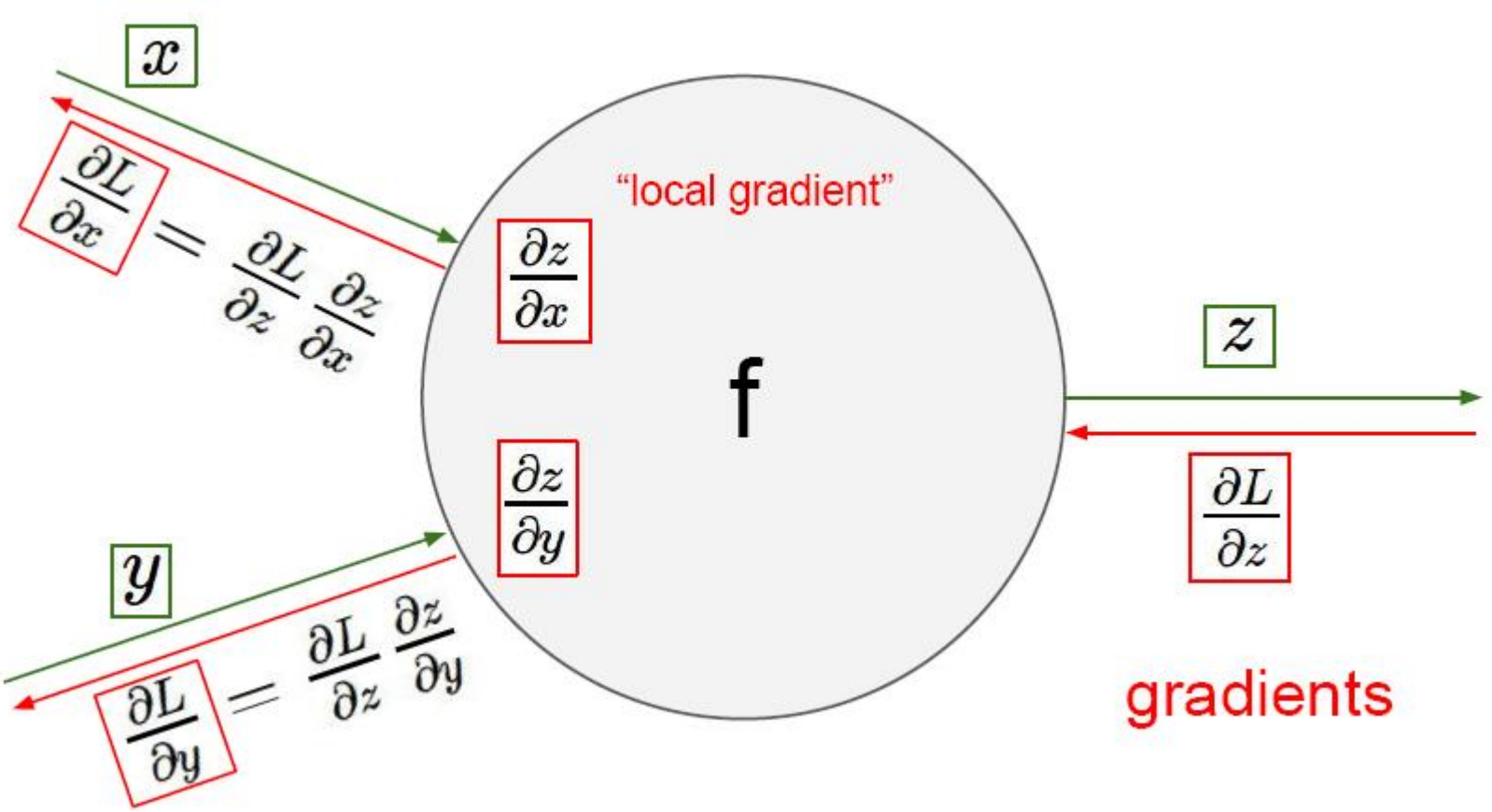
# Backpropagation of gradient in general



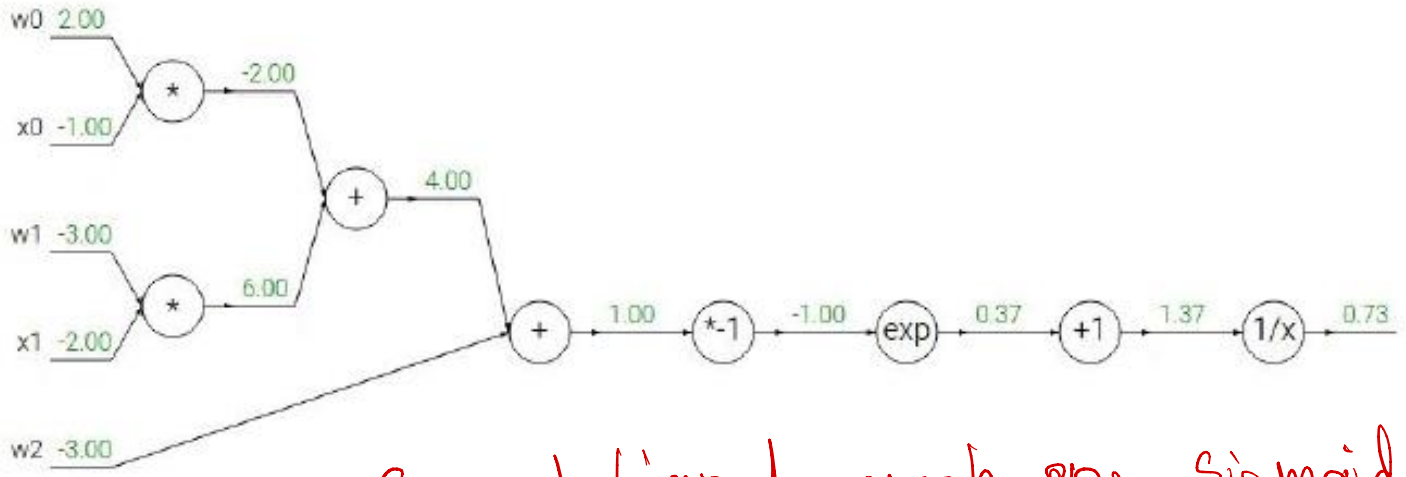
# Backpropagation of gradient in general



# Backpropagation of gradient in general

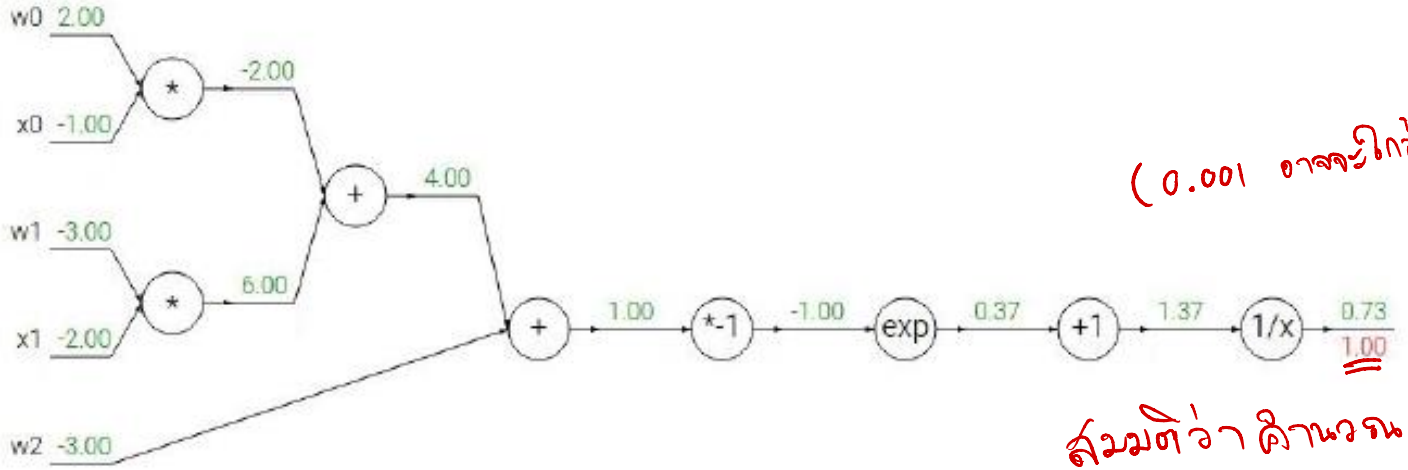


Another example: $$f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$




Computational graph for Sigmoid

Another example: $$f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$



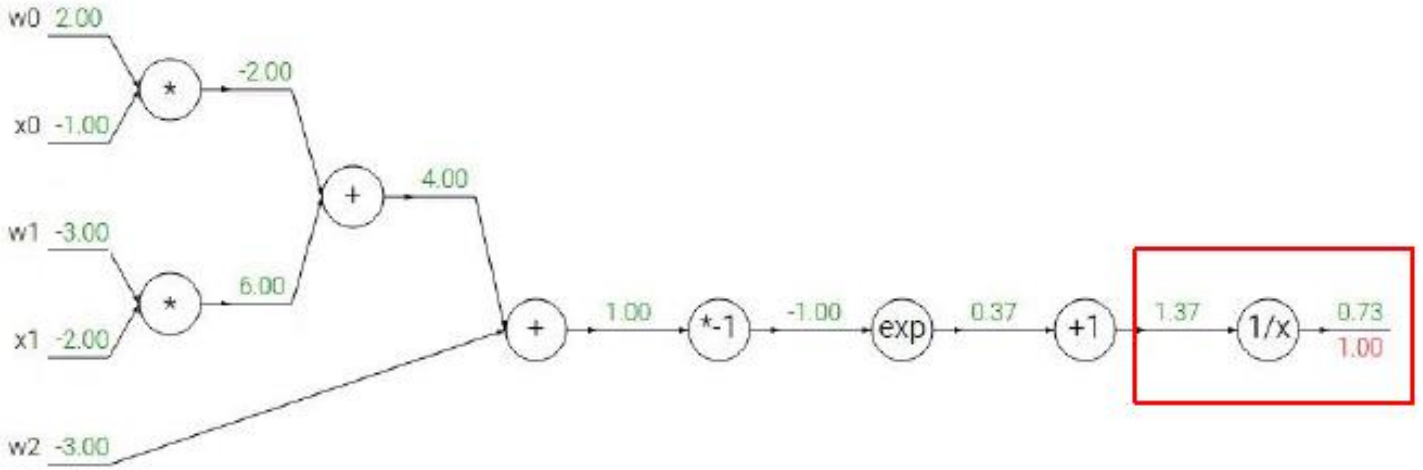

(0.001 อาจจะใกล้ค่าจริงกว่า)

สมมติว่า คำนวณย้อนมาเรื่อยๆ แล้ว loss = 1.00

$$f(x) = e^x \rightarrow \frac{df}{dx} = e^x \qquad f(x) = \frac{1}{x} \rightarrow \frac{df}{dx} = -1/x^2$$

$$f_a(x) = ax \rightarrow \frac{df}{dx} = a \qquad f_c(x) = c + x \rightarrow \frac{df}{dx} = 1$$

Another example: $$f(w,x) = \frac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$$



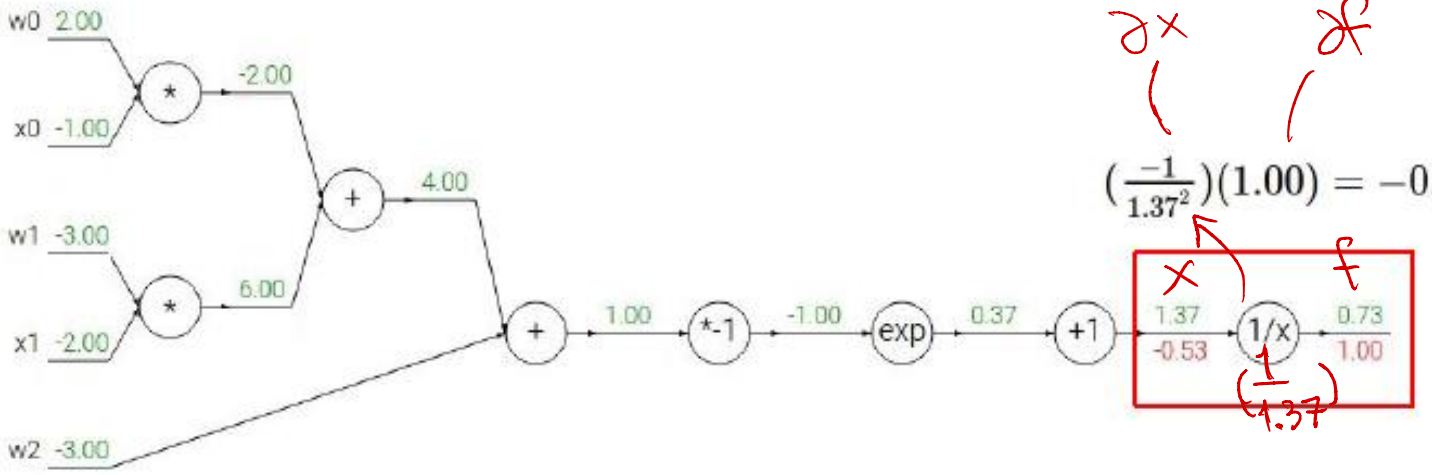

$$f(x) = e^x \quad \rightarrow \quad \frac{df}{dx} = e^x$$

$$f_a(x) = ax \quad \rightarrow \quad \frac{df}{dx} = a$$

$$f(x) = \frac{1}{x} \quad \rightarrow \quad \frac{df}{dx} = -1/x^2$$

$$f_c(x) = c + x \quad \rightarrow \quad \frac{df}{dx} = 1$$

Another example: $$f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$

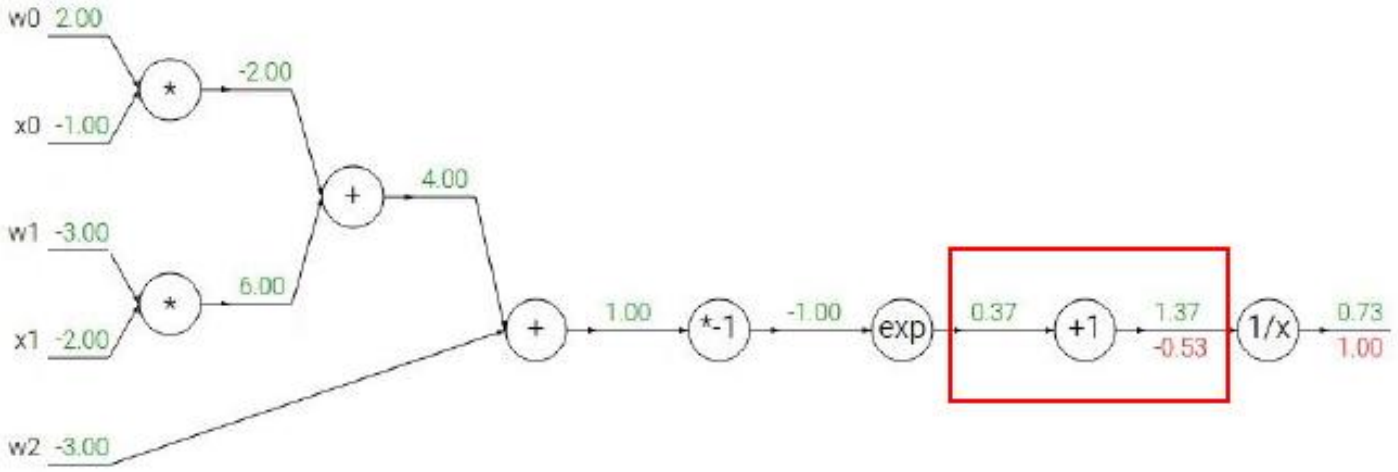



$$\frac{\partial f}{\partial x} \cdot \frac{\partial L}{\partial f} = \frac{\partial L}{\partial x}$$

$$(\frac{-1}{1.37^2})(1.00) = -0.53$$

| | | | | | |
|---|---|---|---|---|---|
| $f(x) = e^x$ | $\rightarrow$ | $\frac{df}{dx} = e^x$ | $f(x) = \frac{1}{x}$ | $\rightarrow$ | $\frac{df}{dx} = -1/x^2$ |
| $f_a(x) = ax$ | $\rightarrow$ | $\frac{df}{dx} = a$ | $f_c(x) = c + x$ | $\rightarrow$ | $\frac{df}{dx} = 1$ |

Another example: $$f(w,x) = \frac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$$



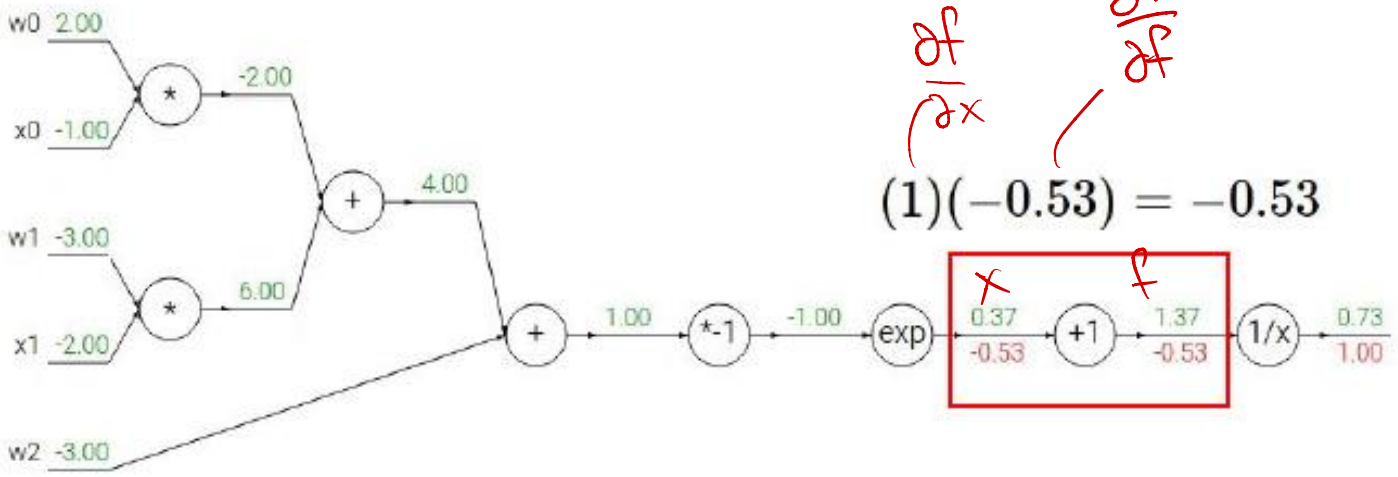

$$f(x) = e^x \quad \rightarrow \quad \frac{df}{dx} = e^x$$

$$f_a(x) = ax \quad \rightarrow \quad \frac{df}{dx} = a$$

$$f(x) = \frac{1}{x} \quad \rightarrow \quad \frac{df}{dx} = -1/x^2$$

$$f_c(x) = c + x \quad \rightarrow \quad \frac{df}{dx} = 1$$

Another example: $$f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$




$$(1)(-0.53) = -0.53$$

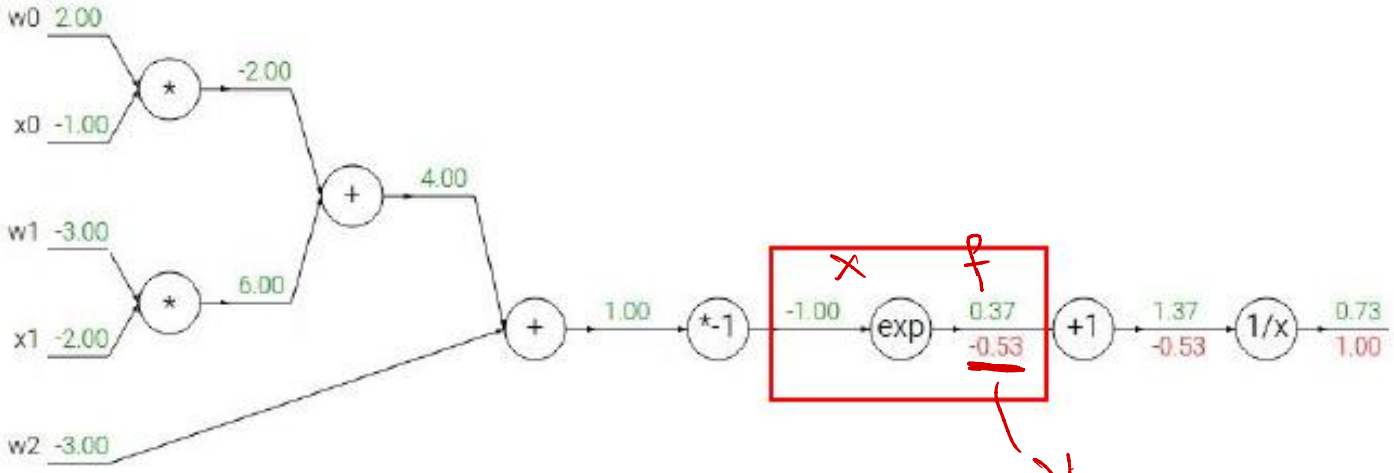
$$f(x) = e^x \rightarrow \frac{df}{dx} = e^x$$

$$f_a(x) = ax \rightarrow \frac{df}{dx} = a$$

$$f(x) = \frac{1}{x} \rightarrow \frac{df}{dx} = -1/x^2$$

$$f_c(x) = c + x \rightarrow \frac{df}{dx} = 1$$

Another example: $$f(w,x) = \frac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$$



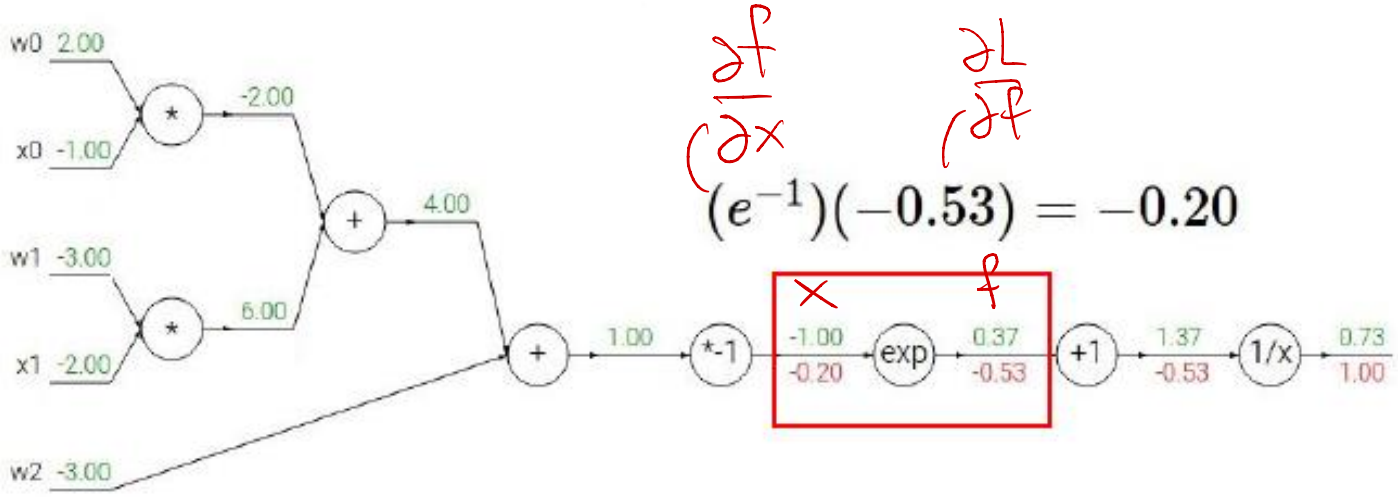

$$f(x) = e^x \rightarrow \frac{df}{dx} = e^x$$

$$f_a(x) = ax \rightarrow \frac{df}{dx} = a$$

$$f(x) = \frac{1}{x} \rightarrow \frac{df}{dx} = -1/x^2$$

$$f_c(x) = c + x \rightarrow \frac{df}{dx} = 1$$

Another example: $$f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$

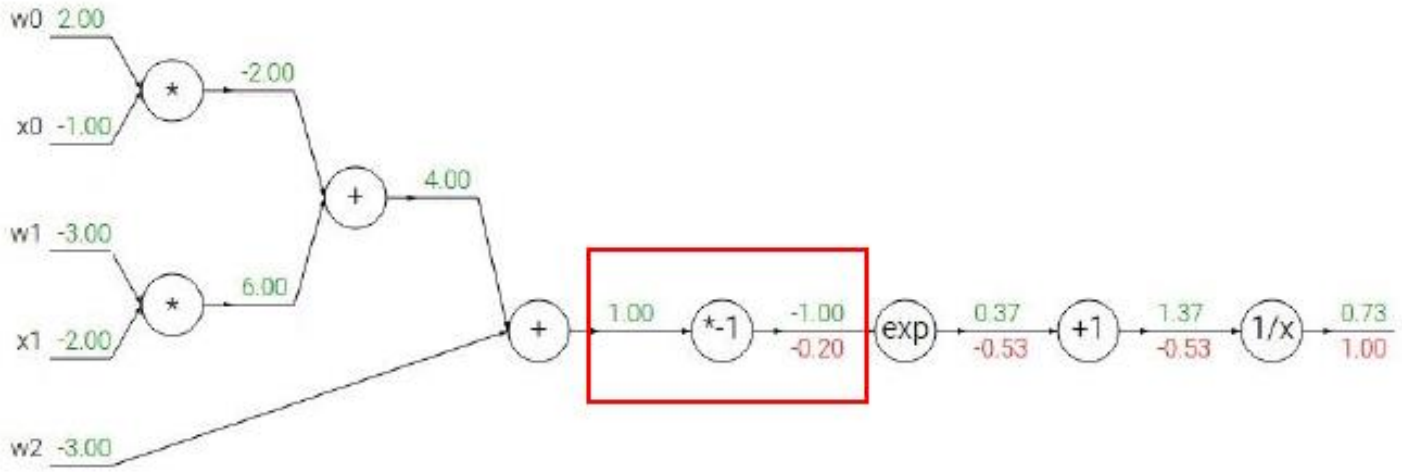



$$(e^{-1})(-0.53) = -0.20$$

| | | |
|---|---|---|
| $f(x) = e^x$ | $\rightarrow$ | $\frac{df}{dx} = e^x$ |
| $f_a(x) = ax$ | $\rightarrow$ | $\frac{df}{dx} = a$ |

| | | |
|---|---|---|
| $f(x) = \frac{1}{x}$ | $\rightarrow$ | $\frac{df}{dx} = -1/x^2$ |
| $f_c(x) = c + x$ | $\rightarrow$ | $\frac{df}{dx} = 1$ |

Another example: $$f(w,x) = \frac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$$



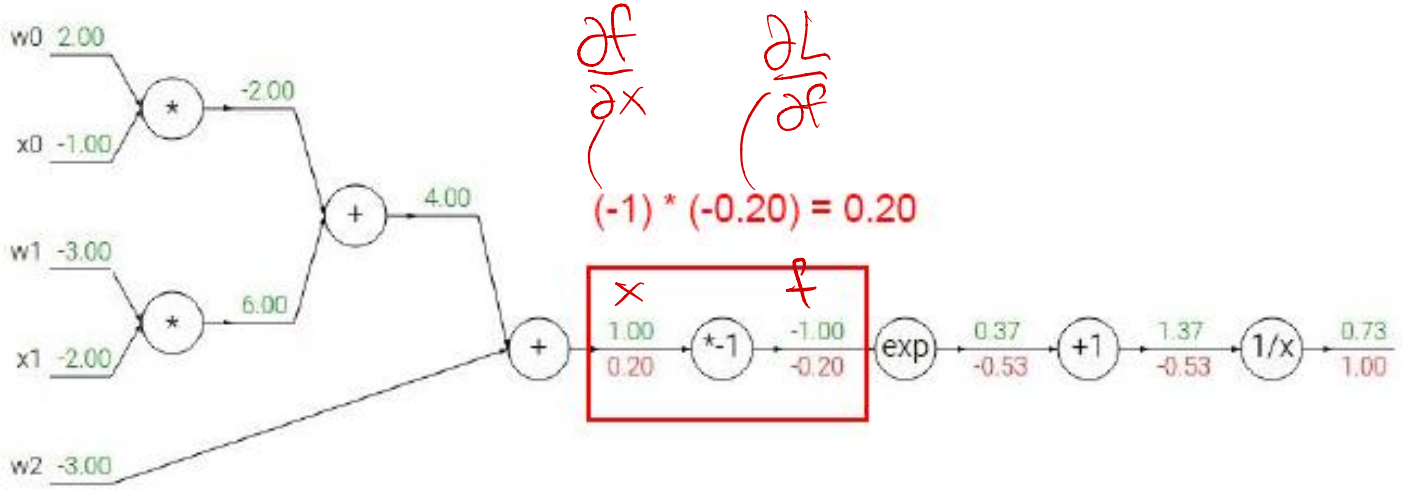

$$f(x) = e^x \quad \rightarrow \quad \frac{df}{dx} = e^x$$

$$f_a(x) = ax \quad \rightarrow \quad \frac{df}{dx} = a$$

$$f(x) = \frac{1}{x} \quad \rightarrow \quad \frac{df}{dx} = -1/x^2$$

$$f_c(x) = c + x \quad \rightarrow \quad \frac{df}{dx} = 1$$

Another example: $$f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$



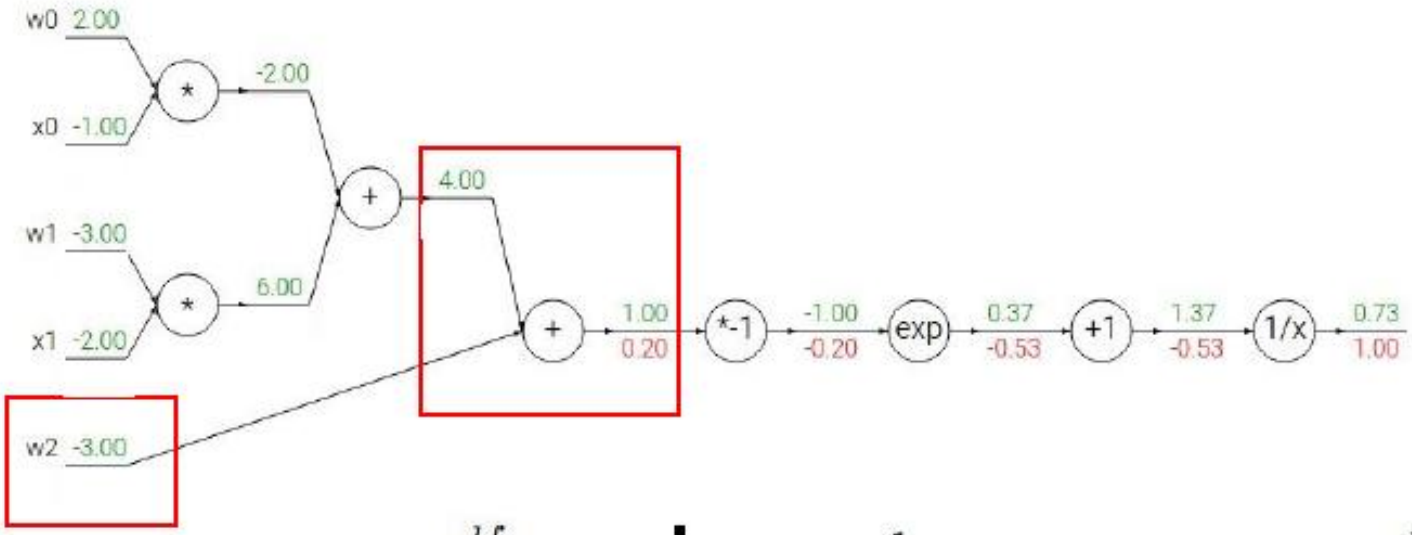

$$f(x) = e^x \rightarrow \frac{df}{dx} = e^x$$

$$f_a(x) = ax \rightarrow \frac{df}{dx} = a$$

$$f(x) = \frac{1}{x} \rightarrow \frac{df}{dx} = -1/x^2$$

$$f_c(x) = c + x \rightarrow \frac{df}{dx} = 1$$

Another example: $$f(w,x) = \frac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$$



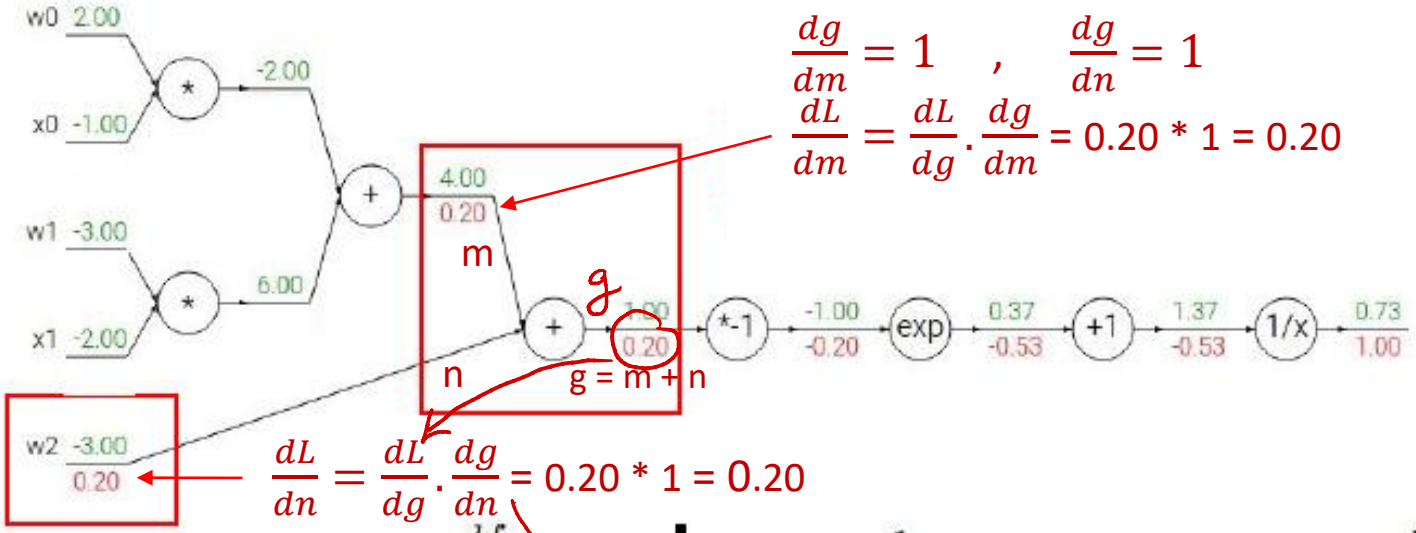

$$f(x) = e^x \quad \rightarrow \quad \frac{df}{dx} = e^x$$

$$f_a(x) = ax \quad \rightarrow \quad \frac{df}{dx} = a$$

$$f(x) = \frac{1}{x} \quad \rightarrow \quad \frac{df}{dx} = -1/x^2$$

$$f_c(x) = c + x \quad \rightarrow \quad \frac{df}{dx} = 1$$

Another example: $f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$




$\frac{dg}{dm} = 1$ , $\frac{dg}{dn} = 1$

$\frac{dL}{dm} = \frac{dL}{dg} \cdot \frac{dg}{dm}$ = 0.20 * 1 = 0.20

$\frac{dL}{dn} = \frac{dL}{dg} \cdot \frac{dg}{dn}$ = 0.20 * 1 = 0.20

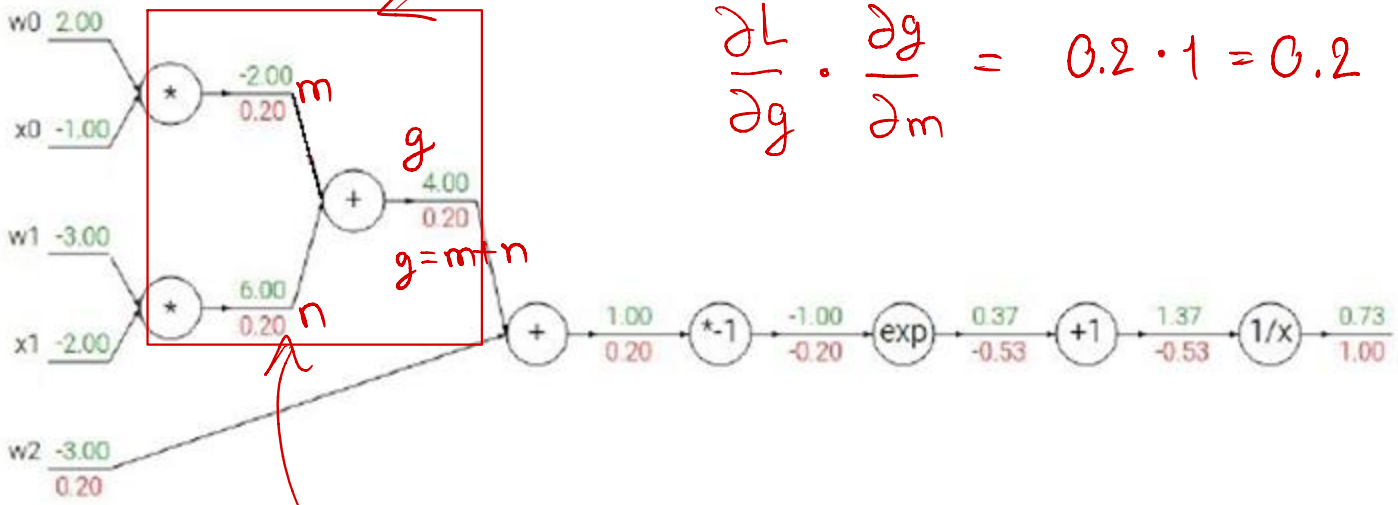
$f(x) = e^x \rightarrow \frac{df}{dx} = e^x$

$f_a(x) = ax \rightarrow \frac{df}{dx} = a$

$f(x) = \frac{1}{x} \rightarrow \frac{df}{dx} = -1/x^2$

$f_c(x) = c + x \rightarrow \frac{df}{dx} = 1$

Another example: $$f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$

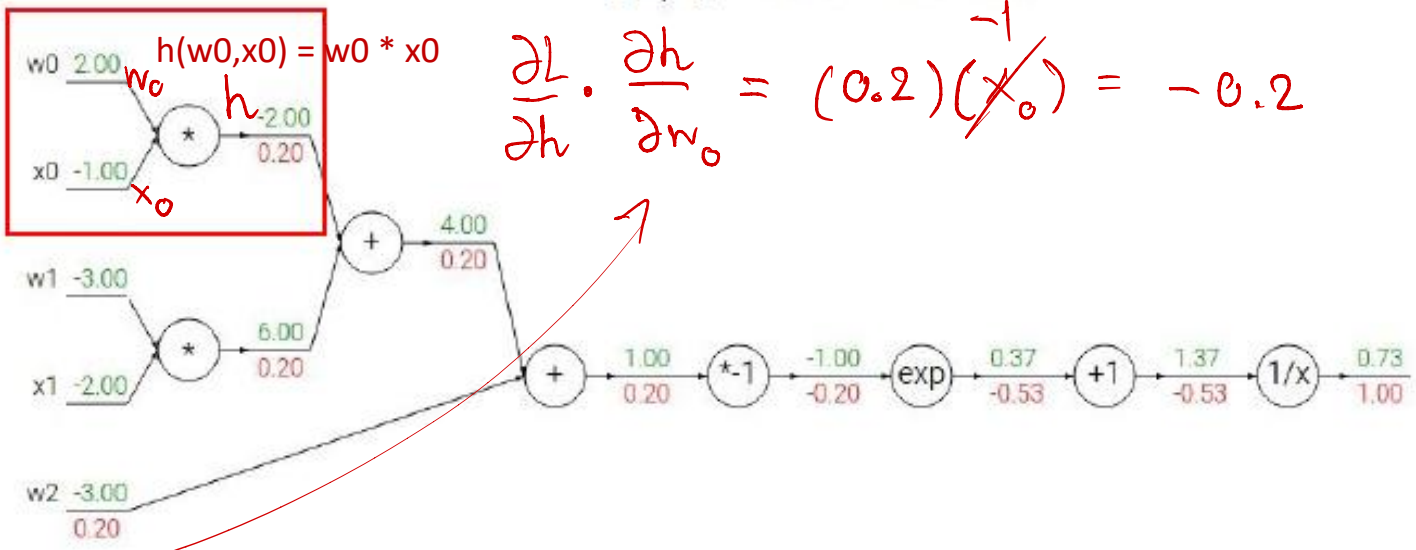



$\frac{\partial L}{\partial g} \cdot \frac{\partial g}{\partial m} = 0.2 \cdot 1 = 0.2$

$\frac{\partial L}{\partial g} \cdot \frac{\partial g}{\partial n} = 0.2 \cdot 1 = 0.2$

| | | | | | |
|---|---|---|---|---|---|
| $f(x) = e^x$ | $\rightarrow$ | $\frac{df}{dx} = e^x$ | $f(x) = \frac{1}{x}$ | $\rightarrow$ | $\frac{df}{dx} = -1/x^2$ |
| $f_a(x) = ax$ | $\rightarrow$ | $\frac{df}{dx} = a$ | $f_c(x) = c + x$ | $\rightarrow$ | $\frac{df}{dx} = 1$ |

Another example: $$f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$



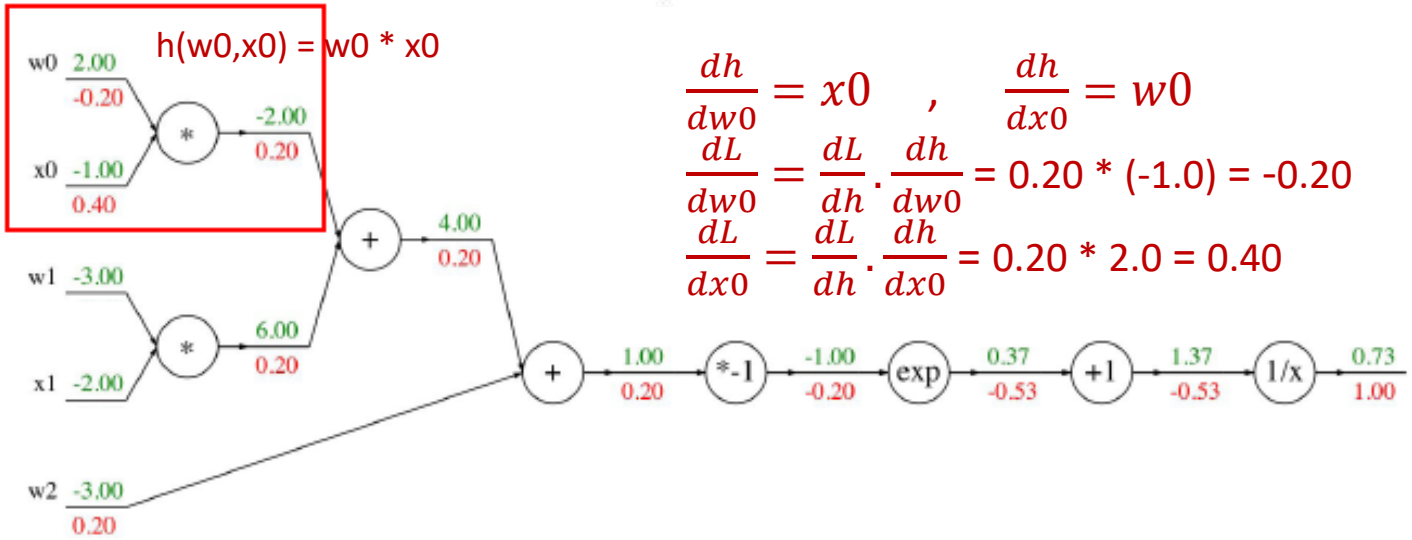

$$\frac{\partial L}{\partial h} \cdot \frac{\partial h}{\partial w_0} = (0.2)(x_0) = -0.2$$

$$f(x) = e^x \rightarrow \frac{df}{dx} = e^x \qquad f(x) = \frac{1}{x} \rightarrow \frac{df}{dx} = -1/x^2$$

$$f_a(x) = ax \rightarrow \frac{df}{dx} = a \qquad f_c(x) = c + x \rightarrow \frac{df}{dx} = 1$$

$$\frac{\partial L}{\partial h} \cdot \frac{\partial h}{\partial x_0} = (0.2)(w_0) = 0.4$$

Another example: $f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$




$\frac{dh}{dw0} = x0$ , $\frac{dh}{dx0} = w0$

$\frac{dL}{dw0} = \frac{dL}{dh} \cdot \frac{dh}{dw0}$ = 0.20 * (-1.0) = -0.20

$\frac{dL}{dx0} = \frac{dL}{dh} \cdot \frac{dh}{dx0}$ = 0.20 * 2.0 = 0.40

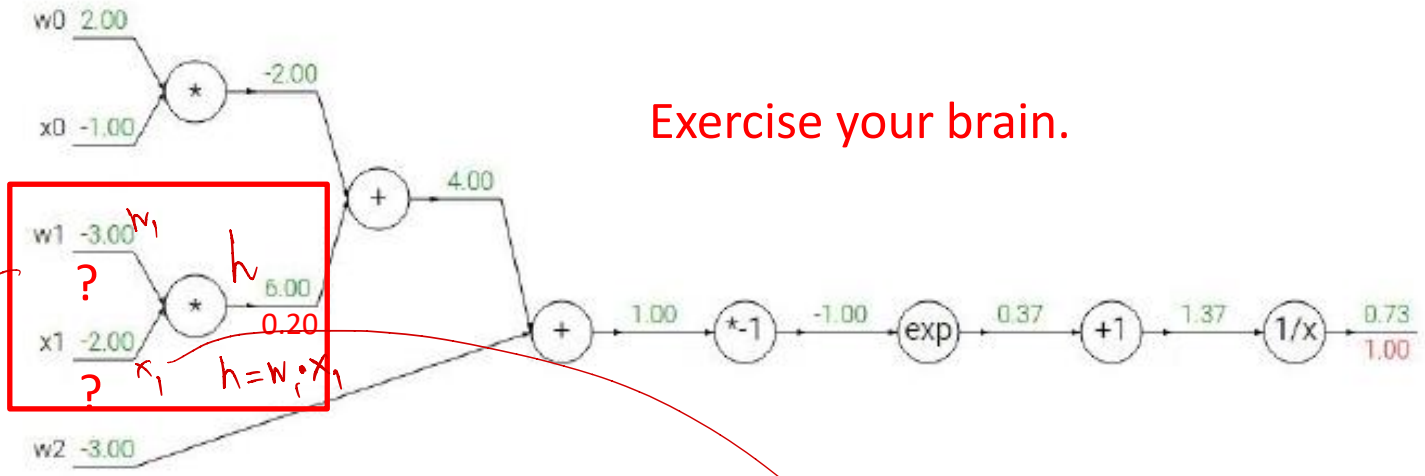
$f(x) = e^x \rightarrow \frac{df}{dx} = e^x$

$f_a(x) = ax \rightarrow \frac{df}{dx} = a$

$f(x) = \frac{1}{x} \rightarrow \frac{df}{dx} = -1/x^2$

$f_c(x) = c + x \rightarrow \frac{df}{dx} = 1$

Another example: $$f(w,x) = \frac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$$




Exercise your brain.

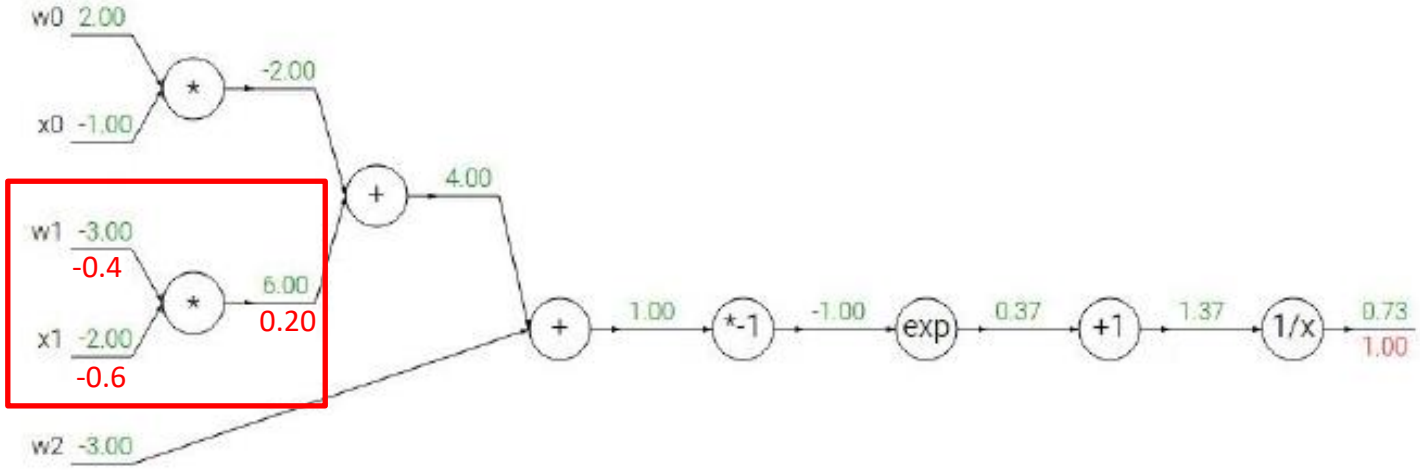
$$f(x) = e^x \rightarrow \frac{df}{dx} = e^x \qquad f(x) = \frac{1}{x} \rightarrow \frac{df}{dx} = -1/x^2$$

$$f_a(x) = ax \rightarrow \frac{df}{dx} = a \qquad f_c(x) = c + x \rightarrow \frac{df}{dx} = 1$$

$\frac{\partial L}{\partial h} \cdot \frac{\partial h}{\partial w_1} = (0.2)(x_1) = -0.4$

$\frac{\partial L}{\partial h} \cdot \frac{\partial h}{\partial x_1} = (0.2)(w_1) = -0.6$

Another example: $$f(w,x) = \frac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$$




$$f(x) = e^x \quad \rightarrow \quad \frac{df}{dx} = e^x$$

$$f_a(x) = ax \quad \rightarrow \quad \frac{df}{dx} = a$$

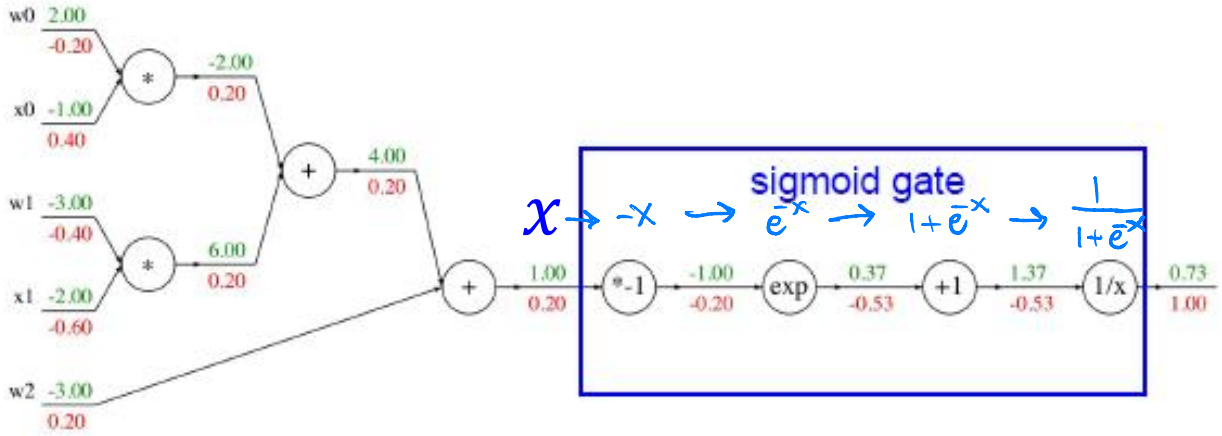
$$f(x) = \frac{1}{x} \quad \rightarrow \quad \frac{df}{dx} = -1/x^2$$

$$f_c(x) = c + x \quad \rightarrow \quad \frac{df}{dx} = 1$$

$$f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$

$$\boxed{\sigma(x) = \frac{1}{1+e^{-x}}}$$ sigmoid function

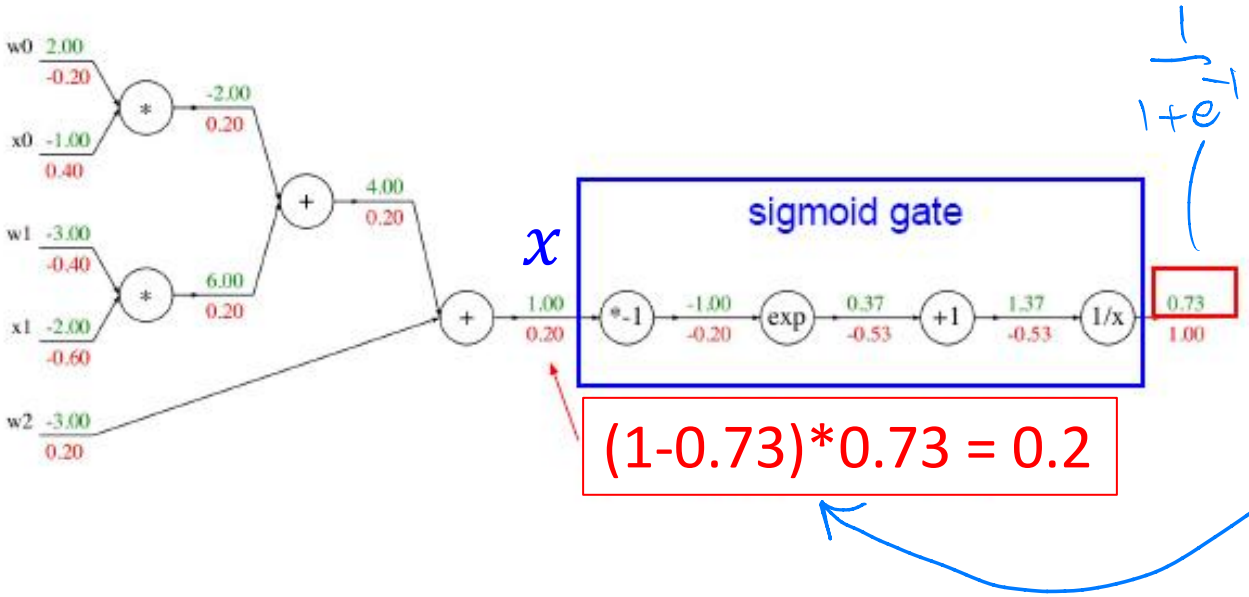
$$\frac{d\sigma(x)}{dx} = \frac{e^{-x}}{(1+e^{-x})^2} = \left(\frac{1+e^{-x}-1}{1+e^{-x}}\right)\left(\frac{1}{1+e^{-x}}\right) = (1-\sigma(x))\,\sigma(x)$$



$$f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$

$$\boxed{\sigma(x) = \frac{1}{1+e^{-x}}}$$ sigmoid function

$$\frac{d\sigma(x)}{dx} = \frac{e^{-x}}{(1+e^{-x})^2} = \left(\frac{1+e^{-x}-1}{1+e^{-x}}\right)\left(\frac{1}{1+e^{-x}}\right) = (1-\sigma(x))\,\sigma(x)$$



# Exercise

In the context of artificial neural networks, the **rectifier** is an activation function defined as the positive part of its argument:

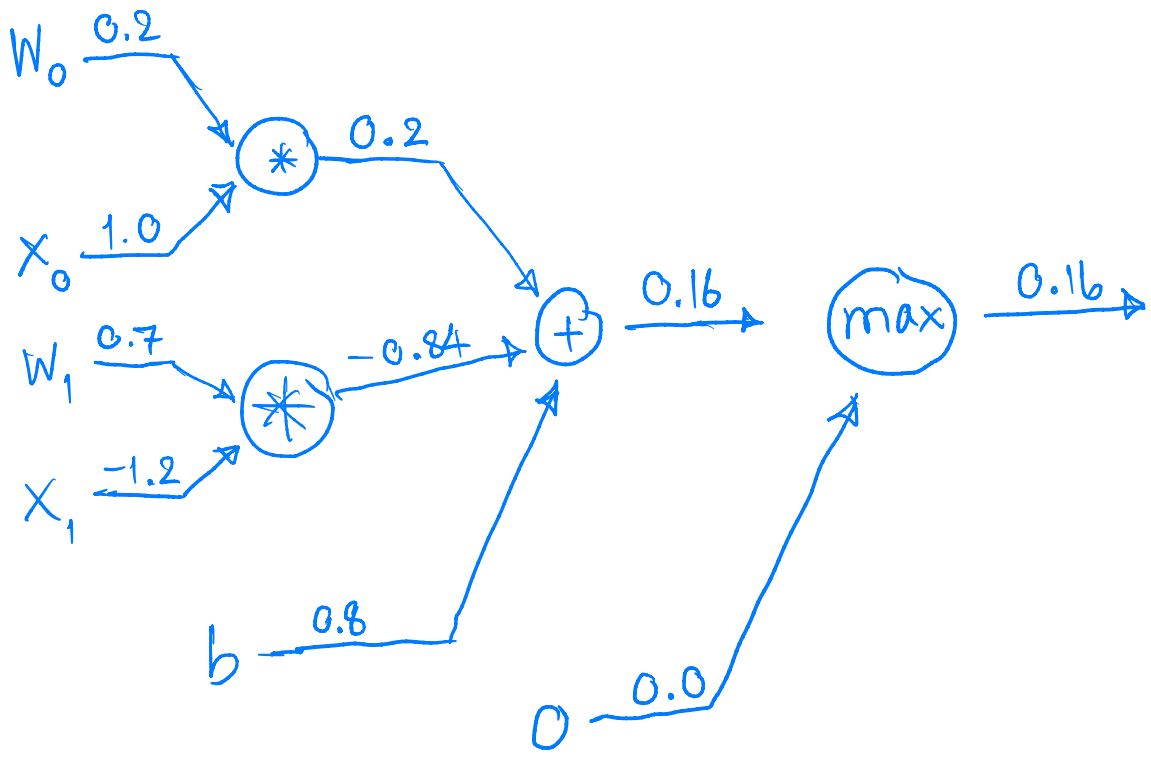
$f(x) = \max(0, x)$ ; this is also know as ReLU function

Let $x = w_0x_0 + w_1x_1 + b$

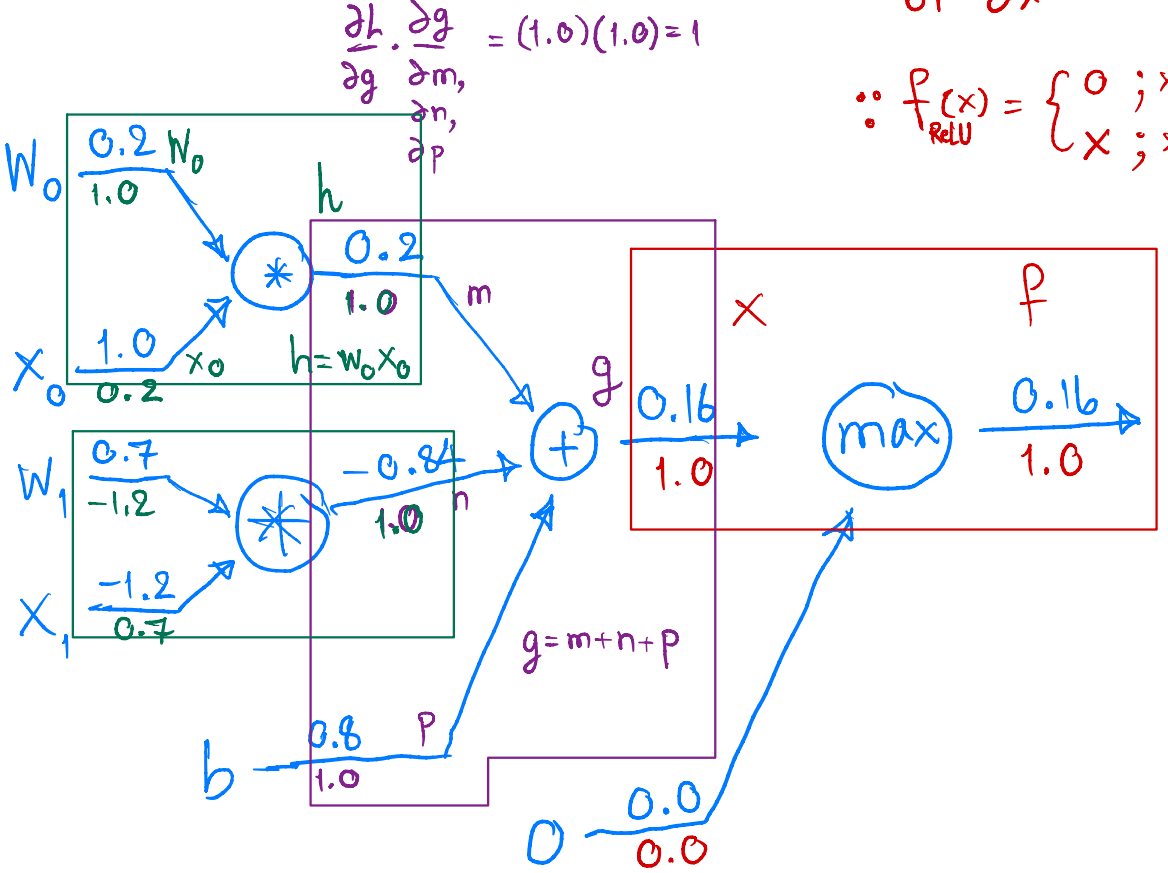
$w_0 = 0.2$, $w_1 = 0.7$, $x_0 = 1.0$, $x_1 = -1.2$, $b = 0.8$

1. Draw the complete computational graph for f(x).
2. Assume that $\frac{\partial L}{\partial f} = 1.0$
3. Calculate backpropagate gradient of L at every branch of the graph.

# Computational graph



# Gradients




$$\frac{\partial L}{\partial h} \cdot \frac{\partial h}{\partial w_0} = (1.0)(x_0) = 1.0$$

$$\frac{\partial L}{\partial h} \cdot \frac{\partial h}{\partial x_0} = (1.0)(w_0) = 0.2$$

$$\frac{\partial L}{\partial h} \cdot \frac{\partial h}{\partial w_1} = (1.0)(x_1) = -1.2$$

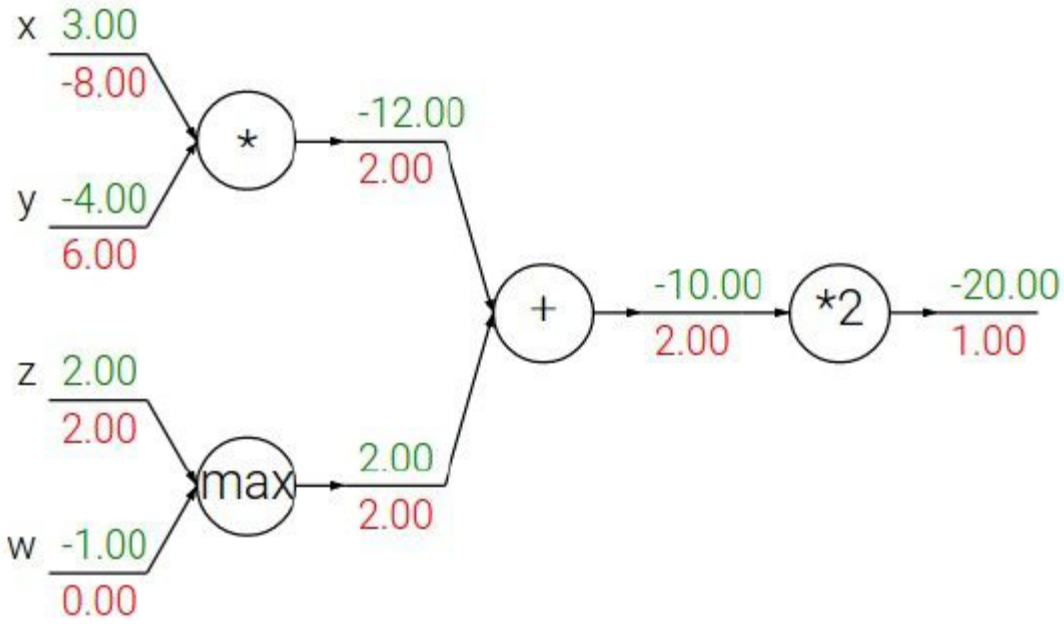
$$\frac{\partial L}{\partial h} \cdot \frac{\partial h}{\partial x_1} = (1.0)(w_1) = 0.7$$

# Patterns in backward flow

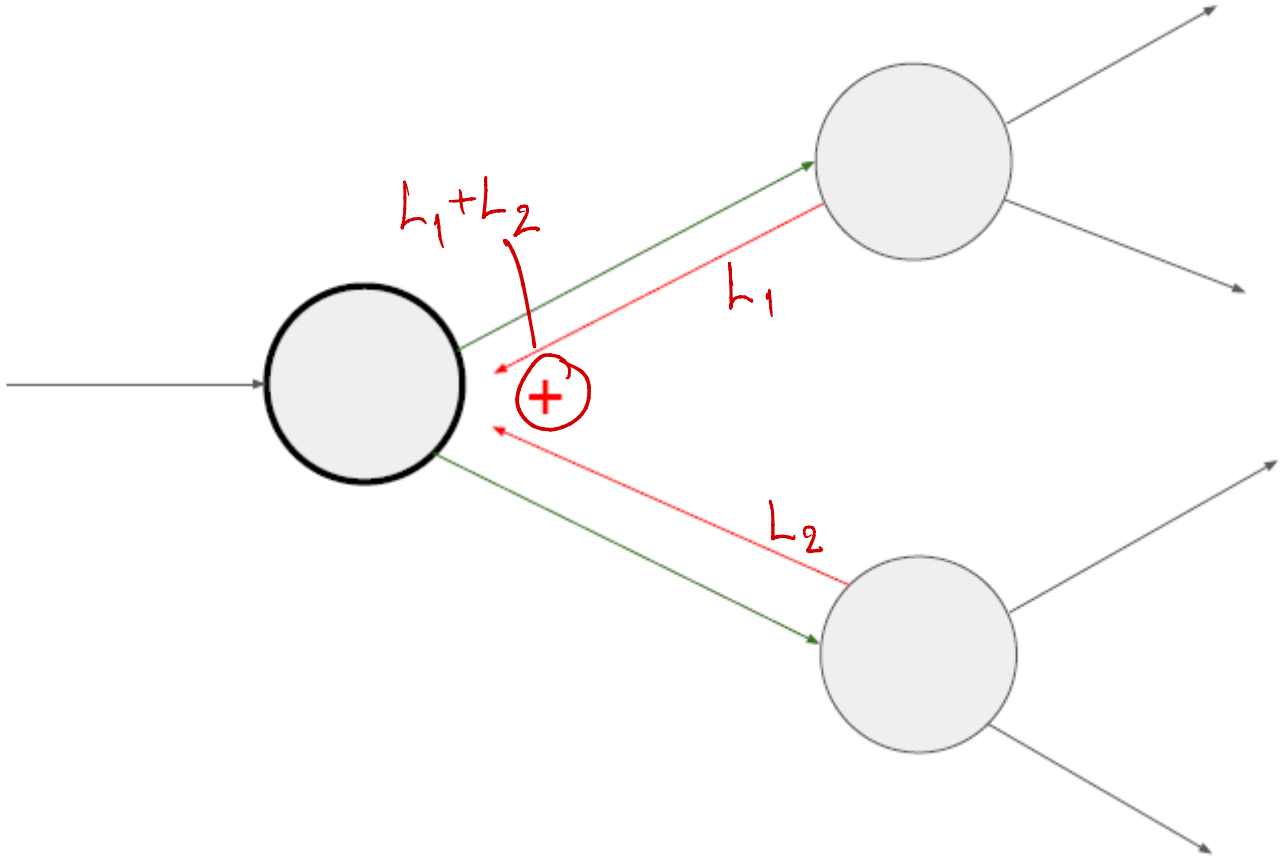
**add** gate: gradient distributor

**max** gate: gradient router

**mul** gate: gradient switcher



# Gradients add at branches



# Revisit: Backpropagation for multilayer NN

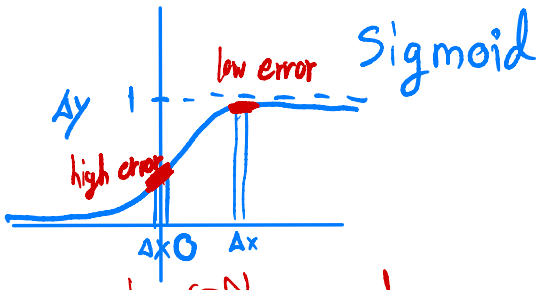
- To propagate error signals, we start at the output layer and work backward to the hidden layer.

- The error signal at the output of neuron k at iteration p is defined by:

Target Actual

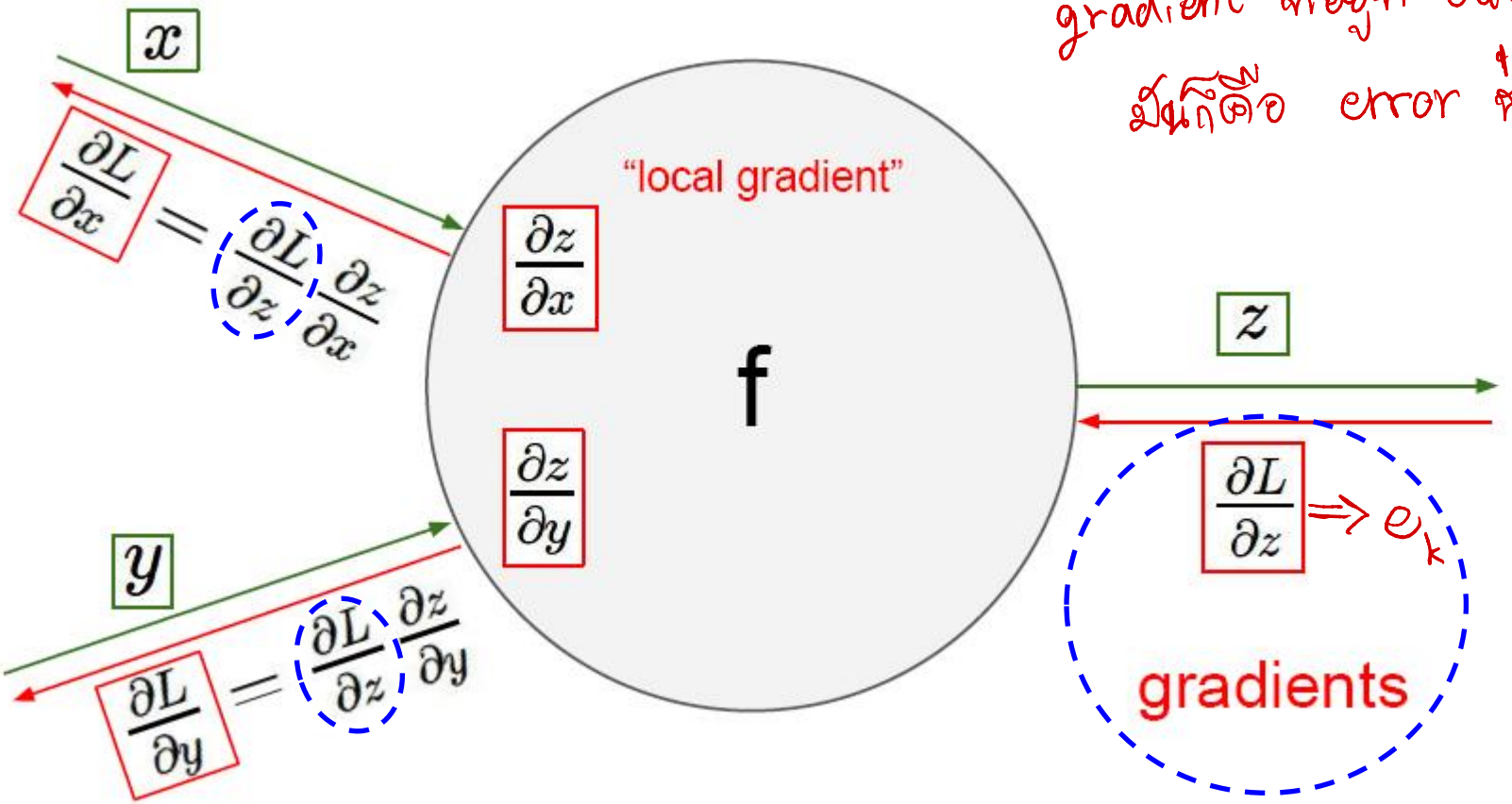
$$e_k(p) = yd_k(p) - y_k(p)$$

Where $yd_k(p)$ is the desired output of neuron k at iteration p.

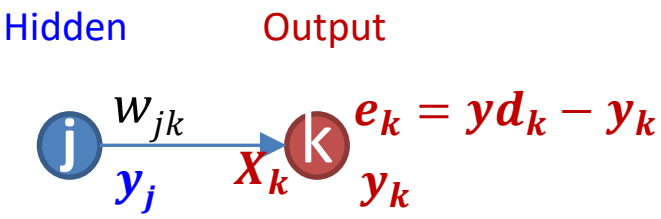



ของแขนไม่ทางการว่ามันคือ slope

Recall gradients backpropagation from page 18.



- To update weights, the weight correction ($\Delta w$) is adjusted to the previous weights.




$$w_{jk}(p+1) = w_{jk}(p) - \Delta w_{jk}(p),$$

$$\Delta w_{jk}(p) = \alpha . y_j(p) . \delta_k(p),$$

$$\delta_k(p) = \frac{\partial y_k(p)}{\partial X_k(p)} . e_k(p),$$

How do we get these terms?

แบบทางการ

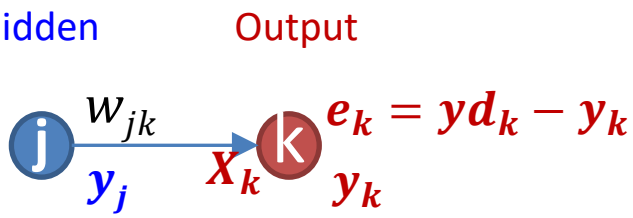
- To update weights, the weight correction ($\Delta w$) is adjusted to the previous weights.

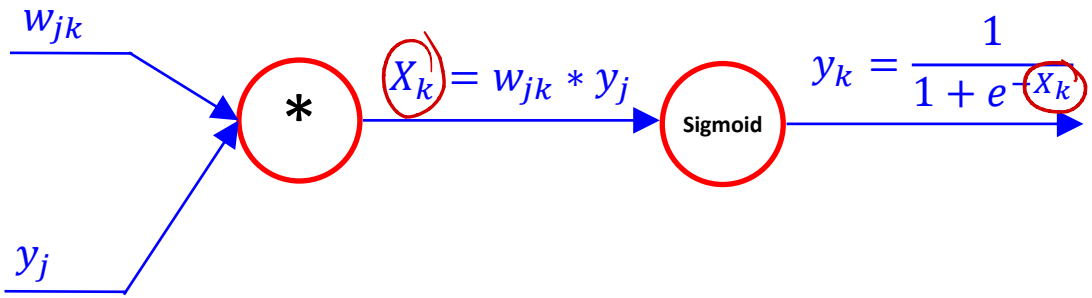
$$w_{jk}(p+1) = w_{jk}(p) - \Delta w_{jk}(p),$$

$$\Delta w_{jk}(p) = \alpha . y_j(p) . \delta_k(p),$$

$$\delta_k(p) = \frac{\partial y_k(p)}{\partial X_k(p)} . e_k(p),$$

How do we get these terms?



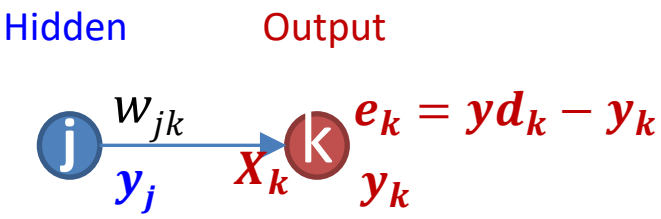



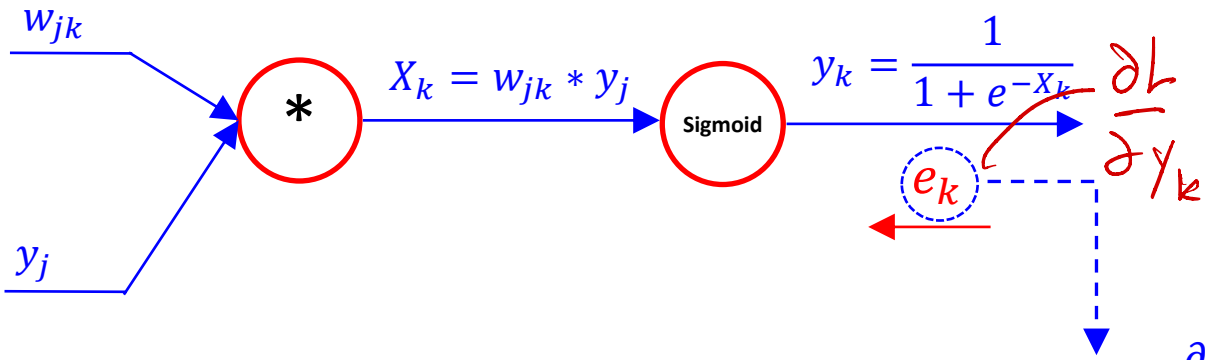
- To update weights, the weight correction ($\Delta w$) is adjusted to the previous weights.

$$w_{jk}(p+1) = w_{jk}(p) - \Delta w_{jk}(p),$$

$$\Delta w_{jk}(p) = \alpha . y_j(p) . \delta_k(p),$$

$$\delta_k(p) = \frac{\partial y_k(p)}{\partial X_k(p)} . e_k(p),$$




This is equivalent to $\frac{\partial e_k}{\partial y_k}$

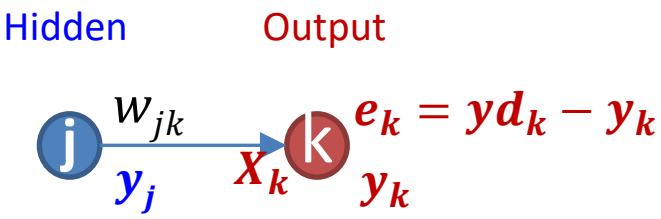
(or $\frac{\partial L}{\partial z}$ in page 18)

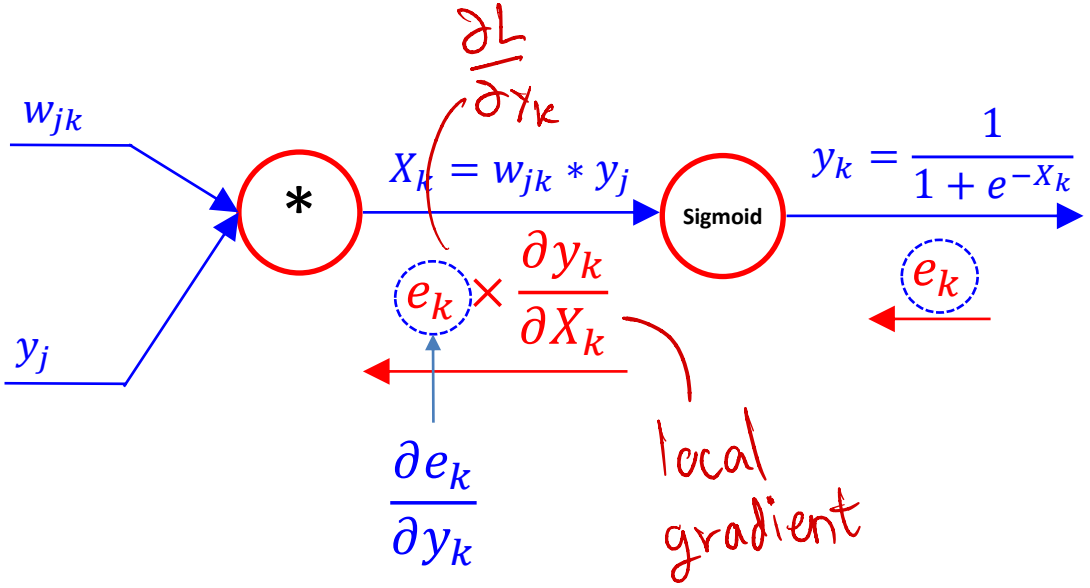
- To update weights, the weight correction ($\Delta w$) is adjusted to the previous weights.

$$w_{jk}(p+1) = w_{jk}(p) - \Delta w_{jk}(p),$$

$$\Delta w_{jk}(p) = \alpha . y_j(p) . \delta_k(p),$$

$$\delta_k(p) = \frac{\partial y_k(p)}{\partial X_k(p)} . e_k(p),$$



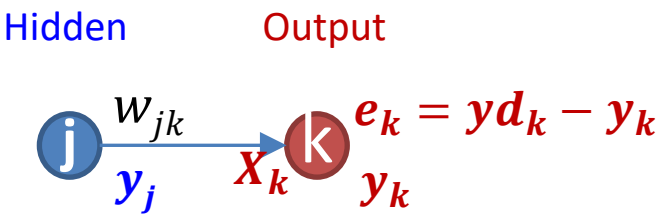



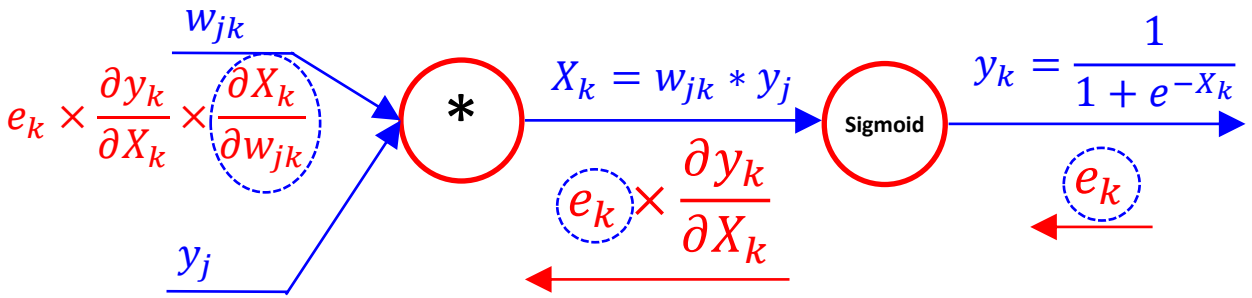
- To update weights, the weight correction ($\Delta w$) is adjusted to the previous weights.

$$w_{jk}(p+1) = w_{jk}(p) - \Delta w_{jk}(p),$$

$$\Delta w_{jk}(p) = \alpha . y_j(p) . \delta_k(p),$$

$$\delta_k(p) = \frac{\partial y_k(p)}{\partial X_k(p)} . e_k(p),$$



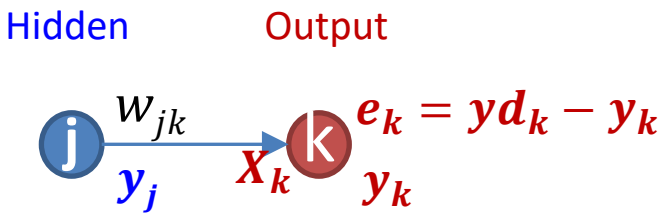



- To update weights, the weight correction ($\Delta w$) is adjusted to the previous weights.

$$w_{jk}(p+1) = w_{jk}(p) - \Delta w_{jk}(p),$$

$$\Delta w_{jk}(p) = \alpha . y_j(p) . \delta_k(p),$$

$$\delta_k(p) = \frac{\partial y_k(p)}{\partial X_k(p)} . e_k(p),$$




คุณ.ที่มาแผนทางการ ※