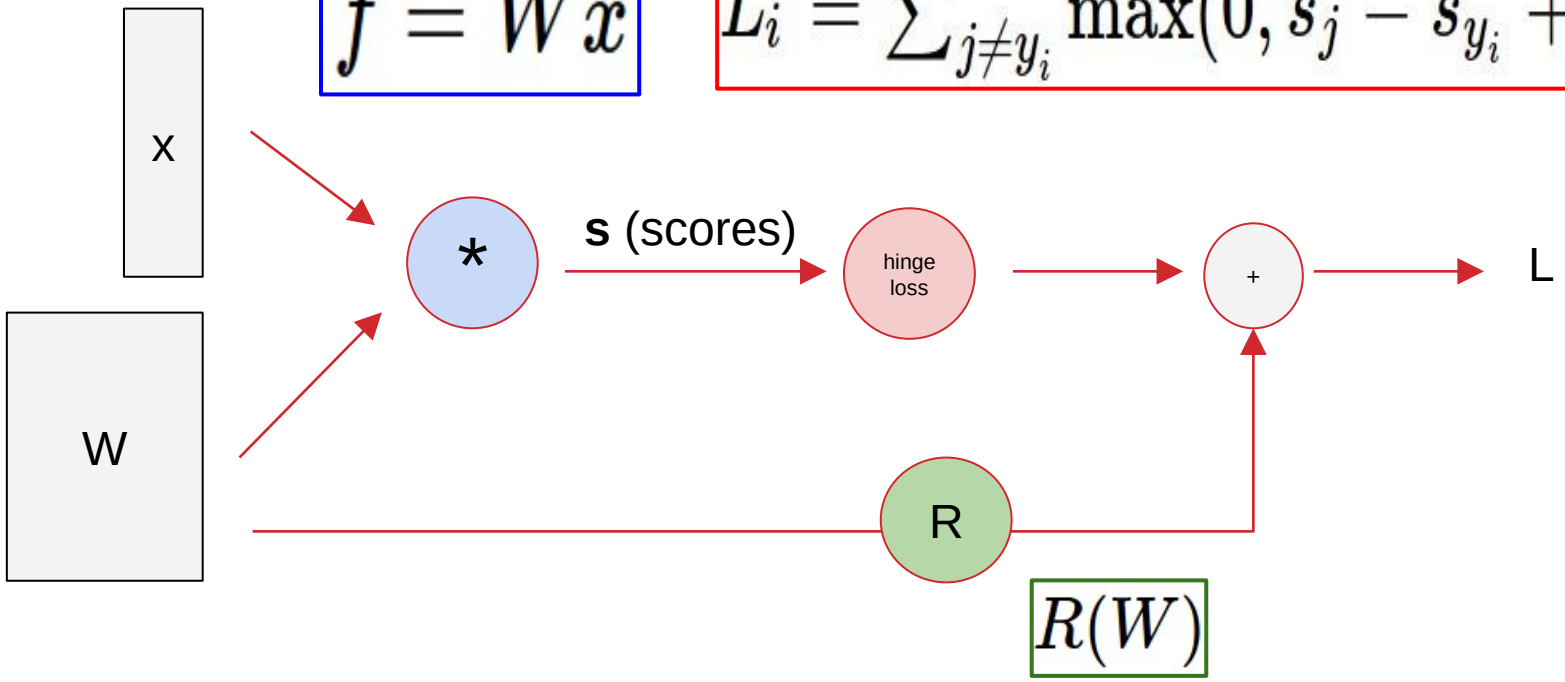
PROPAGAÇÃO RETRÓGRADA

Zenilton K. G. Patrocínio Jr
zenilton@pucminas.br

Grafo de Computação da Função de Perda

$$f = Wx$$

$$L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)$$

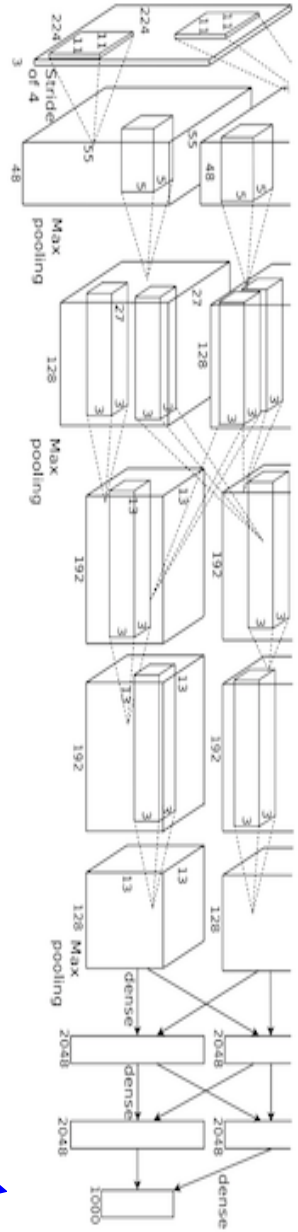


Rede Convolucional (AlexNet)

Imagem de entrada

Pesos

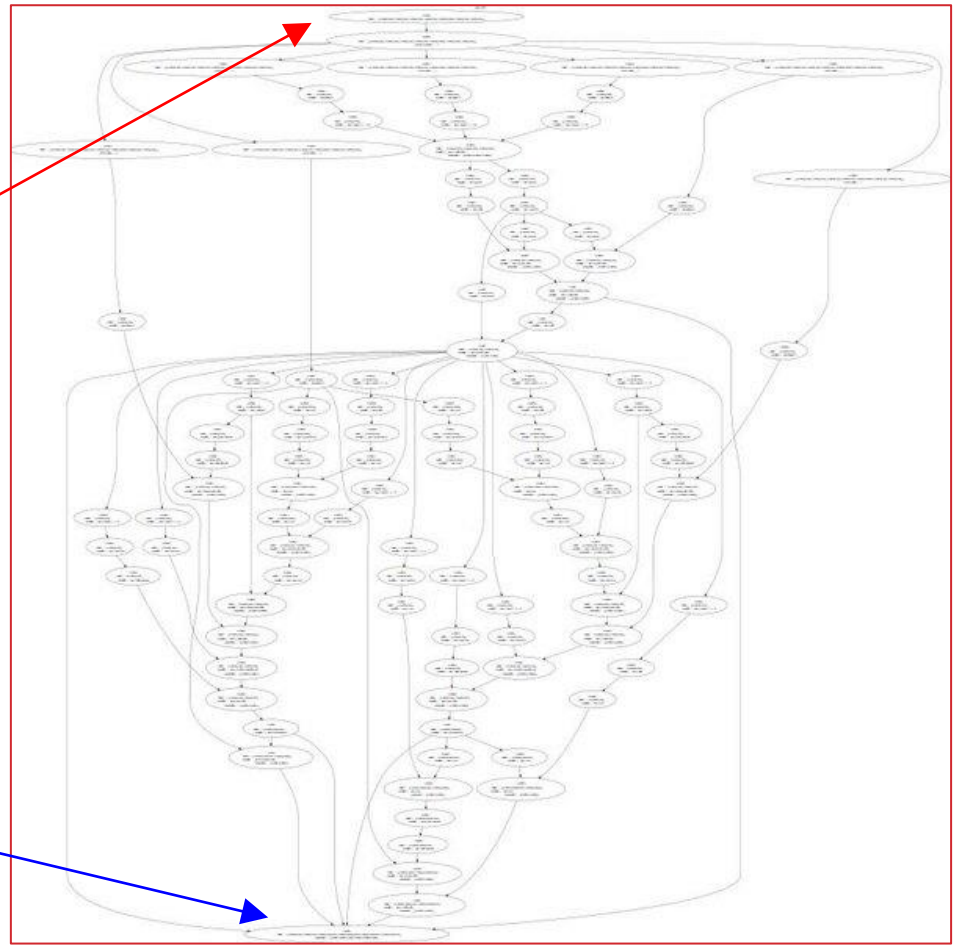
Perda



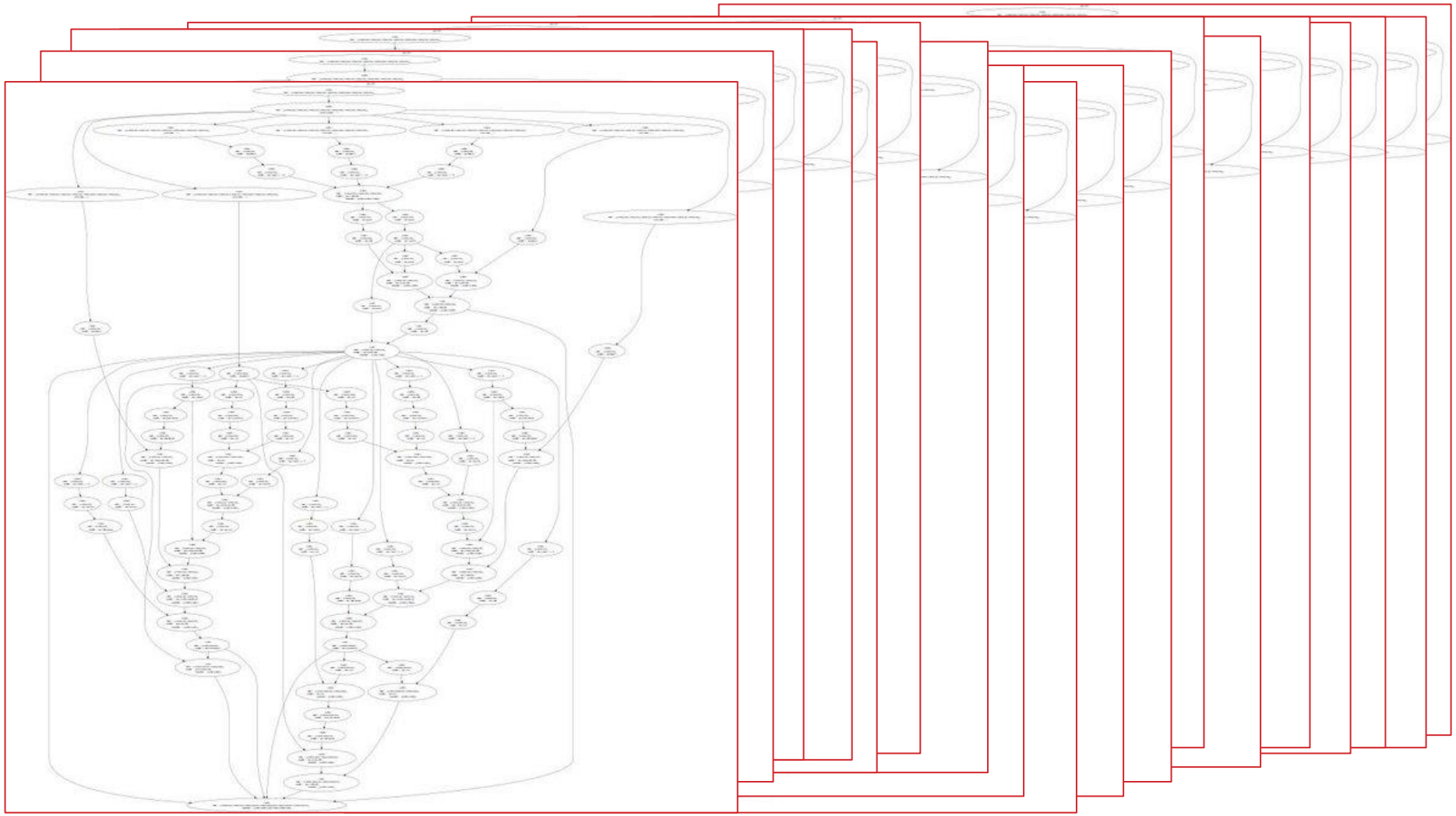
Máquina Neural de Turing

Fita de entrada

Perda

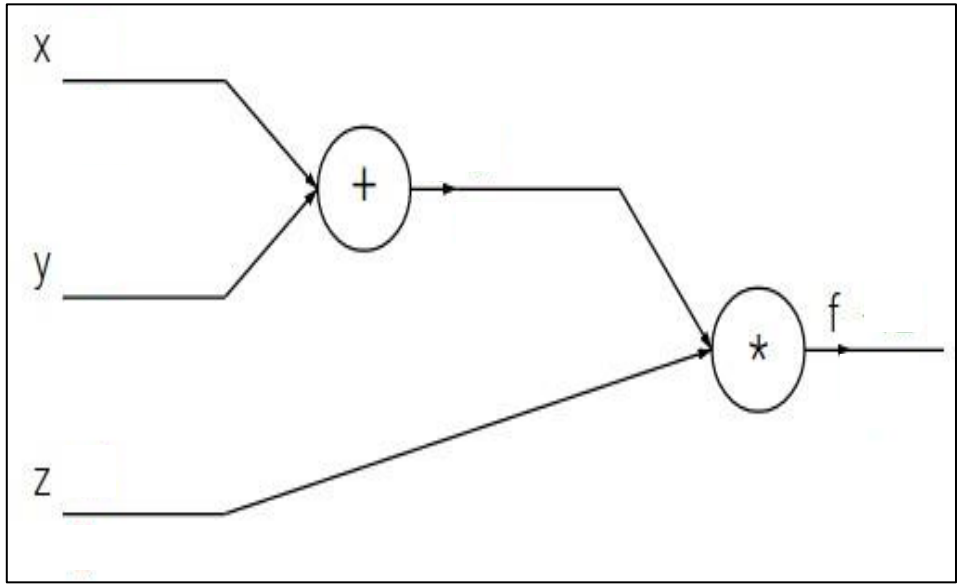


Neural Turing Machine



Diferenciação de um Grafo de Computação

Seja $f(x, y, z) = (x + y) \times z$

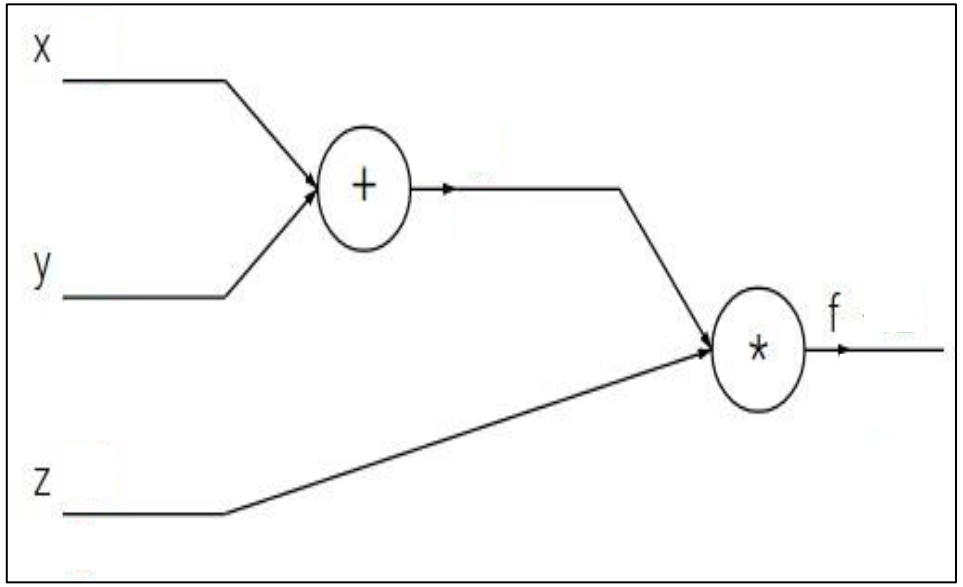


Diferenciação de um Grafo de Computação

Seja $f(x, y, z) = (x + y) \times z$

Então, pode-se dizer que

$$f(x, y, z) = g(h(x, y), z)$$



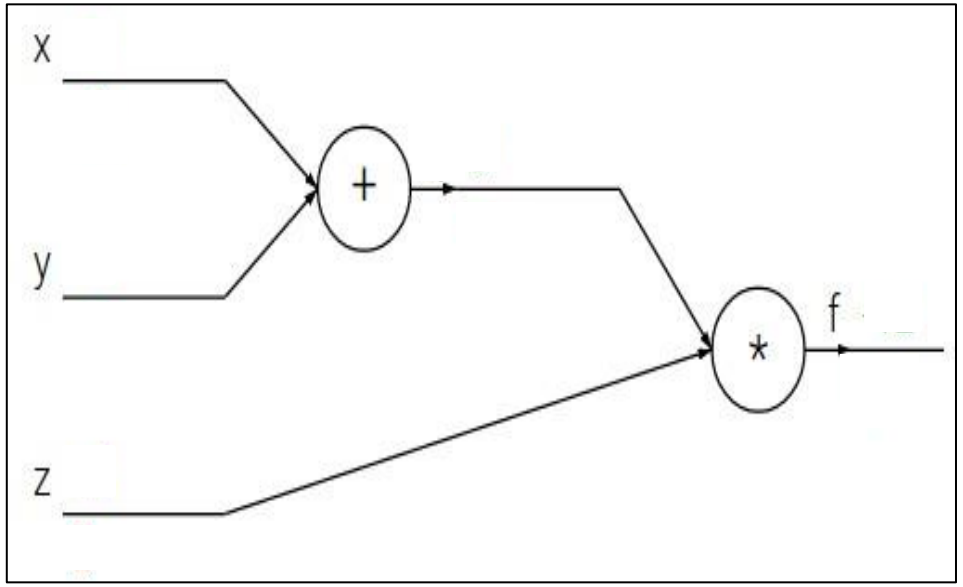
Diferenciação de um Grafo de Computação

Seja $f(x, y, z) = (x + y) \times z$

Então, pode-se dizer que

$$f(x, y, z) = g(h(x, y), z)$$

em $h(x, y) = x + y$



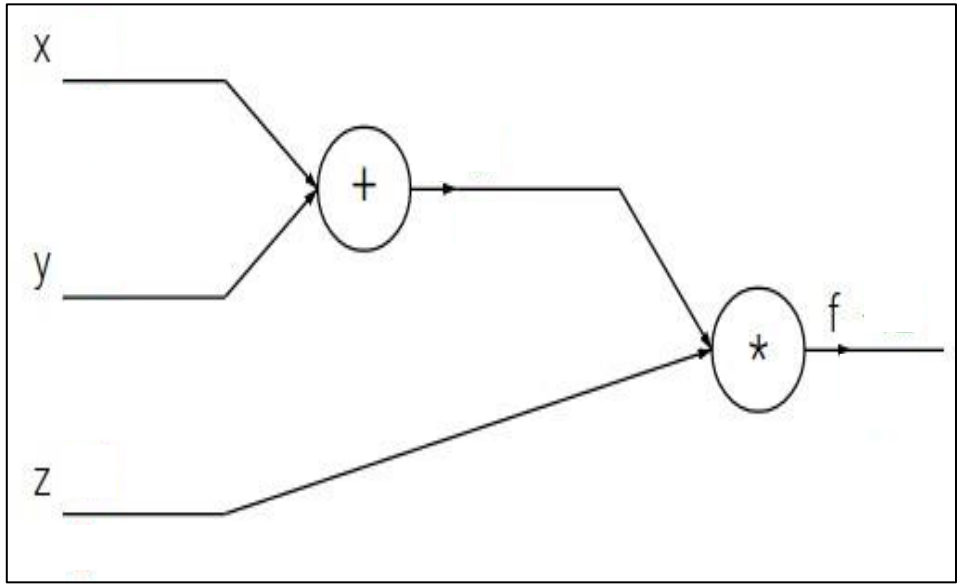
Diferenciação de um Grafo de Computação

Seja $f(x, y, z) = (x + y) \times z$

Então, pode-se dizer que

$$f(x, y, z) = g(h(x, y), z)$$

em $h(x, y) = x + y$ e $g(a, b) = a \times b$



Diferenciação de um Grafo de Computação

Seja $f(x, y, z) = (x + y) \times z$

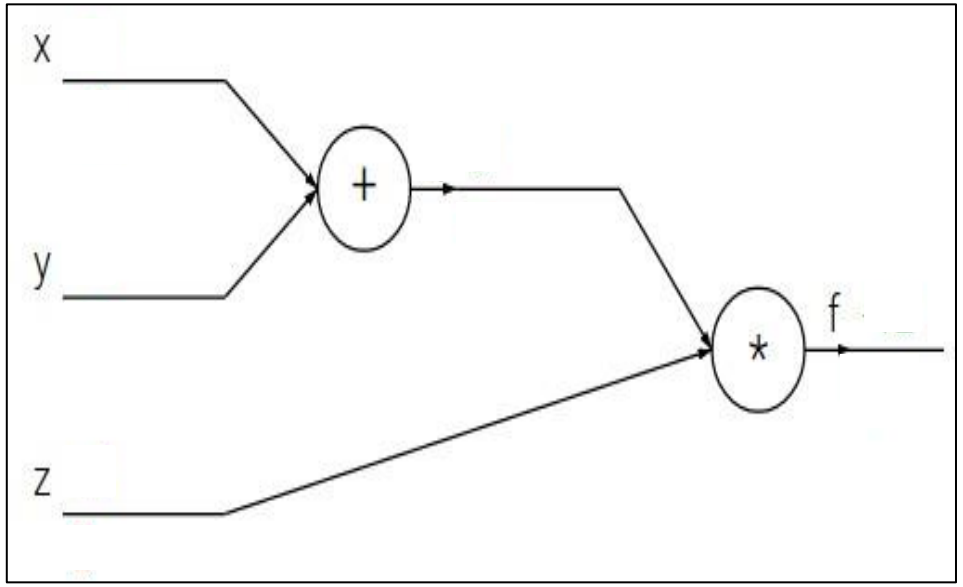
Então, pode-se dizer que

$$f(x, y, z) = g(h(x, y), z)$$

em $h(x, y) = x + y$ e $g(a, b) = a \times b$

Pela **regra da cadeia**, sabe-se que

$$\frac{df}{dx} = \frac{dg}{dh} \times \frac{dh}{dx}$$



Diferenciação de um Grafo de Computação

Seja $f(x, y, z) = (x + y) \times z$

Então, pode-se dizer que

$$f(x, y, z) = g(h(x, y), z)$$

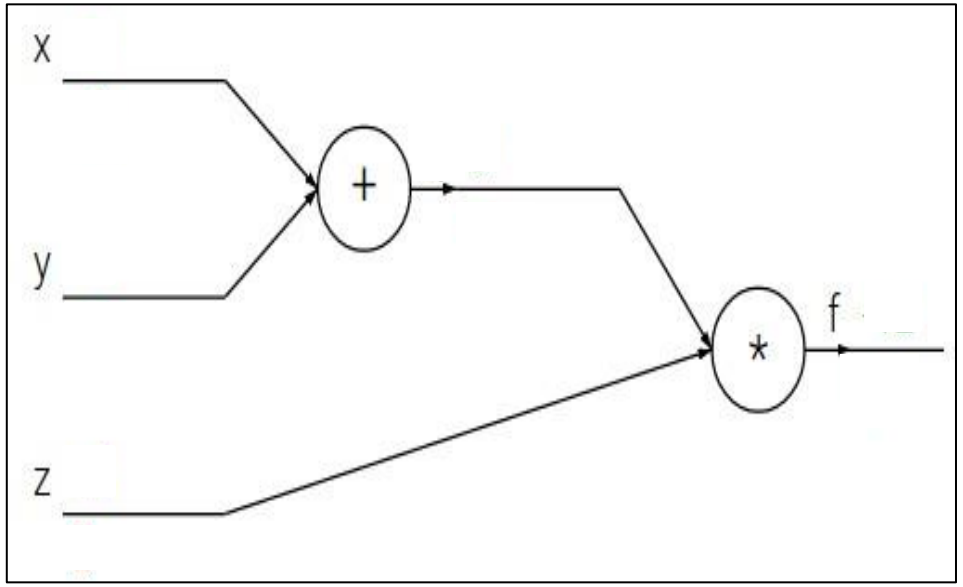
em $h(x, y) = x + y$ e $g(a, b) = a \times b$

Pela **regra da cadeia**, sabe-se que

$$\frac{df}{dx} = \frac{dg}{dh} \times \frac{dh}{dx}$$

e

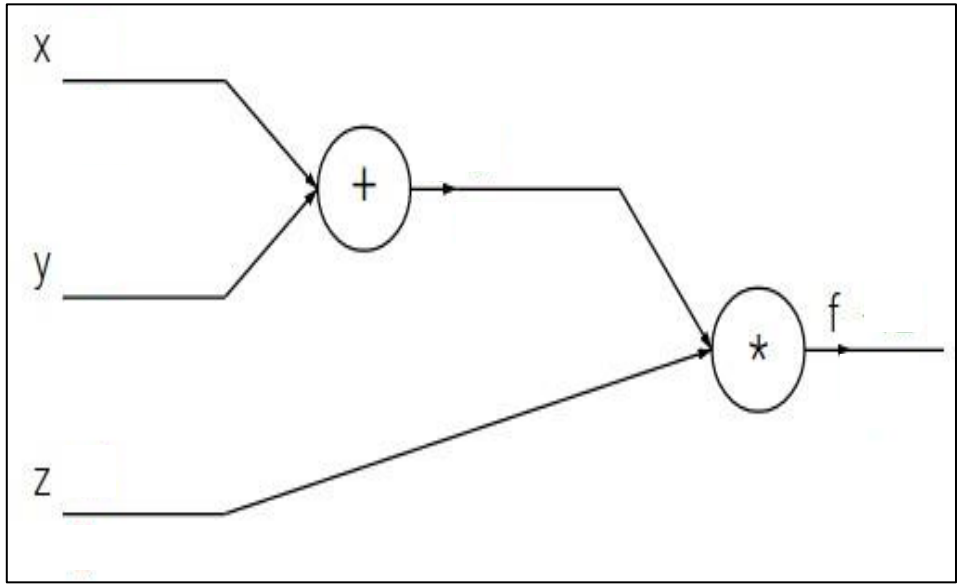
$$\frac{df}{dy} = \frac{dg}{dh} \times \frac{dh}{dy}$$



Diferenciação de um Grafo de Computação

Seja $f(x, y, z) = (x + y) \times z$

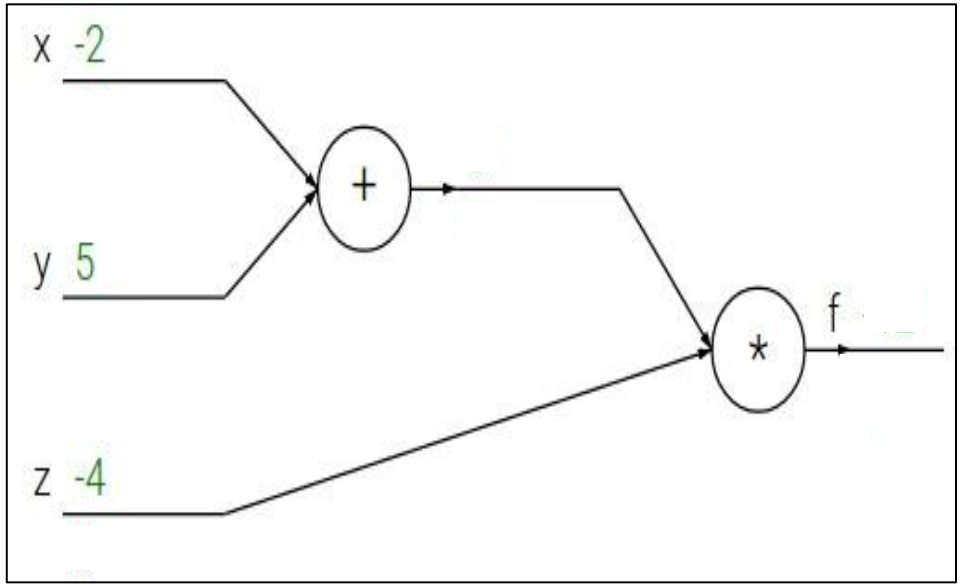
Por exemplo: x = -2, y = 5, z = -4



Diferenciação de um Grafo de Computação

Seja $f(x, y, z) = (x + y) \times z$

Por exemplo: x = -2, y = 5, z = -4

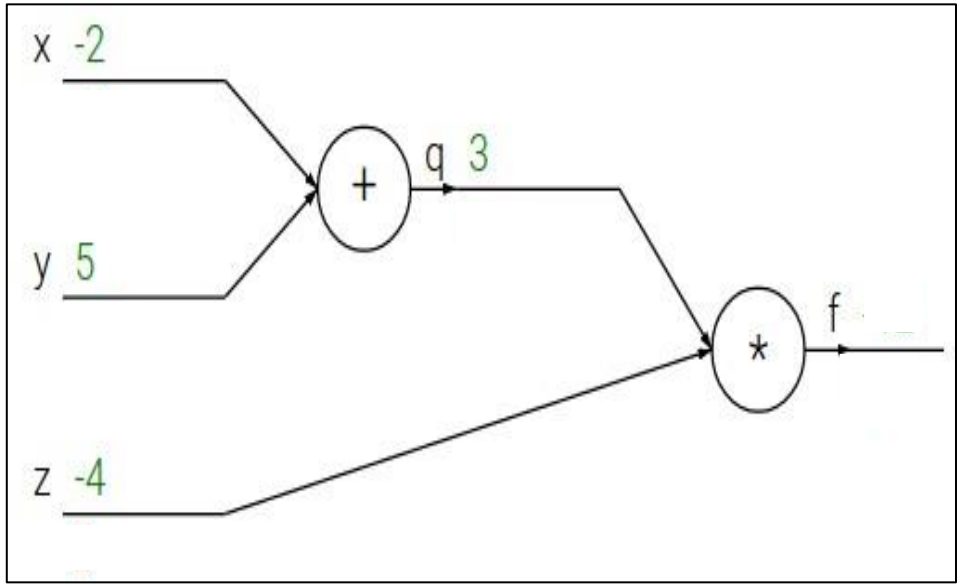


Diferenciação de um Grafo de Computação

Seja $f(x, y, z) = (x + y) \times z$

Por exemplo: x = -2, y = 5, z = -4

$$q = x + y$$



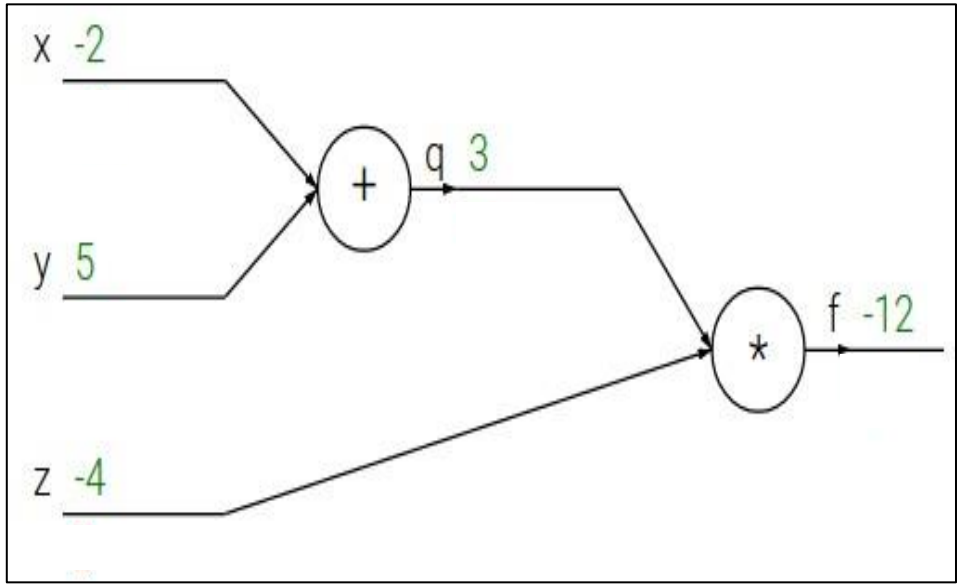
Diferenciação de um Grafo de Computação

Seja $f(x, y, z) = (x + y) \times z$

Por exemplo: x = -2, y = 5, z = -4

$$q = x + y$$

$$f = qz$$



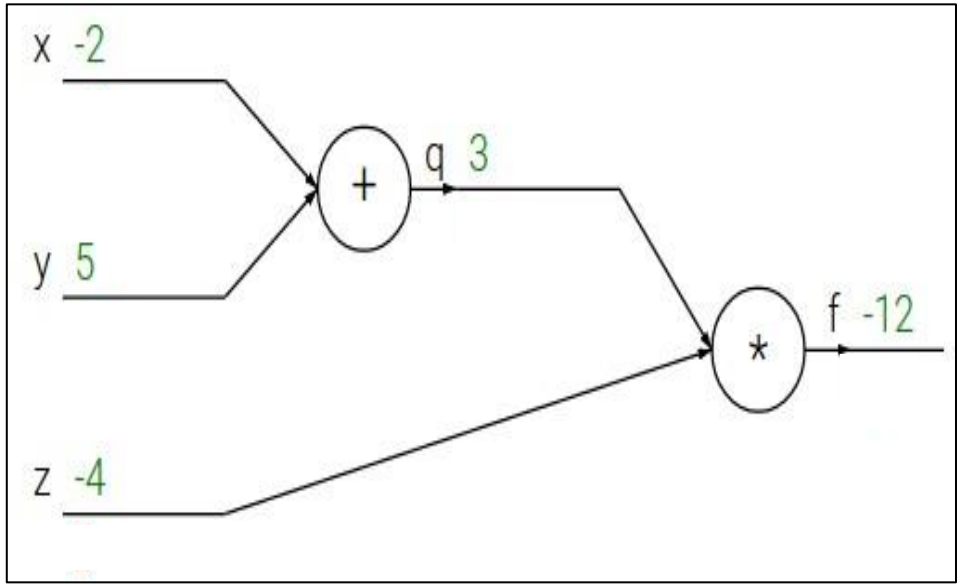
Diferenciação de um Grafo de Computação

Seja $f(x, y, z) = (x + y) \times z$

Por exemplo: x = -2, y = 5, z = -4

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

$$f = qz$$



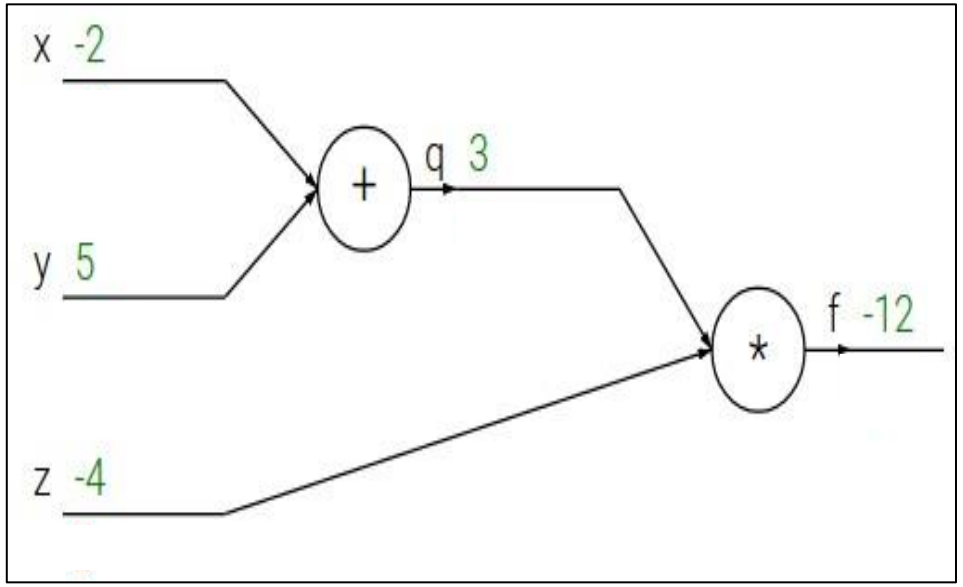
Diferenciação de um Grafo de Computação

Seja $f(x, y, z) = (x + y) \times z$

Por exemplo: x = -2, y = 5, z = -4

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

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



Diferenciação de um Grafo de Computação

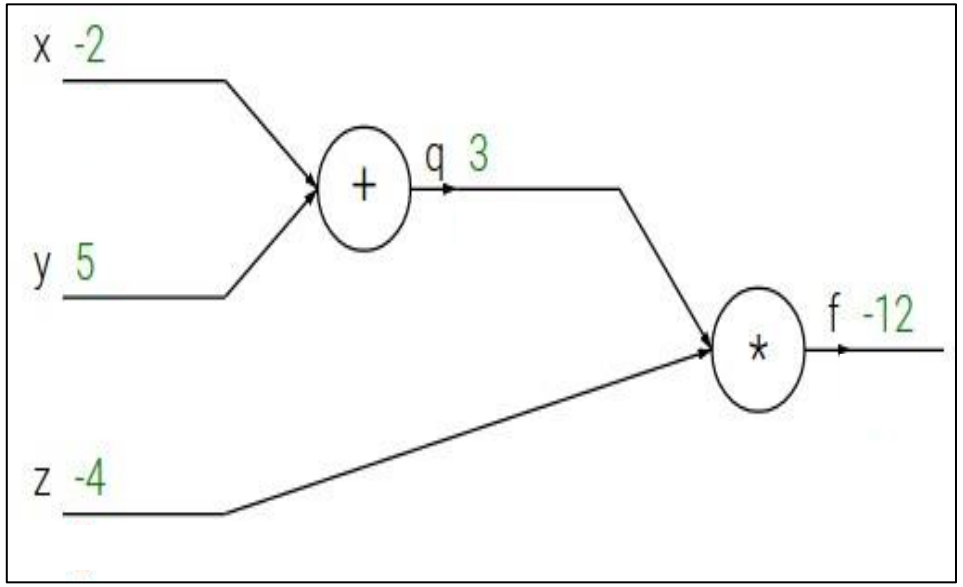
Seja $f(x, y, z) = (x + y) \times z$

Por exemplo: x = -2, y = 5, z = -4

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

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

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



Diferenciação de um Grafo de Computação

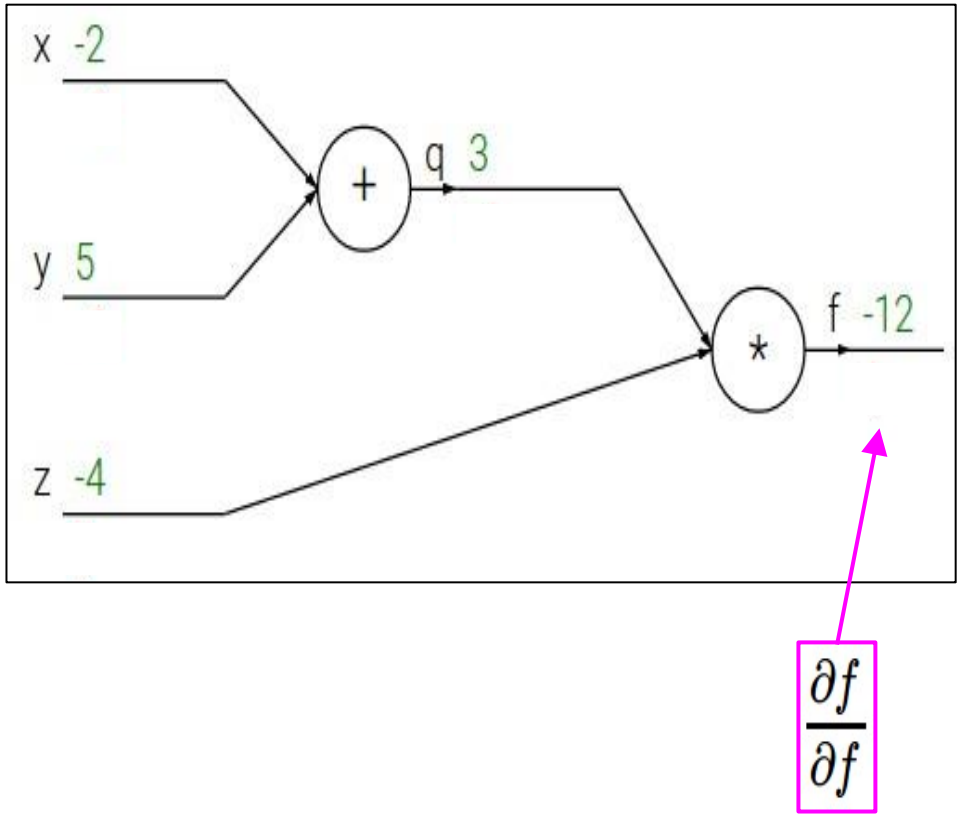
Seja $f(x, y, z) = (x + y) \times z$

Por exemplo: x = -2, y = 5, z = -4

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

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

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



Diferenciação de um Grafo de Computação

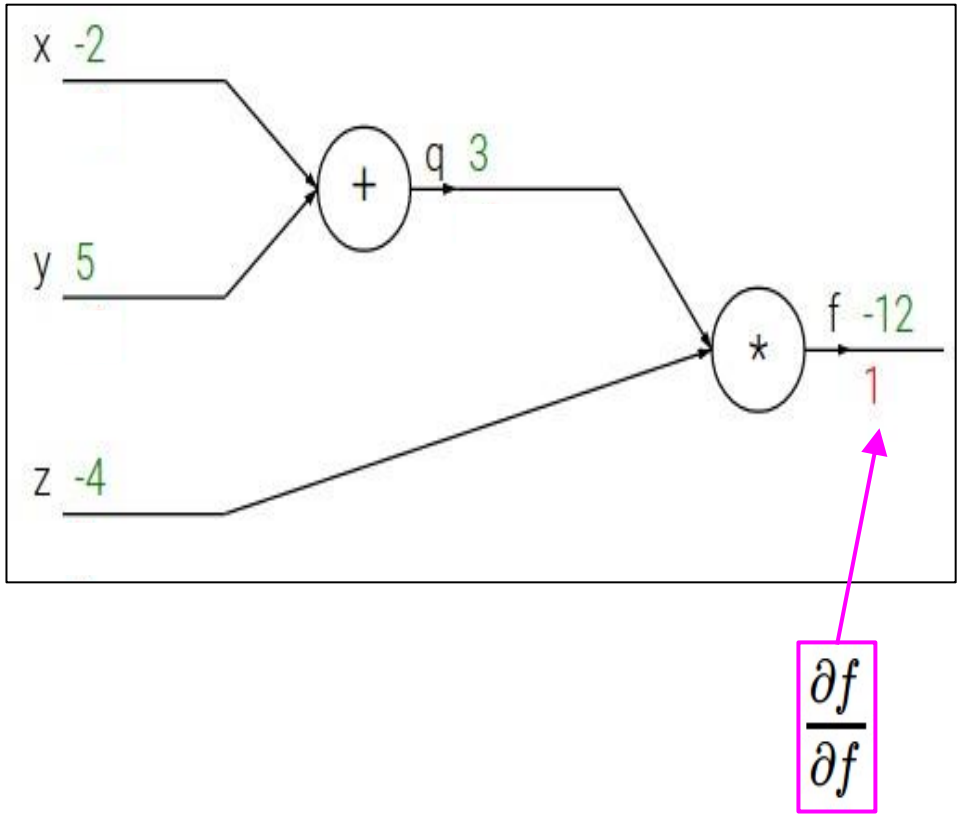
Seja $f(x, y, z) = (x + y) \times z$

Por exemplo: x = -2, y = 5, z = -4

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

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

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



Diferenciação de um Grafo de Computação

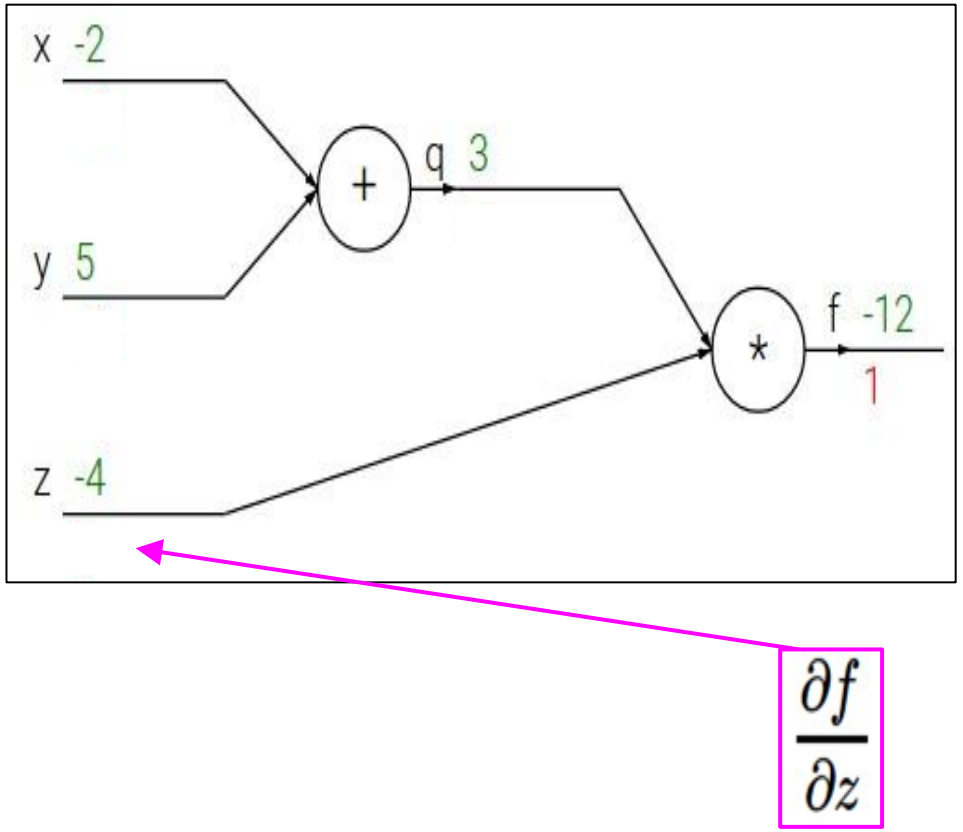
Seja $f(x, y, z) = (x + y) \times z$

Por exemplo: x = -2, y = 5, z = -4

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

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

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



Diferenciação de um Grafo de Computação

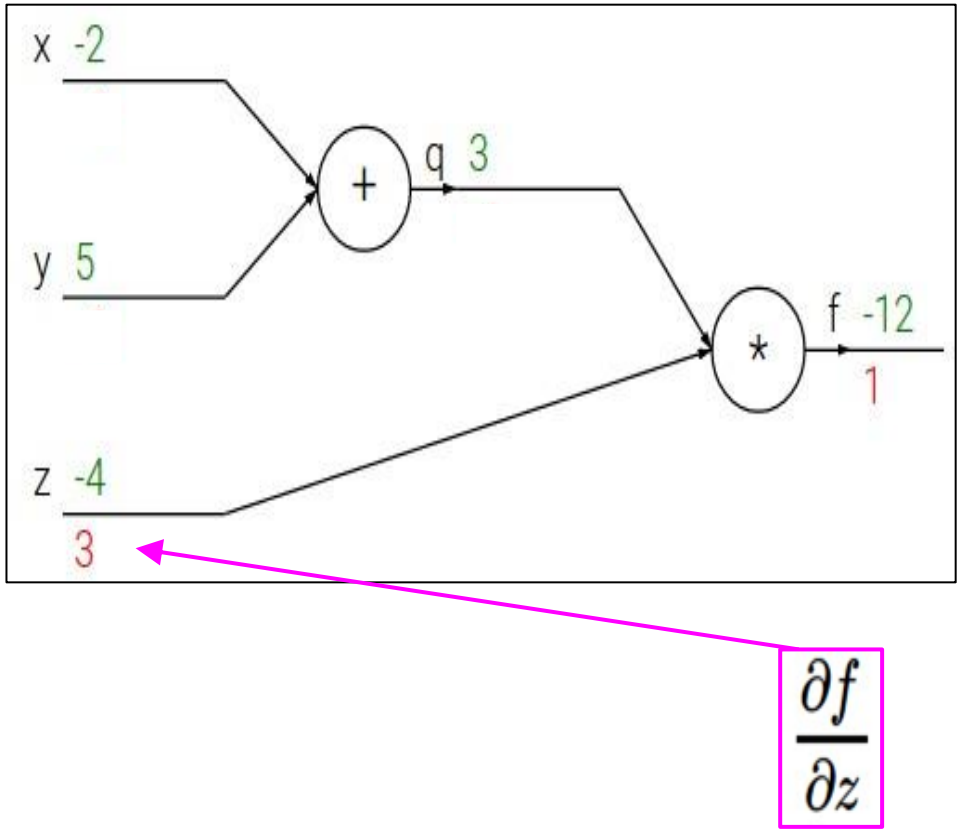
Seja $f(x, y, z) = (x + y) \times z$

Por exemplo: x = -2, y = 5, z = -4

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

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

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



Diferenciação de um Grafo de Computação

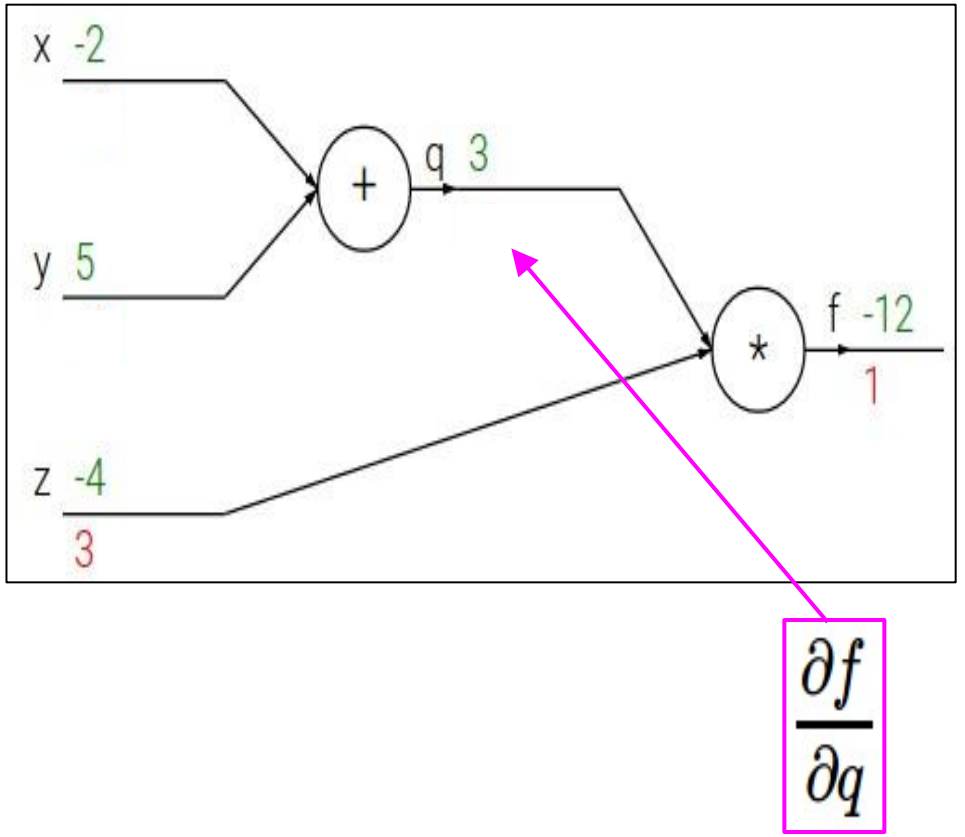
Seja $f(x, y, z) = (x + y) \times z$

Por exemplo: x = -2, y = 5, z = -4

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

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

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



Diferenciação de um Grafo de Computação

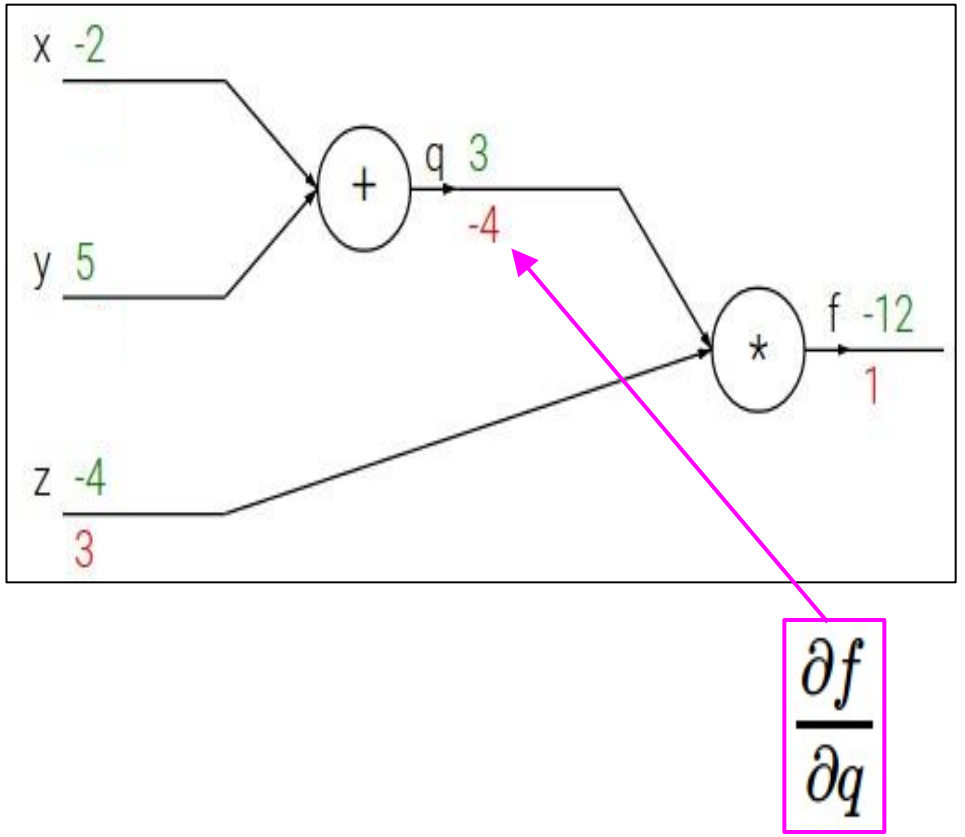
Seja $f(x, y, z) = (x + y) \times z$

Por exemplo: x = -2, y = 5, z = -4

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

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

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



Diferenciação de um Grafo de Computação

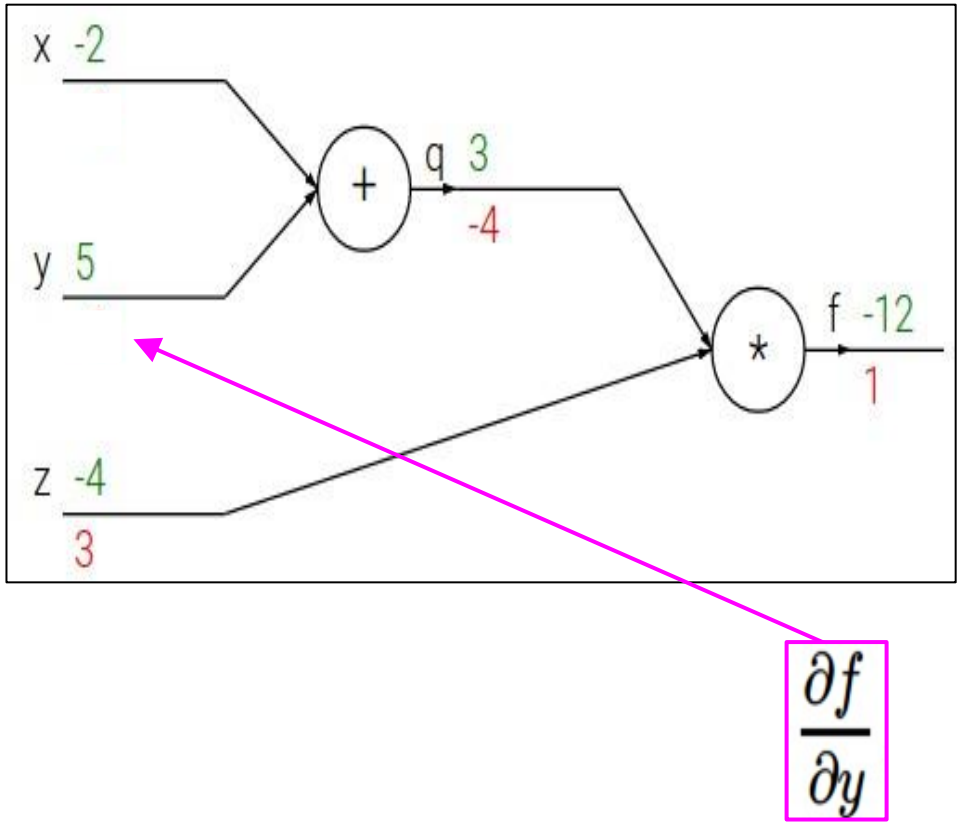
Seja $f(x, y, z) = (x + y) \times z$

Por exemplo: x = -2, y = 5, z = -4

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

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

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



Diferenciação de um Grafo de Computação

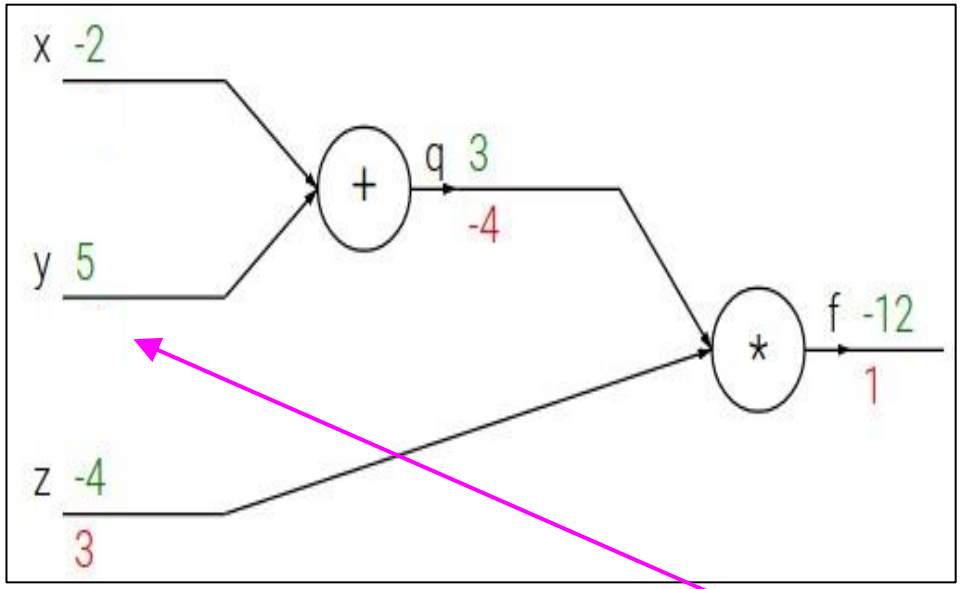
Seja $f(x, y, z) = (x + y) \times z$

Por exemplo: x = -2, y = 5, z = -4

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

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

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



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

Regra da Cadeia:

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

Diferenciação de um Grafo de Computação

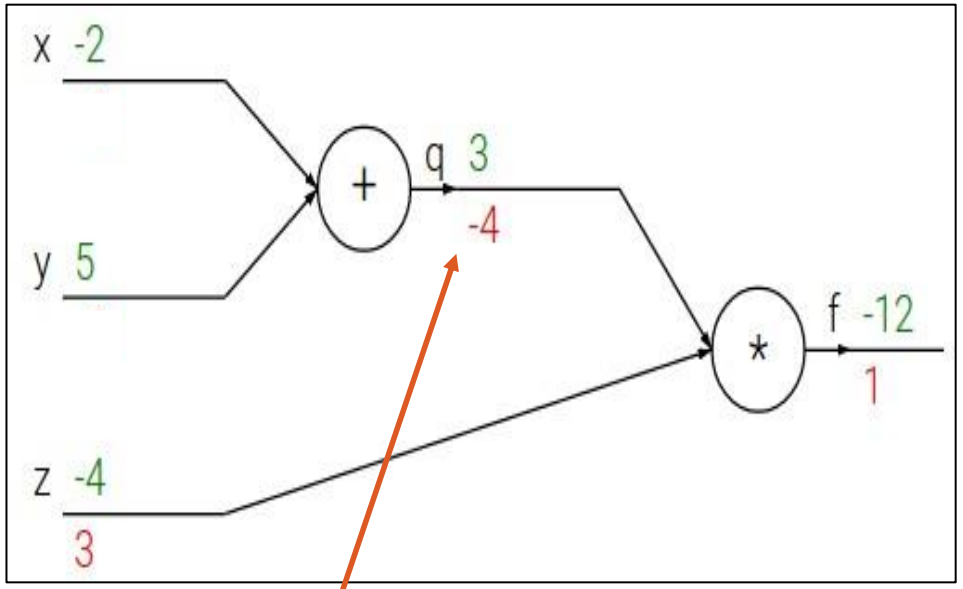
Seja $f(x, y, z) = (x + y) \times z$

Por exemplo: x = -2, y = 5, z = -4

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

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

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



Regra da Cadeia:

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

Diferenciação de um Grafo de Computação

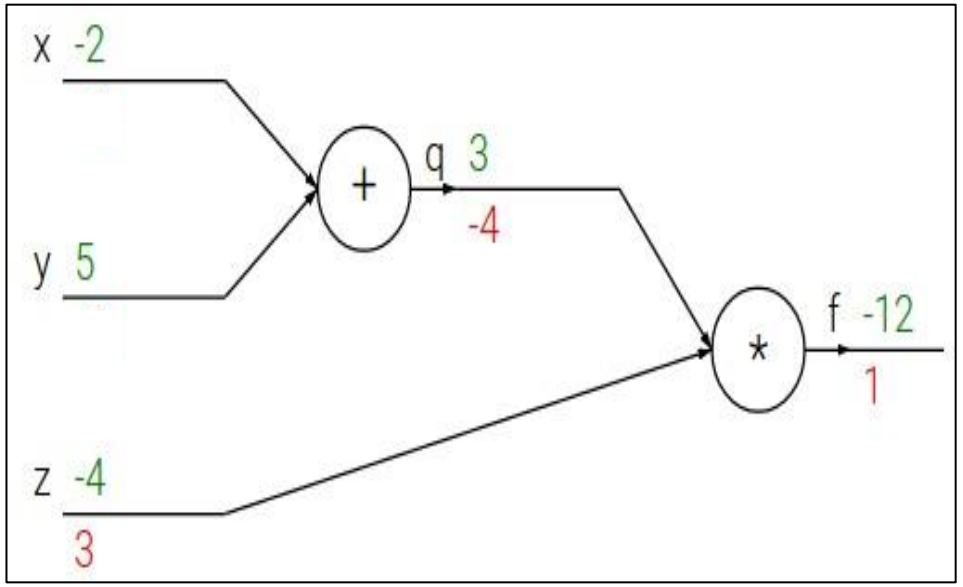
Seja $f(x, y, z) = (x + y) \times z$

Por exemplo: x = -2, y = 5, z = -4

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

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

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



Regra da Cadeia:

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

Diferenciação de um Grafo de Computação

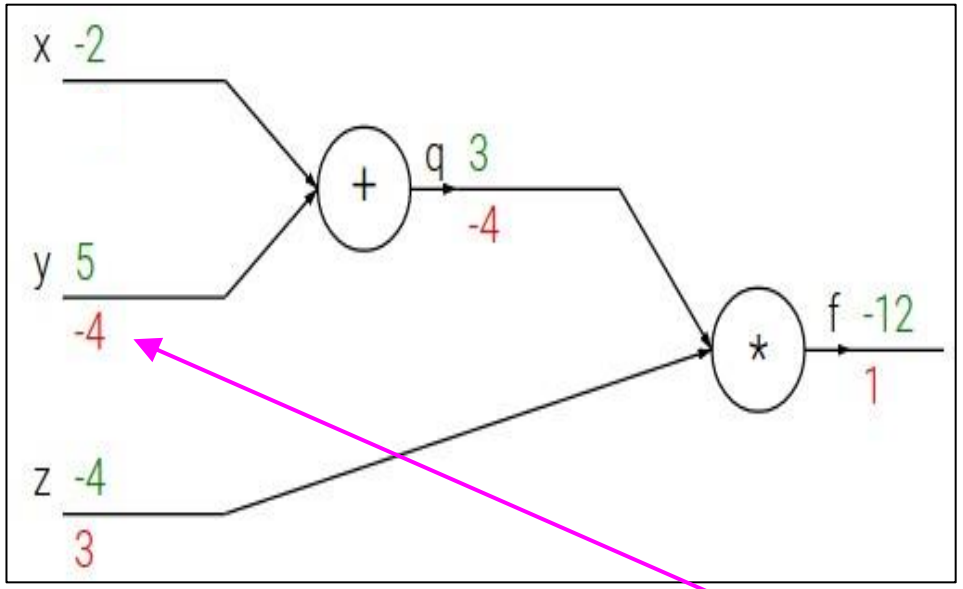
Seja $f(x, y, z) = (x + y) \times z$

Por exemplo: x = -2, y = 5, z = -4

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

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

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



Regra da Cadeia:

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

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

Diferenciação de um Grafo de Computação

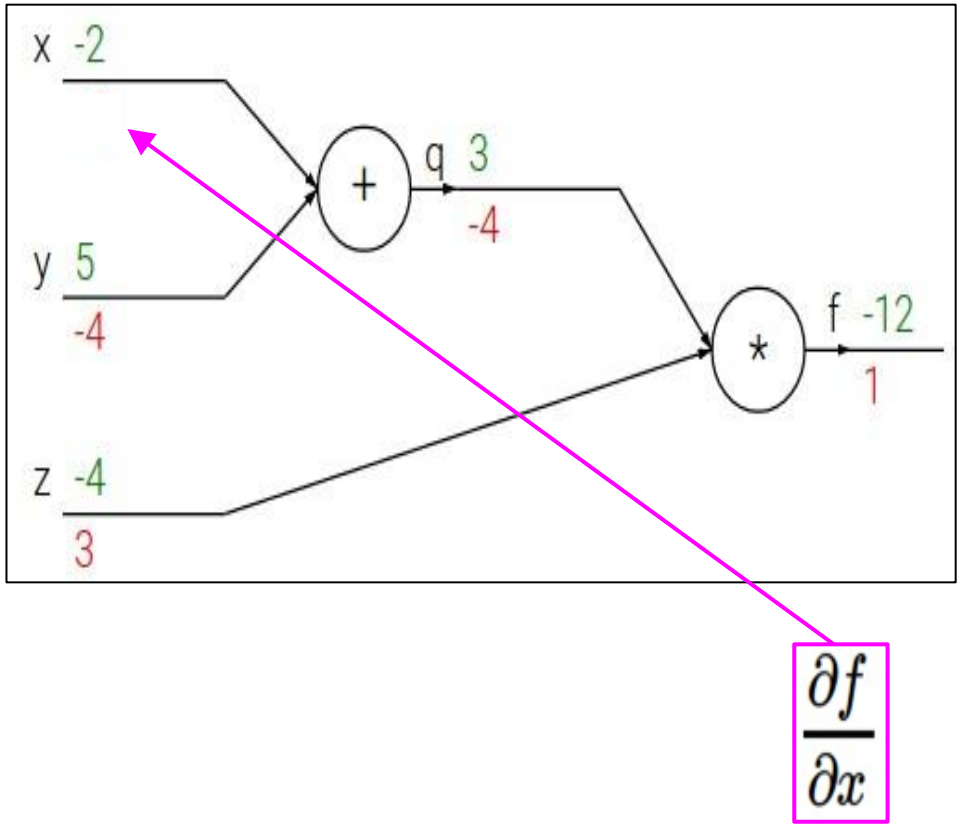
Seja $f(x, y, z) = (x + y) \times z$

Por exemplo: x = -2, y = 5, z = -4

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

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

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



Diferenciação de um Grafo de Computação

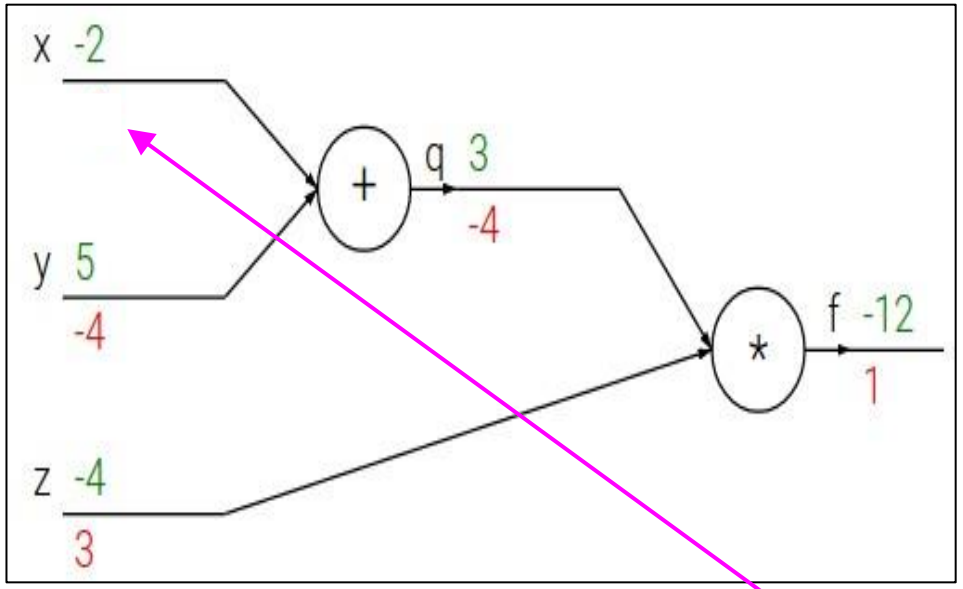
Seja $f(x, y, z) = (x + y) \times z$

Por exemplo: x = -2, y = 5, z = -4

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

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

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



Regra da Cadeia:

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

Diferenciação de um Grafo de Computação

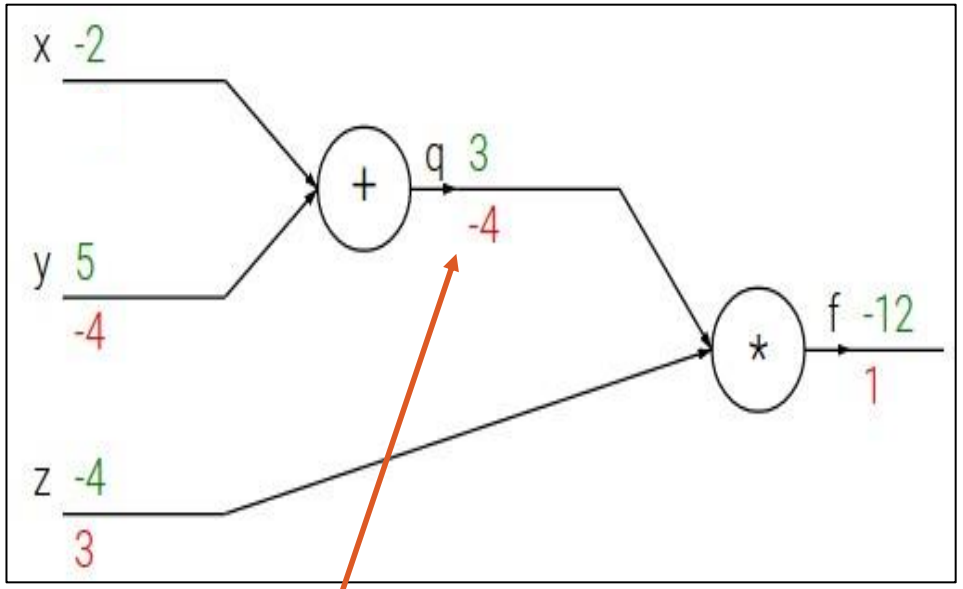
Seja $f(x, y, z) = (x + y) \times z$

Por exemplo: x = -2, y = 5, z = -4

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

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

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



Regra da Cadeia:

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

Diferenciação de um Grafo de Computação

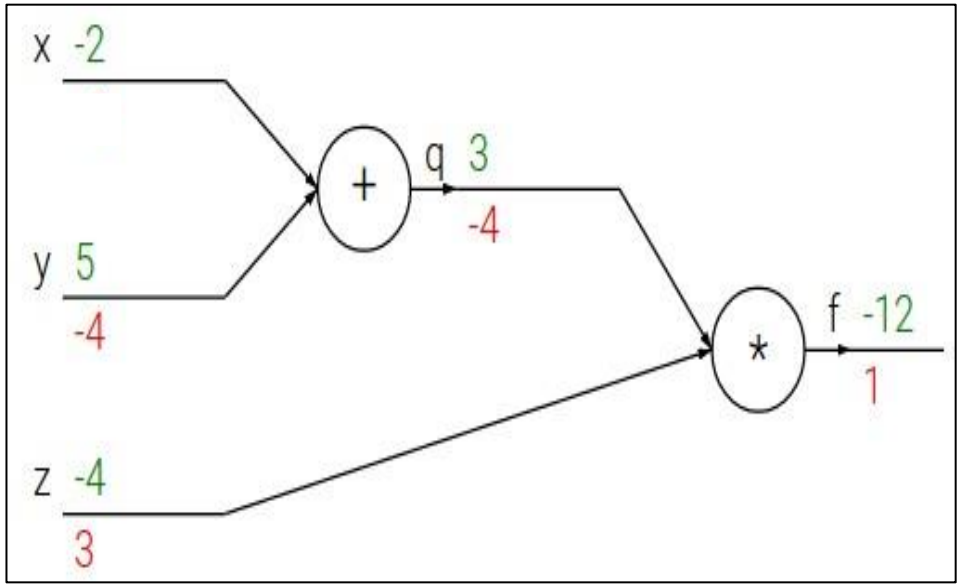
Seja $f(x, y, z) = (x + y) \times z$

Por exemplo: x = -2, y = 5, z = -4

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

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

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



Regra da Cadeia:

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

Diferenciação de um Grafo de Computação

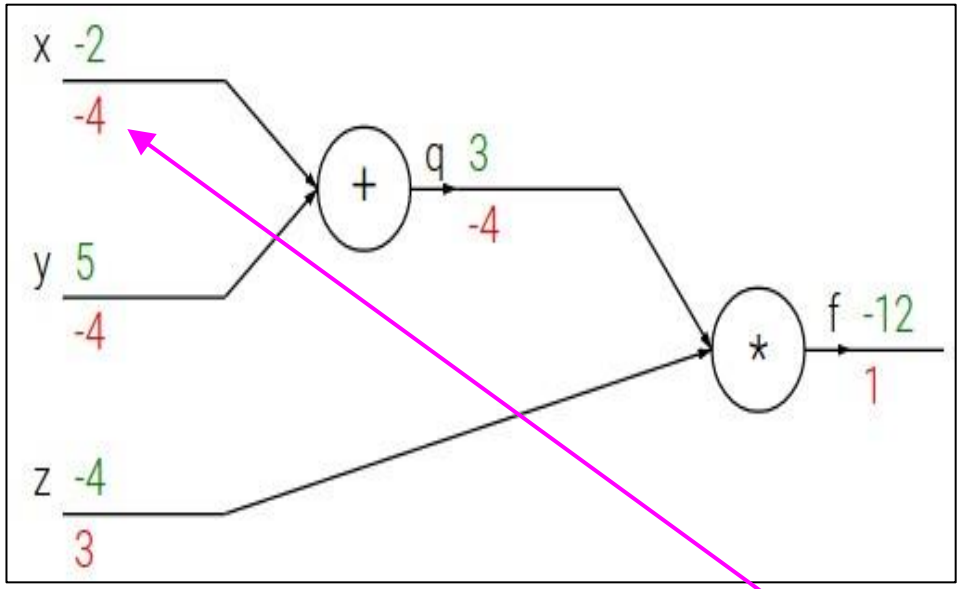
Seja $f(x, y, z) = (x + y) \times z$

Por exemplo: x = -2, y = 5, z = -4

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

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

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

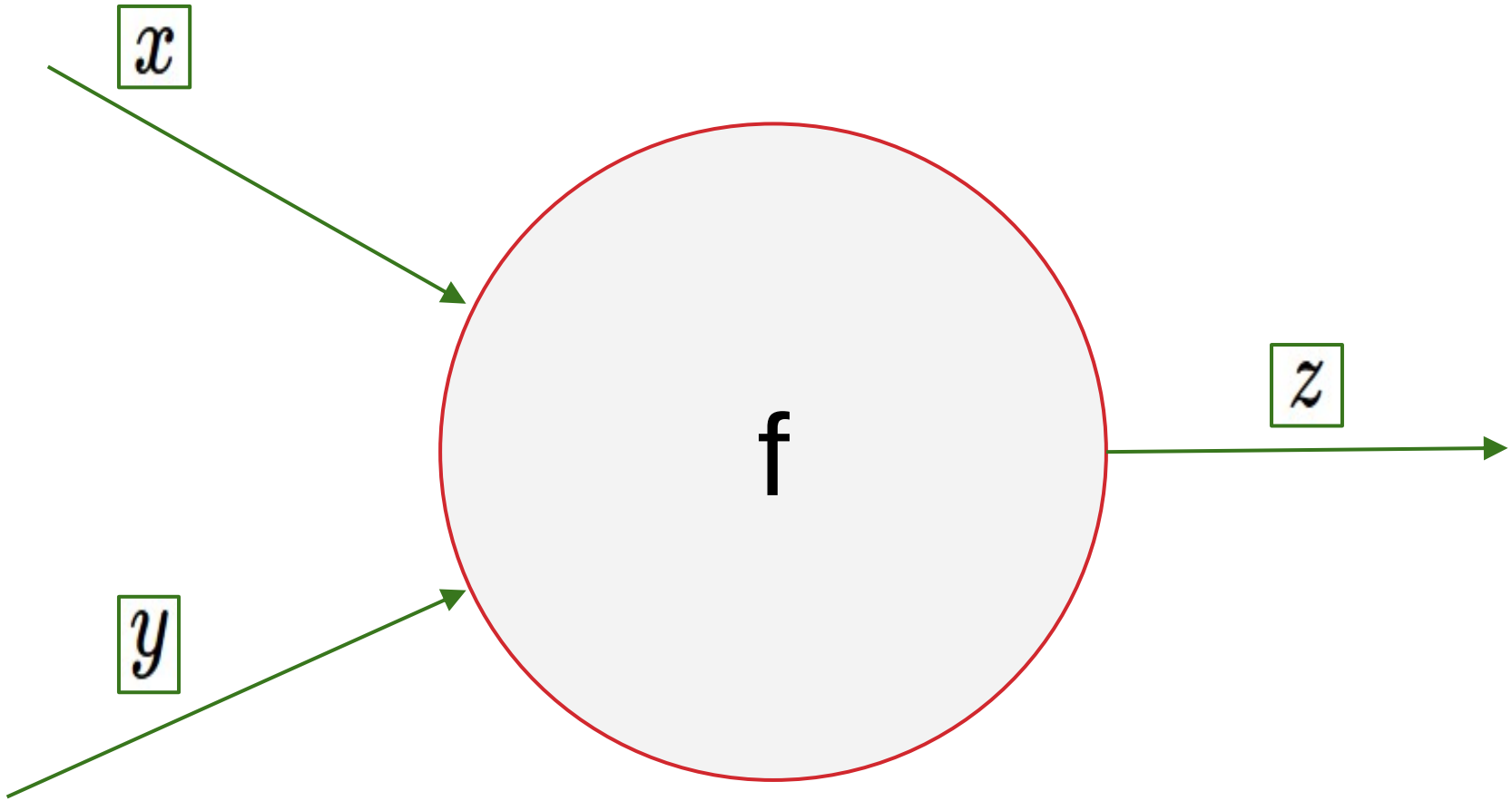


Regra da Cadeia:

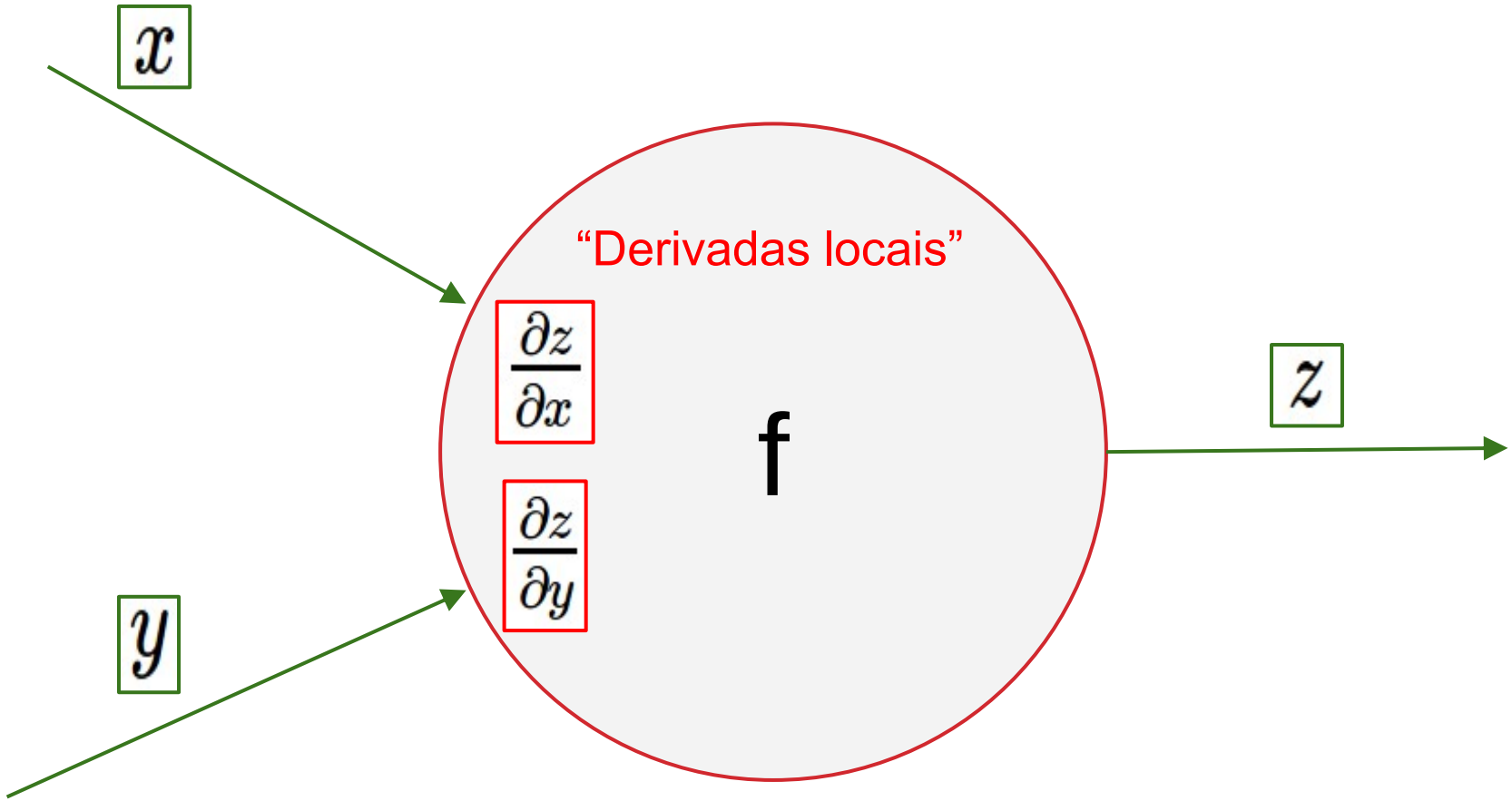
$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial q} \frac{\partial q}{\partial x}$$

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

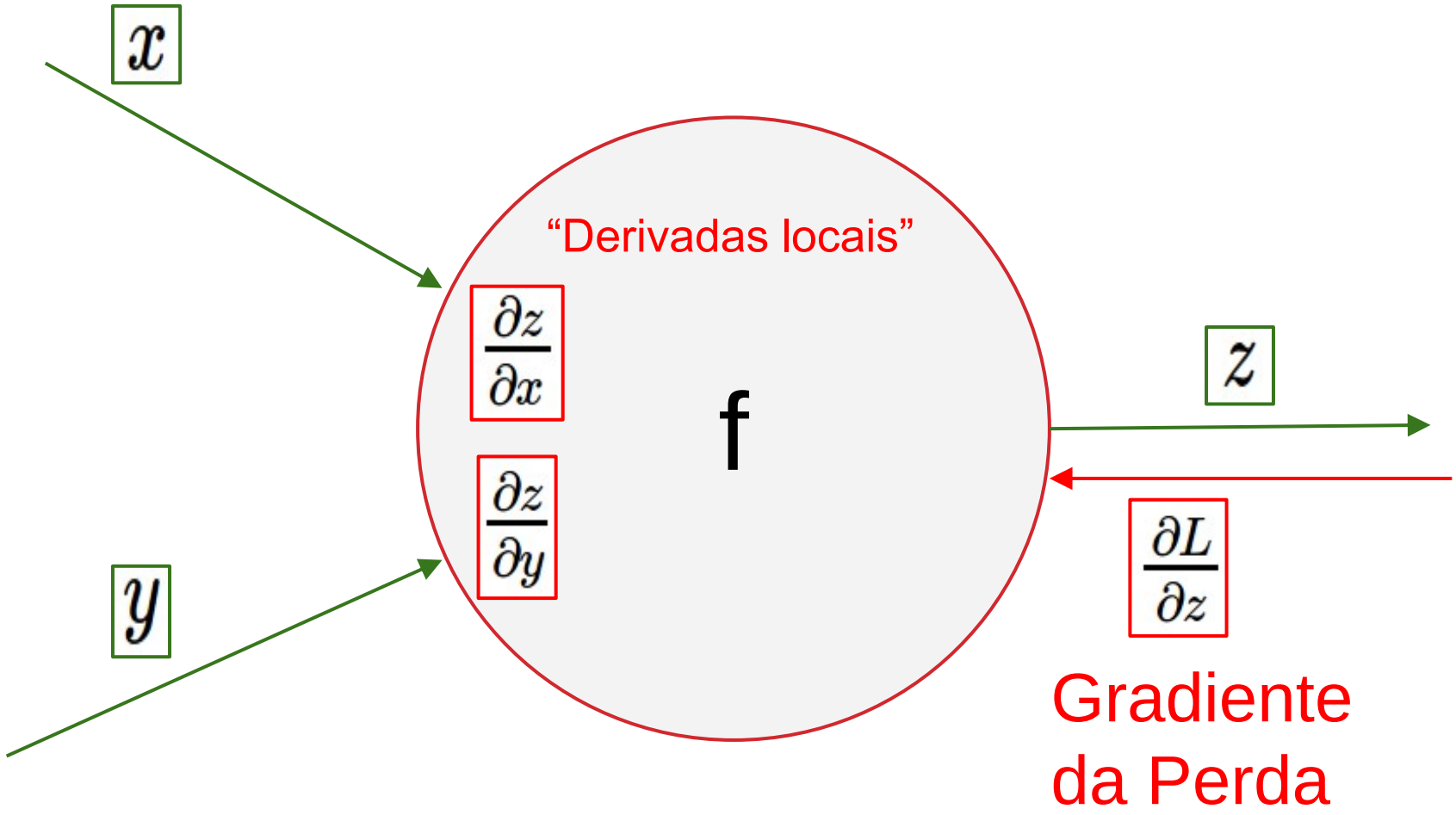
Passo Retrógrado (*Backward Pass*)



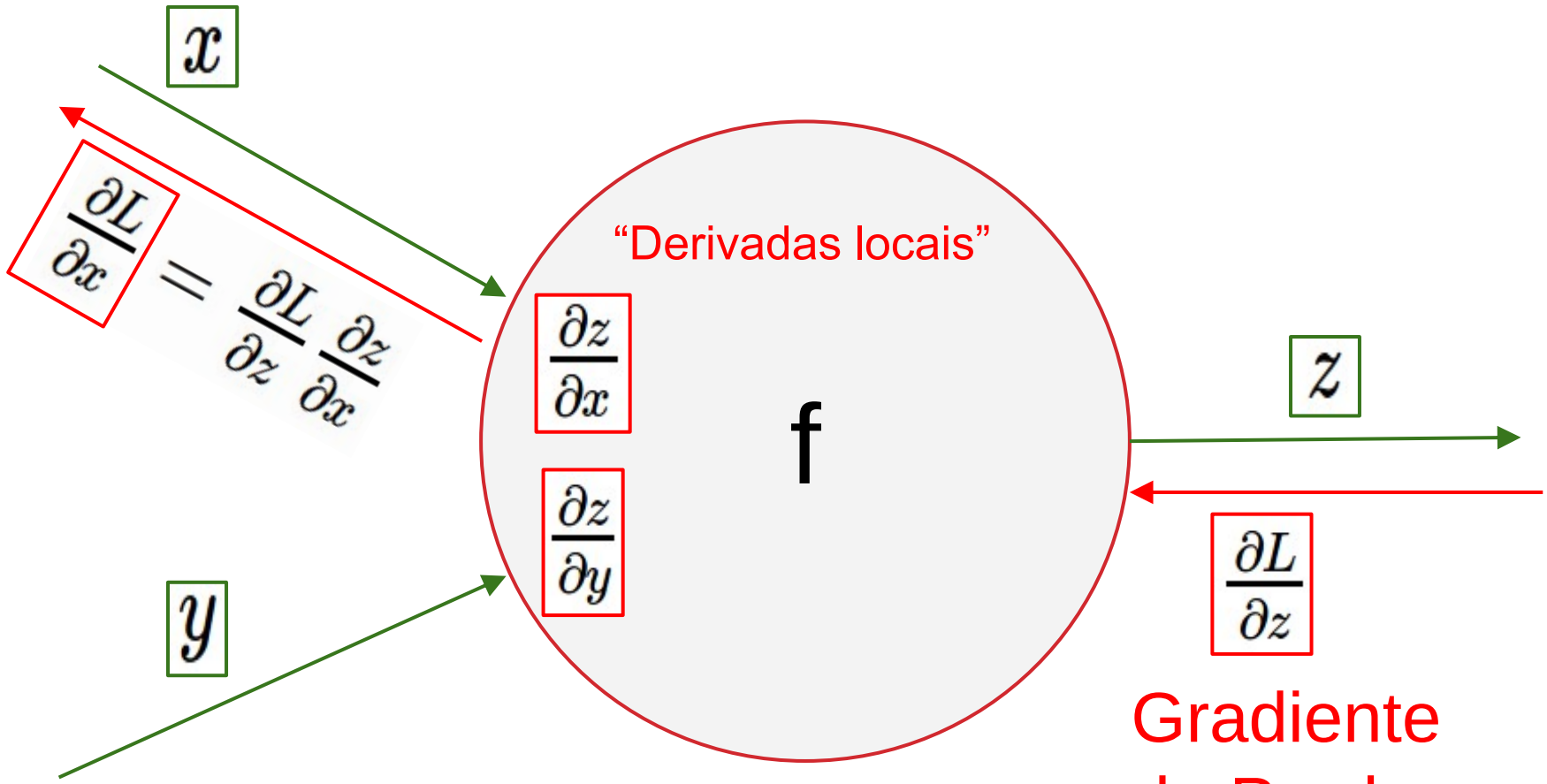
Passo Retrógrado (*Backward Pass*)



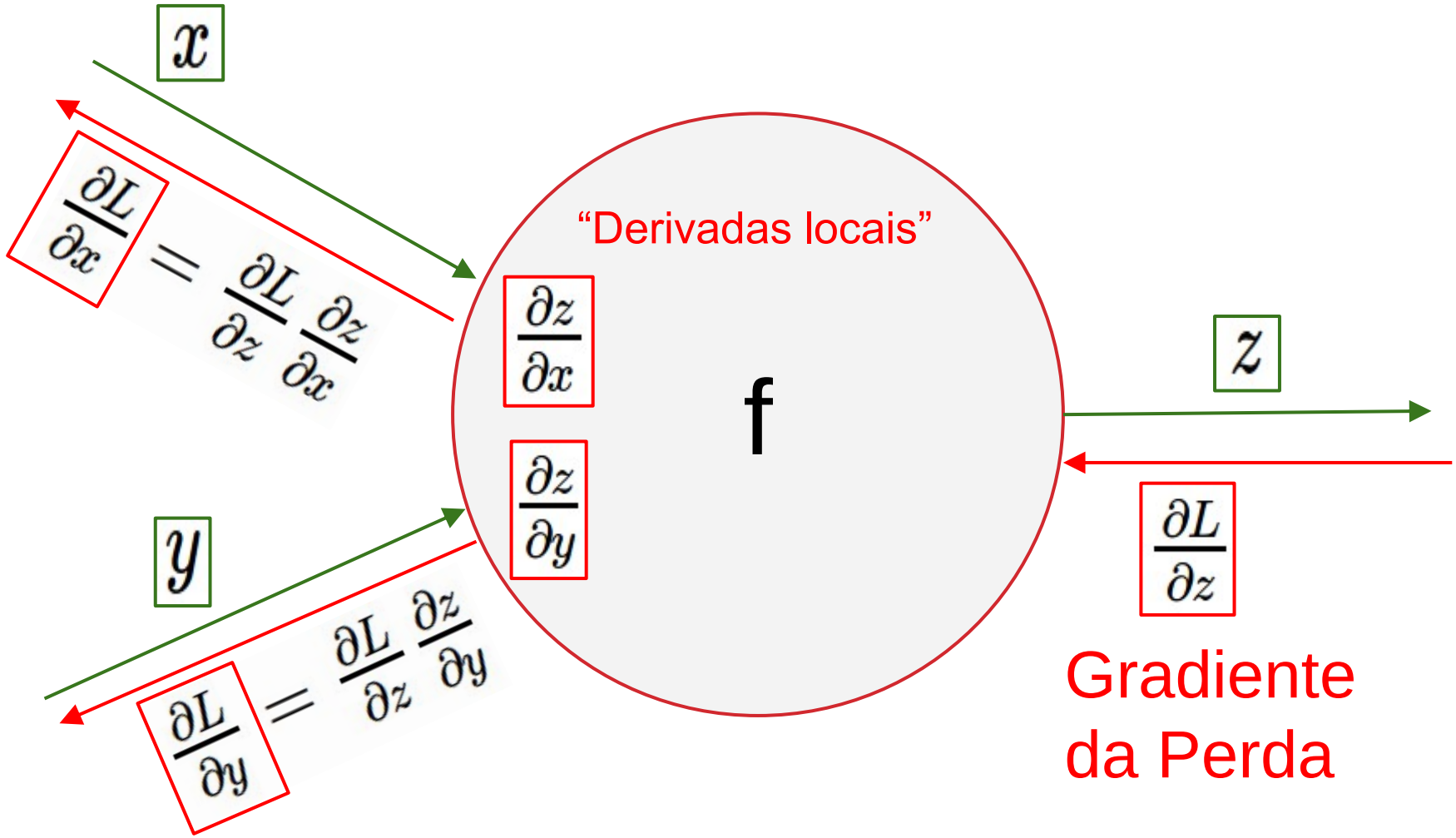
Passo Retrógrado (*Backward Pass*)



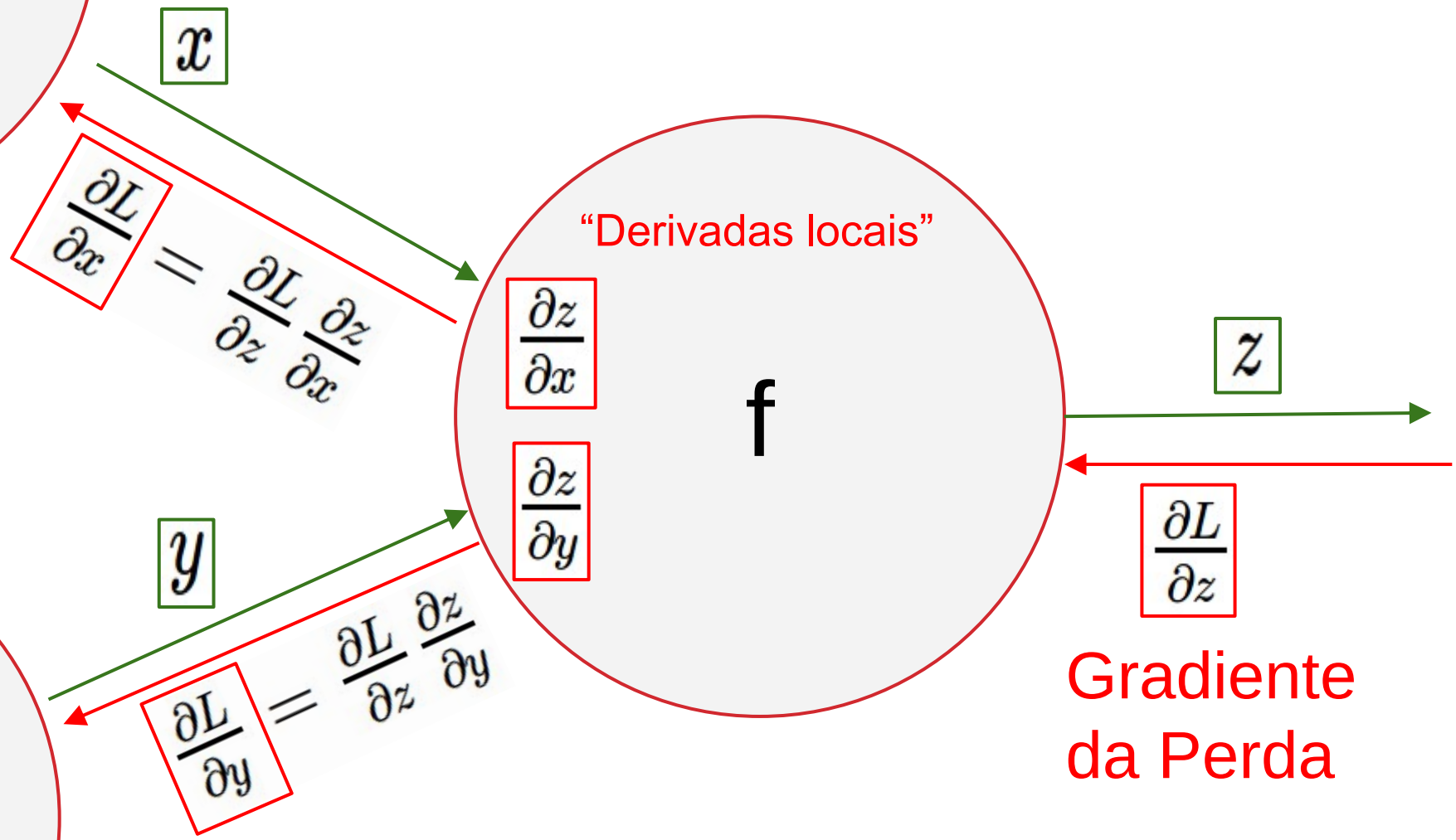
Passo Retrógrado (*Backward Pass*)



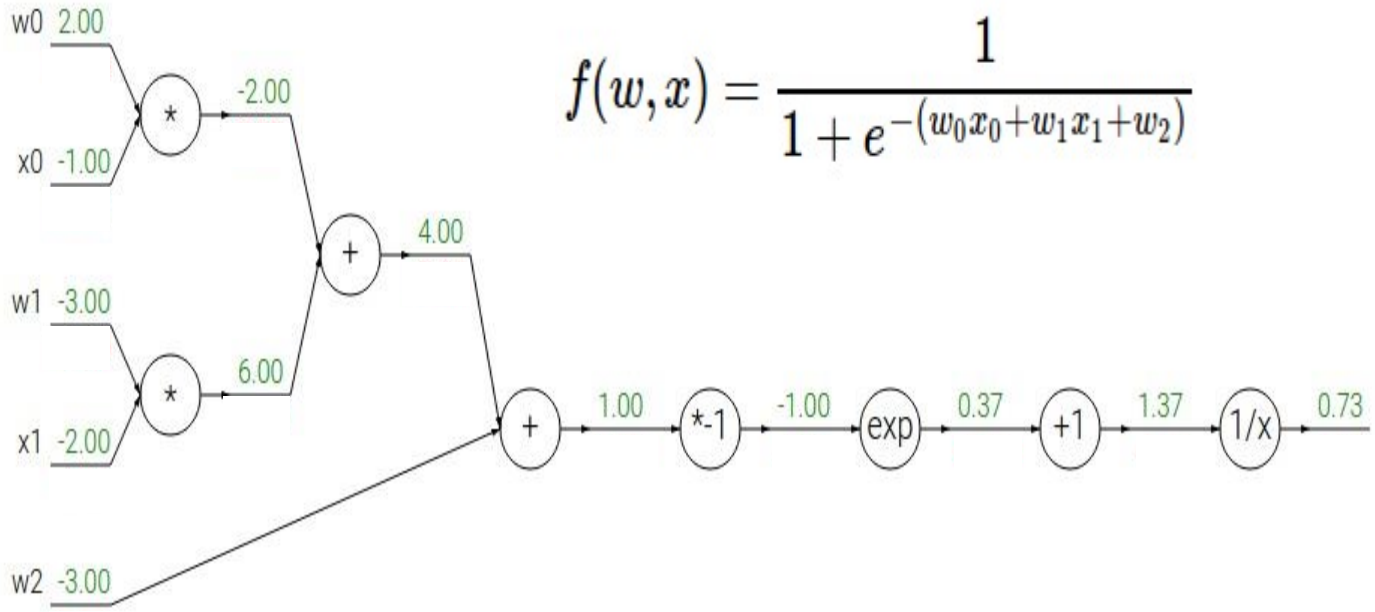
Passo Retrógrado (*Backward Pass*)



Passo Retrógrado (*Backward Pass*)

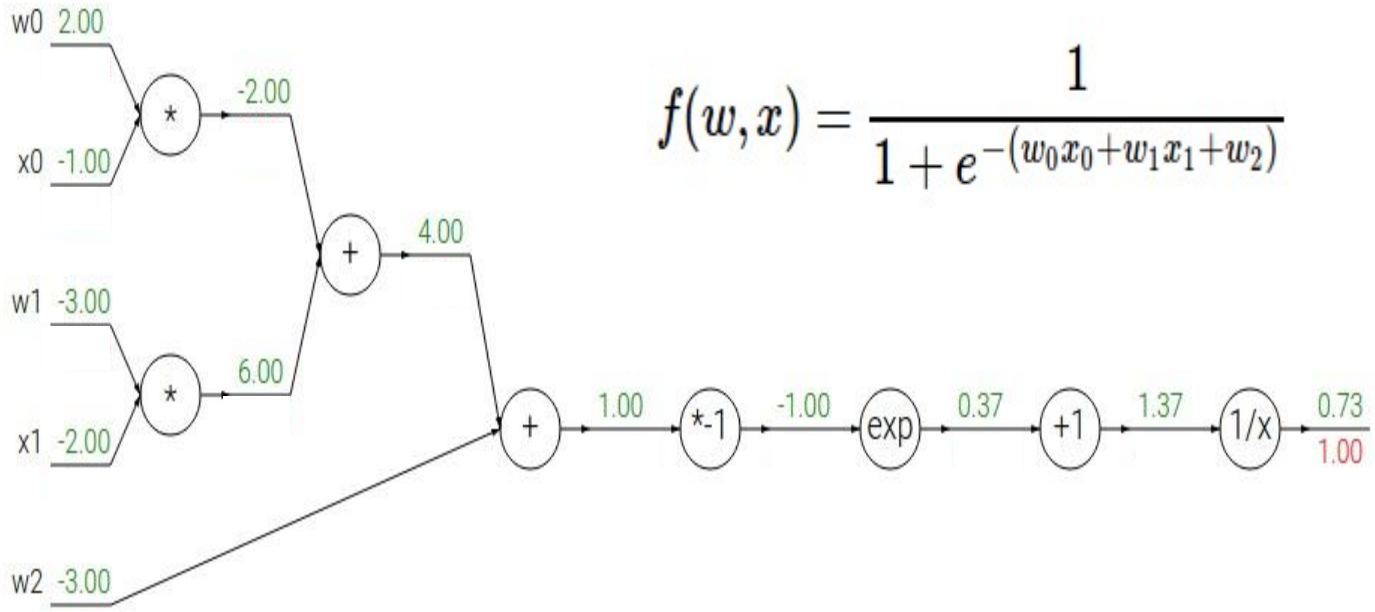


Outro Exemplo – Passo Retrógrado



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

Outro Exemplo – Passo Retrógrado



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

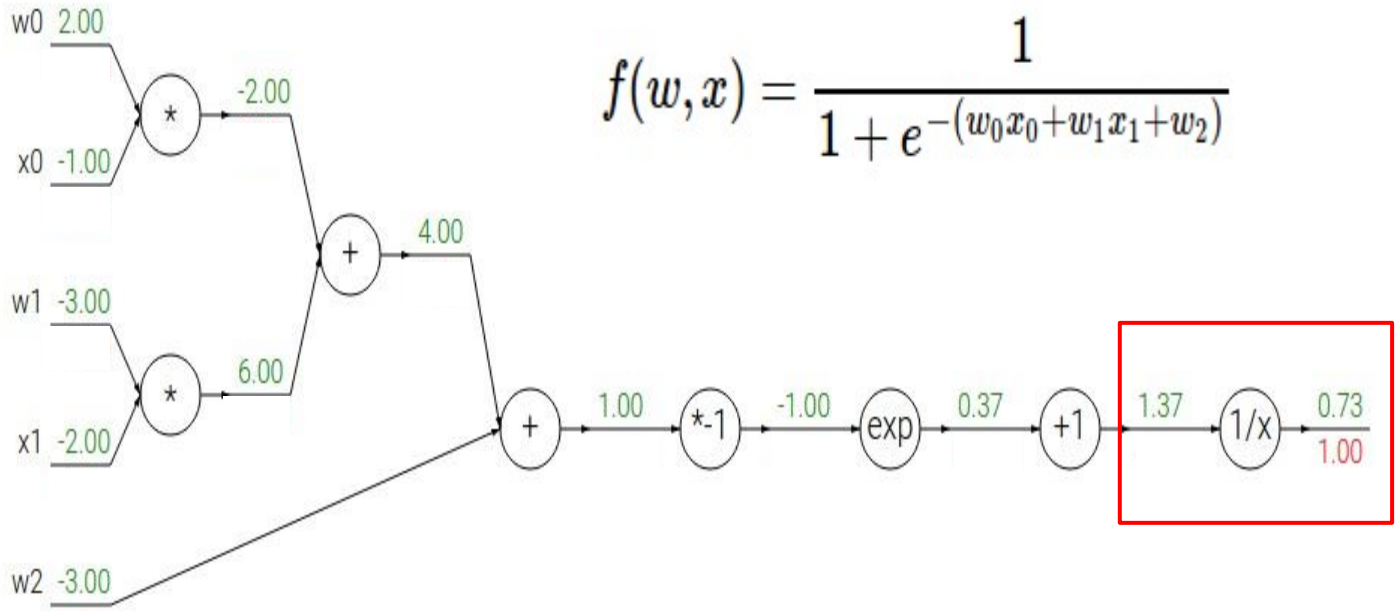
$$f(x) = e^x \quad \rightarrow \quad \frac{df}{dx} = e^x$$

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

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

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

Outro Exemplo – Passo Retrógrado



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

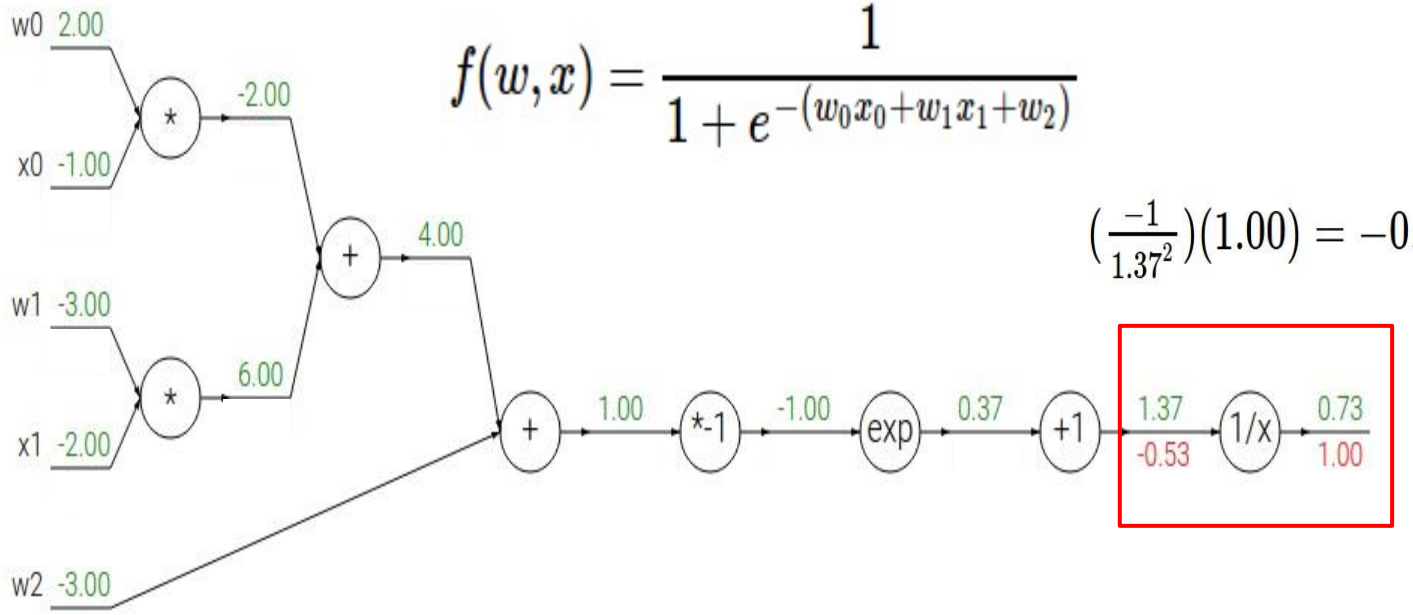
$$f(x) = e^x \quad \rightarrow \quad \frac{df}{dx} = e^x$$

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

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

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

Outro Exemplo – Passo Retrógrado



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

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

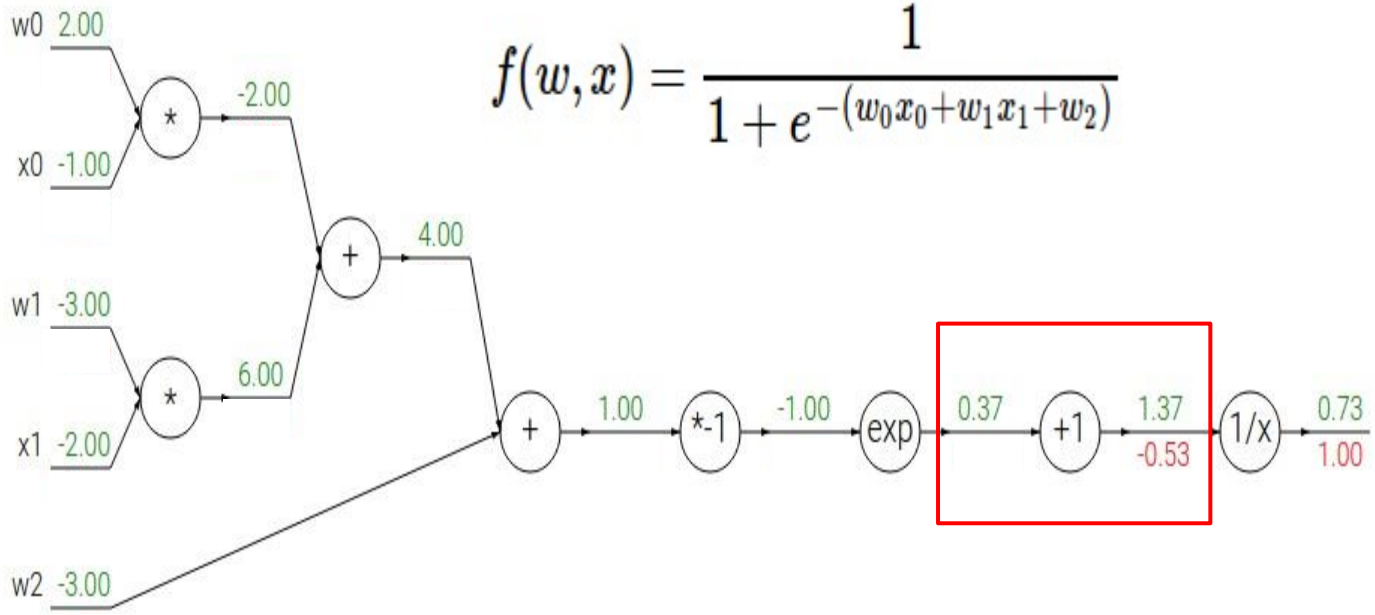
$$f(x) = e^x \quad \rightarrow \quad \frac{df}{dx} = e^x$$

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

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

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

Outro Exemplo – Passo Retrógrado



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

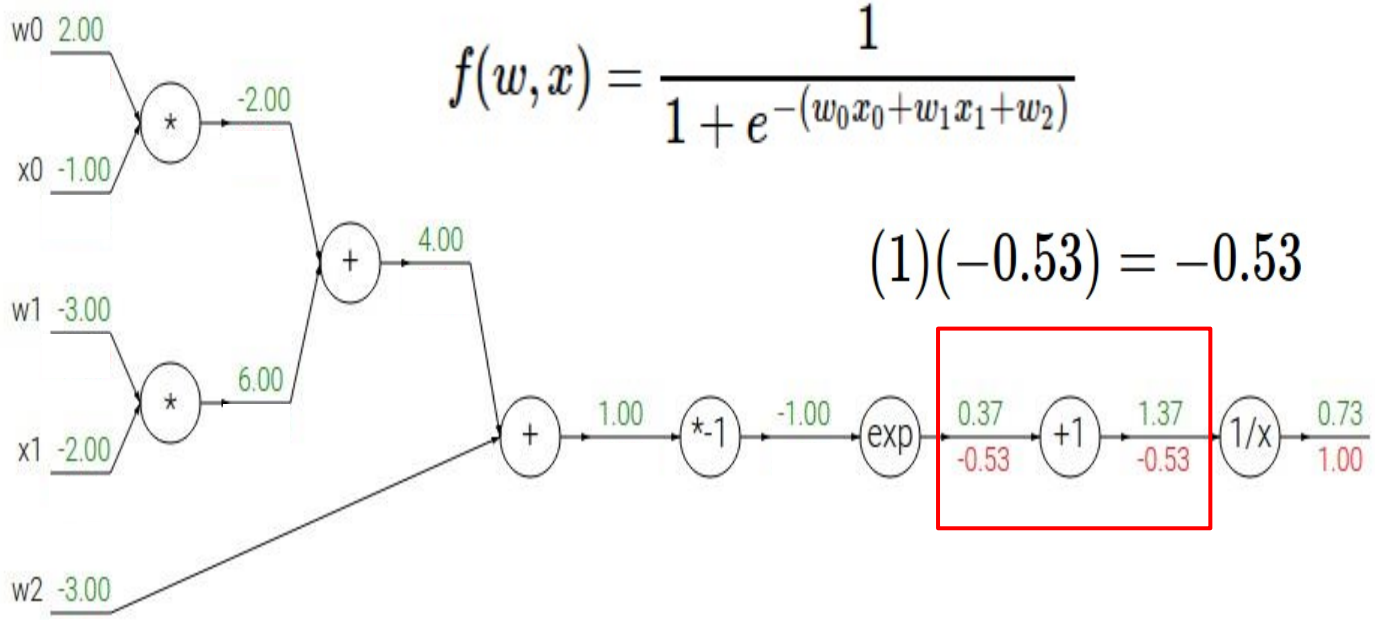
$$f(x) = e^x \quad \rightarrow \quad \frac{df}{dx} = e^x$$

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

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

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

Outro Exemplo – Passo Retrógrado

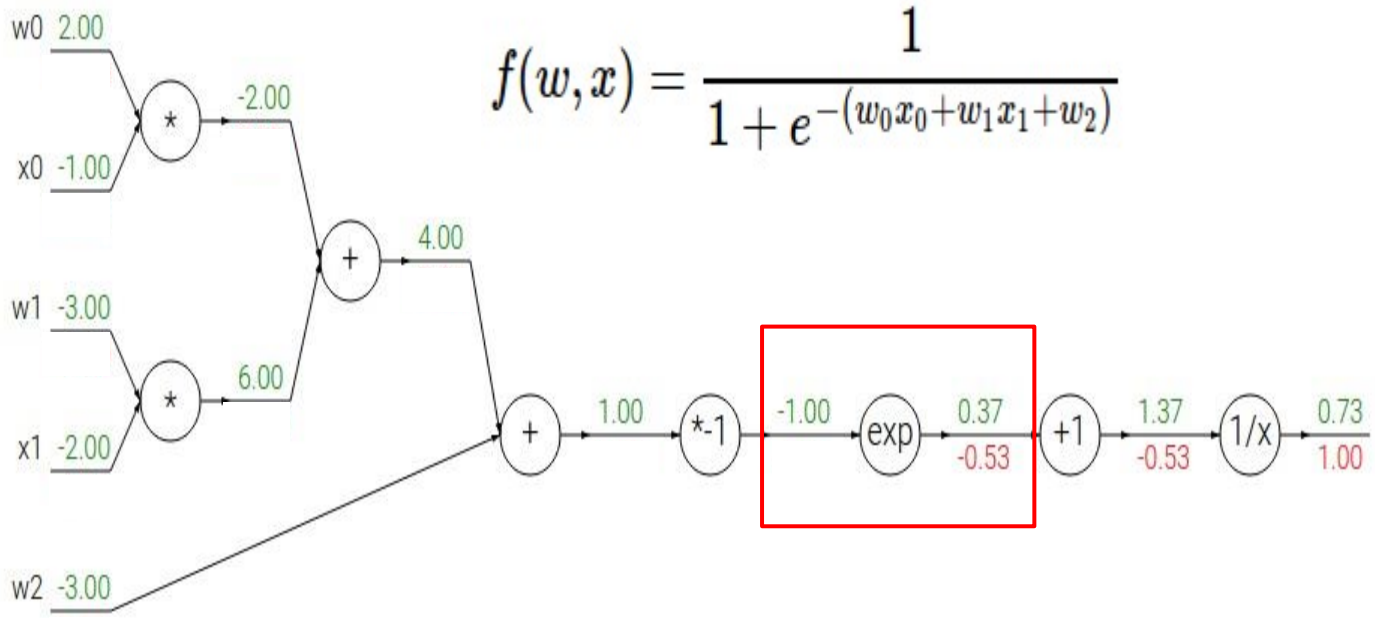


$$f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$

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

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

Outro Exemplo – Passo Retrógrado



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

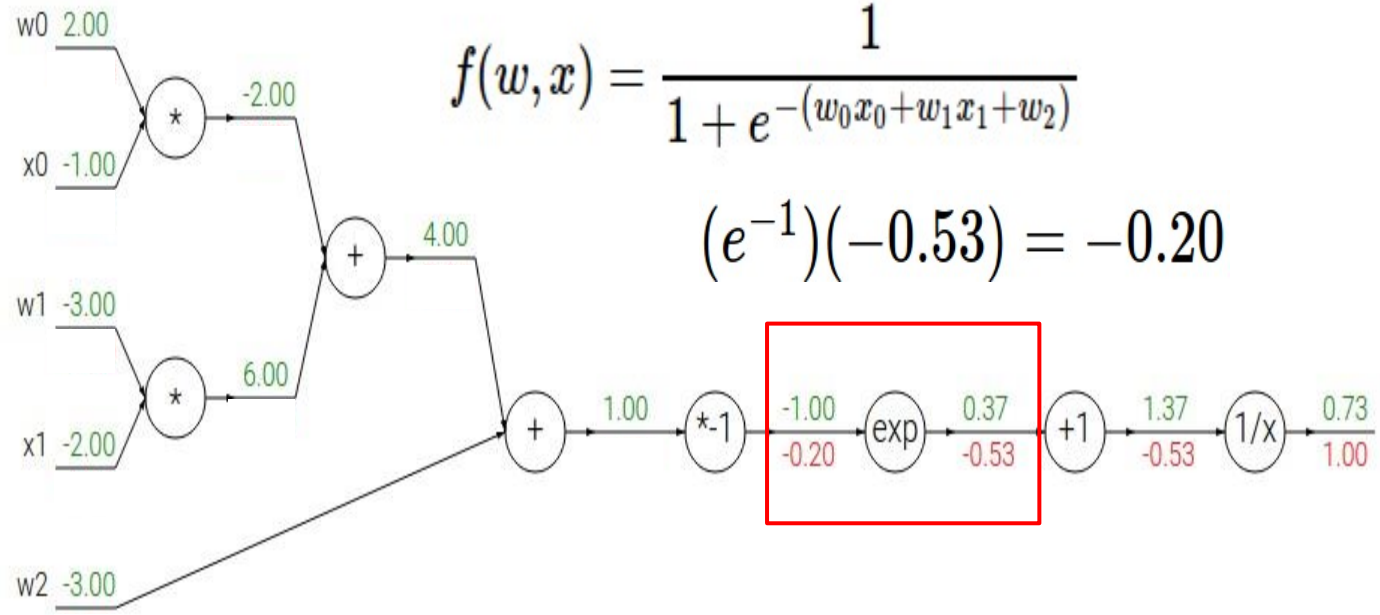
$$f(x) = e^x \quad \rightarrow \quad \frac{df}{dx} = e^x$$

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

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

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

Outro Exemplo – Passo Retrógrado



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

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

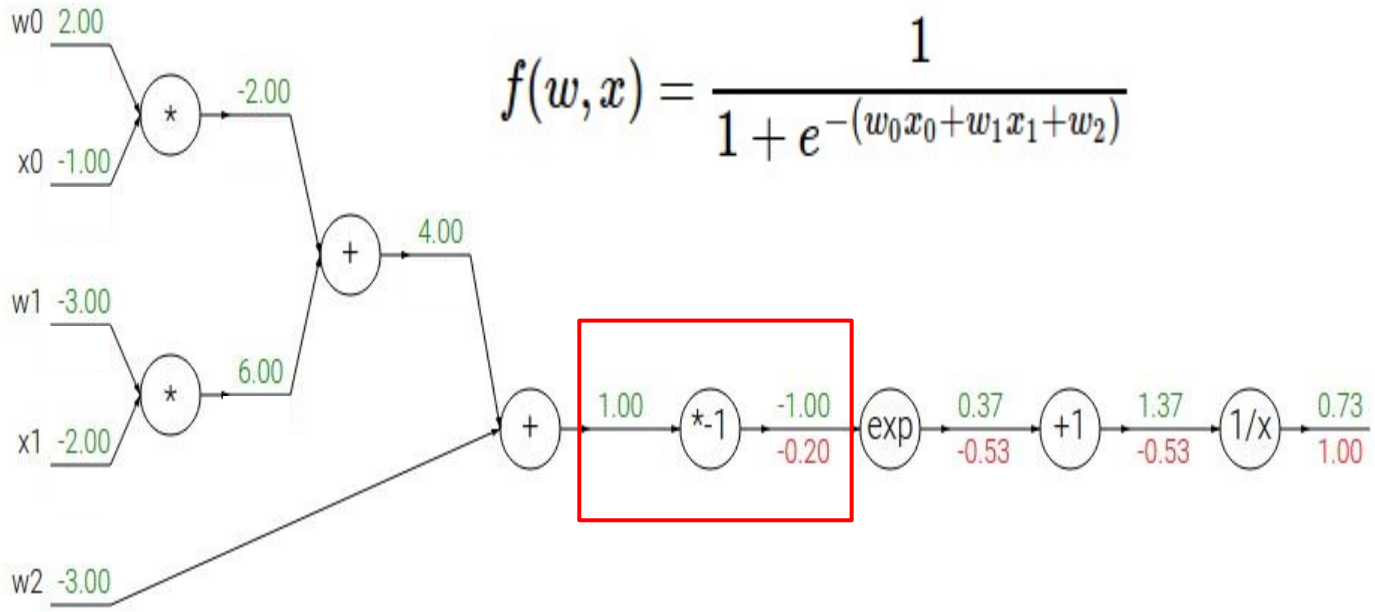
$$f(x) = e^x \quad \rightarrow \quad \frac{df}{dx} = e^x$$

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

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

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

Outro Exemplo – Passo Retrógrado



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

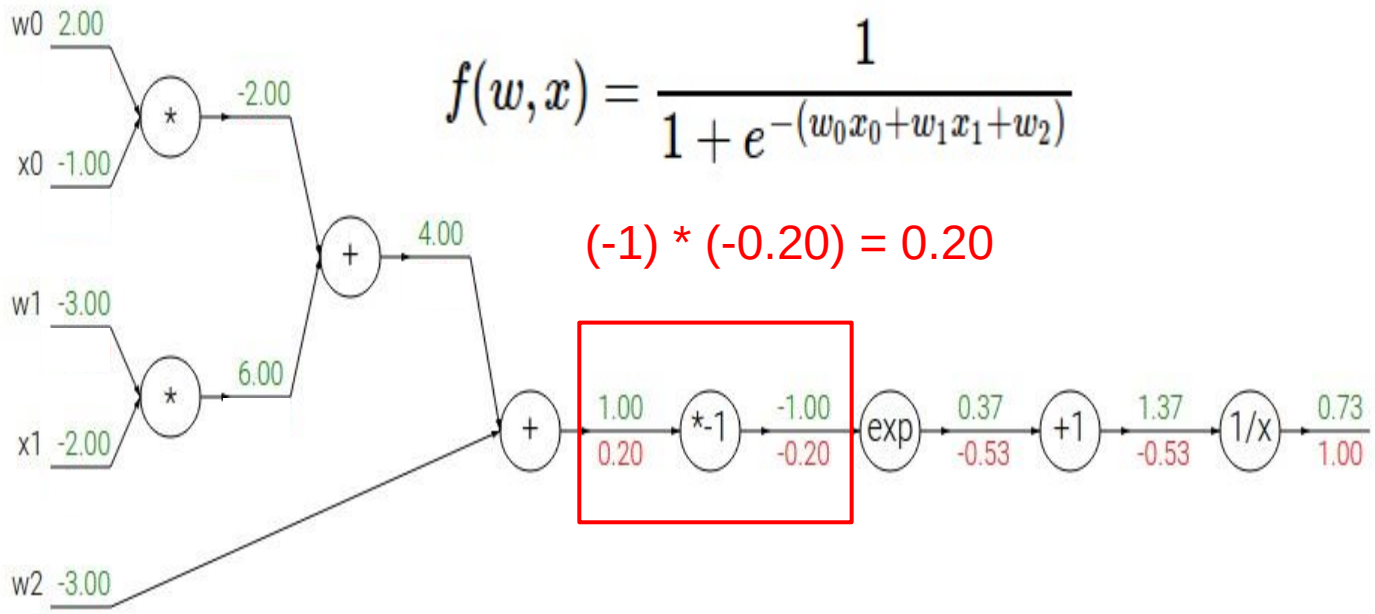
$$f(x) = e^x \quad \rightarrow \quad \frac{df}{dx} = e^x$$

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

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

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

Outro Exemplo – Passo Retrógrado



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

(-1) * (-0.20) = 0.20

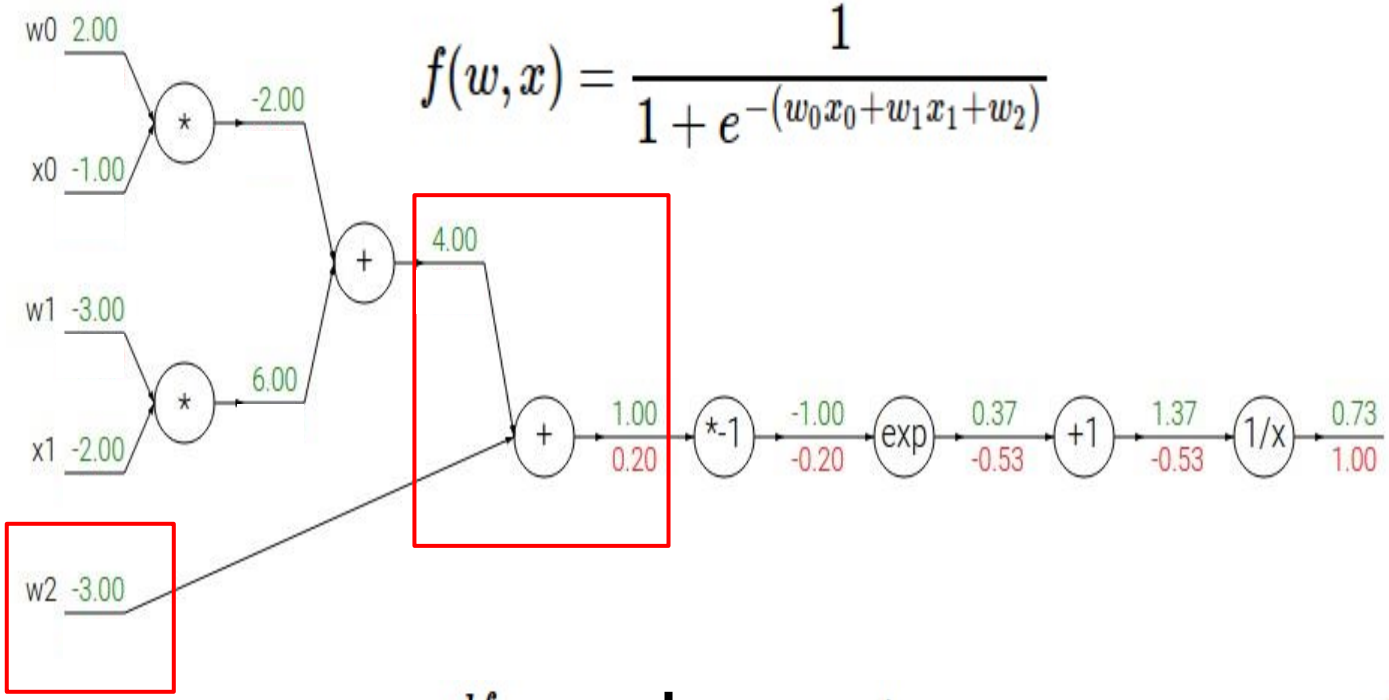
$$f(x) = e^x \quad \rightarrow \quad \frac{df}{dx} = e^x$$

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

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

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

Outro Exemplo – Passo Retrógrado



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

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

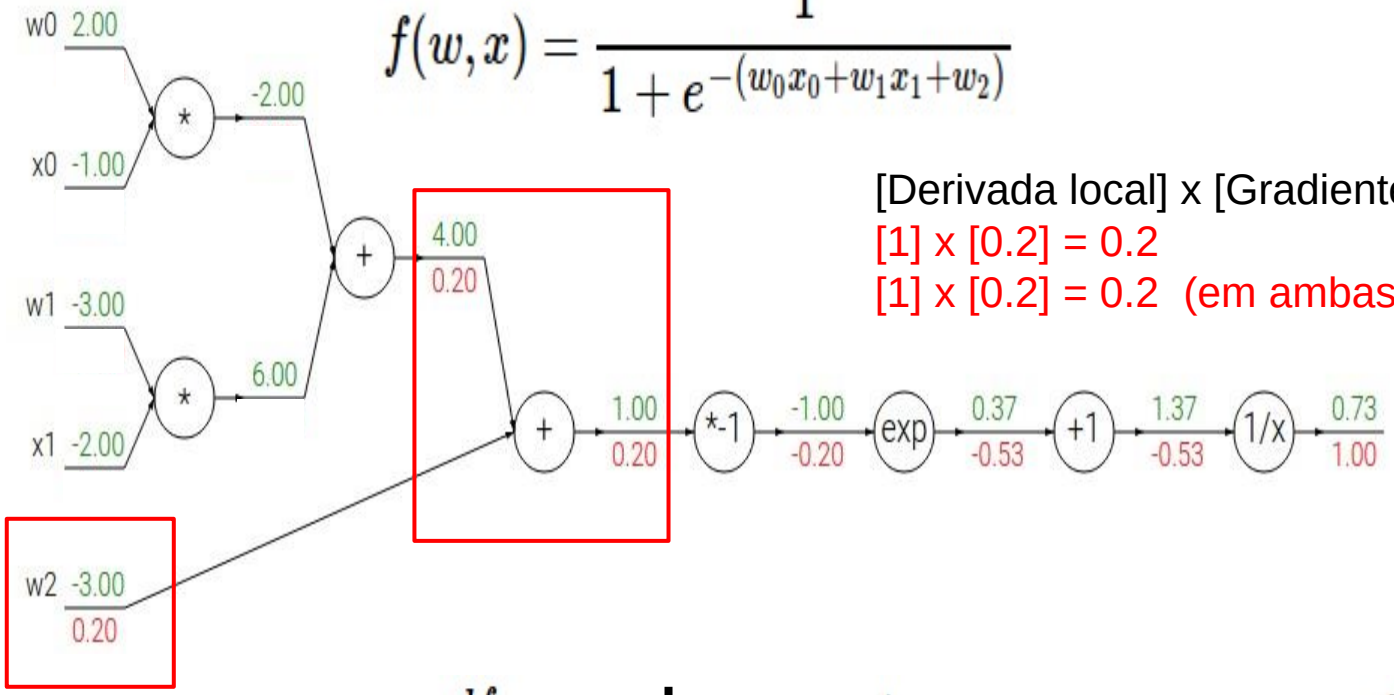
$$f_a(x) = ax \quad \rightarrow \quad \frac{df}{dx} = a$$

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

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

Outro Exemplo – Passo Retrógrado

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

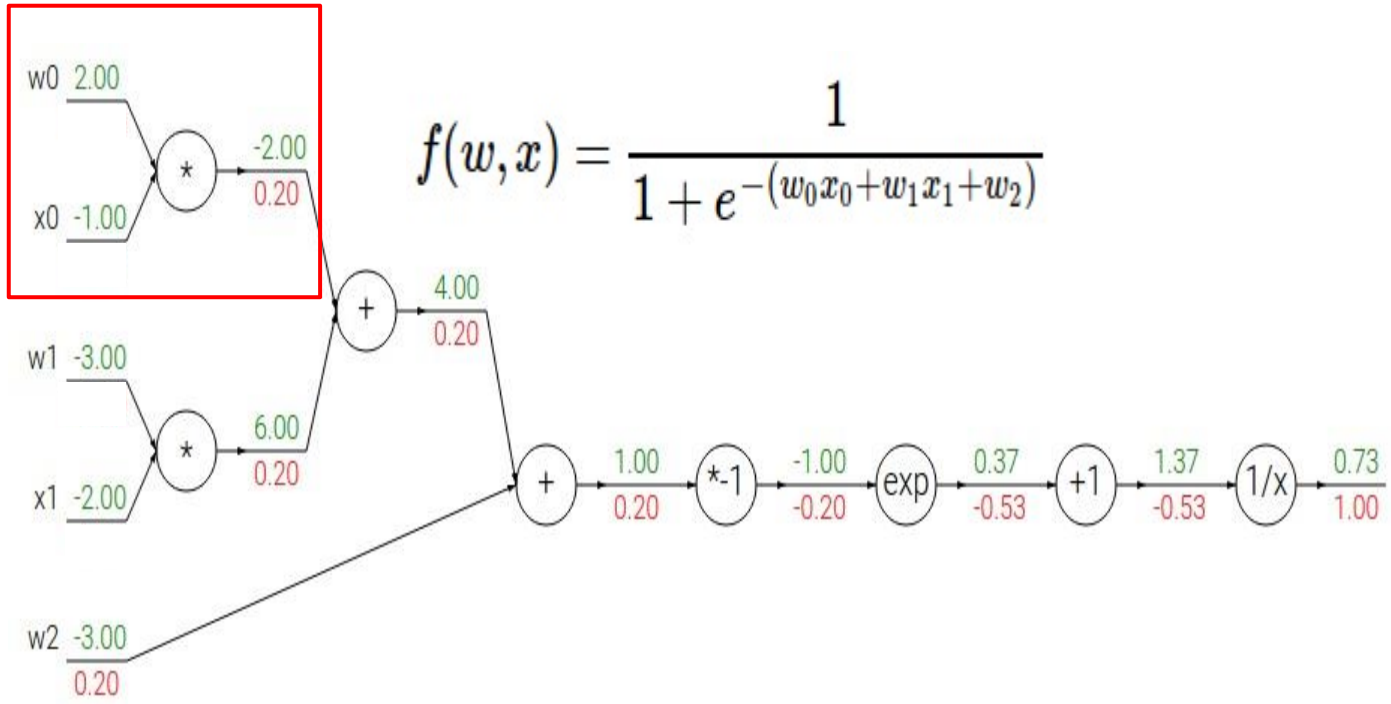


[Derivada local] x [Gradiente]
[1] x [0.2] = 0.2
[1] x [0.2] = 0.2 (em ambas entradas!)

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

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

Outro Exemplo – Passo Retrógrado



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

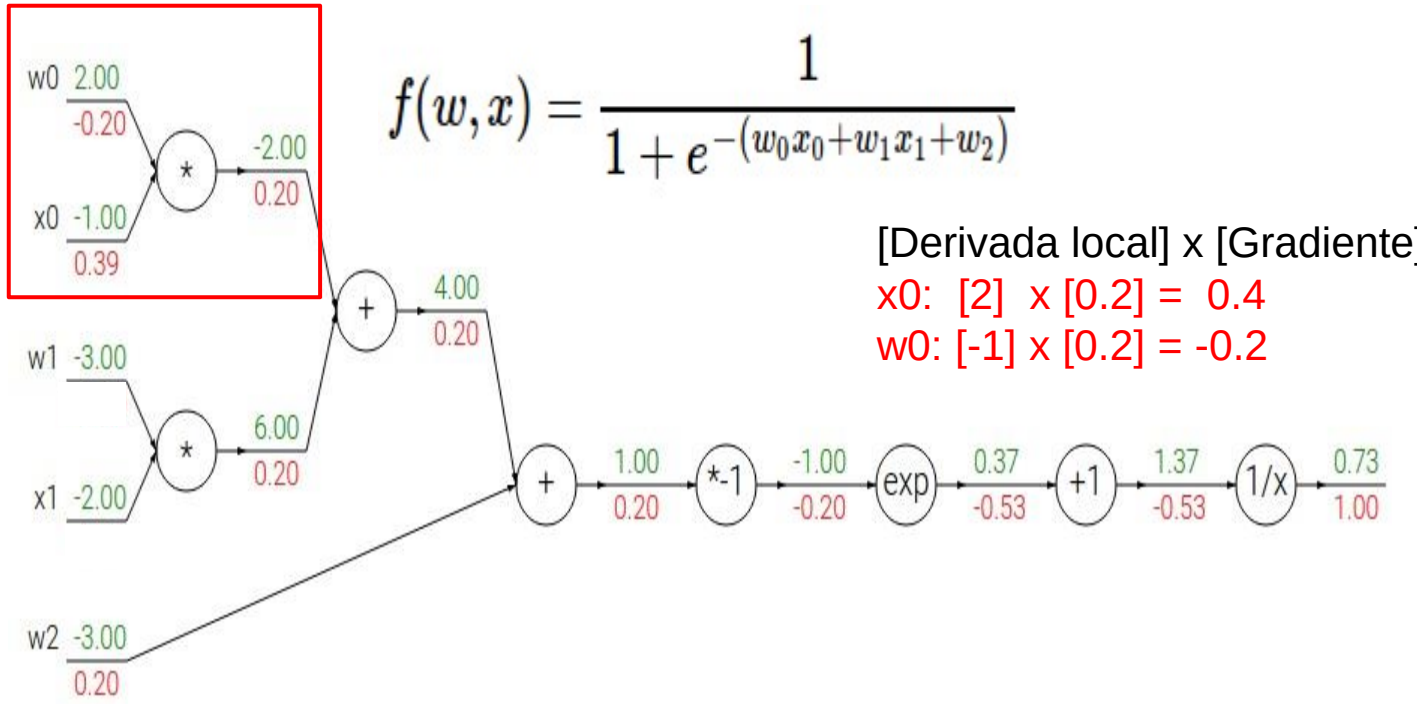
$$f(x) = e^x \quad \rightarrow \quad \frac{df}{dx} = e^x$$

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

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

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

Outro Exemplo – Passo Retrógrado



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

[Derivada local] x [Gradiente]

x0: [2] x [0.2] = 0.4

w0: [-1] x [0.2] = -0.2

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

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

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

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

Outro Exemplo – Passo Retrógrado

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

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

Função sigmoide

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

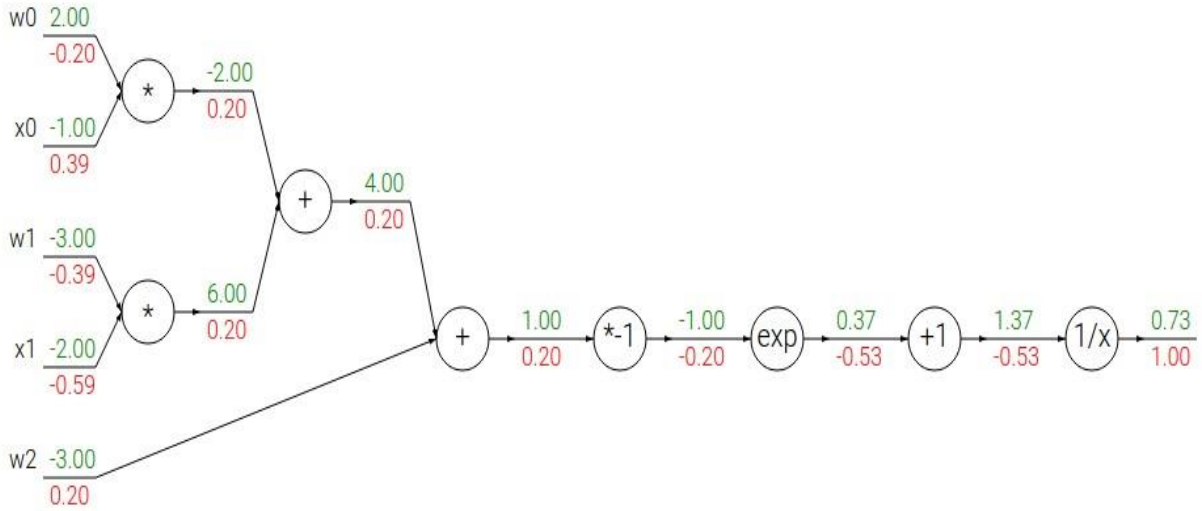
Outro Exemplo – Passo Retrógrado

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

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

Função sigmoide

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



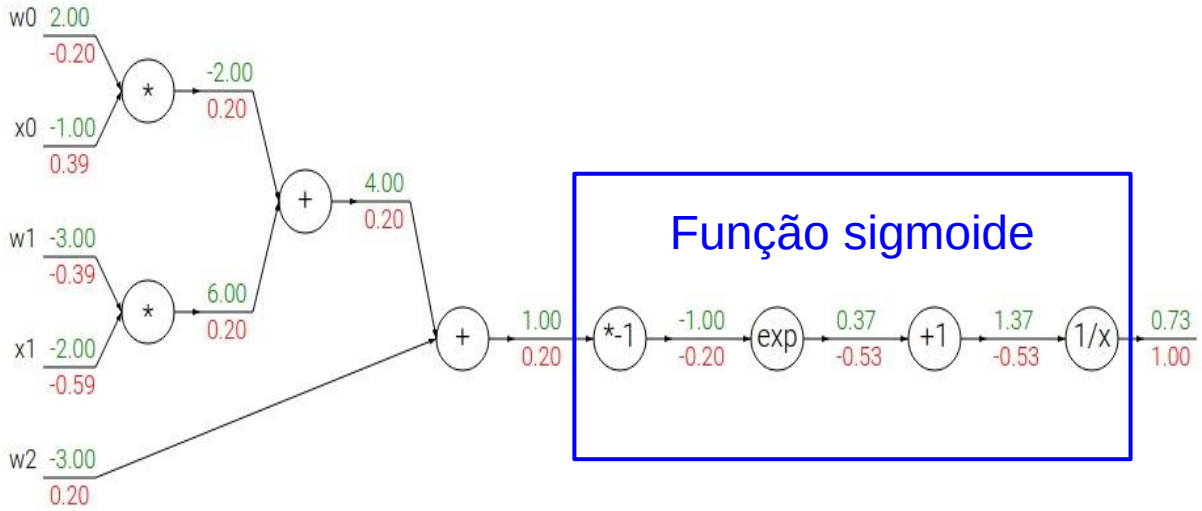
Outro Exemplo – Passo Retrógrado

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

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

Função sigmoide

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



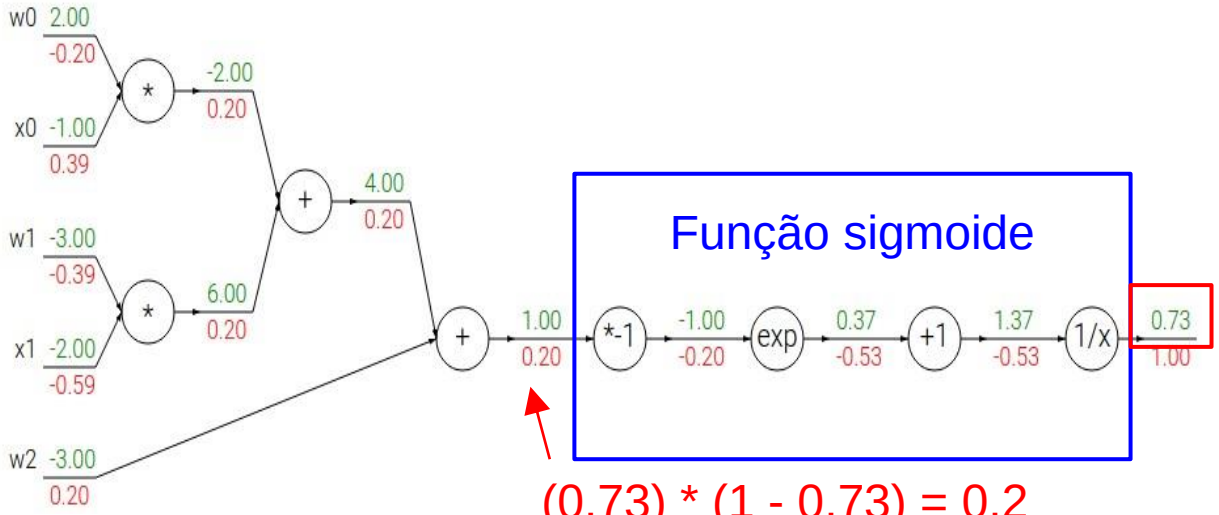
Outro Exemplo – Passo Retrógrado

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

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

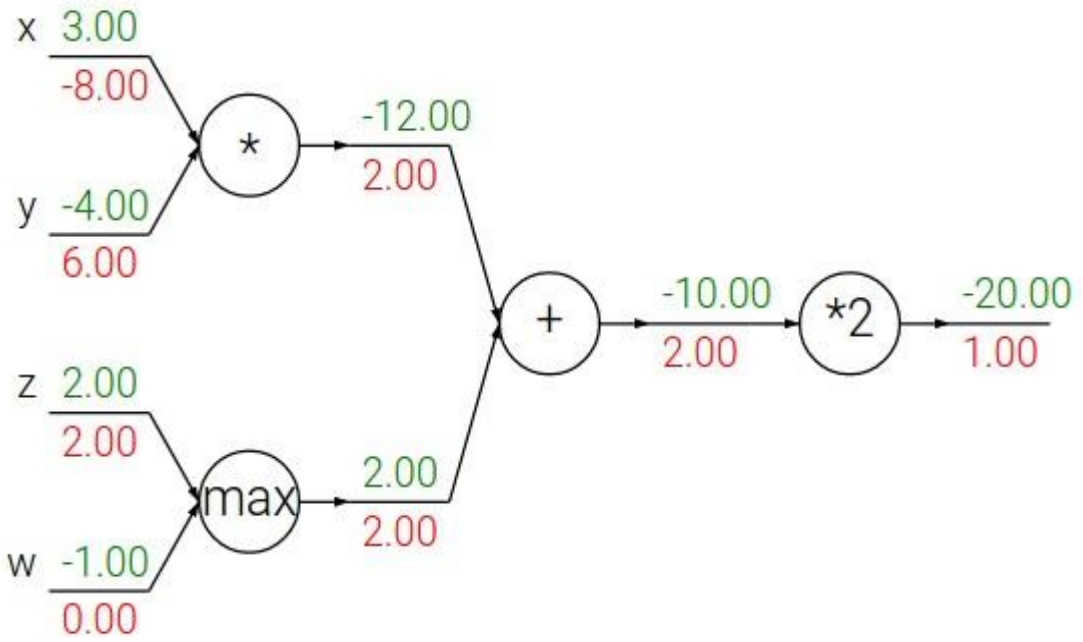
Função sigmoide

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



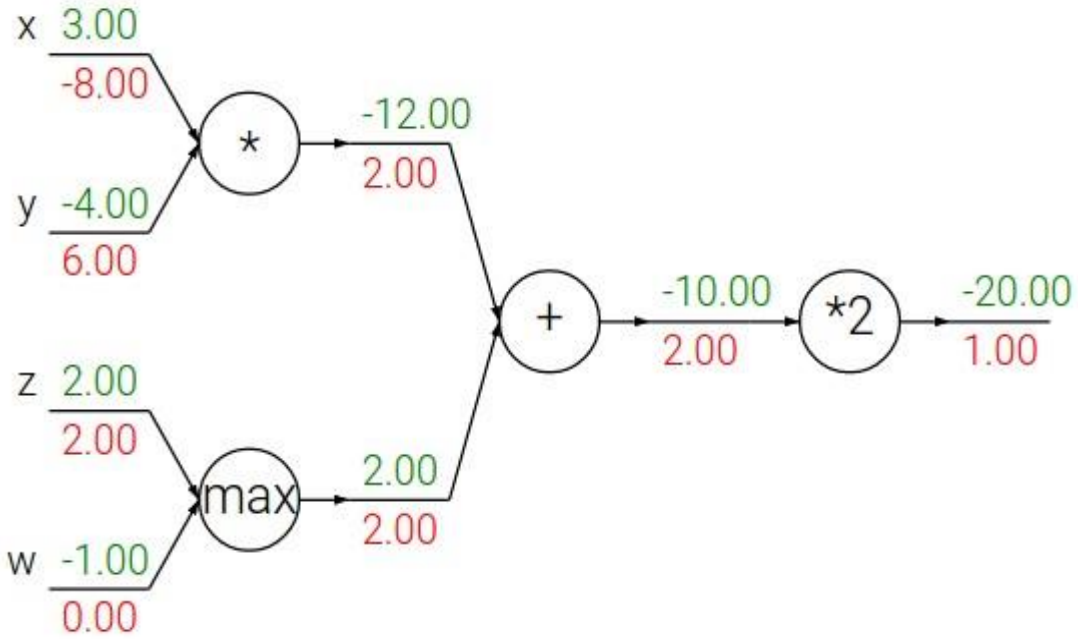
(0.73) * (1 - 0.73) = 0.2

Padrões no Fluxo Reverso



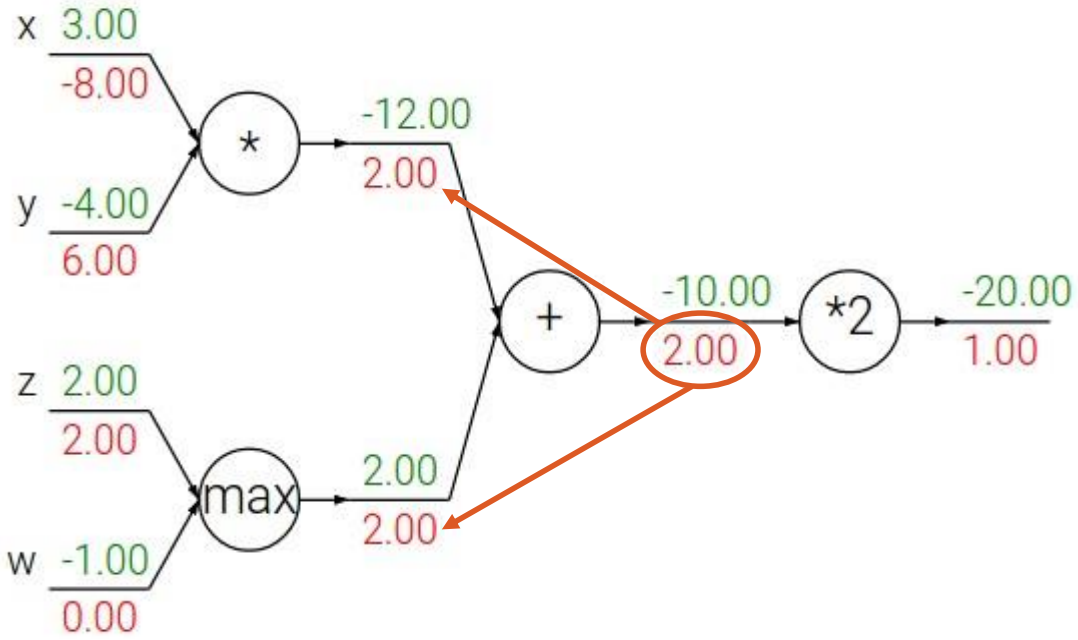
Padrões no Fluxo Reverso

Adição : distribuidor de gradiente



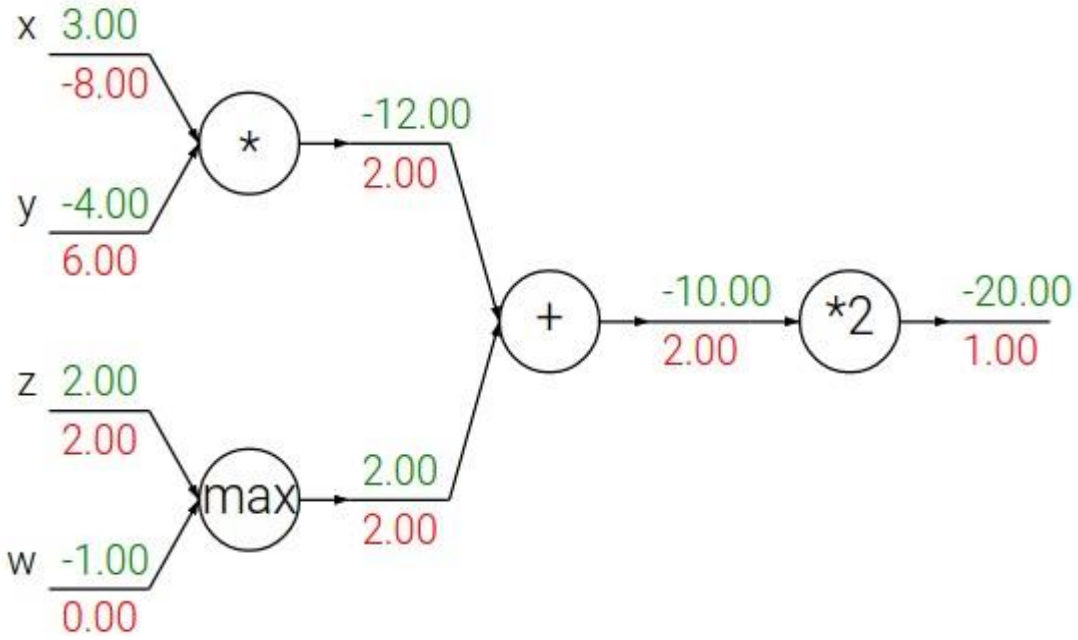
Padrões no Fluxo Reverso

Adição : distribuidor de gradiente



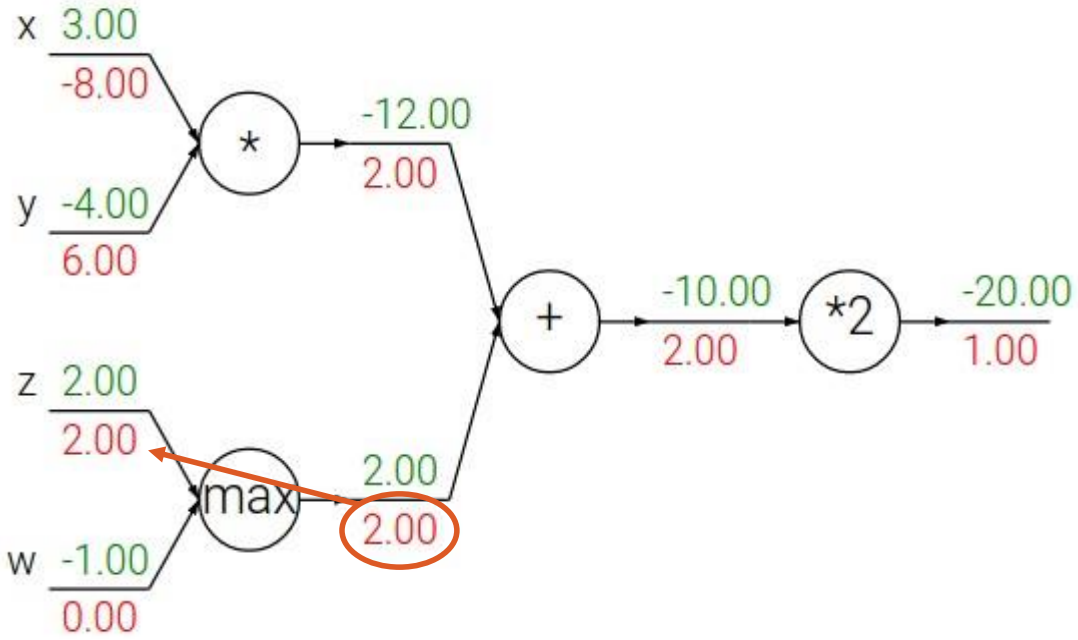
Padrões no Fluxo Reverso

Adição : distribuidor de gradiente
Max : direcionador de gradiente



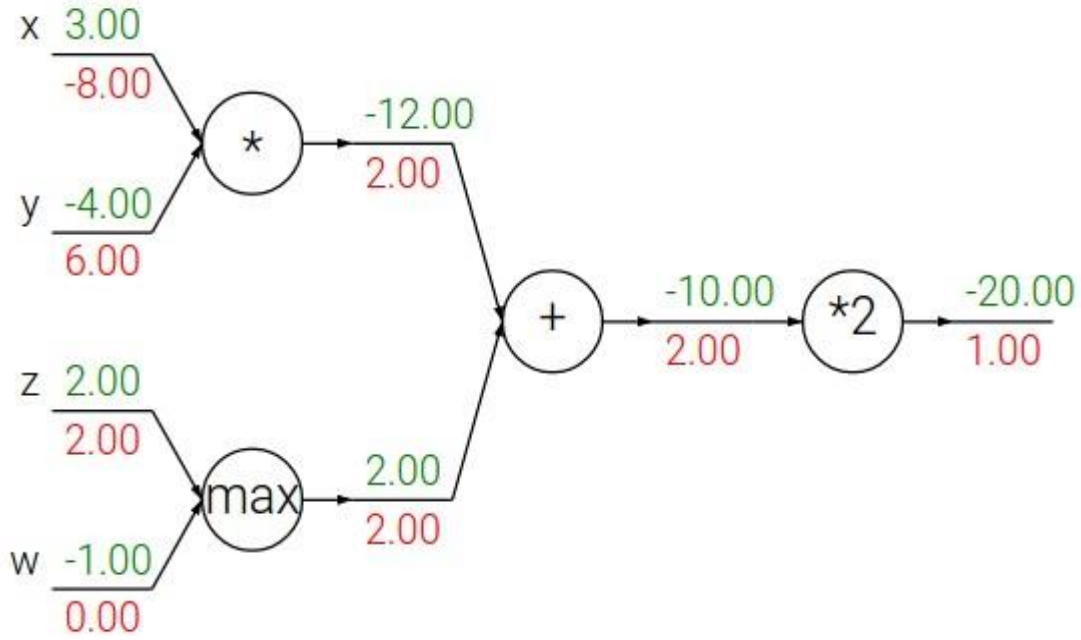
Padrões no Fluxo Reverso

Adição : distribuidor de gradiente
Max : direcionador de gradiente



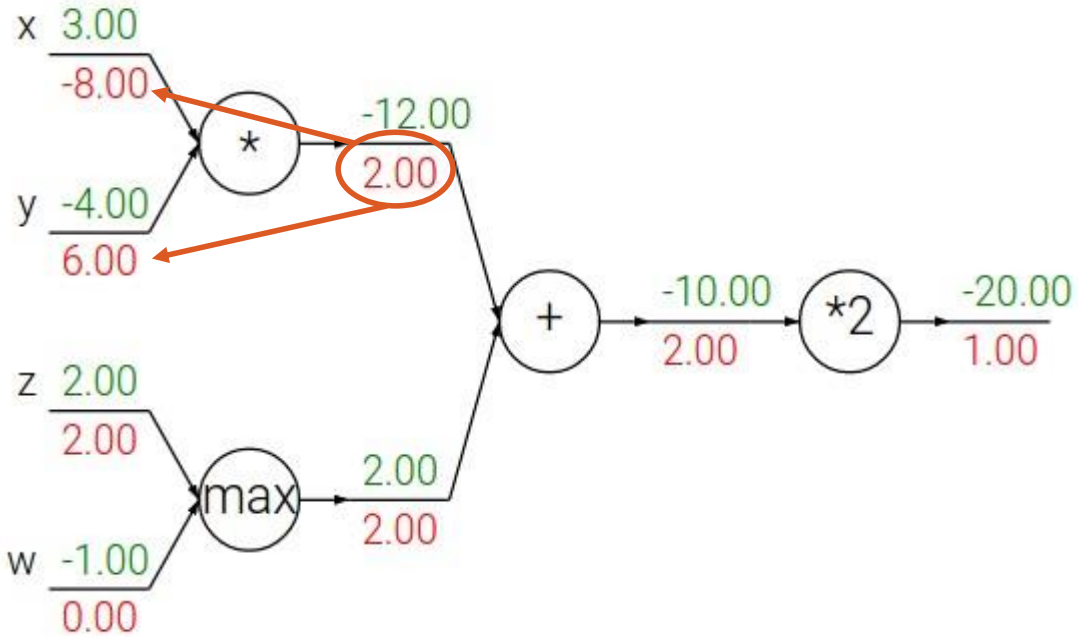
Padrões no Fluxo Reverso

Adição : distribuidor de gradiente
Max : direcionador de gradiente
Produto: “comutador” de gradiente



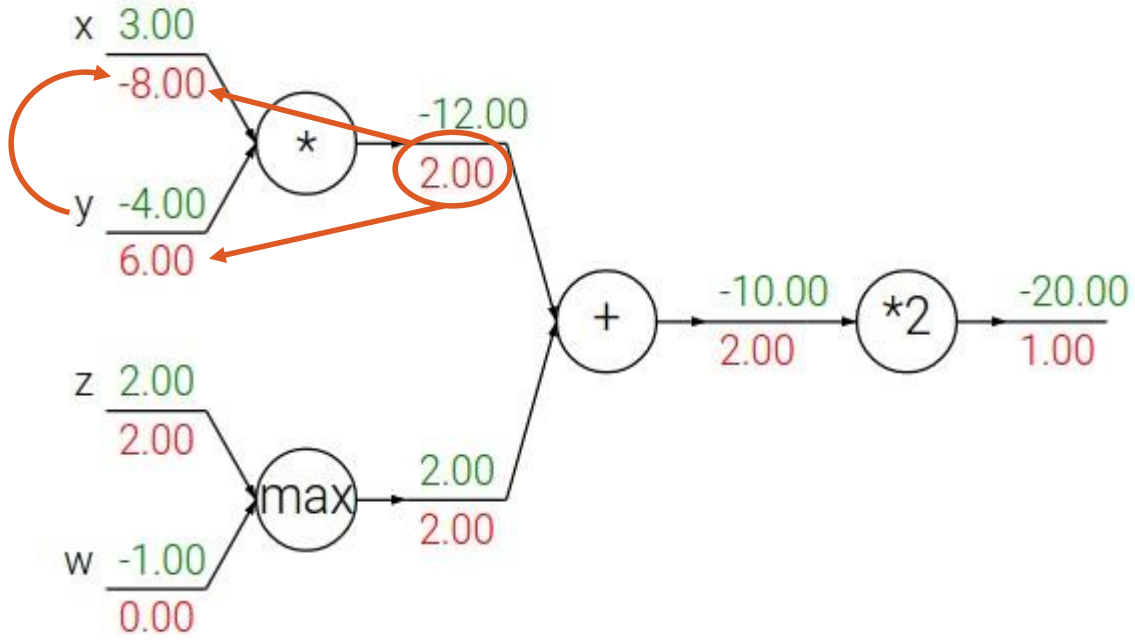
Padrões no Fluxo Reverso

Adição : distribuidor de gradiente
Max : direcionador de gradiente
Produto: “comutador” de gradiente



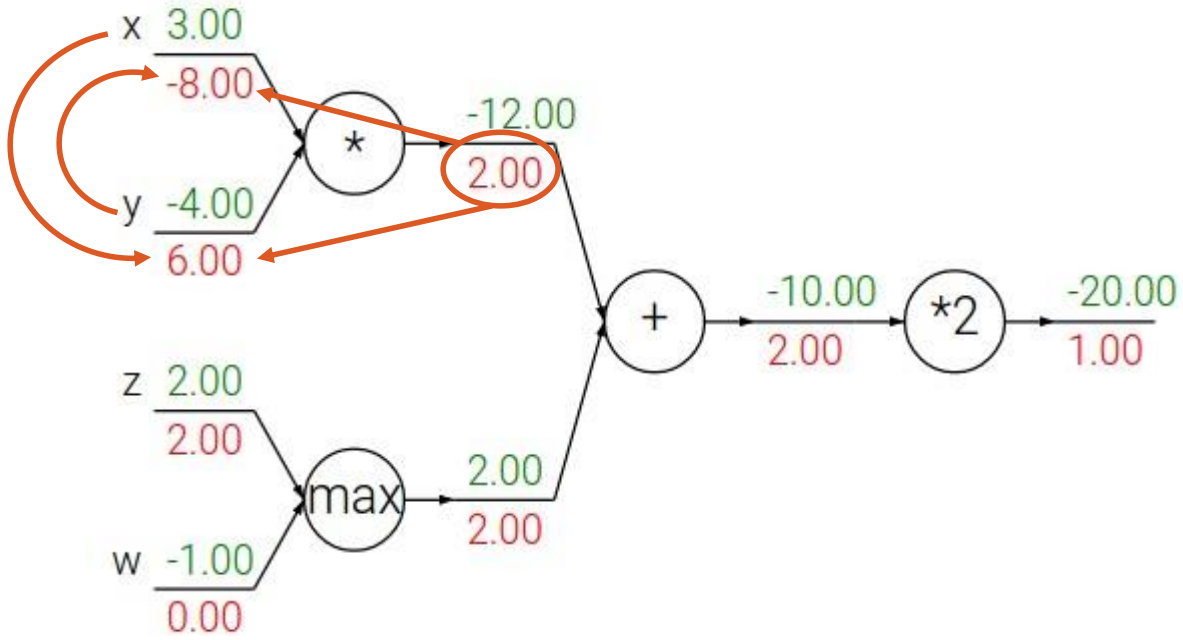
Padrões no Fluxo Reverso

Adição : distribuidor de gradiente
Max : direcionador de gradiente
Produto: “comutador” de gradiente

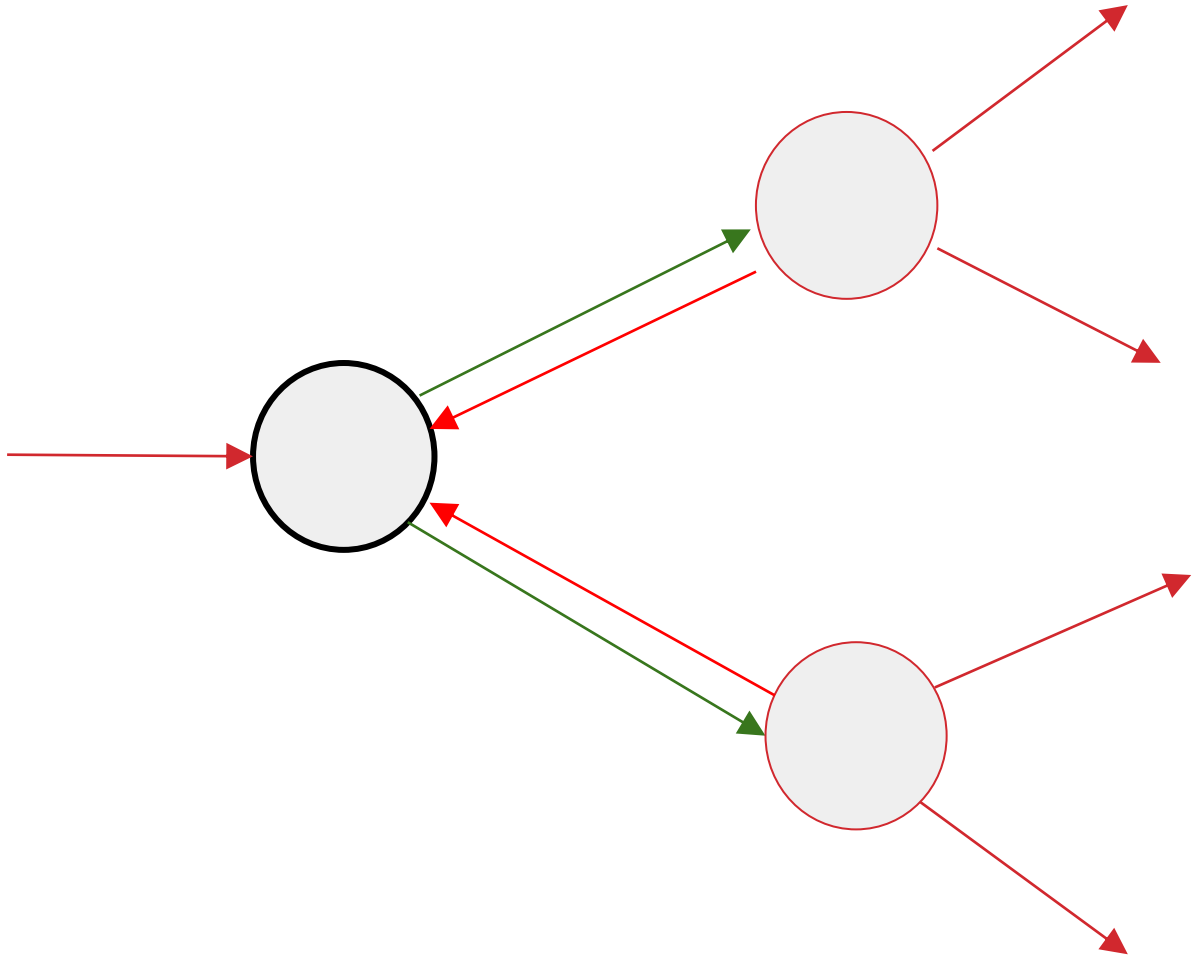


Padrões no Fluxo Reverso

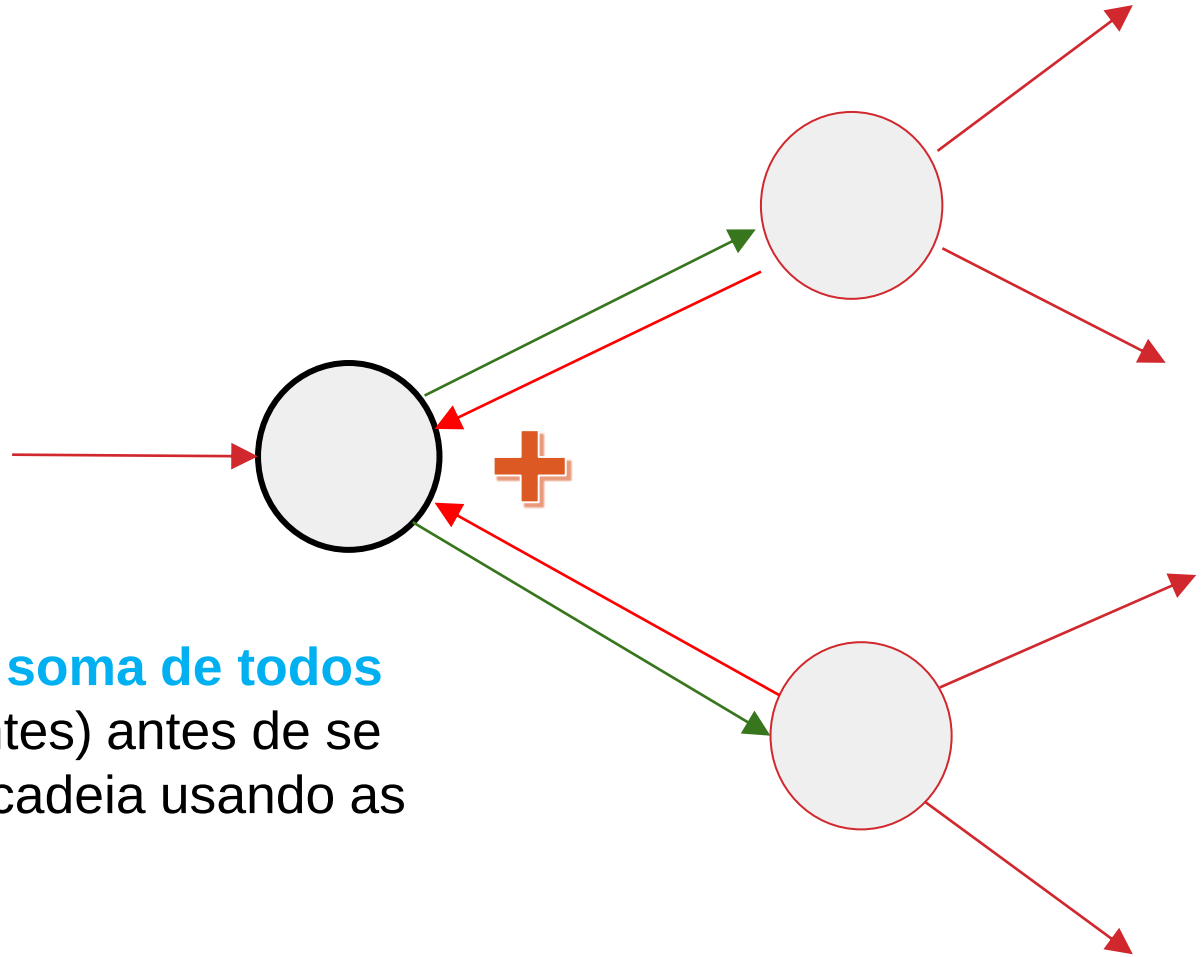
Adição : distribuidor de gradiente
Max : direcionador de gradiente
Produto: “comutador” de gradiente



Gradientes nas Ramificações

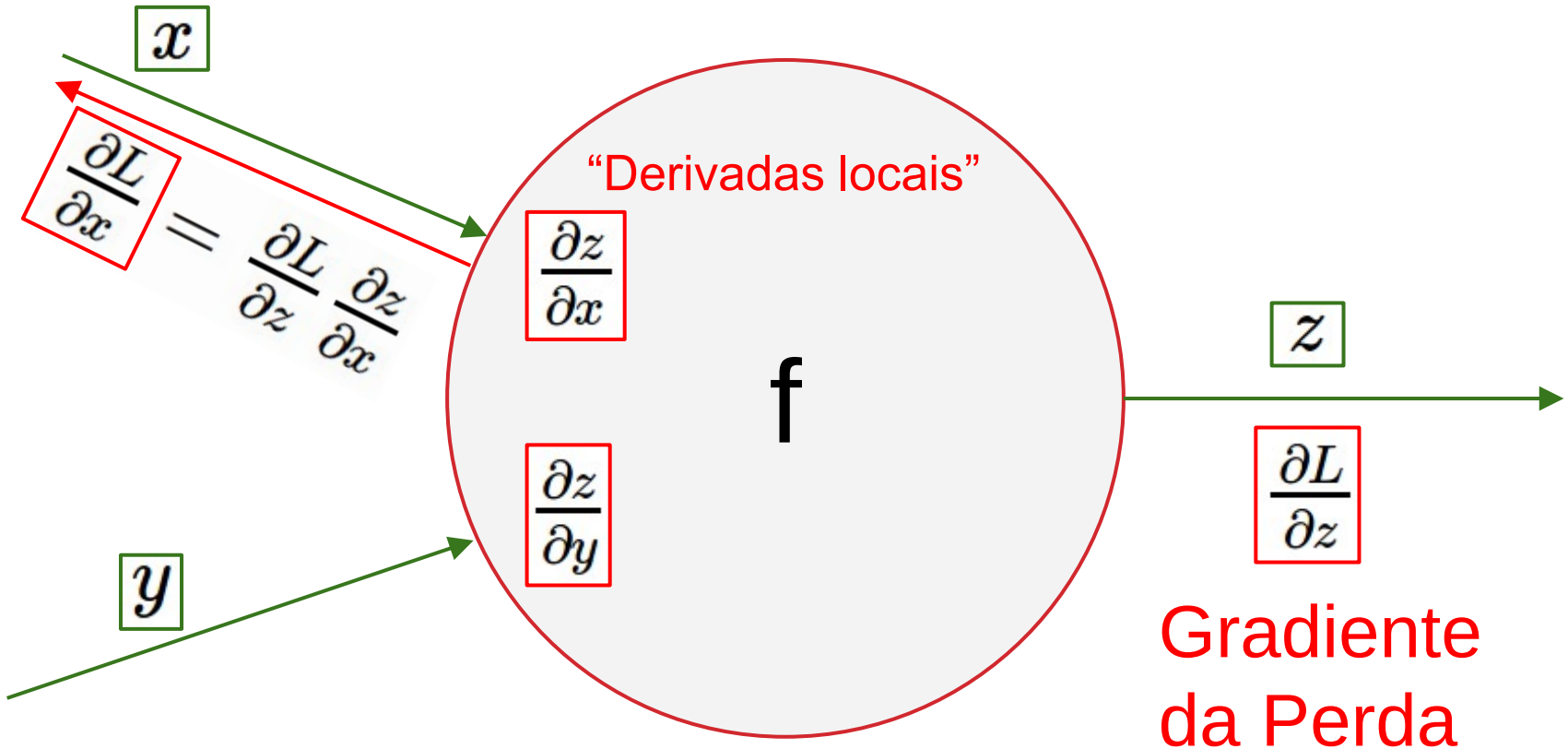


Gradientes nas Ramificações

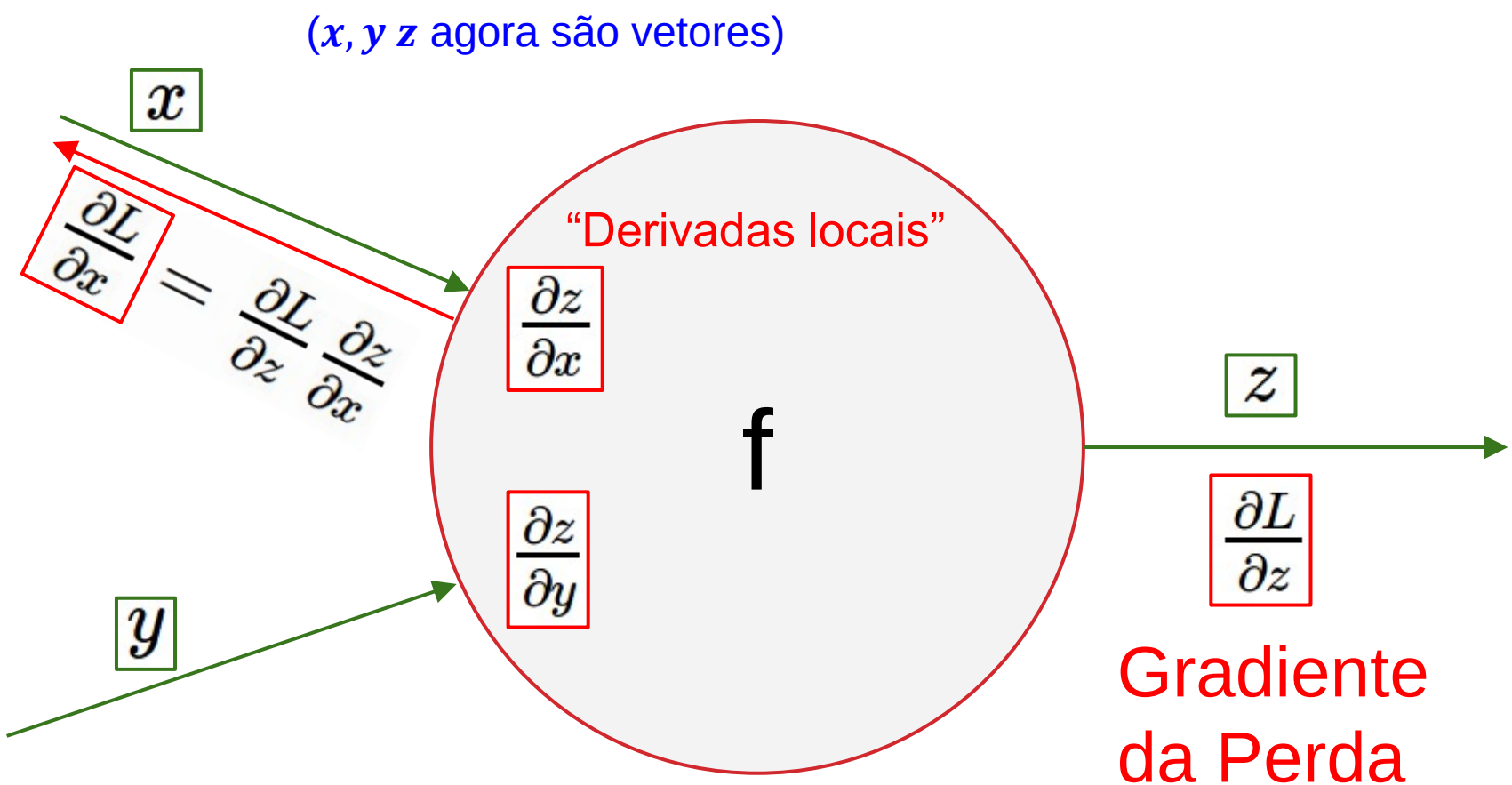


Deve-se realizar a **soma de todos os fluxos** (gradientes) antes de se aplicar a regra da cadeia usando as derivadas locais

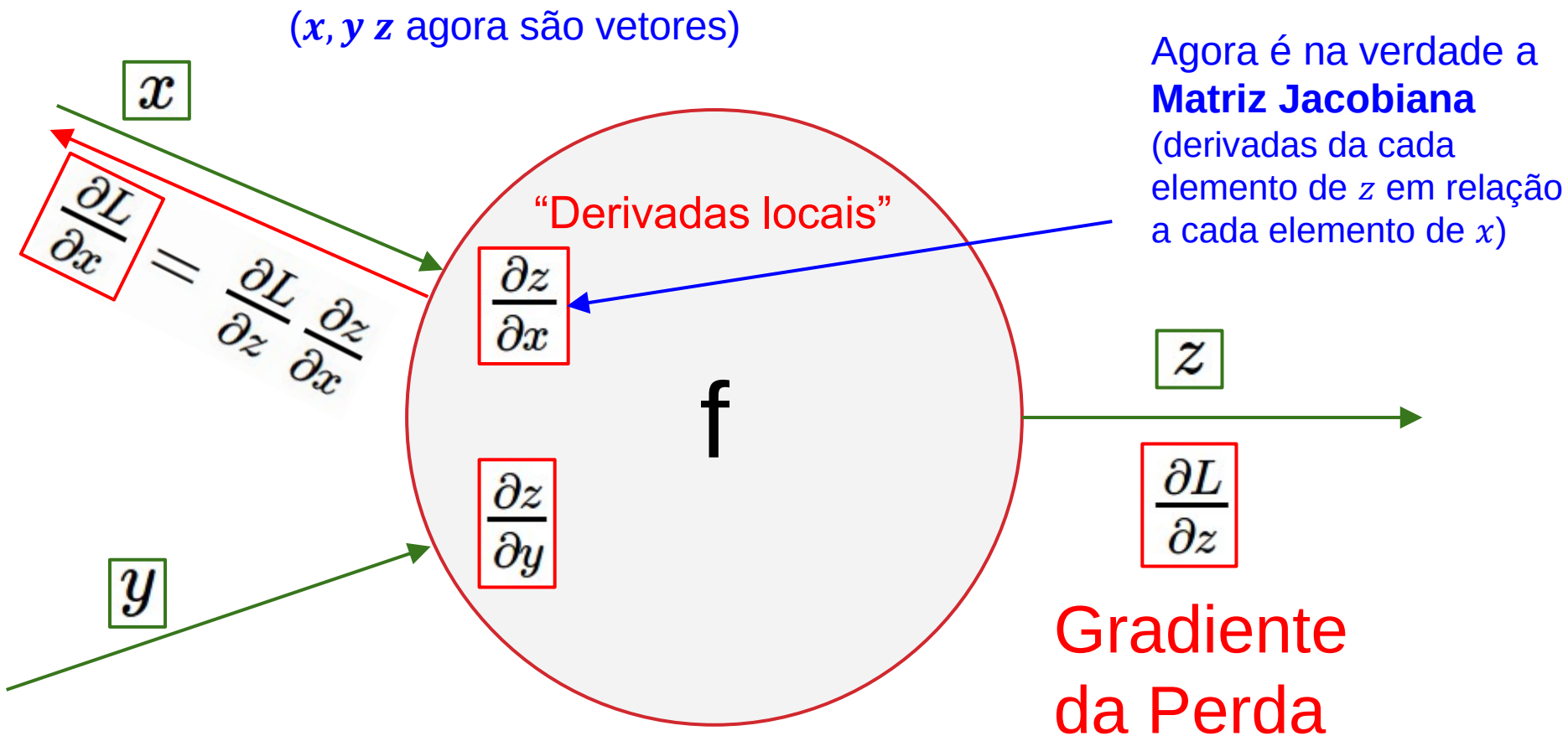
Gradientes para Dados Multidimensionais



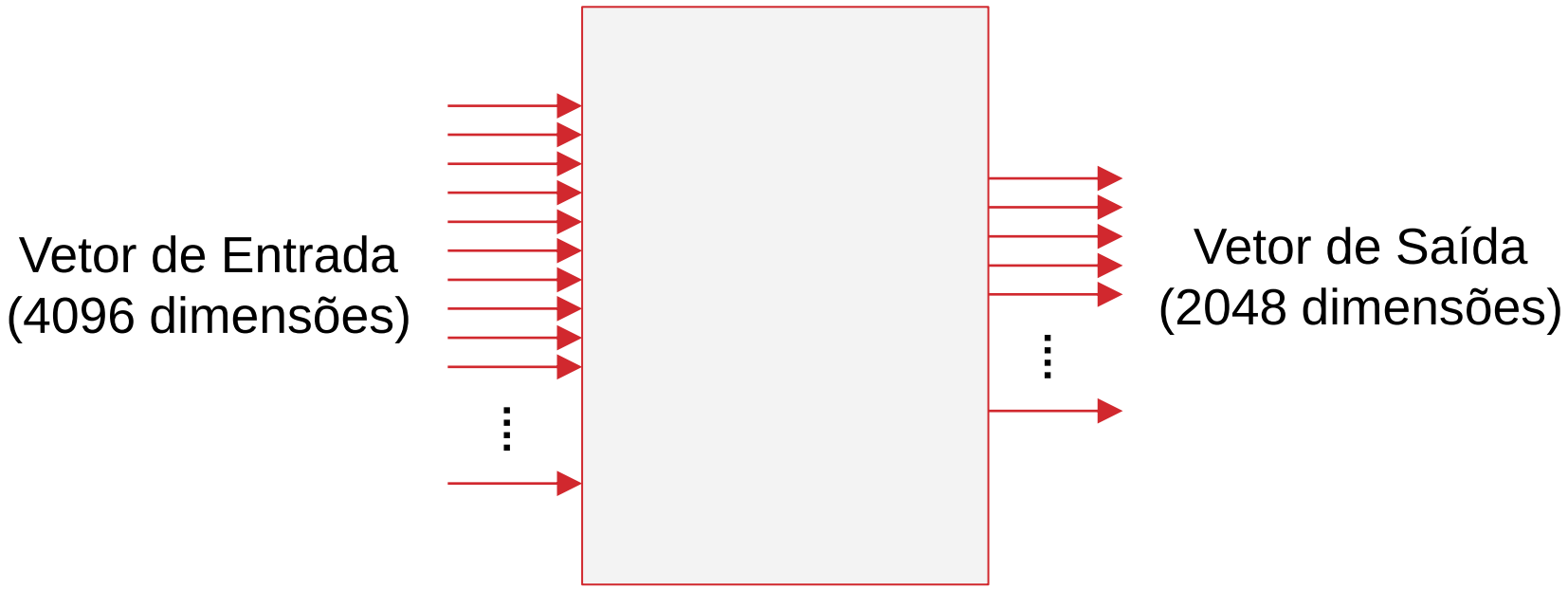
Gradientes para Dados Multidimensionais



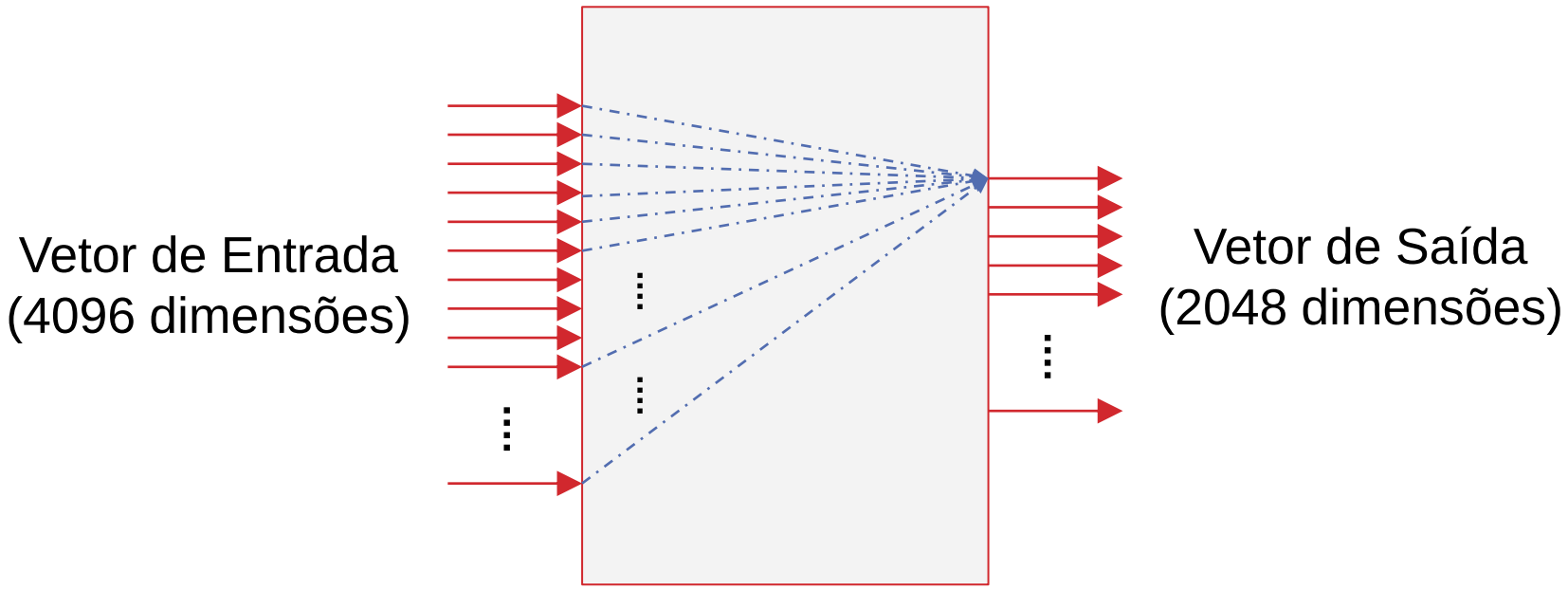
Gradientes para Dados Multidimensionais



Gradientes para Dados Multidimensionais



Gradientes para Dados Multidimensionais

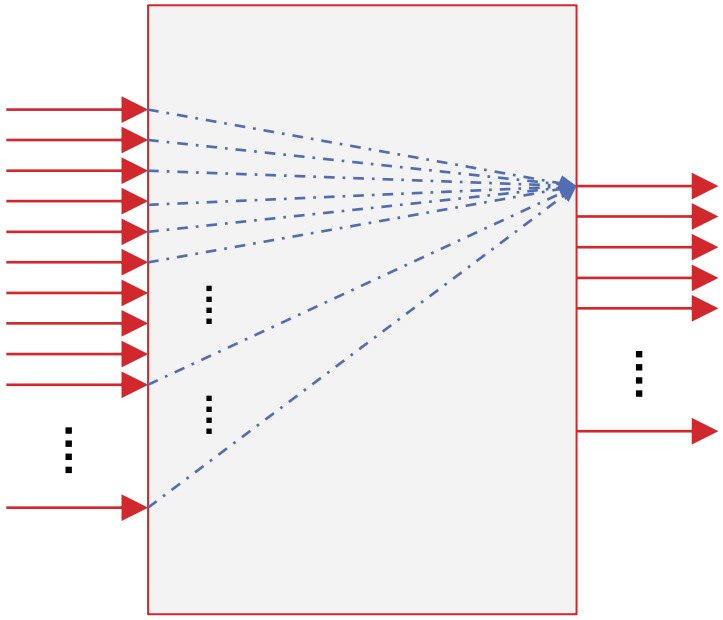


Gradientes para Dados Multidimensionais

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

Matriz Jacobiana

Vetor de Entrada
(4096 dimensões)



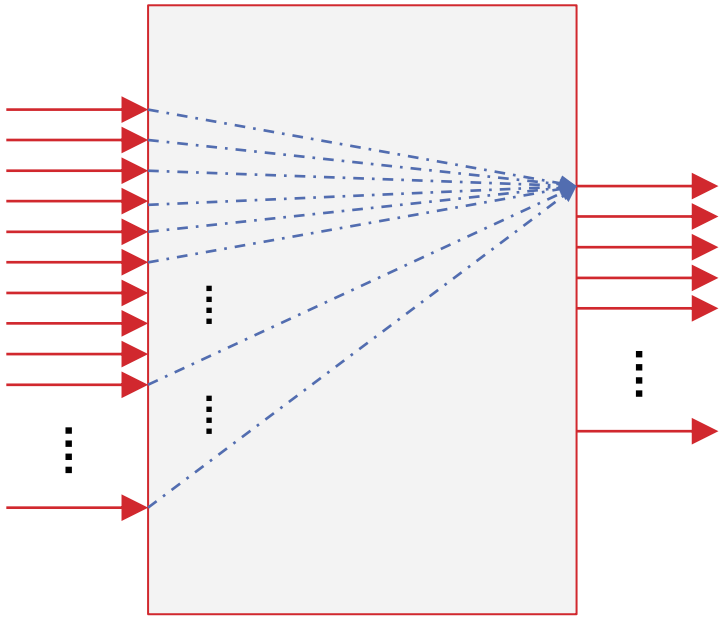
Vetor de Saída
(2048 dimensões)

Gradientes para Dados Multidimensionais

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

Matriz Jacobiana

Vetor de Entrada
(4096 dimensões)



Vetor de Saída
(2048 dimensões)

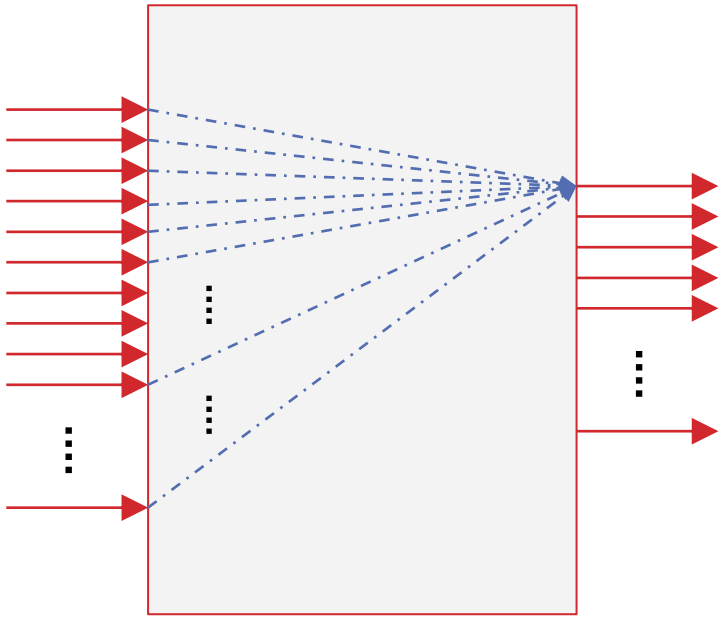
P1: qual o tamanho
da matriz Jacobiana?

Gradientes para Dados Multidimensionais

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

Matriz Jacobiana

Vetor de Entrada
(4096 dimensões)



Vetor de Saída
(2048 dimensões)

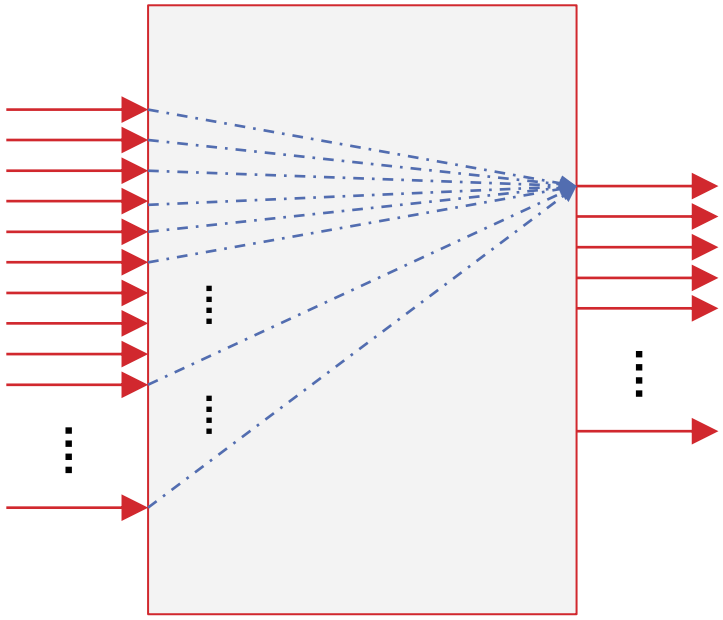
P1: qual o tamanho
da matriz Jacobiana?
$\Rightarrow$ 2048 × 4096 !

Gradientes para Dados Multidimensionais

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

Matriz Jacobiana

Vetor de Entrada
(4096 dimensões)



Vetor de Saída
(2048 dimensões)

P1: qual o tamanho
da matriz Jacobiana?
$\Rightarrow$ 2048 × 4096 !

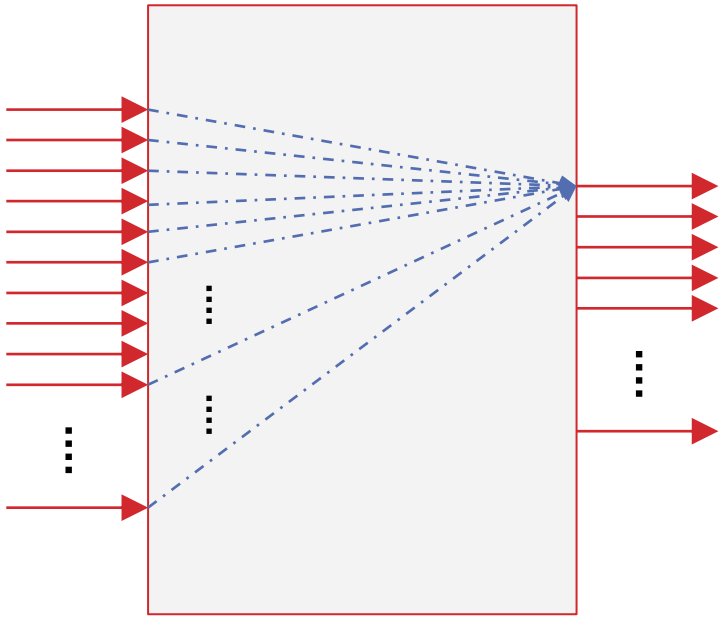
P2: qual a aparência
dessa matriz?

Gradientes para Dados Multidimensionais

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

Matriz Jacobiana

Vetor de Entrada
(4096 dimensões)



Vetor de Saída
(2048 dimensões)

P1: qual o tamanho
da matriz Jacobiana?
$\Rightarrow$ 2048 × 4096 !

P2: qual a aparência
dessa matriz?

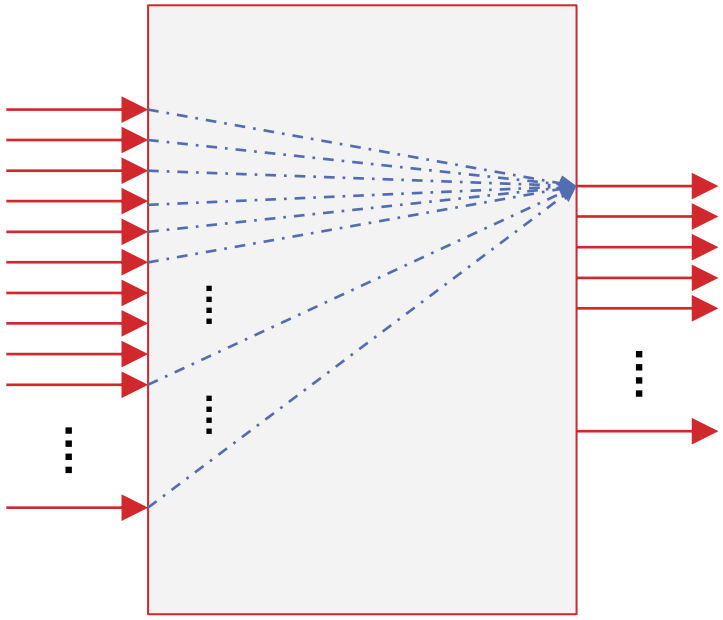
$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_m} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_k}{\partial x_1} & \frac{\partial f_k}{\partial x_2} & \cdots & \frac{\partial f_k}{\partial x_m} \end{bmatrix}$$

Gradientes para Dados Multidimensionais

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

Matriz Jacobiana

Vetor de Entrada
(4096 dimensões)



Vetor de Saída
(2048 dimensões)

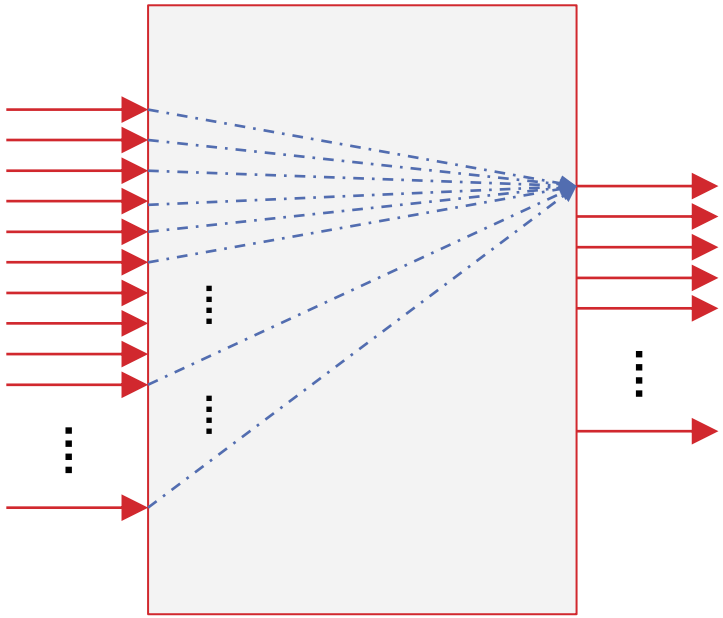
Na prática, processa-se todo um lote ou "*minibatch*" (p.ex. 100 amostras) de uma só vez

Gradientes para Dados Multidimensionais

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

Matriz Jacobiana

Vetor de Entrada
(4096 dimensões)



Vetor de Saída
(2048 dimensões)

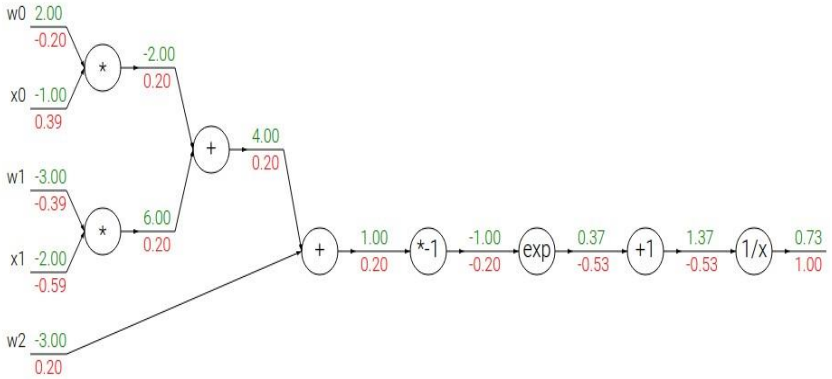
Na prática, processa-se todo um lote ou "*minibatch*" (p.ex. 100 amostras) de uma só vez

Assim, as dimensões da matriz Jacobiana desse exemplo seriam

2.048 × 4.096 × 100

≈ 839 milhões de pesos :(

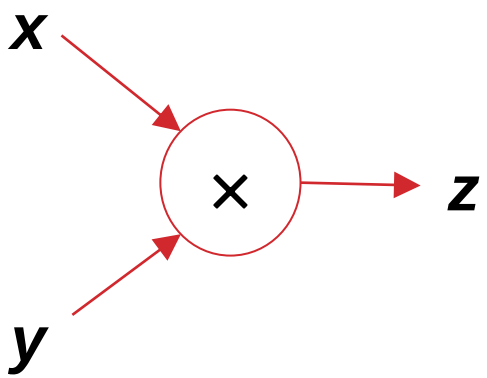
Implementação – Forward / Backward API



Para um grafo ou rede (*pseudocódigo*)

```
class ComputationalGraph(object):
    #...
    def forward(inputs):
        # 1. [pass inputs to input gates...]
        # 2. forward the computational graph:
        for gate in self.graph.nodes_topologically_sorted():
            gate.forward()
        return loss # the final gate in the graph outputs the loss
    def backward():
        for gate in reversed(self.graph.nodes_topologically_sorted()):
            gate.backward() # little piece of backprop (chain rule applied)
        return inputs_gradients
```

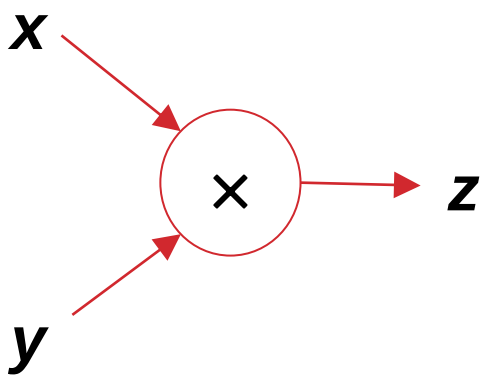
Implementação – Forward / Backward API



(*x*, *y*, *z* escalares)

```
class MultiplyGate(object):
    def forward(x,y):
        z = x*y
        return z
    def backward(dz):
        # dx = ... #todo
        # dy = ... #todo
        return [dx, dy]
```

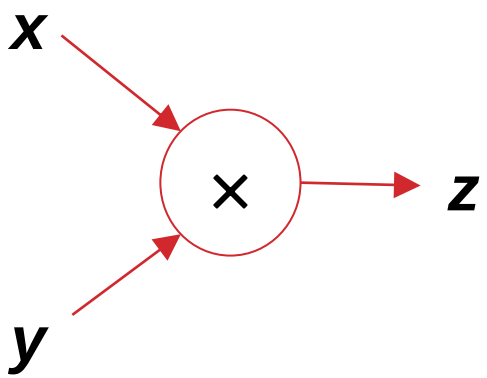
Implementação – Forward / Backward API



(*x*, *y*, *z* escalares)

```
class MultiplyGate(object):
    def forward(x,y):
        z = x*y
        return z
    def backward(dz):
        # dx = ... #todo
        # dy = ... #todo
        return [dx, dy]
```

Implementação – Forward / Backward API



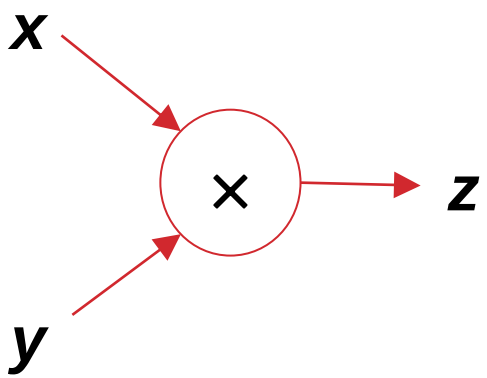
(*x*, *y*, *z* escalares)

```
class MultiplyGate(object):
    def forward(x,y):
        z = x*y
        return z
    def backward(dz):
        # dx = ... #todo
        # dy = ... #todo
        return [dx, dy]
```

$\frac{\partial L}{\partial z}$

$\frac{\partial L}{\partial x}$

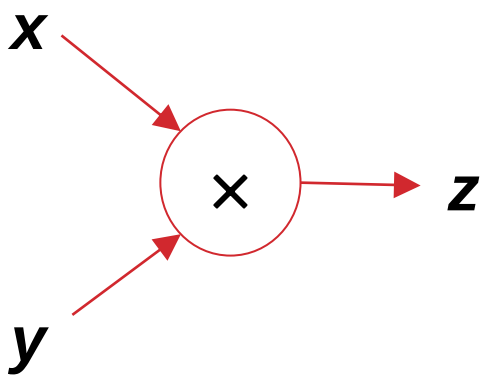
Implementação – Forward / Backward API



(*x*, *y*, *z* escalares)

```
class MultiplyGate(object):
    def forward(x,y):
        z = x*y
        self.x = x # must keep these around!
        self.y = y
        return z
    def backward(dz):
        dx = self.y * dz # [dz/dx * dL/dz]
        dy = self.x * dz # [dz/dy * dL/dz]
        return [dx, dy]
```

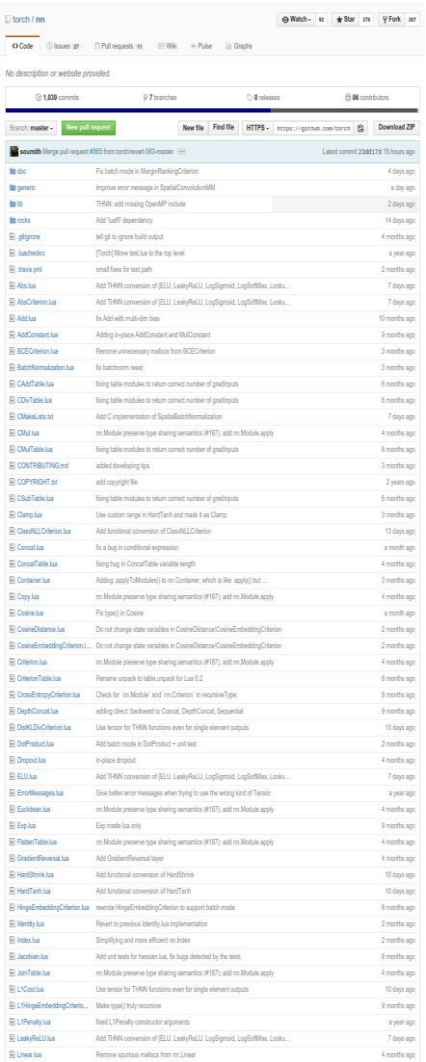
Implementação – Forward / Backward API

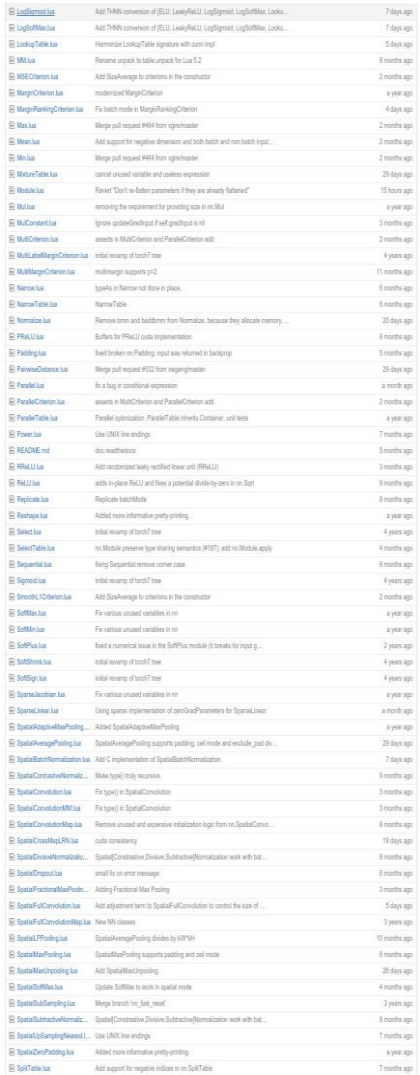


(*x*, *y*, *z* escalares)

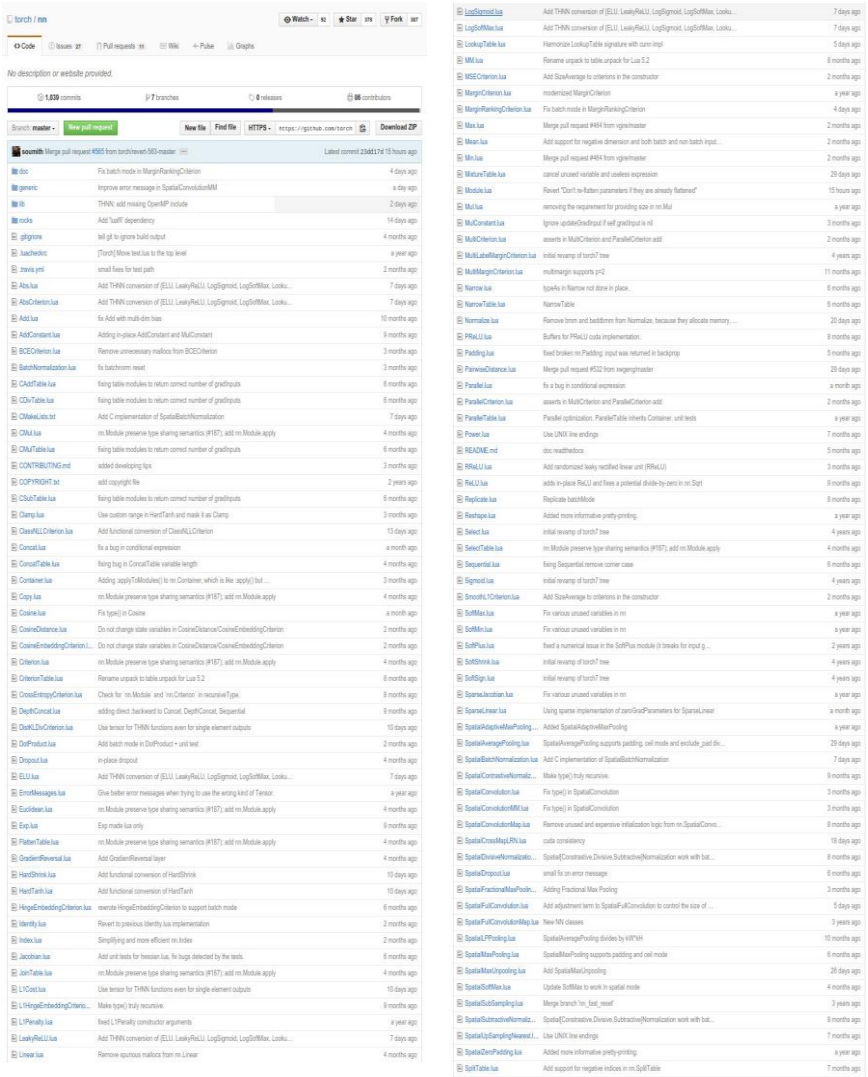
```
class MultiplyGate(object):
    def forward(x,y):
        z = x*y
        self.x = x # must keep these around!
        self.y = y
        return z
    def backward(dz):
        dx = self.y * dz # [dz/dx * dL/dz]
        dy = self.x * dz # [dz/dy * dL/dz]
        return [dx, dy]
```

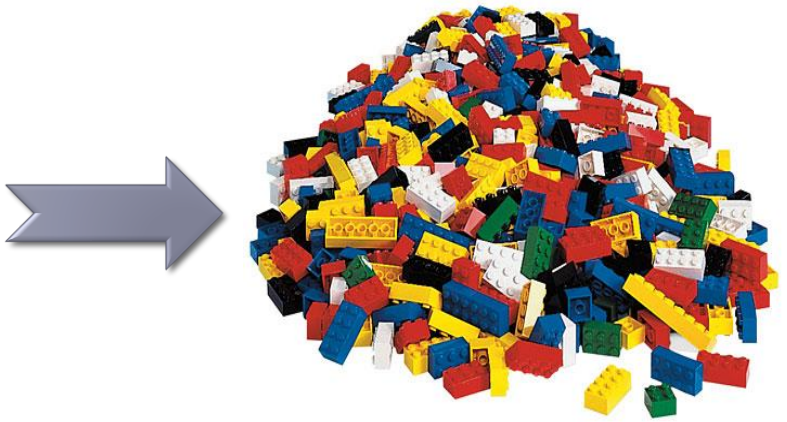
Biblioteca de Componentes / Camadas





Biblioteca de Componentes / Camadas





Biblioteca de Componentes / Camadas