Tema 2.3
Optimización II: Backpropagation

Miguel Ángel Martínez del Amor

Deep Learning

Departamento Ciencias de la Computación e Inteligencia ARtificial

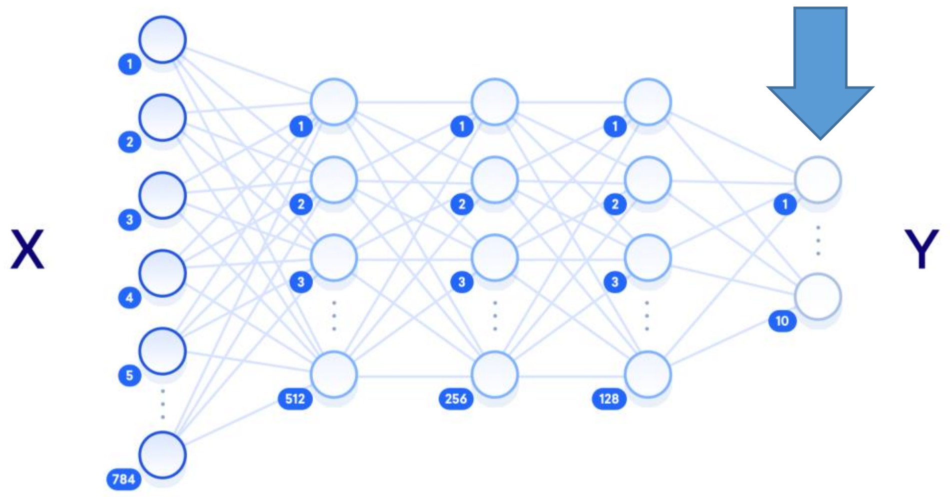
Universidad de Sevilla

Contenido

- Propagación del gradiente
- Grafo computacional
- Vectorización

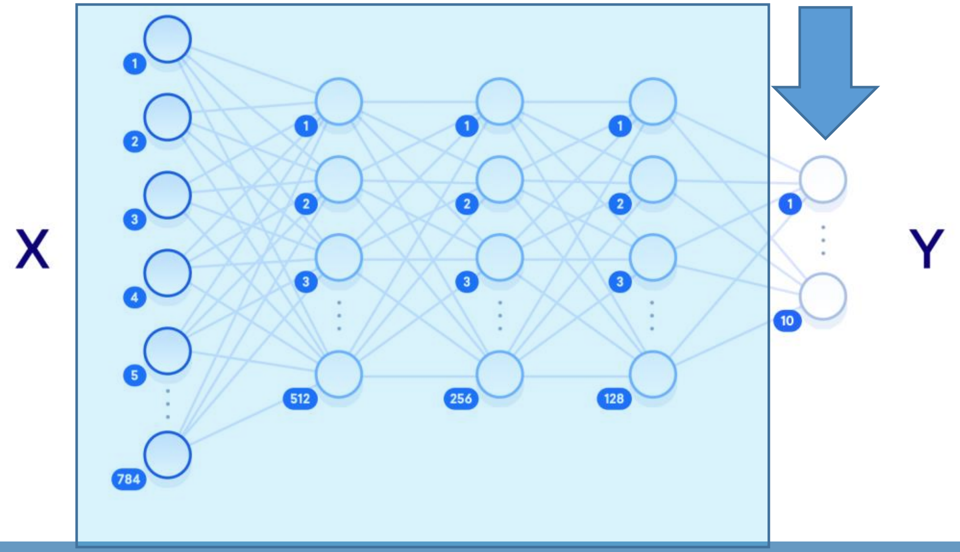
Descenso por gradiente

- Las actualizaciones se pueden aplicar directamente para actualizar pesos en modelos como regresión lineal y logística (perceptrón con función de activación sigmoide).
- ¿Cómo proceder con red neuronal multicapa?



Descenso por gradiente

- Las actualizaciones se pueden aplicar directamente para actualizar pesos en modelos como regresión lineal y logística (perceptrón con función de activación sigmoide).
- ¿Cómo proceder con red neuronal multicapa?

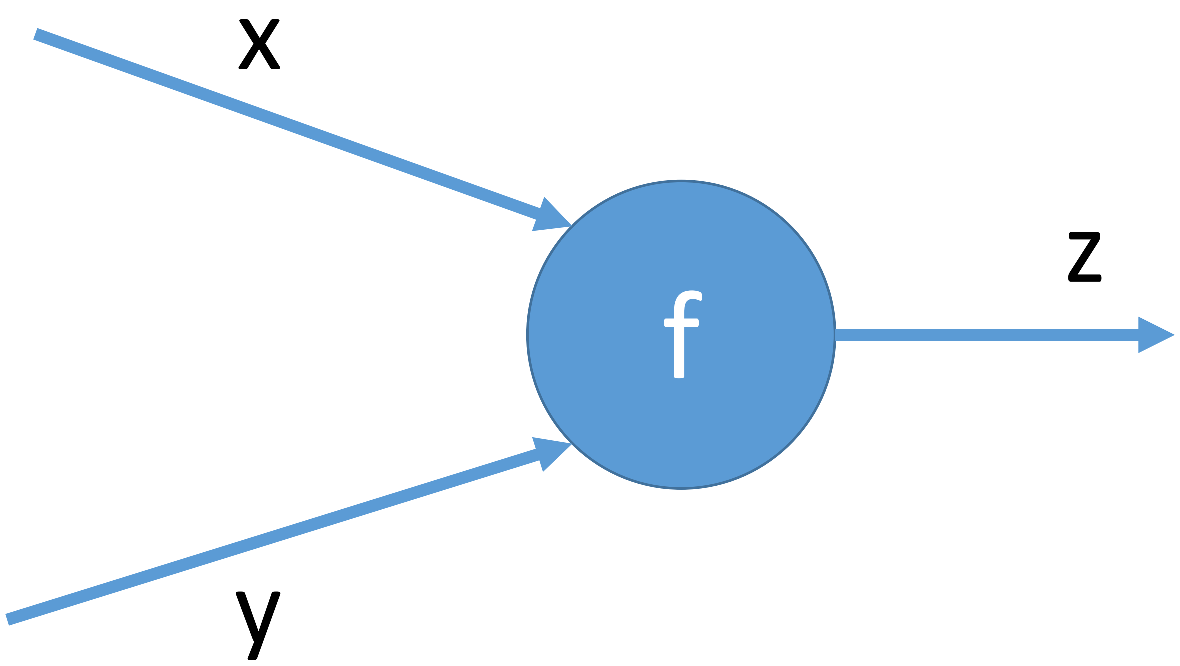


Descenso por gradiente

- Las actualizaciones se pueden aplicar directamente para actualizar pesos en modelos como regresión lineal y logística (perceptrón con función de activación sigmoide).
- ¿Cómo proceder con red neuronal multicapa?
- En la capa de salida, usar la actualización de pesos con el gradiente sobre la función de coste.
- **Problema**: en las capas ocultas desconocemos los valores esperados.
- **Solución**: propagar los gradientes, algoritmo de retropropagación (**backpropagation)**

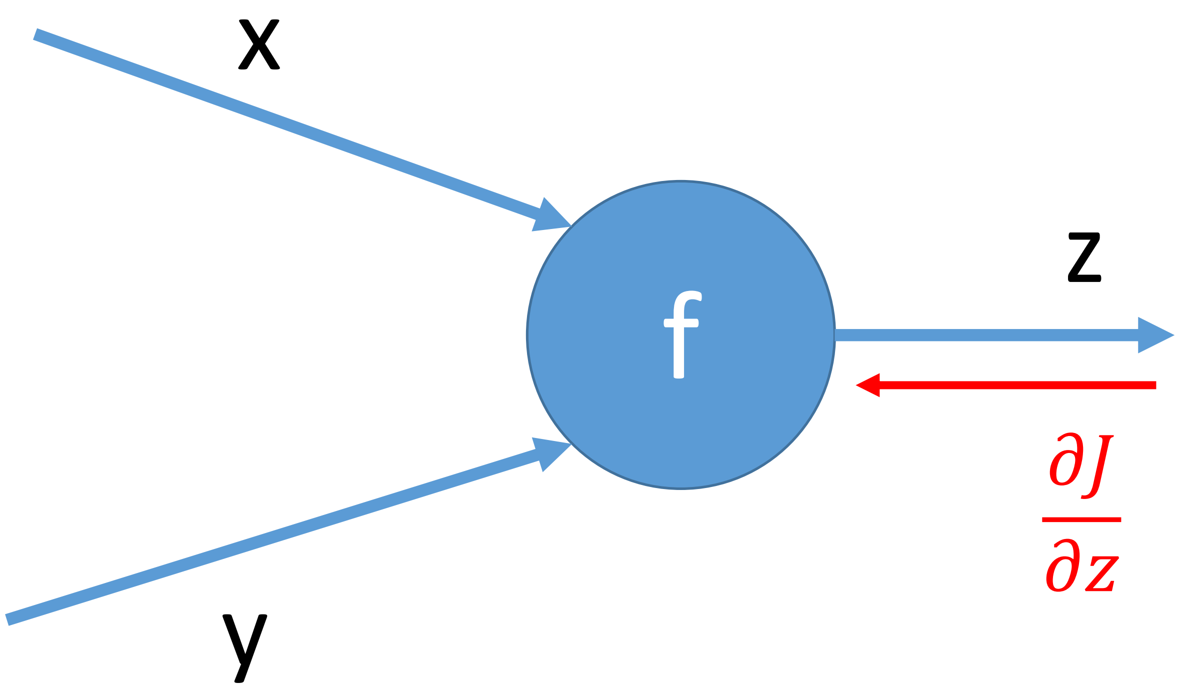
Propagación de gradiente

- Propagación hacia adelante (forward):



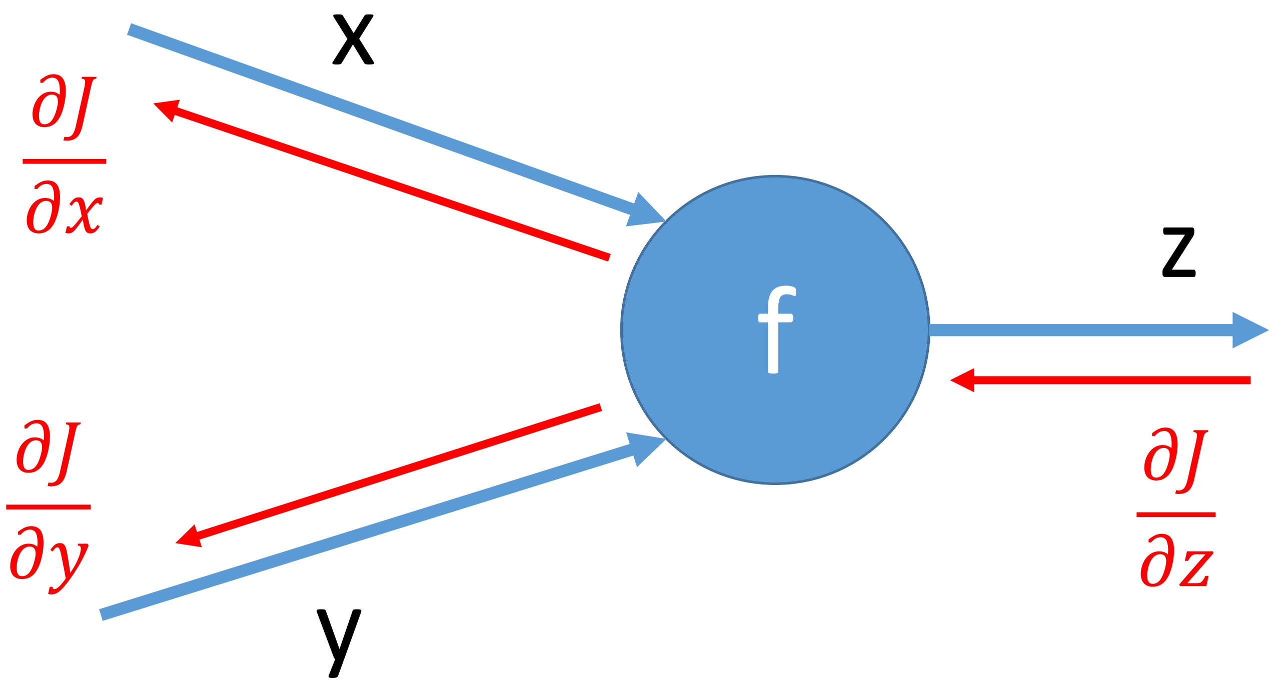
Propagación de gradiente

- Propagación hacia atrás (backward):



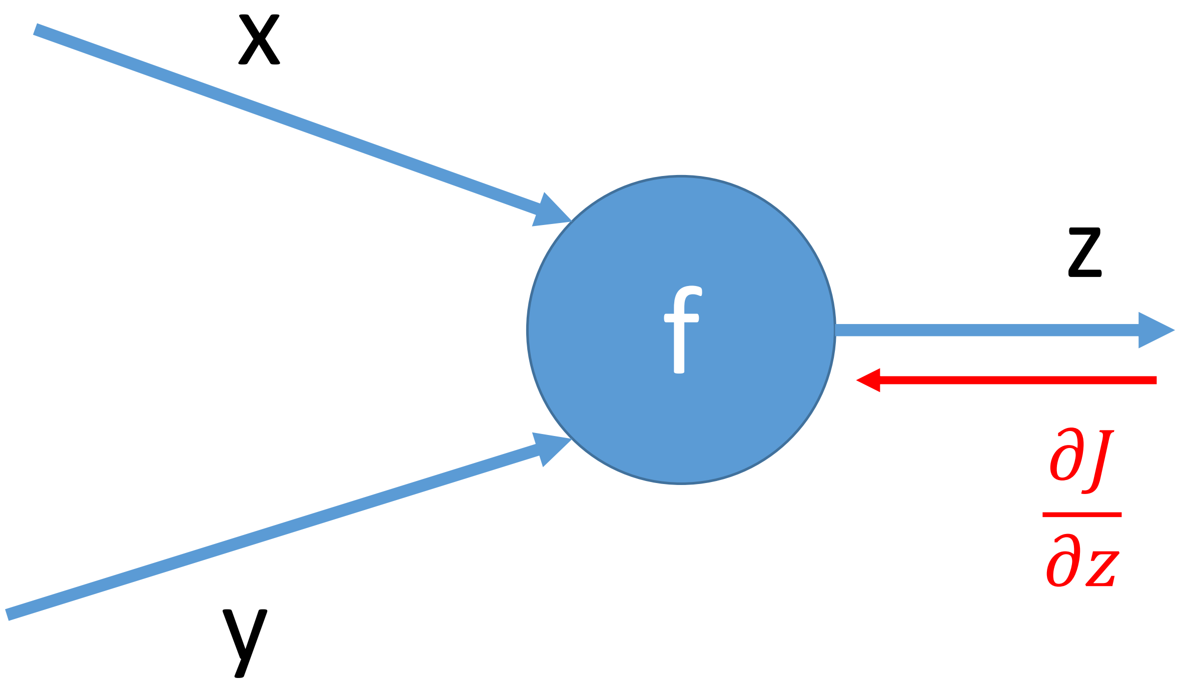
Propagación de gradiente

- Propagación hacia atrás (backward):



Propagación de gradiente

- Regla de la cadena:
  - Clave del algoritmo de backpropagation.

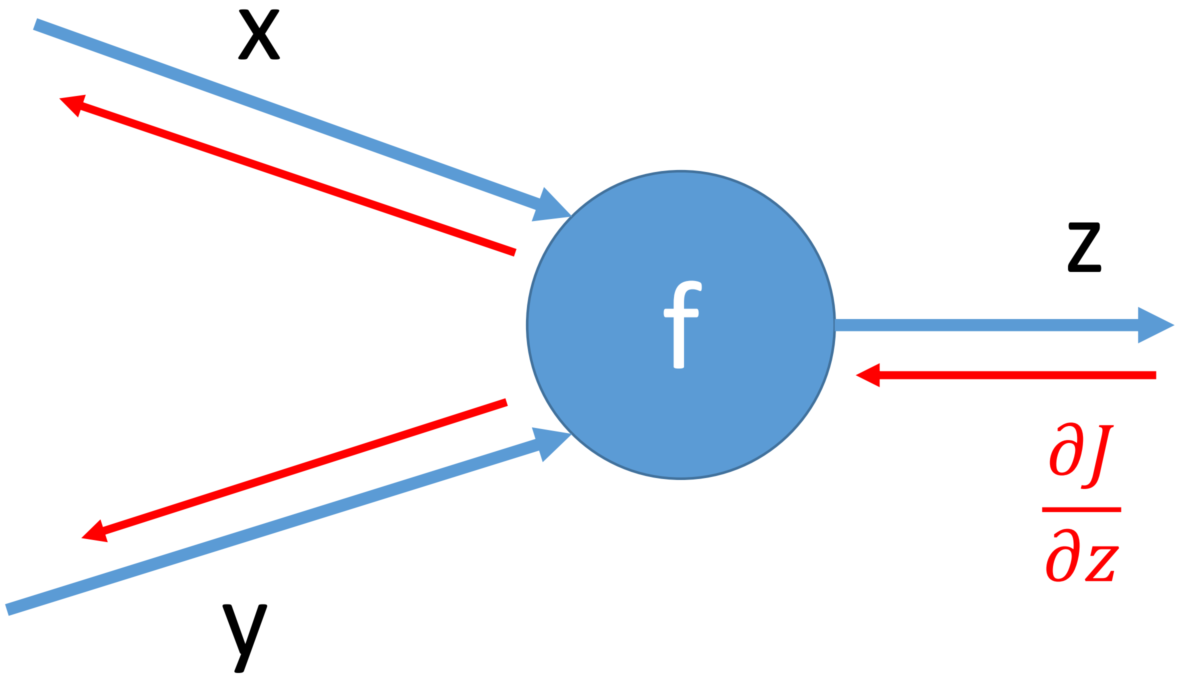


Propagación de gradiente

- Regla de la cadena:
  - Clave del algoritmo de backpropagation.

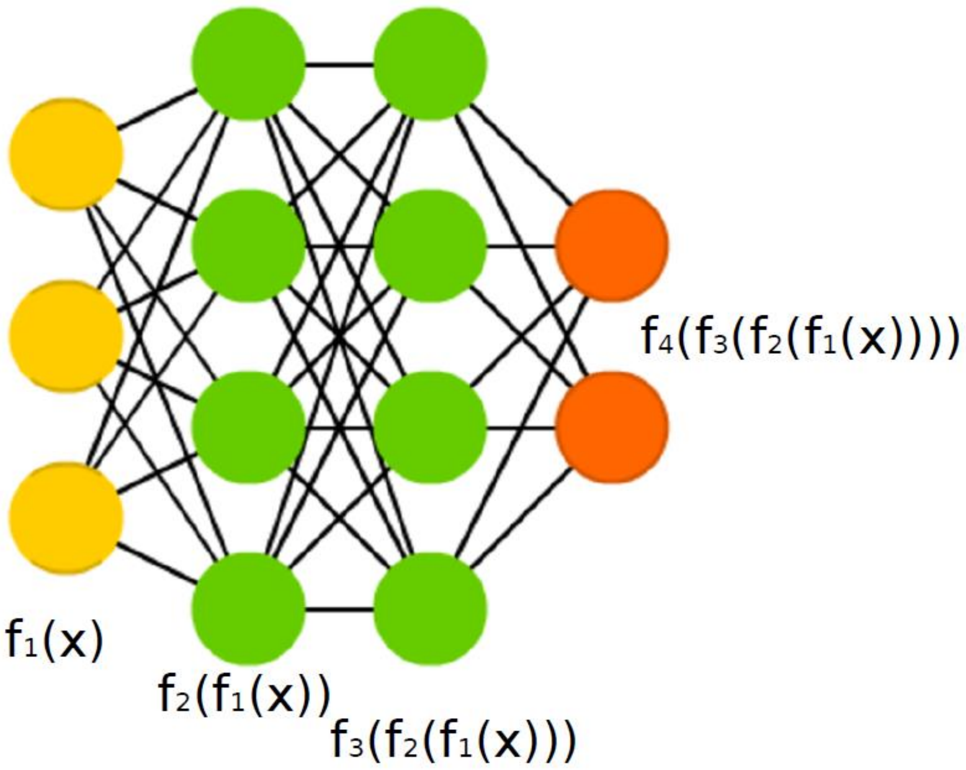
$$\frac{\partial J}{\partial x} = \frac{\partial J}{\partial z}\frac{\partial z}{\partial x}$$

$$\frac{\partial J}{\partial y} = \frac{\partial J}{\partial z}\frac{\partial z}{\partial y}$$



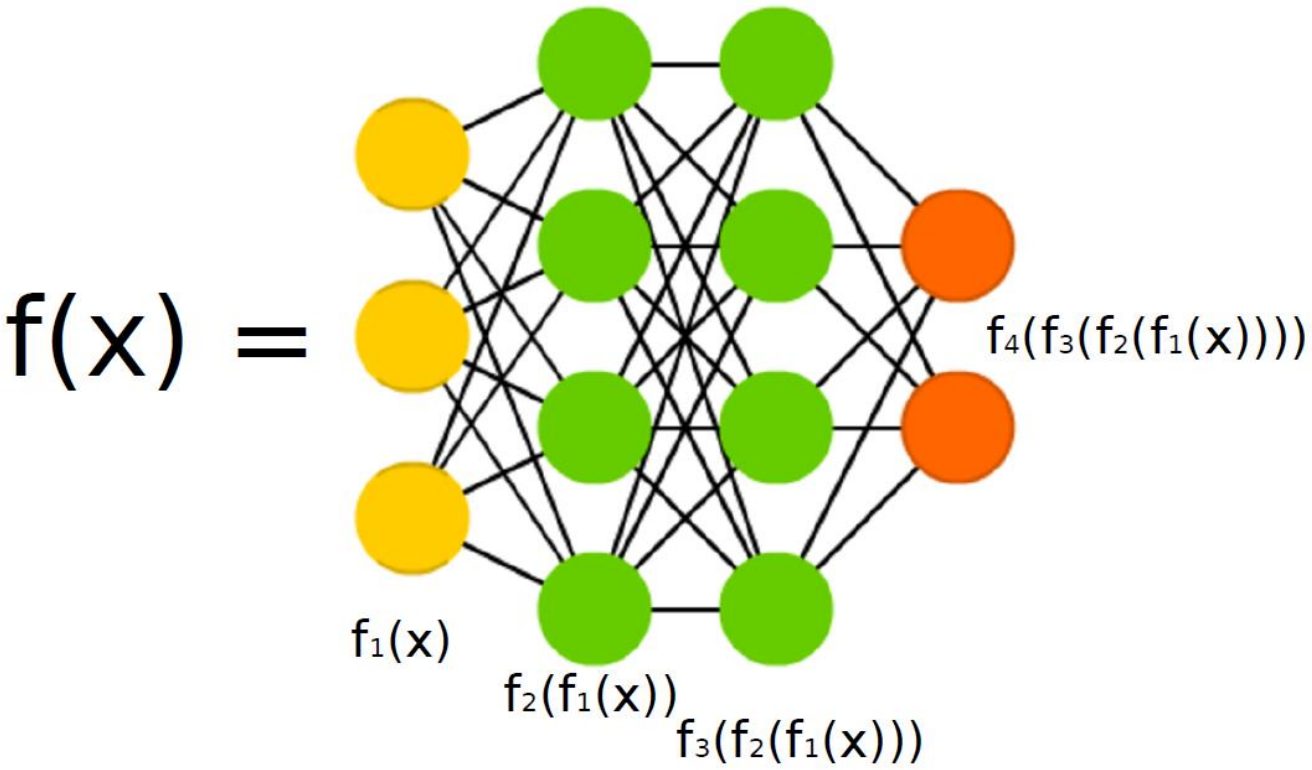
Propagación de gradiente

f(x) =



$$\frac{\partial f}{\partial x} = \frac{\partial f_4}{\partial x}$$

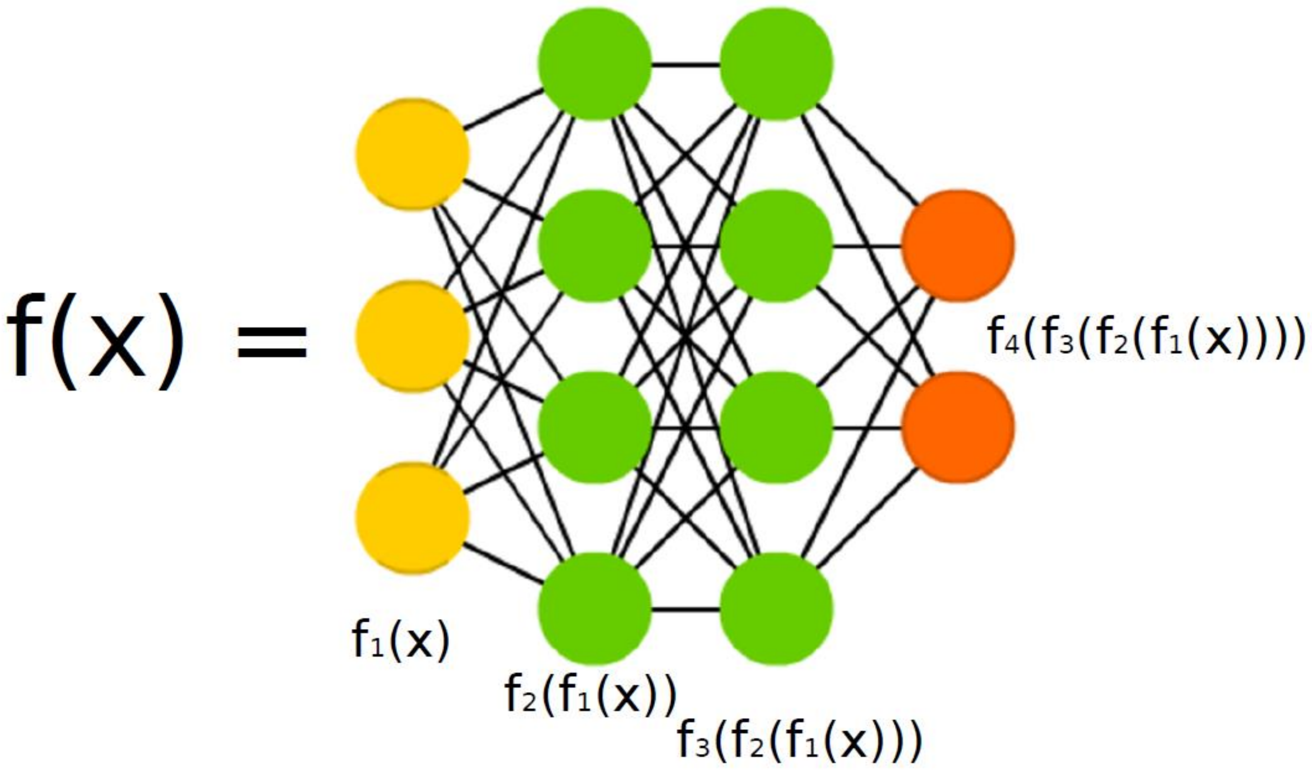
Propagación de gradiente



$$\frac{\partial f}{\partial x} = \frac{\partial f_4}{\partial x}$$

$$\frac{\partial f_4}{\partial x} = \frac{\partial f_4}{\partial f_3}\frac{\partial f_3}{\partial x}$$

Propagación de gradiente



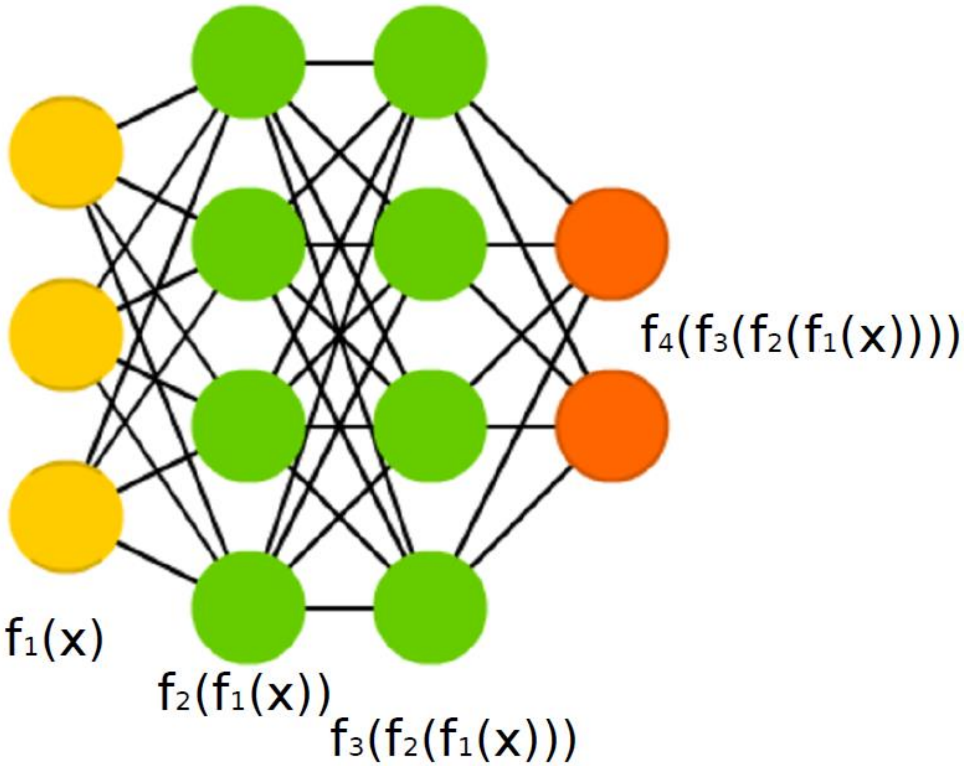
$$\frac{\partial f}{\partial x} = \frac{\partial f_4}{\partial x}$$

$$\frac{\partial f_4}{\partial x} = \frac{\partial f_4}{\partial f_3}\frac{\partial f_3}{\partial x}$$

$$\frac{\partial f_4}{\partial x} = \frac{\partial f_4}{\partial f_3}\frac{\partial f_3}{\partial f_2}\frac{\partial f_2}{\partial x}$$

Propagación de gradiente

$f(x) =$



$$\frac{\partial f}{\partial x} = \frac{\partial f_4}{\partial x}$$

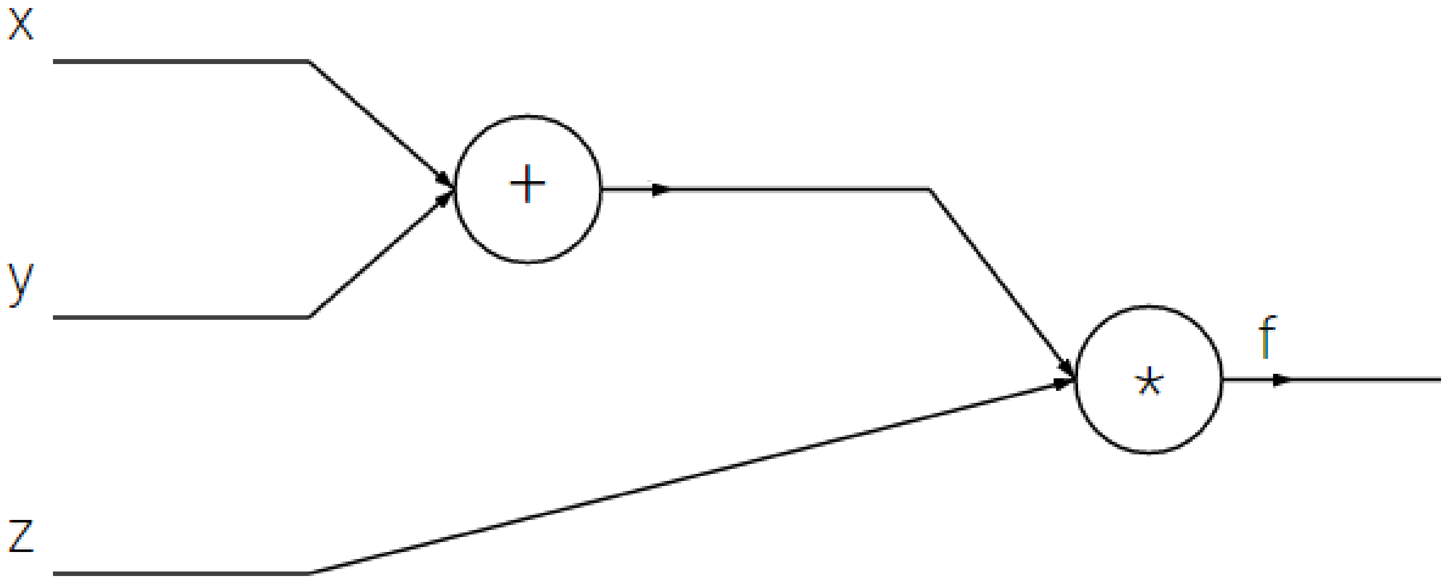
$$\frac{\partial f_4}{\partial x} = \frac{\partial f_4}{\partial f_3}\frac{\partial f_3}{\partial x}$$

$$\frac{\partial f_4}{\partial x} = \frac{\partial f_4}{\partial f_3}\frac{\partial f_3}{\partial f_2}\frac{\partial f_2}{\partial x}$$

$$\frac{\partial f_4}{\partial x} = \frac{\partial f_4}{\partial f_3}\frac{\partial f_3}{\partial f_2}\frac{\partial f_2}{\partial f_1}\frac{\partial f_1}{\partial x}$$

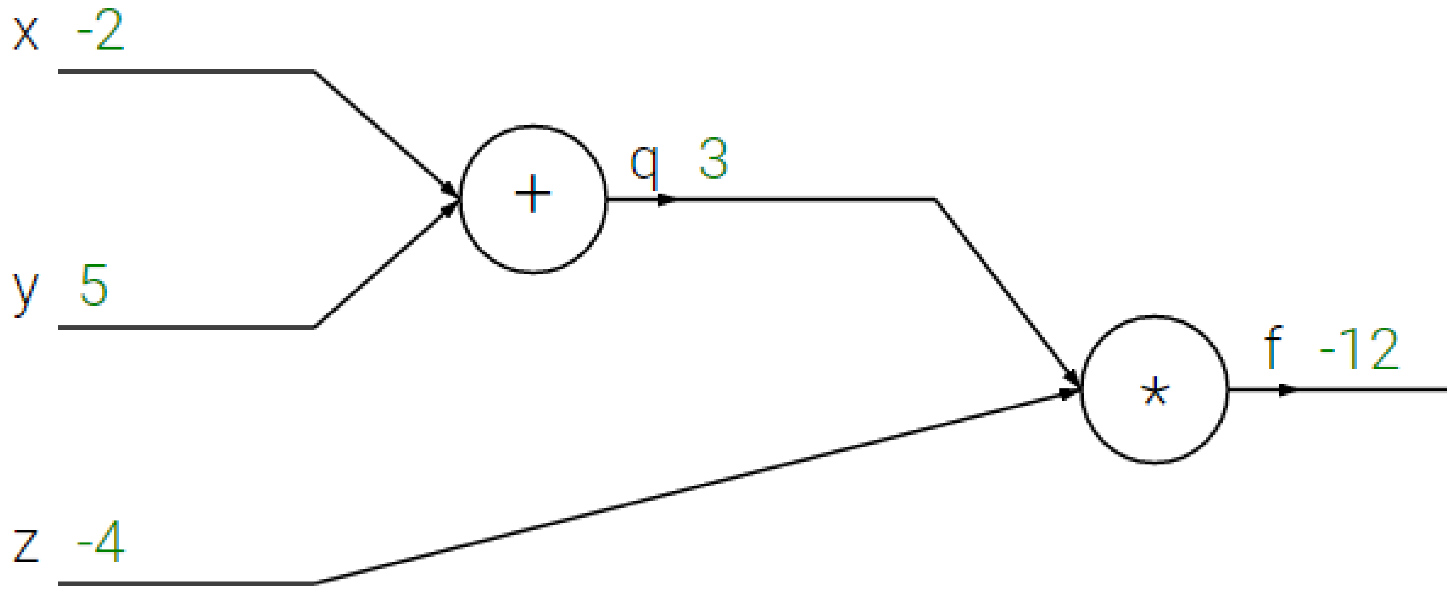
Backpropagation: Grafo de computación

- Veamos un ejemplo: $f(x, y, z) = (x + y)z$
- Sea x=-2, y=5, z=-4
- Sea $q = (x + y)$
- $f = qz$
- $\frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$
- $\frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$
- Buscamos: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



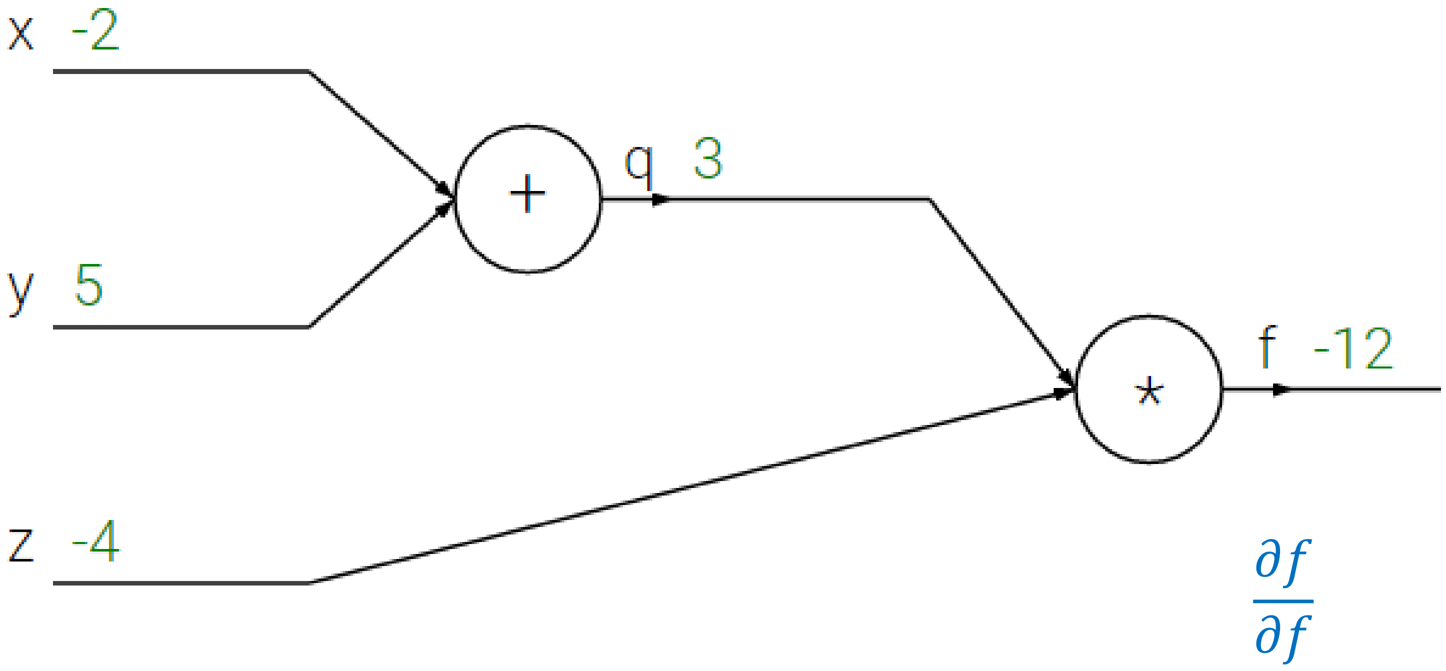
Backpropagation: Grafo de computación

- Veamos un ejemplo: $f(x, y, z) = (x + y)z$
- Sea x=-2, y=5, z=-4
- Sea $q = (x + y)$
- $f = qz$
- $\frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$
- $\frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$
- Buscamos: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



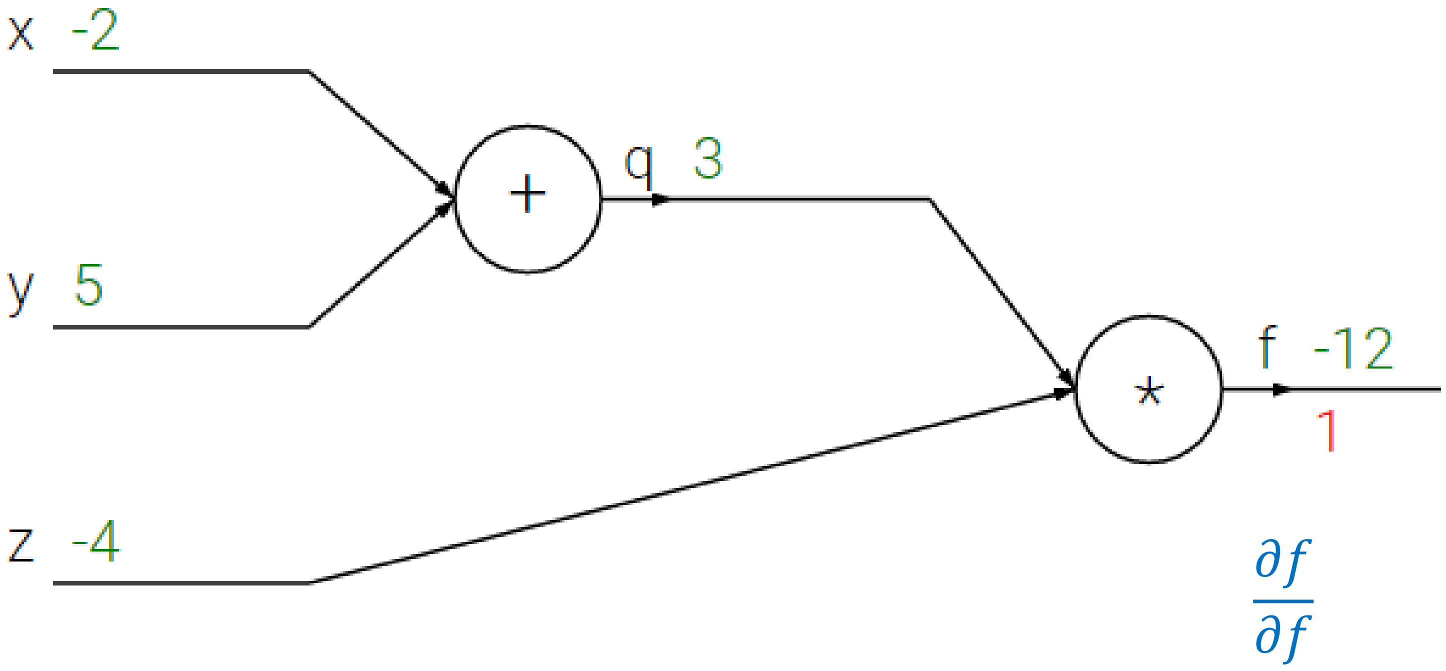
Backpropagation: Grafo de computación

- Veamos un ejemplo: $f(x, y, z) = (x + y)z$
- Sea x=-2, y=5, z=-4
- Sea $q = (x + y)$
- $f = qz$
- $\frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$
- $\frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$
- Buscamos: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



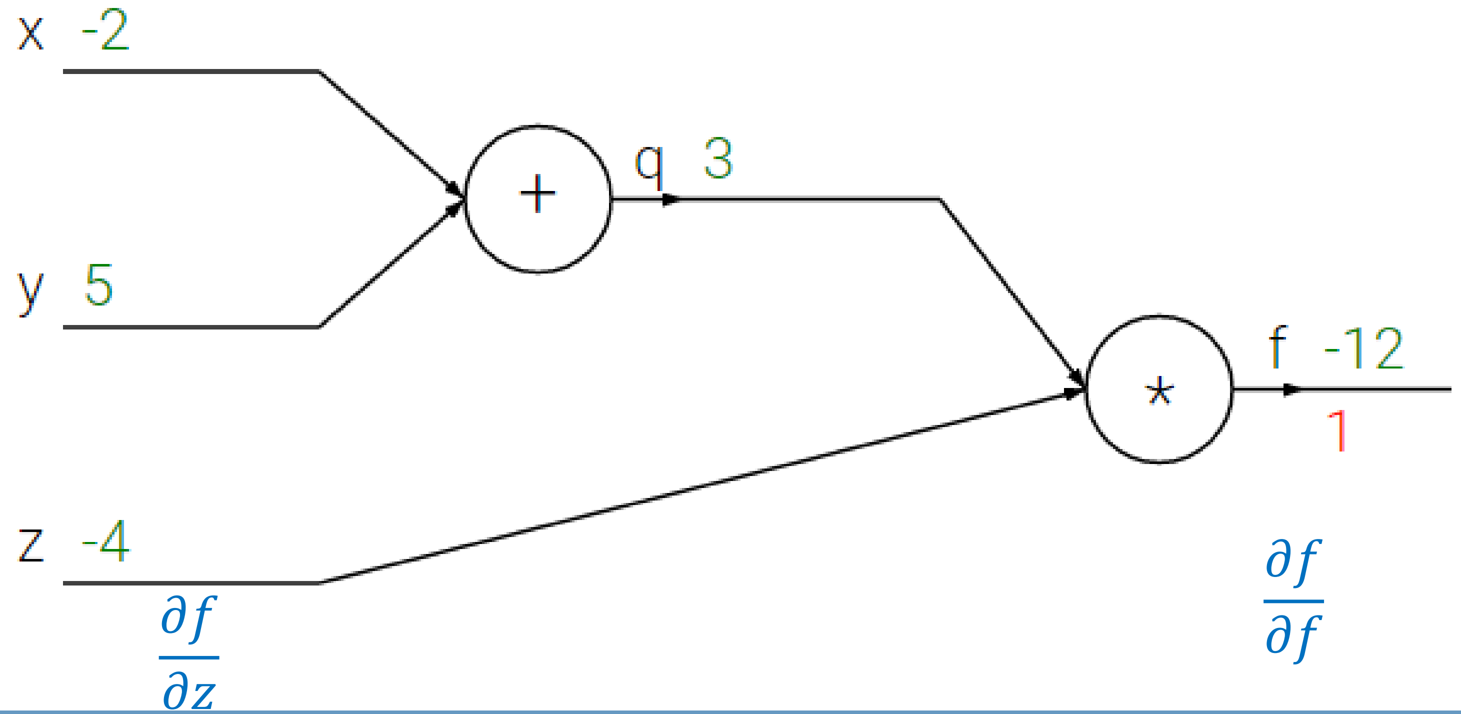
Backpropagation: Grafo de computación

- Veamos un ejemplo: $f(x, y, z) = (x + y)z$
- Sea x=-2, y=5, z=-4
- Sea $q = (x + y)$
- $f = qz$
- $\frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$
- $\frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$
- Buscamos: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



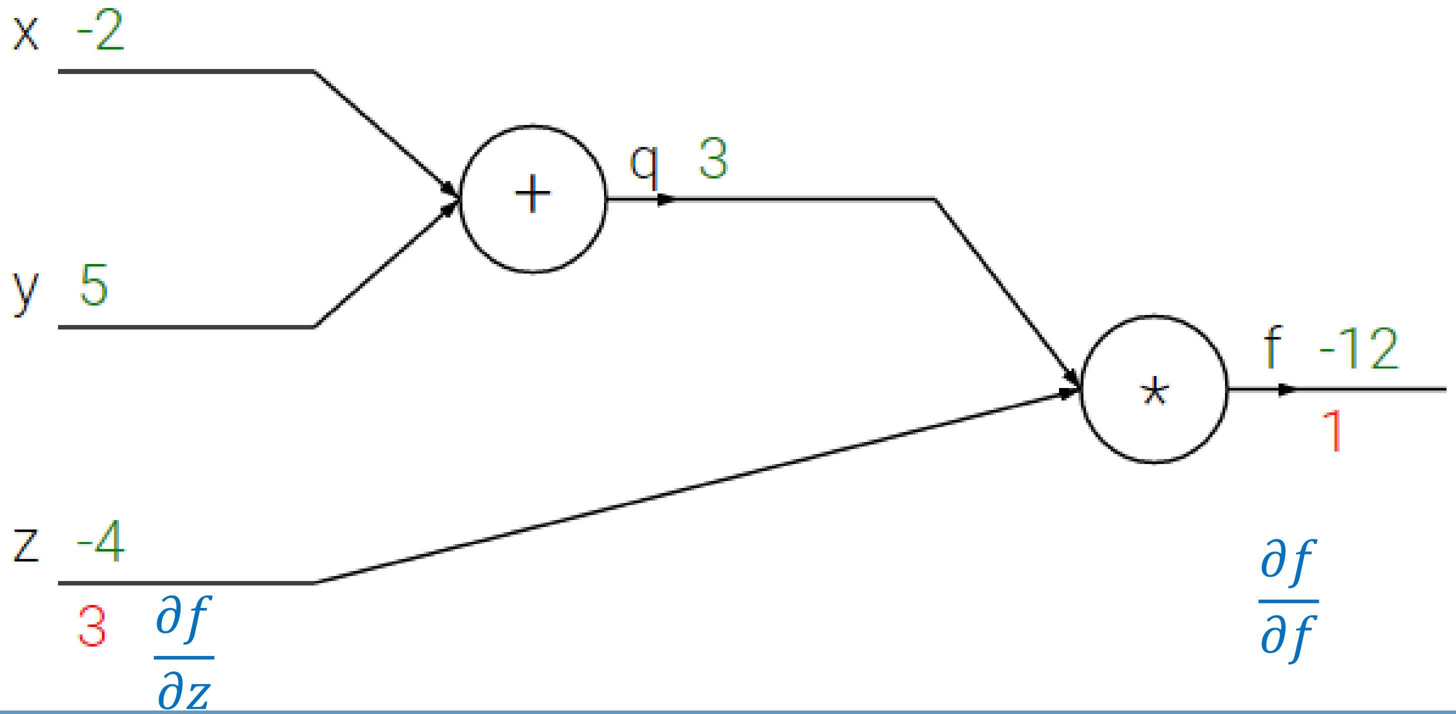
Backpropagation: Grafo de computación

- Veamos un ejemplo: $f(x, y, z) = (x + y)z$
- Sea x=-2, y=5, z=-4
- Sea $q = (x + y)$
- $f = qz$
- $\frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$
- $\frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$
- Buscamos: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



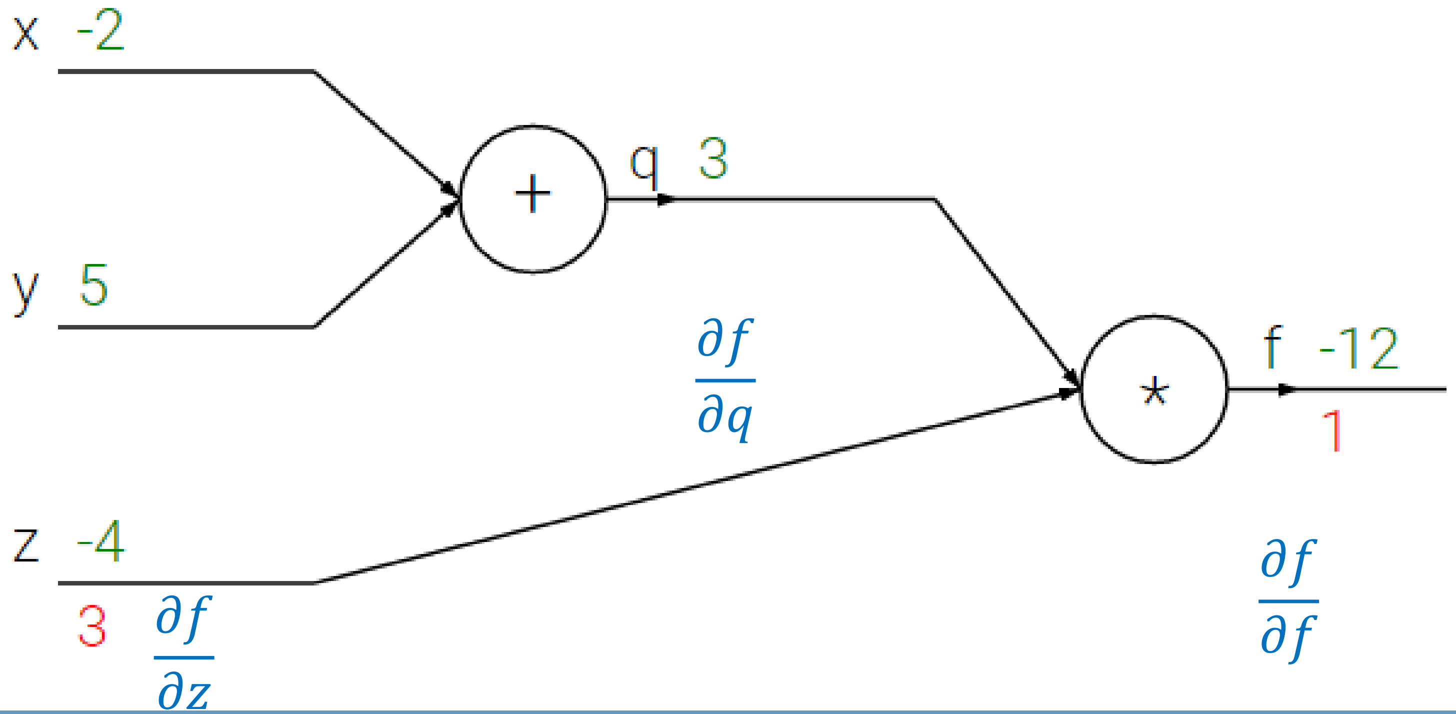
Backpropagation: Grafo de computación

- Veamos un ejemplo: $f(x, y, z) = (x + y)z$
- Sea x=-2, y=5, z=-4
- Sea $q = (x + y)$
- $f = qz$
- $\frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$
- $\frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$
- Buscamos: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



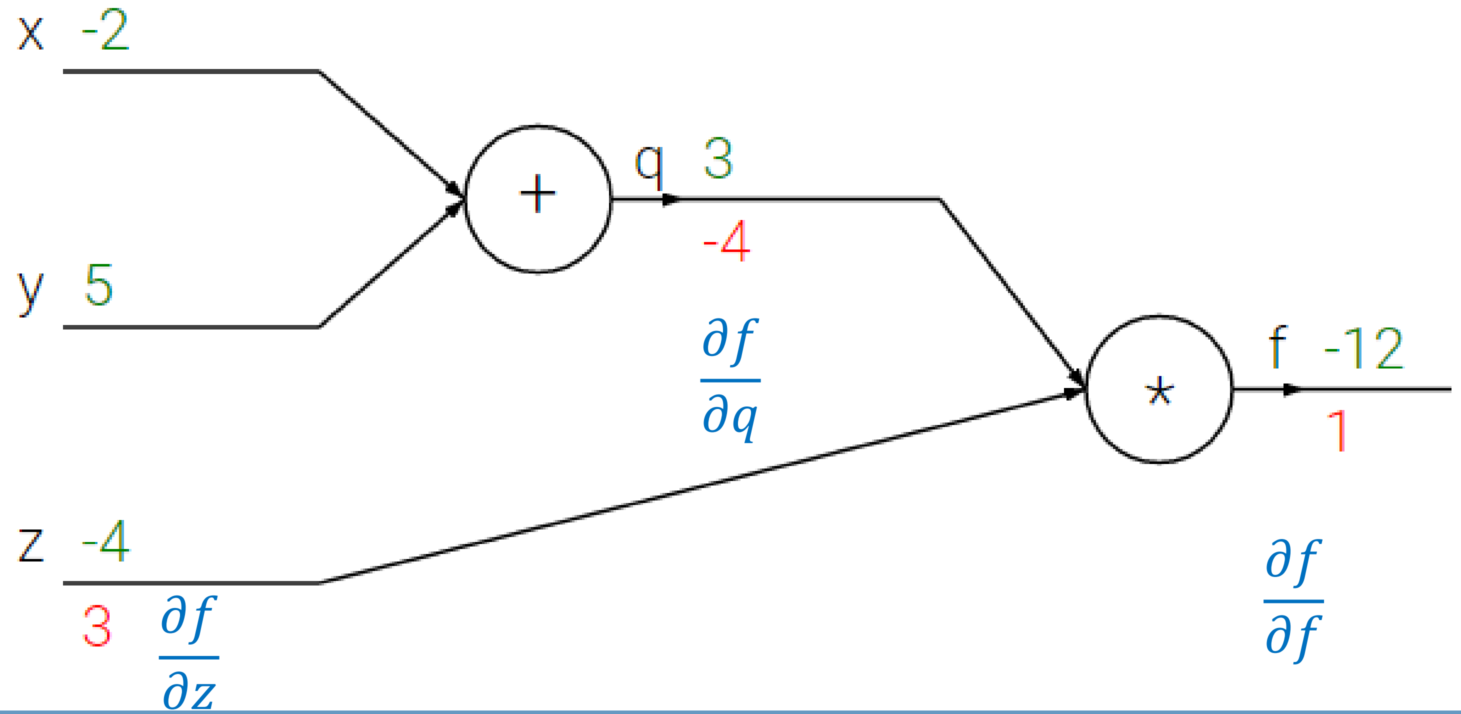
Backpropagation: Grafo de computación

- Veamos un ejemplo: $f(x, y, z) = (x + y)z$
- Sea x=-2, y=5, z=-4
- Sea $q = (x + y)$
- $f = qz$
- $\frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$
- $\frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$
- Buscamos: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



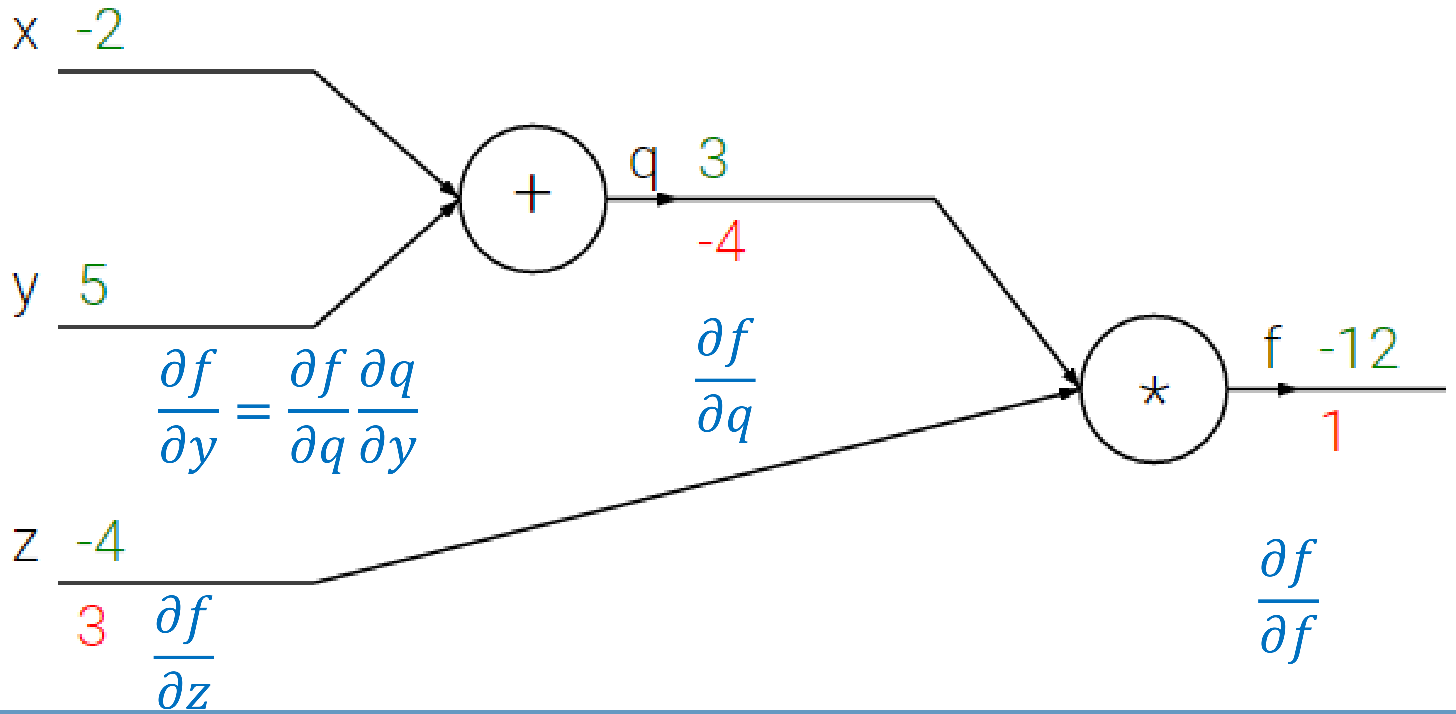
Backpropagation: Grafo de computación

- Veamos un ejemplo: $f(x, y, z) = (x + y)z$
- Sea x=-2, y=5, z=-4
- Sea $q = (x + y)$
- $f = qz$
- $\frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$
- $\frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$
- Buscamos: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



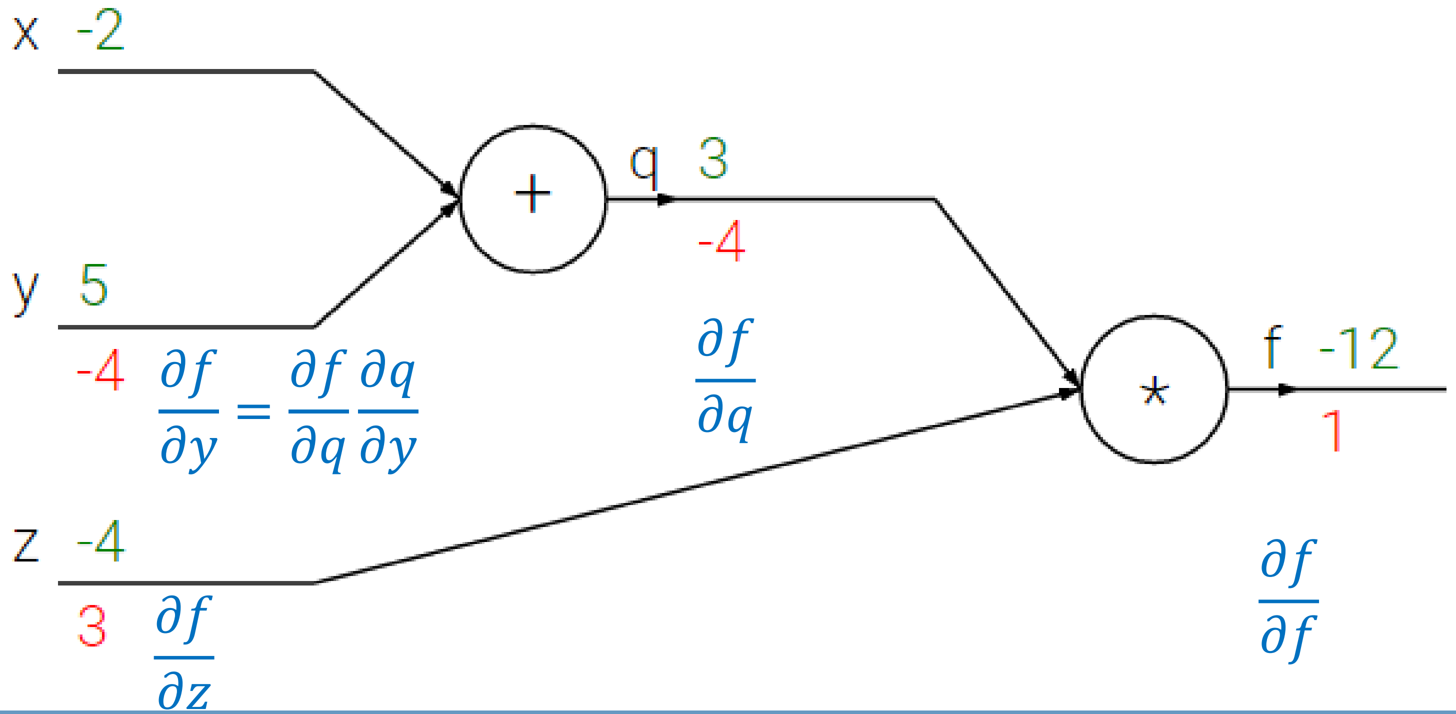
Backpropagation: Grafo de computación

- Veamos un ejemplo: $f(x, y, z) = (x + y)z$
- Sea x=-2, y=5, z=-4
- Sea $q = (x + y)$
- $f = qz$
- $\frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$
- $\frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$
- Buscamos: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



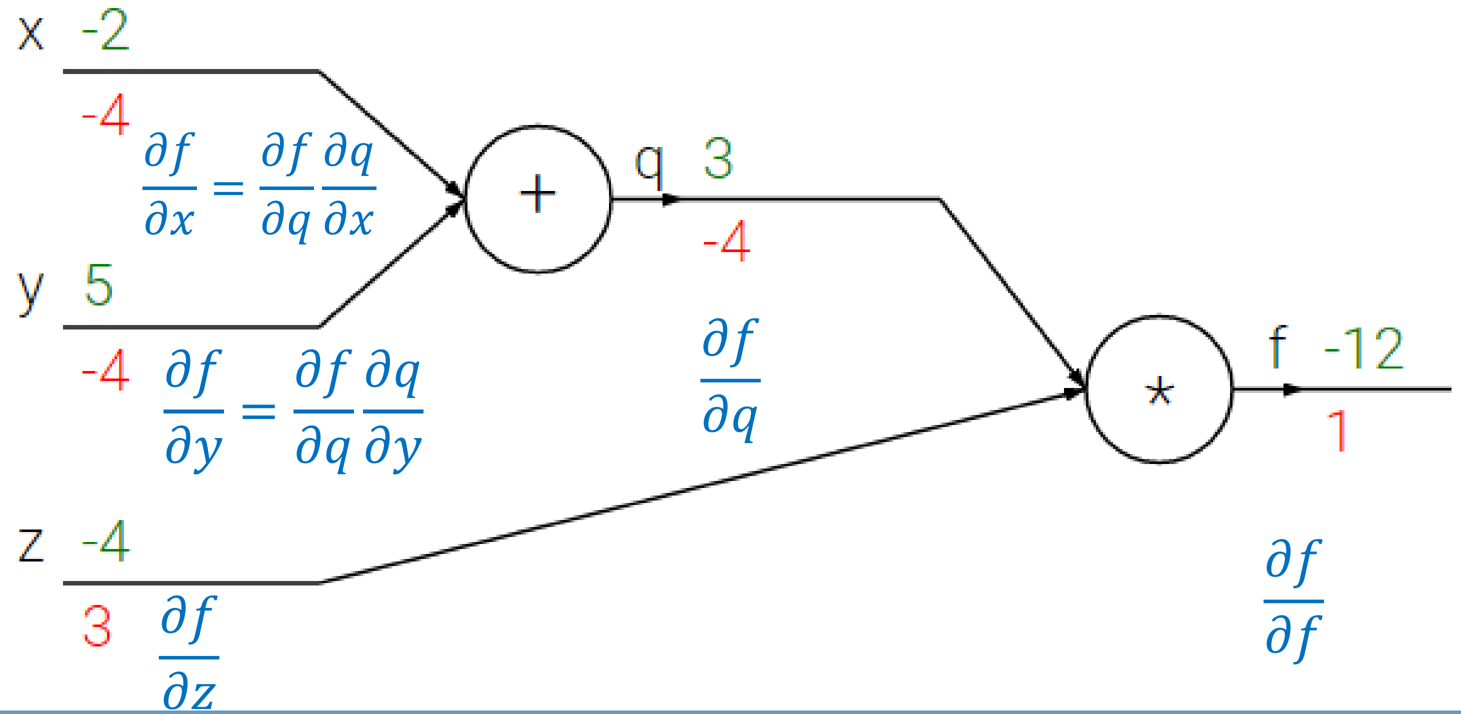
Backpropagation: Grafo de computación

- Veamos un ejemplo: $f(x, y, z) = (x + y)z$
- Sea x=-2, y=5, z=-4
- Sea $q = (x + y)$
- $f = qz$
- $\frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$
- $\frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$
- Buscamos: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



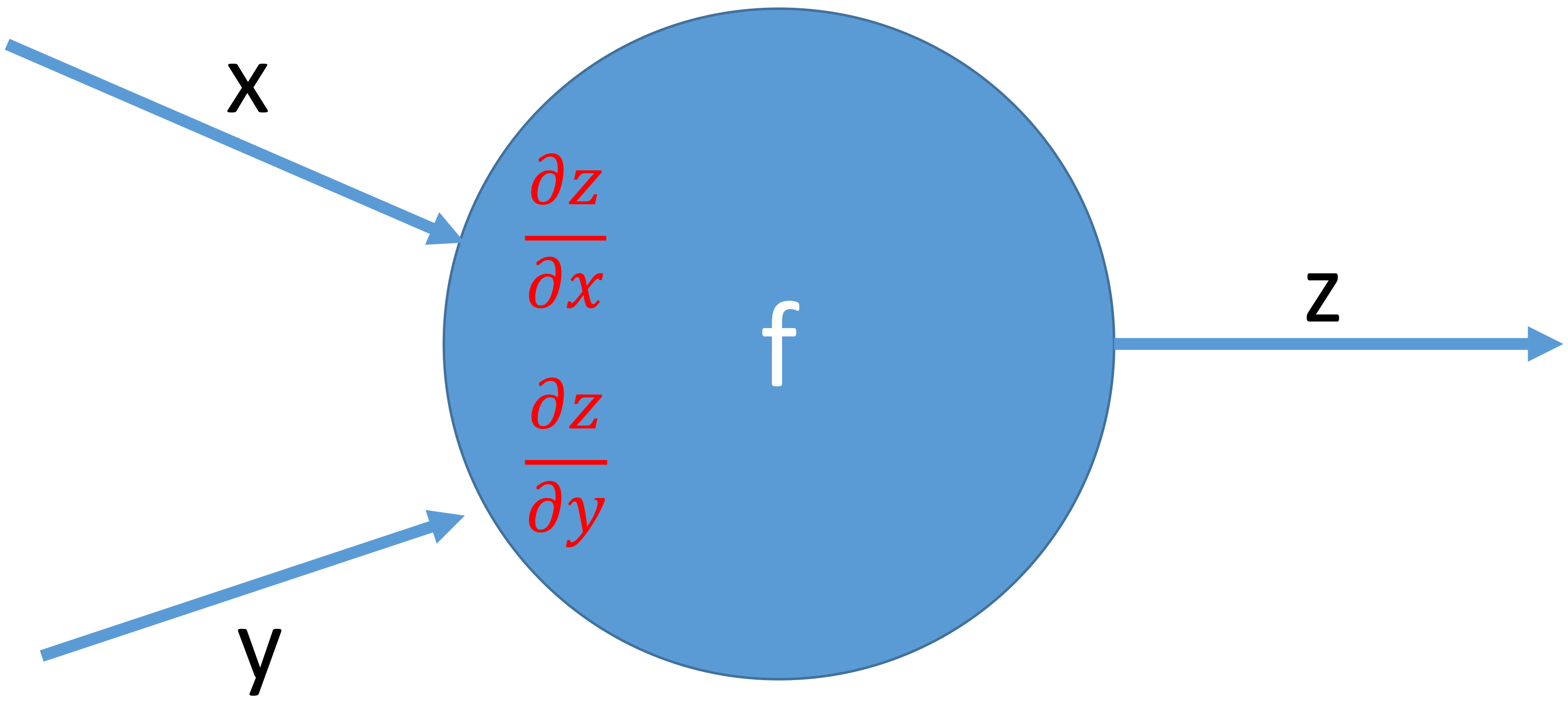
Backpropagation: Grafo de computación

- Veamos un ejemplo: $f(x, y, z) = (x + y)z$
- Sea x=-2, y=5, z=-4
- Sea $q = (x + y)$
- $f = qz$
- $\frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$
- $\frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$
- Buscamos: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



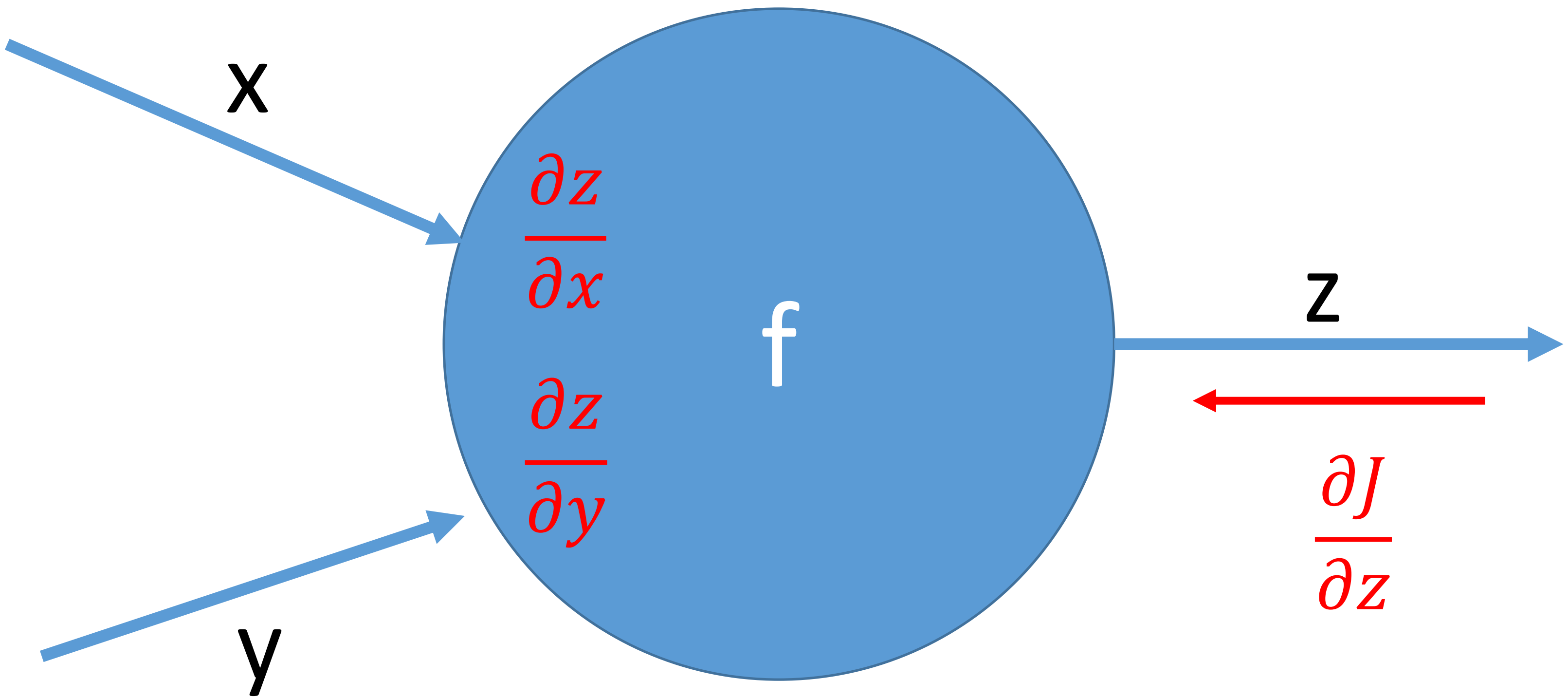
Propagación de gradiente

- Cálculo de gradientes locales mientras forward propagation.



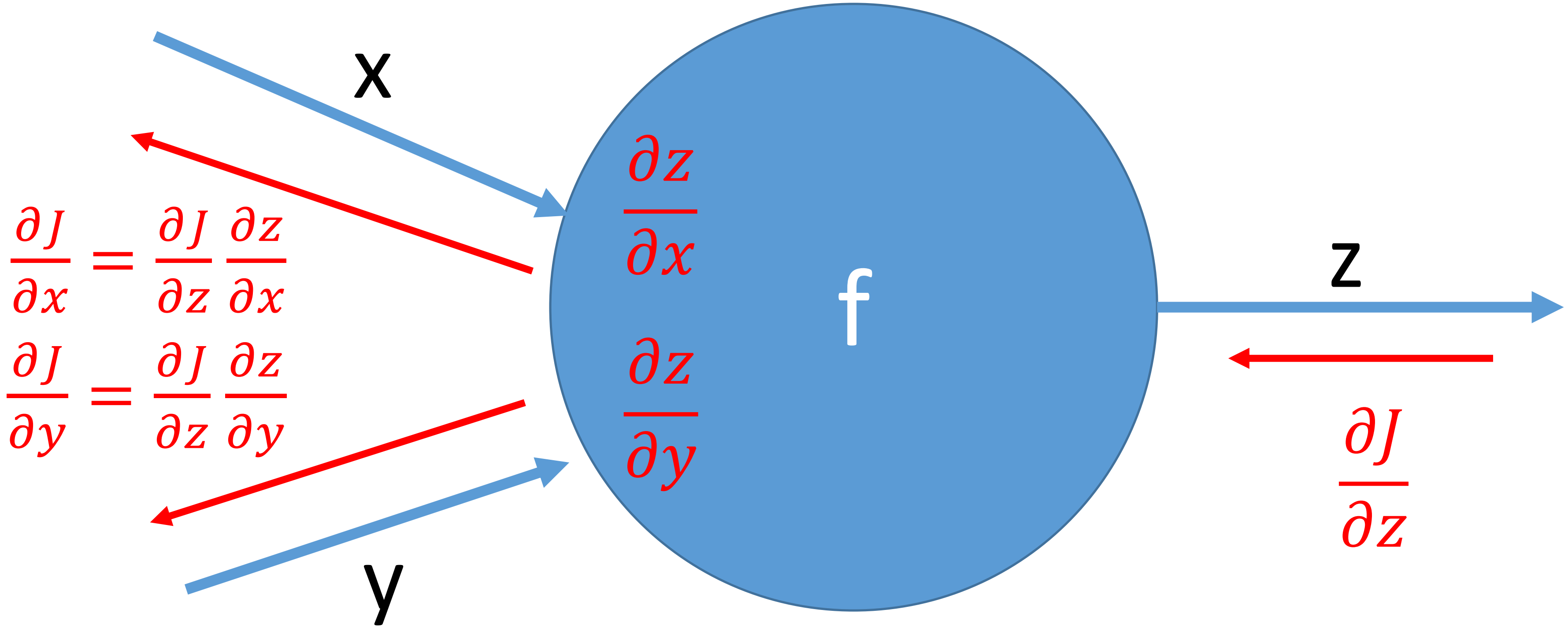
Propagación de gradiente

- Cálculo de gradientes locales mientras forward propagation.



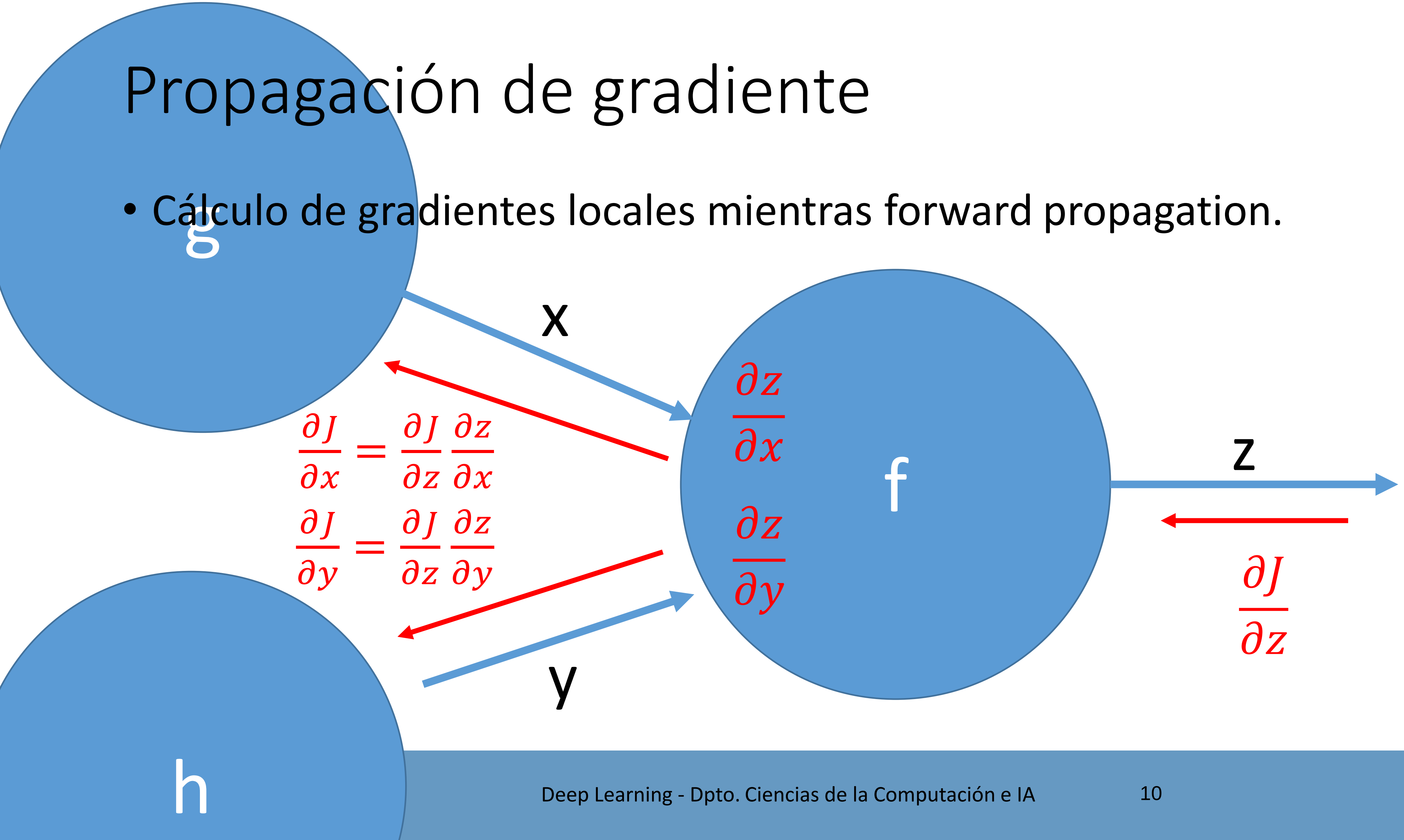
Propagación de gradiente

- Cálculo de gradientes locales mientras forward propagation.

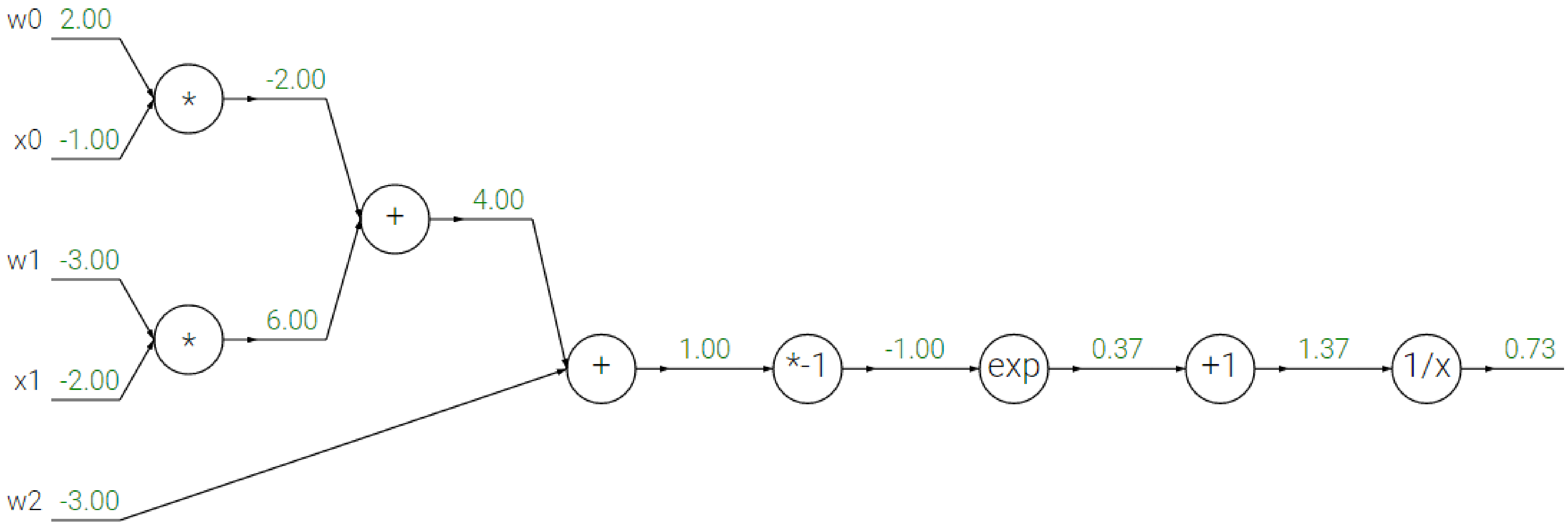


Propagación de gradiente

- Cálculo de gradientes locales mientras forward propagation.

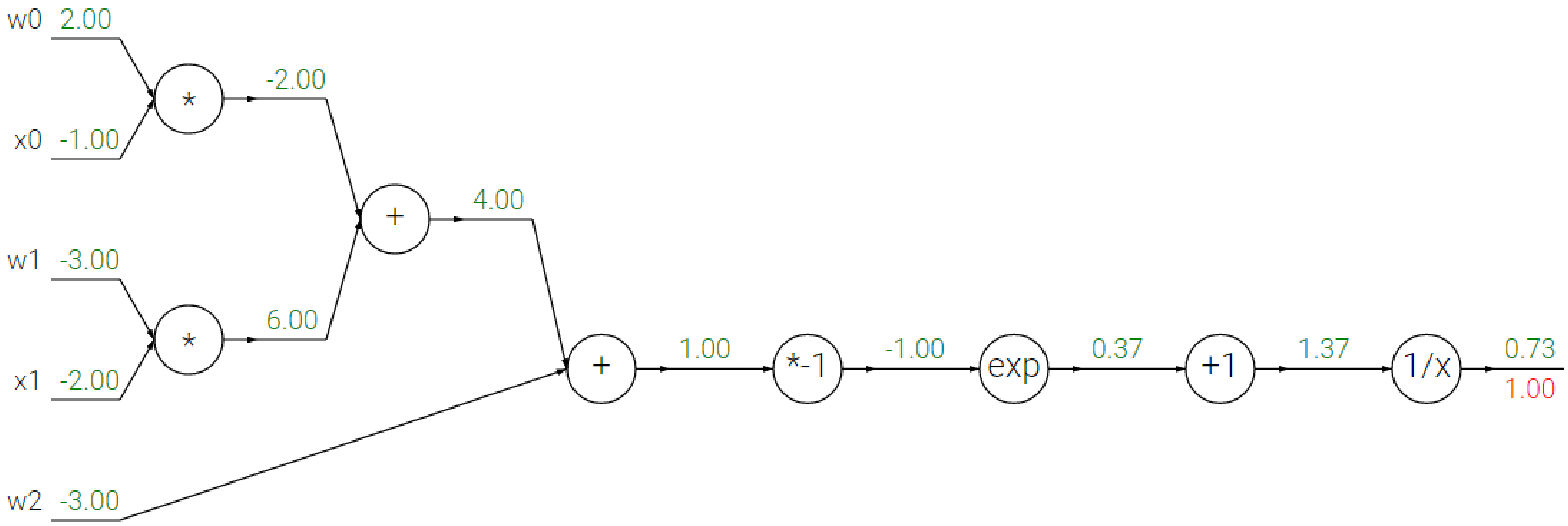


Backpropagation: Grafo de computación



Backpropagation: Grafo de computación

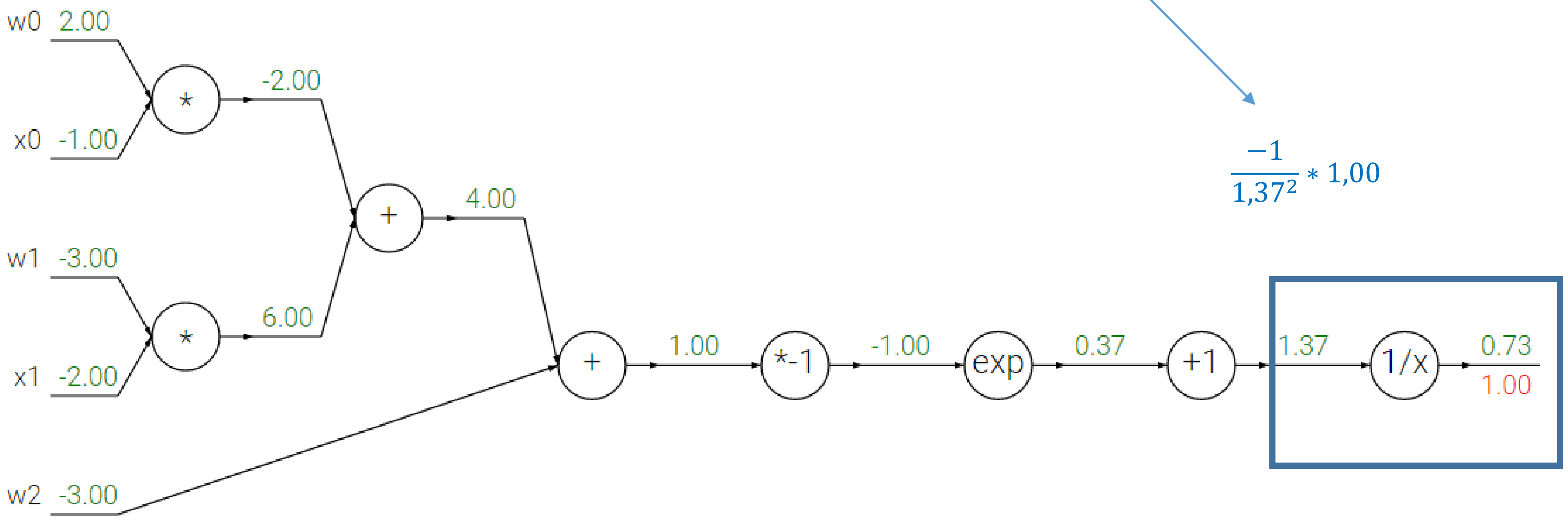
- Ej. con función sigmoide (perceptrón): $f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$
- $f(x) = e^x \rightarrow \frac{\partial f}{\partial x} = e^x$ | $f(x) = ax \rightarrow \frac{\partial f}{\partial x} = a$ | $f(x) = \frac{1}{x} \rightarrow \frac{\partial f}{\partial x} = \frac{-1}{x^2}$ | $f(x) = c + x \rightarrow \frac{\partial f}{\partial x} = 1$



Backpropagation: Grafo de computación

- Ej. con función sigmoide (perceptrón): $f(w,x) = \frac{1}{1+e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$
- $f(x) = e^x \rightarrow \frac{\partial f}{\partial x} = e^x$ | $f(x) = ax \rightarrow \frac{\partial f}{\partial x} = a$ | $f(x) = \frac{1}{x} \rightarrow \frac{\partial f}{\partial x} = \frac{-1}{x^2}$ | $f(x) = c + x \rightarrow \frac{\partial f}{\partial x} = 1$

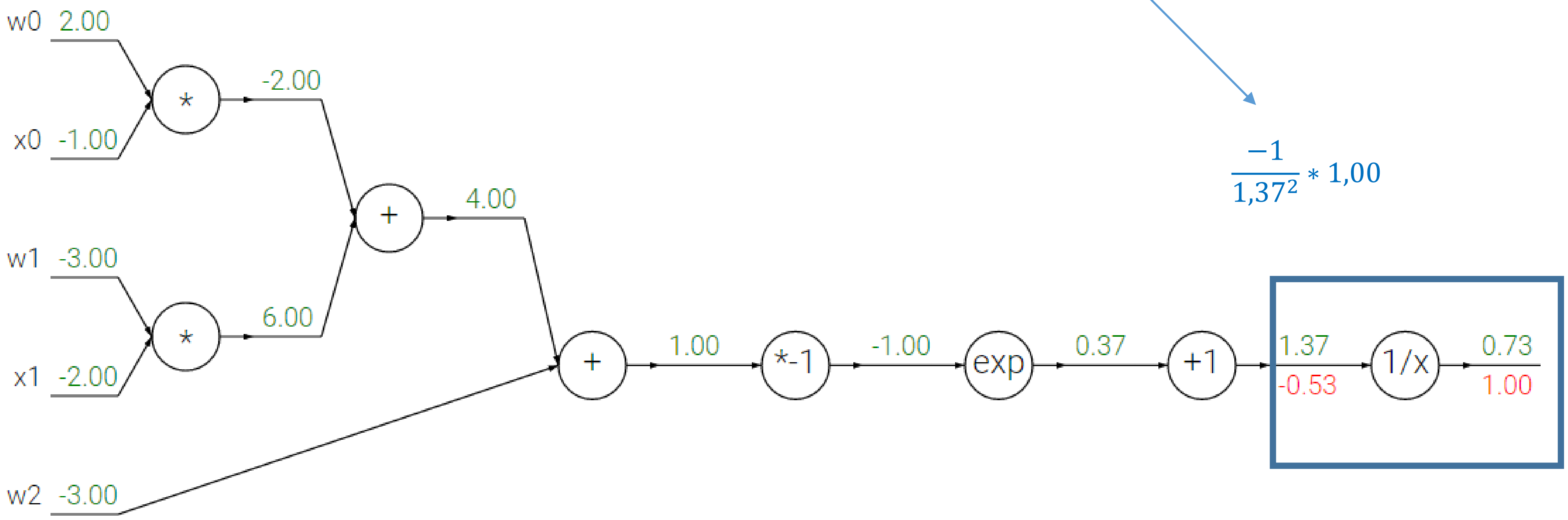
$\frac{-1}{1{,}37^2} * 1{,}00$



Backpropagation: Grafo de computación

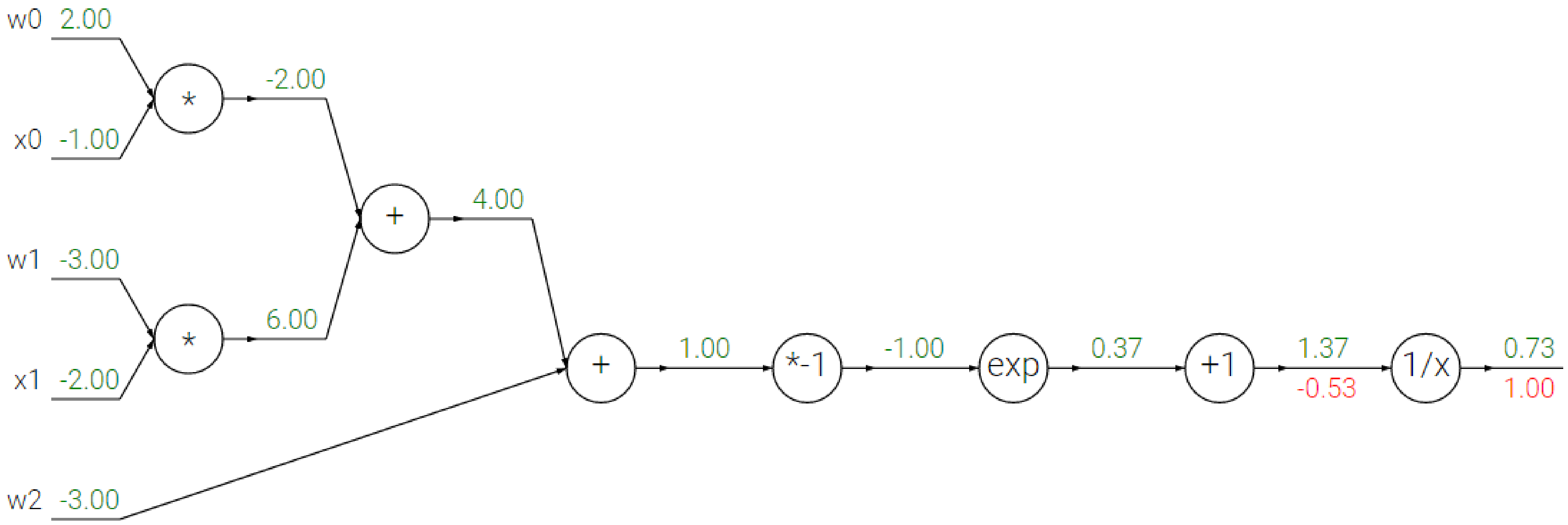
- Ej. con función sigmoide (perceptrón): $f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$
- $f(x) = e^x \rightarrow \frac{\partial f}{\partial x} = e^x$ | $f(x) = ax \rightarrow \frac{\partial f}{\partial x} = a$ | $f(x) = \frac{1}{x} \rightarrow \frac{\partial f}{\partial x} = \frac{-1}{x^2}$ | $f(x) = c + x \rightarrow \frac{\partial f}{\partial x} = 1$

$\frac{-1}{1{,}37^2} * 1{,}00$



Backpropagation: Grafo de computación

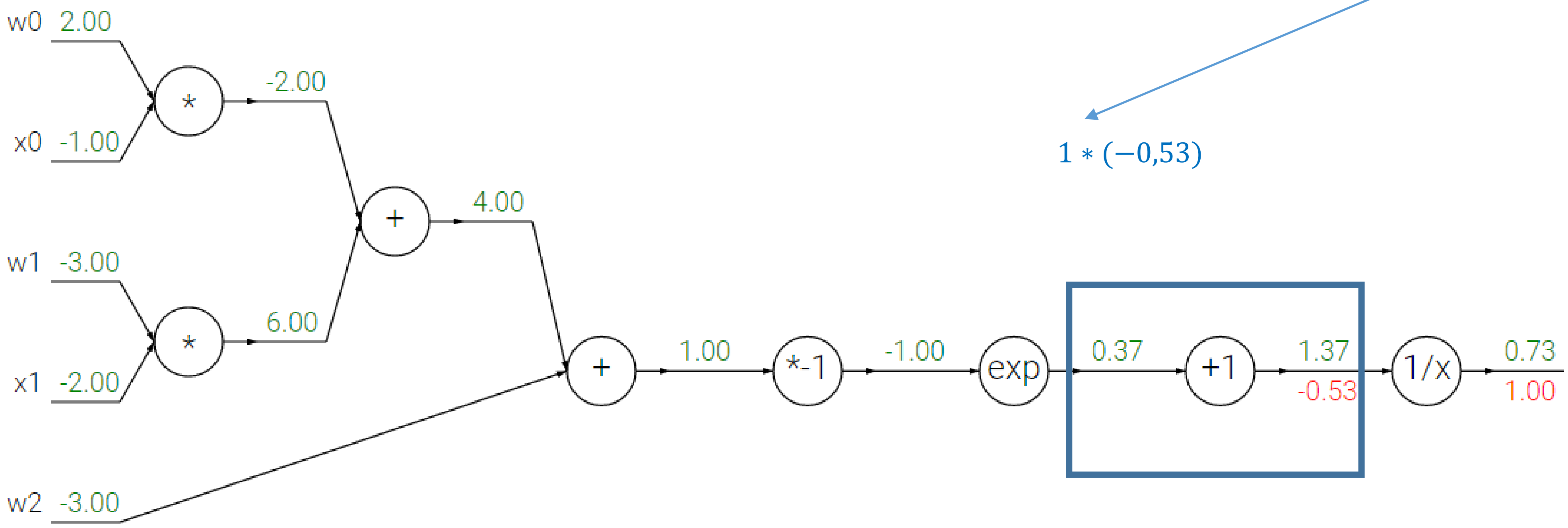
- Ej. con función sigmoide (perceptrón): $f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$
- $f(x) = e^x \rightarrow \frac{\partial f}{\partial x} = e^x$ | $f(x) = ax \rightarrow \frac{\partial f}{\partial x} = a$ | $f(x) = \frac{1}{x} \rightarrow \frac{\partial f}{\partial x} = \frac{-1}{x^2}$ | $f(x) = c + x \rightarrow \frac{\partial f}{\partial x} = 1$



Backpropagation: Grafo de computación

- Ej. con función sigmoide (perceptrón): $f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$
- $f(x) = e^x \rightarrow \frac{\partial f}{\partial x} = e^x$ | $f(x) = ax \rightarrow \frac{\partial f}{\partial x} = a$ | $f(x) = \frac{1}{x} \rightarrow \frac{\partial f}{\partial x} = \frac{-1}{x^2}$ | $f(x) = c + x \rightarrow \frac{\partial f}{\partial x} = 1$

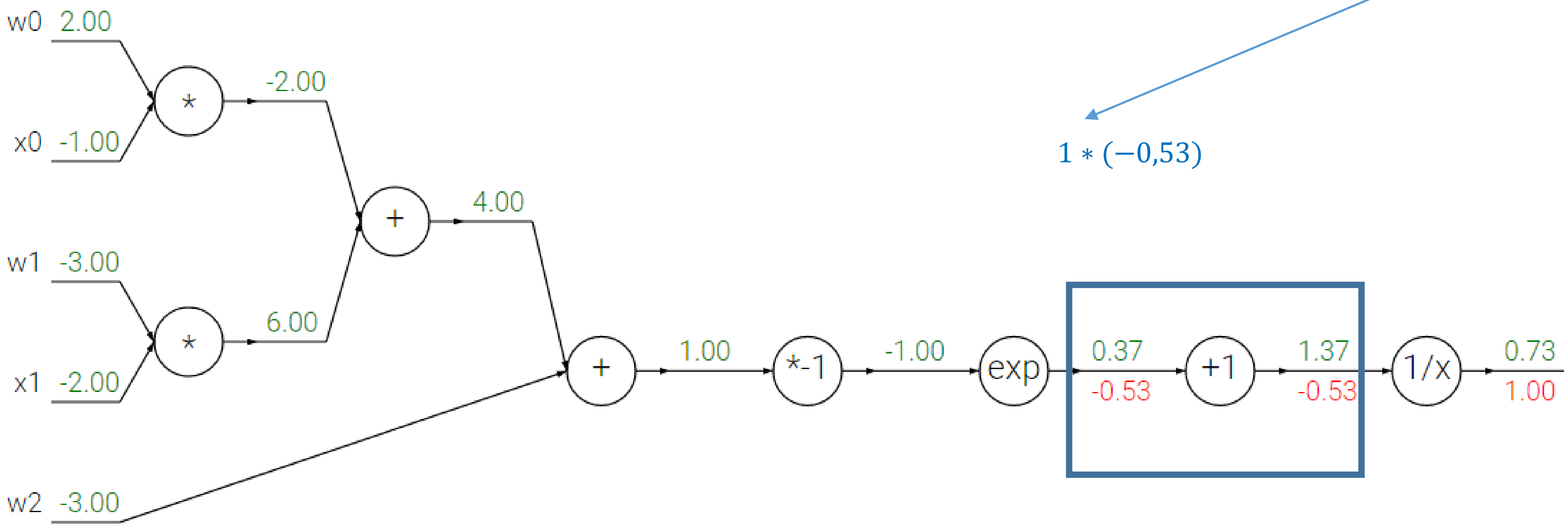
$1 * (-0{,}53)$



Backpropagation: Grafo de computación

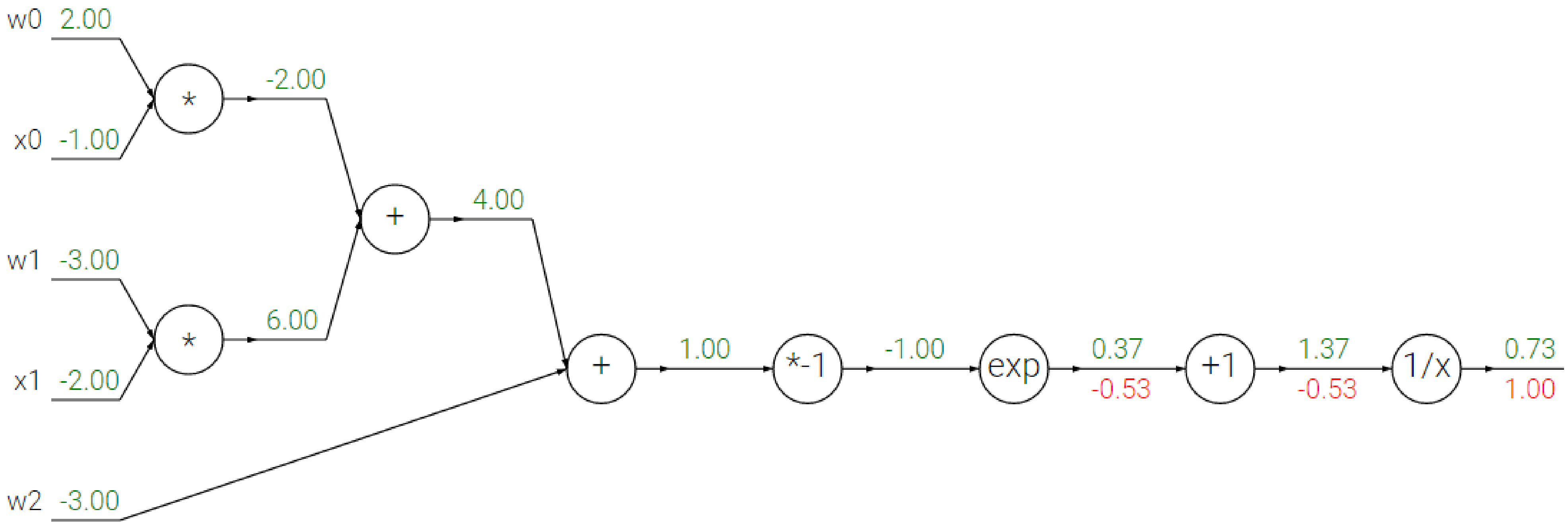
- Ej. con función sigmoide (perceptrón): $f(w,x) = \dfrac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$
- $f(x) = e^x \rightarrow \dfrac{\partial f}{\partial x} = e^x$ | $f(x) = ax \rightarrow \dfrac{\partial f}{\partial x} = a$ | $f(x) = \dfrac{1}{x} \rightarrow \dfrac{\partial f}{\partial x} = \dfrac{-1}{x^2}$ | $f(x) = c + x \rightarrow \dfrac{\partial f}{\partial x} = 1$

$1 * (-0{,}53)$



Backpropagation: Grafo de computación

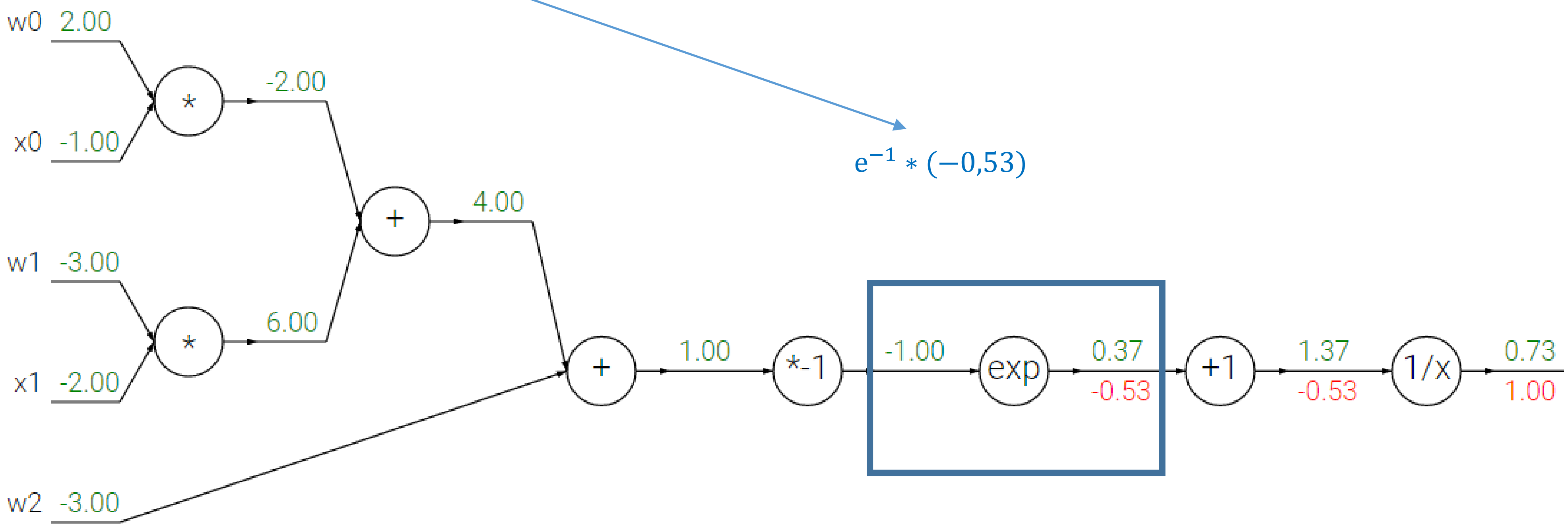
- Ej. con función sigmoide (perceptrón): $f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$
- $f(x) = e^x \rightarrow \frac{\partial f}{\partial x} = e^x$ | $f(x) = ax \rightarrow \frac{\partial f}{\partial x} = a$ | $f(x) = \frac{1}{x} \rightarrow \frac{\partial f}{\partial x} = \frac{-1}{x^2}$ | $f(x) = c + x \rightarrow \frac{\partial f}{\partial x} = 1$



Backpropagation: Grafo de computación

- Ej. con función sigmoide (perceptrón): $f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$
- $f(x) = e^x \rightarrow \frac{\partial f}{\partial x} = e^x$ | $f(x) = ax \rightarrow \frac{\partial f}{\partial x} = a$ | $f(x) = \frac{1}{x} \rightarrow \frac{\partial f}{\partial x} = \frac{-1}{x^2}$ | $f(x) = c + x \rightarrow \frac{\partial f}{\partial x} = 1$

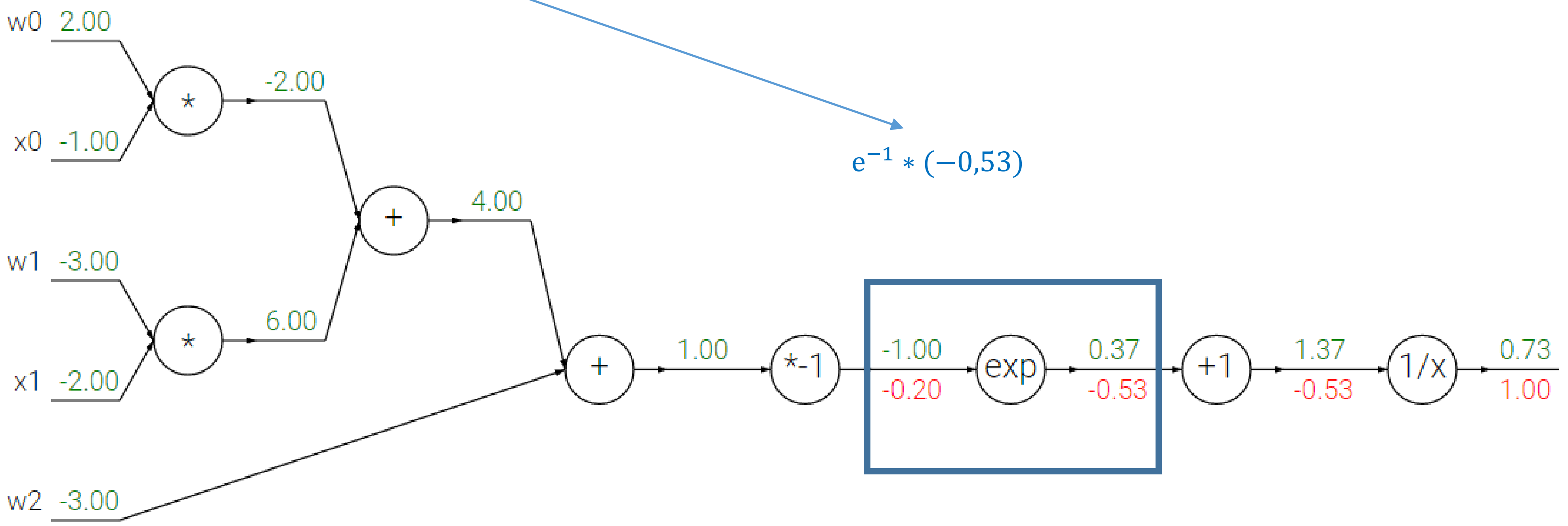
$\mathrm{e}^{-1} * (-0{,}53)$



Backpropagation: Grafo de computación

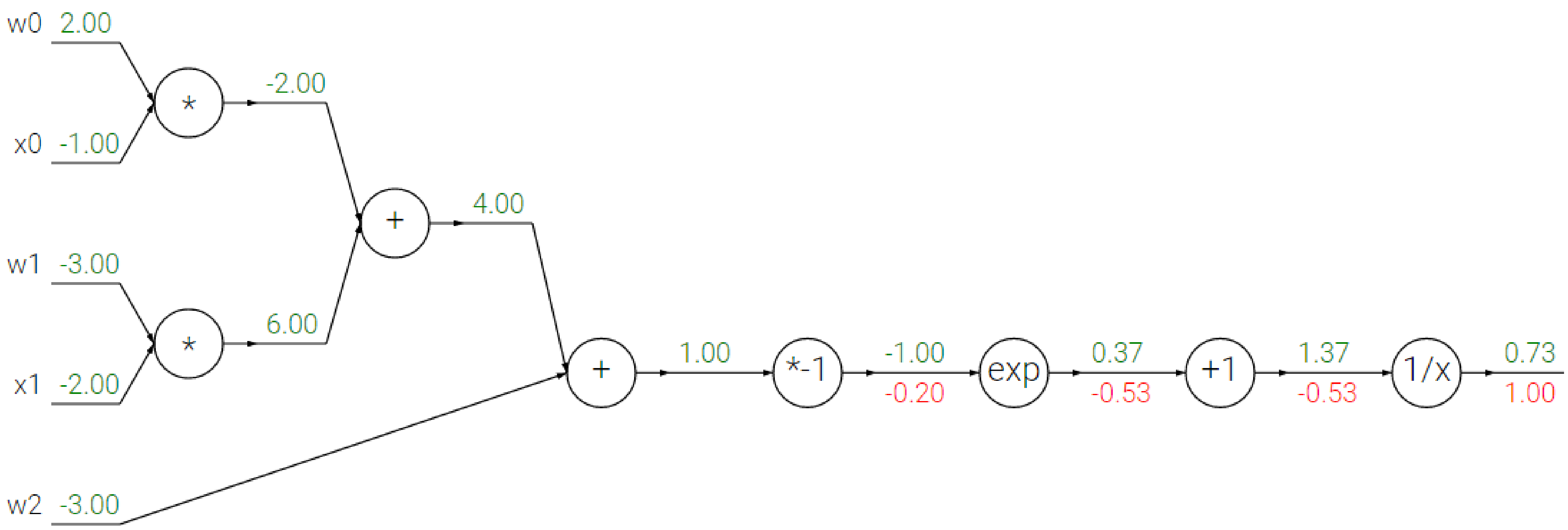
- Ej. con función sigmoide (perceptrón): $f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$
- $f(x) = e^x \rightarrow \frac{\partial f}{\partial x} = e^x$ | $f(x) = ax \rightarrow \frac{\partial f}{\partial x} = a$ | $f(x) = \frac{1}{x} \rightarrow \frac{\partial f}{\partial x} = \frac{-1}{x^2}$ | $f(x) = c + x \rightarrow \frac{\partial f}{\partial x} = 1$

$e^{-1} * (-0{,}53)$



Backpropagation: Grafo de computación

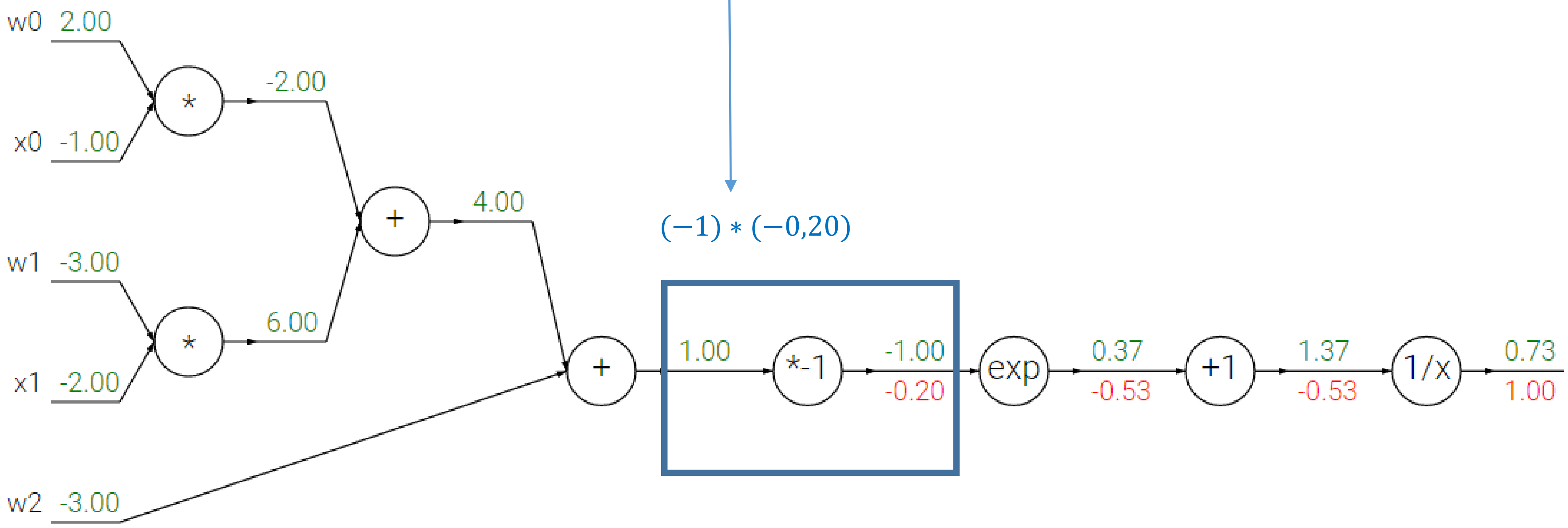
- Ej. con función sigmoide (perceptrón): $f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$
- $f(x) = e^x \rightarrow \frac{\partial f}{\partial x} = e^x$ | $f(x) = ax \rightarrow \frac{\partial f}{\partial x} = a$ | $f(x) = \frac{1}{x} \rightarrow \frac{\partial f}{\partial x} = \frac{-1}{x^2}$ | $f(x) = c + x \rightarrow \frac{\partial f}{\partial x} = 1$



Backpropagation: Grafo de computación

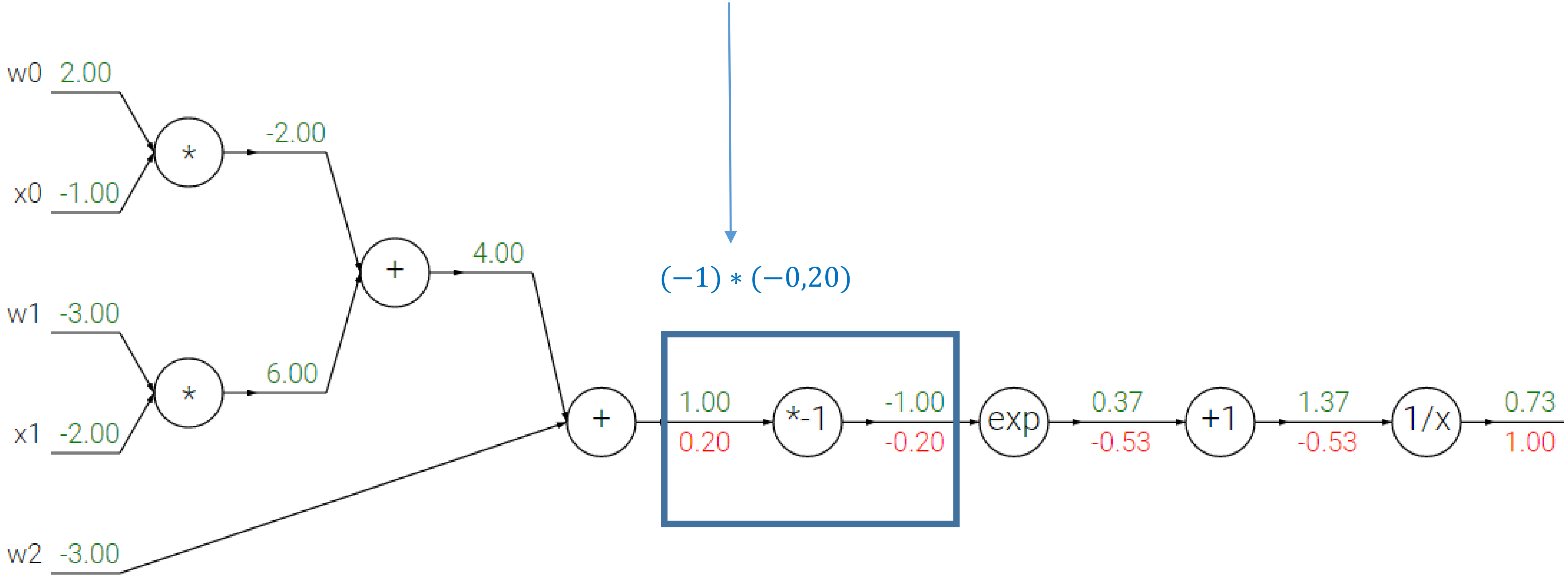
- Ej. con función sigmoide (perceptrón): $f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$
- $f(x) = e^x \rightarrow \frac{\partial f}{\partial x} = e^x$ | $f(x) = ax \rightarrow \frac{\partial f}{\partial x} = a$ | $f(x) = \frac{1}{x} \rightarrow \frac{\partial f}{\partial x} = \frac{-1}{x^2}$ | $f(x) = c + x \rightarrow \frac{\partial f}{\partial x} = 1$

$(-1) * (-0{,}20)$



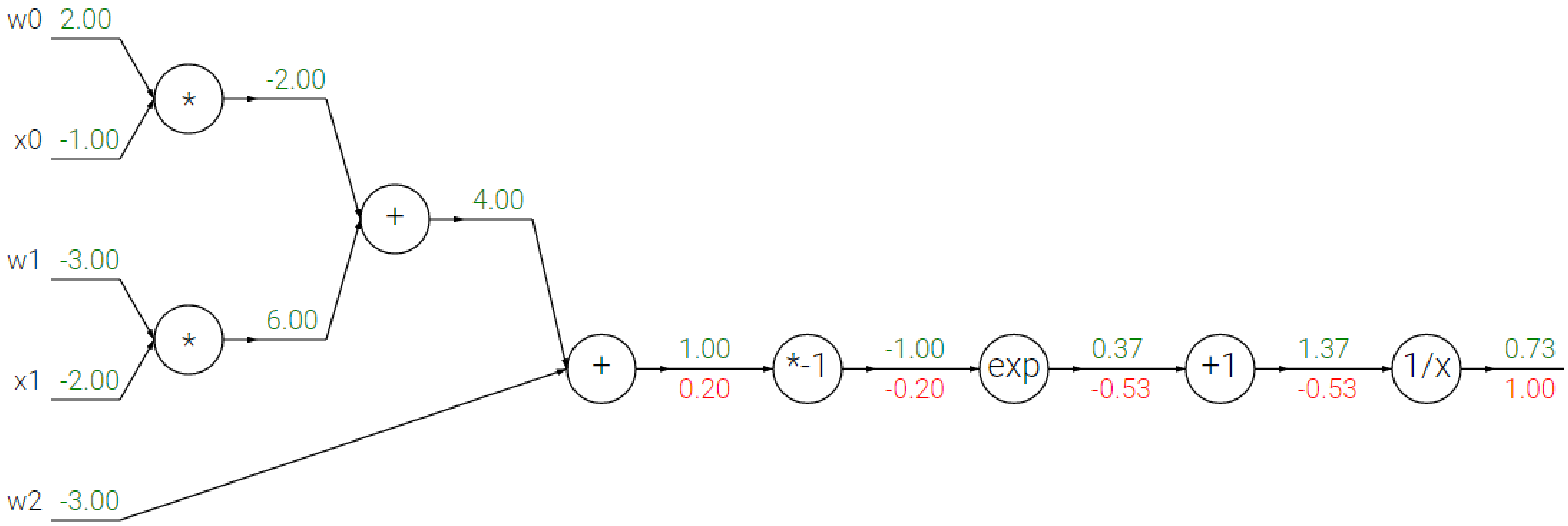
Backpropagation: Grafo de computación

- Ej. con función sigmoide (perceptrón): $f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$
- $f(x) = e^x \rightarrow \frac{\partial f}{\partial x} = e^x$ | $f(x) = ax \rightarrow \frac{\partial f}{\partial x} = a$ | $f(x) = \frac{1}{x} \rightarrow \frac{\partial f}{\partial x} = \frac{-1}{x^2}$ | $f(x) = c + x \rightarrow \frac{\partial f}{\partial x} = 1$



Backpropagation: Grafo de computación

- Ej. con función sigmoide (perceptrón): $f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$
- $f(x) = e^x \rightarrow \frac{\partial f}{\partial x} = e^x$ | $f(x) = ax \rightarrow \frac{\partial f}{\partial x} = a$ | $f(x) = \frac{1}{x} \rightarrow \frac{\partial f}{\partial x} = \frac{-1}{x^2}$ | $f(x) = c + x \rightarrow \frac{\partial f}{\partial x} = 1$



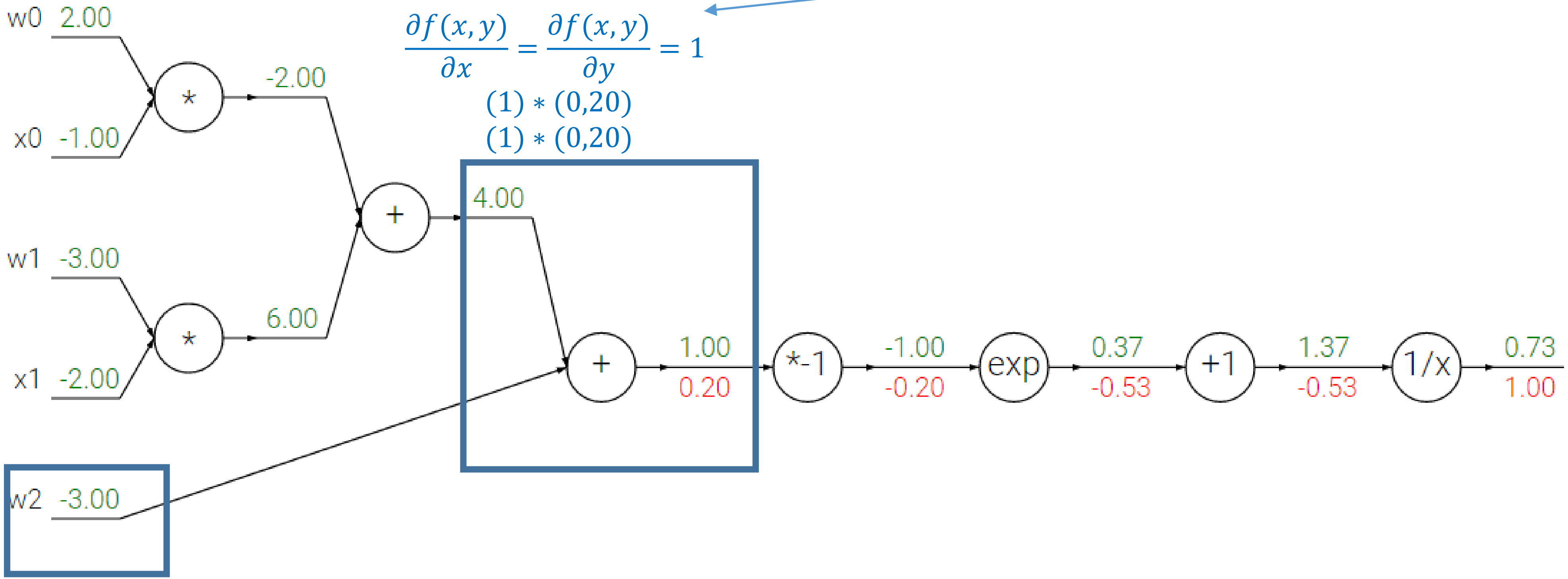
Backpropagation: Grafo de computación

- Ej. con función sigmoide (perceptrón): $f(w,x) = \frac{1}{1+e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$
- $f(x) = e^x \rightarrow \frac{\partial f}{\partial x} = e^x$ | $f(x) = ax \rightarrow \frac{\partial f}{\partial x} = a$ | $f(x) = \frac{1}{x} \rightarrow \frac{\partial f}{\partial x} = \frac{-1}{x^2}$ | $f(x) = c + x \rightarrow \frac{\partial f}{\partial x} = 1$

$$\frac{\partial f(x,y)}{\partial x} = \frac{\partial f(x,y)}{\partial y} = 1$$

$(1) * (0{,}20)$

$(1) * (0{,}20)$



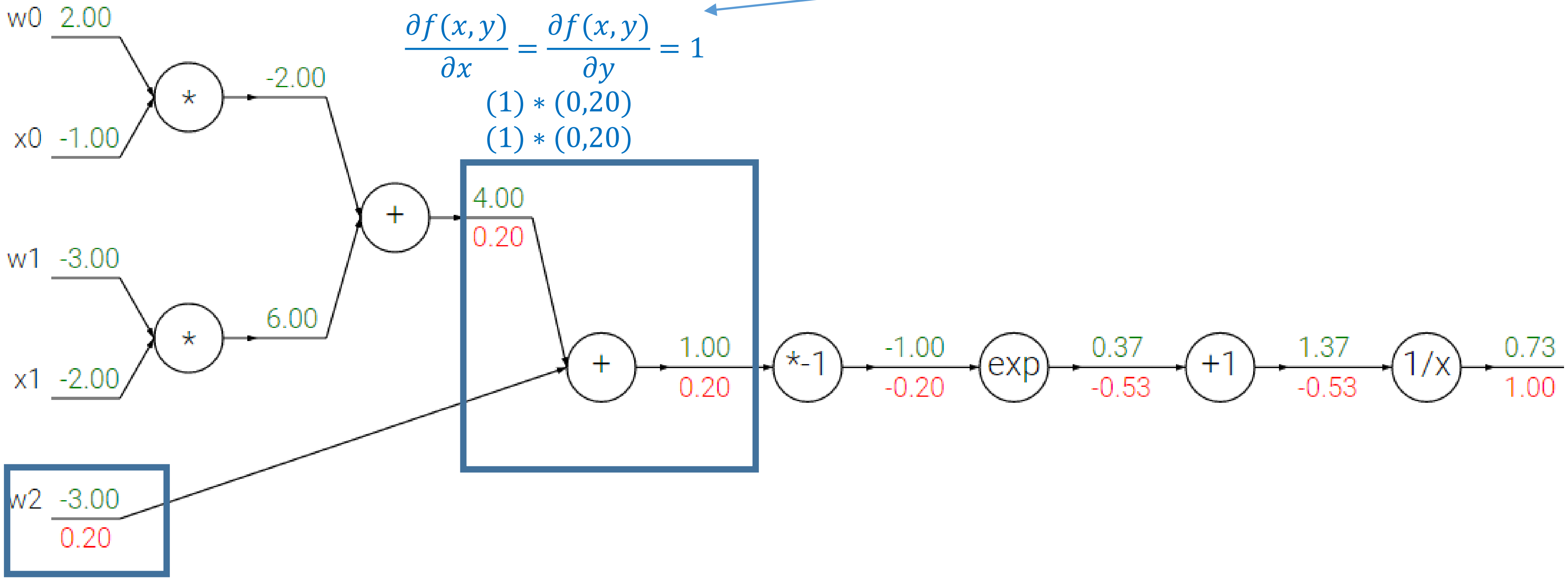
Backpropagation: Grafo de computación

- Ej. con función sigmoide (perceptrón): $f(w,x) = \frac{1}{1+e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$
- $f(x) = e^x \rightarrow \frac{\partial f}{\partial x} = e^x$ | $f(x) = ax \rightarrow \frac{\partial f}{\partial x} = a$ | $f(x) = \frac{1}{x} \rightarrow \frac{\partial f}{\partial x} = \frac{-1}{x^2}$ | $f(x) = c + x \rightarrow \frac{\partial f}{\partial x} = 1$

$$\frac{\partial f(x,y)}{\partial x} = \frac{\partial f(x,y)}{\partial y} = 1$$

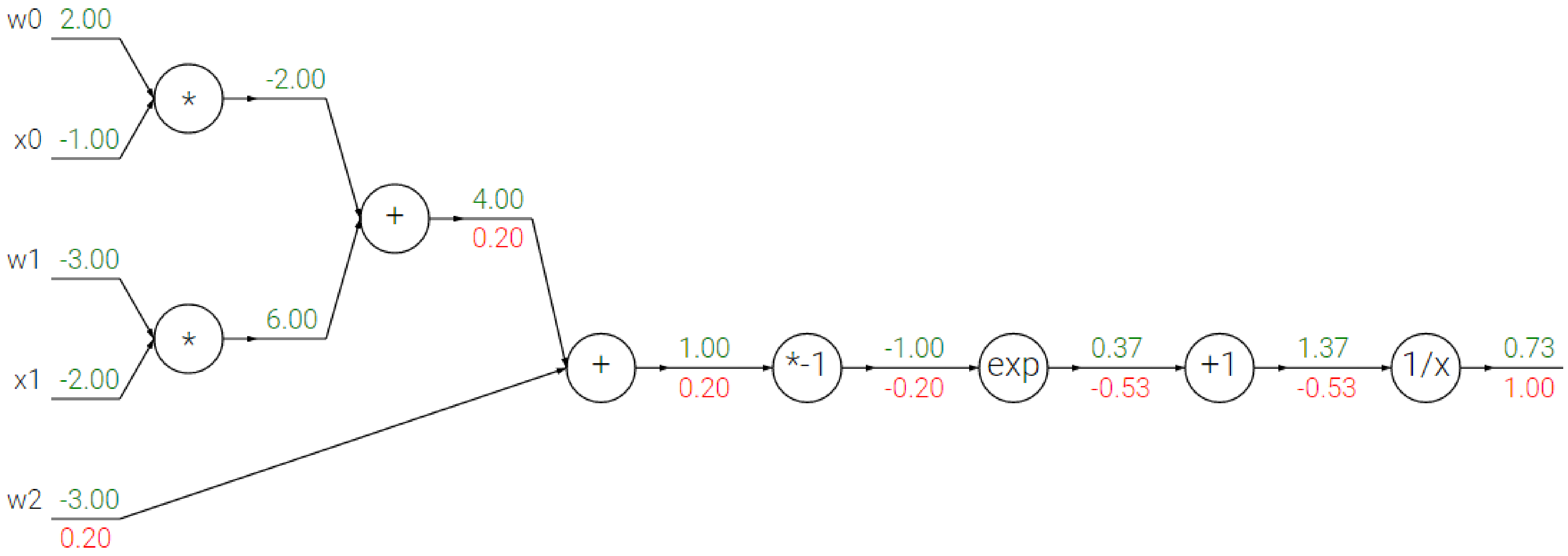
$$(1) * (0{,}20)$$

$$(1) * (0{,}20)$$



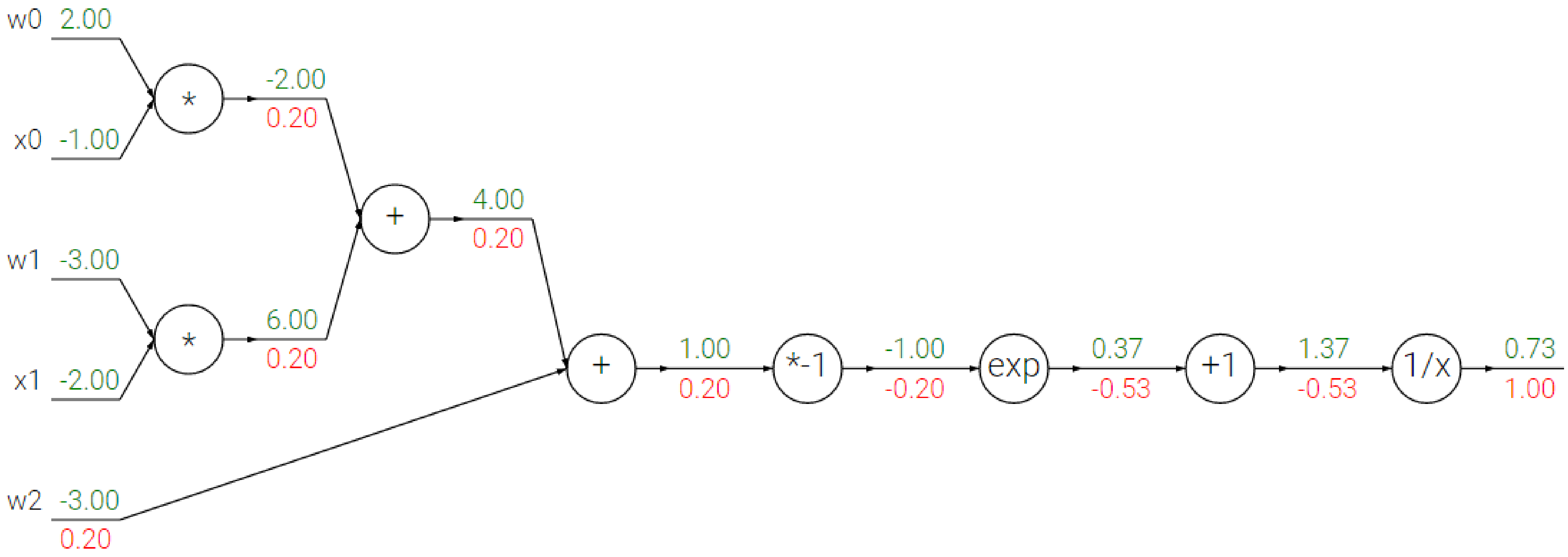
Backpropagation: Grafo de computación

- Ej. con función sigmoide (perceptrón): $f(w,x) = \dfrac{1}{1+e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$
- $f(x) = e^x \rightarrow \dfrac{\partial f}{\partial x} = e^x$ | $f(x) = ax \rightarrow \dfrac{\partial f}{\partial x} = a$ | $f(x) = \dfrac{1}{x} \rightarrow \dfrac{\partial f}{\partial x} = \dfrac{-1}{x^2}$ | $f(x) = c + x \rightarrow \dfrac{\partial f}{\partial x} = 1$



Backpropagation: Grafo de computación

- Ej. con función sigmoide (perceptrón): $f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$
- $f(x) = e^x \rightarrow \frac{\partial f}{\partial x} = e^x$ | $f(x) = ax \rightarrow \frac{\partial f}{\partial x} = a$ | $f(x) = \frac{1}{x} \rightarrow \frac{\partial f}{\partial x} = \frac{-1}{x^2}$ | $f(x) = c + x \rightarrow \frac{\partial f}{\partial x} = 1$

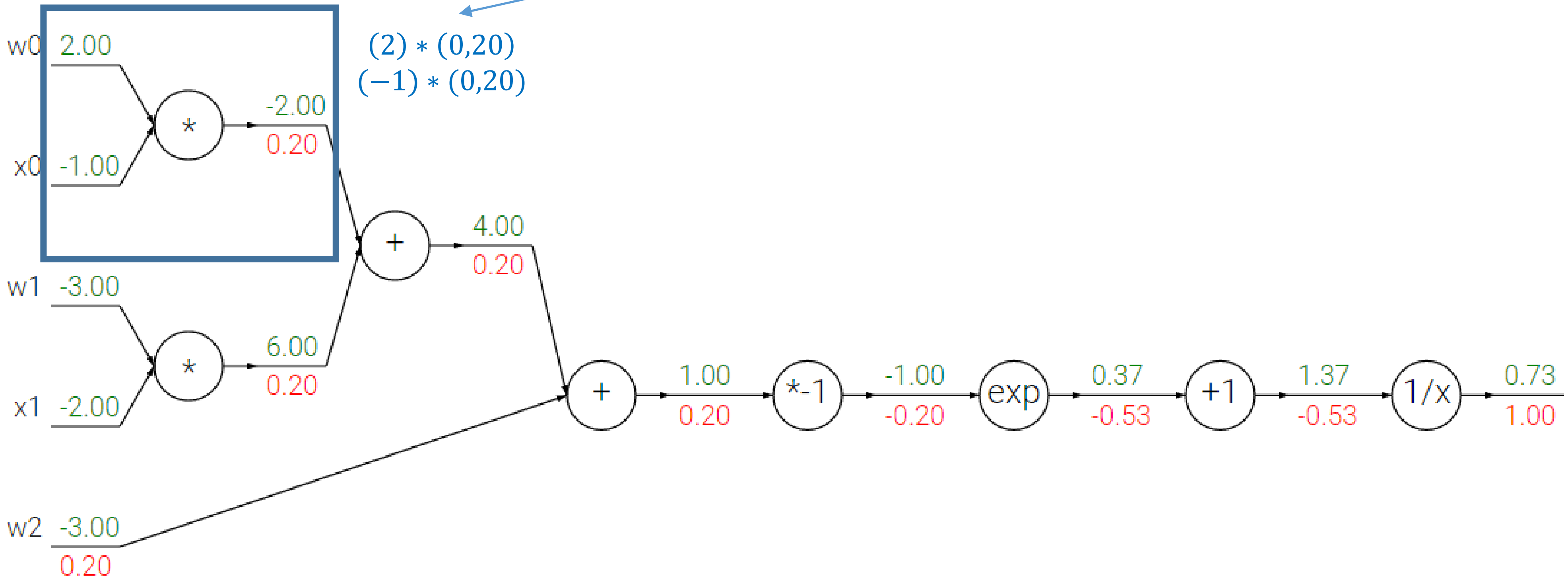


Backpropagation: Grafo de computación

- Ej. con función sigmoide (perceptrón): $f(w,x) = \dfrac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$
- $f(x) = e^x \rightarrow \dfrac{\partial f}{\partial x} = e^x$ | $f(x) = ax \rightarrow \dfrac{\partial f}{\partial x} = a$ | $f(x) = \dfrac{1}{x} \rightarrow \dfrac{\partial f}{\partial x} = \dfrac{-1}{x^2}$ | $f(x) = c + x \rightarrow \dfrac{\partial f}{\partial x} = 1$

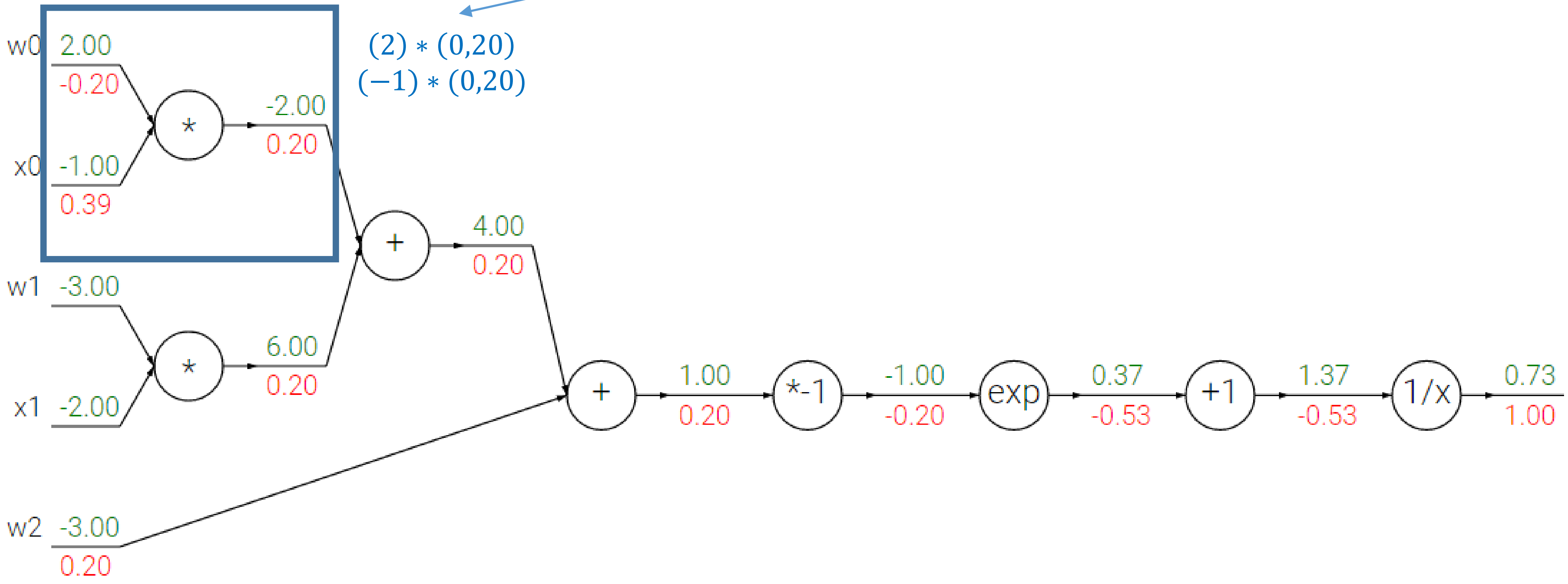
$(2) * (0{,}20)$

$(-1) * (0{,}20)$



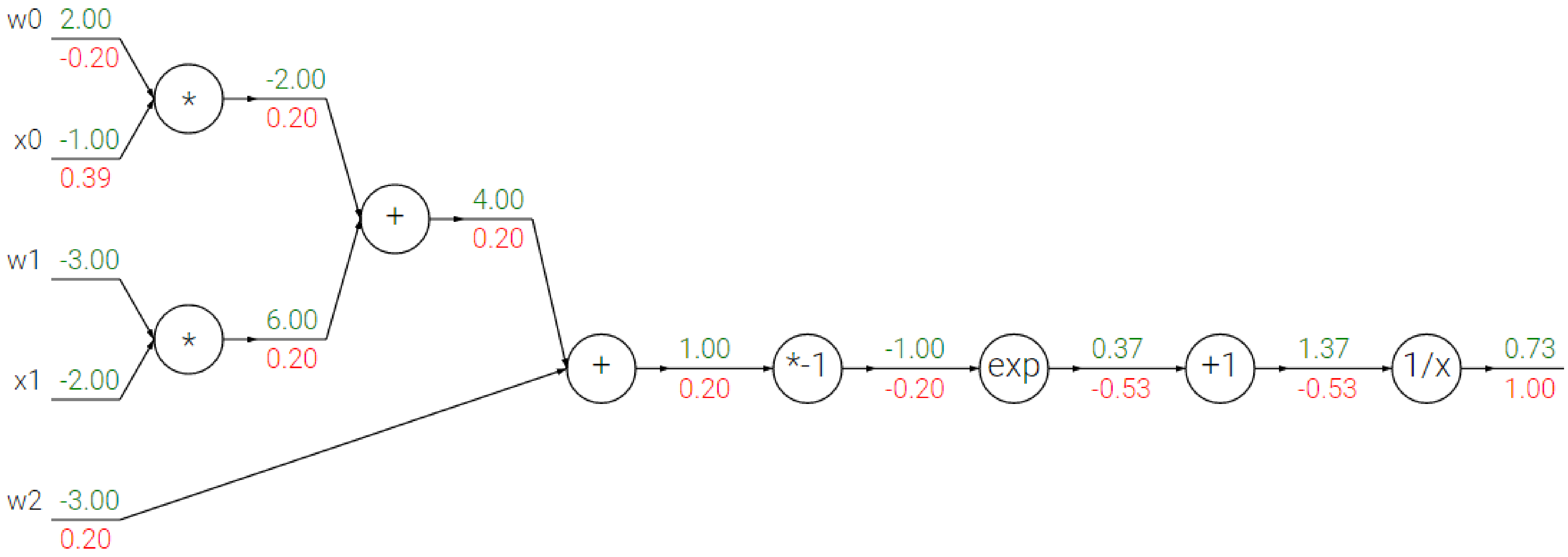
Backpropagation: Grafo de computación

- Ej. con función sigmoide (perceptrón): $f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$
- $f(x) = e^x \rightarrow \frac{\partial f}{\partial x} = e^x$ | $f(x) = ax \rightarrow \frac{\partial f}{\partial x} = a$ | $f(x) = \frac{1}{x} \rightarrow \frac{\partial f}{\partial x} = \frac{-1}{x^2}$ | $f(x) = c + x \rightarrow \frac{\partial f}{\partial x} = 1$



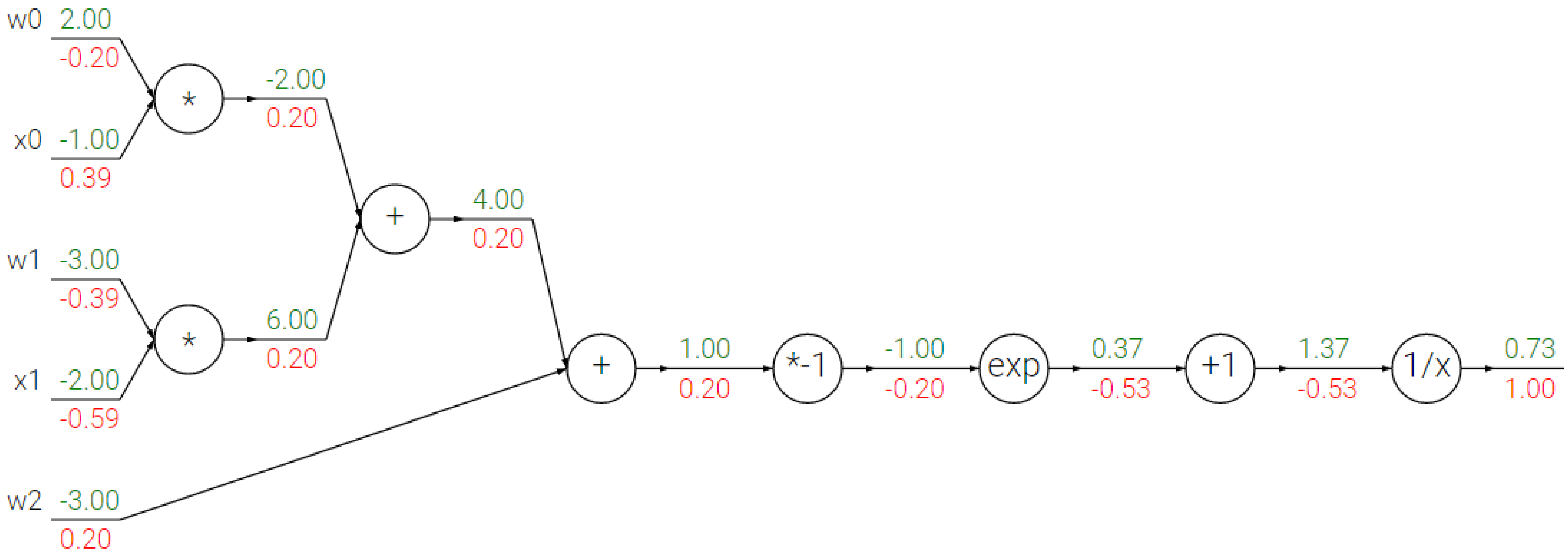
Backpropagation: Grafo de computación

- Ej. con función sigmoide (perceptrón): $f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$
- $f(x) = e^x \rightarrow \frac{\partial f}{\partial x} = e^x$ | $f(x) = ax \rightarrow \frac{\partial f}{\partial x} = a$ | $f(x) = \frac{1}{x} \rightarrow \frac{\partial f}{\partial x} = \frac{-1}{x^2}$ | $f(x) = c + x \rightarrow \frac{\partial f}{\partial x} = 1$



Backpropagation: Grafo de computación

- Ej. con función sigmoide (perceptrón): $f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$
- $f(x) = e^x \rightarrow \frac{\partial f}{\partial x} = e^x$ | $f(x) = ax \rightarrow \frac{\partial f}{\partial x} = a$ | $f(x) = \frac{1}{x} \rightarrow \frac{\partial f}{\partial x} = \frac{-1}{x^2}$ | $f(x) = c + x \rightarrow \frac{\partial f}{\partial x} = 1$

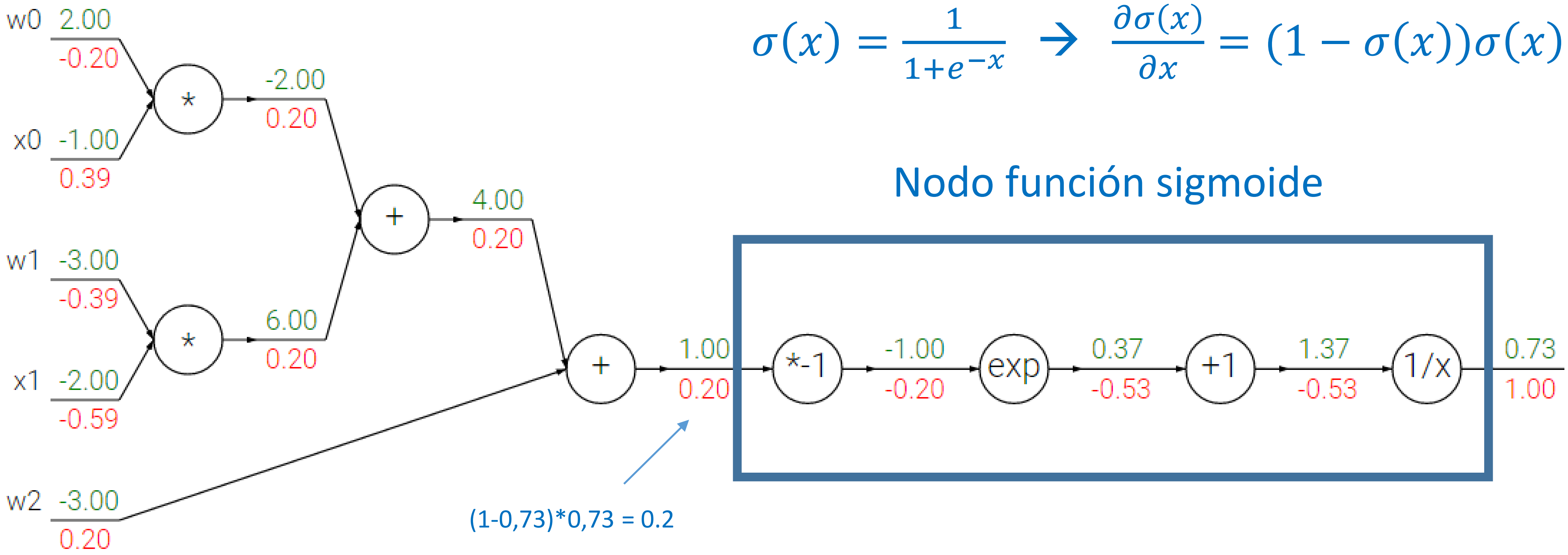


Backpropagation: Grafo de computación

- Se pueden definir nodos para funciones conocidas para ahorrar pasos de computación.

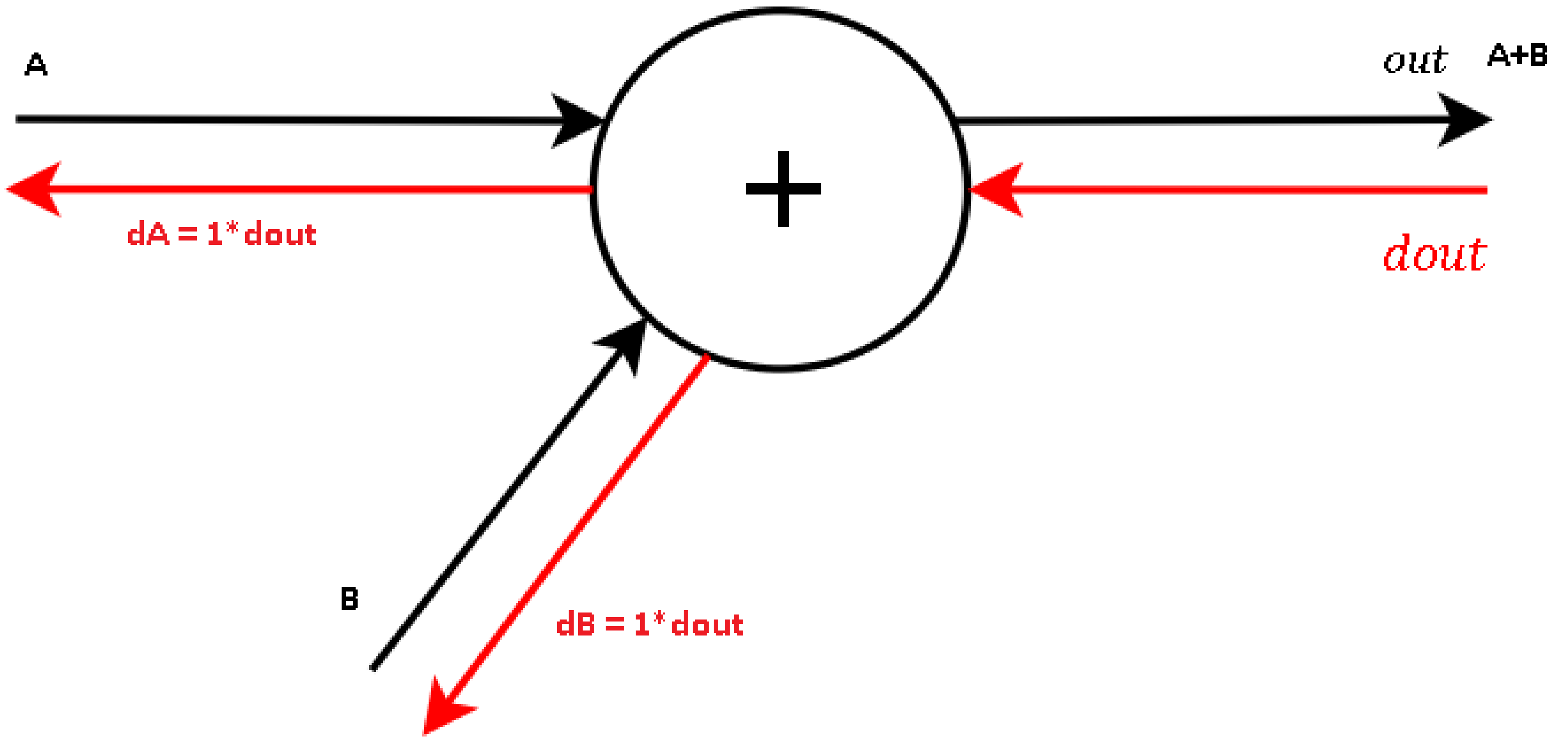
$$\sigma(x) = \frac{1}{1+e^{-x}} \rightarrow \frac{\partial \sigma(x)}{\partial x} = (1-\sigma(x))\sigma(x)$$

Nodo función sigmoide



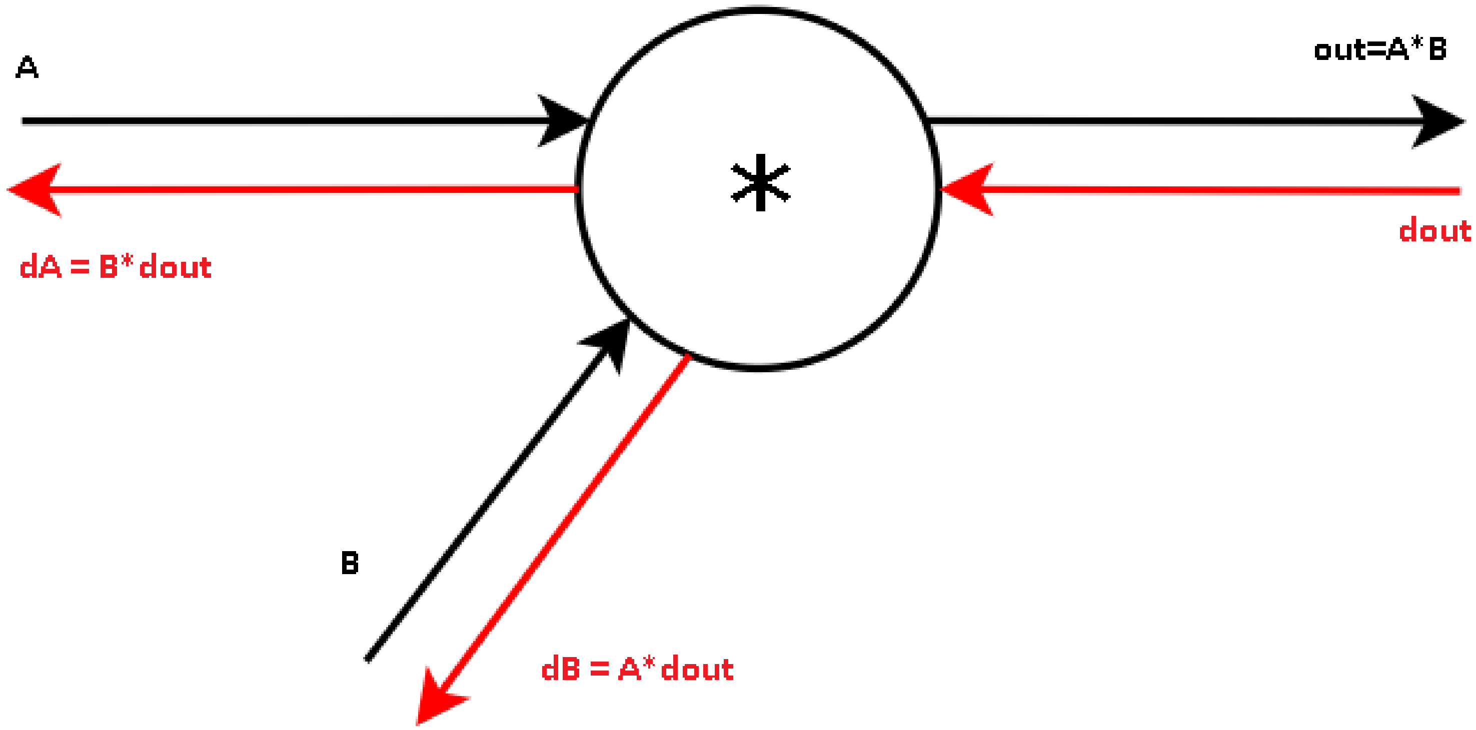
Backpropagation: Grafo de computación

- Bloques básicos
  - Suma (distribuidor de gradiente)



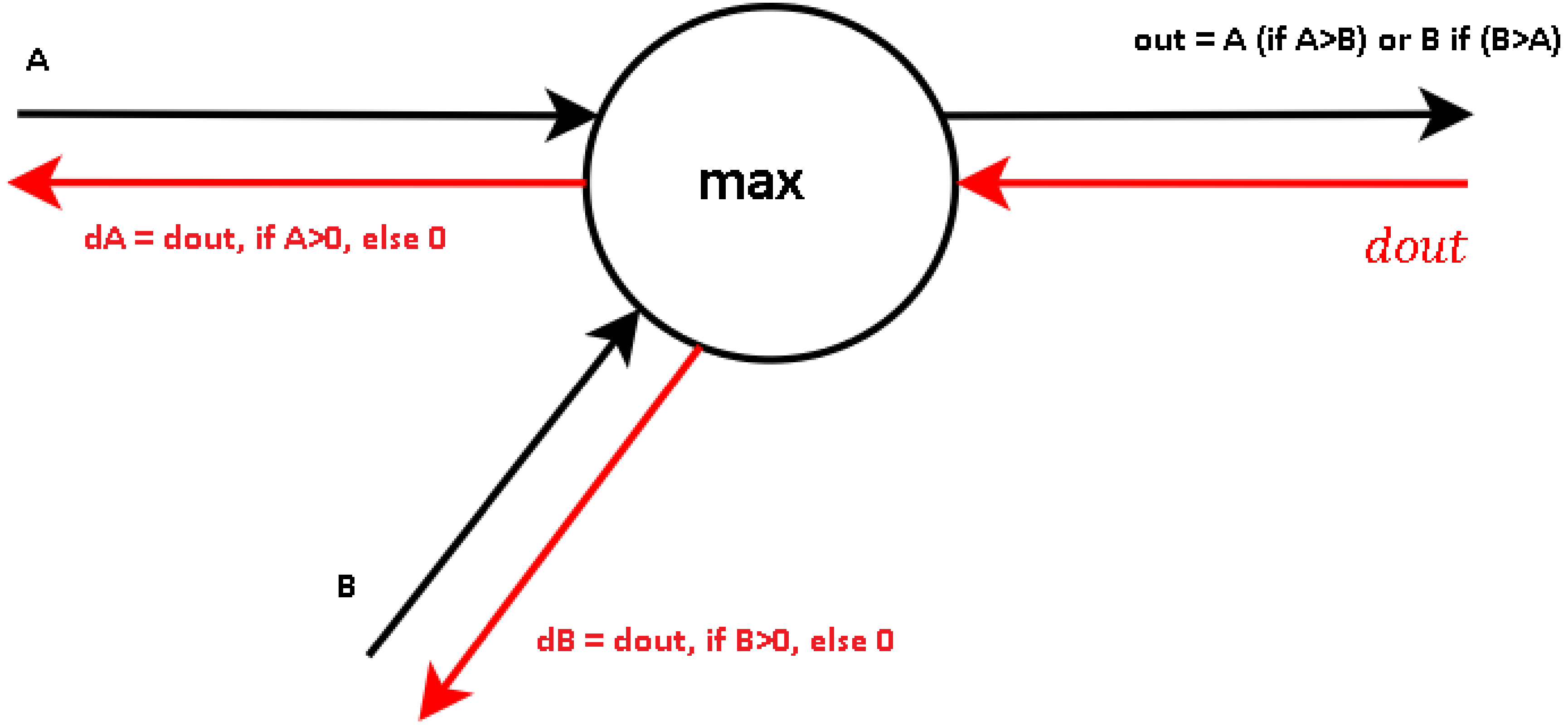
Backpropagation: Grafo de computación

- Bloques básicos
  - Multiplicación (intercambiador de gradiente)



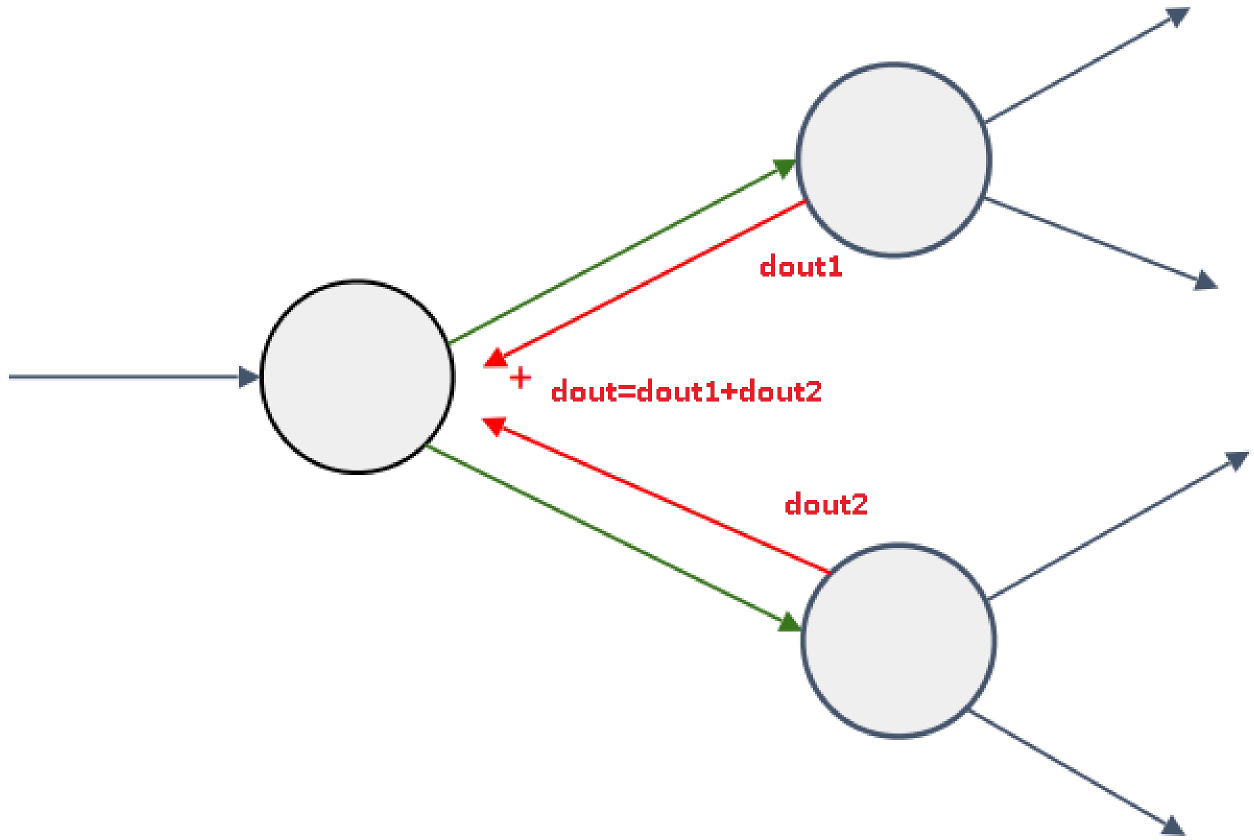
Backpropagation: Grafo de computación

- Bloques básicos
  - Máximo (enrutador de gradiente)



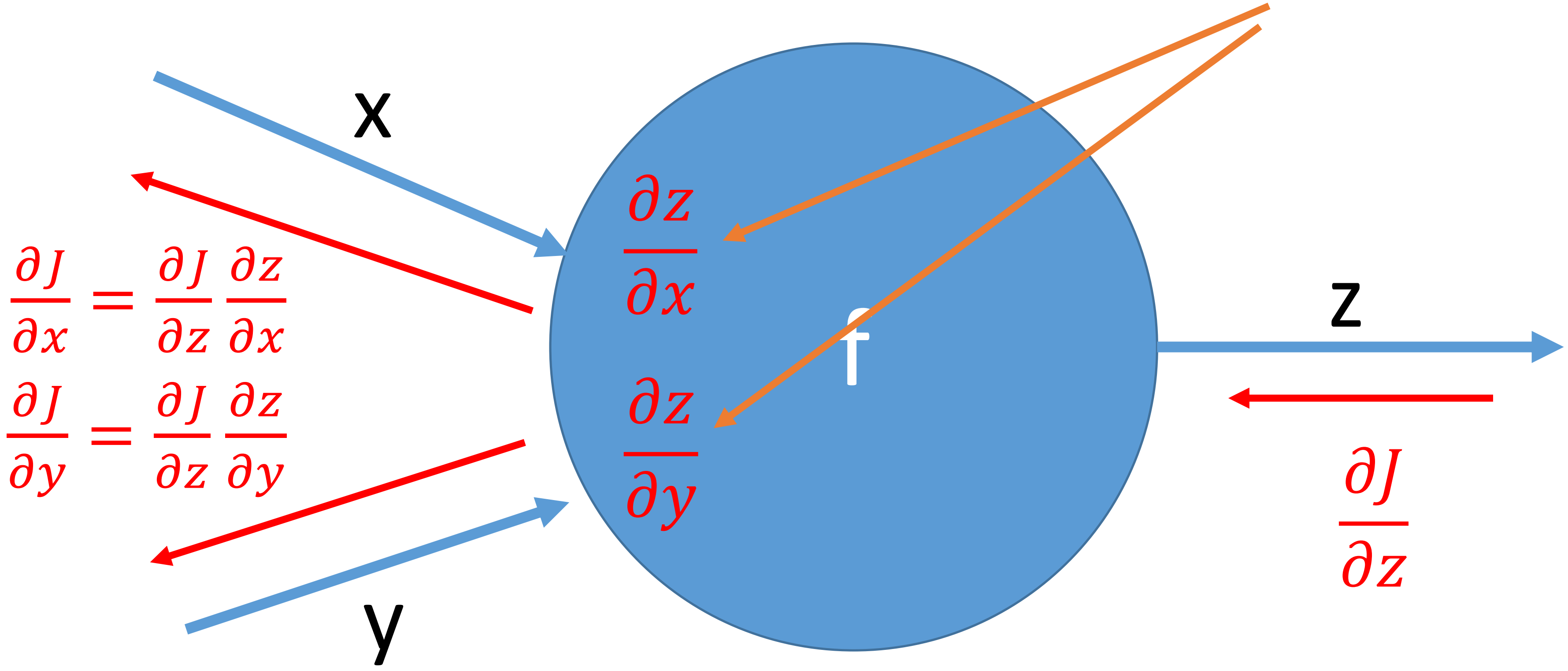
Backpropagation: Grafo de computación

- Bloques básicos
  - Ramas (sumador de gradientes)



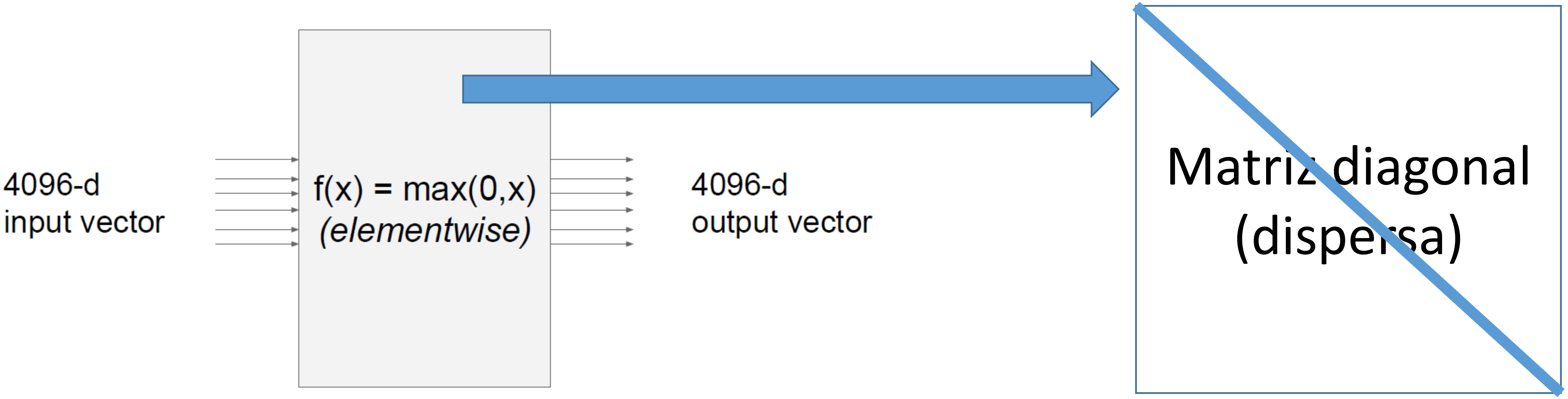
Propagación de gradiente con vectorización

- ¿Qué ocurre cuando las datos son vectores? Matrices jacobianas.



Propagación de gradiente con vectorización

- ¿Qué ocurre cuando las datos son vectores? Matrices jacobianas.



Recapitulación

- La base del algoritmo de backpropagation es la **regla de la cadena**.
- La **propagación del gradiente** es la clave para actualizar los pesos en las **capas ocultas**.
- Si vemos la red como un **grafo computacional**, podemos entender la intuitivamente cómo funciona el algoritmo.
- Es importante entender cómo se propagan los gradientes para evitar el **problema del desvanecimiento del gradiente**.