Neural networks practice

week 3

Soma Kontár

Where we are...

- we have used multiple layers (eg. learning XOR)
- we have formulated and used cost functions
- we have trained perceptrons and simple neural networks
- we have done all this in closed form (normal equations)
- what if we can't / don't want to optimize in closed form?

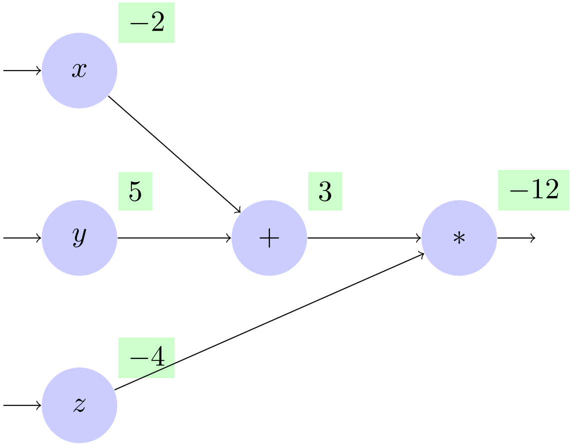
Back-propagation

- **forward propagation (forward pass)**: we feed inputs to our computational graph, and after the input data has flowed through the whole graph, it produces a scalar cost
- **back-propagation (backwards pass)**: from the scalar cost we calculate the gradient of each node $\rightarrow$ how much each element affects the output

A simple example

Forward propagation

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



A simple example

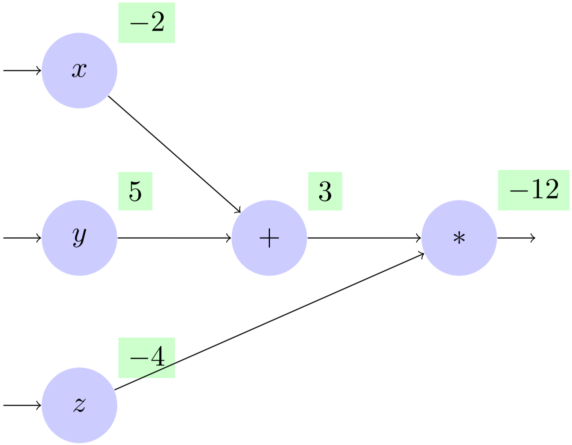
Back-propagation

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

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

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

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



A simple example

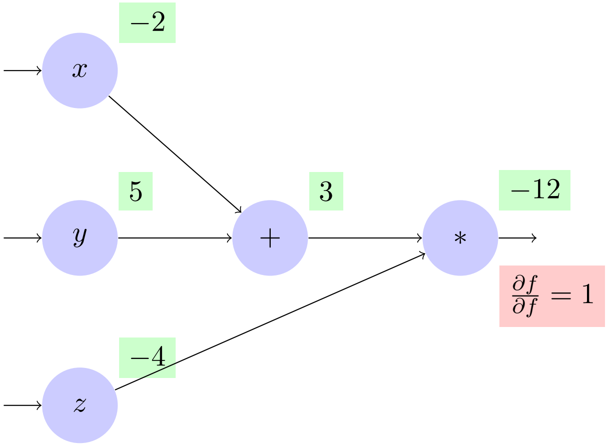
Back-propagation

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

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

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

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



A simple example

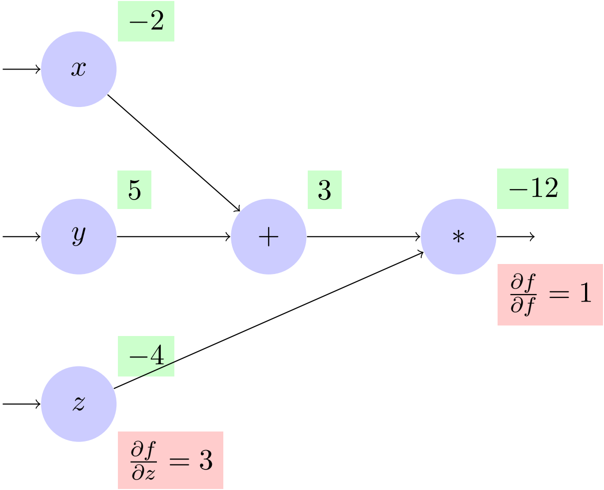
Back-propagation

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

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

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

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



A simple example

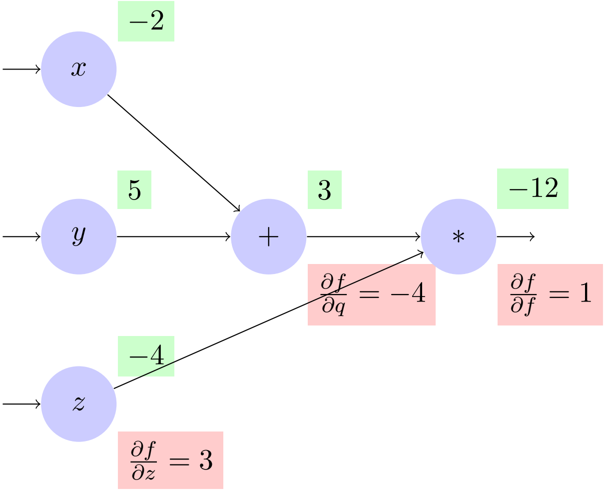
Back-propagation

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

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

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

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



A simple example

Back-propagation

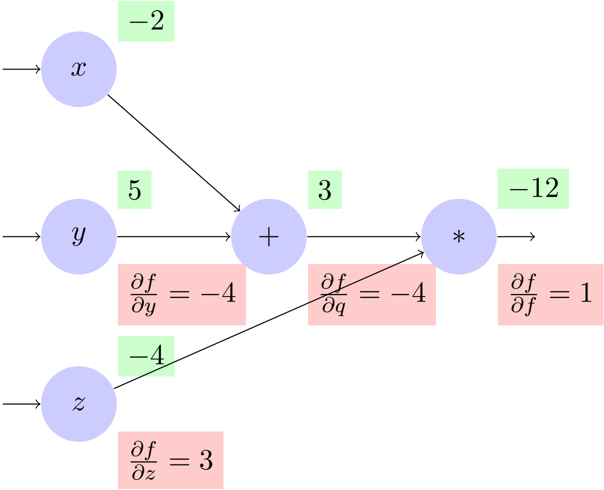
$$f(x, y, z) = (x + y)z$$

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

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

Chain rule: $\dfrac{\partial f}{\partial y} = \dfrac{\partial f}{\partial q}\dfrac{\partial q}{\partial y}$

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



A simple example

Back-propagation

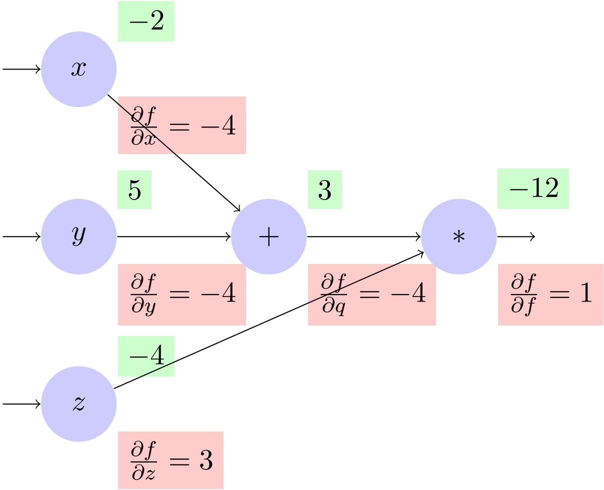
$$f(x, y, z) = (x + y)z$$

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

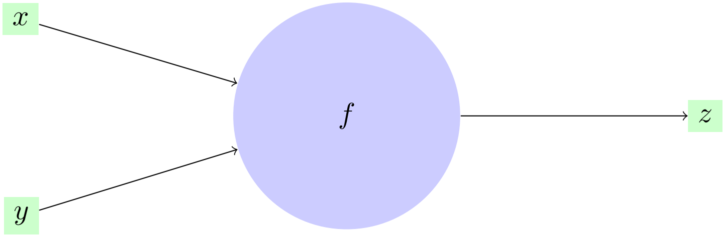
$$f = qz \qquad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Chain rule: $\frac{\partial f}{\partial y} = \frac{\partial f}{\partial q}\frac{\partial q}{\partial y}$

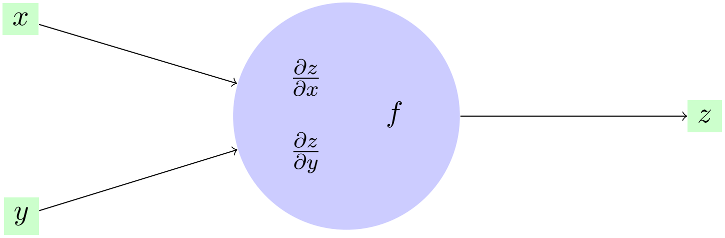
We want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



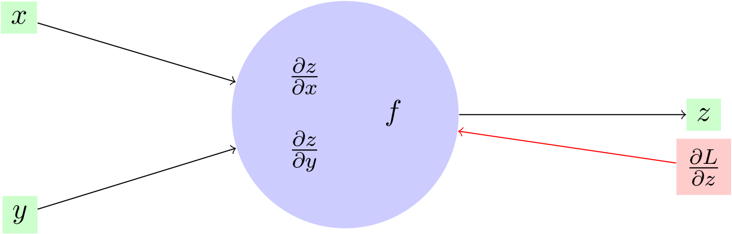
Global and local gradients



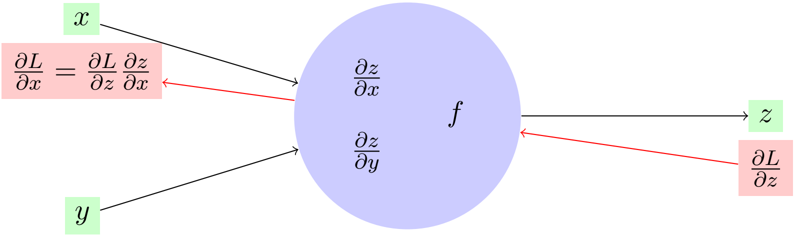
Global and local gradients



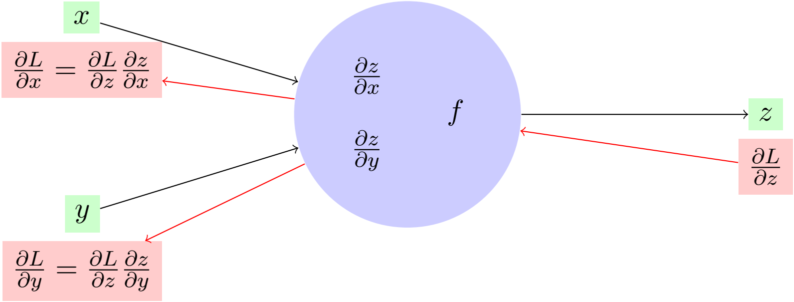
Global and local gradients



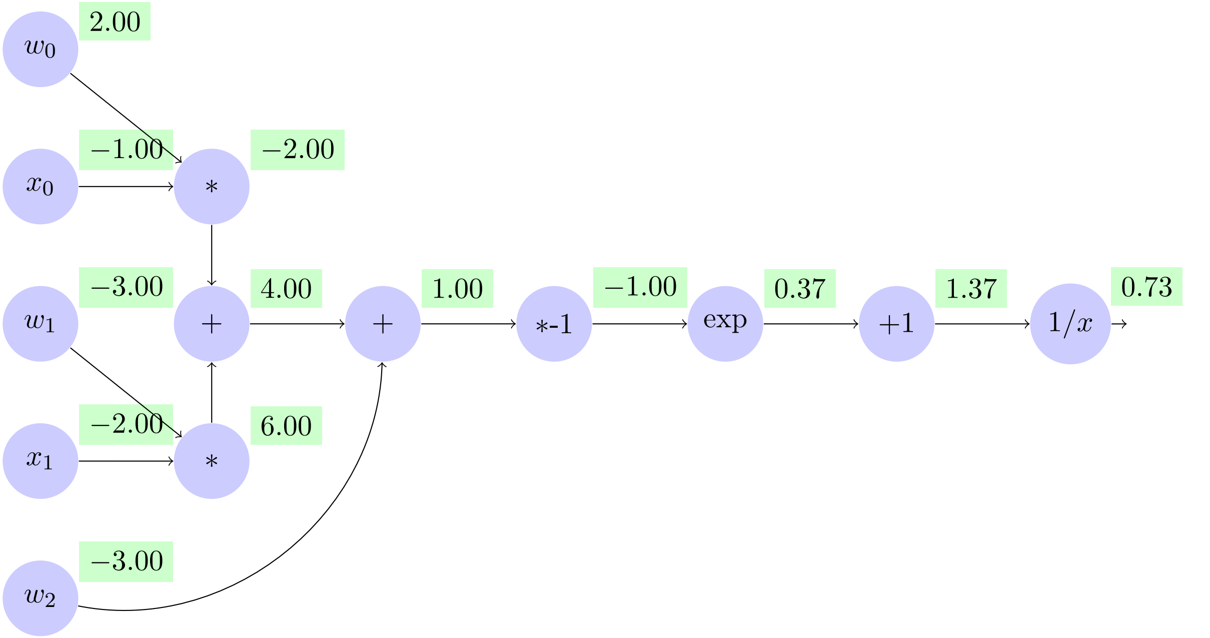
Global and local gradients



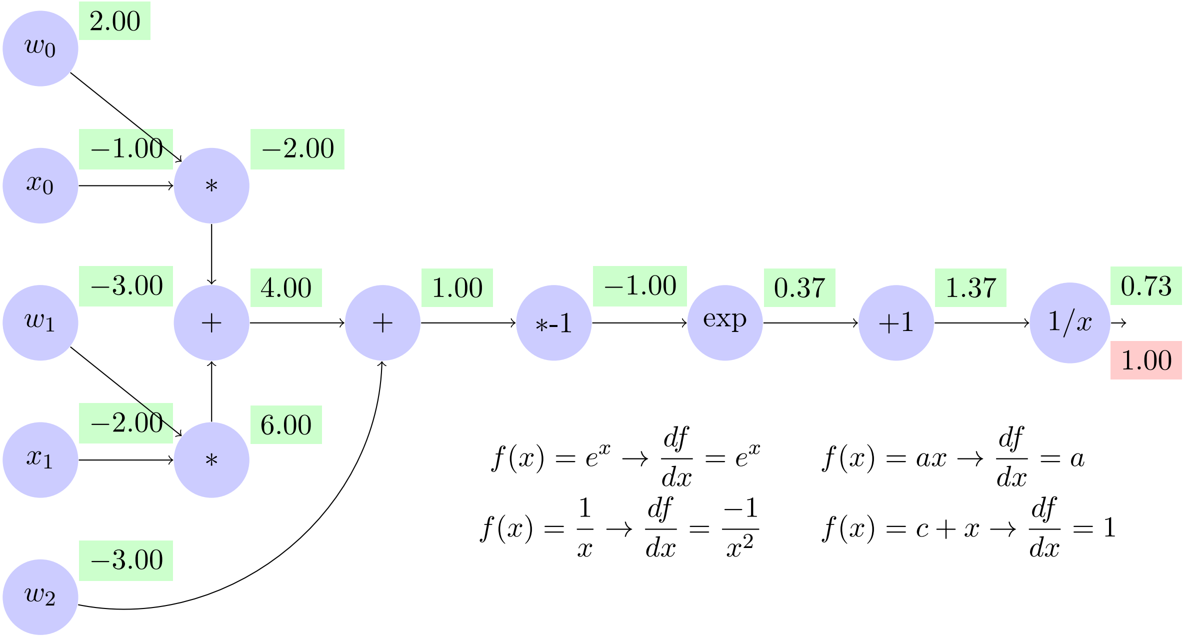
Global and local gradients



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



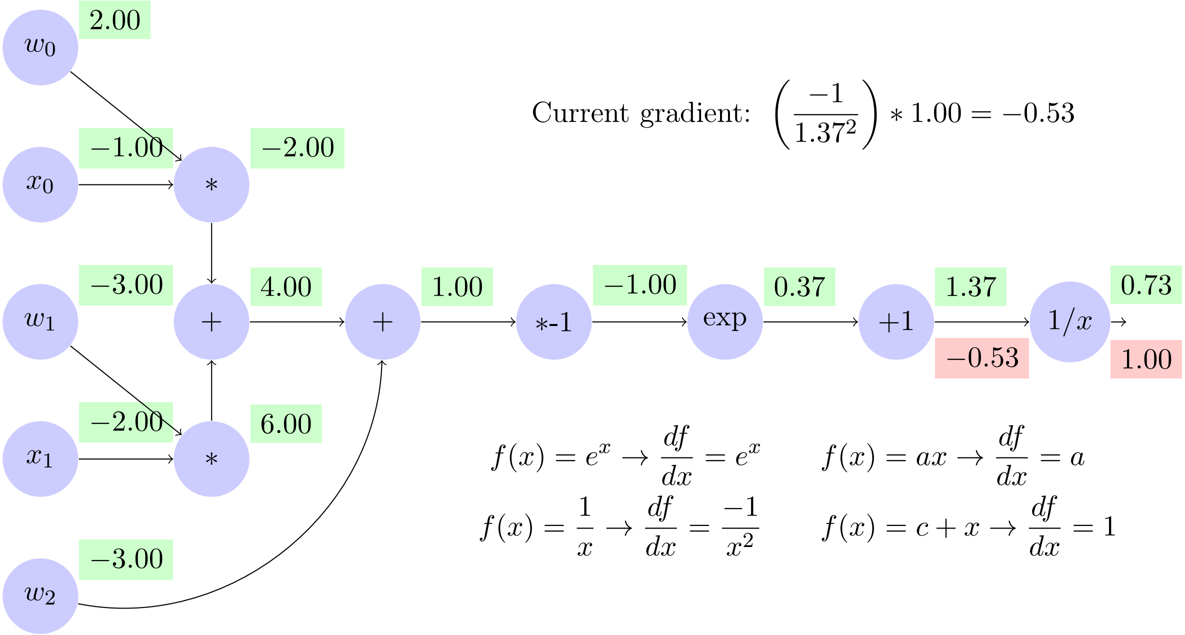
Another example, $f(w,x) = \frac{1}{1+e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$



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

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

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

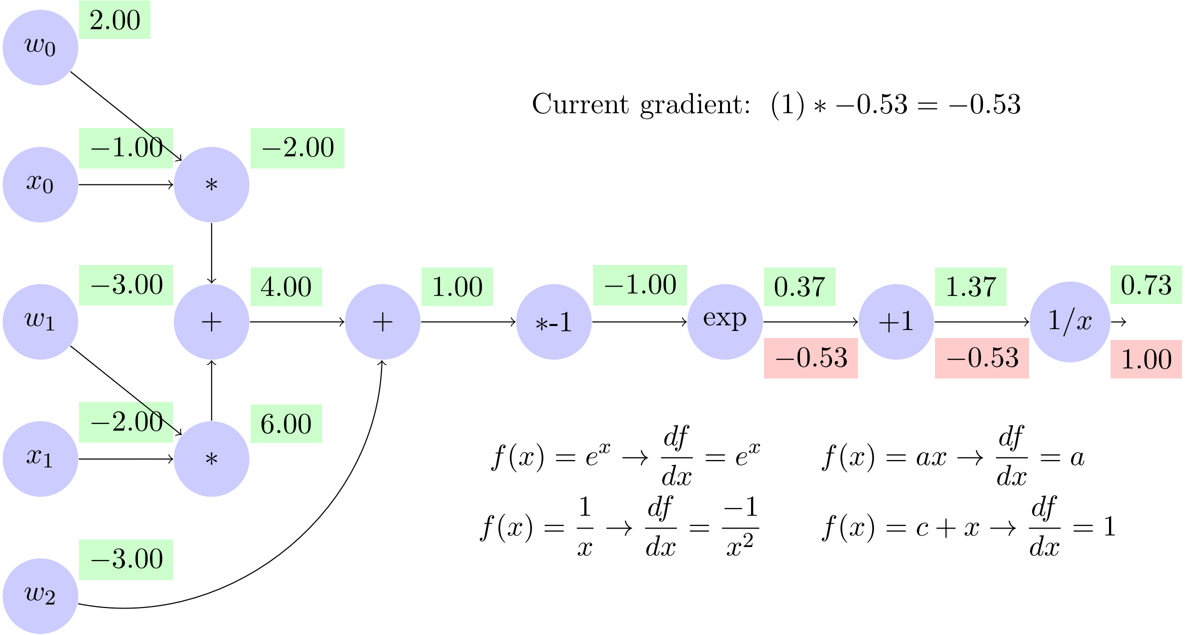


Current gradient: $\left(\frac{-1}{1.37^2}\right) * 1.00 = -0.53$

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

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

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

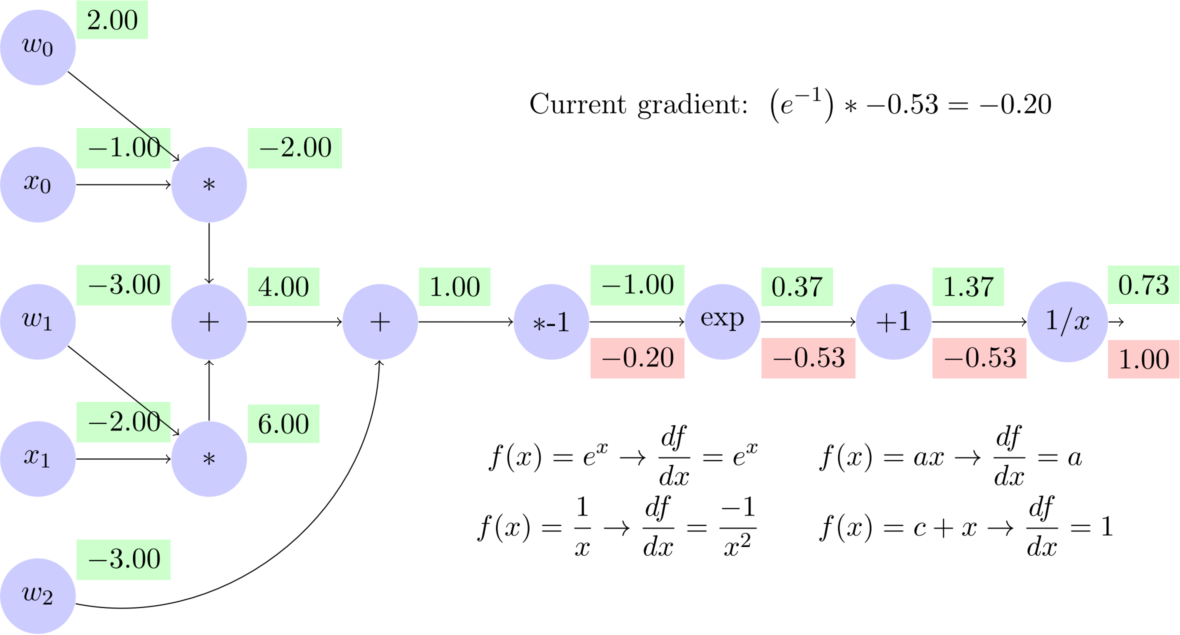


Current gradient: $(1) * -0.53 = -0.53$

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

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

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

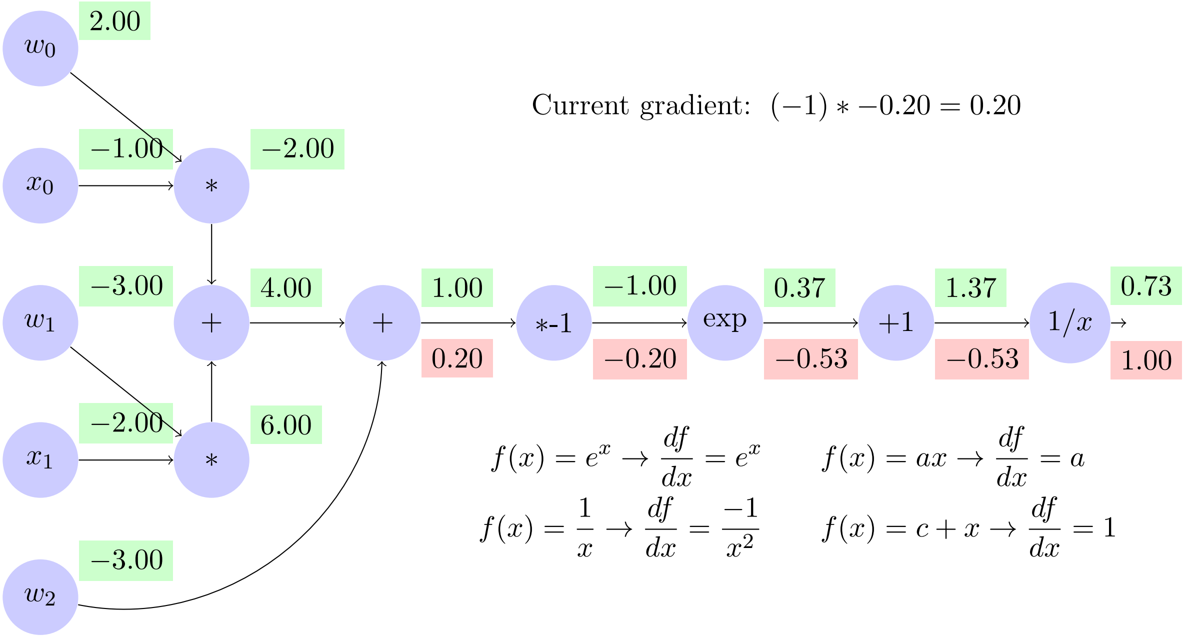


Current gradient: $(e^{-1}) * -0.53 = -0.20$

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

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

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

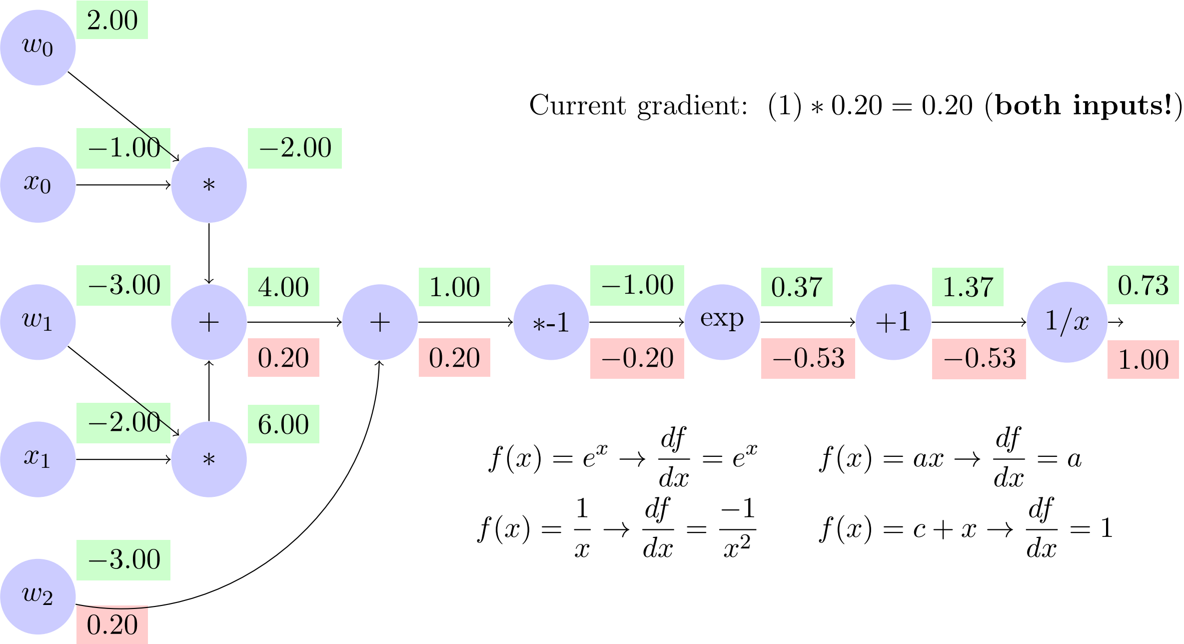


Current gradient: $(-1) * -0.20 = 0.20$

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

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

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

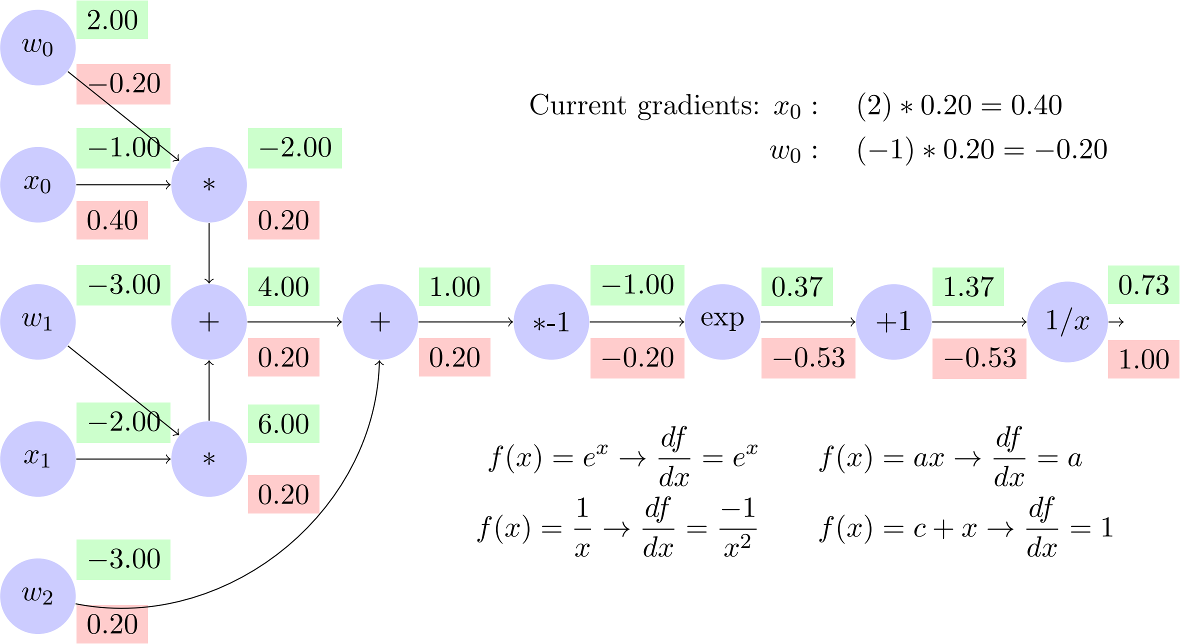


Current gradient: $(1) * 0.20 = 0.20$ (**both inputs!**)

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

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

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



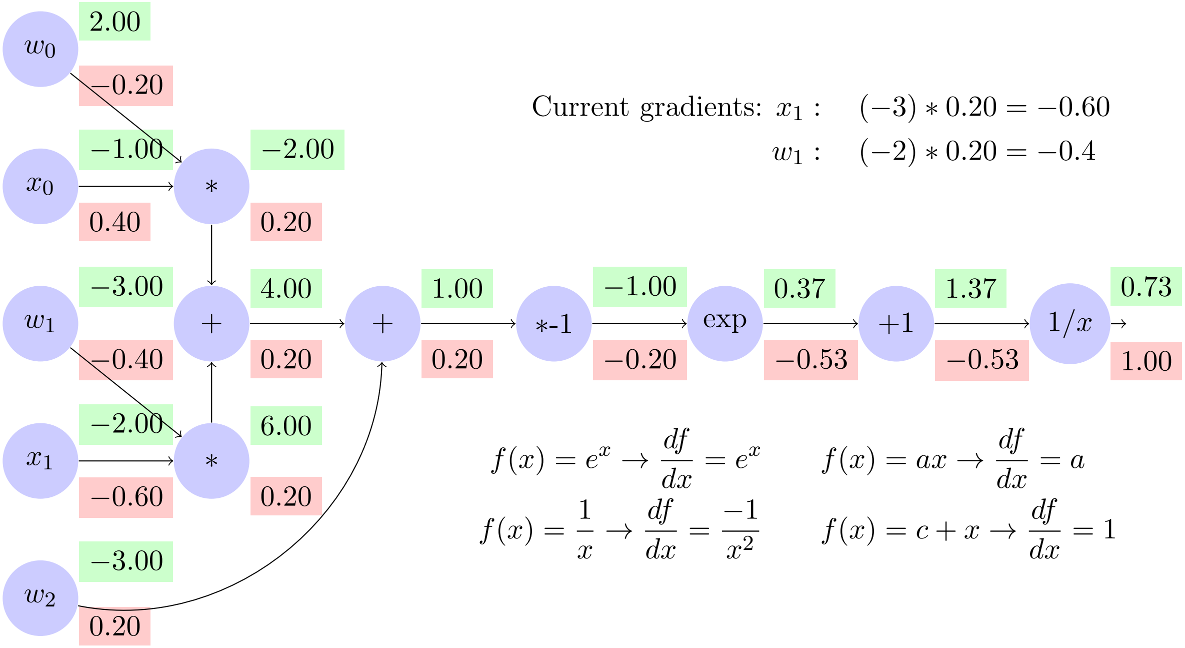
Current gradients: $x_0: \quad (2) * 0.20 = 0.40$

$w_0: \quad (-1) * 0.20 = -0.20$

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

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

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



Current gradients: $x_1 : \quad (-3) * 0.20 = -0.60$

$w_1 : \quad (-2) * 0.20 = -0.4$

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

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

Now, let's implement this in actual code and see if the results are the same!