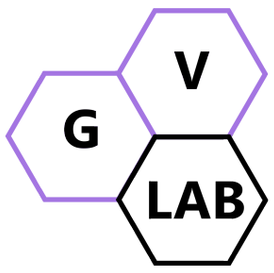
Non-linear Layer

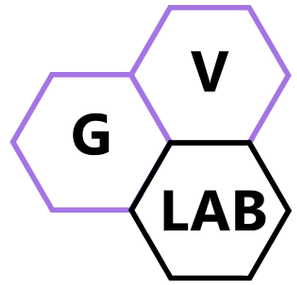
Dr. Thanh-Sach LE
LTSACH@hcmut.edu.vn



GVLab:
Graphics and Vision Laboratory

Faculty of Computer Science and Engineering,
HCMUT

2 Contents

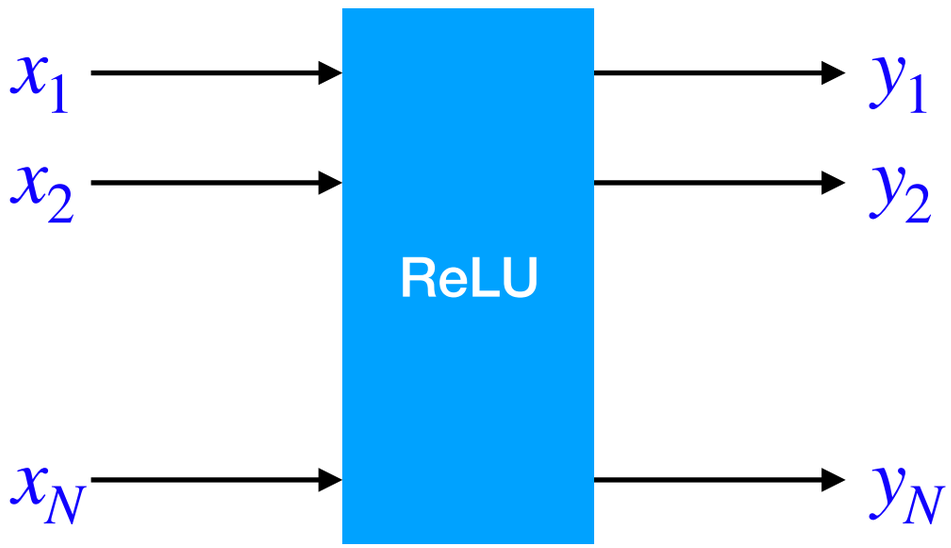


3 Goal of non-linear layer

- Add non-linear capacity to networks
- Without non-linear layers
  - Networks can approximate linear functions (X —> Y)

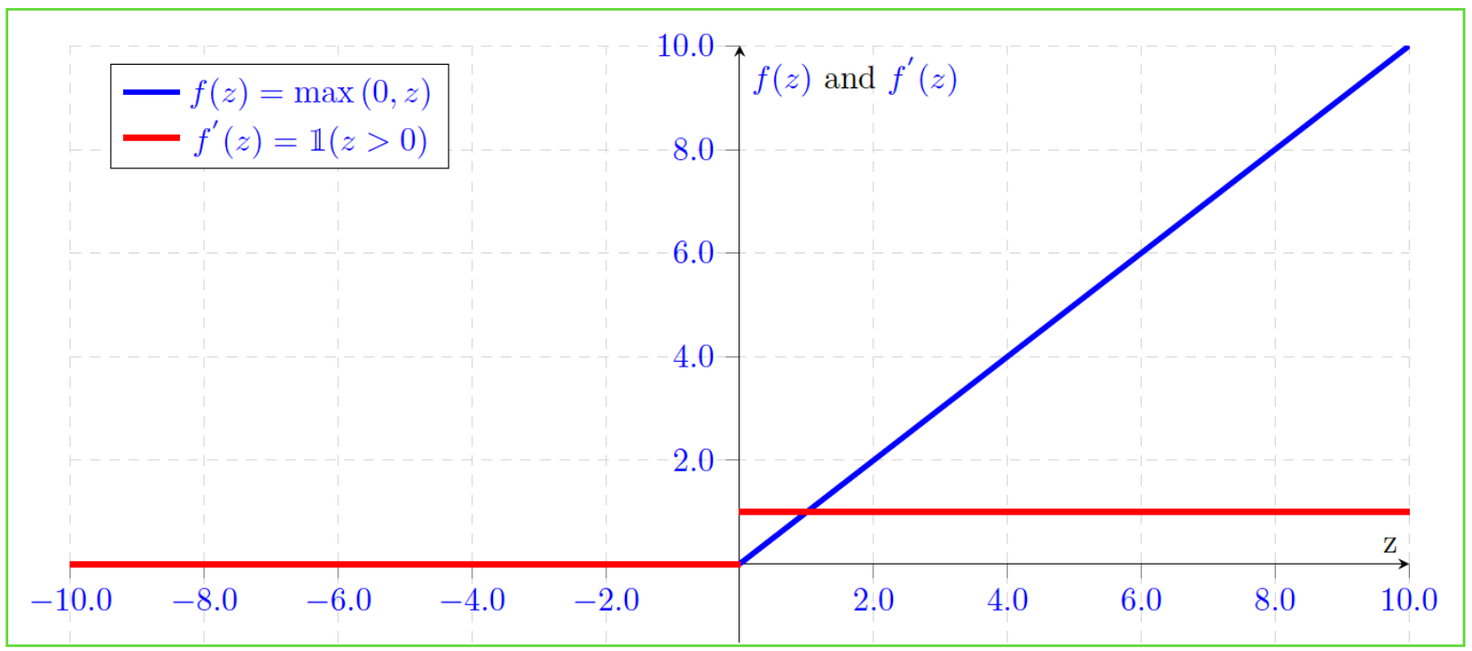
4

ReLU
Forward-pass



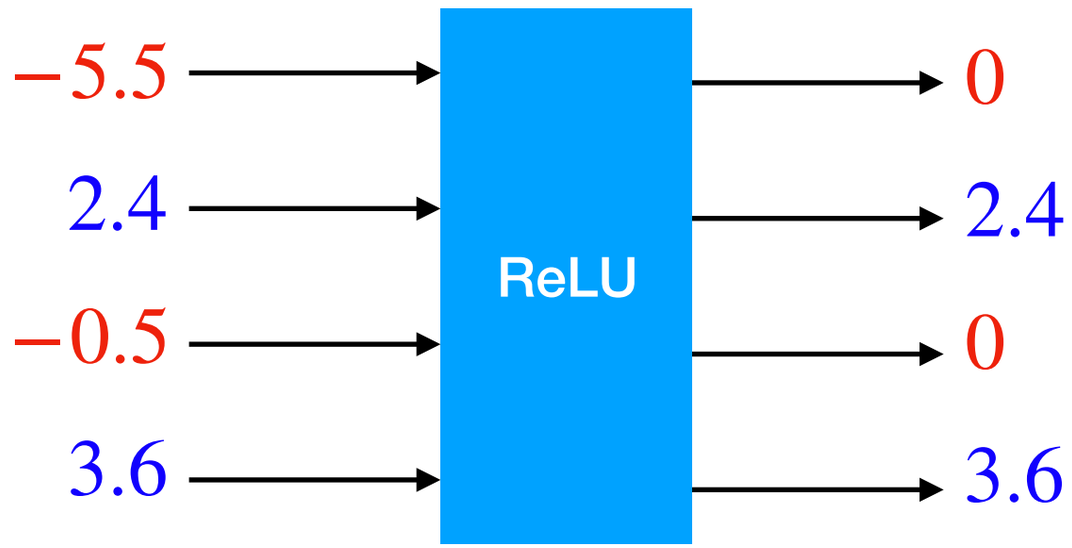
$$\partial b$$

$$y_i = x_i, \textbf{ if } x_i \geq 0$$
$$= 0, \textbf{ otherwise}$$



5

ReLU
Forward-pass

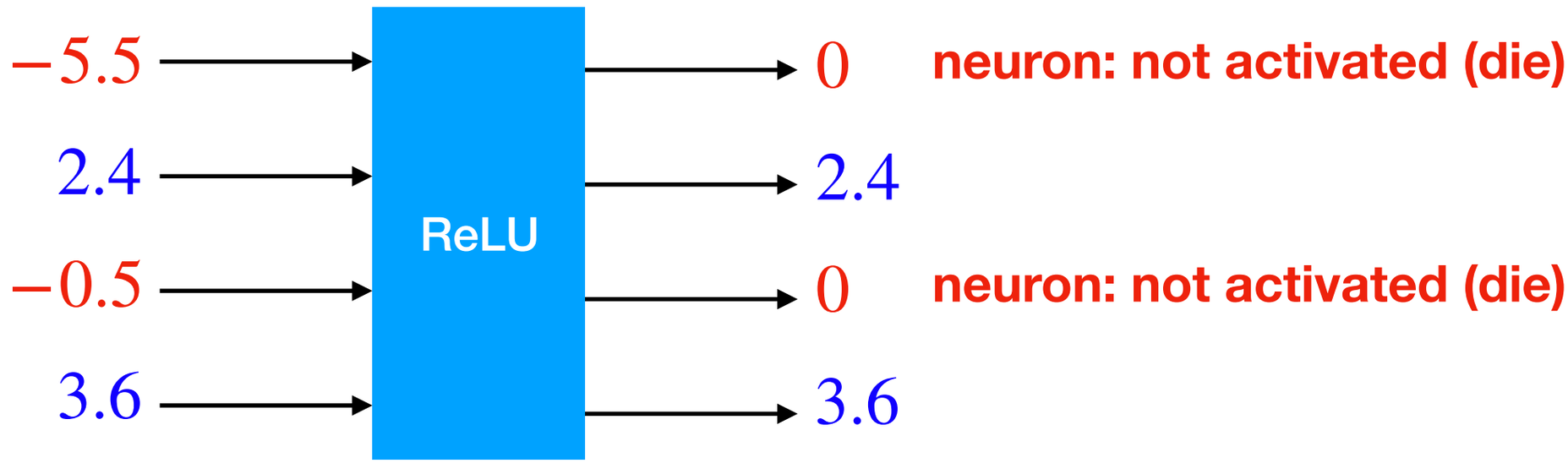


∂b

$$y_i = x_i, \textbf{ if } x_i \geq 0$$
$$= 0, \textbf{ otherwise}$$

6

ReLU
Forward-pass



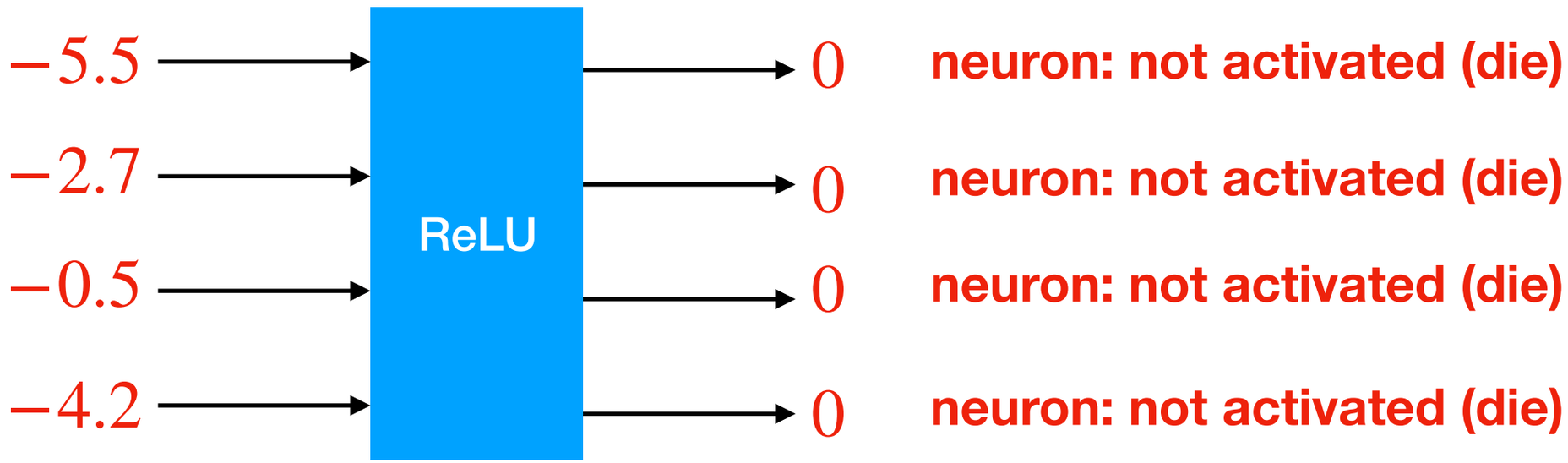
∂b

$$y_i = x_i, \textbf{ if } x_i \geq 0$$
$$= 0, \textbf{ otherwise}$$

deactivated neuron don’t affect or contribute to neurons in next layers

7

ReLU
Forward-pass

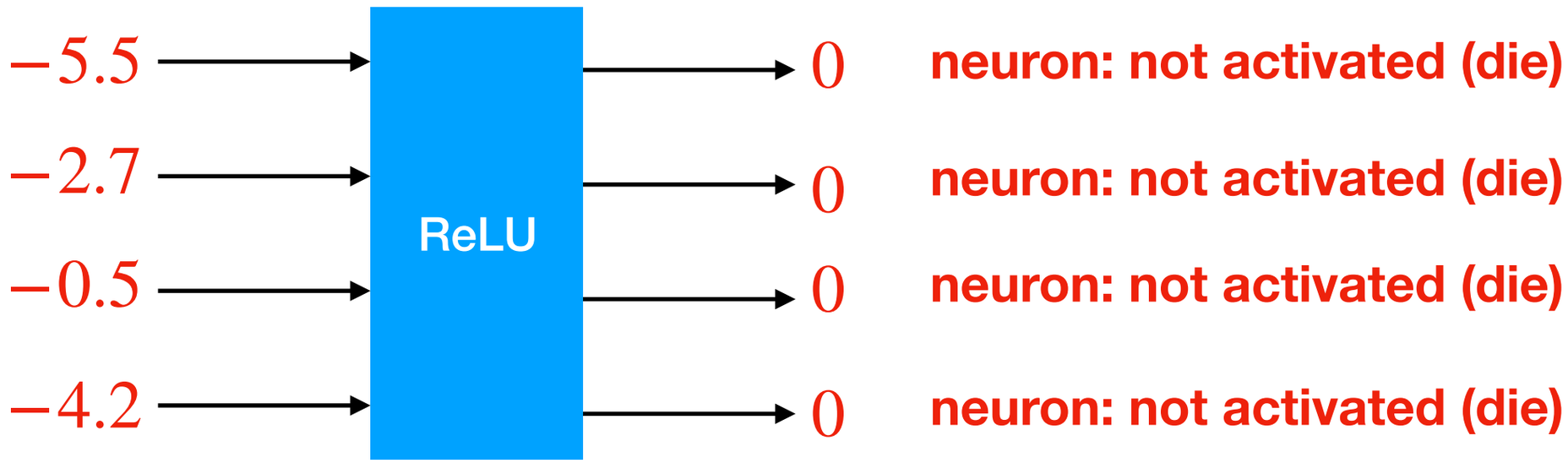


∂b

$$y_i = x_i, \textbf{ if } x_i \geq 0$$
$$= 0, \textbf{ otherwise}$$

What happened if all input to ReLU are negative?

ReLU
Forward-pass



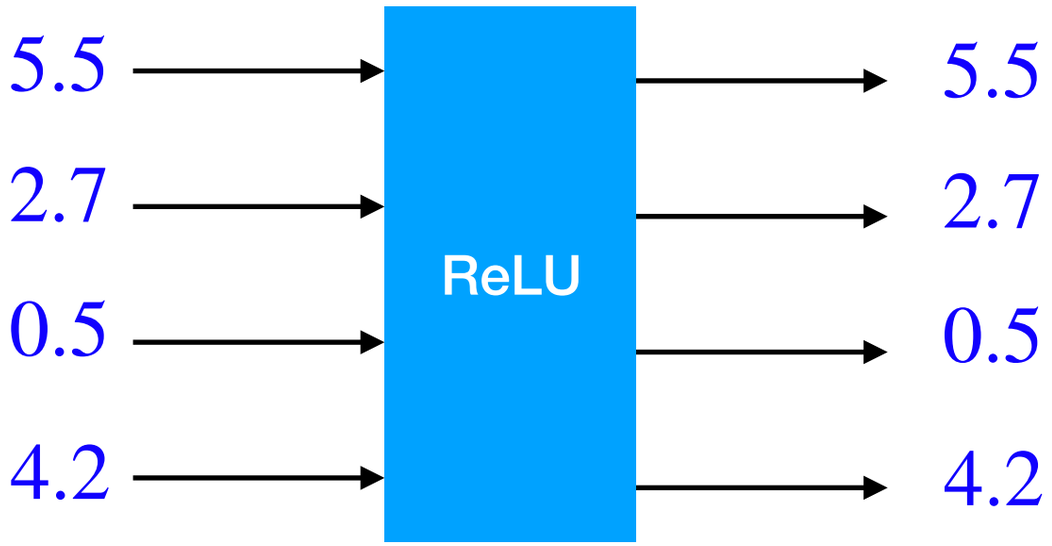
∂b

$$y_i = x_i, \textbf{ if } x_i \geq 0$$
$$= 0, \textbf{ otherwise}$$

What happened if all input to ReLU are negative?

Network can’t learn anything!

ReLU
Forward-pass



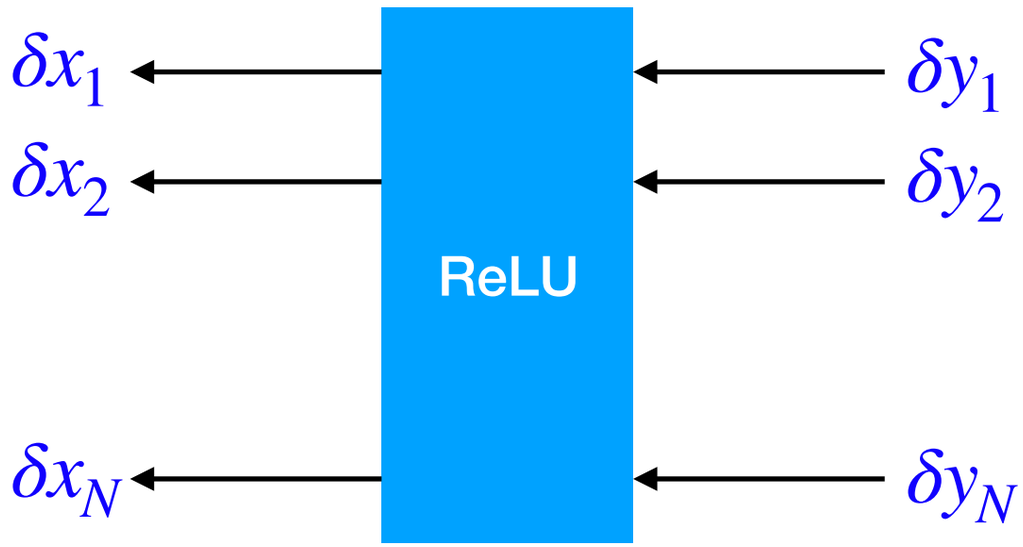
$$\frac{\partial b}{}$$

$$y_i = x_i, \textbf{ if } x_i \geq 0$$
$$= 0, \textbf{ otherwise}$$

What happened if all input to ReLU are positive?

ReLU layer is meaningless.
Convolution layer + ReLU = linear
Fully-connected layer + ReLU = linear

ReLU
Backward-pass

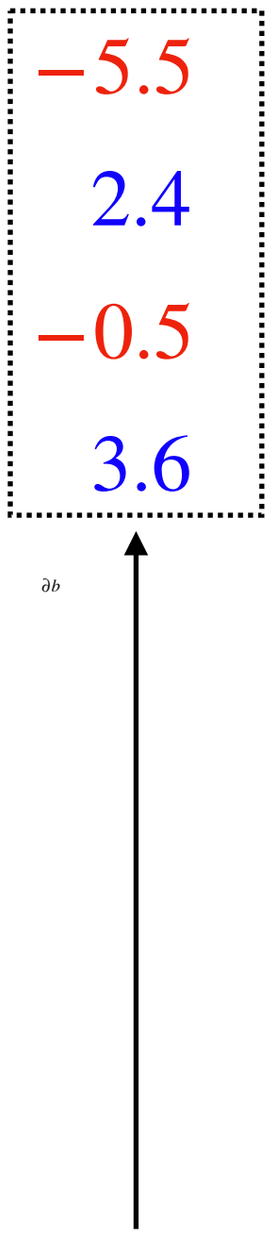


$$\frac{\partial b}{}$$

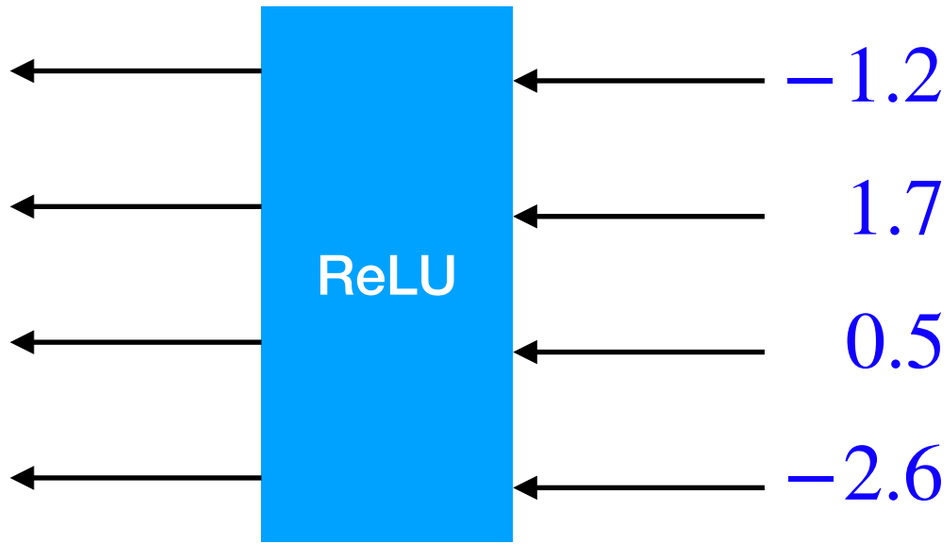
$$\delta x_i = \delta y_i \times \frac{dy_i}{dx_i} = \delta y_i, \textbf{ if } x_i \geq 0$$

$$= 0, \textbf{ otherwise}$$

ReLU
Backward-pass



?

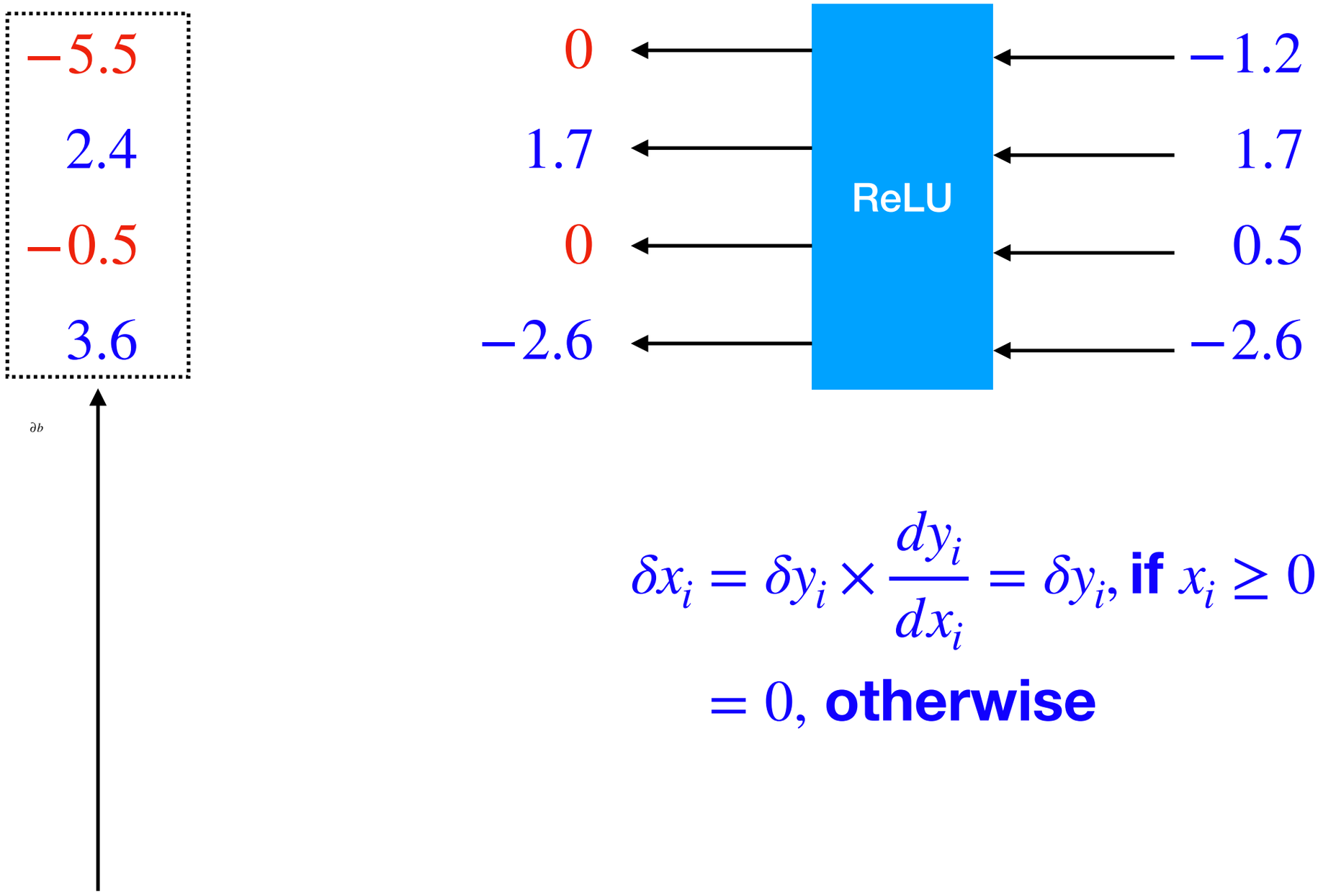


$$\delta x_i = \delta y_i \times \frac{dy_i}{dx_i} = \delta y_i, \textbf{ if } x_i \geq 0$$

$$= 0, \textbf{ otherwise}$$

cached during forward-pass

ReLU
Backward-pass

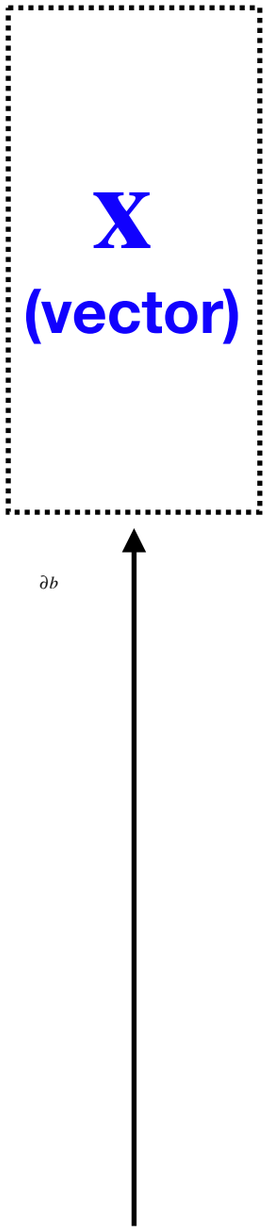


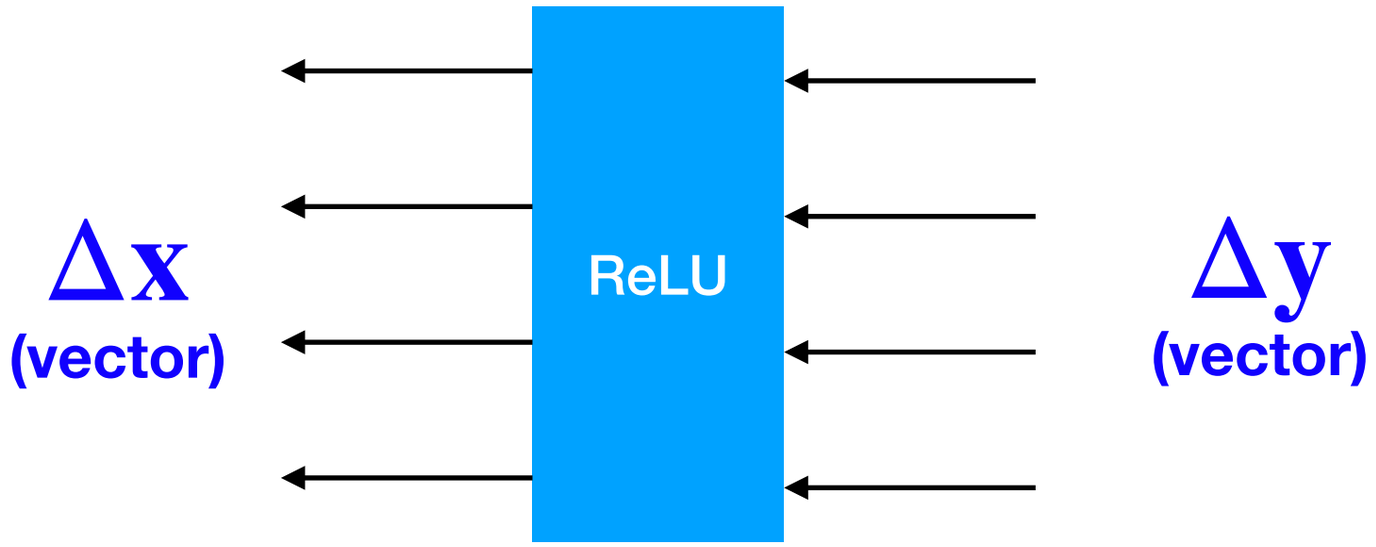
$$\delta x_i = \delta y_i \times \frac{dy_i}{dx_i} = \delta y_i, \textbf{ if } x_i \geq 0$$

$$= 0, \textbf{ otherwise}$$

cached during forward-pass

ReLU
Backward-pass





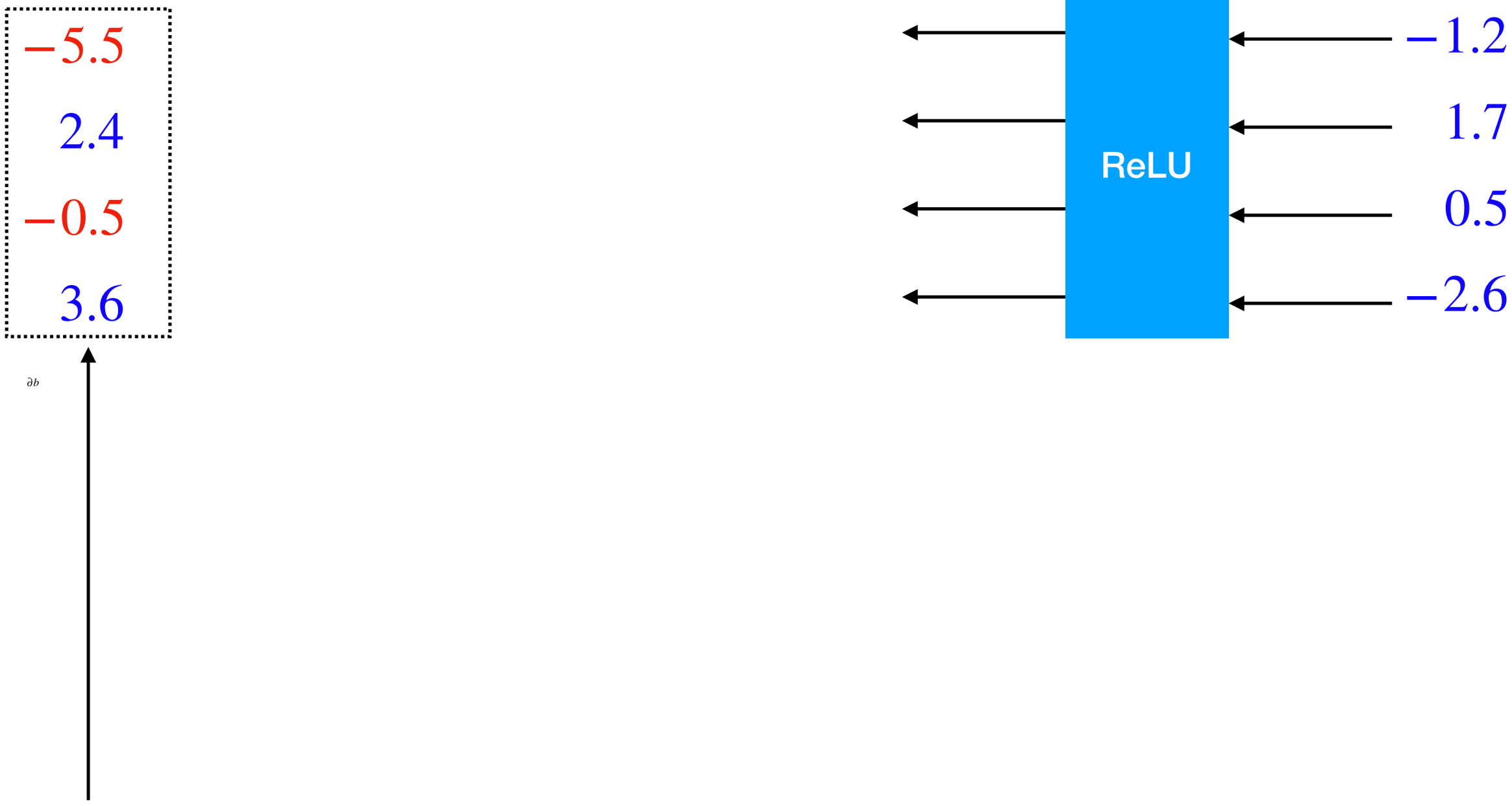
$$\Delta \mathbf{x} = \Delta \mathbf{y} \bullet \mathbf{m}$$

$$m_i = 1, \text{ if } x_i \geq 0$$
$$= 0, \text{ otherwise}$$

- is element-wise multiplication

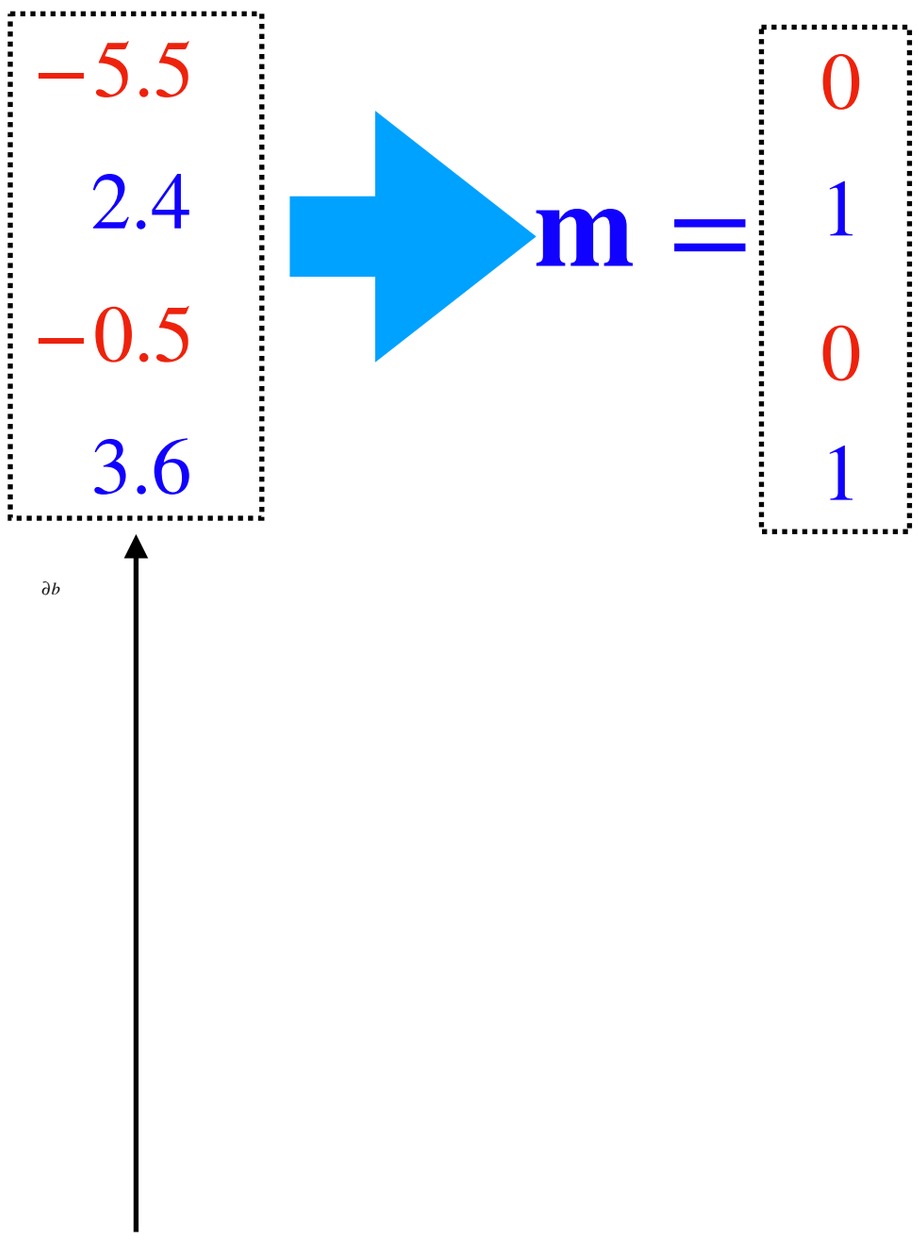
cached during forward-pass

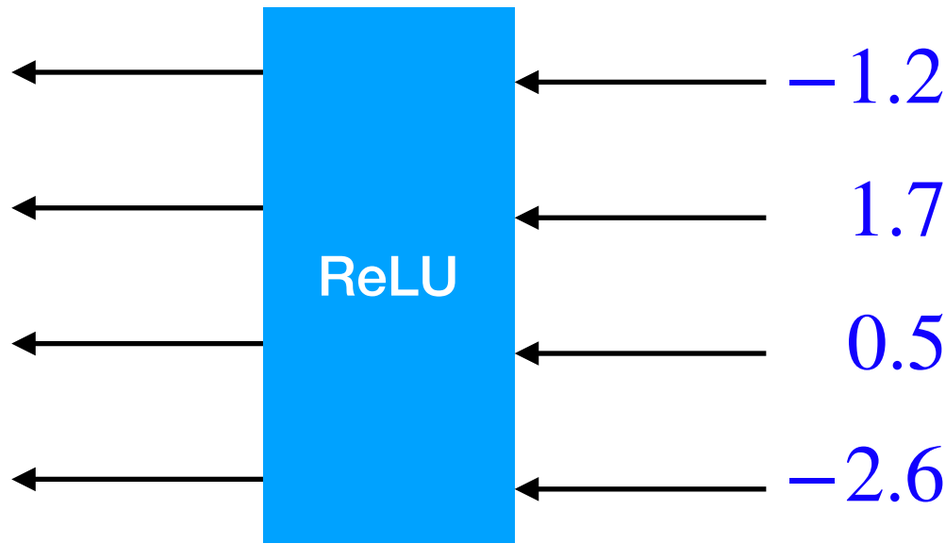
ReLU
Backward-pass



cached during forward-pass

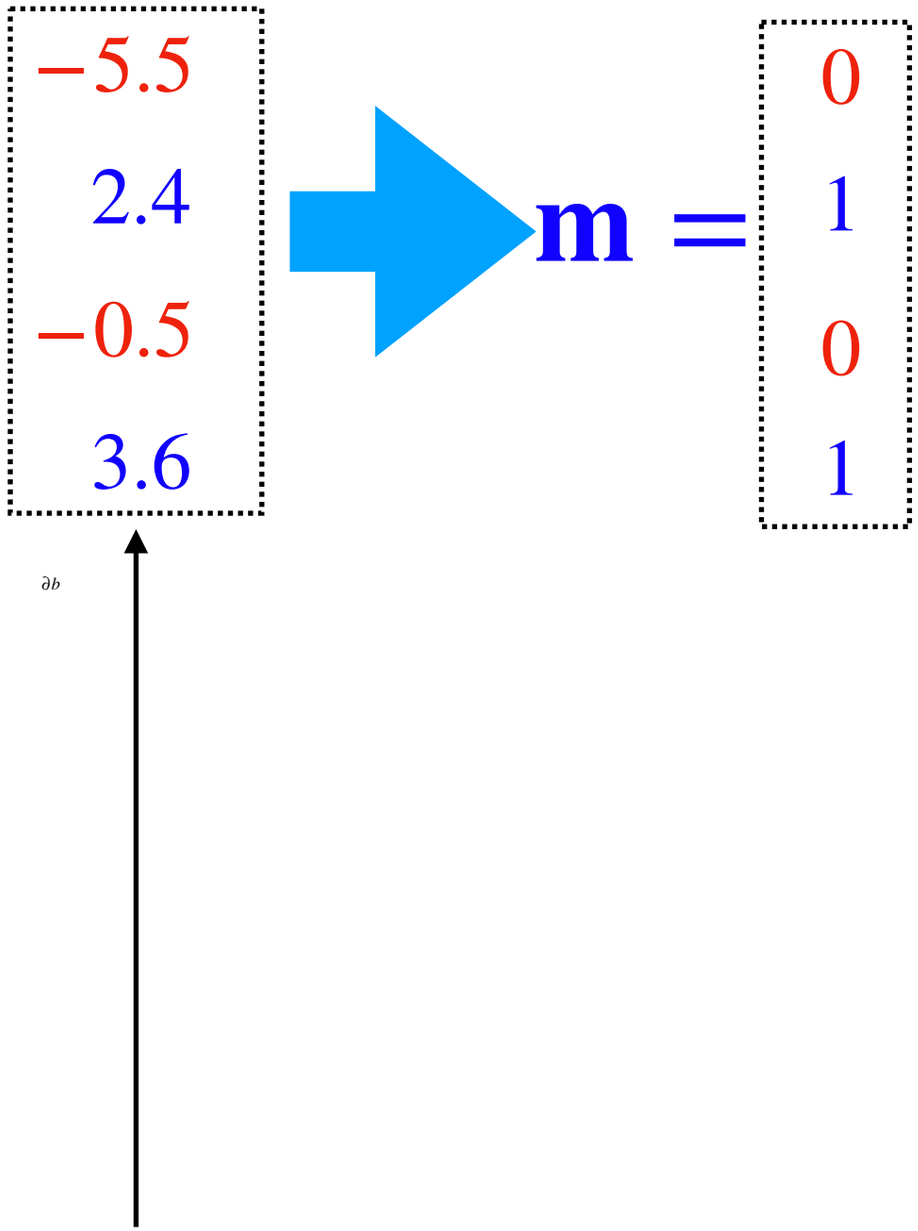
ReLU
Backward-pass





cached during forward-pass

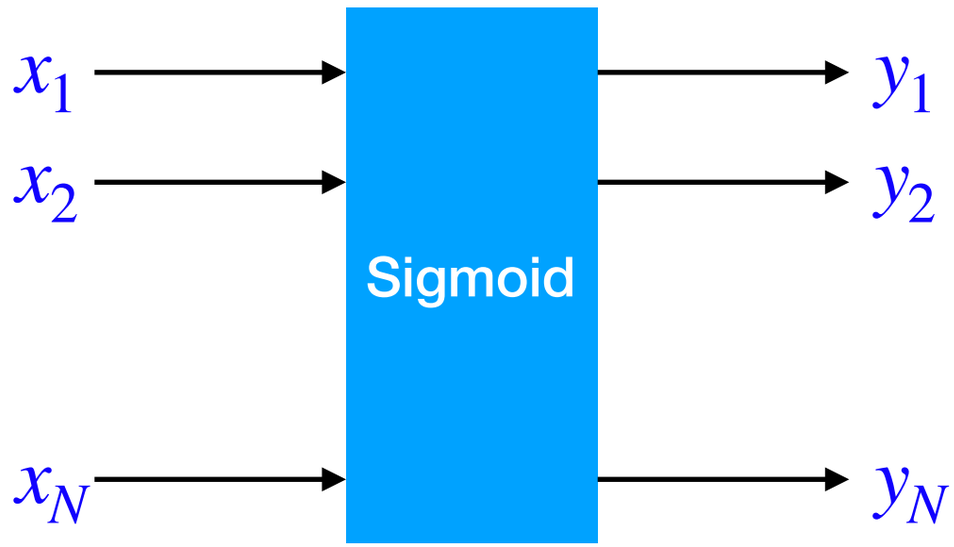
ReLU
Backward-pass



$$\Delta x = \Delta y \bullet m$$

cached during forward-pass

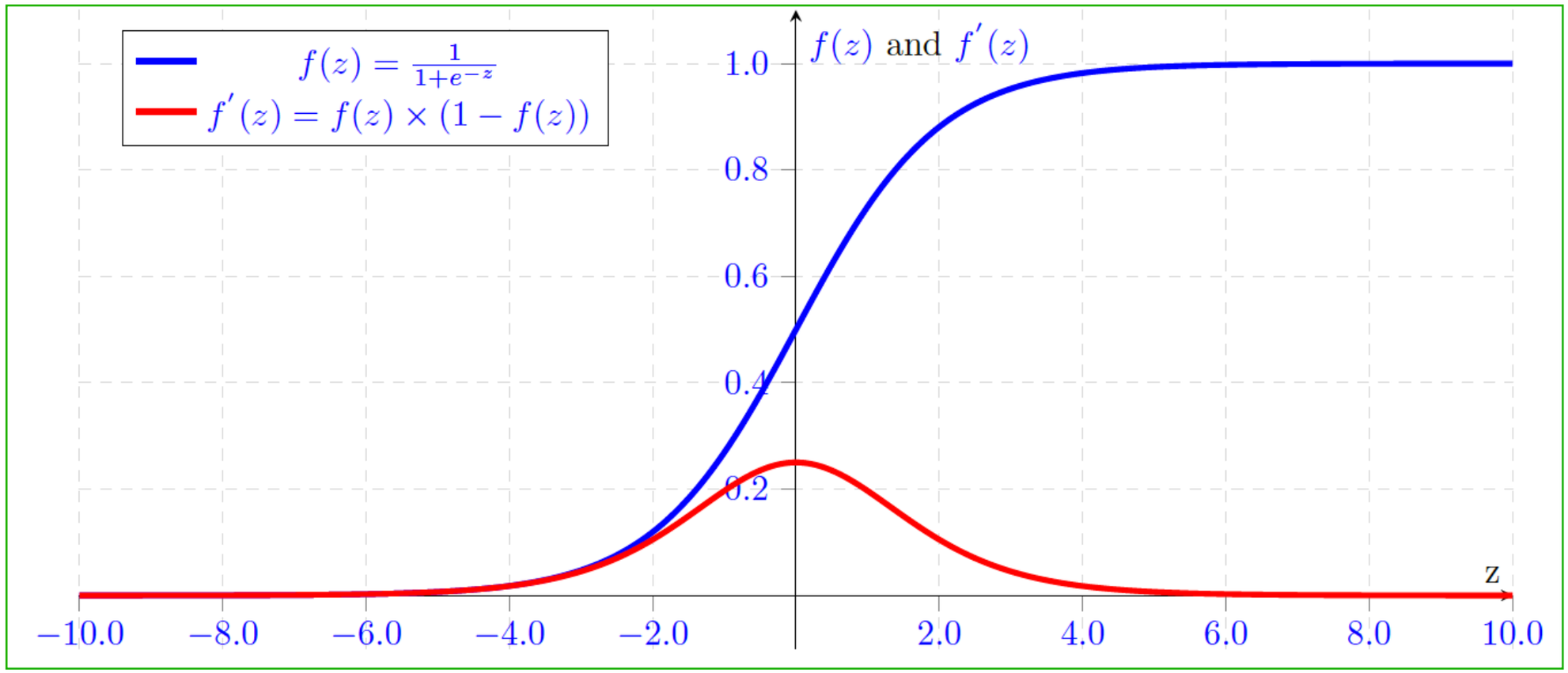
Sigmoid
Forward-pass



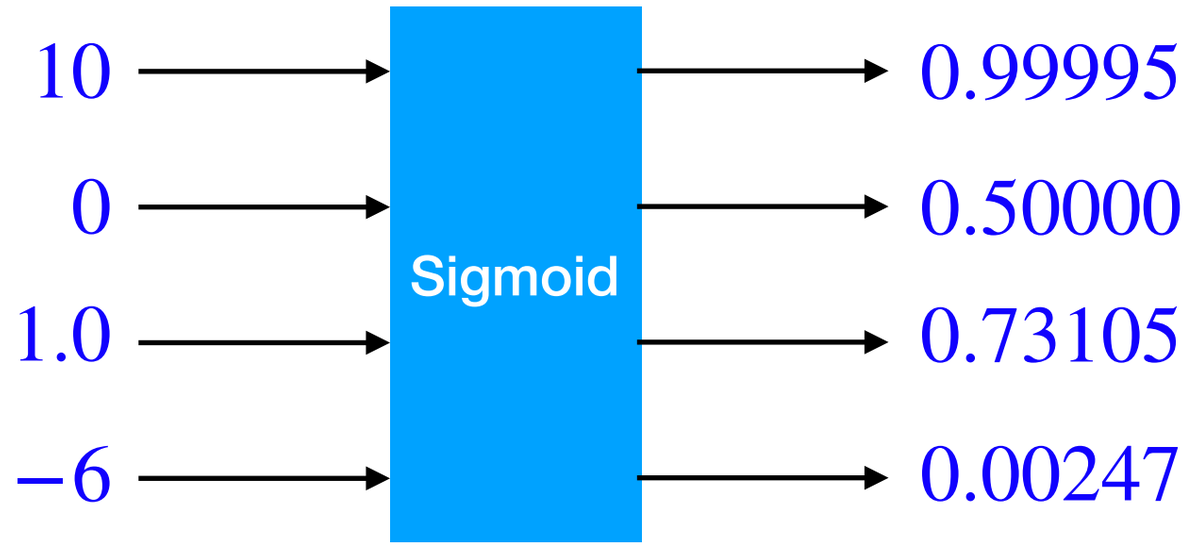
$$y_i = \frac{1}{1 + e^{-x_i}}$$

$$y_i^{'} = y_i(1 - y_i)$$

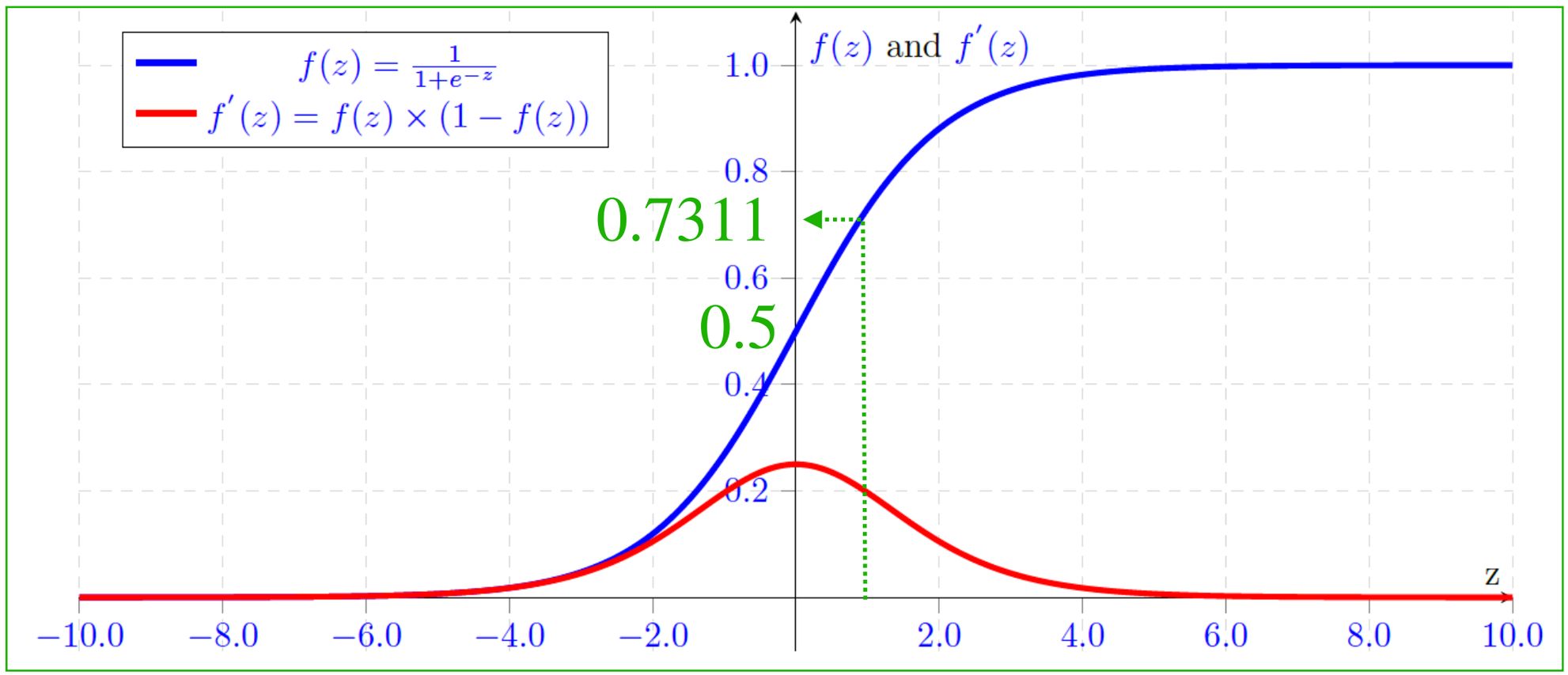
∂b



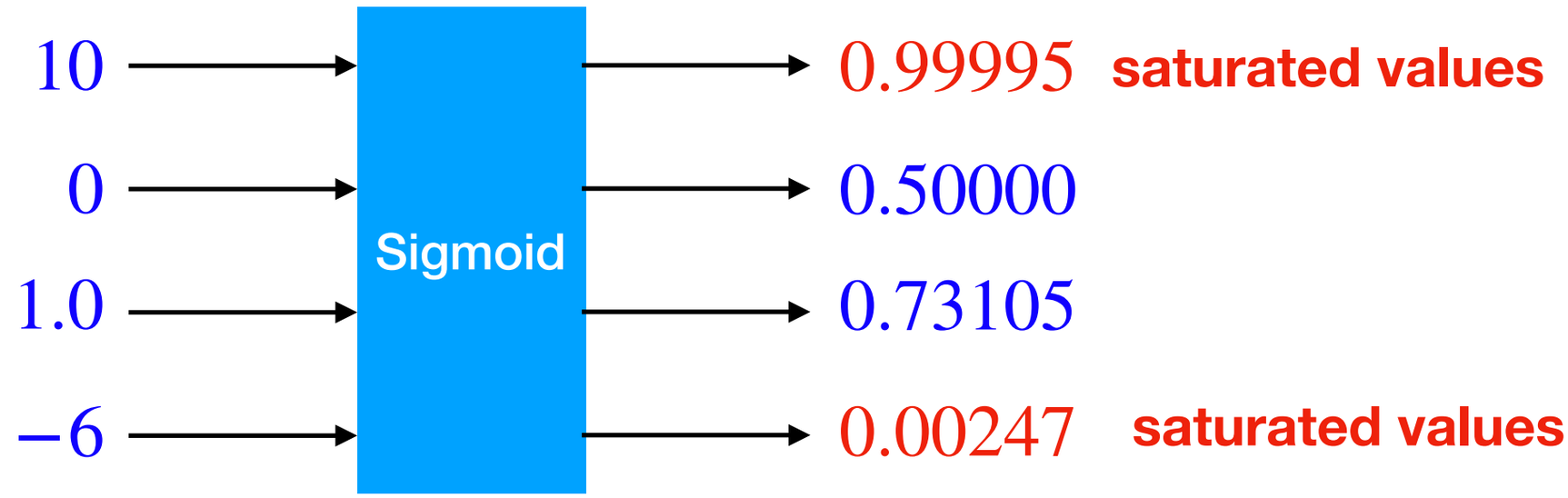
Sigmoid
Forward-pass



∂b



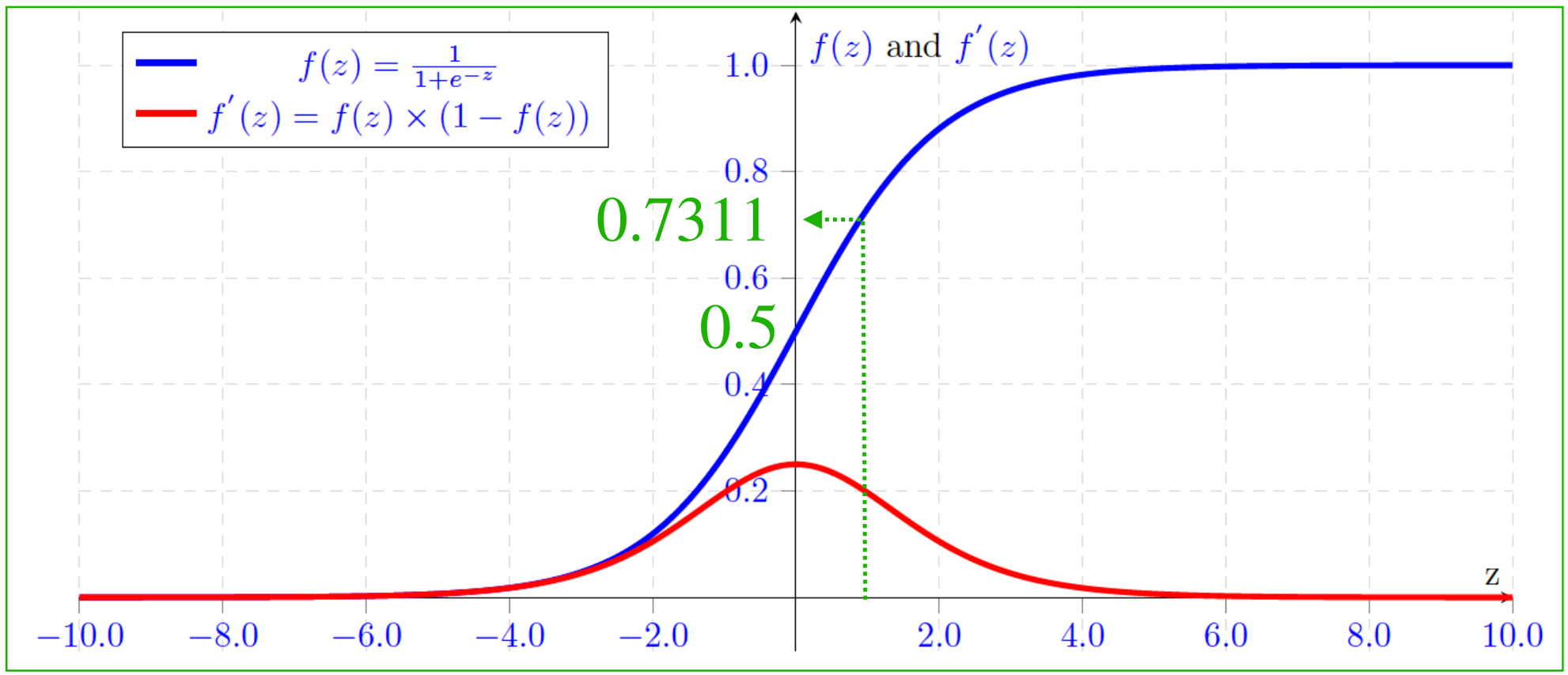
Sigmoid
Forward-pass



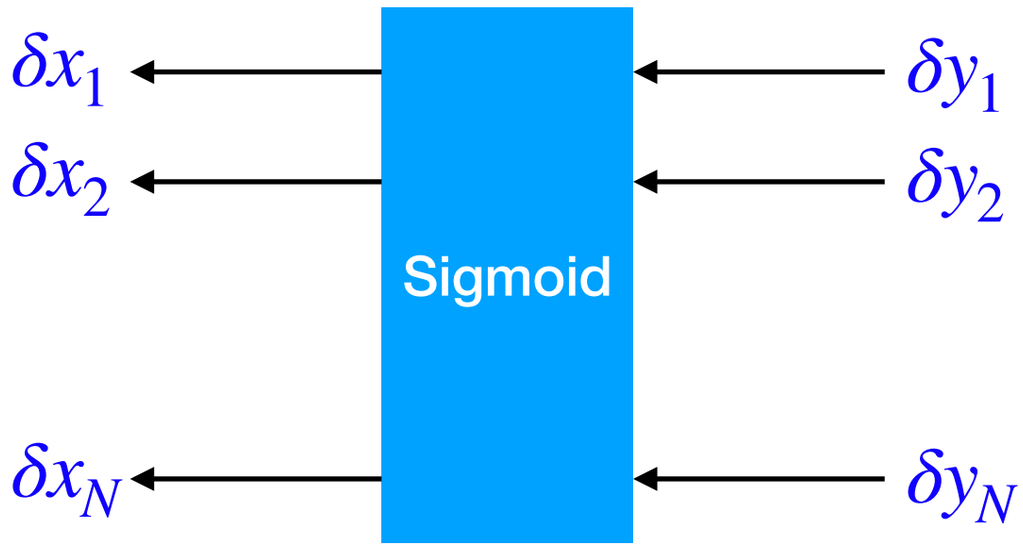
saturated values

saturated values

∂b



Sigmoid
Backward-pass



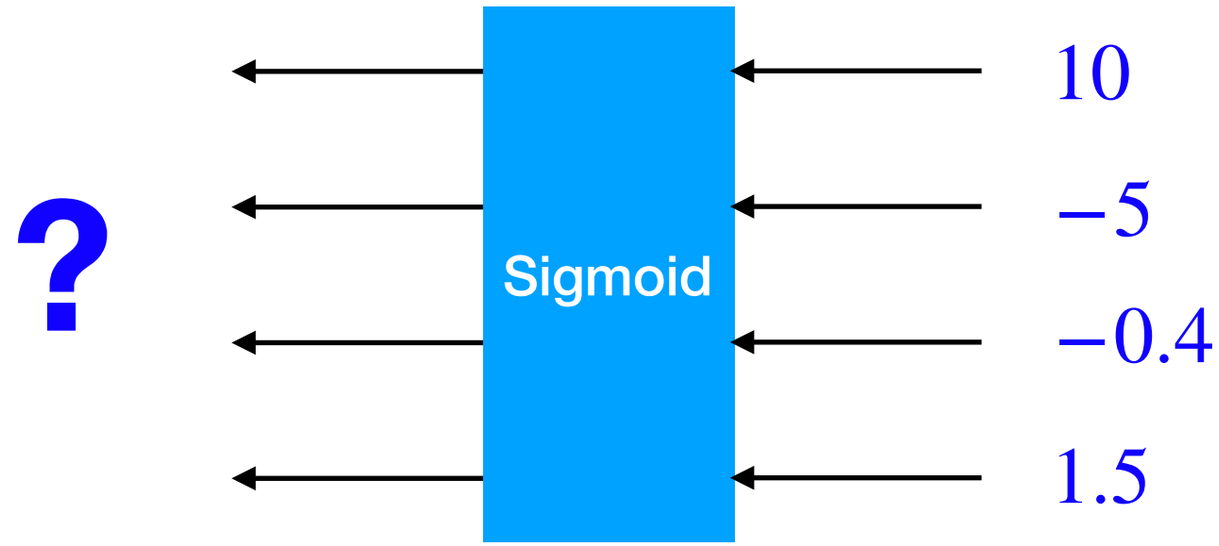
$$y_i = \frac{1}{1 + e^{-x_i}}$$

$$y_i^{'} = y_i(1 - y_i)$$

$$\frac{\partial b}{}$$

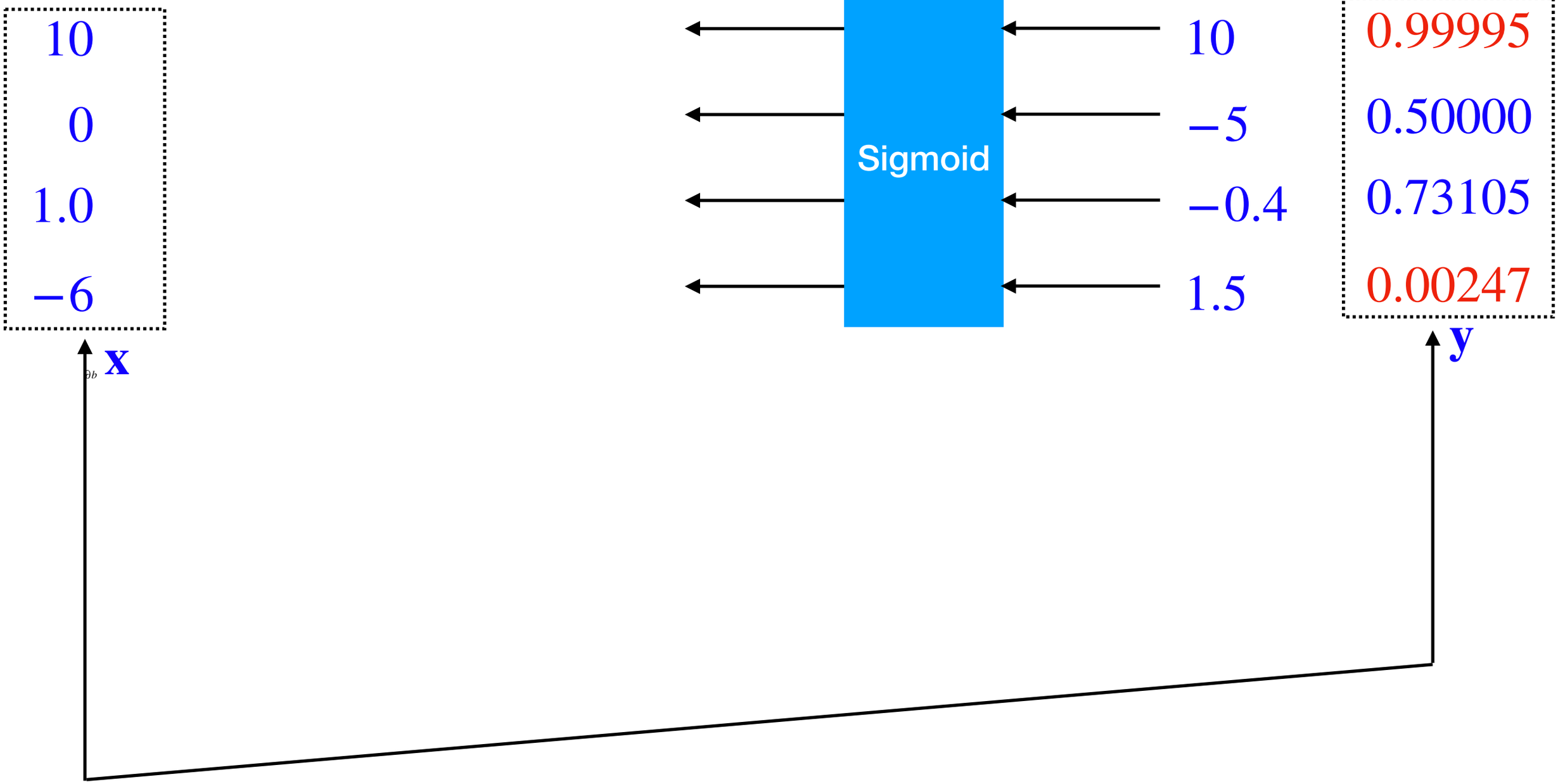
$$\delta x_i = \delta y_i \times \frac{dy_i}{dx_i} = \delta y_i \times y_i \times (1 - y_i)$$

Sigmoid
Backward-pass



∂b

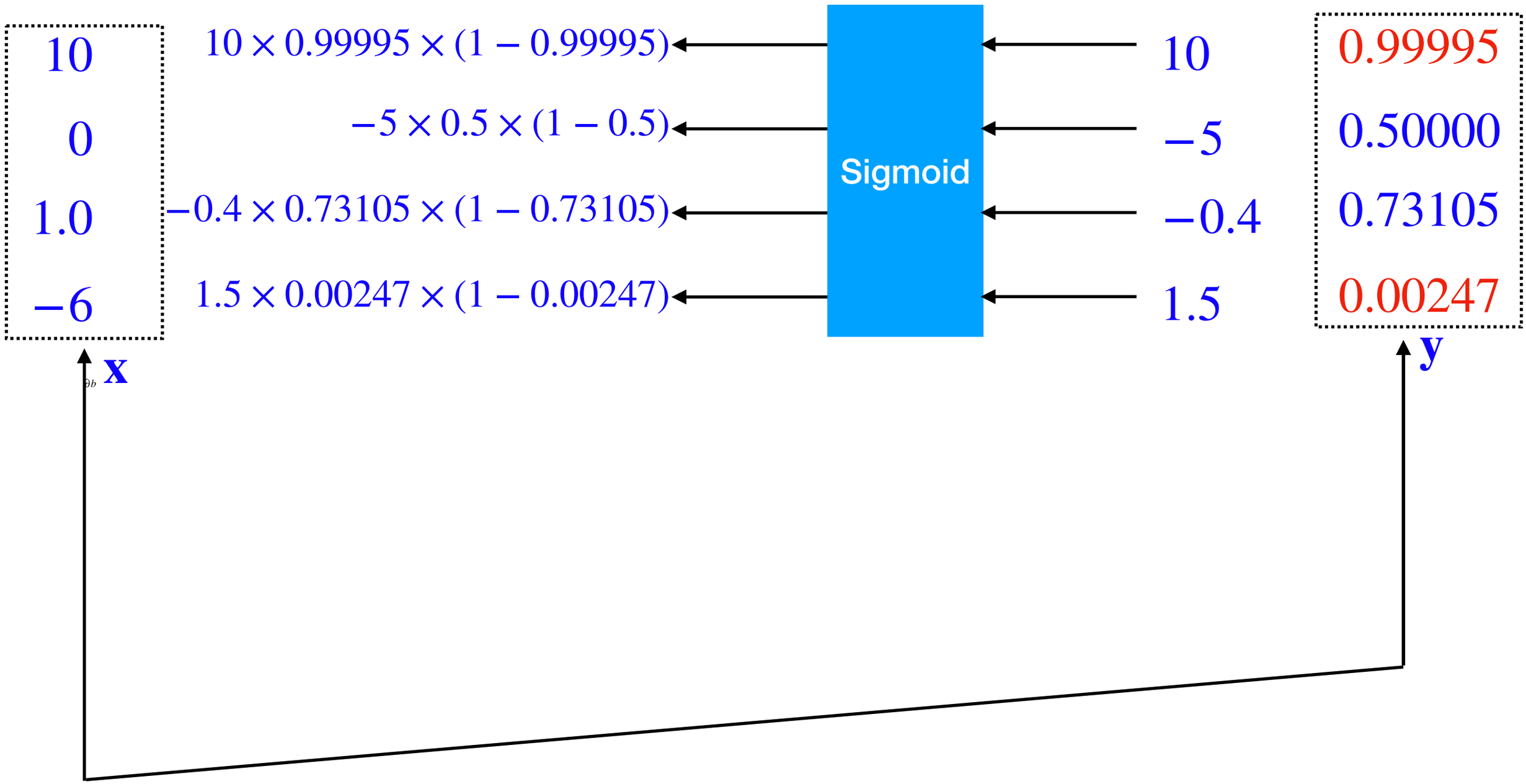
Sigmoid
Backward-pass



cached during forward-pass

$$\delta x_i = \delta y_i \times \frac{dy_i}{dx_i} = \delta y_i \times y_i \times (1 - y_i)$$

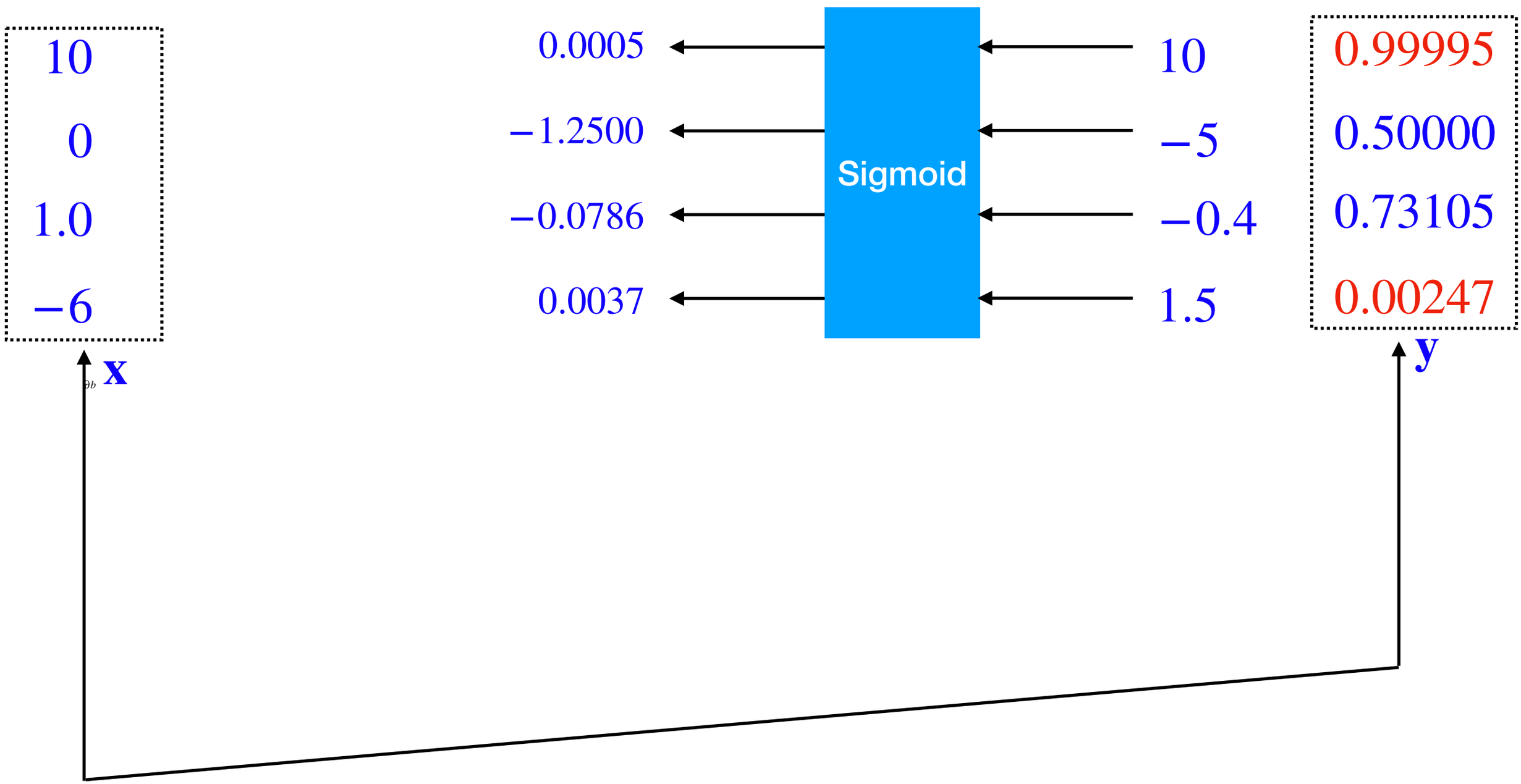
Sigmoid
Backward-pass



cached during forward-pass

$$\delta x_i = \delta y_i \times \frac{dy_i}{dx_i} = \delta y_i \times y_i \times (1 - y_i)$$

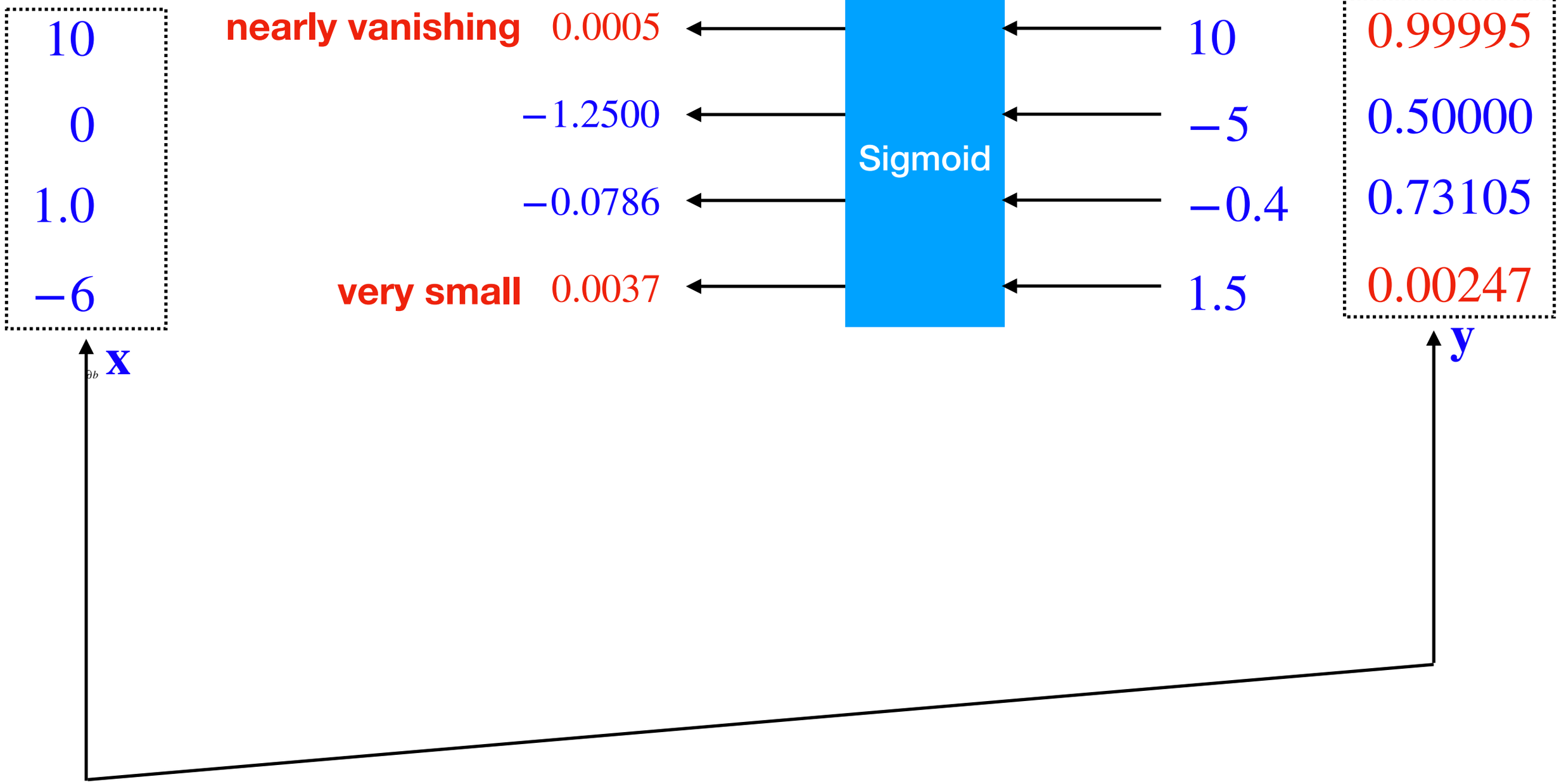
Sigmoid
Backward-pass



cached during forward-pass

$$\delta x_i = \delta y_i \times \frac{dy_i}{dx_i} = \delta y_i \times y_i \times (1 - y_i)$$

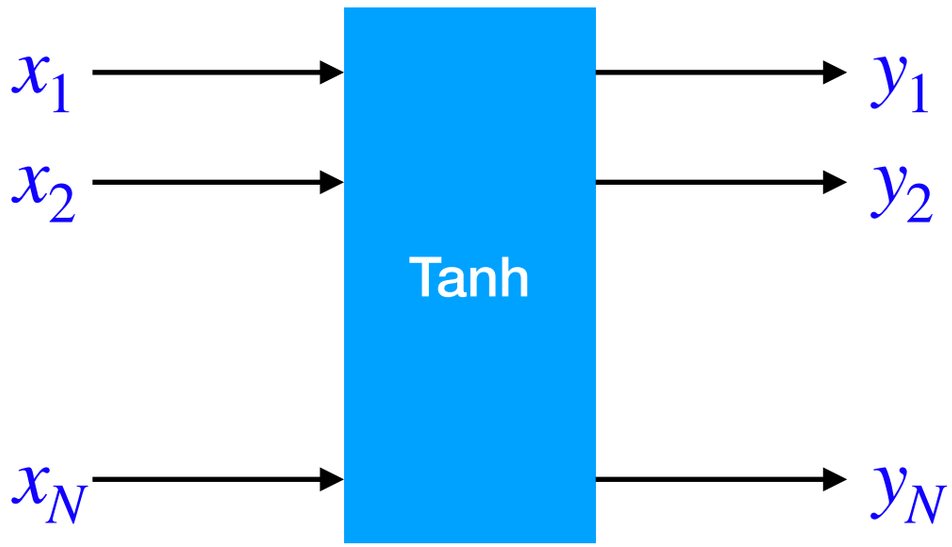
Sigmoid
Backward-pass



cached during forward-pass

$$\delta x_i = \delta y_i \times \frac{dy_i}{dx_i} = \delta y_i \times y_i \times (1 - y_i)$$

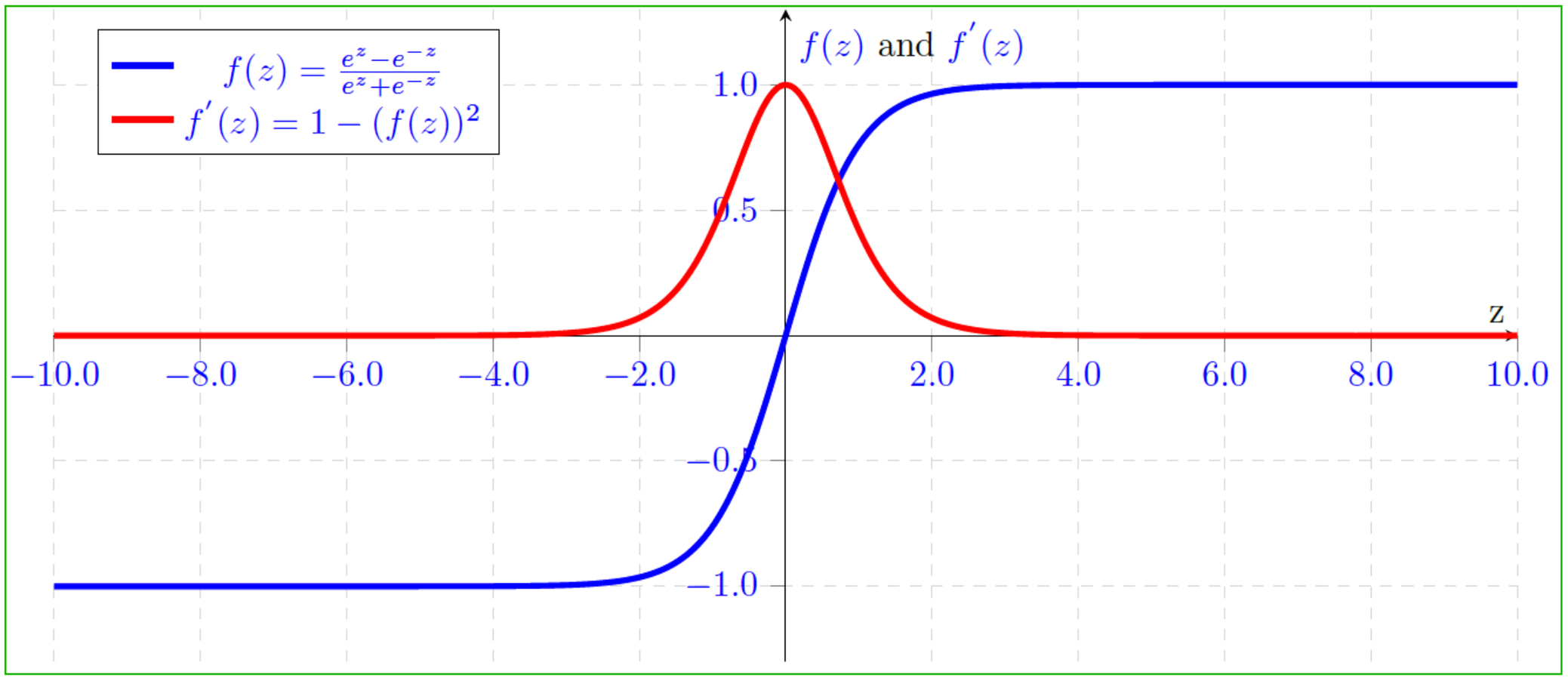
Tanh
Forward-pass



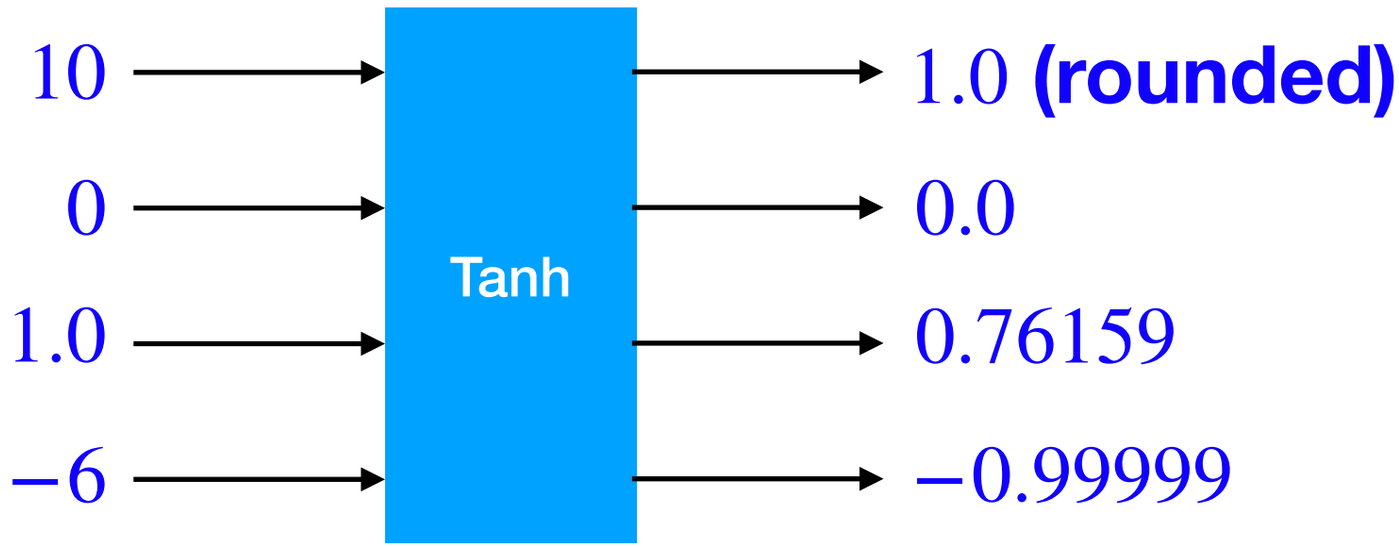
$$y_i = \frac{e^{x_i} - e^{-x_i}}{e^{x_i} + e^{-x_i}}$$

$$y_i^{'} = 1 - y_i^2$$

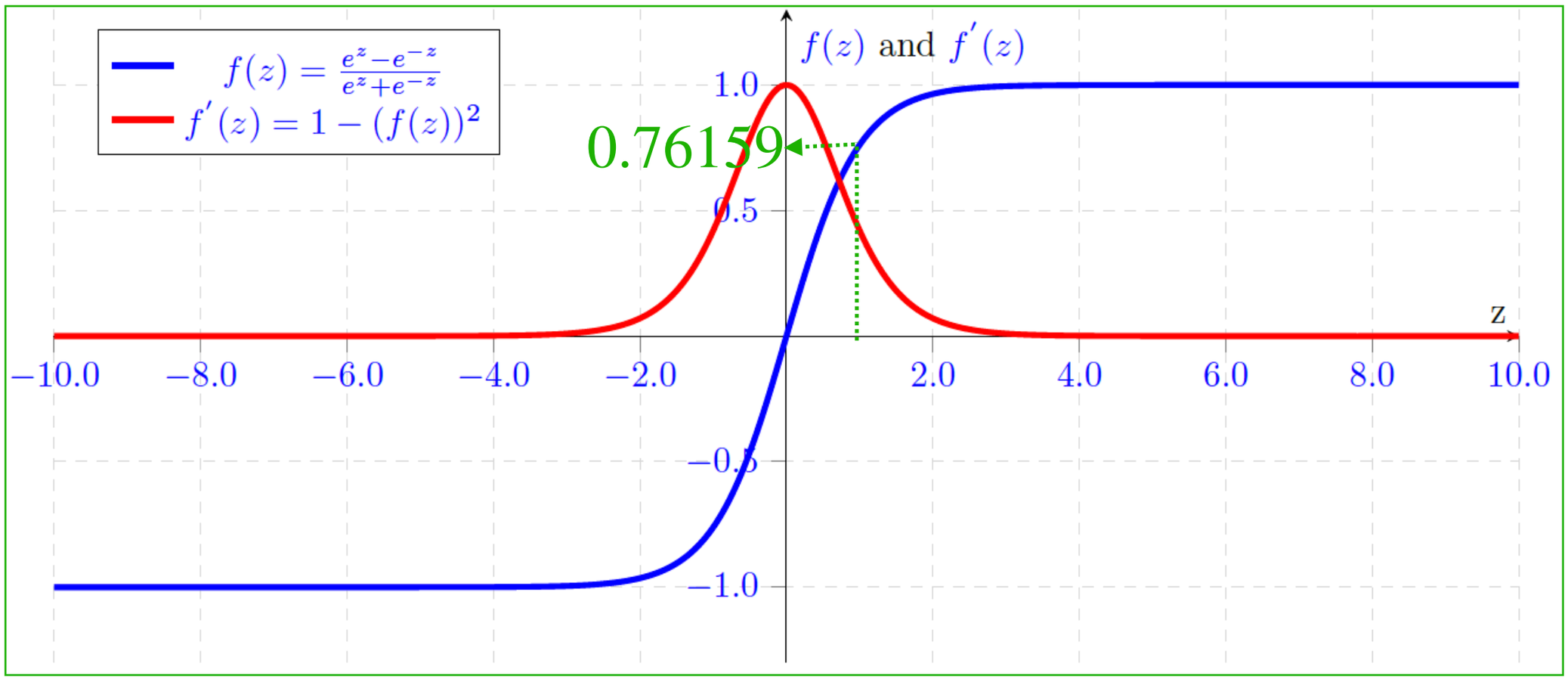
∂b



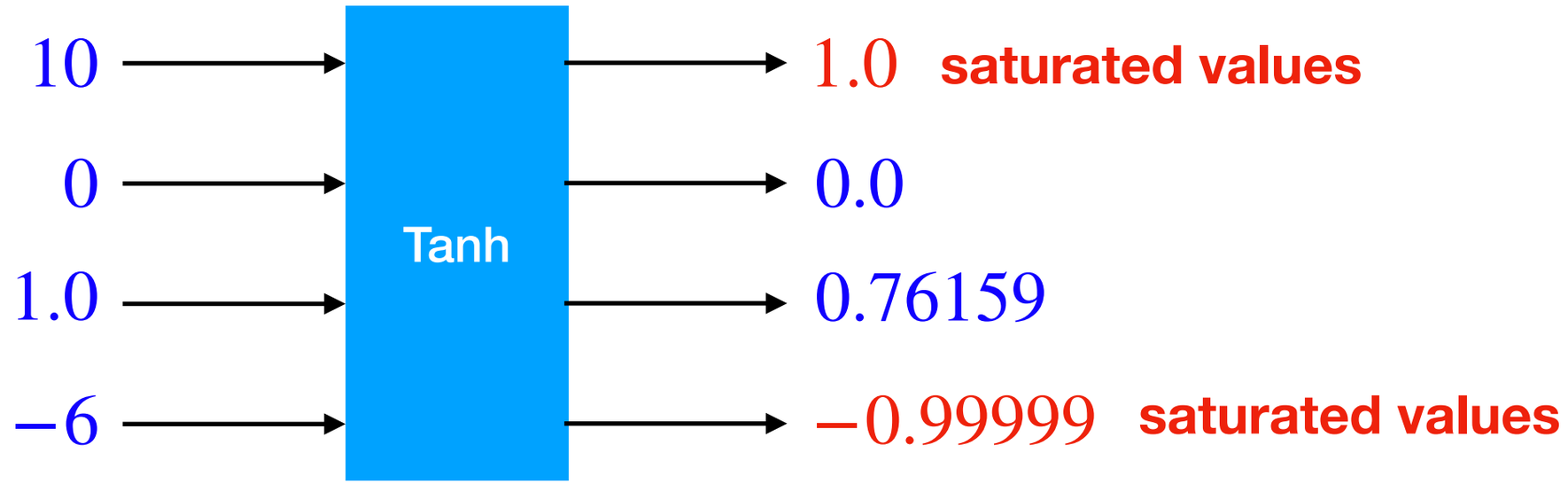
Tanh
Forward-pass



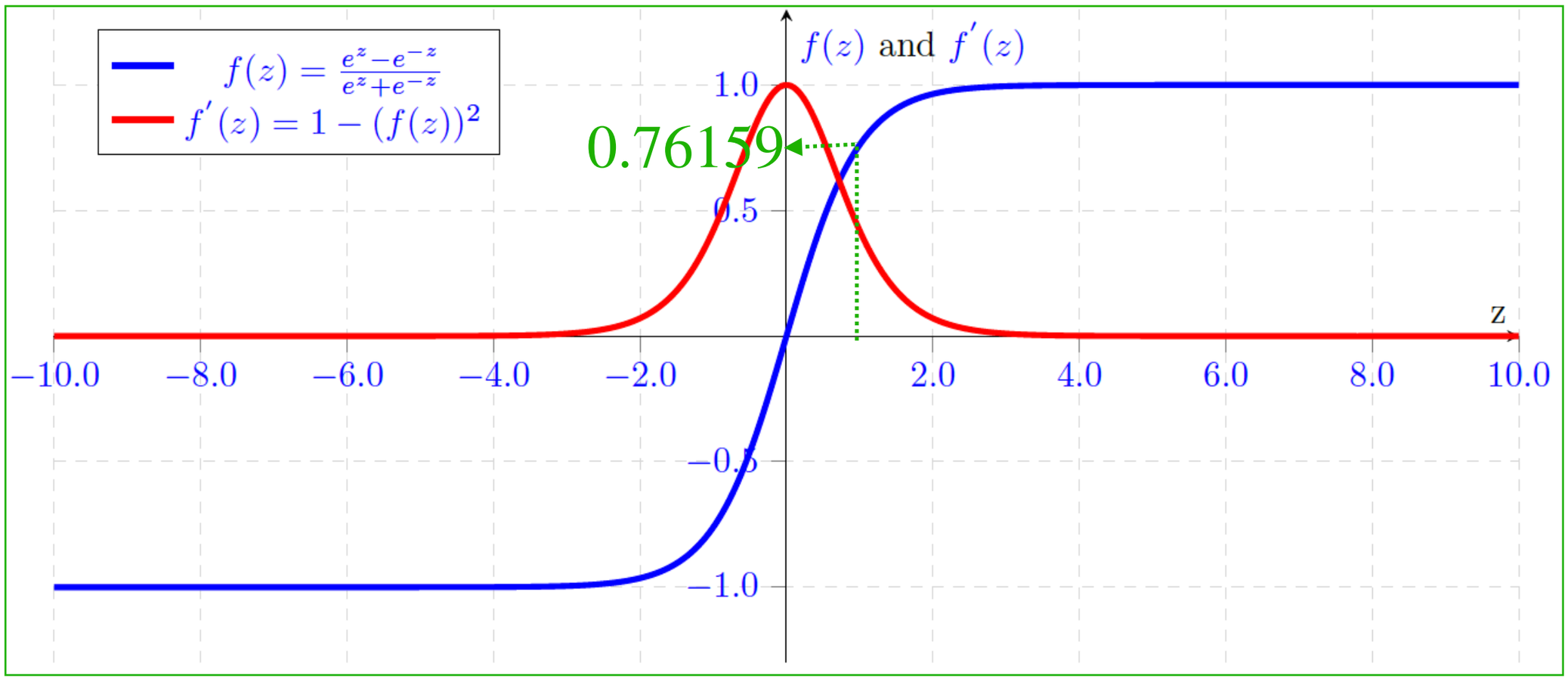
∂b



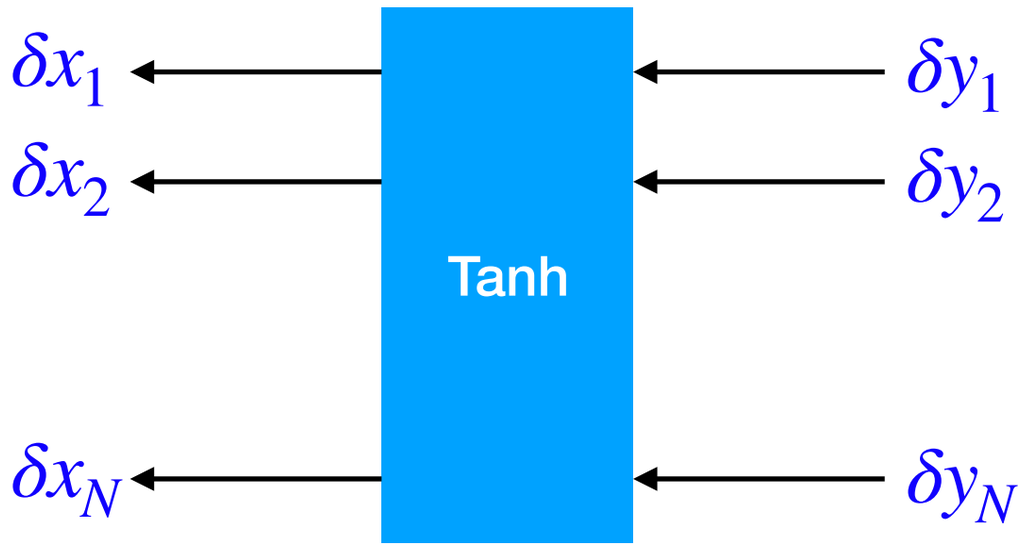
Tanh
Forward-pass



∂b



Tanh
Backward-pass



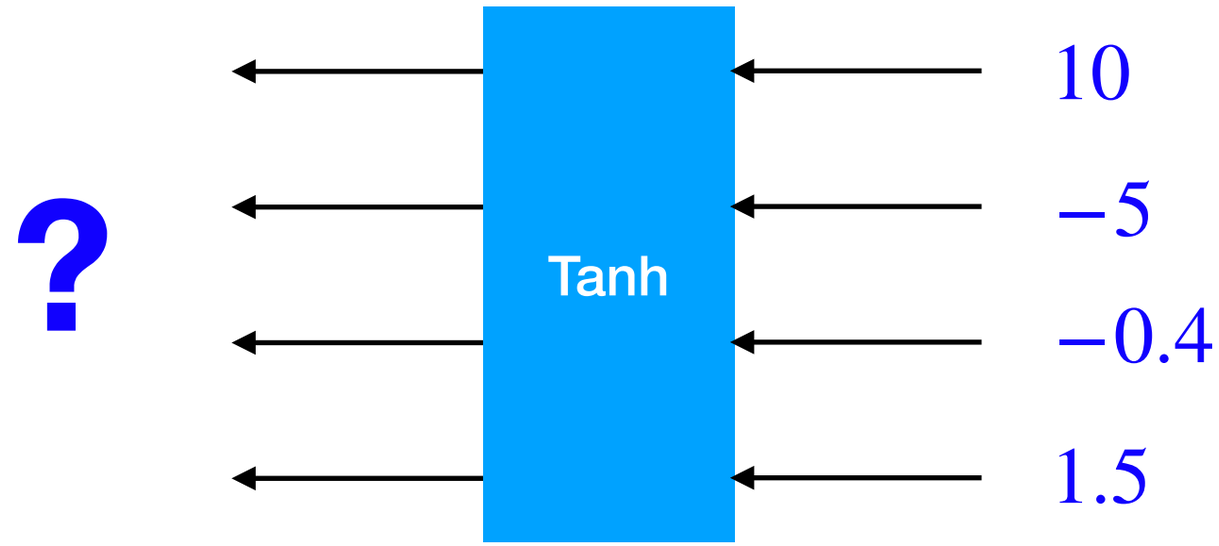
$$y_i = \frac{e^{x_i} - e^{-x_i}}{e^{x_i} + e^{-x_i}}$$

$$y_i^{'} = 1 - y_i^2$$

$$\frac{\partial b}{}$$

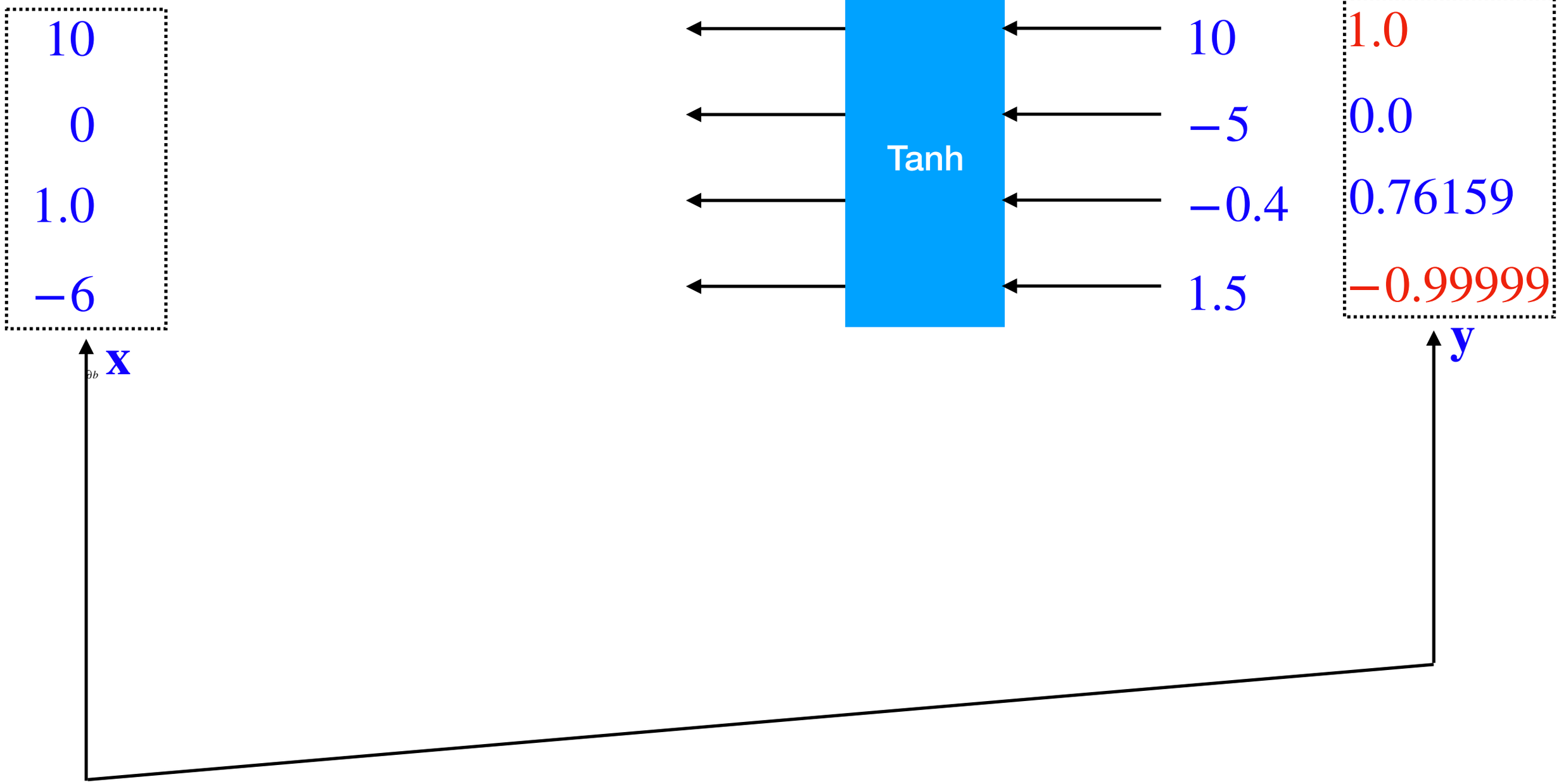
$$\delta x_i = \delta y_i \times \frac{dy_i}{dx_i} = \delta y_i \times (1 - y_i^2)$$

Tanh
Backward-pass



$$\frac{\partial b}{\partial x_i} = \delta y_i \times \frac{dy_i}{dx_i} = \delta y_i \times (1 - y_i^2)$$

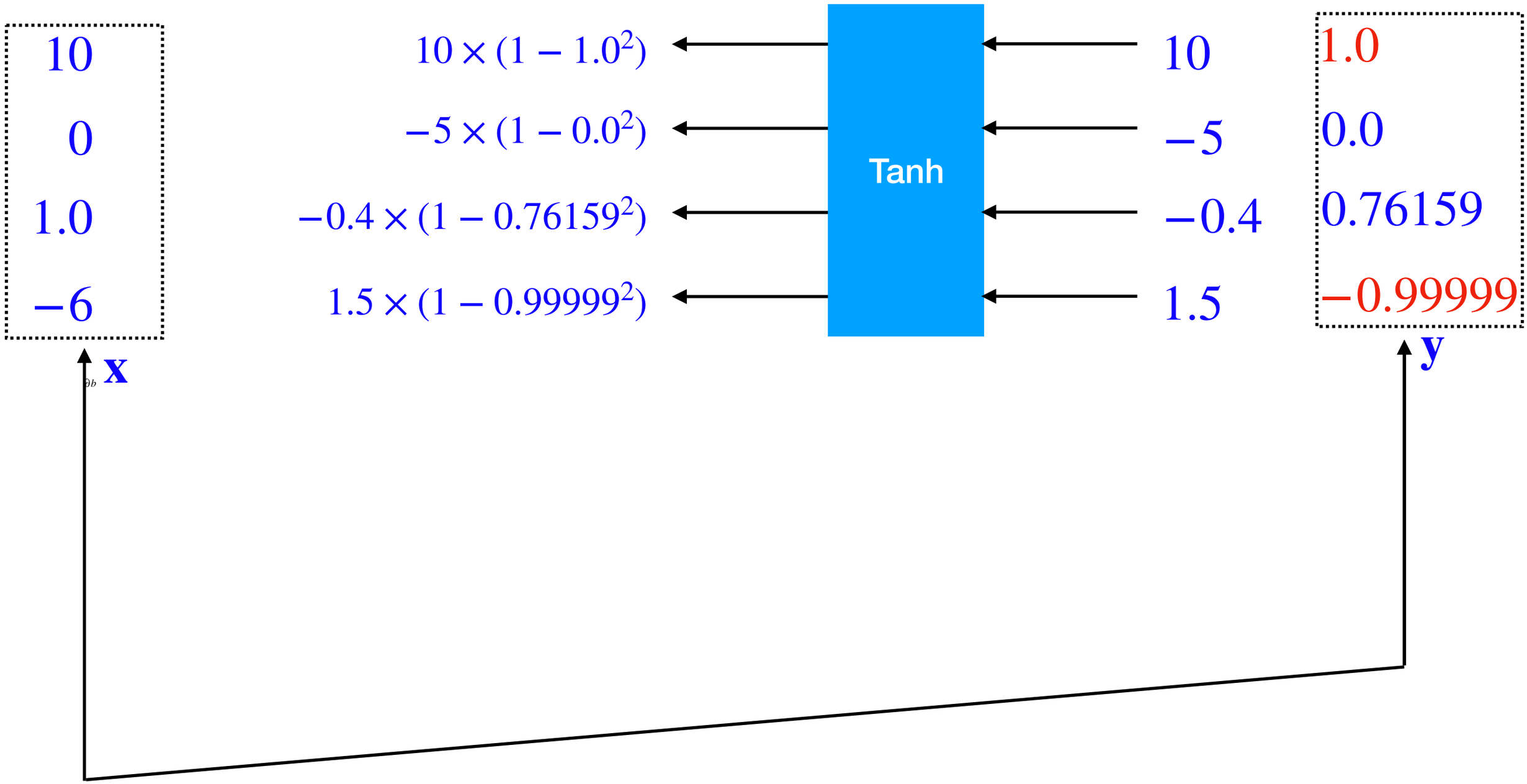
Tanh
Backward-pass



cached during forward-pass

$$\delta x_i = \delta y_i \times \frac{dy_i}{dx_i} = \delta y_i \times (1 - y_i^2)$$

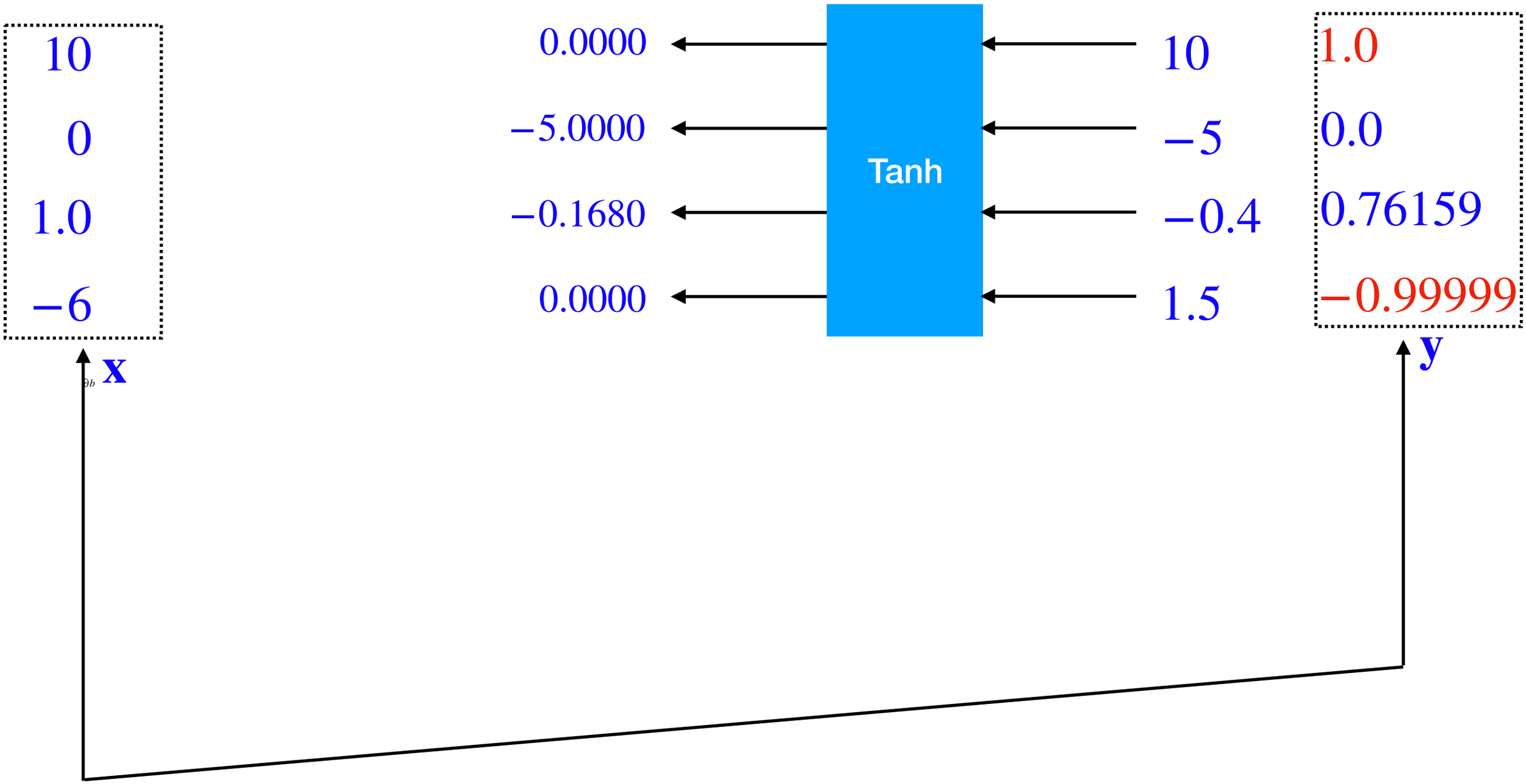
Tanh
Backward-pass



cached during forward-pass

$$\delta x_i = \delta y_i \times \frac{dy_i}{dx_i} = \delta y_i \times (1 - y_i^2)$$

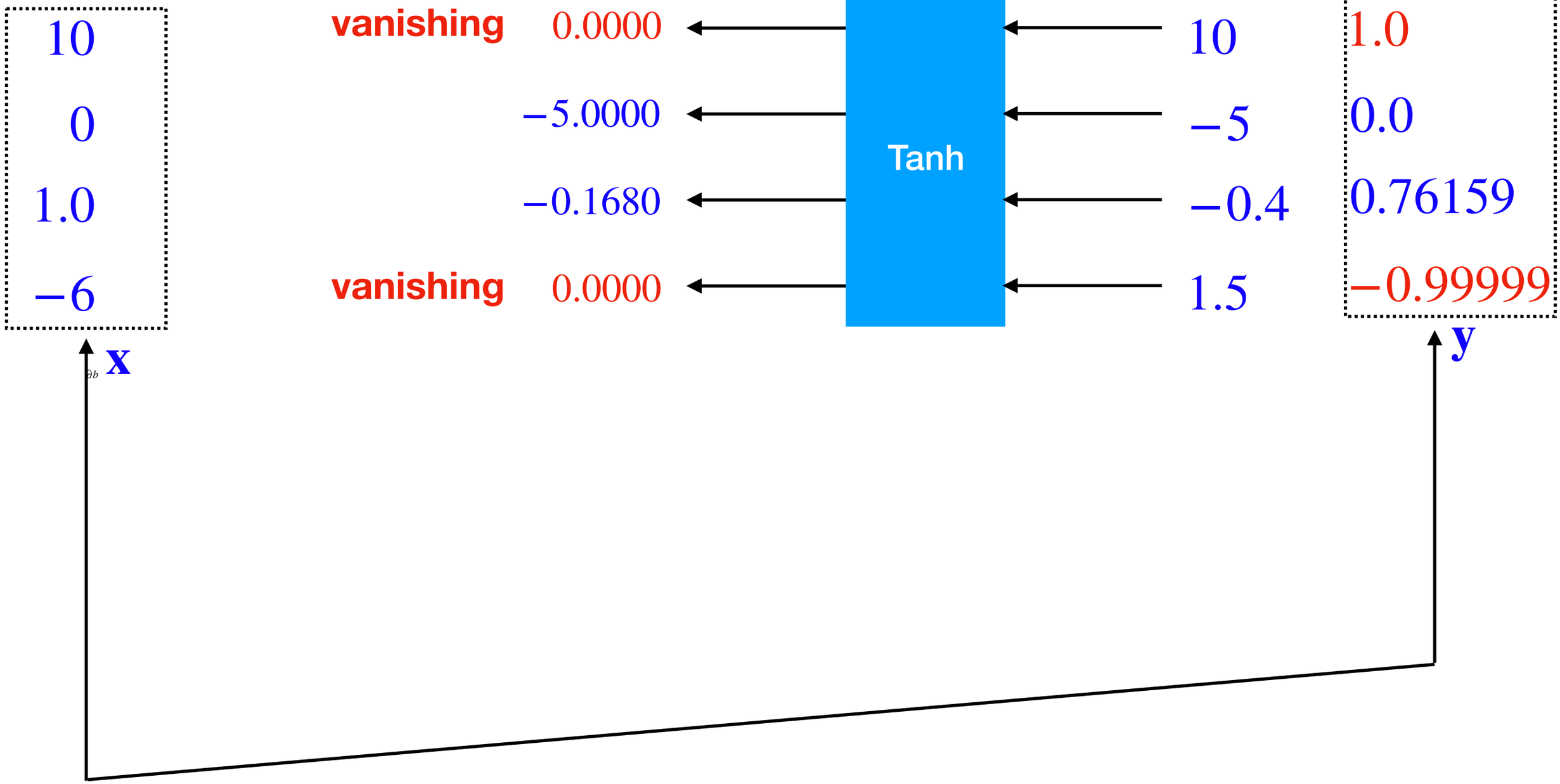
Tanh
Backward-pass



cached during forward-pass

$$\delta x_i = \delta y_i \times \frac{dy_i}{dx_i} = \delta y_i \times (1 - y_i^2)$$

Tanh
Backward-pass

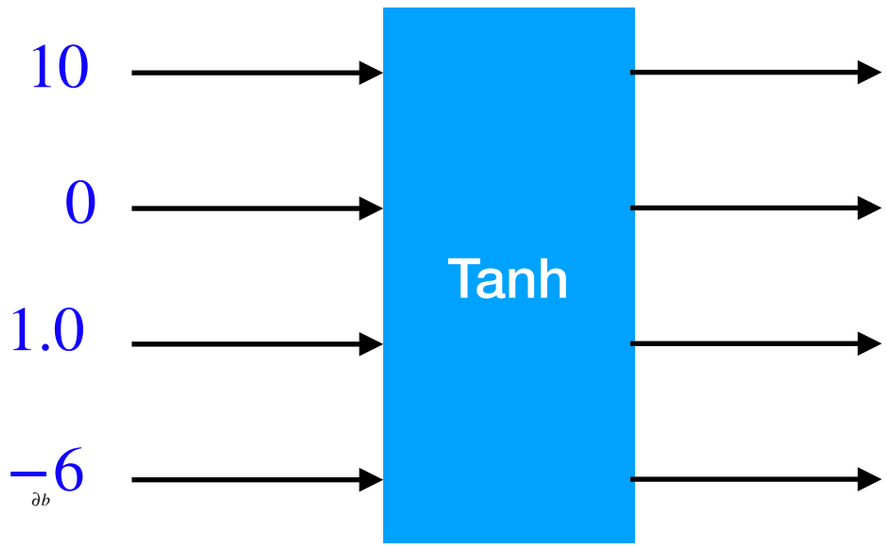


cached during forward-pass

$$\delta x_i = \delta y_i \times \frac{dy_i}{dx_i} = \delta y_i \times (1 - y_i^2)$$

Summary

	Sigmoid	Tanh	ReLU
10	0.99995	1.0	10
0	0.50000	0.0	0.0
1.0	0.73105	0.76159	1.0
-6	0.00247	−0.99999	0



ReLU	Tanh	Sigmoid
10	0.0000	0.0005
−5	−5.0000	−1.2500
−0.4	−0.1680	−0.0786
0	0.0000	0.0037