

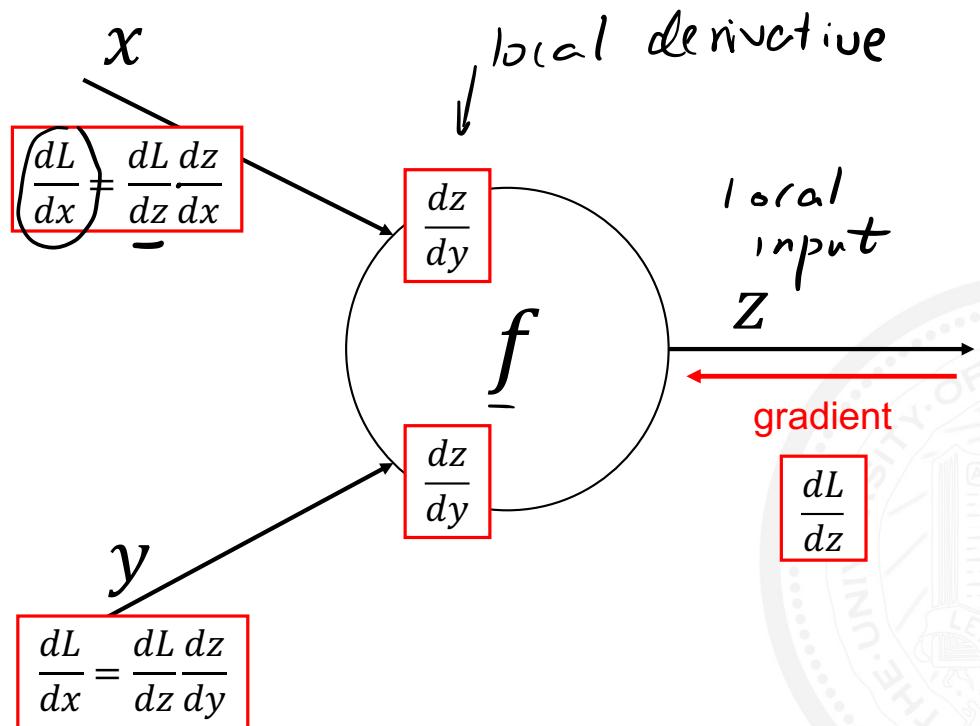
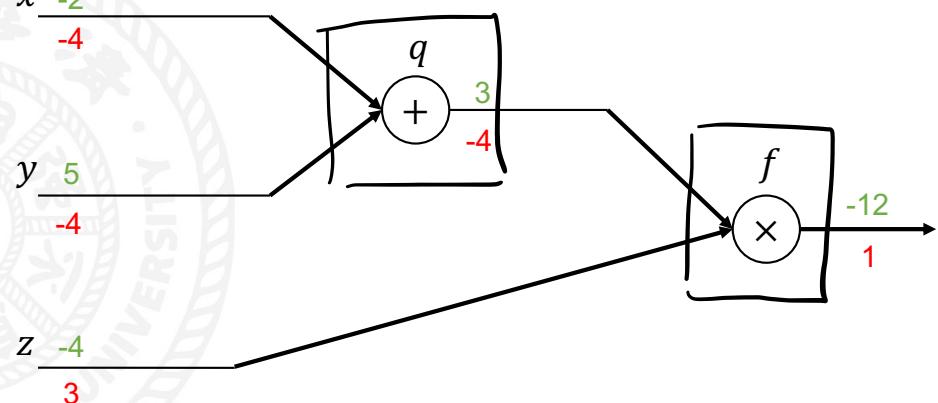
# Backpropagation



TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute

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

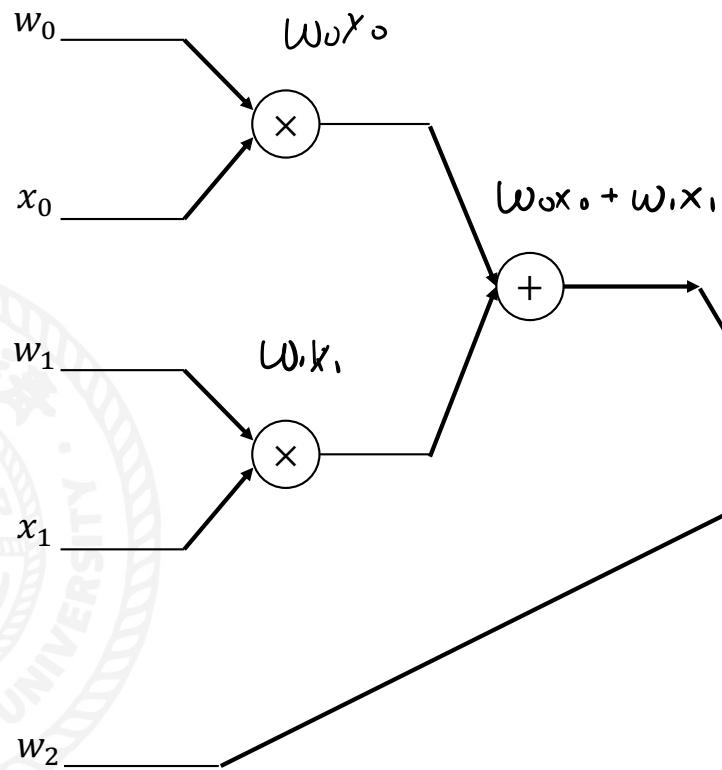
We want  $\frac{df}{dx}, \frac{df}{dy}, \frac{df}{dz}$



# Backpropagation

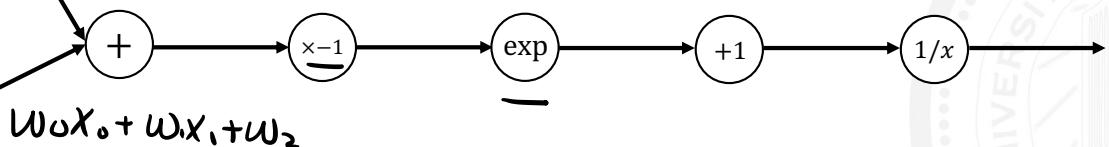


TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute



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

$\begin{bmatrix} x_0 \\ x_1 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix}$



# Backpropagation



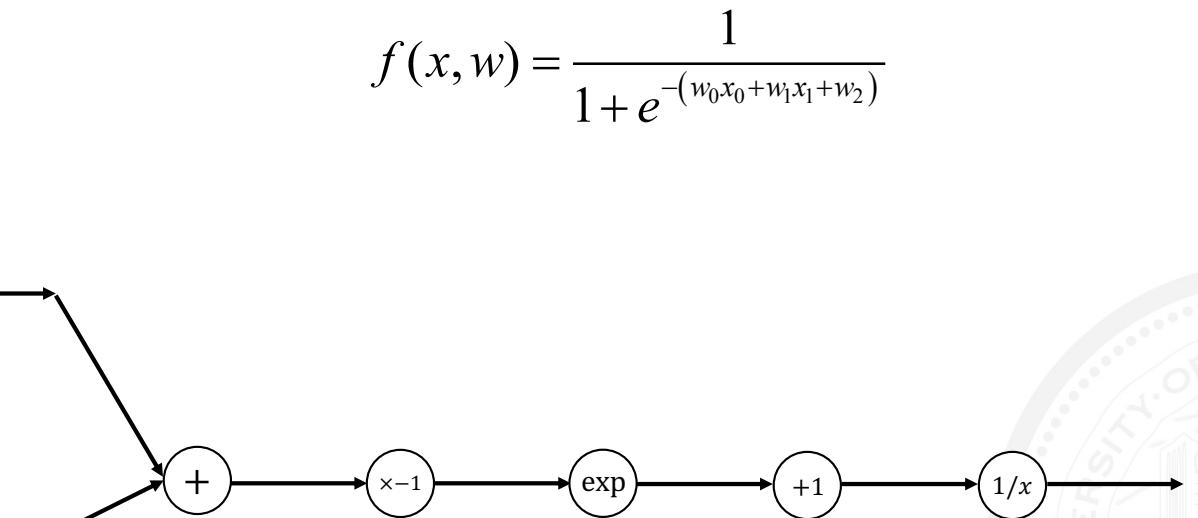
TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute

$$w_0 \underset{2.0}{\text{---}} \xrightarrow{f(x) = ax} \times \quad x_0 \underset{-1.0}{\text{---}}$$

$$w_1 \underset{-3.0}{\text{---}} \xrightarrow{f(x) = ax} \times \quad x_1 \underset{-2.0}{\text{---}}$$

$$w_2 \underset{-3.0}{\text{---}}$$

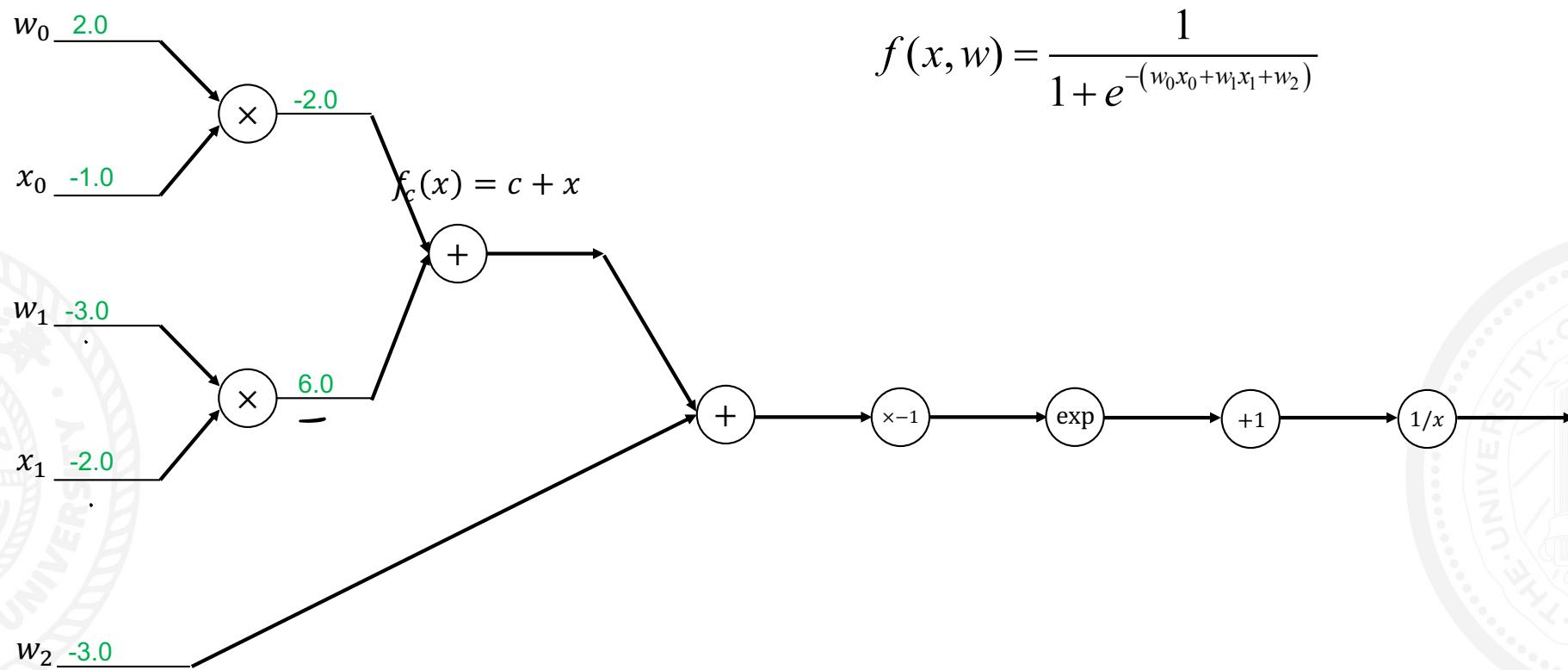
$$f(x, w) = \frac{1}{1 + e^{-(w_0x_0 + w_1x_1 + w_2)}}$$



# Backpropagation



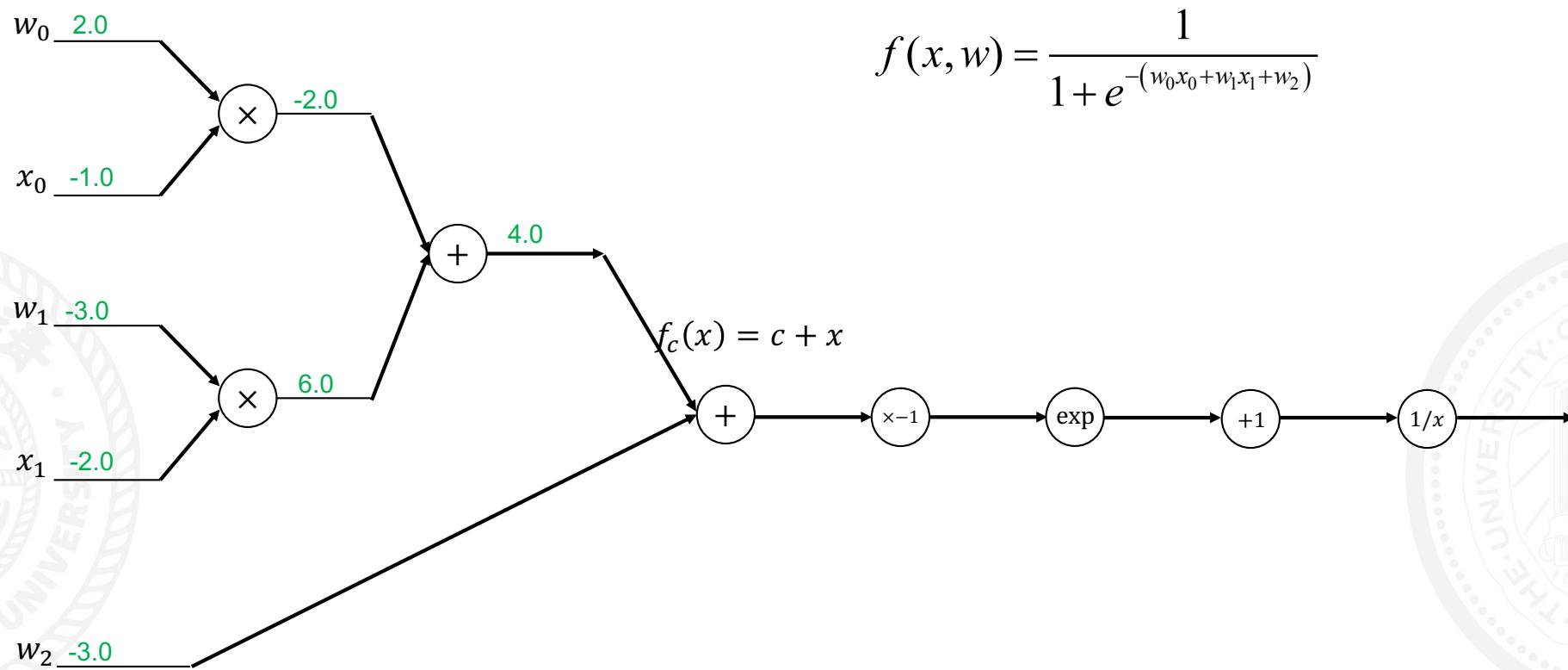
TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute



# Backpropagation



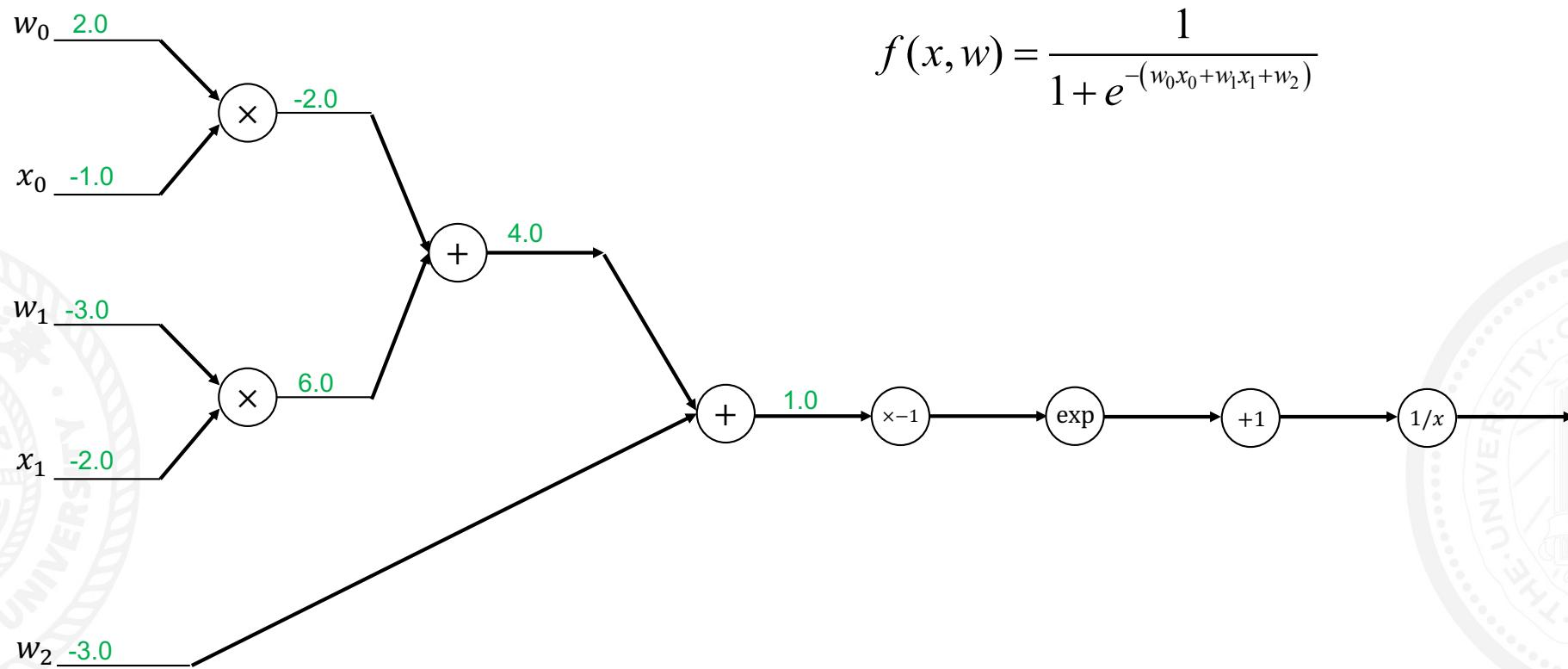
TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute



# Backpropagation



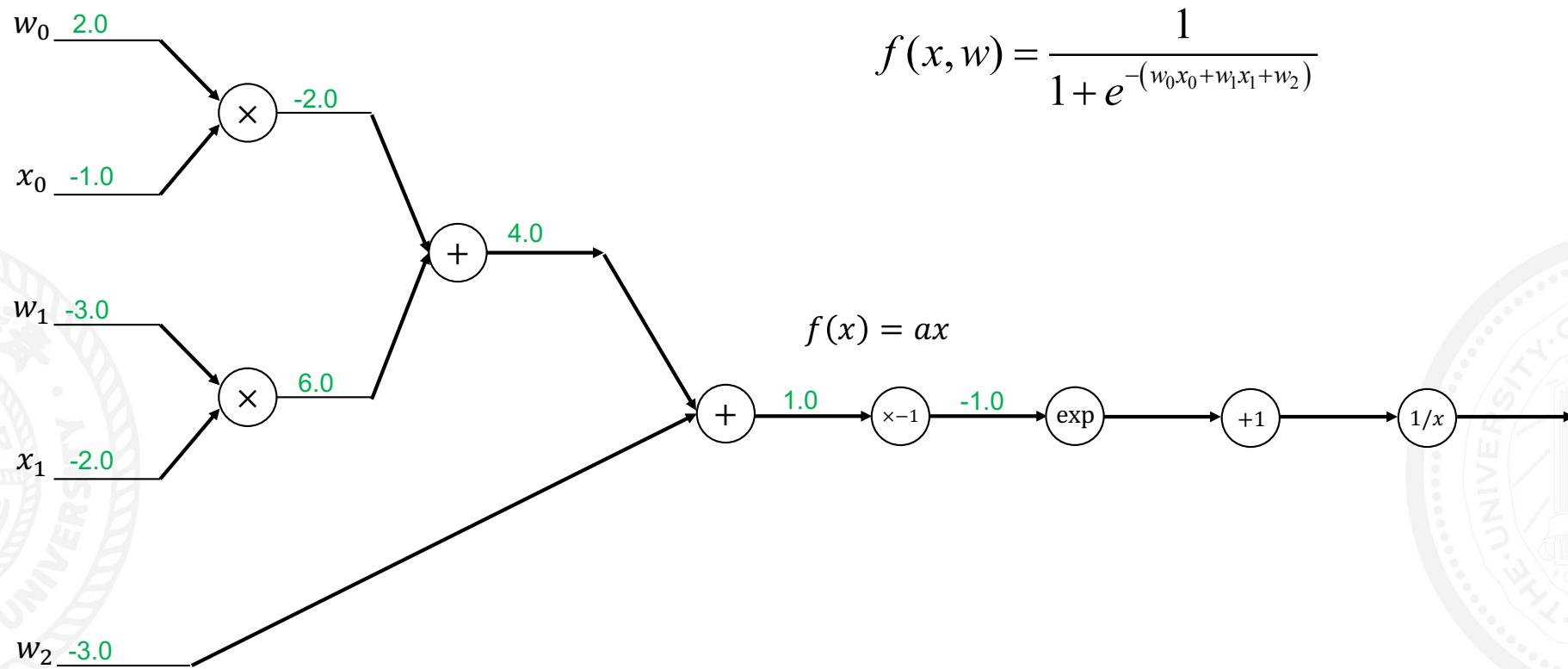
TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute



# Backpropagation



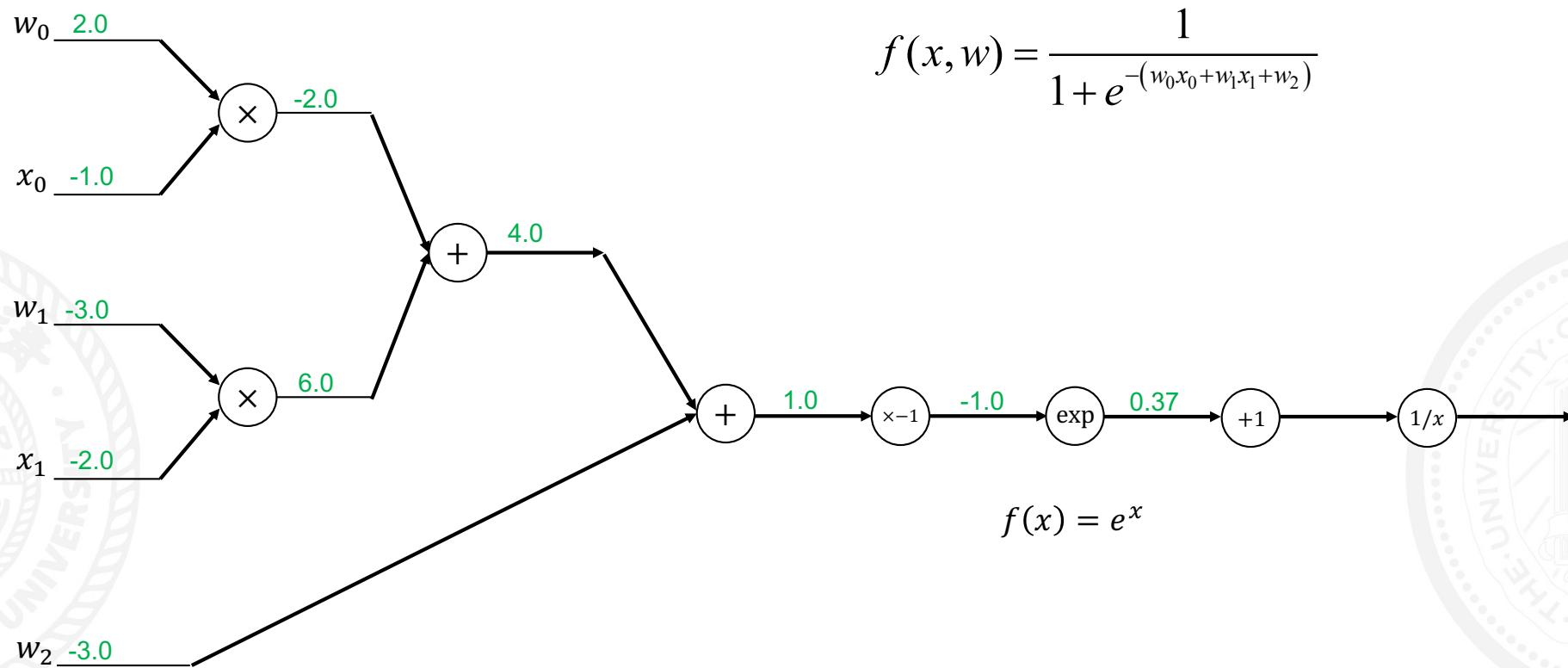
TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute



# Backpropagation



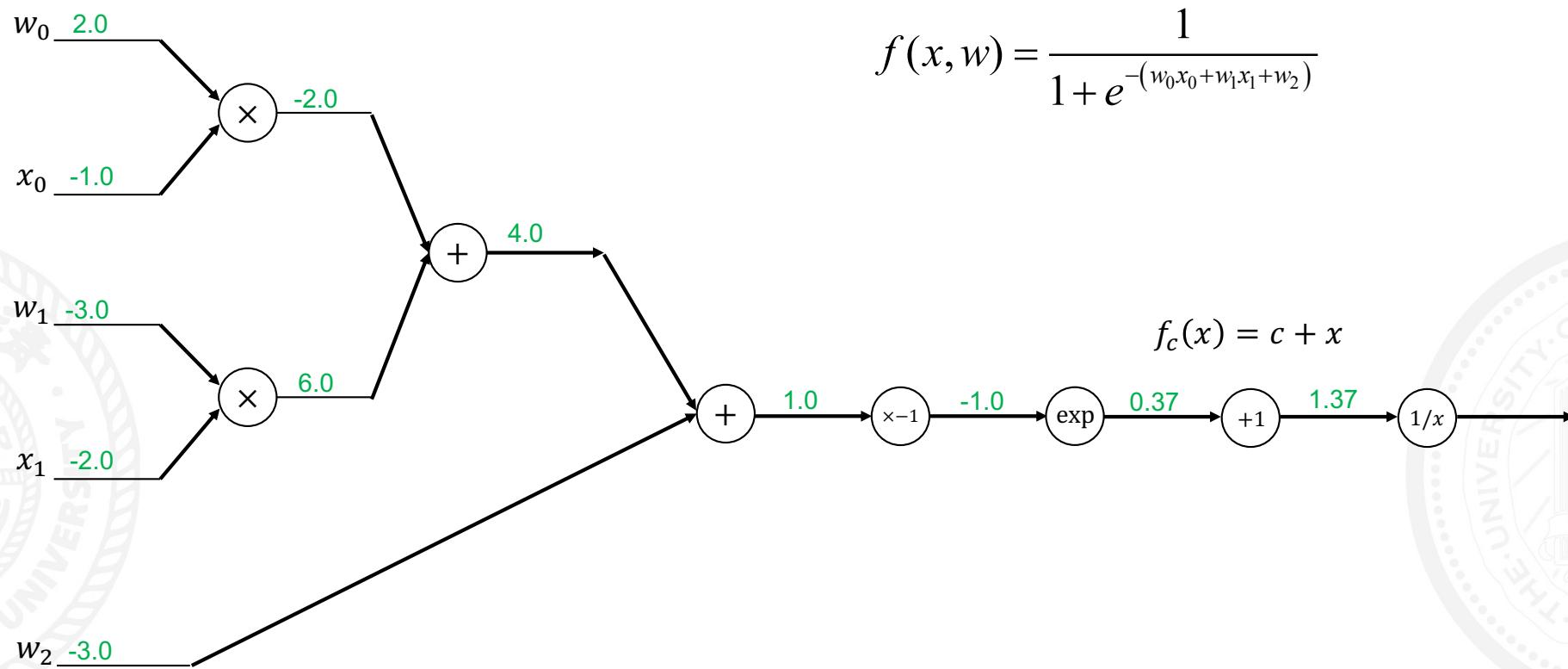
TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute



# Backpropagation



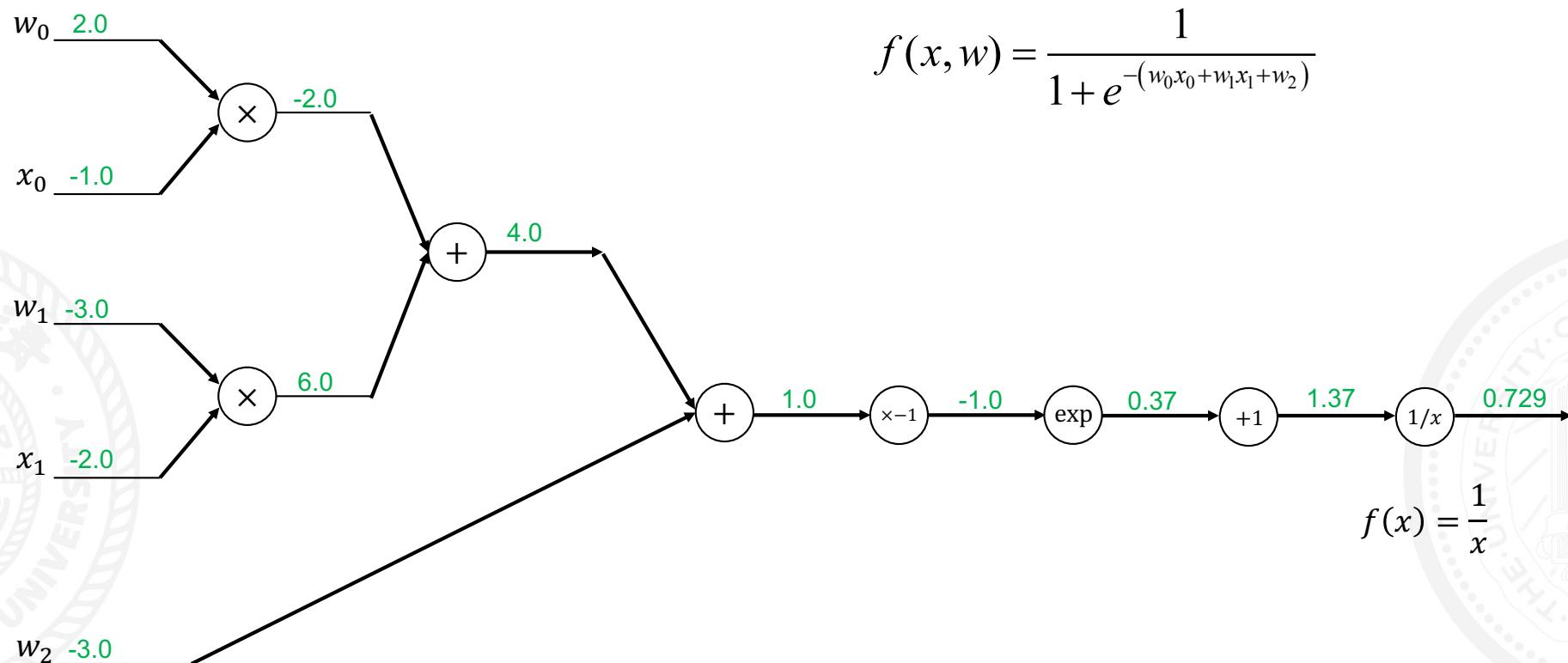
TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute



# Backpropagation



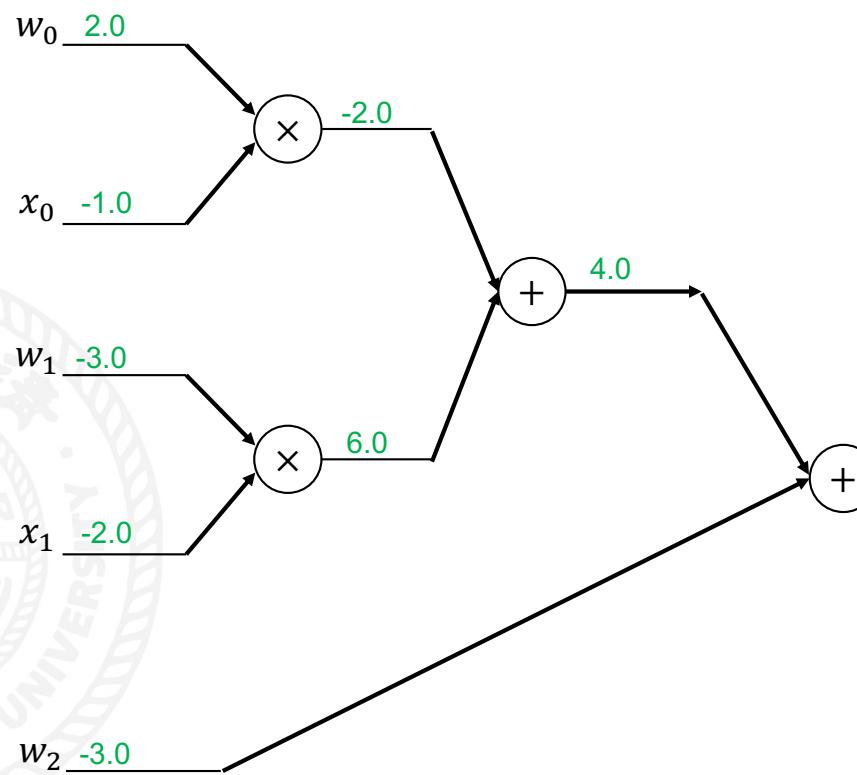
TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute



# Backpropagation



TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute



$$f(x, w) = \frac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}} \quad \left| \begin{matrix} x = \begin{bmatrix} -1 \\ -2 \\ -3 \end{bmatrix} & w = \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix} \end{matrix} \right.$$

$$\frac{dL}{df} = 1$$

$$f(x) = \frac{1}{x}$$

$$\frac{dL}{dx} = \frac{1}{x^2}$$

$$\frac{\partial f}{\partial x} = -1/x^2$$

$$= -\frac{1}{1.37^2}$$

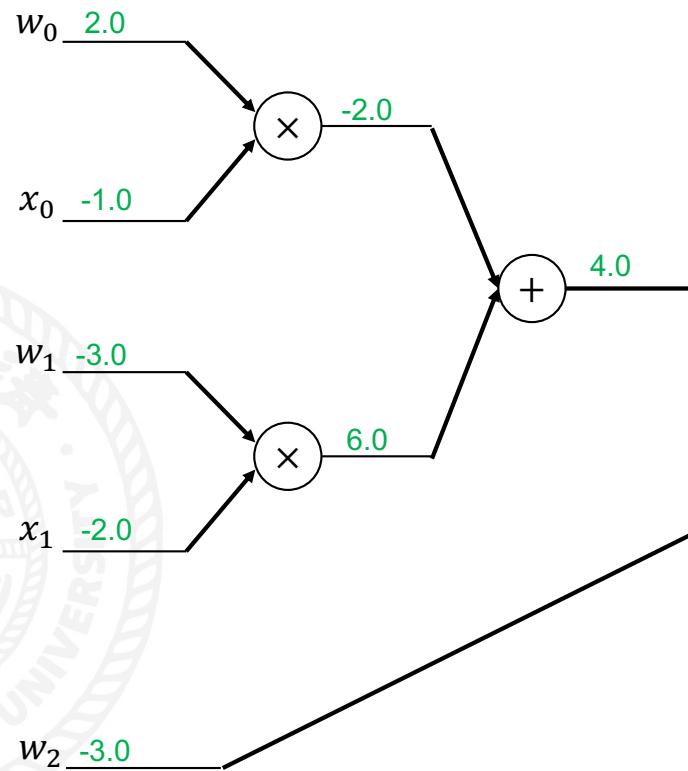
$$\frac{dL}{dx} = \frac{dL}{df} \cdot \frac{df}{dx}$$

$$= -0.53 \cdot 1 = -0.53$$

# Backpropagation



TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute



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

$$f_c(x) = c + x$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x} &= \left( \frac{\partial f_c}{\partial x} = 1 \right) \cdot \frac{\partial \mathcal{L}}{\partial f} \\ &= -0.53 \end{aligned}$$

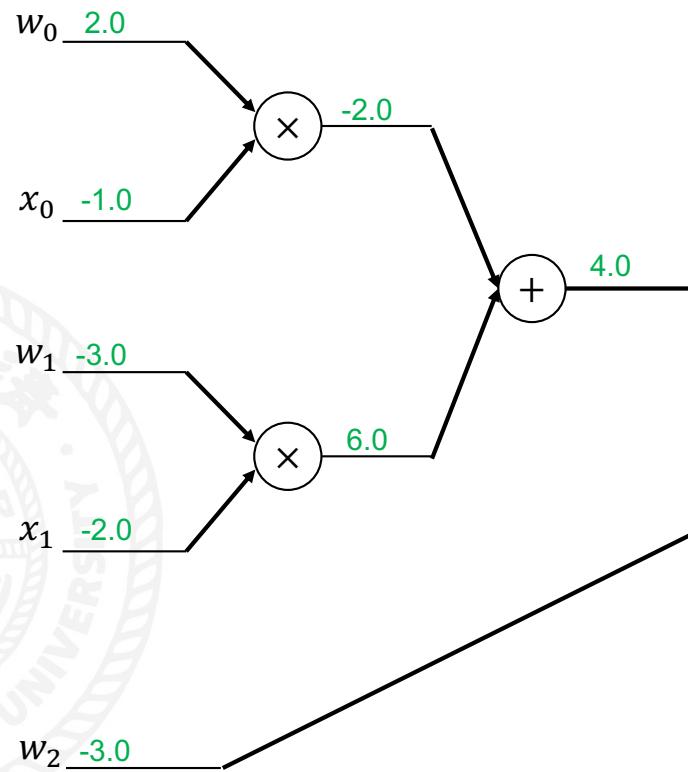
$$0.729$$

$$1.0$$

## Backpropagation



 TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute



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

$$f(x) = e^x$$

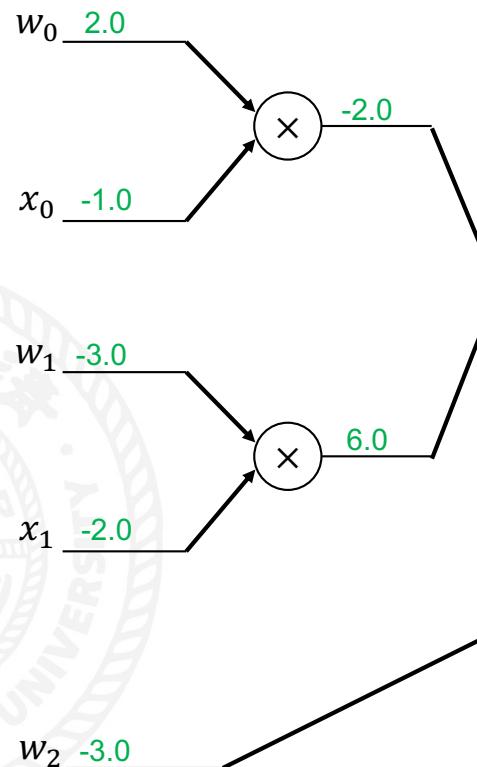
$$\frac{\partial f}{\partial x} = \left( e^x \right) = e^{-1.0} = \frac{1}{e}$$

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial f} \frac{\partial f}{\partial x} = -0.53 \cdot \frac{1}{e}.$$

# Backpropagation



TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute



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

$$f(x) = ax$$

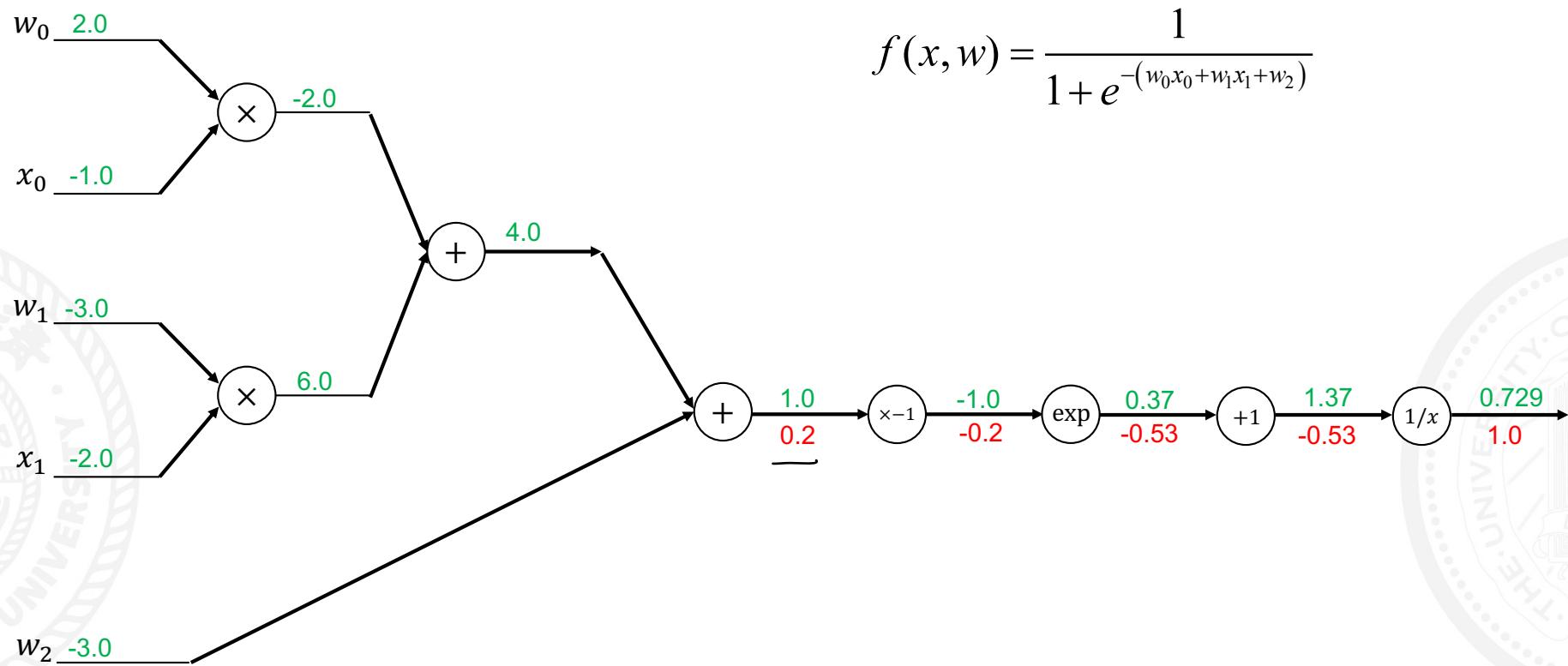
$$\frac{\partial f}{\partial x} = \alpha$$

-1 • (-0.2)

# Backpropagation



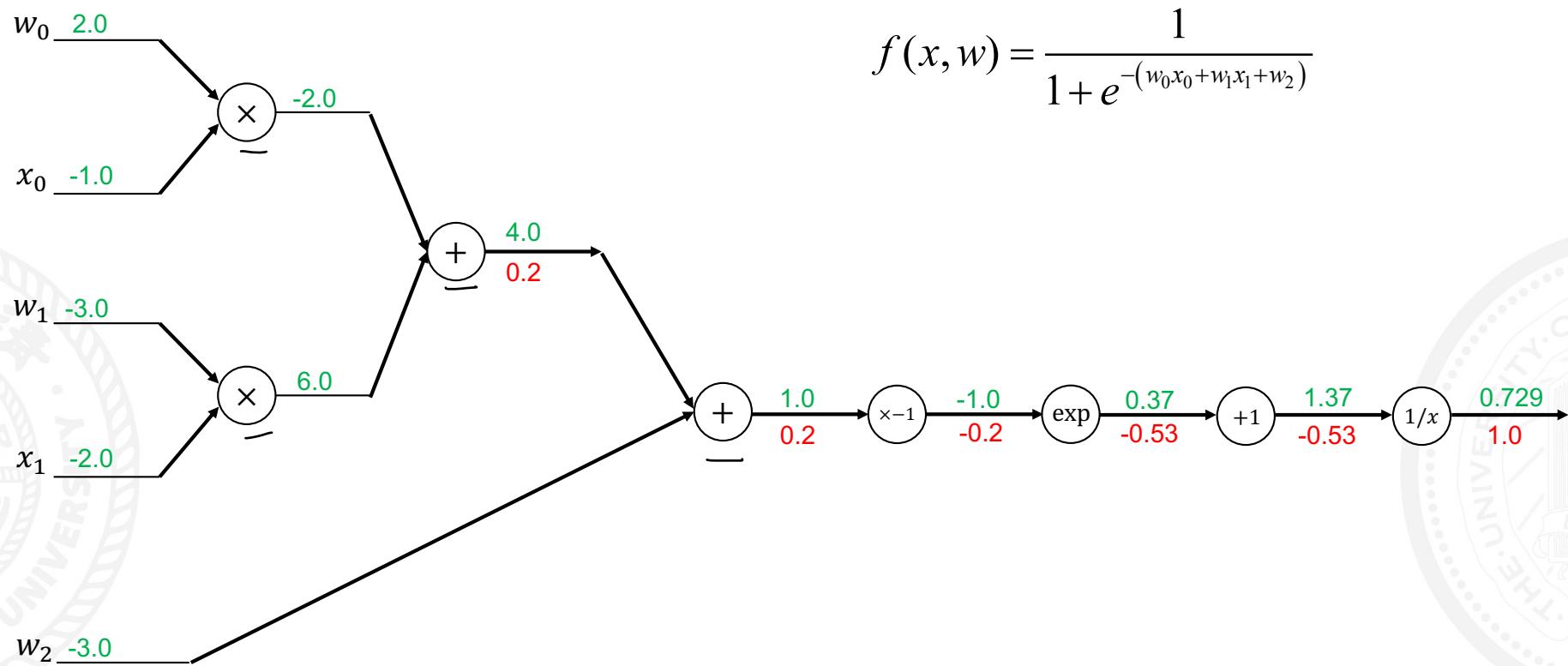
TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute



# Backpropagation



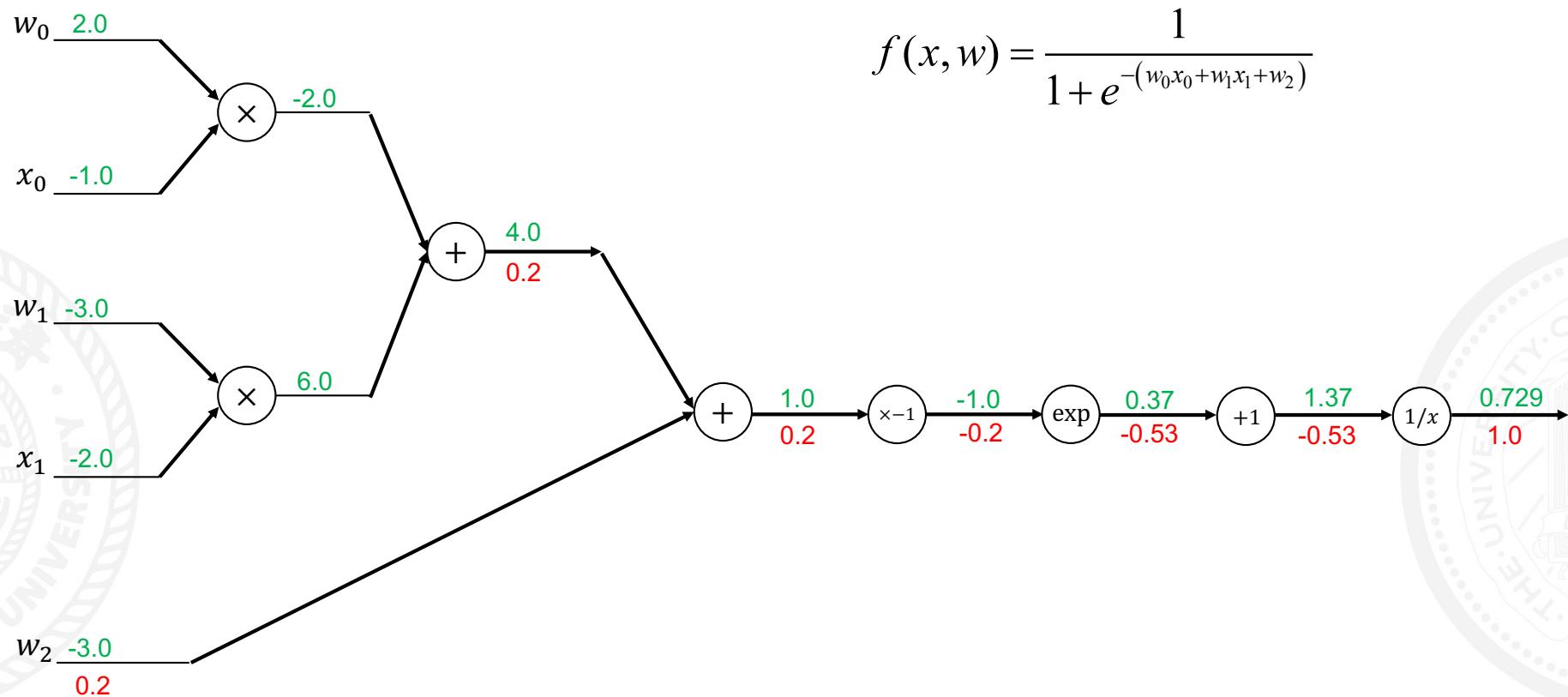
TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute



# Backpropagation



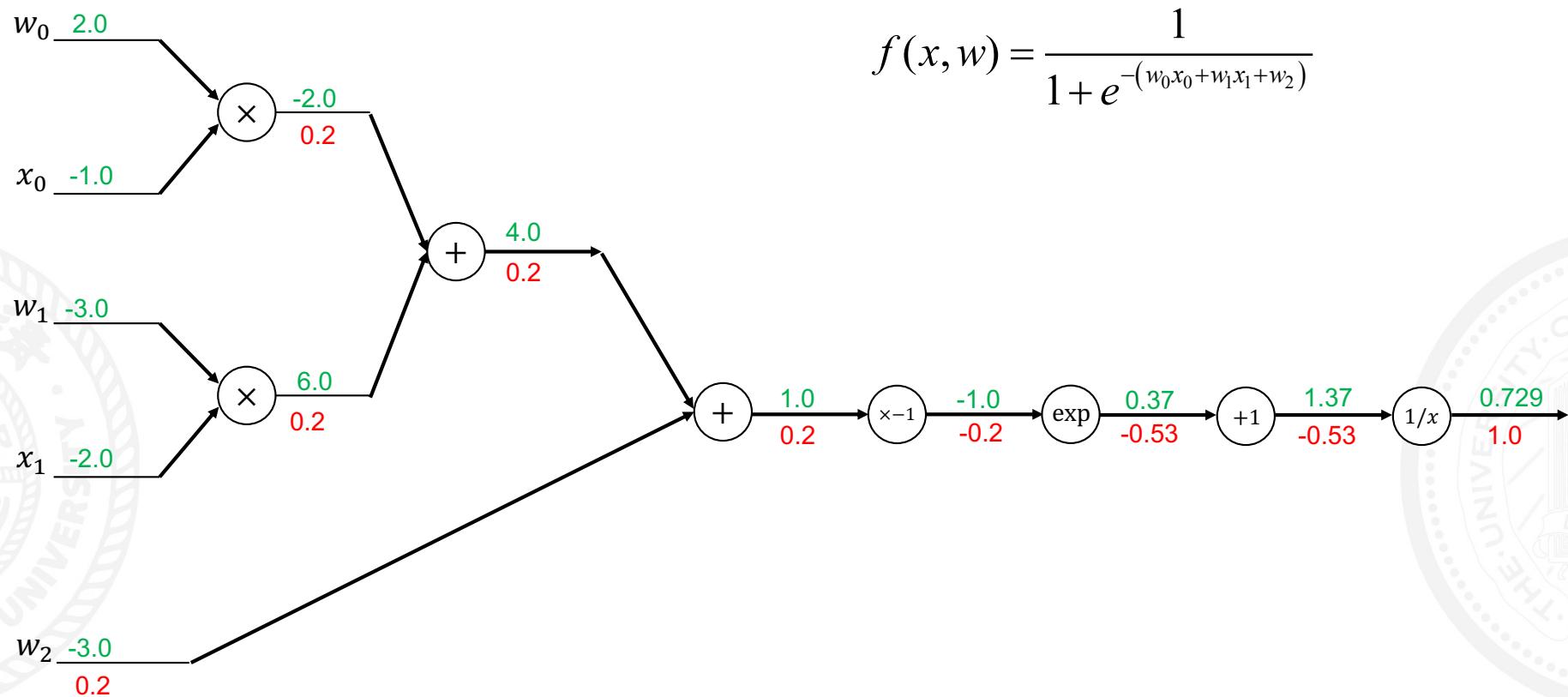
TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute



# Backpropagation



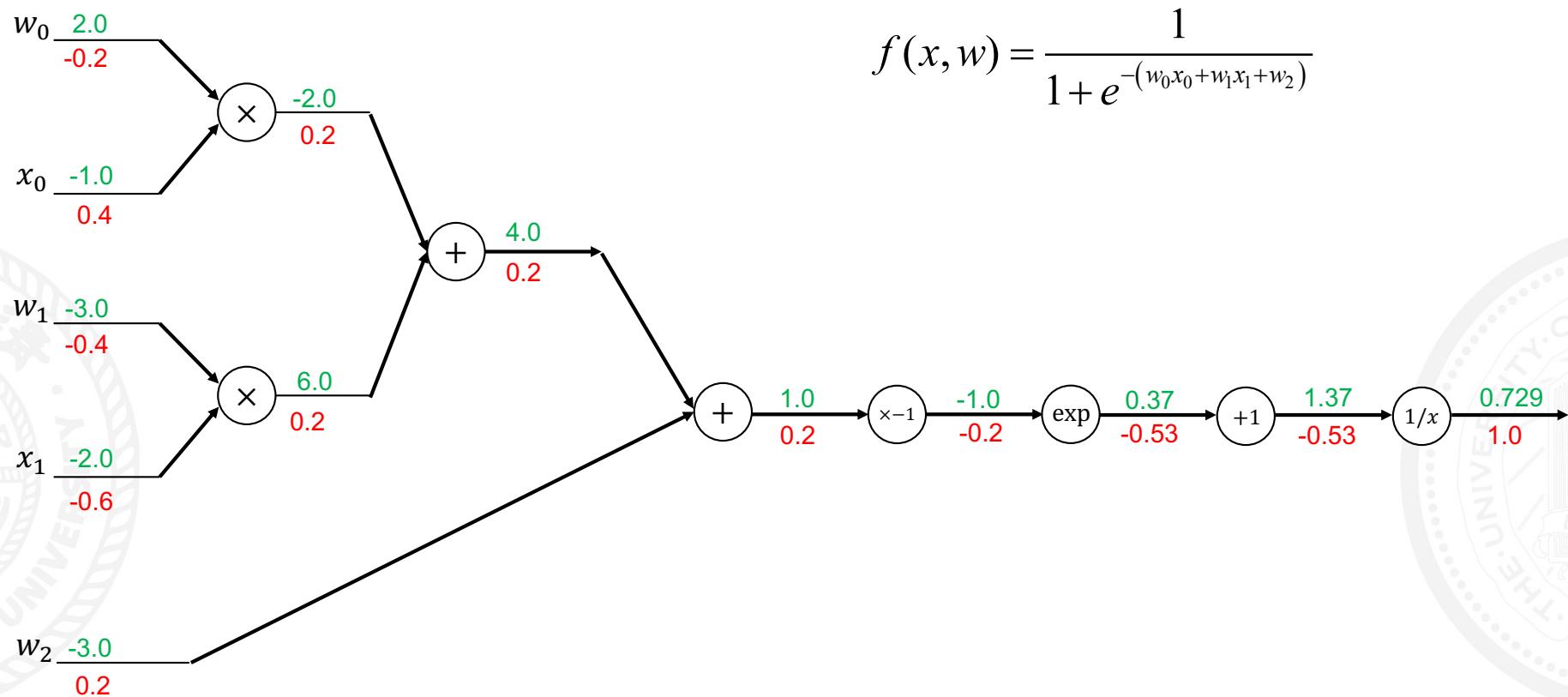
TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute



# Backpropagation



TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute

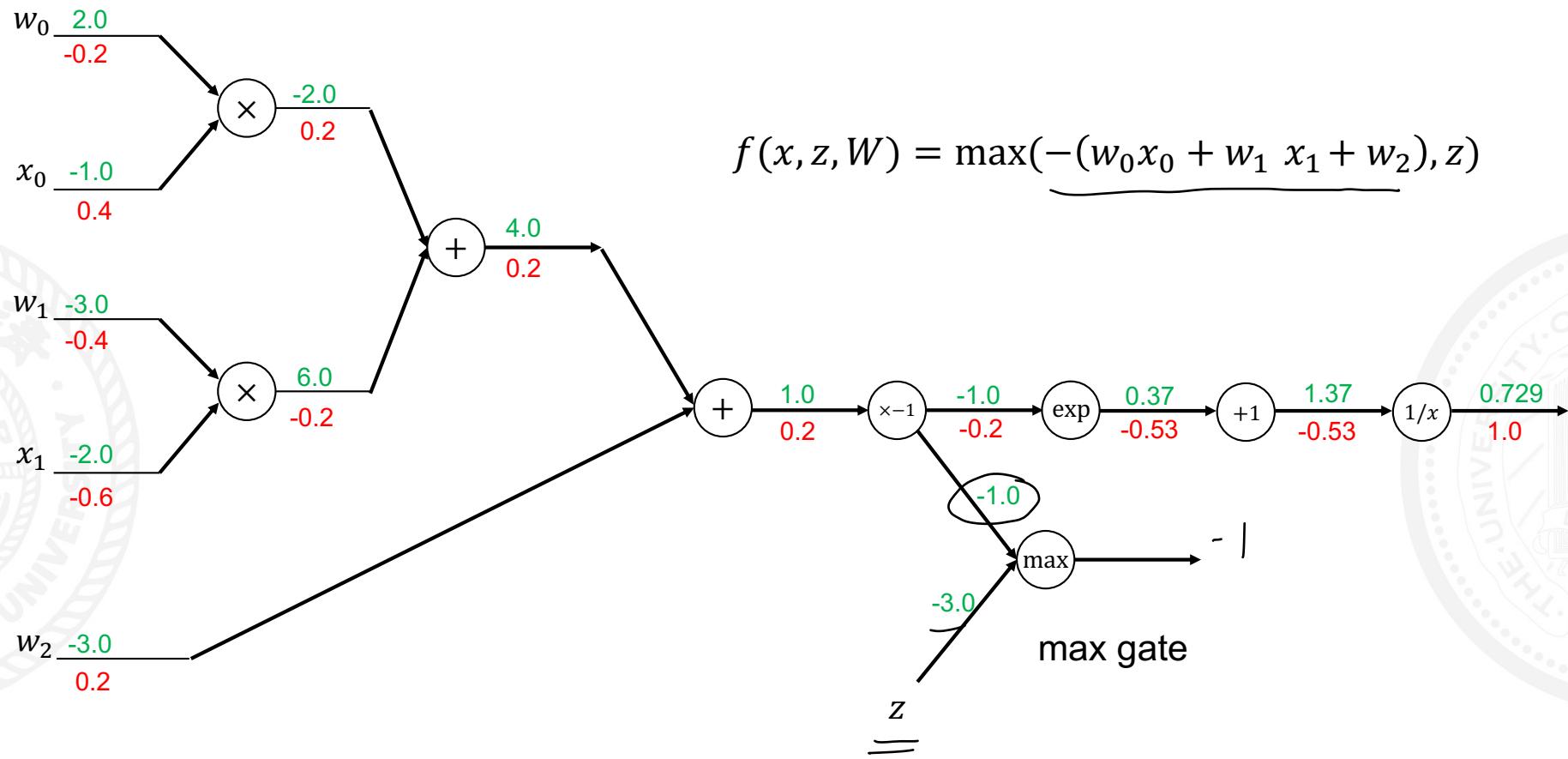


# Backpropagation



TBSI

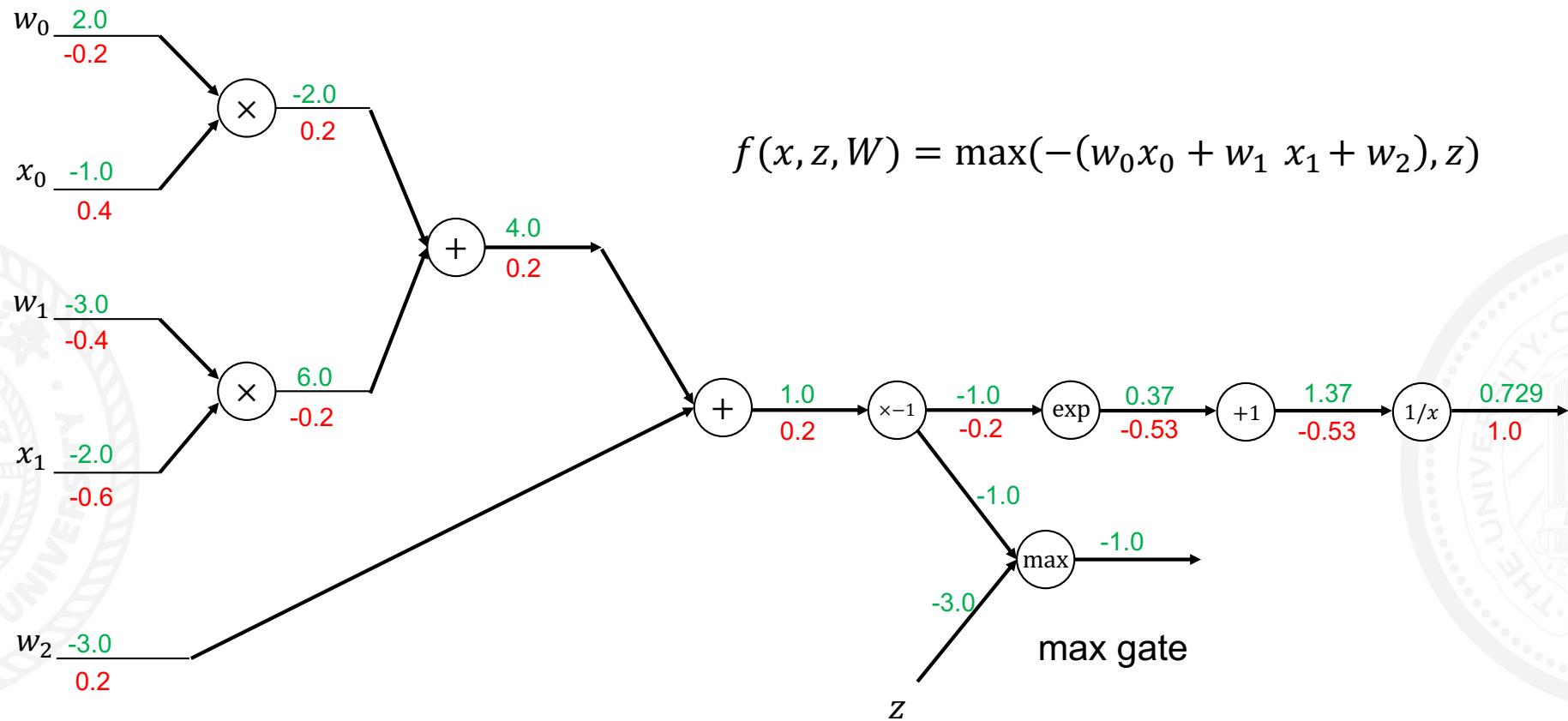
Tsinghua-Berkeley Shenzhen Institute



# Backpropagation



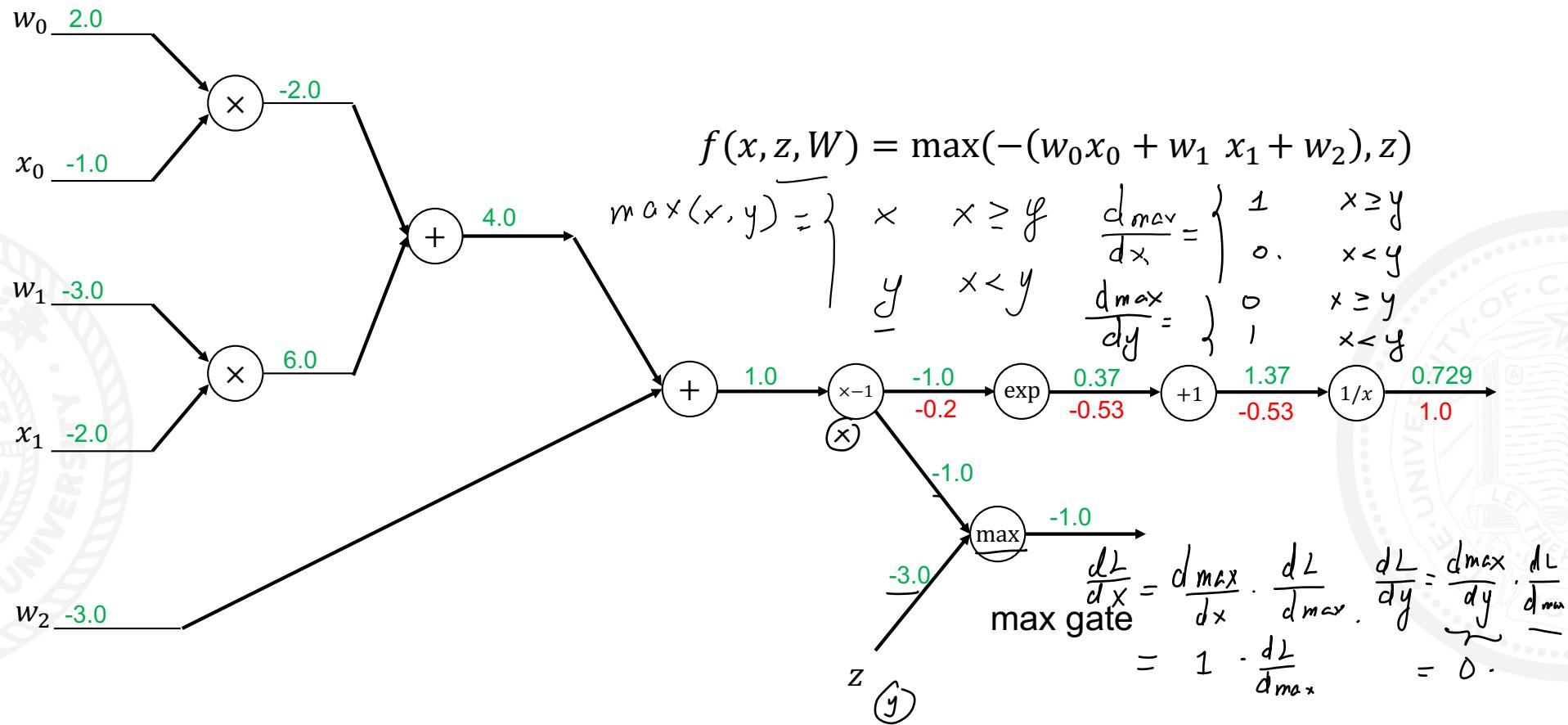
TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute



# Backpropagation



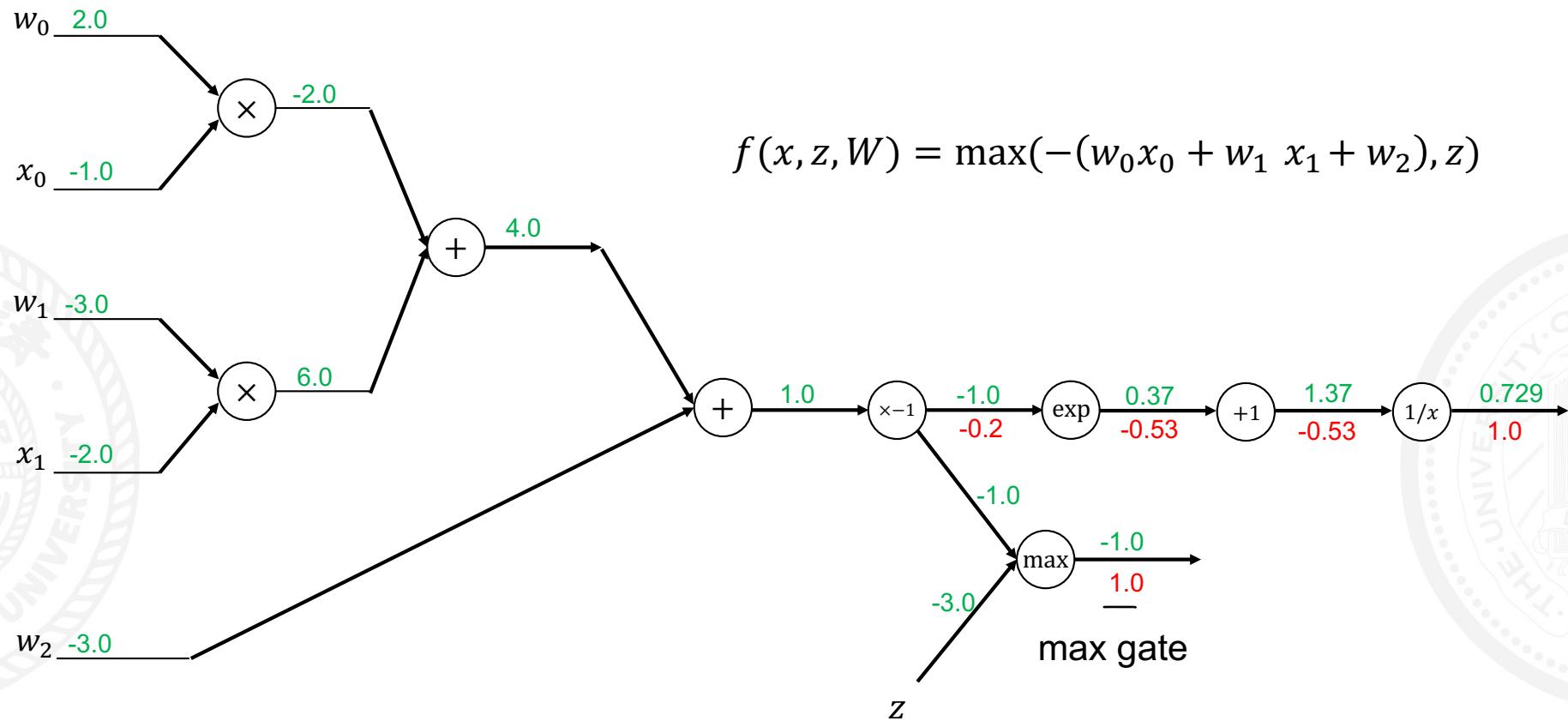
TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute



# Backpropagation



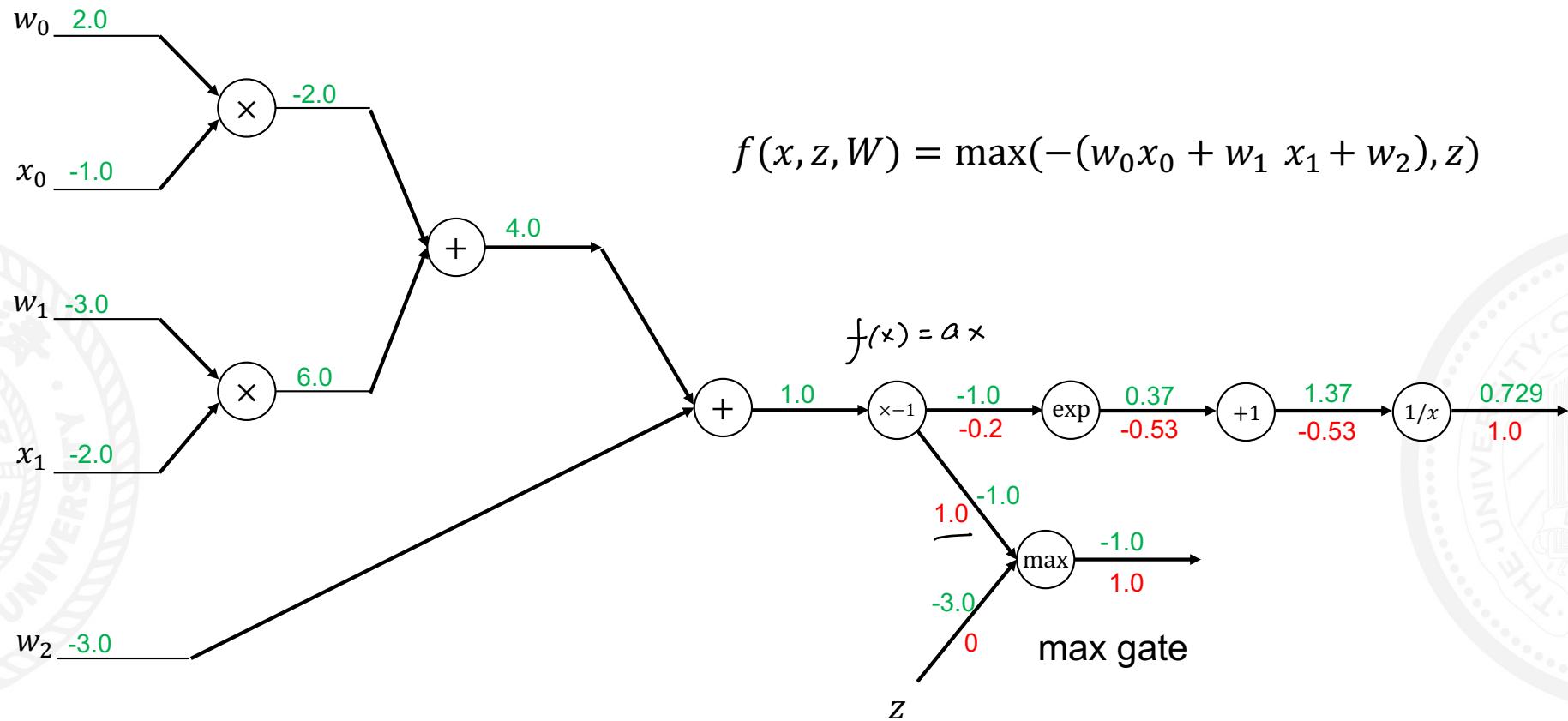
TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute



# Backpropagation



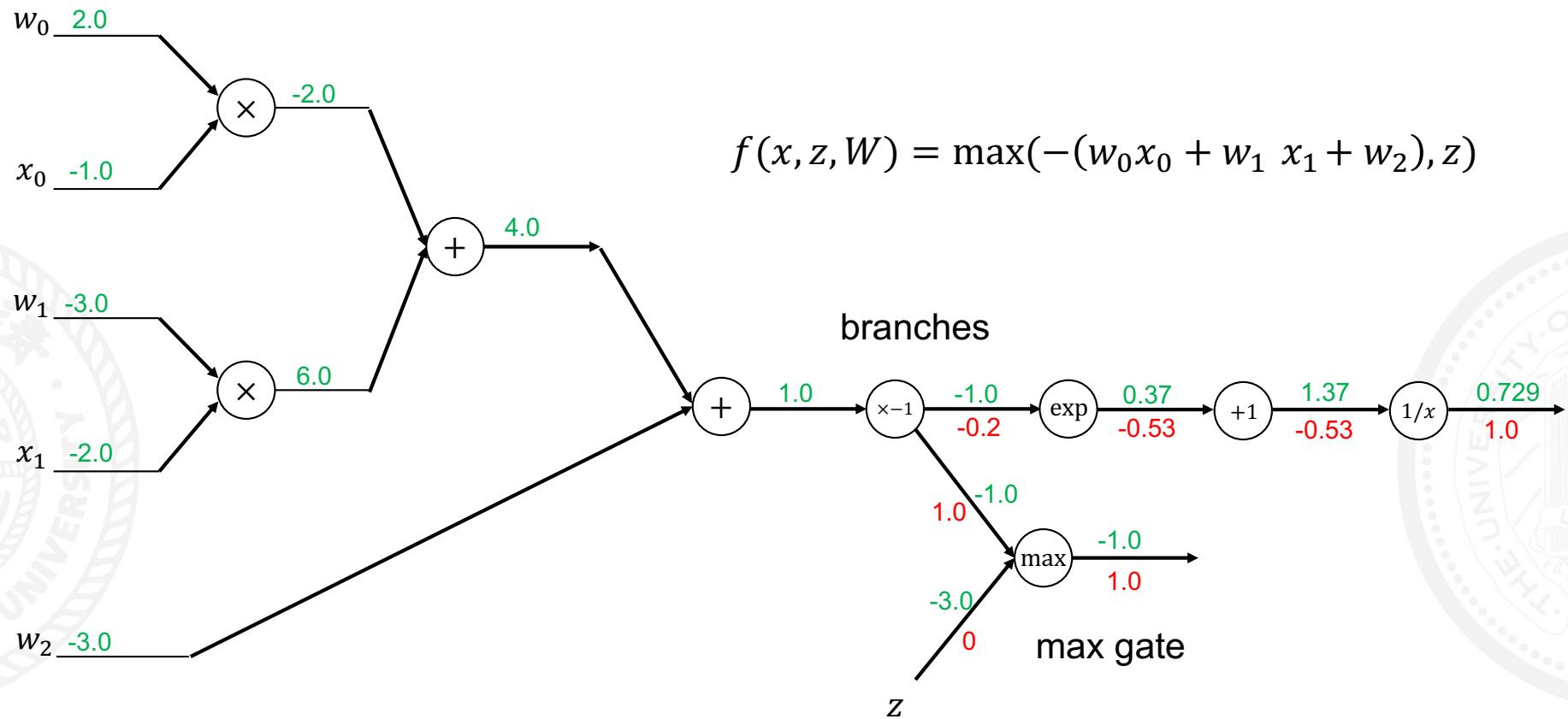
TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute



## Backpropagation



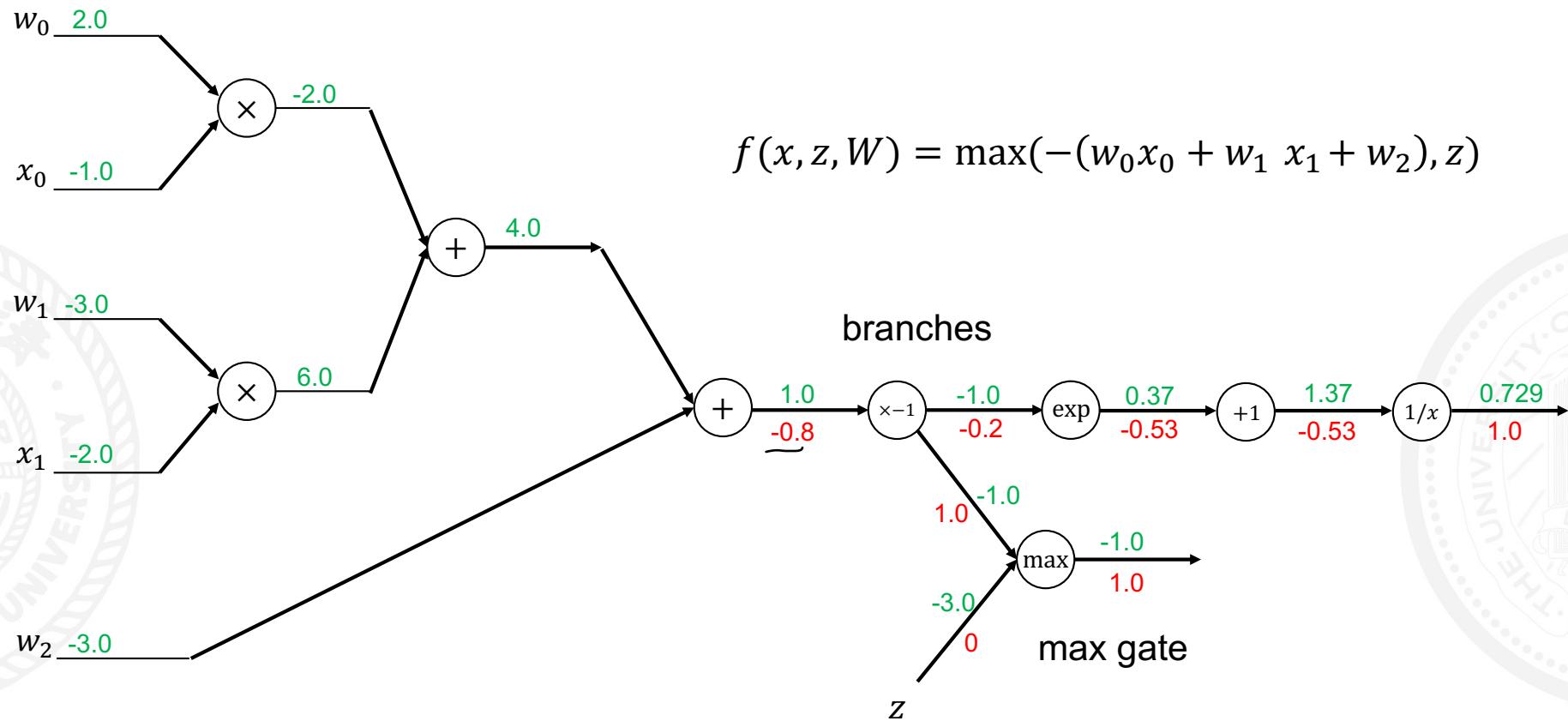
TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute



# Backpropagation



TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute



# Backpropagation



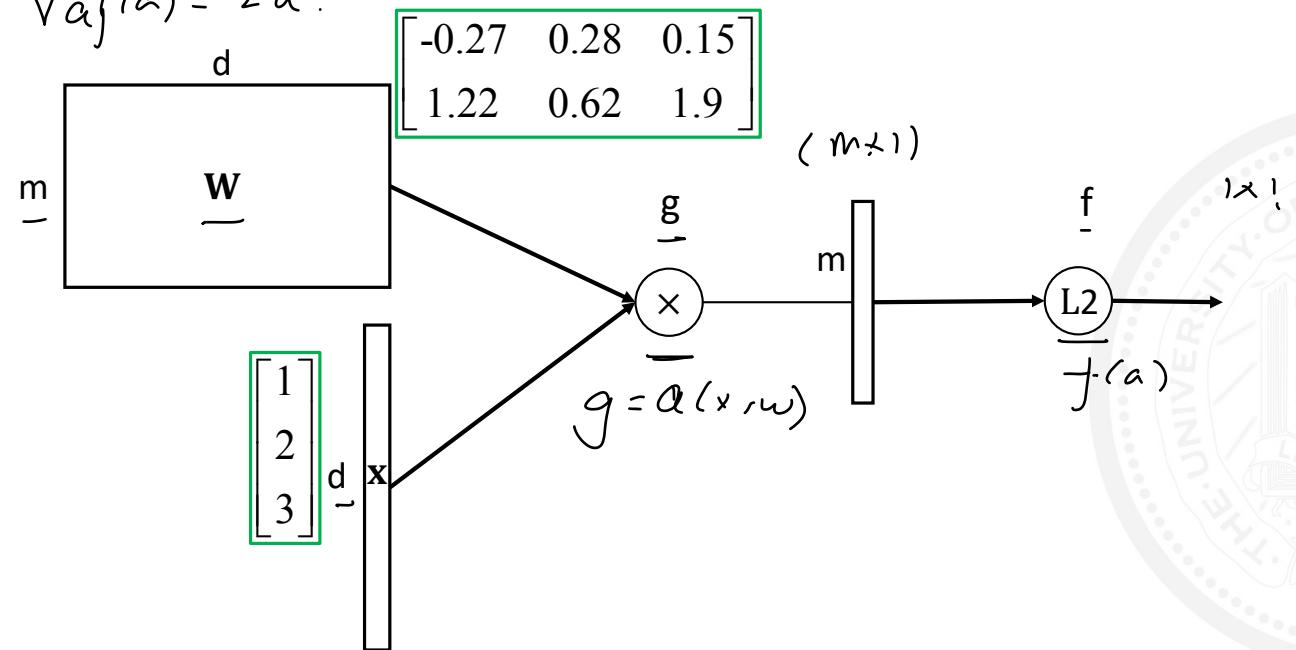
TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute

Vectorized example

$$f(\mathbf{x}, \mathbf{w}) = \|\mathbf{w} \cdot \mathbf{x}\|^2$$

$$\begin{array}{l} \textcircled{1} \quad \underline{a}(\mathbf{x}, \mathbf{w}) = \mathbf{x} \cdot \mathbf{w} \\ \textcircled{2} \quad \underline{f}(a) = \|a\|^2 = a^T a \\ \nabla_a f(a) = 2a \end{array}$$

$$\begin{aligned} \nabla_{\mathbf{w}} a(\mathbf{x}, \mathbf{w}) &= \mathbf{x}^T \\ \nabla_{\mathbf{x}} a(\mathbf{x}, \mathbf{w}) &= \mathbf{w} \end{aligned}$$



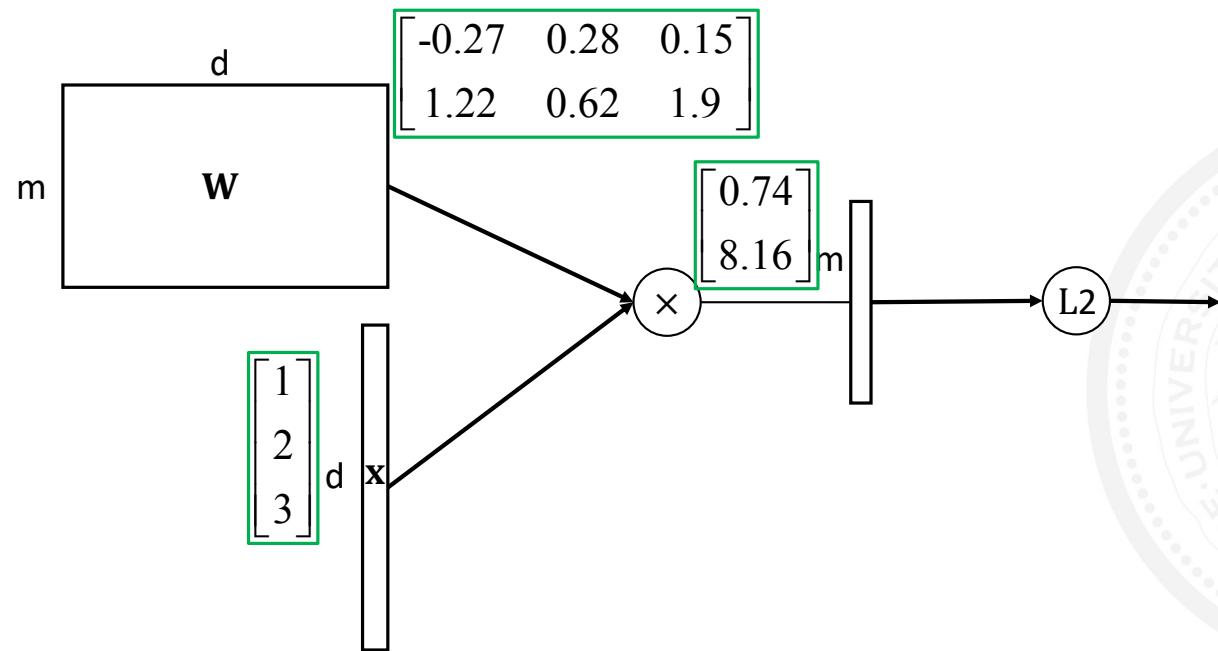
# Backpropagation



TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute

Vectorized example

$$f(\mathbf{x}, \mathbf{W}) = \|\mathbf{W} \cdot \mathbf{x}\|^2$$



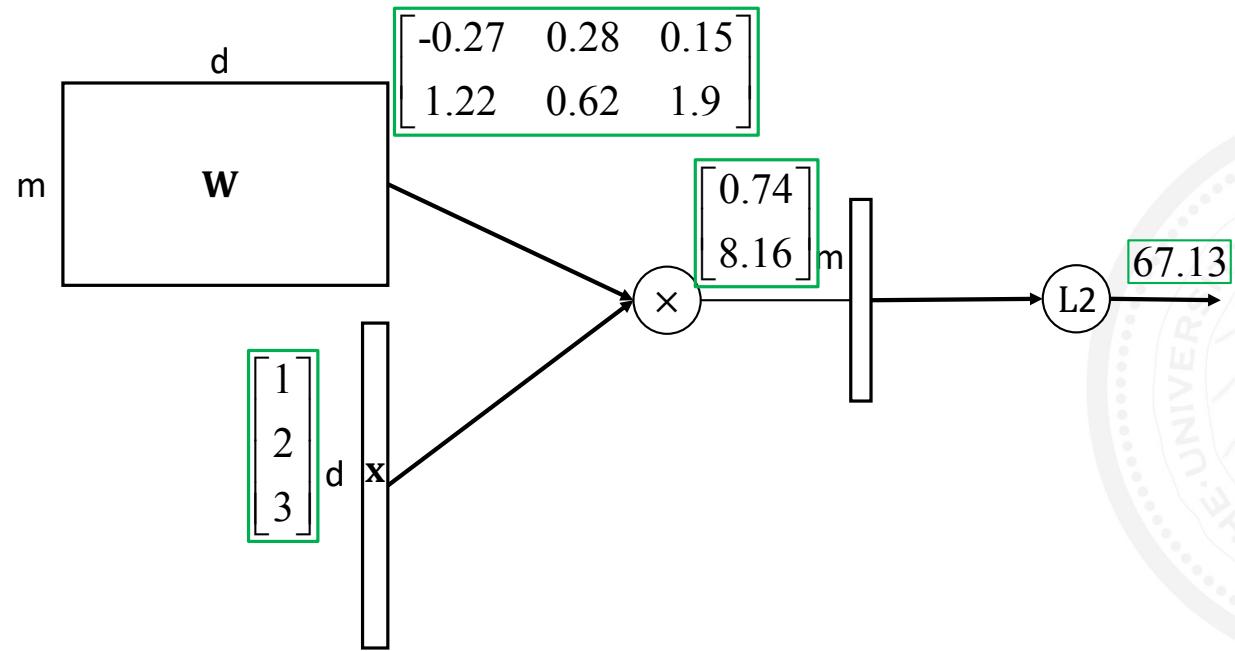
# Backpropagation



TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute

Vectorized example

$$f(\mathbf{x}, \mathbf{W}) = \|\mathbf{W} \cdot \mathbf{x}\|^2$$



# Backpropagation



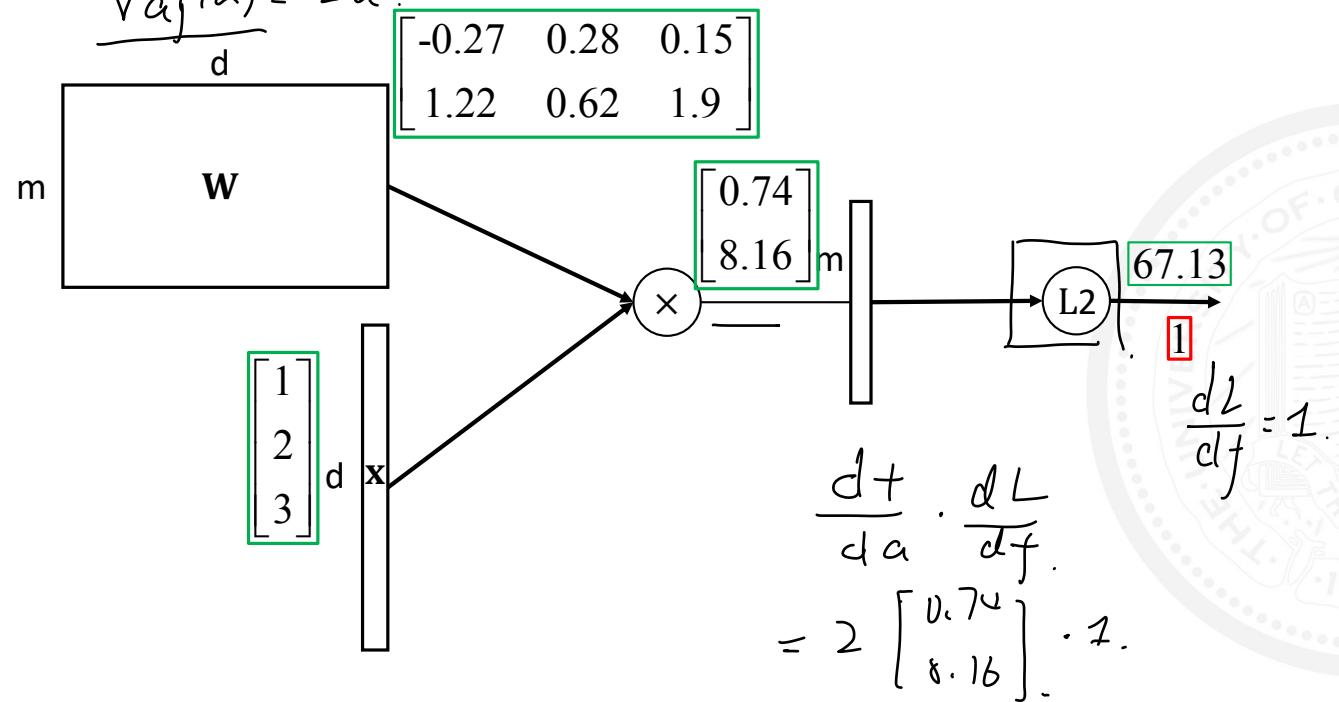
TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute

Vectorized example

$$f(\mathbf{x}, \mathbf{W}) = \|\mathbf{W} \cdot \mathbf{x}\|^2$$

$$\begin{aligned} ① \quad & \underline{a(x, w)} = X w \\ ② \quad & \underline{f(a)} = \|a\|^2 = a^T a \\ & \underline{\nabla_a f(a)} = 2 a \end{aligned}$$

$$\begin{aligned} \nabla_w a(x, w) &= X^T \\ \nabla_x a(x, w) &= W \end{aligned}$$



# Backpropagation



TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute

Vectorized example

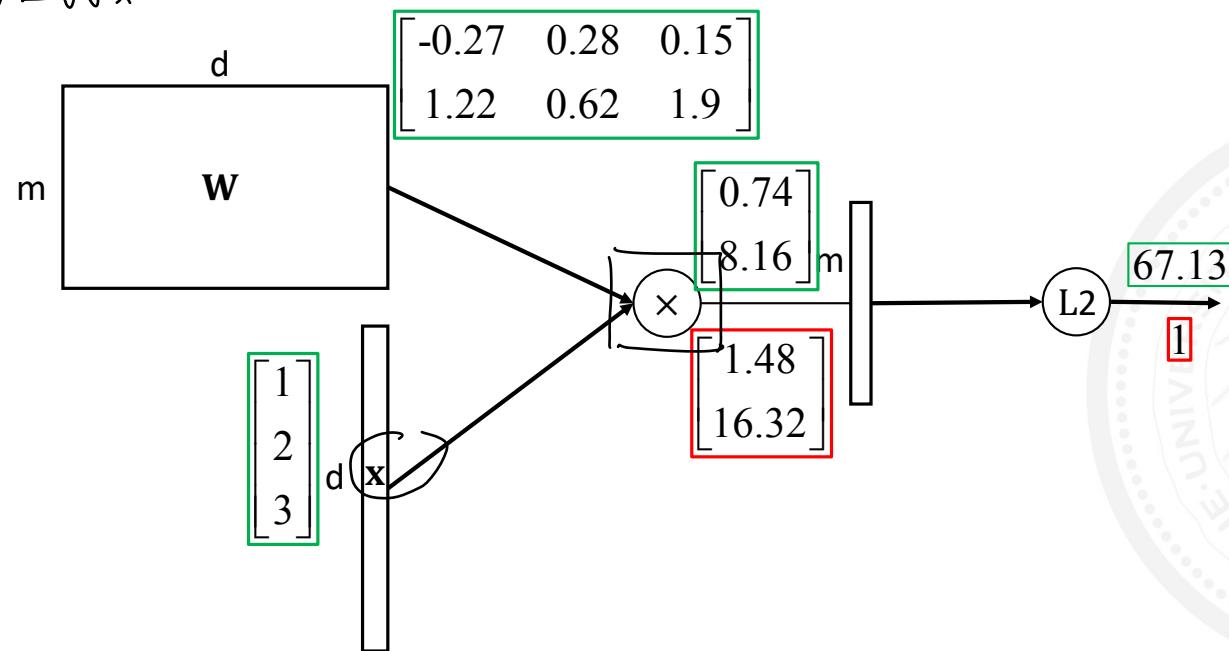
$$\widehat{f(\mathbf{x}, \mathbf{W})} = \|\mathbf{W} \cdot \mathbf{x}\|^2$$

$$a(\mathbf{w}, \mathbf{x}) = \mathbf{w} \cdot \mathbf{x}$$

$$\frac{df}{d\mathbf{w}} = \left( \frac{d}{da} \right) \left( \frac{c}{d\mathbf{w}} \right)$$

$$= \begin{pmatrix} 1.48 \\ 16.32 \end{pmatrix} \mathbf{x}^\top$$

$$= \begin{pmatrix} 1.48 \\ 16.32 \end{pmatrix} (1 \quad 2 \quad 3).$$



# Backpropagation



TBSI 清华-伯克利深圳学院  
Tsinghua-Berkeley Shenzhen Institute

Vectorized example

$$f(\mathbf{x}, \mathbf{W}) = \|\mathbf{W} \cdot \mathbf{x}\|^2$$

$$\frac{df}{d\omega} \downarrow$$
$$\begin{bmatrix} 1.48 \\ 16.32 \end{bmatrix} \times [1 \ 2 \ 3] = \begin{bmatrix} 1.48 & 2.96 & 4.44 \\ 16.32 & 32.64 & 48.96 \end{bmatrix}$$

