BACKPROPAGATION AND NEURAL NETWORKS

– ThS. Đoàn Chánh Thống
– ThS. Nguyễn Cường Phát
– ThS. Nguyễn Hữu Lợi
– ThS. Trương Quốc Dũng
– ThS. Nguyễn Thành Hiệp
– ThS. Võ Duy Nguyên
– ThS. Nguyễn Văn Toàn
– ThS. Lê Ngô Thục Vi
– TS. Nguyễn Duy Khánh
– TS. Nguyễn Tấn Trần Minh Khang

Administrative

– Assignment 1 due Thursday April 20, 11:59pm on Canvas
– Bài tập 1 sẽ nộp vào Thứ Năm, ngày 20 tháng 4, 23:59 trên Canvas

Administrative

- Project: TA specialities and some project ideas are posted on Piazza.
- Dự án: Trợ giảng gợi ý và một số ý tưởng dự án được đăng tải trên Piazza.

Administrative

– Google Cloud: All registered students will receive an email this week with instructions on how to redeem $100 in credits.

– Google Cloud: Tất cả sinh viên đã đăng ký sẽ nhận được email trong tuần này kèm theo hướng dẫn về cách đổi tín dụng $100.

Where we are...

– The loss is just a function of W:

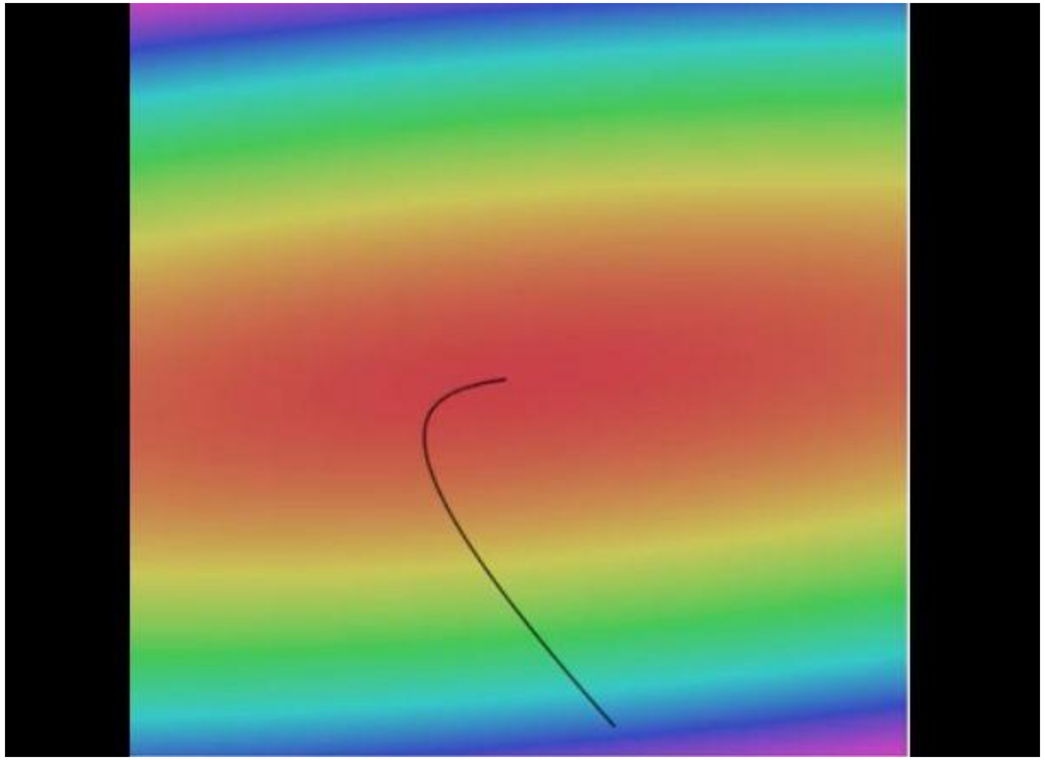
+ $s^{(i)} = f\left(x^{(i)}; W\right) = Wx^{(i)}$ Scores function

+ $L^{(i)} = \sum_{j \neq y^{(i)}} max\left(0, s_j - s_{y^{(i)}} + 1\right)$ SVM LOSS

+ $L = \frac{1}{N}\sum_{i=0}^{N-1} L^{(i)} + \sum_k W_k^2$ Data loss + Regularization

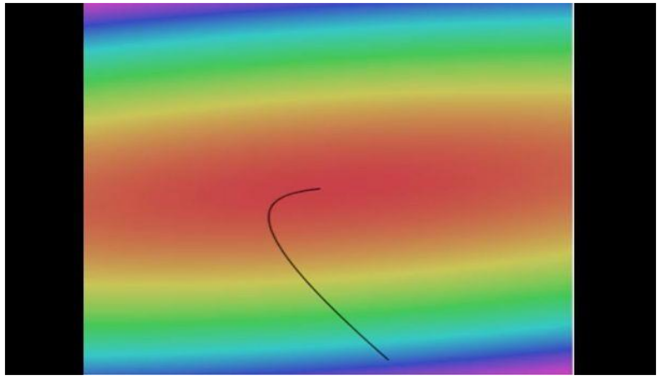
– Want $\nabla_W L$

Optimization



Optimization

```
# Vanilla Gradient Descent
while True:
    weights_grad = d_gradient(loss_fun, data, weights)
    weights += - step_size * weights_grad
    # perform parameter update
```



Gradient descent

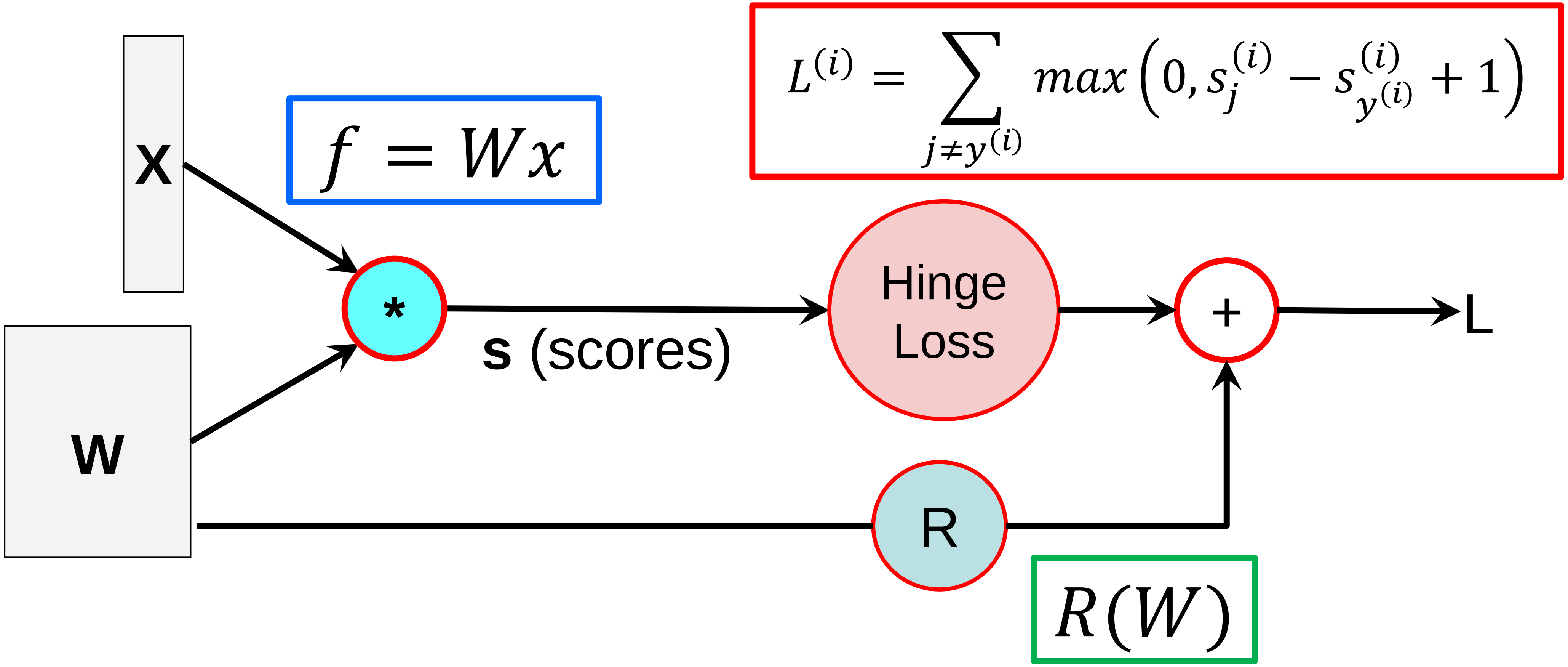
$$\frac{df(x)}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

- Numerical gradient: approximate, slow, easy to write.
- Analytic gradient: exact, fast, error – prone.

➢ In practice: Always use analytic gradient but check implementation with numerical gradient.

COMPUTATIONAL GRAPHS

Computational graphs

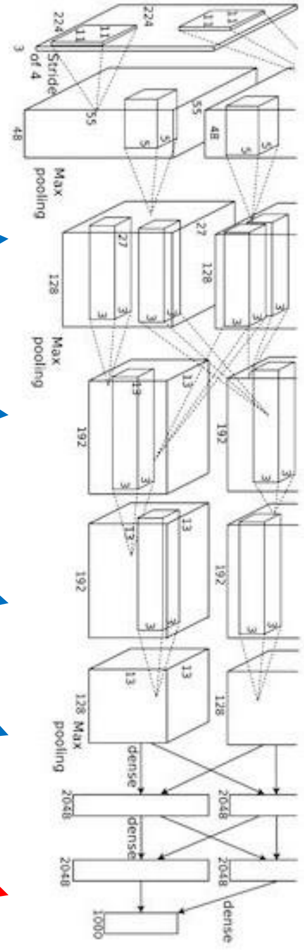


Convolutional network (AlexNet)

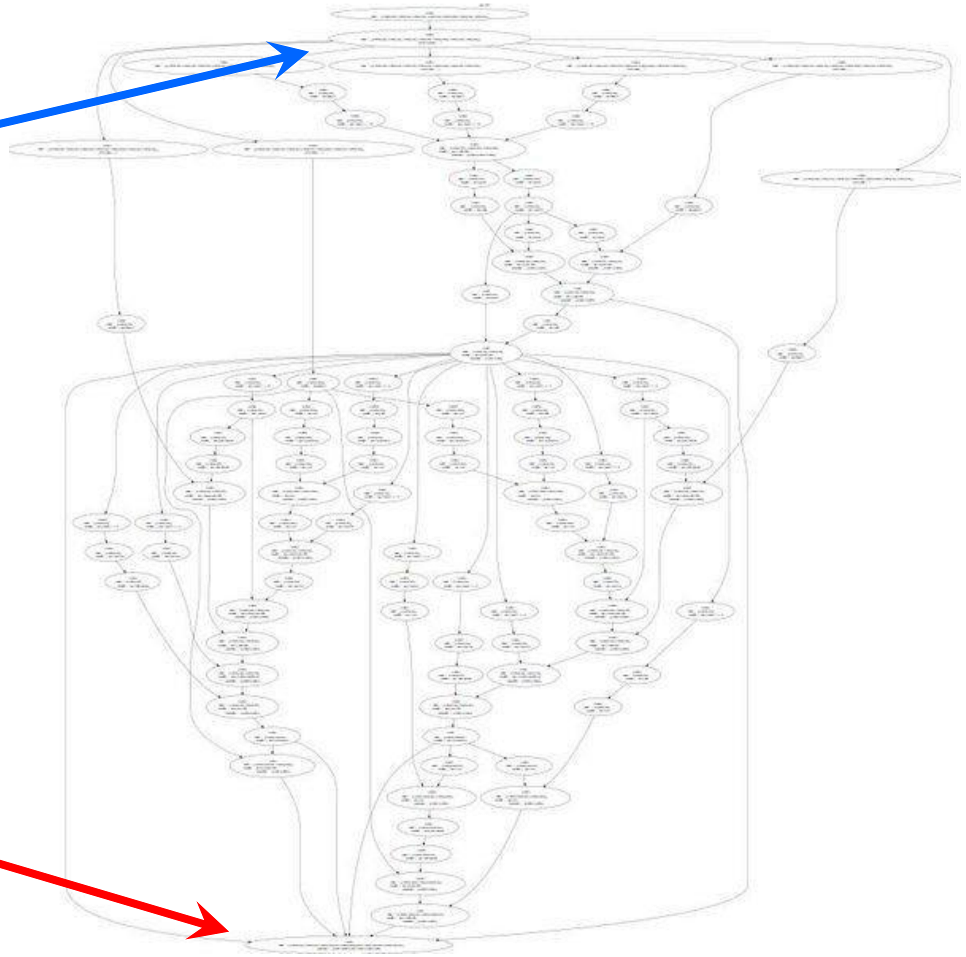
Input image

Weights

loss



Neural Turing Machine

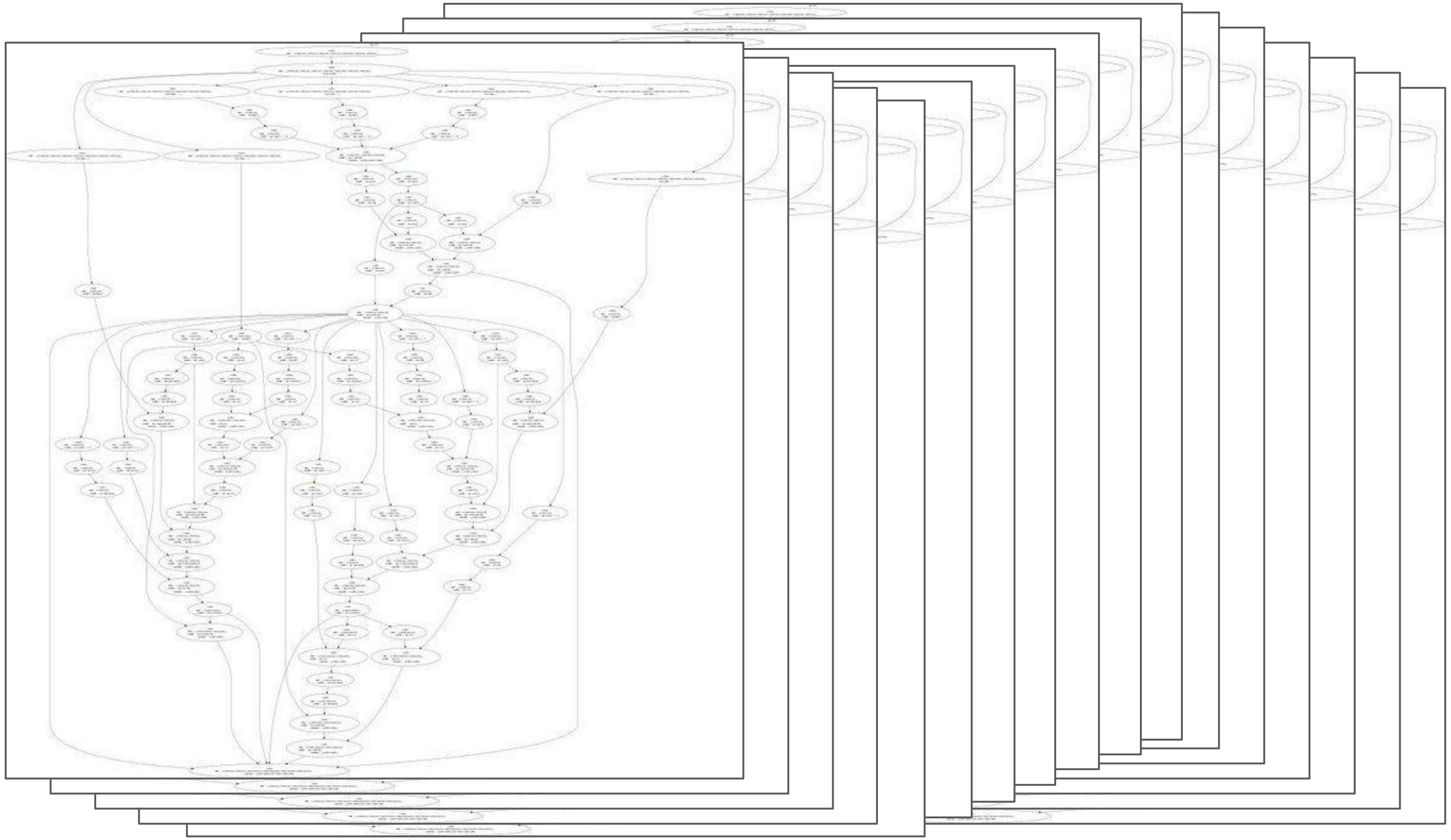


input image

loss

UIT VNUHCM University of Information Technology

Neural Turing Machine



A simple example

BACKPROPAGATION

Backpropagation – A simple example

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

E.g.

$x = -2,$

$y = 5,$

$z = -4$

Backpropagation – A simple example

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

– **Want:** $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$

E.g.

$x = -2,$

$y = 5,$

$z = -4$

Backpropagation – A simple example

$$\boldsymbol{f(x, y, z) = (x + y)z}$$

e.g. $x = -2, y = 5, z = -4$

– **Want:** $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$

Backpropagation – A simple example

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

e.g. $x = -2, y = 5, z = -4$

– **Want:** $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$

$x \longrightarrow$

$y \longrightarrow$

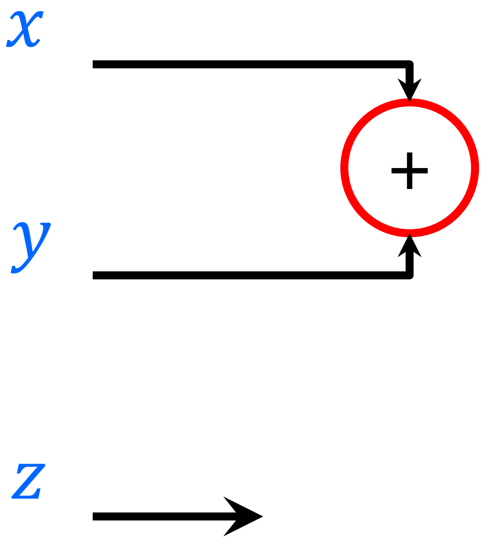
$z \longrightarrow$

Backpropagation – A simple example

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

e.g. $x = -2, y = 5, z = -4$

– **Want:** $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$

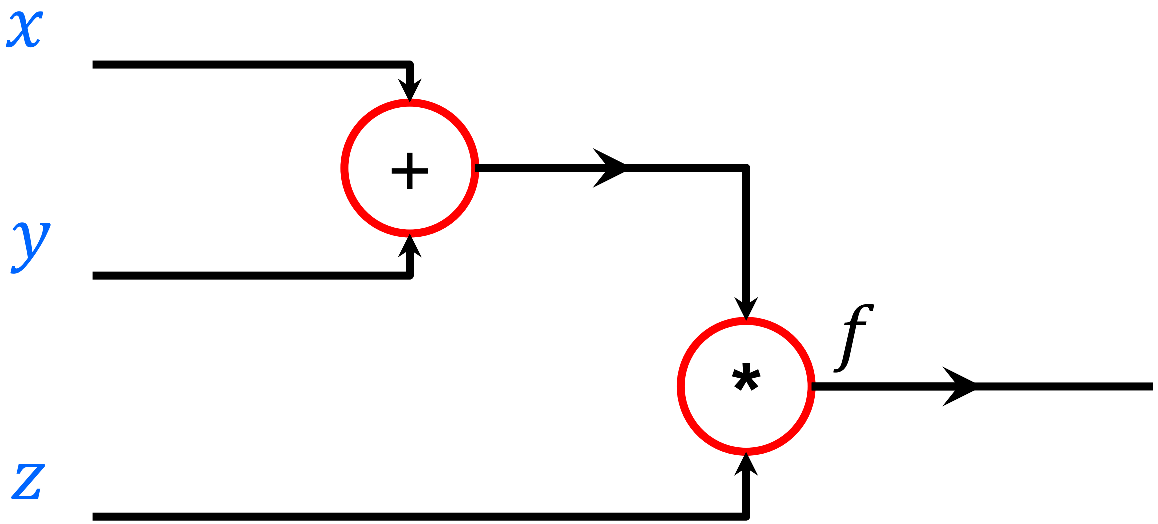


Backpropagation – A simple example

$$\boldsymbol{f(x,y,z) = (x+y)z}$$

e.g. $x = -2, y = 5, z = -4$

– **Want:** $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$

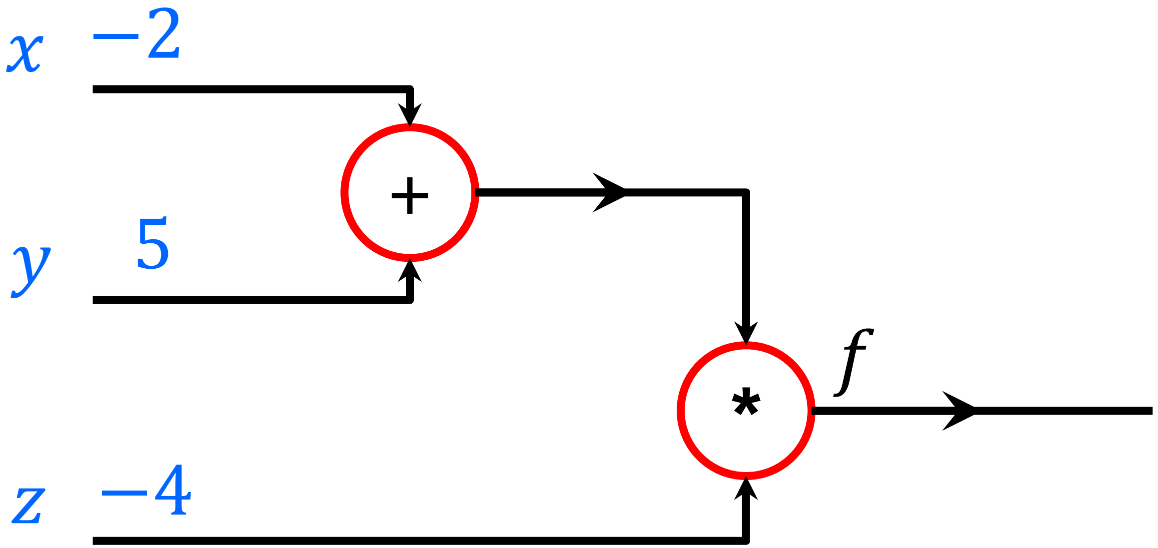


Backpropagation – A simple example

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

e.g. $x = -2, y = 5, z = -4$

– **Want:** $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



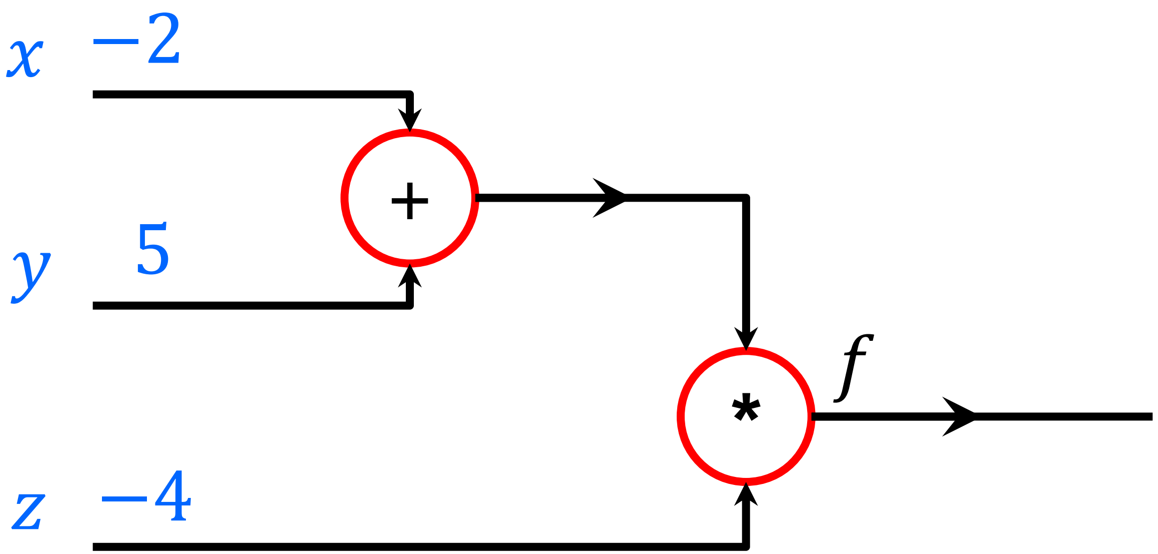
Backpropagation – A simple example

$$\boldsymbol{f(x, y, z) = (x + y)z}$$

e.g. $x = -2, y = 5, z = -4$

$q = x + y,$

– **Want:** $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



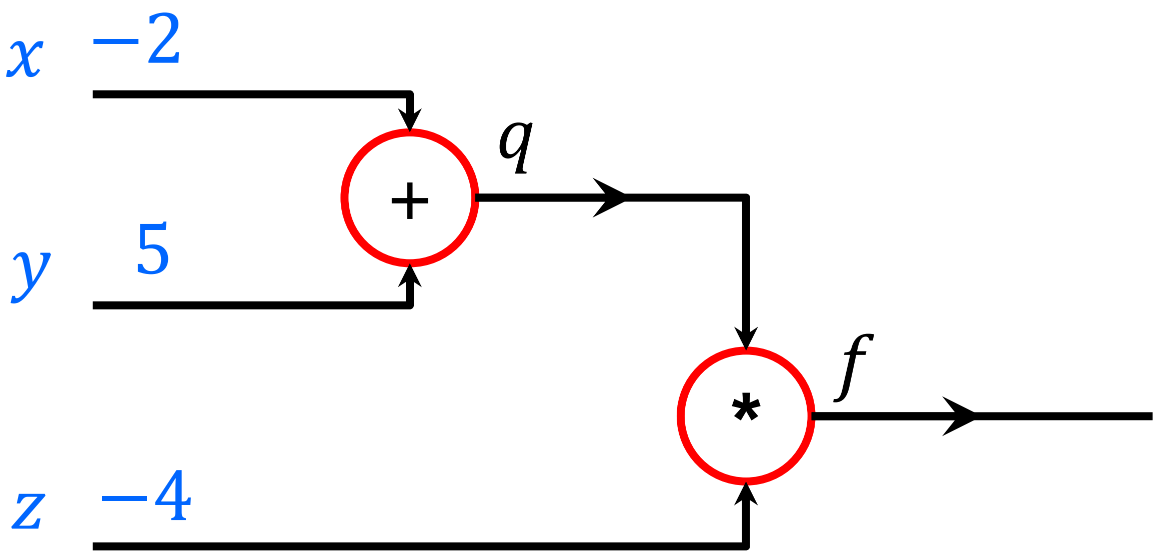
Backpropagation – A simple example

$$\boldsymbol{f(x,y,z) = (x+y)z}$$

e.g. $x = -2, y = 5, z = -4$

$q = x + y,$

– **Want:** $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



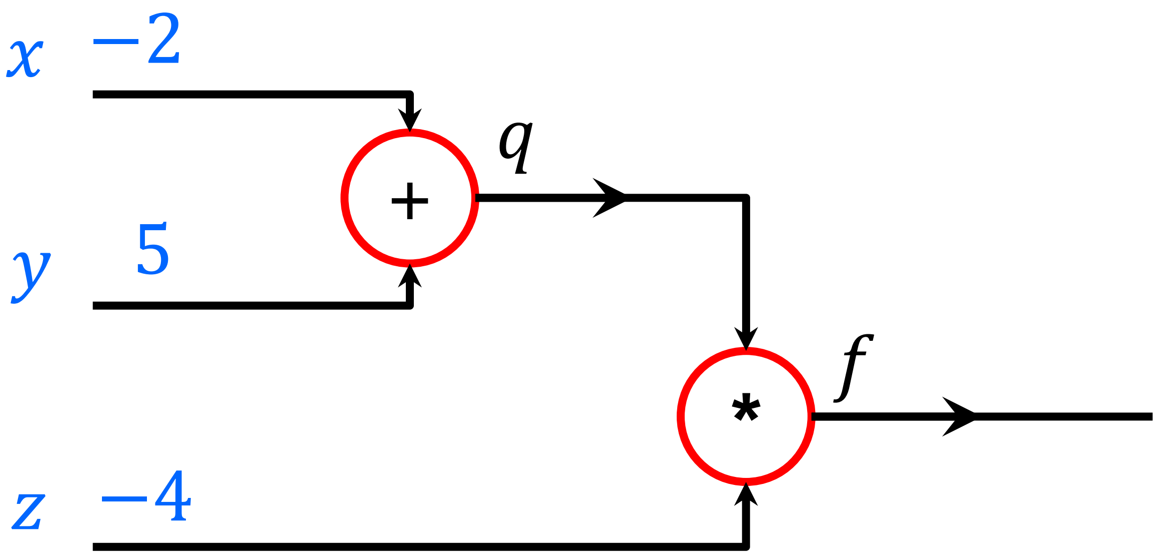
Backpropagation – A simple example

$$\boldsymbol{f(x,y,z) = (x+y)z}$$

e.g. $x = -2, y = 5, z = -4$

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

– **Want:** $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



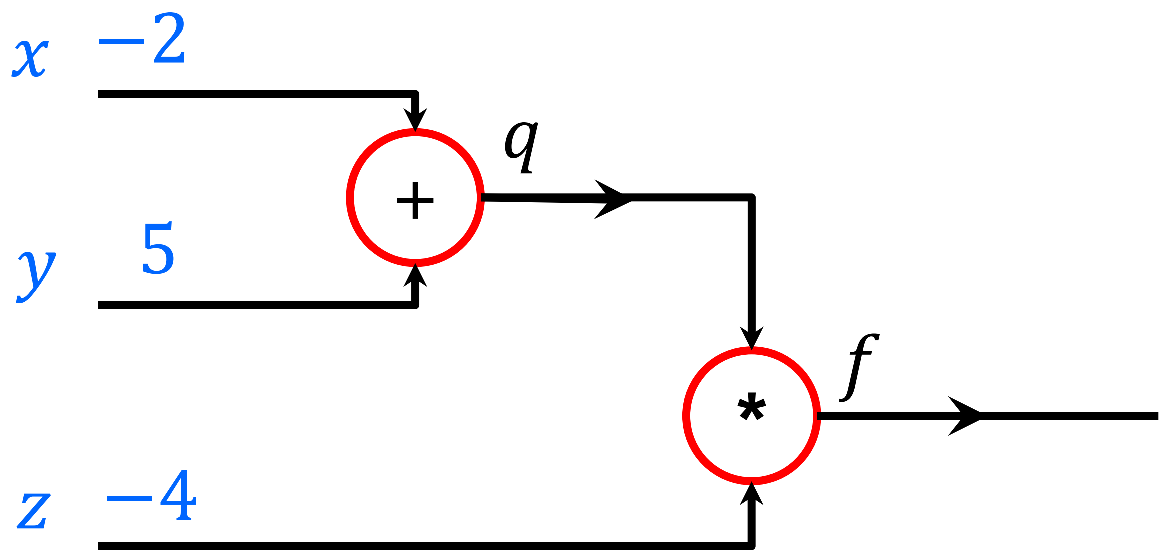
Backpropagation – A simple example

$$\boldsymbol{f(x, y, z) = (x + y)z}$$

e.g. $x = -2, y = 5, z = -4$

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

– **Want:** $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



Backpropagation – A simple example

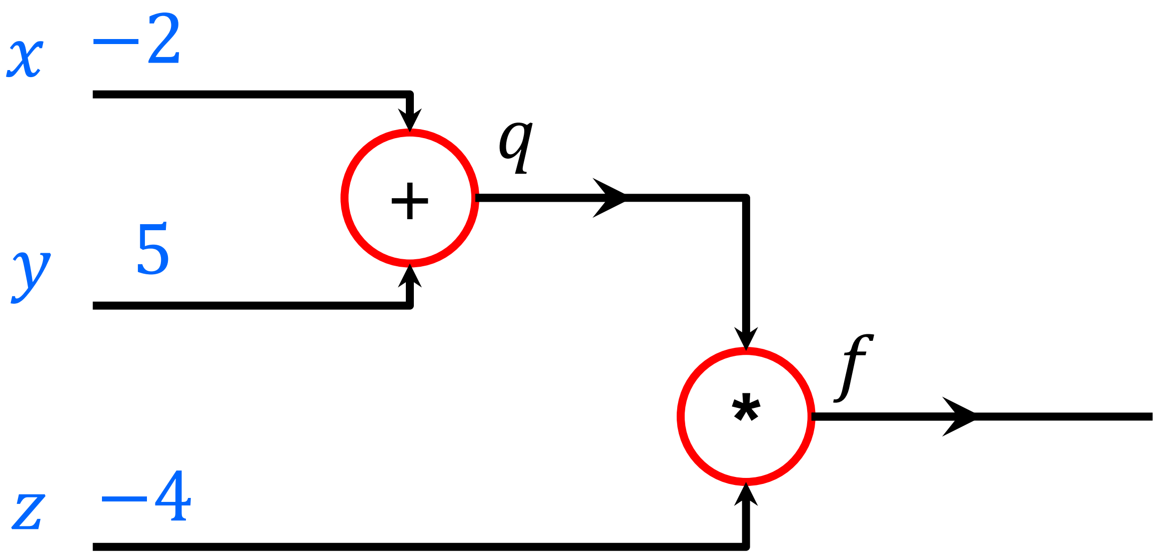
$$\boldsymbol{f(x,y,z) = (x+y)z}$$

e.g. $x = -2, y = 5, z = -4$

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

$f = qz$,

– **Want:** $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



Backpropagation – A simple example

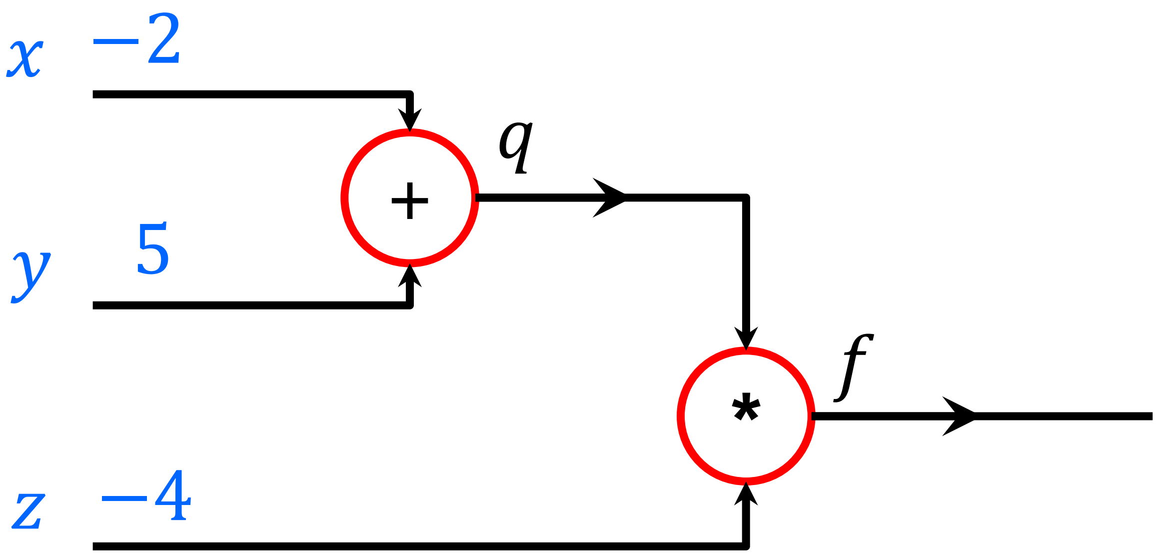
$$\boldsymbol{f(x,y,z) = (x+y)z}$$

e.g. $x = -2, y = 5, z = -4$

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

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

– **Want:** $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



Backpropagation – A simple example

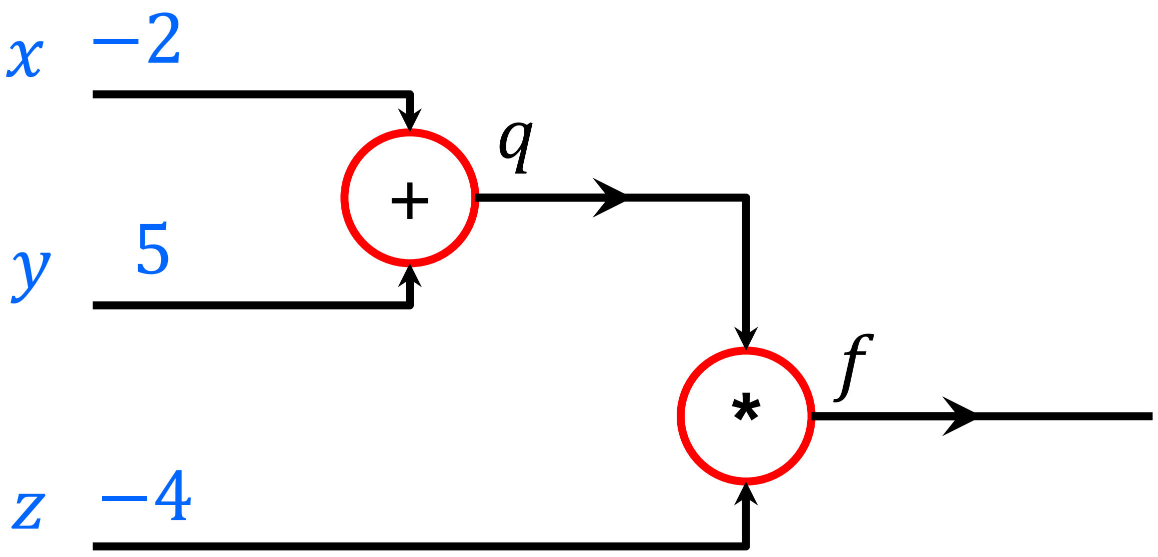
$$\boldsymbol{f(x, y, z) = (x + y)z}$$

e.g. $x = -2, y = 5, z = -4$

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

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

– **Want:** $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



Backpropagation – A simple example

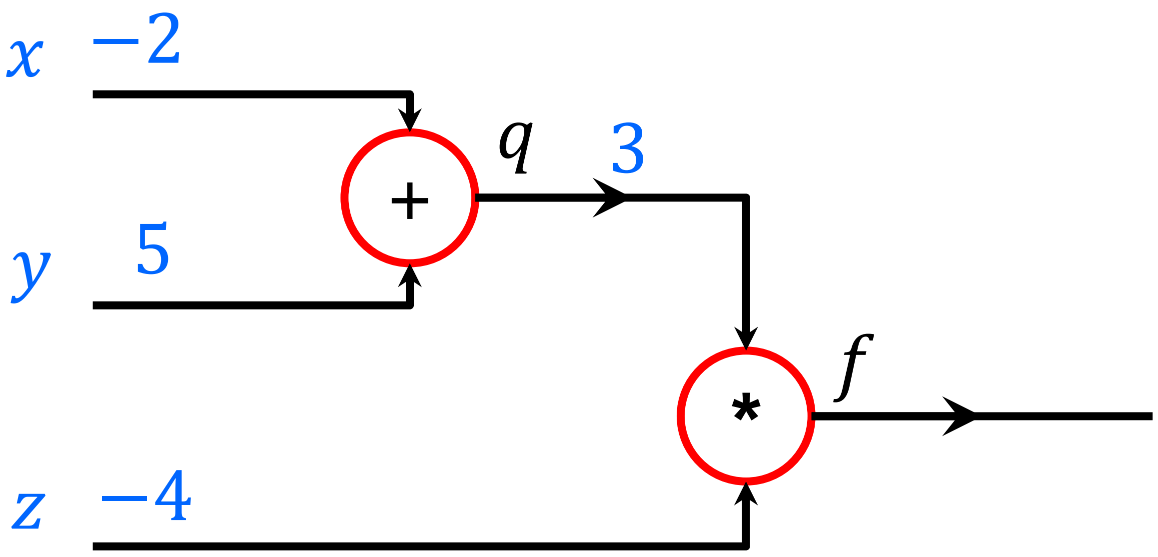
$$\boldsymbol{f(x, y, z) = (x + y)z}$$

e.g. $x = -2, y = 5, z = -4$

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

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

– **Want:** $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



Backpropagation – A simple example

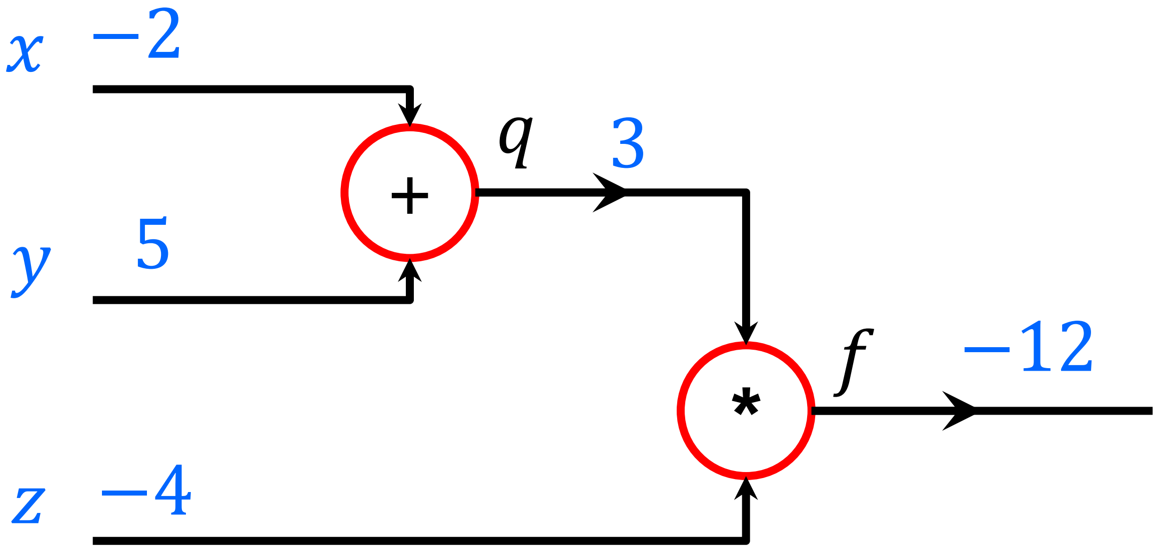
$$f(x, y, z) = (x + y)z$$

e.g. $x = -2, y = 5, z = -4$

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

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

– **Want:** $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



Backpropagation – A simple example

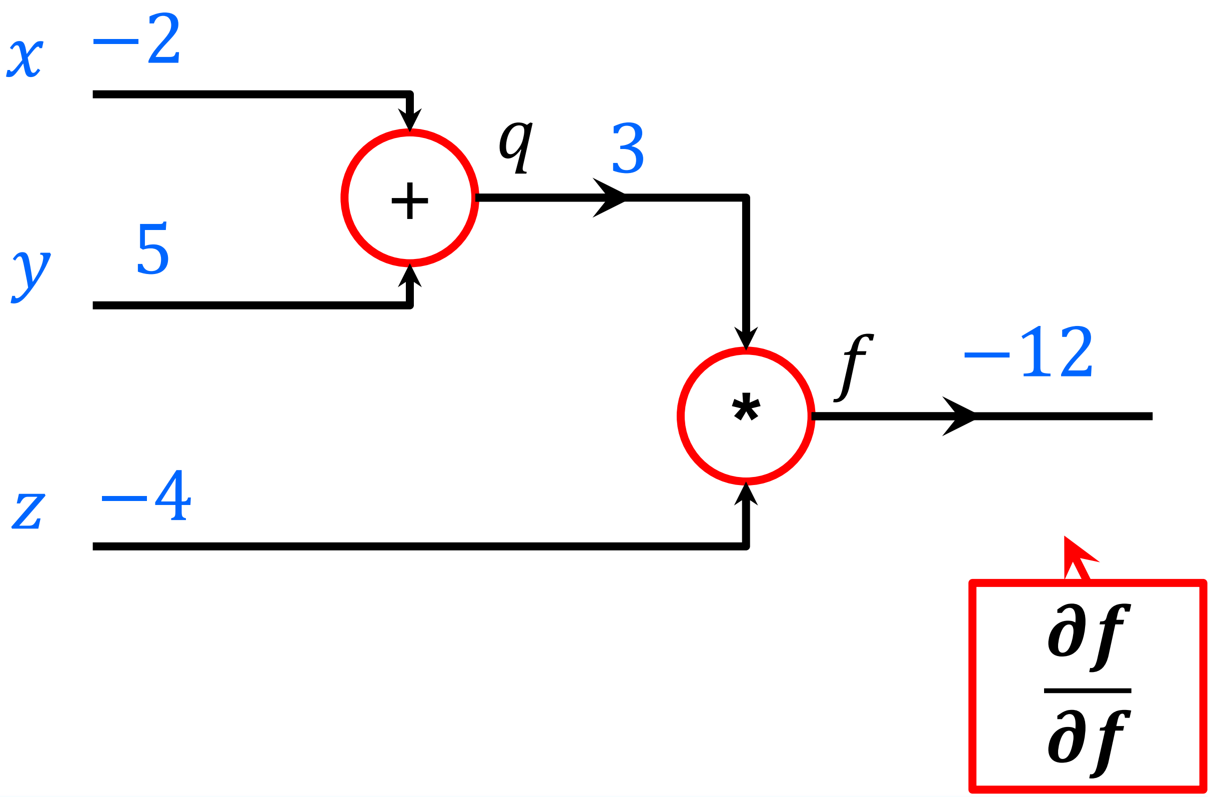
$$\boldsymbol{f(x, y, z) = (x + y)z}$$

e.g. $x = -2, y = 5, z = -4$

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

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

– **Want:** $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



Backpropagation – A simple example

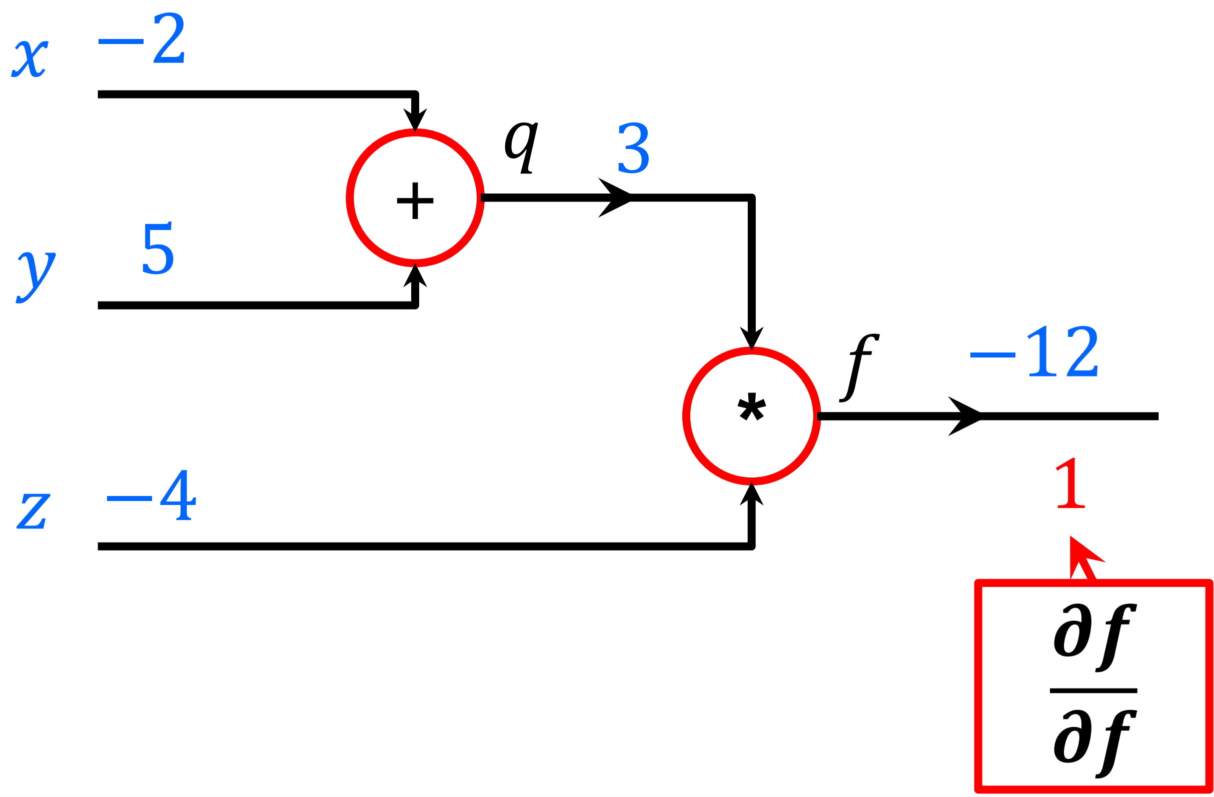
$$f(x, y, z) = (x + y)z$$

e.g. $x = -2, y = 5, z = -4$

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

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

– Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



Backpropagation – A simple example

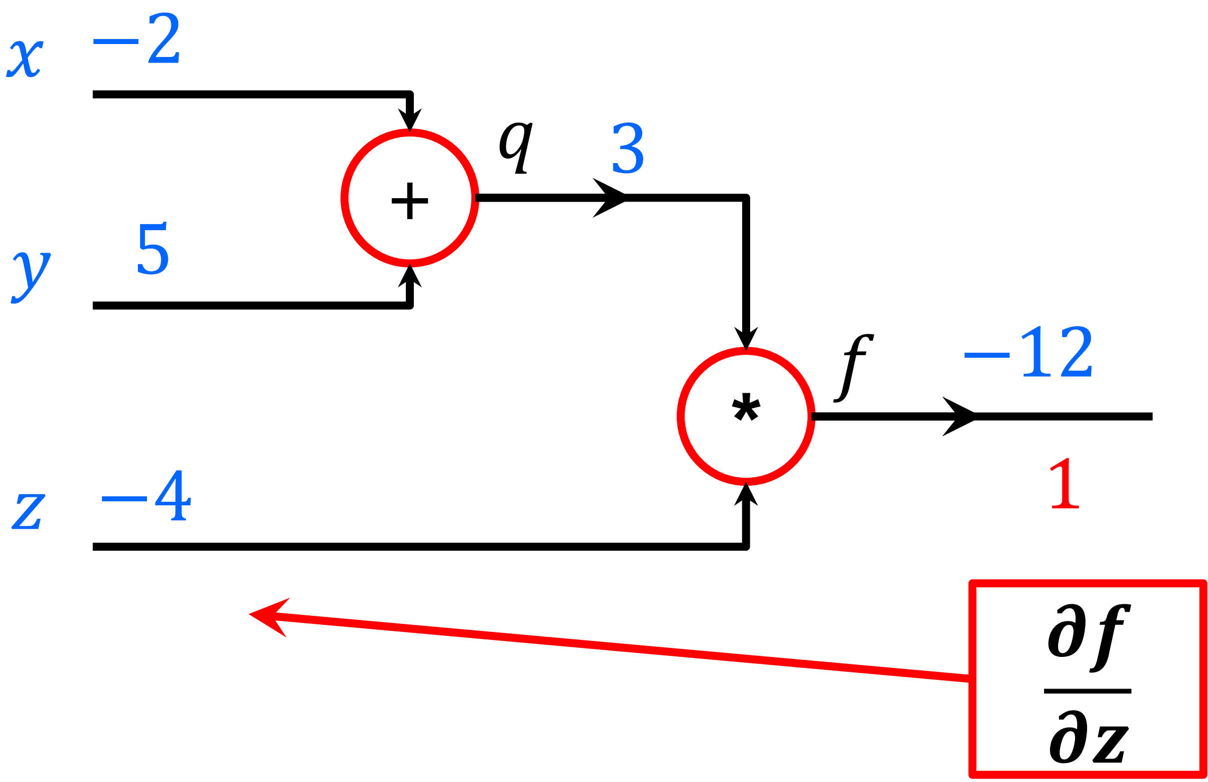
$$\boldsymbol{f(x,y,z) = (x+y)z}$$

e.g. $x = -2, y = 5, z = -4$

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

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

– **Want:** $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



Backpropagation – A simple example

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

e.g. $x = -2, y = 5, z = -4$

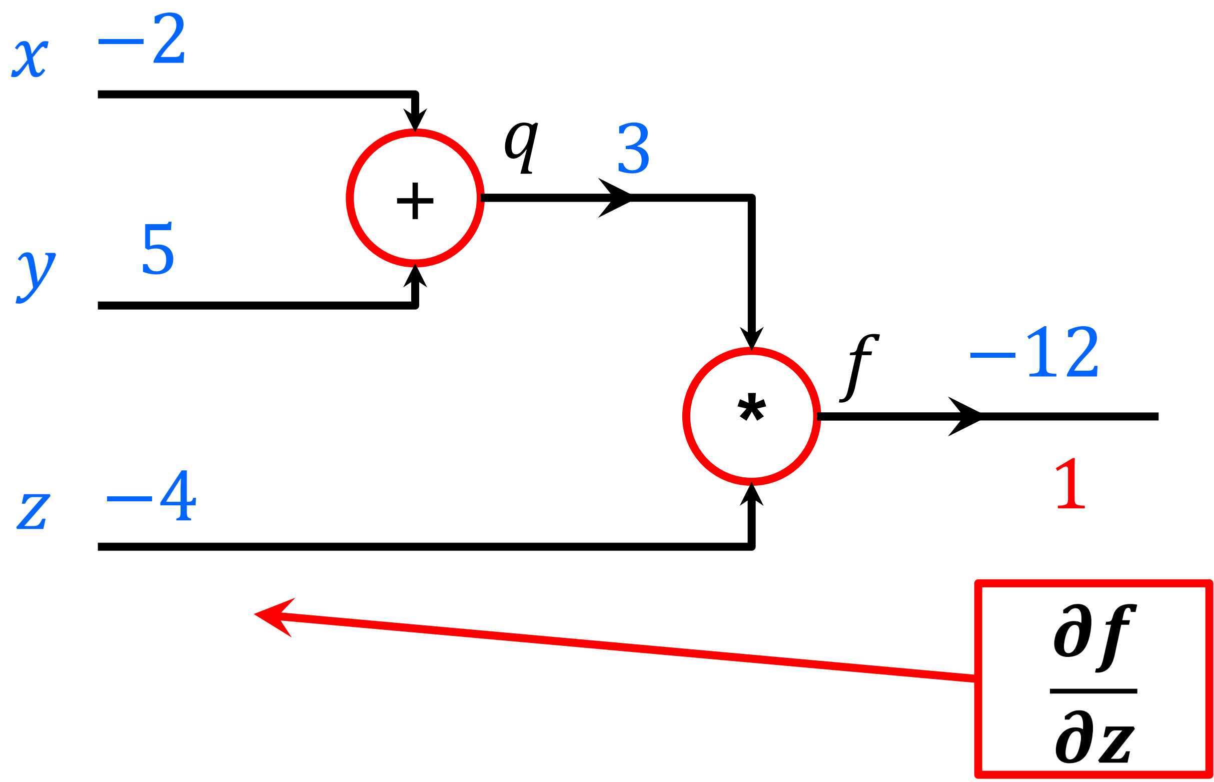
$q = x + y,\qquad \frac{\partial q}{\partial x} = 1,\qquad \frac{\partial q}{\partial y} = 1$

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

– Chain rule

$$\frac{\partial f}{\partial z} = \frac{\partial f}{\partial f}\frac{\partial f}{\partial z}$$

– Want: $\frac{\partial f}{\partial x}, \quad \frac{\partial f}{\partial y}, \quad \frac{\partial f}{\partial z}$



Backpropagation – A simple example

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

e.g. $x = -2, y = 5, z = -4$

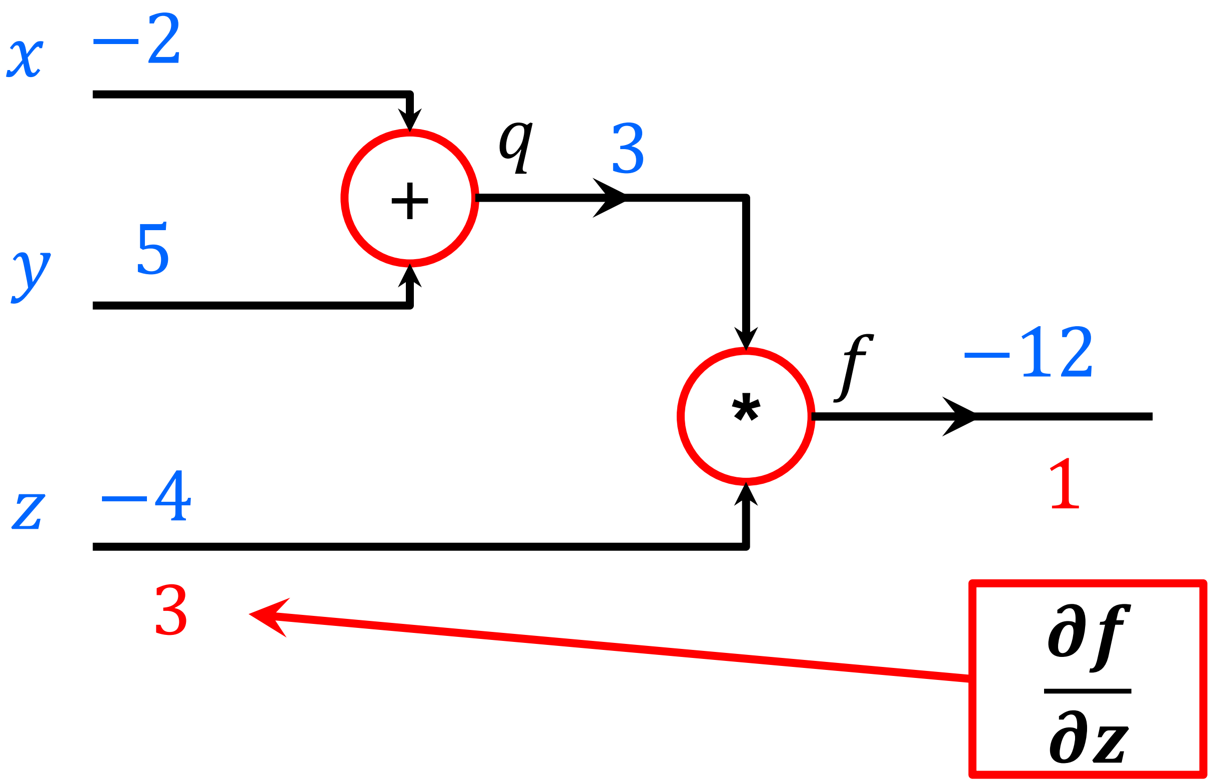
$q = x + y,\quad \frac{\partial q}{\partial x} = 1,\quad \frac{\partial q}{\partial y} = 1$

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

– Chain rule

$$\frac{\partial f}{\partial z} = \frac{\partial f}{\partial f}\frac{\partial f}{\partial z}$$

– Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



Backpropagation – A simple example

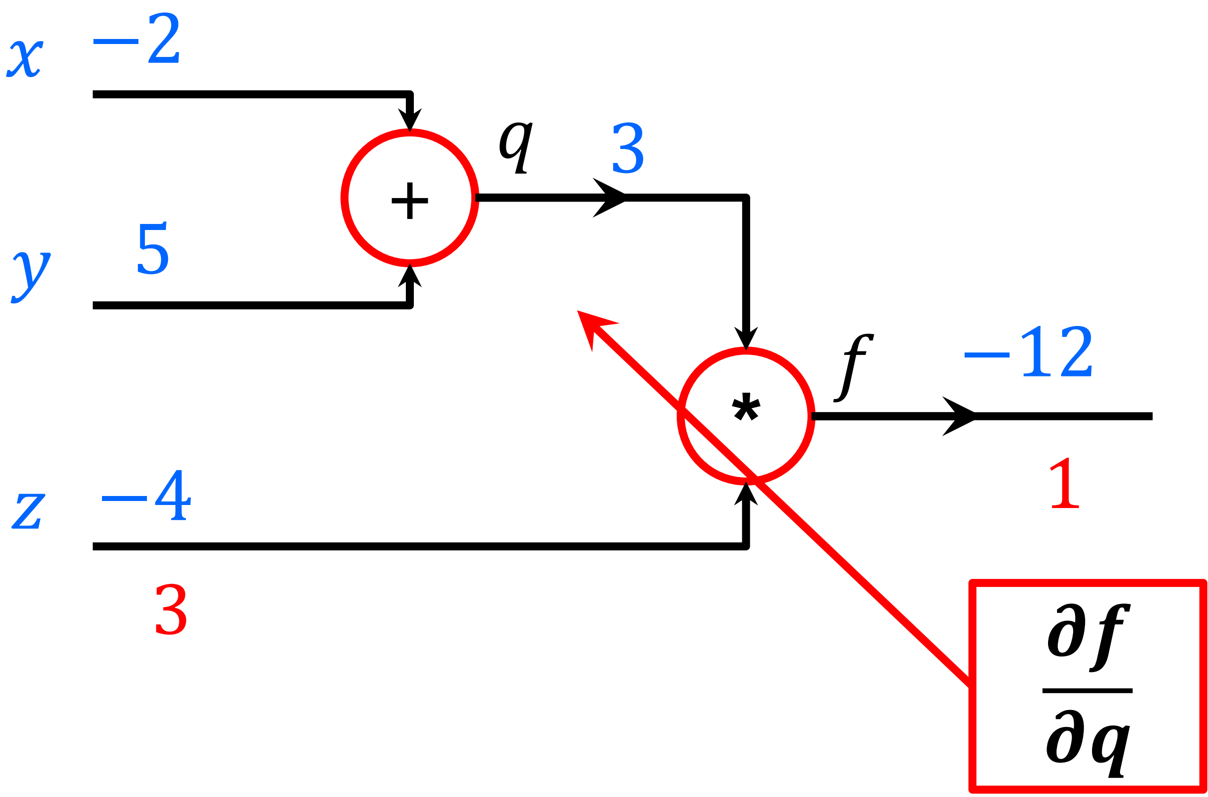
$$f(x, y, z) = (x + y)z$$

e.g. $x = -2, y = 5, z = -4$

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

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

– Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



Backpropagation – A simple example

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

e.g. $x = -2, y = 5, z = -4$

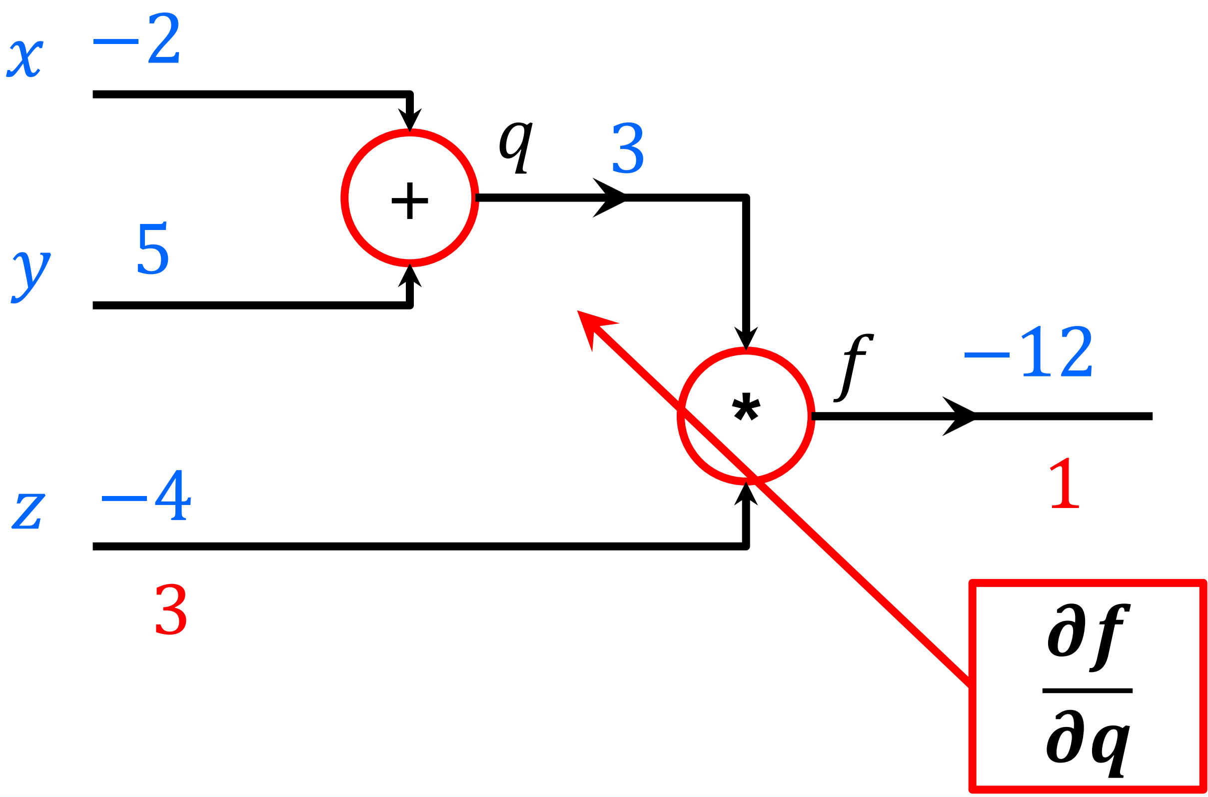
$q = x + y,\qquad \frac{\partial q}{\partial x} = 1,\qquad \frac{\partial q}{\partial y} = 1$

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

– Chain rule

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

– Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



Backpropagation – A simple example

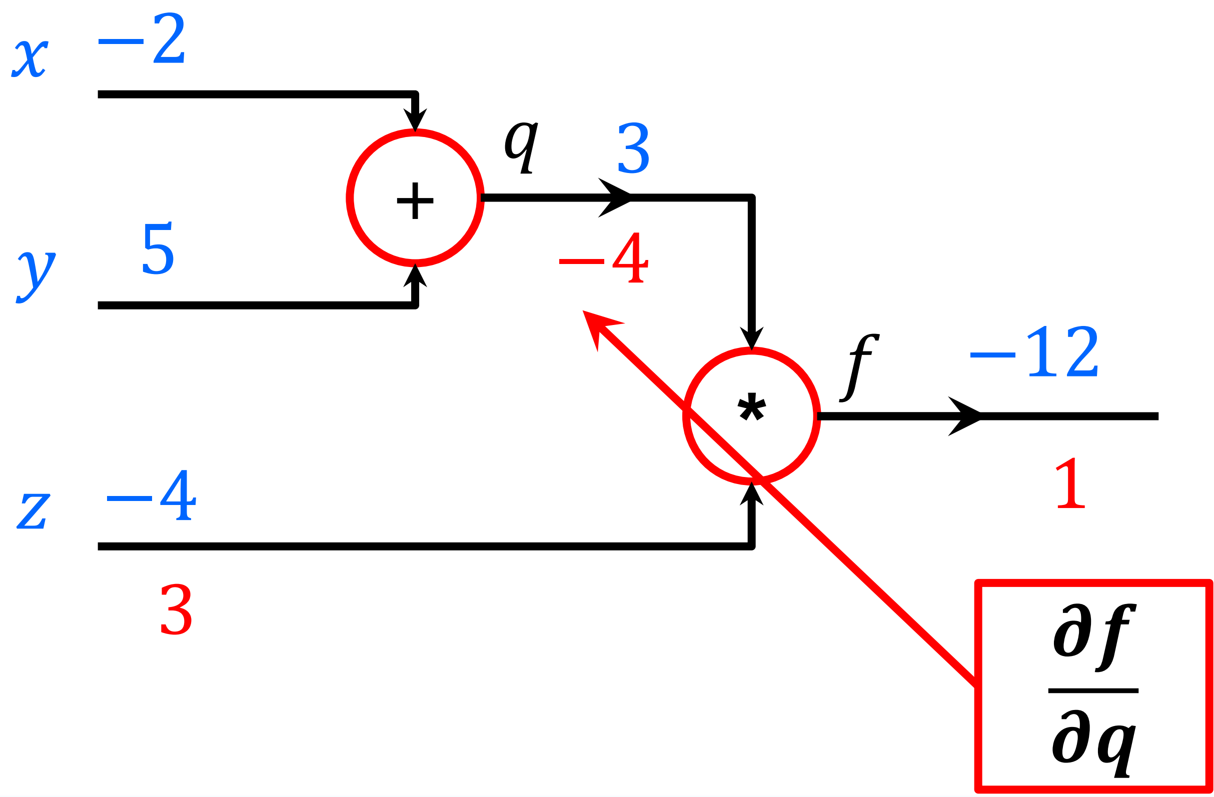
$$f(x, y, z) = (x + y)z$$

e.g. $x = -2, y = 5, z = -4$

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

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

– Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



– Chain rule

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

Backpropagation – A simple example

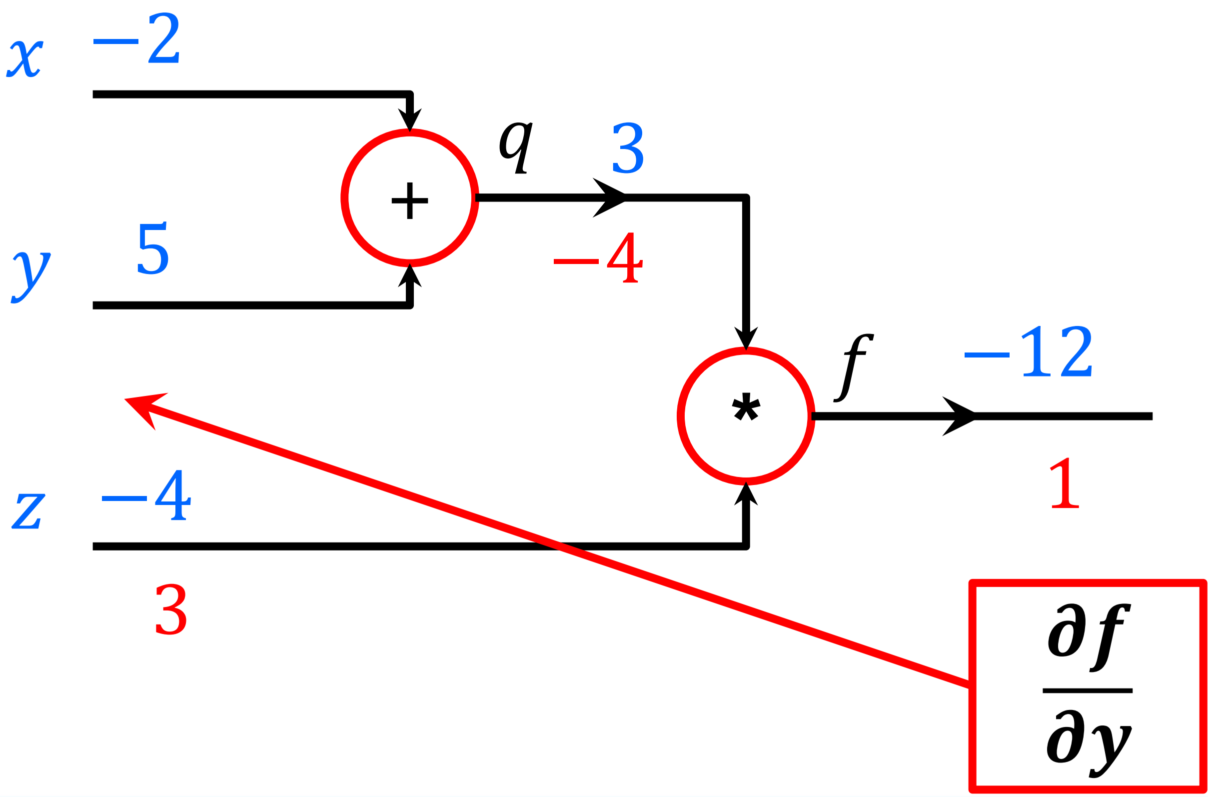
$$\boldsymbol{f(x,y,z) = (x+y)z}$$

e.g. $x = -2, y = 5, z = -4$

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

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

– **Want:** $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



Backpropagation – A simple example

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

e.g. $x=-2, y=5, z=-4$

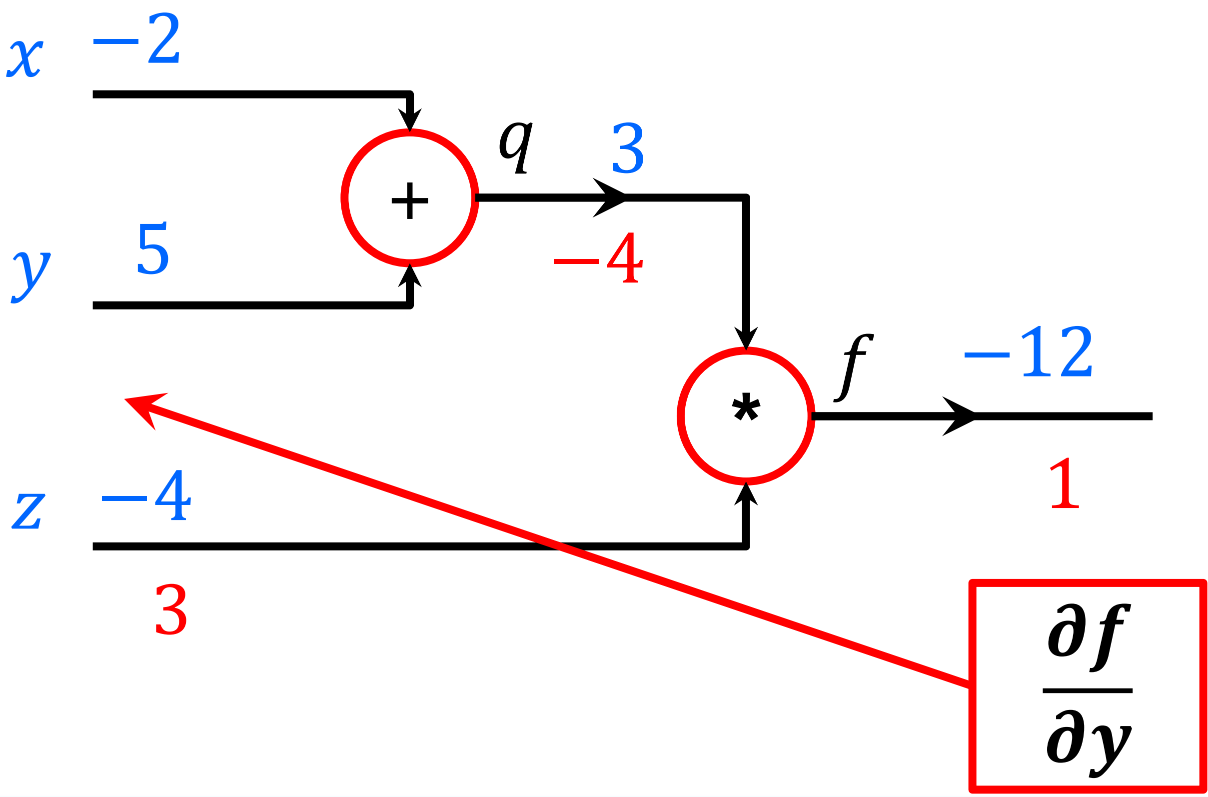
$q = x + y,\quad \frac{\partial q}{\partial x} = 1,\quad \frac{\partial q}{\partial y} = 1$

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

– Chain rule

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

– Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



Backpropagation – A simple example

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

e.g. $x = -2, y = 5, z = -4$

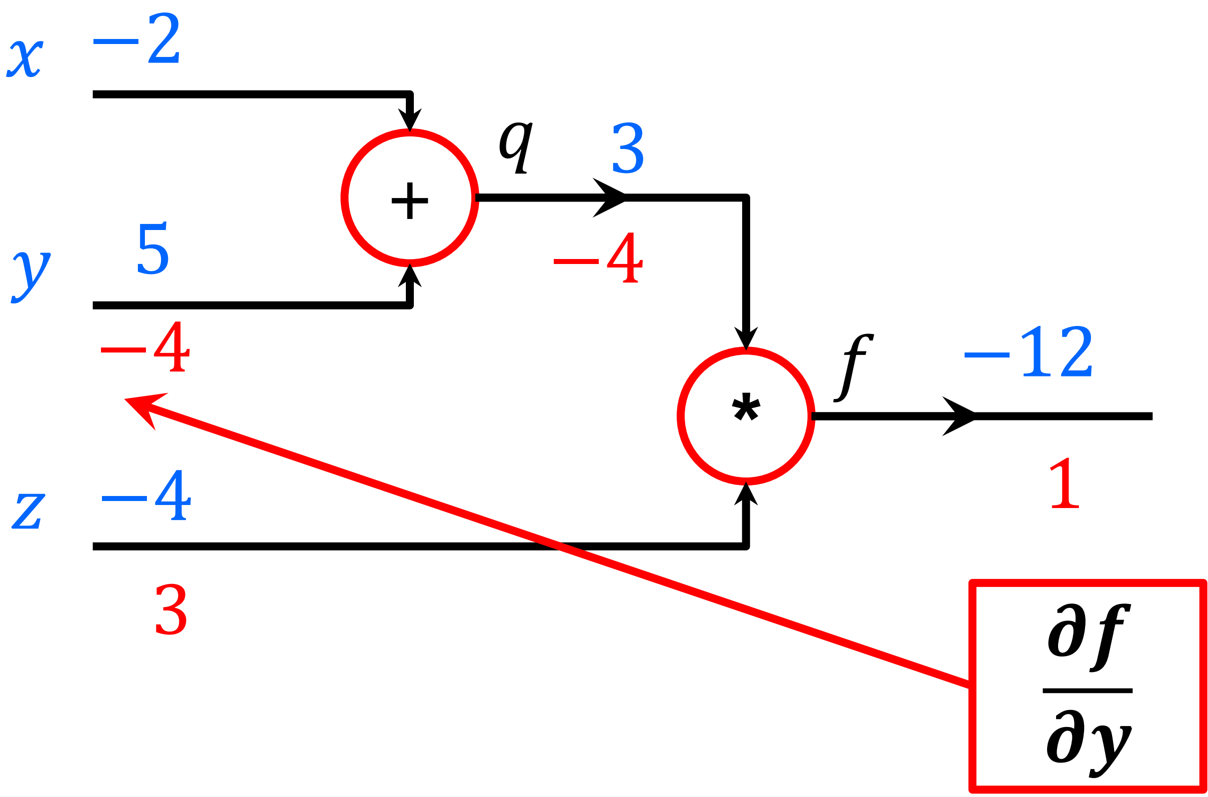
$q = x + y,\quad \frac{\partial q}{\partial x} = 1,\quad \frac{\partial q}{\partial y} = 1$

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

– Chain rule

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

– Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



Backpropagation – A simple example

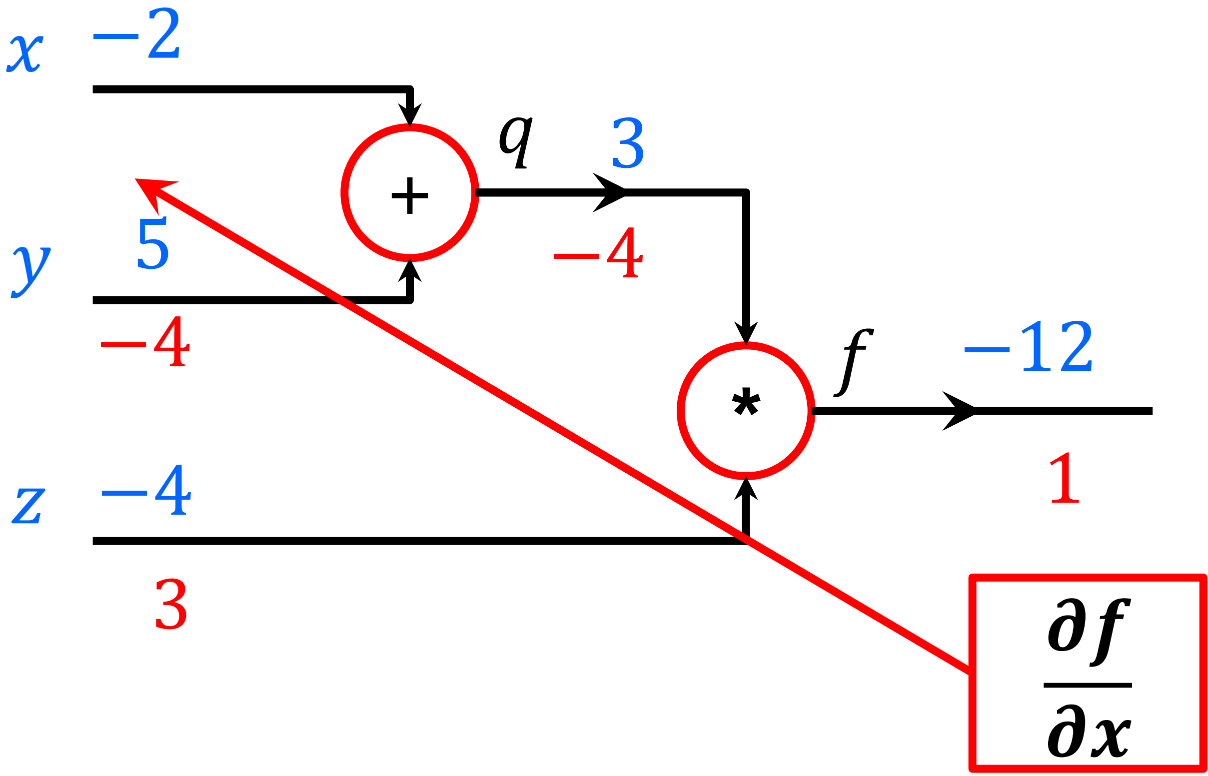
$$f(x, y, z) = (x + y)z$$

e.g. $x = -2, y = 5, z = -4$

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

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

– Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



Backpropagation – A simple example

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

e.g. $x = -2, y = 5, z = -4$

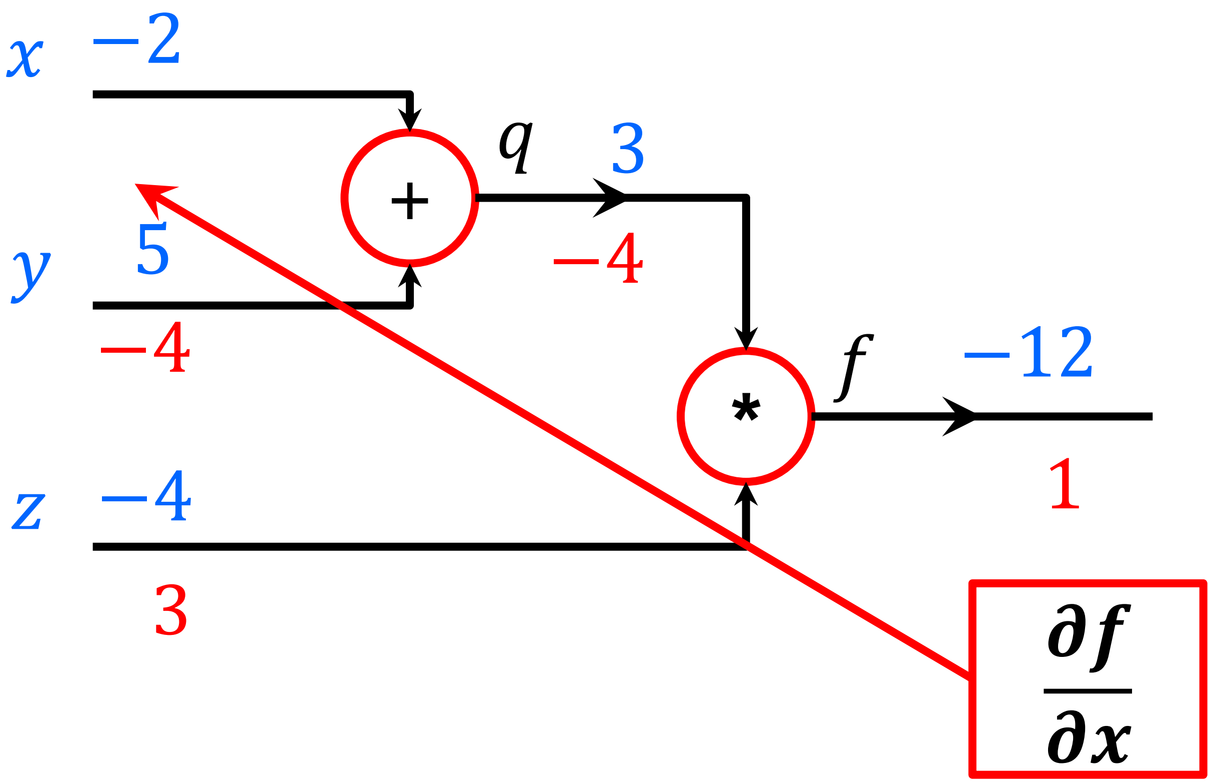
$q = x + y,\quad \frac{\partial q}{\partial x} = 1,\quad \frac{\partial q}{\partial y} = 1$

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

– Chain rule

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

– Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



Backpropagation – A simple example

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

e.g. $x = -2, y = 5, z = -4$

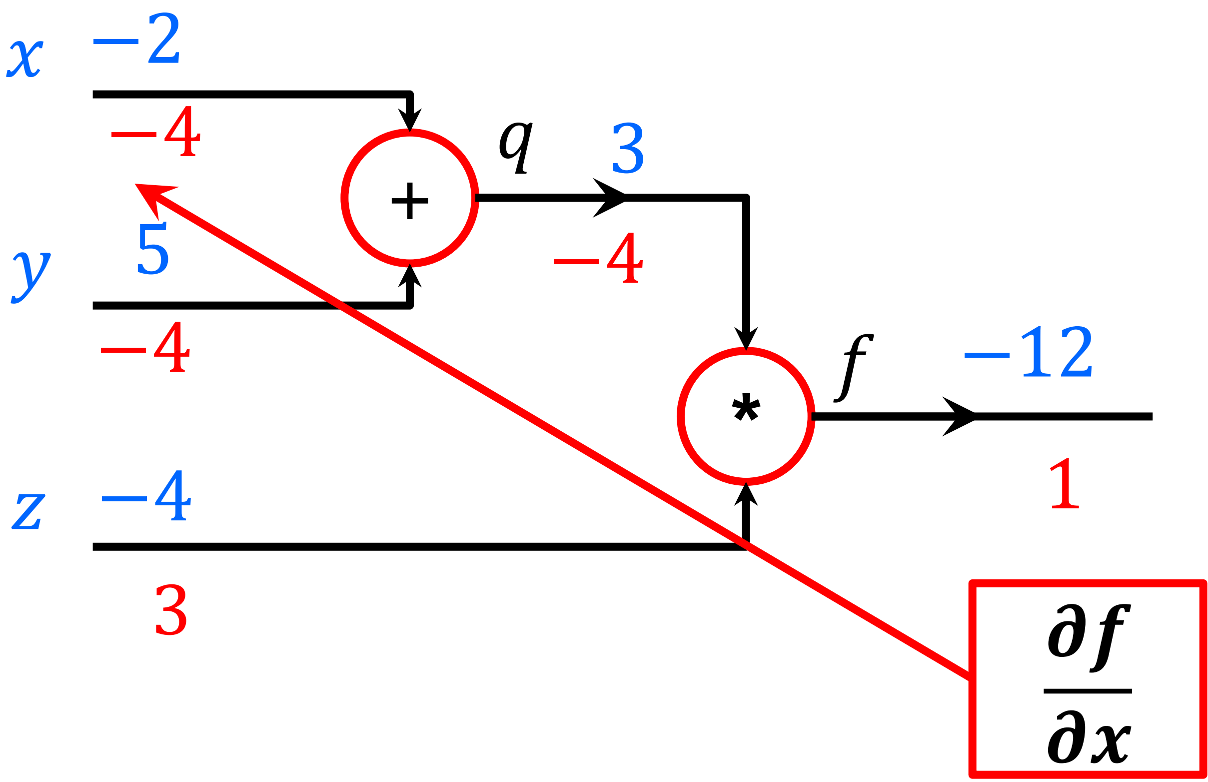
$q = x + y,\qquad \frac{\partial q}{\partial x} = 1,\qquad \frac{\partial q}{\partial y} = 1$

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

– Chain rule

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

– Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



Backpropagation – A simple example

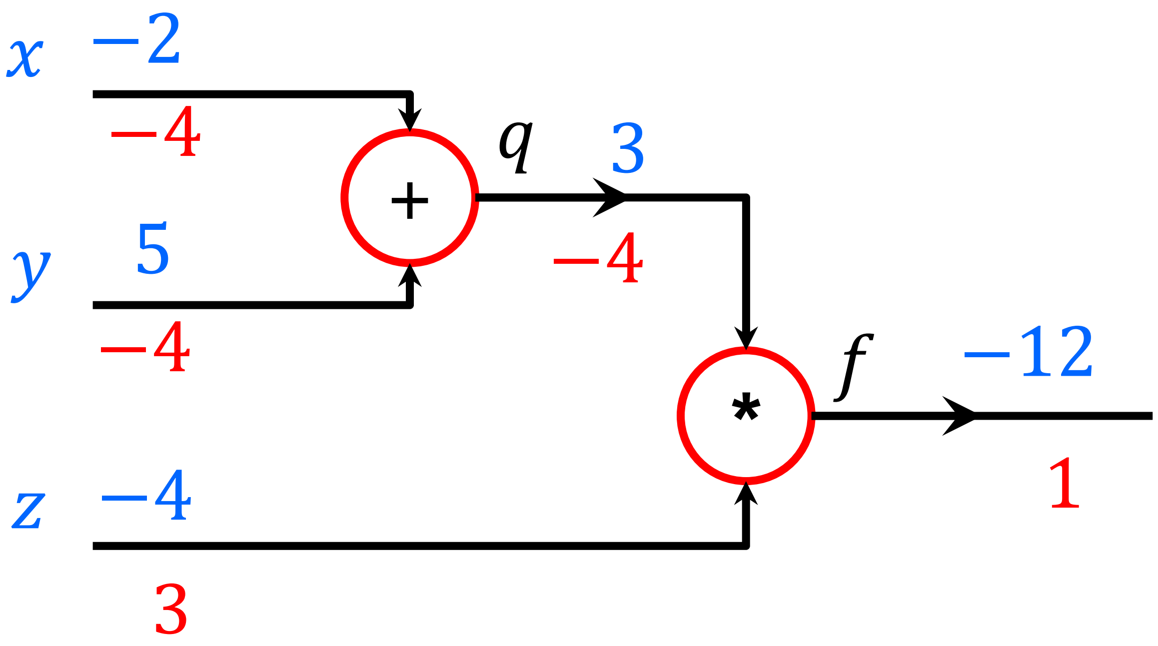
$$\boldsymbol{f(x,y,z) = (x+y)z}$$

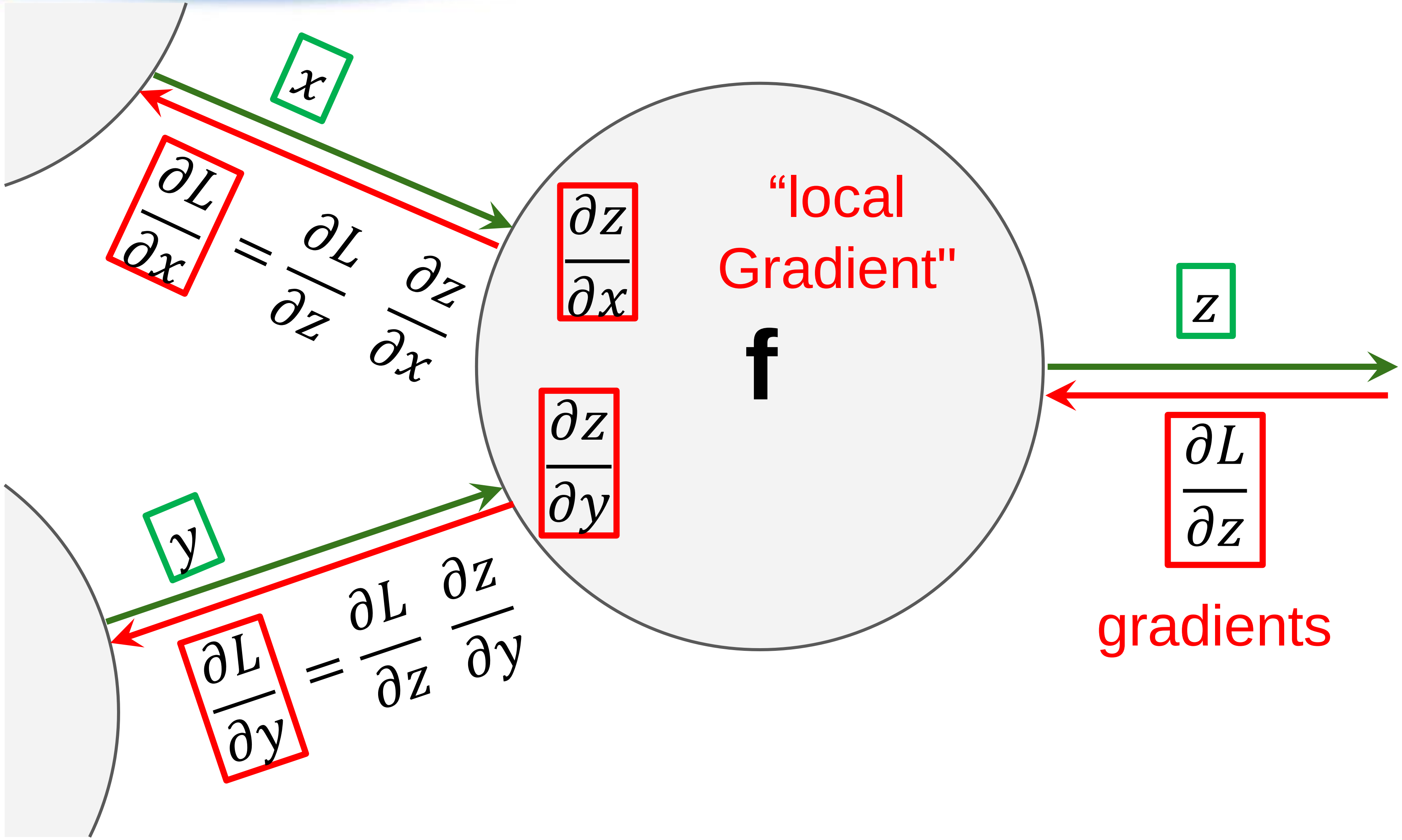
e.g. $x = -2, y = 5, z = -4$

– Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$

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

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





x
$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial z}\frac{\partial z}{\partial x}$
$\frac{\partial z}{\partial x}$
"local Gradient"
f
$\frac{\partial z}{\partial y}$
z
$\frac{\partial L}{\partial z}$
gradients
y
$\frac{\partial L}{\partial y} = \frac{\partial L}{\partial z}\frac{\partial z}{\partial y}$

Another example

BACKPROPAGATION

Backpropagation – Another example

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

E.g:

$w_0 = 2 \qquad x_0 = -1$

$w_1 = -3 \qquad x_1 = -2$

$w_2 = -3$

Want: $\frac{\partial f}{\partial w_0} \quad \frac{\partial f}{\partial x_0} \quad \frac{\partial f}{\partial w_1} \quad \frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial w_2}$

Backpropagation – Another example

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

– Remind

$$f(x) = e^x \qquad \frac{\partial f}{\partial x} = e^x$$

$$f_a(x) = ax \qquad \frac{\partial f}{\partial x} = a$$

$$f(x) = \frac{1}{x} \qquad \frac{\partial f}{\partial x} = -\frac{1}{x^2}$$

$$f_c(x) = c + x \qquad \frac{\partial f}{\partial x} = 1$$

Backpropagation – Another example

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

$w_0 \longrightarrow$

$x_0 \longrightarrow$

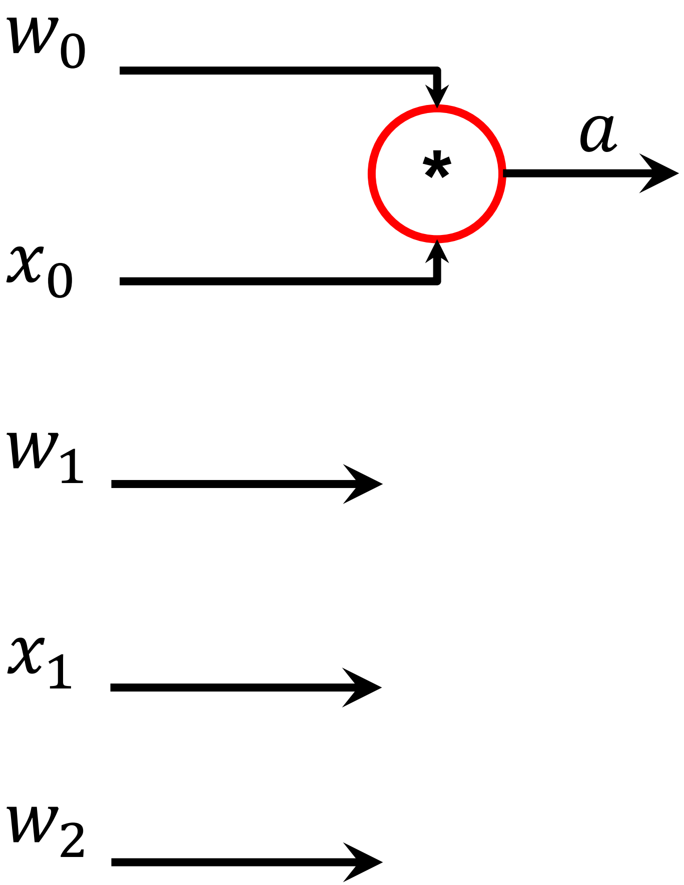
$w_1 \longrightarrow$

$x_1 \longrightarrow$

$w_2 \longrightarrow$

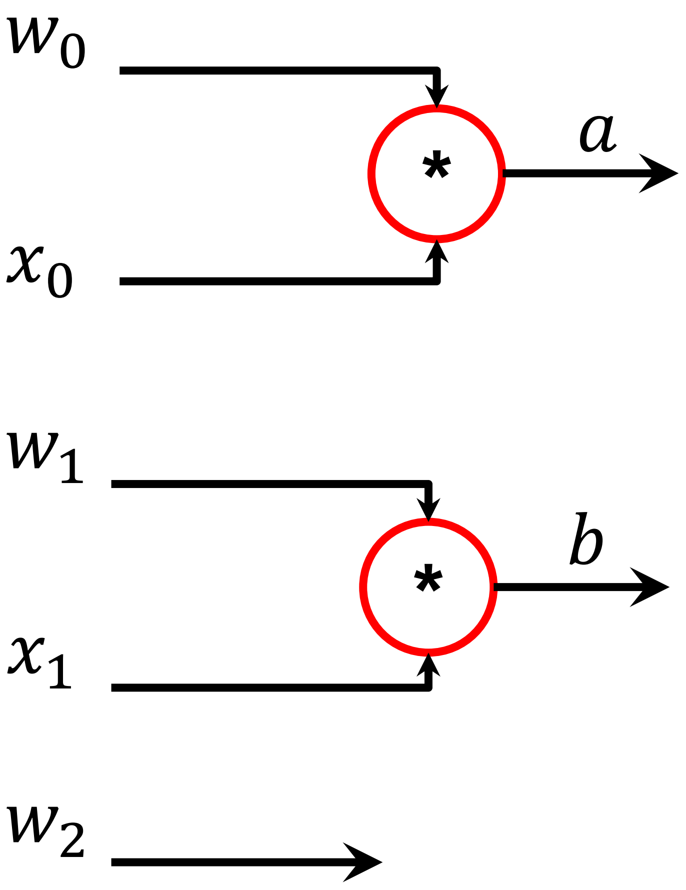
Backpropagation – Another example

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



Backpropagation – Another example

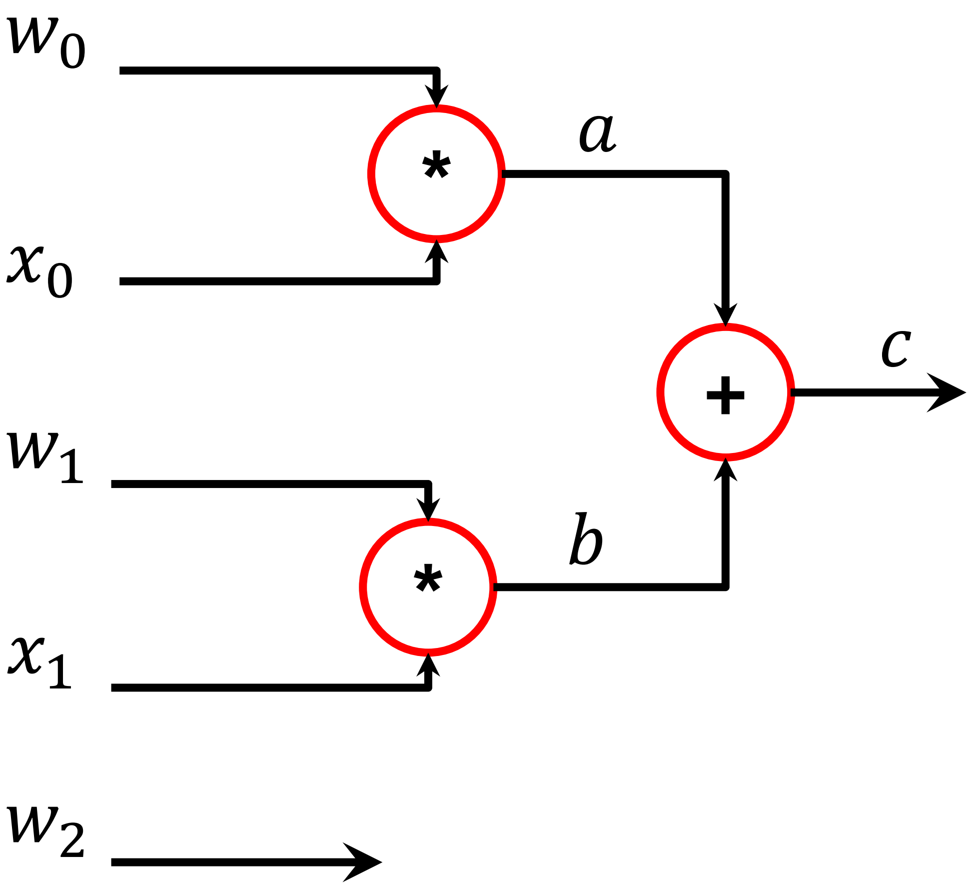
$$f(w,x) = \frac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$$



University of Information Technology

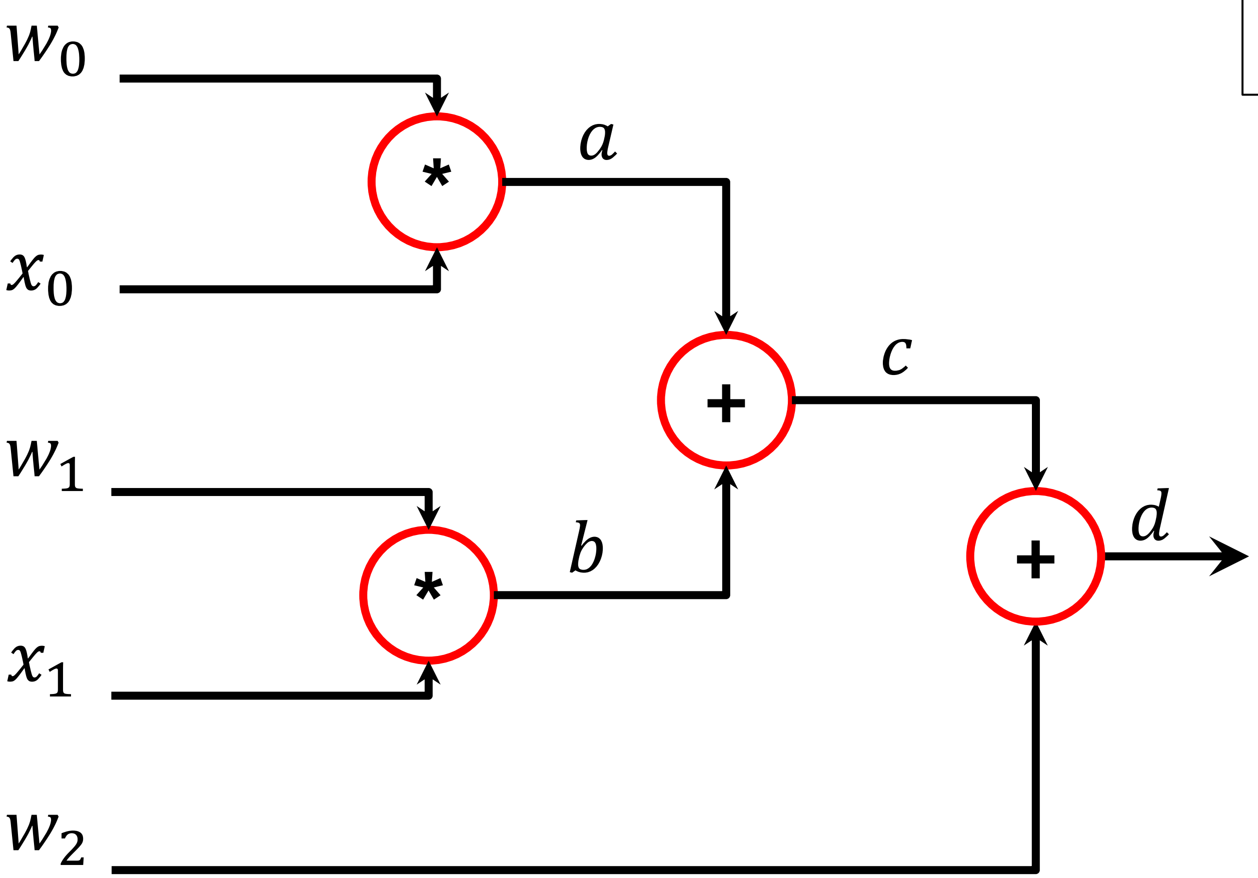
Backpropagation – Another example

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



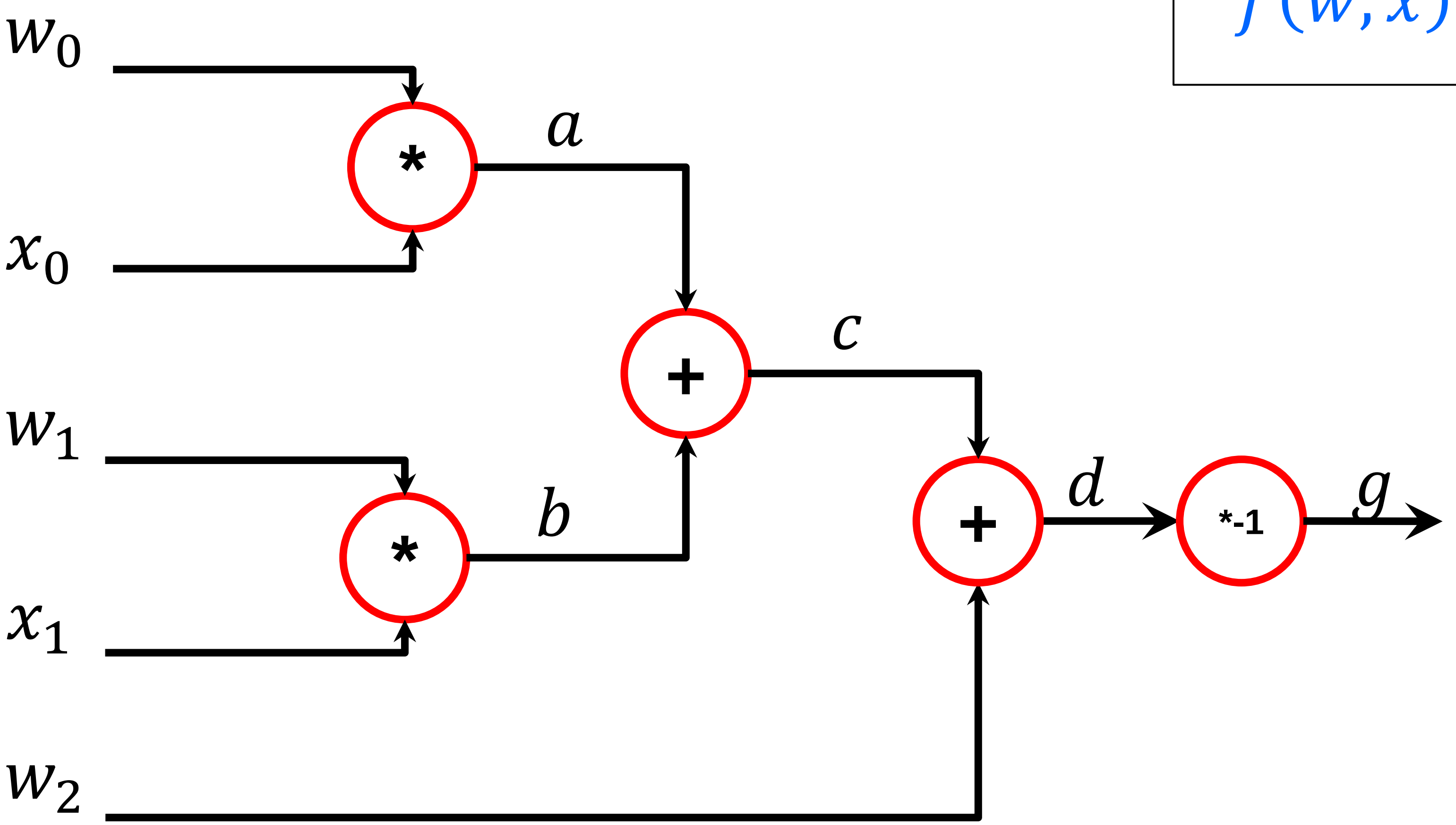
Backpropagation – Another example

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



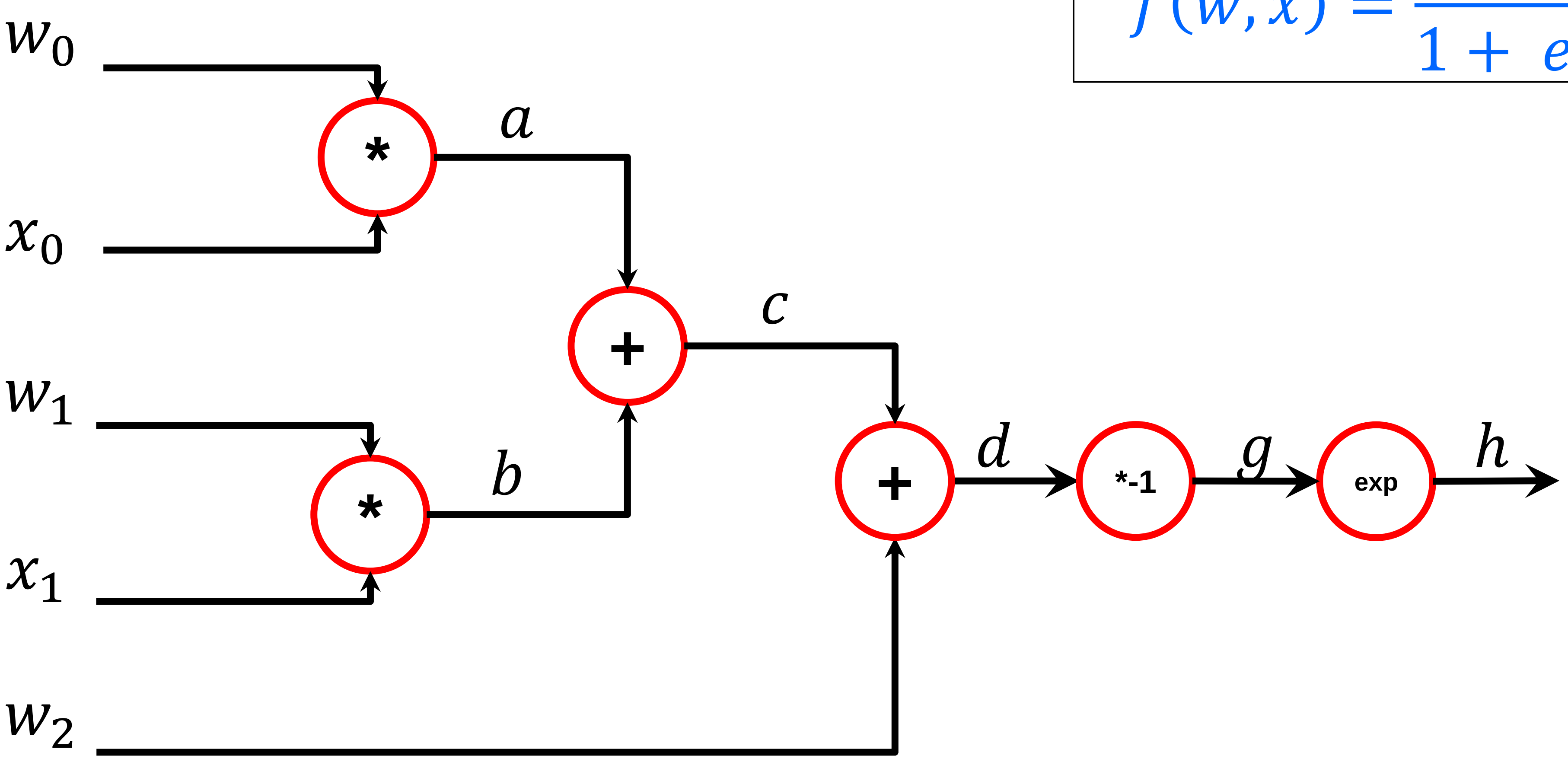
Backpropagation – Another example

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



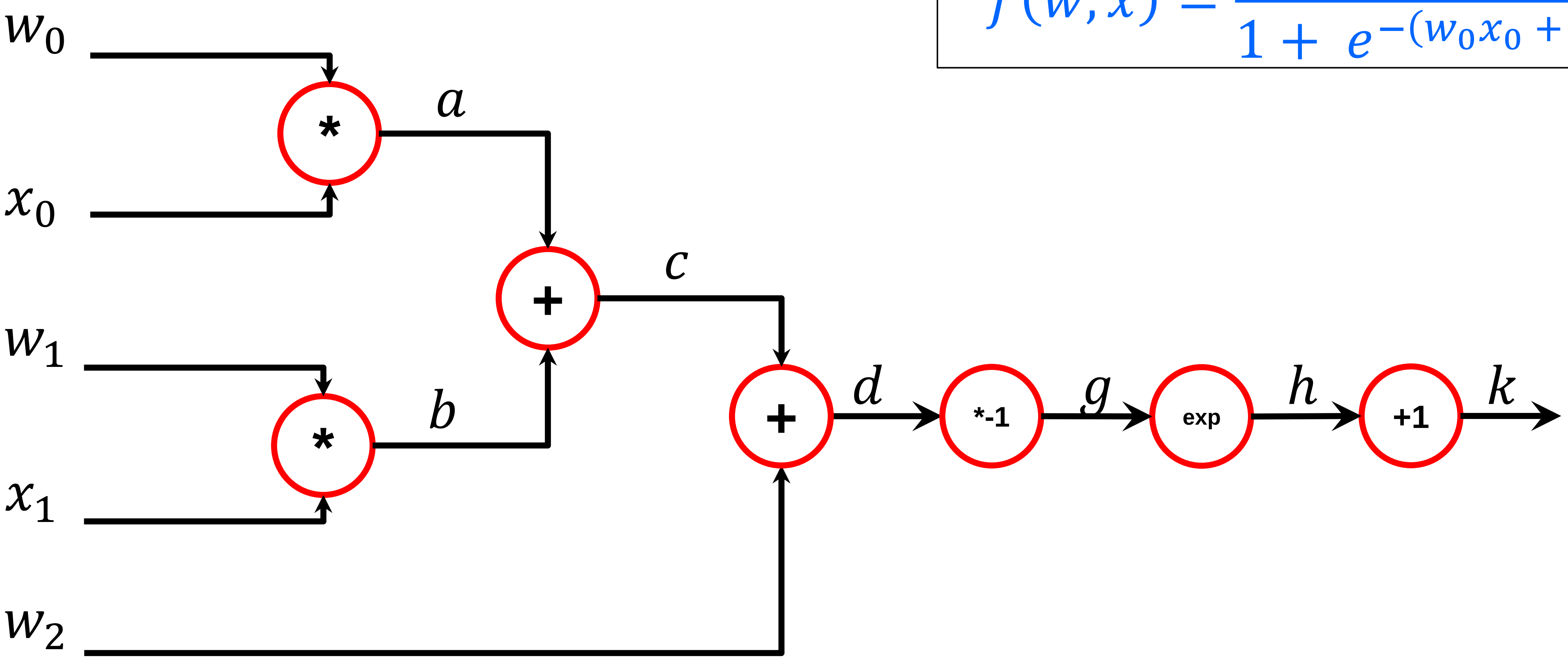
Backpropagation – Another example

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



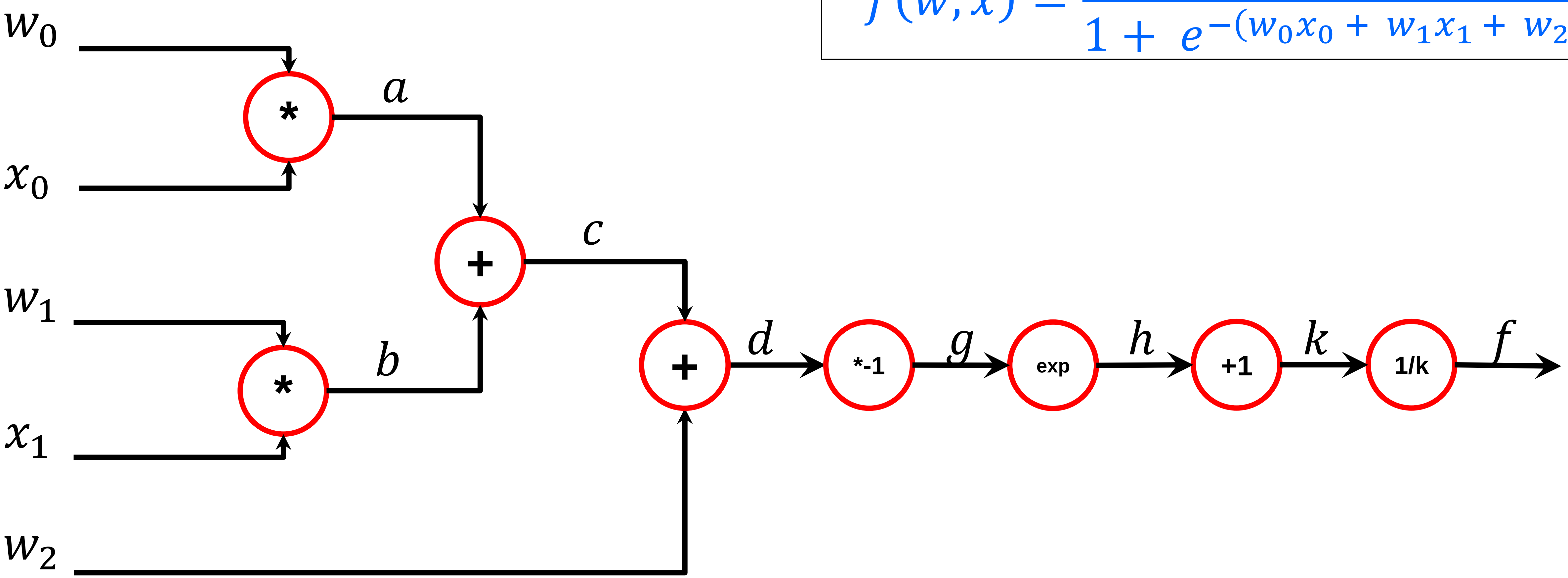
Backpropagation – Another example

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



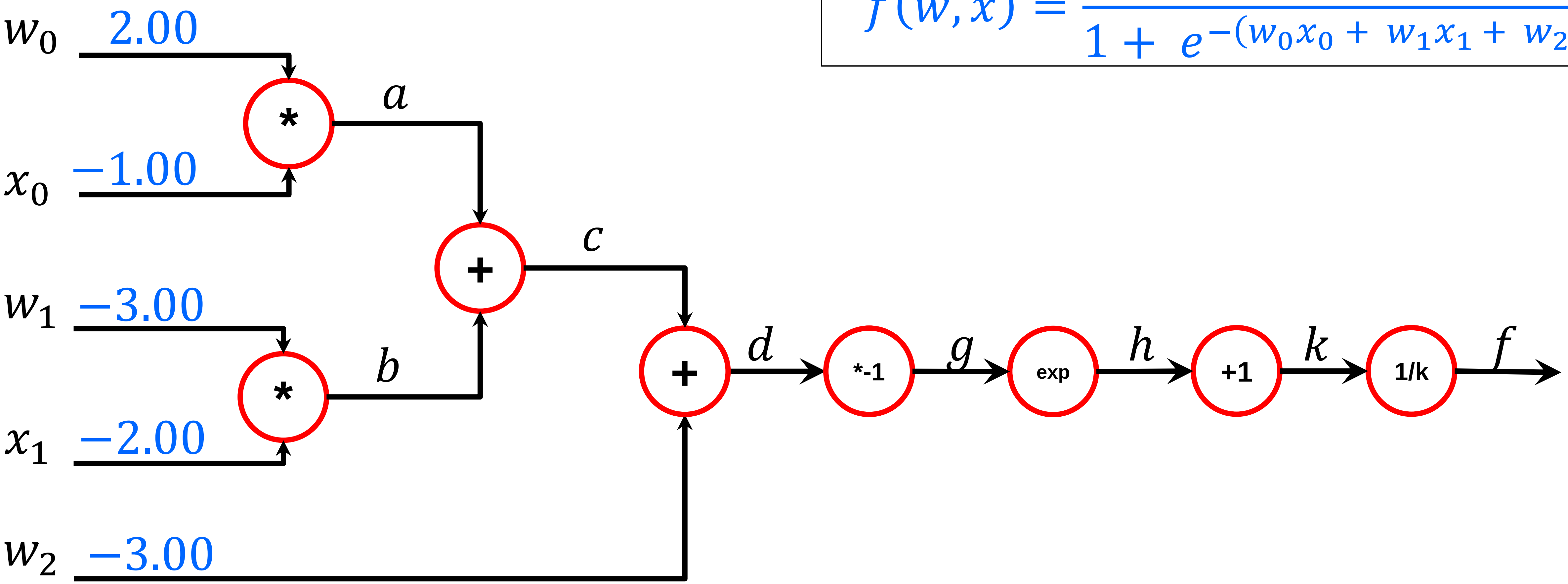
Backpropagation – Another example

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



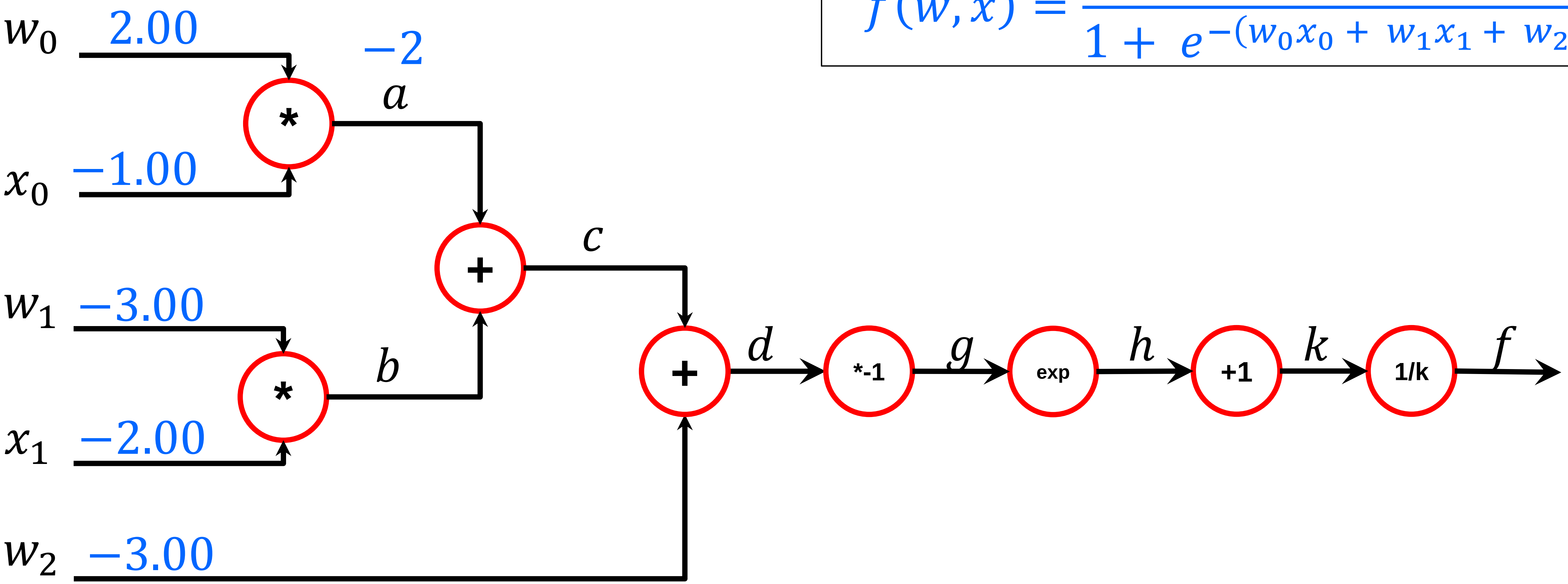
Backpropagation – Another example

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



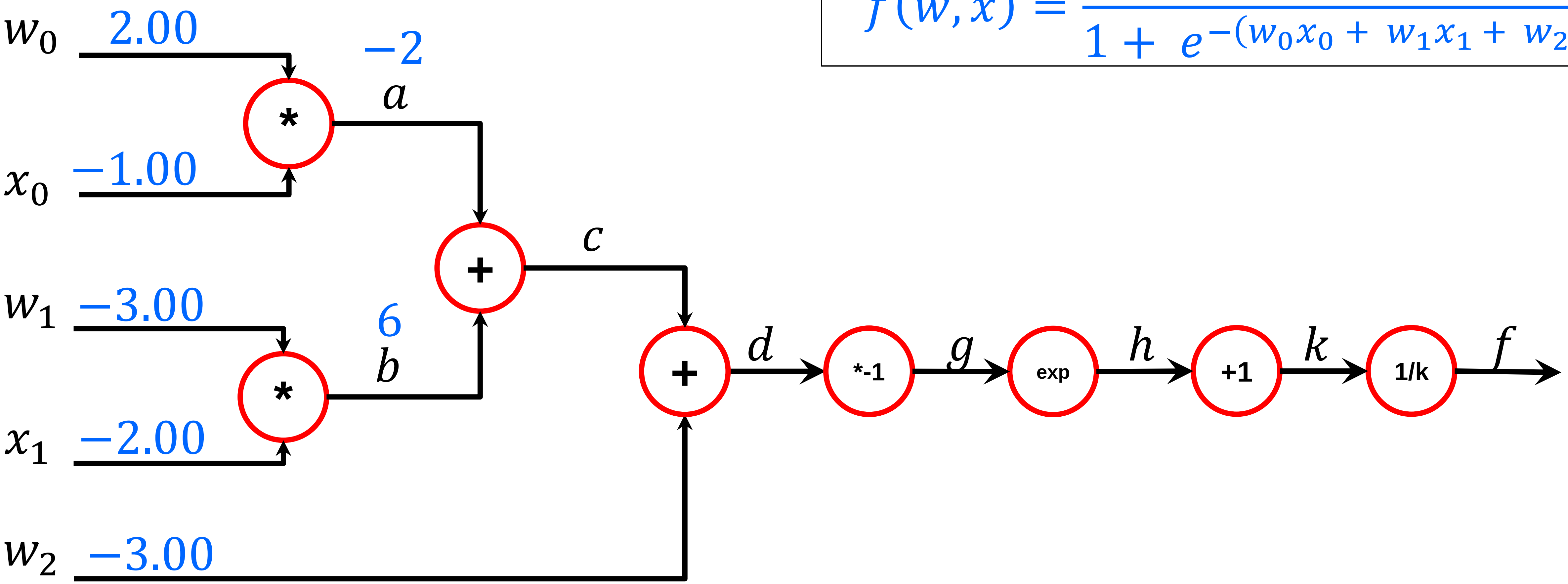
Backpropagation – Another example

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



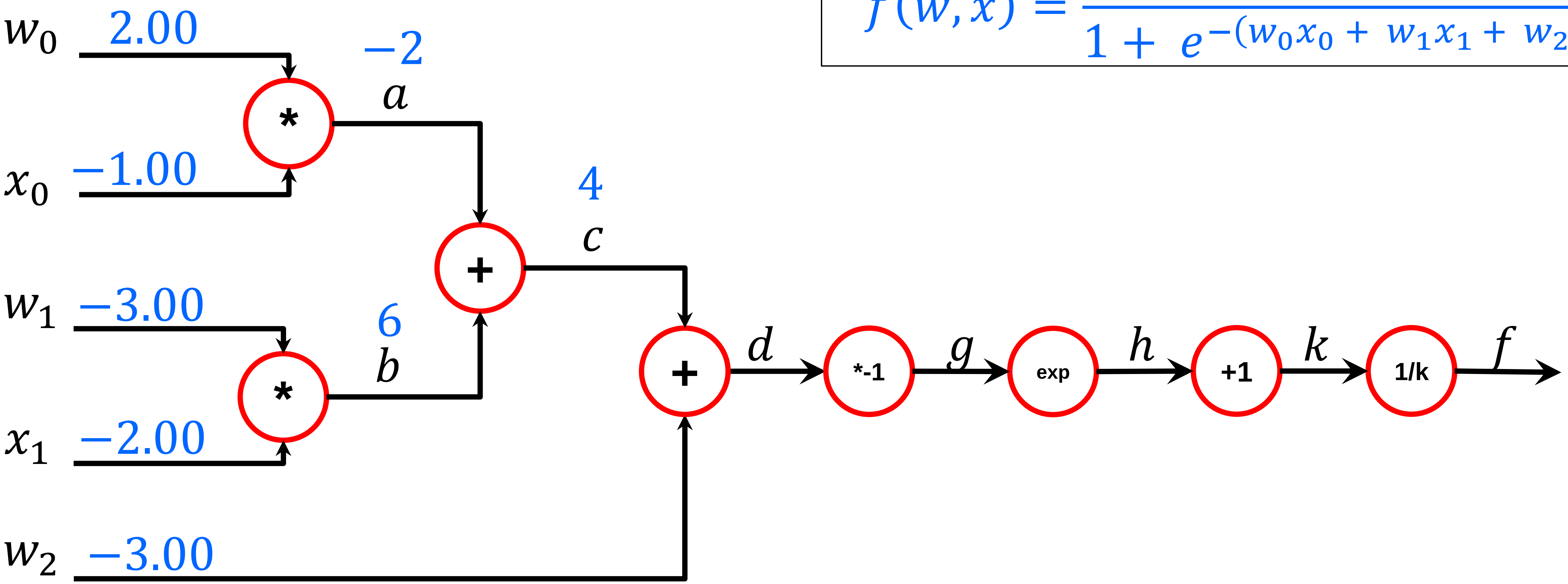
Backpropagation – Another example

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



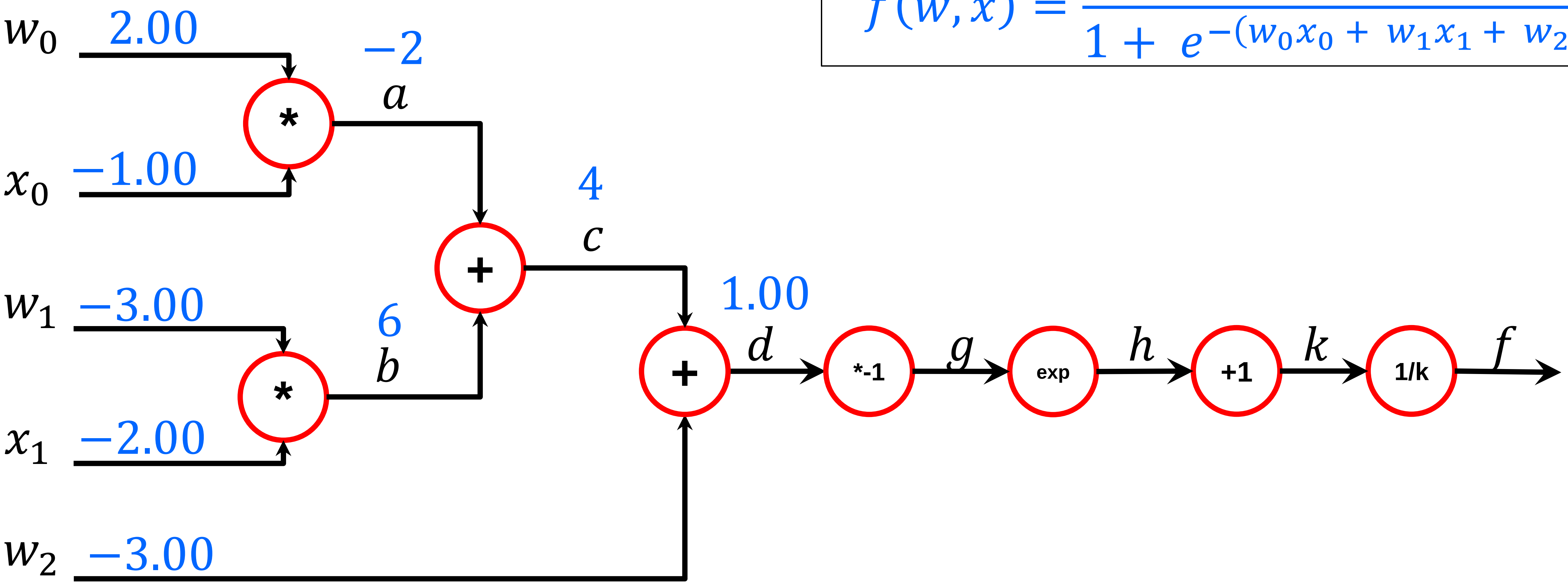
Backpropagation – Another example

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



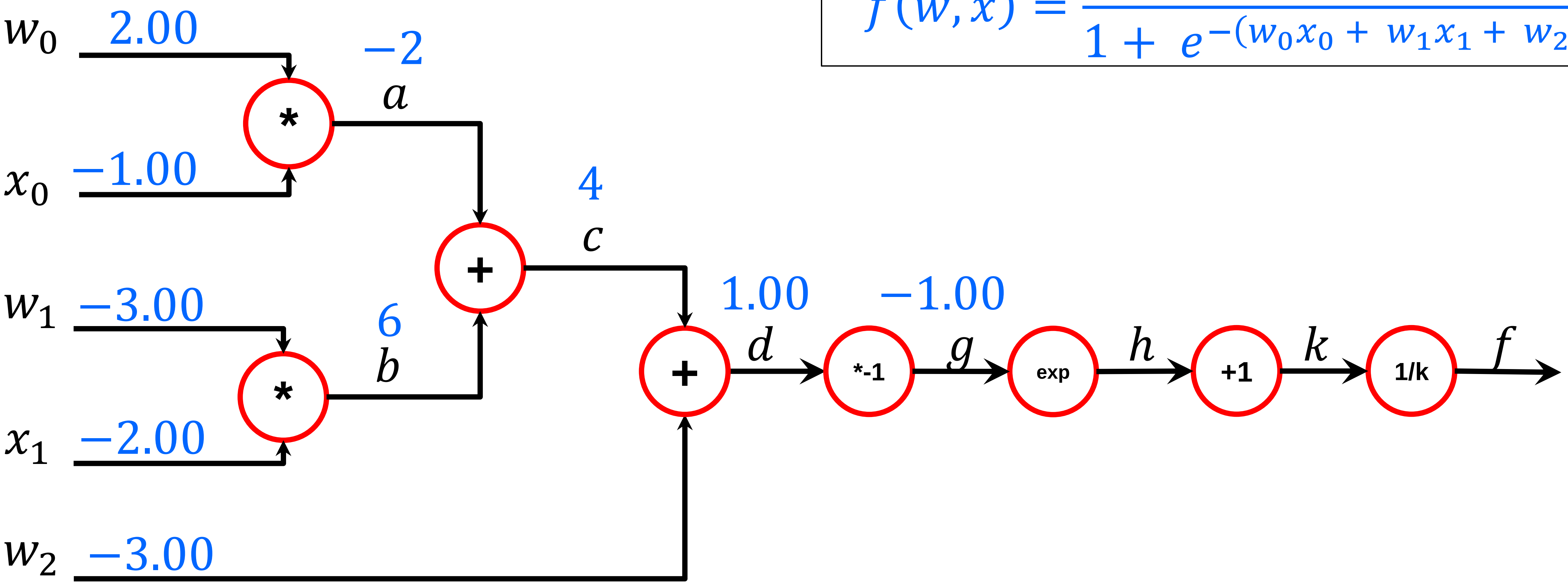
Backpropagation – Another example

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



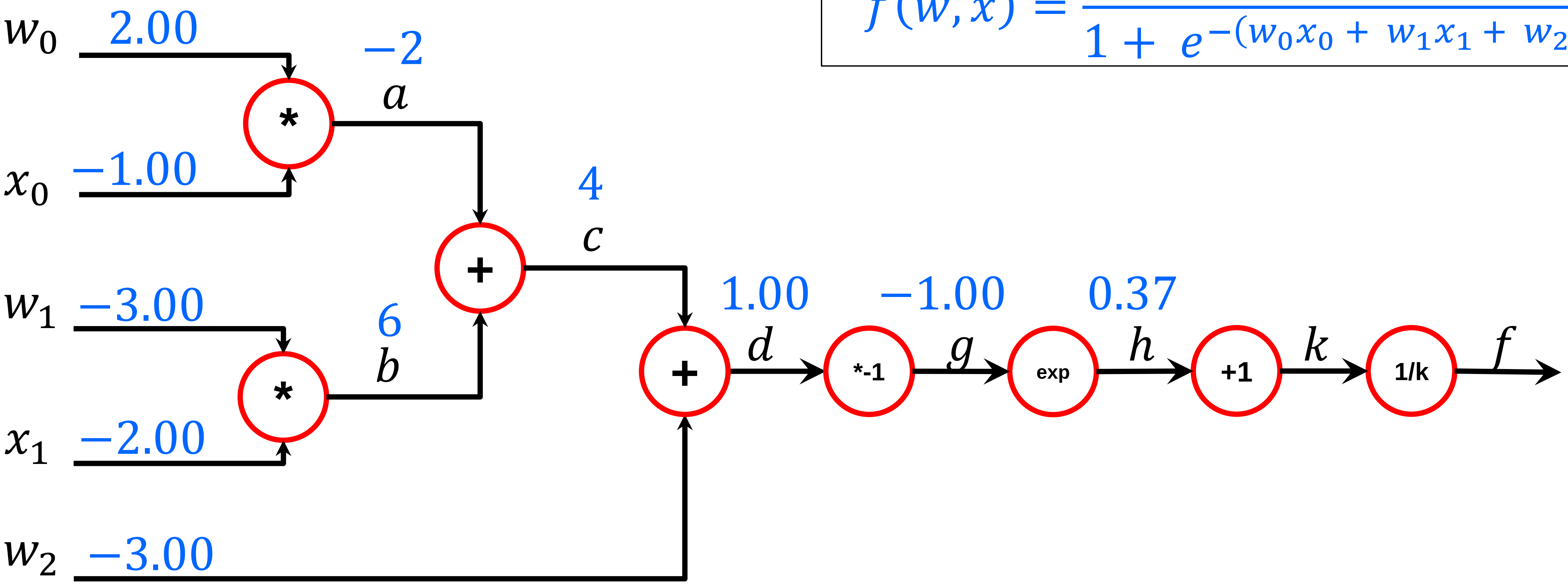
Backpropagation – Another example

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



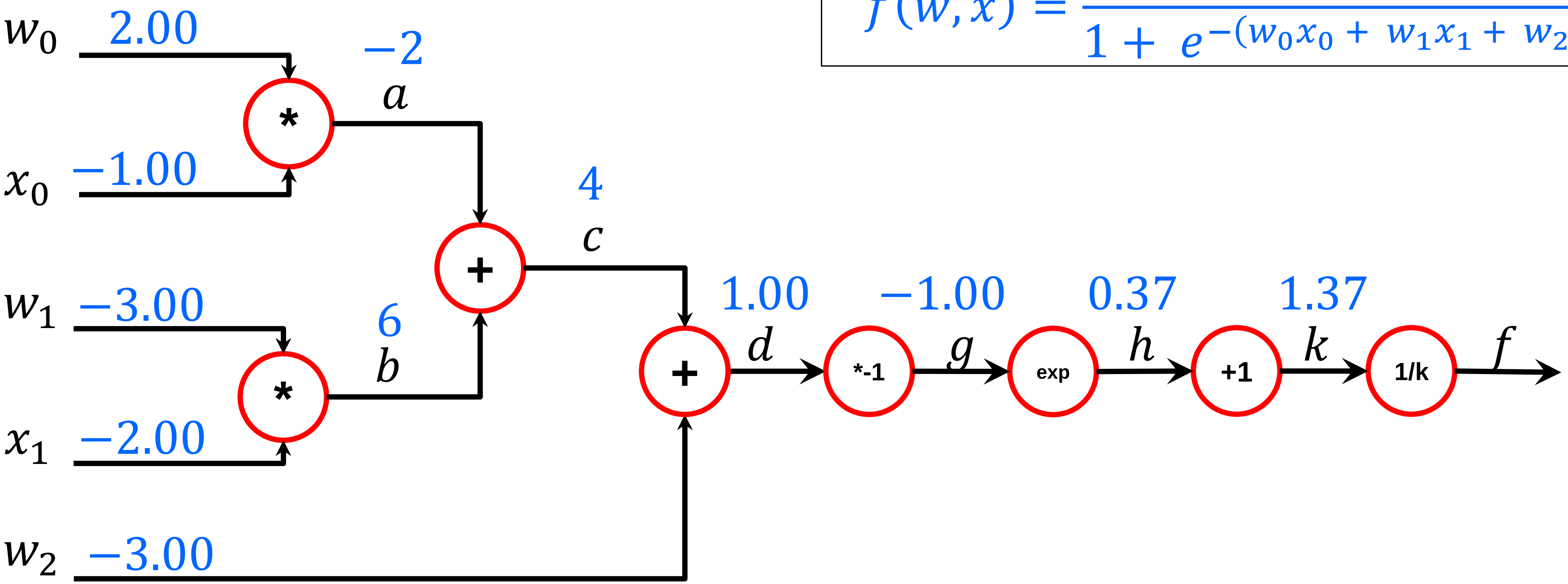
Backpropagation – Another example

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



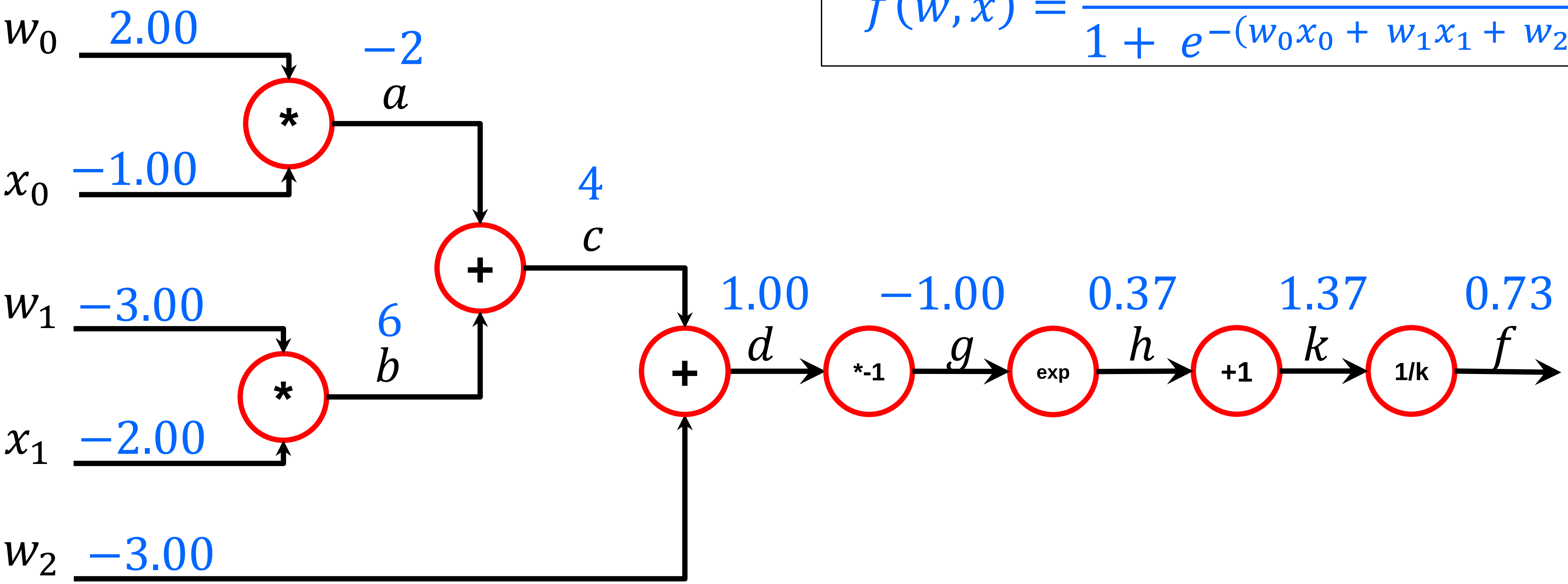
Backpropagation – Another example

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



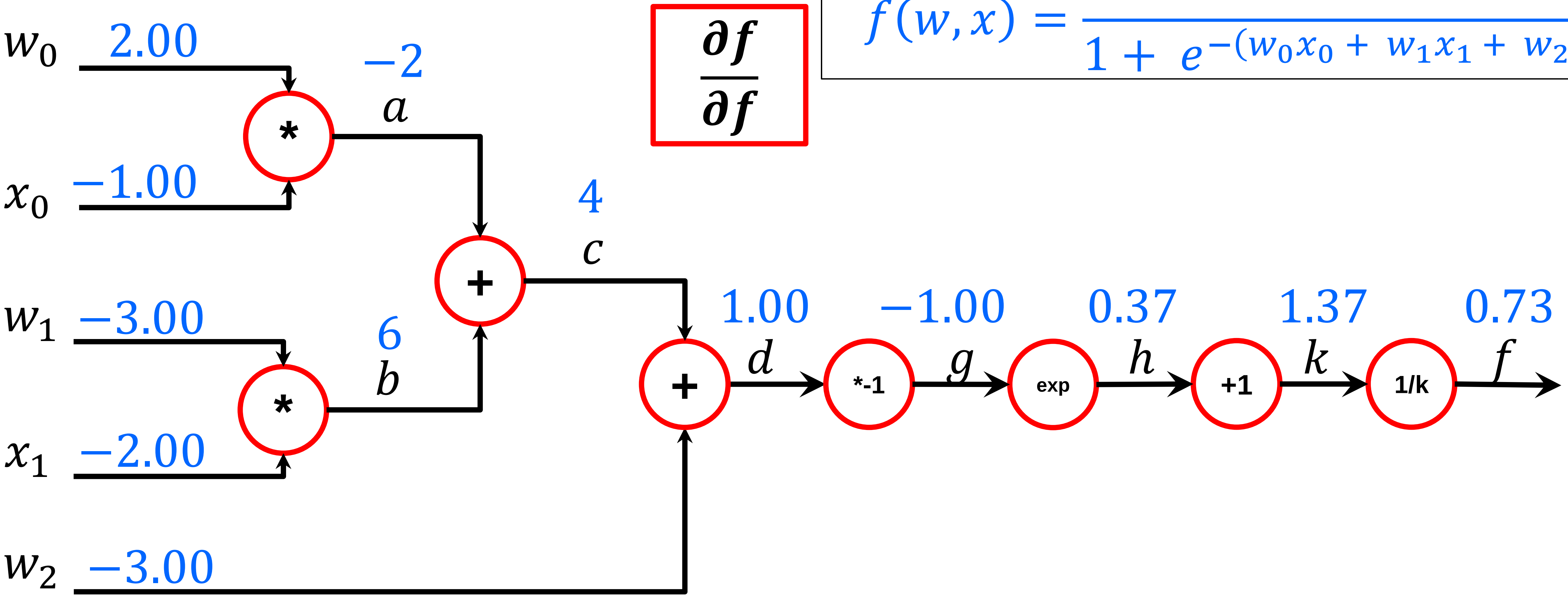
Backpropagation – Another example

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



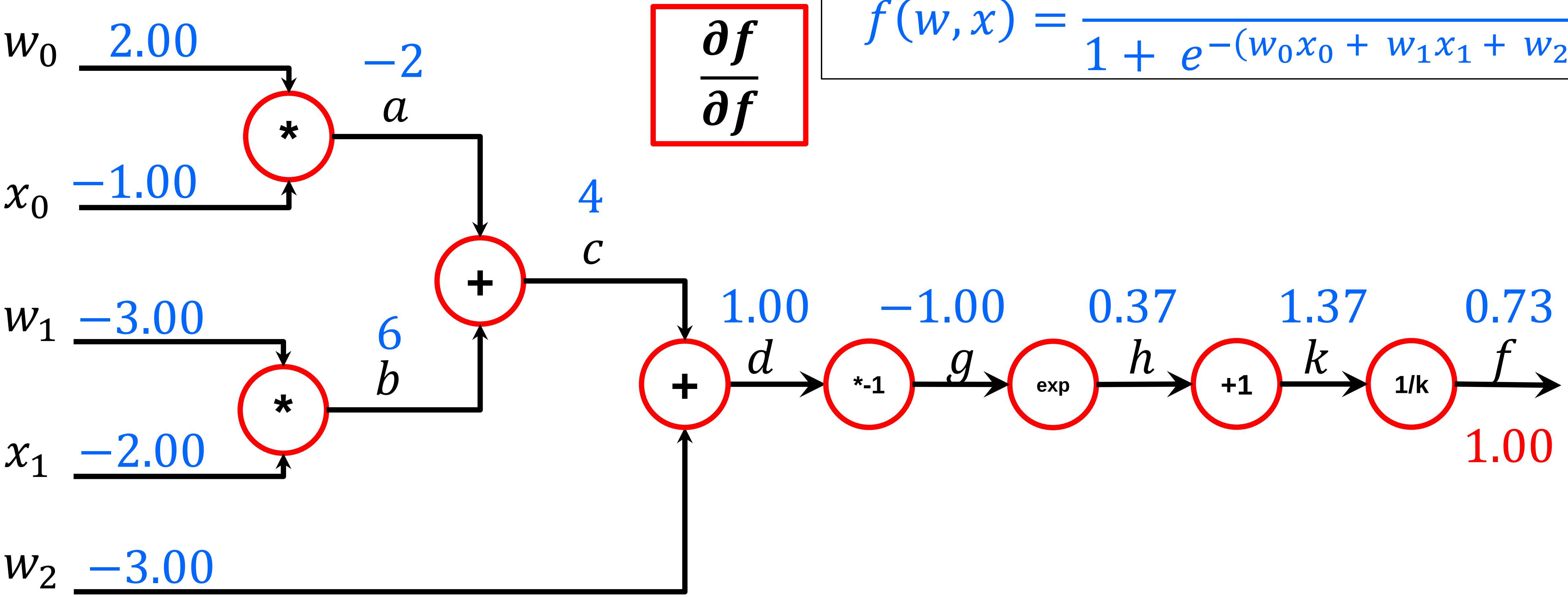
Backpropagation – Another example

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



Backpropagation – Another example

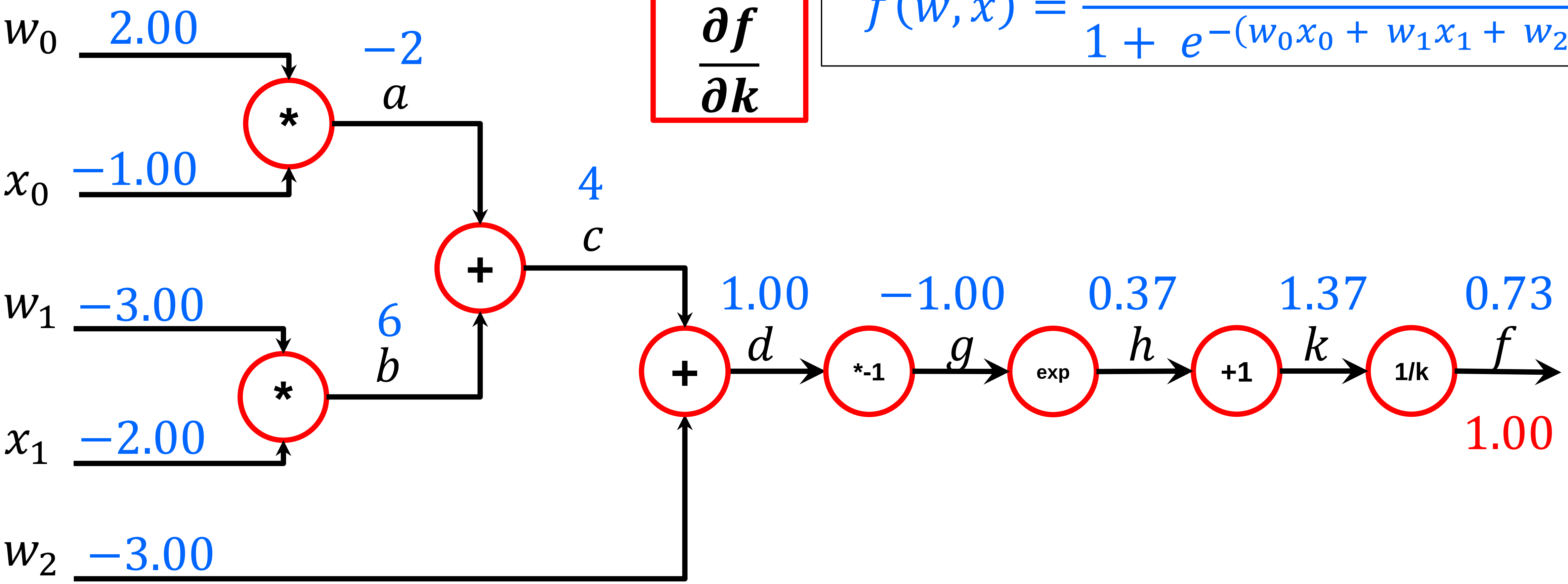
$$f(w,x) = \frac{1}{1 + e^{-(w_0x_0 + w_1x_1 + w_2)}}$$



Backpropagation – Another example

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

$$\frac{\partial f}{\partial k}$$

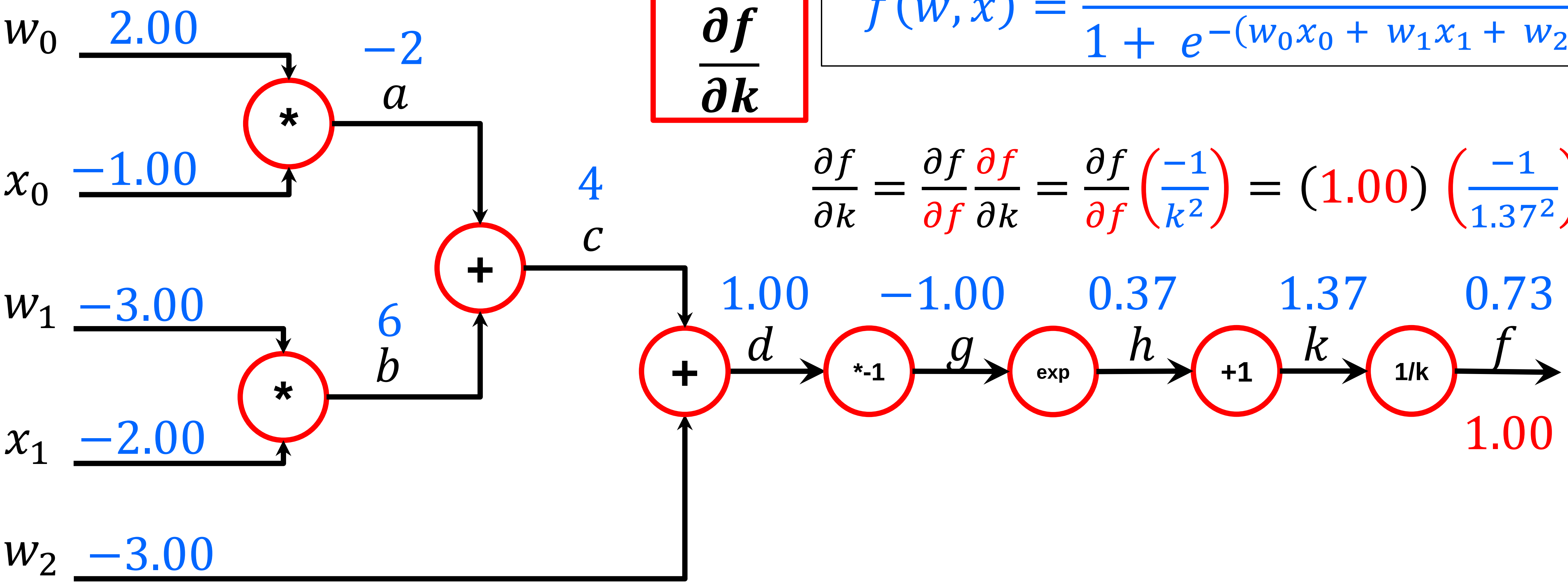


Backpropagation – Another example

$$\boxed{\frac{\partial f}{\partial k}}$$

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

$$\frac{\partial f}{\partial k} = \frac{\partial f}{\partial f}\frac{\partial f}{\partial k} = \frac{\partial f}{\partial f}\left(\frac{-1}{k^2}\right) = (1.00)\left(\frac{-1}{1.37^2}\right)$$

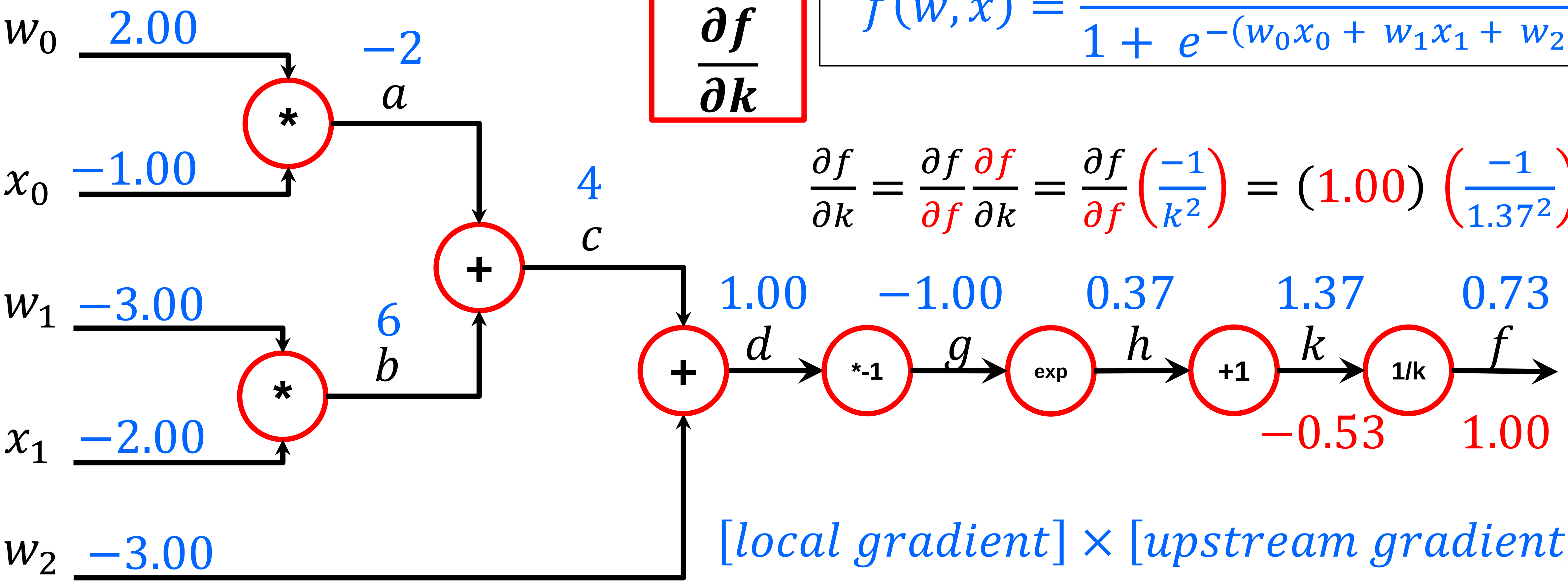


Backpropagation – Another example

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

$$\frac{\partial f}{\partial k}$$

$$\frac{\partial f}{\partial k} = \frac{\partial f}{\partial f}\frac{\partial f}{\partial k} = \frac{\partial f}{\partial f}\left(\frac{-1}{k^2}\right) = (1.00)\left(\frac{-1}{1.37^2}\right)$$

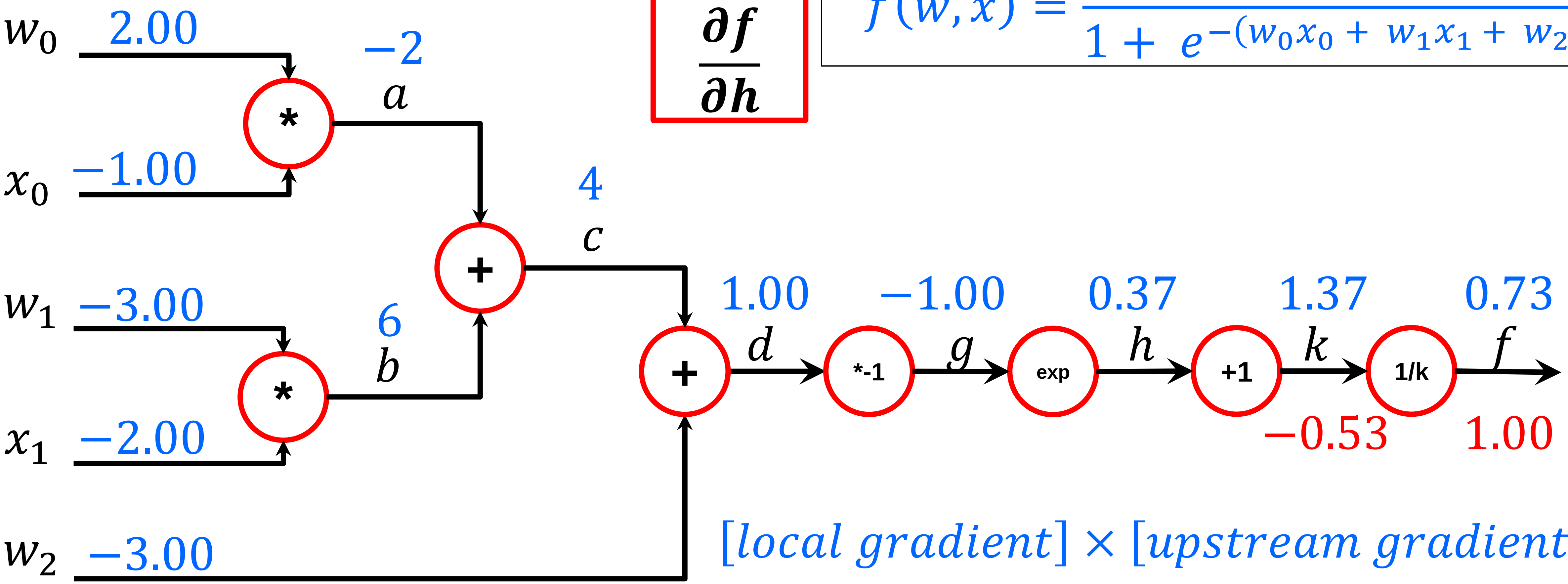


$[local\ gradient] \times [upstream\ gradient]$

Backpropagation – Another example

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

$$\frac{\partial f}{\partial h}$$



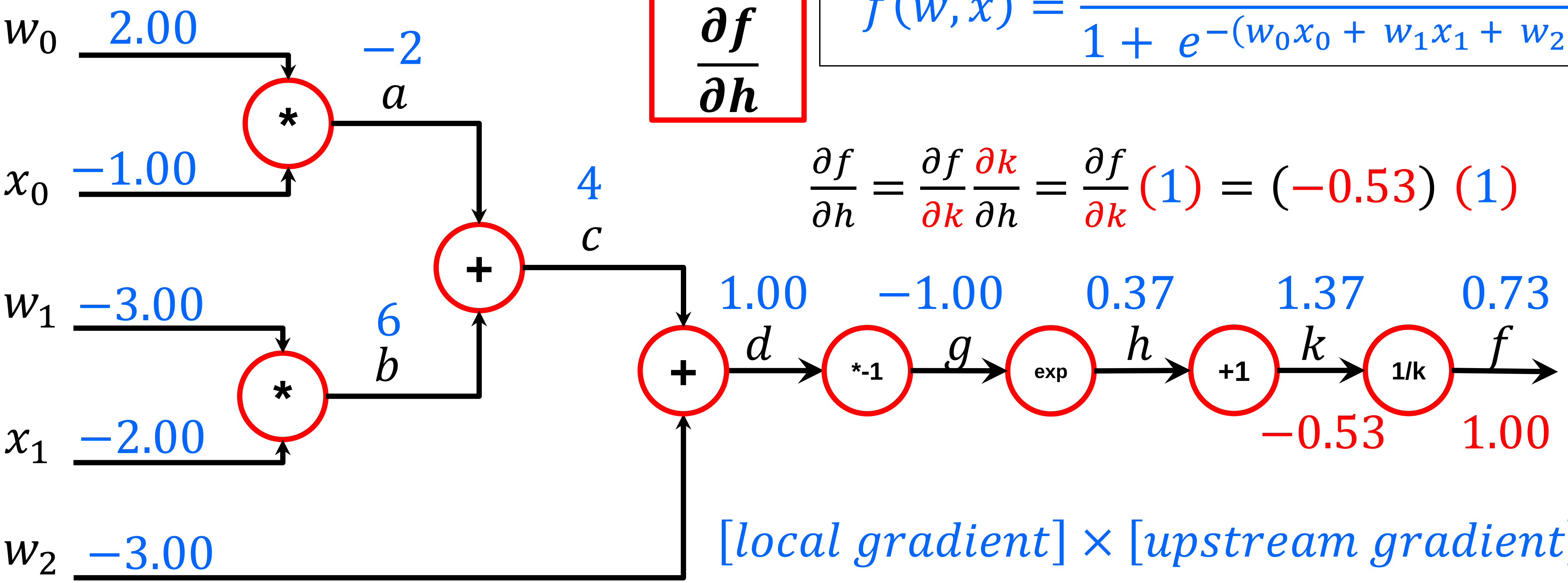
$$[local\ gradient] \times [upstream\ gradient]$$

Backpropagation – Another example

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

$$\boxed{\frac{\partial f}{\partial h}}$$

$$\frac{\partial f}{\partial h} = \frac{\partial f}{\partial k}\frac{\partial k}{\partial h} = \frac{\partial f}{\partial k}(1) = (-0.53)\ (1)$$



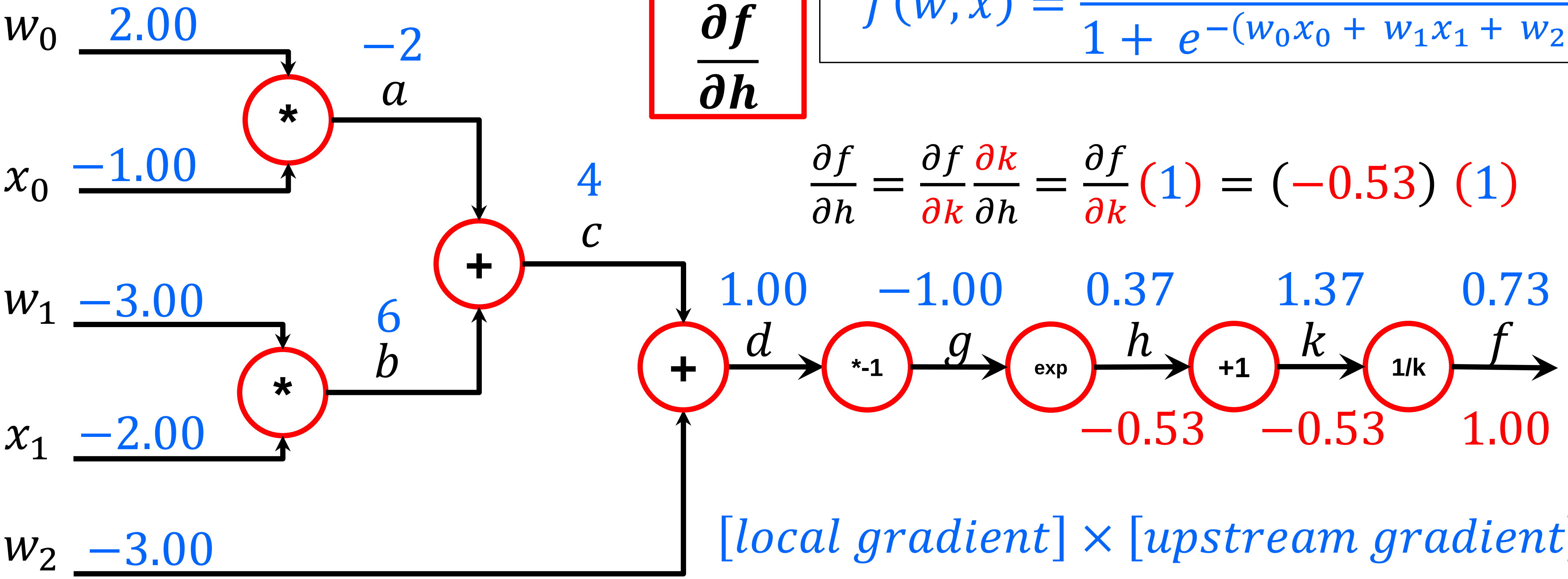
$$[local\ gradient] \times [upstream\ gradient]$$

Backpropagation – Another example

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

$$\frac{\partial f}{\partial h}$$

$$\frac{\partial f}{\partial h} = \frac{\partial f}{\partial k}\frac{\partial k}{\partial h} = \frac{\partial f}{\partial k}(1) = (-0.53)\ (1)$$

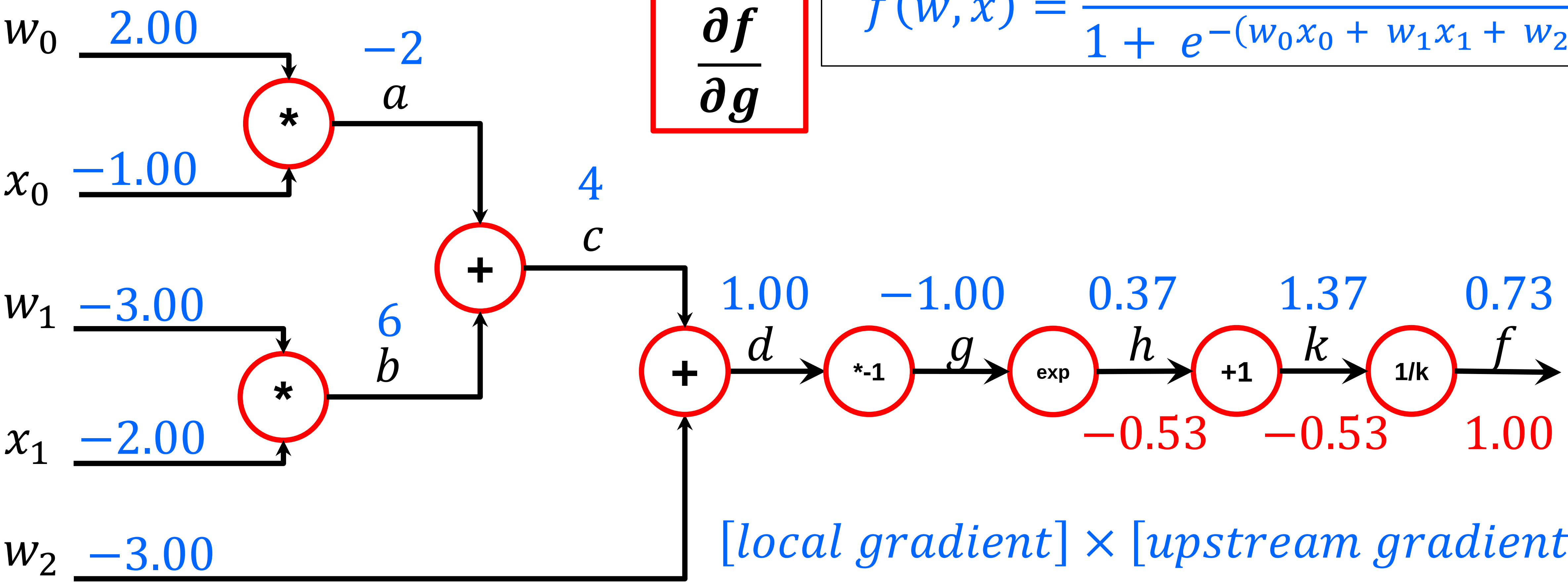


$$[local\ gradient] \times [upstream\ gradient]$$

Backpropagation – Another example

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

$$\frac{\partial f}{\partial g}$$



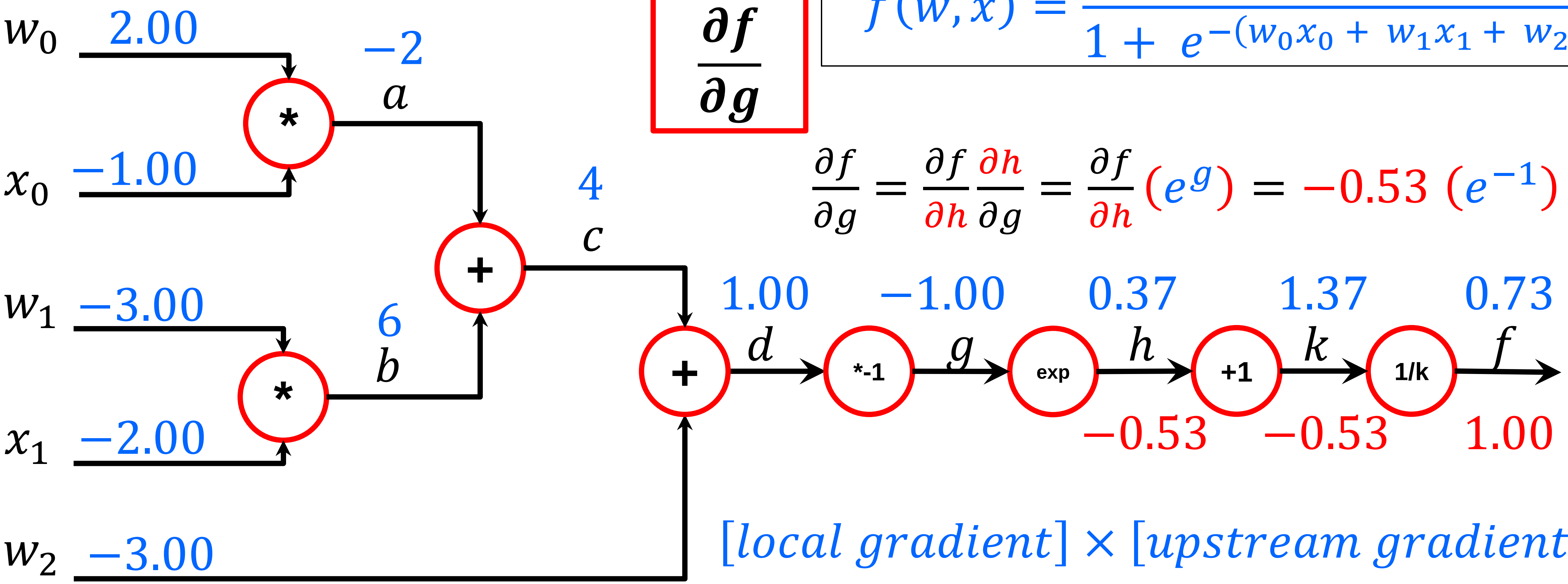
$[local\ gradient] \times [upstream\ gradient]$

Backpropagation – Another example

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

$$\frac{\partial f}{\partial g}$$

$$\frac{\partial f}{\partial g} = \frac{\partial f}{\partial h}\frac{\partial h}{\partial g} = \frac{\partial f}{\partial h}(e^g) = -0.53\,(e^{-1})$$



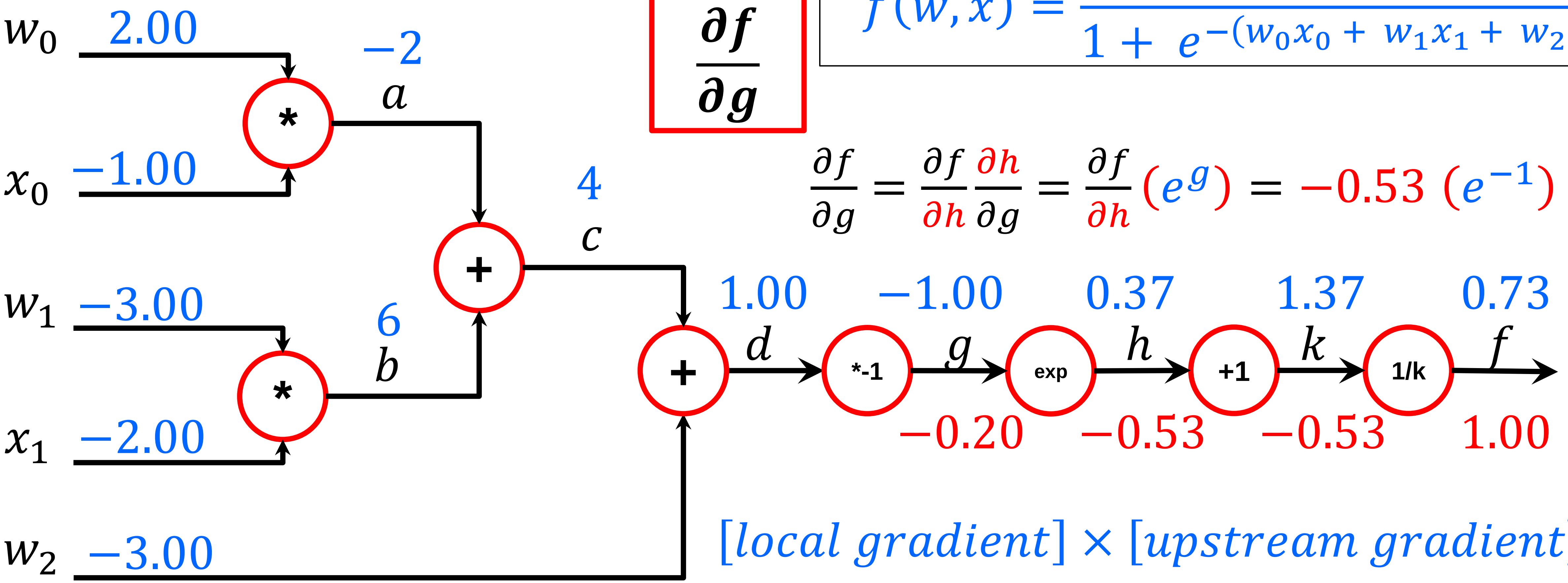
$$[local\ gradient] \times [upstream\ gradient]$$

Backpropagation – Another example

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

$$\boxed{\frac{\partial f}{\partial g}}$$

$$\frac{\partial f}{\partial g} = \frac{\partial f}{\partial h}\frac{\partial h}{\partial g} = \frac{\partial f}{\partial h}(e^{g}) = -0.53\,(e^{-1})$$

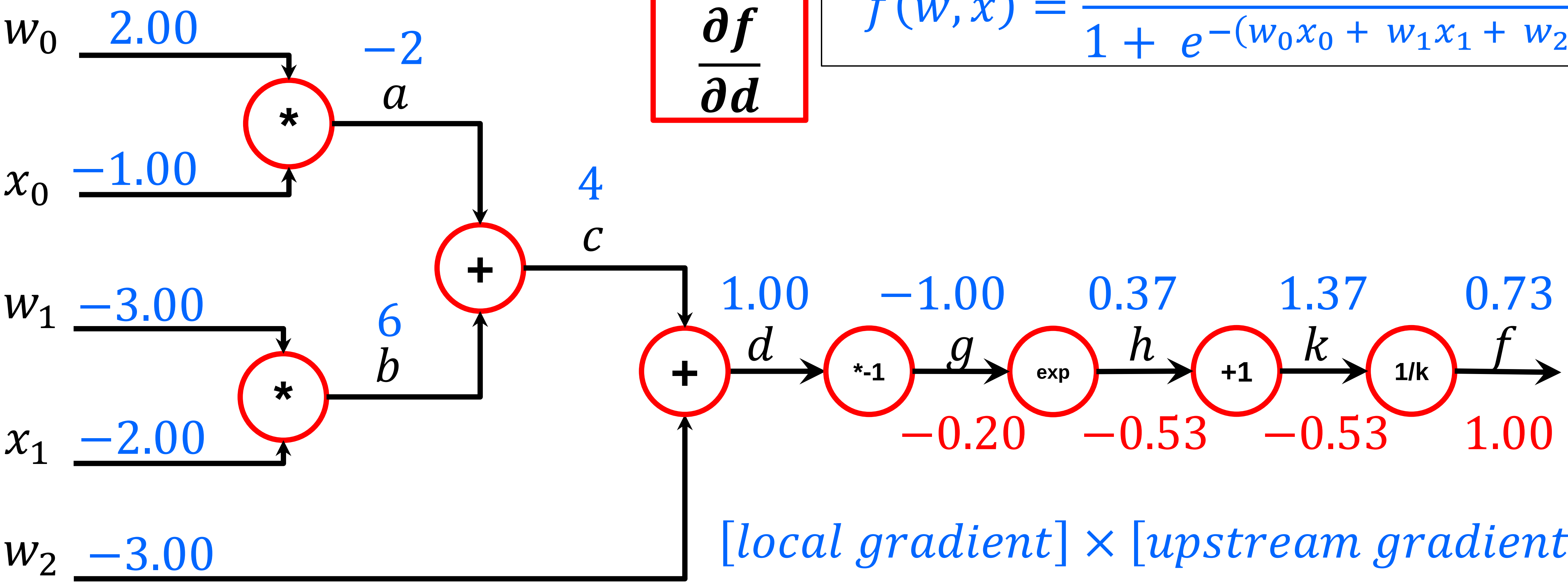


$[local\ gradient] \times [upstream\ gradient]$

Backpropagation – Another example

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

$$\boxed{\frac{\partial f}{\partial d}}$$



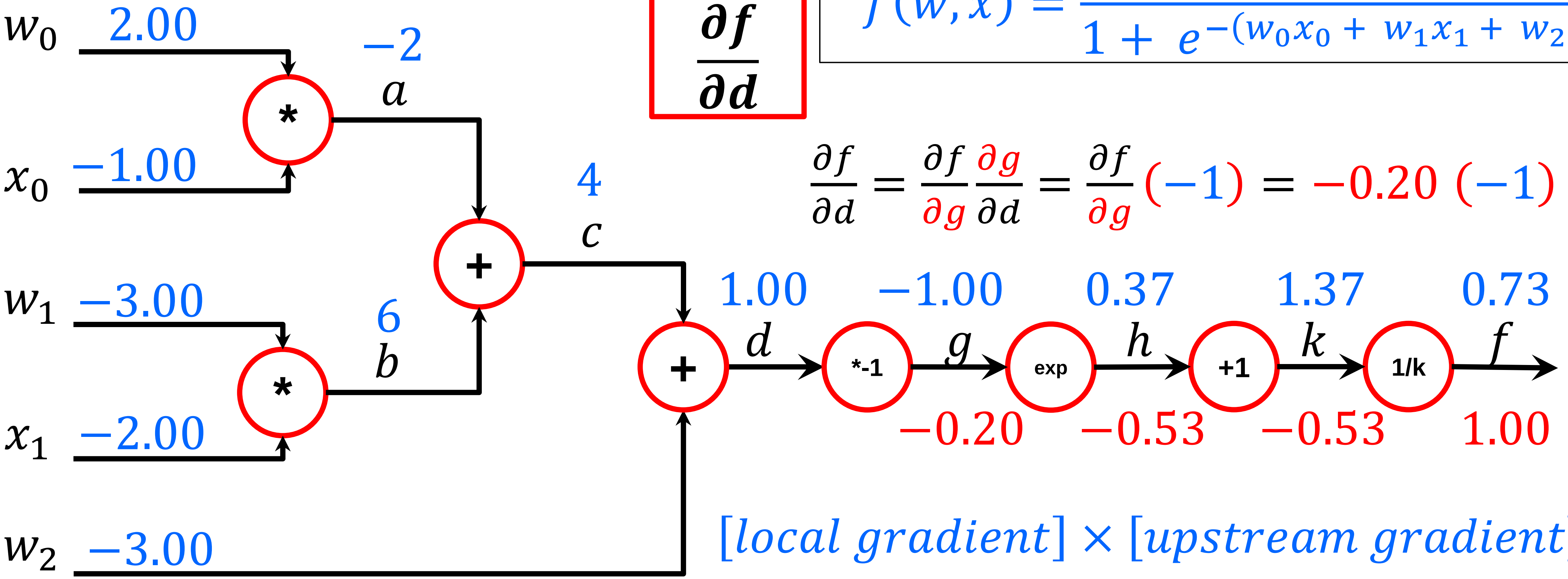
$[local\ gradient] \times [upstream\ gradient]$

Backpropagation – Another example

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

$$\boxed{\frac{\partial f}{\partial d}}$$

$$\frac{\partial f}{\partial d} = \frac{\partial f}{\partial g}\frac{\partial g}{\partial d} = \frac{\partial f}{\partial g}(-1) = -0.20\,(-1)$$



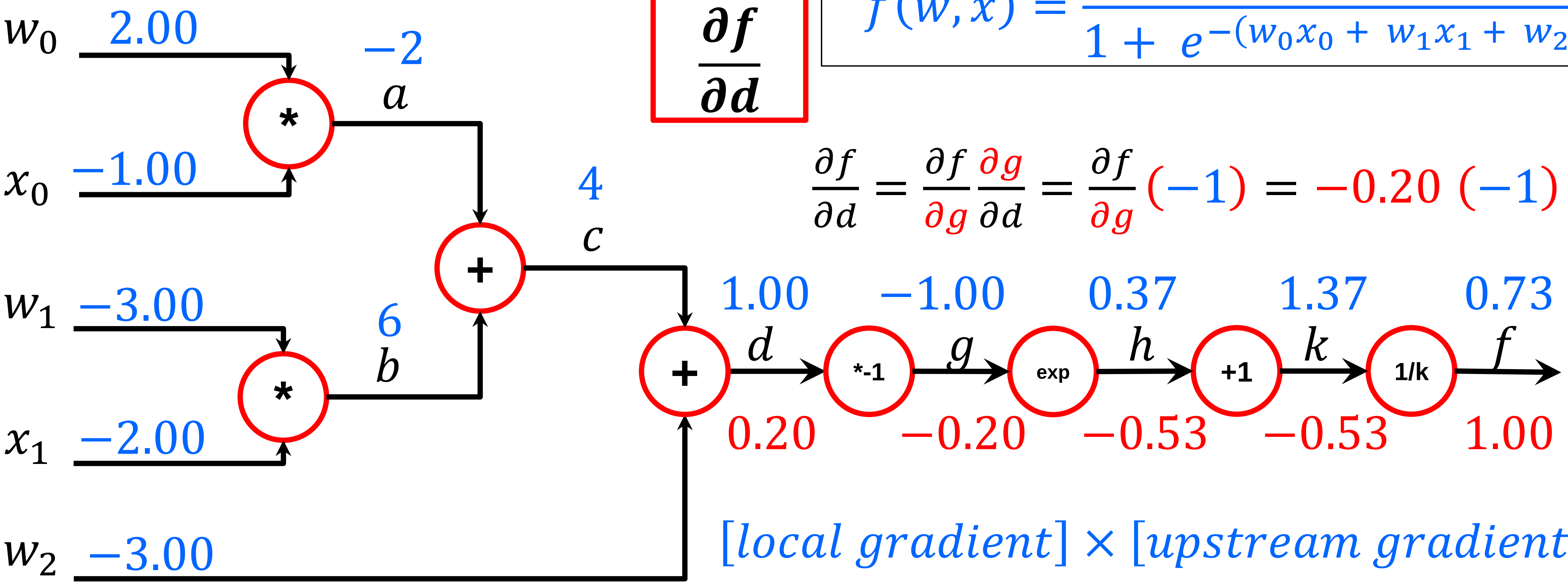
$$[local\ gradient] \times [upstream\ gradient]$$

Backpropagation – Another example

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

$$\frac{\partial f}{\partial d}$$

$$\frac{\partial f}{\partial d} = \frac{\partial f}{\partial g}\frac{\partial g}{\partial d} = \frac{\partial f}{\partial g}(-1) = -0.20\,(-1)$$

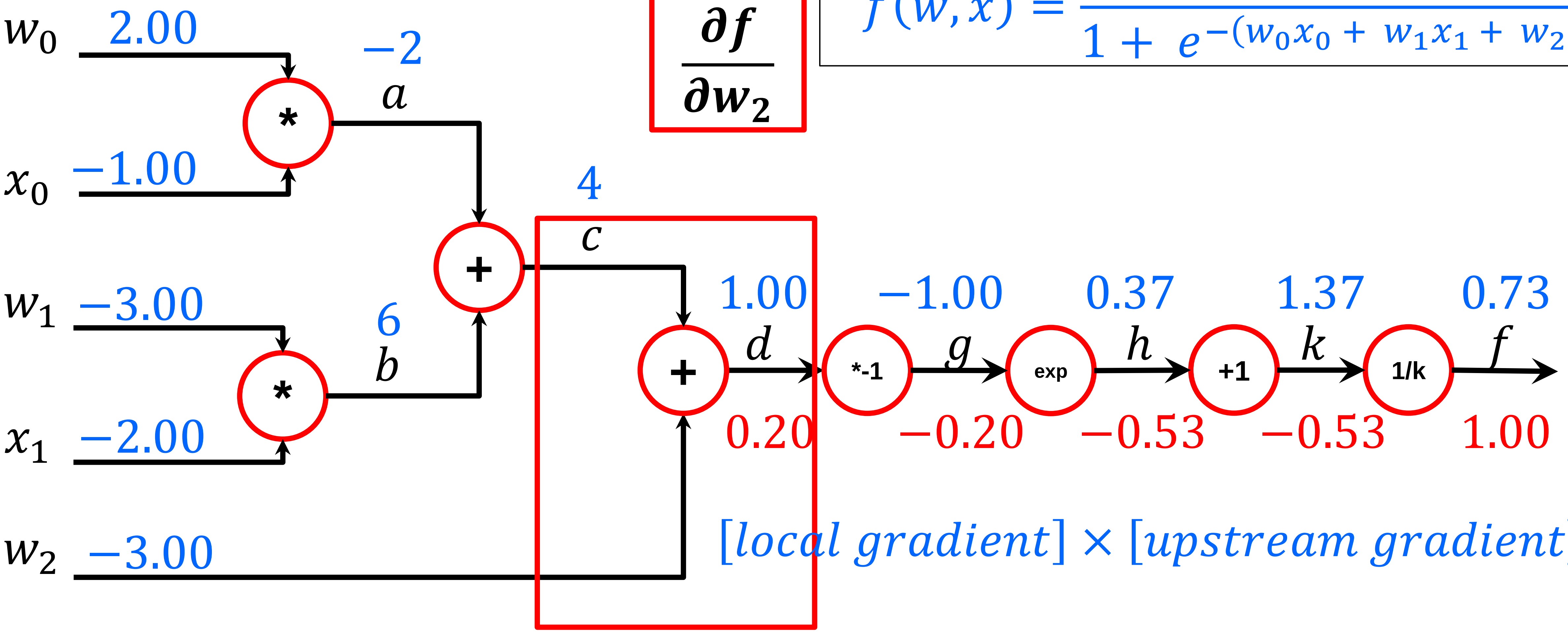


$[local\ gradient] \times [upstream\ gradient]$

Backpropagation – Another example

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

$$\frac{\partial f}{\partial w_2}$$

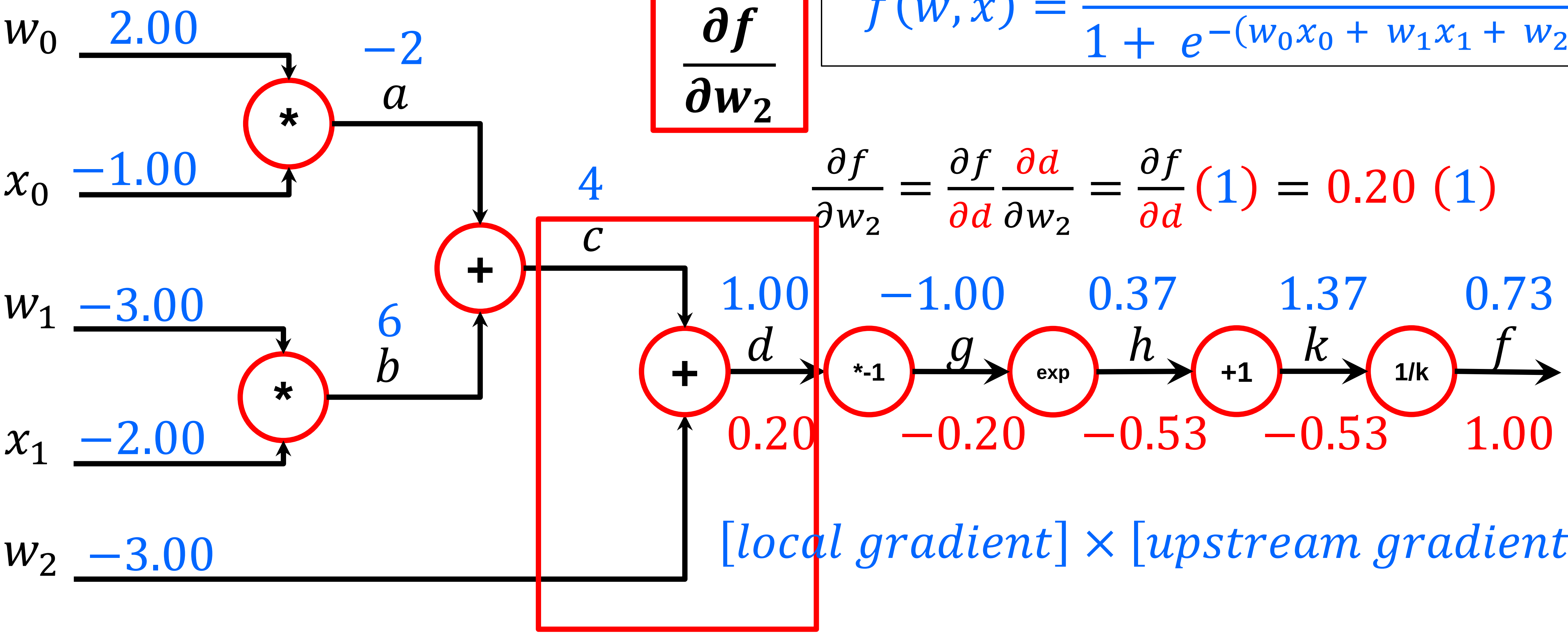


$[local\ gradient] \times [upstream\ gradient]$

Backpropagation – Another example

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

$$\boxed{\frac{\partial f}{\partial w_2}}$$



$$\frac{\partial f}{\partial w_2} = \frac{\partial f}{\partial d}\frac{\partial d}{\partial w_2} = \frac{\partial f}{\partial d}(1) = 0.20\ (1)$$

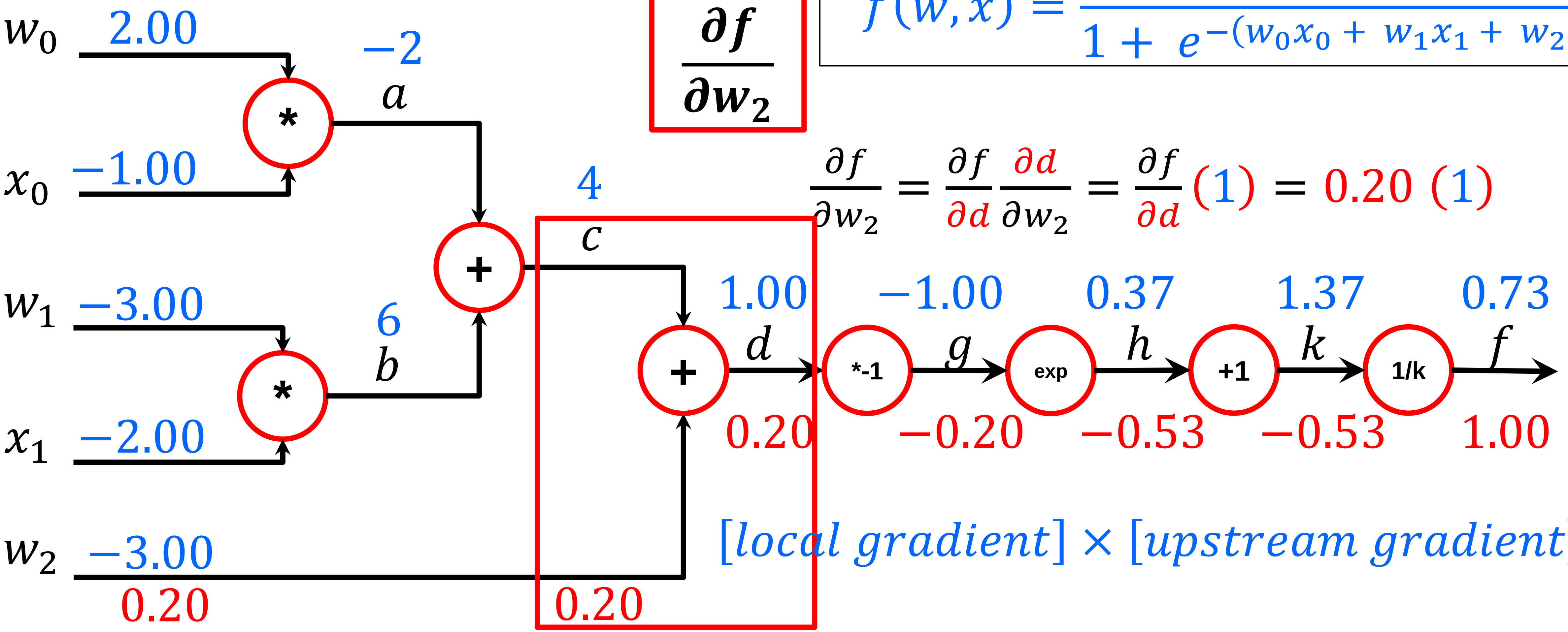
$$[\mathit{local\ gradient}] \times [\mathit{upstream\ gradient}]$$

Backpropagation – Another example

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

$$\boxed{\frac{\partial f}{\partial w_2}}$$

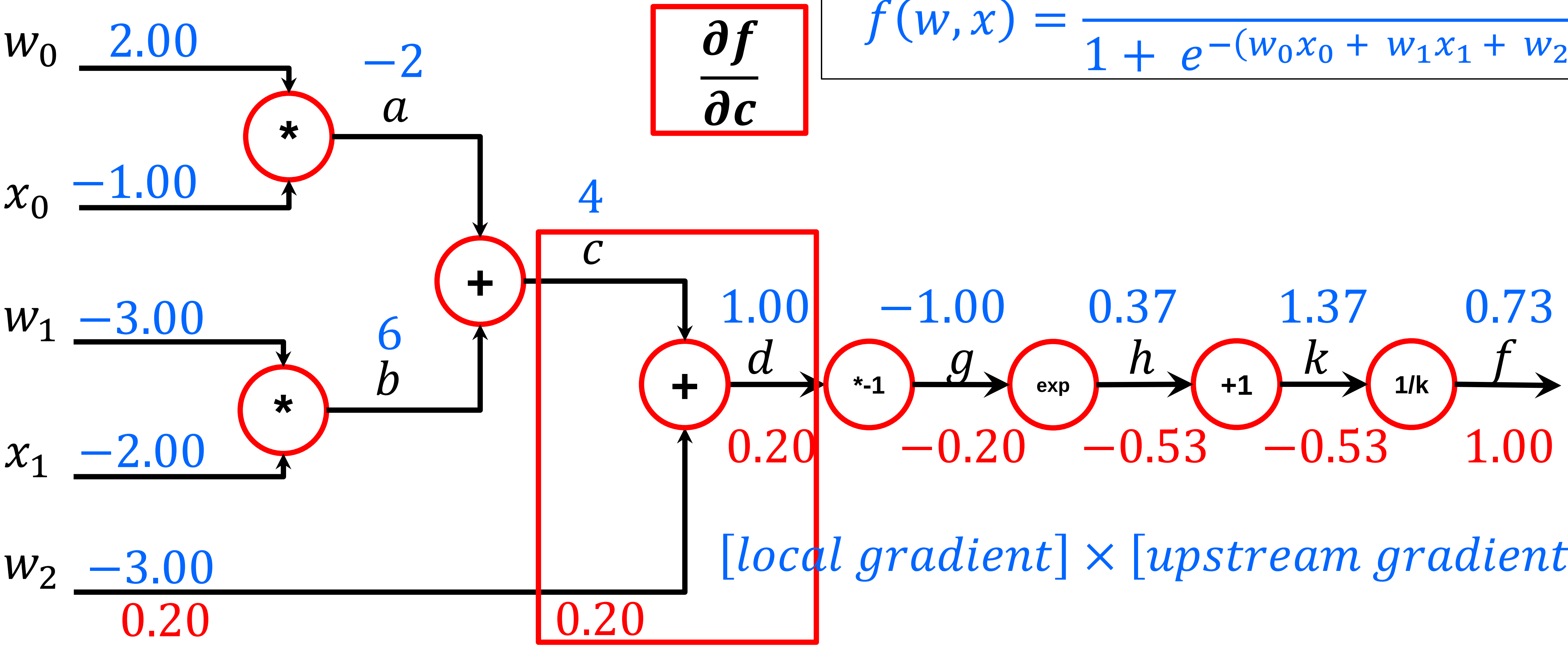
$$\frac{\partial f}{\partial w_2} = \frac{\partial f}{\partial d}\frac{\partial d}{\partial w_2} = \frac{\partial f}{\partial d}(1) = 0.20\ (1)$$



$[local\ gradient] \times [upstream\ gradient]$

Backpropagation – Another example

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



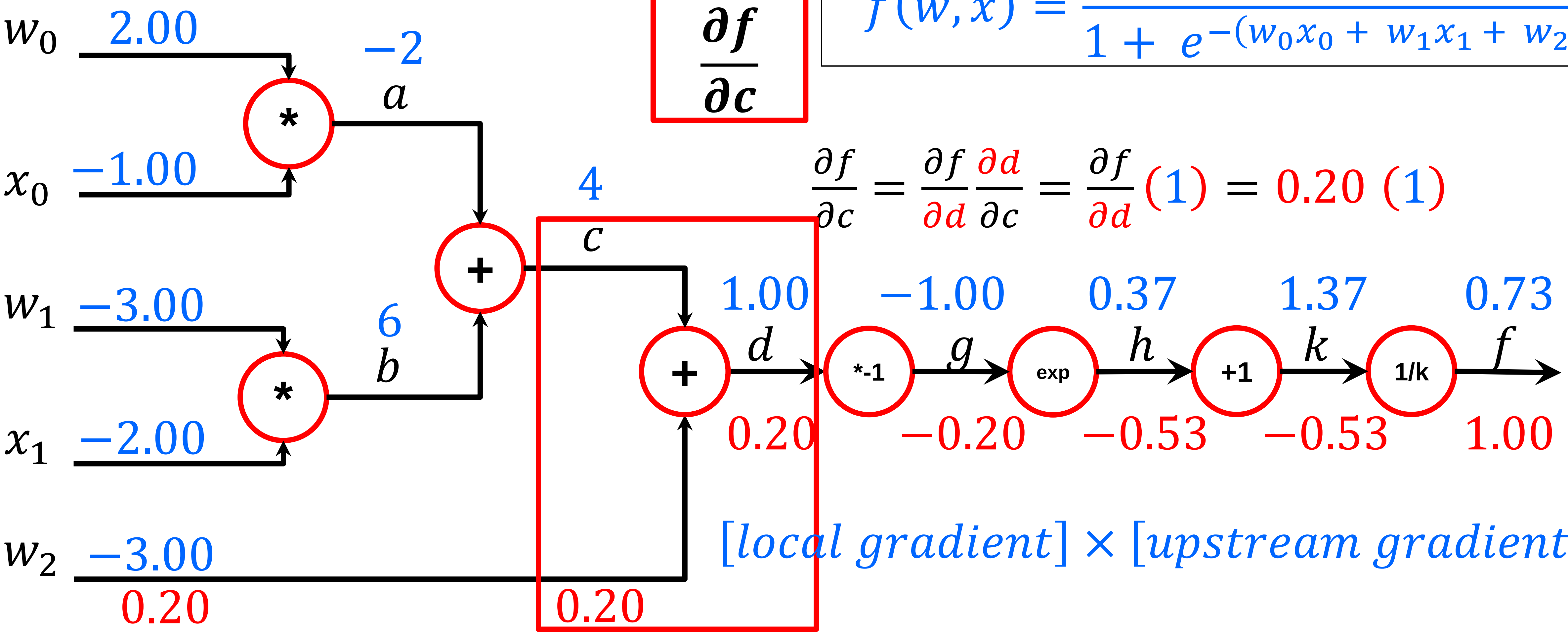
$[local\ gradient] \times [upstream\ gradient]$

Backpropagation – Another example

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

$$\frac{\partial f}{\partial c}$$

$$\frac{\partial f}{\partial c} = \frac{\partial f}{\partial d}\frac{\partial d}{\partial c} = \frac{\partial f}{\partial d}(1) = 0.20\ (1)$$



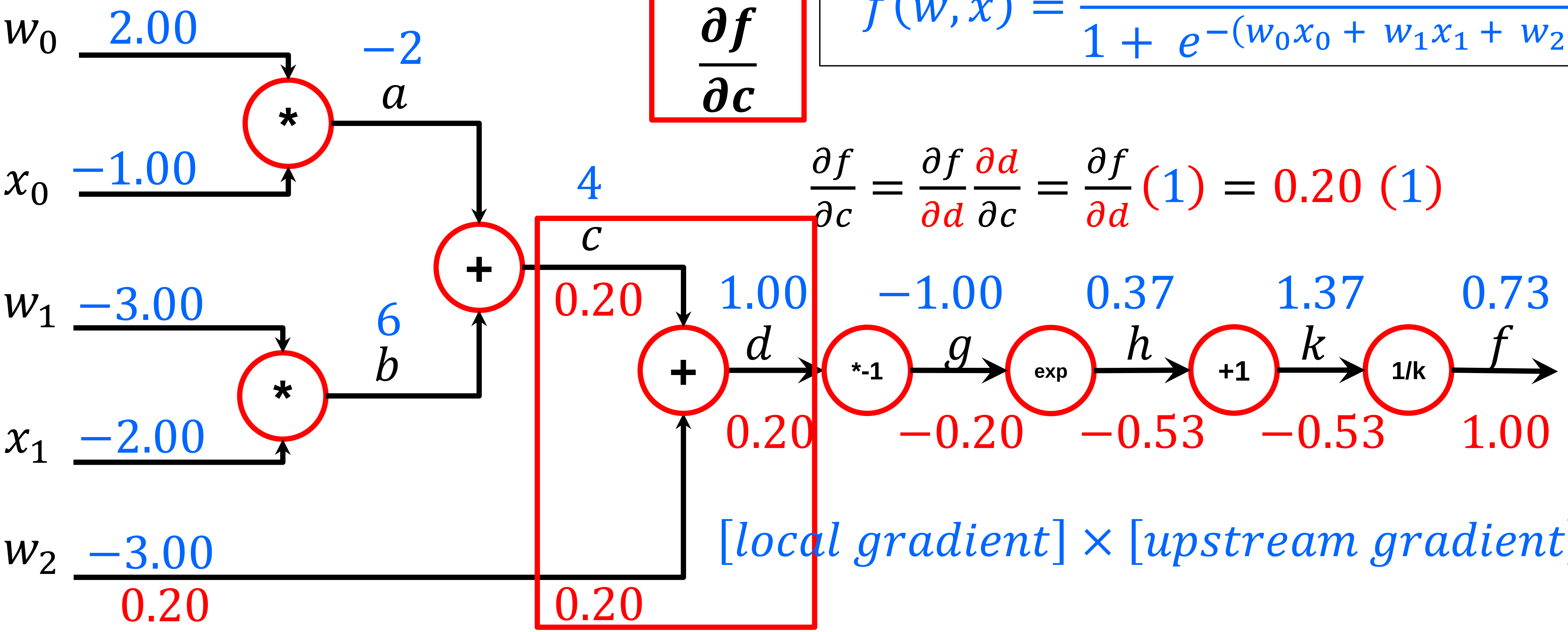
$[local\ gradient] \times [upstream\ gradient]$

Backpropagation – Another example

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

$$\boxed{\frac{\partial f}{\partial c}}$$

$$\frac{\partial f}{\partial c} = \frac{\partial f}{\partial d}\frac{\partial d}{\partial c} = \frac{\partial f}{\partial d}(1) = 0.20\ (1)$$

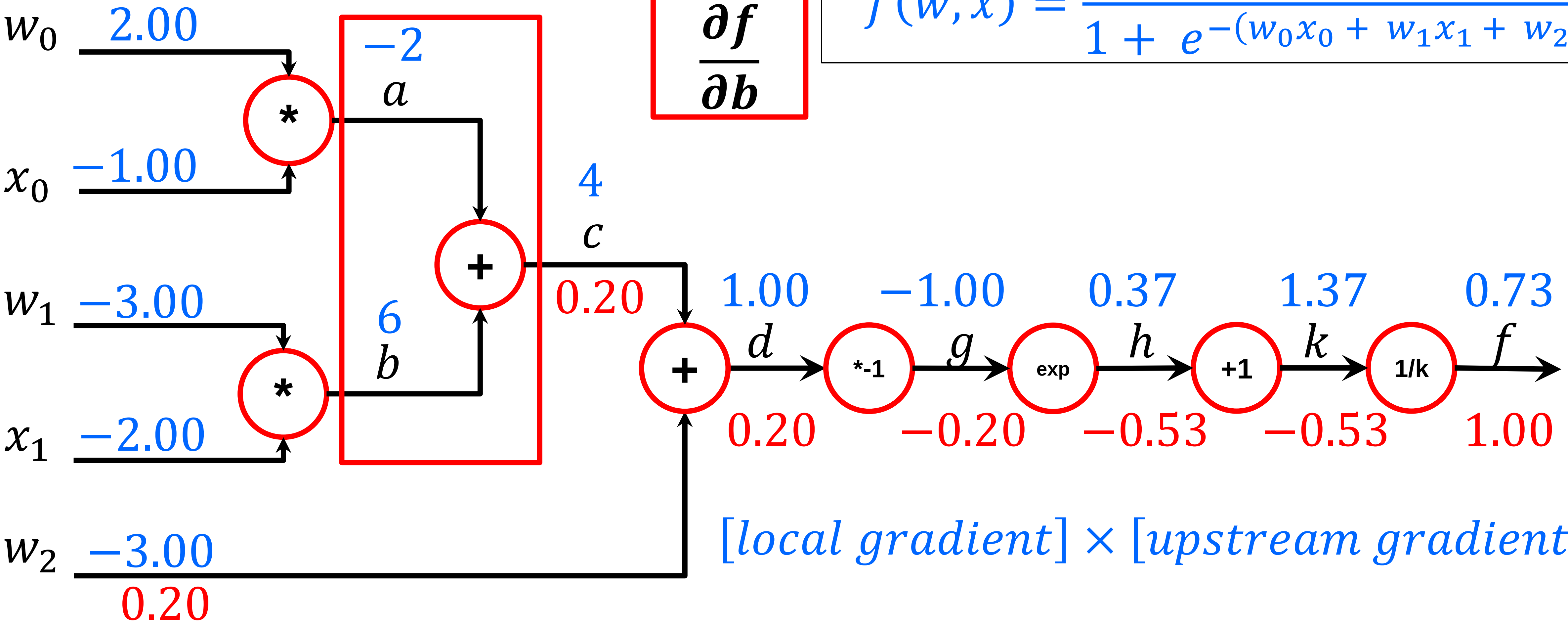


$$[local\ gradient] \times [upstream\ gradient]$$

Backpropagation – Another example

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

$$\boxed{\frac{\partial f}{\partial b}}$$



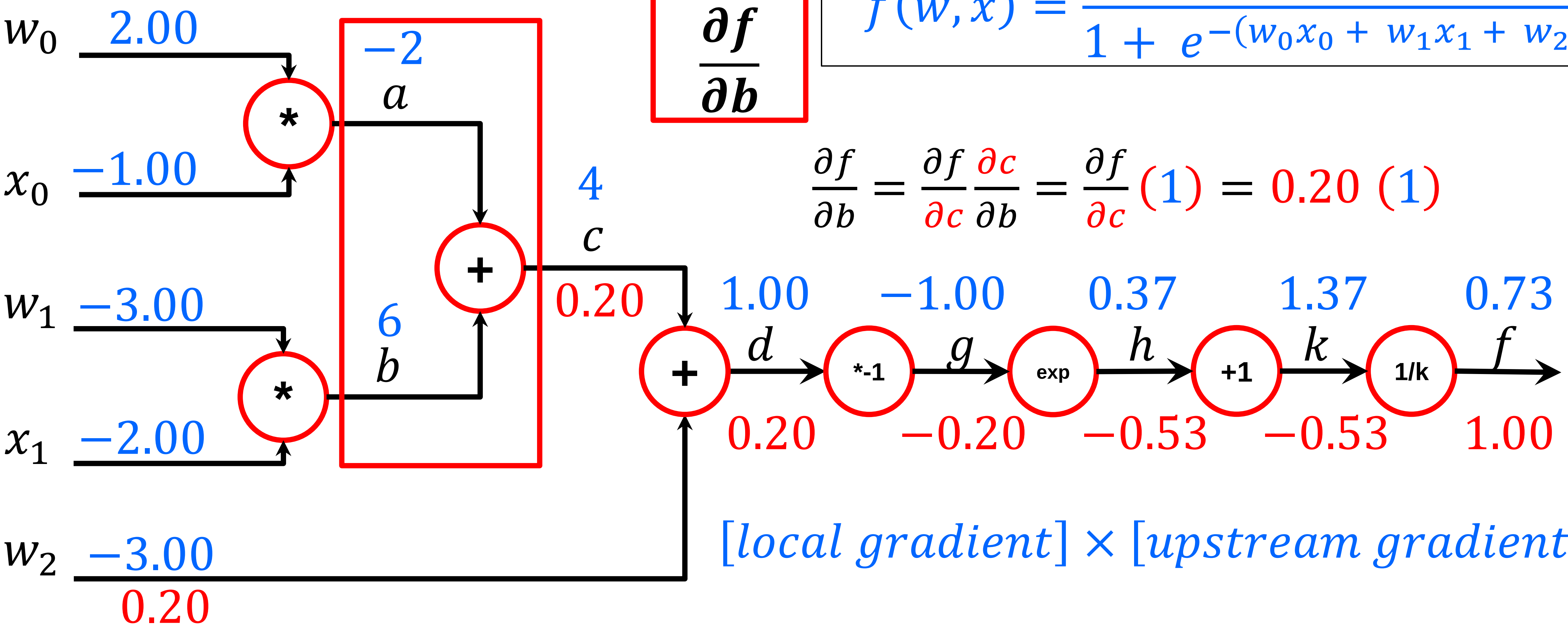
$[local\ gradient] \times [upstream\ gradient]$

Backpropagation – Another example

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

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

$$\frac{\partial f}{\partial b} = \frac{\partial f}{\partial c}\frac{\partial c}{\partial b} = \frac{\partial f}{\partial c}(1) = 0.20\,(1)$$



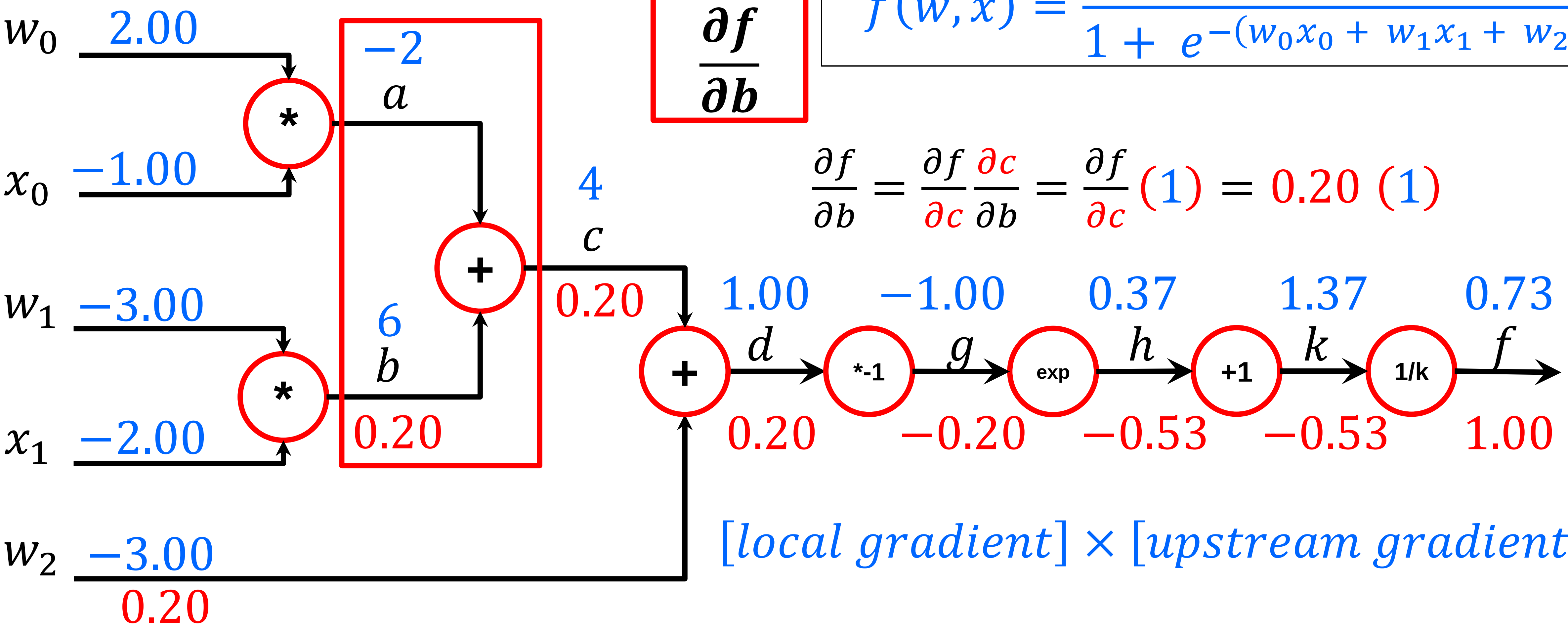
$[local\ gradient] \times [upstream\ gradient]$

Backpropagation – Another example

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

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

$$\frac{\partial f}{\partial b} = \frac{\partial f}{\partial c}\frac{\partial c}{\partial b} = \frac{\partial f}{\partial c}(1) = 0.20\ (1)$$

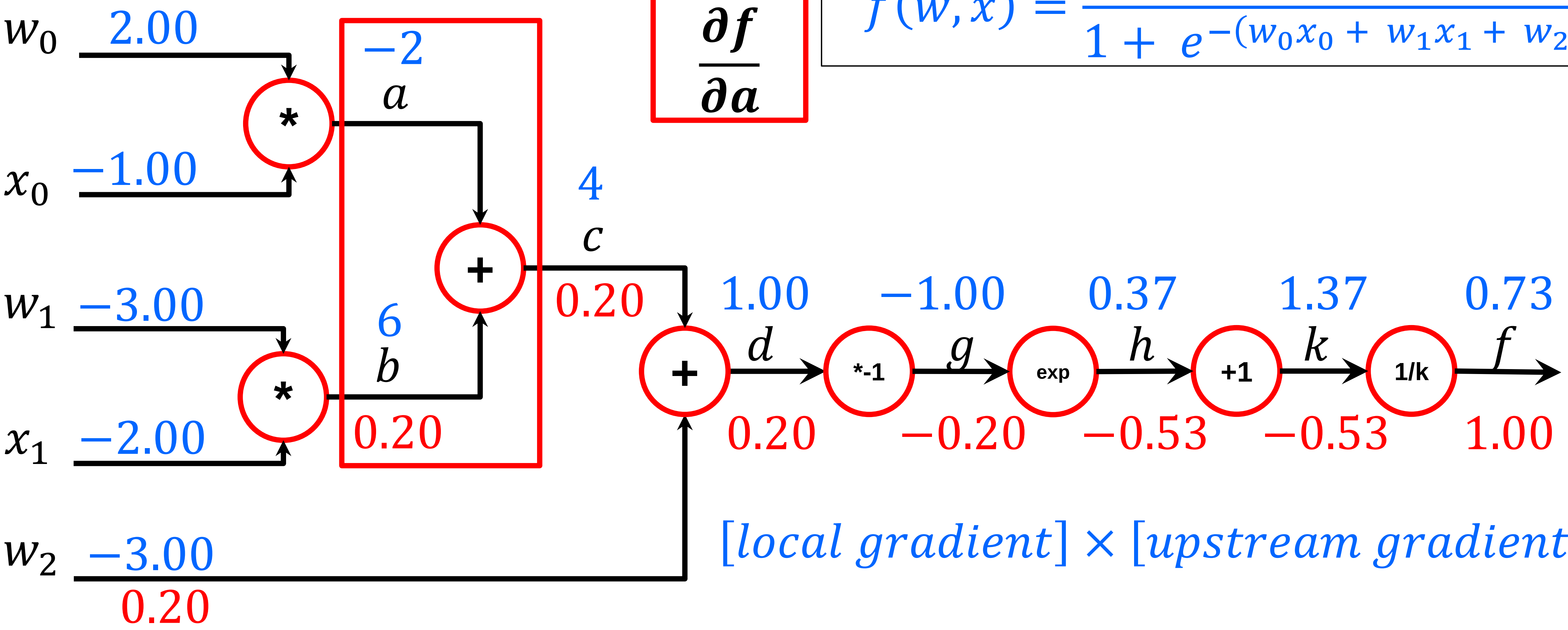


$$[local\ gradient] \times [upstream\ gradient]$$

Backpropagation – Another example

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

$$\frac{\partial f}{\partial a}$$

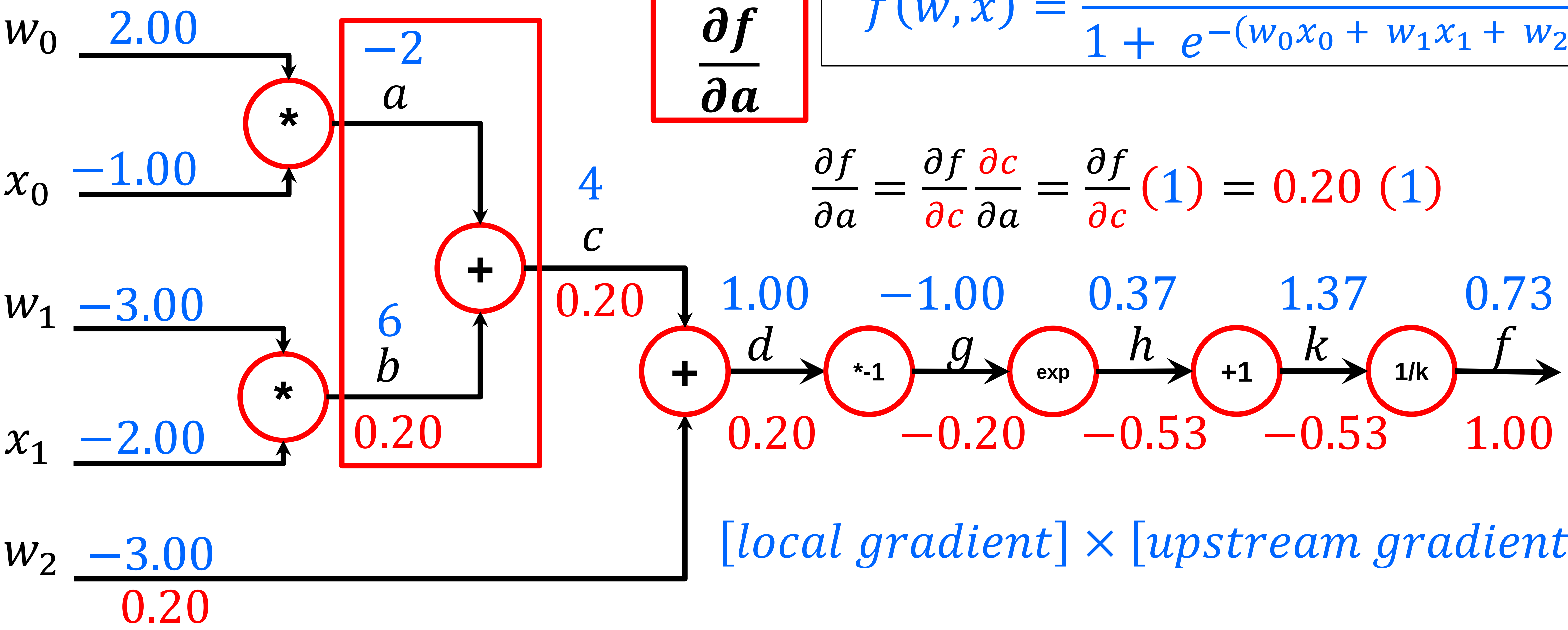


$$[local\ gradient] \times [upstream\ gradient]$$

Backpropagation – Another example

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

$$\boxed{\frac{\partial f}{\partial a}}$$



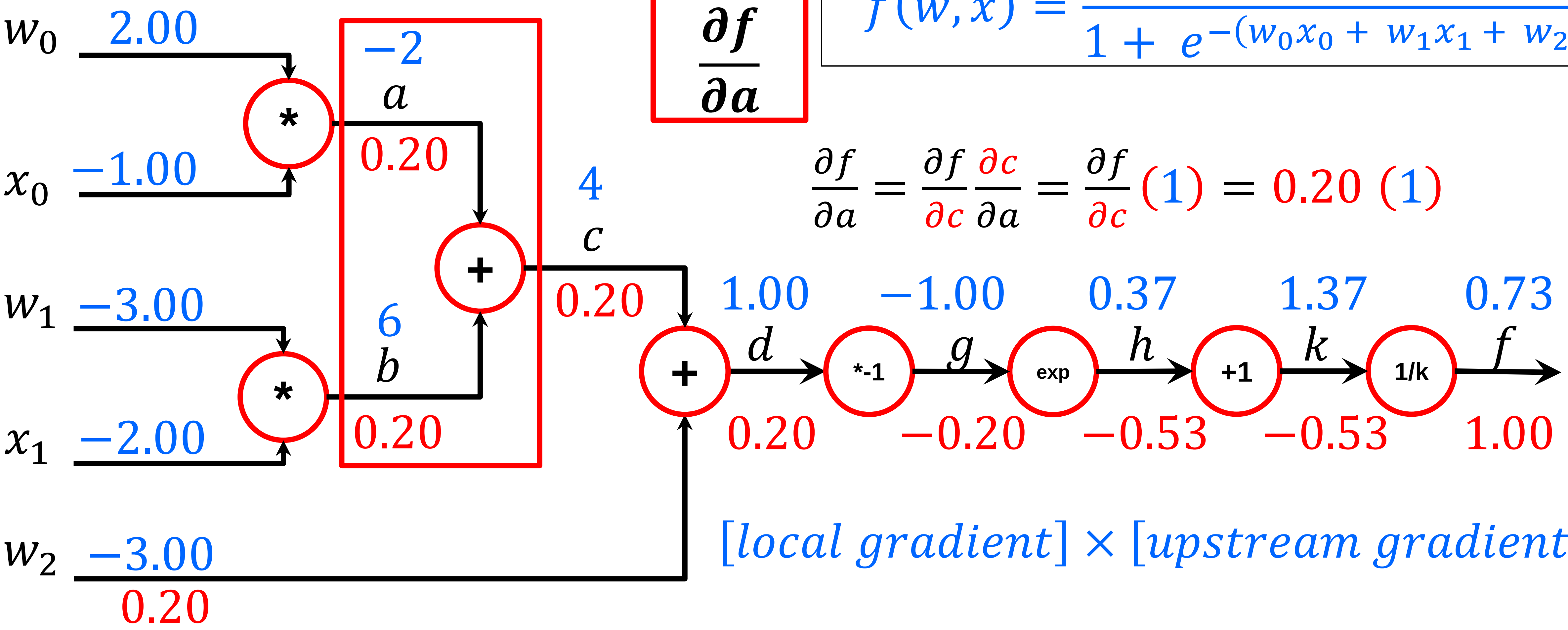
$$\frac{\partial f}{\partial a} = \frac{\partial f}{\partial c}\frac{\partial c}{\partial a} = \frac{\partial f}{\partial c}(1) = 0.20\ (1)$$

$$[local\ gradient] \times [upstream\ gradient]$$

Backpropagation – Another example

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

$$\boxed{\frac{\partial f}{\partial a}}$$



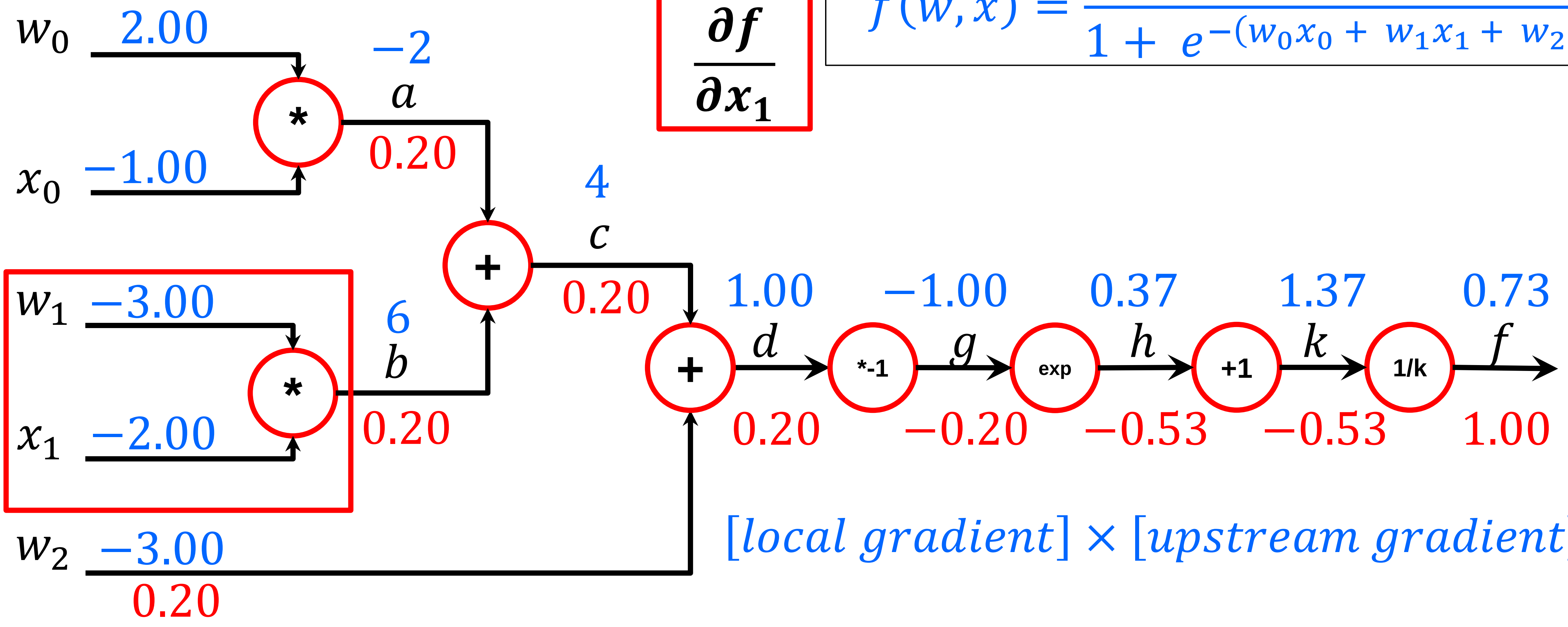
$$\frac{\partial f}{\partial a} = \frac{\partial f}{\partial c}\frac{\partial c}{\partial a} = \frac{\partial f}{\partial c}(1) = 0.20\ (1)$$

$$[local\ gradient] \times [upstream\ gradient]$$

Backpropagation – Another example

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

$$\frac{\partial f}{\partial x_1}$$

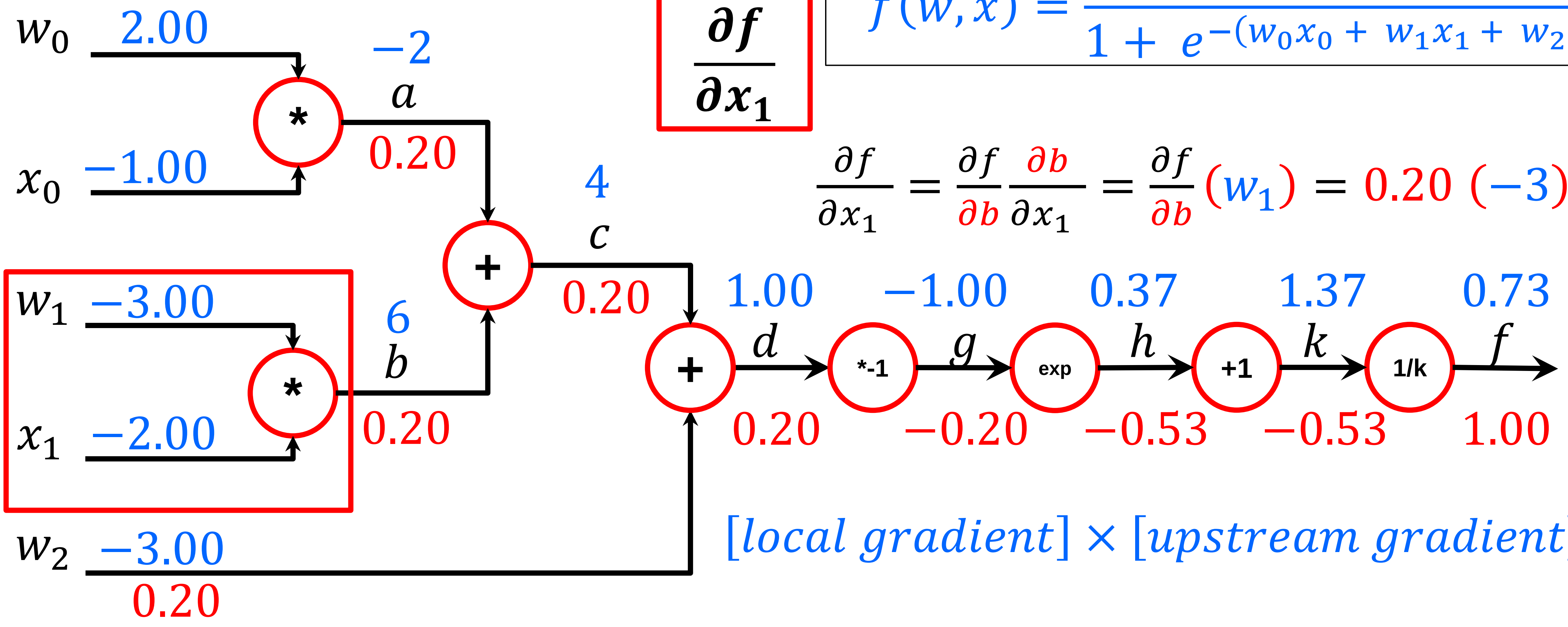


$[local\ gradient] \times [upstream\ gradient]$

Backpropagation – Another example

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

$$\boxed{\frac{\partial f}{\partial x_1}}$$



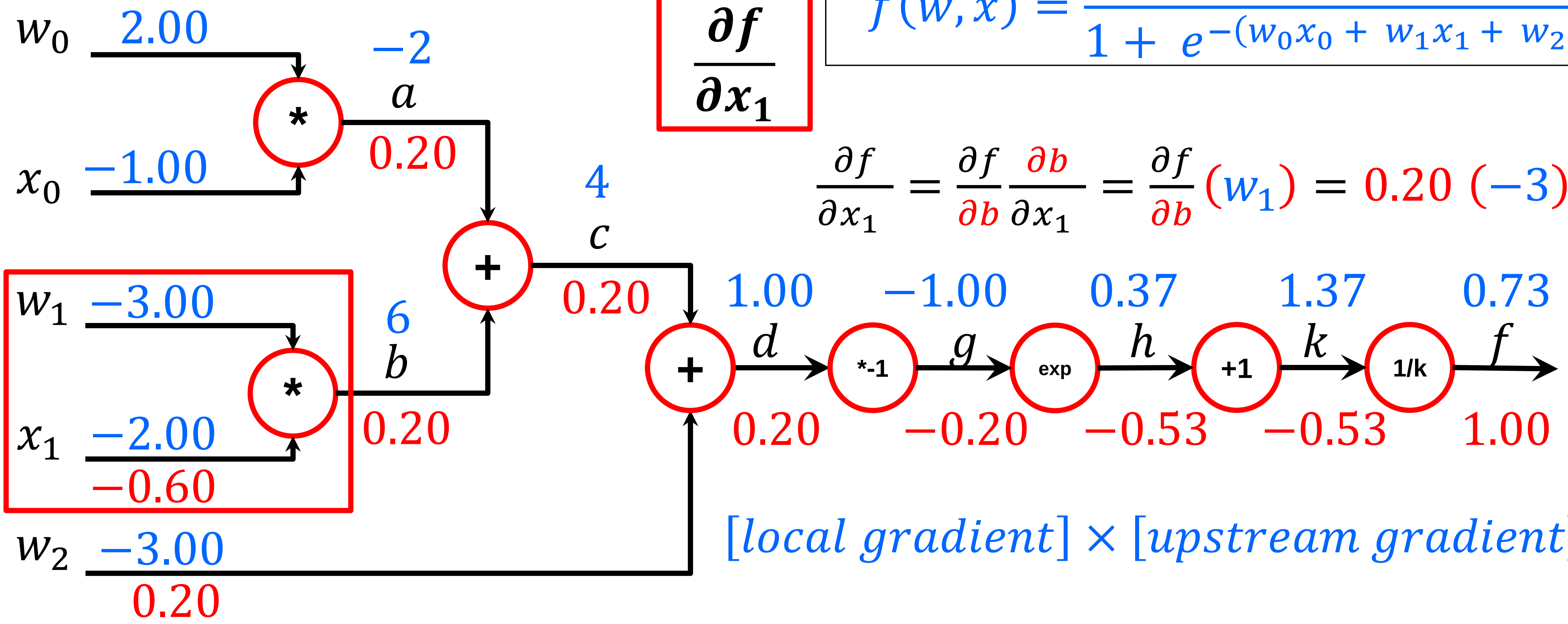
$$\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial b}\frac{\partial b}{\partial x_1} = \frac{\partial f}{\partial b}(w_1) = 0.20\,(-3)$$

$[local\ gradient] \times [upstream\ gradient]$

Backpropagation – Another example

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

$$\boxed{\frac{\partial f}{\partial x_1}}$$



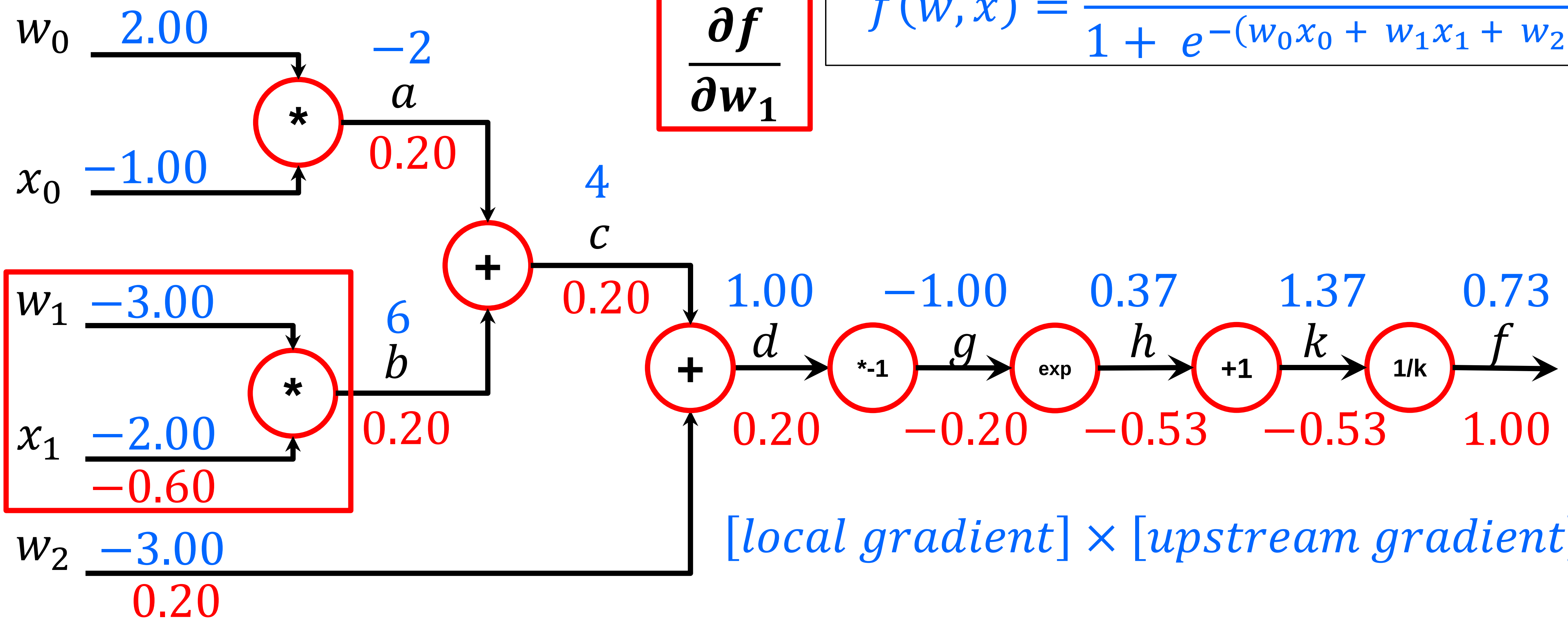
$$\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial b}\frac{\partial b}{\partial x_1} = \frac{\partial f}{\partial b}(w_1) = 0.20\,(-3)$$

$$[local\ gradient] \times [upstream\ gradient]$$

Backpropagation – Another example

$$\boxed{\frac{\partial f}{\partial w_1}}$$

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



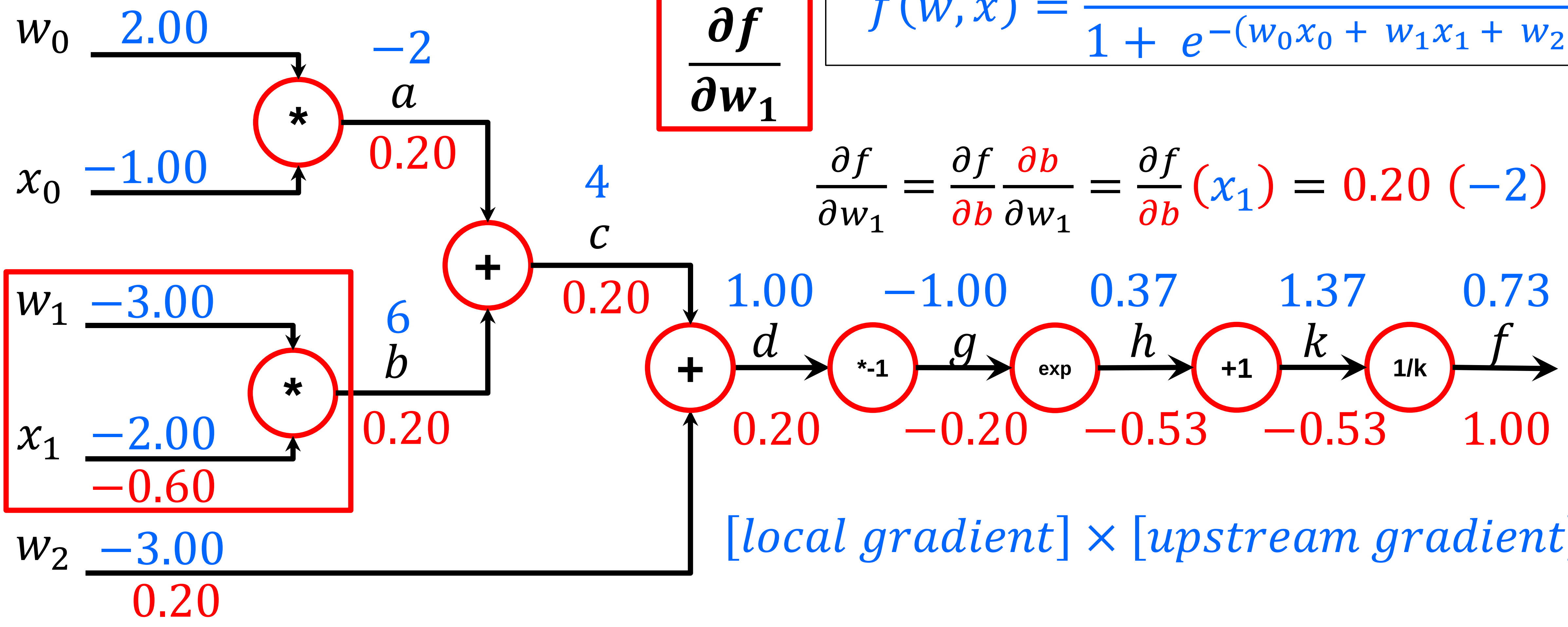
$[local\ gradient] \times [upstream\ gradient]$

Backpropagation – Another example

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

$$\boxed{\frac{\partial f}{\partial w_1}}$$

$$\frac{\partial f}{\partial w_1} = \frac{\partial f}{\partial b}\frac{\partial b}{\partial w_1} = \frac{\partial f}{\partial b}(x_1) = 0.20\ (-2)$$



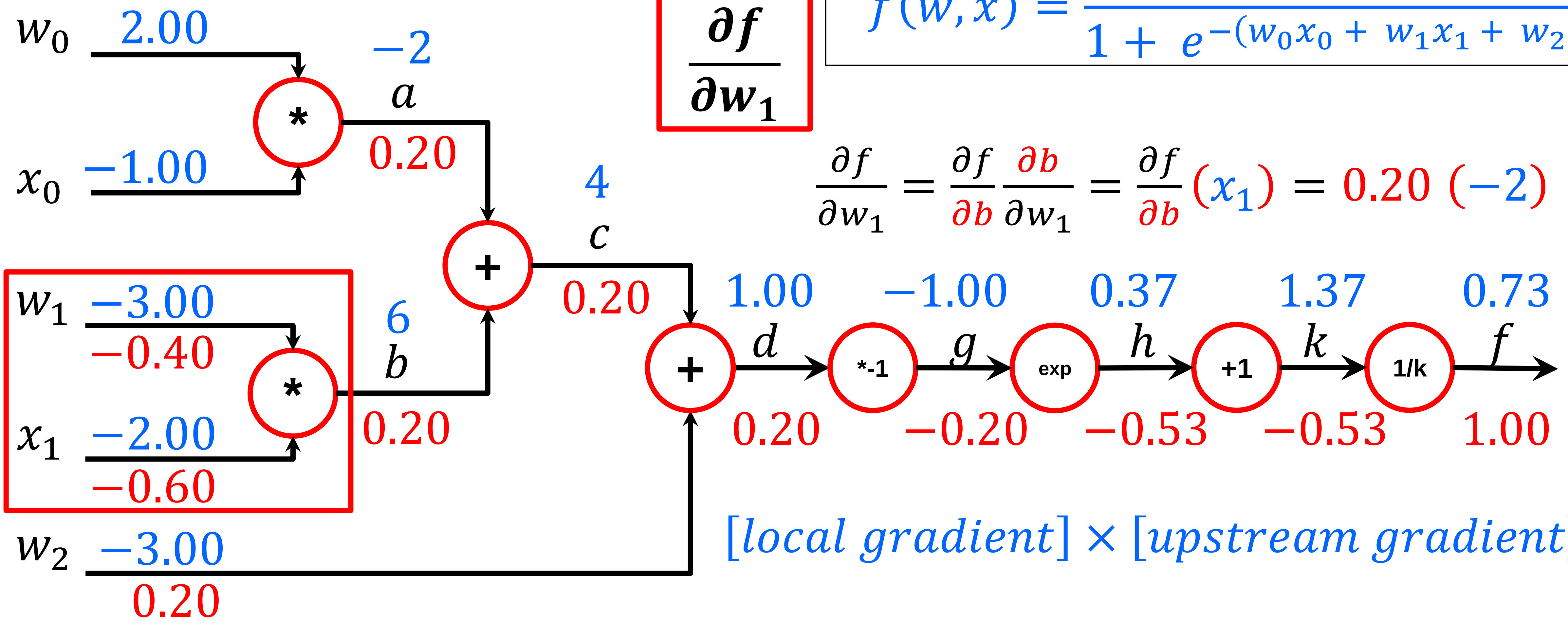
$$[local\ gradient] \times [upstream\ gradient]$$

Backpropagation – Another example

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

$$\boxed{\frac{\partial f}{\partial w_1}}$$

$$\frac{\partial f}{\partial w_1} = \frac{\partial f}{\partial b}\frac{\partial b}{\partial w_1} = \frac{\partial f}{\partial b}(x_1) = 0.20\ (-2)$$

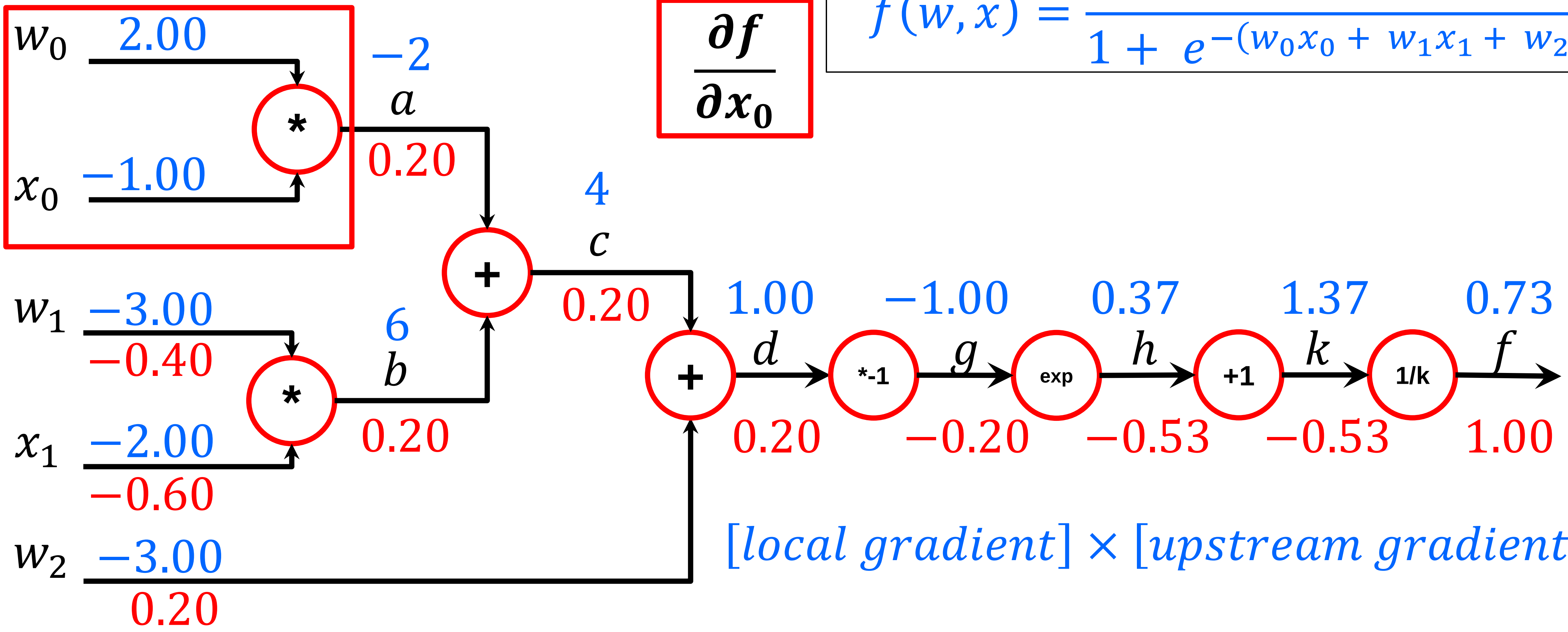


$$[local\ gradient] \times [upstream\ gradient]$$

Backpropagation – Another example

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

$$\frac{\partial f}{\partial x_0}$$

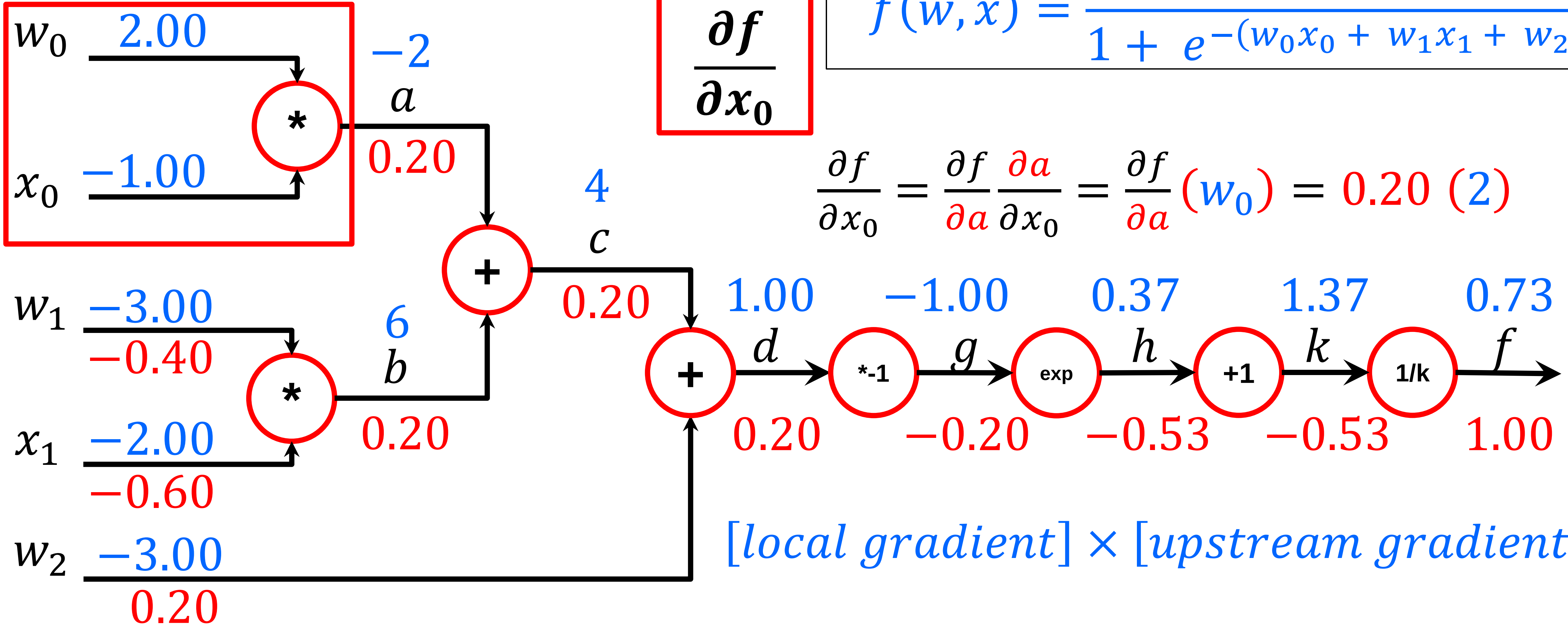


$[local\ gradient] \times [upstream\ gradient]$

Backpropagation – Another example

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

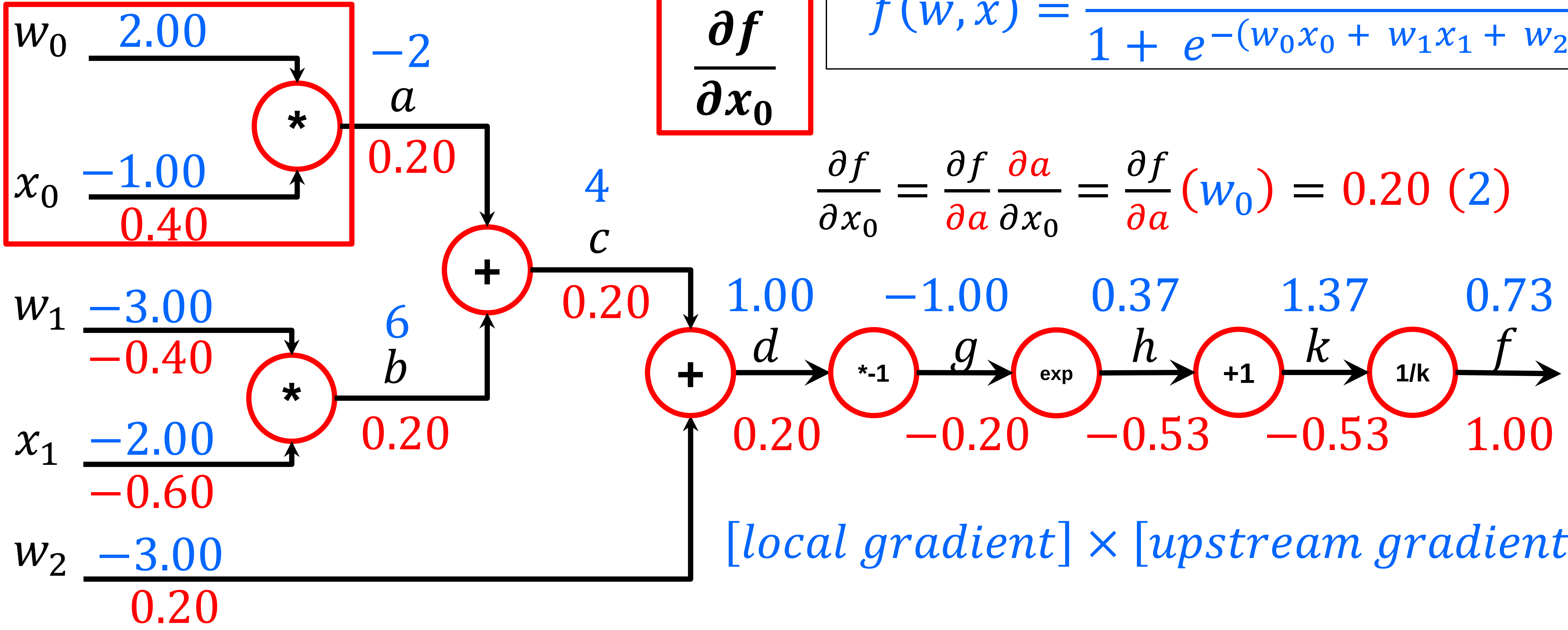
$$\frac{\partial f}{\partial x_0}$$



$$\frac{\partial f}{\partial x_0} = \frac{\partial f}{\partial a}\frac{\partial a}{\partial x_0} = \frac{\partial f}{\partial a}(w_0) = 0.20\ (2)$$

$[local\ gradient] \times [upstream\ gradient]$

Backpropagation – Another example



$$\frac{\partial f}{\partial x_0}$$

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

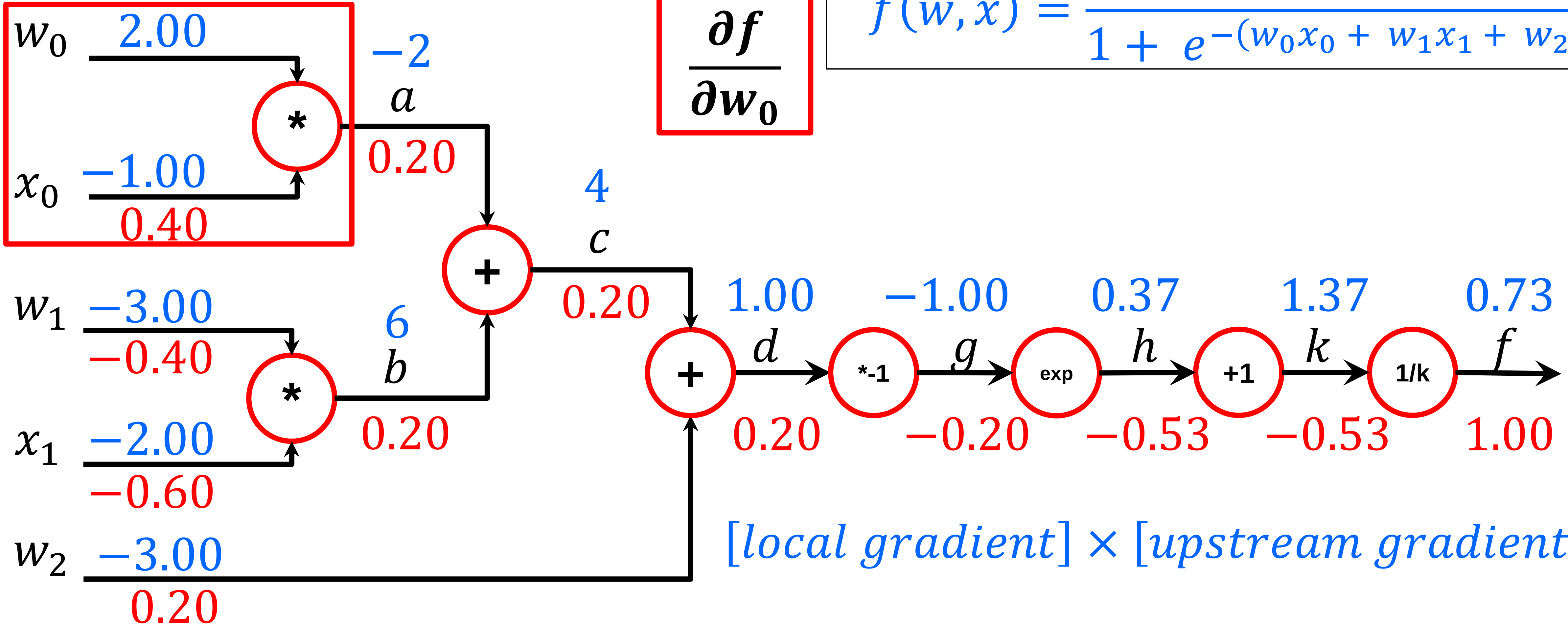
$$\frac{\partial f}{\partial x_0} = \frac{\partial f}{\partial a}\frac{\partial a}{\partial x_0} = \frac{\partial f}{\partial a}(w_0) = 0.20\ (2)$$

$[local\ gradient] \times [upstream\ gradient]$

Backpropagation – Another example

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

$$\boxed{\frac{\partial f}{\partial w_0}}$$

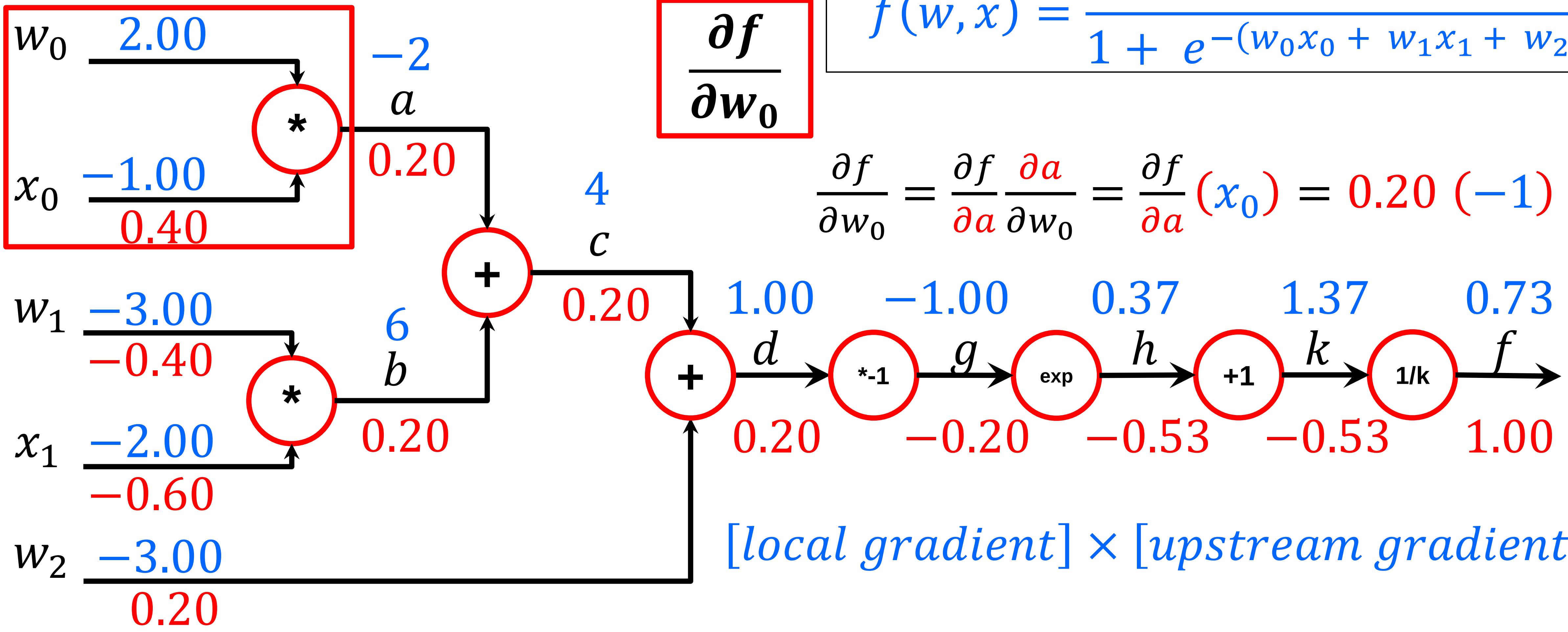


$[local\ gradient] \times [upstream\ gradient]$

Backpropagation – Another example

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

$$\frac{\partial f}{\partial w_0}$$



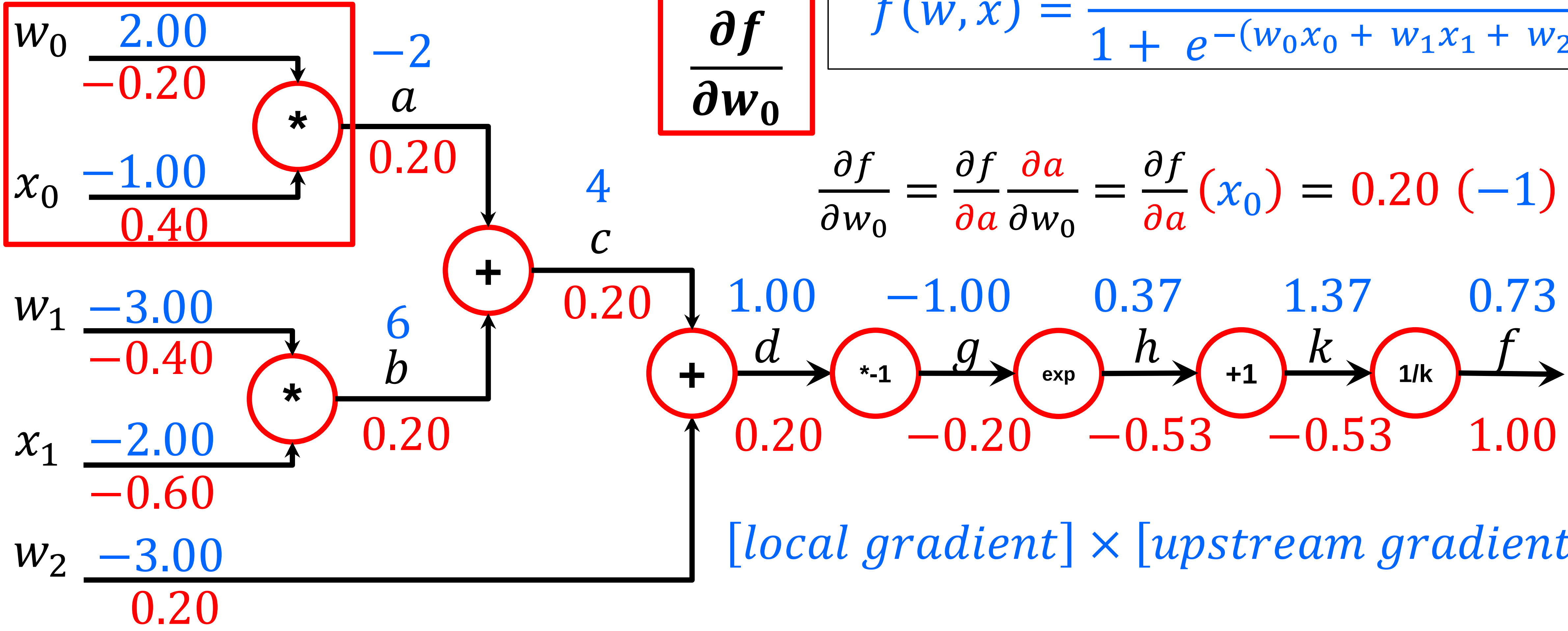
$$\frac{\partial f}{\partial w_0} = \frac{\partial f}{\partial a}\frac{\partial a}{\partial w_0} = \frac{\partial f}{\partial a}(x_0) = 0.20\ (-1)$$

$[local\ gradient] \times [upstream\ gradient]$

Backpropagation – Another example

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

$$\frac{\partial f}{\partial w_0}$$

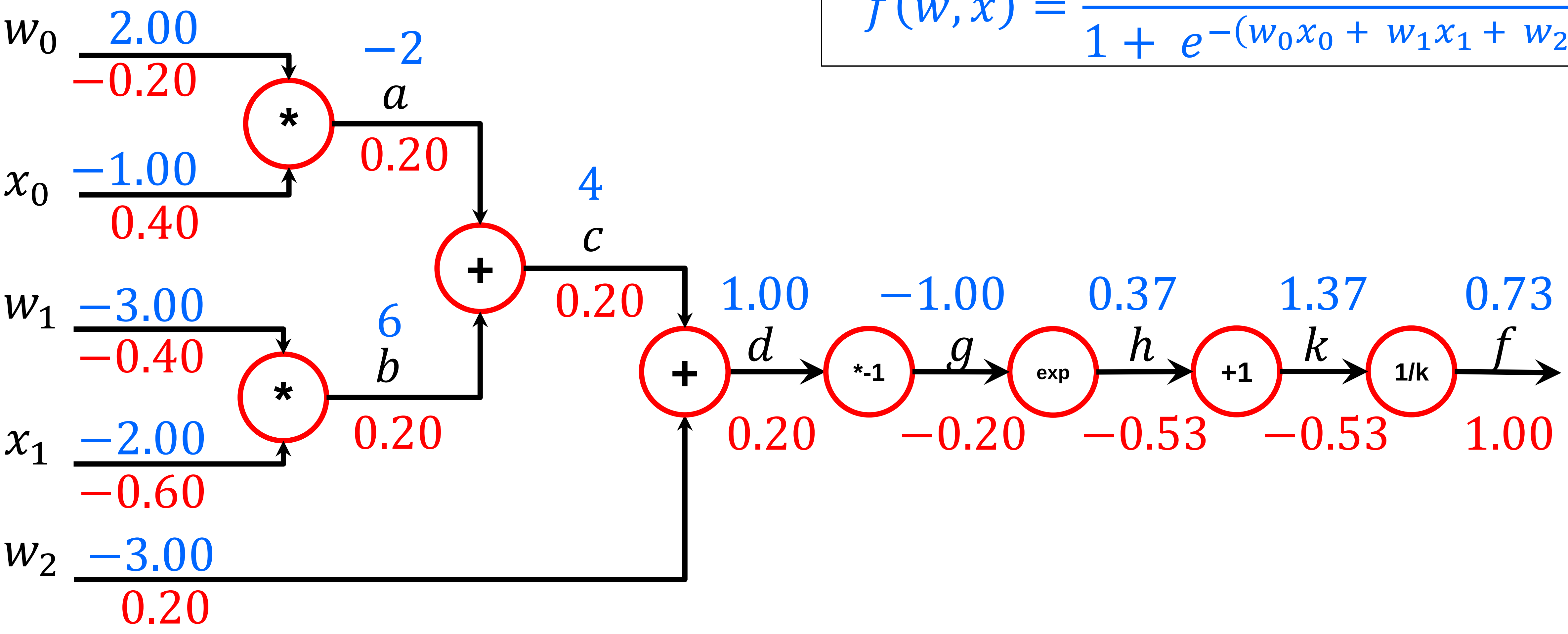


$$\frac{\partial f}{\partial w_0} = \frac{\partial f}{\partial a}\frac{\partial a}{\partial w_0} = \frac{\partial f}{\partial a}(x_0) = 0.20\ (-1)$$

$$[local\ gradient] \times [upstream\ gradient]$$

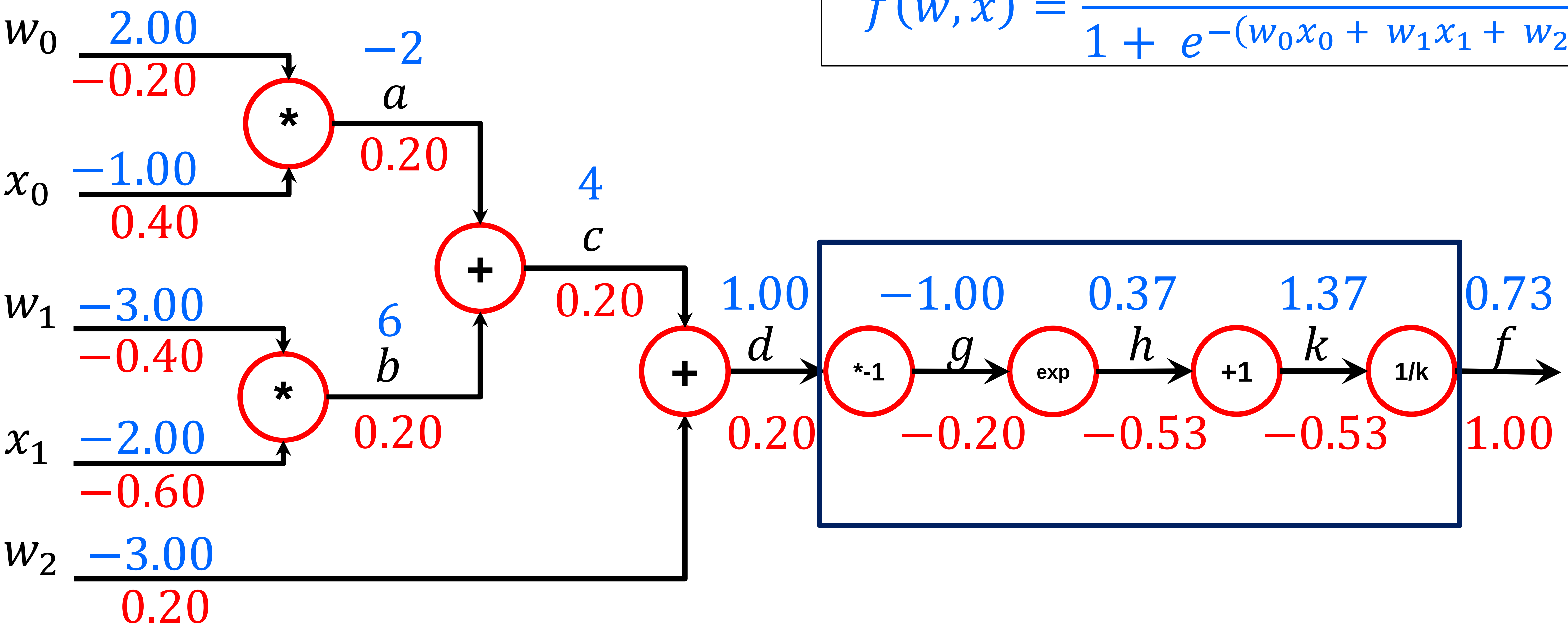
Backpropagation – Another example

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



Backpropagation – Another example

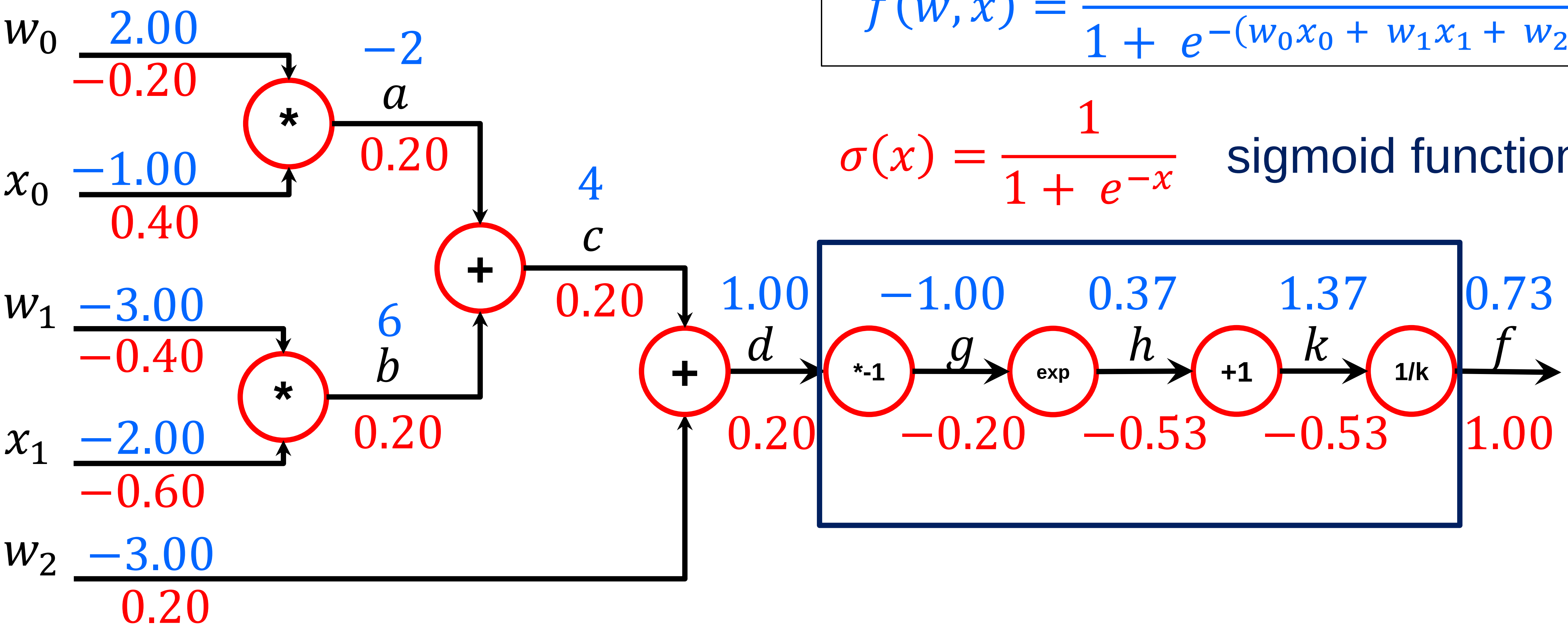
$$f(w,x) = \frac{1}{1 + e^{-(w_0x_0 + w_1x_1 + w_2)}}$$



Backpropagation – Another example

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

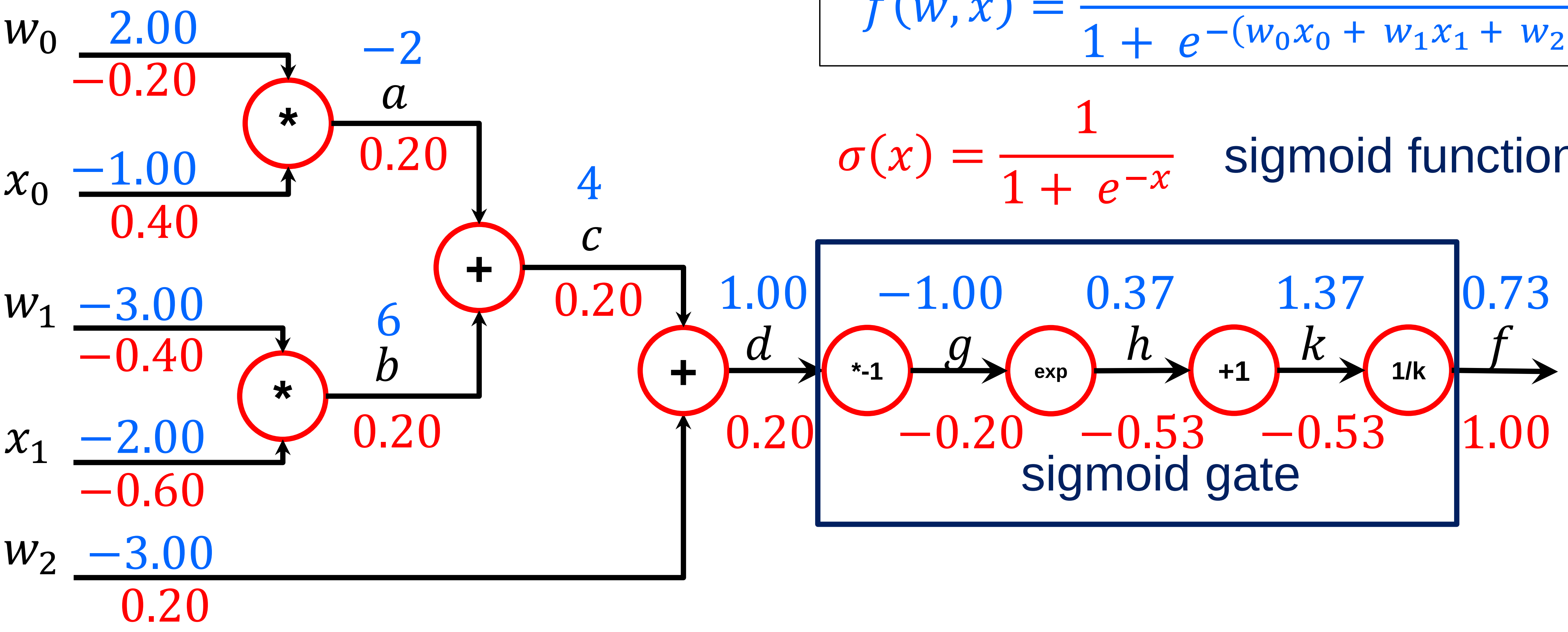
$$\sigma(x) = \frac{1}{1 + e^{-x}}$$ sigmoid function



Backpropagation – Another example

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

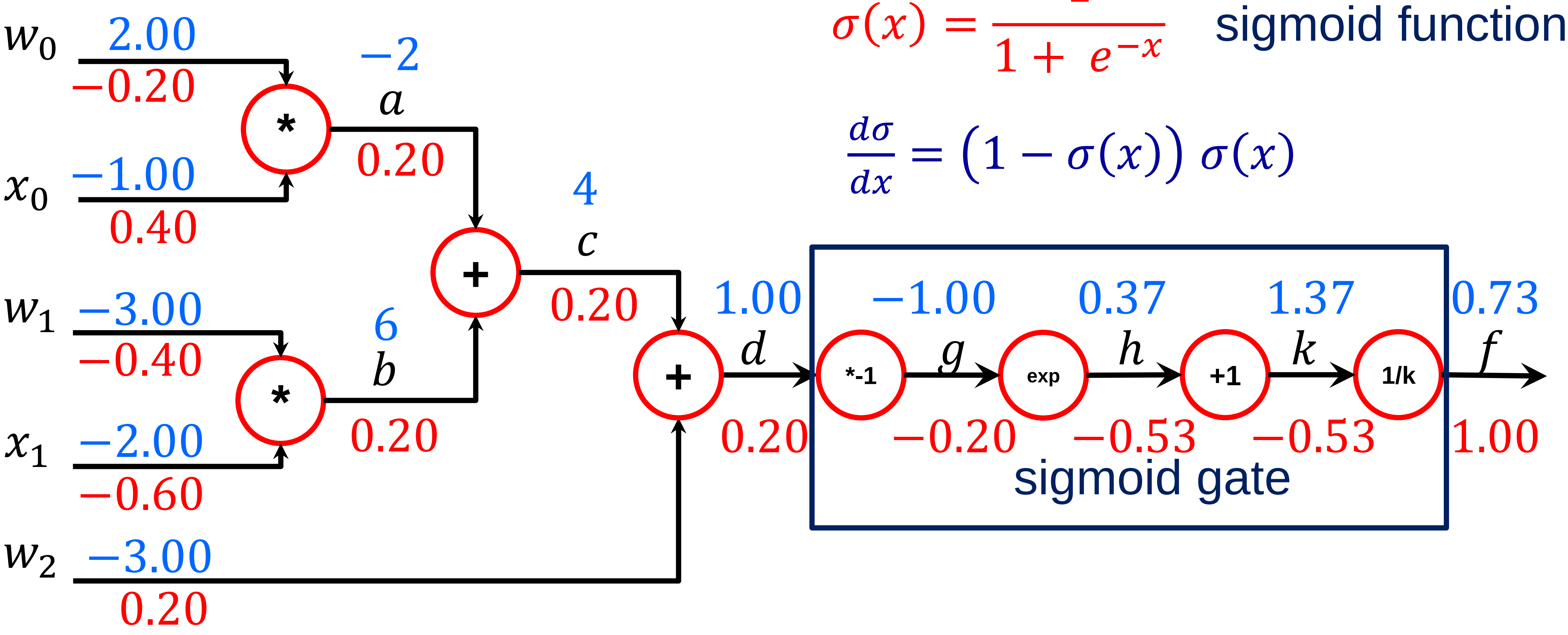
$$\sigma(x) = \frac{1}{1 + e^{-x}}$$ sigmoid function



Backpropagation – Another example

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

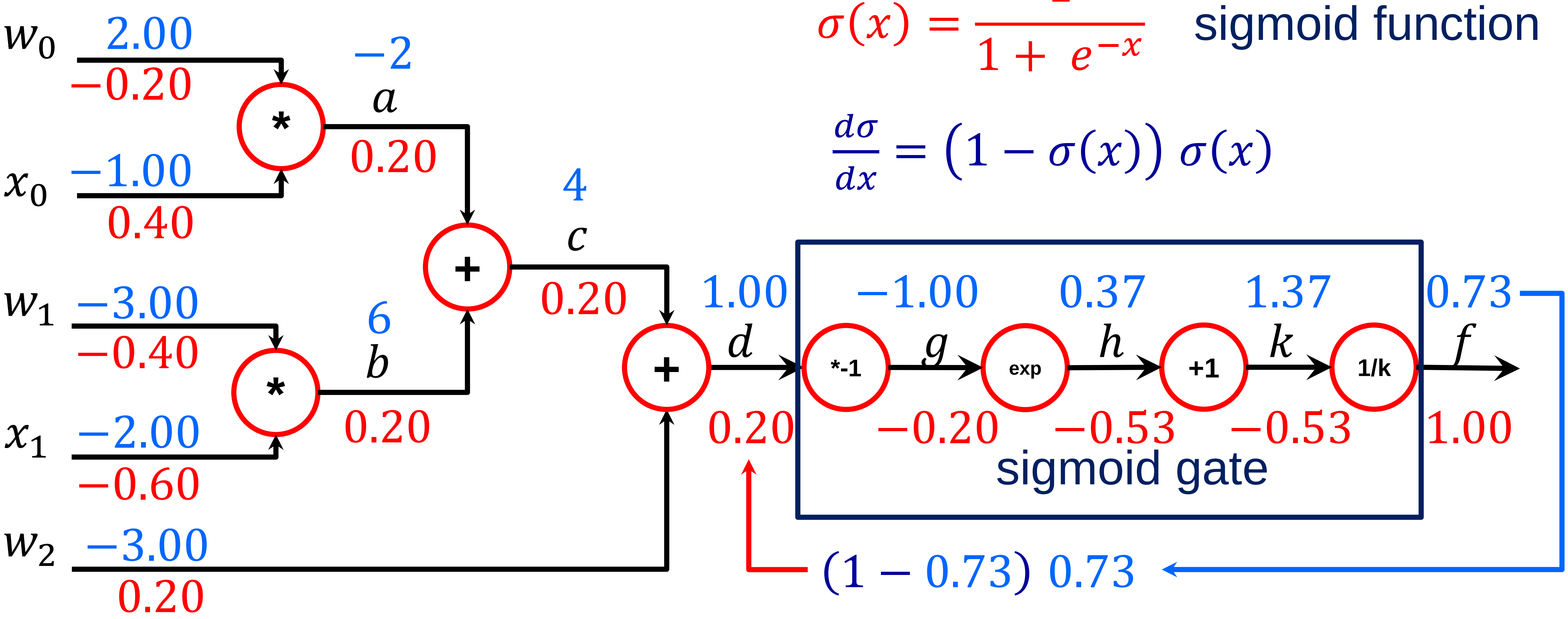
$$\frac{d\sigma}{dx} = \left(1-\sigma(x)\right)\sigma(x)$$



Backpropagation – Another example

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

$$\frac{d\sigma}{dx} = (1-\sigma(x))\,\sigma(x)$$



More example

BACKPROPAGATION

Backpropagation – More example

$$\boldsymbol{f(x, y, z, w) = \big((x \times y) + max(z, w)\big) \times 2}$$

E.g:

$x = 3$

$y = -4$

$z = 2$

$\mathrm{w} = -1$

Want: $\frac{\partial f}{\partial x}$ $\frac{\partial f}{\partial y}$ $\frac{\partial f}{\partial z}$ $\frac{\partial f}{\partial w}$

Backpropagation – More example

$$f(x,y,z,w) = \big((x \times y) + max(z,w)\big) \times 2$$

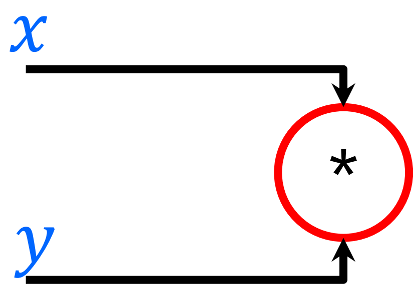
E.g:

$x = 3$

$y = -4$

$z = 2$

$\mathrm{w} = -1$



Want: $\frac{\partial f}{\partial x}$ $\frac{\partial f}{\partial y}$ $\frac{\partial f}{\partial z}$ $\frac{\partial f}{\partial w}$

Backpropagation – More example

$$f(x,y,z,w) = \big((x \times y) + max(z,w)\big) \times 2$$

E.g:

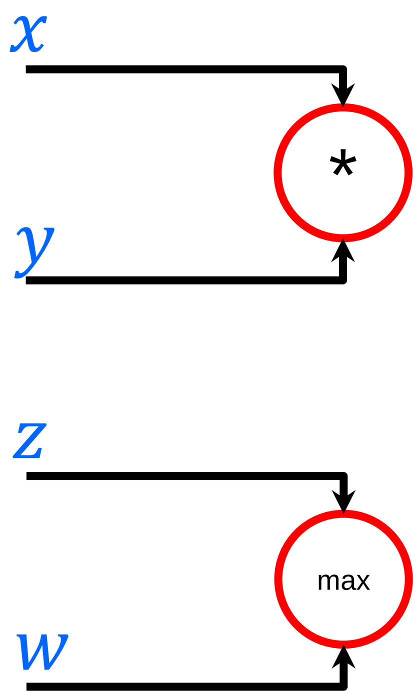
$x = 3$

$y = -4$

$z = 2$

$\mathrm{w} = -1$

Want: $\frac{\partial f}{\partial x}$ $\frac{\partial f}{\partial y}$ $\frac{\partial f}{\partial z}$ $\frac{\partial f}{\partial w}$



Backpropagation – More example

$$\boldsymbol{f(x,y,z,w) = \big((x \times y) + max(z,w)\big) \times 2}$$

E.g:

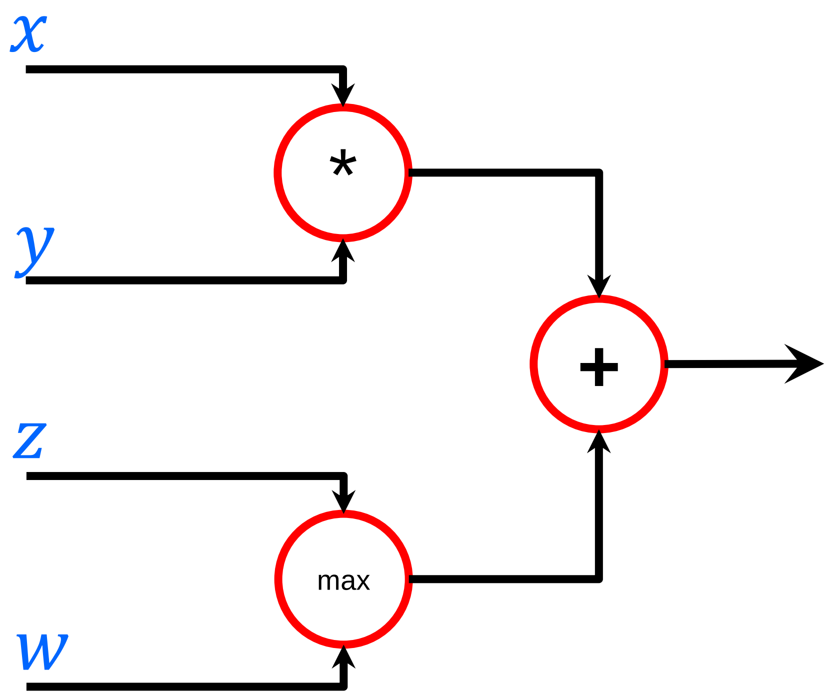
$x = 3$

$y = -4$

$z = 2$

$\mathrm{w} = -1$

Want: $\frac{\partial f}{\partial x}$ $\frac{\partial f}{\partial y}$ $\frac{\partial f}{\partial z}$ $\frac{\partial f}{\partial w}$



Backpropagation – More example

$$f(x, y, z, w) = \big((x \times y) + max(z, w)\big) \times 2$$

E.g:

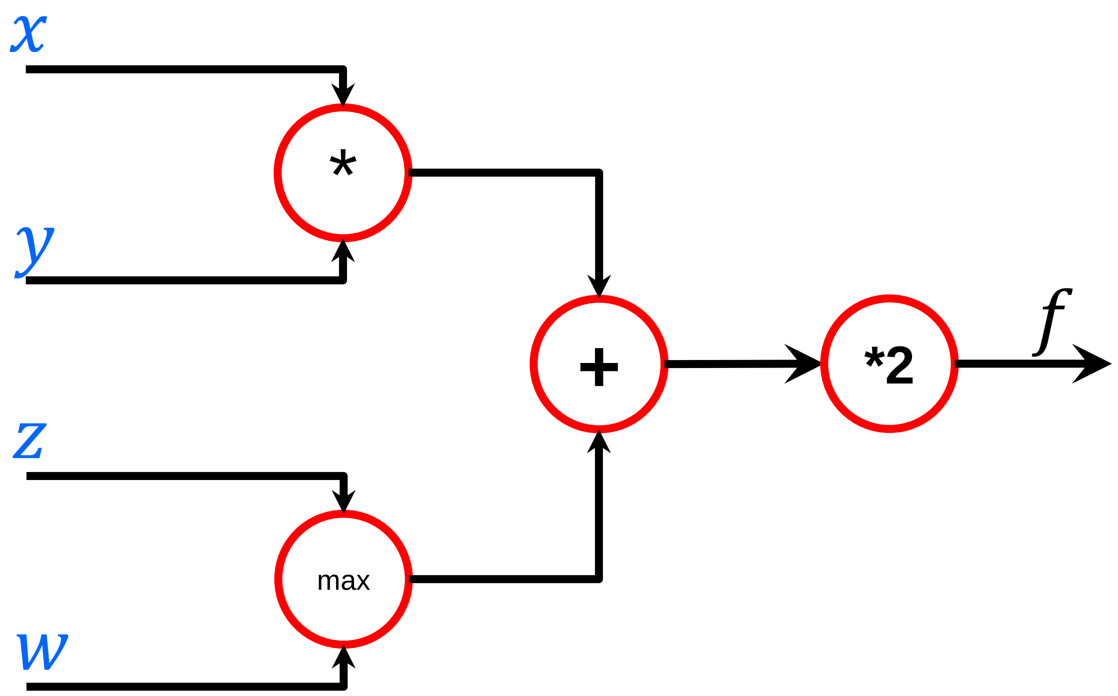
$x = 3$

$y = -4$

$z = 2$

$\mathrm{w} = -1$

Want: $\frac{\partial f}{\partial x}$ $\frac{\partial f}{\partial y}$ $\frac{\partial f}{\partial z}$ $\frac{\partial f}{\partial w}$



Backpropagation – More example

$$f(x, y, z, w) = \left((x \times y) + max(z, w)\right) \times 2$$

E.g:

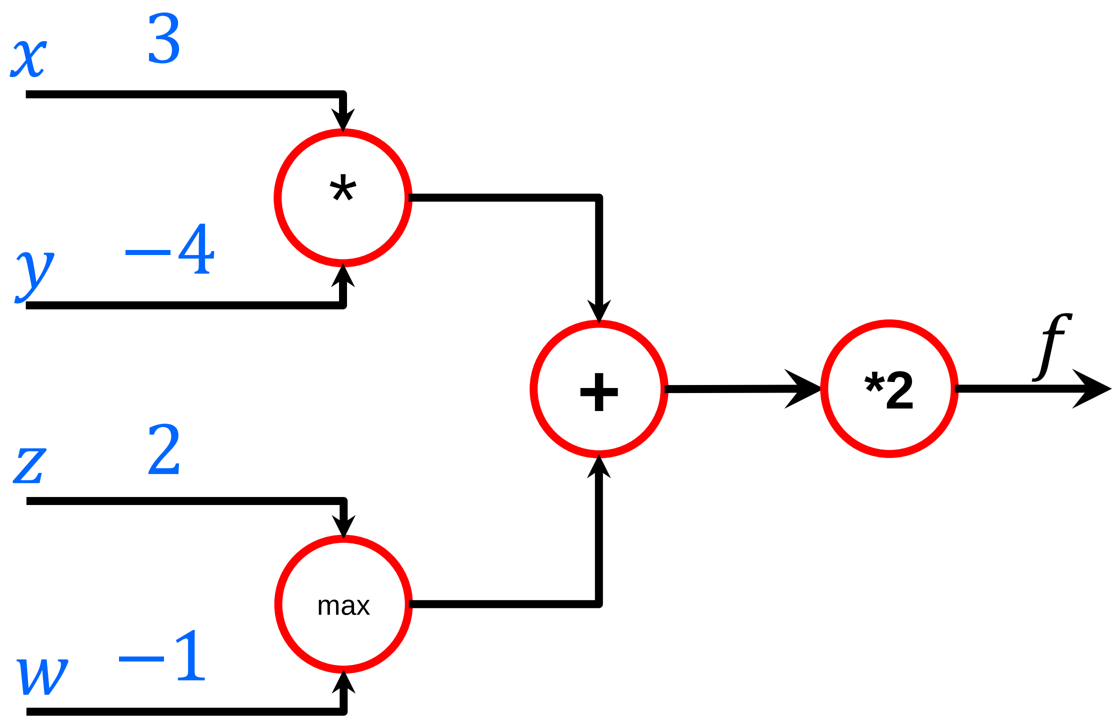
$x = 3$

$y = -4$

$z = 2$

$w = -1$

Want: $\frac{\partial f}{\partial x}$ $\frac{\partial f}{\partial y}$ $\frac{\partial f}{\partial z}$ $\frac{\partial f}{\partial w}$



Backpropagation – More example

$$f(x, y, z, w) = \big((x \times y) + max(z, w)\big) \times 2$$

E.g:

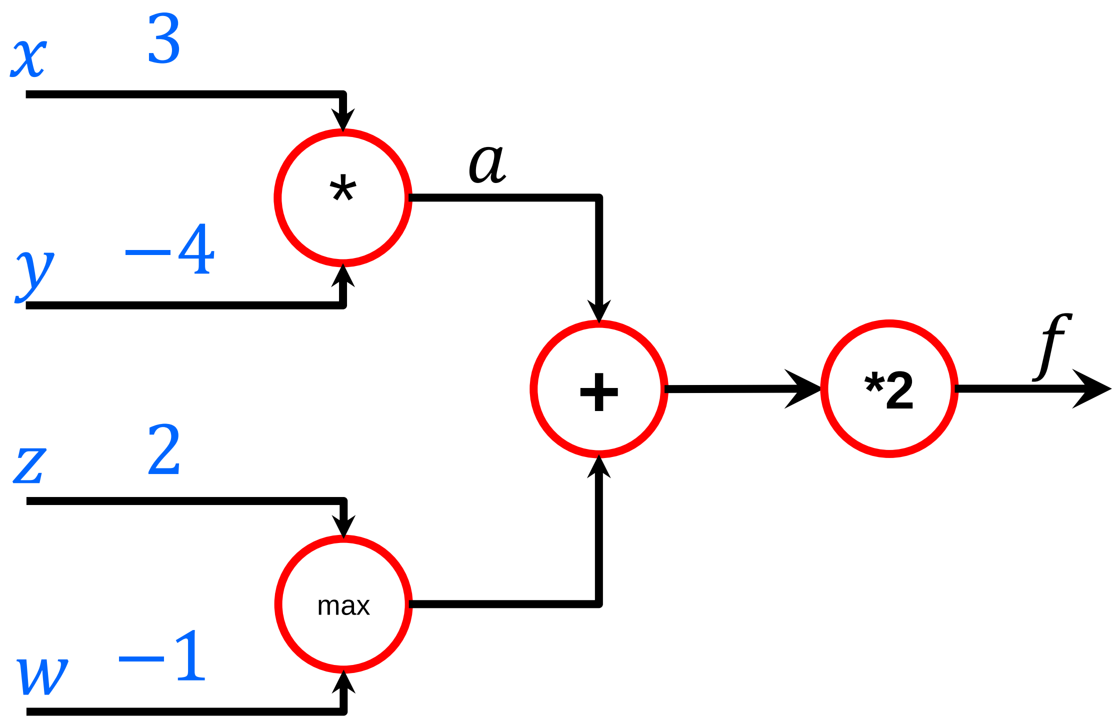
$x = 3$

$y = -4$

$z = 2$

$w = -1$

Want: $\frac{\partial f}{\partial x}$ $\frac{\partial f}{\partial y}$ $\frac{\partial f}{\partial z}$ $\frac{\partial f}{\partial w}$



Backpropagation – More example

$$f(x, y, z, w) = \left((x \times y) + max(z, w)\right) \times 2$$

E.g:

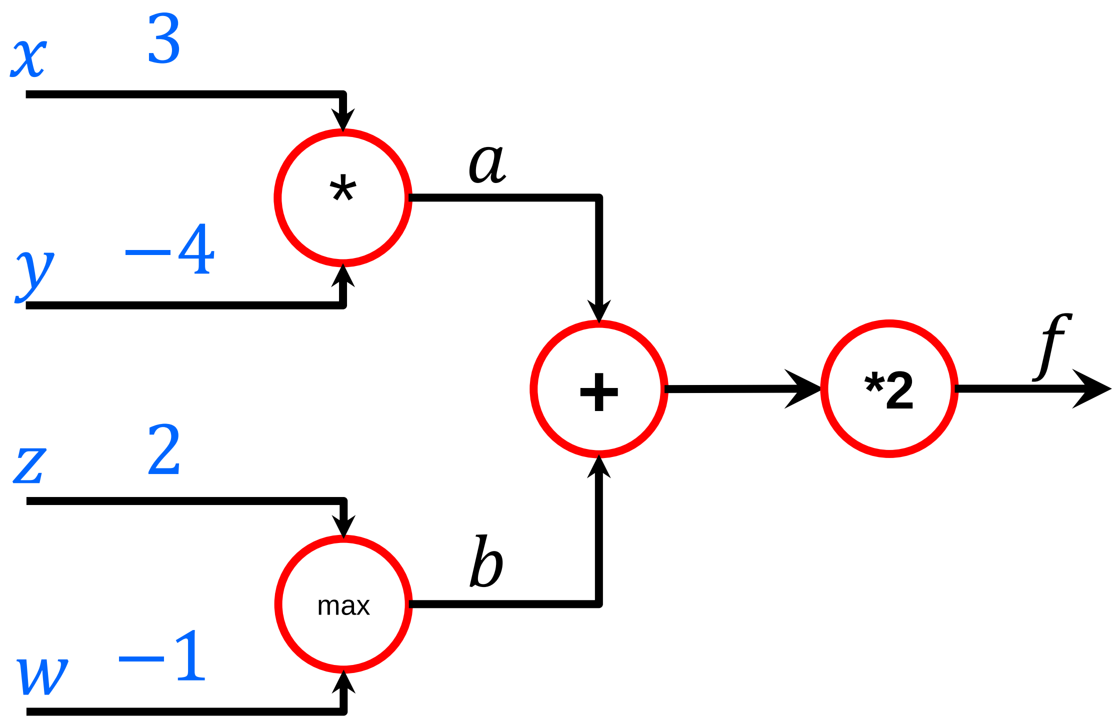
$x = 3$

$y = -4$

$z = 2$

$w = -1$

Want: $\frac{\partial f}{\partial x}$ $\frac{\partial f}{\partial y}$ $\frac{\partial f}{\partial z}$ $\frac{\partial f}{\partial w}$



Backpropagation – More example

$$f(x,y,z,w) = \left((x \times y) + max(z,w)\right) \times 2$$

E.g:

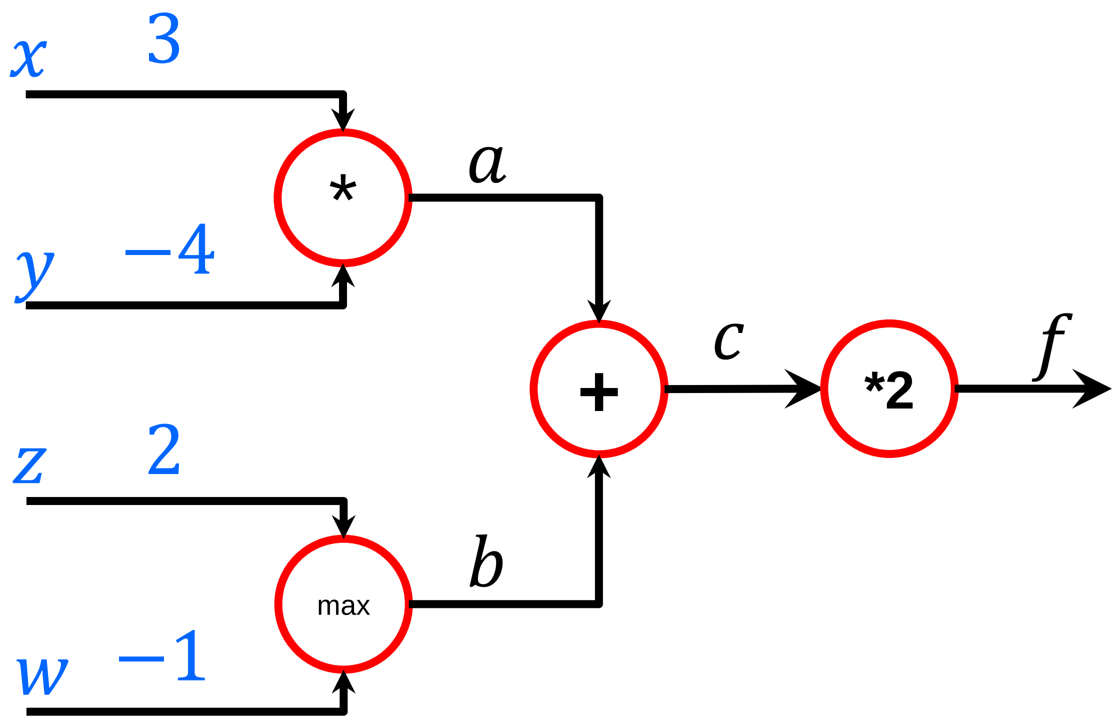
$x = 3$

$y = -4$

$z = 2$

$\mathrm{w} = -1$

Want: $\frac{\partial f}{\partial x}$ $\frac{\partial f}{\partial y}$ $\frac{\partial f}{\partial z}$ $\frac{\partial f}{\partial w}$



Backpropagation – More example

$$f(x,y,z,w) = \big((x \times y) + max(z,w)\big) \times 2$$

E.g:

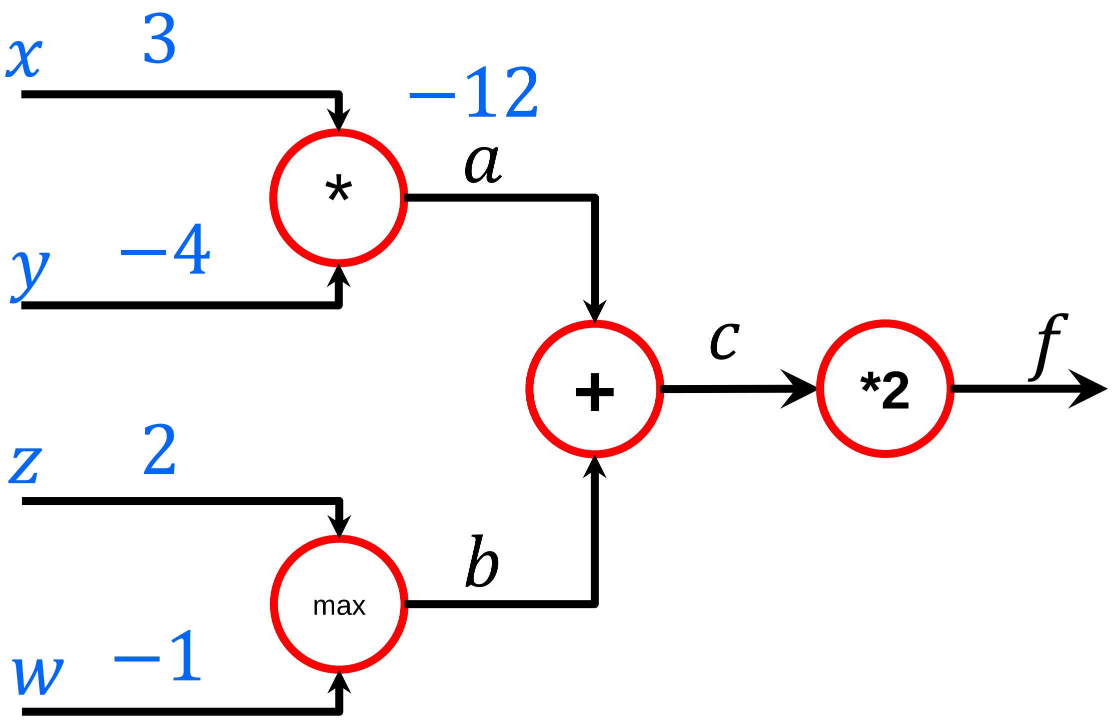
$x = 3$

$y = -4$

$z = 2$

$w = -1$

Want: $\frac{\partial f}{\partial x}$ $\frac{\partial f}{\partial y}$ $\frac{\partial f}{\partial z}$ $\frac{\partial f}{\partial w}$



Backpropagation – More example

$$\boldsymbol{f(x,y,z,w) = \big((x \times y) + max(z,w)\big) \times 2}$$

E.g:

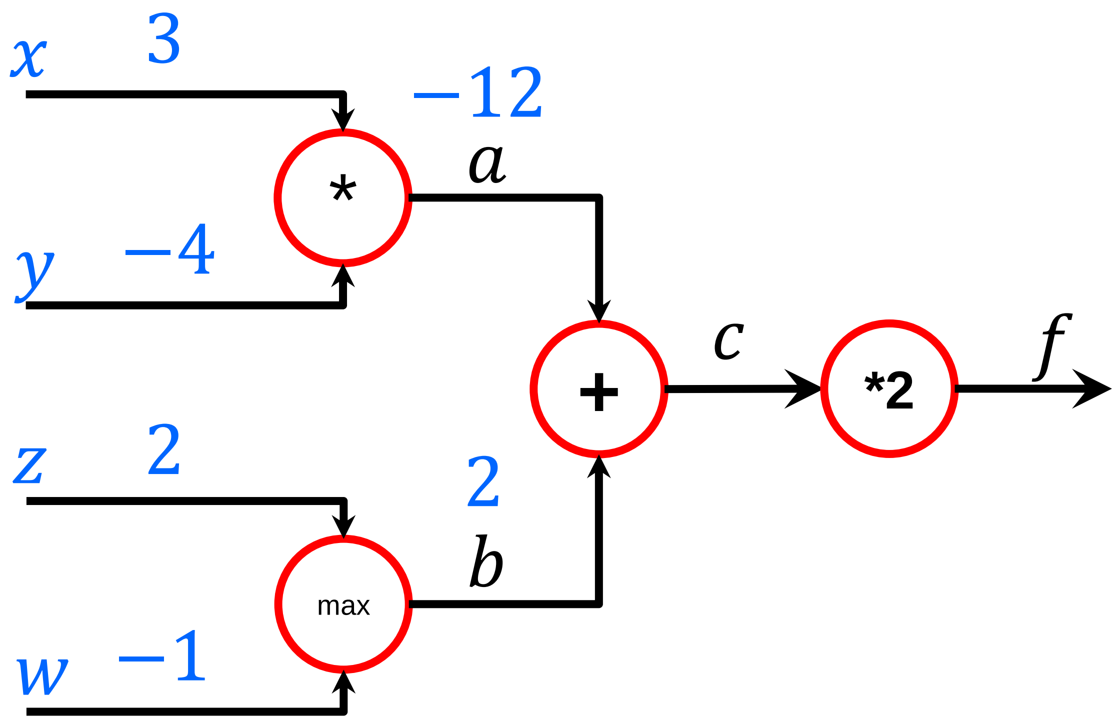
$x = 3$

$y = -4$

$z = 2$

$\mathrm{w} = -1$

Want: $\frac{\partial f}{\partial x}$ $\frac{\partial f}{\partial y}$ $\frac{\partial f}{\partial z}$ $\frac{\partial f}{\partial w}$



Backpropagation – More example

$$f(x, y, z, w) = \left((x \times y) + max(z, w)\right) \times 2$$

E.g:

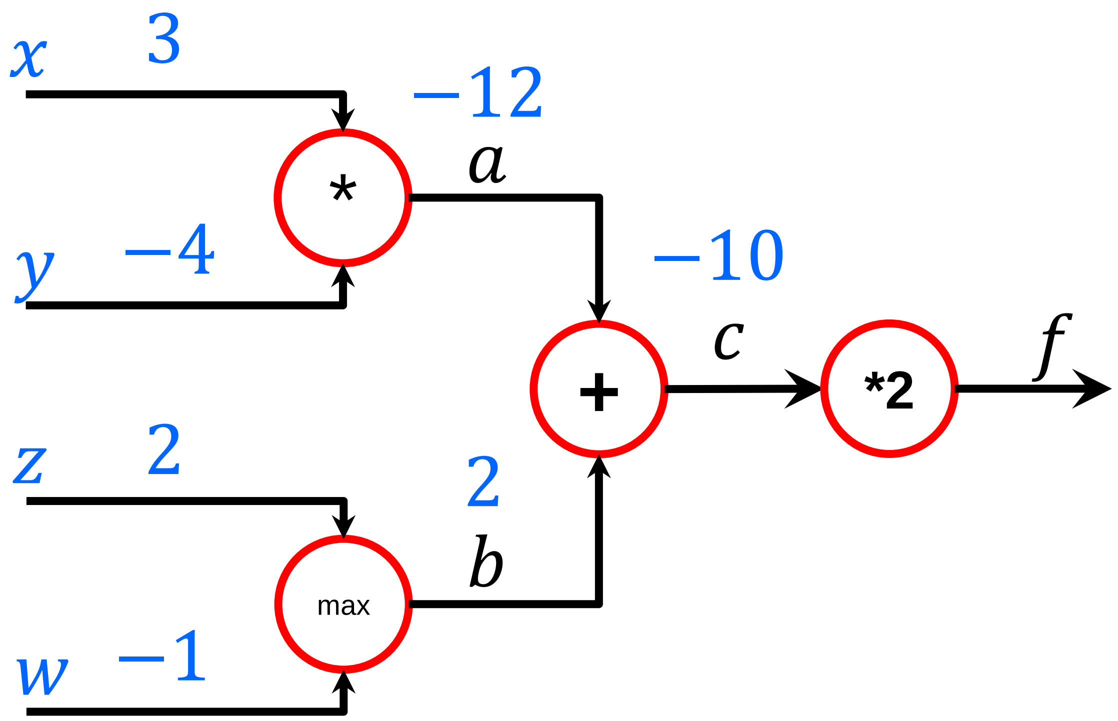
$x = 3$

$y = -4$

$z = 2$

$w = -1$

Want: $\frac{\partial f}{\partial x}$ $\frac{\partial f}{\partial y}$ $\frac{\partial f}{\partial z}$ $\frac{\partial f}{\partial w}$



Backpropagation – More example

$$f(x, y, z, w) = \big((x \times y) + max(z, w)\big) \times 2$$

E.g:

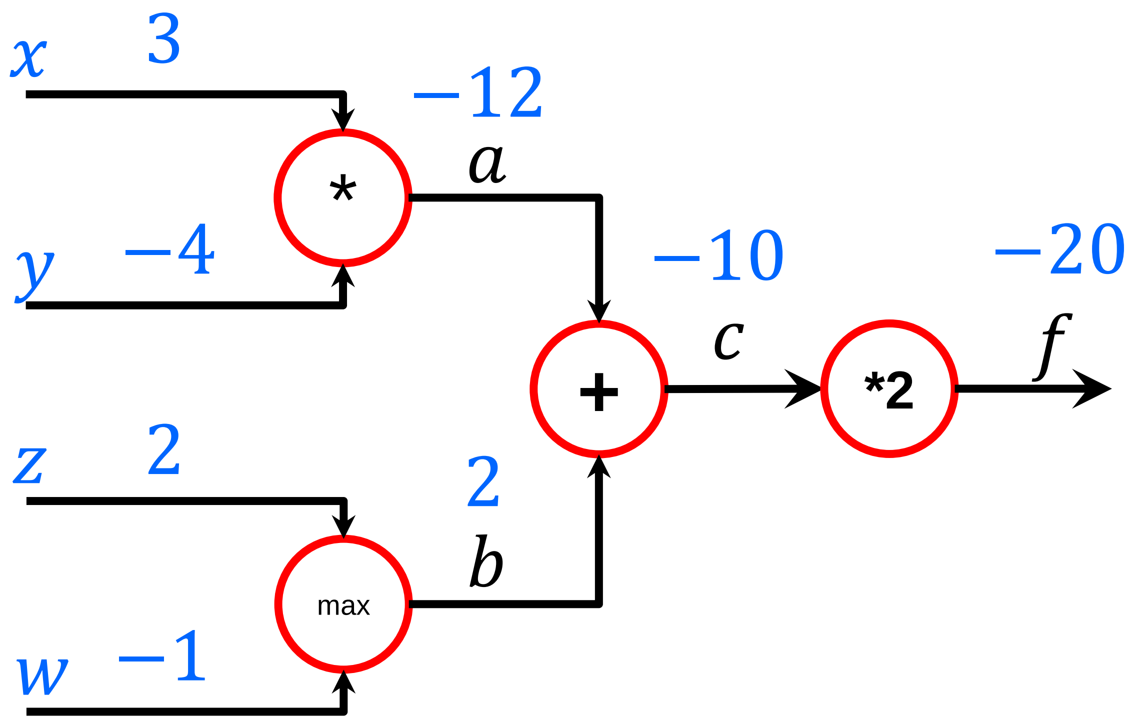
$x = 3$

$y = -4$

$z = 2$

$w = -1$

Want: $\frac{\partial f}{\partial x}$ $\frac{\partial f}{\partial y}$ $\frac{\partial f}{\partial z}$ $\frac{\partial f}{\partial w}$



Backpropagation – More example

$$\boldsymbol{f(x, y, z, w) = \big((x \times y) + max(z, w)\big) \times 2}$$

E.g:

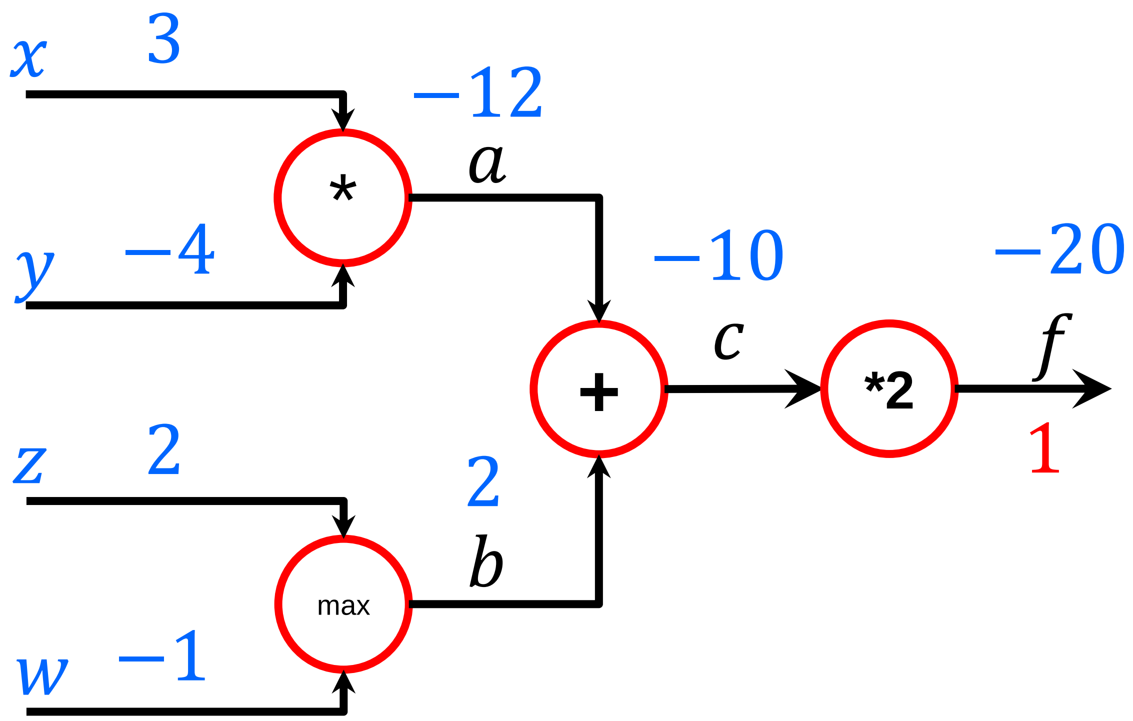
$x = 3$

$y = -4$

$z = 2$

$\mathrm{w} = -1$

Want: $\frac{\partial f}{\partial x}$ $\frac{\partial f}{\partial y}$ $\frac{\partial f}{\partial z}$ $\frac{\partial f}{\partial w}$



Backpropagation – More example

$$f(x,y,z,w) = \big((x \times y) + max(z,w)\big) \times 2$$

E.g:

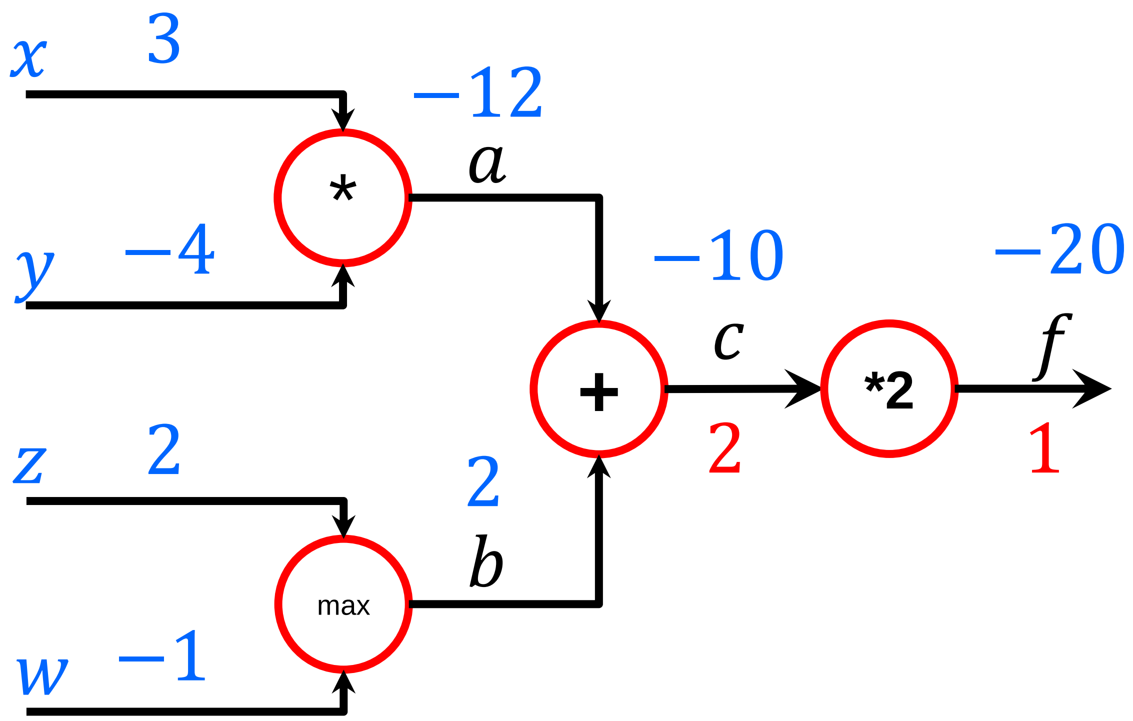
$x = 3$

$y = -4$

$z = 2$

$\mathrm{w} = -1$

Want: $\frac{\partial f}{\partial x}$ $\frac{\partial f}{\partial y}$ $\frac{\partial f}{\partial z}$ $\frac{\partial f}{\partial w}$



Backpropagation – More example

$$f(x,y,z,w) = \big((x \times y) + max(z,w)\big) \times 2$$

E.g:

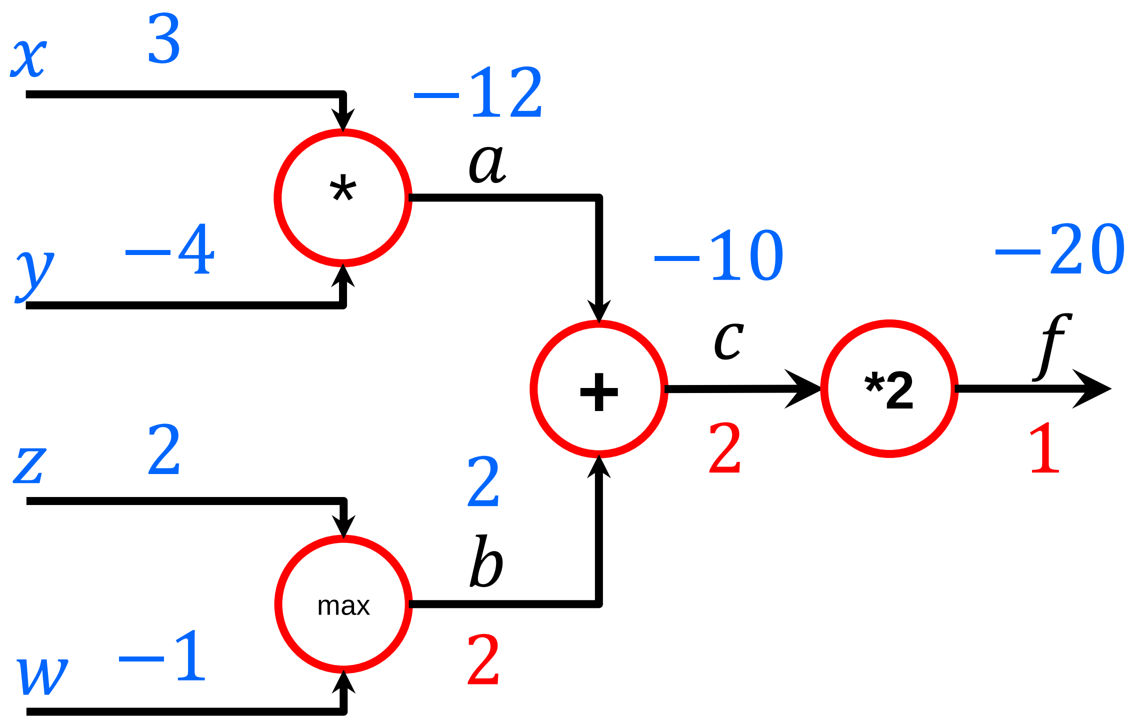
$x = 3$

$y = -4$

$z = 2$

$w = -1$

Want: $\frac{\partial f}{\partial x}$ $\frac{\partial f}{\partial y}$ $\frac{\partial f}{\partial z}$ $\frac{\partial f}{\partial w}$



Backpropagation – More example

$$f(x,y,z,w) = \big((x \times y) + max(z,w)\big) \times 2$$

E.g:

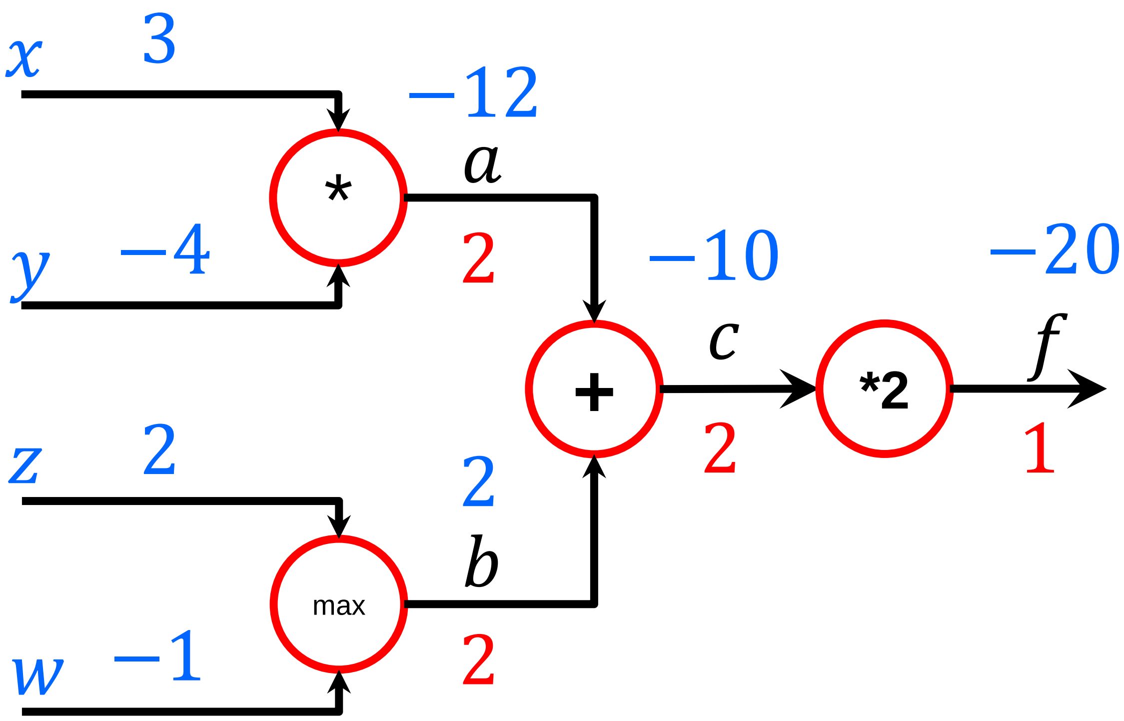
$x = 3$

$y = -4$

$z = 2$

$w = -1$

Want: $\frac{\partial f}{\partial x}$ $\frac{\partial f}{\partial y}$ $\frac{\partial f}{\partial z}$ $\frac{\partial f}{\partial w}$



Backpropagation – More example

$$f(x,y,z,w) = \big((x \times y) + max(z,w)\big) \times 2$$

E.g:

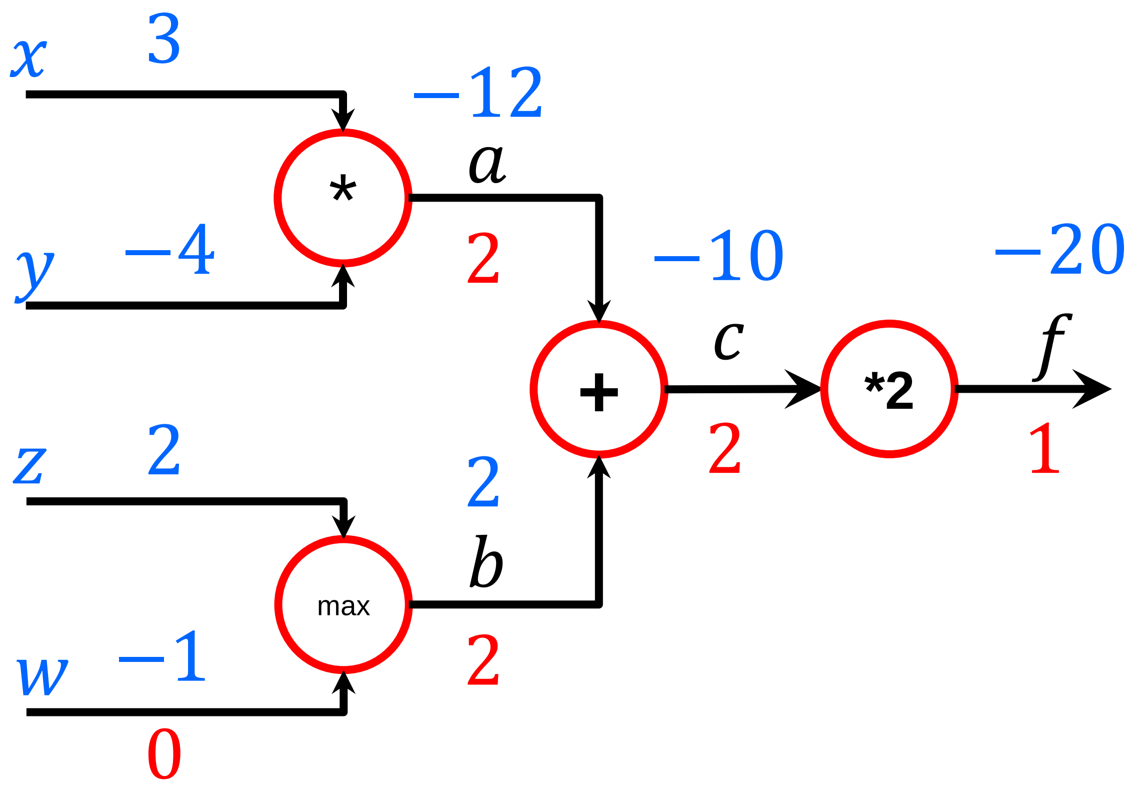
$x = 3$

$y = -4$

$z = 2$

$w = -1$

Want: $\frac{\partial f}{\partial x}$ $\frac{\partial f}{\partial y}$ $\frac{\partial f}{\partial z}$ $\frac{\partial f}{\partial w}$



Backpropagation – More example

$$f(x, y, z, w) = \big((x \times y) + max(z, w)\big) \times 2$$

E.g:

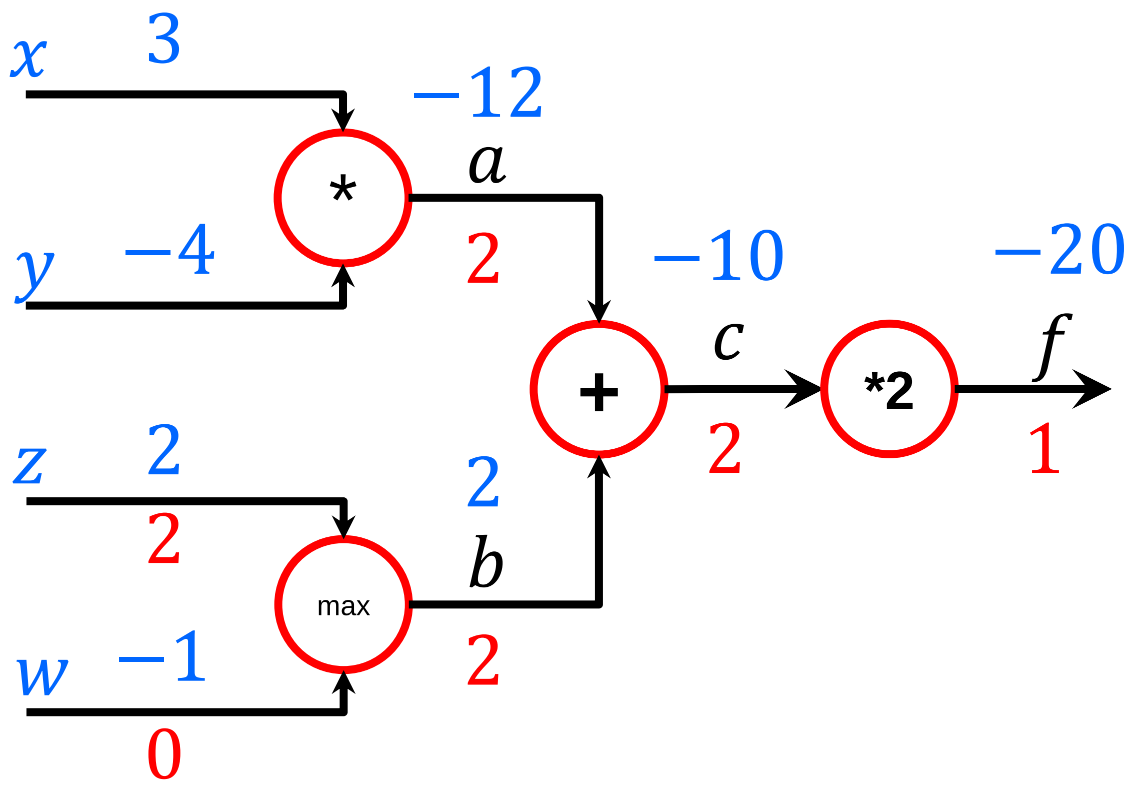
$x = 3$

$y = -4$

$z = 2$

$w = -1$

Want: $\frac{\partial f}{\partial x}$ $\frac{\partial f}{\partial y}$ $\frac{\partial f}{\partial z}$ $\frac{\partial f}{\partial w}$



Backpropagation – More example

$$f(x,y,z,w) = \big((x \times y) + max(z,w)\big) \times 2$$

E.g:

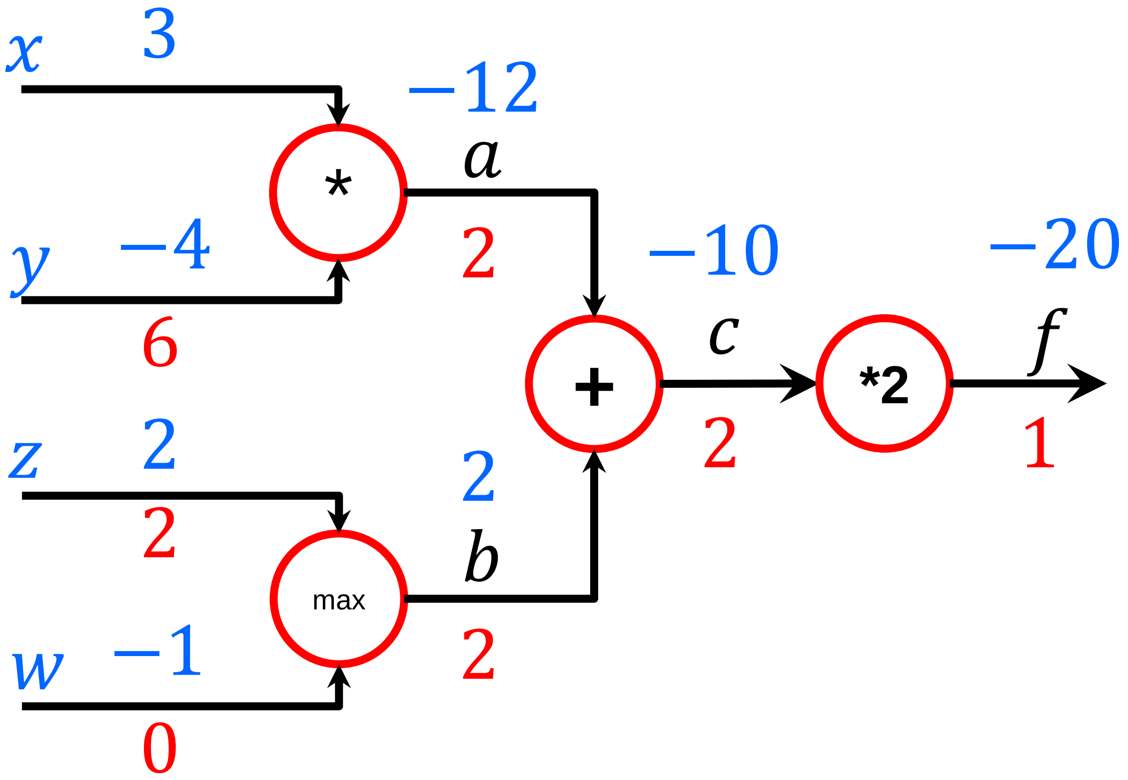
$x = 3$

$y = -4$

$z = 2$

$\mathrm{w} = -1$

Want: $\frac{\partial f}{\partial x}$ $\frac{\partial f}{\partial y}$ $\frac{\partial f}{\partial z}$ $\frac{\partial f}{\partial w}$



Backpropagation – More example

$$f(x,y,z,w) = \big((x \times y) + max(z,w)\big) \times 2$$

E.g:

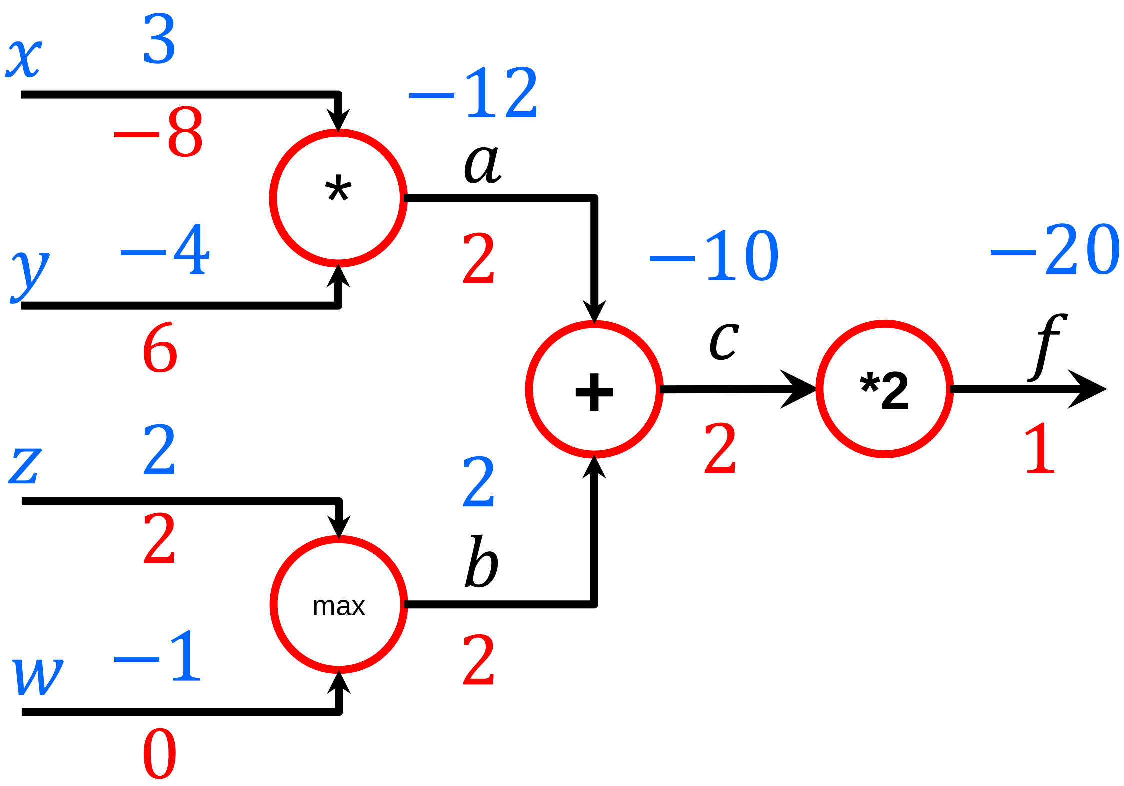
$x = 3$

$y = -4$

$z = 2$

$w = -1$

Want: $\frac{\partial f}{\partial x}$ $\frac{\partial f}{\partial y}$ $\frac{\partial f}{\partial z}$ $\frac{\partial f}{\partial w}$

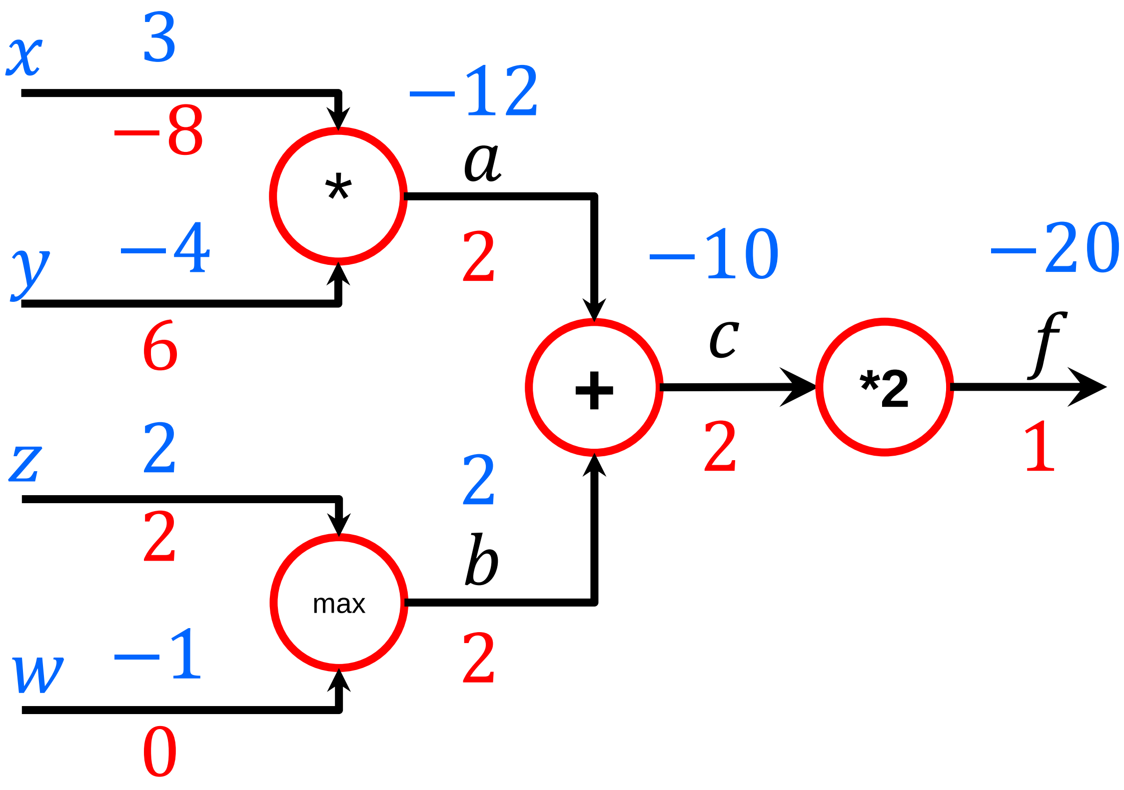


Patterns in backward flow

BACKPROPAGATION

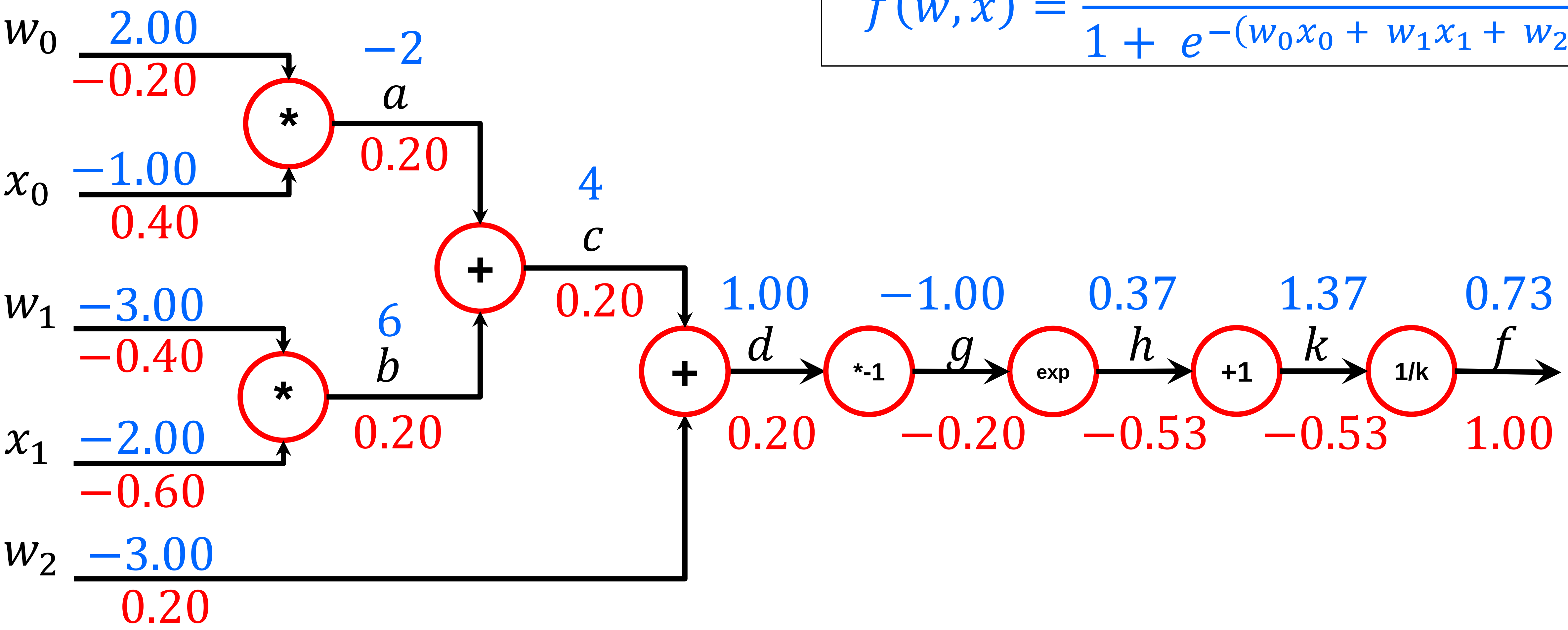
Backpropagation – Patterns

- Add gate: gradient distributor.
- Add gate: phân phối đạo hàm.
- Max gate: gradient router.
- Max gate: định tuyến đạo hàm.
- Mul gate: gradient switcher.
- Mul gate: chuyển đổi đạo hàm.



Backpropagation – Patterns

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



Backpropagation – Patterns

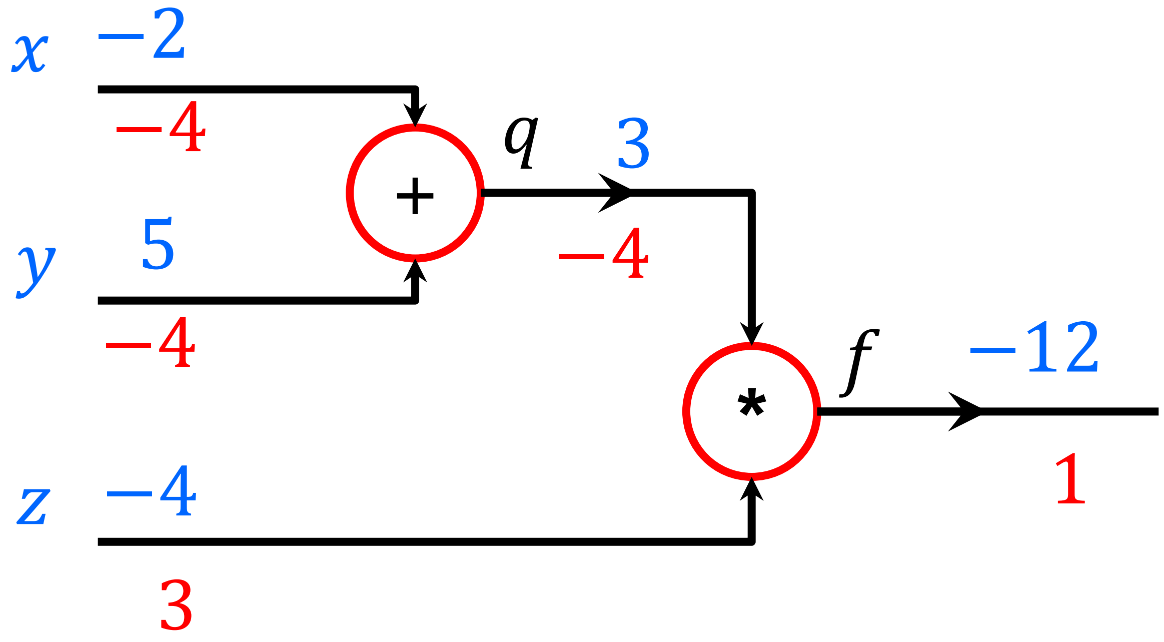
$$f(x, y, z) = (x + y)z$$

e.g. $x = -2, y = 5, z = -4$

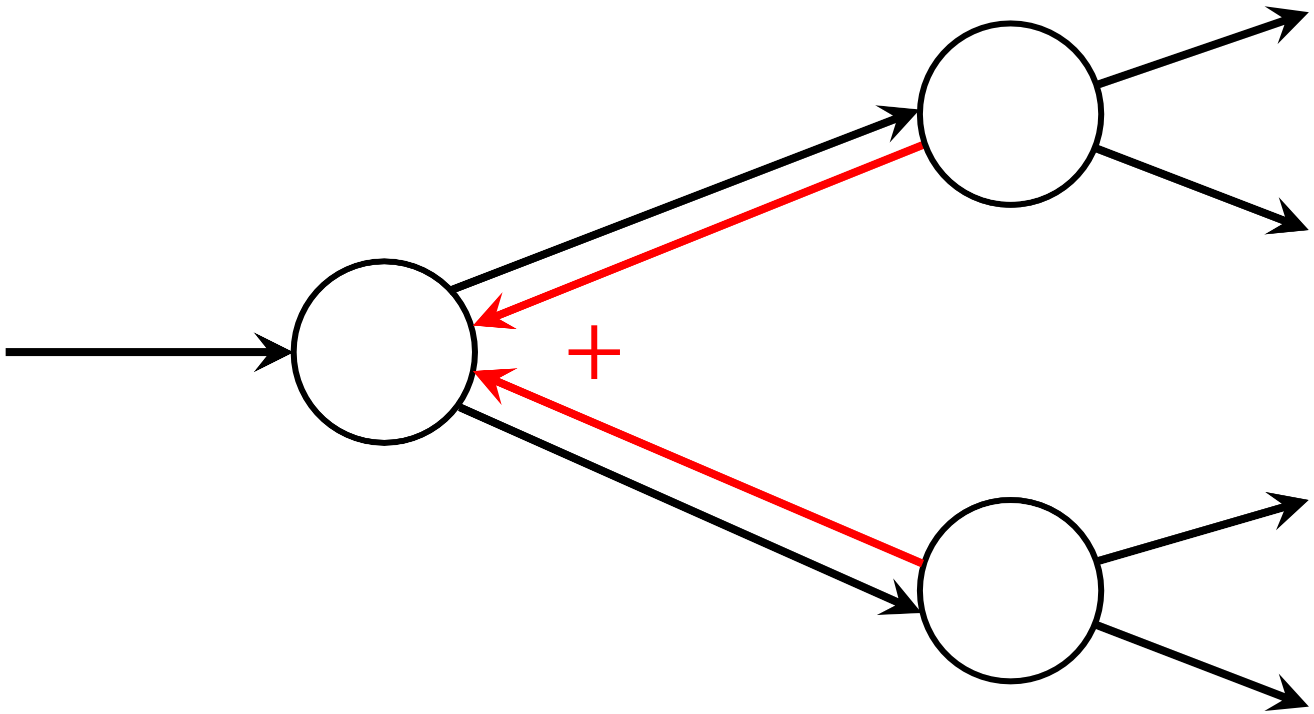
$q = x + y,$ $\quad \frac{\partial q}{\partial x} = 1, \quad \frac{\partial q}{\partial y} = 1$

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

– **Want:** $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$

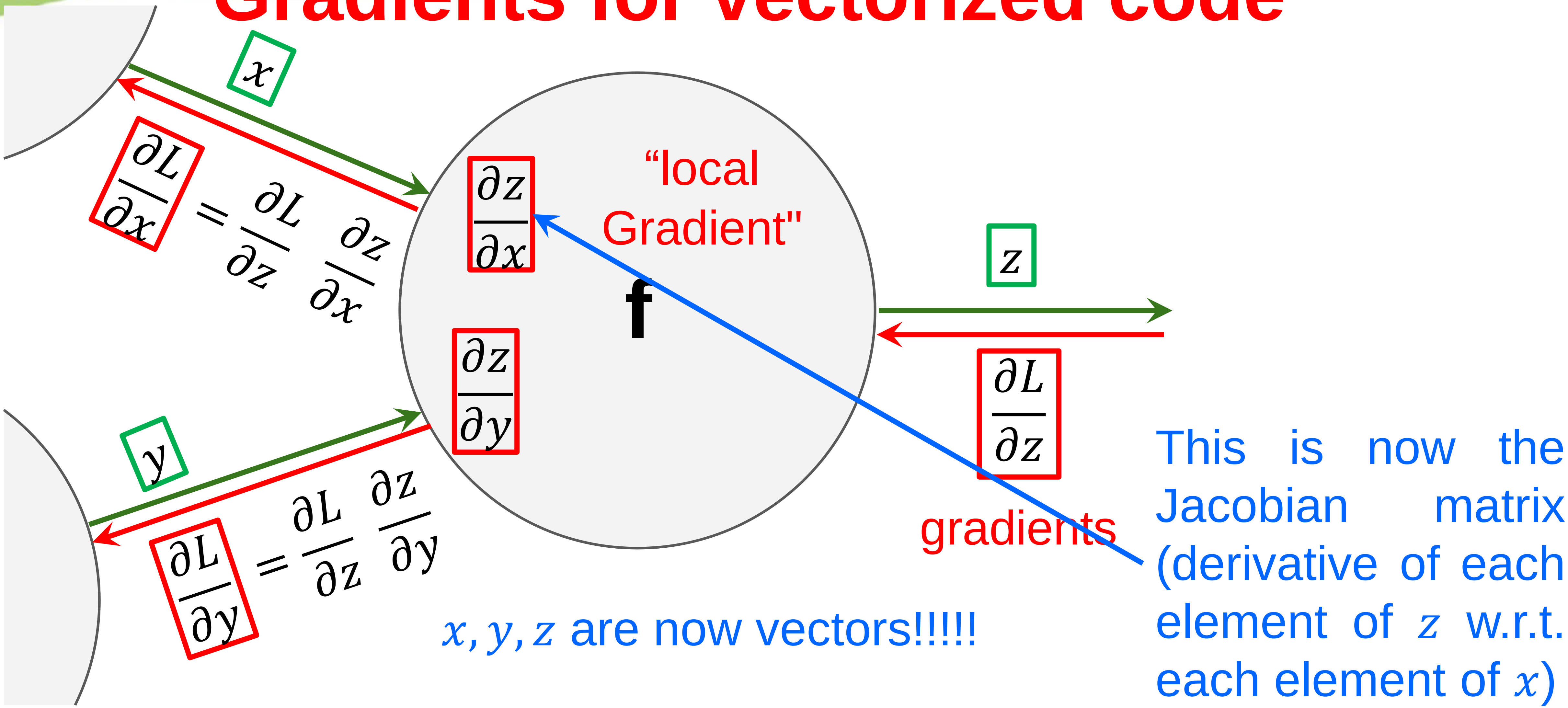


Gradients add at branches



$$\frac{\partial f}{\partial x} = \sum_{i=1}^{n} \frac{\partial f}{\partial q_i} \cdot \frac{\partial q_i}{\partial x}$$

Gradients for vectorized code



Jacobian matrix

- Take $f:\mathbb{R}^n \to \mathbb{R}^m$ a function such that its first-order partial deriavatives exist on $\mathbb{R}^n$.
- It takes a point $x \in \mathbb{R}^n$ as input and produces the vector z $= f(x) \in \mathbb{R}^m$ as output.
  + $x\ (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$
  + $z\ (z_1, z_2, \cdots, z_m) \in \mathbb{R}^m$
- Then the Jacobian matrix of f is

$$J = \begin{bmatrix} \frac{\partial z}{\partial x_1} & \cdots & \frac{\partial z}{\partial x_n} \end{bmatrix}$$

$$J = \begin{bmatrix} \frac{\partial z_1}{\partial x_1} & \frac{\partial z_1}{\partial x_2} & \cdots & \frac{\partial z_1}{\partial x_{n-1}} & \frac{\partial z_1}{\partial x_n} \\ \frac{\partial z_2}{\partial x_1} & \frac{\partial z_2}{\partial x_2} & \cdots & \frac{\partial z_2}{\partial x_{n-1}} & \frac{\partial z_2}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{\partial z_m}{\partial x_1} & \frac{\partial z_m}{\partial x_2} & \cdots & \frac{\partial z_m}{\partial x_{n-1}} & \frac{\partial z_m}{\partial x_n} \end{bmatrix}$$

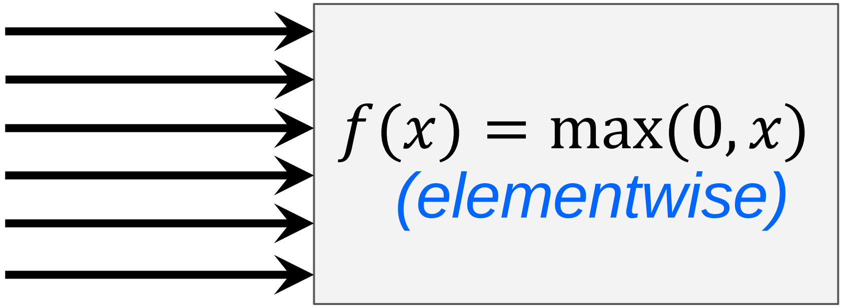
Vectorized operations

$$f(x) = \max(0, x)$$

(elementwise)

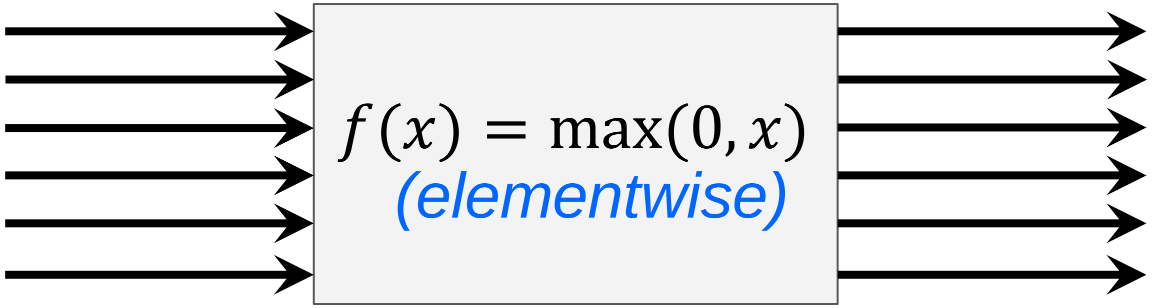
Vectorized operations

$4096-d$
input vector



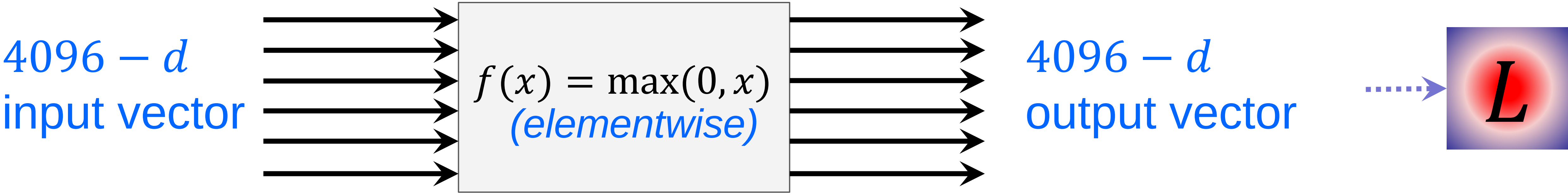
Vectorized operations

$4096 - d$
input vector

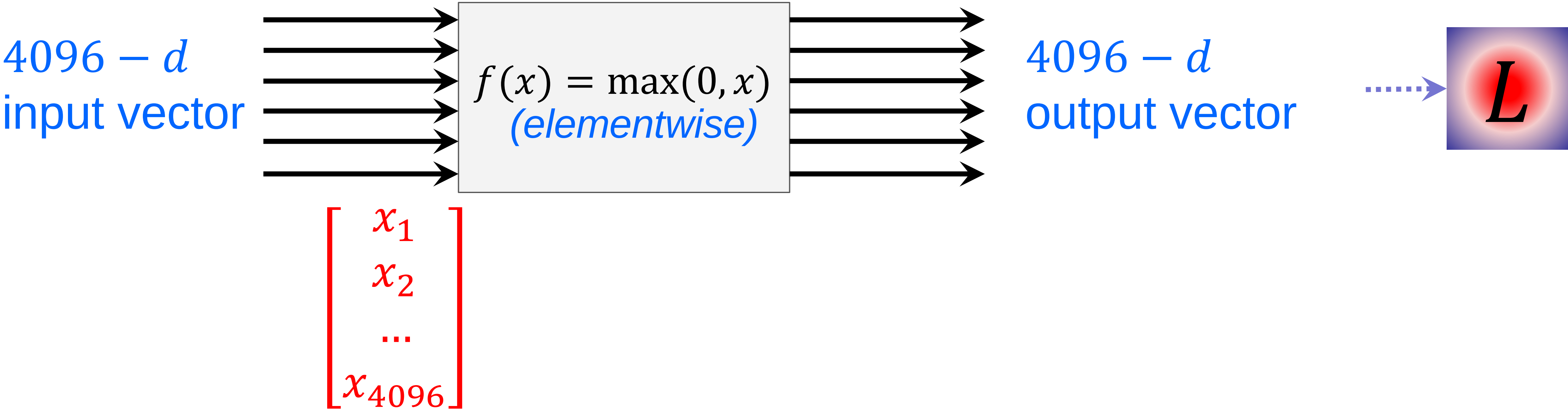


$4096 - d$
output vector

Vectorized operations

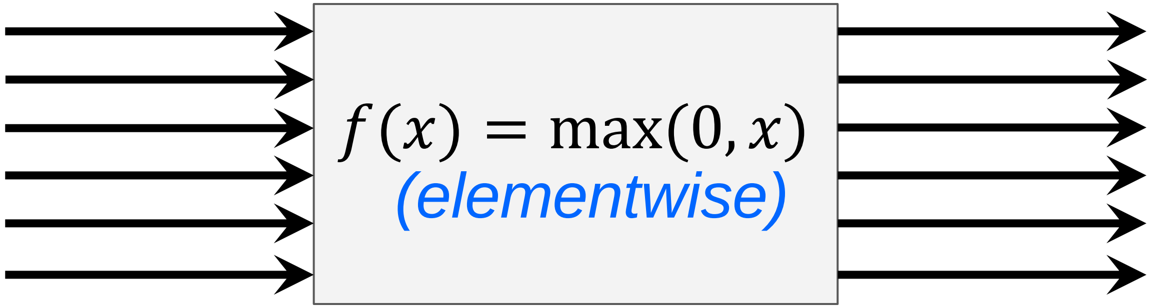


Vectorized operations



Vectorized operations

$4096-d$
input vector



$4096-d$
output vector

L

$$\begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_{4096} \end{bmatrix} \qquad \begin{bmatrix} max(0, x_1) \\ max(0, x_2) \\ \dots \\ max(0, x_{4096}) \end{bmatrix}$$

Vectorized operations

$4096 - d$
input vector

$f(x) = \max(0, x)$
(elementwise)

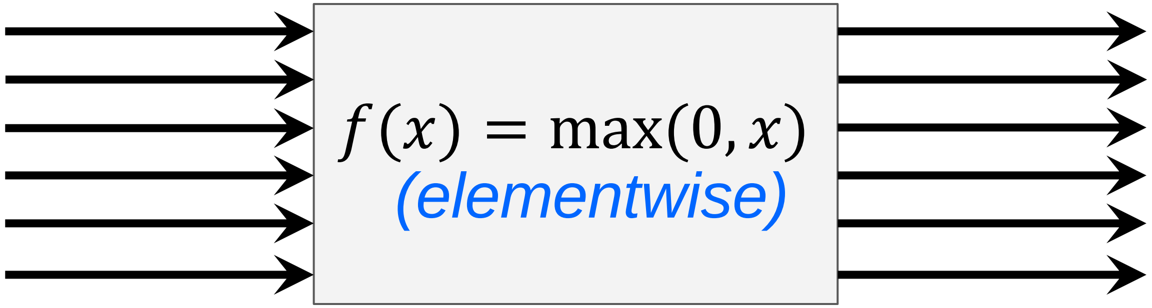
$4096 - d$
output vector

L

$$\begin{bmatrix} x_1 \\ x_2 \\ \ldots \\ x_{4096} \end{bmatrix} \qquad \begin{bmatrix} max(0, x_1) \\ max(0, x_2) \\ \ldots \\ max(0, x_{4096}) \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ \ldots \\ z_{4096} \end{bmatrix}$$

Vectorized operations

$4096 - d$
input vector



$4096 - d$
output vector

L

$$\begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_{4096} \end{bmatrix} \qquad \begin{bmatrix} max(0, x_1) \\ max(0, x_2) \\ \dots \\ max(0, x_{4096}) \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ \dots \\ z_{4096} \end{bmatrix}$$

Vectorized operations

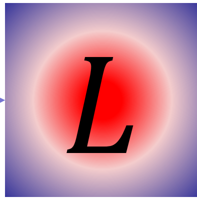
$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial z} \boxed{\frac{\partial z}{\partial x}}$$

Jacobian matrix

$4096 - d$
input vector

$f(x) = \max(0, x)$
(elementwise)

$4096 - d$
output vector



$$\begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_{4096} \end{bmatrix}$$

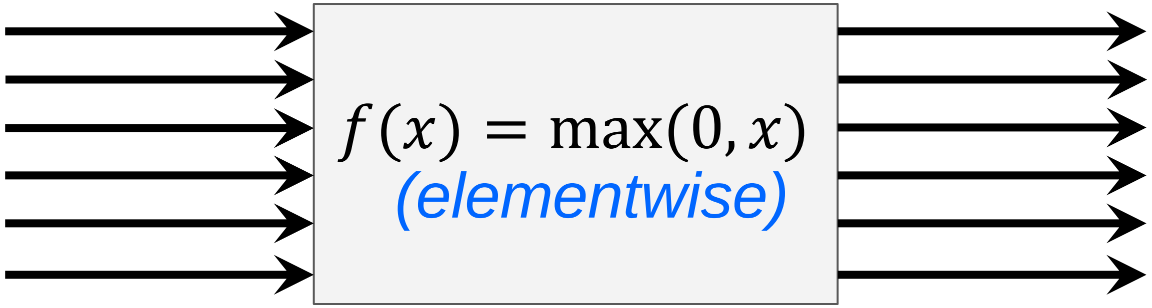
$$\begin{bmatrix} max(0, x_1) \\ max(0, x_2) \\ \dots \\ max(0, x_{4096}) \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ \dots \\ z_{4096} \end{bmatrix}$$

Vectorized operations

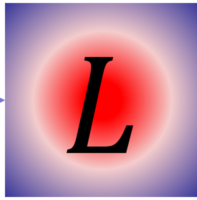
$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial z} \boxed{\frac{\partial z}{\partial x}}$$

Jacobian matrix

$4096 - d$
input vector



$4096 - d$
output vector



Q: what is the size of the Jacobian matrix?

$$\begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_{4096} \end{bmatrix} \qquad \begin{bmatrix} max(0, x_1) \\ max(0, x_2) \\ \dots \\ max(0, x_{4096}) \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ \dots \\ z_{4096} \end{bmatrix}$$

Vectorized operations

$$J = \begin{bmatrix} \frac{\partial z_1}{\partial x_1} & \frac{\partial z_1}{\partial x_2} & \cdots & \frac{\partial z_1}{\partial x_{4095}} & \frac{\partial z_1}{\partial x_{4096}} \\ \frac{\partial z_2}{\partial x_1} & \frac{\partial z_2}{\partial x_2} & \cdots & \frac{\partial z_2}{\partial x_{4095}} & \frac{\partial z_2}{\partial x_{4096}} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{\partial z_{4096}}{\partial x_1} & \frac{\partial z_{4096}}{\partial x_2} & \cdots & \frac{\partial z_{4095}}{\partial x_{4095}} & \frac{\partial z_{4096}}{\partial x_{4096}} \end{bmatrix}$$

Vectorized operations

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

Jacobian matrix

$4096-d$ input vector

$f(x) = \max(0, x)$ *(elementwise)*

$4096-d$ output vector

L

Q: what is the size of the Jacobian matrix?

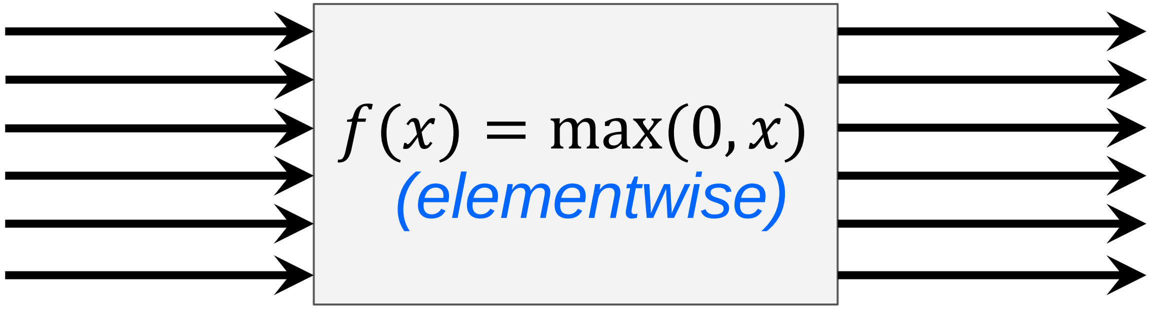
$$\begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_{4096} \end{bmatrix} \qquad \begin{bmatrix} max(0, x_1) \\ max(0, x_2) \\ \dots \\ max(0, x_{4096}) \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ \dots \\ z_{4096} \end{bmatrix}$$

Vectorized operations

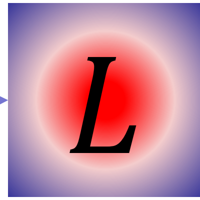
$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial z}\boxed{\frac{\partial z}{\partial x}}$$

Jacobian matrix

$4096-d$
input vector



$4096-d$
output vector



Q: what is the size of the Jacobian matrix?

A: $[4096 \times 4096]$

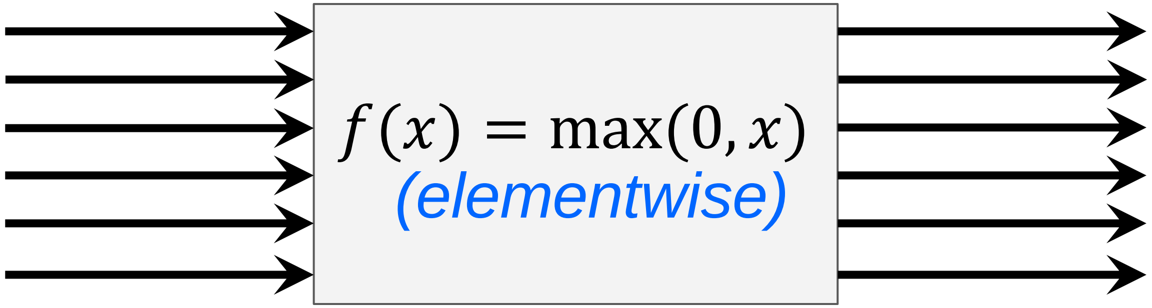
$$\begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_{4096} \end{bmatrix} \qquad \begin{bmatrix} max(0, x_1) \\ max(0, x_2) \\ \dots \\ max(0, x_{4096}) \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ \dots \\ z_{4096} \end{bmatrix}$$

Vectorized operations

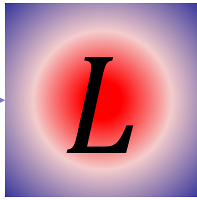
$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial z} \boxed{\frac{\partial z}{\partial x}}$$

Jacobian matrix

$4096-d$
input vector



$4096-d$
output vector



in practice we process an entire minibatch (e.g. 100) of examples at one time: i.e. Jacobian would technically be a $[409{,}600 \times 409{,}600]$ matrix.

A vectorized example

- A vectorized example:

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

$$f(x, W) = \sum_{i=1}^{n} (W \cdot x)_i^2$$

- Where:
  + $x \in \mathbb{R}^n$
  + $W \in \mathbb{R}^{n \times n}$

- Want:

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

- Where:

$$W = \begin{bmatrix} 0.1 & 0.5 \\ -0.3 & 0.8 \end{bmatrix}$$

$$x = \begin{bmatrix} 0.2 \\ 0.4 \end{bmatrix}$$

A vectorized example

– A vectorized example:

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

$$f(x,W) = \sum_{i=1}^{n} (W \cdot x)_i^2$$

– Where:

+ $x \in \mathbb{R}^n$

+ $W \in \mathbb{R}^{n \times n}$

– Want: $\frac{\partial f}{\partial x}$ $\frac{\partial f}{\partial W}$

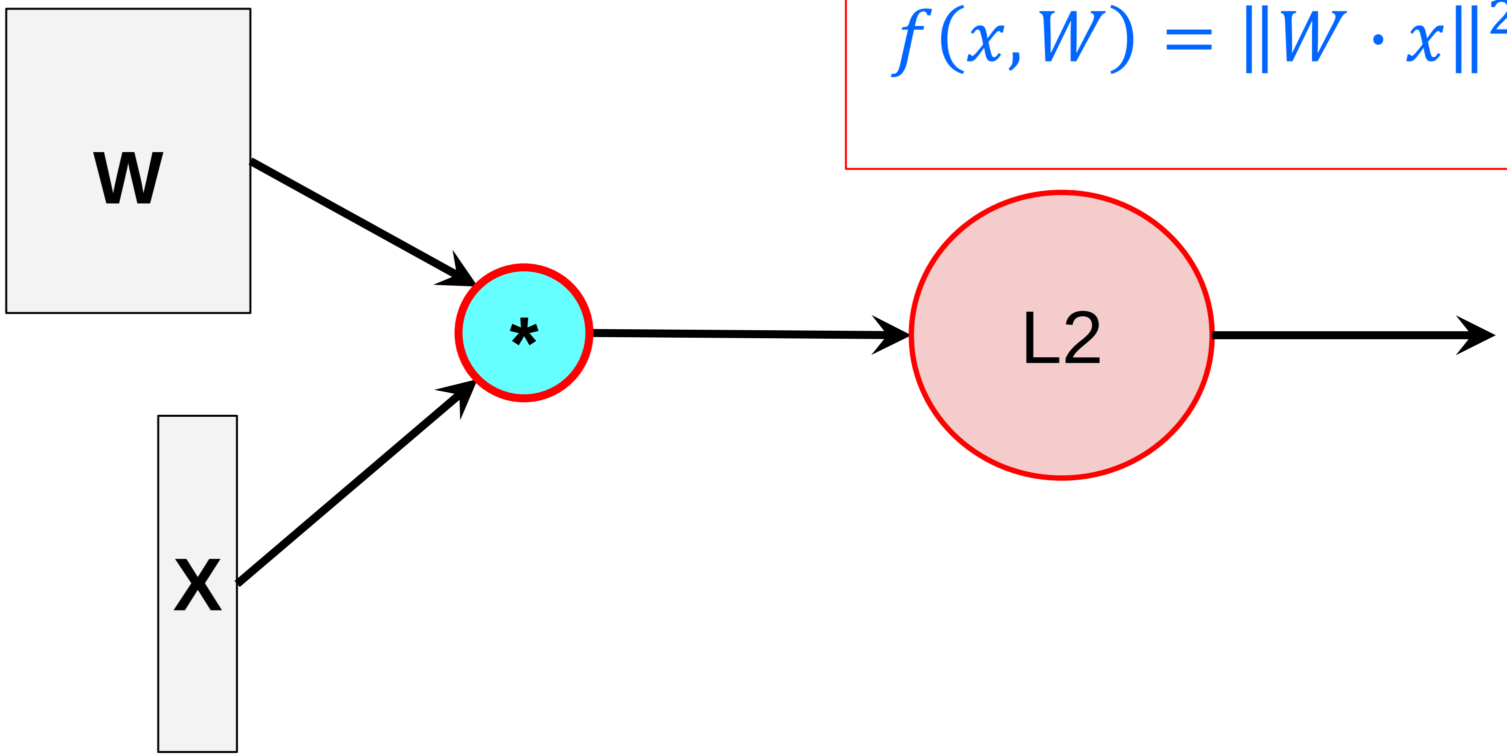
+ $x = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} \in \mathbb{R}^n$

+ $x^T = [x_1 \quad x_2 \quad \dots \quad x_n] \in \mathbb{R}^n$

+ $W = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1(n-1)} & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2(n-1)} & w_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ w_{n1} & w_{n2} & \dots & w_{n(n-1)} & w_{nn} \end{bmatrix}$

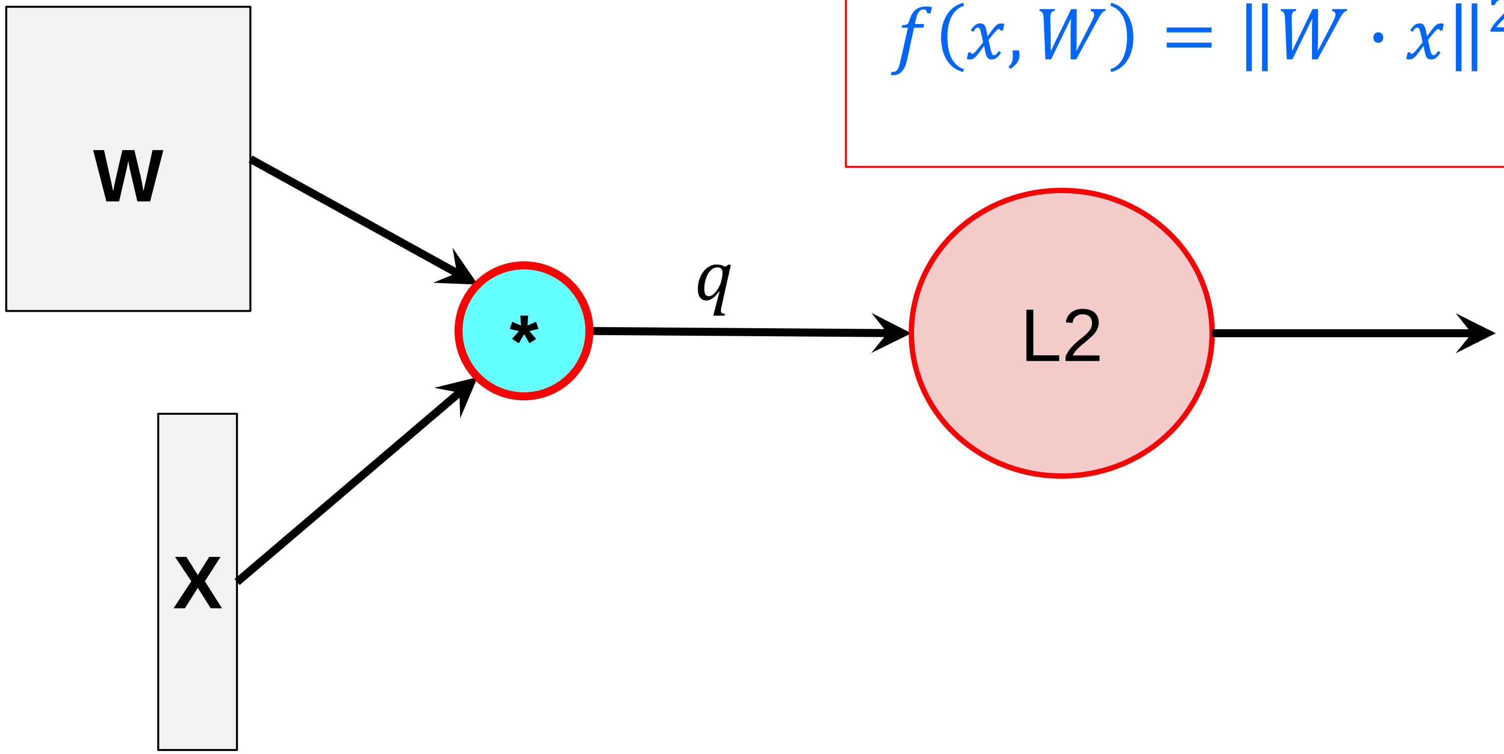
A vectorized example

$$f(x,W) = \|W \cdot x\|^2 = \sum_{i=1}^{n} (W \cdot x)_i^2$$



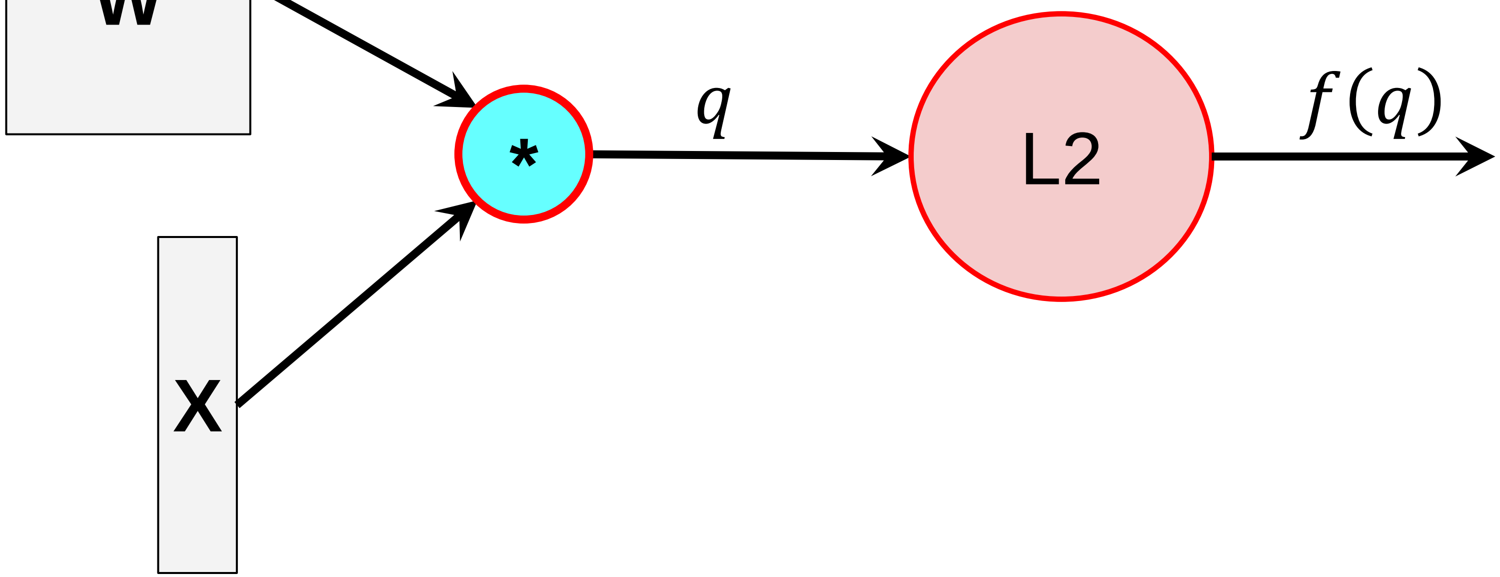
A vectorized example

$$f(x, W) = \|W \cdot x\|^2 = \sum_{i=1}^{n} (W \cdot x)_i^2$$



A vectorized example

$$f(x, W) = \|W \cdot x\|^2 = \sum_{i=1}^{n} (W \cdot x)_i^2$$



A vectorized example

– A vectorized example:

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

$$f(x,W) = \sum_{i=1}^{n} (W \cdot x)_i^2$$

– Where:

+ $x \in \mathbb{R}^n$

+ $W \in \mathbb{R}^{n \times n}$

– Want: $\frac{\partial f}{\partial x}$ $\frac{\partial f}{\partial W}$

+ $x = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} \in \mathbb{R}^n$

+ $x^T = [x_1 \quad x_2 \quad \dots \quad x_n] \in \mathbb{R}^n$

+ $W = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1(n-1)} & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2(n-1)} & w_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ w_{n1} & w_{n2} & \dots & w_{n(n-1)} & w_{nn} \end{bmatrix}$

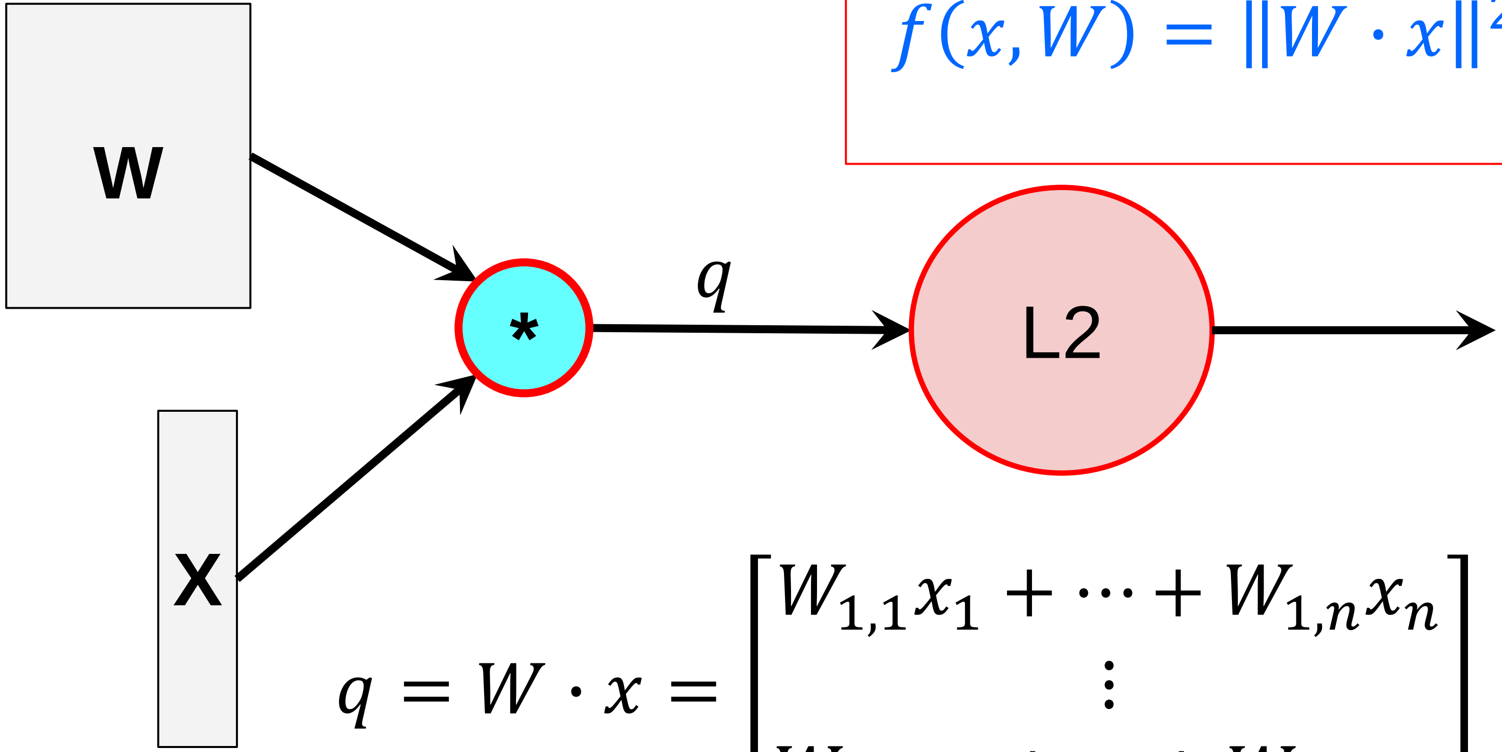
A vectorized example

$$q = W \cdot x = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1(n-1)} & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2(n-1)} & w_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ w_{n1} & w_{n2} & \dots & w_{n(n-1)} & w_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}$$

$$q = W \cdot x = \begin{bmatrix} W_{1,1}x_1 + \cdots + W_{1,n}x_n \\ \vdots \\ W_{n,1}x_1 + \cdots + W_{n,n}x_n \end{bmatrix}$$

A vectorized example

$$f(x, W) = \|W \cdot x\|^2 = \sum_{i=1}^{n} (W \cdot x)_i^2$$



$$q = W \cdot x = \begin{bmatrix} W_{1,1}x_1 + \cdots + W_{1,n}x_n \\ \vdots \\ W_{n,1}x_1 + \cdots + W_{n,n}x_n \end{bmatrix} \in \mathbb{R}^n$$

A vectorized example

– A vectorized example:

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

$$f(x, W) = \sum_{i=1}^{n} (W \cdot x)_i^2$$

– Where:

+ $x \in \mathbb{R}^n$

+ $W \in \mathbb{R}^{n \times n}$

– Want: $\frac{\partial f}{\partial x}$ $\frac{\partial f}{\partial W}$

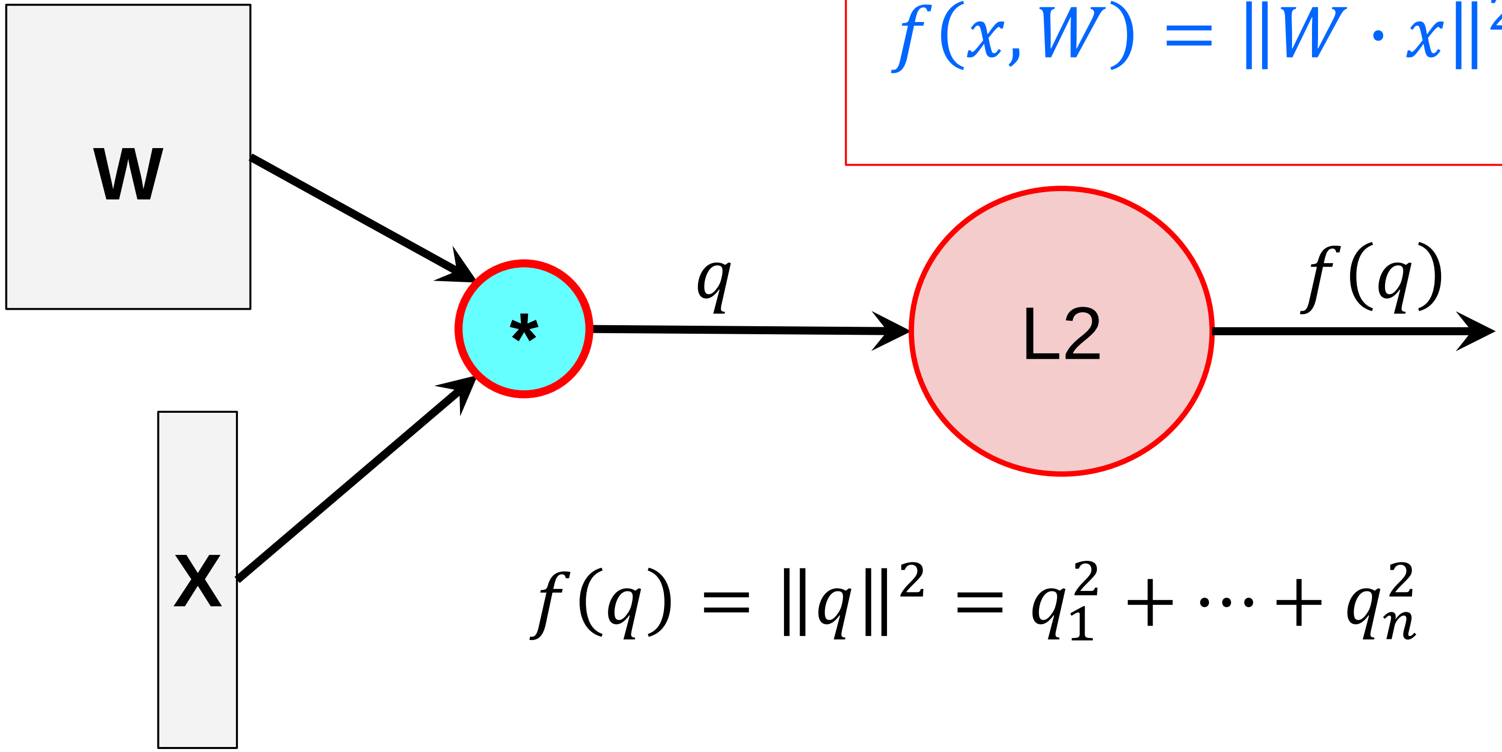
+ $x^T \; [x_1 \quad x_2 \quad \dots \quad x_n 0] \in \mathbb{R}^n$

+ $W = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1(n-1)} & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2(n-1)} & w_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ w_{n1} & w_{n2} & \dots & w_{n(n-1)} & w_{nn} \end{bmatrix}$

+ $q = W \cdot x = \begin{bmatrix} W_{1,1}x_1 + \cdots + W_{1,n}x_n \\ \vdots \\ W_{n,1}x_1 + \cdots + W_{n,n}x_n \end{bmatrix}$

A vectorized example

$$f(x, W) = \|W \cdot x\|^2 = \sum_{i=1}^{n} (W \cdot x)_i^2$$



$$f(q) = \|q\|^2 = q_1^2 + \cdots + q_n^2 \qquad f(q) \in \mathbb{R}$$

A vectorized example

– A vectorized example:

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

$$f(x,W) = \sum_{i=1}^{n} (W \cdot x)_i^2$$

– Where:

+ $x \in \mathbb{R}^n$

+ $W \in \mathbb{R}^{n\times n}$

– Want: $\frac{\partial f}{\partial x}$ $\frac{\partial f}{\partial W}$

+ $x^T \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \in \mathbb{R}^n$

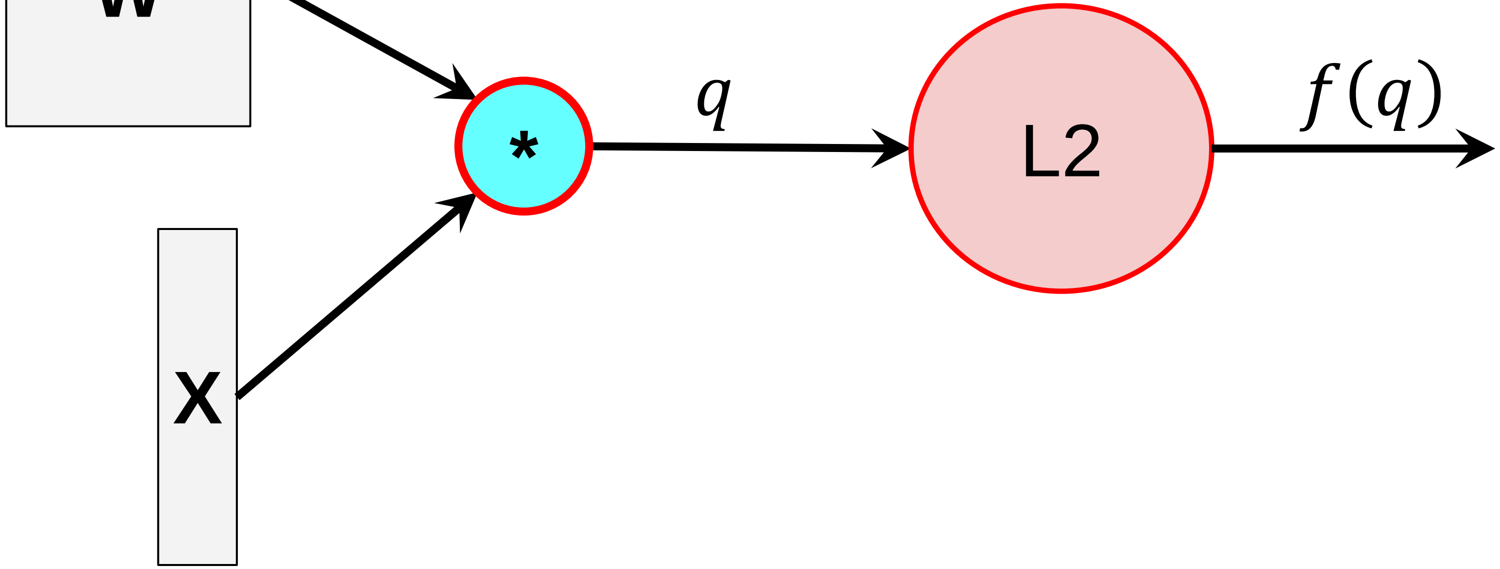
+ $W = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1(n-1)} & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2(n-1)} & w_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ w_{n1} & w_{n2} & \dots & w_{n(n-1)} & w_{nn} \end{bmatrix}$

+ $q = W \cdot x = \begin{bmatrix} W_{1,1}x_1 + \cdots + W_{1,n}x_n \\ \vdots \\ W_{n,1}x_1 + \cdots + W_{n,n}x_n \end{bmatrix}$

+ $f(q) = \|q\|^2 = q_1^2 + \cdots + q_n^2$

A vectorized example

$$f(x, W) = \|W \cdot x\|^2 = \sum_{i=1}^{n} (W \cdot x)_i^2$$

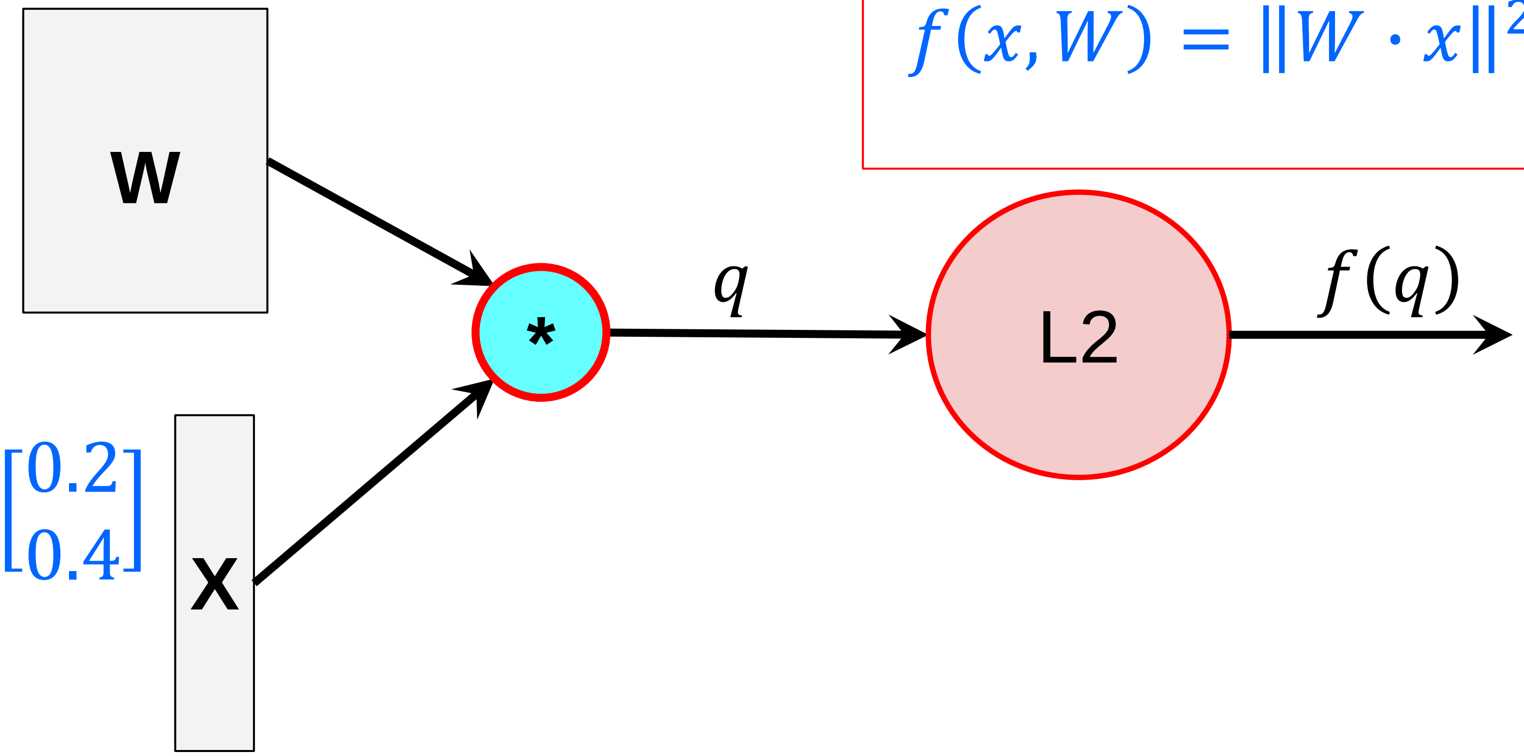


A vectorized example

$$f(x, W) = \|W \cdot x\|^2 = \sum_{i=1}^{n} (W \cdot x)_i^2$$

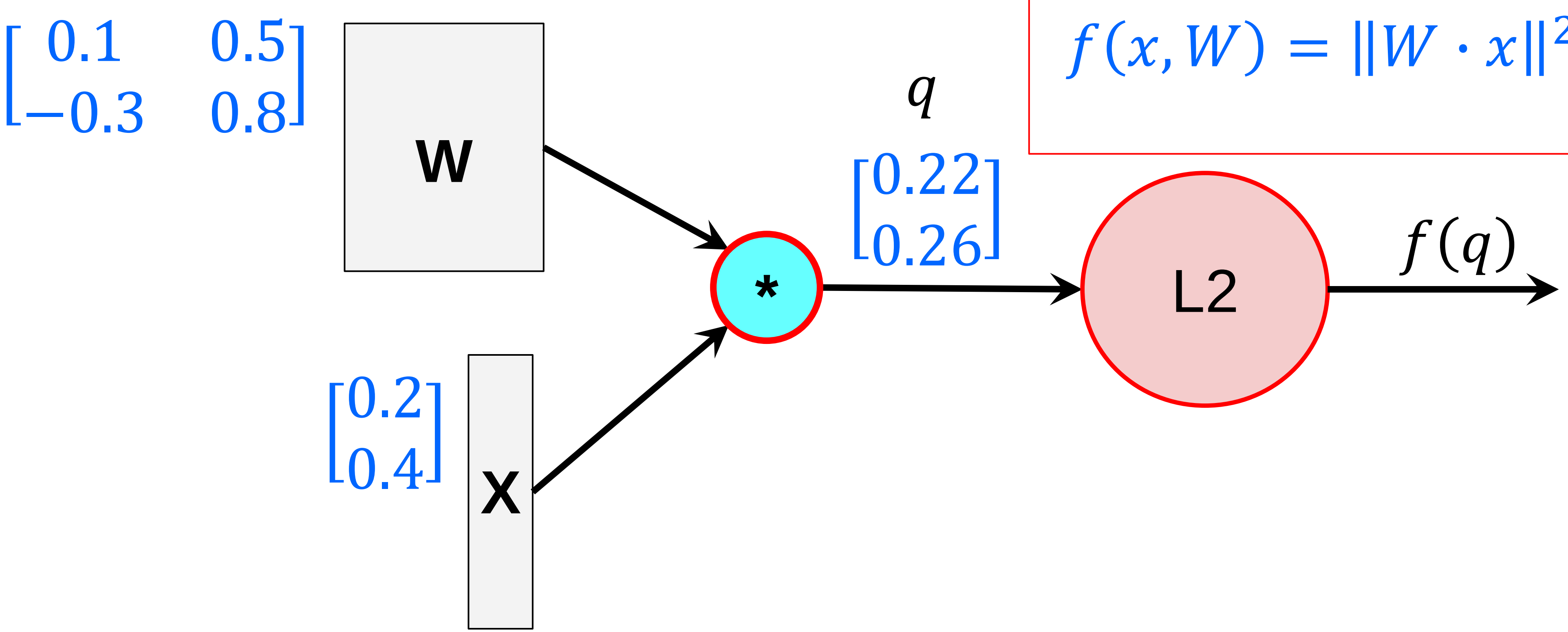
$$\begin{bmatrix} 0.1 & 0.5 \\ -0.3 & 0.8 \end{bmatrix}$$

$$\begin{bmatrix} 0.2 \\ 0.4 \end{bmatrix}$$



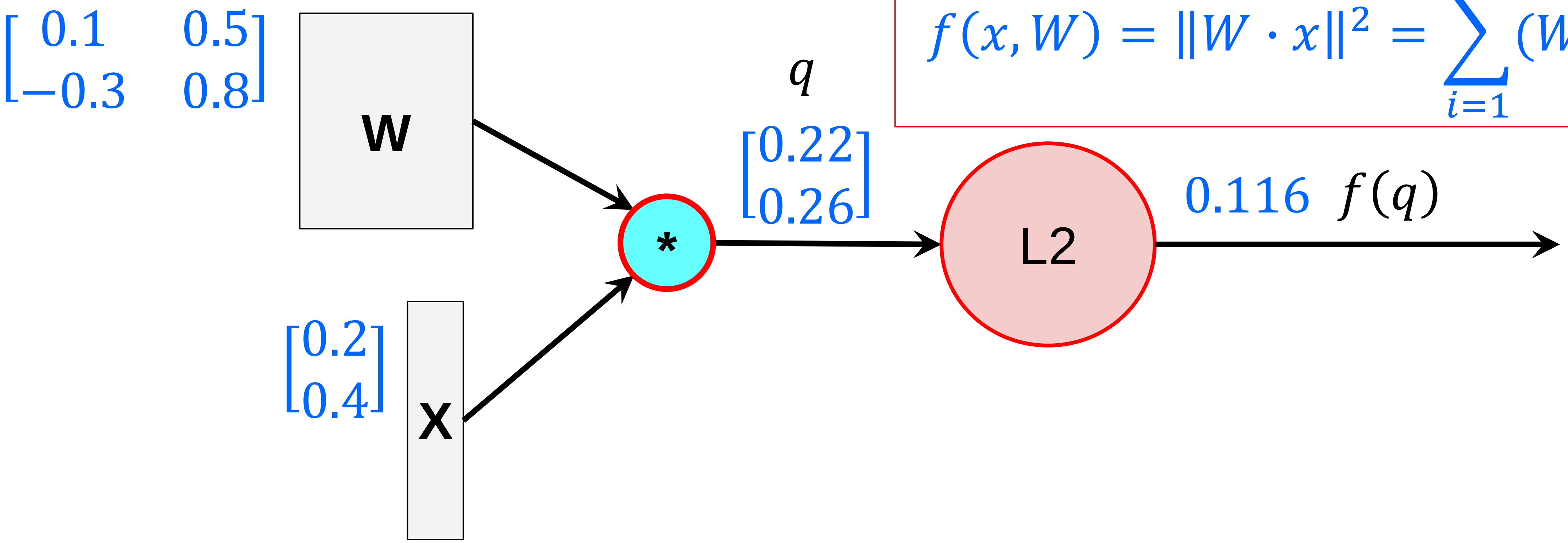
A vectorized example

$$f(x, W) = \|W \cdot x\|^2 = \sum_{i=1}^{n} (W \cdot x)_i^2$$



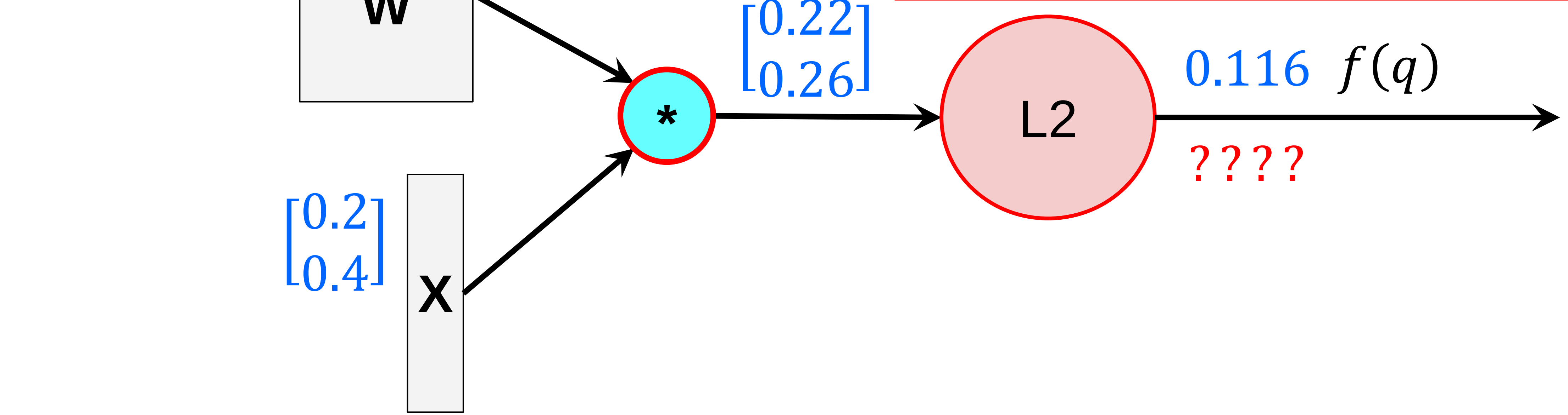
A vectorized example

$$f(x, W) = \|W \cdot x\|^2 = \sum_{i=1}^{n} (W \cdot x)_i^2$$



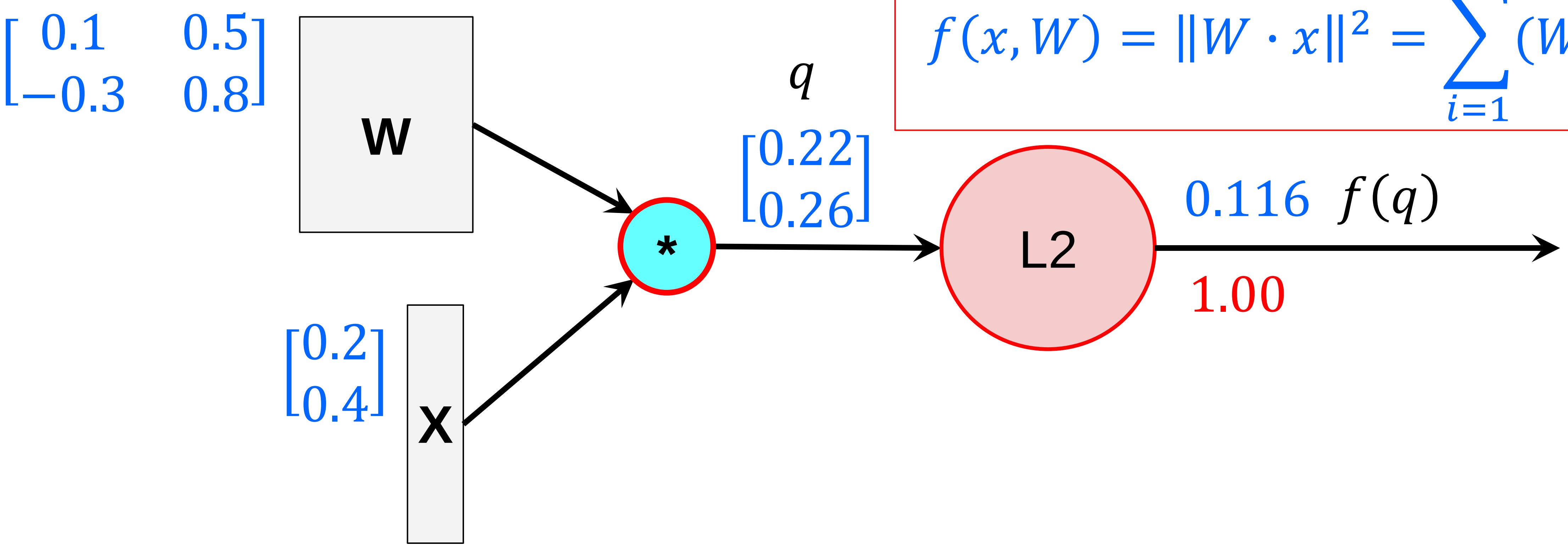
A vectorized example

$$f(x, W) = \|W \cdot x\|^2 = \sum_{i=1}^{n} (W \cdot x)_i^2$$



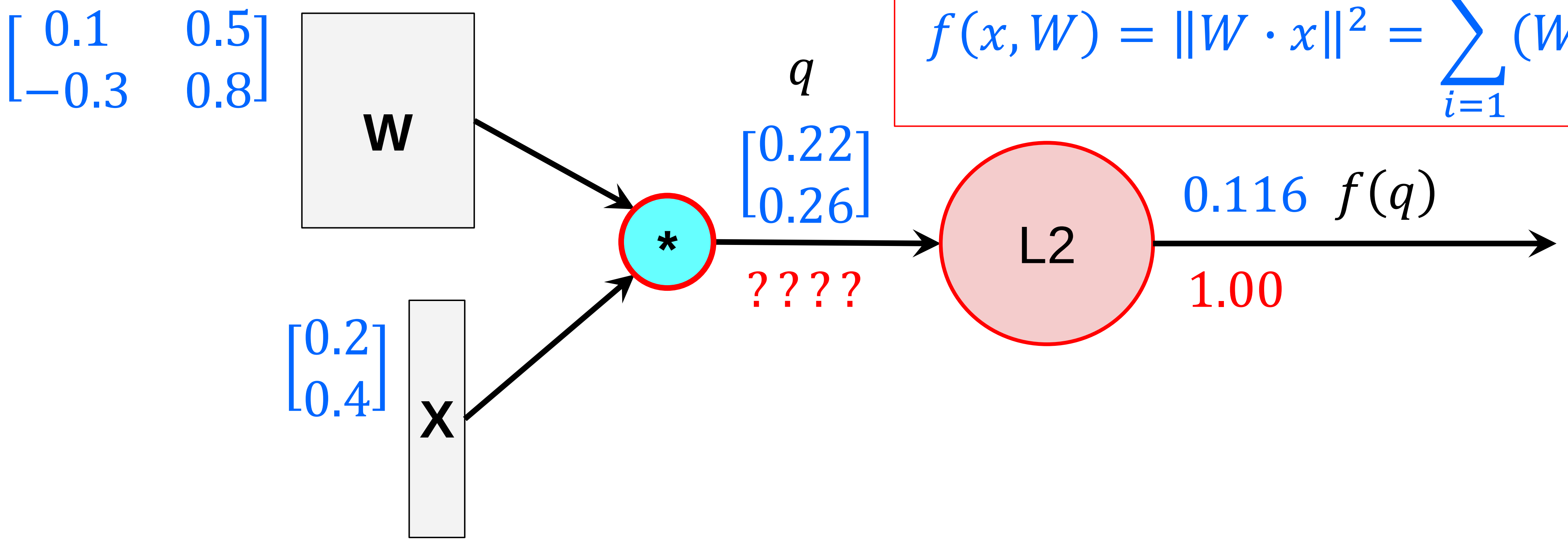
A vectorized example

$$f(x, W) = \|W \cdot x\|^2 = \sum_{i=1}^{n} (W \cdot x)_i^2$$



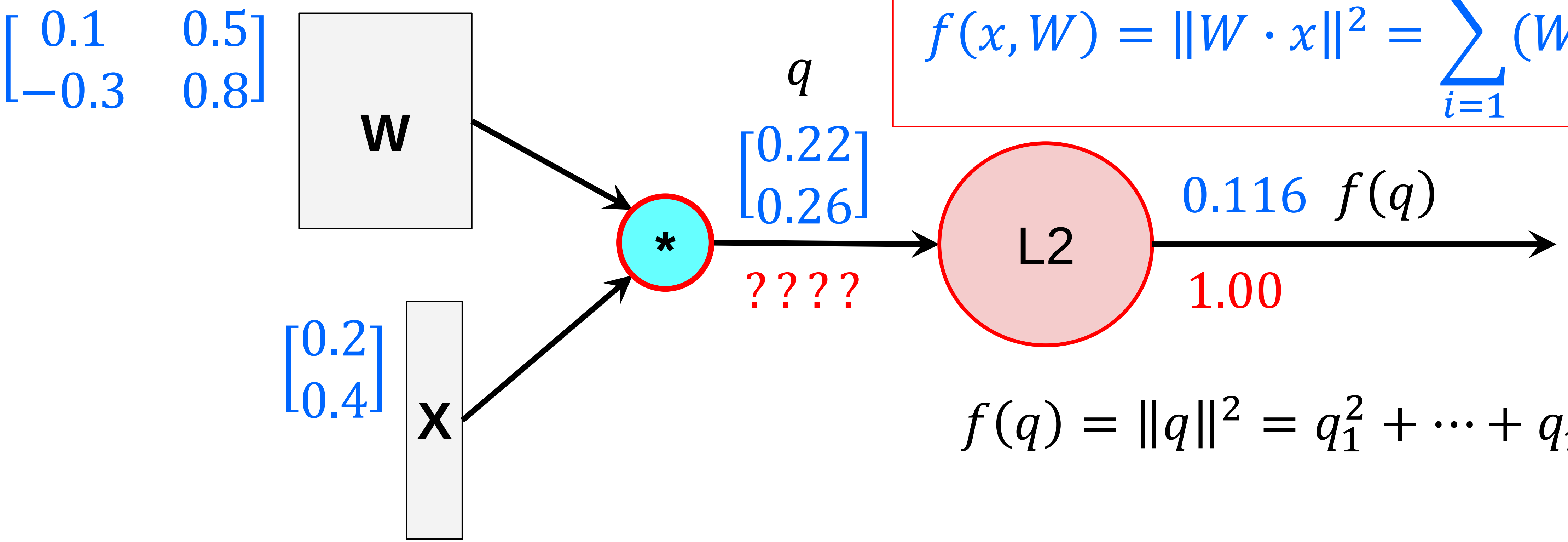
A vectorized example

$$f(x, W) = \|W \cdot x\|^2 = \sum_{i=1}^{n} (W \cdot x)_i^2$$



A vectorized example

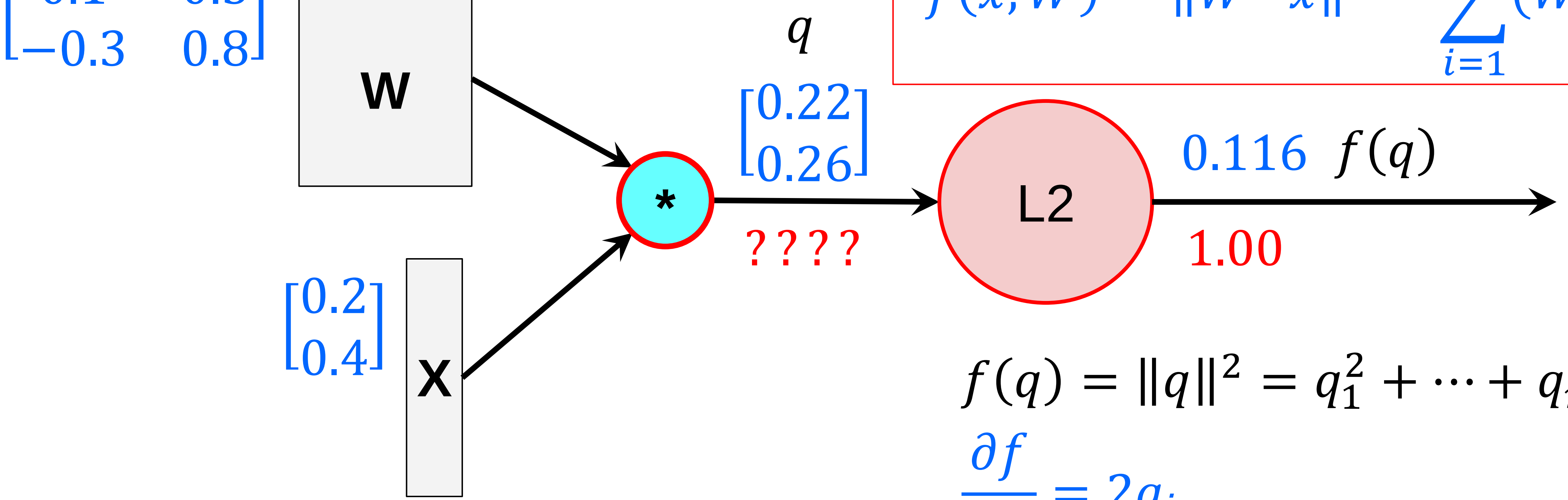
$$f(x, W) = \|W \cdot x\|^2 = \sum_{i=1}^{n} (W \cdot x)_i^2$$



$$f(q) = \|q\|^2 = q_1^2 + \cdots + q_n^2$$

A vectorized example

$$f(x, W) = \|W \cdot x\|^2 = \sum_{i=1}^{n} (W \cdot x)_i^2$$

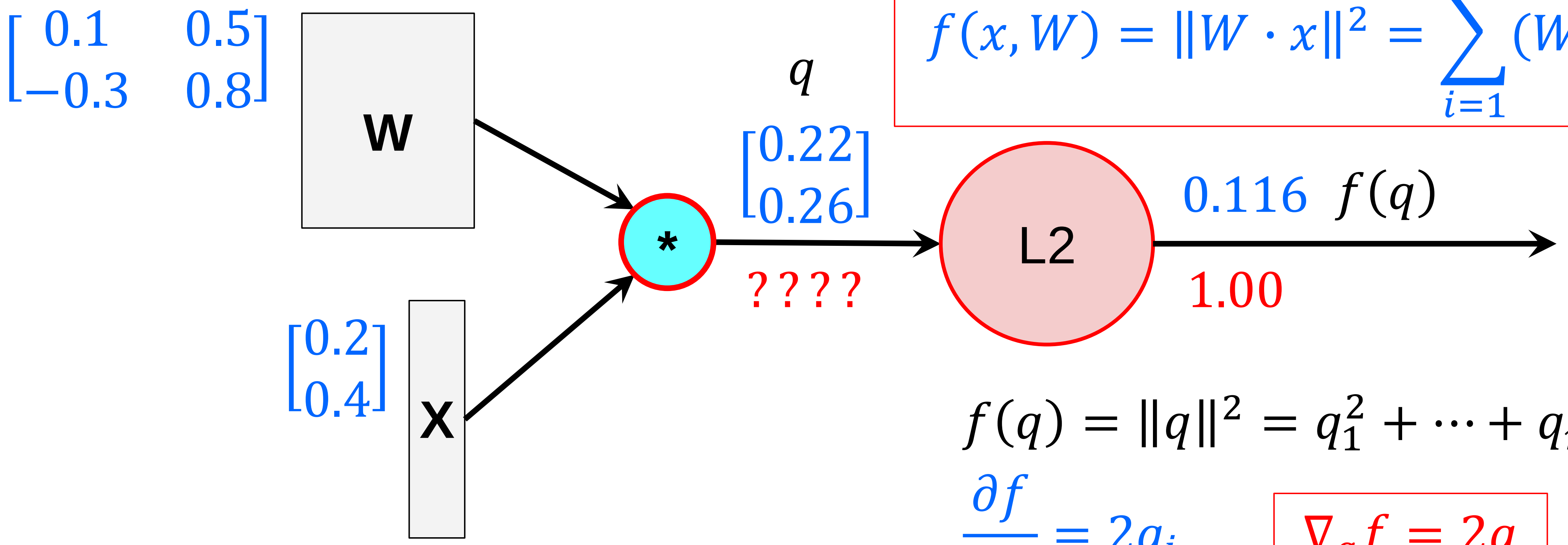


$$f(q) = \|q\|^2 = q_1^2 + \cdots + q_n^2$$

$$\frac{\partial f}{\partial q_i} = 2q_i$$

A vectorized example

$$f(x, W) = \|W \cdot x\|^2 = \sum_{i=1}^{n} (W \cdot x)_i^2$$



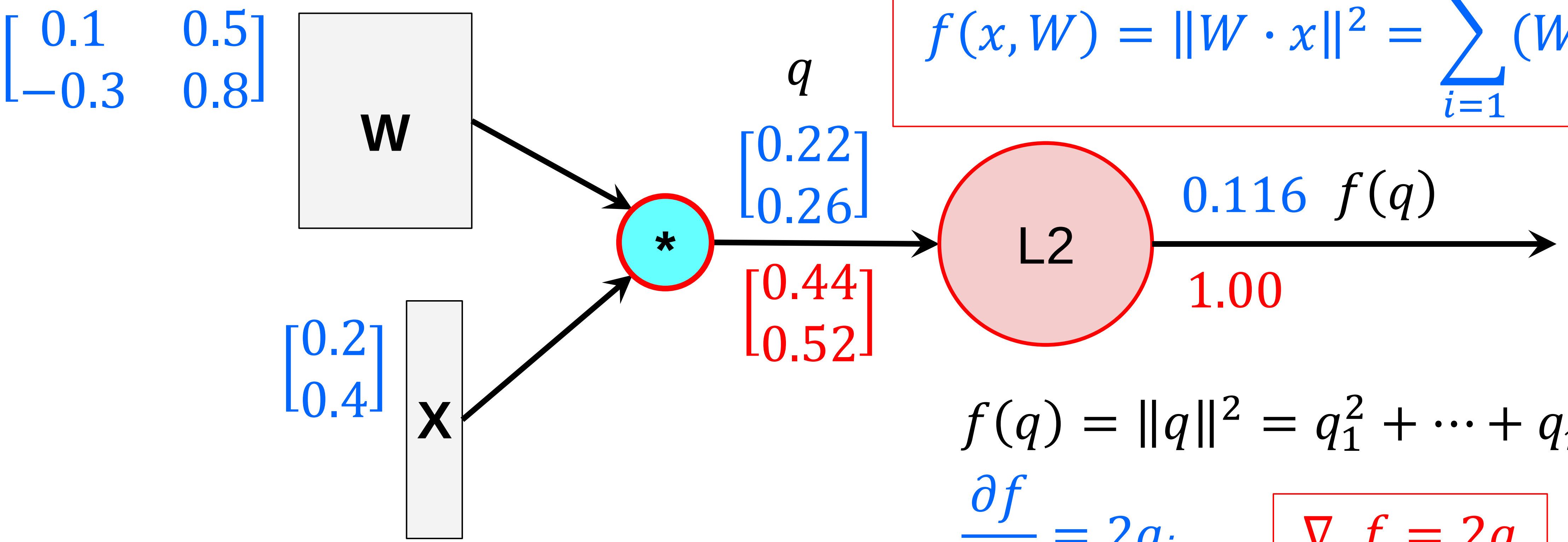
$$f(q) = \|q\|^2 = q_1^2 + \cdots + q_n^2$$

$$\frac{\partial f}{\partial q_i} = 2q_i$$

$$\boxed{\nabla_q f = 2q}$$

A vectorized example

$$f(x,W) = \|W \cdot x\|^2 = \sum_{i=1}^{n} (W \cdot x)_i^2$$



$$f(q) = \|q\|^2 = q_1^2 + \cdots + q_n^2$$

$$\frac{\partial f}{\partial q_i} = 2q_i$$

$$\boxed{\nabla_q f = 2q}$$

A vectorized example

– Bài toán: Tìm đạo hàm của L đối với W và đối với x khi biết đạo hàm của L đối với q và $q = W \cdot x$

– Giả thiết:

+ $q = W \cdot x$,

+ $\frac{\partial L}{\partial q}$.

– Cần tính:

+ $\frac{\partial L}{\partial W}$ và $\frac{\partial L}{\partial x}$

A vectorized example

– Đạo hàm của L đối với W:

+ Để tìm $\frac{\partial L}{\partial W}$, ta sử dụng quy tắc chuỗi: $\frac{\partial L}{\partial W} = \frac{\partial L}{\partial q} \cdot \frac{\partial q}{\partial W}$

+ Vì $q = W \cdot x$, do đó: $\frac{\partial q}{\partial W} = x$

+ Với $\frac{\partial L}{\partial q}$, là một vector cột, ta cần nhân $\frac{\partial L}{\partial q}$ với chuyển vị của x, kết quả là: $\frac{\partial L}{\partial W} = \frac{\partial L}{\partial q} \cdot x^\top$

+ Điều này tạo ra một ma trận mà mỗi phần tử là tích của các phần tử tương ứng của $\frac{\partial L}{\partial q}$ và $x^\top$.

A vectorized example

– Đạo hàm của L đối với x:

+ Để tìm $\frac{\partial L}{\partial x}$, ta sử dụng quy tắc chuỗi:

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial q} \cdot \frac{\partial q}{\partial x}$$

+ Vì $q = W \cdot x$, ta có: $\frac{\partial q}{\partial x} = W$

+ Với $\frac{\partial L}{\partial q}$, ta nhân nó với W, kết quả là:

$$\frac{\partial L}{\partial x} = W^{\top} \cdot \frac{\partial L}{\partial q}$$

A vectorized example

– Tóm lại: Để tìm đạo hàm của L đối với W và x khi biết đạo hàm của L đối với q (biết $\frac{\partial L}{\partial q}$) và $q = W \cdot x$

+ Đạo hàm của L đối với W:

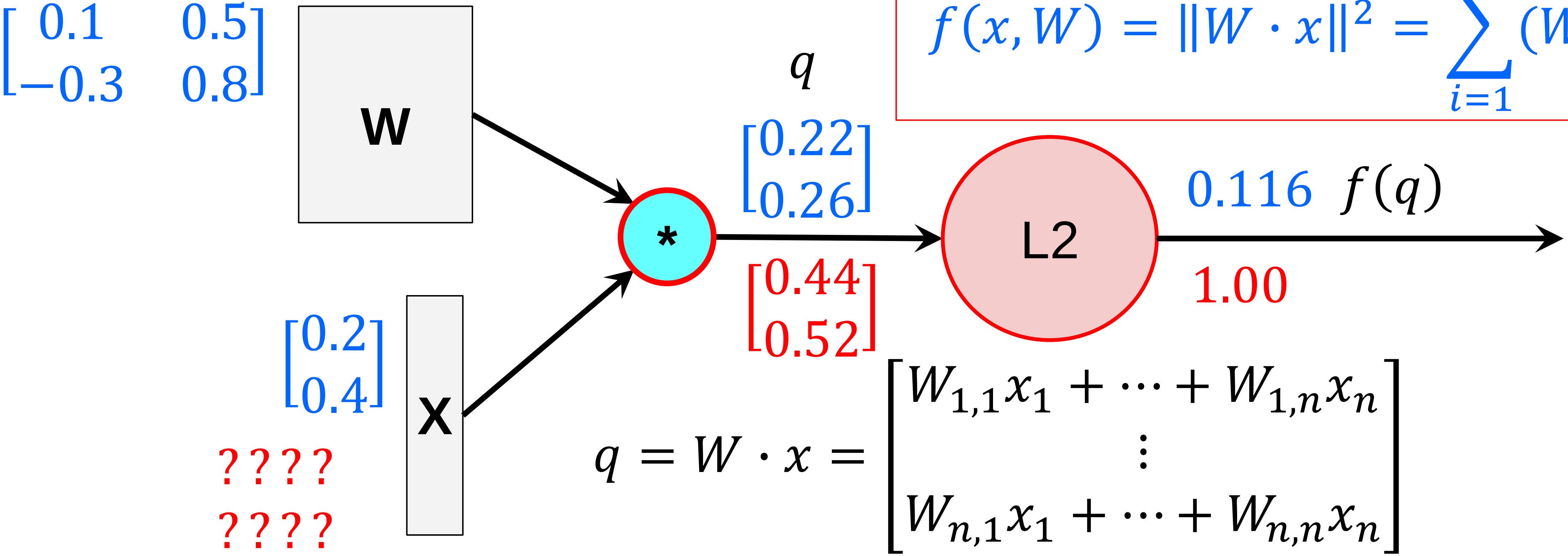
$$\frac{\partial L}{\partial W} = \frac{\partial L}{\partial q} \cdot x^{\top}$$

+ Đạo hàm của L đối với x:

$$\frac{\partial L}{\partial x} = W^{\top} \cdot \frac{\partial L}{\partial q}$$

A vectorized example

$$f(x, W) = \|W \cdot x\|^2 = \sum_{i=1}^{n} (W \cdot x)_i^2$$



$$q = W \cdot x = \begin{bmatrix} W_{1,1}x_1 + \cdots + W_{1,n}x_n \\ \vdots \\ W_{n,1}x_1 + \cdots + W_{n,n}x_n \end{bmatrix}$$

A vectorized example

– A vectorized example:

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

$$f(x,W) = \sum_{i=1}^{n} (W \cdot x)_i^2$$

+ $x = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} \in \mathbb{R}^n$

+ $x^T = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \in \mathbb{R}^n$

– Where:

+ $x \in \mathbb{R}^n$

+ $W \in \mathbb{R}^{n \times n}$

+ $W = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1(n-1)} & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2(n-1)} & w_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ w_{n1} & w_{n2} & \dots & w_{n(n-1)} & w_{nn} \end{bmatrix}$

– Want: $\frac{\partial f}{\partial x}$ $\frac{\partial f}{\partial W}$

A vectorized example

$$q = W \cdot x$$

– Want: $\frac{\partial q}{\partial x}$

$$q = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1(n-1)} & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2(n-1)} & w_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ w_{n1} & w_{n2} & \dots & w_{n(n-1)} & w_{nn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}$$

$$q = \begin{bmatrix} w_{1,1}x_1 + \cdots + w_{1,n}x_n \\ \vdots \\ w_{n,1}x_1 + \cdots + w_{n,n}x_n \end{bmatrix} = \begin{bmatrix} q_1 \\ q_2 \\ \dots \\ q_n \end{bmatrix}$$

A vectorized example

– We have

$$q = \begin{bmatrix} q_1 \\ \vdots \\ q_n \end{bmatrix} = \begin{bmatrix} w_{1,1}x_1 + \cdots + w_{1,n}x_n \\ \vdots \\ w_{n,1}x_1 + \cdots + w_{n,n}x_n \end{bmatrix}$$

– So that

$$\frac{\partial q}{\partial x_1} = \begin{bmatrix} w_{1,1} \\ w_{2,1} \\ \cdots \\ w_{n,1} \end{bmatrix} \frac{\partial q}{\partial x_2} = \begin{bmatrix} w_{1,2} \\ w_{2,2} \\ \cdots \\ w_{n,2} \end{bmatrix} \frac{\partial q}{\partial x_i} = \begin{bmatrix} w_{1,i} \\ w_{2,i} \\ \cdots \\ w_{n,i} \end{bmatrix} \frac{\partial q}{\partial x_{(n-1)}} = \begin{bmatrix} w_{1,(n-1)} \\ w_{2,(n-1)} \\ \cdots \\ w_{n,(n-1)} \end{bmatrix} \frac{\partial q}{\partial x_n} = \begin{bmatrix} w_{1,n} \\ w_{2,n} \\ \cdots \\ w_{n,n} \end{bmatrix}$$

A vectorized example

– So that

$$\frac{\partial q}{\partial x_1} = \begin{bmatrix} w_{1,1} \\ w_{2,1} \\ \cdots \\ w_{n,1} \end{bmatrix} \frac{\partial q}{\partial x_2} = \begin{bmatrix} w_{1,2} \\ w_{2,2} \\ \cdots \\ w_{n,2} \end{bmatrix} \frac{\partial q}{\partial x_i} = \begin{bmatrix} w_{1,i} \\ w_{2,i} \\ \cdots \\ w_{n,i} \end{bmatrix} \frac{\partial q}{\partial x_{(n-1)}} = \begin{bmatrix} w_{1,(n-1)} \\ w_{2,(n-1)} \\ \cdots \\ w_{n,(n-1)} \end{bmatrix} \frac{\partial q}{\partial x_n} = \begin{bmatrix} w_{1,n} \\ w_{2,n} \\ \cdots \\ w_{n,n} \end{bmatrix}$$

– Thus,

$$\frac{\partial q_k}{\partial x_i} = w_{k,i}$$

A vectorized example

– So we have:

$$\frac{\partial q_k}{\partial x_i} = w_{k,i}$$

– We have:

$$\frac{\partial f}{\partial x_i} = \sum_{k=1}^{n} \frac{\partial f}{\partial q_k} \frac{\partial q_k}{\partial x_i} = \sum_{k=1}^{n} (2q_k)(w_{k,i})$$

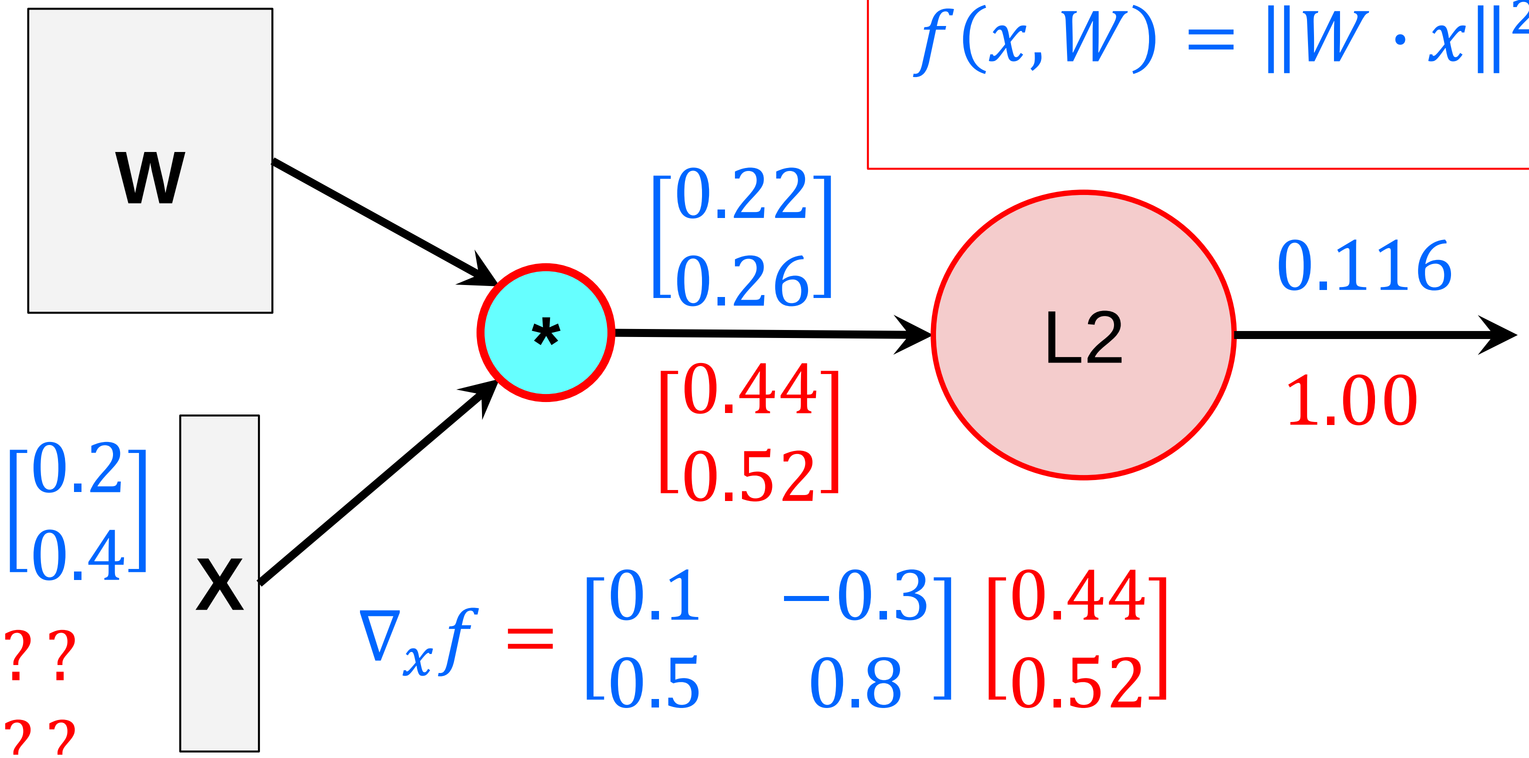
– Thus:

$$\nabla_x f = W^T \cdot \nabla_q f$$

A vectorized example

$$f(x,W) = \|W \cdot x\|^2 = \sum_{i=1}^{n} (W \cdot x)_i^2$$

$\begin{bmatrix} 0.1 & 0.5 \\ -0.3 & 0.8 \end{bmatrix}$



$\begin{bmatrix} 0.2 \\ 0.4 \end{bmatrix}$

? ? ? ?
? ? ? ?

$\begin{bmatrix} 0.22 \\ 0.26 \end{bmatrix}$

$\begin{bmatrix} 0.44 \\ 0.52 \end{bmatrix}$

0.116

1.00

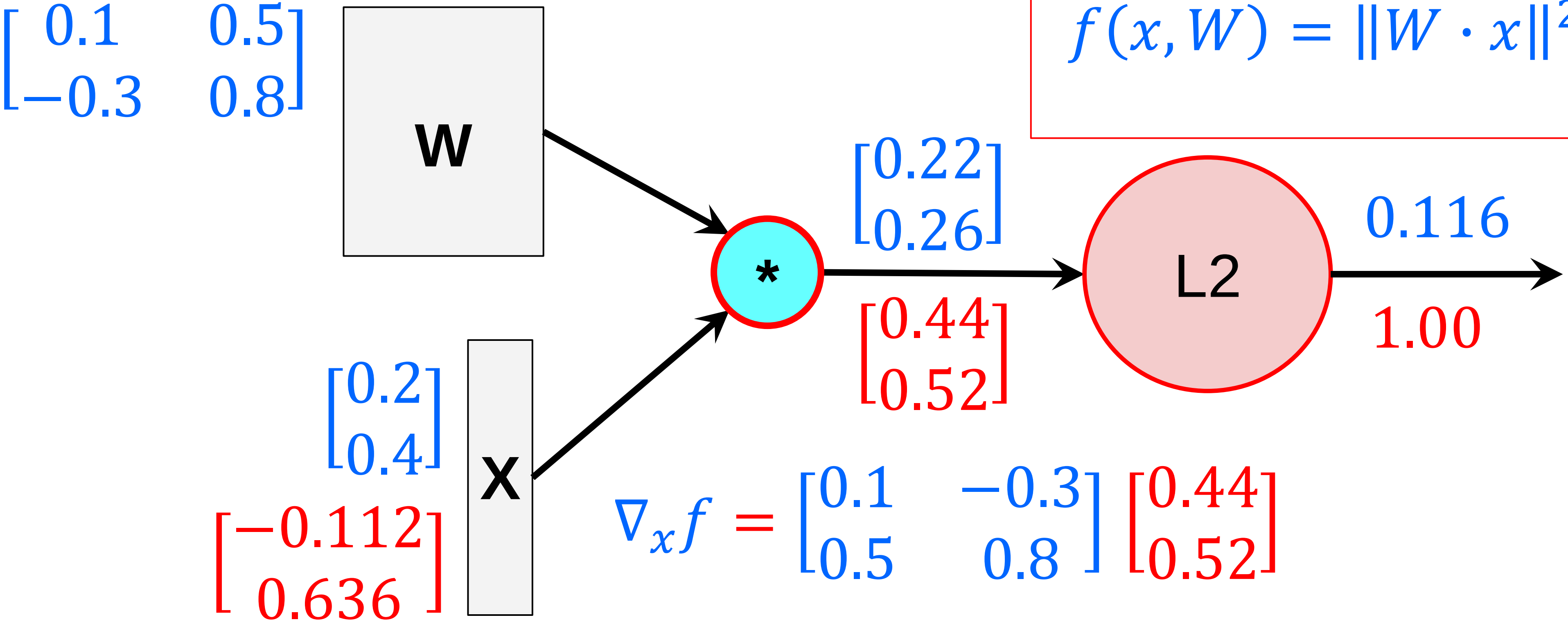
$$\nabla_x f = \begin{bmatrix} 0.1 & -0.3 \\ 0.5 & 0.8 \end{bmatrix} \begin{bmatrix} 0.44 \\ 0.52 \end{bmatrix}$$

$$\frac{\partial q_k}{\partial x_i} = W_{k,i}$$

$$\nabla_x f = W^T \cdot \nabla_q f$$

A vectorized example

$$f(x, W) = \|W \cdot x\|^2 = \sum_{i=1}^{n} (W \cdot x)_i^2$$



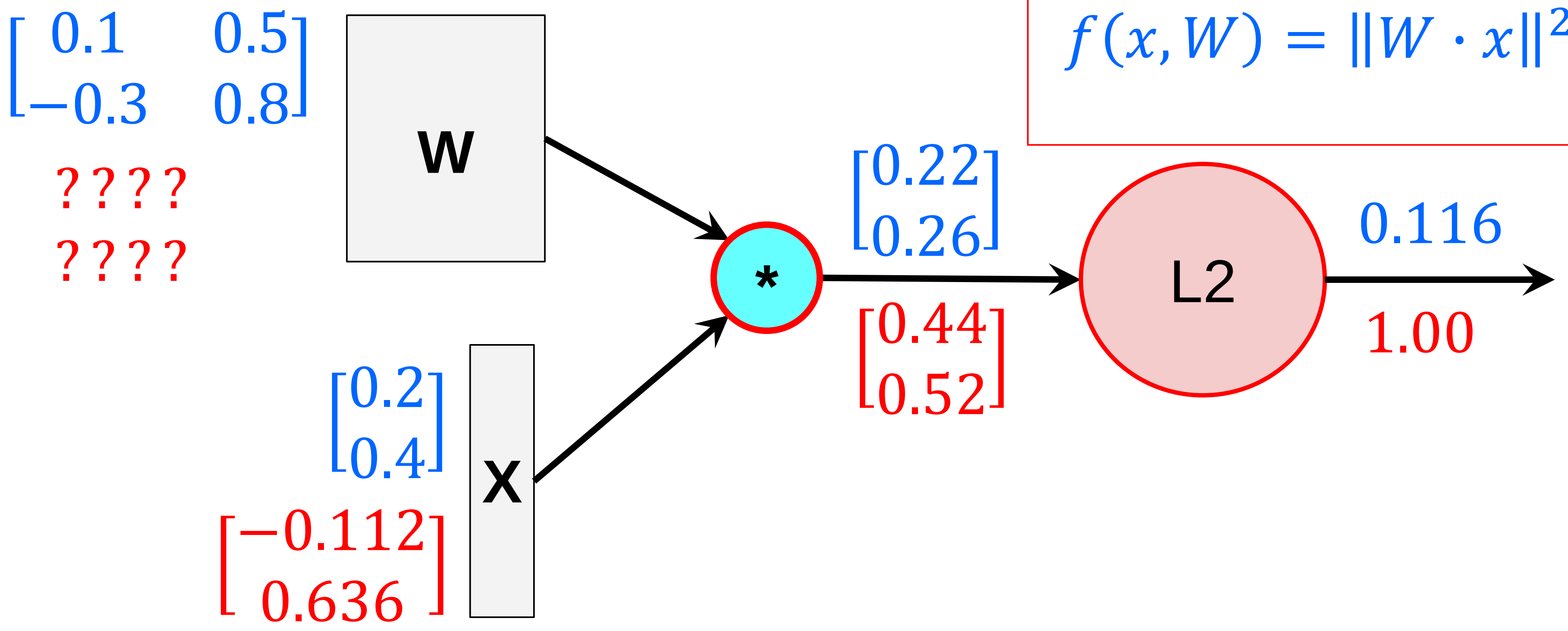
$$\nabla_x f = \begin{bmatrix} 0.1 & -0.3 \\ 0.5 & 0.8 \end{bmatrix} \begin{bmatrix} 0.44 \\ 0.52 \end{bmatrix}$$

$$\frac{\partial q_k}{\partial x_i} = w_{k,i}$$

$$\nabla_x f = W^T \cdot \nabla_q f$$

A vectorized example

$$f(x, W) = \|W \cdot x\|^2 = \sum_{i=1}^{n} (W \cdot x)_i^2$$



A vectorized example

– A vectorized example:

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

$$f(x, W) = \sum_{i=1}^{n} (W \cdot x)_i^2$$

– Where:

+ $x \in \mathbb{R}^n$

+ $W \in \mathbb{R}^{n \times n}$

– Want: $\frac{\partial f}{\partial x}$ $\frac{\partial f}{\partial W}$

+ $x = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} \in \mathbb{R}^n$

+ $x^T = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \in \mathbb{R}^n$

+ $W = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1(n-1)} & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2(n-1)} & w_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ w_{n1} & w_{n2} & \dots & w_{n(n-1)} & w_{nn} \end{bmatrix}$

A vectorized example

$$q = W \cdot x$$

– Want: $\frac{\partial q}{\partial W}$

$$q = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1(n-1)} & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2(n-1)} & w_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ w_{n1} & w_{n2} & \dots & w_{n(n-1)} & w_{nn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}$$

$$q = \begin{bmatrix} w_{1,1}x_1 + \cdots + w_{1,n}x_n \\ \vdots \\ w_{n,1}x_1 + \cdots + w_{n,n}x_n \end{bmatrix}$$

A vectorized example

- $q_1 = w_{1,1}x_1 + w_{1,2}x_2 + \cdots + w_{1,n}x_n$

$$\frac{\partial q_1}{\partial w_{1,1}} = x_1,\ \frac{\partial q_1}{\partial w_{1,2}} = x_2,\ \ldots,\ \frac{\partial q_1}{\partial w_{1,n}} = x_n$$

$$\frac{\partial q_1}{\partial w_{2,1}} = 0,\ \frac{\partial q_1}{\partial w_{2,2}} = 0,\ \ldots,\ \frac{\partial q_1}{\partial w_{2,n}} = 0$$

… … … … … … … … … … … … … … … … … … …

$$\frac{\partial q_1}{\partial w_{(n-1),1}} = 0,\ \frac{\partial q_1}{\partial w_{(n-1),2}} = 0,\ \ldots,\ \frac{\partial q_1}{\partial w_{(n-1),n}} = 0$$

$$\frac{\partial q_1}{\partial w_{n,1}} = 0,\ \frac{\partial q_1}{\partial w_{n,2}} = 0,\ \ldots,\ \frac{\partial q_1}{\partial w_{n,n}} = 0$$

A vectorized example

$- q_2 = w_{2,1}x_1 + w_{2,2}x_2 + \cdots + w_{2,n}x_n$

$\frac{\partial q_2}{\partial w_{1,1}} = 0,\ \frac{\partial q_2}{\partial w_{1,2}} = 0,\ \ldots,\ \frac{\partial q_2}{\partial w_{1,n}} = 0$

$\frac{\partial q_2}{\partial w_{2,1}} = x_1,\ \frac{\partial q_2}{\partial w_{2,2}} = x_2,\ \ldots,\ \frac{\partial q_2}{\partial w_{2,n}} = x_n$

… … … … … … … … … … … … … … … … … … …

$\frac{\partial q_2}{\partial w_{(n-1),1}} = 0,\ \frac{\partial q_2}{\partial w_{(n-1),2}} = 0,\ \ldots,\ \frac{\partial q_2}{\partial w_{(n-1),n}} = 0$

$\frac{\partial q_2}{\partial w_{n,1}} = 0,\ \frac{\partial q_2}{\partial w_{n,2}} = 0,\ \ldots,\ \frac{\partial q_2}{\partial w_{n,n}} = 0$

A vectorized example

$$- \; q_k = w_{k,1}x_1 + w_{k,2}x_2 + \cdots + w_{k,n}x_n$$

$$\frac{\partial q_k}{\partial w_{1,1}} = 0, \; \frac{\partial q_k}{\partial w_{1,2}} = 0, \; \dots, \; \frac{\partial q_k}{\partial w_{1,n}} = 0$$

… … … … … … … … … … … … … … … … … … …

$$\frac{\partial q_k}{\partial w_{k,1}} = x_1, \; \frac{\partial q_k}{\partial w_{k,2}} = x_2, \; \dots, \; \frac{\partial q_k}{\partial w_{k,n}} = x_n$$

… … … … … … … … … … … … … … … … … … …

$$\frac{\partial q_k}{\partial w_{n,1}} = 0, \; \frac{\partial q_k}{\partial w_{n,2}} = 0, \; \dots, \; \frac{\partial q_k}{\partial w_{n,n}} = 0$$

A vectorized example

- $q_{(n-1)} = w_{(n-1),1}x_1 + w_{(n-1),2}x_2 + \cdots + w_{(n-1),n}x_n$

$$\frac{\partial q_{(n-1)}}{\partial w_{1,1}} = 0, \quad \frac{\partial q_{(n-1)}}{\partial w_{1,2}} = 0, \quad \ldots, \quad \frac{\partial q_{(n-1)}}{\partial w_{1,n}} = 0$$

$$\frac{\partial q_{(n-1)}}{\partial w_{2,1}} = 0, \quad \frac{\partial q_{(n-1)}}{\partial w_{2,2}} = 0, \quad \ldots, \quad \frac{\partial q_{(n-1)}}{\partial w_{2,n}} = 0$$

… … … … … … … … … … … … … … … … … … …

$$\frac{\partial q_{(n-1)}}{\partial w_{(n-1),1}} = x_1, \quad \frac{\partial q_{(n-1)}}{\partial w_{(n-1),2}} = x_2, \quad \ldots, \quad \frac{\partial q_{(n-1)}}{\partial w_{(n-1),n}} = x_n$$

$$\frac{\partial q_{(n-1)}}{\partial w_{n,1}} = 0, \quad \frac{\partial q_{(n-1)}}{\partial w_{n,2}} = 0, \quad \ldots, \quad \frac{\partial q_{(n-1)}}{\partial w_{n,n}} = 0$$

A vectorized example

- $q_n = w_{n,1}x_1 + w_{n,2}x_2 + \cdots + w_{n,n}x_n$

$$\frac{\partial q_n}{\partial w_{1,1}} = 0,\ \frac{\partial q_n}{\partial w_{1,2}} = 0,\ \ldots,\ \frac{\partial q_n}{\partial w_{1,n}} = 0$$

$$\frac{\partial q_n}{\partial w_{2,1}} = 0,\ \frac{\partial q_n}{\partial w_{2,2}} = 0,\ \ldots,\ \frac{\partial q_n}{\partial w_{2,n}} = 0$$

… … … … … … … … … … … … … … … … … … …

$$\frac{\partial q_n}{\partial w_{(n-1),1}} = 0,\ \frac{\partial q_n}{\partial w_{(n-1),2}} = 0,\ \ldots,\ \frac{\partial q_n}{\partial w_{(n-1),n}} = 0$$

$$\frac{\partial q_n}{\partial w_{n,1}} = x_1,\ \frac{\partial q_n}{\partial w_{n,2}} = x_2,\ \ldots,\ \frac{\partial q_n}{\partial w_{n,n}} = x_n$$

A vectorized example

$$q_k = w_{k,1}x_1 + w_{k,2}x_2 + \cdots + w_{k,n}x_n$$

$$\frac{\partial q_k}{\partial w_{k,1}} = x_1, \qquad \frac{\partial q_k}{\partial w_{k,2}} = x_2, \qquad \dots, \qquad \frac{\partial q_k}{\partial w_{k,n}} = x_n$$

– So we have:

$$\frac{\partial q_k}{\partial w_{i,j}} = \begin{cases} x_j & if\ k = i \\ 0 & if\ k \neq i \end{cases} = 1_{k=i}x_j$$

– We have:

$$\frac{\partial f}{\partial w_{i,j}} = \sum_{k=1}^{n} \frac{\partial f}{\partial q_k}\frac{\partial q_k}{\partial w_{i,j}} = \sum_{k=1}^{n}(2q_k)(1_{k=i}x_j) = 2q_i x_j$$

A vectorized example

– We have:

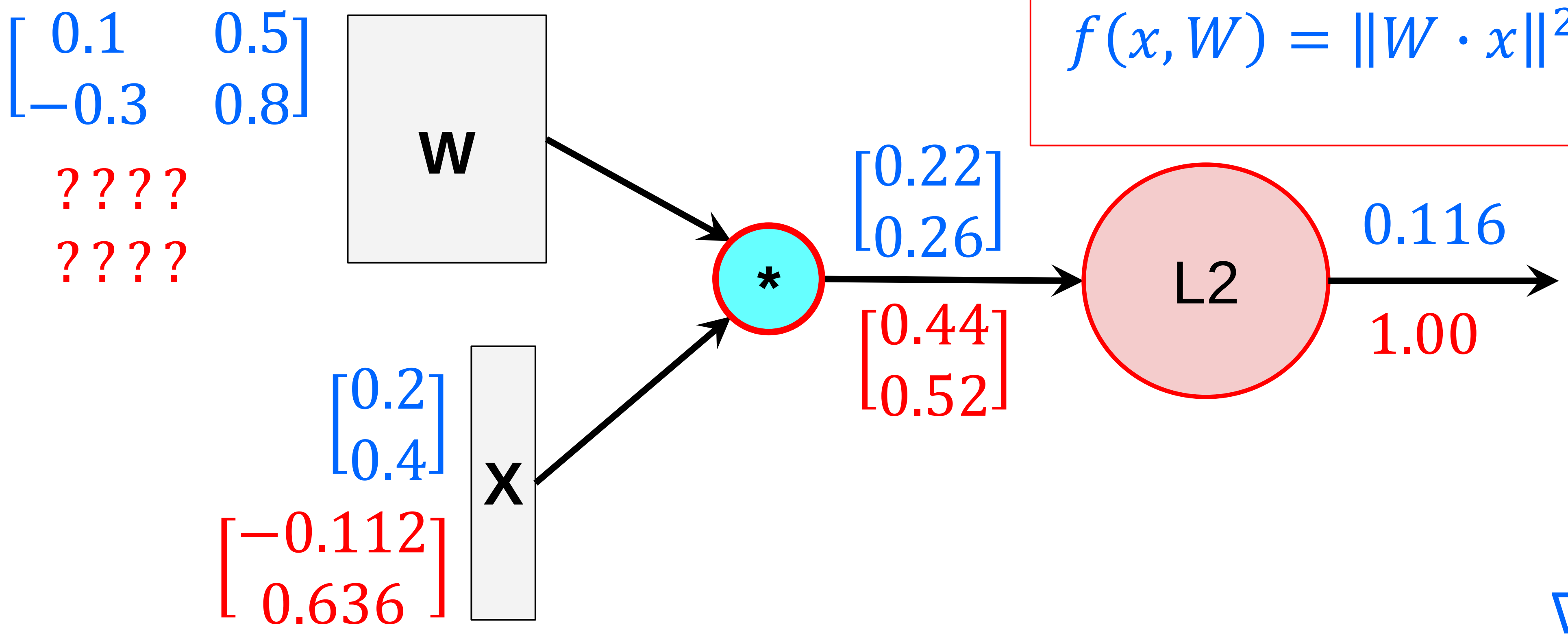
$$\frac{\partial f}{\partial w_{i,j}} = \sum_{k=1}^{n} \frac{\partial f}{\partial q_k} \frac{\partial q_k}{\partial w_{i,j}} = \sum_{k=1}^{n} (2q_k)\left(1_{k=i} x_j\right) = 2q_i x_j$$

– Thus:

$$\nabla_W f = \nabla_q f \cdot x^T$$

A vectorized example

$$f(x, W) = \|W \cdot x\|^2 = \sum_{i=1}^{n} (W \cdot x)_i^2$$

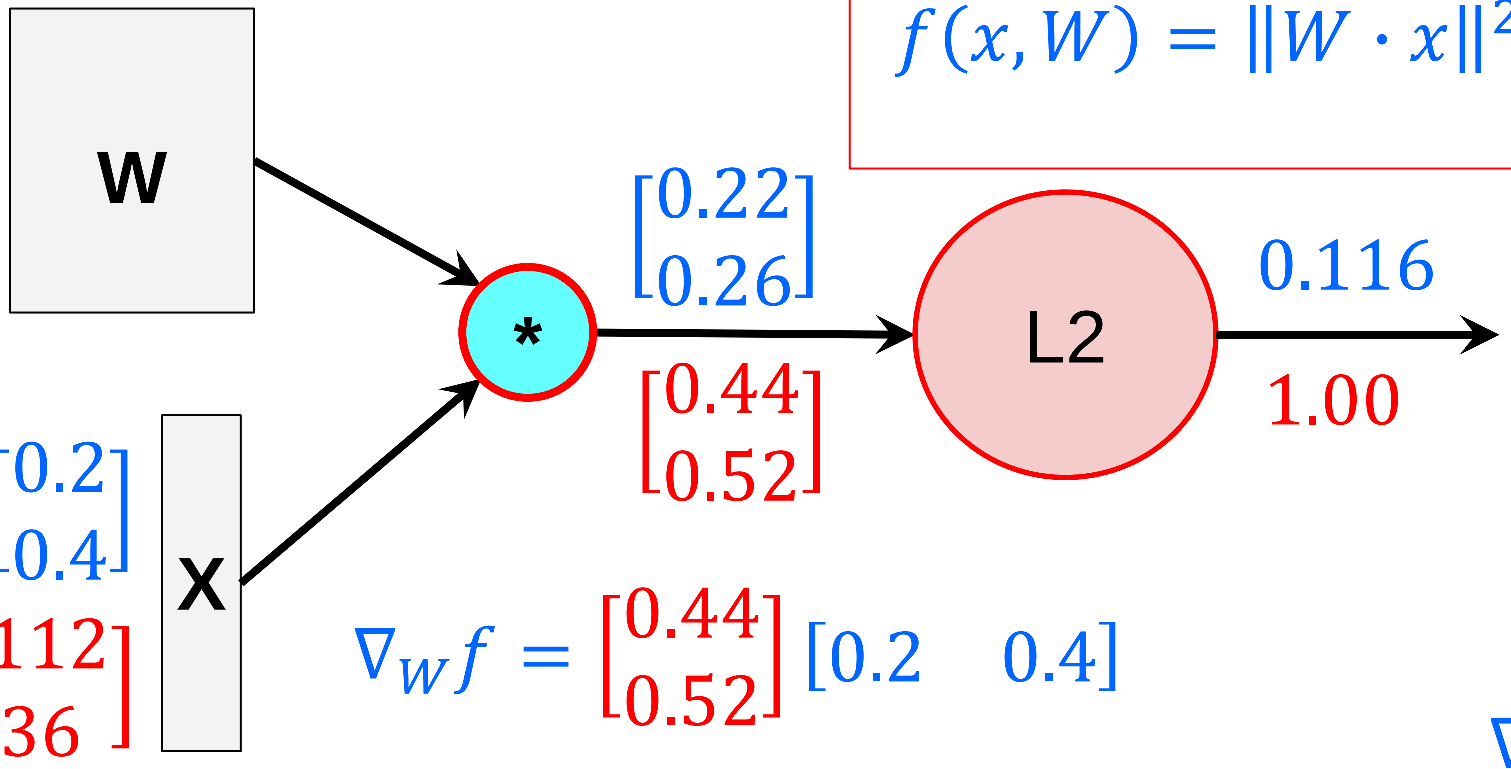


$$\frac{\partial f}{\partial w_{i,j}} = 2 q_i x_j$$

$$\nabla_W f = \nabla_q f \cdot x^T$$

A vectorized example

$$f(x, W) = \|W \cdot x\|^2 = \sum_{i=1}^{n} (W \cdot x)_i^2$$



$$\begin{bmatrix} 0.1 & 0.5 \\ -0.3 & 0.8 \end{bmatrix}$$

? ? ? ?
? ? ? ?

$$\begin{bmatrix} 0.2 \\ 0.4 \end{bmatrix}$$

$$\begin{bmatrix} -0.112 \\ 0.636 \end{bmatrix}$$

$$\begin{bmatrix} 0.22 \\ 0.26 \end{bmatrix}$$

$$\begin{bmatrix} 0.44 \\ 0.52 \end{bmatrix}$$

0.116

1.00

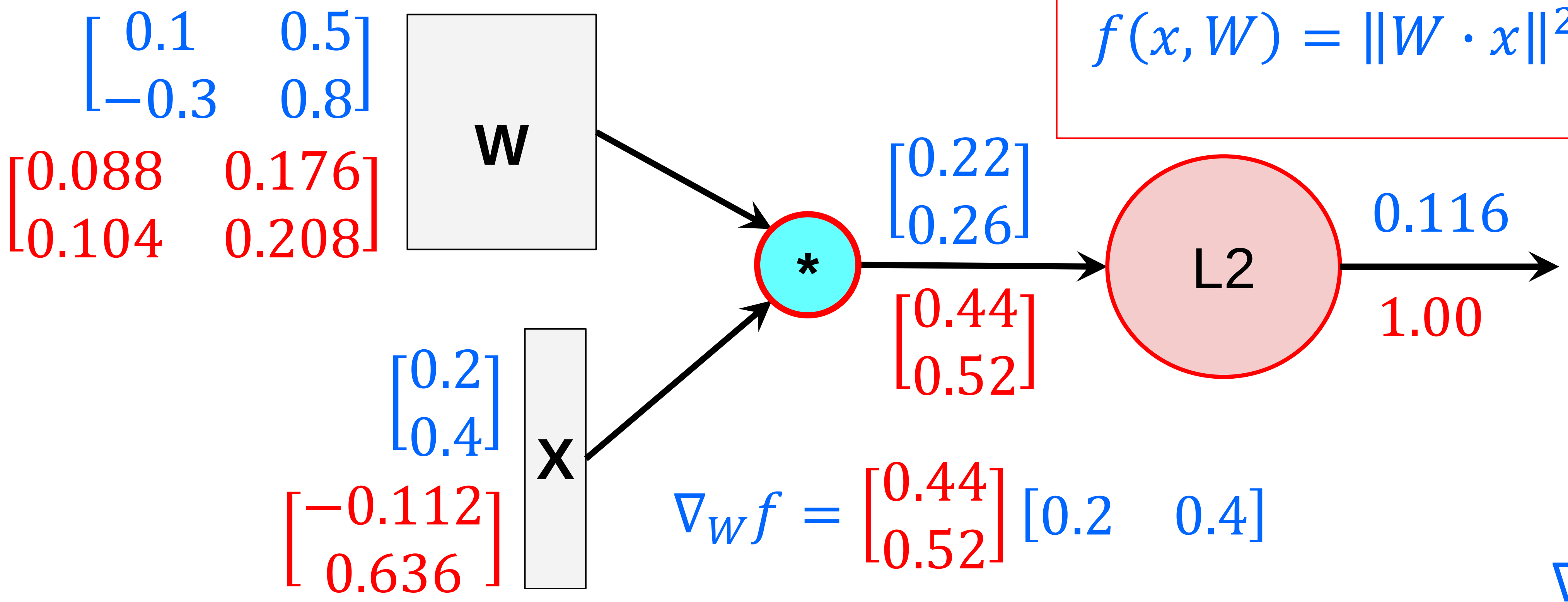
$$\nabla_W f = \begin{bmatrix} 0.44 \\ 0.52 \end{bmatrix} \begin{bmatrix} 0.2 & 0.4 \end{bmatrix}$$

$$\frac{\partial f}{\partial w_{i,j}} = 2 q_i x_j$$

$$\nabla_W f = \nabla_q f \cdot x^T$$

A vectorized example

$$f(x, W) = \|W \cdot x\|^2 = \sum_{i=1}^{n} (W \cdot x)_i^2$$



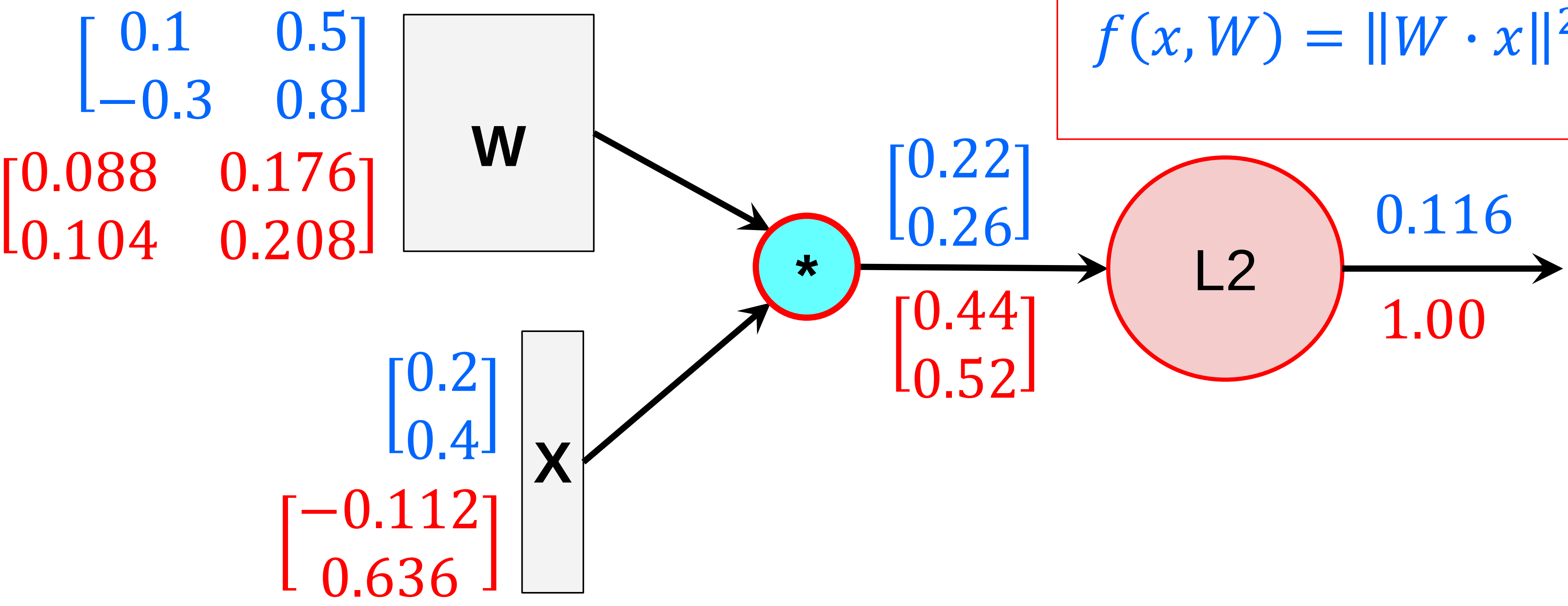
$$\nabla_W f = \begin{bmatrix} 0.44 \\ 0.52 \end{bmatrix} \begin{bmatrix} 0.2 & 0.4 \end{bmatrix}$$

$$\frac{\partial f}{\partial w_{i,j}} = 2 q_i x_j$$

$$\nabla_W f = \nabla_q f \cdot x^T$$

A vectorized example

$$f(x,W) = \|W \cdot x\|^2 = \sum_{i=1}^{n} (W \cdot x)_i^2$$

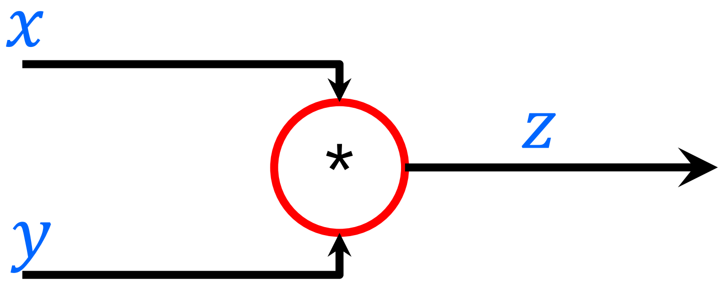


Modularized implementation

– Graph (or Net) object (rough psuedo code)

```
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
```

Modularized implementation



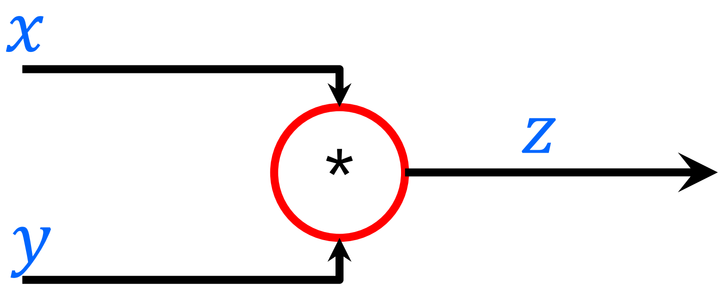
x, y, z are scalars

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

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

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

Modularized implementation



x, y, z are scalars

```
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]
```

Example: Caffe layers

Branch: master ▾ caffe / src / caffe / layers / — Create new file | Upload files | Find file | History

shelhamer committed on GitHub Merge pull request #4630 from BlGene/load_hdf5_fix ... — Latest commit e687a71 21 days ago

..		
absval_layer.cpp	dismantle layer headers	a year ago
absval_layer.cu	dismantle layer headers	a year ago
accuracy_layer.cpp	dismantle layer headers	a year ago
argmax_layer.cpp	dismantle layer headers	a year ago
base_conv_layer.cpp	enable dilated deconvolution	a year ago
base_data_layer.cpp	Using default from proto for prefetch	3 months ago
base_data_layer.cu	Switched multi-GPU to NCCL	3 months ago
batch_norm_layer.cpp	Add missing spaces besides equal signs in batch_norm_layer.cpp	4 months ago
batch_norm_layer.cu	dismantle layer headers	a year ago
batch_reindex_layer.cpp	dismantle layer headers	a year ago
batch_reindex_layer.cu	dismantle layer headers	a year ago
bias_layer.cpp	Remove incorrect cast of gemm int arg to Dtype in BiasLayer	a year ago
bias_layer.cu	Separation and generalization of ChannelwiseAffineLayer into BiasLayer	a year ago
bnll_layer.cpp	dismantle layer headers	a year ago
bnll_layer.cu	dismantle layer headers	a year ago
concat_layer.cpp	dismantle layer headers	a year ago
concat_layer.cu	dismantle layer headers	a year ago
contrastive_loss_layer.cpp	dismantle layer headers	a year ago
contrastive_loss_layer.cu	dismantle layer headers	a year ago
conv_layer.cpp	add support for 2D dilated convolution	a year ago
conv_layer.cu	dismantle layer headers	a year ago
crop_layer.cpp	remove redundant operations in Crop layer (#5138)	2 months ago
crop_layer.cu	remove redundant operations in Crop layer (#5138)	2 months ago
cudnn_conv_layer.cpp	dismantle layer headers	a year ago
cudnn_conv_layer.cu	Add cuDNN v5 support, drop cuDNN v3 support	11 months ago

cudnn_lcn_layer.cpp	dismantle layer headers	a year ago
cudnn_lcn_layer.cu	dismantle layer headers	a year ago
cudnn_lrn_layer.cpp	dismantle layer headers	a year ago
cudnn_lrn_layer.cu	dismantle layer headers	a year ago
cudnn_pooling_layer.cpp	dismantle layer headers	a year ago
cudnn_pooling_layer.cu	dismantle layer headers	a year ago
cudnn_relu_layer.cpp	Add cuDNN v5 support, drop cuDNN v3 support	11 months ago
cudnn_relu_layer.cu	Add cuDNN v5 support, drop cuDNN v3 support	11 months ago
cudnn_sigmoid_layer.cpp	Add cuDNN v5 support, drop cuDNN v3 support	11 months ago
cudnn_sigmoid_layer.cu	Add cuDNN v5 support, drop cuDNN v3 support	11 months ago
cudnn_softmax_layer.cpp	dismantle layer headers	a year ago
cudnn_softmax_layer.cu	dismantle layer headers	a year ago
cudnn_tanh_layer.cpp	Add cuDNN v5 support, drop cuDNN v3 support	11 months ago
cudnn_tanh_layer.cu	Add cuDNN v5 support, drop cuDNN v3 support	11 months ago
data_layer.cpp	Switched multi-GPU to NCCL	3 months ago
deconv_layer.cpp	enable dilated deconvolution	a year ago
deconv_layer.cu	dismantle layer headers	a year ago
dropout_layer.cpp	supporting N-D Blobs in Dropout layer Reshape	a year ago
dropout_layer.cu	dismantle layer headers	a year ago
dummy_data_layer.cpp	dismantle layer headers	a year ago
eltwise_layer.cpp	dismantle layer headers	a year ago
eltwise_layer.cu	dismantle layer headers	a year ago
elu_layer.cpp	ELU layer with basic tests	a year ago
elu_layer.cu	ELU layer with basic tests	a year ago
embed_layer.cpp	dismantle layer headers	a year ago
embed_layer.cu	dismantle layer headers	a year ago
euclidean_loss_layer.cpp	dismantle layer headers	a year ago
euclidean_loss_layer.cu	dismantle layer headers	a year ago
exp_layer.cpp	Solving issue with exp layer with base e	a year ago
exp_layer.cu	dismantle layer headers	a year ago

Caffe Sigmoid Layer

```
#include <cmath>
#include <vector>

#include "caffe/layers/sigmoid_layer.hpp"

namespace caffe {

template <typename Dtype>
inline Dtype sigmoid(Dtype x) {
  return 1. / (1. + exp(-x));
}

template <typename Dtype>
void SigmoidLayer<Dtype>::Forward_cpu(const vector<Blob<Dtype>*>& bottom,
    const vector<Blob<Dtype>*>& top) {
  const Dtype* bottom_data = bottom[0]->cpu_data();
  Dtype* top_data = top[0]->mutable_cpu_data();
  const int count = bottom[0]->count();
  for (int i = 0; i < count; ++i) {
    top_data[i] = sigmoid(bottom_data[i]);
  }
}

template <typename Dtype>
void SigmoidLayer<Dtype>::Backward_cpu(const vector<Blob<Dtype>*>& top,
    const vector<bool>& propagate_down,
    const vector<Blob<Dtype>*>& bottom) {
  if (propagate_down[0]) {
    const Dtype* top_data = top[0]->cpu_data();
    const Dtype* top_diff = top[0]->cpu_diff();
    Dtype* bottom_diff = bottom[0]->mutable_cpu_diff();
    const int count = bottom[0]->count();
    for (int i = 0; i < count; ++i) {
      const Dtype sigmoid_x = top_data[i];
      bottom_diff[i] = top_diff[i] * sigmoid_x * (1. - sigmoid_x);
    }
  }
}

#ifdef CPU_ONLY
STUB_GPU(SigmoidLayer);
#endif

INSTANTIATE_CLASS(SigmoidLayer);


}  // namespace caffe
```

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

$$(1-\sigma(x))\sigma(x) \quad * \; top\_diff$$ (chain rule)

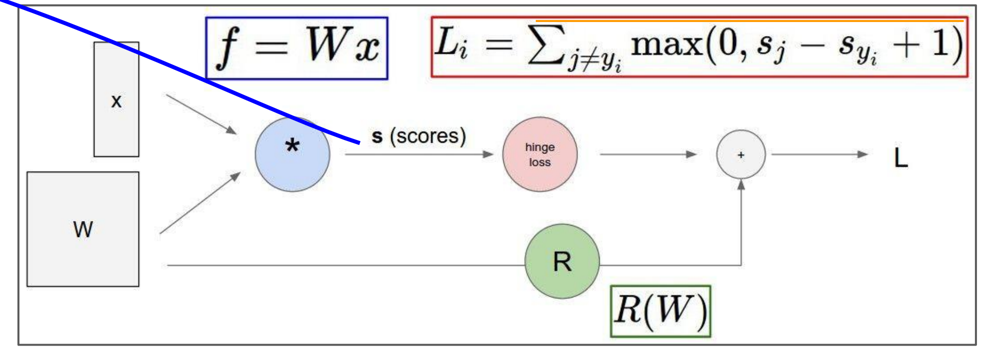
In Assignment 1: Writing SVM / Softmax

– Stage your forward/backward computation!

– E.g. for the SVM:

margins

```
# receive W (weights), X (data)
# forward pass (we have 8 lines)
scores = #...
margins = #...
data_loss = #...
reg_loss = #...
loss = data_loss + reg_loss
# backward pass (we have 5 lines)
dmargins = # ... (optionally, we go direct to dscores)
dscores = #...
dW = #...
```



Summary so far...

- Neural nets will be very large: impractical to write down gradient formula by hand for all parameters
- Mạng lưới nơ ron rất lớn: việc viết công thức gradient bằng tay cho tất cả các tham số là không thực tế
- Backpropagation = recursive application of the chain rule along a computational graph to compute the gradients of all inputs/parameters/intermediates
- Lan truyền ngược = ứng dụng đệ quy của quy tắc chuỗi dọc theo biểu đồ tính toán để tính toán độ dốc của tất cả đầu vào/tham số/trung gian.

Summary so far...

- Implementations maintain a graph structure, where the nodes implement the forward() / backward() API
- Việc triển khai đồ thị tính toán, trong đó các nút forward() và nút backward();

Summary so far...

- forward: compute result of an operation and save any intermediates needed for gradient computation in memory
- Truyền tiến: tính toán kết quả của một thao tác và lưu mọi giá trị trung gian cần thiết để tính toán đạo hàm trong bộ nhớ
- backward: apply the chain rule to compute the gradient of the loss function with respect to the inputs
- Truyền ngược: áp dụng quy tắc chuỗi để tính đạo hàm của hàm mất mát đối với từng đầu vào

Next: Neural Networks

Neural networks: without the brain stuff

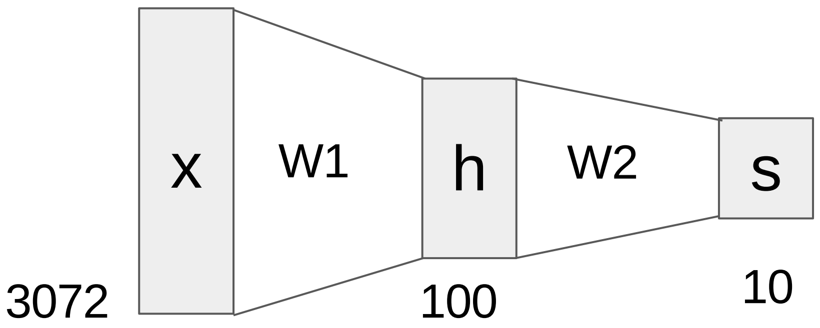
– (Before) Linear score function: $f = Wx$

Neural networks: without the brain stuff

- (Before) Linear score function: $f = Wx$
- (Now) 2-layer Neural Network: $f = W_2 max(0, W_1 x)$

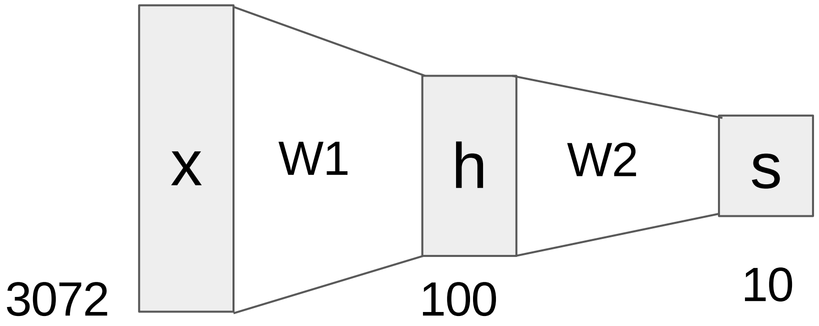
Neural networks: without the brain stuff

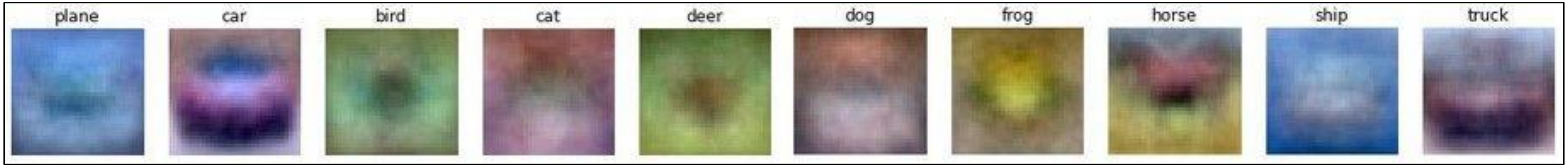
- (Before) Linear score function: $f = Wx$
- (Now) 2-layer Neural Network: $f = W_2 max(0, W_1 x)$



Neural networks: without the brain stuff

- (Before) Linear score function: $f = Wx$
- (Now) 2-layer Neural Network: $f = W_2 max(0, W_1 x)$





Neural networks: without the brain stuff

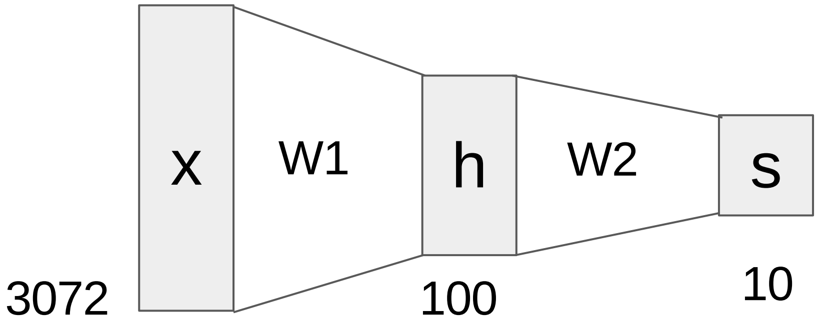
- (Before) Linear score function: $f = Wx$
- (Now) 2-layer Neural Network: $f = W_2 max(0, W_1 x)$

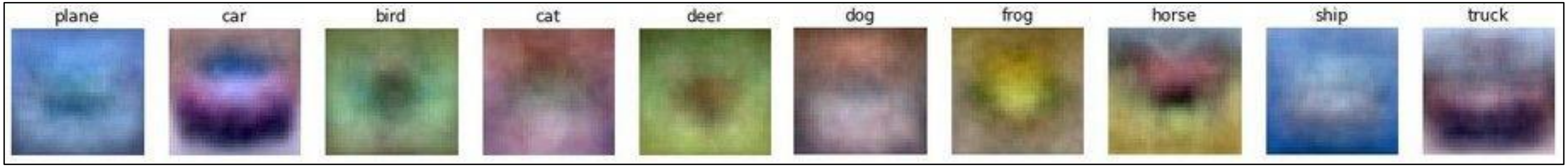
  or 3-layer Neural Network

$$f = W_3 max\big(0, W_2 max(0, W_1 x)\big)$$

Neural networks: without the brain stuff

- (Before) Linear score function: $f = Wx$
- (Now) 2-layer Neural Network: $f = W_2 max(0, W_1 x)$

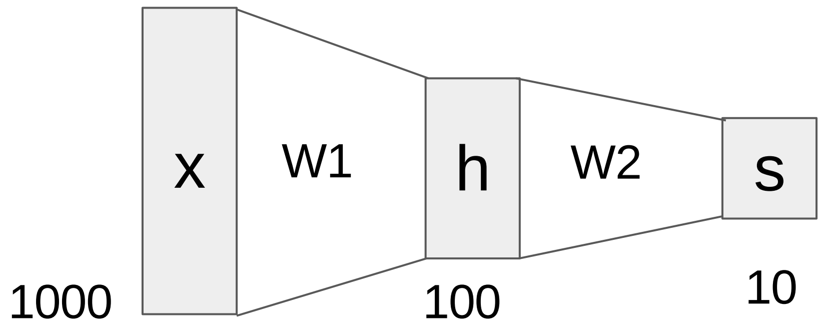




2-layer Neural Network

– Giả thiết:
  - + $n = 64.$
  - + $d_{in} = 1000.$
  - + $h = 100.$
  - + $d_{out} = 10.$
  - + $X(64 \times 1000)$
  - + $Y(64 \times 10)$
  - + $W_1(1000 \times 100)$
  - + $W_2(100 \times 10)$

– Full implementation of training a 2-layer Neural Network needs ~20 lines.



2-layer Neural Network

– Truyền tiến:

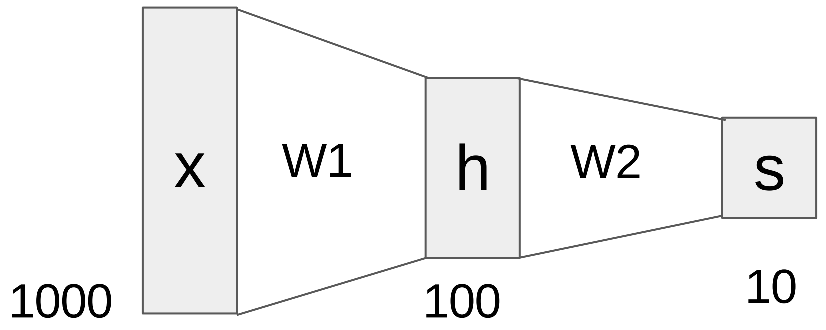
+ $q = X \cdot W_1$ (64×100)

+ $h = \frac{1}{1+e^{-q}}$ (64×100)

+ $\hat{Y} = h \cdot W_2$ (64×10)

+ $L = \sum(\hat{Y} - Y)^2$ $(scalar)$

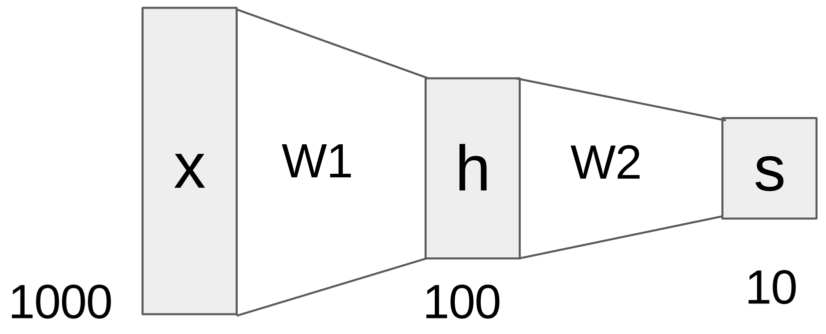
– Full implementation of training a 2-layer Neural Network needs ~20 lines.



2-layer Neural Network

- Truyền ngược:
  + $L = \sum(\hat{Y} - Y)^2 \qquad (scalar)$
  + $d\hat{Y} = 2 \cdot (\hat{Y} - Y)$
  + $\frac{\partial L}{\partial \hat{Y}} = 2 \cdot (\hat{Y} - Y) \cdot (\hat{Y} - Y)'$
  + $\frac{\partial L}{\partial \hat{Y}} = 2 \cdot (\hat{Y} - Y) \cdot 1$
  + $\frac{\partial L}{\partial \hat{Y}} = 2 \cdot (\hat{Y} - Y)$

- Full implementation of training a 2-layer Neural Network needs ~20 lines.



2-layer Neural Network

– Truyền ngược:

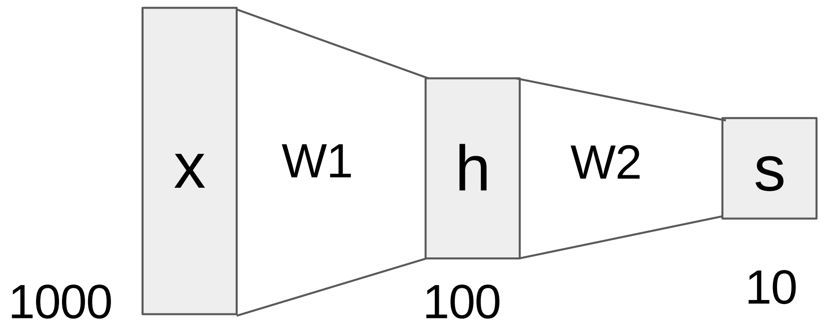
$+\ \hat{Y} = h \cdot W_2 \qquad (64 \times 10)$

$+\ dW_2 = h^\top \cdot d\hat{Y}$

$+\ \frac{\partial L}{\partial W_2} = \frac{\partial \hat{Y}}{\partial W_2} \cdot \frac{\partial L}{\partial \hat{Y}}$

$+\ \frac{\partial L}{\partial W_2} = h^\top \cdot \frac{\partial L}{\partial \hat{Y}}$

– Full implementation of training a 2-layer Neural Network needs ~20 lines.



2-layer Neural Network

– Truyền ngược:

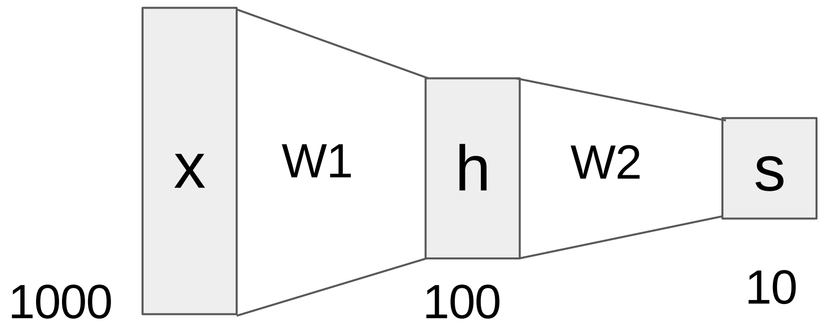
+ $\hat{Y} = h \cdot W_2$ (64×10)

+ $dh = d\hat{Y} \cdot {W_2}^{\top}$

+ $\frac{\partial L}{\partial h} = \frac{\partial L}{\partial \hat{Y}} \cdot \frac{\partial \hat{Y}}{\partial h}$

+ $\frac{\partial L}{\partial h} = \frac{\partial L}{\partial \hat{Y}} \cdot {W_2}^{\top}$

– Full implementation of training a 2-layer Neural Network needs ~20 lines.



2-layer Neural Network

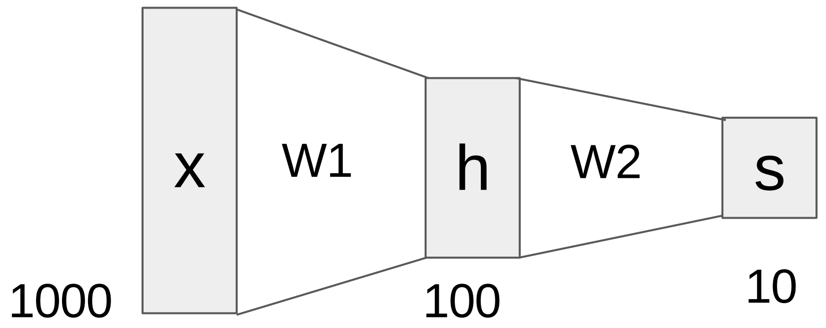
– Truyền ngược:

+ $\hat{Y} = h \cdot W_2$ (64×10)

+ $dW_2 = h^{\top} \cdot d\hat{Y}$

+ $dh = d\hat{Y} \cdot {W_2}^{\top}$

– Full implementation of training a 2-layer Neural Network needs ~20 lines.



2-layer Neural Network

– Truyền ngược:

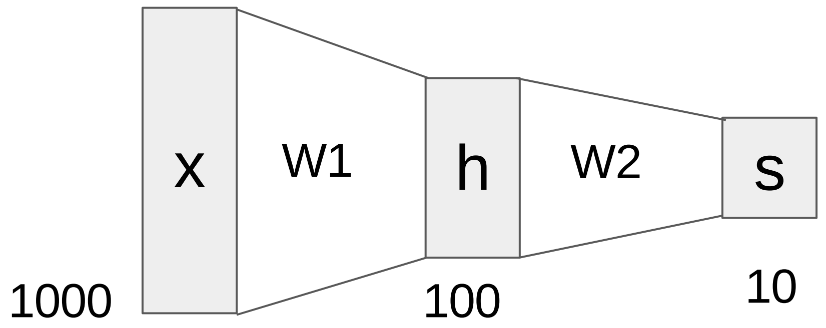
+ $h = \frac{1}{1+e^{-q}}$ (64×100)

+ $dq = dh \cdot (1 - h) \cdot h$

+ $\frac{\partial L}{\partial q} = \frac{\partial L}{\partial h} \cdot \frac{\partial h}{\partial q}$

+ $\frac{\partial L}{\partial q} = \frac{\partial L}{\partial h} \cdot (1 - h) \cdot h$

– Full implementation of training a 2-layer Neural Network needs ~20 lines.



2-layer Neural Network

– Truyền ngược:

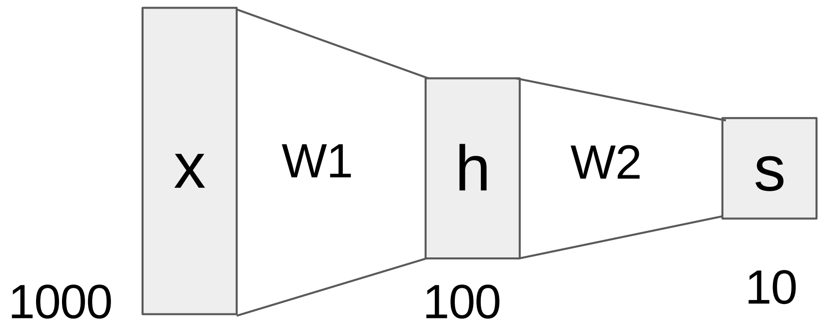
+ $q = X \cdot W_1$ (64×100)

+ $dX = dq \cdot W_1{}^{\top}$

+ $\frac{\partial L}{\partial X} = \frac{\partial L}{\partial q} \cdot \frac{\partial q}{\partial X}$

+ $\frac{\partial L}{\partial X} = \frac{\partial L}{\partial q} \cdot W_1{}^{\top}$

– Full implementation of training a 2-layer Neural Network needs ~20 lines.



2-layer Neural Network

– Truyền ngược:

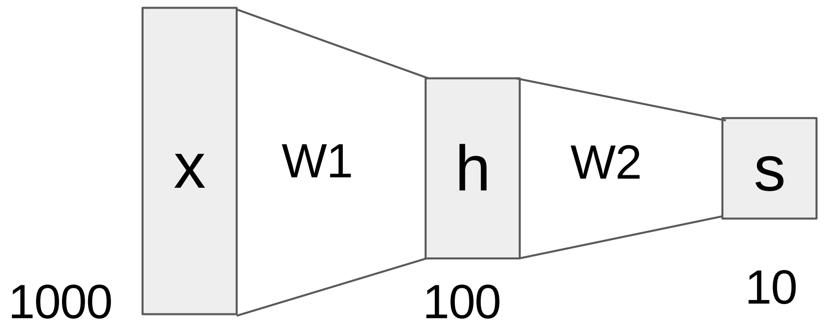
+ $q = X \cdot W_1$ (64×100)

+ $dW_1 = X^\top \cdot dq$

+ $\frac{\partial L}{\partial W_1} = \frac{\partial q}{\partial W_1} \cdot \frac{\partial L}{\partial q}$

+ $\frac{\partial L}{\partial W_1} = X^\top \cdot \frac{\partial L}{\partial q}$

– Full implementation of training a 2-layer Neural Network needs ~20 lines.



2-layer Neural Network

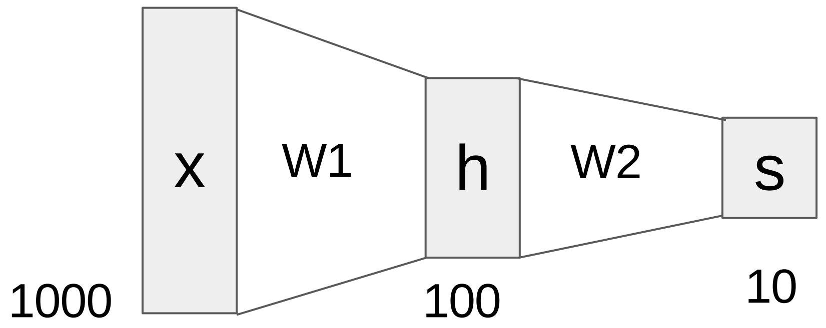
– Truyền ngược:

+ $q = X \cdot W_1$ (64×100)

+ $dX = dq \cdot W_1{}^\top$

+ $dW_1 = X^\top \cdot dq$

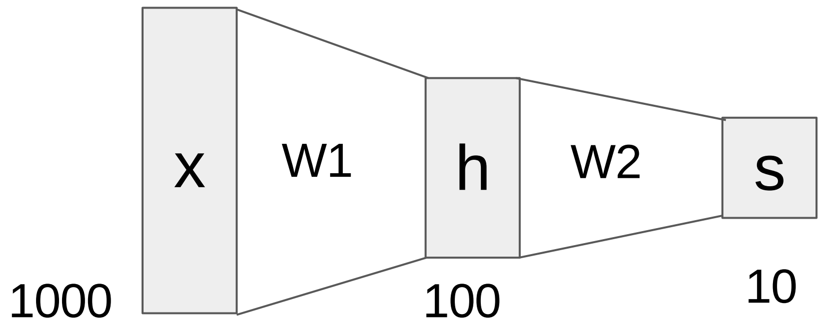
– Full implementation of training a 2-layer Neural Network needs ~20 lines.



2-layer Neural Network

– Cập nhật trọng số:
  + $W_1 -= \eta \cdot dW_1$
  + $W_2 -= \eta \cdot dW_2$

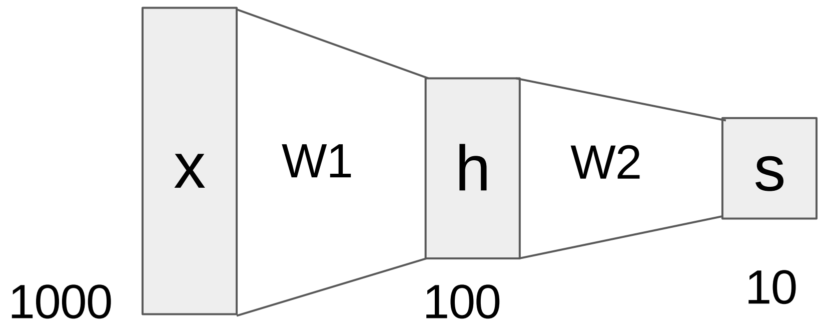
– Full implementation of training a 2-layer Neural Network needs ~20 lines.



2-layer Neural Network

```
import numpy as np
from numpy.random import randn

N, D_in, H, D_out = 64, 1000, 100, 10
x, y = randn(N, D_in), randn(N, D_out)
w1, w2 = randn(D_in, H), randn(H, D_out)

for t in range(2000):
  h = 1 / (1 + np.exp(-x.dot(w1)))
  y_pred = h.dot(w2)
  loss = np.square(y_pred - y).sum()
  print(t, loss)

  grad_y_pred = 2.0 * (y_pred - y)
  grad_w2 = h.T.dot(grad_y_pred)
  grad_h = grad_y_pred.dot(w2.T)
  grad_w1 = x.T.dot(grad_h * h * (1 - h))

  w1 -= 1e-4 * grad_w1
  w2 -= 1e-4 * grad_w2
```

– Full implementation of training a 2-layer Neural Network needs ~20 lines.

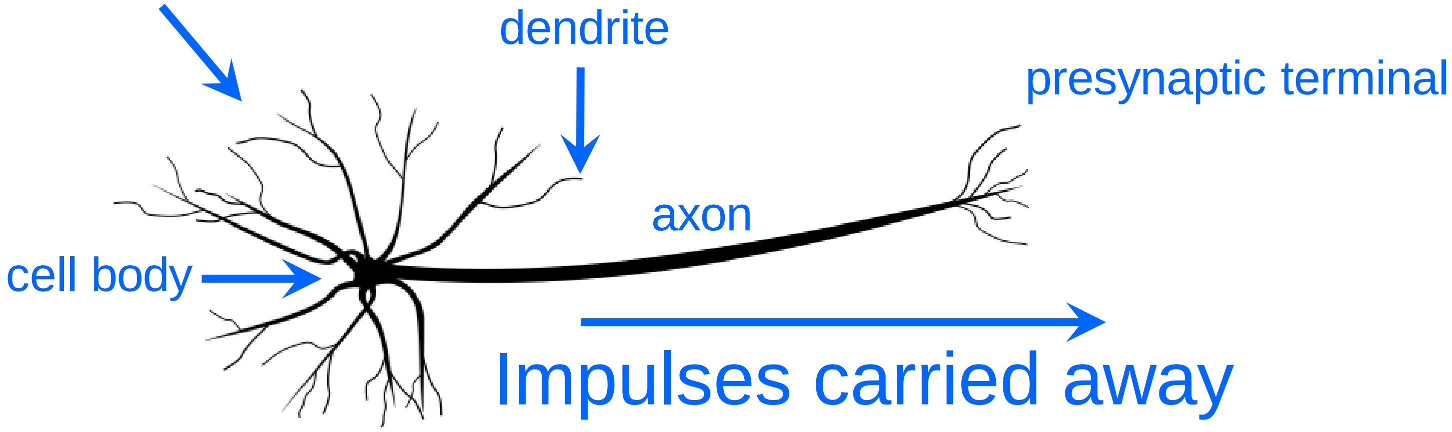


In Assignment 2: Writing a 2-layer net

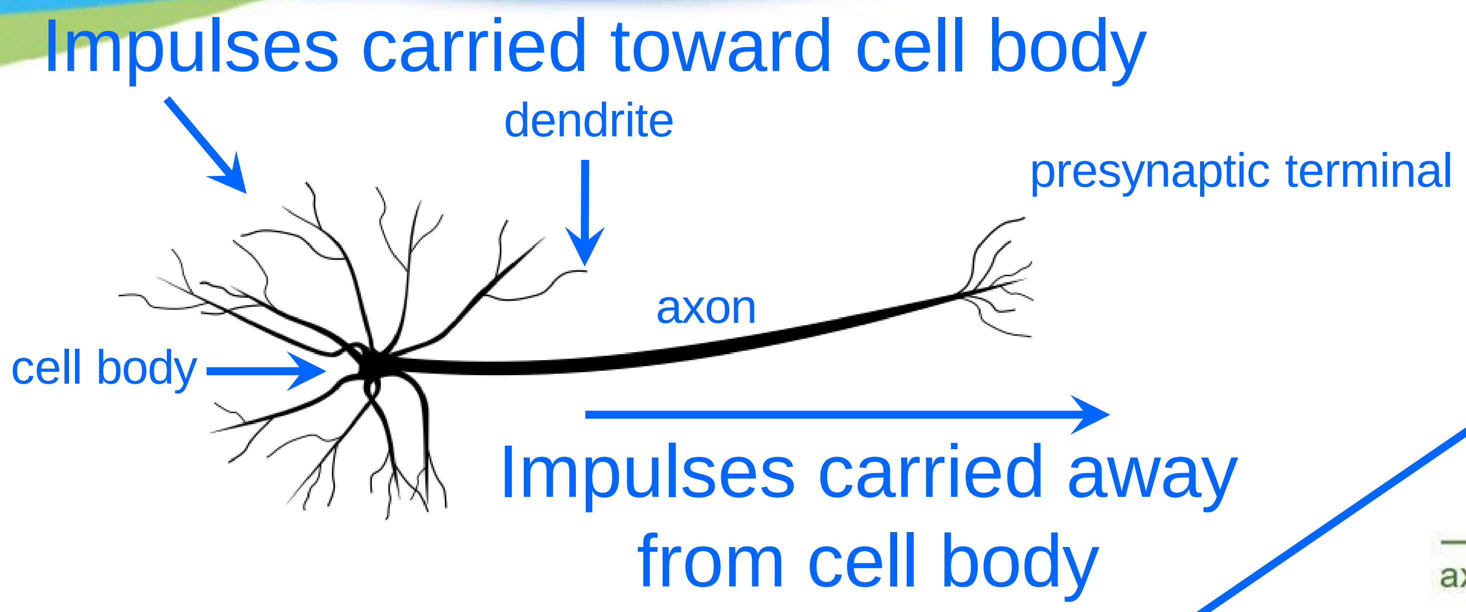
```
# receive W1,W2,b1,b2 (weights/biases), X (data)
# forward pass:
h1 = #... function of X,W1,b1
scores = #... function of h1,W2,b2
loss = #... (several lines of code to evaluate Softmax loss)
# backward pass:
dscores = #...
dh1,dW2,db2 = #...
dW1,db1 = #...
```

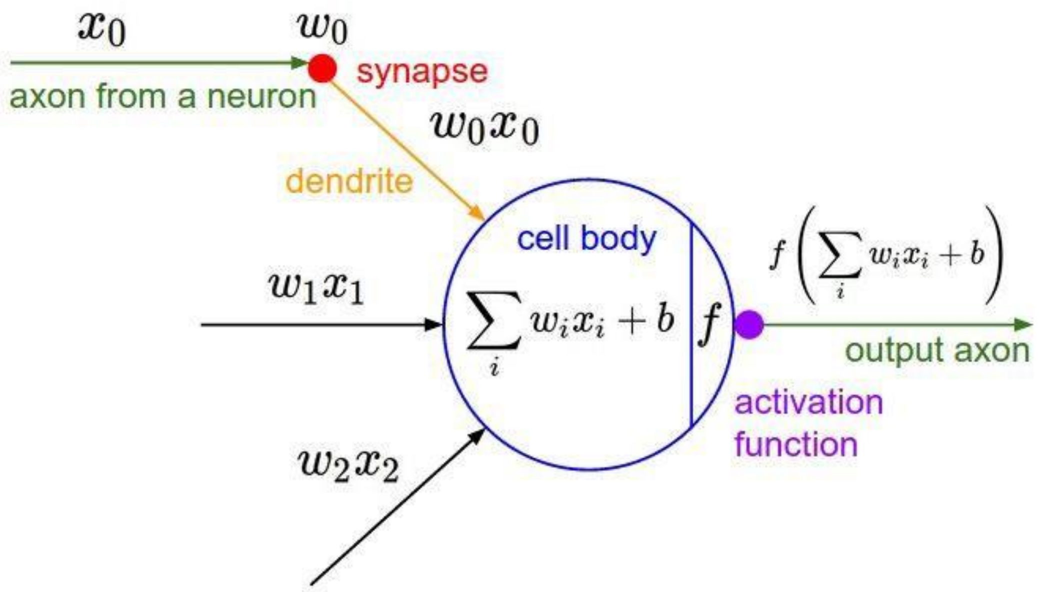
Neural networks

Impulses carried toward cell body



Impulses carried away from cell body


Impulses carried toward cell body
dendrite
presynaptic terminal
axon
cell body
Impulses carried away from cell body


x_0
w_0
synapse
axon from a neuron
$w_0 x_0$
dendrite
cell body
$w_1 x_1$
$\sum_i w_i x_i + b$
f
$f\left(\sum_i w_i x_i + b\right)$
output axon
activation function
$w_2 x_2$

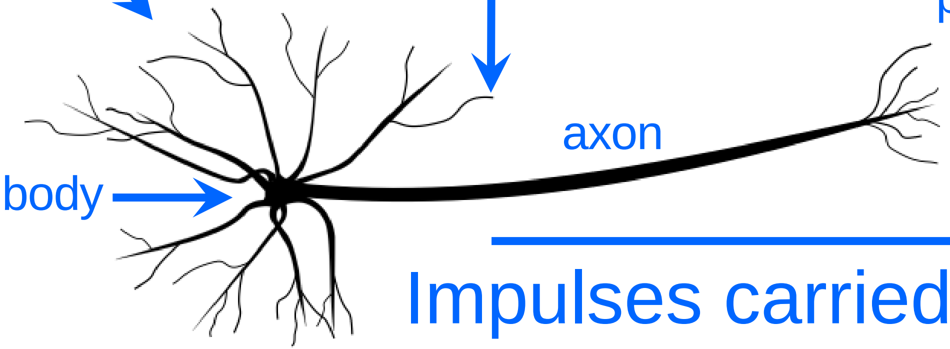
Impulses carried toward cell body

dendrite

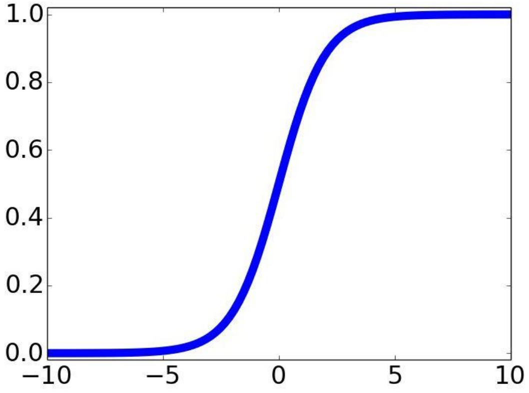
presynaptic terminal

axon

cell body

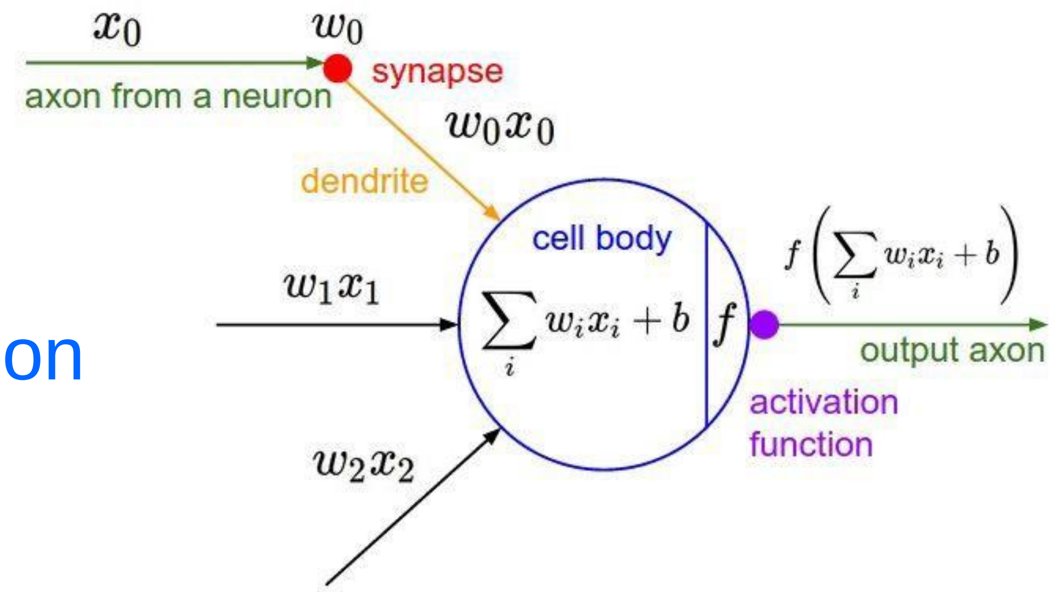


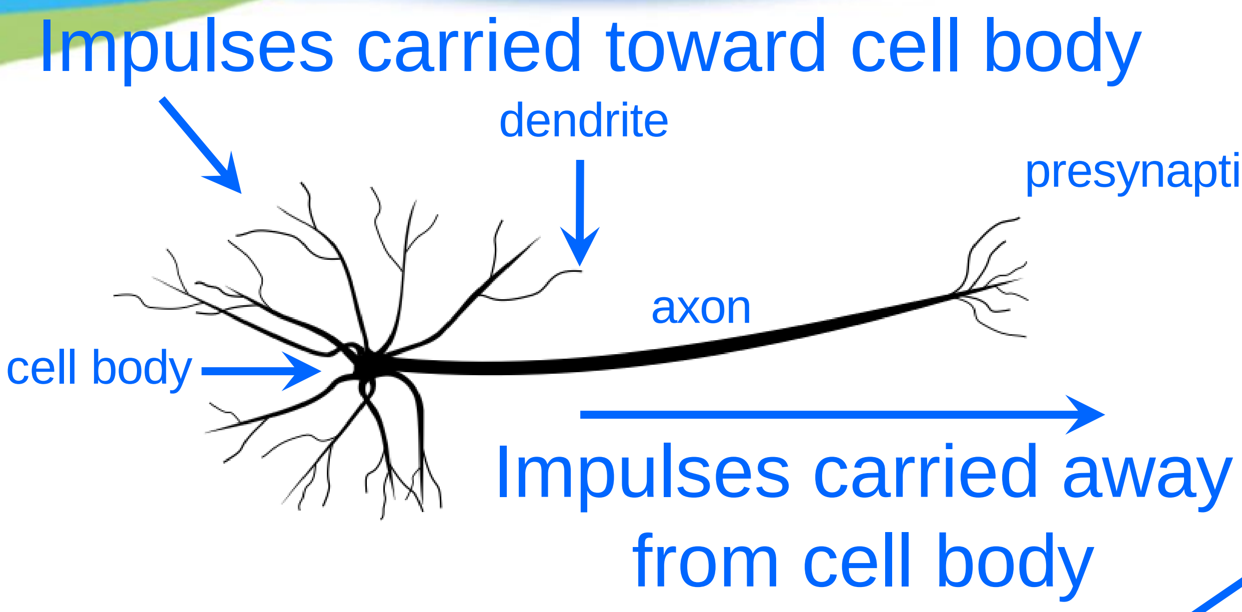
Impulses carried away from cell body



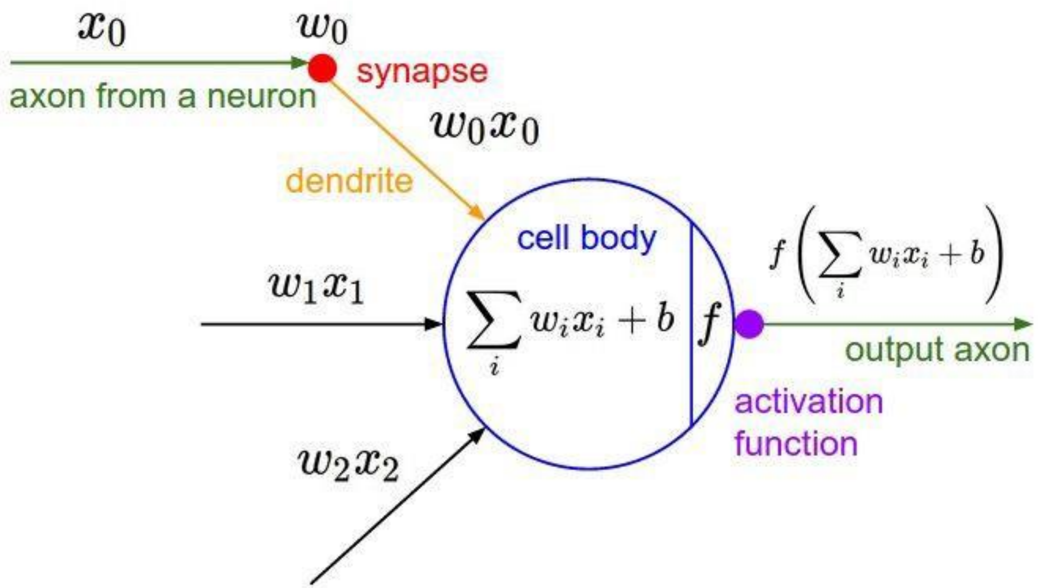
sigmoid activation function

$$\frac{1}{1+e^{-x}}$$






```python
class Neuron:
  # ...
  def neuron_tick(inputs):
    """ assume inputs and weights are 1-D numpy arrays and bias is a number """
    cell_body_sum = np.sum(inputs * self.weights) + self.bias
    firing_rate = 1.0 / (1.0 + math.exp(-cell_body_sum)) # sigmoid activation function
    return firing_rate
```

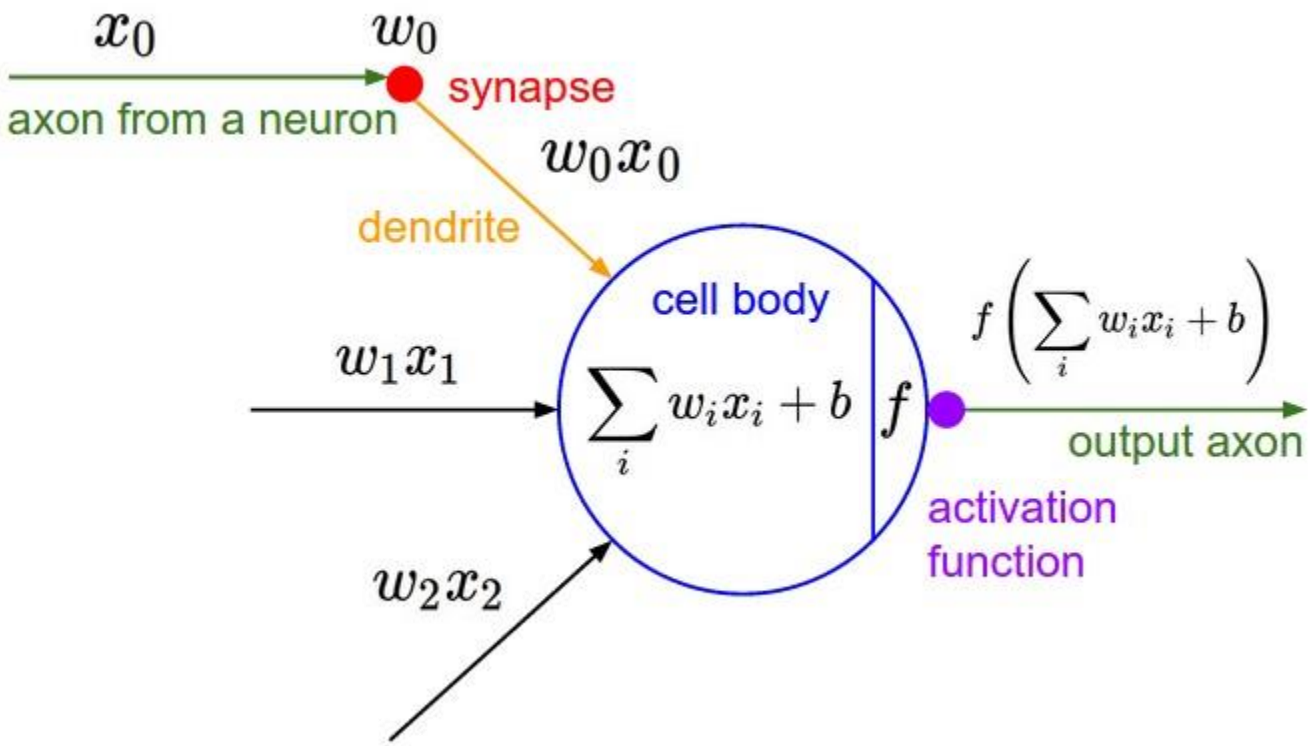


Be very careful with your brain analogies!

- Biological Neurons:
  - Many different types
  - Dendrites can perform complex non-linear computations
  - Synapses are not a single weight but a complex non-linear dynamical system
  - Rate code may not be adequate

- [Dendritic Computation. London and Hausser]

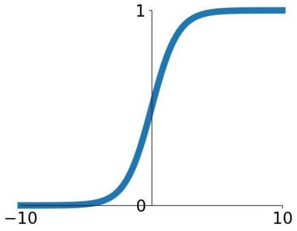
Activation function



Activation function

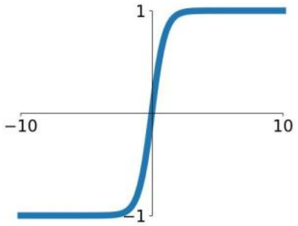
– Sigmoid

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



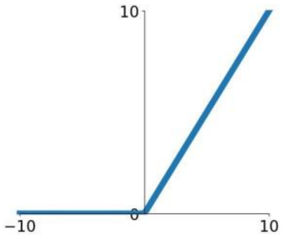
– tanh

$$tanh(x)$$



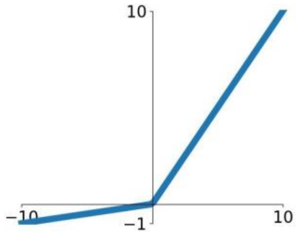
– ReLU

$$max(0, x)$$



– Leaky ReLU

$$max(0.1x, x)$$

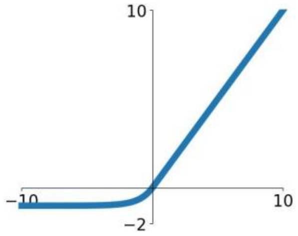


– Maxout

$$max(w_1^T x + b_1, w_2^T x + b_2)$$

– ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



Hàm ReLU

– Hàm $ReLU$ (Rectified Linear Unit) là một hàm kích hoạt phổ biến được sử dụng trong các mạng nơ-ron sâu.

– Hàm $ReLU$ giúp mô hình học cách biểu diễn dữ liệu phức tạp hơn bằng cách thêm tính phi tuyến tính vào mô hình.

– Hàm $ReLU$ được định nghĩa như sau:

$$ReLU(x) = \max(0, x)$$

Hàm ReLu

- Đặc điểm của hàm $ReLU$
  - Giá trị đầu ra: Hàm $ReLU$ trả về giá trị của đầu vào nếu đầu vào lớn hơn 0; nếu đầu vào nhỏ hơn hoặc bằng 0, hàm trả về 0.
  - Tính phi tuyến: $ReLU$ thêm tính phi tuyến vào mô hình, cho phép học các biểu diễn phức tạp hơn.
  - Đơn giản: Hàm $ReLU$ rất đơn giản để tính toán và không yêu cầu các phép toán phức tạp.

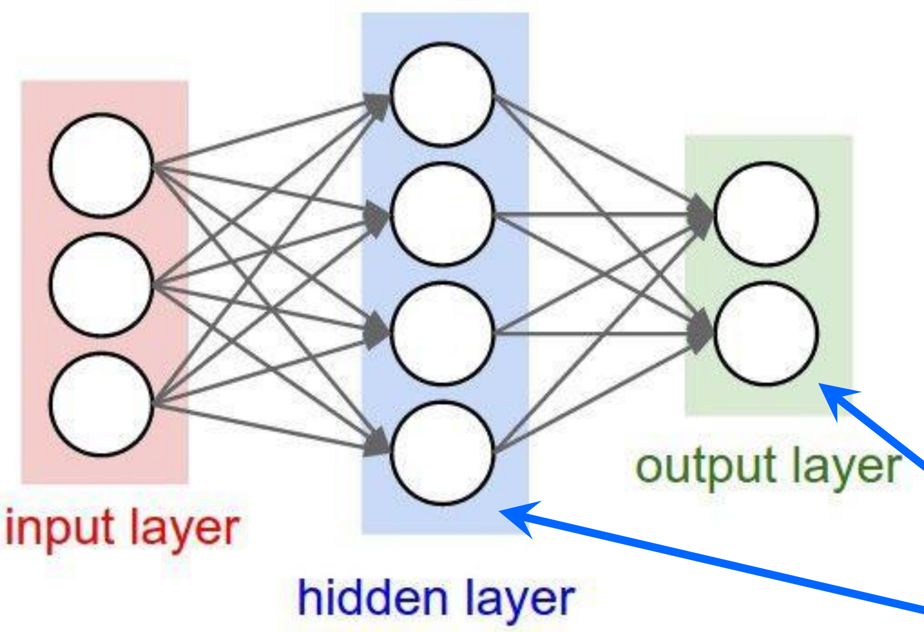
Hàm ReLu

- Lợi ích của hàm $ReLU$
  - Tốc độ tính toán nhanh: $ReLU$ chỉ cần thực hiện một phép so sánh đơn giản và không yêu cầu các phép toán phức tạp như các hàm sigmoid hay tanh.
  - Khắc phục vấn đề gradient biến mất: So với các hàm sigmoid hoặc tanh, $ReLU$ giúp khắc phục vấn đề gradient biến mất, làm cho quá trình huấn luyện mạng sâu hiệu quả hơn.

Hàm ReLu

- Nhược điểm của hàm $ReLU$
  - Dying $ReLU$: Một vấn đề của ReLU là khi giá trị đầu vào nhỏ hơn 0, gradient của hàm là 0. Điều này có thể dẫn đến việc một số nơ-ron bị chết và không bao giờ được kích hoạt lại trong quá trình huấn luyện.
  - Không trơn tru: $ReLU$ không phải là một hàm trơn tại điểm 0, điều này có thể gây ra một số vấn đề trong quá trình tối ưu hóa.

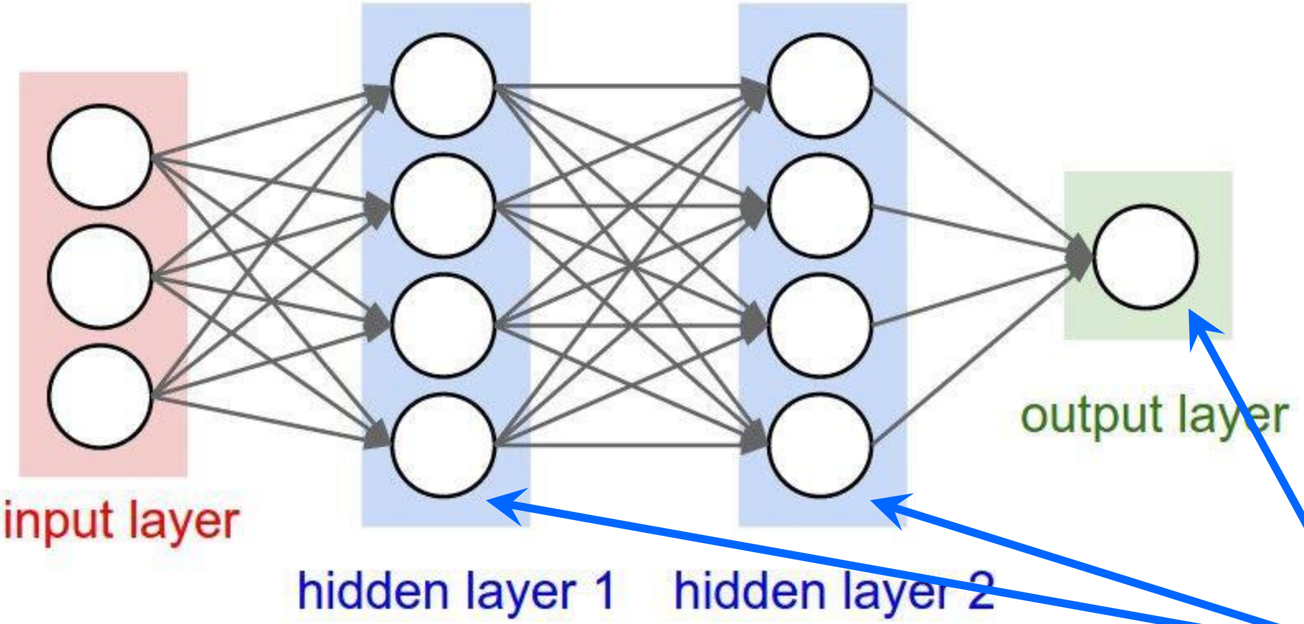
Neural networks: Architectures



“Fully-connected” layers”

– 2-layer Neural Net, or
– 1-hidden-layer Neural Net

Neural networks: Architectures



– 3-layer Neural Net, or
– 2-hidden-layer Neural Net

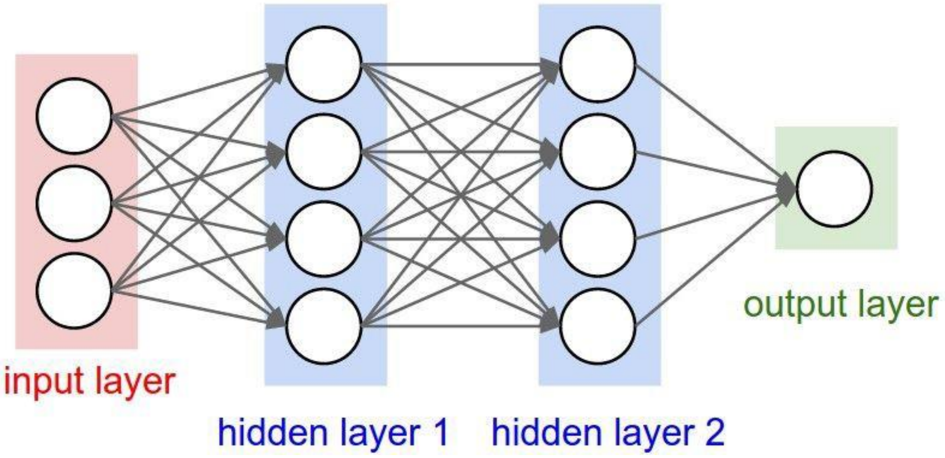
"Fully-connected" layers"

Example feed-forward computation of a neural network

```
class Neuron:
  # ...
  def neuron_tick(inputs):
    """ assume inputs and weights are 1-D numpy arrays and bias is a number """
    cell_body_sum = np.sum(inputs * self.weights) + self.bias
    firing_rate = 1.0 / (1.0 + math.exp(-cell_body_sum)) # sigmoid activation function
    return firing_rate
```

– We can efficiently evaluate an entire layer of neurons.

Example feed-forward computation of a neural network




```python
# forward-pass of a 3-layer neural network:
f = lambda x: 1.0/(1.0 + np.exp(-x)) # activation function (use sigmoid)
x = np.random.randn(3, 1) # random input vector of three numbers (3x1)
h1 = f(np.dot(W1, x) + b1) # calculate first hidden layer activations (4x1)
h2 = f(np.dot(W2, h1) + b2) # calculate second hidden layer activations (4x1)
out = np.dot(W3, h2) + b3 # output neuron (1x1)
```

Summary

- We arrange neurons into fully-connected layers
- Chúng tôi sắp xếp các nơ-ron thành các lớp được kết nối đầy đủ
- The abstraction of a layer has the nice property that it allows us to use efficient vectorized code (e.g. matrix multiplies)
- Sự trừu tượng hóa của một lớp có một đặc tính hay là nó cho phép chúng ta sử dụng mã lệnh vector hóa hiệu quả (ví dụ: nhân ma trận).

Summary

– Neural networks are not really neural.
– Mạng nơ rơn không thực sự là mạng thần kinh trong bộ não con người.
– Next time: Convolutional Neural Networks.
– Bài giảng tiếp theo: Mạng nơ ron tích chập.

NEURAL NETWORKS

Content

Neural networks

WHAT IS ARTIFICIAL NEURAL NETWORK?

What is Artificial Neural Network?

- Artificial Neural Network ANN is an efficient computing system whose central theme is borrowed from the analogy of biological neural networks.
- Mạng thần kinh nhân tạo ANN là một hệ thống tính toán hiệu quả được lấy ý tưởng của mạng thần kinh sinh học.

What is Artificial Neural Network?

- ANNs are also named as “artificial neural systems” or “parallel distributed processing systems,” or “connectionist systems.”
- ANN cũng được đặt tên là “hệ thống nơ-ron nhân tạo” hoặc “hệ thống xử lý phân tán song song” hoặc “hệ thống kết nối”.

What is Artificial Neural Network?

– ANN acquires a large collection of units that are interconnected in some pattern to allow communication between the units. These units, also referred to as nodes or neurons, are simple processors which operate in parallel.

– ANN có một tập hợp lớn các đơn vị được kết nối với nhau theo một số kiểu để cho phép giao tiếp giữa các đơn vị. Những đơn vị này được gọi là node hoặc nơ-ron, là những đơn vị xử lý đơn giản hoạt động song song.

What is Artificial Neural Network?

– Every neuron is connected with other neuron through a connection link.

– Mọi nơron đều được kết nối với nơron khác thông qua một liên kết kết nối.

What is Artificial Neural Network?

– Each connection link is associated with a weight that has information about the input signal.

– Mỗi kết nối được liên kết với một trọng số có thông tin.

What is Artificial Neural Network?

- This is the most useful information for neurons to solve a particular problem because the weight usually excites or inhibits the signal that is being communicated.
- Đây là thông tin hữu ích nhất cho các nơ-ron để giải quyết một vấn đề vì trọng số thường kích thích hoặc ức chế tín hiệu.

What is Artificial Neural Network?

- Each neuron has an internal state, which is called an activation signal.
- Mỗi tế bào thần kinh có một trạng thái bên trong, trạng thái này được gọi là tín hiệu kích hoạt.

What is Artificial Neural Network?

– Output signals, which are produced after combining the input signals and activation rule, may be sent to other units.

– Tín hiệu đầu ra, được tạo ra sau khi kết hợp các tín hiệu đầu vào và quy tắc kích hoạt, có thể được gửi đến các nơ-ron khác.

Neural networks

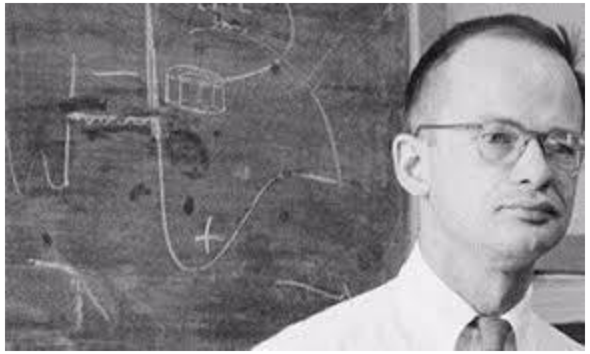
A BRIEF HISTORY OF ANN

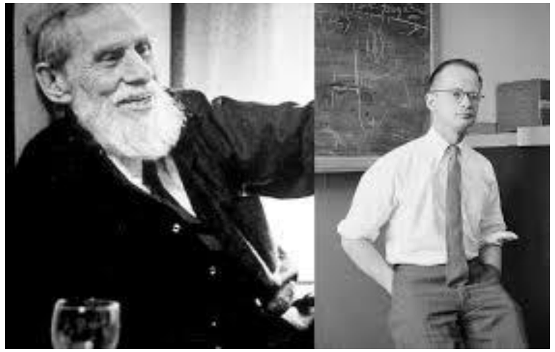
A Brief History of ANN

- The history of ANN can be divided into the following three eras
- Lịch sử của ANN có thể được chia thành ba thời kỳ như sau
  + ANN during 1940s to 1960s
  + ANN during 1960s to 1980s
  + ANN from 1980s till Present

ANN during 1940s to 1960s

- 1943 – The concept of neural network started with the work of physiologist, Warren McCulloch, and mathematician, Walter Pitts, when in 1943
- 1943 – Khái niệm mạng nơ-ron bắt đầu từ công trình nghiên cứu của nhà sinh lý học – Warren McCulloch và nhà toán học – Walter Pitts, vào năm 1943.

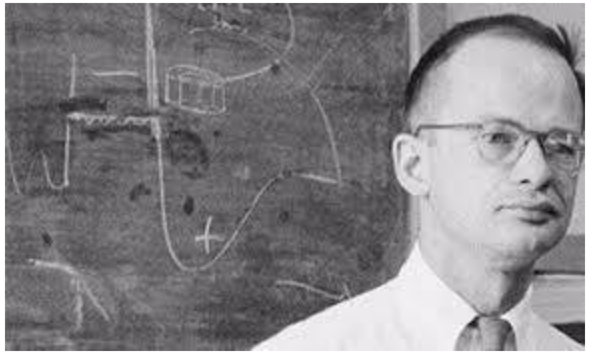


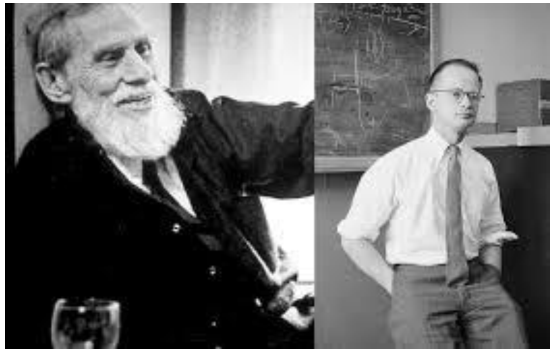


ANN during 1940s to 1960s

– Warren McCulloch and Walter Pitts modeled a simple neural network using electrical circuits in order to describe how neurons in the brain might work.

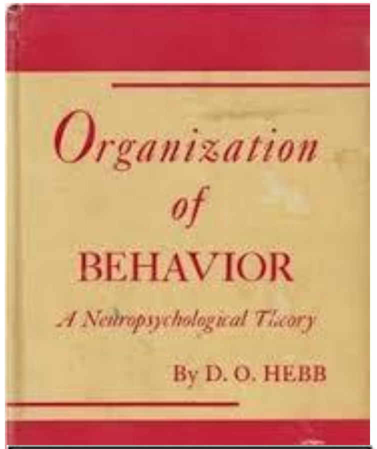
– Hai ông đã mô hình hóa một mạng nơ-ron đơn giản sử dụng các mạch điện để mô tả cách các nơ-ron hoạt động trong não người.





ANN during 1940s to 1960s

- 1949 – Donald Hebb's book, The Organization of Behavior, put forth the fact that repeated activation of one neuron by another increases its strength each time they are used.
- 1949 – Cuốn sách của Donald Hebb có tên, Tổ chức của Hành vi, đưa ra bằng chứng là việc kích hoạt lặp đi lặp lại một nơ-ron thần kinh sẽ làm tăng sức mạnh của nơ-ron đó.



ANN during 1940s to 1960s

– 1956 − An associative memory network was introduced by Taylor.
– 1956 – Mạng bộ nhớ kết hợp được giới thiệu bởi Taylor.

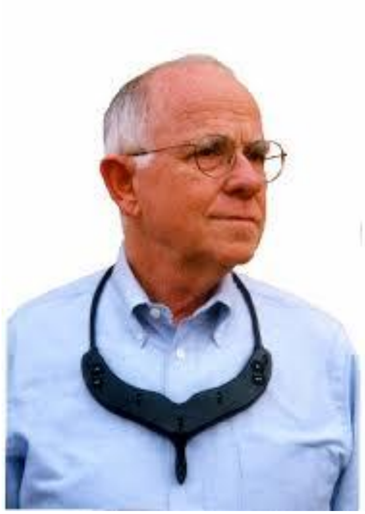
ANN during 1940s to 1960s

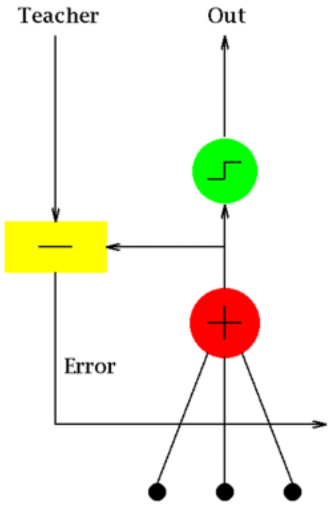
- 1958 – A learning method for McCulloch and Pitts neuron model named Perceptron was invented by Rosenblatt.
- 1958 – Phương pháp học Perceptron được phát minh bởi Rosenblatt trên mô hình mạng neuron của McCulloch và Pitts.

ANN during 1940s to 1960s

– 1960 – Bernard Widrow and Marcian Hoff developed models called "ADALINE" and “MADALINE.”

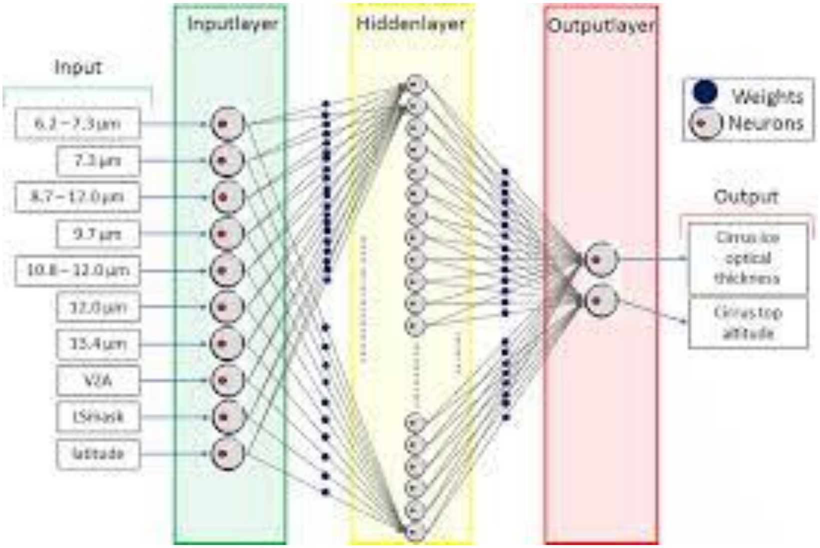
– 1960 – Bernard Widrow và Marcian Hoff phát triển các mô hình được gọi là "ADALINE" và "MADALINE."





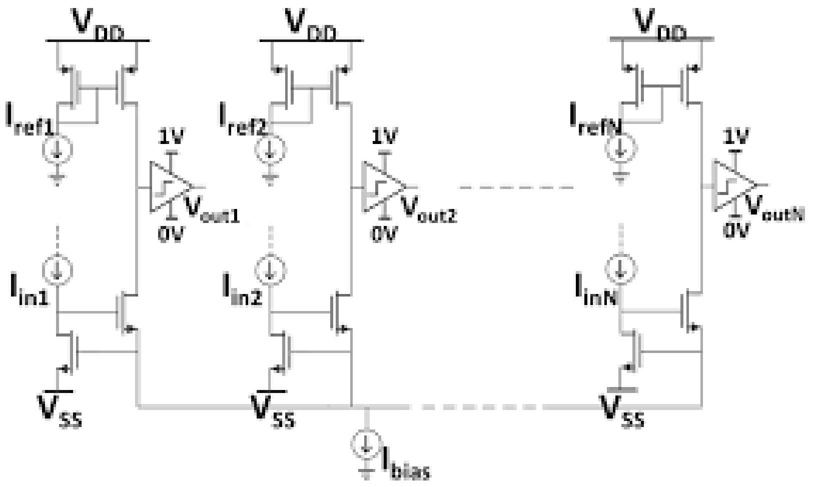
ANN during 1960s to 1980s

- 1961 – Rosenblatt made an unsuccessful attempt but proposed the “backpropagation” scheme for multilayer networks.
- 1961 - Rosenblatt đã thực hiện một nỗ lực không thành công nhưng đã đề xuất phương pháp “backpropagation” cho các mạng đa lớp.



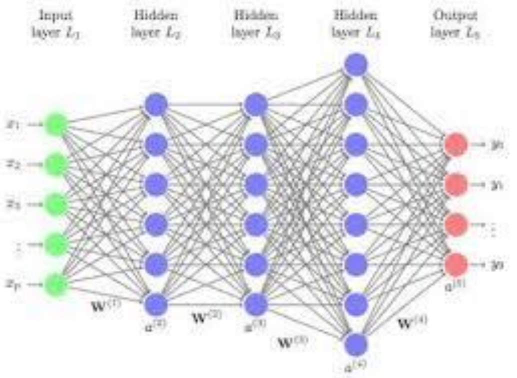
ANN during 1960s to 1980s

- 1964 – Taylor constructed a winner-take-all circuit with inhibitions among output units.
- 1964 – Taylor đã xây dựng một mạch thắng – lấy – tất – cả với sự ức chế giữa các đơn vị đầu ra.



ANN during 1960s to 1980s

- 1969 – Multilayer perceptron MLP was invented by Minsky and Papert.
- 1969 – Mạng đa lớp MLP được phát minh bởi Minsky và Papert.

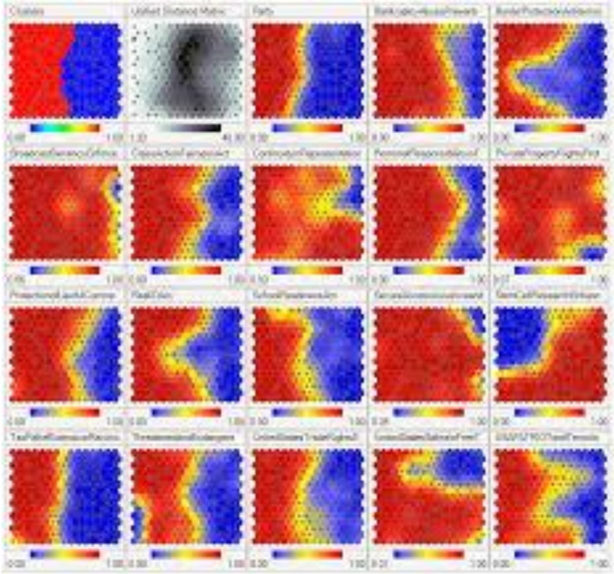


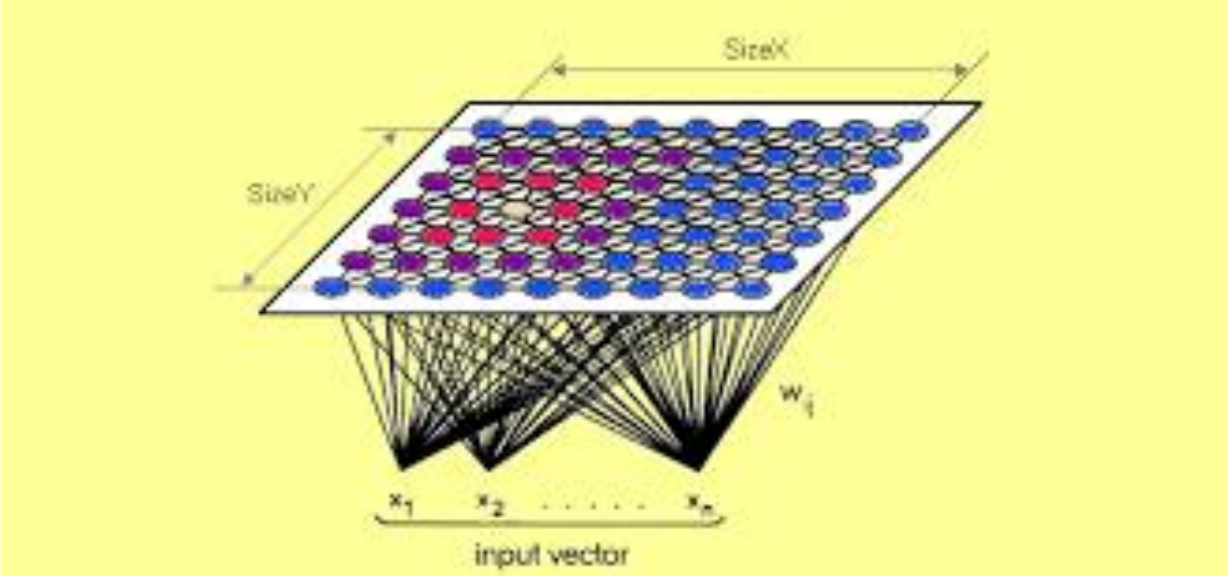


ANN during 1960s to 1980s

– 1971 – Kohonen developed Associative memories.

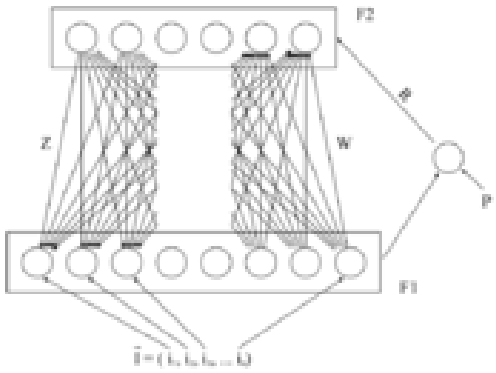
– 1971 – Kohonen phát triển Associative memories.





ANN during 1960s to 1980s

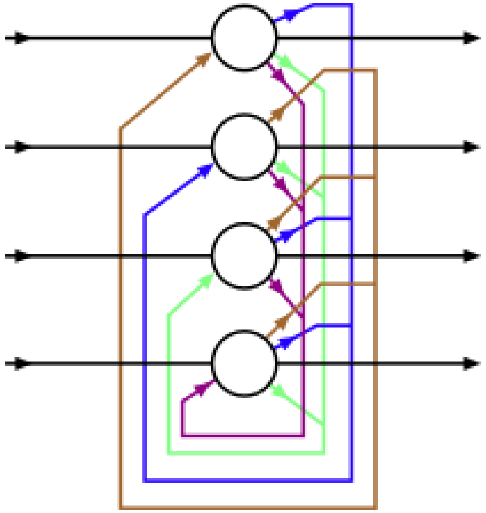
- 1976 – Stephen Grossberg and Gail Carpenter developed Adaptive resonance theory.
- 1976 - Stephen Grossberg và Gail Carpenter phát triển lý thuyết cộng hưởng thích nghi.



ANN from 1980s till Present

– 1982 – The major development was Hopfield's Energy approach.

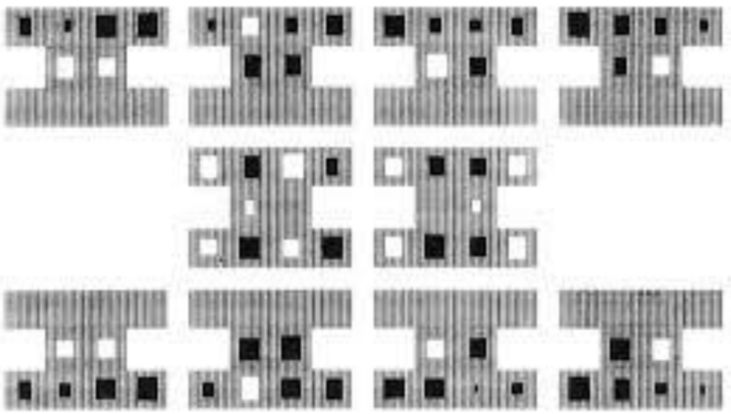
– Năm 1982 – Phương pháp Hopfield's Energy.



ANN from 1980s till Present

– 1985 – Boltzmann machine was developed by Ackley, Hinton, and Sejnowski.

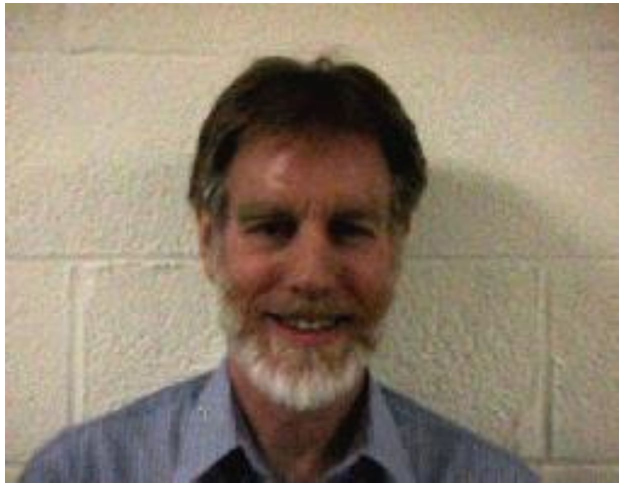
– 1985 – Boltzmann machine được phát triển bởi Ackley, Hinton và Sejnowski.



ANN from 1980s till Present

- 1986 – Rumelhart, Hinton, and Williams introduced Generalised Delta Rule.
- 1986 – Rumelhart, Hinton và Williams giới thiệu Generalised Delta Rule.

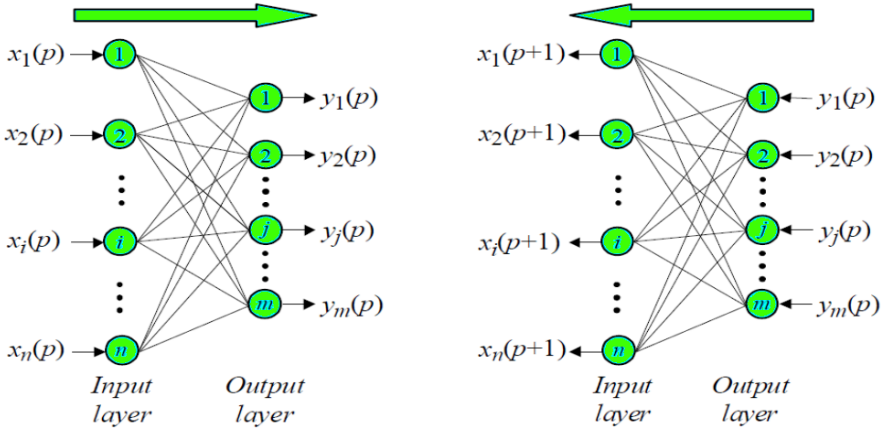




ANN from 1980s till Present

– 1988 – Kosko developed Binary Associative Memory BAM and also gave the concept of Fuzzy Logic in ANN.

– 1988 – Kosko phát triển Bộ nhớ Liên kết Nhị phân – BAM và cũng đưa ra khái niệm Logic mờ trong ANN.



(a) Forward direction.

(b) Backward direction.

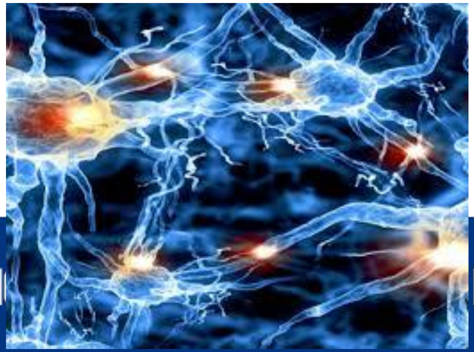
Neural networks

BIOLOGICAL NEURON

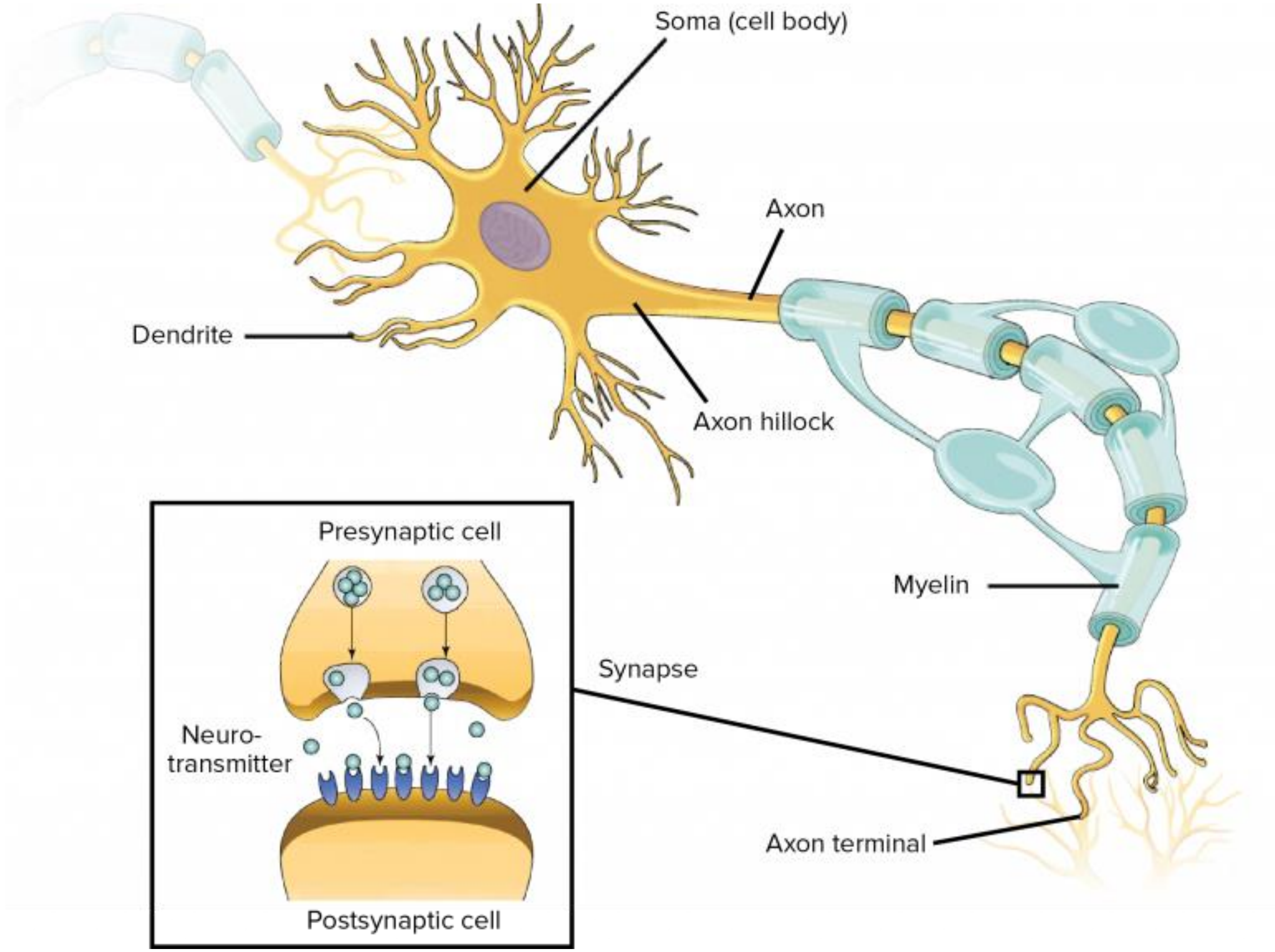
Neuron

- A nerve cell neuron is a special biological cell that processes information. According to an estimation, there are huge number of neurons, approximately 10^{11} with numerous interconnections, approximately 10^{15}.
- Tế bào thần kinh là một tế bào sinh học đặc biệt xử lý thông tin. Theo một ước tính, não bộ con người có một số lượng rất lớn từ 10^{11} tới 10^{15} tế bào thần kinh và các tế bào thần kinh có nhiều liên kết với nhau.

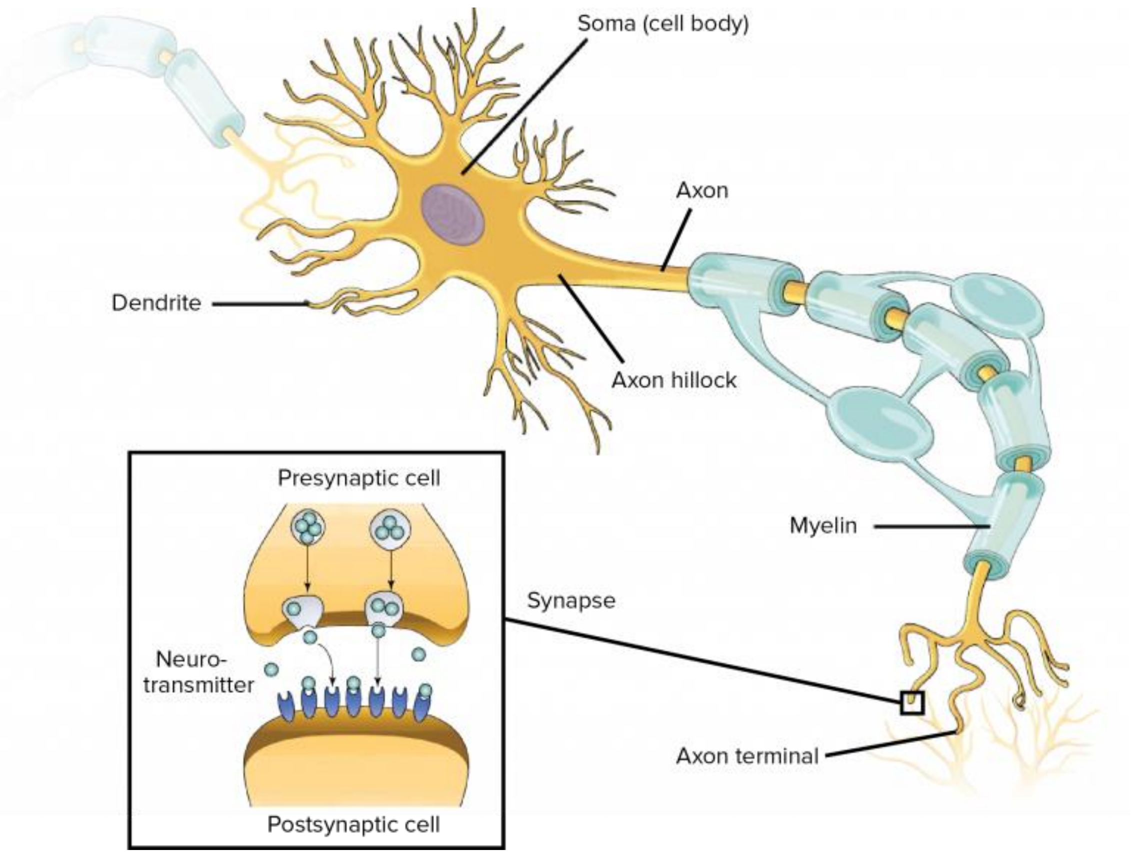





Schematic Diagram

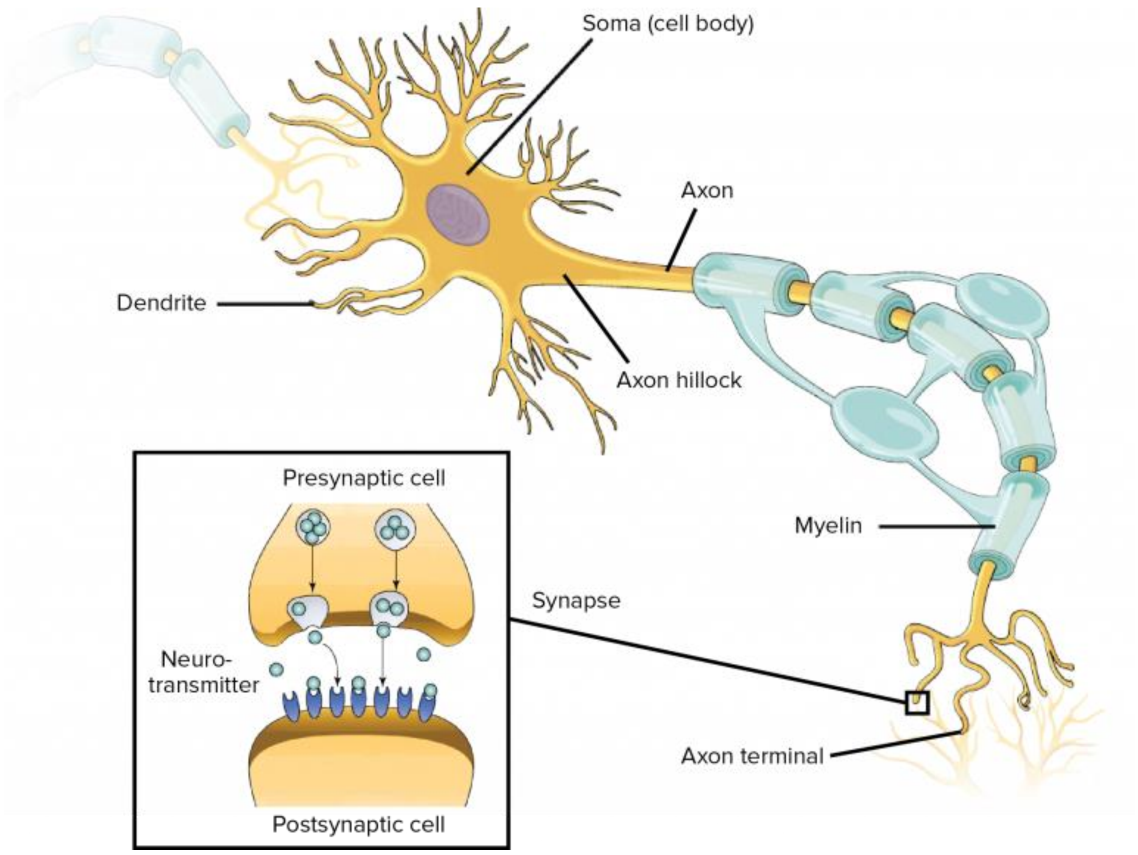


Working of a Biological Neuron



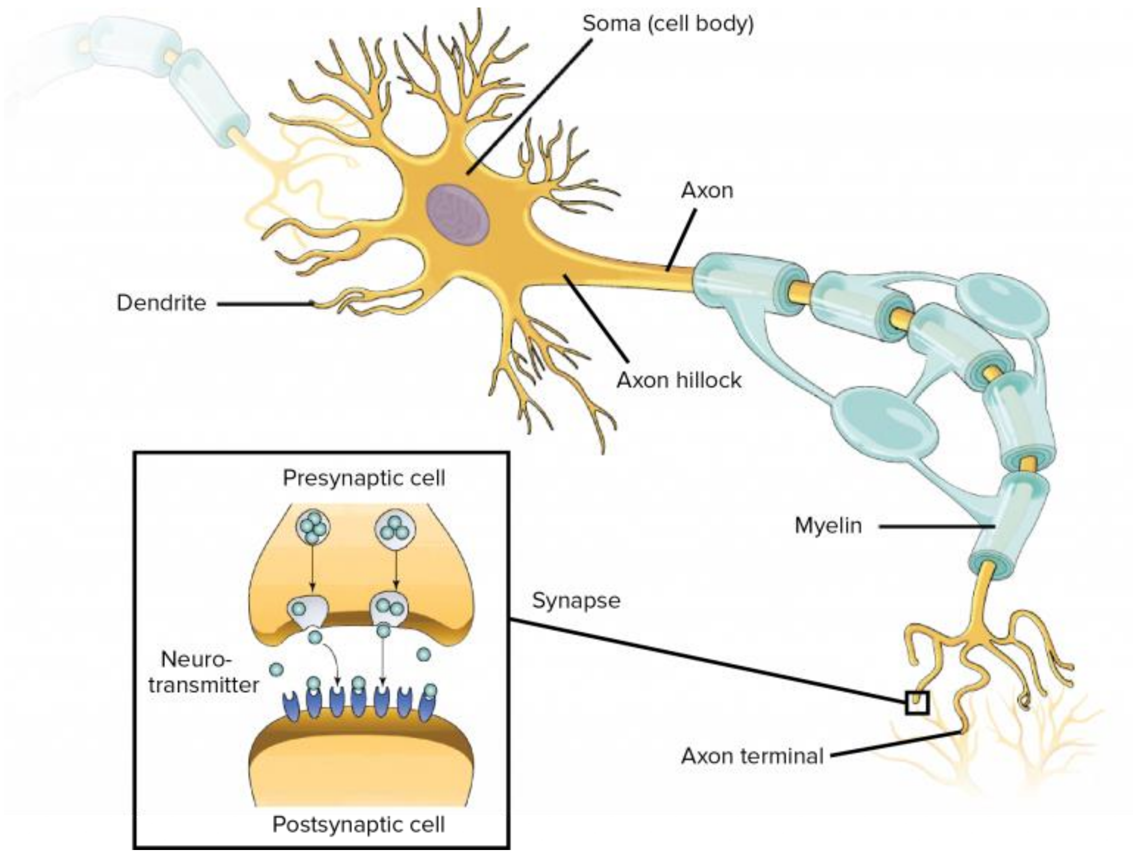
- As shown in the above diagram, a typical neuron consists of the following four parts with the help of which we can explain its working
- Như được minh họa trong sơ đồ trên, một nơ-ron điển hình bao gồm bốn phần có thể giải thích hoạt động của nó

Working of a Biological Neuron



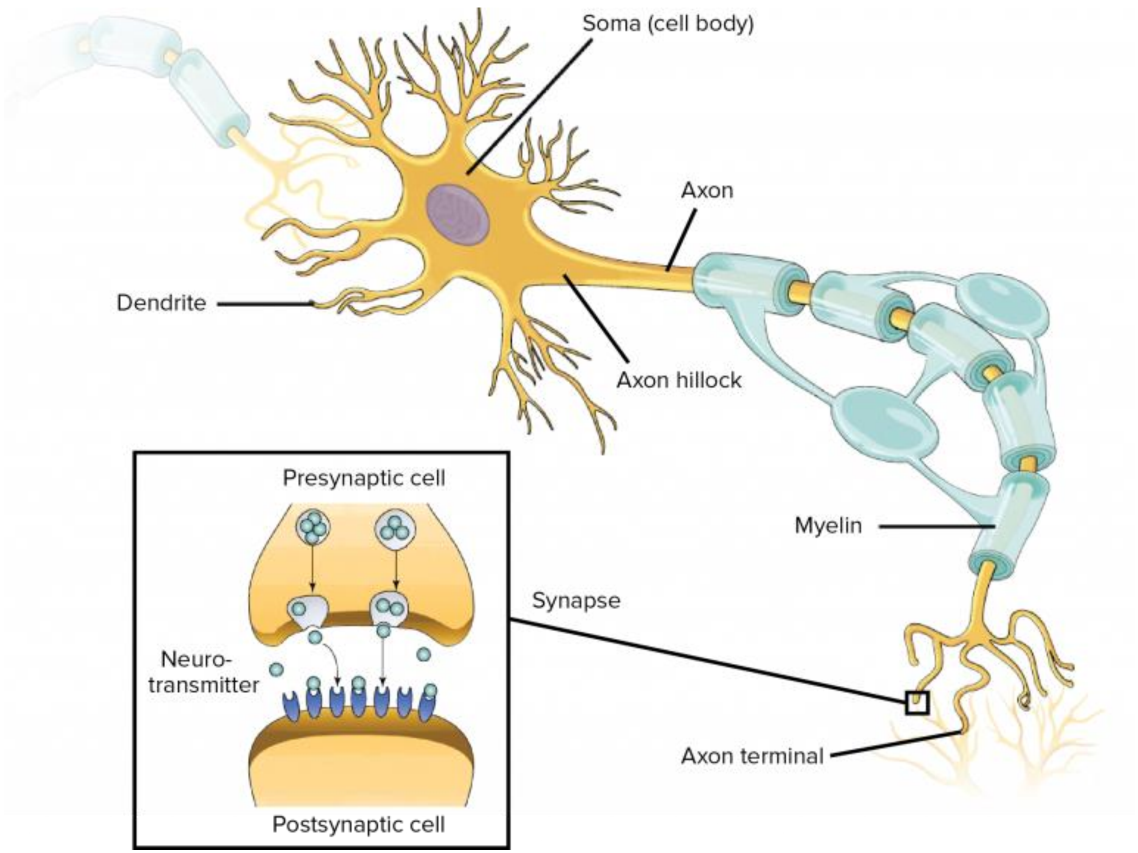
- Dendrites − They are tree-like branches, responsible for receiving the information from other neurons it is connected to.
- Đuôi gai – Dendrites – là những nhánh giống như rễ cây, nhận thông tin từ các tế bào thần kinh khác mà nơ ron có kết nối.

Working of a Biological Neuron



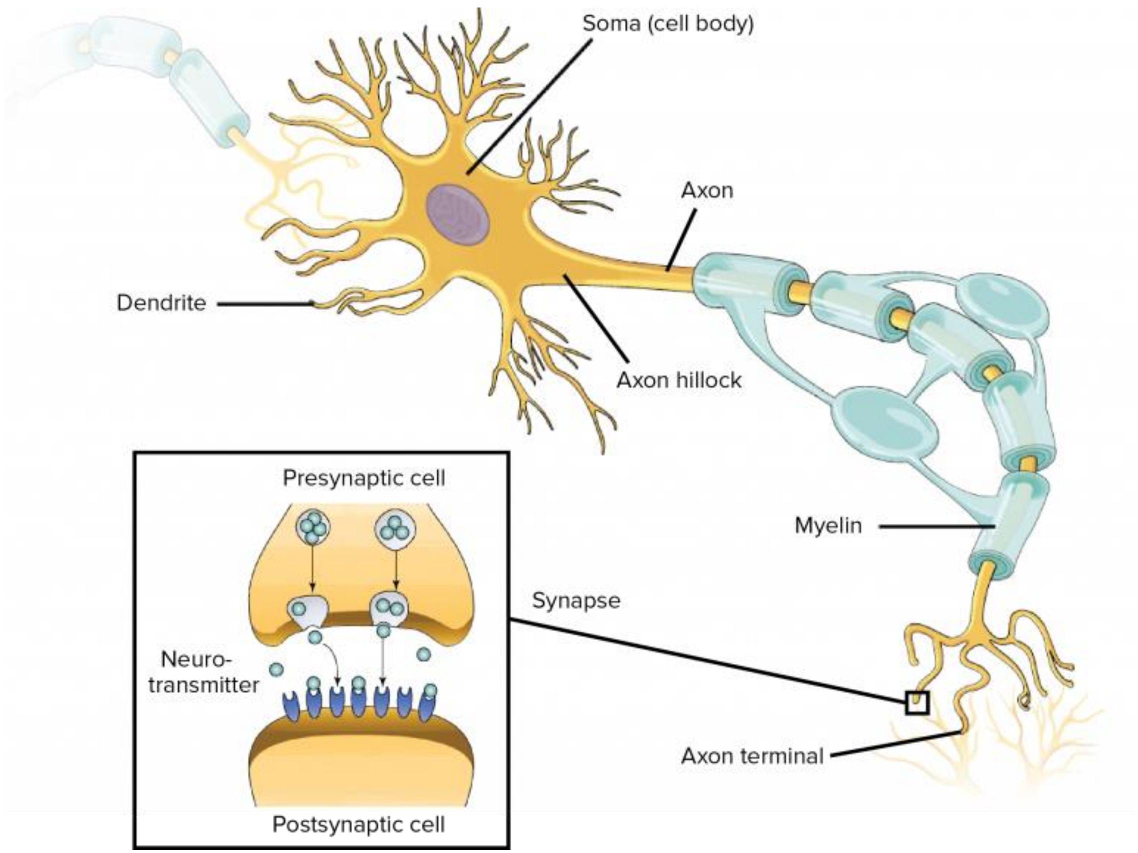
+ Soma − It is the cell body of the neuron and is responsible for processing of information, they have received from dendrites.
+ Thể tế bào – Soma – là thân tế bào của nơ ron và thể tế bào xử lý các thông tin nhận được từ các đuôi gai.

Working of a Biological Neuron



+ Axon − It is just like a cable through which neurons send the information.
+ Sợi trục – Axon – Giống như một sợi cáp thông qua đó các tế bào thần kinh gửi thông tin cho nhau.

Working of a Biological Neuron



+ Synapses − It is the connection between the axon and other neuron dendrites.
+ Khớp thần kinh – Synapses – là kết nối giữa sợi trục và các đuôi gai thần kinh khác.

Neural networks

ANN VERSUS BNN

ANN versus BNN

- The similarities between Artificial Neural Network (ANN) and Biological Neural Network (BNN).
- Điểm giống nhau giữa Mạng nơron nhân tạo (ANN) và Mạng nơron sinh học (BNN).

Biological Neural Network	Artificial Neural Network
Soma – Thể tế bào	Node – Nút
Dendrites – Đuôi gai	Input – Đầu vào
Synapse – Khớp thần kinh	Weights – Trọng số
Axon – Sợi trục	Output – Đầu ra

ANN versus BNN

– The following table shows the comparison between ANN and BNN based on some criteria mentioned.

– Bảng sau đây cho thấy sự so sánh giữa ANN và BNN dựa trên một số tiêu chí đã đề cập.

+ Processing – Khả năng xử lý

+ Size – Cỡ

+ Learning – Khả năng học

+ Fault tolerance – Khả năng chịu lỗi

+ Storage capacity – Khả năng lưu trữ

ANN versus BNN

Criteria	BNN	ANN
Processing	Massively parallel, slow but superior than ANN	Massively parallel, fast but inferior than BNN
Khả năng xử lý	Song song, chậm nhưng vượt trội hơn ANN	Song song, nhanh nhưng kém hơn BNN

ANN versus BNN

Criteria	BNN	ANN
Size	10^{11} neurons and 10^{15} interconnections	10^{2} to 10^{4} nodes
Cỡ	10^{11} nơ-ron và 10^{15} liên kết	10^{2} đến 10^{4} nodes

ANN versus BNN

Criteria	BNN	ANN
Learning	They can tolerate ambiguity	Very precise, structured and formatted data is required to tolerate ambiguity
Khả năng học	BNN có thể chịu đựng sự mơ hồ	Dữ liệu có cấu trúc và định dạng rất chính xác là bắt buộc để chống lại sự mơ hồ

ANN versus BNN

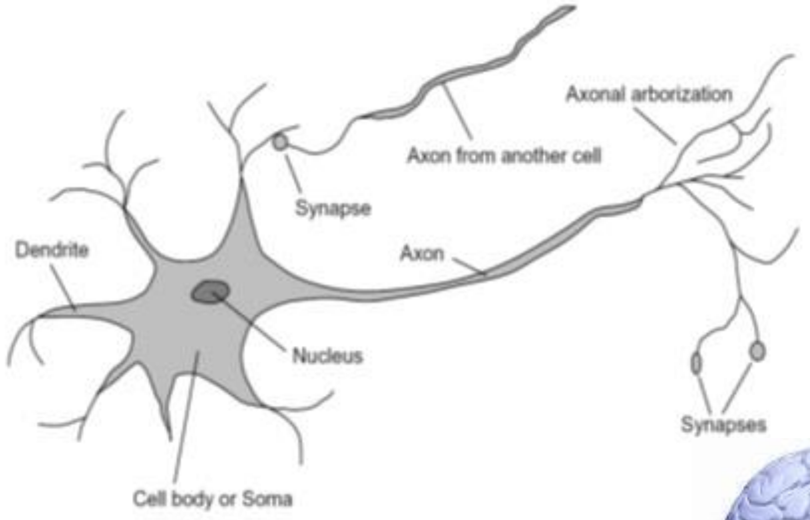
Criteria	BNN	ANN
Fault tolerance	Performance degrades with even partial damage	It is capable of robust performance, hence has the potential to be fault tolerant
Khả năng chịu lỗi	BNN có hiệu suất suy giảm với thậm chí hư hỏng một phần	ANN có khả năng hoạt động mạnh mẽ, do đó có khả năng chịu lỗi

UIT University of Information Technology
VNUHCM

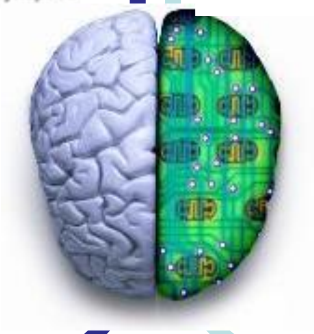
ANN versus BNN

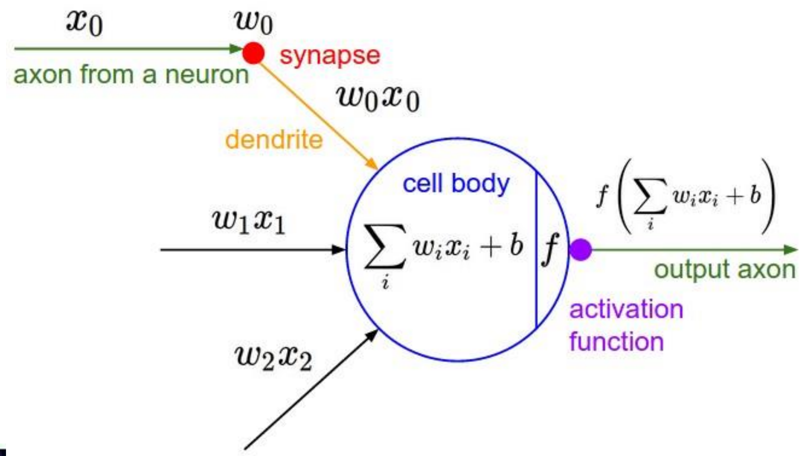
Criteria	BNN	ANN
Storage capacity	Stores the information in the synapse	Stores the information in the continuous memory locations
Khả năng lưu trữ	Lưu trữ thông tin trong khớp thần kinh	Lưu trữ thông tin trong các vị trí bộ nhớ liên tục

ANN versus BNN



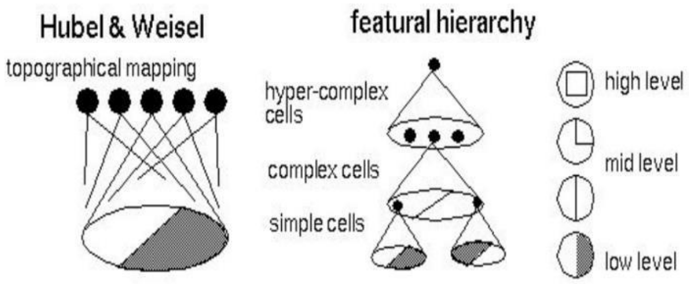
A biological neuron



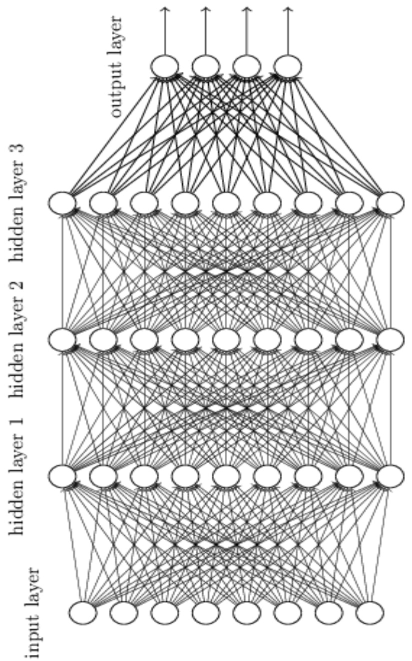


An artificial neuron

ANN versus BNN



Hubel and Weisel's architecture



Multi – layer neural network

Neural networks

ARTIFICIAL NEURAL NETWORKS

Artificial Neural Networks

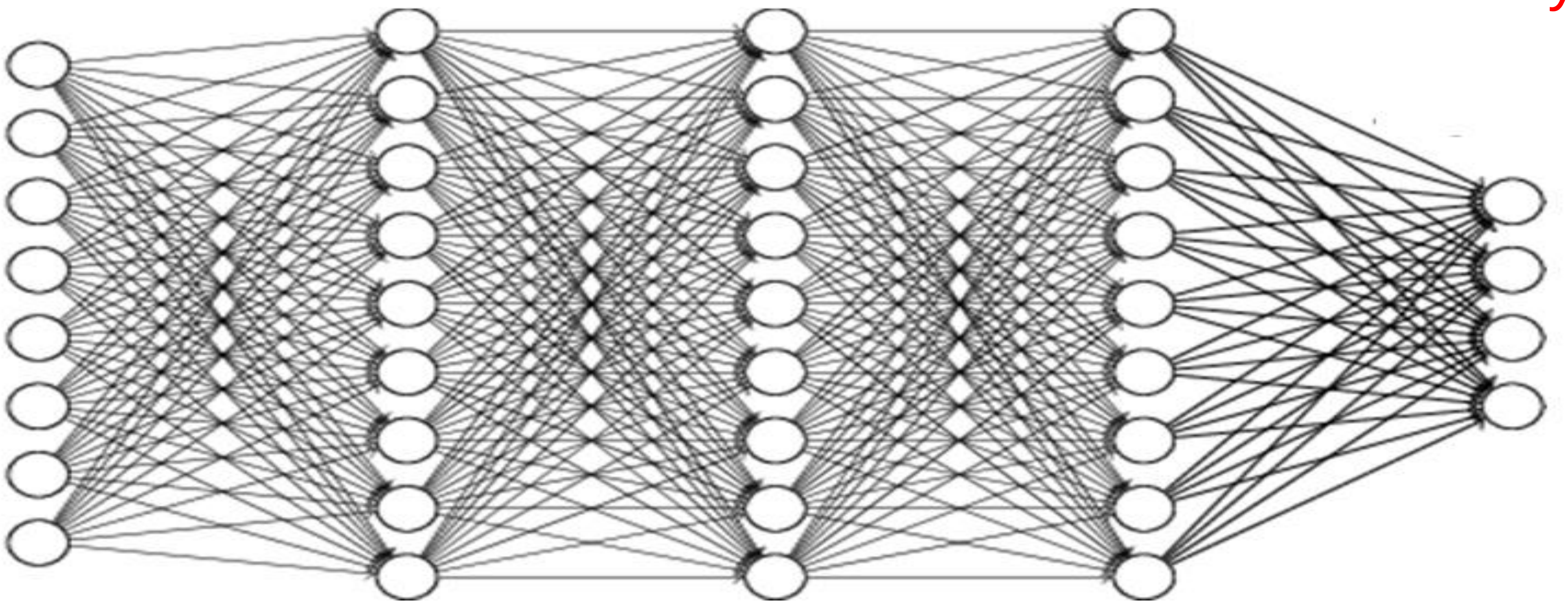
Input layer

Hidden layer 1

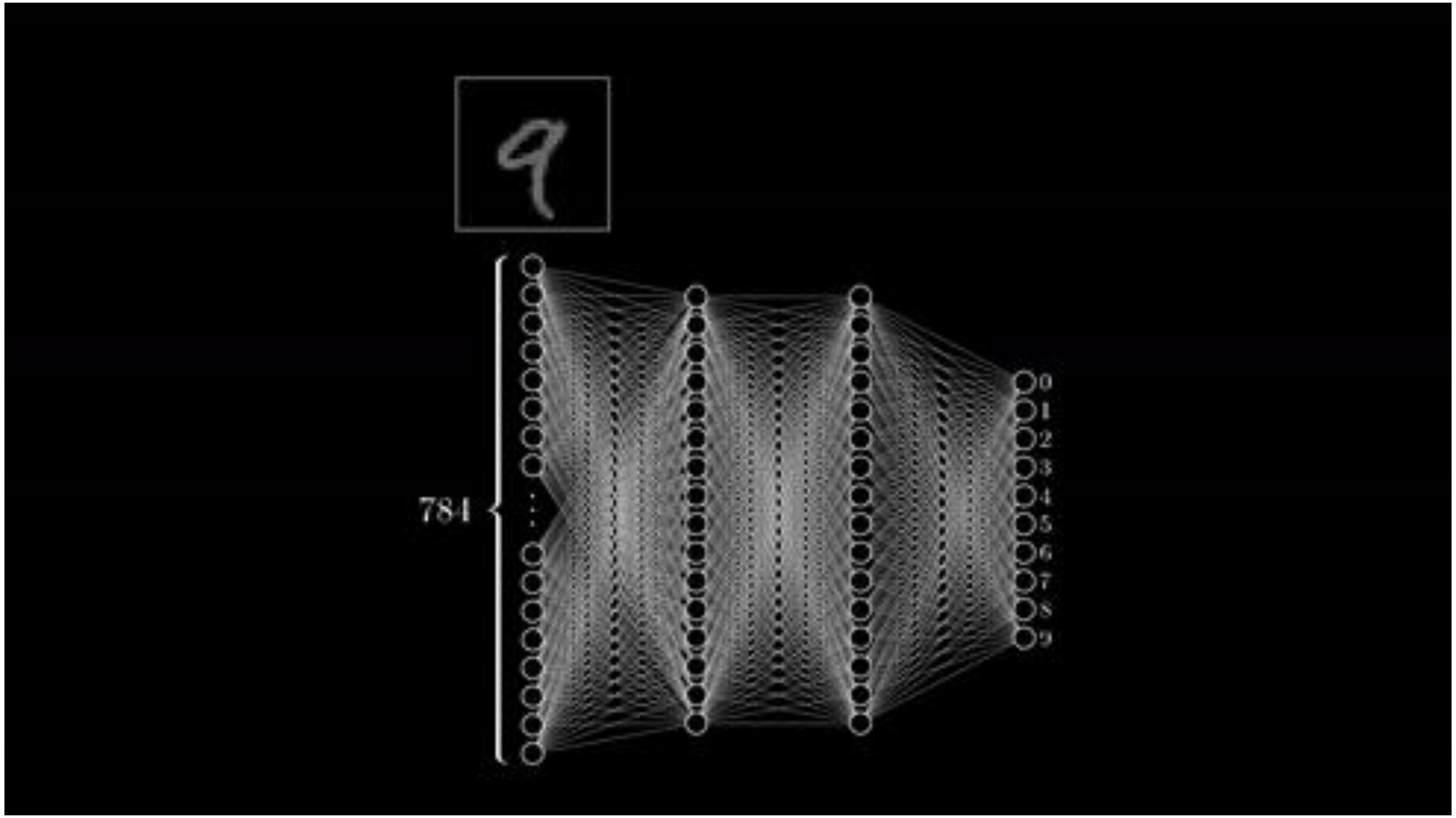
Hidden layer 2

Hidden layer 3

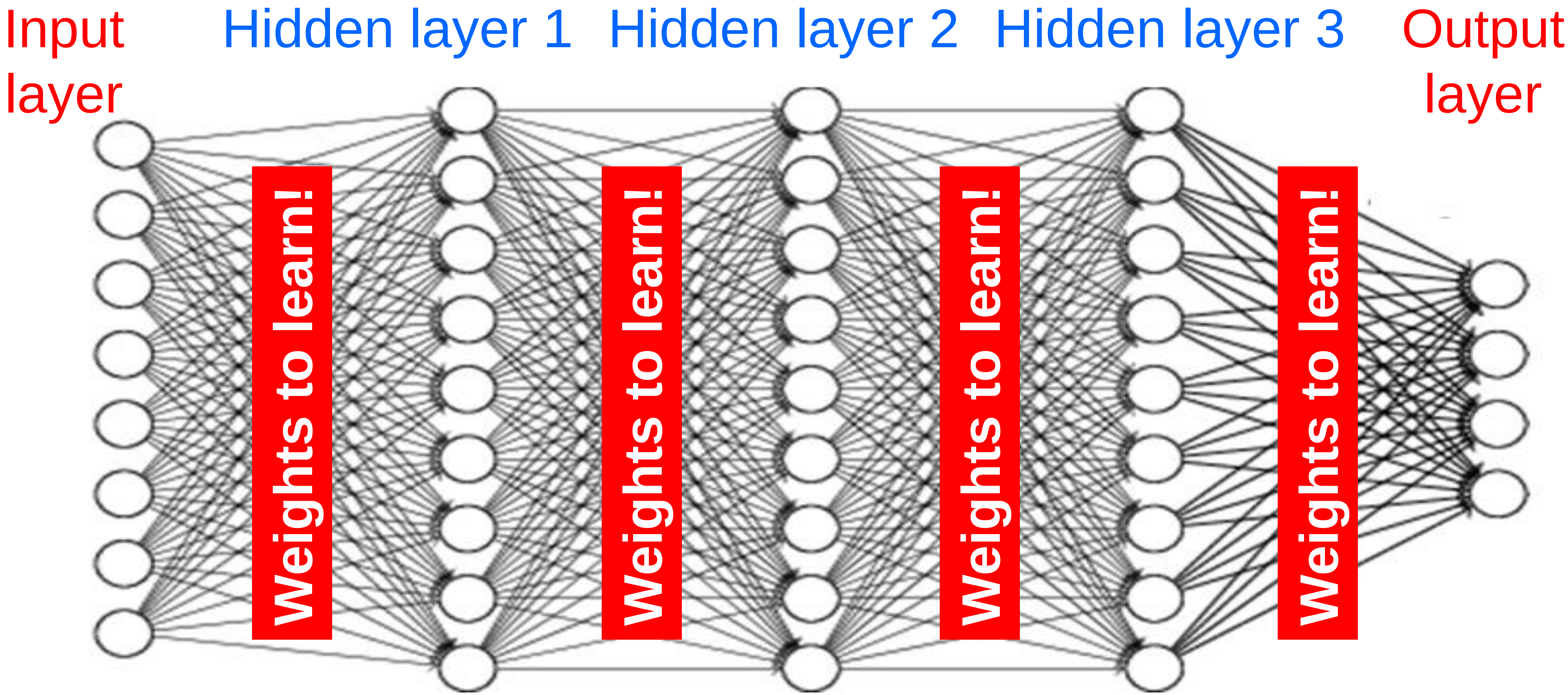
Output layer



Artificial Neural Networks



Artificial Neural Networks

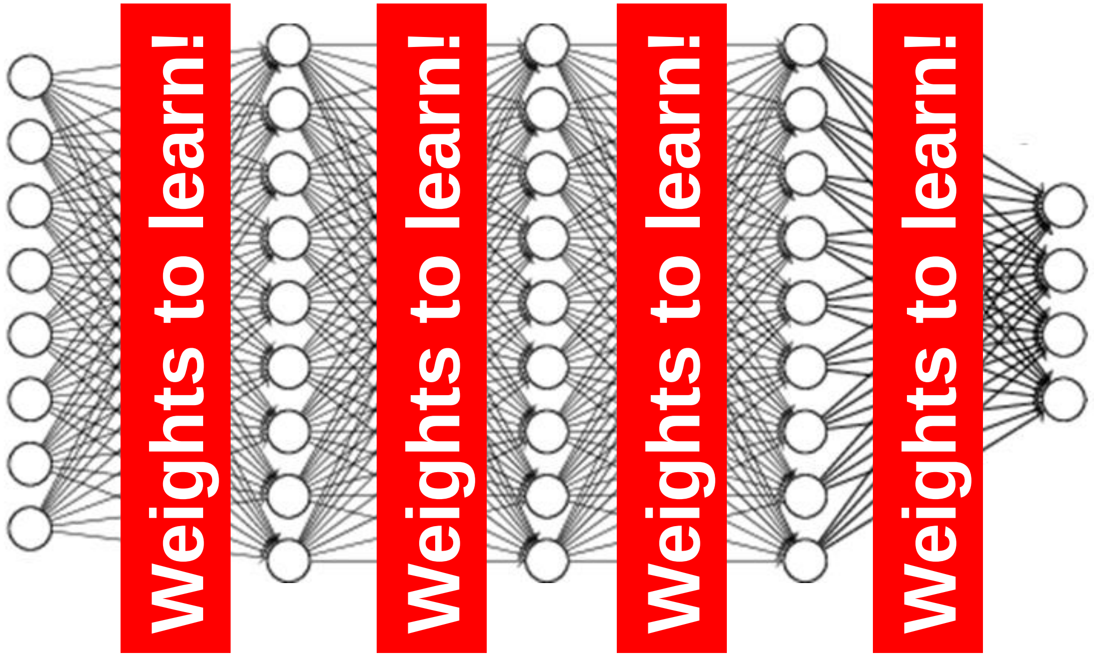


Neural networks

LEARNING PROCESS

Learning process

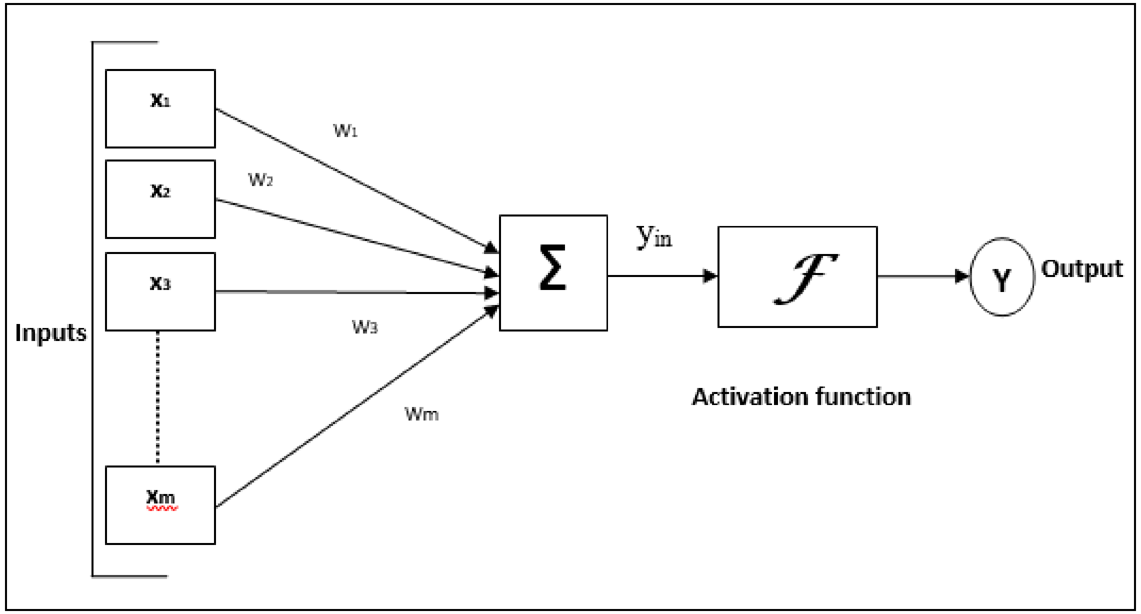
- Bước 01: Initialize network weights
- Bước 02: For each datapoint i in training set:
  - + Bước 2.1: Forward.
  - + Bước 2.2: Backward.
  - + Bước 2.3: Weight update.
- Bước 3: Return the network.



Neural networks

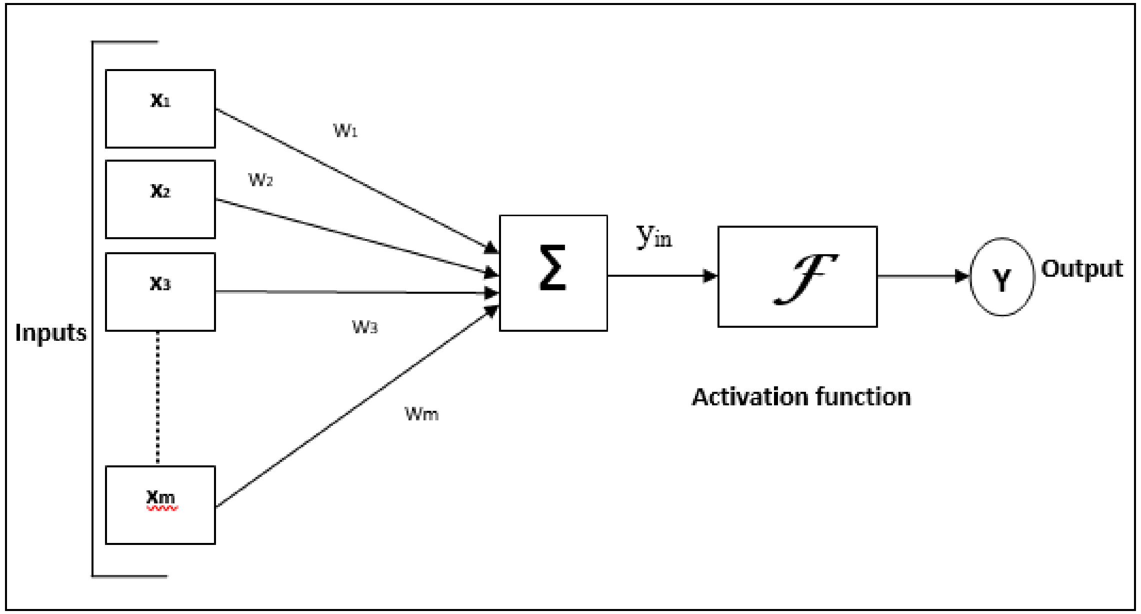
ARTIFICIAL NEURON

Artificial neuron



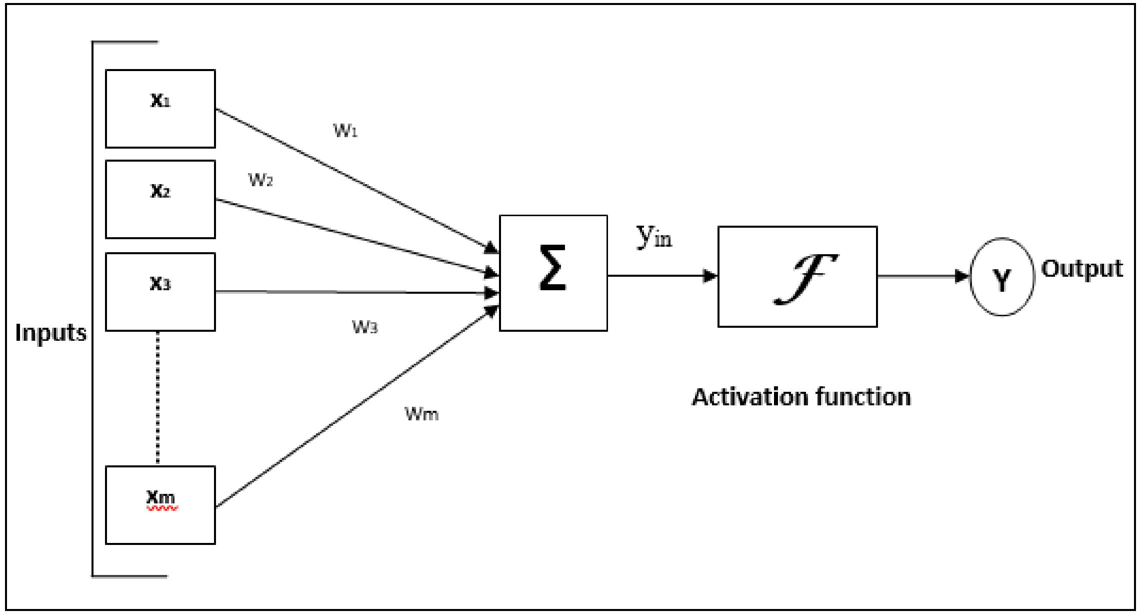
- The following diagram represents the general model of ANN followed by its processing.
- Sơ đồ sau đây thể hiện mô hình chung của ANN theo sau là quá trình xử lý của nó.

Artificial neuron



- For the above general model of artificial neural network, the net input can be calculated as follows.
- Đối với mô hình chung của mạng nơron nhân tạo ở trên, *net input* có thể được tính như sau.

Artificial neuron



- Giá trị y_{in} được tính như sau:

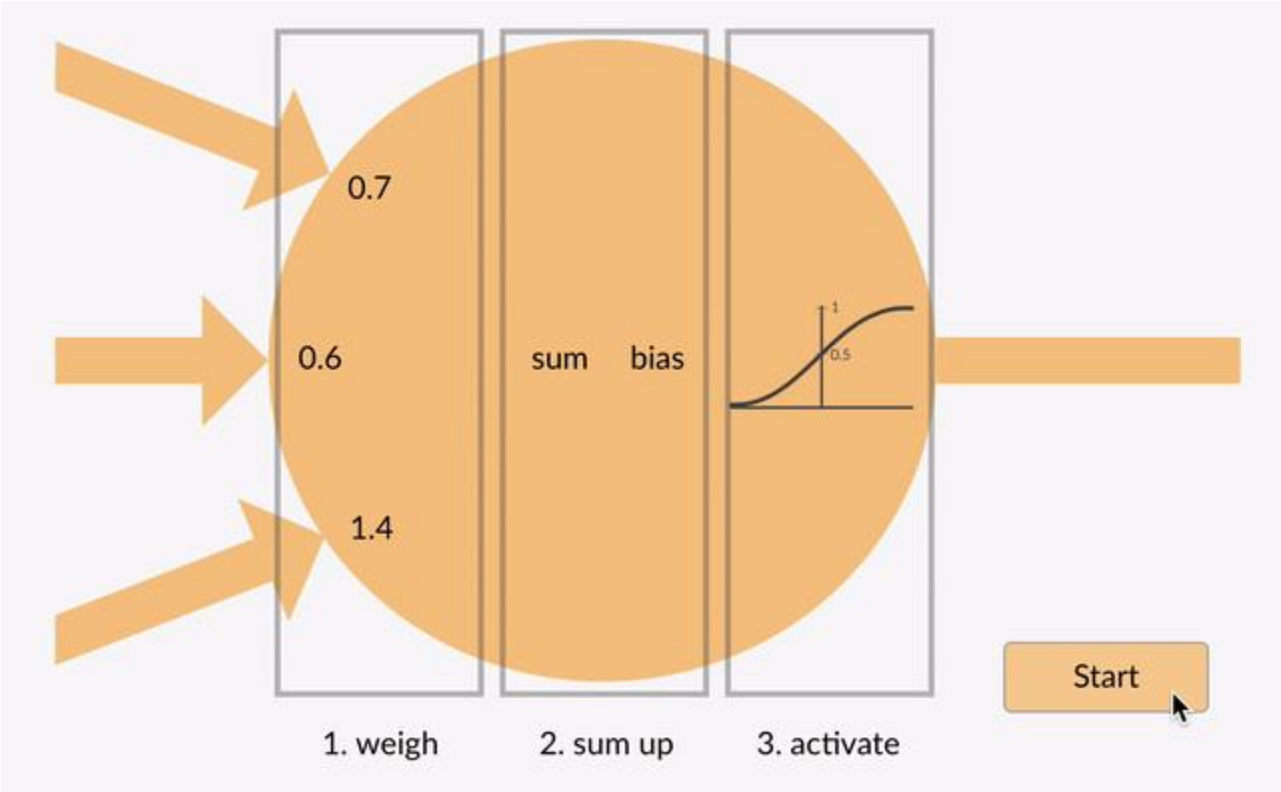
$$y_{in} = x_1 \cdot w_1 + x_2 \cdot w_2 + \cdots + x_m \cdot w_m$$

- $Net\ input$: $y_{in} = \sum_{i=1}^{m} x_i \cdot w_i$
- Đầu ra được tính bằng cách áp dụng hàm kích hoạt trên $net\ input$.

$$Y = F(y_{in})$$

- $Output = function\ netinput$

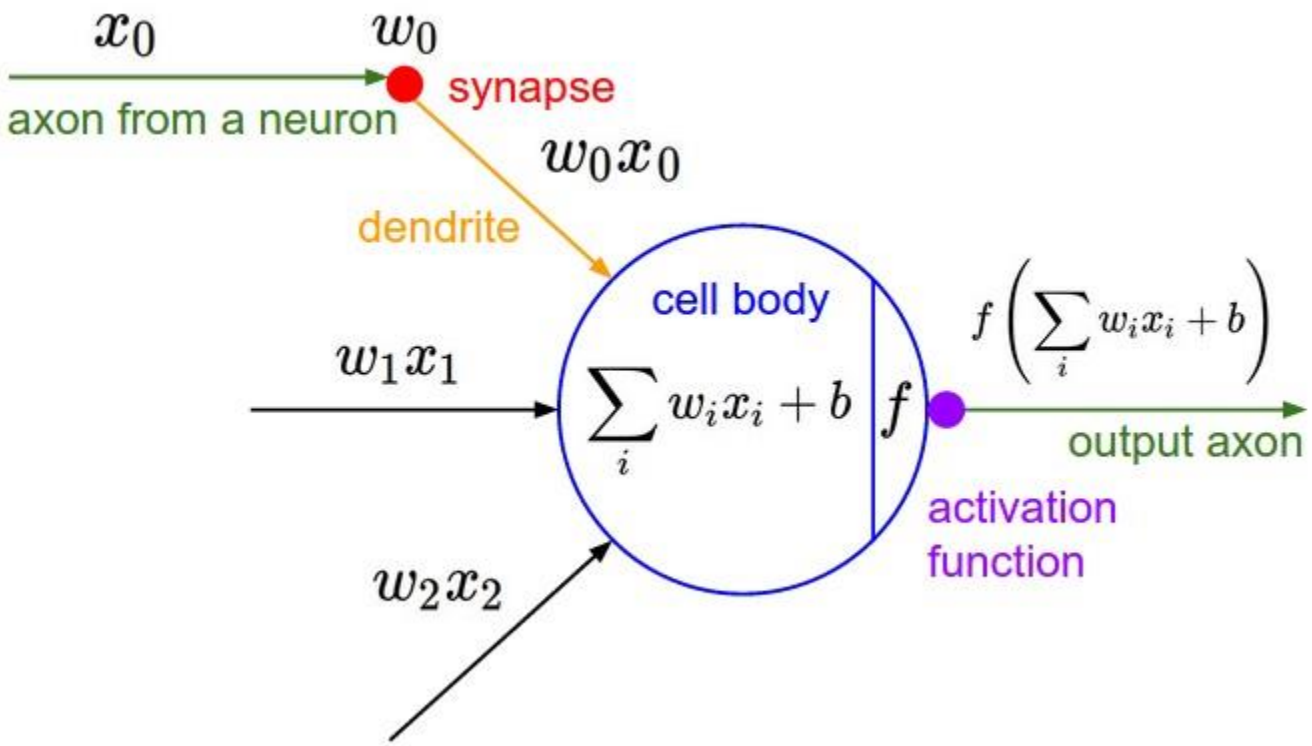
Artificial neuron



Neural networks

ACTIVATION FUNCTION

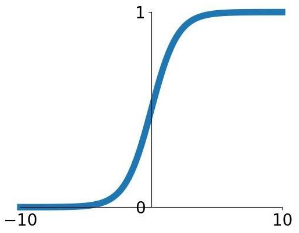
Activation function



Activation function

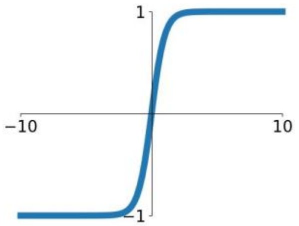
– Sigmoid

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



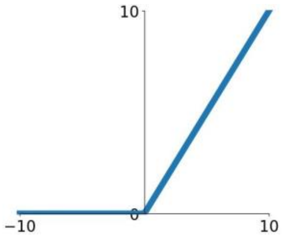
– tanh

$$tanh(x)$$



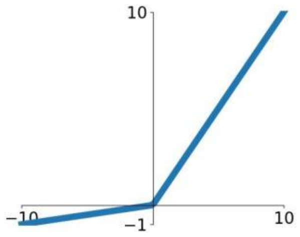
– ReLU

$$max(0, x)$$



– Leaky ReLU

$$max(0.1x, x)$$

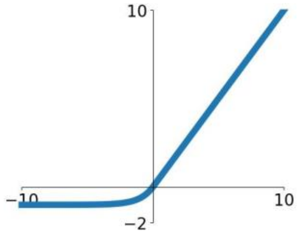


– Maxout

$$max(w_1^T x + b_1, w_2^T x + b_2)$$

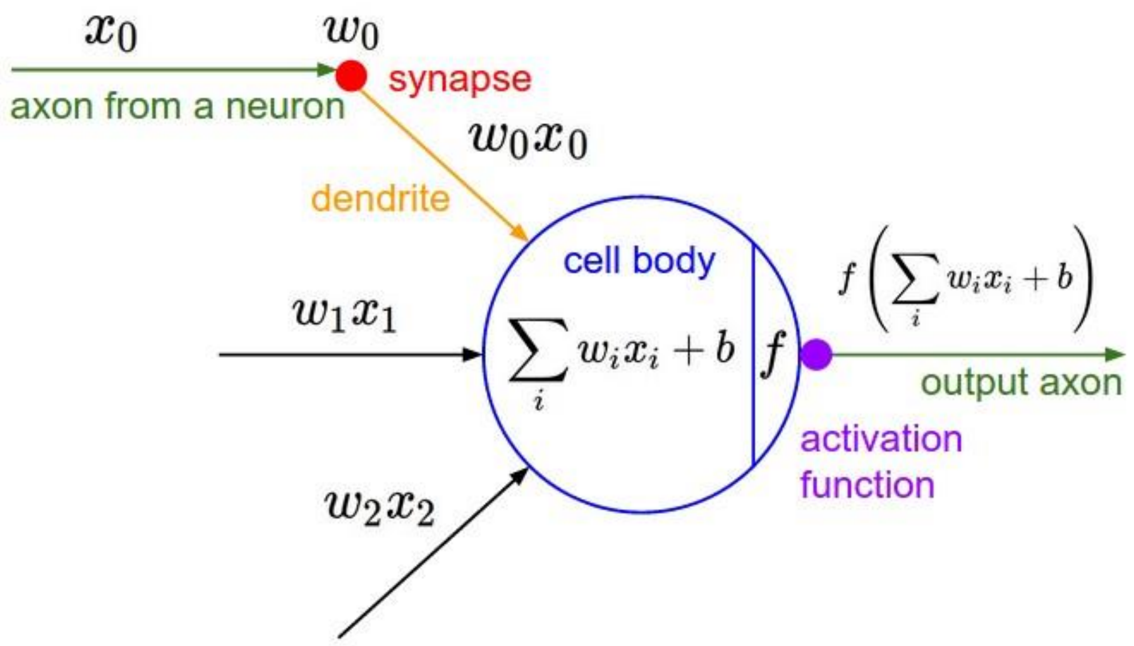
– ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



Activation function

- The output layer neurons most commonly have a different activation function
  + Softmax for class scores (classification).
  + Linear functions for real-valued target (regression).



Neural networks

LEARNING PROCESS

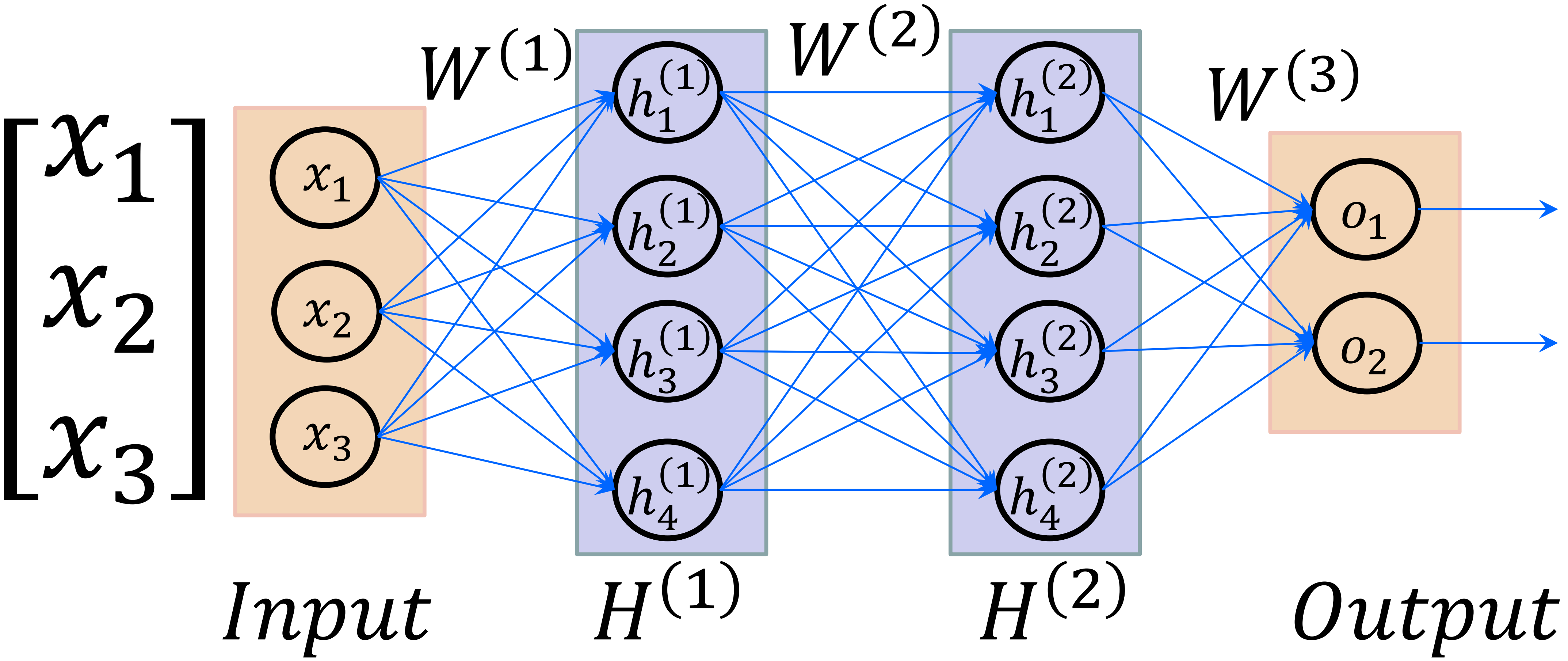
Learning process

- Bước 01: Initialize network weights (often small random values)
- Bước 02: For each datapoint i in training set:
  + Bước 2.1: Forward propagation.
  + Bước 2.2: Backward propagation.
  + Bước 2.3: Weight update.
- Bước 3: Return the network.

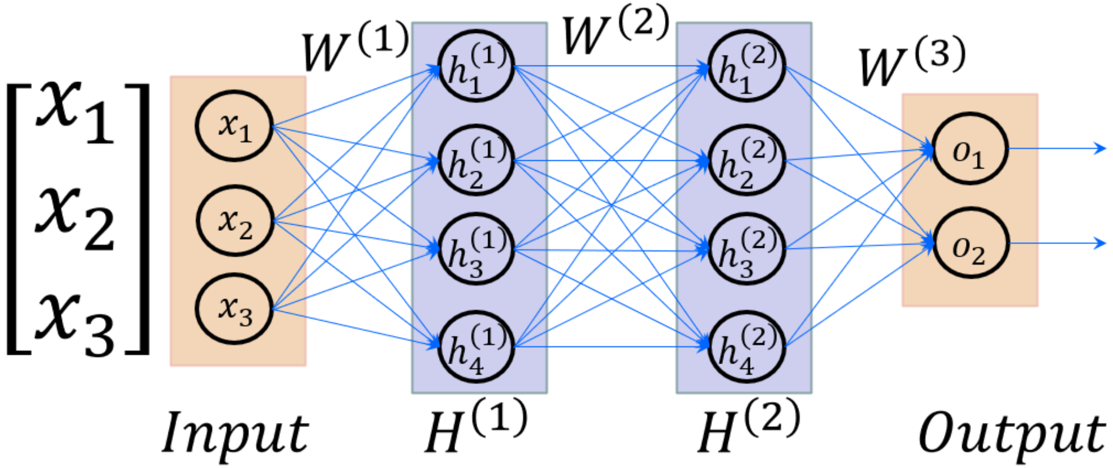
Neural networks

FORWARD PROPAGATION START

Forward propagation

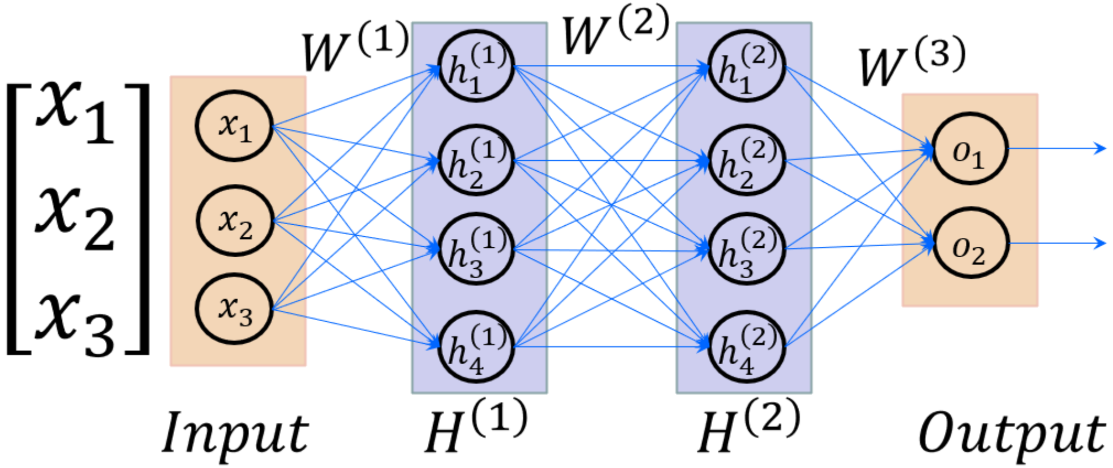


Forward propagation



– Câu hỏi 1: Neural networks này có bao nhiêu lớp?

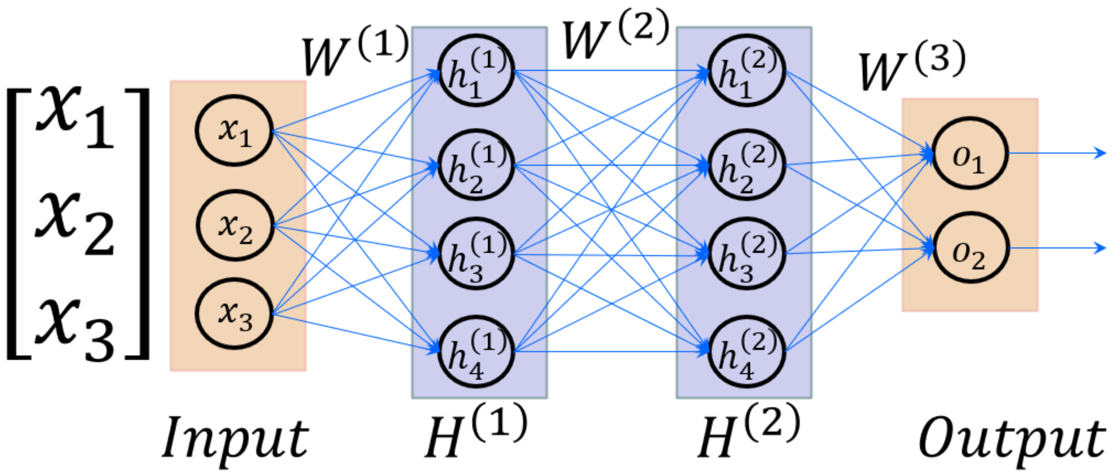
Forward propagation



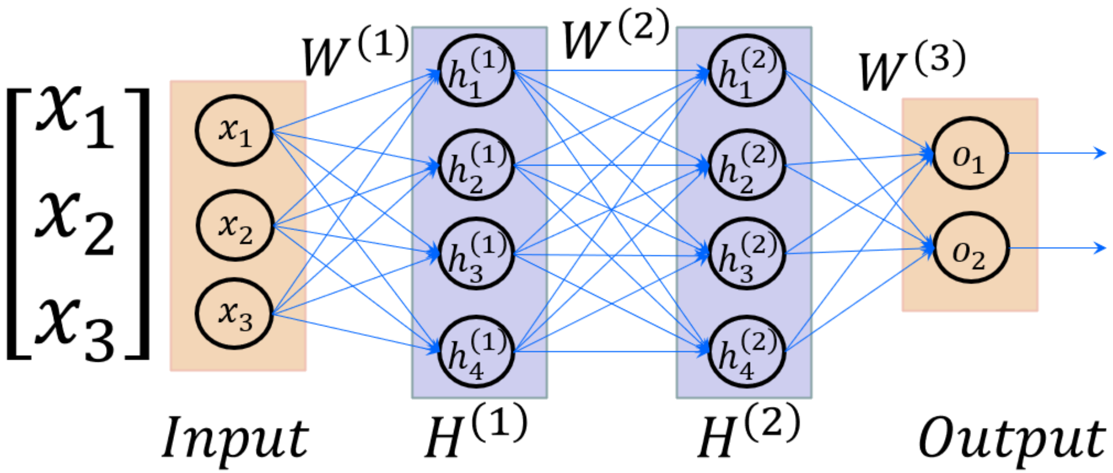
– Câu hỏi 2: Neural networks này lớp input có bao nhiêu phần tử?

Forward propagation

– Câu hỏi 3: Lớp ẩn thứ nhất có bao nhiêu nơ ron?

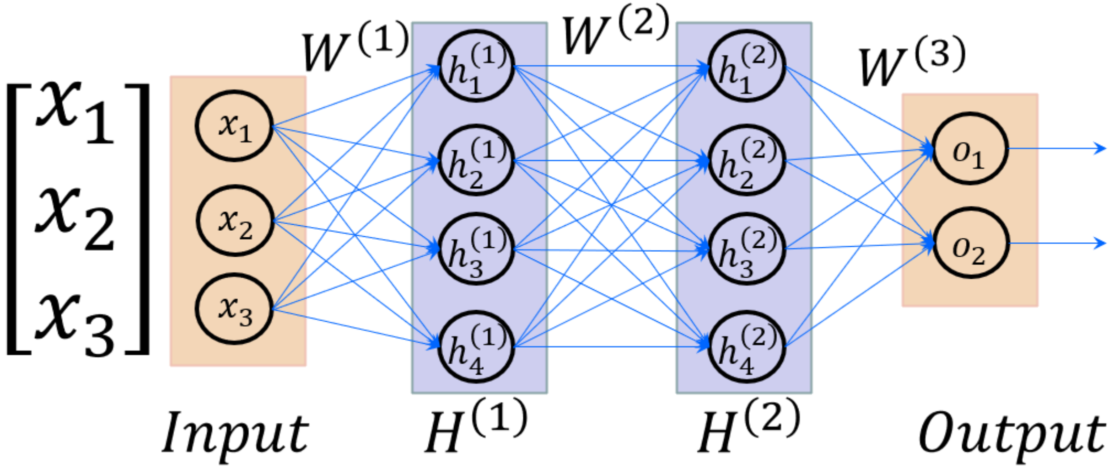


Forward propagation



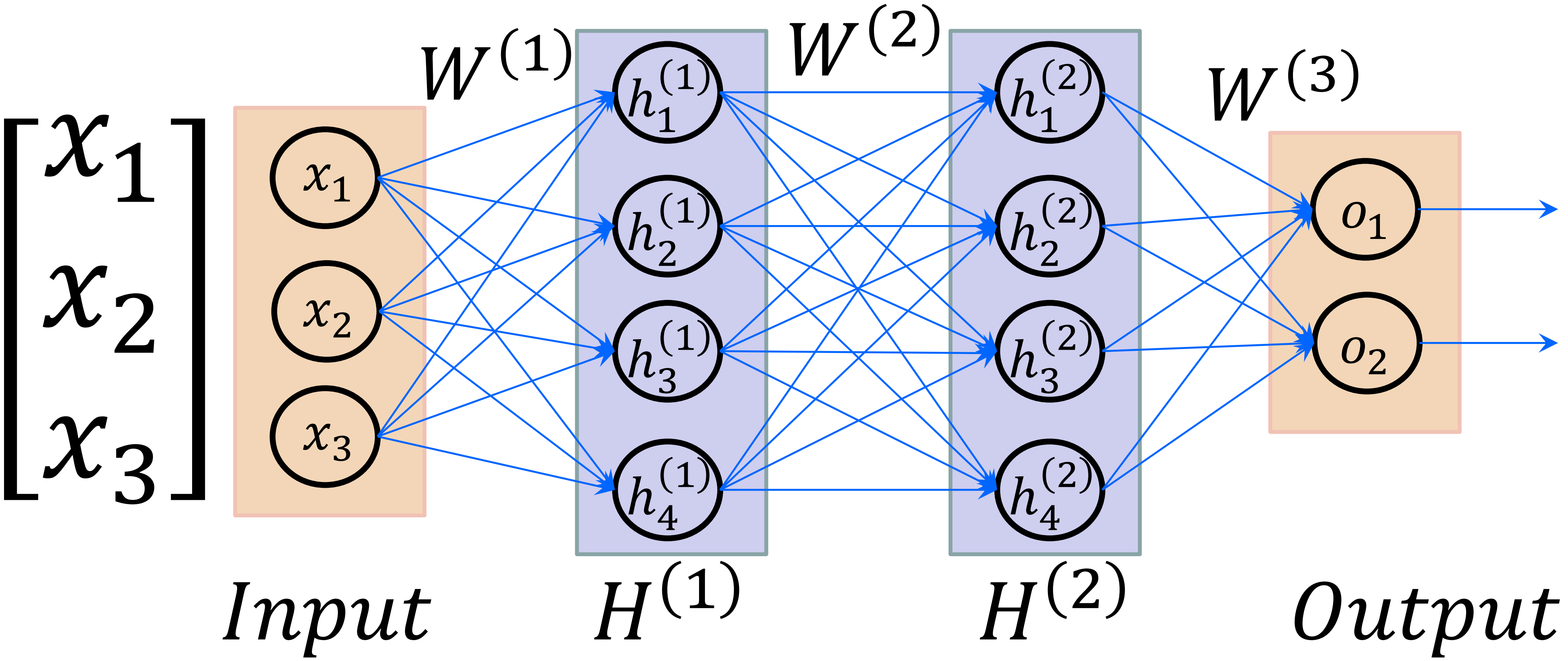
– Câu hỏi 4: Lớp ẩn thứ hai có bao nhiêu nơ ron?

Forward propagation

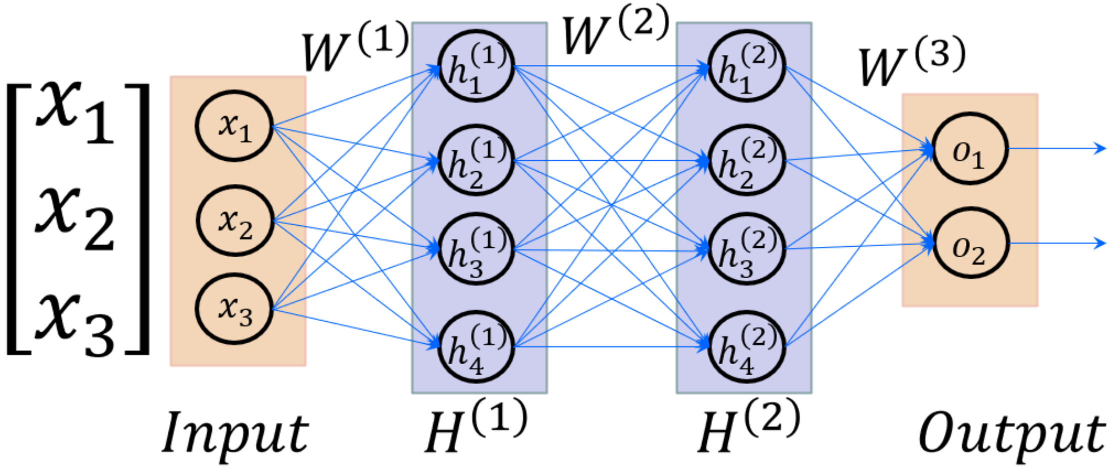


– Câu hỏi 5: Lớp output có bao nhiêu nơ ron?

Forward propagation



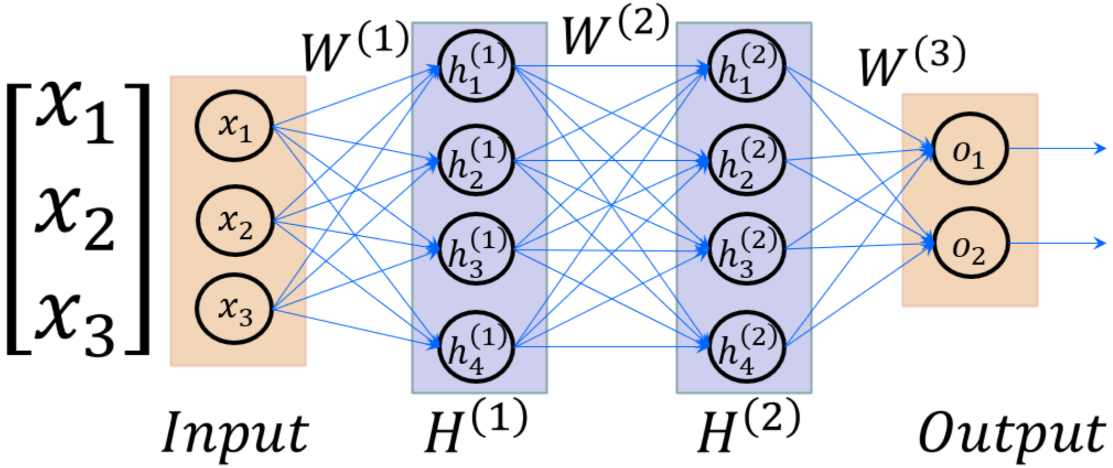
Forward propagation



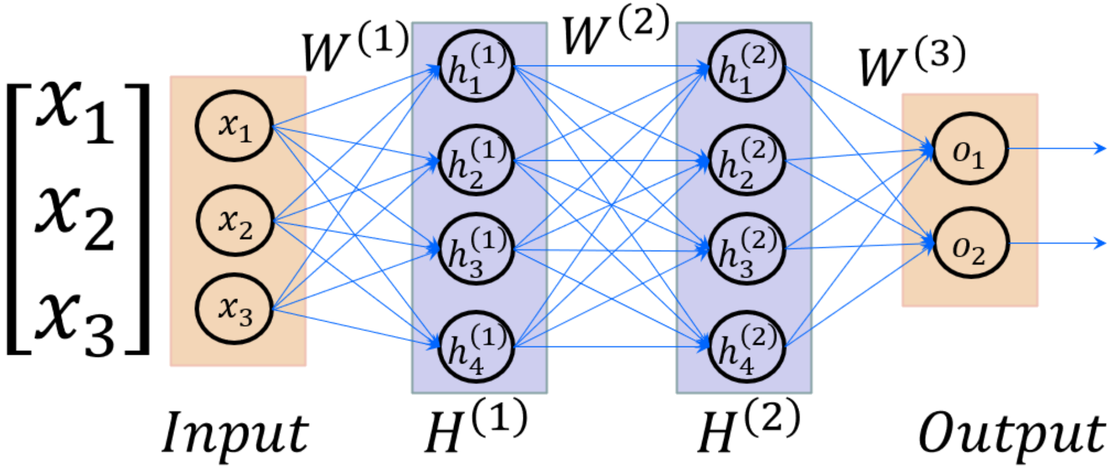
– Câu hỏi 1: Ma trận trọng số $W^{(1)}$ có kích thước bao nhiêu?

Forward propagation

– Câu hỏi 2: Ma trận trọng số $W^{(2)}$ có kích thước bao nhiêu?



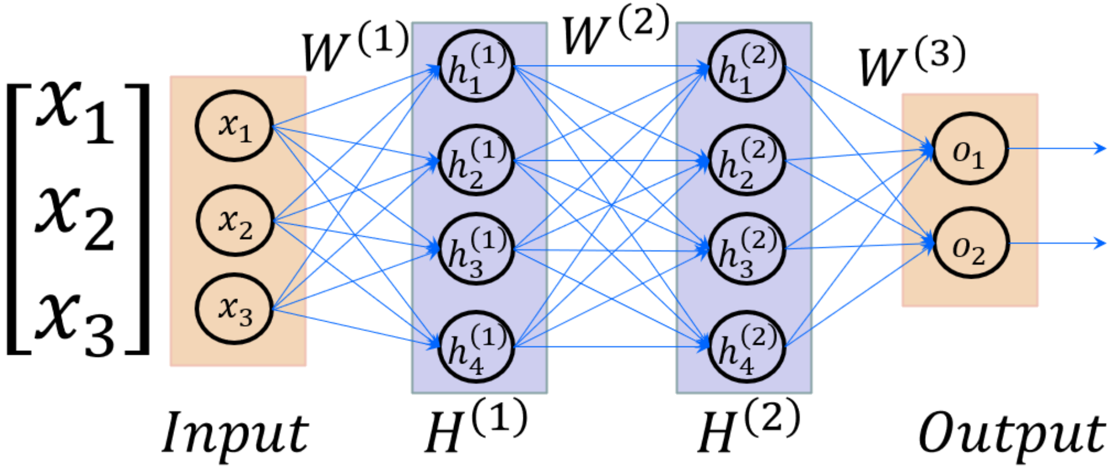
Forward propagation



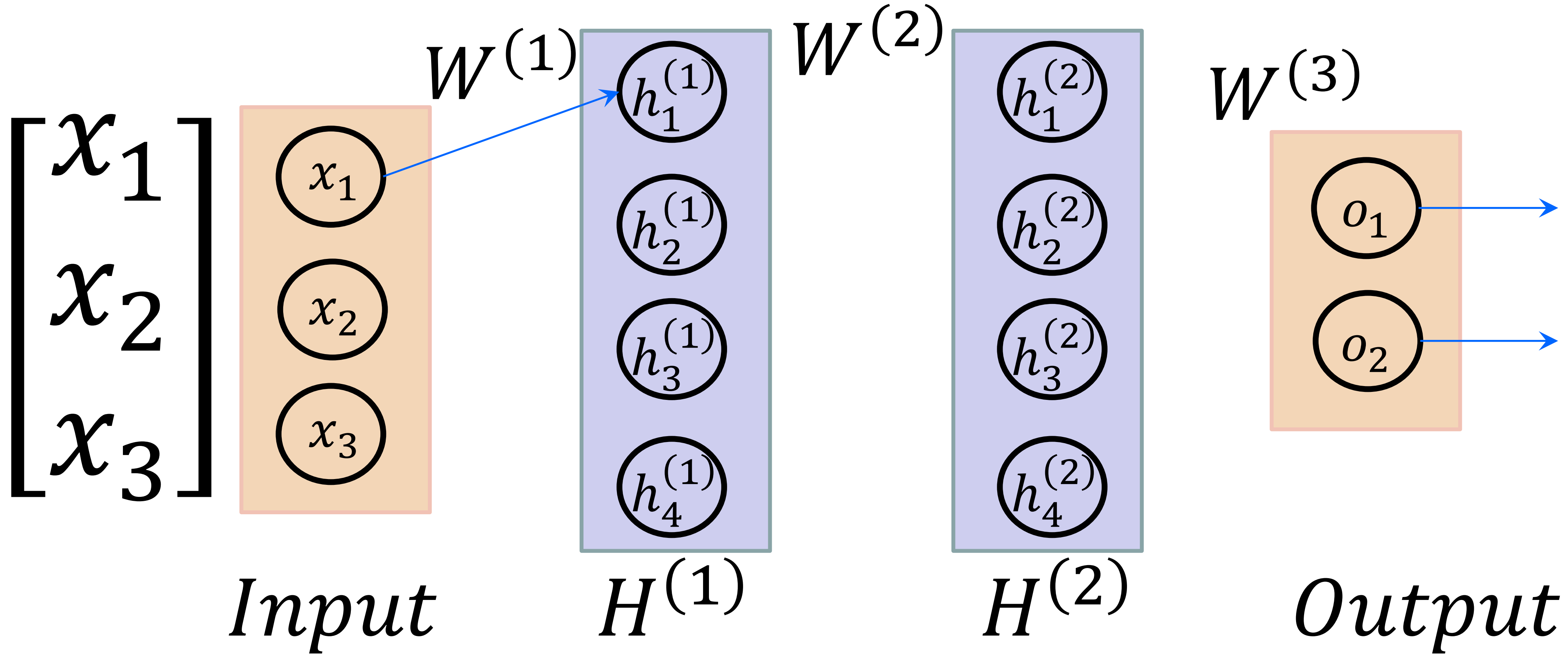
– Câu hỏi 3: Ma trận trọng số $W^{(3)}$ có kích thước bao nhiêu?

Forward propagation

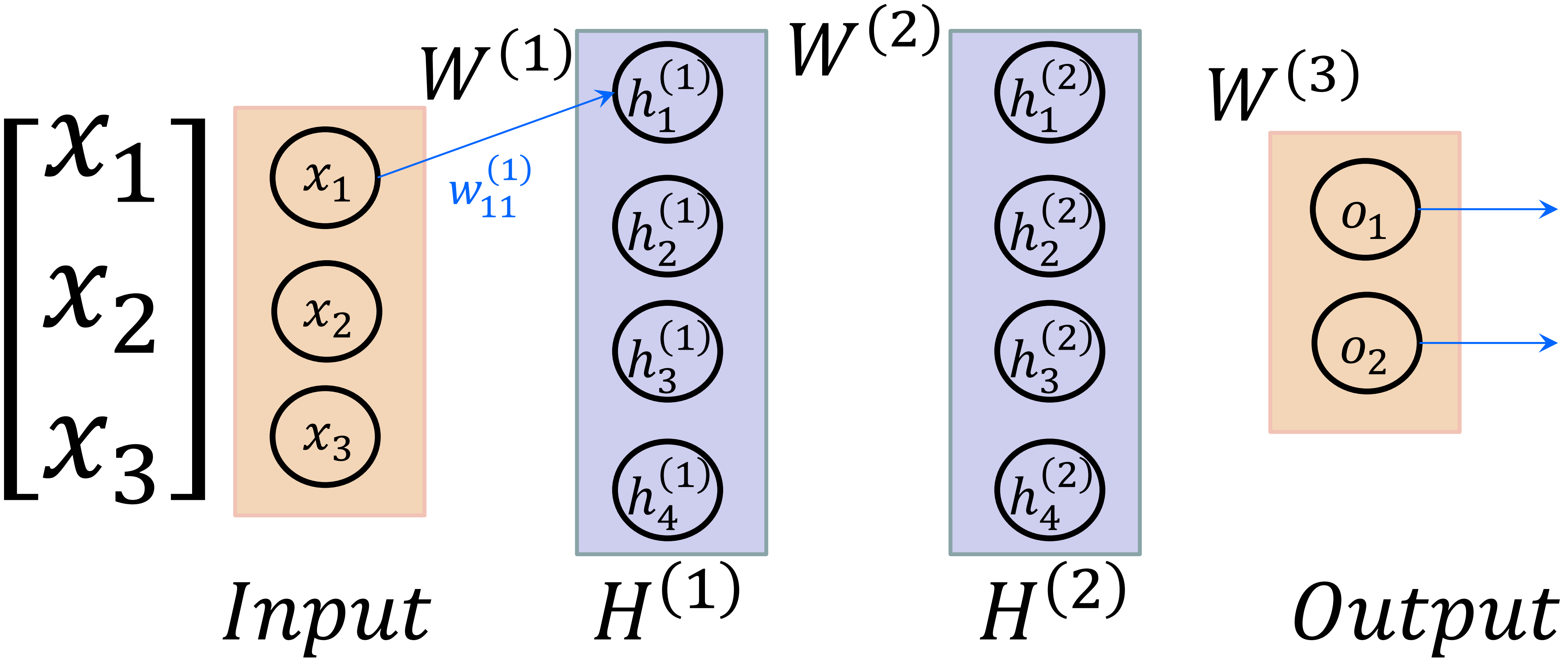
– Câu hỏi 4: Ma trận trọng số $W^{(l)}$ có kích thước bao nhiêu?



Forward propagation



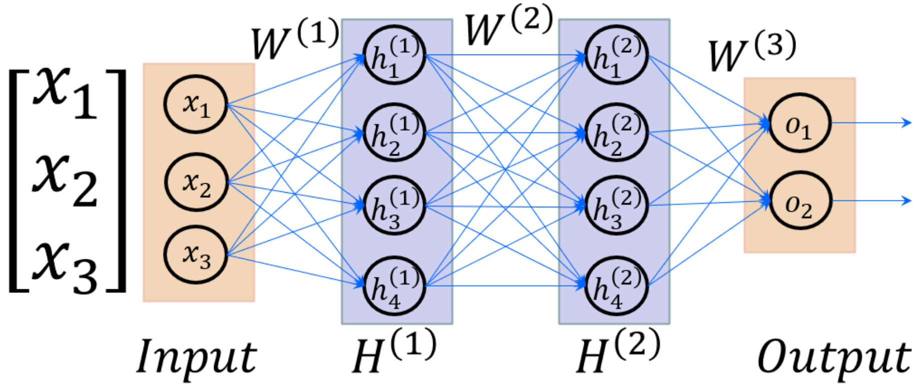
Forward propagation



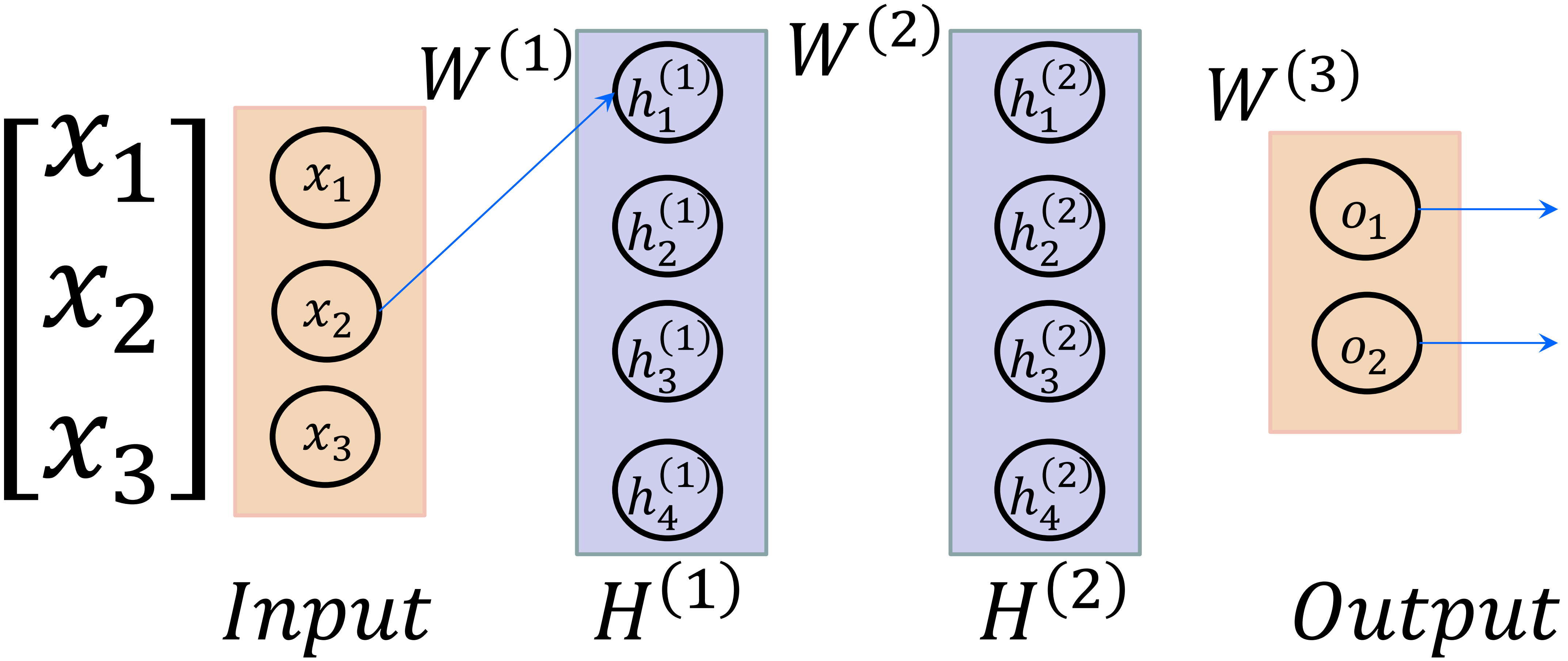
Forward propagation

– Ma trận trọng số $W^{(1)}$

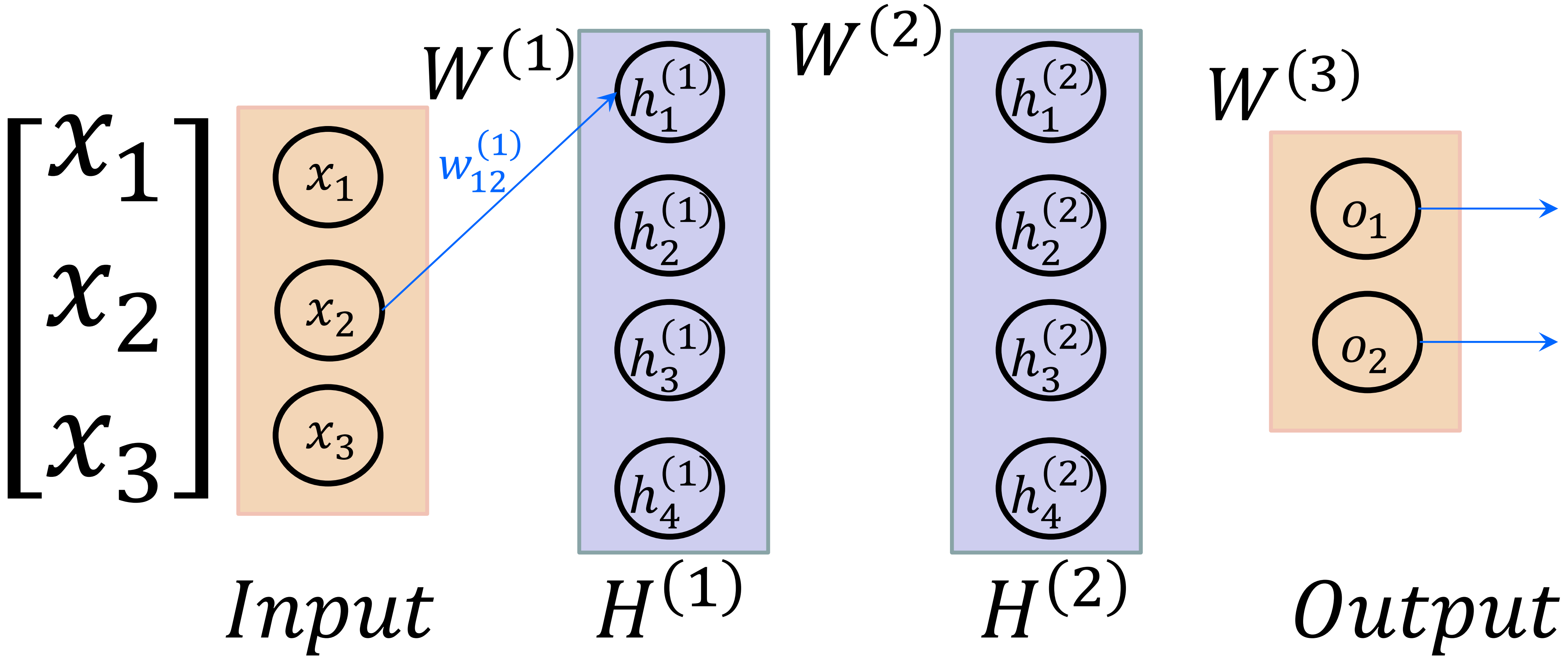
$$W^{(1)} = \begin{bmatrix} w_{11}^{(1)} & & & \\ & & & \\ & & & \\ & & & \end{bmatrix}$$



Forward propagation



Forward propagation

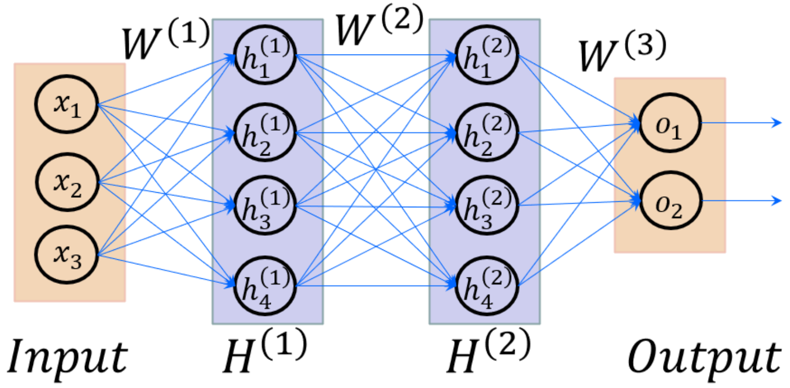


Forward propagation

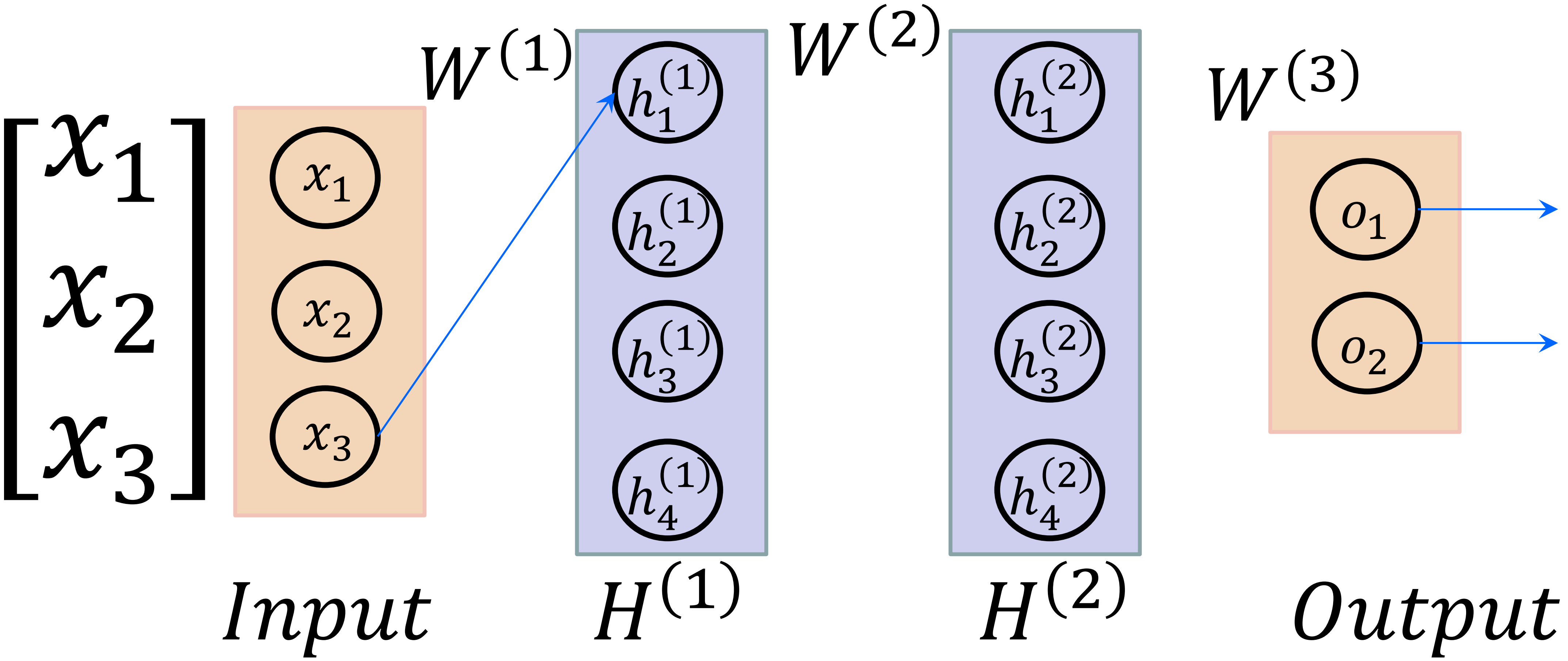
– Ma trận trọng số $W^{(1)}$

$$W^{(1)} = \begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & & \\ & & & \\ & & & \\ & & & \end{bmatrix}$$

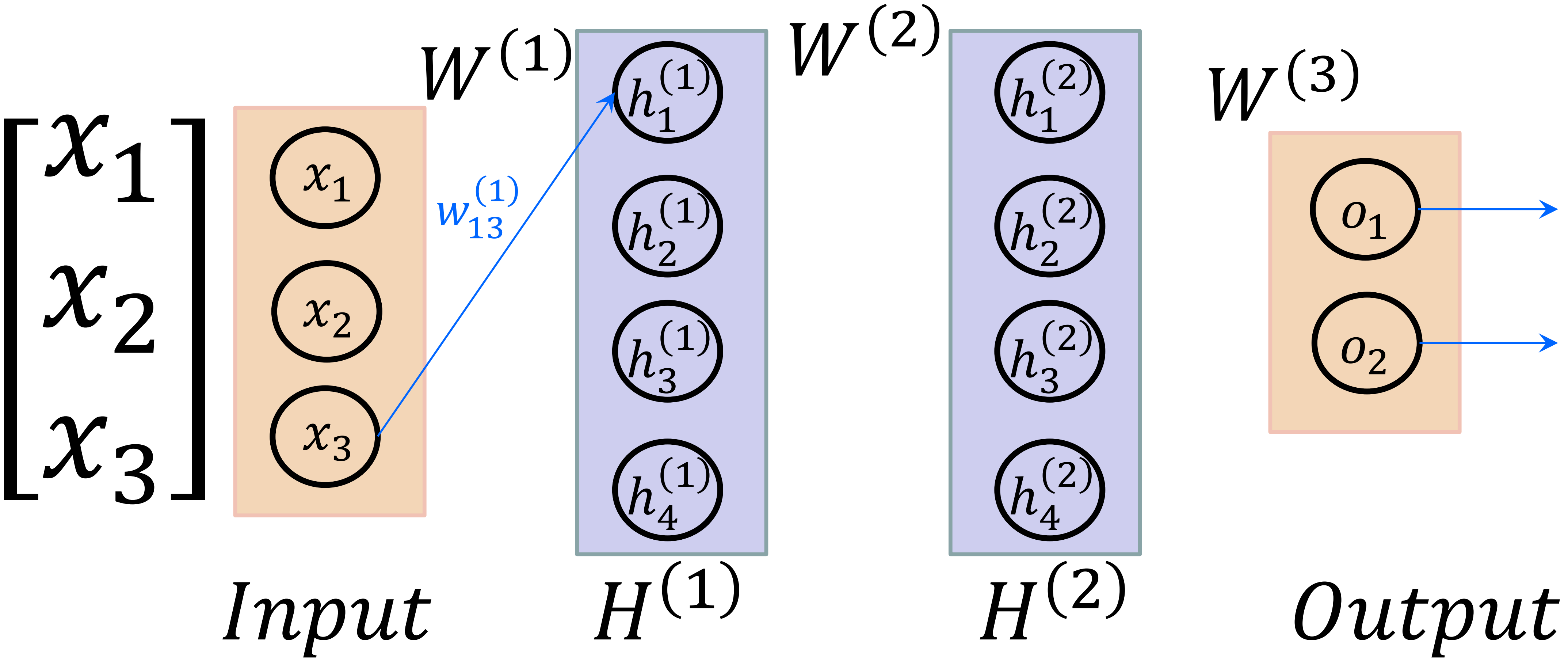
$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$



Forward propagation



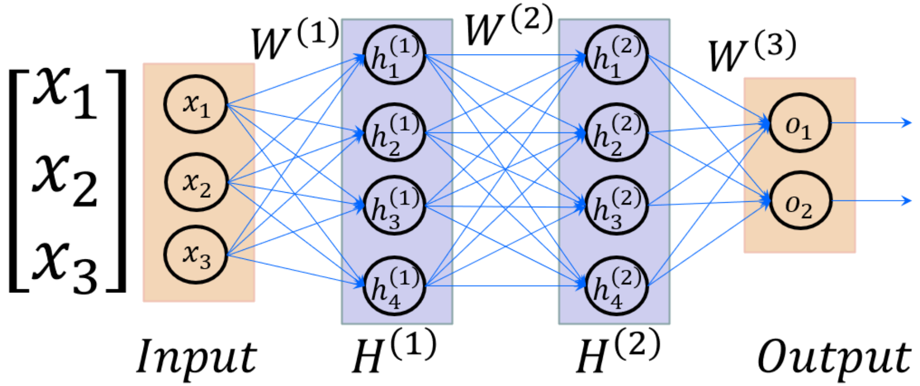
Forward propagation



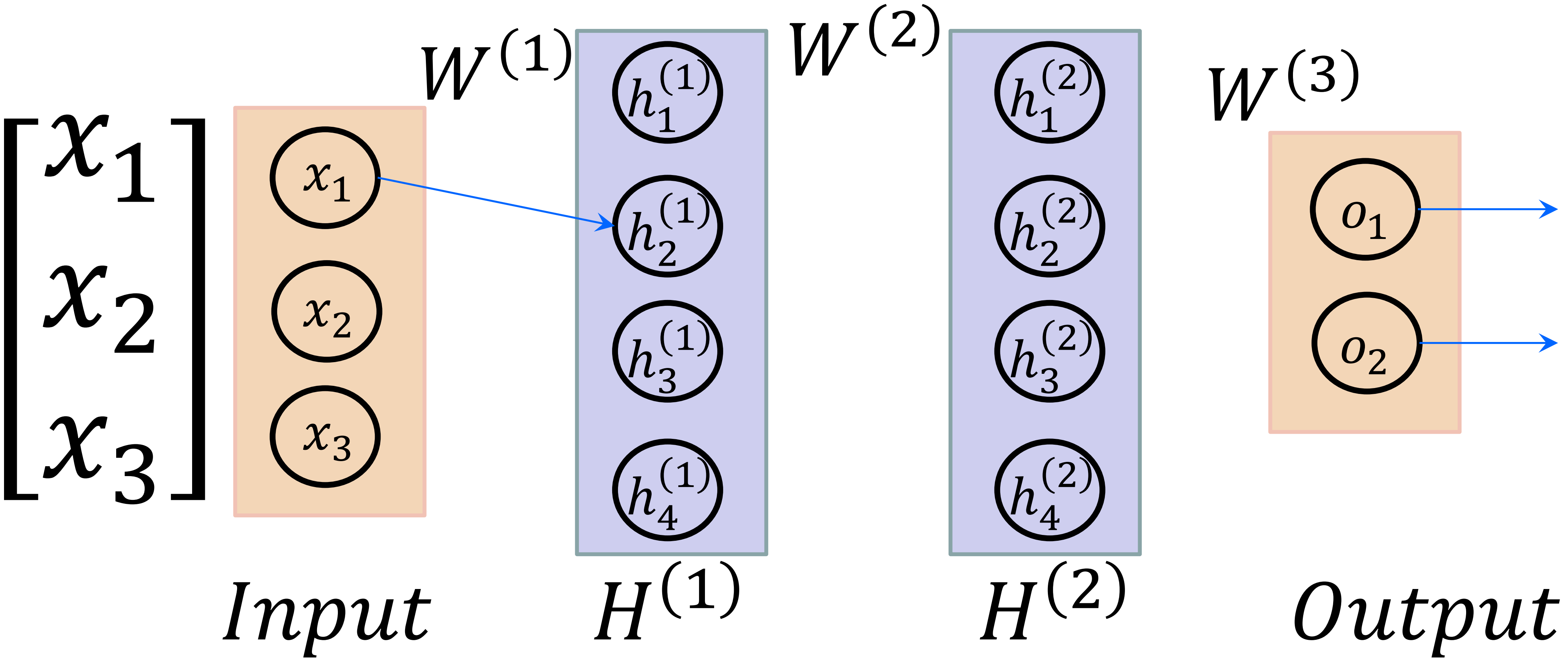
Forward propagation

– Ma trận trọng số $W^{(1)}$

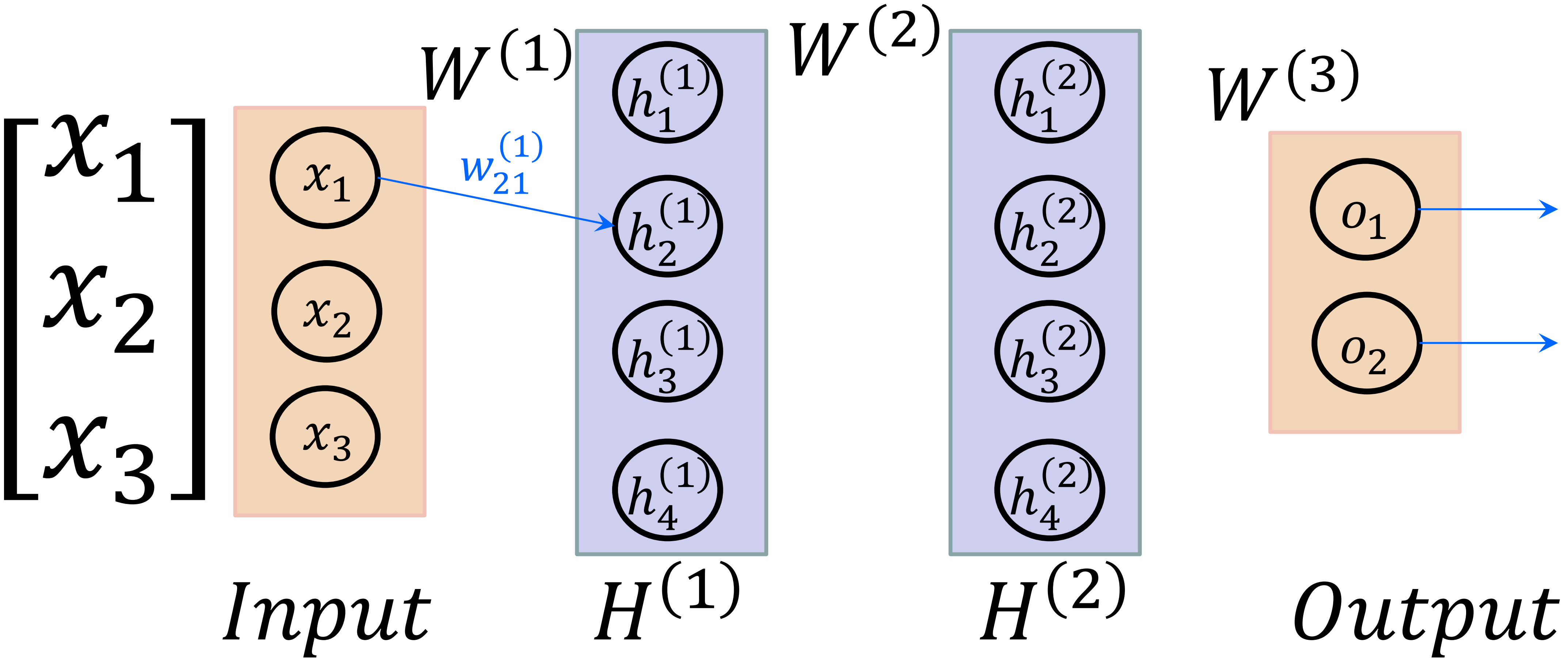
$$W^{(1)} = \begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} \\ & & \\ & & \\ & & \end{bmatrix}$$



Forward propagation



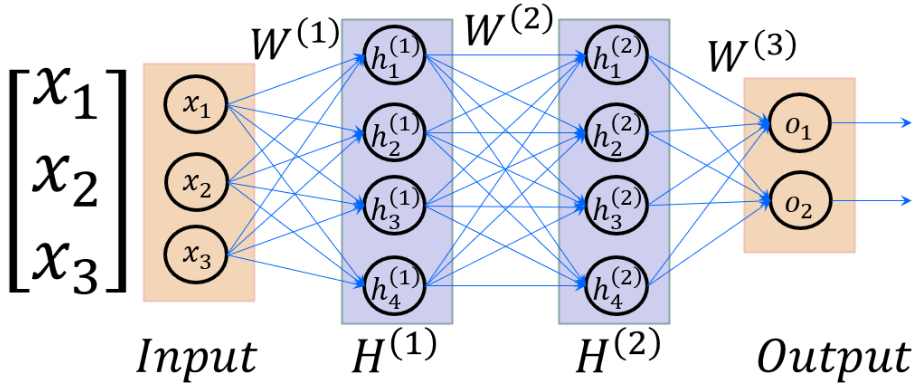
Forward propagation



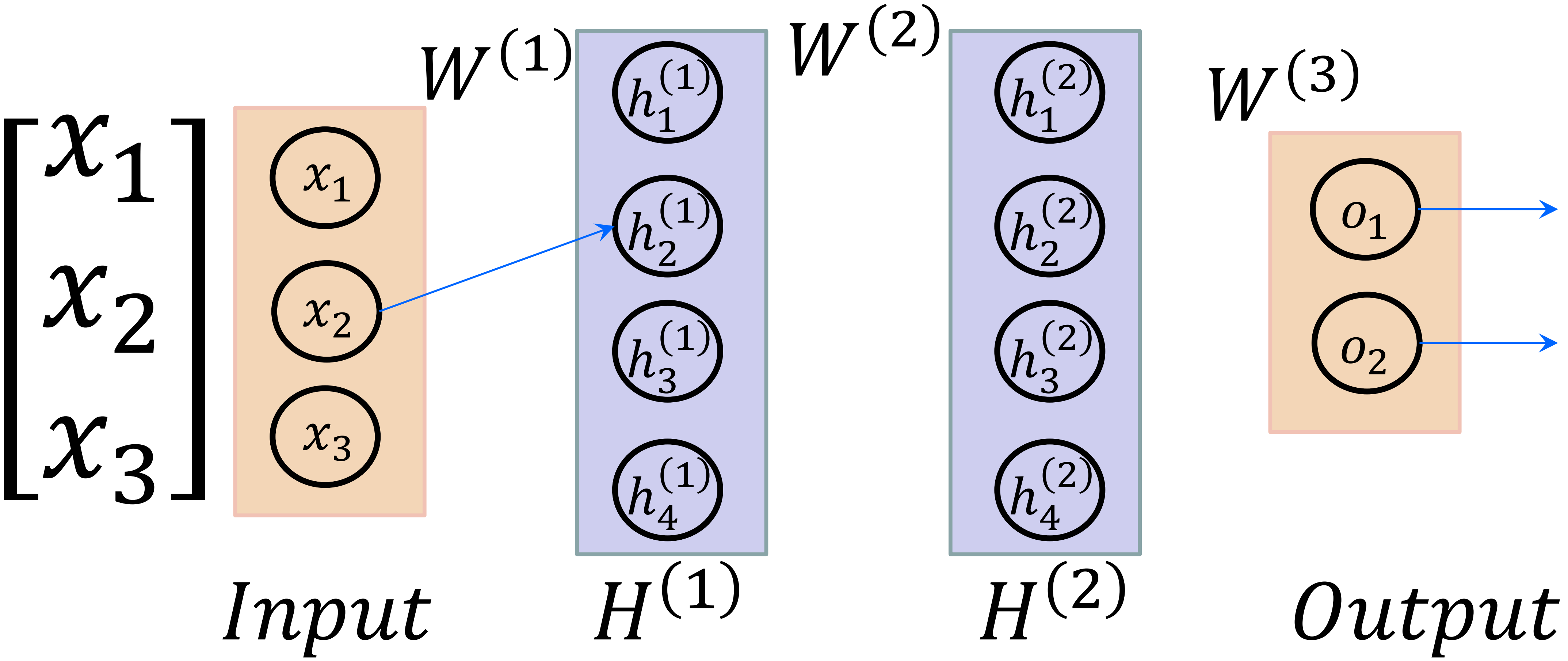
Forward propagation

– Ma trận trọng số $W^{(1)}$

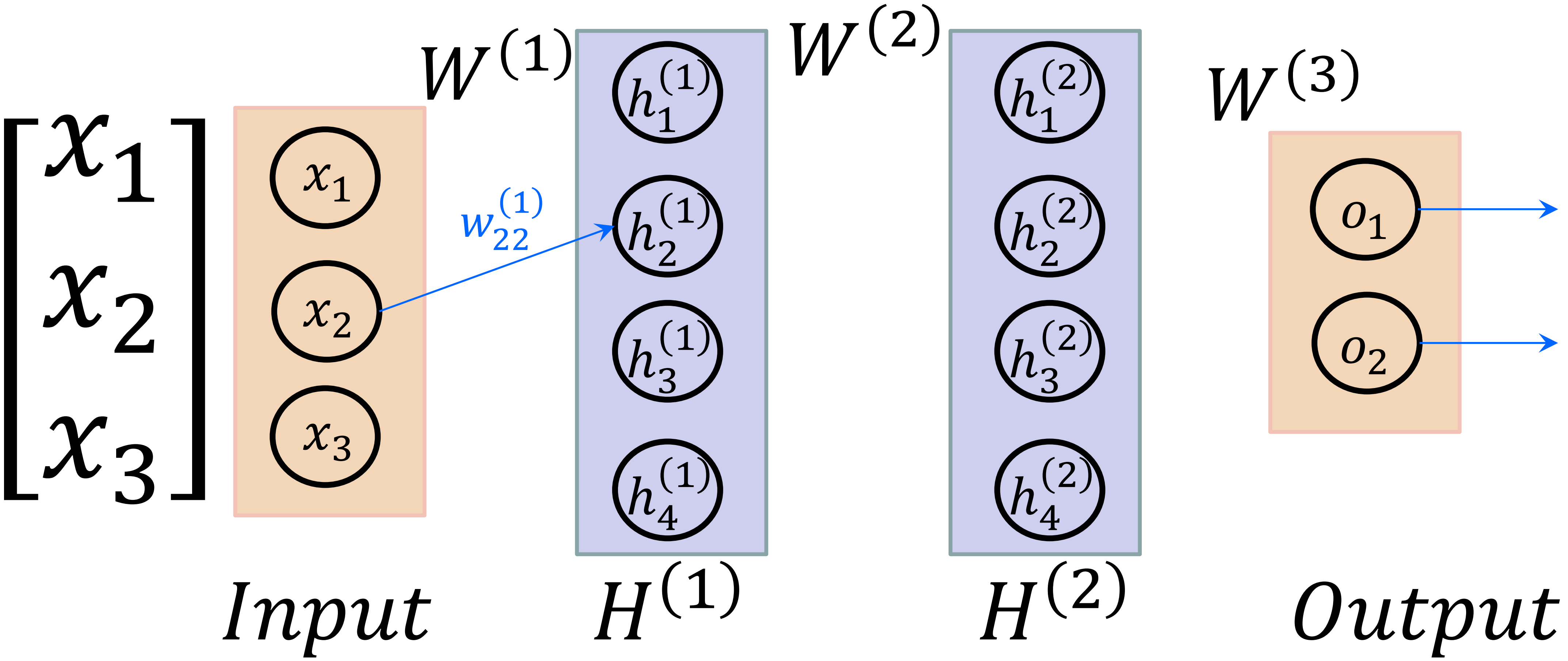
$$W^{(1)} = \begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} \\ w_{21}^{(1)} & & \\ & & \\ & & \end{bmatrix}$$



Forward propagation



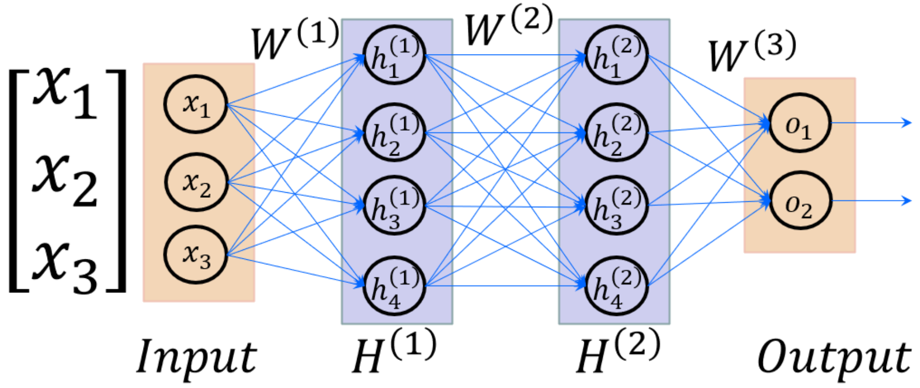
Forward propagation



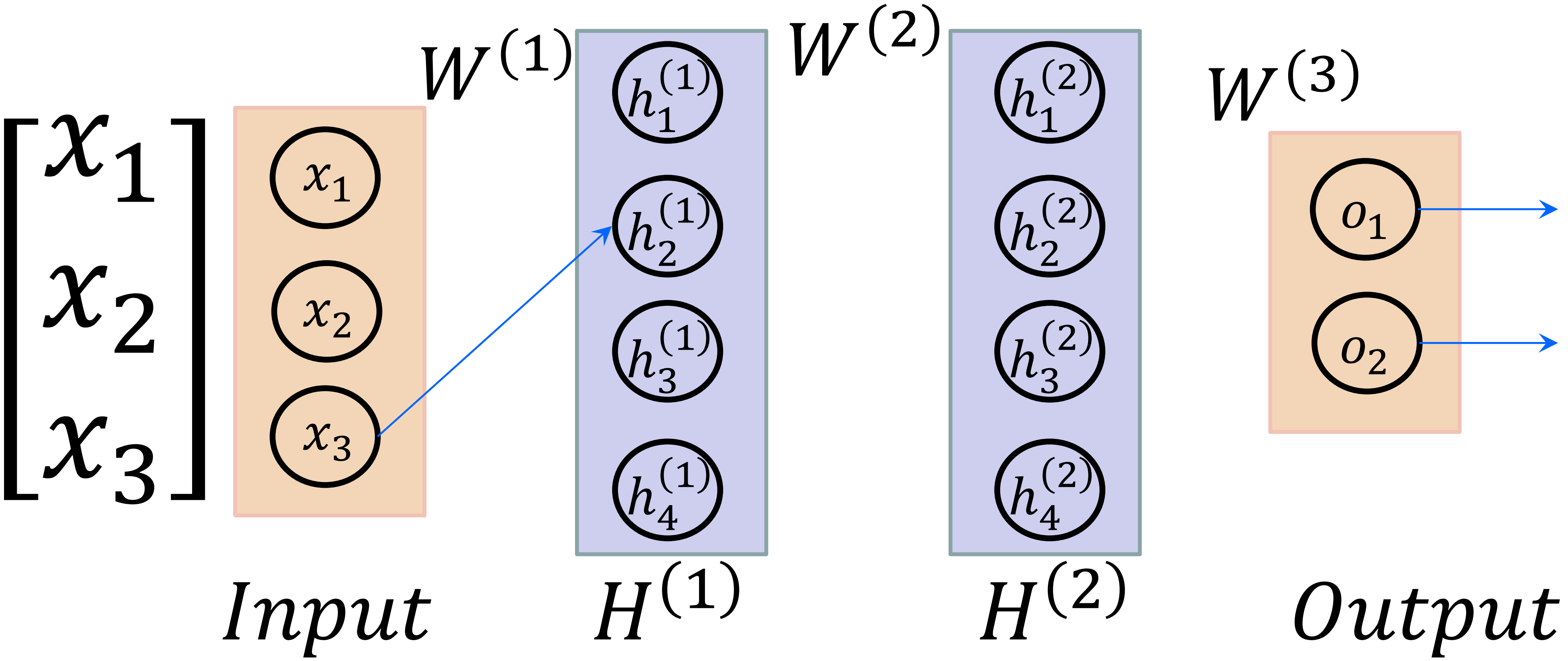
Forward propagation

– Ma trận trọng số $W^{(1)}$

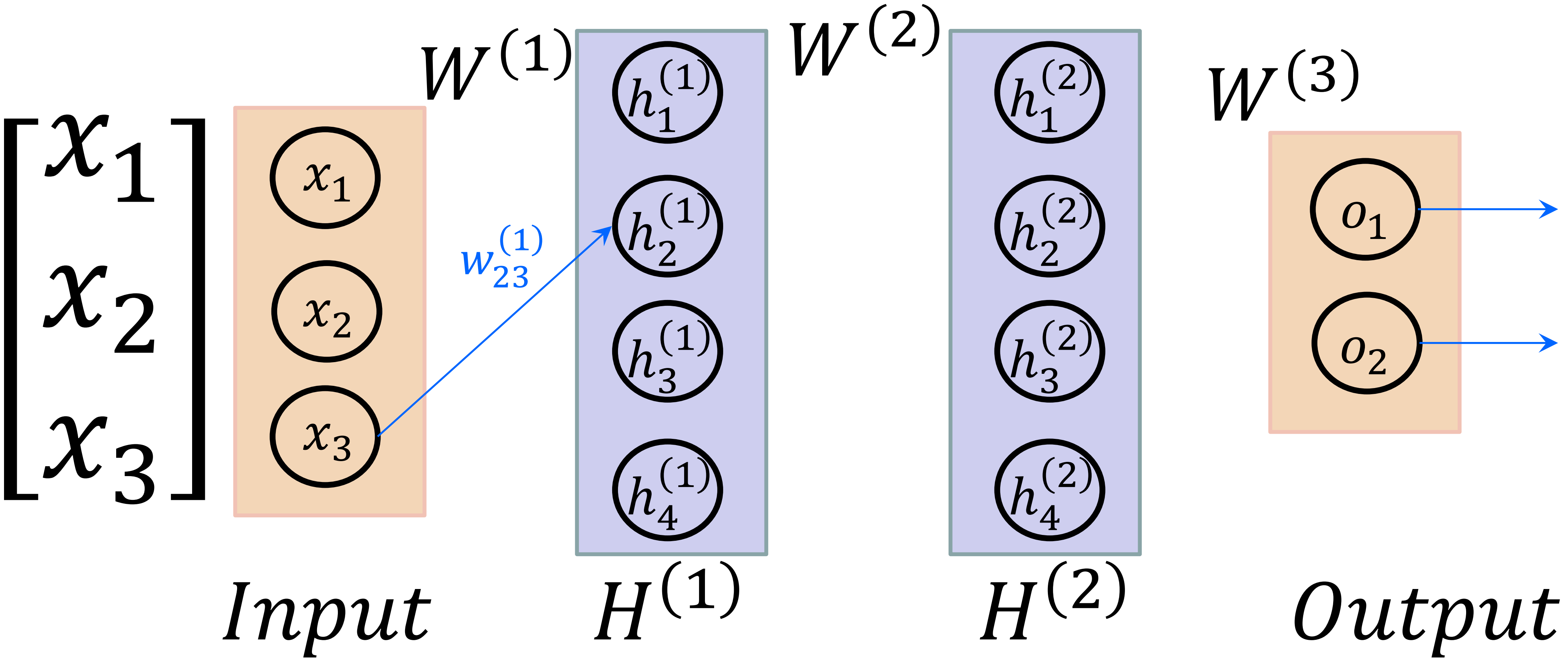
$$W^{(1)} = \begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & \\ & & \\ & & \end{bmatrix}$$



Forward propagation



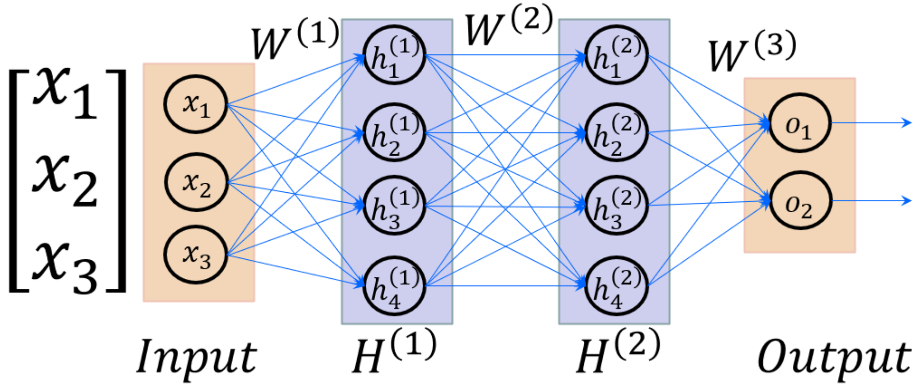
Forward propagation



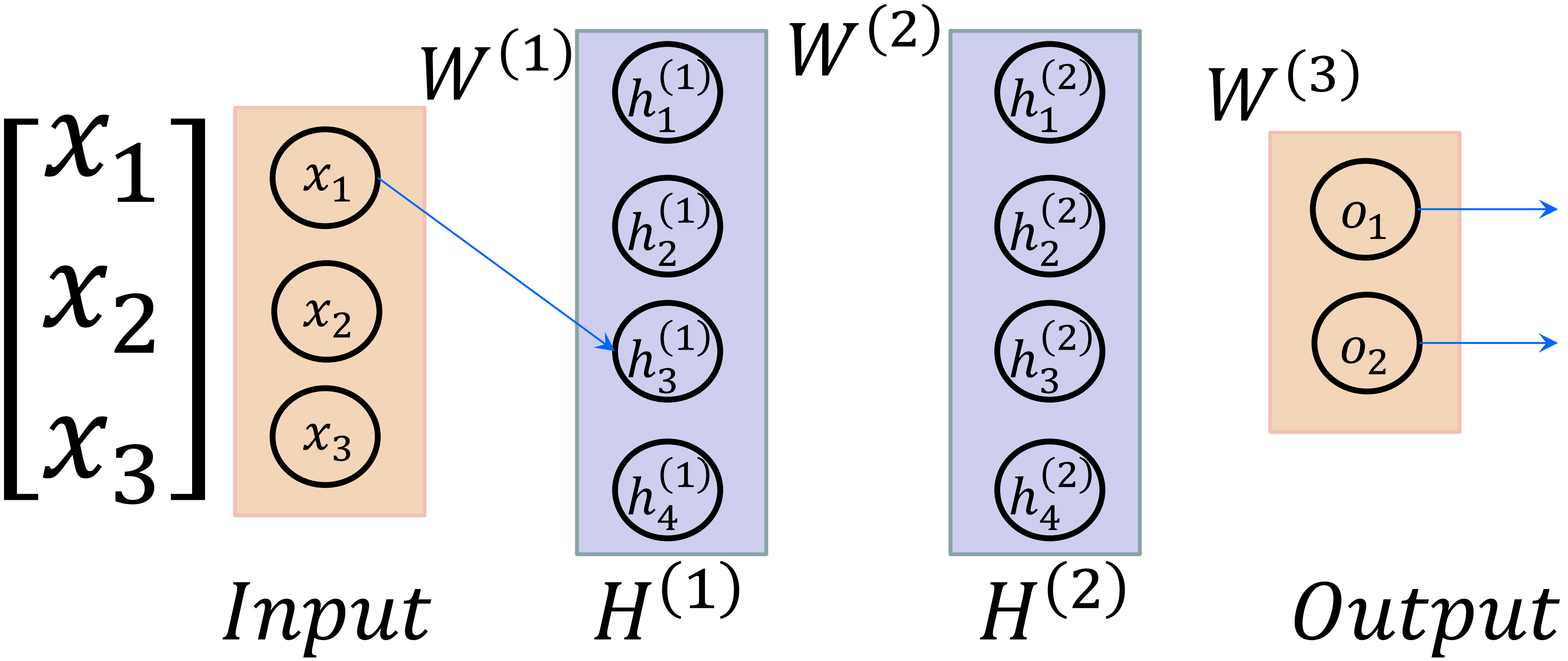
Forward propagation

– Ma trận trọng số $W^{(1)}$

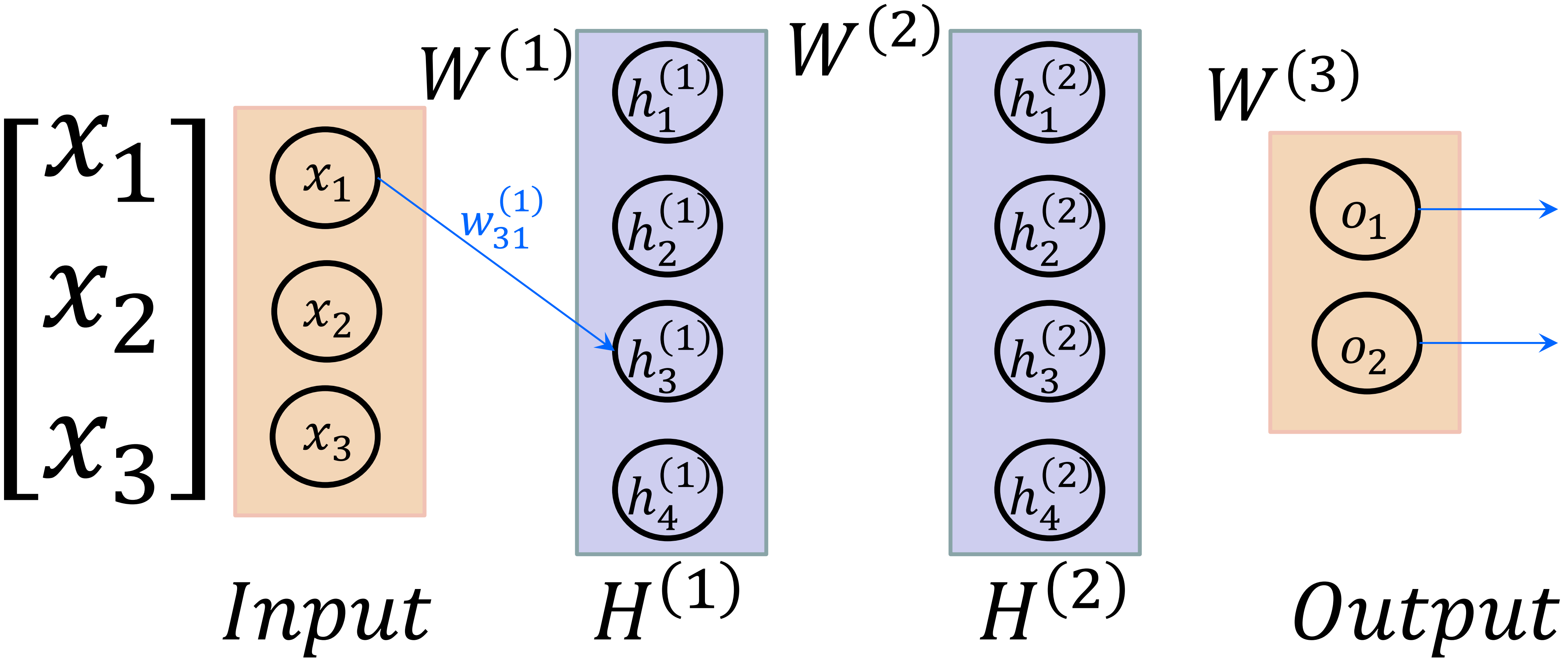
$$W^{(1)} = \begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & w_{23}^{(1)} \\ & & \\ & & \end{bmatrix}$$



Forward propagation



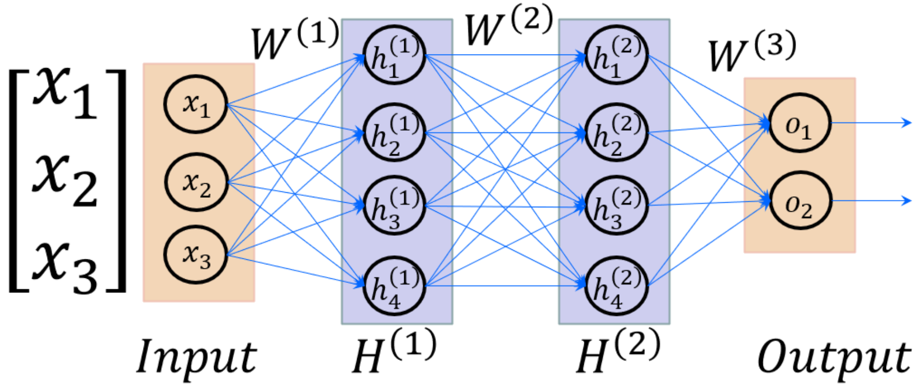
Forward propagation



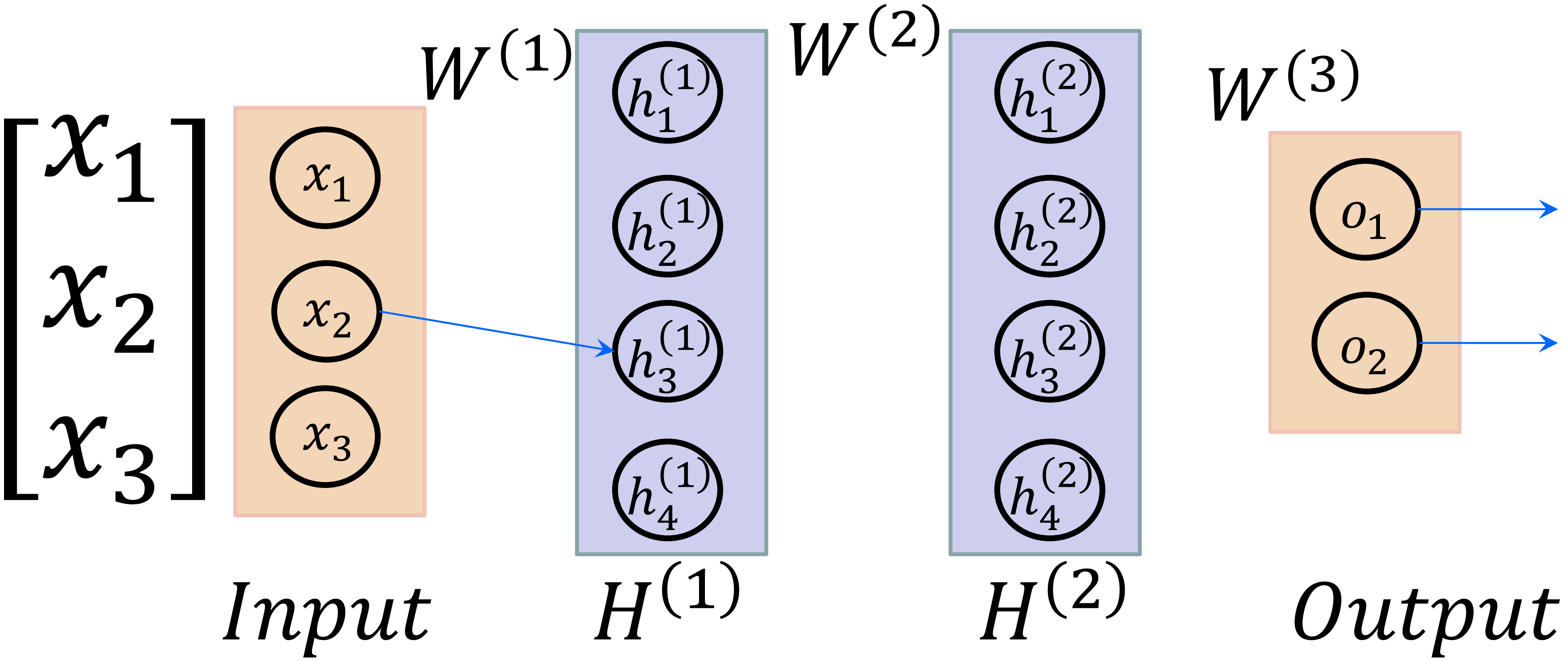
Forward propagation

– Ma trận trọng số $W^{(1)}$

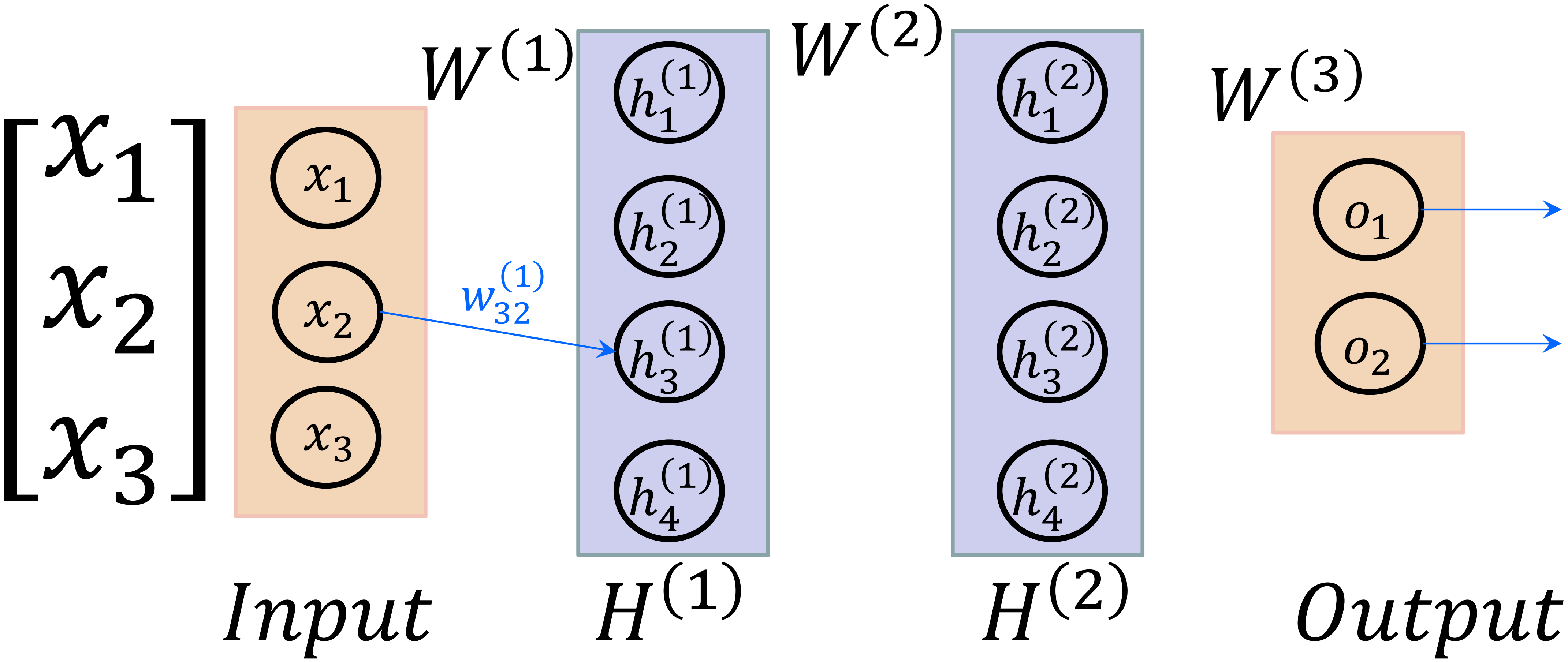
$$W^{(1)} = \begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & w_{23}^{(1)} \\ w_{31}^{(1)} & & \\ & & \end{bmatrix}$$



Forward propagation



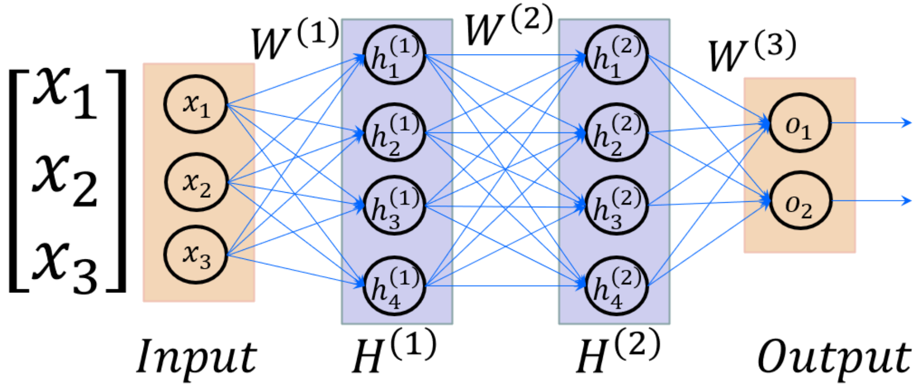
Forward propagation



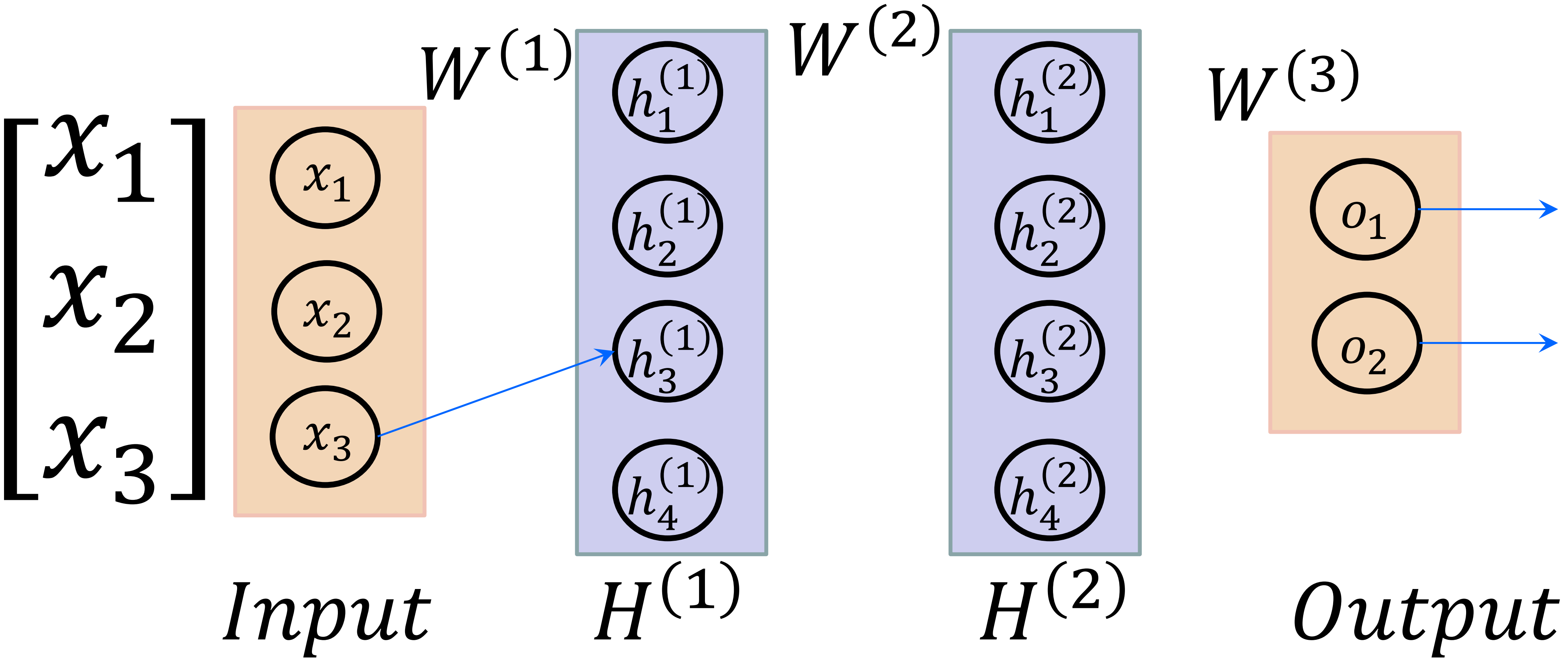
Forward propagation

– Ma trận trọng số $W^{(1)}$

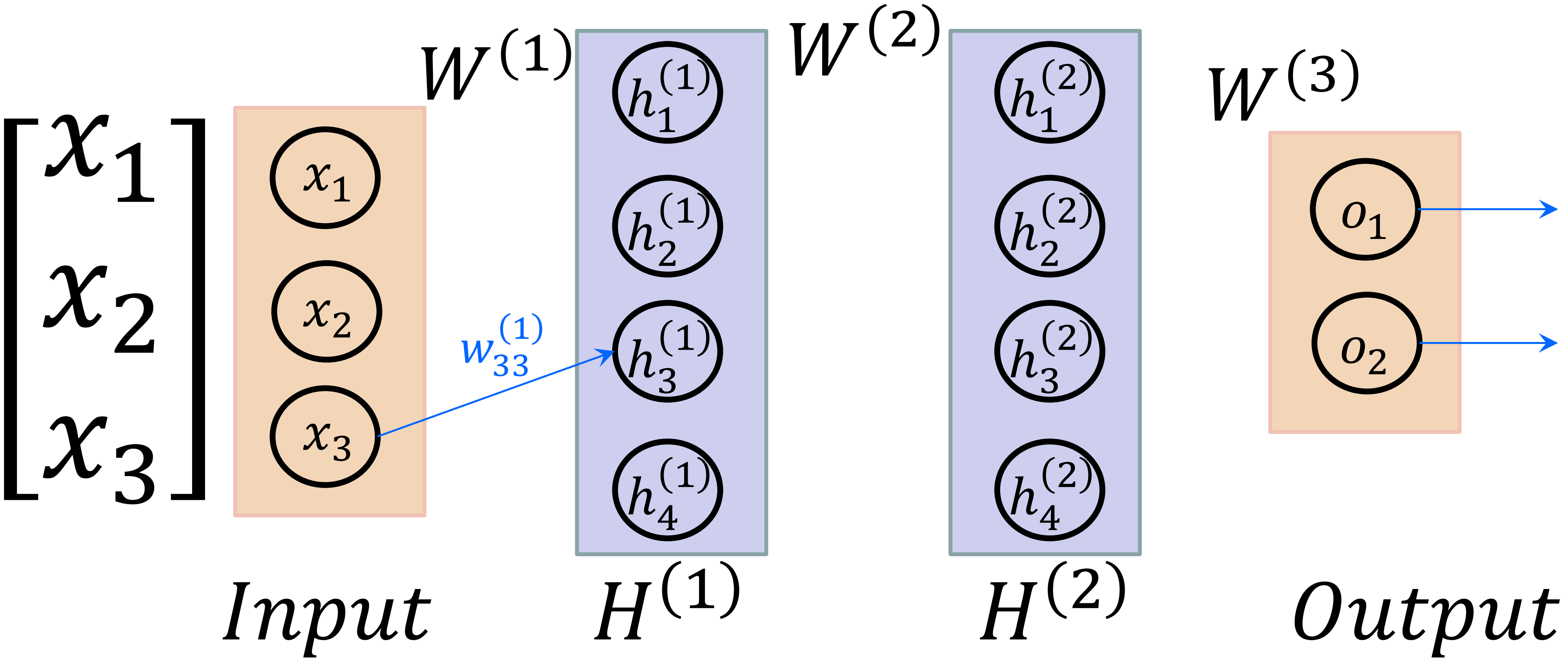
$$W^{(1)} = \begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & w_{23}^{(1)} \\ w_{31}^{(1)} & w_{32}^{(1)} & \\ & & \end{bmatrix}$$



Forward propagation



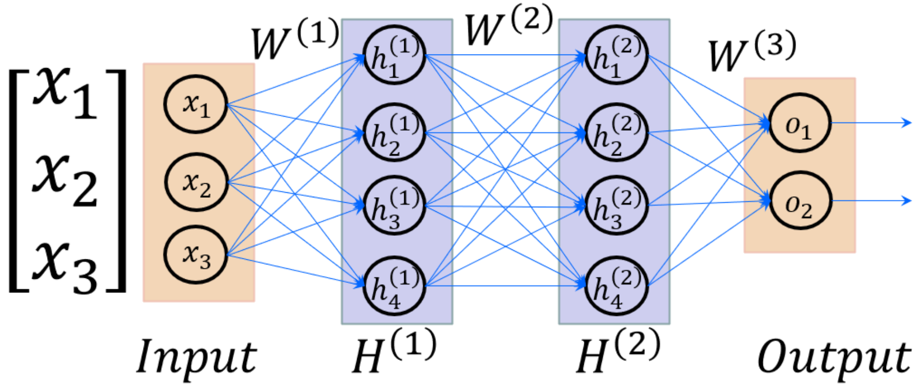
Forward propagation



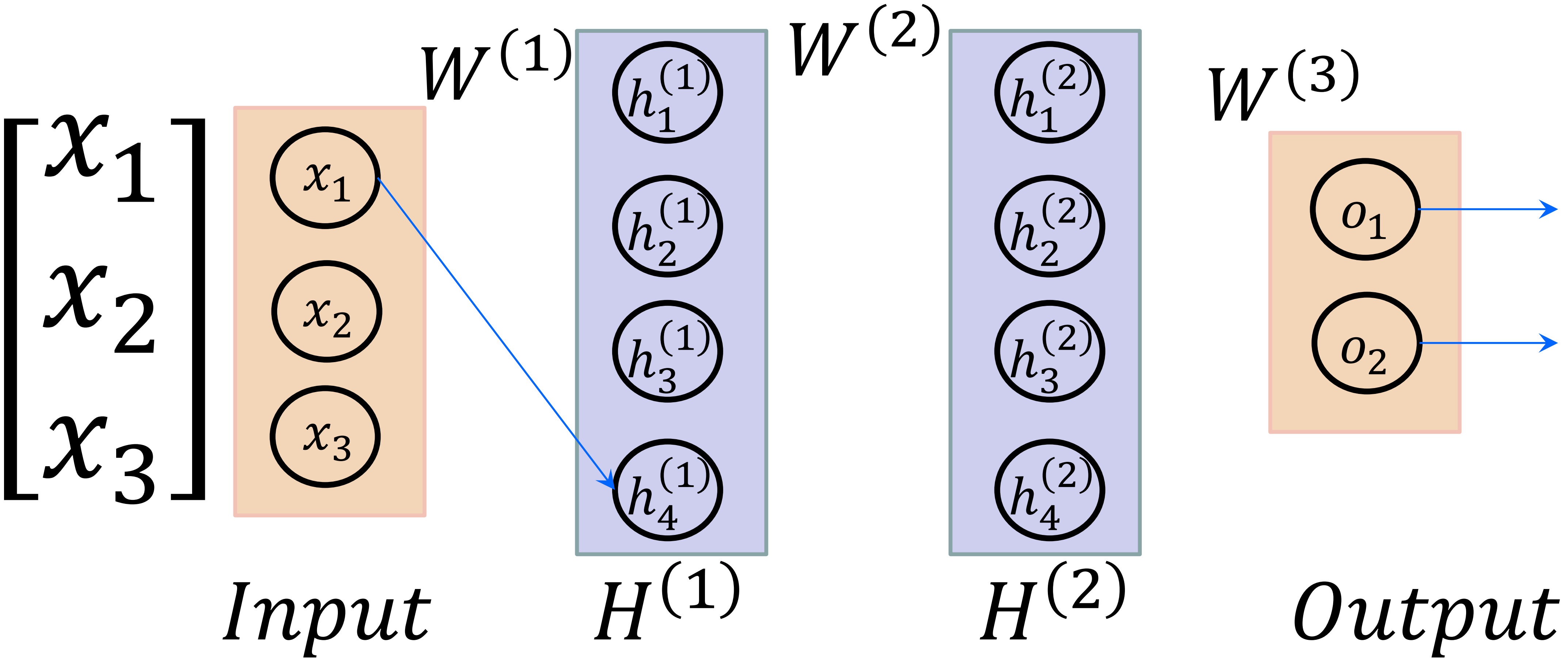
Forward propagation

– Ma trận trọng số $W^{(1)}$

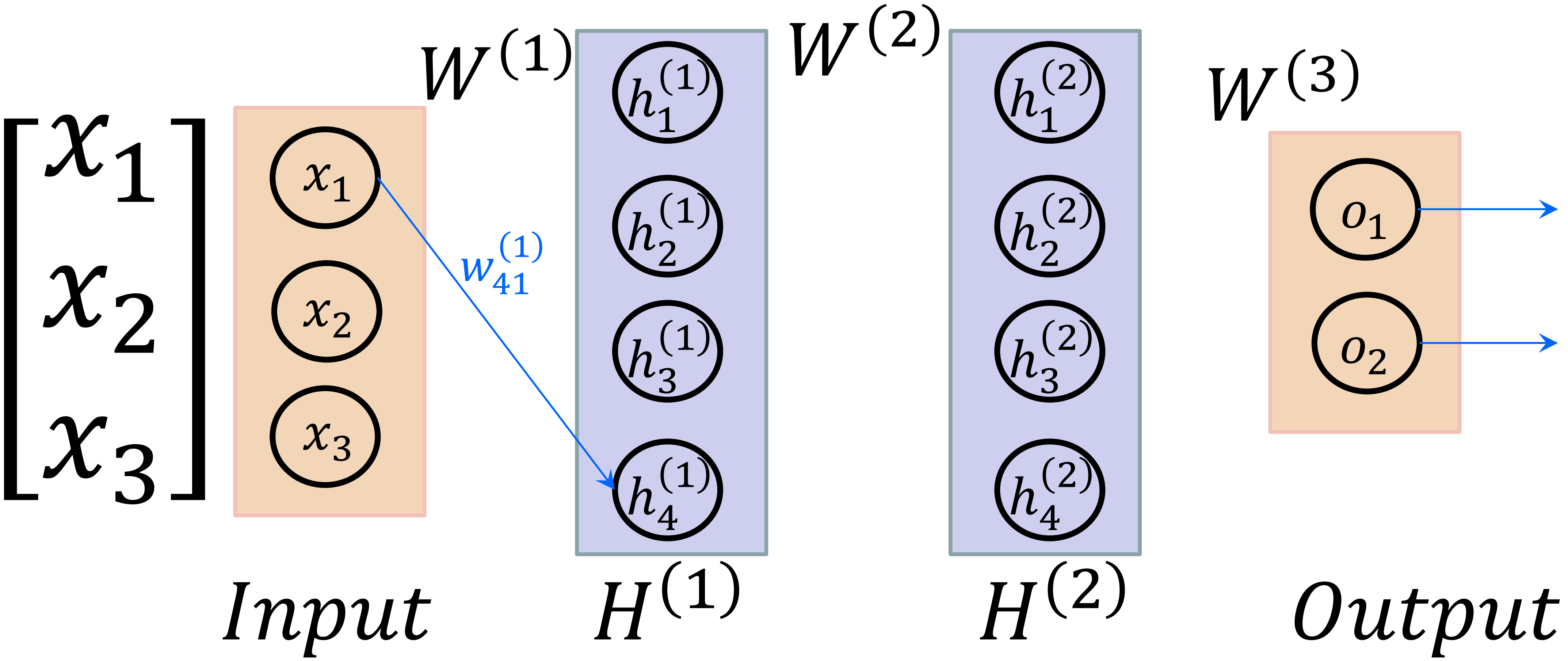
$$W^{(1)} = \begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & w_{23}^{(1)} \\ w_{31}^{(1)} & w_{32}^{(1)} & w_{33}^{(1)} \\ & & \end{bmatrix}$$



Forward propagation



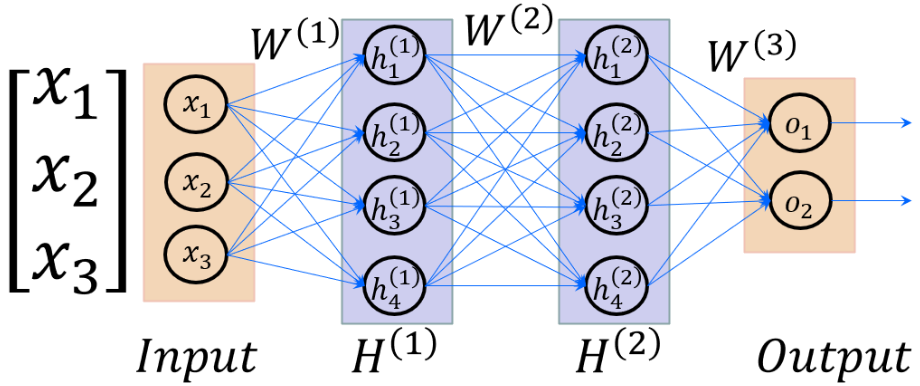
Forward propagation



Forward propagation

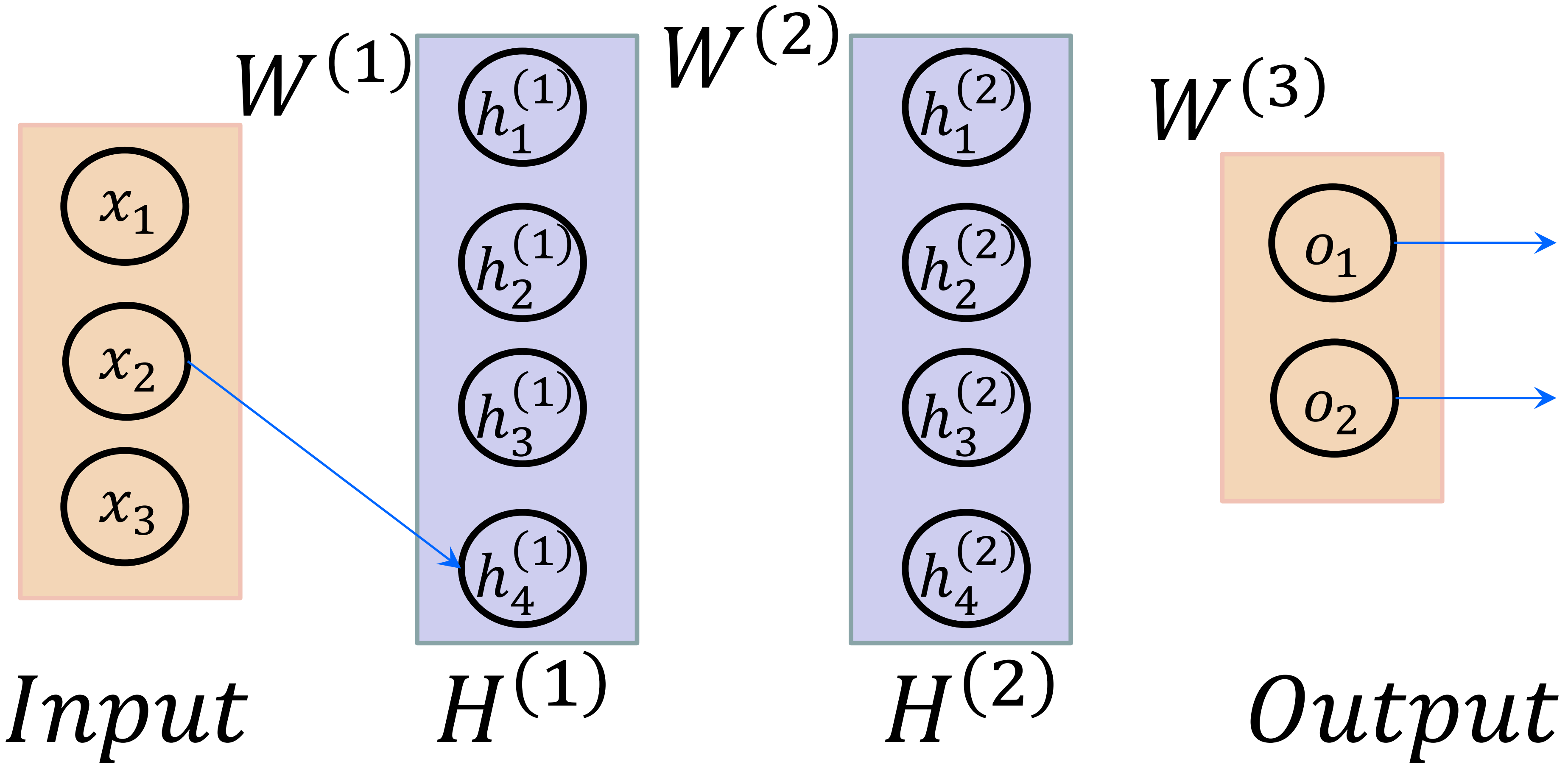
– Ma trận trọng số $W^{(1)}$

$$W^{(1)} = \begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & w_{23}^{(1)} \\ w_{31}^{(1)} & w_{32}^{(1)} & w_{33}^{(1)} \\ w_{41}^{(1)} & & \end{bmatrix}$$

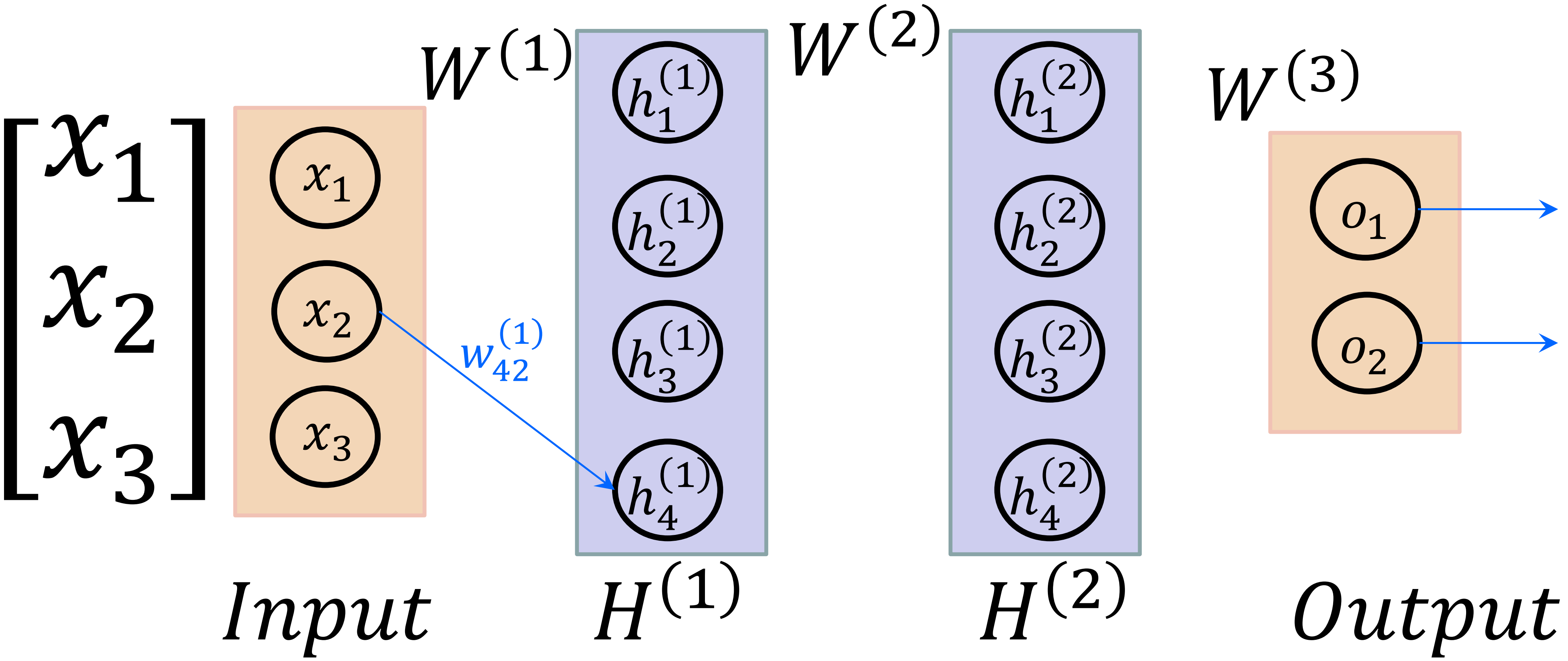


Forward propagation

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$



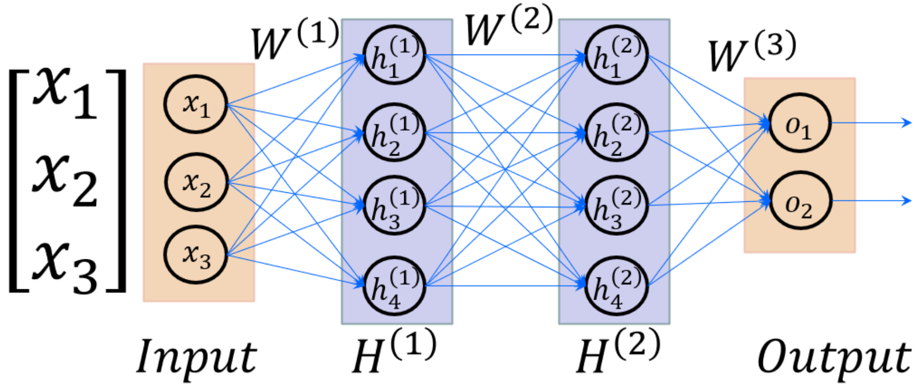
Forward propagation



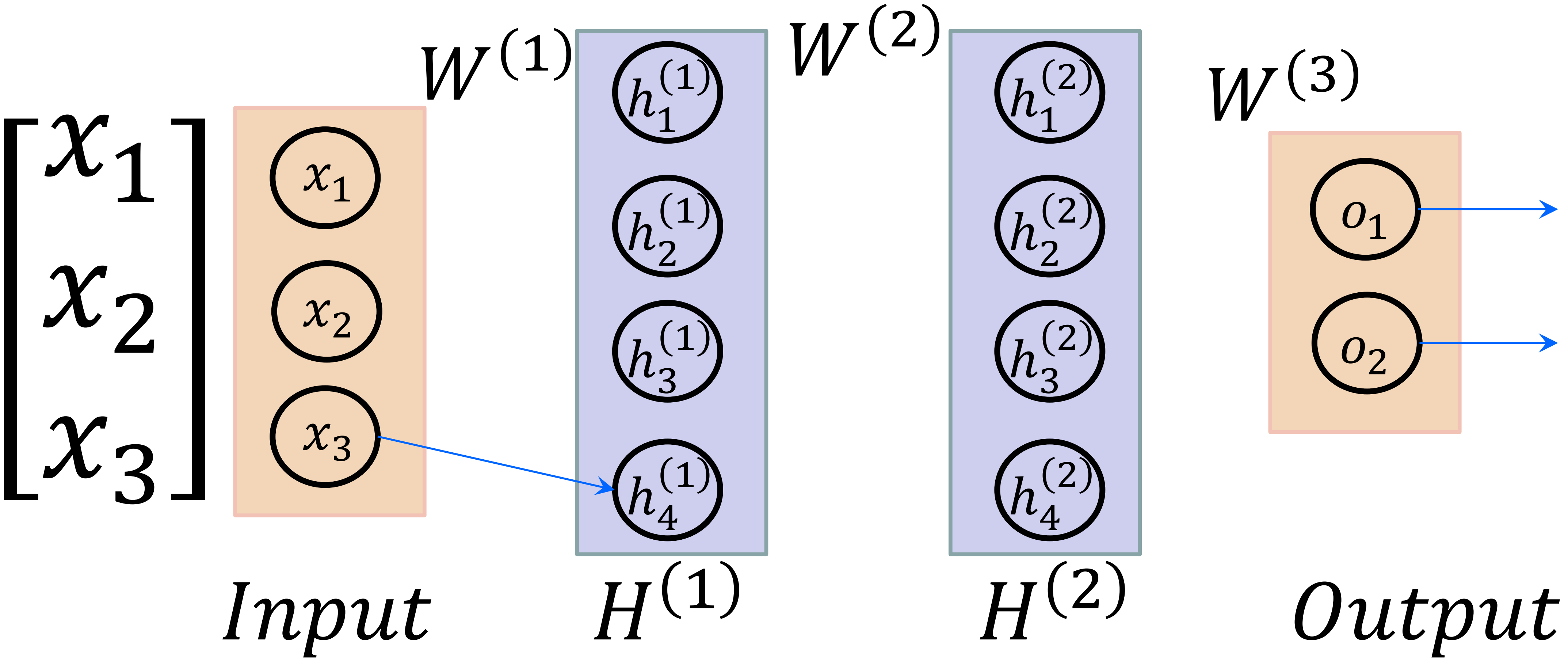
Forward propagation

– Ma trận trọng số $W^{(1)}$

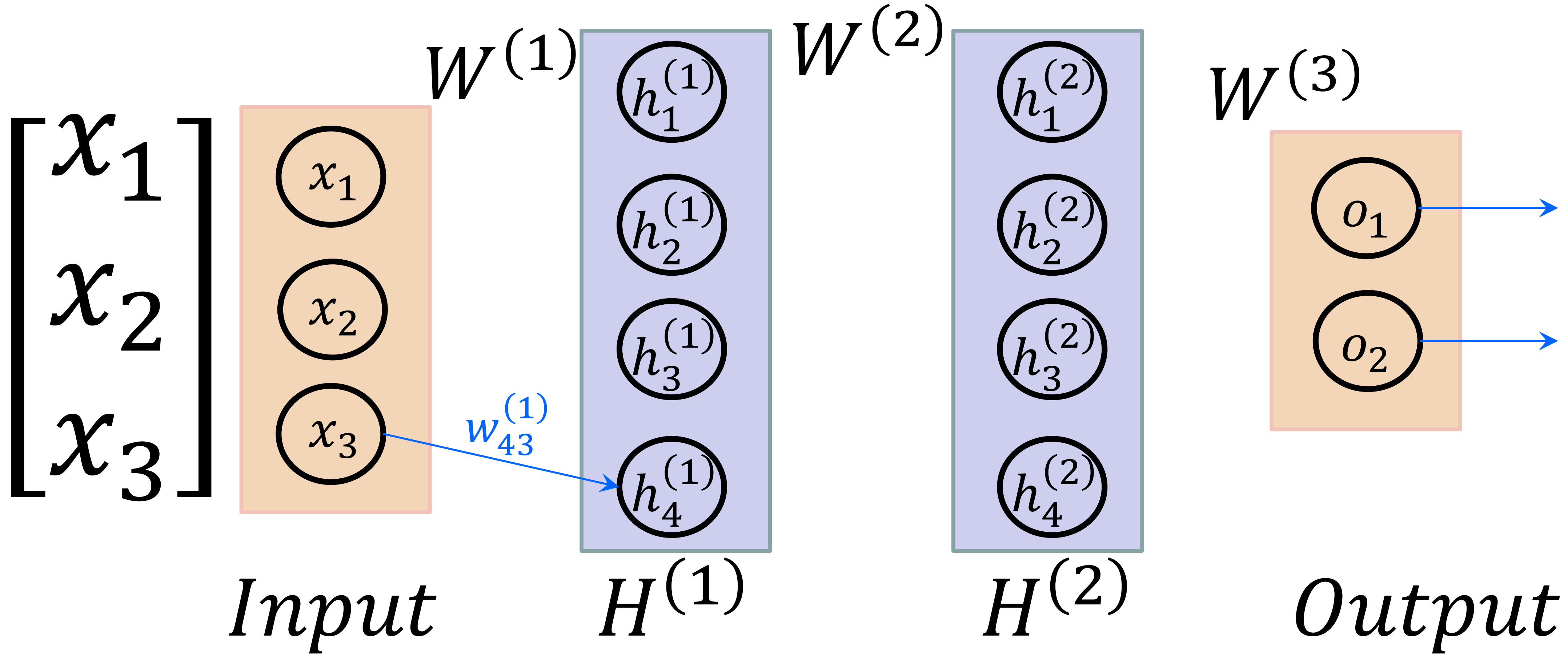
$$W^{(1)} = \begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & w_{23}^{(1)} \\ w_{31}^{(1)} & w_{32}^{(1)} & w_{33}^{(1)} \\ w_{41}^{(1)} & w_{42}^{(1)} & \end{bmatrix}$$



Forward propagation



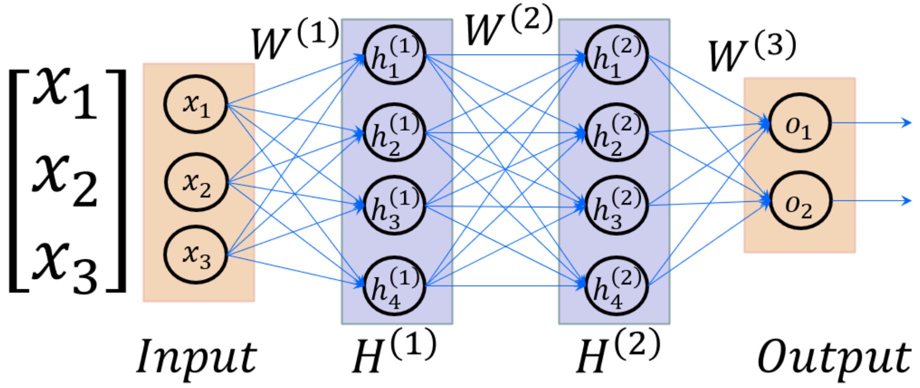
Forward propagation



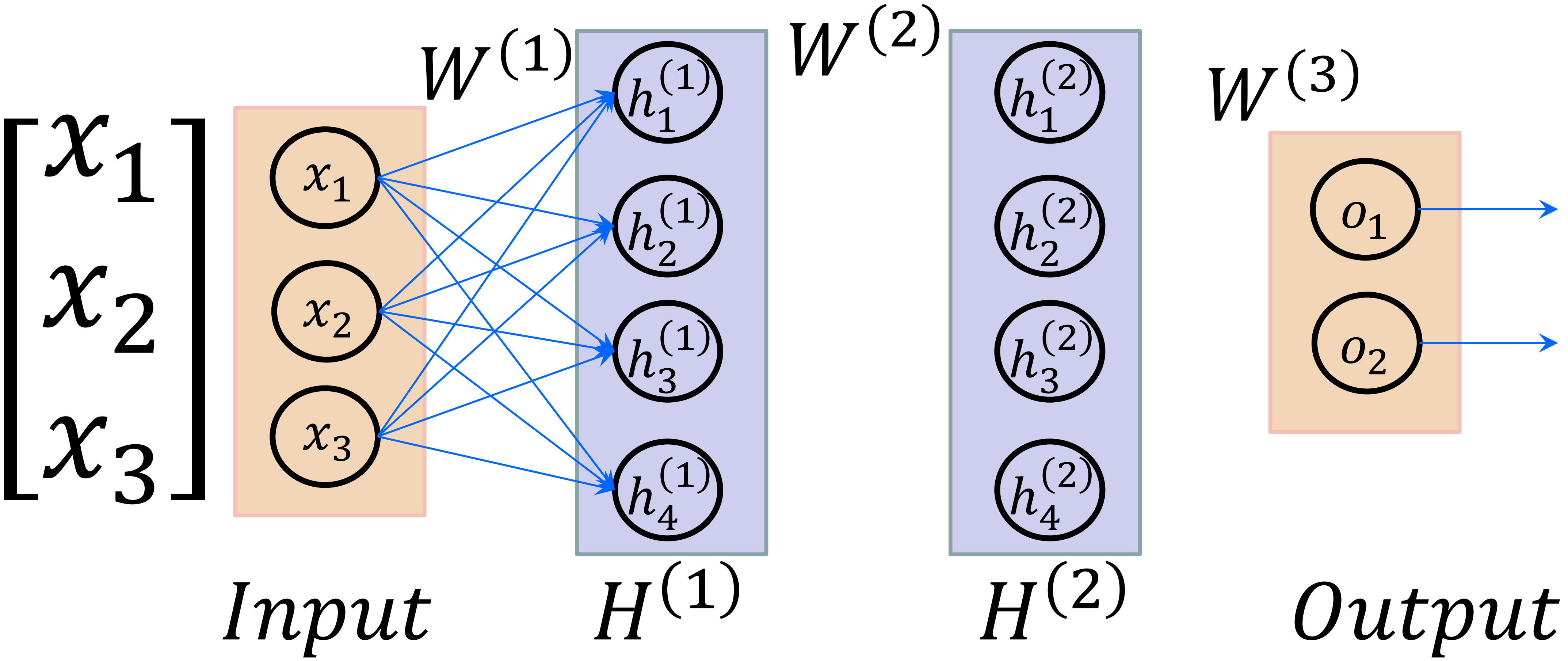
Forward propagation

– Ma trận trọng số $W^{(1)}$

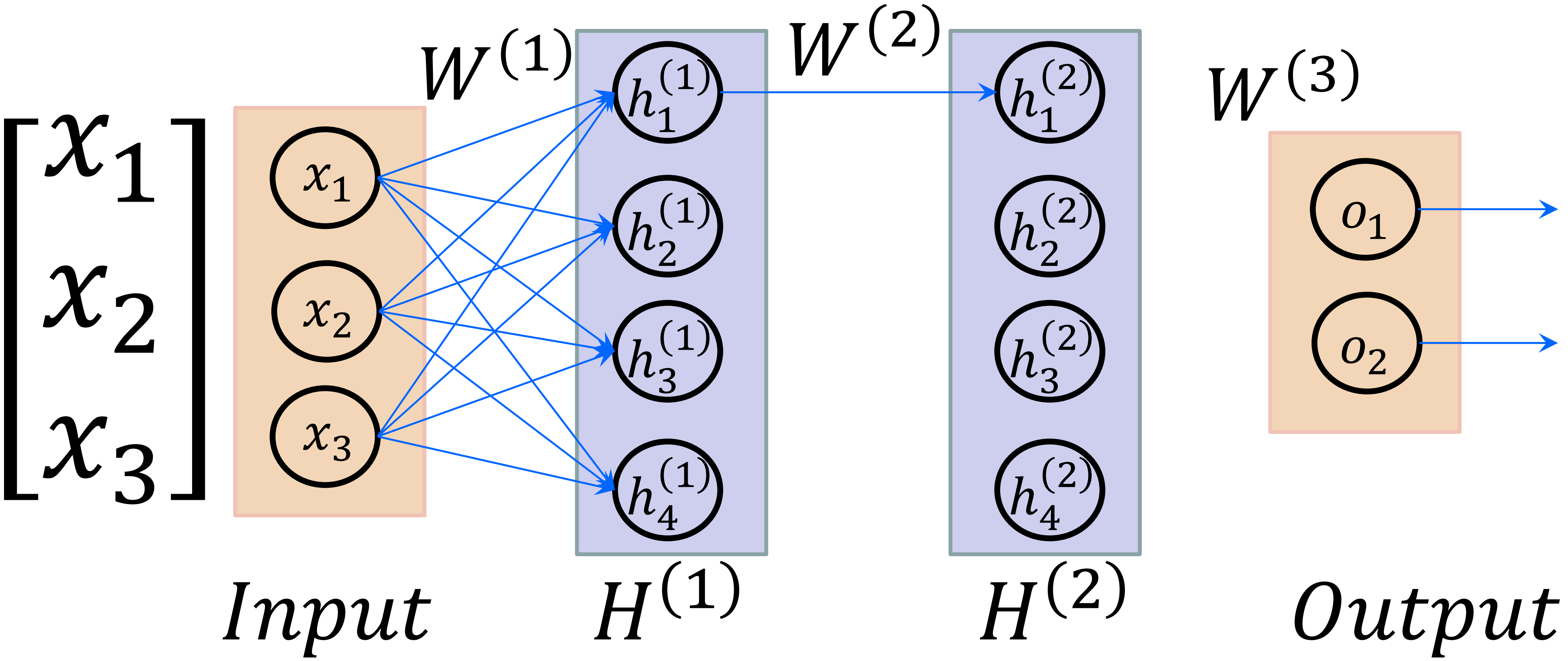
$$W^{(1)} = \begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & w_{23}^{(1)} \\ w_{31}^{(1)} & w_{32}^{(1)} & w_{33}^{(1)} \\ w_{41}^{(1)} & w_{42}^{(1)} & w_{43}^{(1)} \end{bmatrix}$$



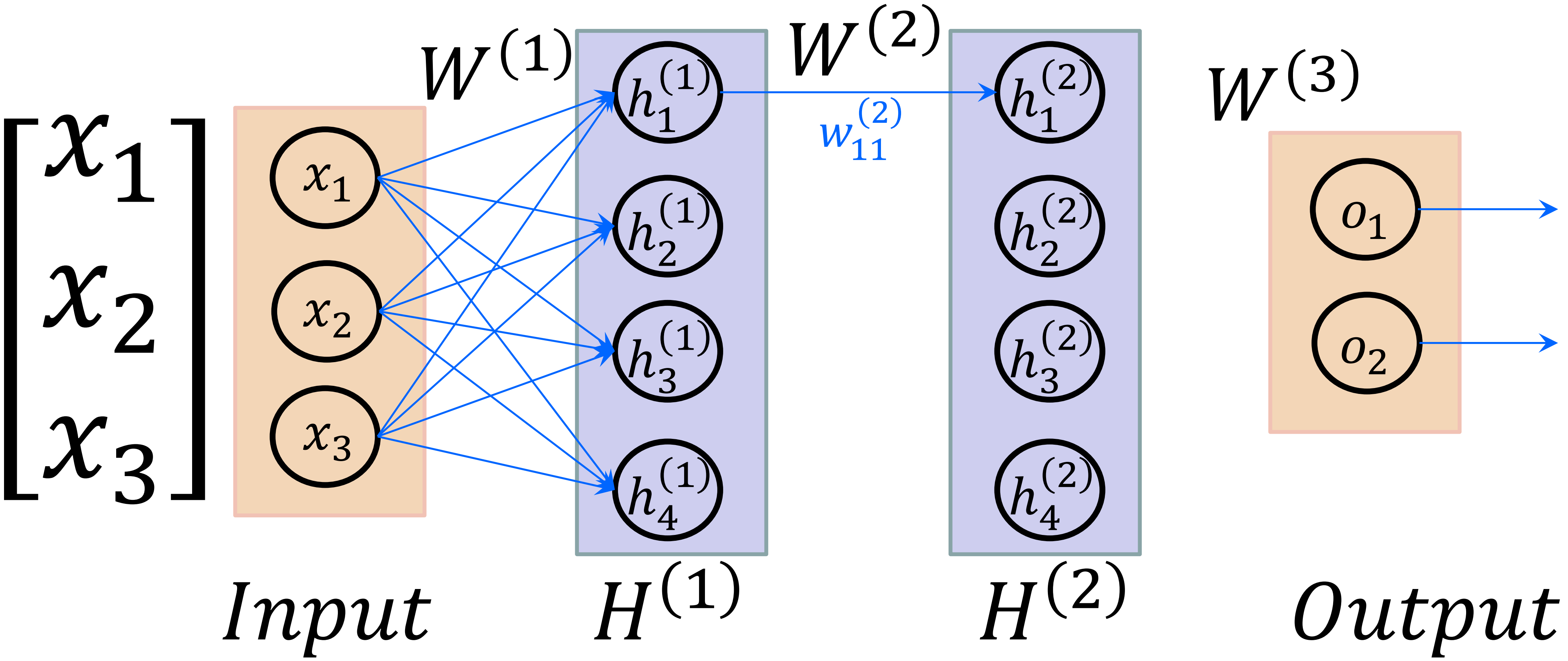
Forward propagation



Forward propagation



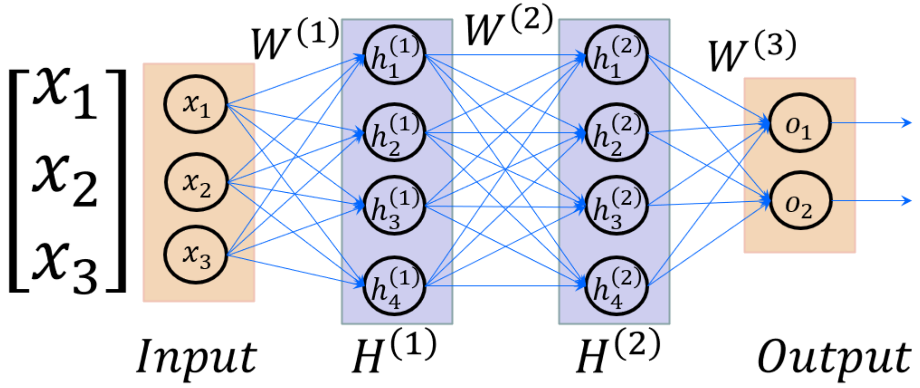
Forward propagation



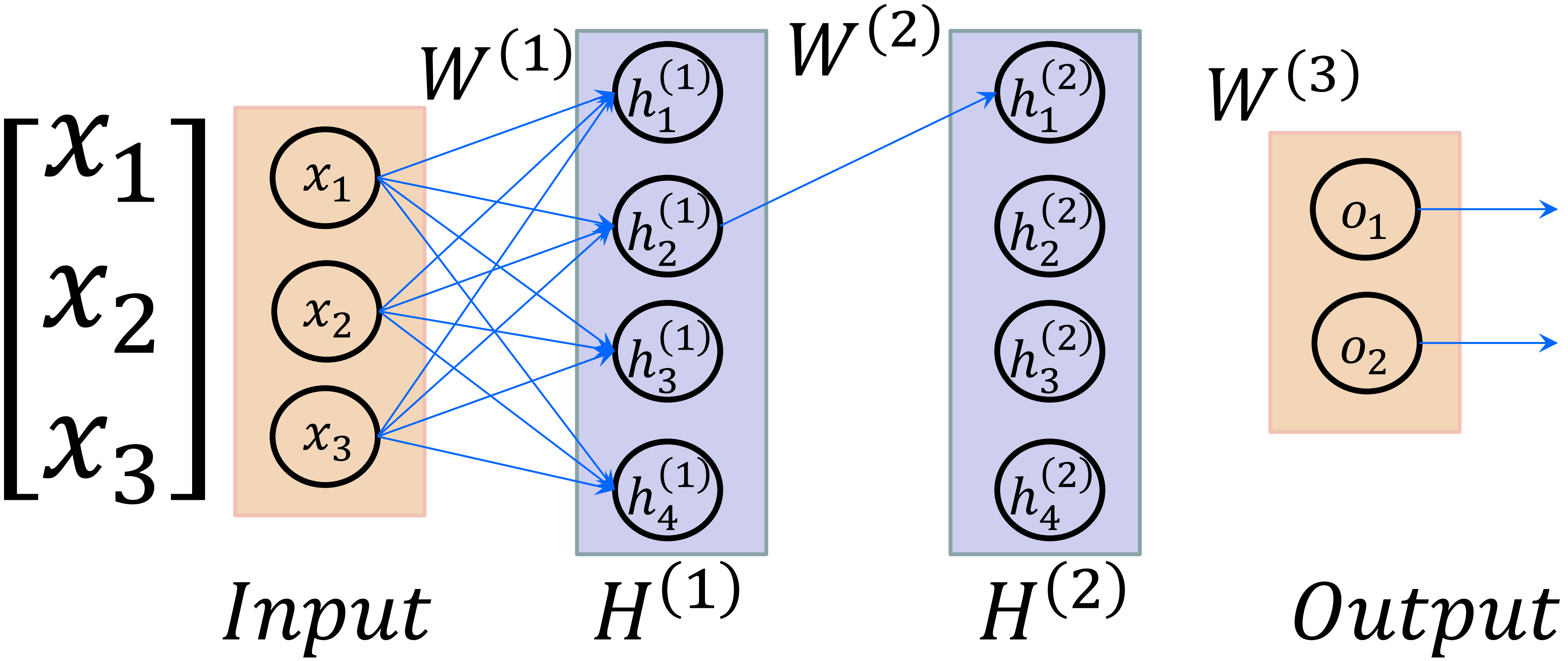
Forward propagation

– Ma trận trọng số $W^{(2)}$

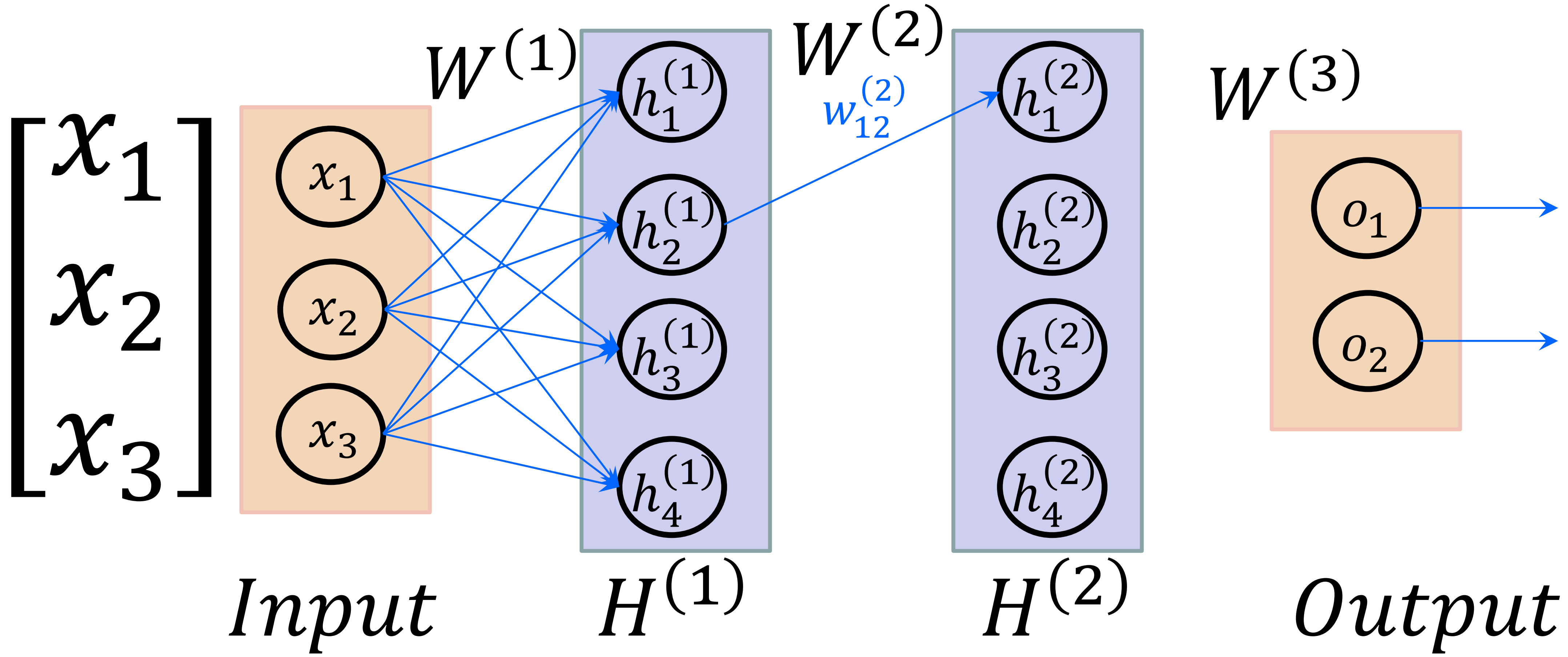
$$W^{(2)} = \begin{bmatrix} w_{11}^{(2)} & & & \\ & & & \\ & & & \\ & & & \end{bmatrix}$$



Forward propagation



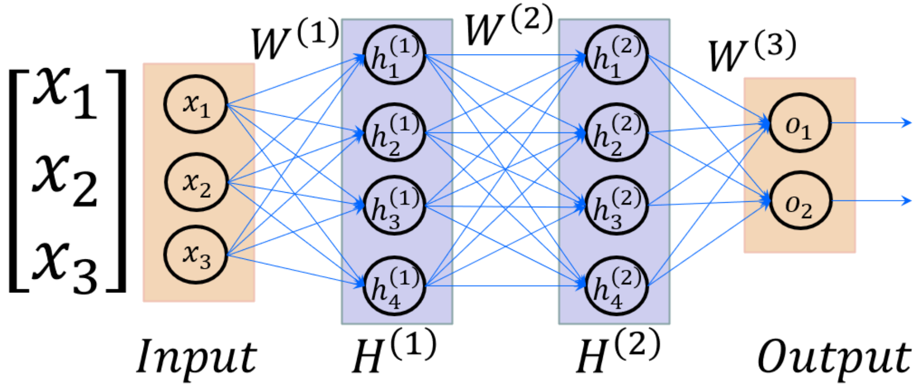
Forward propagation



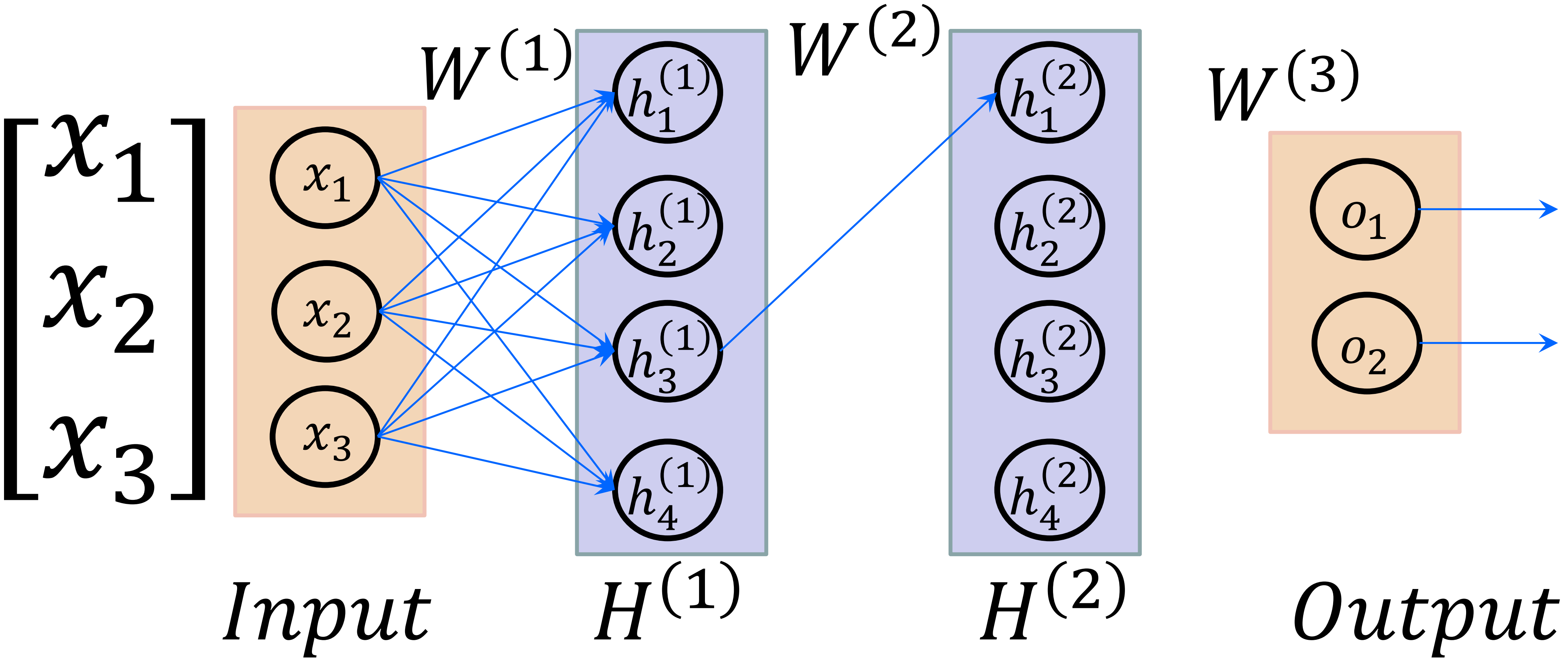
Forward propagation

– Ma trận trọng số $W^{(2)}$

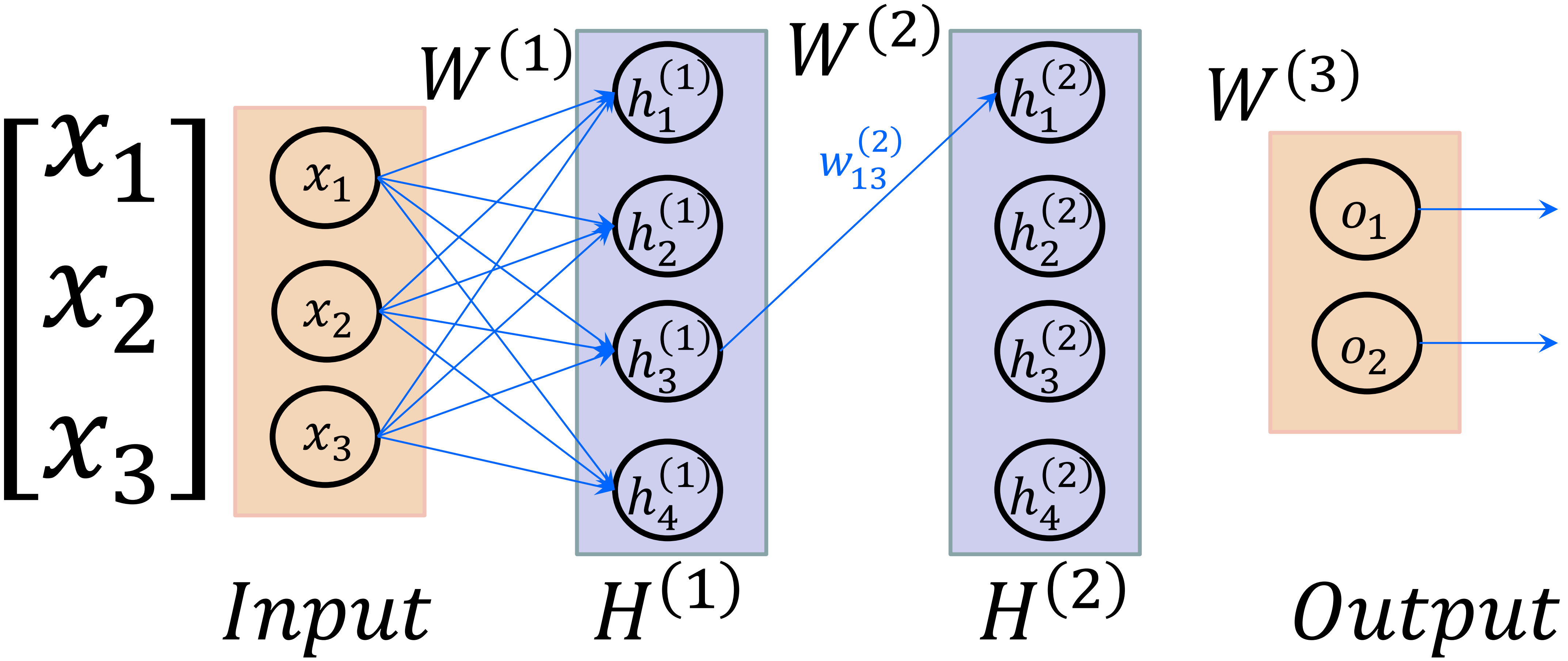
$$W^{(2)} = \begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & & & \\ & & & & \\ & & & & \\ & & & & \end{bmatrix}$$



Forward propagation



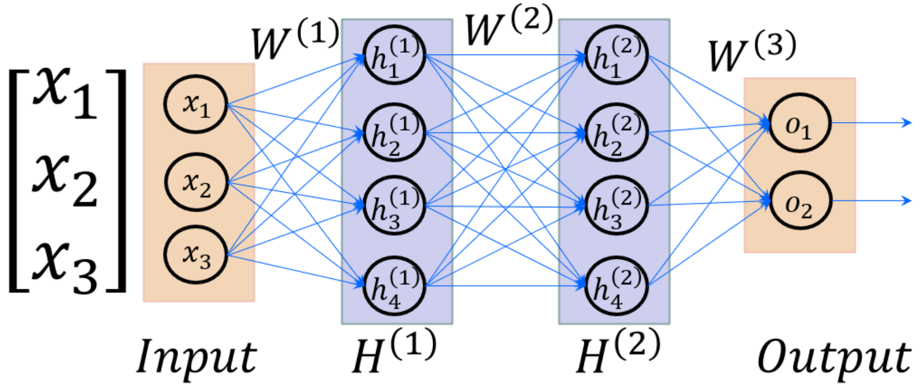
Forward propagation



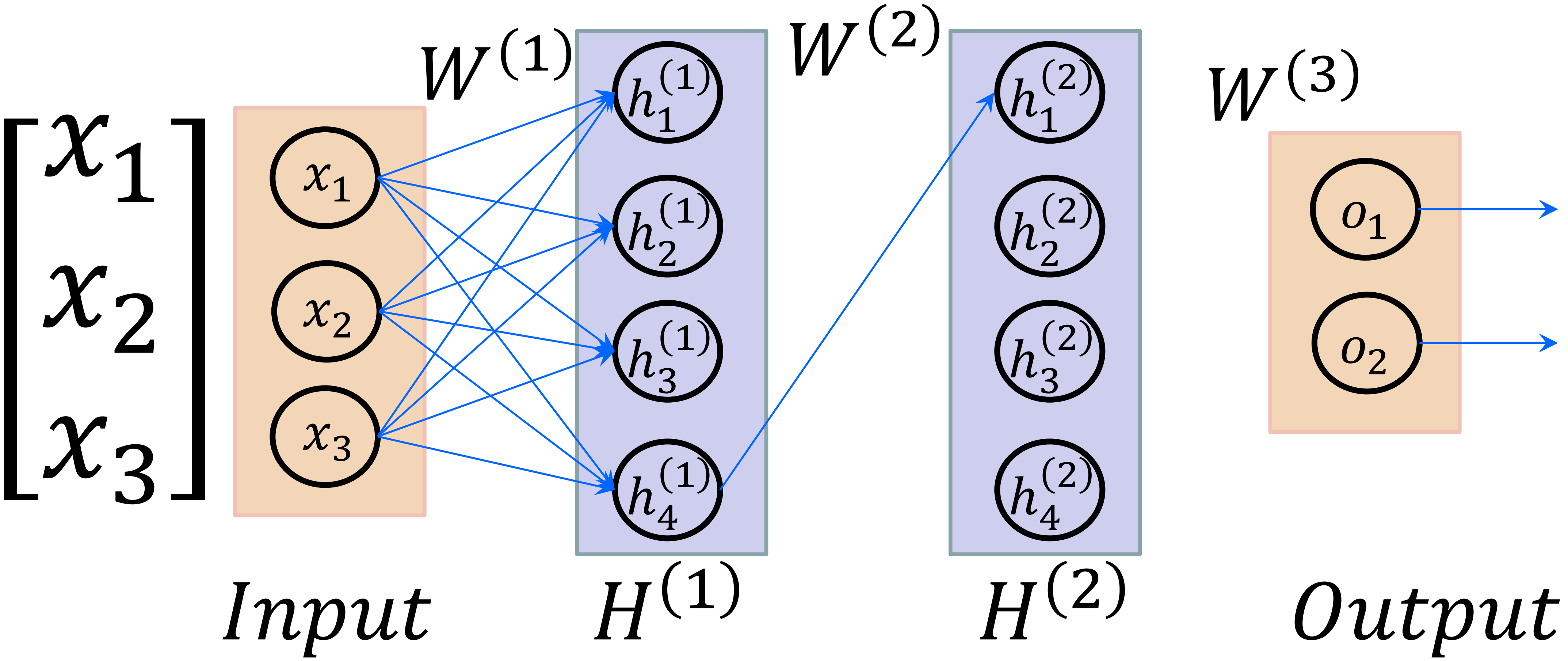
Forward propagation

– Ma trận trọng số $W^{(2)}$

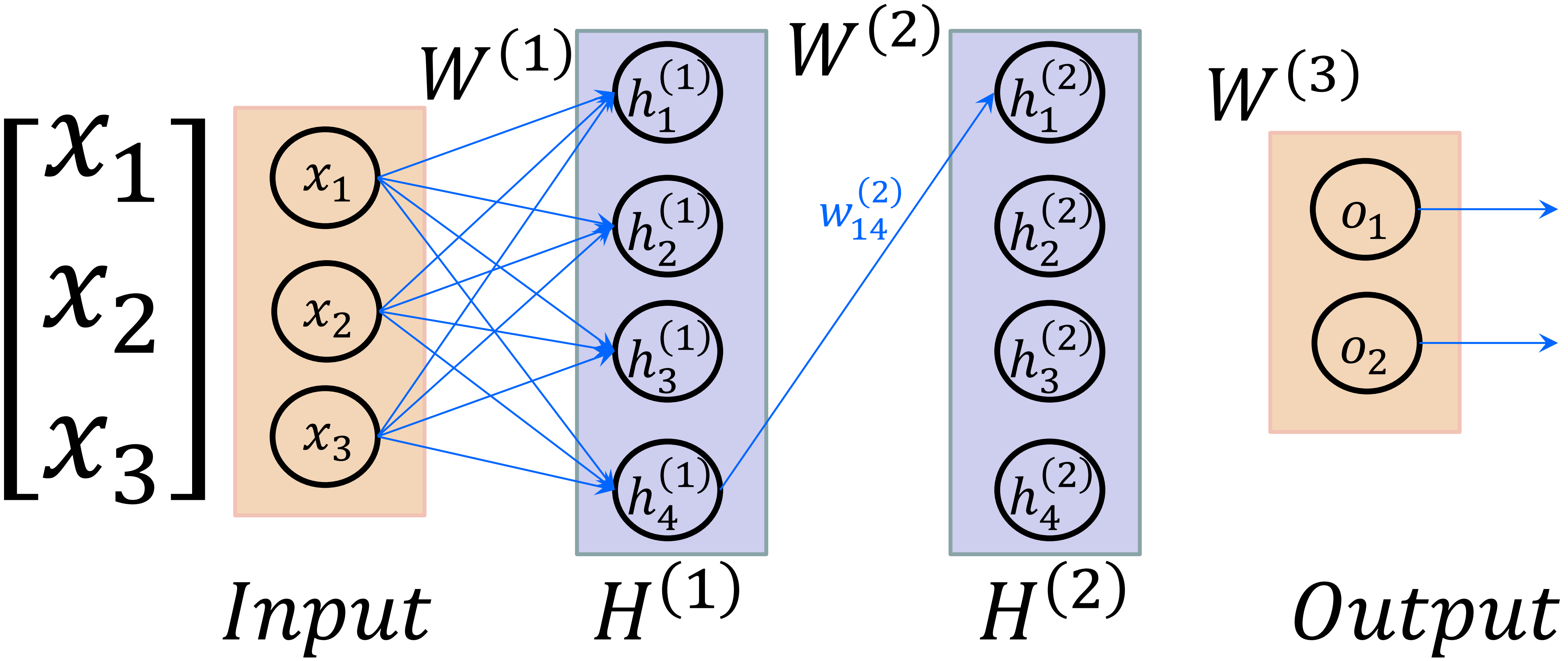
$$W^{(2)} = \begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} & \\ & & & \\ & & & \\ & & & \end{bmatrix}$$



Forward propagation



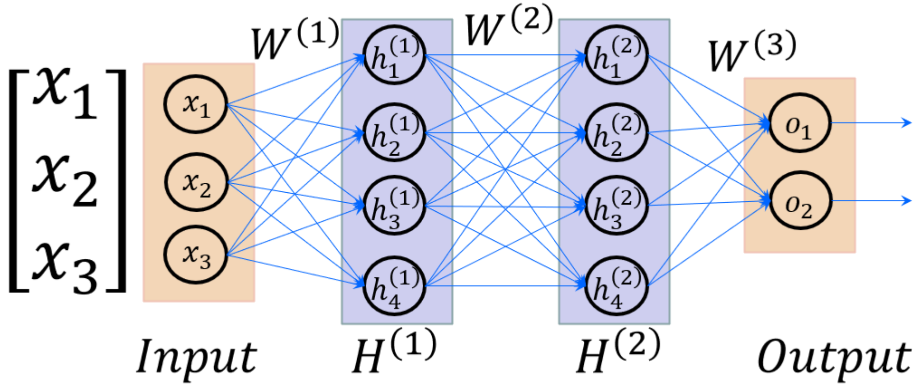
Forward propagation



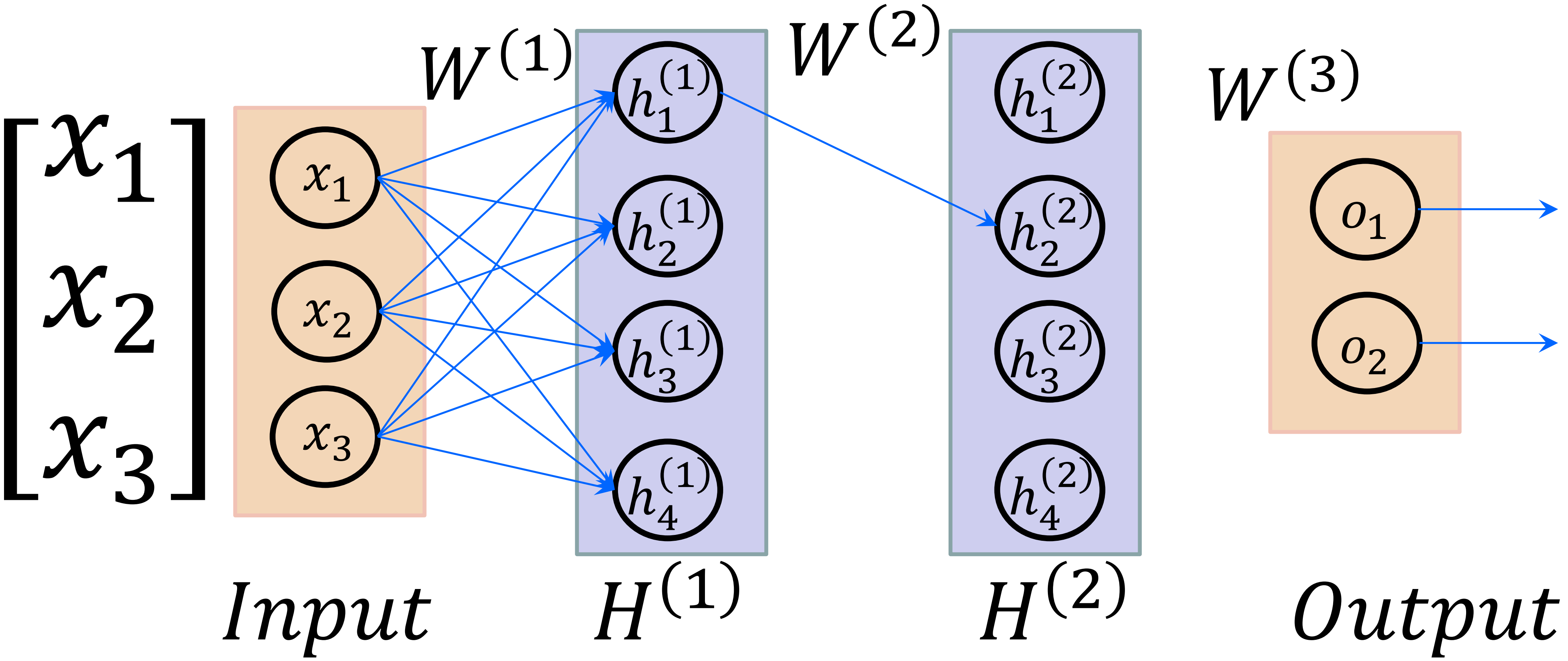
Forward propagation

– Ma trận trọng số $W^{(2)}$

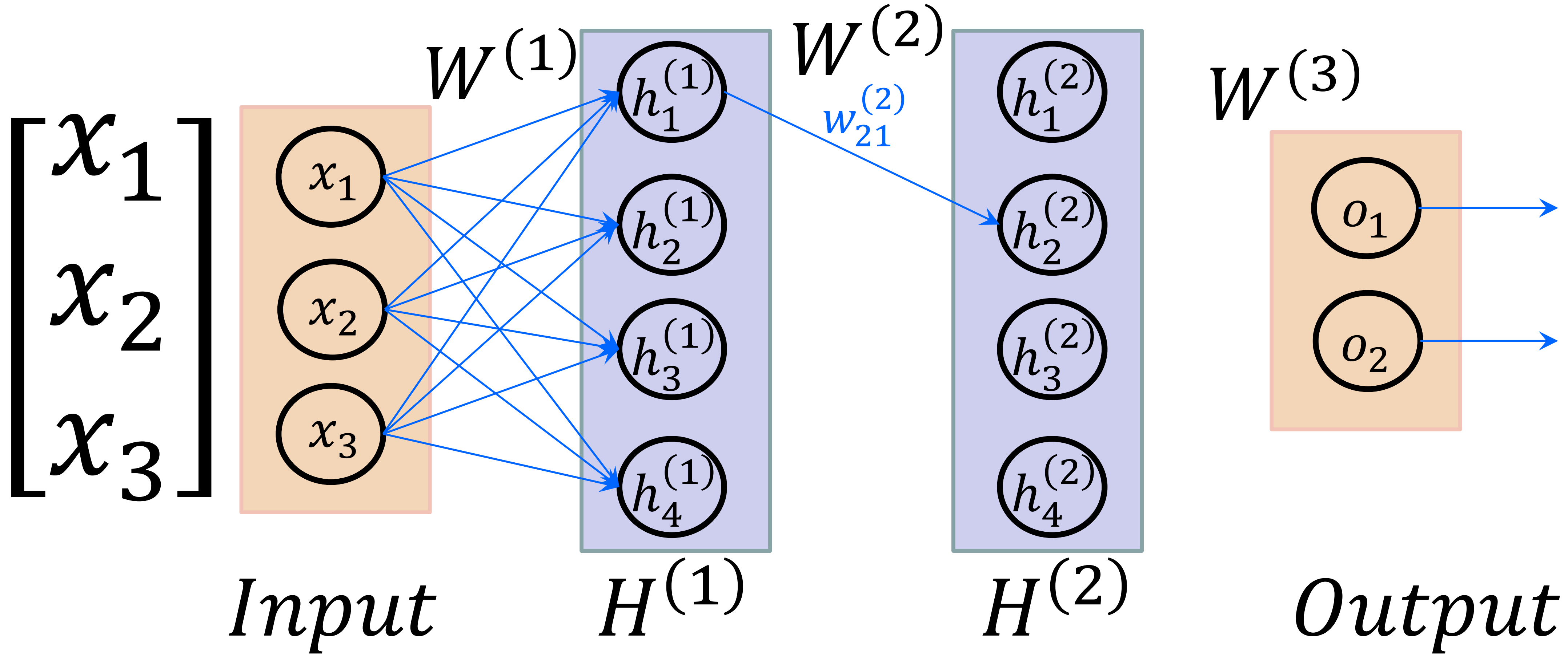
$$W^{(2)} = \begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} & w_{14}^{(2)} \\ & & & \\ & & & \\ & & & \end{bmatrix}$$



Forward propagation



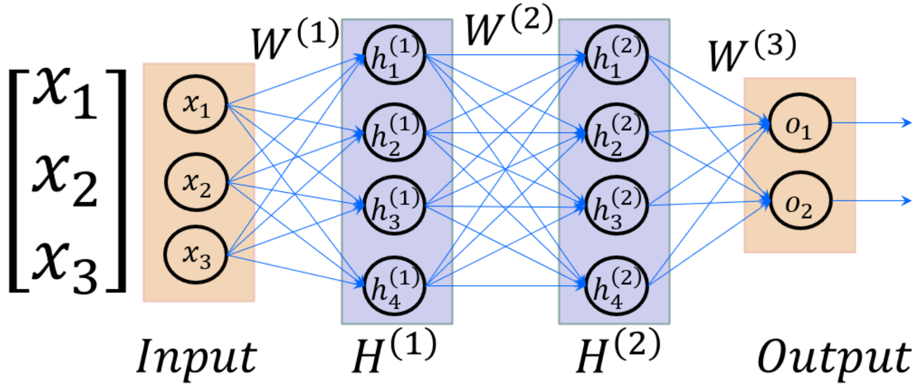
Forward propagation



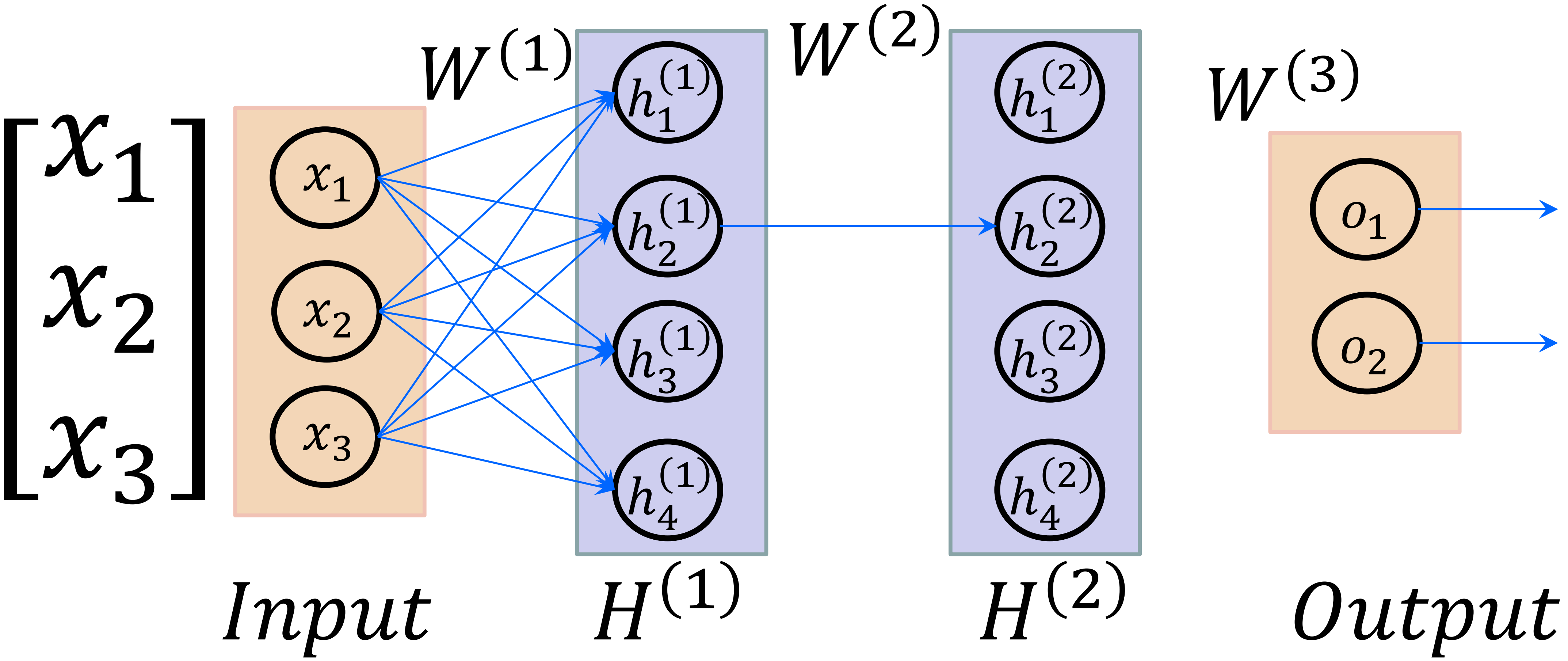
Forward propagation

– Ma trận trọng số $W^{(2)}$

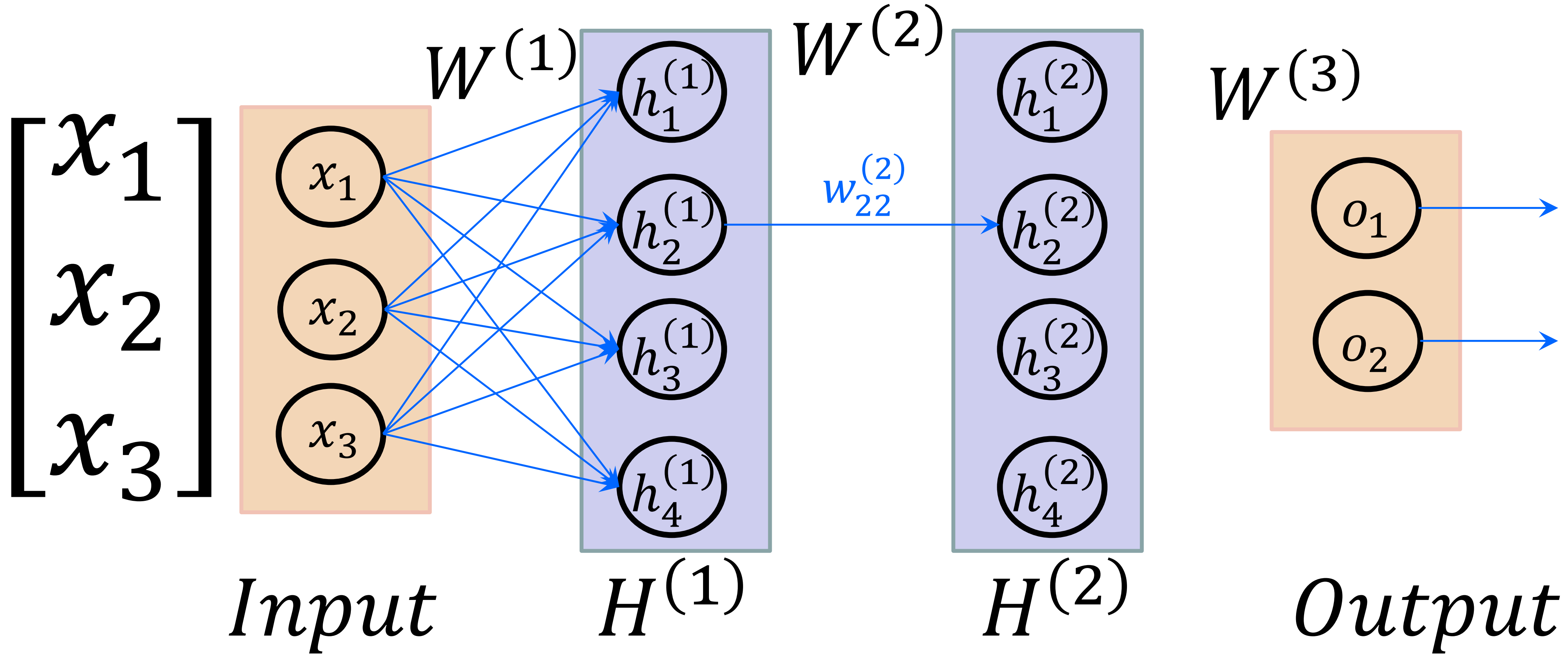
$$W^{(2)} = \begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} & w_{14}^{(2)} \\ w_{21}^{(2)} & & & \\ & & & \\ & & & \end{bmatrix}$$



Forward propagation



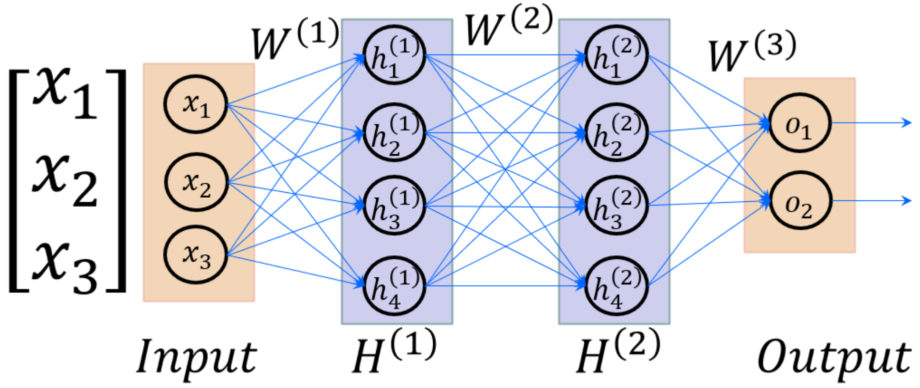
Forward propagation



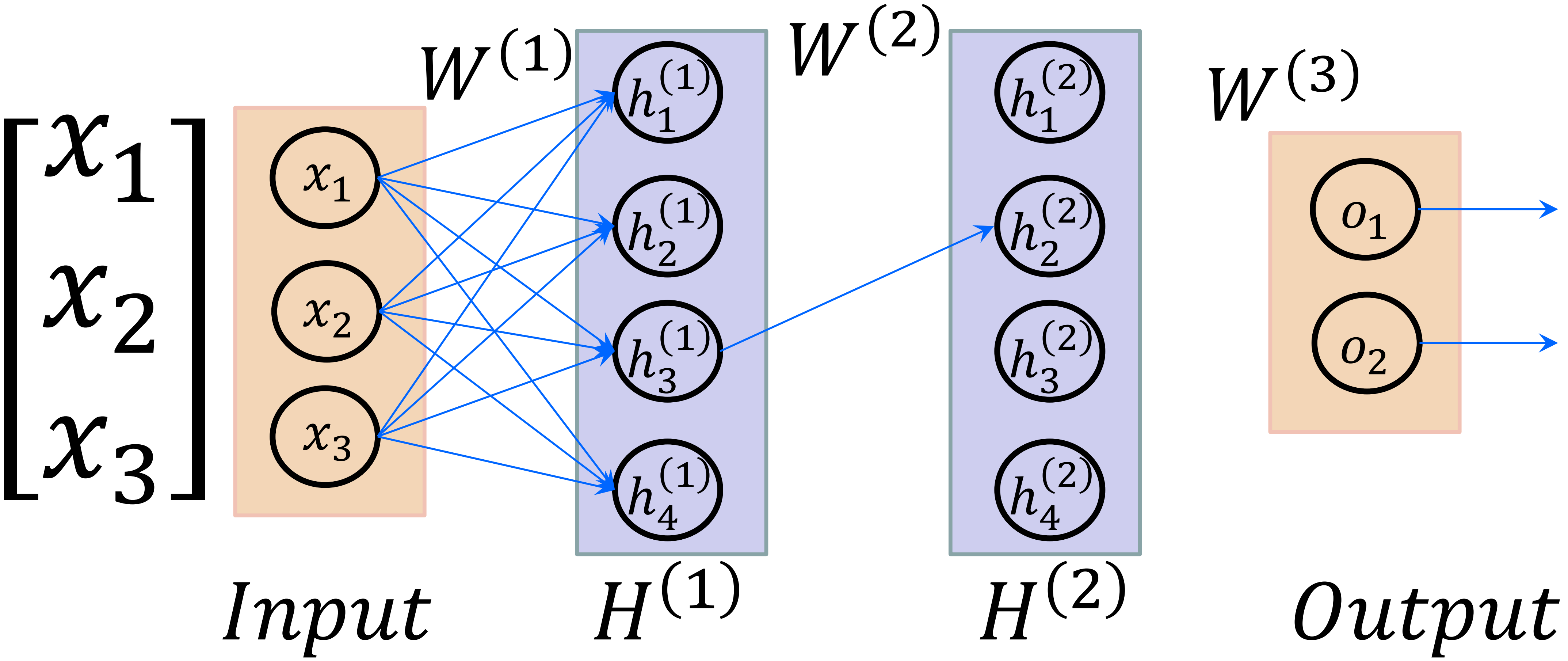
Forward propagation

– Ma trận trọng số $W^{(2)}$

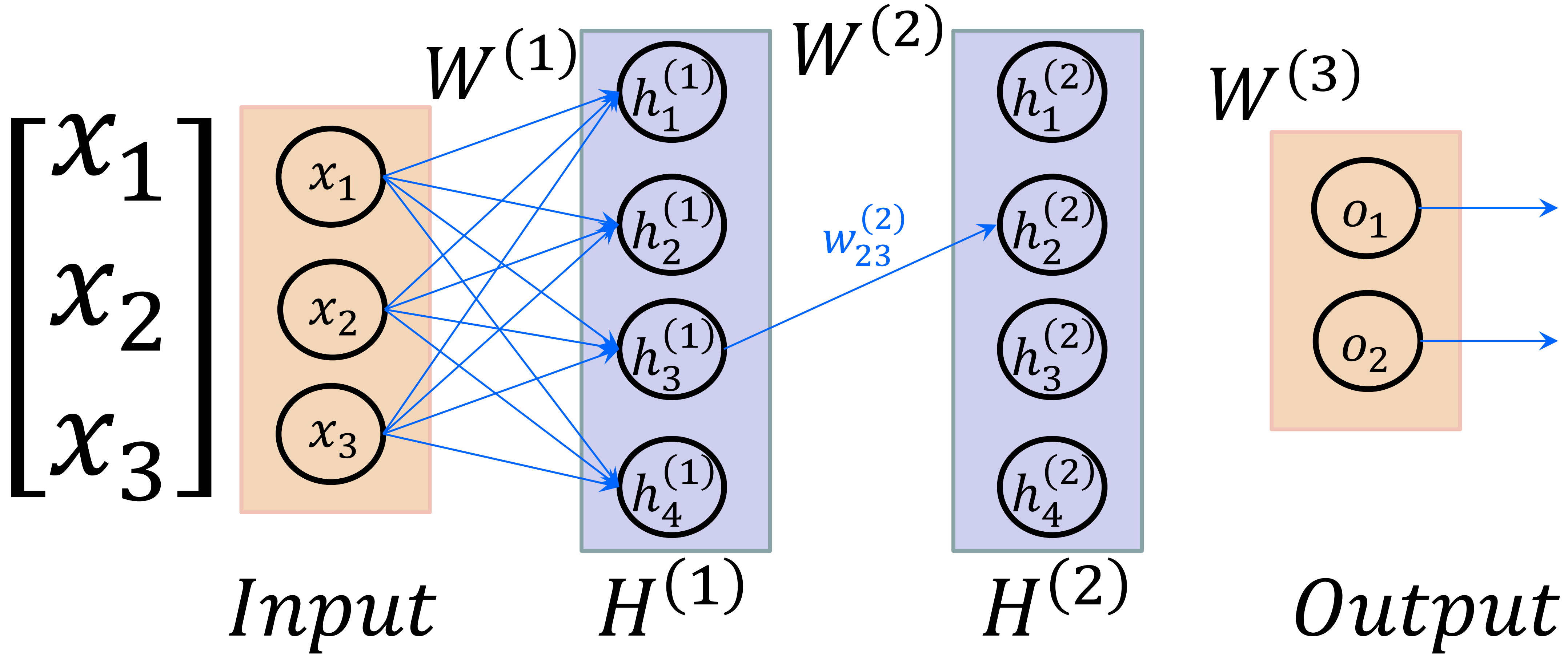
$$W^{(2)} = \begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} & w_{14}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & & \\ & & & \\ & & & \end{bmatrix}$$



Forward propagation



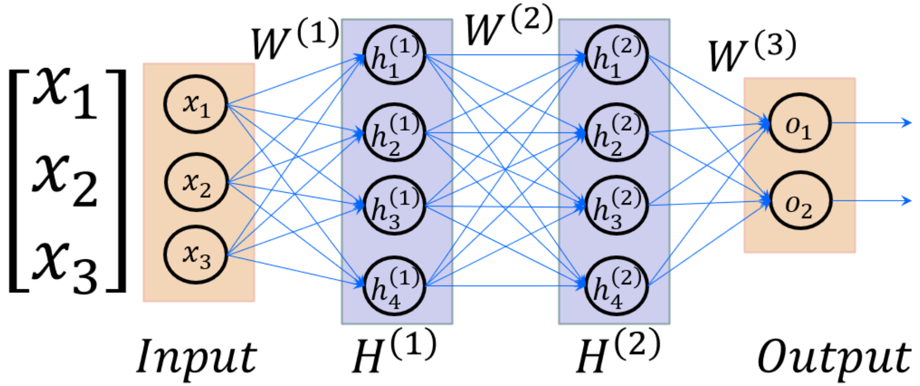
Forward propagation



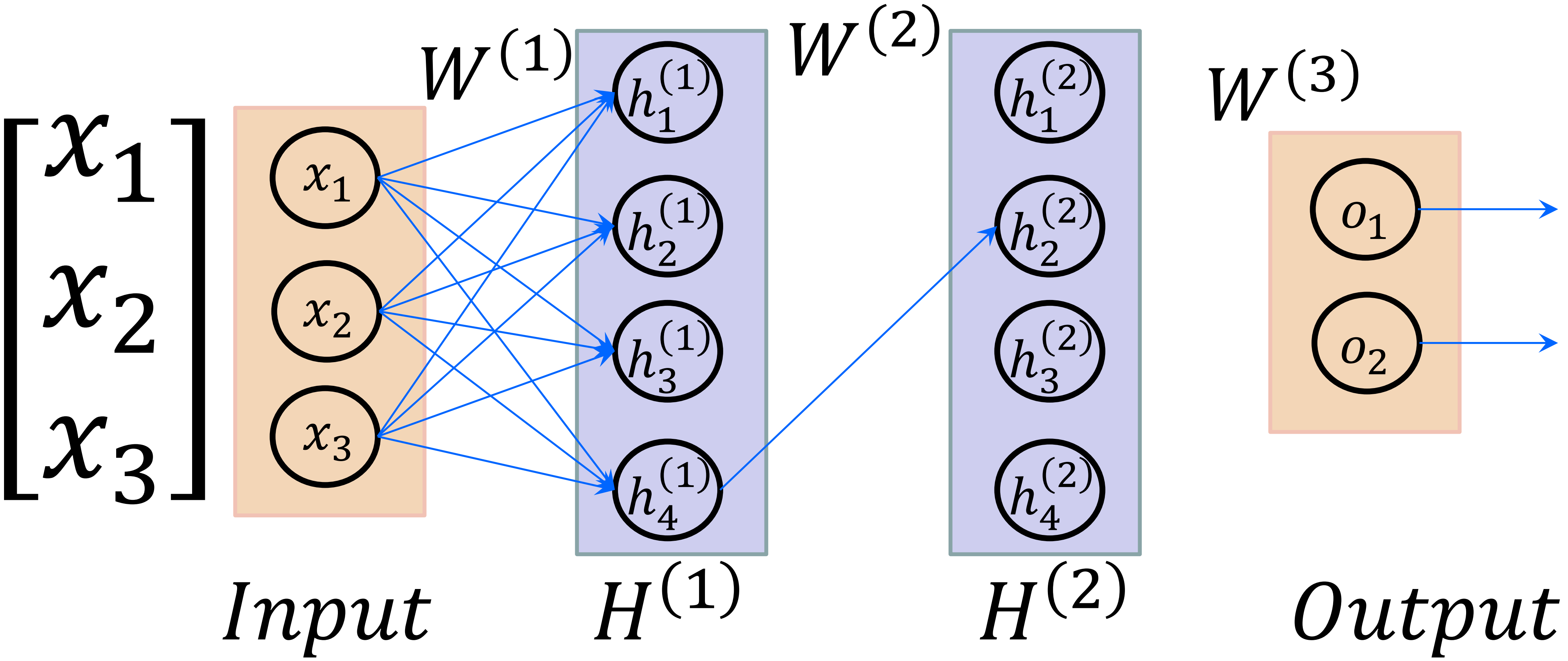
Forward propagation

– Ma trận trọng số $W^{(2)}$

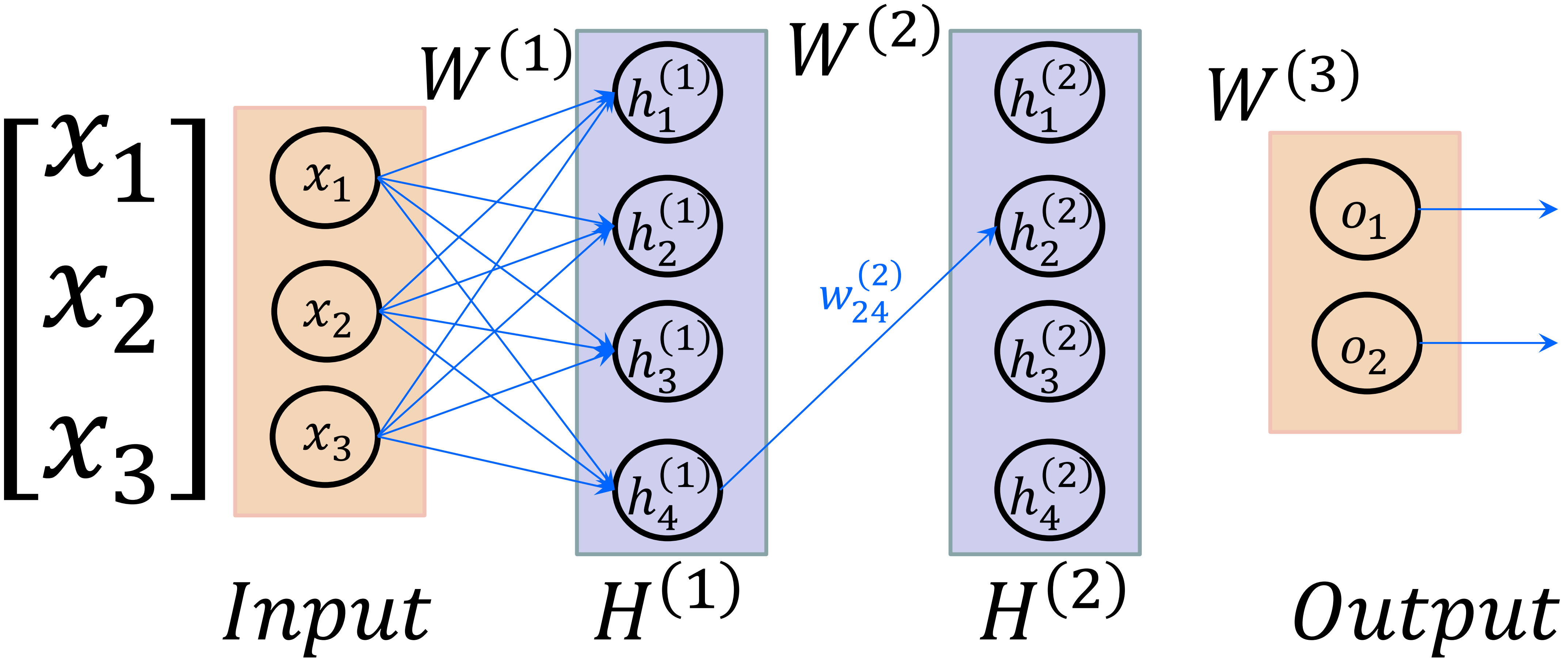
$$W^{(2)} = \begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} & w_{14}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)} & \\ & & & \\ & & & \end{bmatrix}$$



Forward propagation



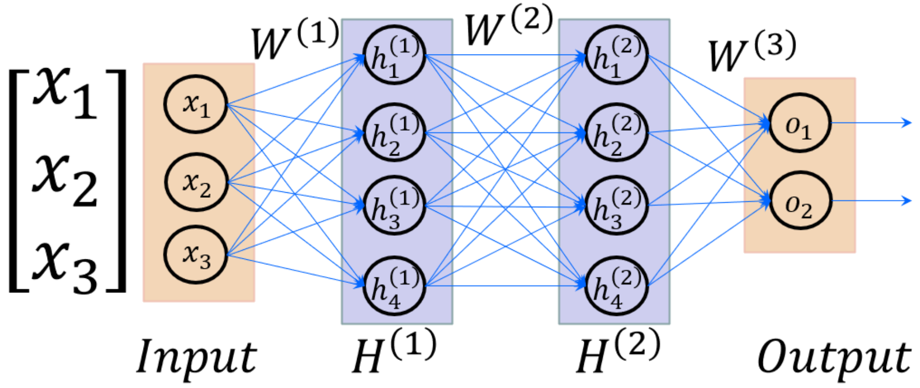
Forward propagation



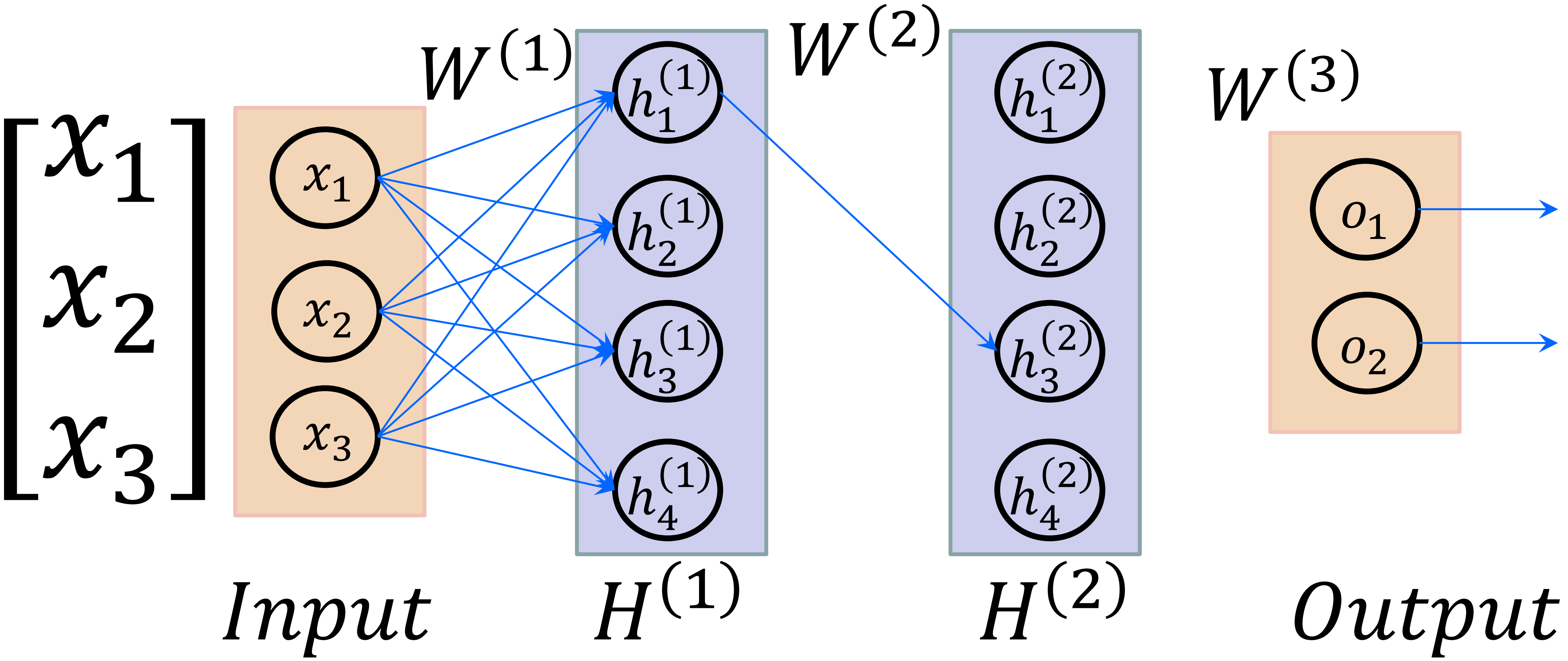
Forward propagation

– Ma trận trọng số $W^{(2)}$

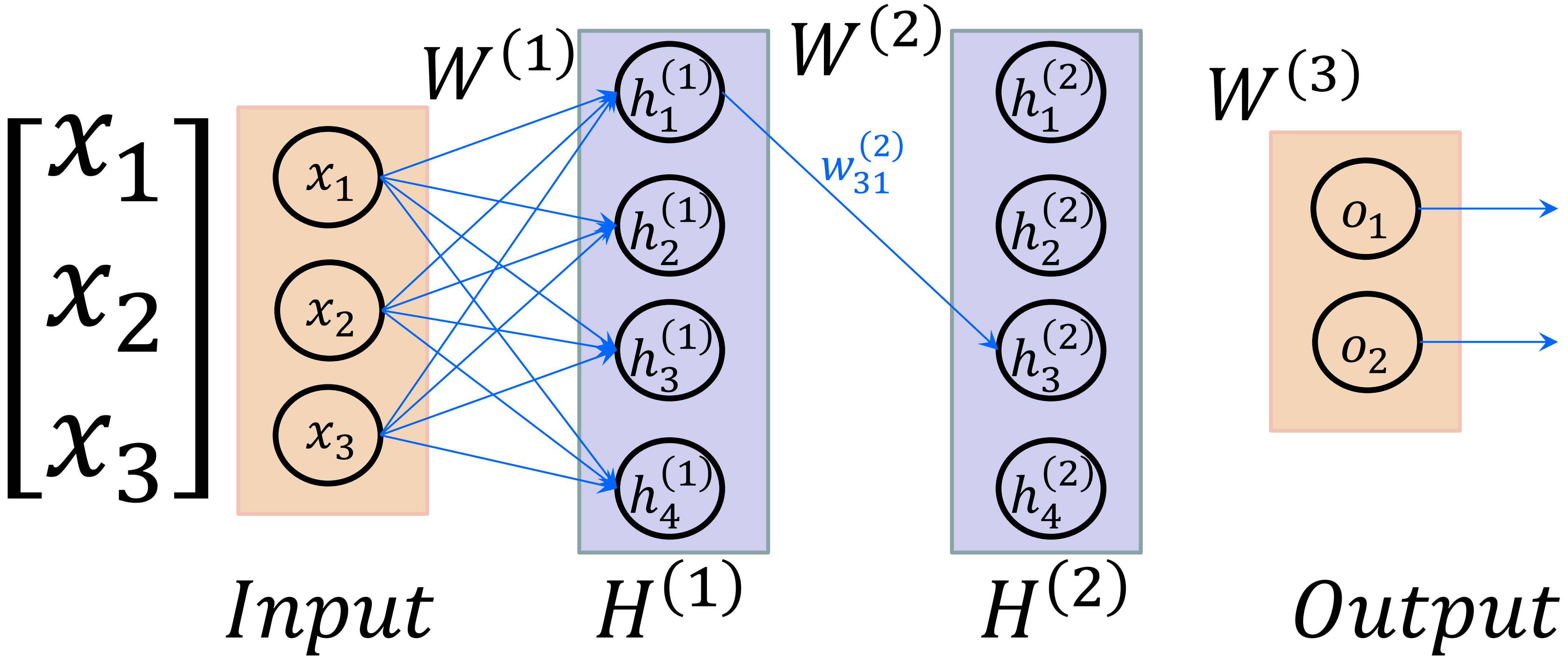
$$W^{(2)} = \begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} & w_{14}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)} & w_{24}^{(2)} \\ & & & \\ & & & \end{bmatrix}$$



Forward propagation



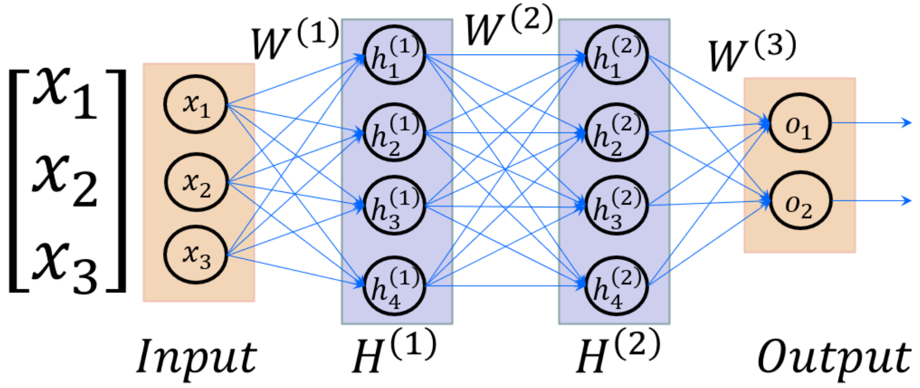
Forward propagation



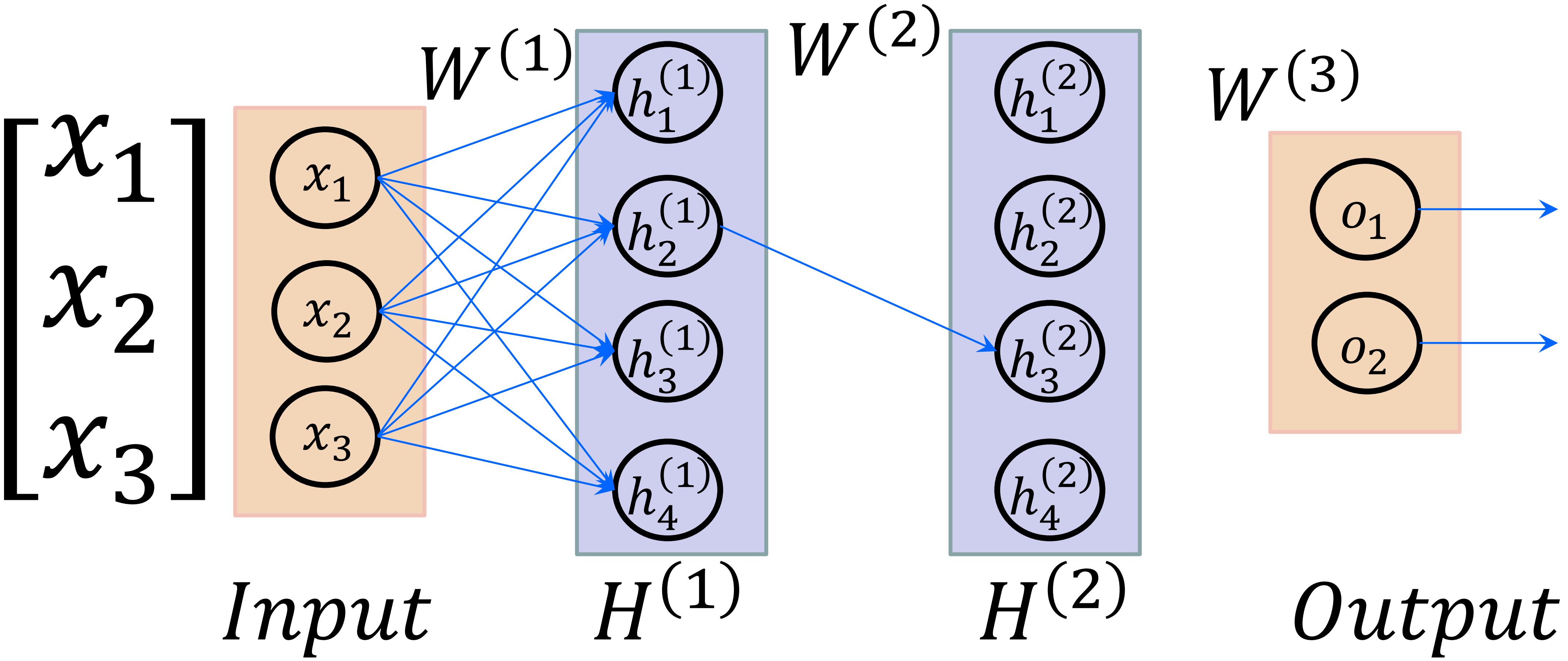
Forward propagation

– Ma trận trọng số $W^{(2)}$

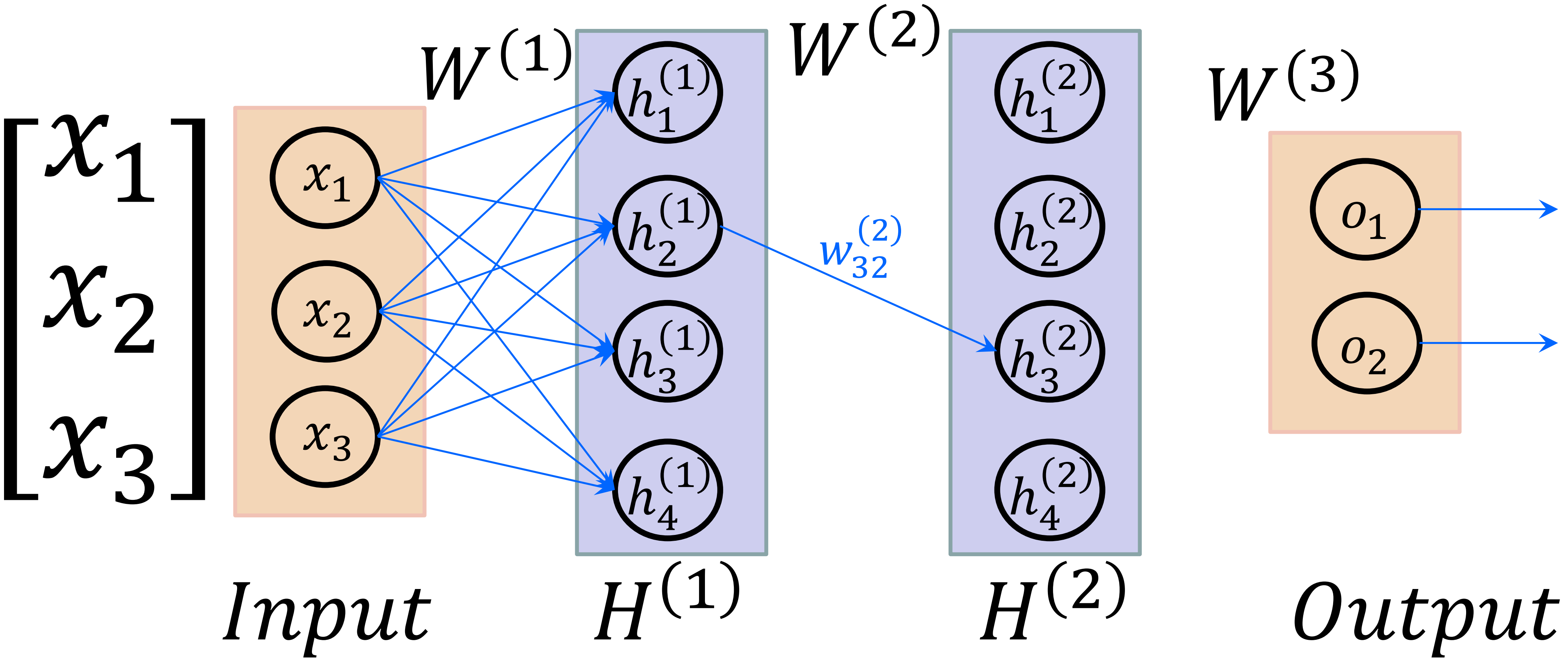
$$W^{(2)} = \begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} & w_{14}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)} & w_{24}^{(2)} \\ w_{31}^{(2)} & & & \\ & & & \end{bmatrix}$$



Forward propagation



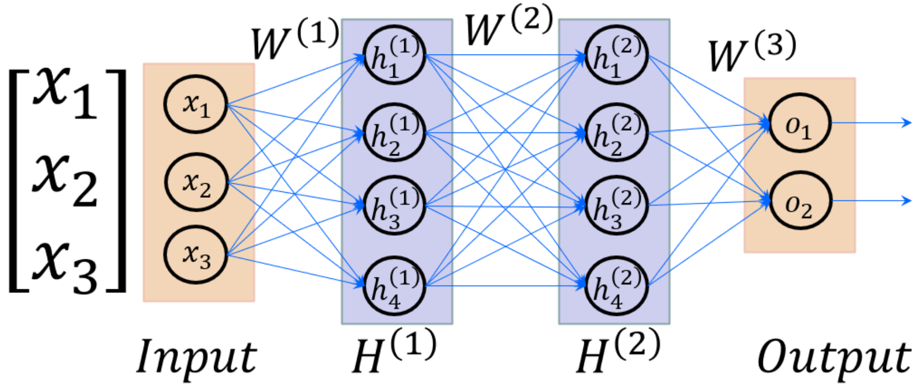
Forward propagation



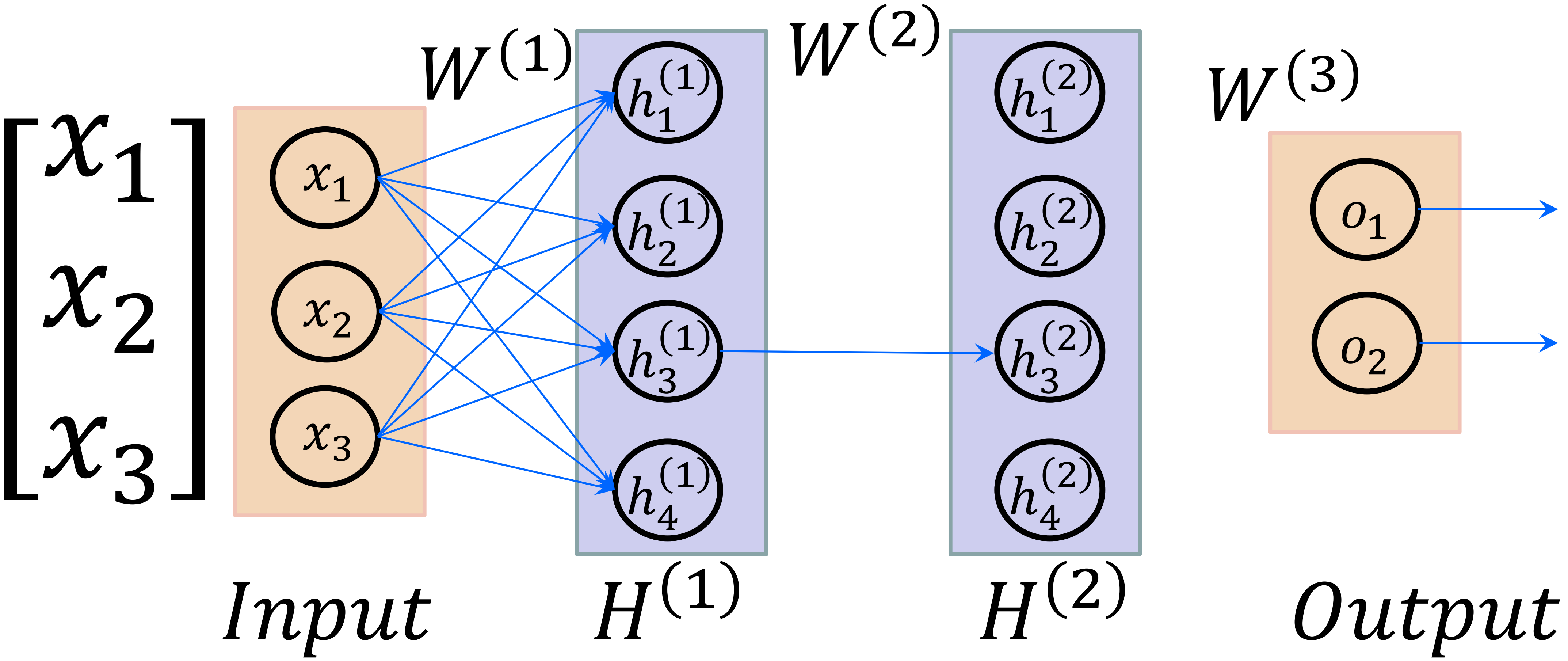
Forward propagation

– Ma trận trọng số $W^{(2)}$

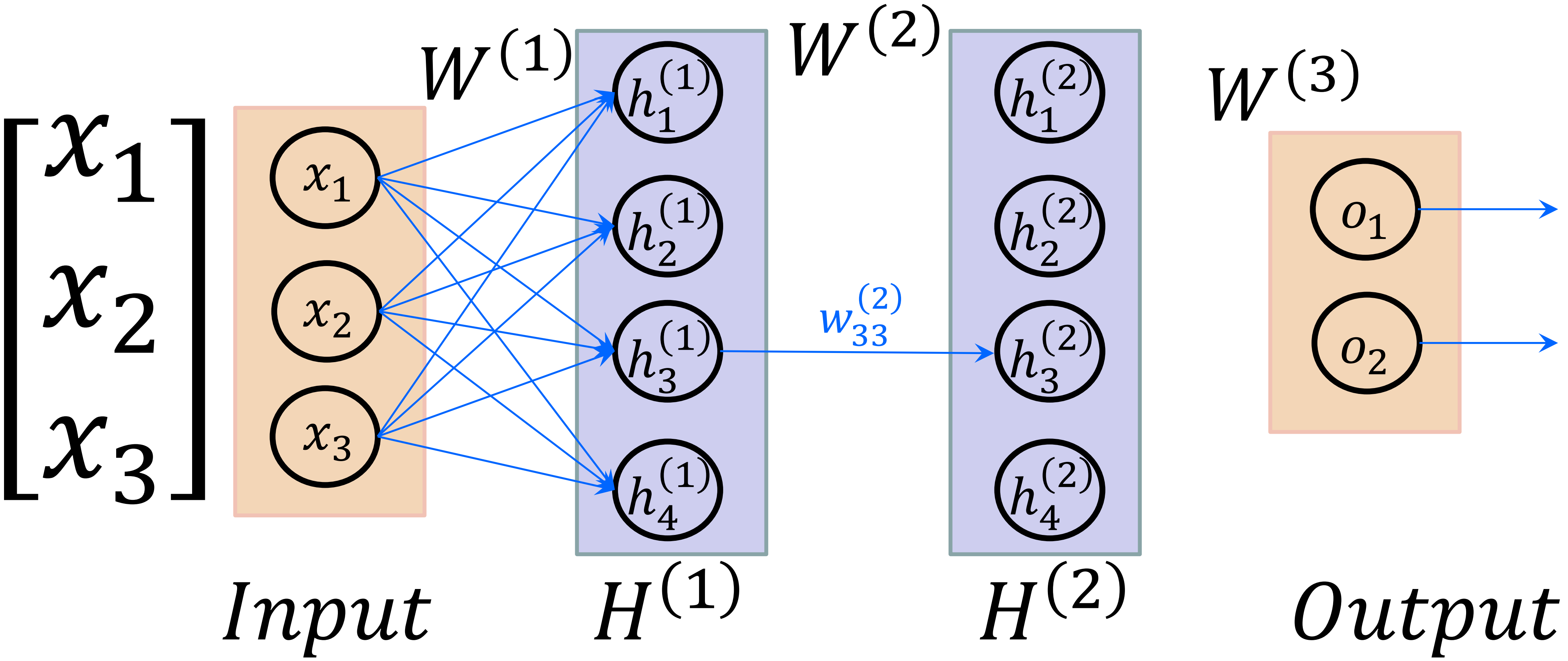
$$W^{(2)} = \begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} & w_{14}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)} & w_{24}^{(2)} \\ w_{31}^{(2)} & w_{32}^{(2)} & & \\ & & & \end{bmatrix}$$



Forward propagation



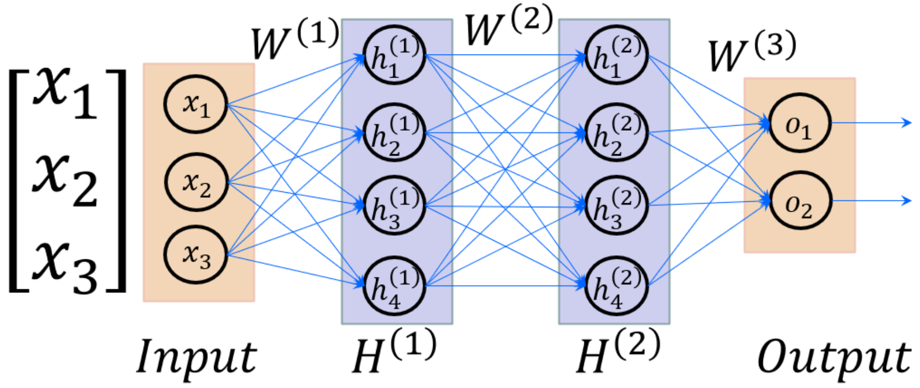
Forward propagation



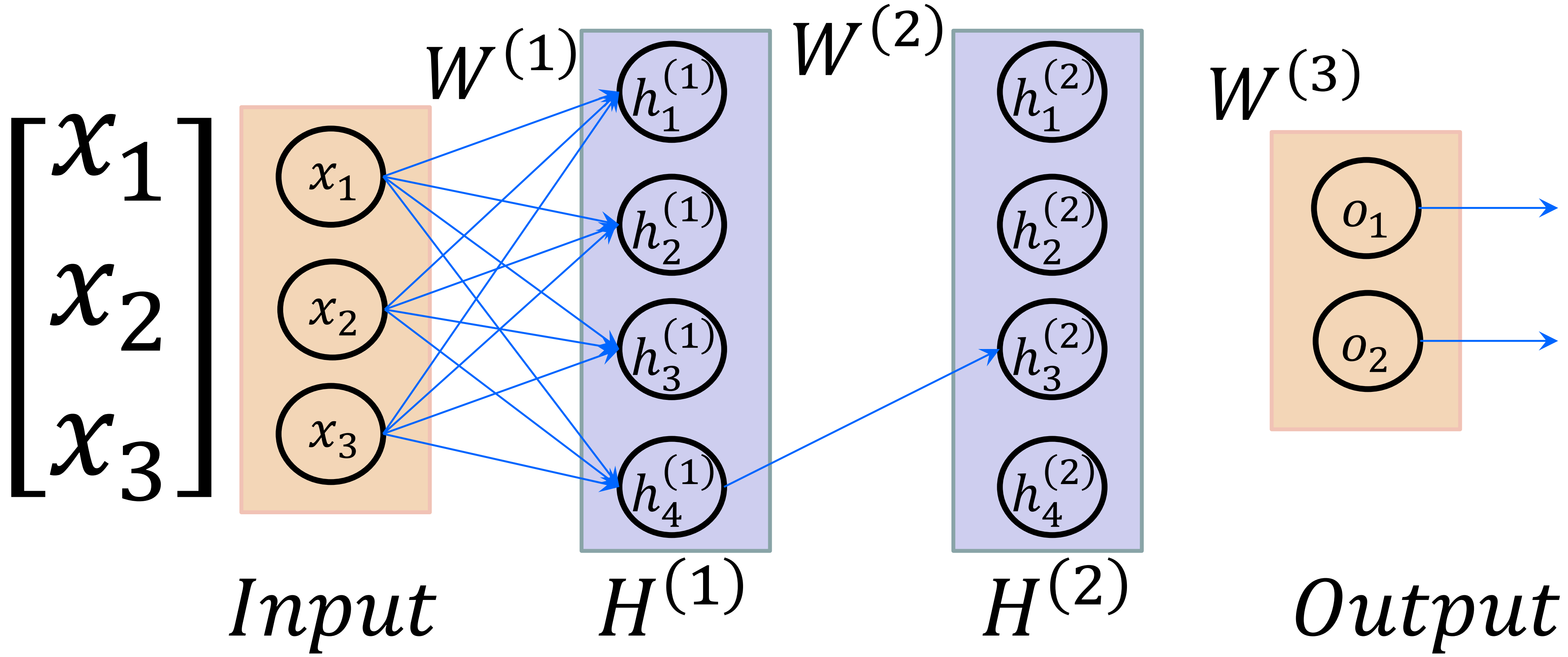
Forward propagation

– Ma trận trọng số $W^{(2)}$

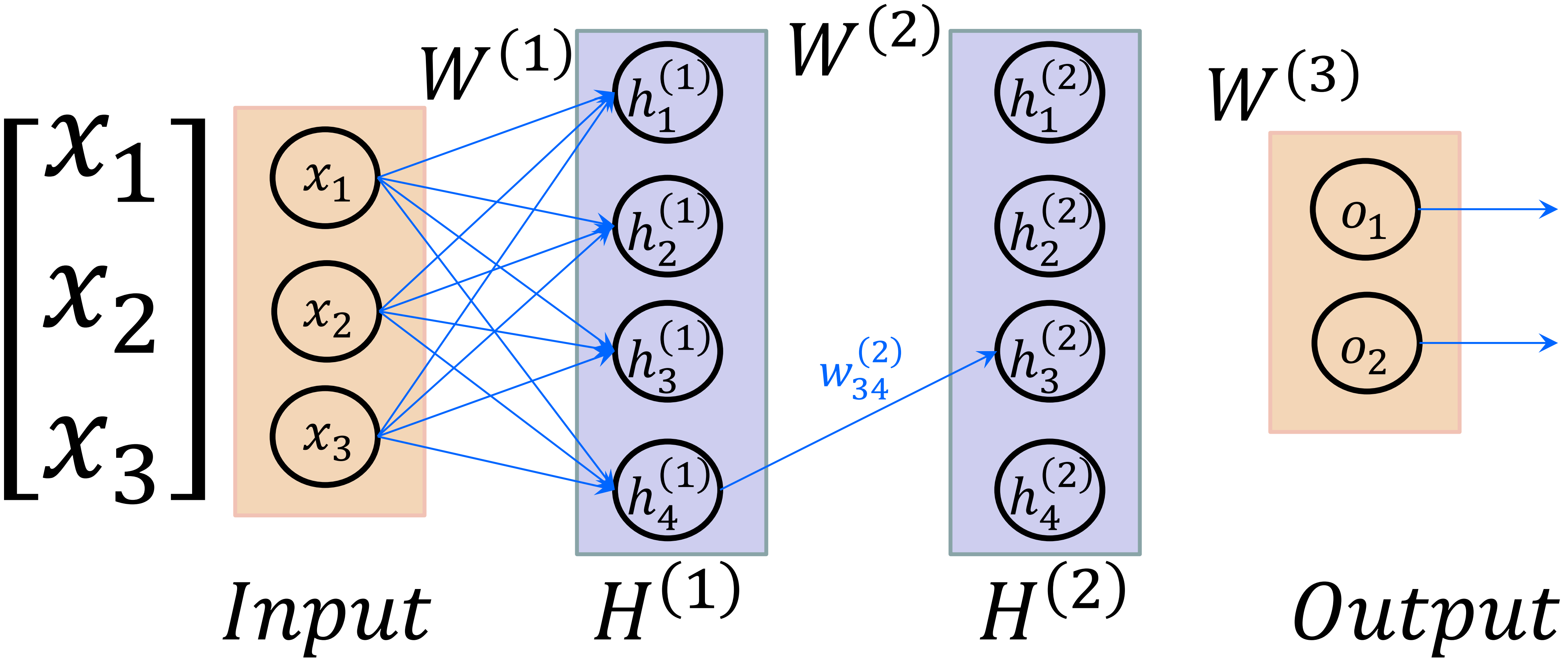
$$W^{(2)} = \begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} & w_{14}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)} & w_{24}^{(2)} \\ w_{31}^{(2)} & w_{32}^{(2)} & w_{33}^{(2)} & \\ & & & \end{bmatrix}$$



Forward propagation



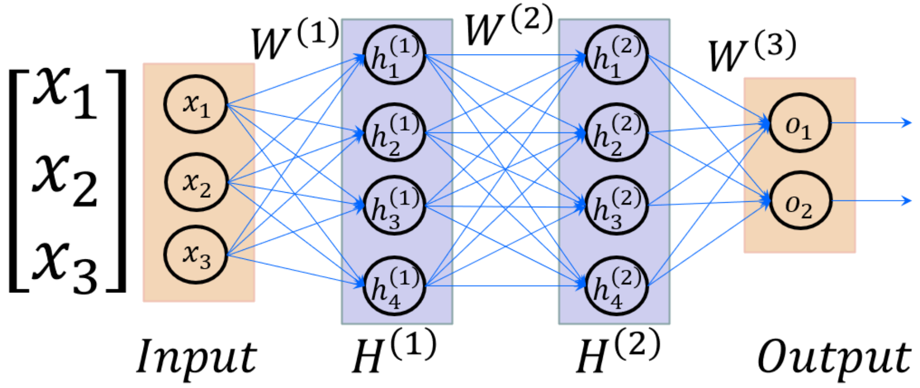
Forward propagation



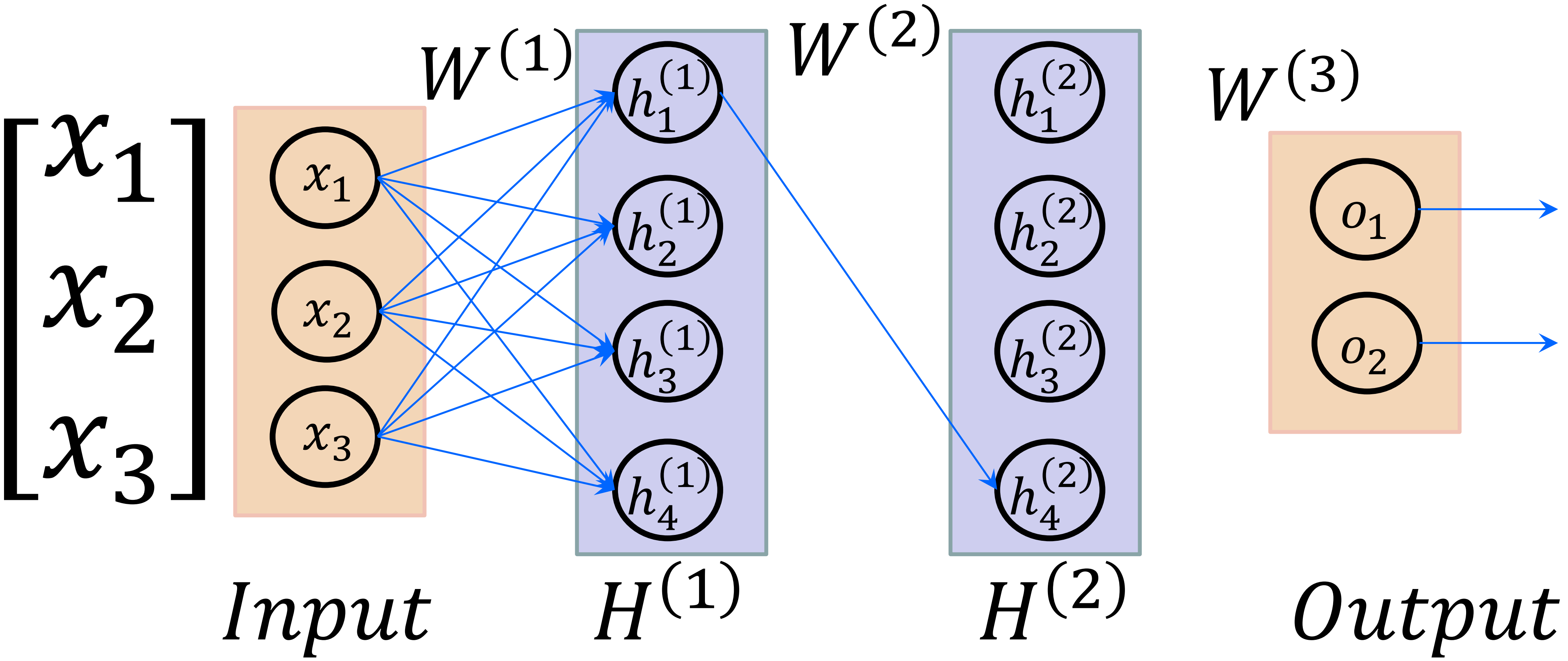
Forward propagation

– Ma trận trọng số $W^{(2)}$

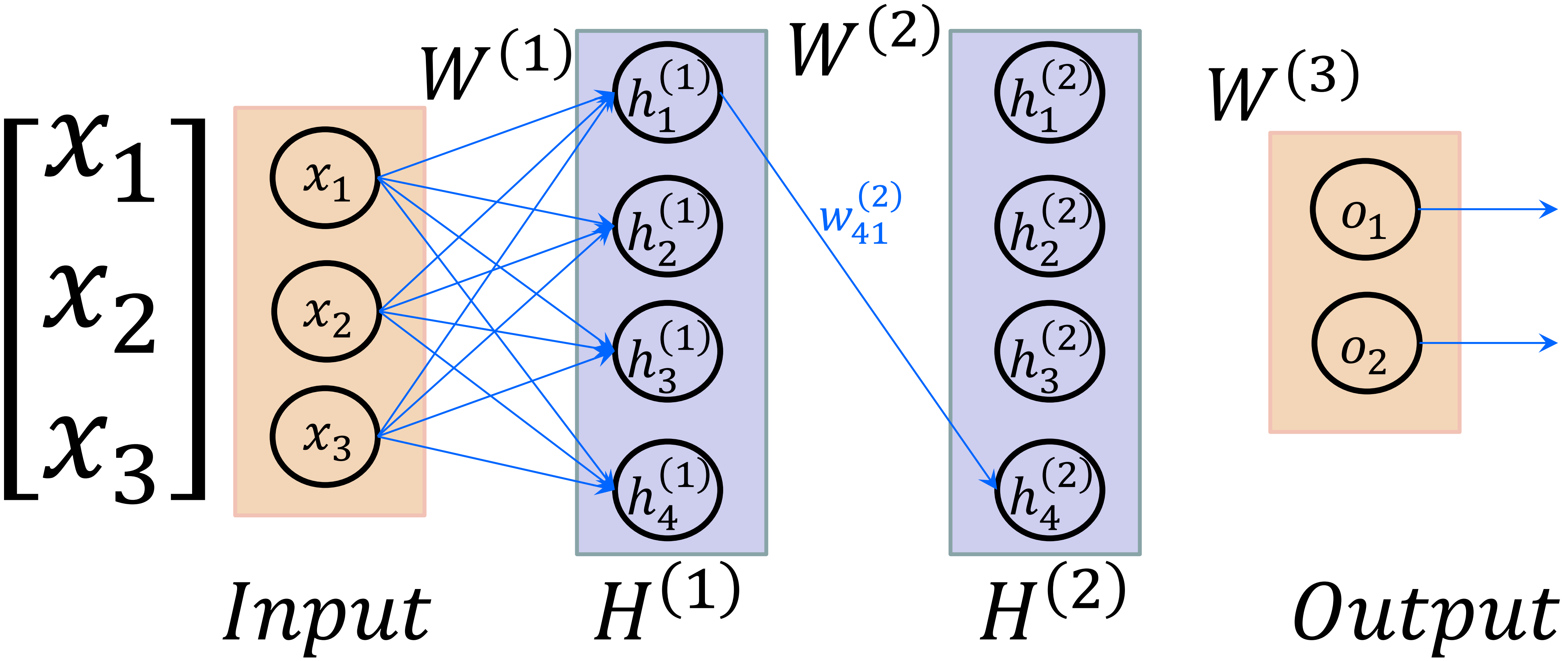
$$W^{(2)} = \begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} & w_{14}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)} & w_{24}^{(2)} \\ w_{31}^{(2)} & w_{32}^{(2)} & w_{33}^{(2)} & w_{34}^{(2)} \\ & & & \end{bmatrix}$$



Forward propagation



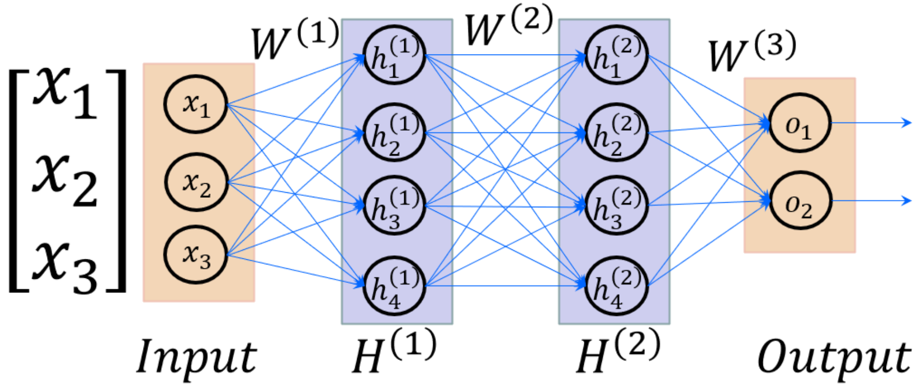
Forward propagation



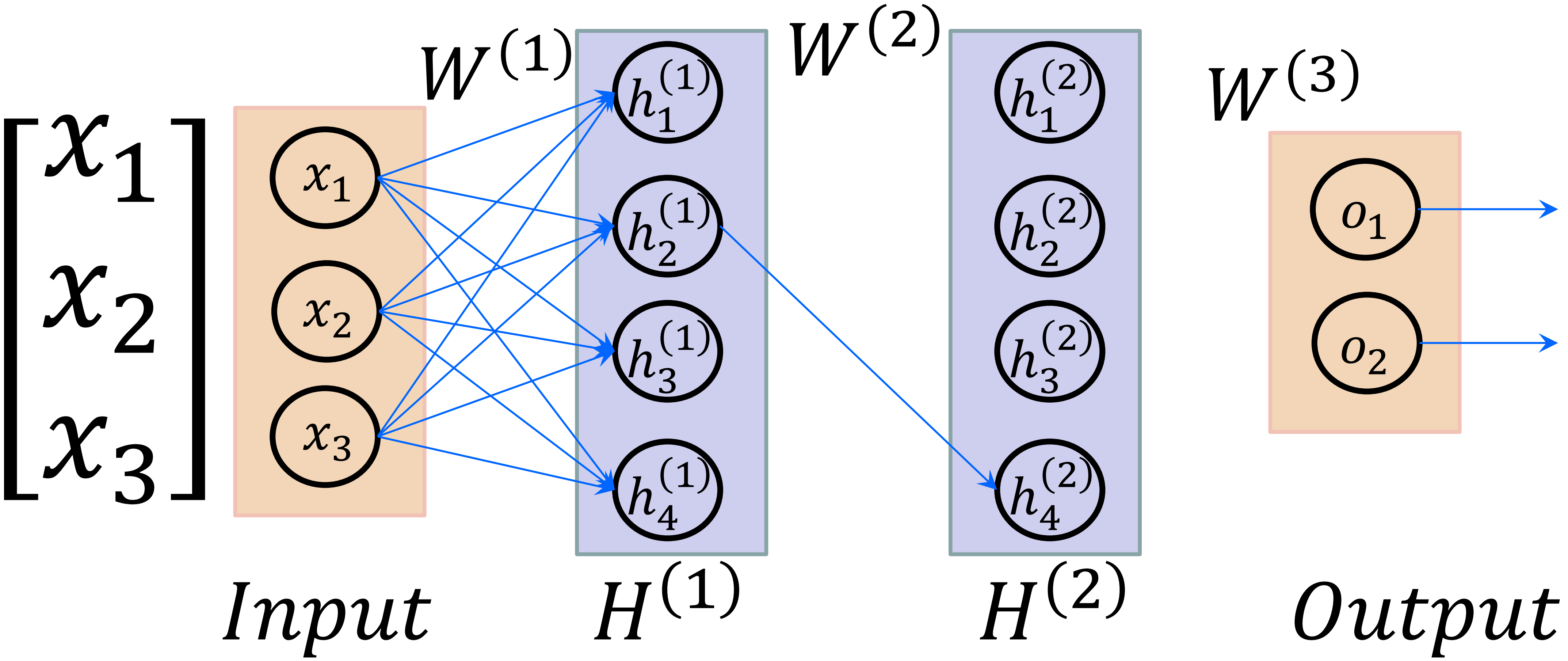
Forward propagation

– Ma trận trọng số $W^{(2)}$

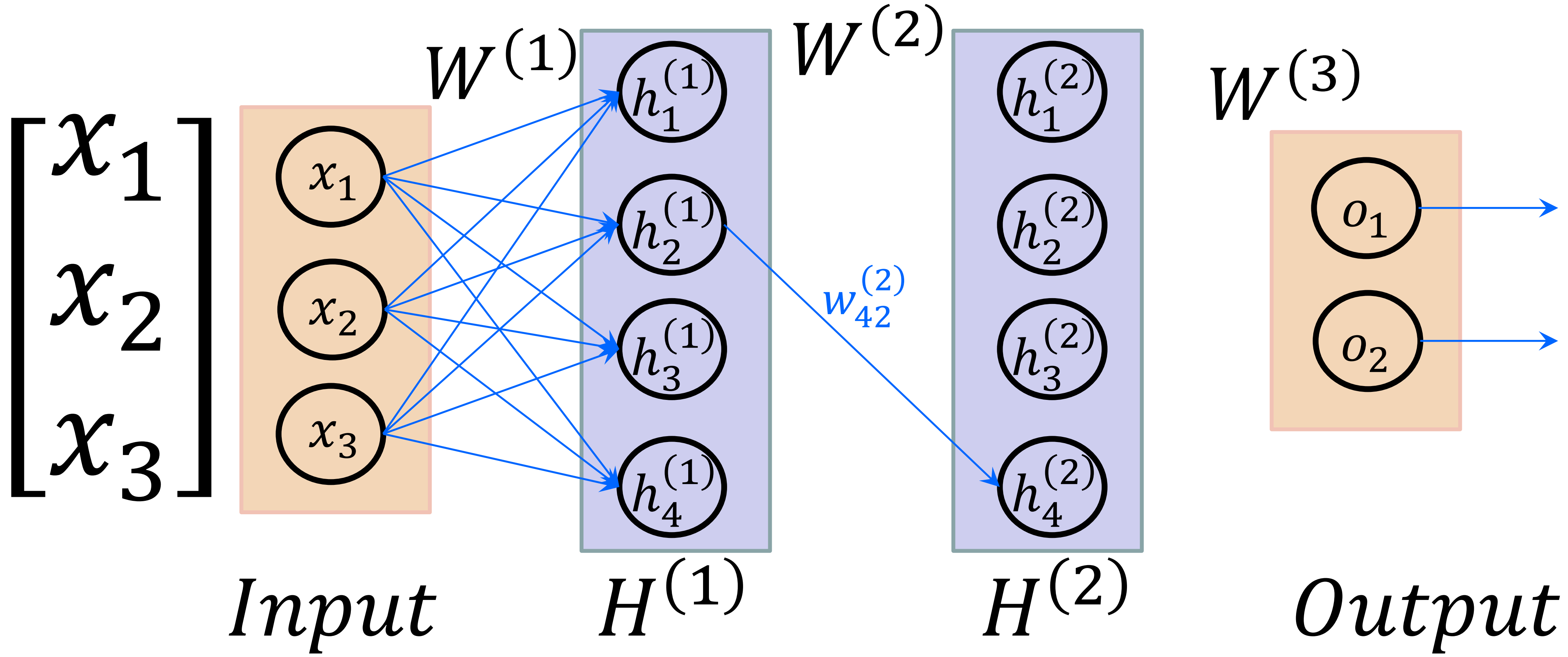
$$W^{(2)} = \begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} & w_{14}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)} & w_{24}^{(2)} \\ w_{31}^{(2)} & w_{32}^{(2)} & w_{33}^{(2)} & w_{34}^{(2)} \\ w_{41}^{(2)} & & & \end{bmatrix}$$



Forward propagation



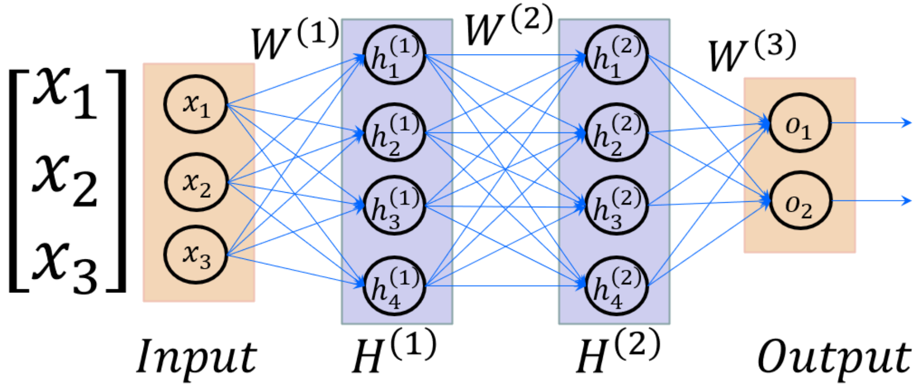
Forward propagation



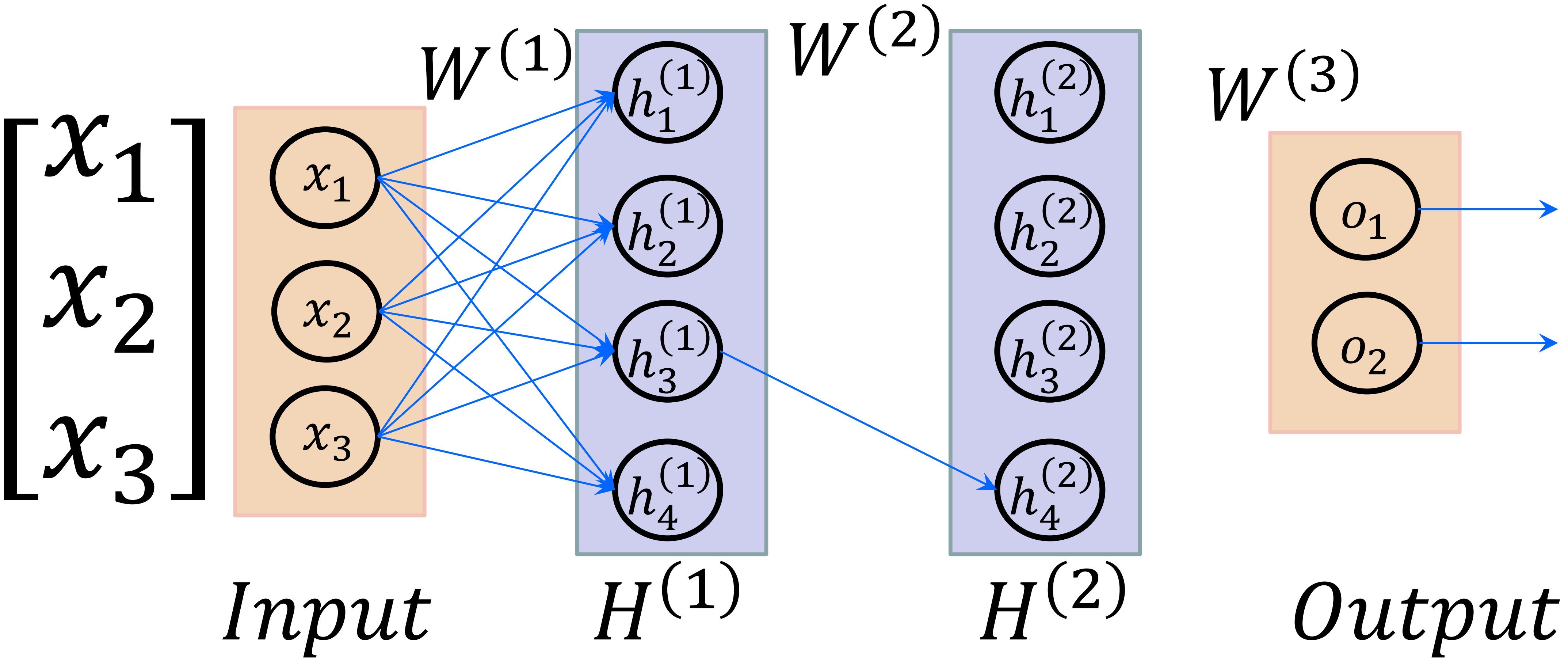
Forward propagation

– Ma trận trọng số $W^{(2)}$

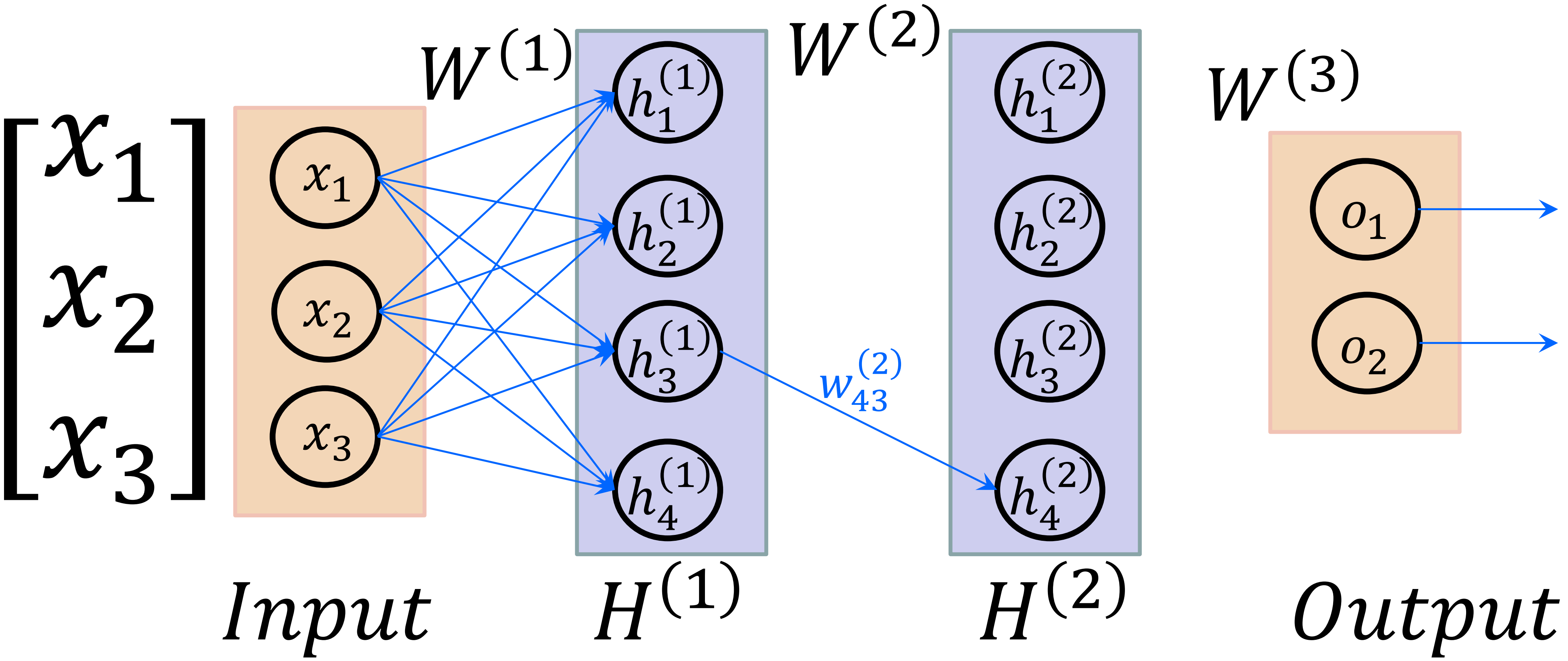
$$W^{(2)} = \begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} & w_{14}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)} & w_{24}^{(2)} \\ w_{31}^{(2)} & w_{32}^{(2)} & w_{33}^{(2)} & w_{34}^{(2)} \\ w_{41}^{(2)} & w_{42}^{(2)} & & \end{bmatrix}$$



Forward propagation



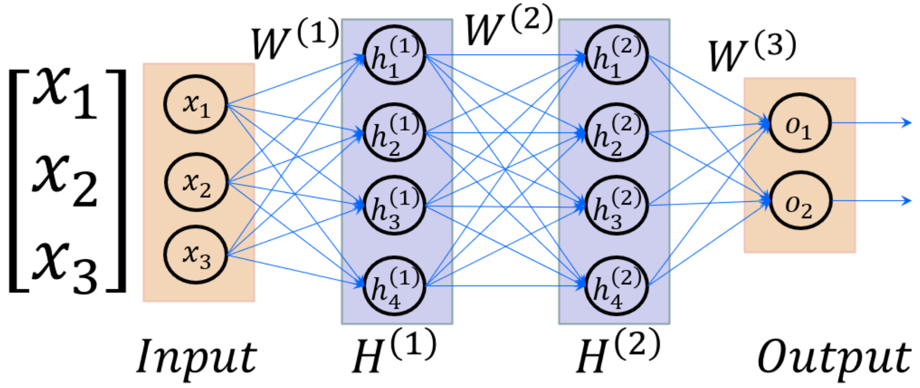
Forward propagation



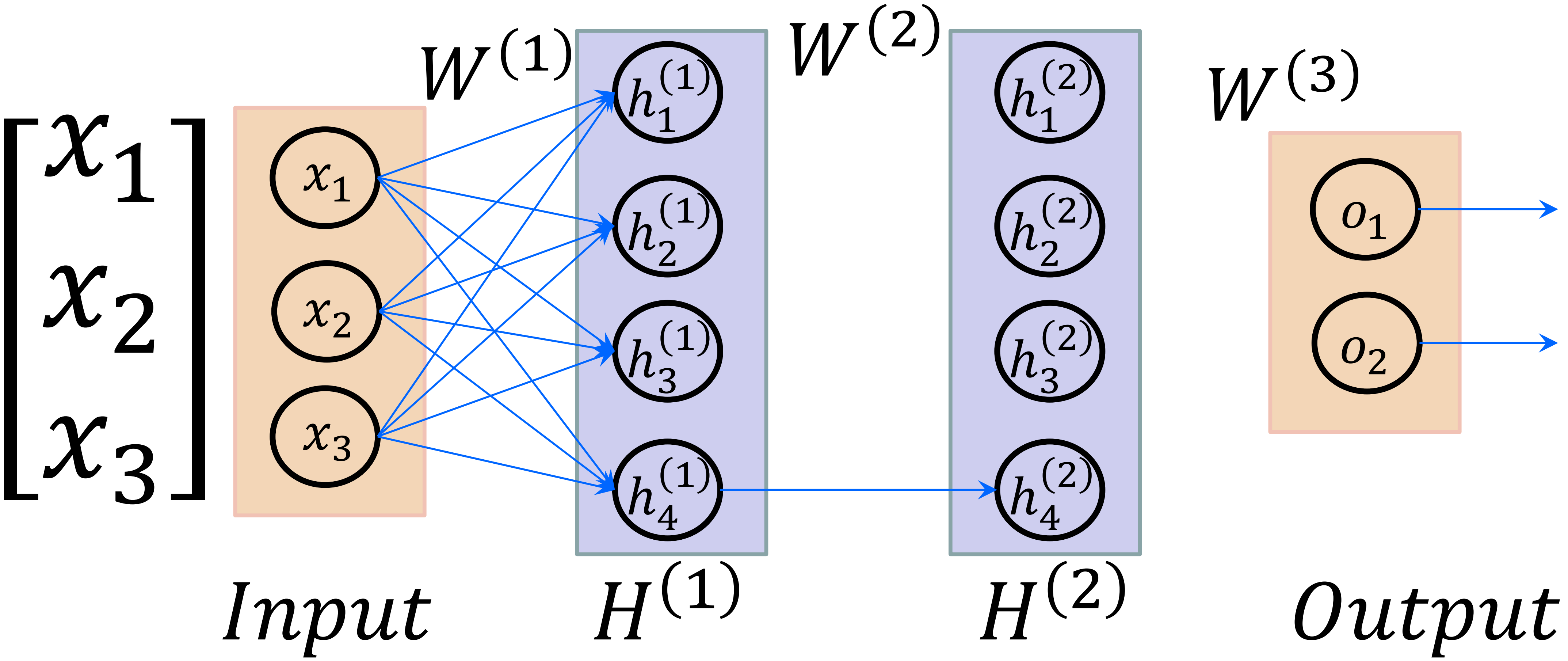
Forward propagation

– Ma trận trọng số $W^{(2)}$

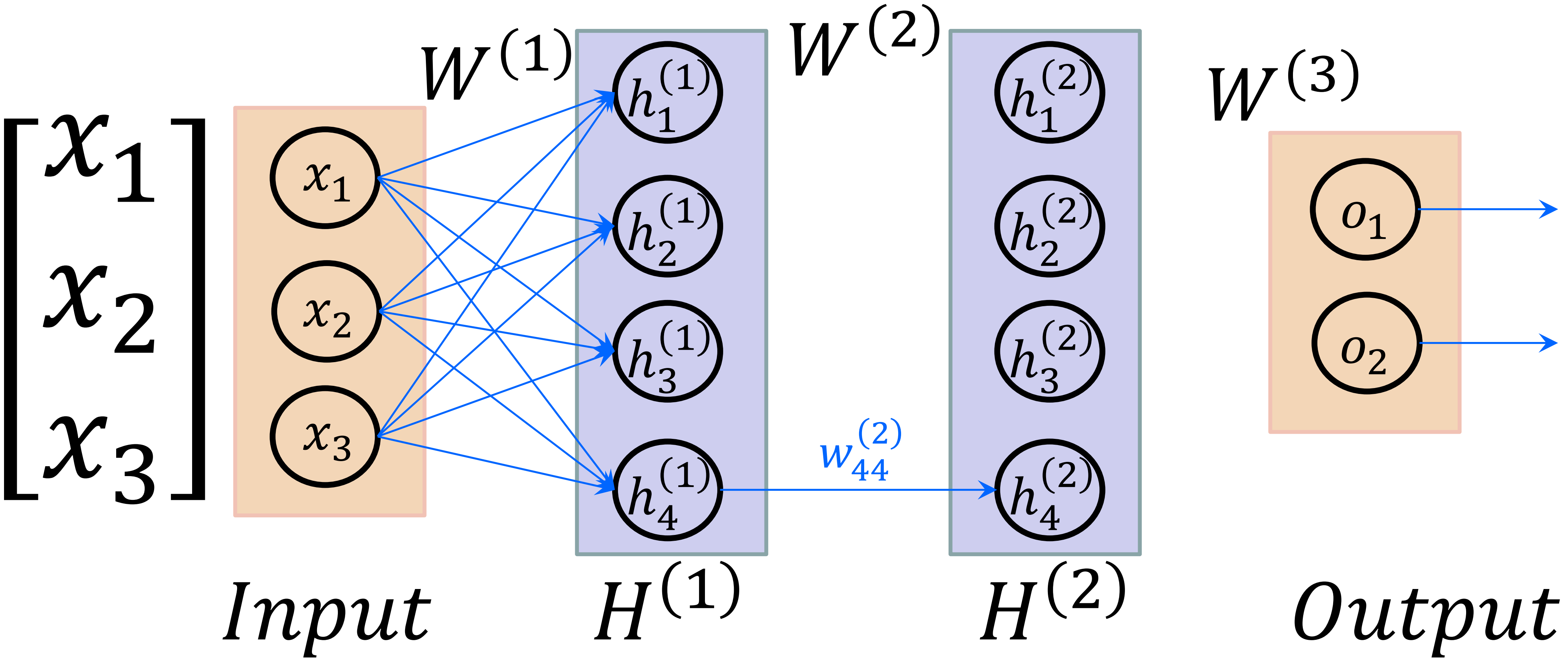
$$W^{(2)} = \begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} & w_{14}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)} & w_{24}^{(2)} \\ w_{31}^{(2)} & w_{32}^{(2)} & w_{33}^{(2)} & w_{34}^{(2)} \\ w_{41}^{(2)} & w_{42}^{(2)} & w_{43}^{(2)} & \end{bmatrix}$$



Forward propagation



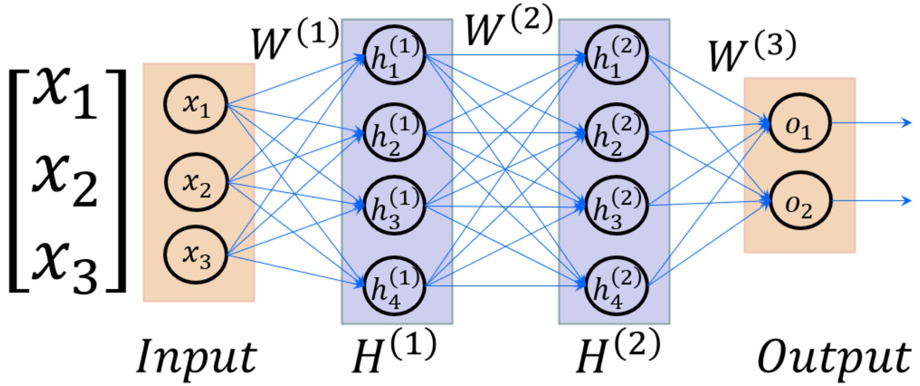
Forward propagation



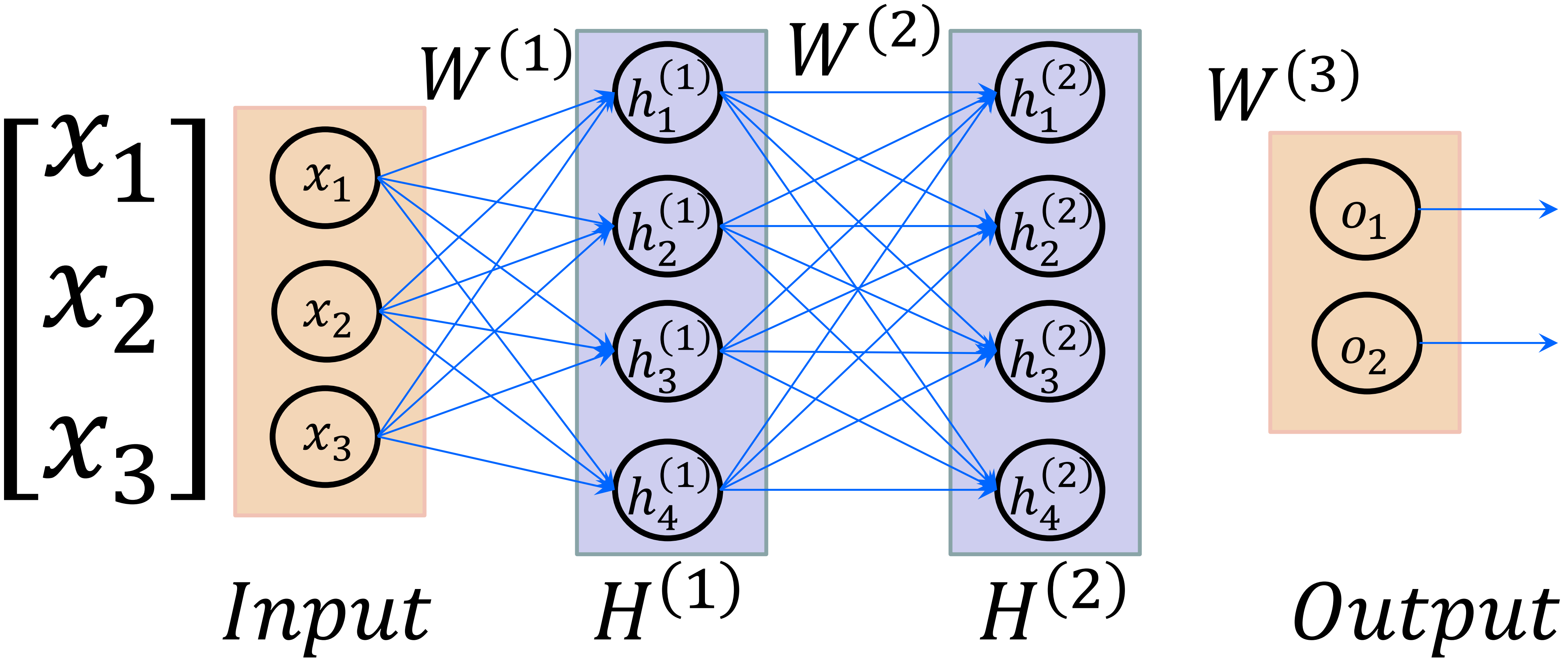
Forward propagation

– Ma trận trọng số $W^{(2)}$

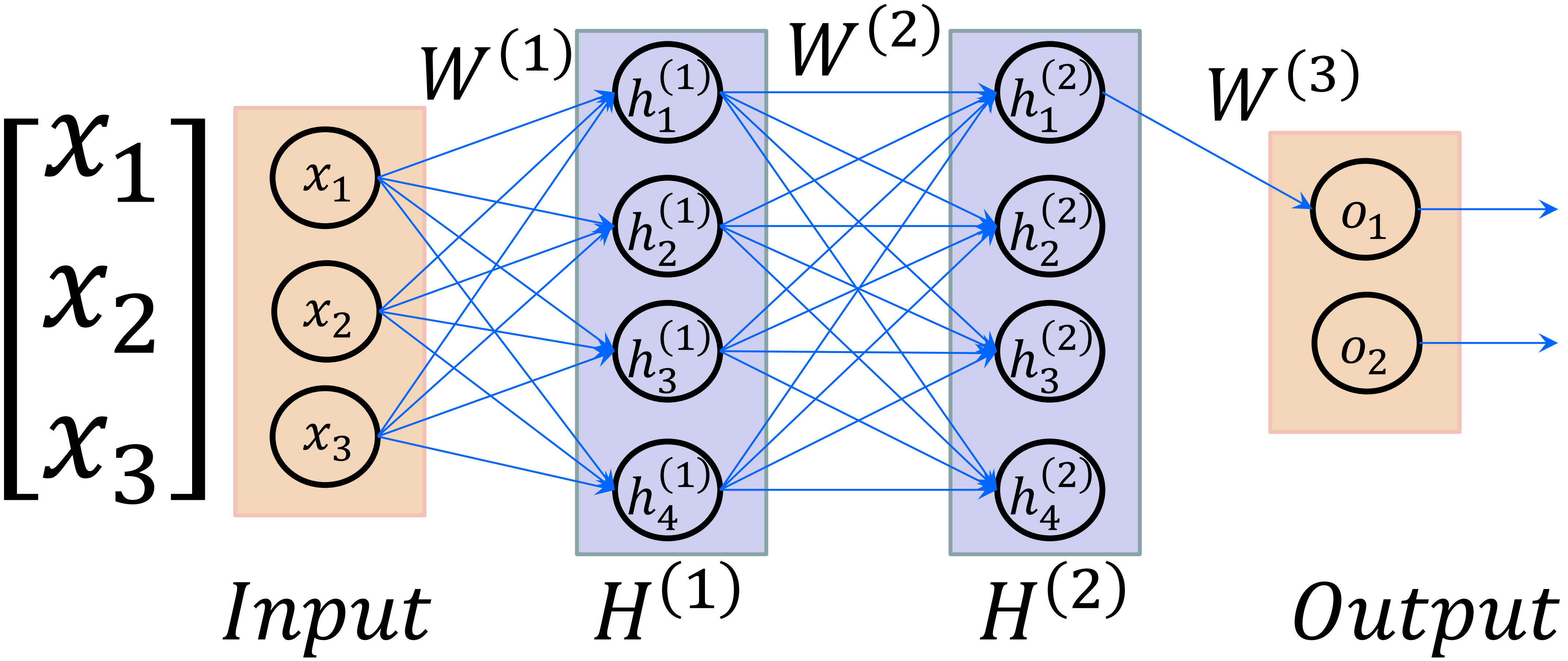
$$W^{(2)} = \begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} & w_{14}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)} & w_{24}^{(2)} \\ w_{31}^{(2)} & w_{32}^{(2)} & w_{33}^{(2)} & w_{34}^{(2)} \\ w_{41}^{(2)} & w_{42}^{(2)} & w_{43}^{(2)} & w_{44}^{(2)} \end{bmatrix}$$



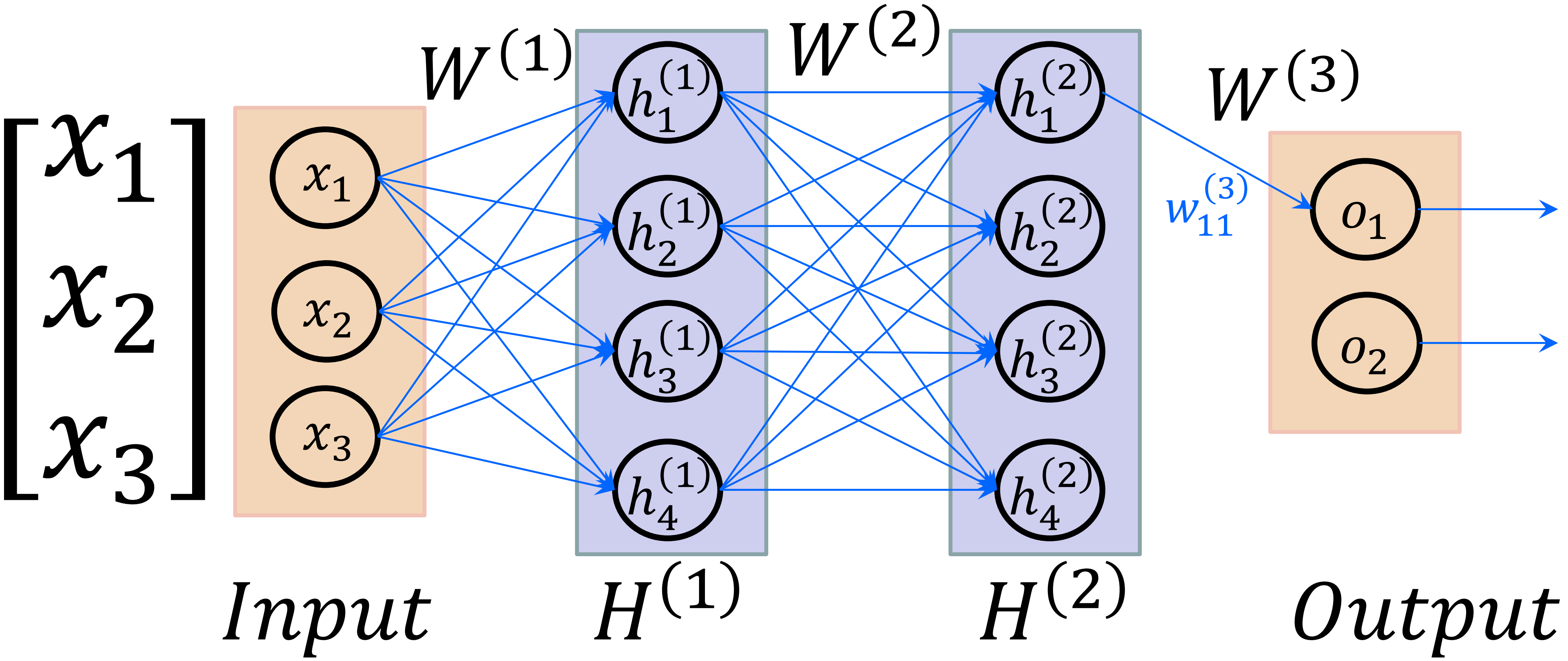
Forward propagation



Forward propagation



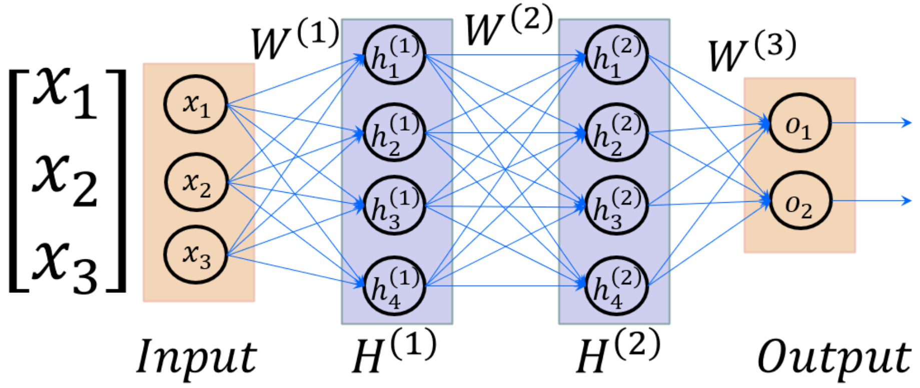
Forward propagation



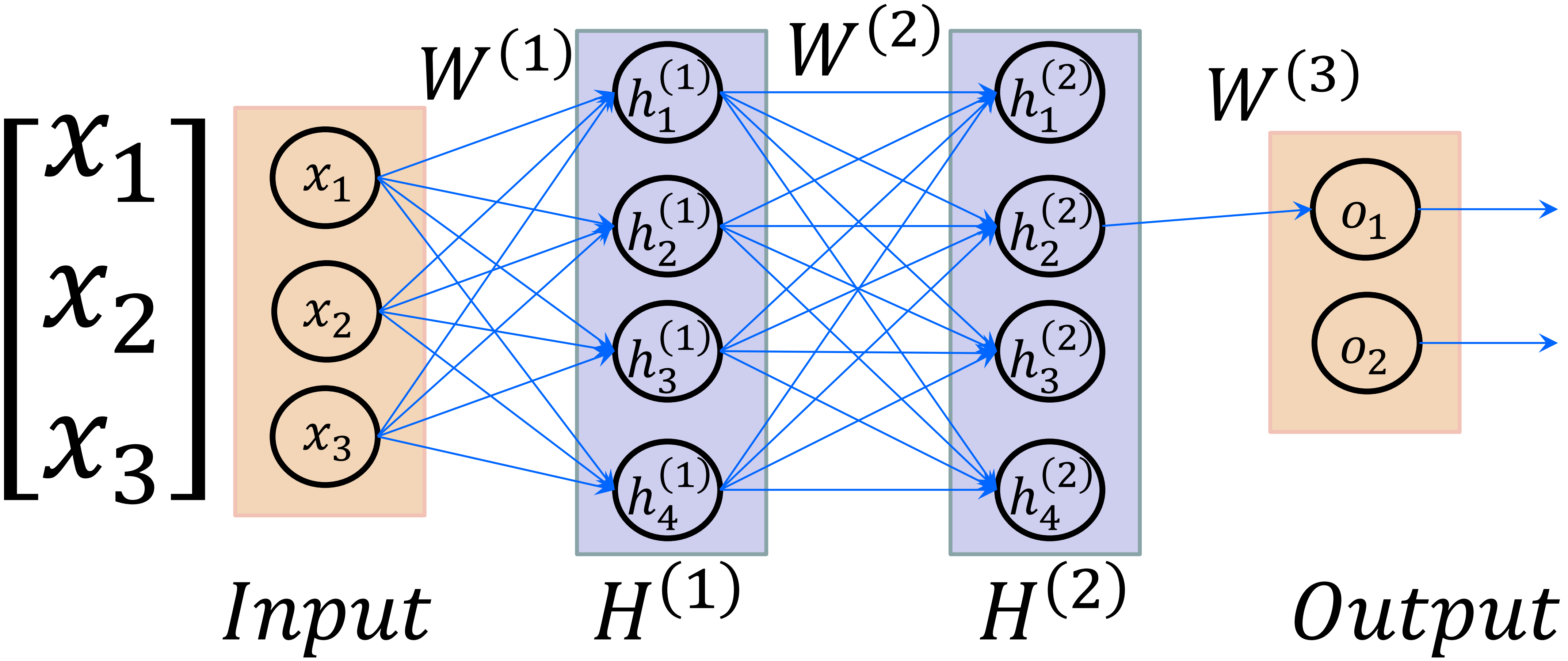
Forward propagation

– Ma trận trọng số $W^{(3)}$

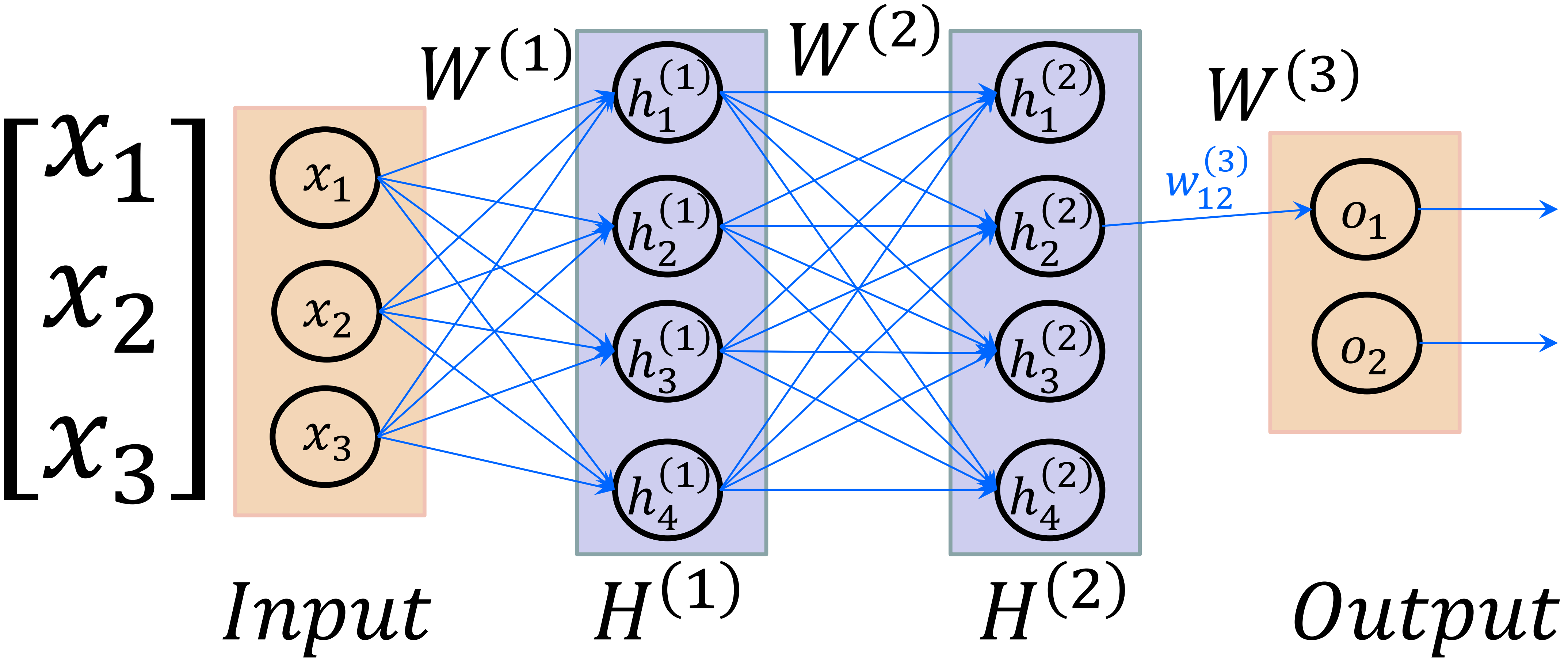
$$W^{(3)} = \begin{bmatrix} w_{11}^{(3)} & & \\ & & \end{bmatrix}$$



Forward propagation



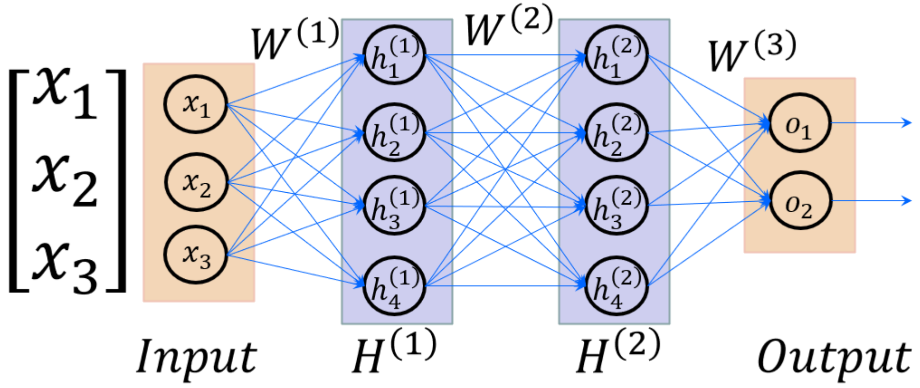
Forward propagation



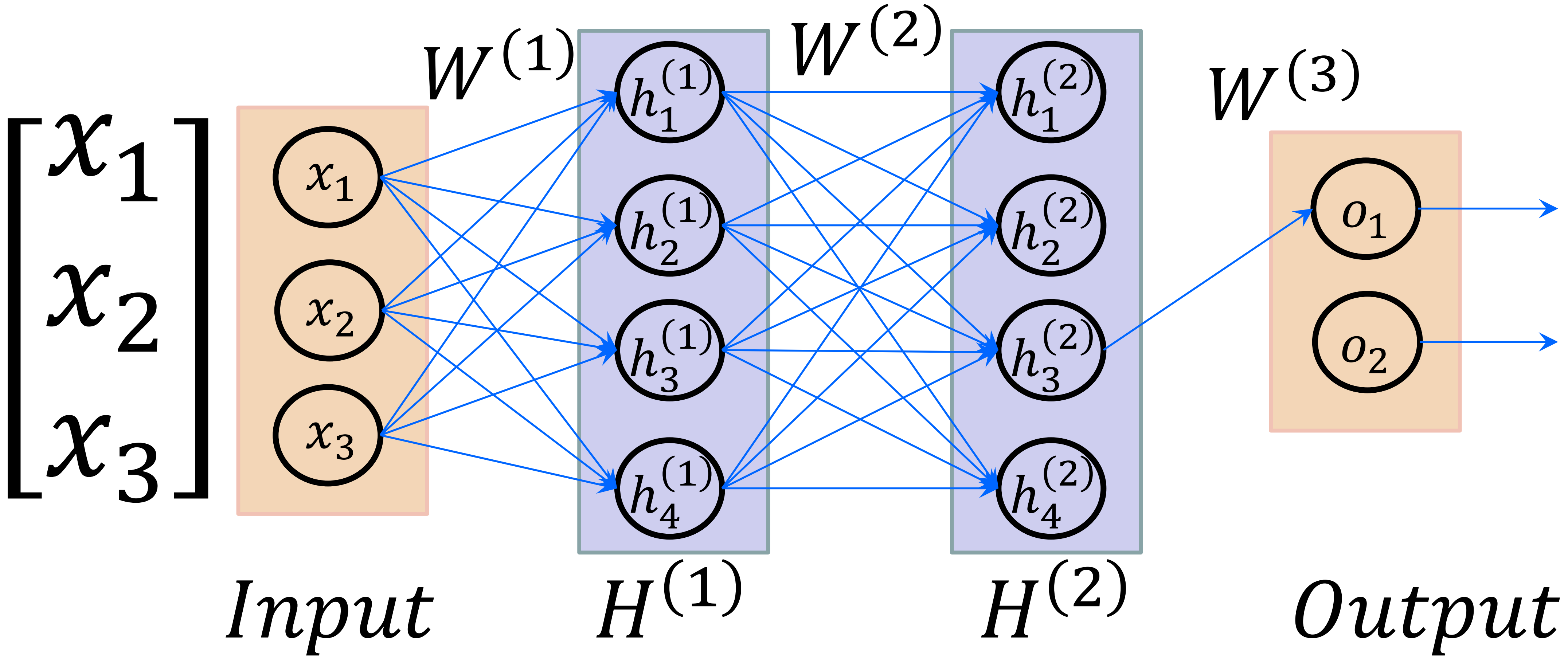
Forward propagation

– Ma trận trọng số $W^{(3)}$

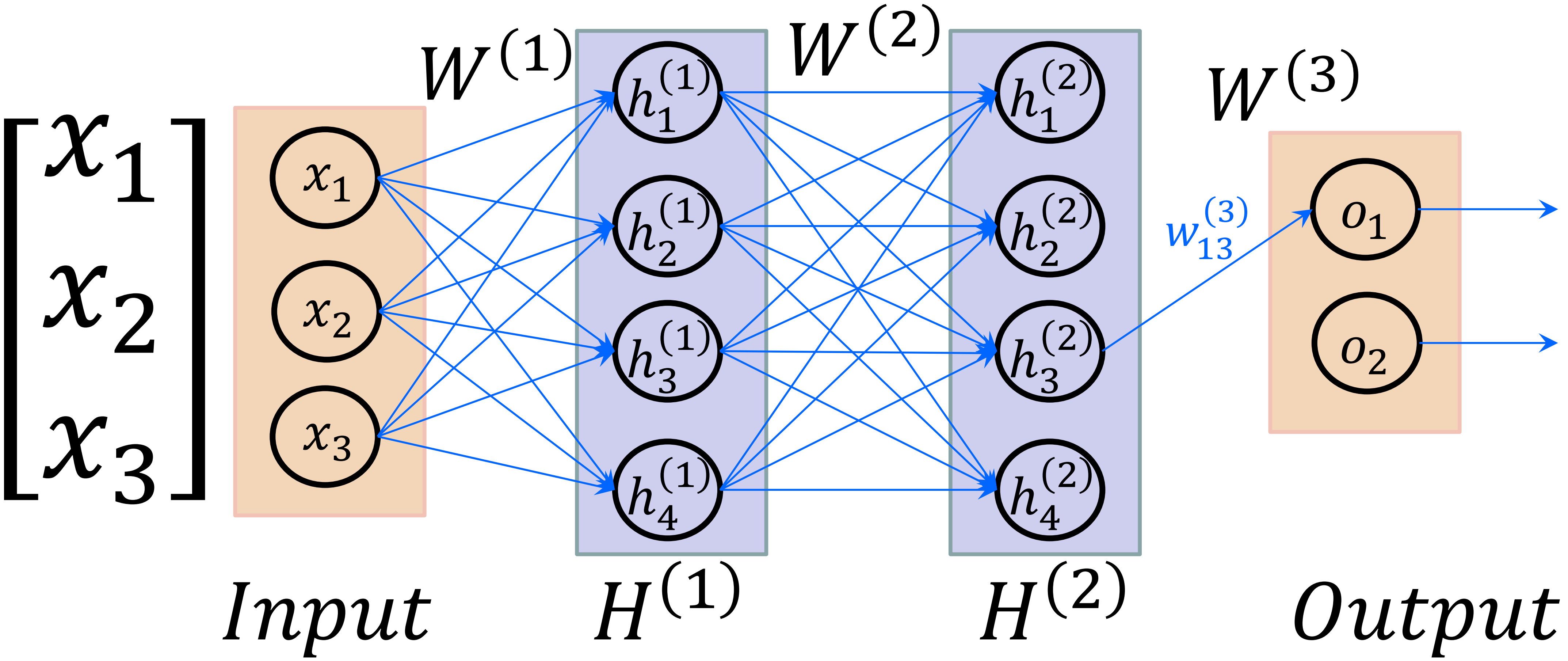
$$W^{(3)} = \begin{bmatrix} w_{11}^{(3)} & w_{12}^{(3)} & & \\ & & & \end{bmatrix}$$



Forward propagation



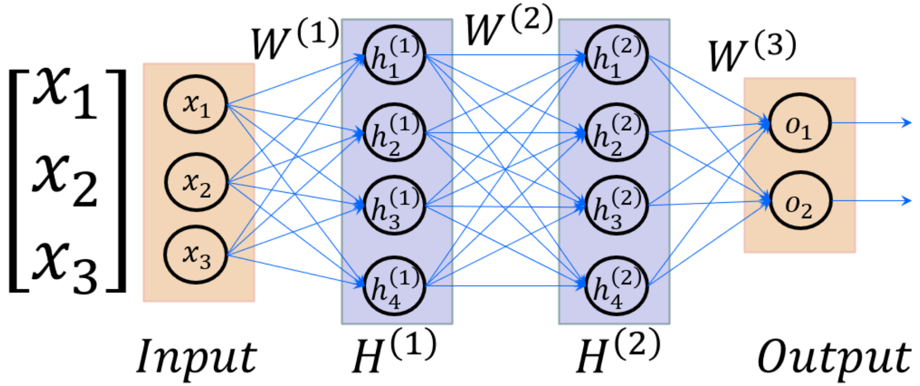
Forward propagation



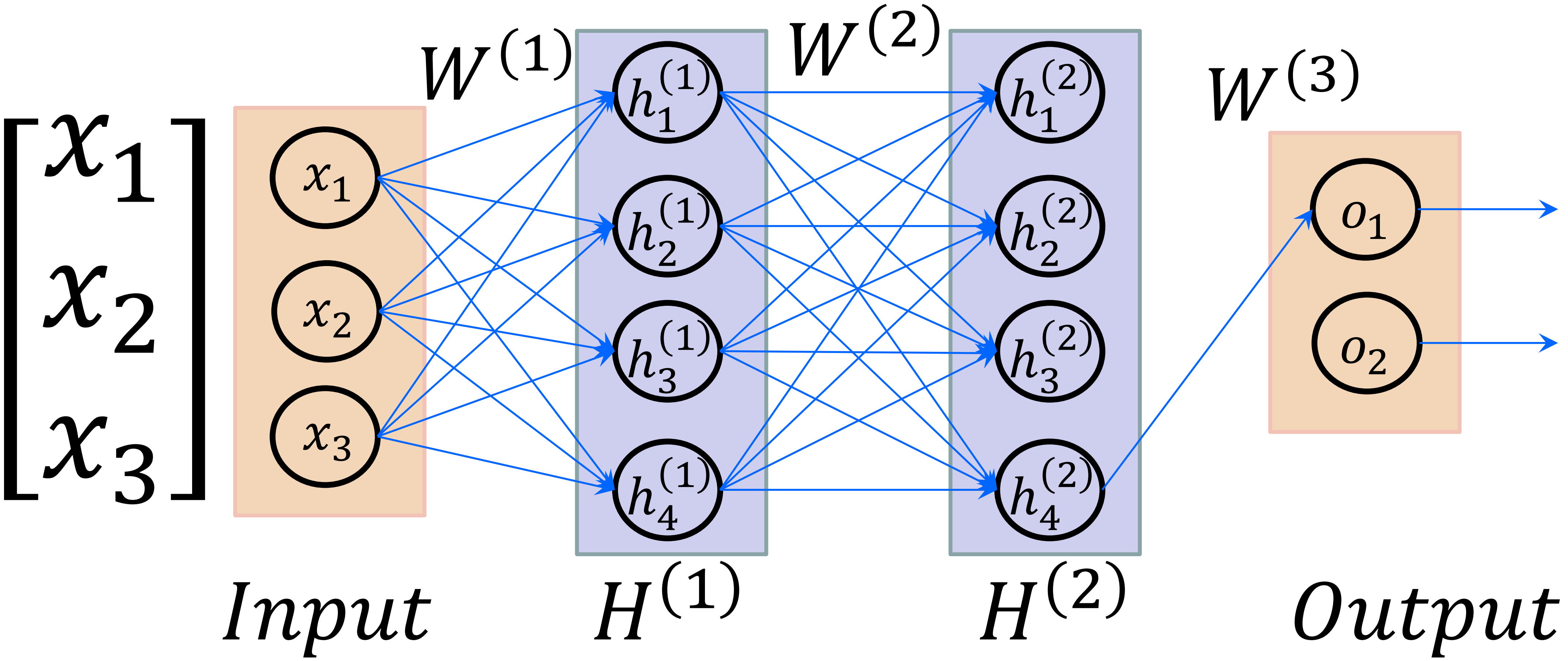
Forward propagation

– Ma trận trọng số $W^{(3)}$

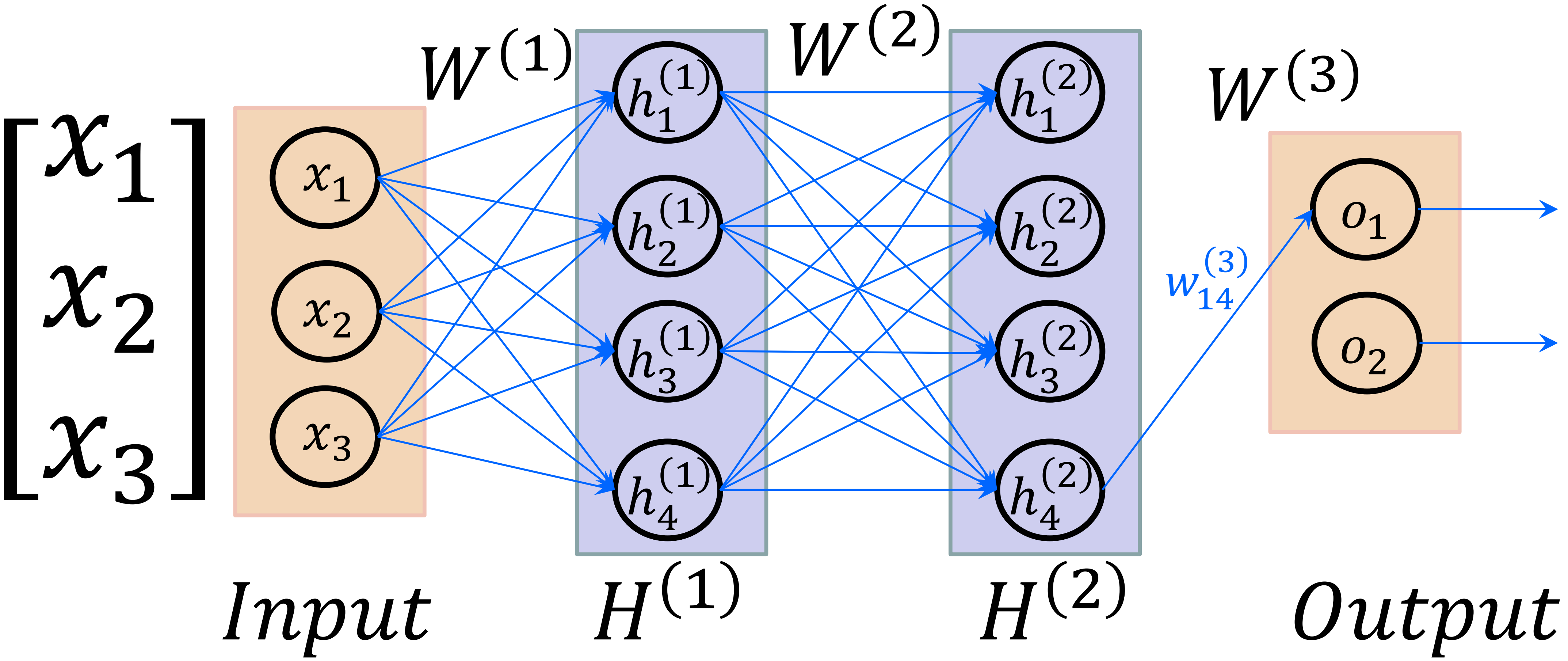
$$W^{(3)} = \begin{bmatrix} w_{11}^{(3)} & w_{12}^{(3)} & w_{13}^{(3)} & \\ & & & \end{bmatrix}$$



Forward propagation



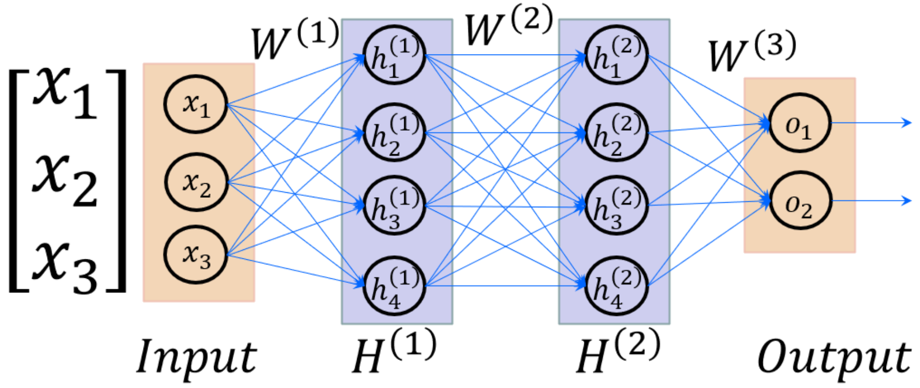
Forward propagation



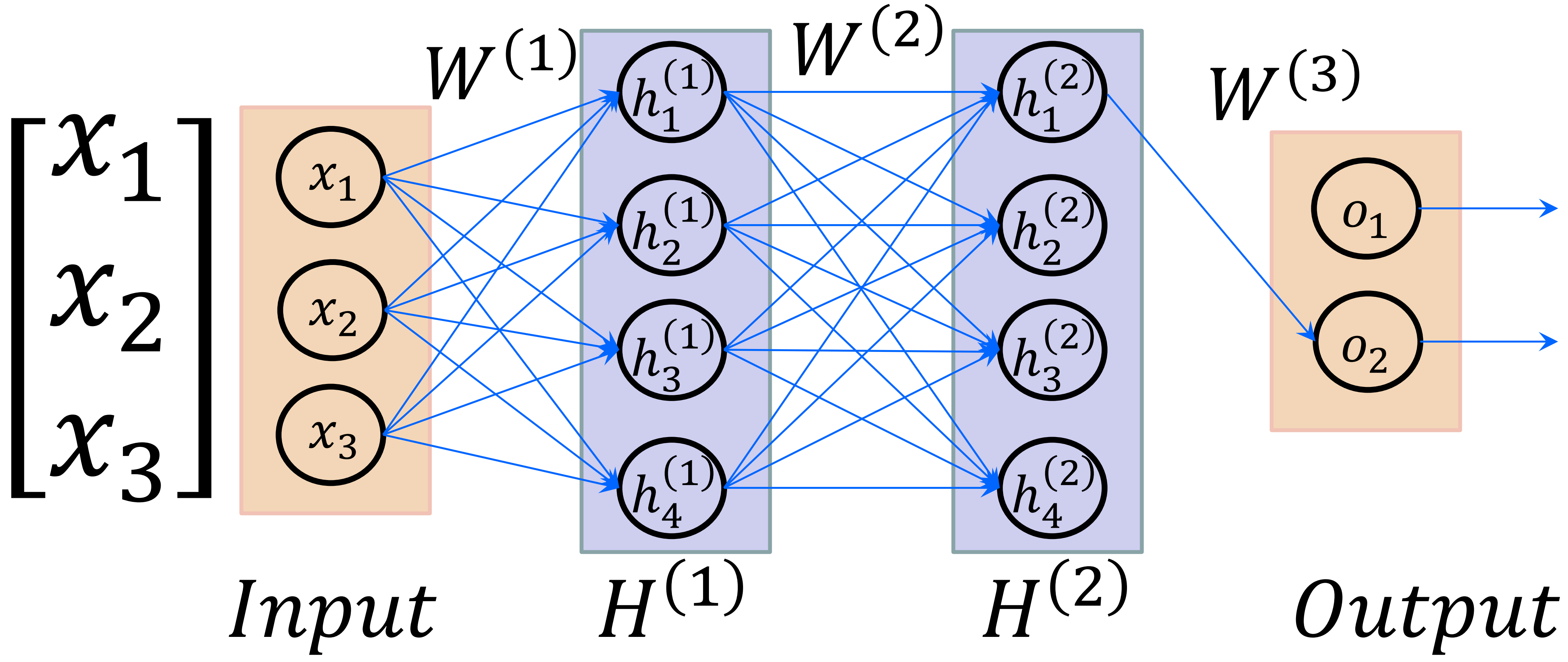
Forward propagation

– Ma trận trọng số $W^{(3)}$

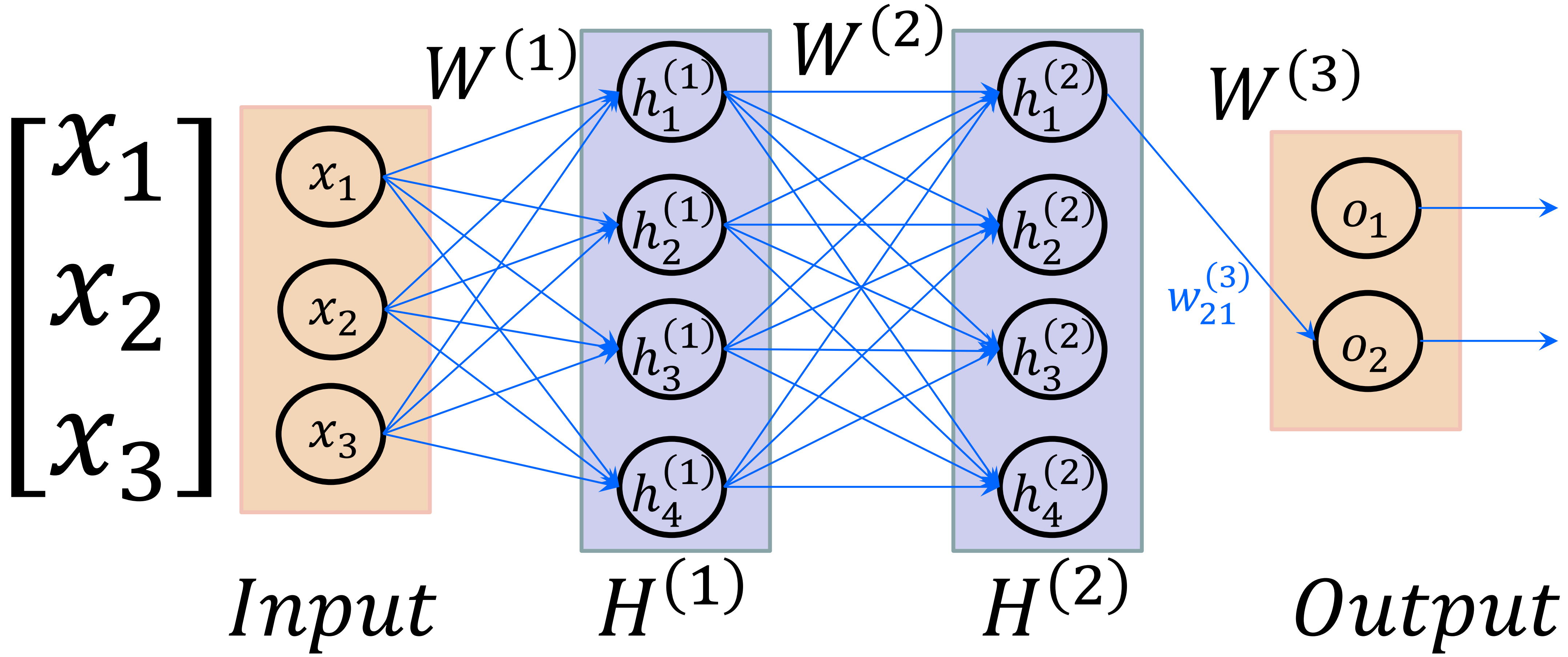
$$W^{(3)} = \begin{bmatrix} w_{11}^{(3)} & w_{12}^{(3)} & w_{13}^{(3)} & w_{14}^{(3)} \\ & & & \end{bmatrix}$$



Forward propagation



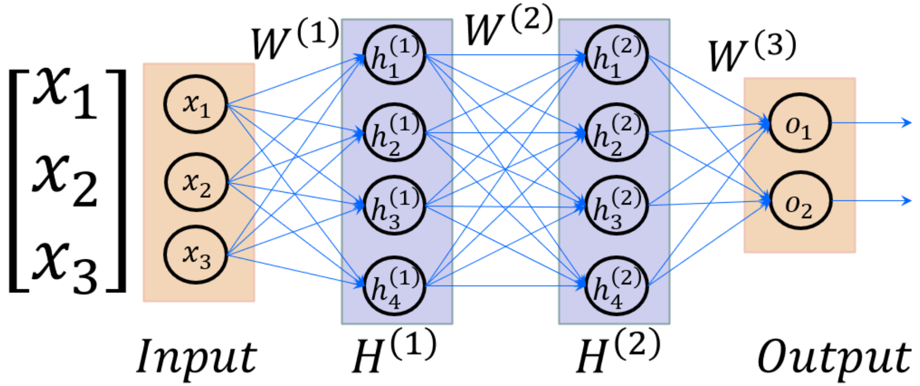
Forward propagation



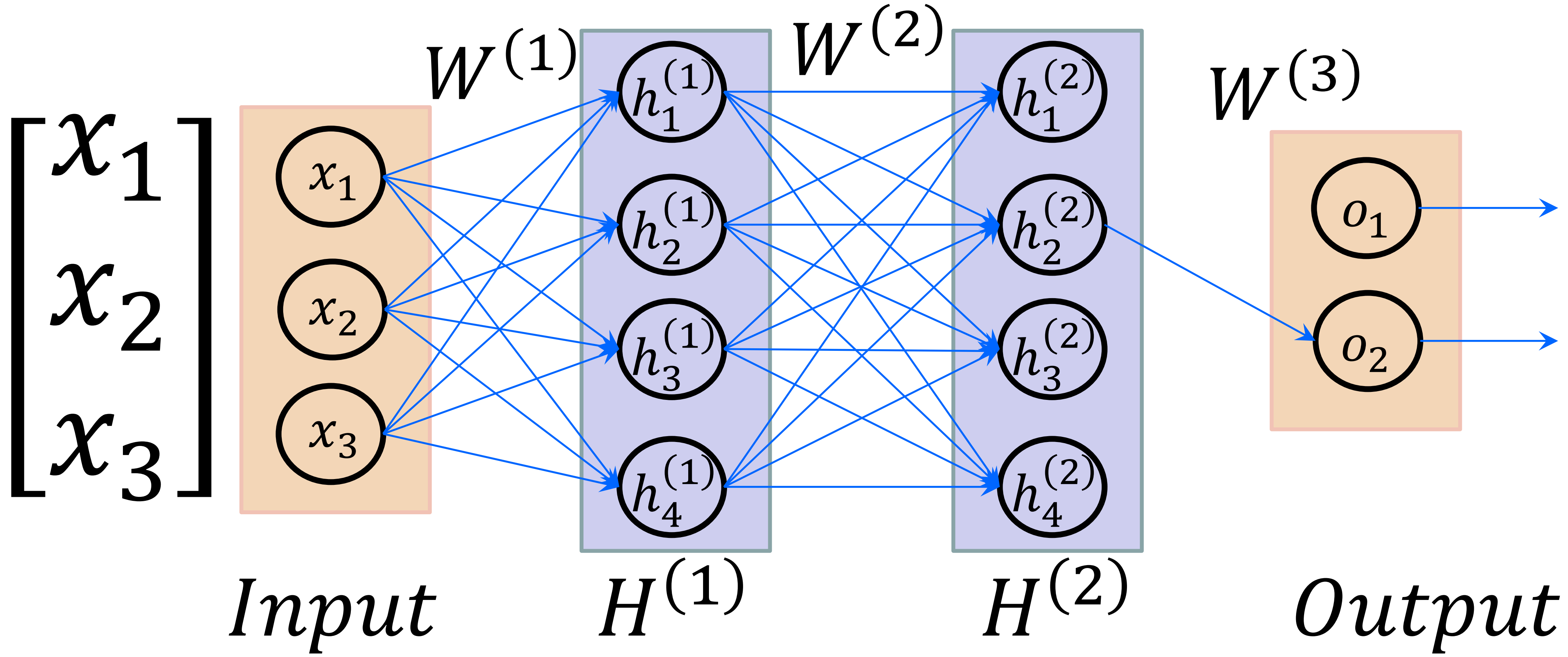
Forward propagation

– Ma trận trọng số $W^{(3)}$

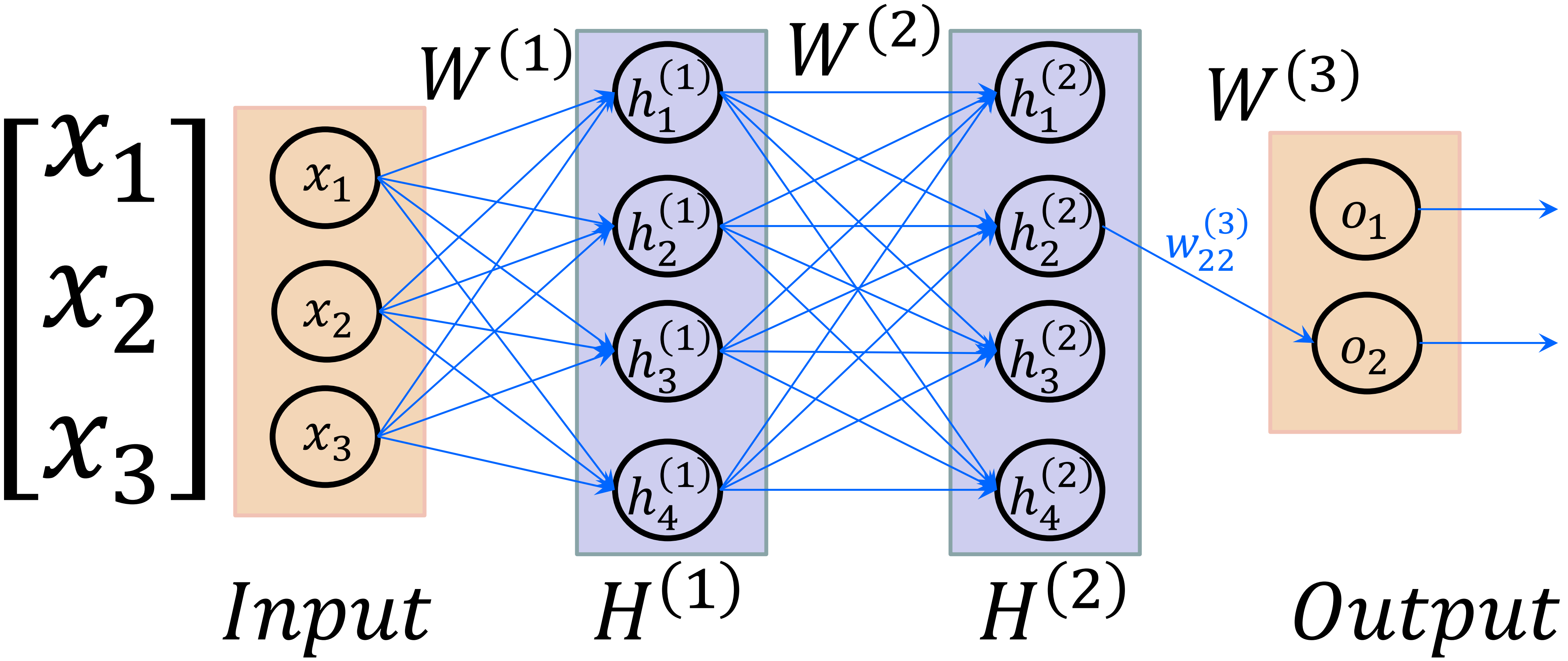
$$W^{(3)} = \begin{bmatrix} w_{11}^{(3)} & w_{12}^{(3)} & w_{13}^{(3)} & w_{14}^{(3)} \\ w_{21}^{(3)} & & & \end{bmatrix}$$



Forward propagation



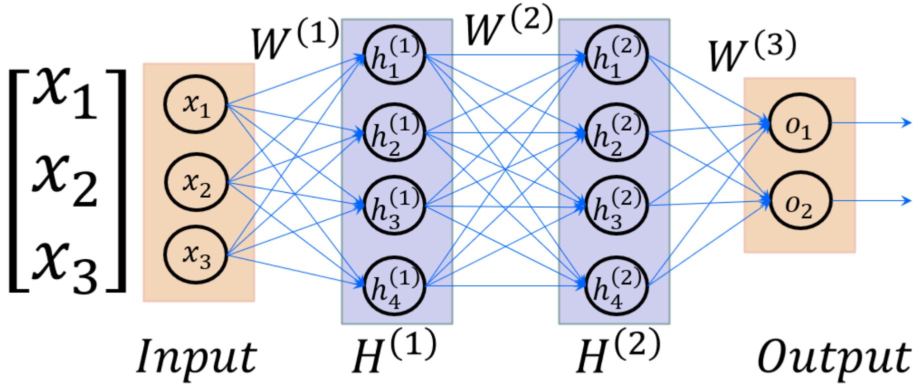
Forward propagation



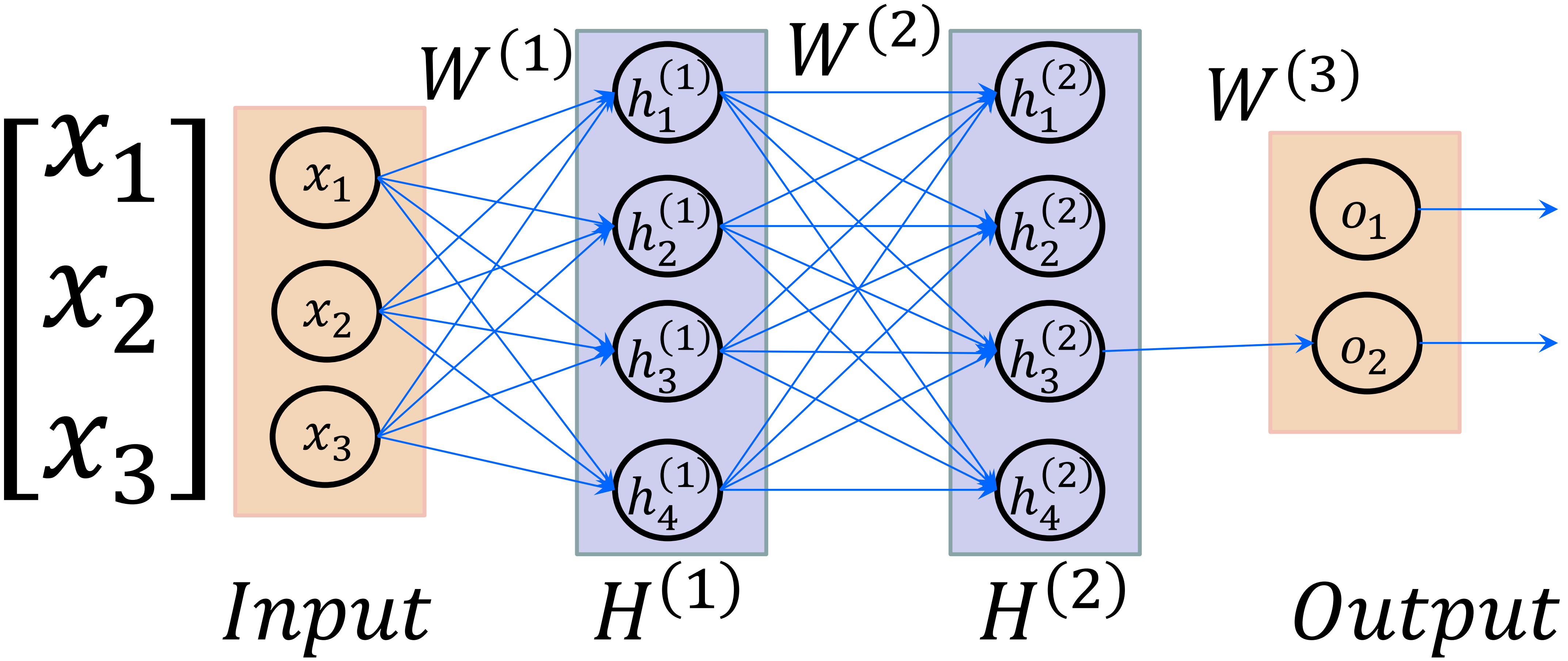
Forward propagation

– Ma trận trọng số $W^{(3)}$

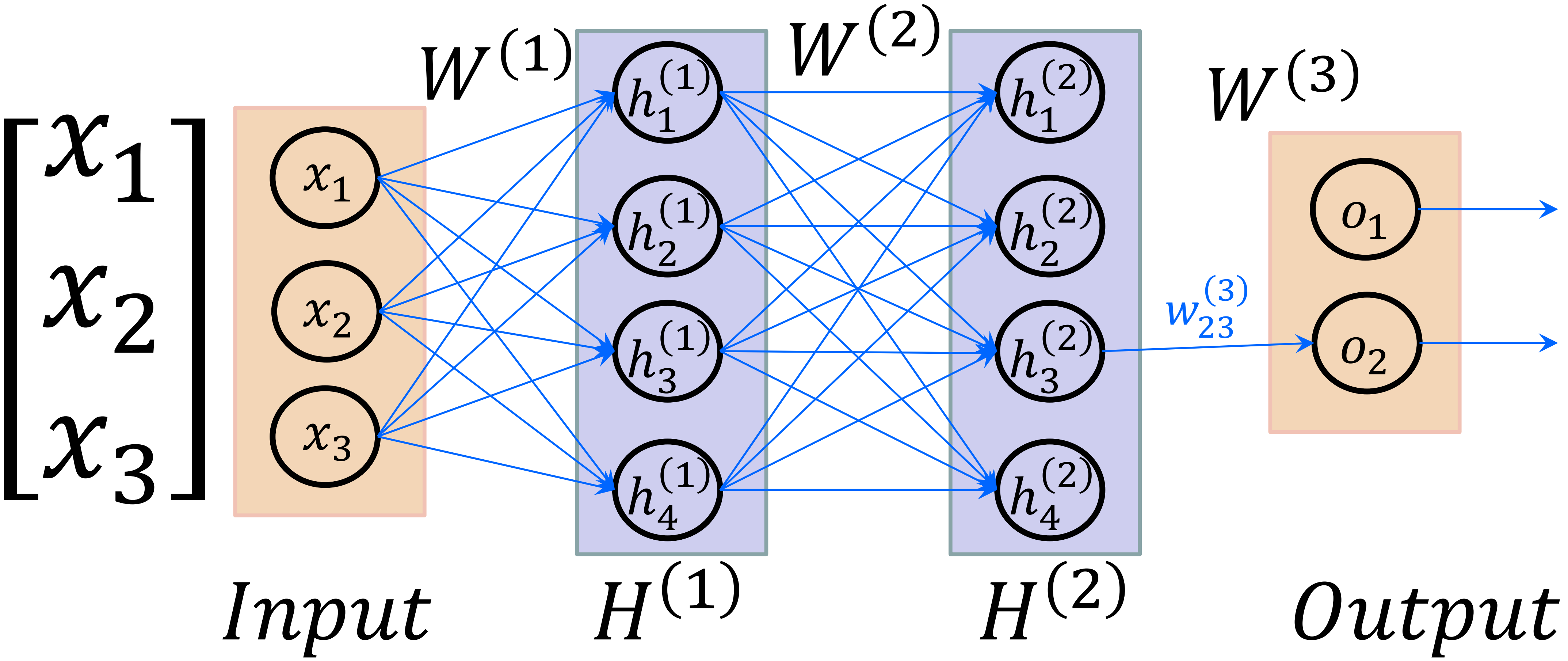
$$W^{(3)} = \begin{bmatrix} w_{11}^{(3)} & w_{12}^{(3)} & w_{13}^{(3)} & w_{14}^{(3)} \\ w_{21}^{(3)} & w_{22}^{(3)} & & \end{bmatrix}$$



Forward propagation



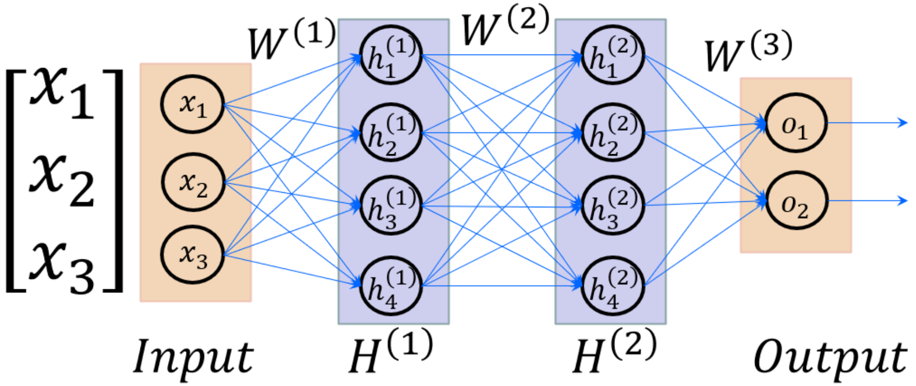
Forward propagation



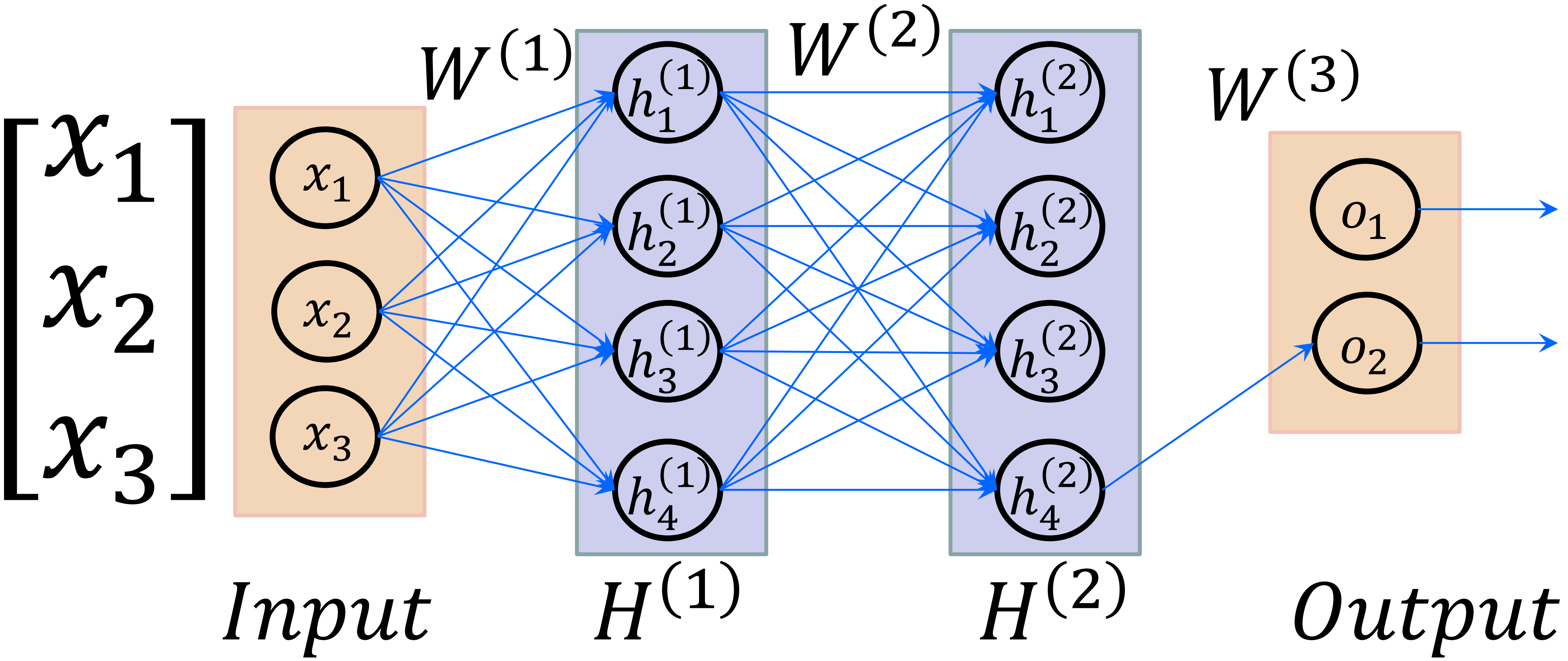
Forward propagation

– Ma trận trọng số $W^{(3)}$

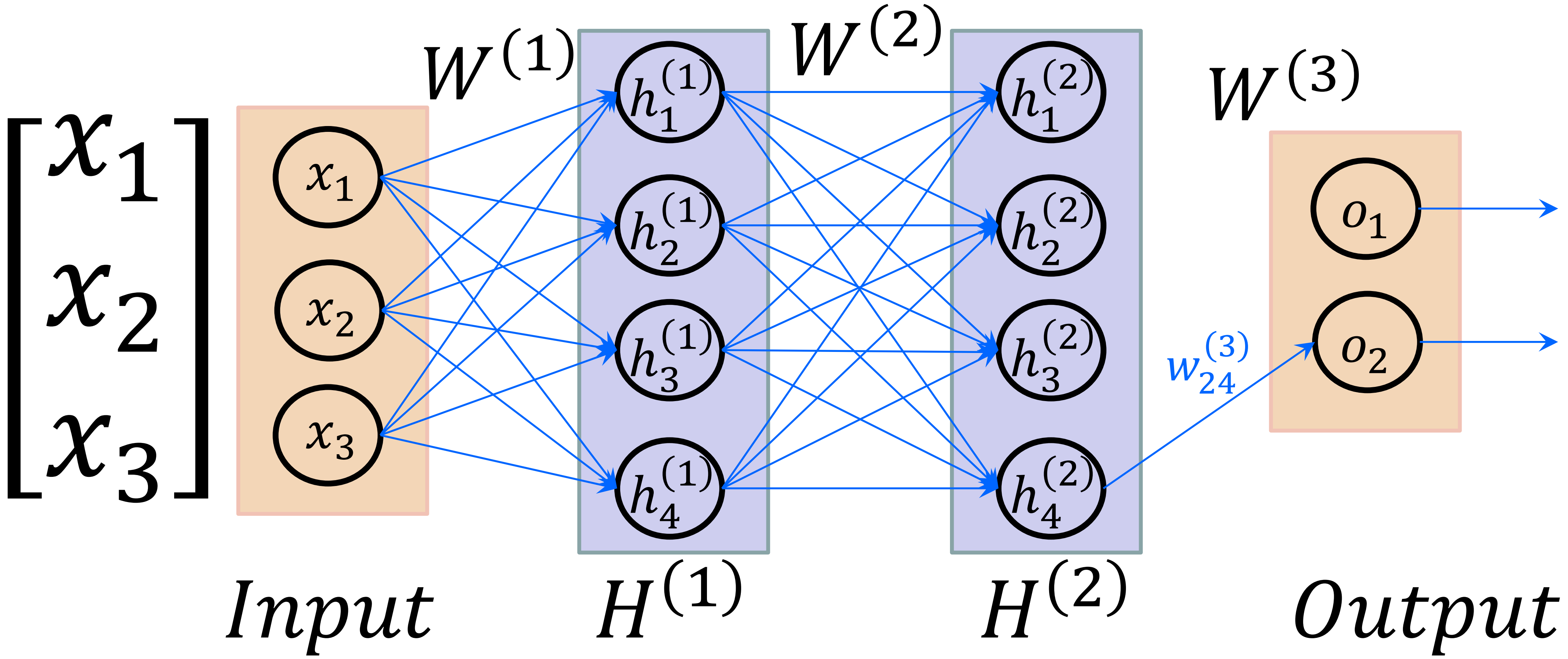
$$W^{(3)} = \begin{bmatrix} w_{11}^{(3)} & w_{12}^{(3)} & w_{13}^{(3)} & w_{14}^{(3)} \\ w_{21}^{(3)} & w_{22}^{(3)} & w_{23}^{(3)} & \end{bmatrix}$$



Forward propagation



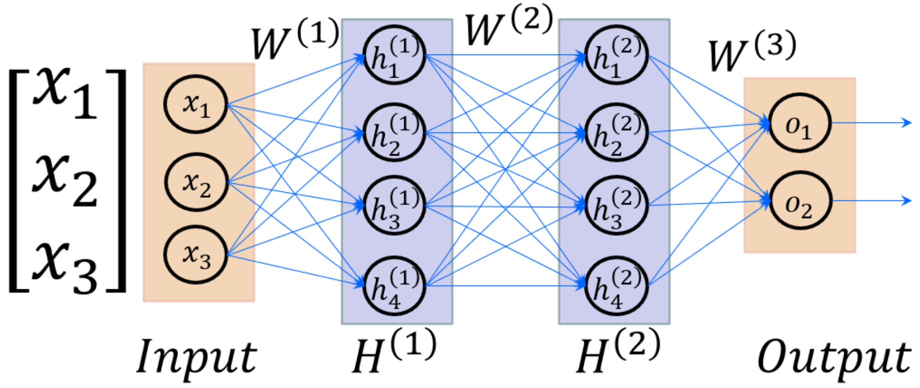
Forward propagation



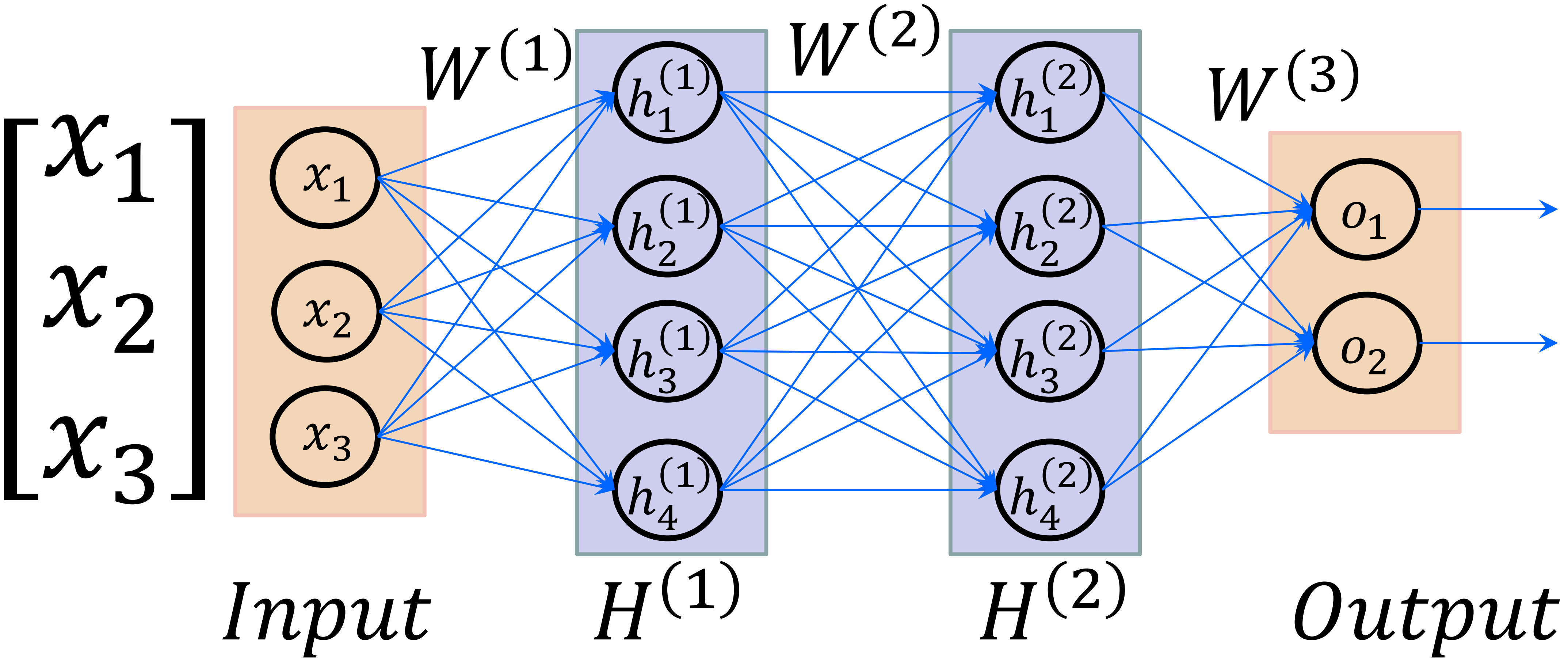
Forward propagation

– Ma trận trọng số $W^{(3)}$

$$W^{(3)} = \begin{bmatrix} w_{11}^{(3)} & w_{12}^{(3)} & w_{13}^{(3)} & w_{14}^{(3)} \\ w_{21}^{(3)} & w_{22}^{(3)} & w_{23}^{(3)} & w_{24}^{(3)} \end{bmatrix}$$



Forward propagation



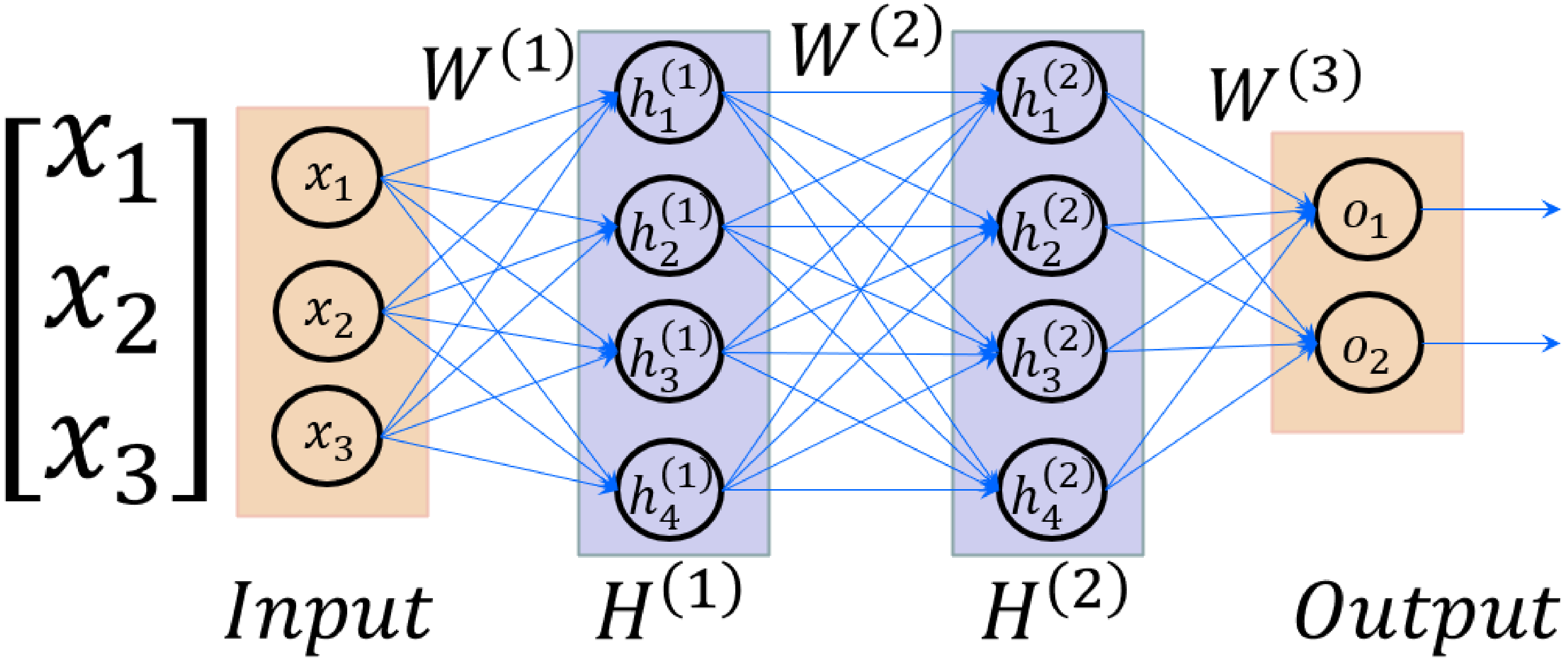
Neural networks

FORWARD PROPAGATION

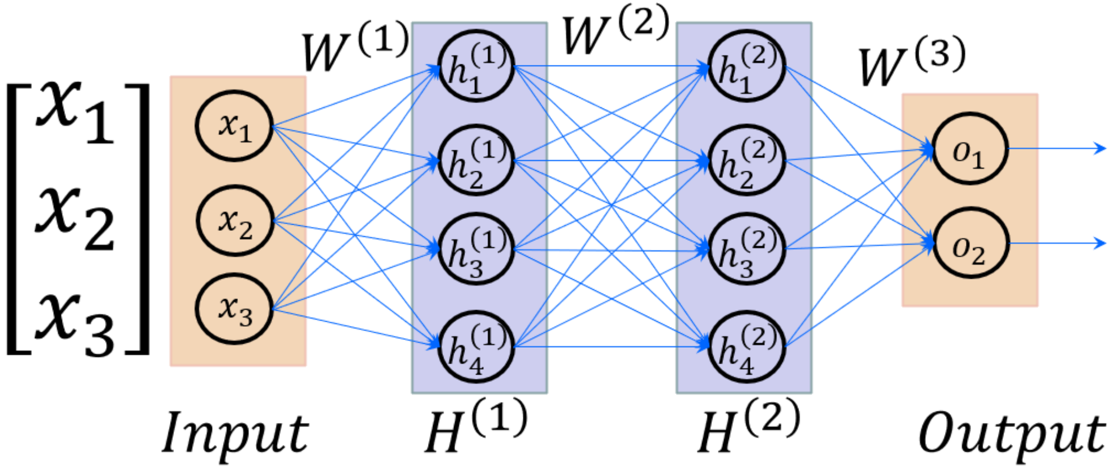
Forward propagation

- Compute output (e.g. probability of a particular class being present in the sample)

Forward propagation

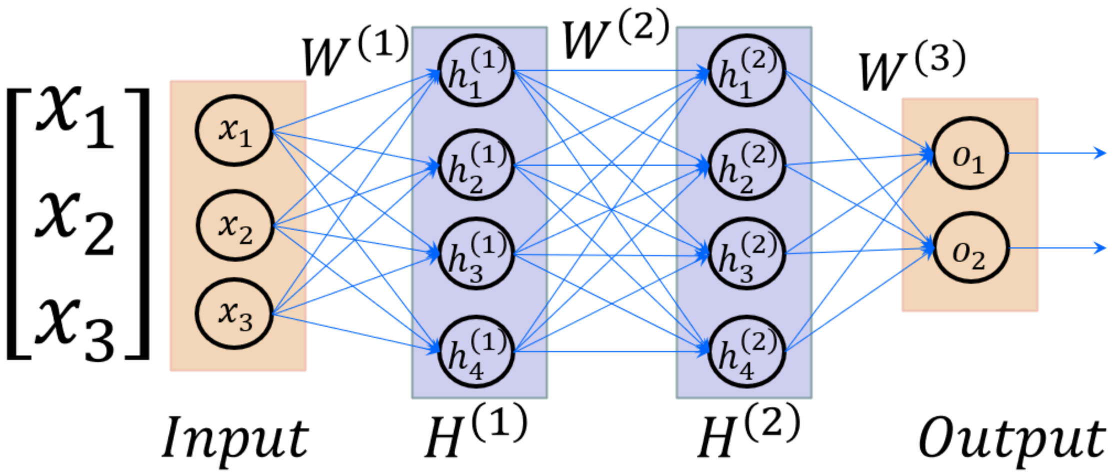


Forward propagation



– Câu hỏi 1: Neural networks này có bao nhiêu lớp?

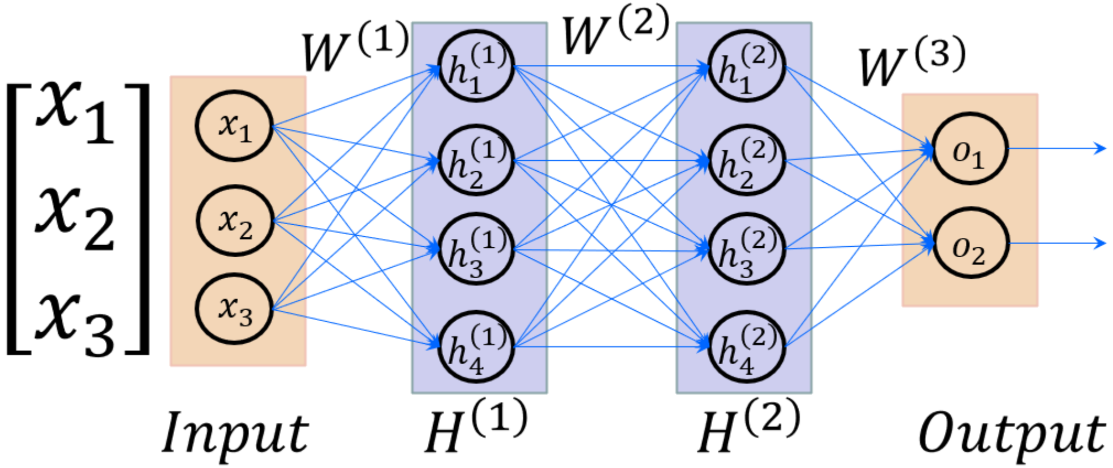
Forward propagation



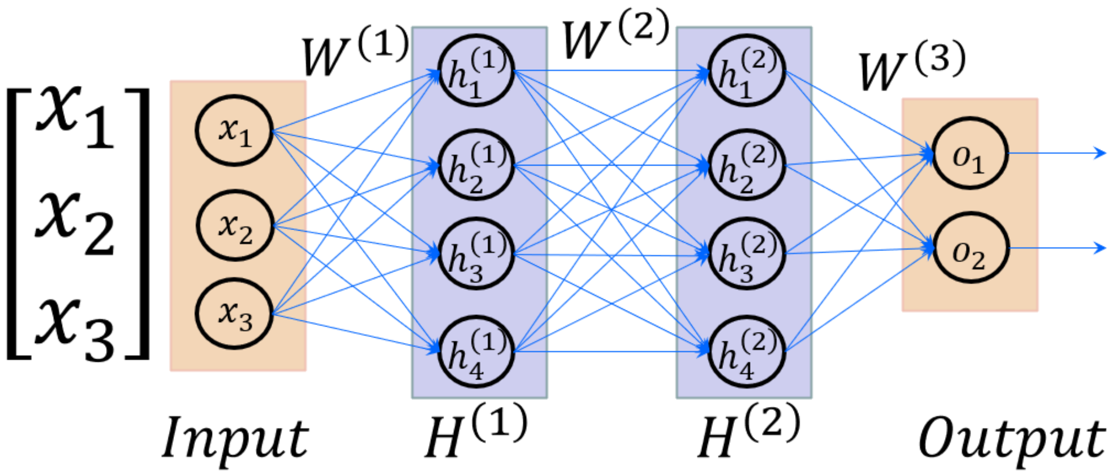
– Câu hỏi 2: Neural networks này lớp input có bao nhiêu phần tử?

Forward propagation

– Câu hỏi 3: Lớp ẩn thứ nhất có bao nhiêu nơ ron?



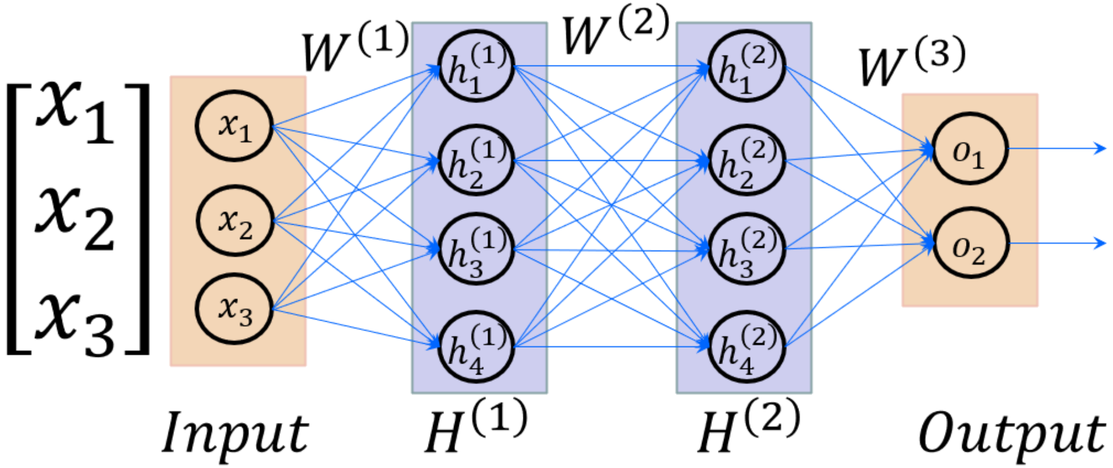
Forward propagation



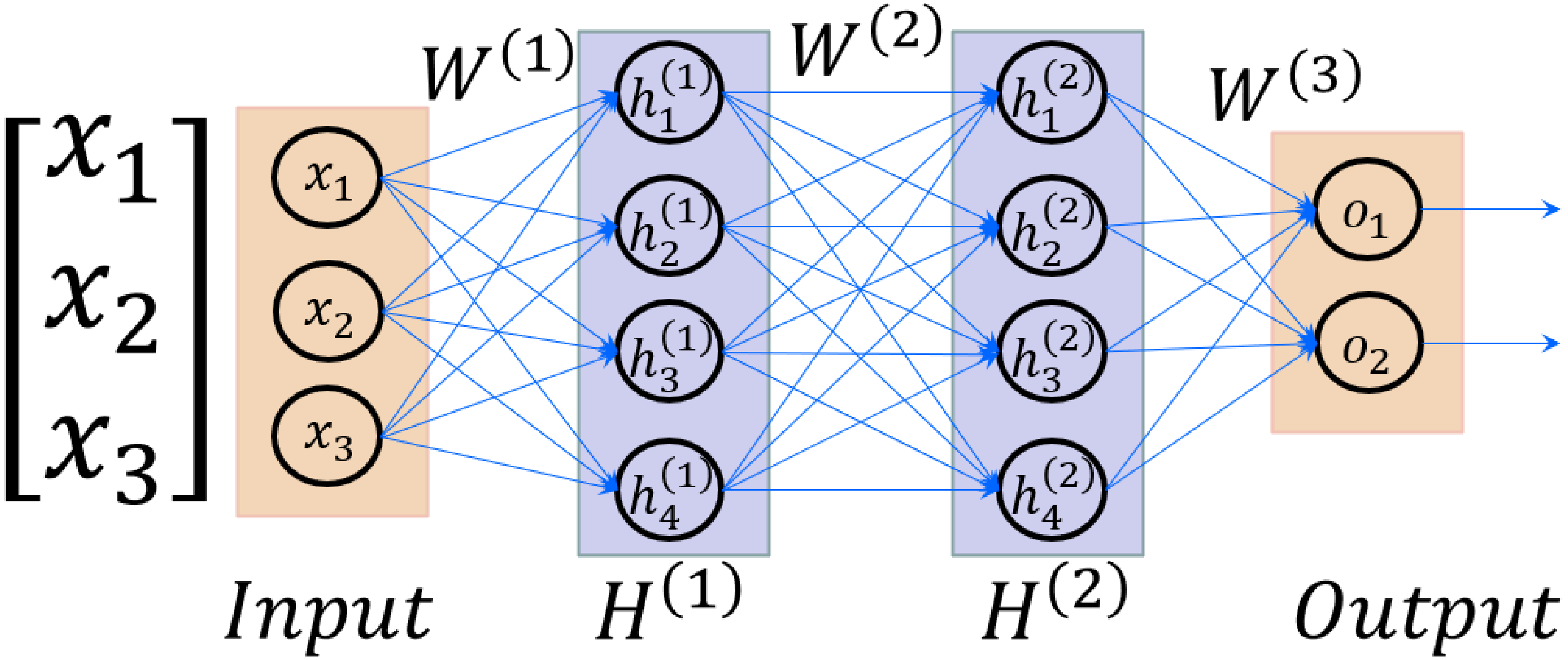
– Câu hỏi 4: Lớp ẩn thứ hai có bao nhiêu nơ ron?

Forward propagation

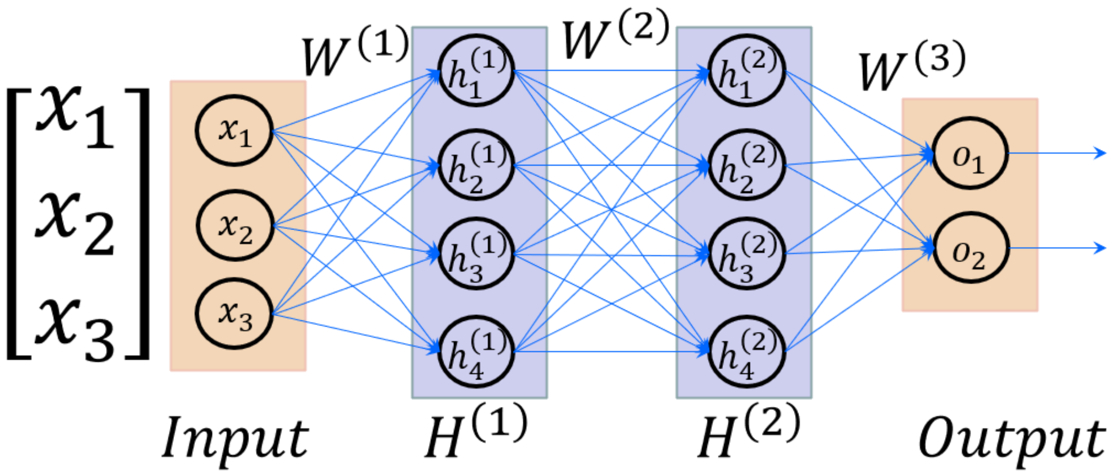
– Câu hỏi 5: Lớp output có bao nhiêu nơ ron?



Forward propagation

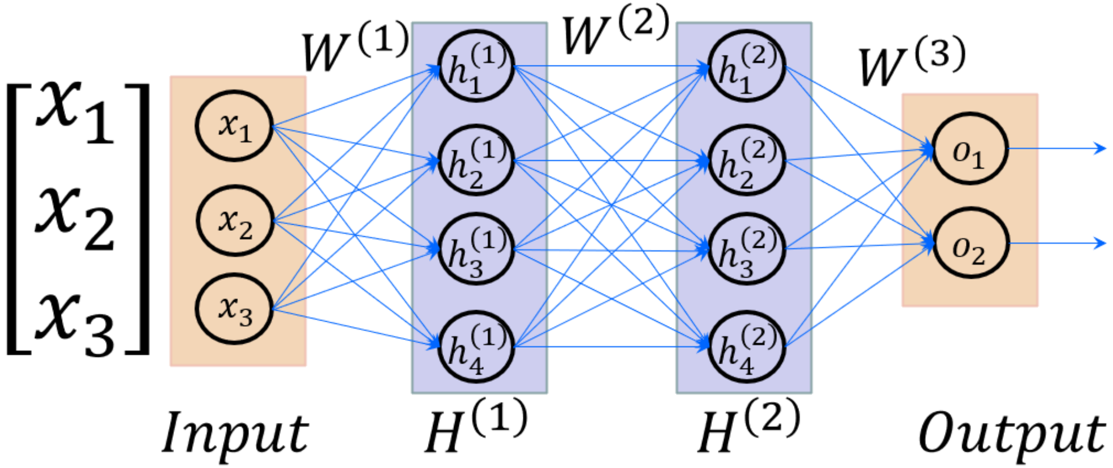


Forward propagation



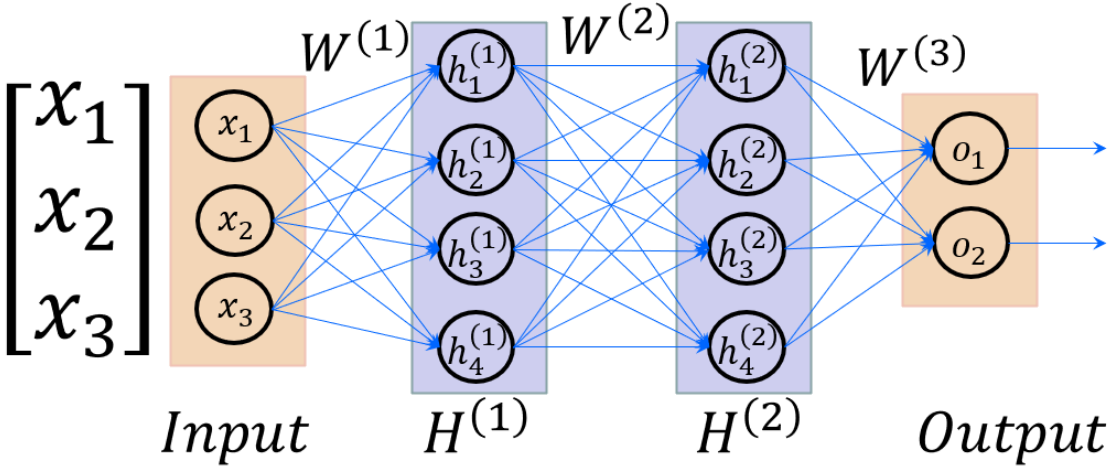
– Câu hỏi 1: Ma trận trọng số $W^{(1)}$ có kích thước bao nhiêu?

Forward propagation



– Câu hỏi 2: Ma trận trọng số $W^{(2)}$ có kích thước bao nhiêu?

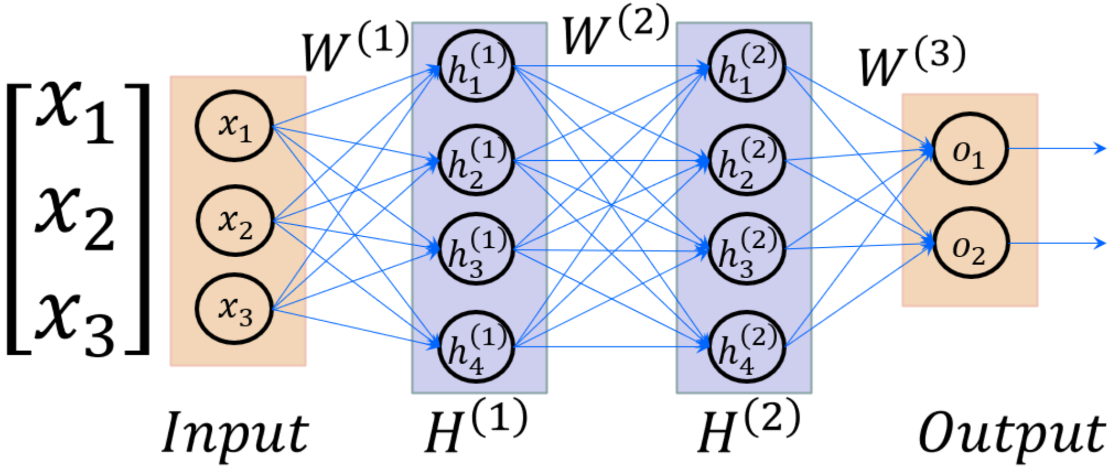
Forward propagation



– Câu hỏi 3: Ma trận trọng số $W^{(3)}$ có kích thước bao nhiêu?

Forward propagation

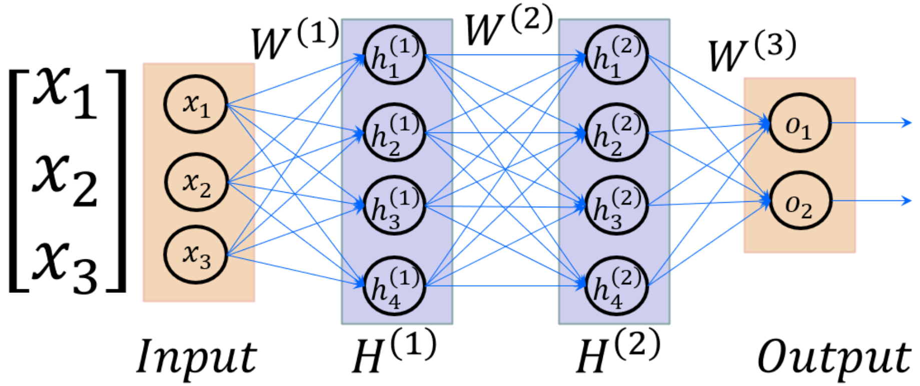
– Câu hỏi 4: Ma trận trọng số $W^{(l)}$ có kích thước bao nhiêu?



Forward propagation

– Input layer – Input của lớp ẩn thứ 1:

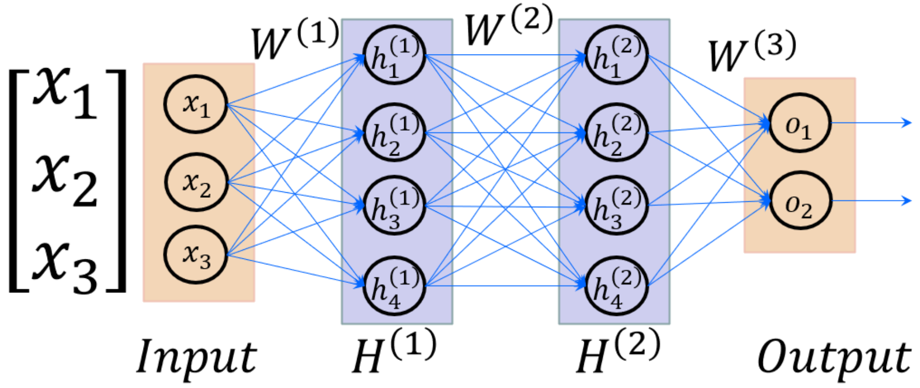
$$input = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$



Forward propagation

– Ma trận trọng số $W^{(1)}$

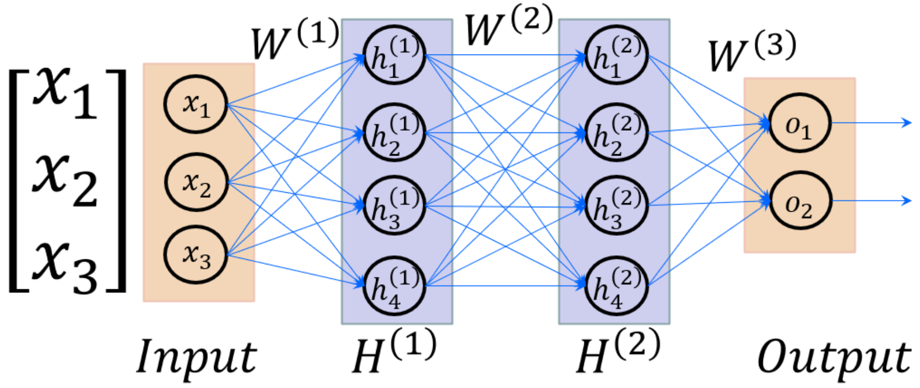
$$W^{(1)} = \begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & w_{23}^{(1)} \\ w_{31}^{(1)} & w_{32}^{(1)} & w_{33}^{(1)} \\ w_{41}^{(1)} & w_{42}^{(1)} & w_{43}^{(1)} \end{bmatrix}$$



Forward propagation

– Hệ số bias của lớp ẩn 1

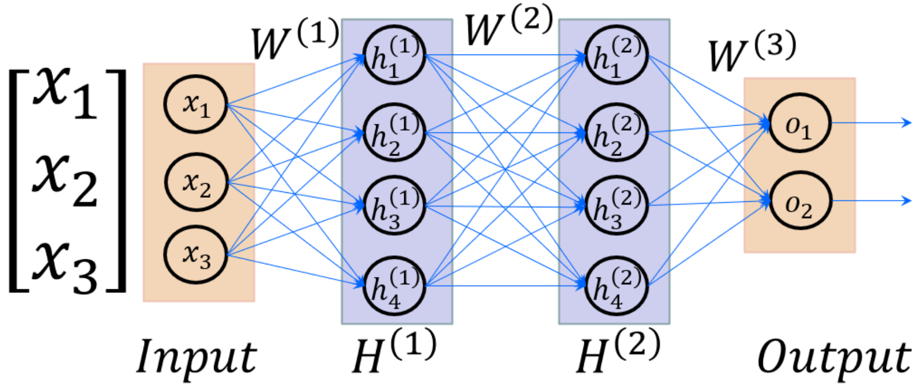
$$b^{(1)} = \begin{bmatrix} b_1^{(1)} \\ b_2^{(1)} \\ b_3^{(1)} \\ b_4^{(1)} \end{bmatrix}$$



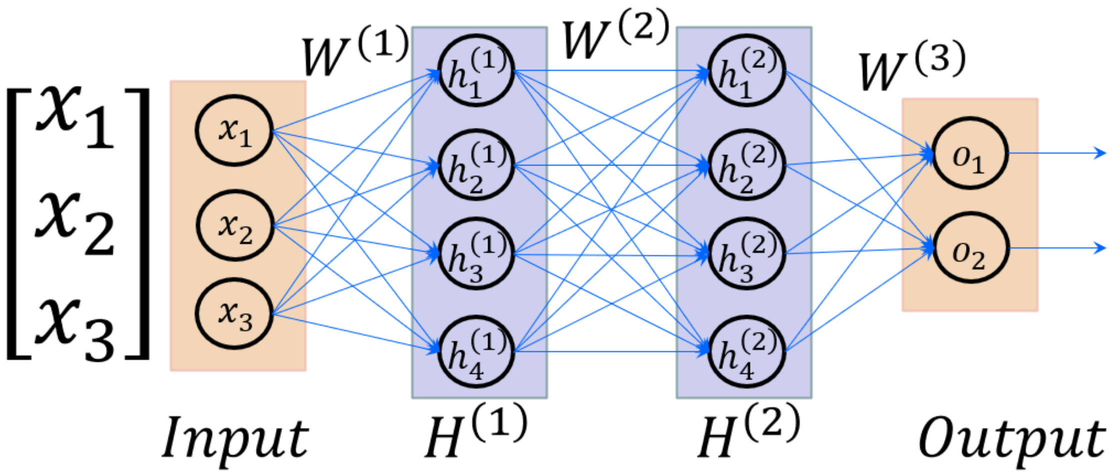
Forward propagation

– Lớp ẩn 1

$$H^{(1)} = \begin{bmatrix} h_1^{(1)} \\ h_2^{(1)} \\ h_3^{(1)} \\ h_4^{(1)} \end{bmatrix}$$



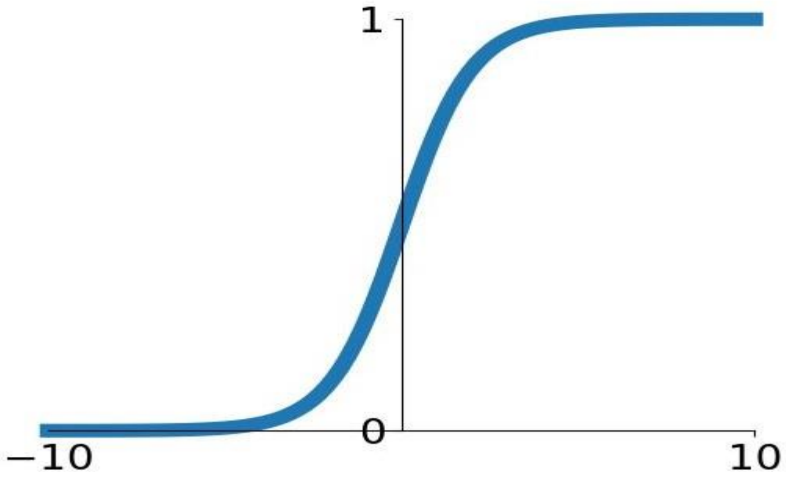
Forward propagation



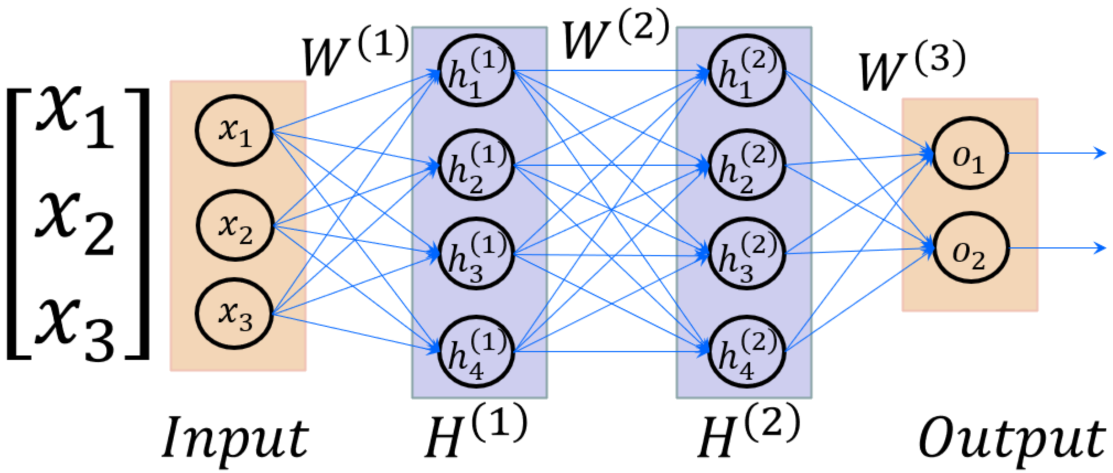
– Activation function: Sigmoid

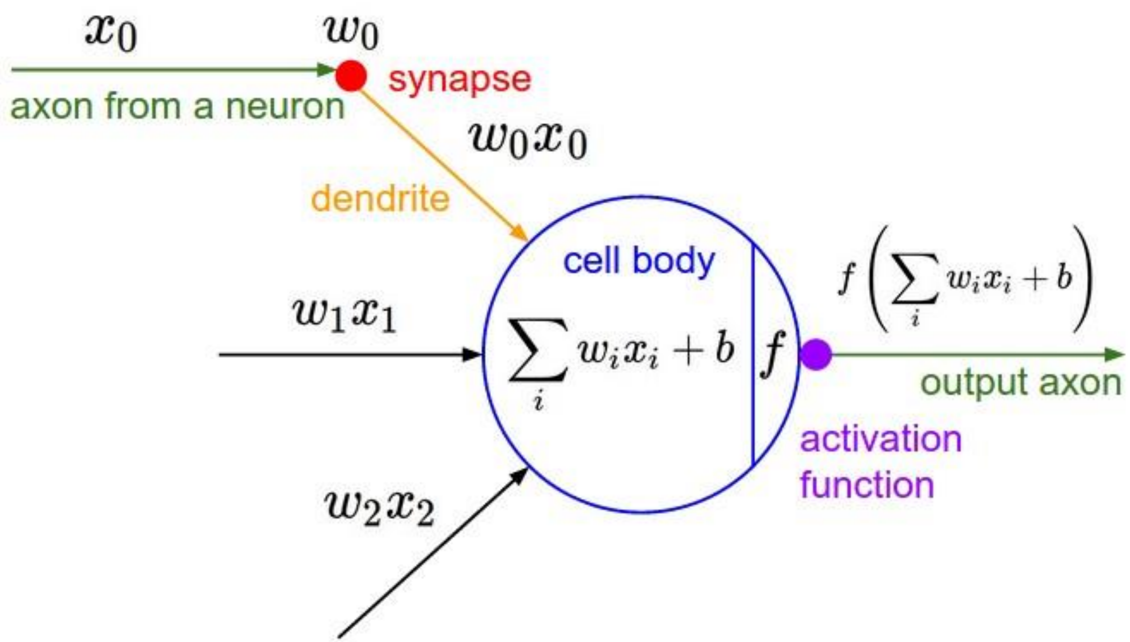
$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

– Đồ thị hàm Sigmoid

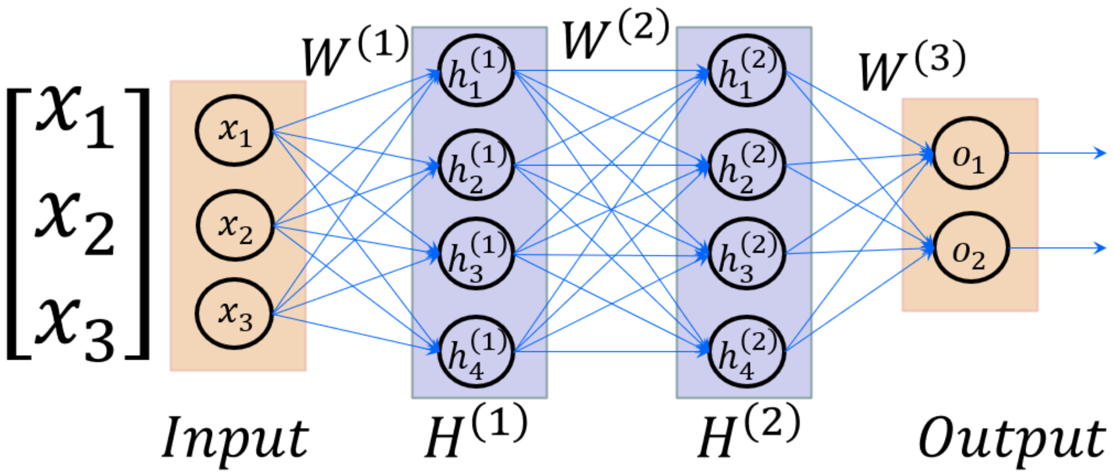


Forward propagation





Forward propagation



– The output layer neurons most commonly have a different activation function
  + Softmax for class scores (classification).
  + Linear functions for real-valued target (regression).

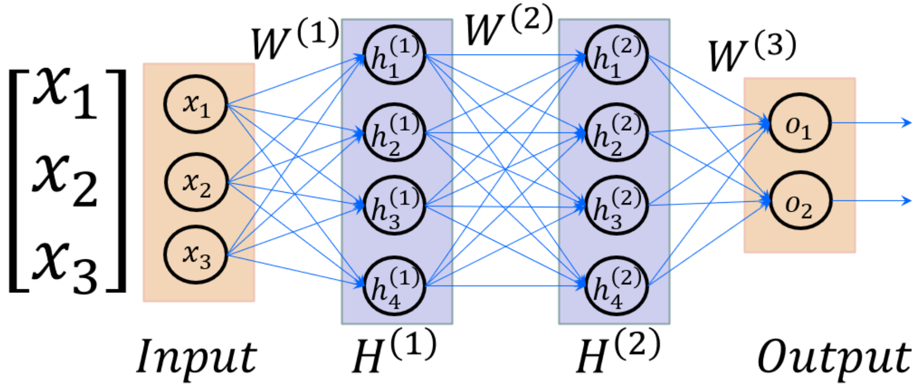
Forward propagation

– Lớp ẩn 1 được tính toán như sau:

$$a^{(1)} = W^{(1)}x + b^{(1)}$$

$$H^{(1)} = \sigma\left(a^{(1)}\right)$$

$$H^{(1)} = \sigma\left(\begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & w_{23}^{(1)} \\ w_{31}^{(1)} & w_{32}^{(1)} & w_{33}^{(1)} \\ w_{41}^{(1)} & w_{42}^{(1)} & w_{43}^{(1)} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} b_1^{(1)} \\ b_2^{(1)} \\ b_3^{(1)} \\ b_4^{(1)} \end{bmatrix}\right) = \begin{bmatrix} h_1^{(1)} \\ h_2^{(1)} \\ h_3^{(1)} \\ h_4^{(1)} \end{bmatrix}$$



$$h_1^{(1)} = \frac{1}{1+e^{-a_1^{(1)}}} = \frac{1}{1+e^{-\left(w_{11}^{(1)}x_1 + w_{12}^{(1)}x_2 + w_{13}^{(1)}x_3 + b_1^{(1)}\right)}}$$

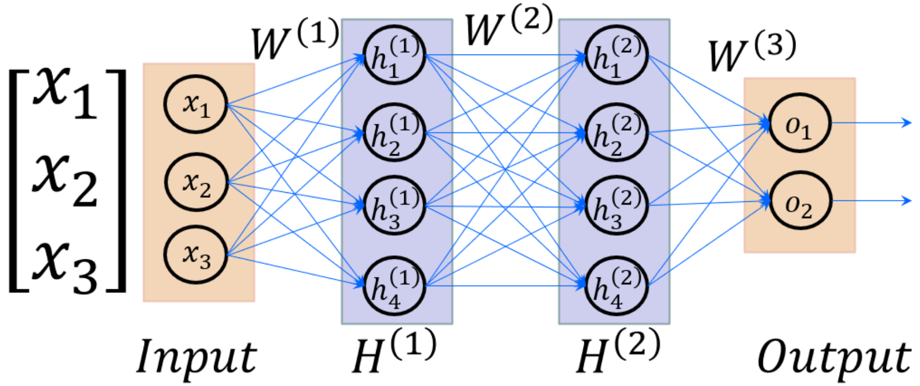
Forward propagation

– Lớp ẩn 1 được tính toán như sau:

$a^{(1)} = W^{(1)}x + b^{(1)}$

$H^{(1)} = \sigma\left(a^{(1)}\right)$

$$H^{(1)} = \sigma\left(\begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & w_{23}^{(1)} \\ w_{31}^{(1)} & w_{32}^{(1)} & w_{33}^{(1)} \\ w_{41}^{(1)} & w_{42}^{(1)} & w_{43}^{(1)} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} b_1^{(1)} \\ b_2^{(1)} \\ b_3^{(1)} \\ b_4^{(1)} \end{bmatrix}\right) = \begin{bmatrix} h_1^{(1)} \\ h_2^{(1)} \\ h_3^{(1)} \\ h_4^{(1)} \end{bmatrix}$$



$$h_2^{(1)} = \frac{1}{1+e^{-a_2^{(1)}}} = \frac{1}{1+e^{-\left(w_{21}^{(1)}x_1 + w_{22}^{(1)}x_2 + w_{23}^{(1)}x_3 + b_2^{(1)}\right)}}$$

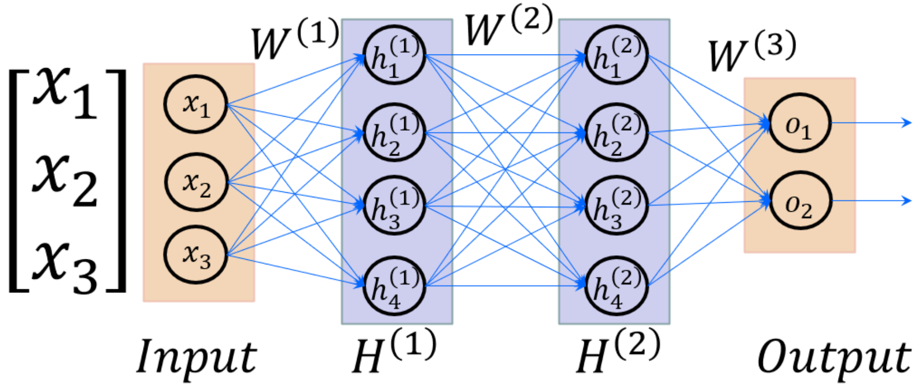
Forward propagation

– Lớp ẩn 1 được tính toán như sau:

$$a^{(1)} = W^{(1)}x + b^{(1)}$$

$$H^{(1)} = \sigma\left(a^{(1)}\right)$$

$$H^{(1)} = \sigma\left(\begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & w_{23}^{(1)} \\ w_{31}^{(1)} & w_{32}^{(1)} & w_{33}^{(1)} \\ w_{41}^{(1)} & w_{42}^{(1)} & w_{43}^{(1)} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} b_1^{(1)} \\ b_2^{(1)} \\ b_3^{(1)} \\ b_4^{(1)} \end{bmatrix}\right) = \begin{bmatrix} h_1^{(1)} \\ h_2^{(1)} \\ h_3^{(1)} \\ h_4^{(1)} \end{bmatrix}$$



$$h_3^{(1)} = \frac{1}{1+e^{-a_3^{(1)}}} = \frac{1}{1+e^{-\left(w_{31}^{(1)}x_1 + w_{32}^{(1)}x_2 + w_{33}^{(1)}x_3 + b_3^{(1)}\right)}}$$

Forward propagation

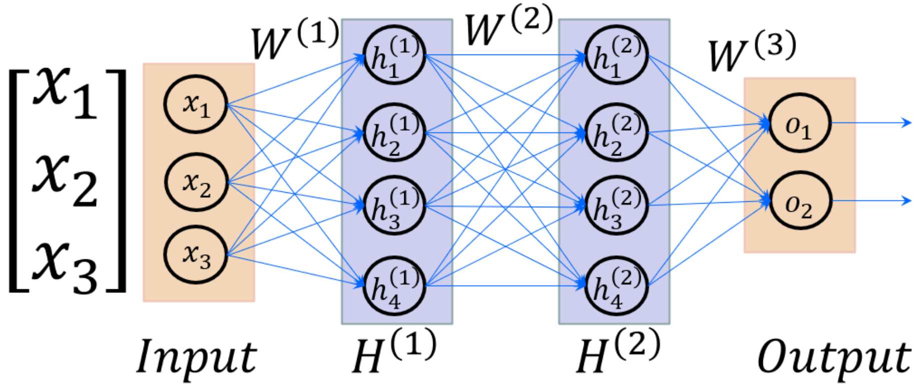
– Lớp ẩn 1 được tính toán như sau:

$$a^{(1)} = W^{(1)}x + b^{(1)}$$

$$H^{(1)} = \sigma\left(a^{(1)}\right)$$

$$H^{(1)} = \sigma\left(\begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & w_{23}^{(1)} \\ w_{31}^{(1)} & w_{32}^{(1)} & w_{33}^{(1)} \\ w_{41}^{(1)} & w_{42}^{(1)} & w_{43}^{(1)} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} b_1^{(1)} \\ b_2^{(1)} \\ b_3^{(1)} \\ b_4^{(1)} \end{bmatrix}\right) = \begin{bmatrix} h_1^{(1)} \\ h_2^{(1)} \\ h_3^{(1)} \\ h_4^{(1)} \end{bmatrix}$$

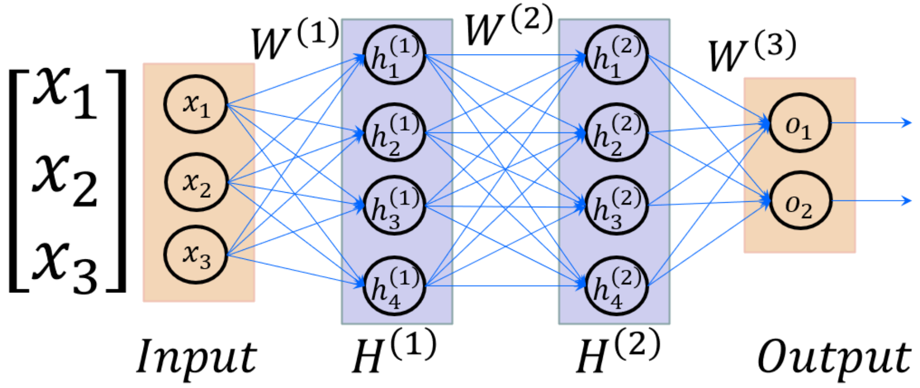
$$h_4^{(1)} = \frac{1}{1+e^{-a_4^{(1)}}} = \frac{1}{1+e^{-\left(w_{41}^{(1)}x_1 + w_{42}^{(1)}x_2 + w_{43}^{(1)}x_3 + b_4^{(1)}\right)}}$$



Forward propagation

– Input của lớp ẩn 2

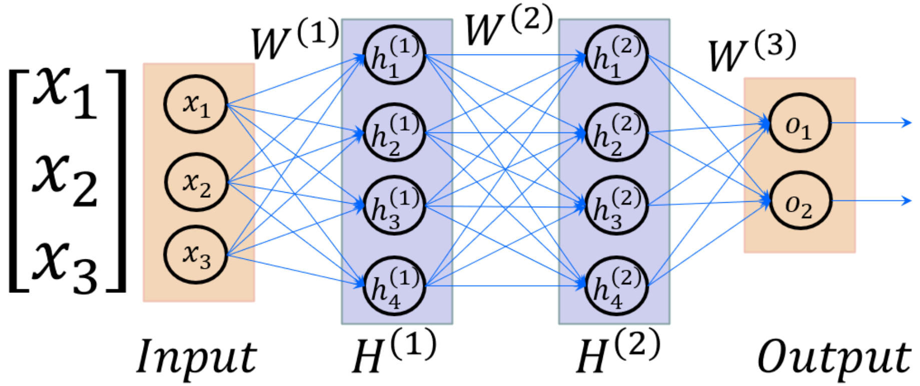
$$H^{(1)} = \begin{bmatrix} h_1^{(1)} \\ h_2^{(1)} \\ h_3^{(1)} \\ h_4^{(1)} \end{bmatrix}$$



Forward propagation

– Ma trận trọng số $W^{(2)}$

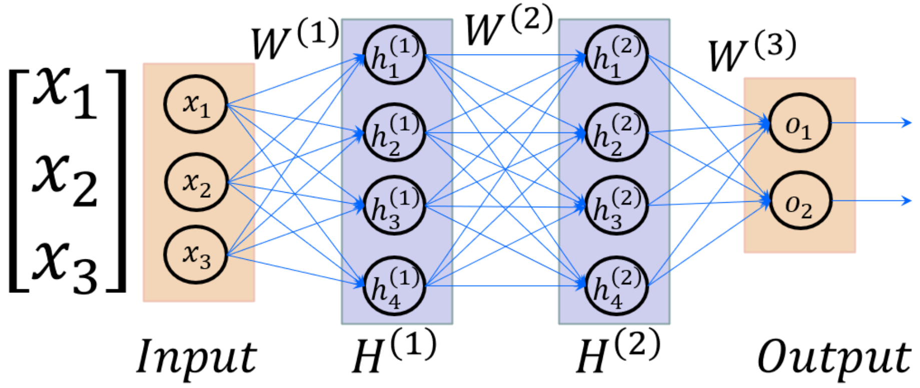
$$W^{(2)} = \begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} & w_{14}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)} & w_{24}^{(2)} \\ w_{31}^{(2)} & w_{32}^{(2)} & w_{33}^{(2)} & w_{34}^{(2)} \\ w_{41}^{(2)} & w_{42}^{(2)} & w_{43}^{(2)} & w_{44}^{(2)} \end{bmatrix}$$



Forward propagation

– Hệ số bias của lớp ẩn 2

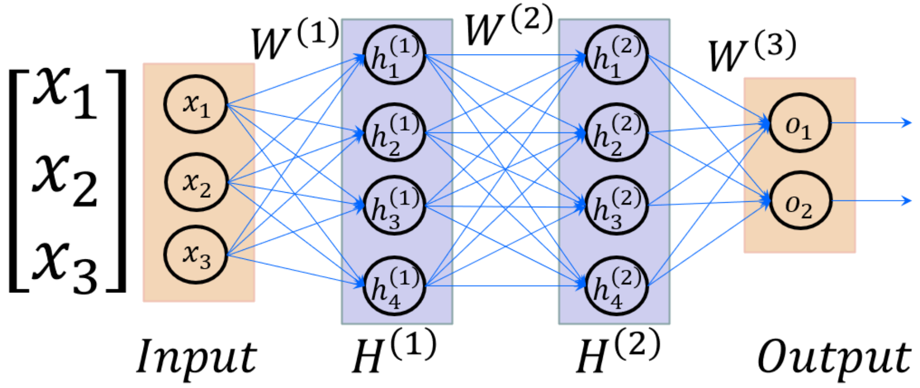
$$b^{(2)} = \begin{bmatrix} b_1^{(2)} \\ b_2^{(2)} \\ b_3^{(2)} \\ b_4^{(2)} \end{bmatrix}$$



Forward propagation

– Lớp ẩn 2

$$H^{(2)} = \begin{bmatrix} h_1^{(2)} \\ h_2^{(2)} \\ h_3^{(2)} \\ h_4^{(2)} \end{bmatrix}$$



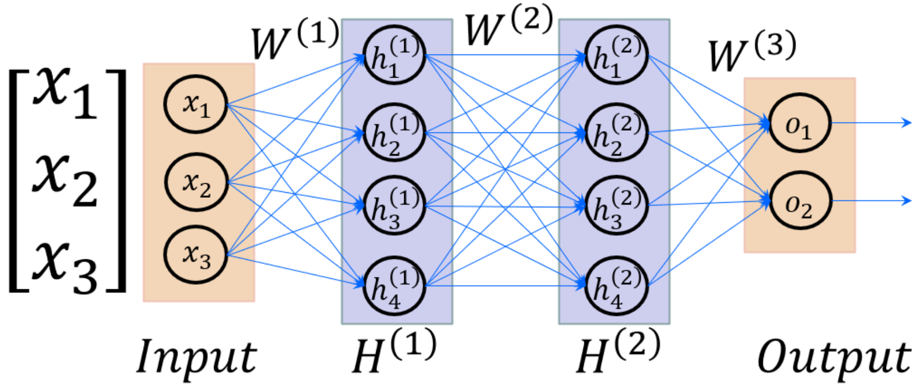
Forward propagation

– Lớp ẩn 2 được tính toán như sau:

$$a^{(2)} = W^{(2)}H^{(1)} + b^{(2)}$$

$$H^{(2)} = \sigma\left(a^{(2)}\right)$$

$$H^{(2)} = \sigma\left(\begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} & w_{14}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)} & w_{24}^{(2)} \\ w_{31}^{(2)} & w_{32}^{(2)} & w_{33}^{(2)} & w_{34}^{(2)} \\ w_{41}^{(2)} & w_{42}^{(2)} & w_{43}^{(2)} & w_{44}^{(2)} \end{bmatrix}\begin{bmatrix} h_1^{(1)} \\ h_2^{(1)} \\ h_3^{(1)} \\ h_4^{(1)} \end{bmatrix} + \begin{bmatrix} b_1^{(2)} \\ b_2^{(2)} \\ b_3^{(2)} \\ b_4^{(2)} \end{bmatrix}\right) = \begin{bmatrix} h_1^{(2)} \\ h_2^{(2)} \\ h_3^{(2)} \\ h_4^{(2)} \end{bmatrix}$$



$$h_1^{(2)} = \frac{1}{1+e^{-a_1^{(2)}}} = \frac{1}{1+e^{-\left(w_{11}^{(2)}h_1^{(1)}+w_{12}^{(2)}h_2^{(1)}+w_{13}^{(2)}h_3^{(1)}+w_{14}^{(2)}h_4^{(1)}+b_1^{(2)}\right)}}$$

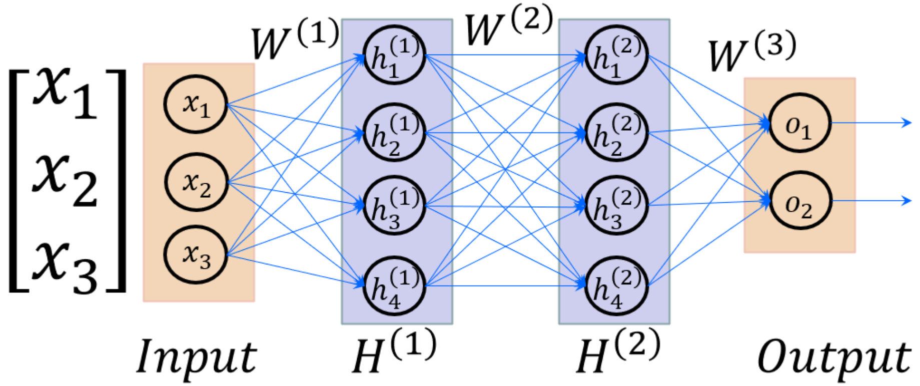
Forward propagation

– Lớp ẩn 2 được tính toán như sau:

$$a^{(2)} = W^{(2)}H^{(1)} + b^{(2)}$$

$$H^{(2)} = \sigma\left(a^{(2)}\right)$$

$$H^{(2)} = \sigma\left(\begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} & w_{14}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)} & w_{24}^{(2)} \\ w_{31}^{(2)} & w_{32}^{(2)} & w_{33}^{(2)} & w_{34}^{(2)} \\ w_{41}^{(2)} & w_{42}^{(2)} & w_{43}^{(2)} & w_{44}^{(2)} \end{bmatrix}\begin{bmatrix} h_1^{(1)} \\ h_2^{(1)} \\ h_3^{(1)} \\ h_4^{(1)} \end{bmatrix} + \begin{bmatrix} b_1^{(2)} \\ b_2^{(2)} \\ b_3^{(2)} \\ b_4^{(2)} \end{bmatrix}\right) = \begin{bmatrix} h_1^{(2)} \\ h_2^{(2)} \\ h_3^{(2)} \\ h_4^{(2)} \end{bmatrix}$$



$$h_2^{(2)} = \frac{1}{1+e^{-a_2^{(2)}}} = \frac{1}{1+e^{-\left(w_{21}^{(2)}h_1^{(1)}+w_{22}^{(2)}h_2^{(1)}+w_{23}^{(2)}h_3^{(1)}+w_{24}^{(2)}h_4^{(1)}+b_2^{(2)}\right)}}$$

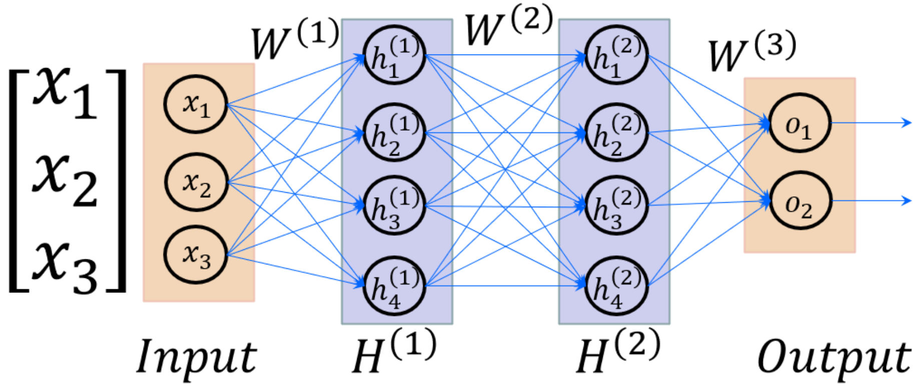
Forward propagation

– Lớp ẩn 2 được tính toán như sau:

$$a^{(2)} = W^{(2)}H^{(1)} + b^{(2)}$$

$$H^{(2)} = \sigma\left(a^{(2)}\right)$$

$$H^{(2)} = \sigma\left(\begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} & w_{14}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)} & w_{24}^{(2)} \\ w_{31}^{(2)} & w_{32}^{(2)} & w_{33}^{(2)} & w_{34}^{(2)} \\ w_{41}^{(2)} & w_{42}^{(2)} & w_{43}^{(2)} & w_{44}^{(2)} \end{bmatrix}\begin{bmatrix} h_1^{(1)} \\ h_2^{(1)} \\ h_3^{(1)} \\ h_4^{(1)} \end{bmatrix} + \begin{bmatrix} b_1^{(2)} \\ b_2^{(2)} \\ b_3^{(2)} \\ b_4^{(2)} \end{bmatrix}\right) = \begin{bmatrix} h_1^{(2)} \\ h_2^{(2)} \\ h_3^{(2)} \\ h_4^{(2)} \end{bmatrix}$$



$$h_3^{(2)} = \frac{1}{1+e^{-a_3^{(2)}}} = \frac{1}{1+e^{-\left(w_{31}^{(2)}h_1^{(1)}+w_{32}^{(2)}h_2^{(1)}+w_{33}^{(2)}h_3^{(1)}+w_{34}^{(2)}h_4^{(1)}+b_3^{(2)}\right)}}$$

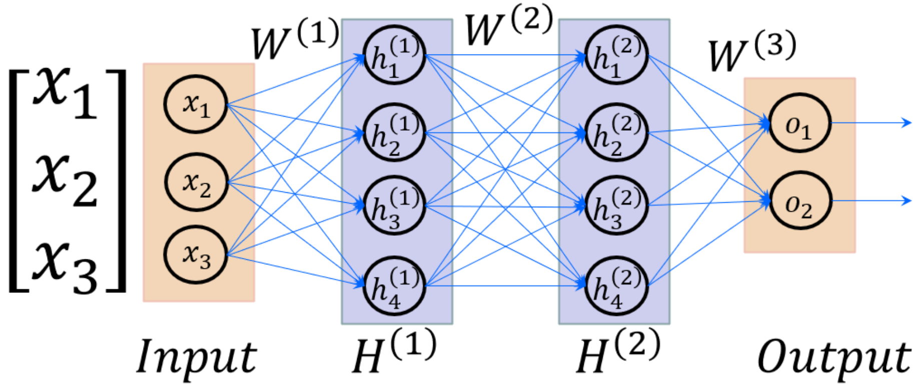
Forward propagation

– Lớp ẩn 2 được tính toán như sau:

$$a^{(2)} = W^{(2)}H^{(1)} + b^{(2)}$$

$$H^{(2)} = \sigma\left(a^{(2)}\right)$$

$$H^{(2)} = \sigma\left(\begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} & w_{14}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)} & w_{24}^{(2)} \\ w_{31}^{(2)} & w_{32}^{(2)} & w_{33}^{(2)} & w_{34}^{(2)} \\ w_{41}^{(2)} & w_{42}^{(2)} & w_{43}^{(2)} & w_{44}^{(2)} \end{bmatrix}\begin{bmatrix} h_1^{(1)} \\ h_2^{(1)} \\ h_3^{(1)} \\ h_4^{(1)} \end{bmatrix} + \begin{bmatrix} b_1^{(2)} \\ b_2^{(2)} \\ b_3^{(2)} \\ b_4^{(2)} \end{bmatrix}\right) = \begin{bmatrix} h_1^{(2)} \\ h_2^{(2)} \\ h_3^{(2)} \\ h_4^{(2)} \end{bmatrix}$$

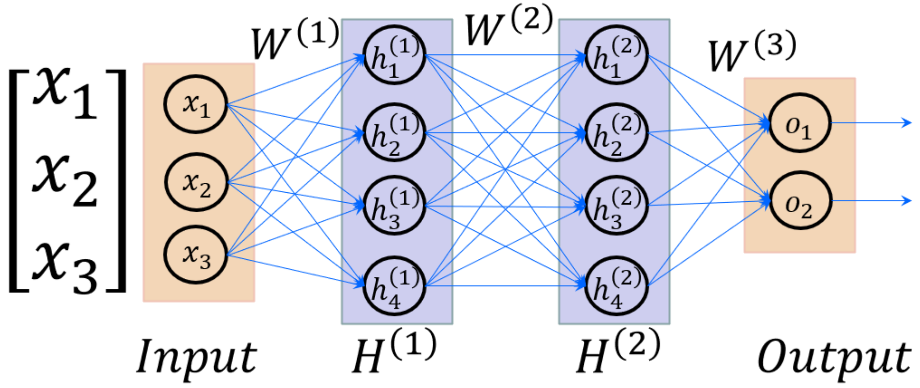


$$h_4^{(2)} = \frac{1}{1+e^{-a_4^{(2)}}} = \frac{1}{1+e^{-\left(w_{41}^{(2)}h_1^{(1)}+w_{42}^{(2)}h_2^{(1)}+w_{43}^{(2)}h_3^{(1)}+w_{44}^{(2)}h_4^{(1)}+b_4^{(2)}\right)}}$$

Forward propagation

– Input của lớp output

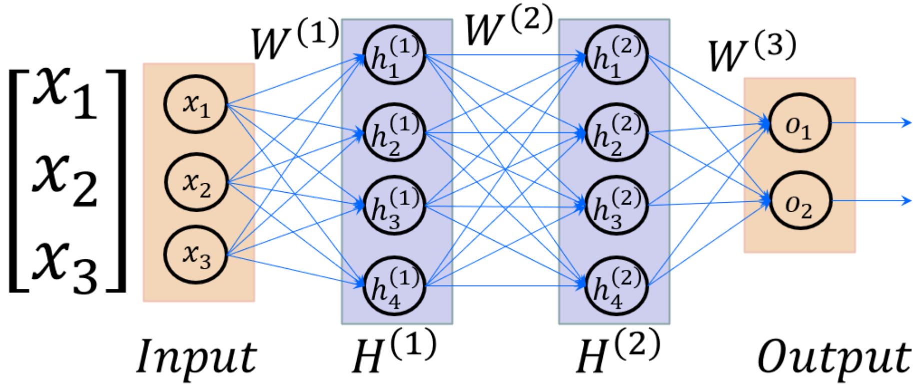
$$H^{(2)} = \begin{bmatrix} h_1^{(2)} \\ h_2^{(2)} \\ h_3^{(2)} \\ h_4^{(2)} \end{bmatrix}$$



Forward propagation

– Ma trận trọng số $W^{(3)}$

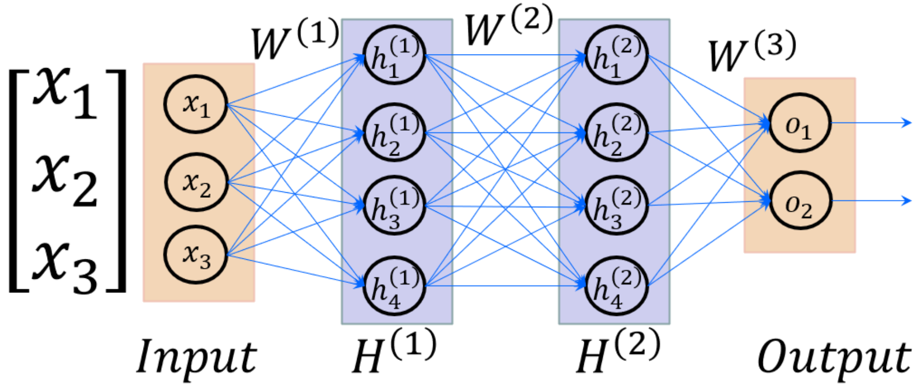
$$W^{(3)} = \begin{bmatrix} w_{11}^{(3)} & w_{12}^{(3)} & w_{13}^{(3)} & w_{14}^{(3)} \\ w_{21}^{(3)} & w_{22}^{(3)} & w_{23}^{(3)} & w_{24}^{(3)} \end{bmatrix}$$



Forward propagation

– Hệ số bias của lớp output

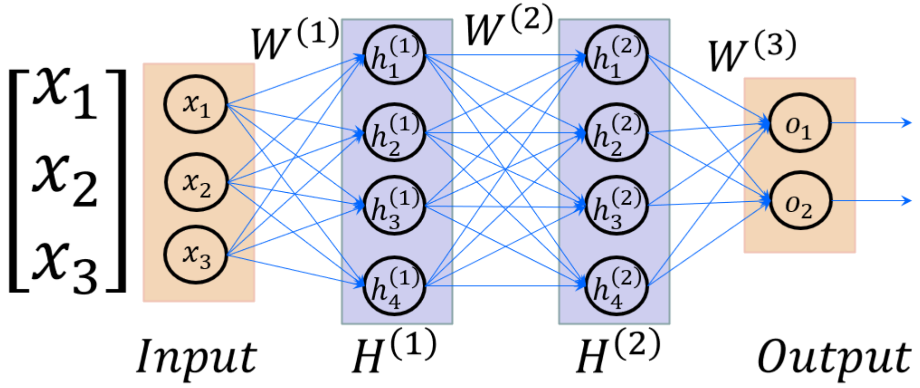
$$b^{(3)} = \begin{bmatrix} b_1^{(3)} \\ b_2^{(3)} \end{bmatrix}$$



Forward propagation

– Lớp output

$$H^{(3)} = \begin{bmatrix} h_1^{(3)} \\ h_1^{(3)} \end{bmatrix} = \begin{bmatrix} o_1 \\ o_2 \end{bmatrix} = [output]$$



Forward propagation

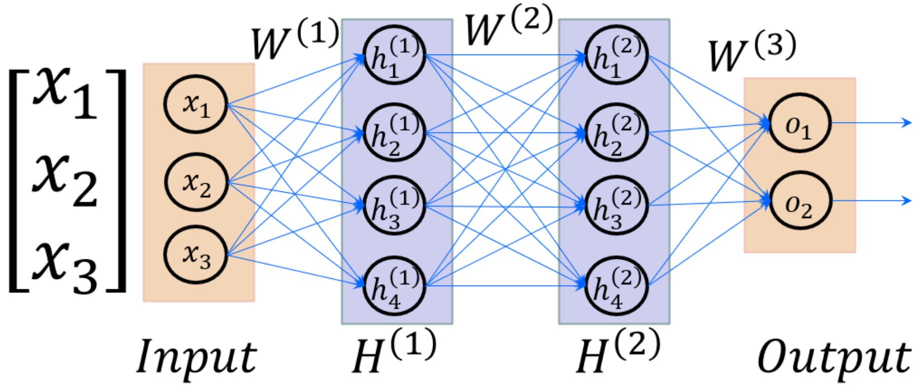
– Lớp output được tính toán như sau:

$$a^{(3)} = W^{(3)} H^{(2)} + b^{(3)}$$

$$H^{(3)} = \sigma\left(a^{(3)}\right)$$

$$H^{(3)} = \sigma\left(\begin{bmatrix} w_{11}^{(3)} & w_{12}^{(3)} & w_{13}^{(3)} & w_{14}^{(3)} \\ w_{21}^{(3)} & w_{22}^{(3)} & w_{23}^{(3)} & w_{24}^{(3)} \end{bmatrix} \begin{bmatrix} h_1^{(2)} \\ h_2^{(2)} \\ h_3^{(2)} \\ h_4^{(2)} \end{bmatrix} + \begin{bmatrix} b_1^{(3)} \\ b_2^{(3)} \end{bmatrix}\right)$$

$$h_1^{(3)} = \frac{1}{1+e^{-a_1^{(3)}}} = \frac{1}{1+e^{-\left(w_{11}^{(3)} h_1^{(2)} + w_{12}^{(3)} h_2^{(2)} + w_{13}^{(3)} h_3^{(2)} + w_{14}^{(3)} h_4^{(2)} + b_1^{(3)}\right)}}$$



Forward propagation

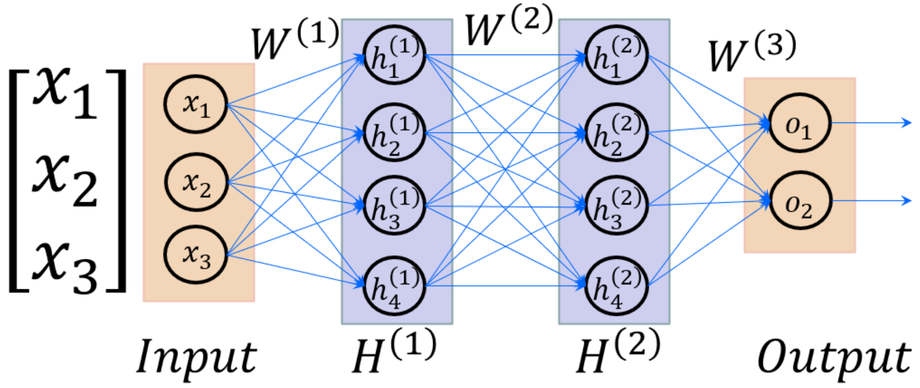
– Lớp output được tính toán như sau:

$$a^{(3)} = W^{(3)}H^{(2)} + b^{(3)}$$

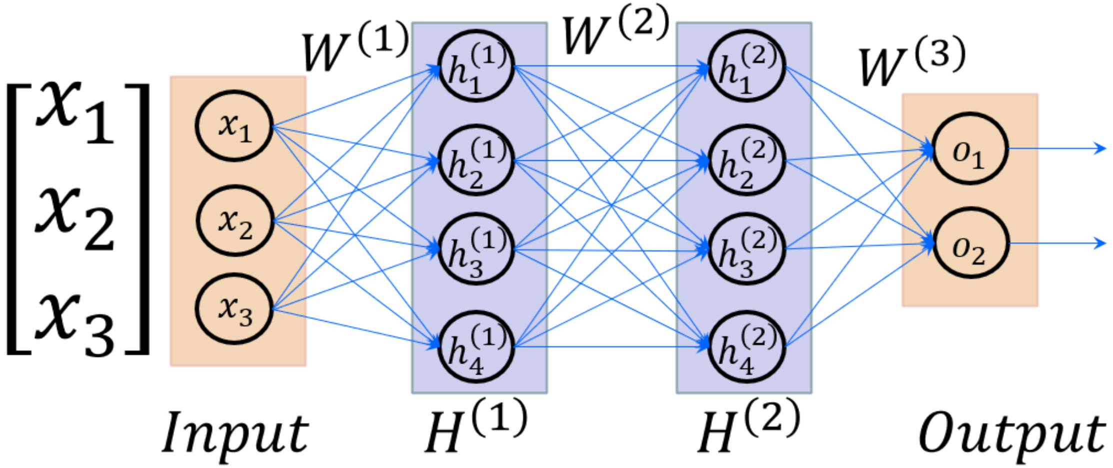
$$H^{(3)} = \sigma\left(a^{(3)}\right)$$

$$H^{(3)} = \sigma\left(\begin{bmatrix} w_{11}^{(3)} & w_{12}^{(3)} & w_{13}^{(3)} & w_{14}^{(3)} \\ w_{21}^{(3)} & w_{22}^{(3)} & w_{23}^{(3)} & w_{24}^{(3)} \end{bmatrix}\begin{bmatrix} h_1^{(2)} \\ h_2^{(2)} \\ h_3^{(2)} \\ h_4^{(2)} \end{bmatrix} + \begin{bmatrix} b_1^{(3)} \\ b_2^{(3)} \end{bmatrix}\right)$$

$$h_2^{(3)} = \frac{1}{1+e^{-a_2^{(3)}}} = \frac{1}{1+e^{-\left(w_{21}^{(3)}h_1^{(2)}+w_{22}^{(3)}h_2^{(2)}+w_{23}^{(3)}h_3^{(2)}+w_{24}^{(3)}h_4^{(2)}+b_2^{(3)}\right)}}$$



Forward propagation



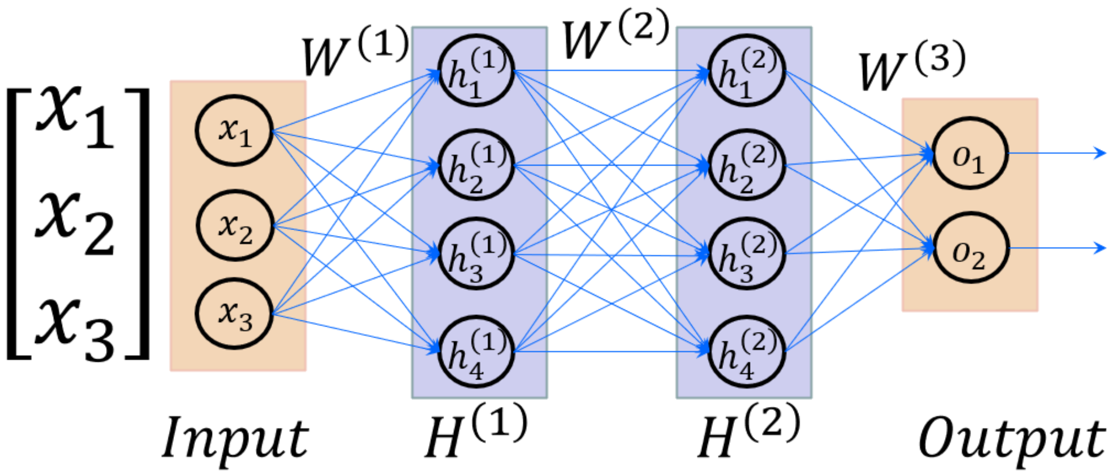
– Tính sai số

+ $E_{O_1} = \frac{1}{2}(target - output)^2$

+ $E_{O_2} = \frac{1}{2}(target - output)^2$

+ $E = E_{O_1} + E_{O_2}$

Forward propagation



– Các tham số phải học của mạng Neural networks.
  + $W^{(1)}, W^{(2)}, W^{(3)}$
  + $b^{(1)}, b^{(2)}, b^{(3)}$

Neural networks

FORWARD PROPAGATION – RUN

UIT VNUHCM University of Information Technology

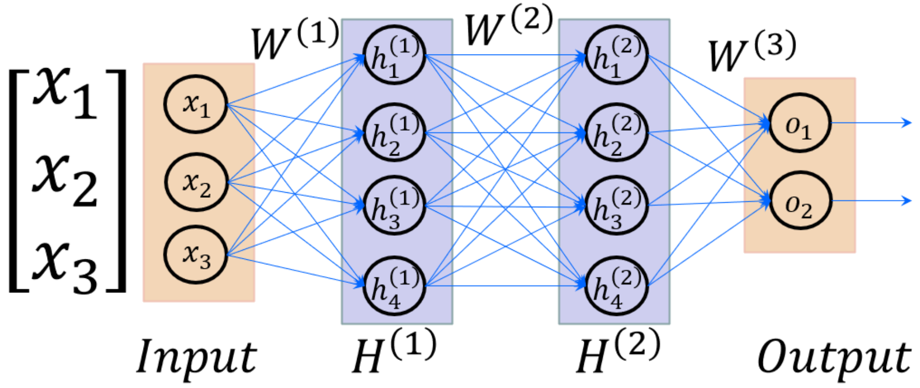
Forward propagation

– Datatrain $(x_1, x_2, x_3, o_1, o_2)$:

+ $(1, 2, 3, 4, 5)$

+ $(3, 2, 1, 5, 6)$

+ $(1, 5, 7, 8, 6)$

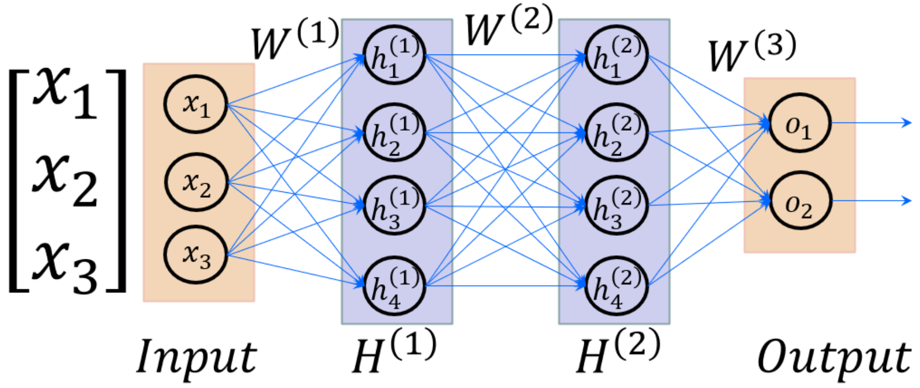


+ $(1, 2, 3, 4, 5)$

Forward propagation

– Input layer – Input của lớp ẩn thứ 1:

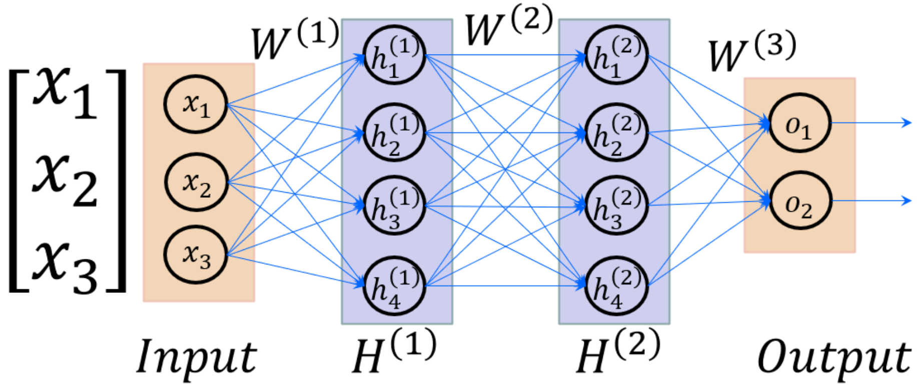
$$input = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$



Forward propagation

– Ma trận trọng số $W^{(1)}$

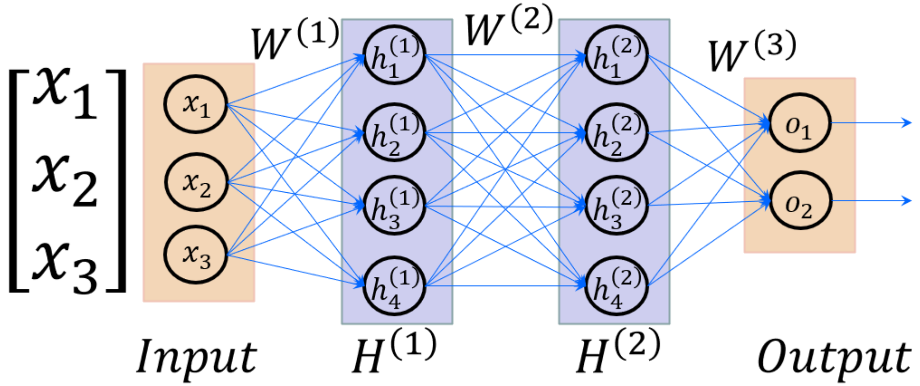
$$W^{(1)} = \begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & w_{23}^{(1)} \\ w_{31}^{(1)} & w_{32}^{(1)} & w_{33}^{(1)} \\ w_{41}^{(1)} & w_{42}^{(1)} & w_{43}^{(1)} \end{bmatrix}$$



Forward propagation

– Ma trận trọng số $W^{(1)}$

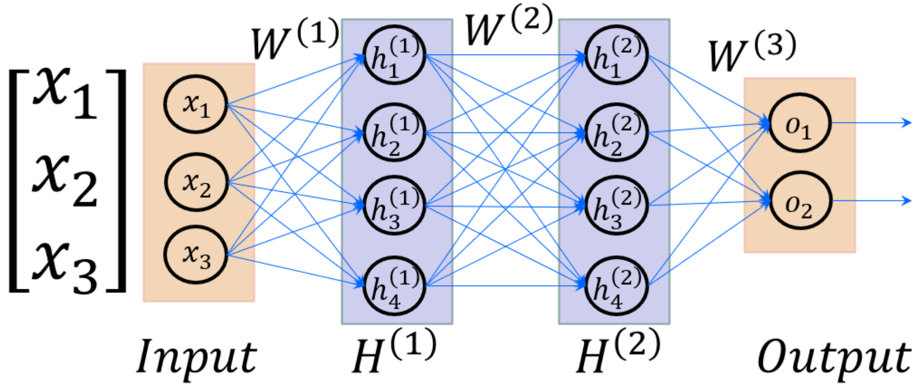
$$W^{(1)} = \begin{bmatrix} 0. & w_{12}^{(1)} & w_{13}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & w_{23}^{(1)} \\ w_{31}^{(1)} & w_{32}^{(1)} & w_{33}^{(1)} \\ w_{41}^{(1)} & w_{42}^{(1)} & w_{43}^{(1)} \end{bmatrix}$$



Forward propagation

– Hệ số bias của lớp ẩn 1

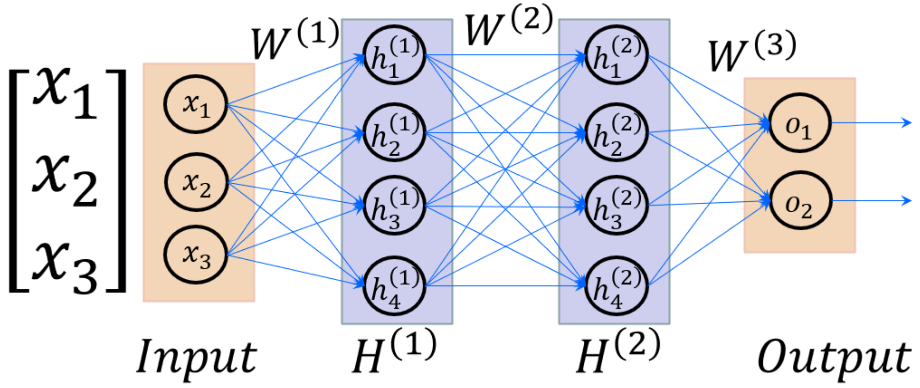
$$b^{(1)} = \begin{bmatrix} b_1^{(1)} \\ b_2^{(1)} \\ b_3^{(1)} \\ b_4^{(1)} \end{bmatrix} = \begin{bmatrix} 0. \\ b_2^{(1)} \\ b_3^{(1)} \\ b_4^{(1)} \end{bmatrix}$$



Forward propagation

– Lớp ẩn 1

$$H^{(1)} = \begin{bmatrix} h_1^{(1)} \\ h_2^{(1)} \\ h_3^{(1)} \\ h_4^{(1)} \end{bmatrix}$$



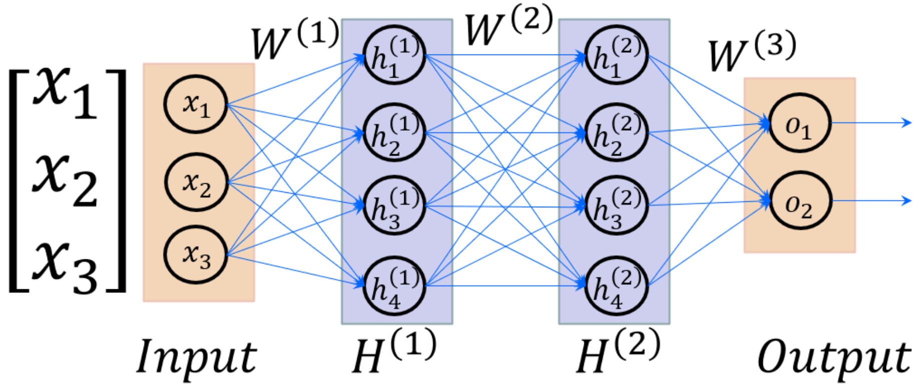
Forward propagation

– Lớp ẩn 1 được tính toán như sau:

$$a^{(1)} = W^{(1)}x + b^{(1)}$$

$$H^{(1)} = \sigma\left(a^{(1)}\right)$$

$$H^{(1)} = \sigma\left(\begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & w_{23}^{(1)} \\ w_{31}^{(1)} & w_{32}^{(1)} & w_{33}^{(1)} \\ w_{41}^{(1)} & w_{42}^{(1)} & w_{43}^{(1)} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} b_1^{(1)} \\ b_2^{(1)} \\ b_3^{(1)} \\ b_4^{(1)} \end{bmatrix}\right) = \begin{bmatrix} h_1^{(1)} \\ h_2^{(1)} \\ h_3^{(1)} \\ h_4^{(1)} \end{bmatrix}$$



$$h_1^{(1)} = \frac{1}{1+e^{-a_1^{(1)}}} = \frac{1}{1+e^{-\left(w_{11}^{(1)}x_1+w_{12}^{(1)}x_2+w_{13}^{(1)}x_3+b_1^{(1)}\right)}}$$

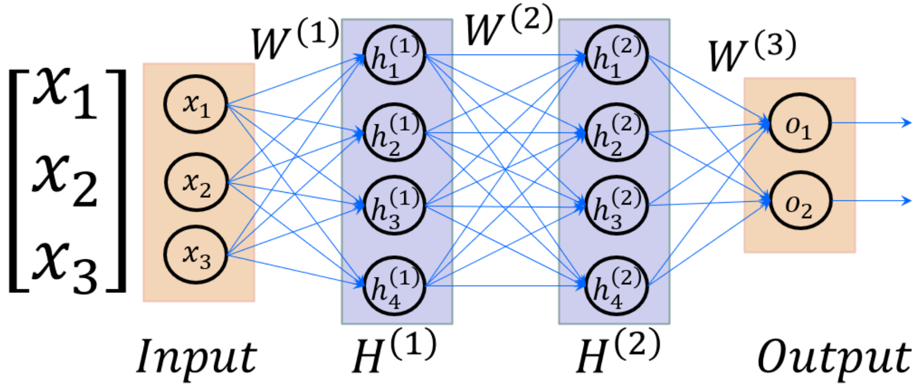
Forward propagation

– Lớp ẩn 1 được tính toán như sau:

$$a^{(1)} = W^{(1)}x + b^{(1)}$$

$$H^{(1)} = \sigma\left(a^{(1)}\right)$$

$$H^{(1)} = \sigma\left(\begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & w_{23}^{(1)} \\ w_{31}^{(1)} & w_{32}^{(1)} & w_{33}^{(1)} \\ w_{41}^{(1)} & w_{42}^{(1)} & w_{43}^{(1)} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} b_1^{(1)} \\ b_2^{(1)} \\ b_3^{(1)} \\ b_4^{(1)} \end{bmatrix}\right) = \begin{bmatrix} h_1^{(1)} \\ h_2^{(1)} \\ h_3^{(1)} \\ h_4^{(1)} \end{bmatrix}$$



$$h_2^{(1)} = \frac{1}{1+e^{-a_2^{(1)}}} = \frac{1}{1+e^{-\left(w_{21}^{(1)}x_1 + w_{22}^{(1)}x_2 + w_{23}^{(1)}x_3 + b_2^{(1)}\right)}}$$

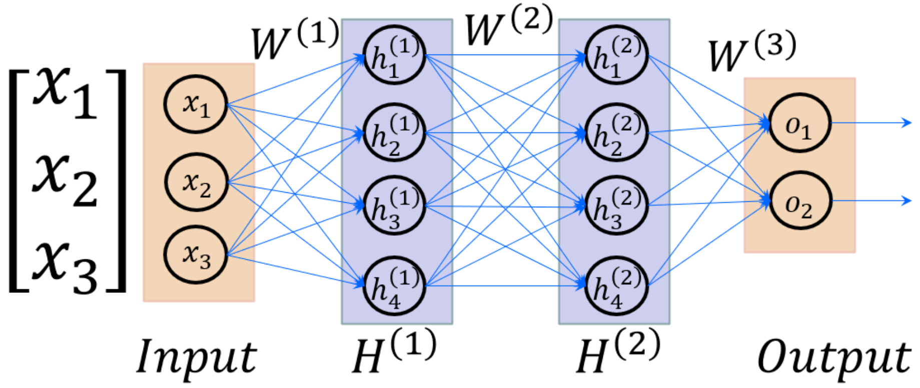
Forward propagation

– Lớp ẩn 1 được tính toán như sau:

$$a^{(1)} = W^{(1)}x + b^{(1)}$$

$$H^{(1)} = \sigma\left(a^{(1)}\right)$$

$$H^{(1)} = \sigma\left(\begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & w_{23}^{(1)} \\ w_{31}^{(1)} & w_{32}^{(1)} & w_{33}^{(1)} \\ w_{41}^{(1)} & w_{42}^{(1)} & w_{43}^{(1)} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} b_1^{(1)} \\ b_2^{(1)} \\ b_3^{(1)} \\ b_4^{(1)} \end{bmatrix}\right) = \begin{bmatrix} h_1^{(1)} \\ h_2^{(1)} \\ h_3^{(1)} \\ h_4^{(1)} \end{bmatrix}$$



$$h_3^{(1)} = \frac{1}{1+e^{-a_3^{(1)}}} = \frac{1}{1+e^{-\left(w_{31}^{(1)}x_1 + w_{32}^{(1)}x_2 + w_{33}^{(1)}x_3 + b_3^{(1)}\right)}}$$

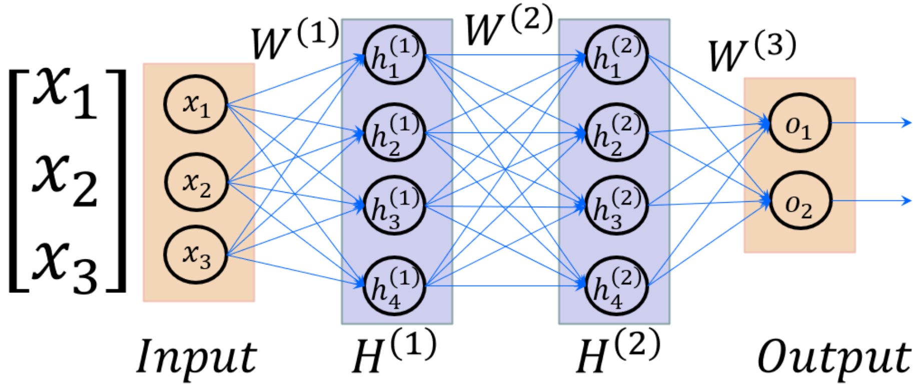
Forward propagation

– Lớp ẩn 1 được tính toán như sau:

$a^{(1)} = W^{(1)}x + b^{(1)}$

$H^{(1)} = \sigma\left(a^{(1)}\right)$

$$H^{(1)} = \sigma\left(\begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & w_{23}^{(1)} \\ w_{31}^{(1)} & w_{32}^{(1)} & w_{33}^{(1)} \\ w_{41}^{(1)} & w_{42}^{(1)} & w_{43}^{(1)} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} b_1^{(1)} \\ b_2^{(1)} \\ b_3^{(1)} \\ b_4^{(1)} \end{bmatrix}\right) = \begin{bmatrix} h_1^{(1)} \\ h_2^{(1)} \\ h_3^{(1)} \\ h_4^{(1)} \end{bmatrix}$$



$$h_4^{(1)} = \frac{1}{1+e^{-a_4^{(1)}}} = \frac{1}{1+e^{-\left(w_{41}^{(1)}x_1 + w_{42}^{(1)}x_2 + w_{43}^{(1)}x_3 + b_4^{(1)}\right)}}$$

Forward propagation

- $a^{(1)} = W^{(1)}x + b^{(1)}$

- $a^{(1)} = \begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & w_{23}^{(1)} \\ w_{31}^{(1)} & w_{32}^{(1)} & w_{33}^{(1)} \\ w_{41}^{(1)} & w_{42}^{(1)} & w_{43}^{(1)} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} b_1^{(1)} \\ b_2^{(1)} \\ b_3^{(1)} \\ b_4^{(1)} \end{bmatrix}$

Forward propagation

$$- a^{(1)} = W^{(1)}x + b^{(1)}$$

$$- a^{(1)} = \begin{bmatrix} 0 & w_{12}^{(1)} & w_{13}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & w_{23}^{(1)} \\ w_{31}^{(1)} & w_{32}^{(1)} & w_{33}^{(1)} \\ w_{41}^{(1)} & w_{42}^{(1)} & w_{43}^{(1)} \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 0 \\ b_2^{(1)} \\ b_3^{(1)} \\ b_4^{(1)} \end{bmatrix}$$

Forward propagation

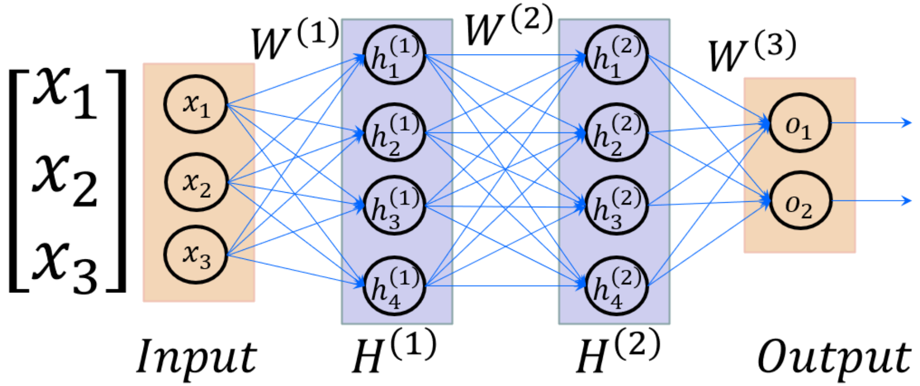
$$- H^{(1)} = \sigma\left(a^{(1)}\right) = \frac{1}{1+e^{-a^{(1)}}}$$

$$- H^{(1)} = \begin{bmatrix} h_1^{(1)} \\ h_2^{(1)} \\ h_3^{(1)} \\ h_4^{(1)} \end{bmatrix} = \begin{bmatrix} 0 \\ h_2^{(1)} \\ h_3^{(1)} \\ h_4^{(1)} \end{bmatrix}$$

Forward propagation

– Input của lớp ẩn 2

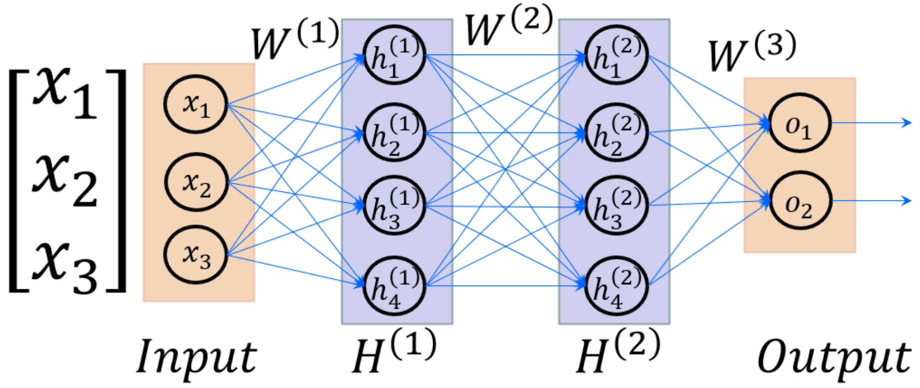
$$H^{(1)} = \begin{bmatrix} h_1^{(1)} \\ h_2^{(1)} \\ h_3^{(1)} \\ h_4^{(1)} \end{bmatrix} = \begin{bmatrix} 0 \\ h_2^{(1)} \\ h_3^{(1)} \\ h_4^{(1)} \end{bmatrix}$$



Forward propagation

– Ma trận trọng số $W^{(2)}$

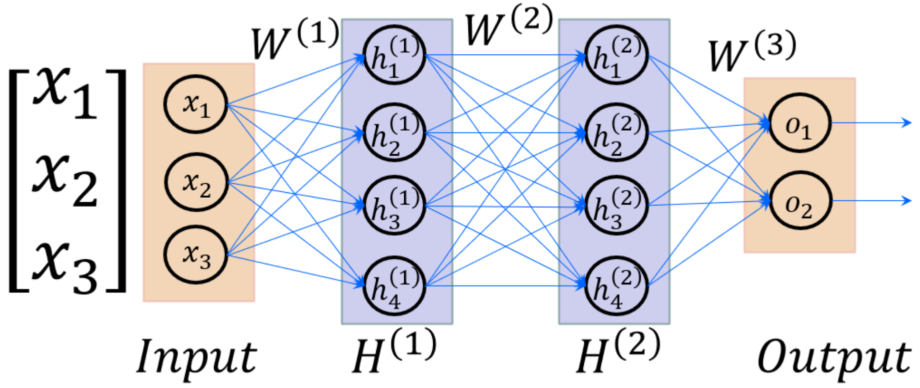
$$W^{(2)} = \begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} & w_{14}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)} & w_{24}^{(2)} \\ w_{31}^{(2)} & w_{32}^{(2)} & w_{33}^{(2)} & w_{34}^{(2)} \\ w_{41}^{(2)} & w_{42}^{(2)} & w_{43}^{(2)} & w_{44}^{(2)} \end{bmatrix}$$



Forward propagation

– Ma trận trọng số $W^{(2)}$

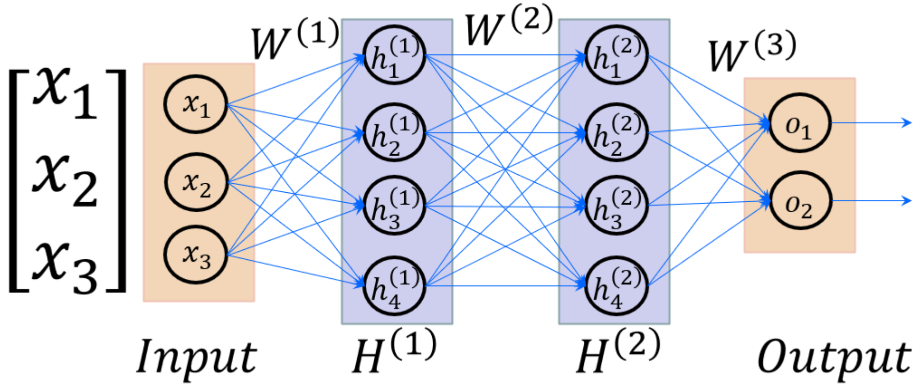
$$W^{(2)} = \begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} & w_{14}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)} & w_{24}^{(2)} \\ w_{31}^{(2)} & w_{32}^{(2)} & w_{33}^{(2)} & w_{34}^{(2)} \\ w_{41}^{(2)} & w_{42}^{(2)} & w_{43}^{(2)} & w_{44}^{(2)} \end{bmatrix}$$



Forward propagation

– Hệ số bias của lớp ẩn 2

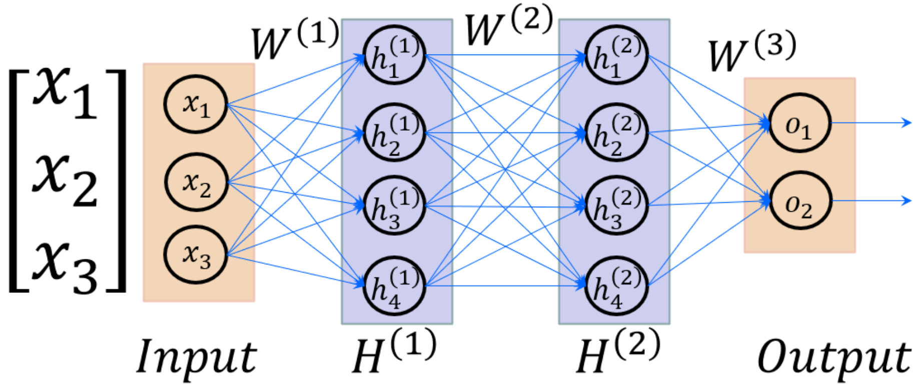
$$b^{(2)} = \begin{bmatrix} b_1^{(2)} \\ b_2^{(2)} \\ b_3^{(2)} \\ b_4^{(2)} \end{bmatrix}$$



Forward propagation

– Hệ số bias của lớp ẩn 2

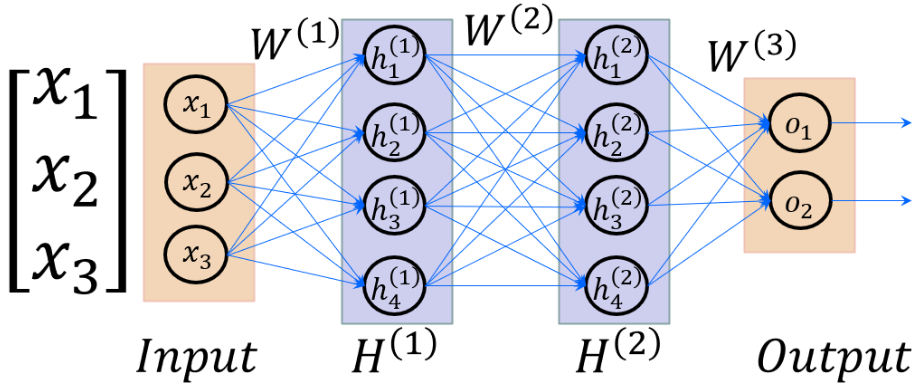
$$b^{(2)} = \begin{bmatrix} b_1^{(2)} \\ b_2^{(2)} \\ b_3^{(2)} \\ b_4^{(2)} \end{bmatrix}$$



Forward propagation

– Lớp ẩn 2

$$H^{(2)} = \begin{bmatrix} h_1^{(2)} \\ h_2^{(2)} \\ h_3^{(2)} \\ h_4^{(2)} \end{bmatrix}$$



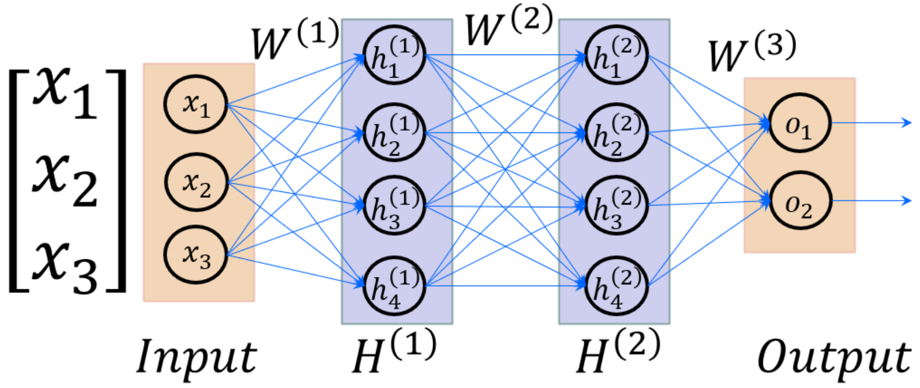
Forward propagation

– Lớp ẩn 2 được tính toán như sau:

$$a^{(2)} = W^{(2)}H^{(1)} + b^{(2)}$$

$$H^{(2)} = \sigma\left(a^{(2)}\right)$$

$$H^{(2)} = \sigma\left(\begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} & w_{14}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)} & w_{24}^{(2)} \\ w_{31}^{(2)} & w_{32}^{(2)} & w_{33}^{(2)} & w_{34}^{(2)} \\ w_{41}^{(2)} & w_{42}^{(2)} & w_{43}^{(2)} & w_{44}^{(2)} \end{bmatrix}\begin{bmatrix} h_1^{(1)} \\ h_2^{(1)} \\ h_3^{(1)} \\ h_4^{(1)} \end{bmatrix} + \begin{bmatrix} b_1^{(2)} \\ b_2^{(2)} \\ b_3^{(2)} \\ b_4^{(2)} \end{bmatrix}\right) = \begin{bmatrix} h_1^{(2)} \\ h_2^{(2)} \\ h_3^{(2)} \\ h_4^{(2)} \end{bmatrix}$$



$$h_1^{(2)} = \frac{1}{1+e^{-a_1^{(2)}}} = \frac{1}{1+e^{-\left(w_{11}^{(2)}h_1^{(1)}+w_{12}^{(2)}h_2^{(1)}+w_{13}^{(2)}h_3^{(1)}+w_{14}^{(2)}h_4^{(1)}+b_1^{(2)}\right)}}$$

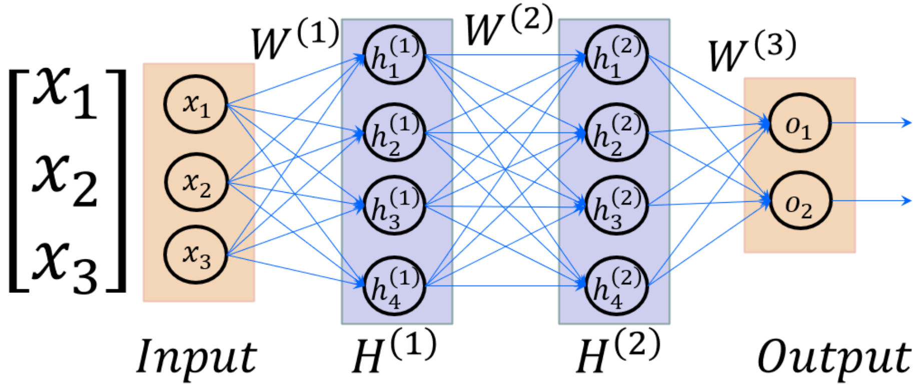
Forward propagation

– Lớp ẩn 2 được tính toán như sau:

$$a^{(2)} = W^{(2)}H^{(1)} + b^{(2)}$$

$$H^{(2)} = \sigma\left(a^{(2)}\right)$$

$$H^{(2)} = \sigma\left(\begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} & w_{14}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)} & w_{24}^{(2)} \\ w_{31}^{(2)} & w_{32}^{(2)} & w_{33}^{(2)} & w_{34}^{(2)} \\ w_{41}^{(2)} & w_{42}^{(2)} & w_{43}^{(2)} & w_{44}^{(2)} \end{bmatrix}\begin{bmatrix} h_1^{(1)} \\ h_2^{(1)} \\ h_3^{(1)} \\ h_4^{(1)} \end{bmatrix} + \begin{bmatrix} b_1^{(2)} \\ b_2^{(2)} \\ b_3^{(2)} \\ b_4^{(2)} \end{bmatrix}\right) = \begin{bmatrix} h_1^{(2)} \\ h_2^{(2)} \\ h_3^{(2)} \\ h_4^{(2)} \end{bmatrix}$$



$$h_2^{(2)} = \frac{1}{1+e^{-a_2^{(2)}}} = \frac{1}{1+e^{-\left(w_{21}^{(2)}h_1^{(1)}+w_{22}^{(2)}h_2^{(1)}+w_{23}^{(2)}h_3^{(1)}+w_{24}^{(2)}h_4^{(1)}+b_2^{(2)}\right)}}$$

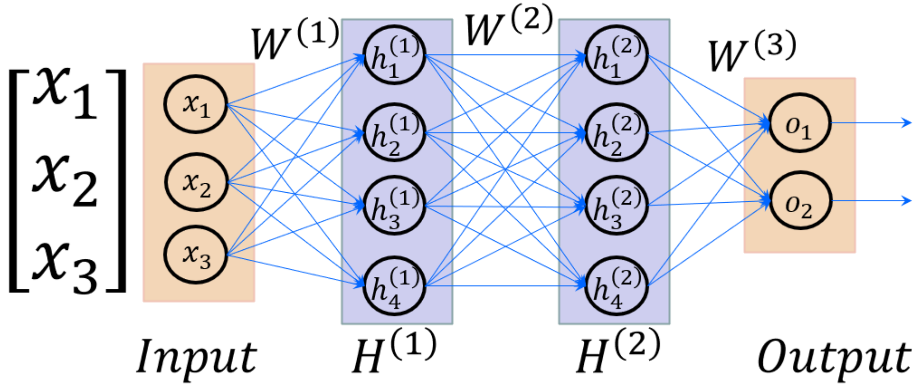
Forward propagation

– Lớp ẩn 2 được tính toán như sau:

$$a^{(2)} = W^{(2)}H^{(1)} + b^{(2)}$$

$$H^{(2)} = \sigma\left(a^{(2)}\right)$$

$$H^{(2)} = \sigma\left(\begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} & w_{14}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)} & w_{24}^{(2)} \\ w_{31}^{(2)} & w_{32}^{(2)} & w_{33}^{(2)} & w_{34}^{(2)} \\ w_{41}^{(2)} & w_{42}^{(2)} & w_{43}^{(2)} & w_{44}^{(2)} \end{bmatrix}\begin{bmatrix} h_1^{(1)} \\ h_2^{(1)} \\ h_3^{(1)} \\ h_4^{(1)} \end{bmatrix} + \begin{bmatrix} b_1^{(2)} \\ b_2^{(2)} \\ b_3^{(2)} \\ b_4^{(2)} \end{bmatrix}\right) = \begin{bmatrix} h_1^{(2)} \\ h_2^{(2)} \\ h_3^{(2)} \\ h_4^{(2)} \end{bmatrix}$$



$$h_3^{(2)} = \frac{1}{1+e^{-a_3^{(2)}}} = \frac{1}{1+e^{-\left(w_{31}^{(2)}h_1^{(1)}+w_{32}^{(2)}h_2^{(1)}+w_{33}^{(2)}h_3^{(1)}+w_{34}^{(2)}h_4^{(1)}+b_3^{(2)}\right)}}$$

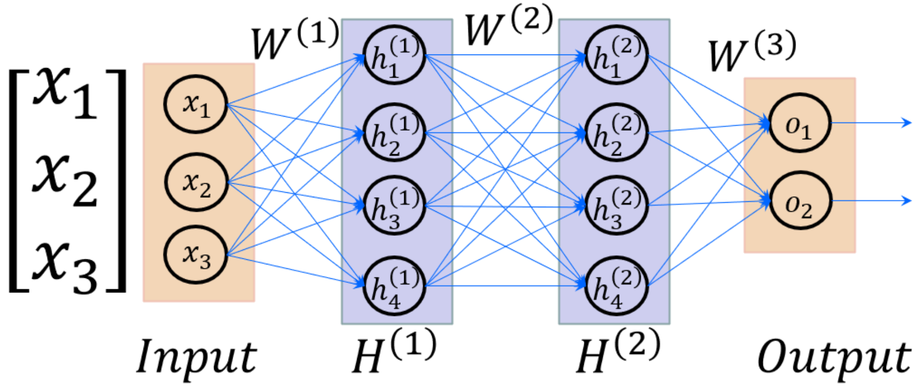
Forward propagation

– Lớp ẩn 2 được tính toán như sau:

$$a^{(2)} = W^{(2)}H^{(1)} + b^{(2)}$$

$$H^{(2)} = \sigma\left(a^{(2)}\right)$$

$$H^{(2)} = \sigma\left(\begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} & w_{14}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)} & w_{24}^{(2)} \\ w_{31}^{(2)} & w_{32}^{(2)} & w_{33}^{(2)} & w_{34}^{(2)} \\ w_{41}^{(2)} & w_{42}^{(2)} & w_{43}^{(2)} & w_{44}^{(2)} \end{bmatrix}\begin{bmatrix} h_1^{(1)} \\ h_2^{(1)} \\ h_3^{(1)} \\ h_4^{(1)} \end{bmatrix} + \begin{bmatrix} b_1^{(2)} \\ b_2^{(2)} \\ b_3^{(2)} \\ b_4^{(2)} \end{bmatrix}\right) = \begin{bmatrix} h_1^{(2)} \\ h_2^{(2)} \\ h_3^{(2)} \\ h_4^{(2)} \end{bmatrix}$$



$$h_4^{(2)} = \frac{1}{1+e^{-a_4^{(2)}}} = \frac{1}{1+e^{-\left(w_{41}^{(2)}h_1^{(1)}+w_{42}^{(2)}h_2^{(1)}+w_{43}^{(2)}h_3^{(1)}+w_{44}^{(2)}h_4^{(1)}+b_4^{(2)}\right)}}$$

Forward propagation

$- a^{(2)} = W^{(2)}H^{(1)} + b^{(2)}$

$$- a^{(2)} = \begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} & w_{14}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)} & w_{24}^{(2)} \\ w_{31}^{(2)} & w_{32}^{(2)} & w_{33}^{(2)} & w_{34}^{(2)} \\ w_{41}^{(2)} & w_{42}^{(2)} & w_{43}^{(2)} & w_{44}^{(2)} \end{bmatrix} \begin{bmatrix} h_1^{(1)} \\ h_2^{(1)} \\ h_3^{(1)} \\ h_4^{(1)} \end{bmatrix} + \begin{bmatrix} b_1^{(2)} \\ b_2^{(2)} \\ b_3^{(2)} \\ b_4^{(2)} \end{bmatrix}$$

Forward propagation

$- \ a^{(2)} = W^{(2)} H^{(1)} + b^{(2)}$

$$- \ a^{(2)} = \begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} & w_{14}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)} & w_{24}^{(2)} \\ w_{31}^{(2)} & w_{32}^{(2)} & w_{33}^{(2)} & w_{34}^{(2)} \\ w_{41}^{(2)} & w_{42}^{(2)} & w_{43}^{(2)} & w_{44}^{(2)} \end{bmatrix} \begin{bmatrix} h_1^{(1)} \\ h_2^{(1)} \\ h_3^{(1)} \\ h_4^{(1)} \end{bmatrix} + \begin{bmatrix} b_1^{(2)} \\ b_2^{(2)} \\ b_3^{(2)} \\ b_4^{(2)} \end{bmatrix}$$

Forward propagation

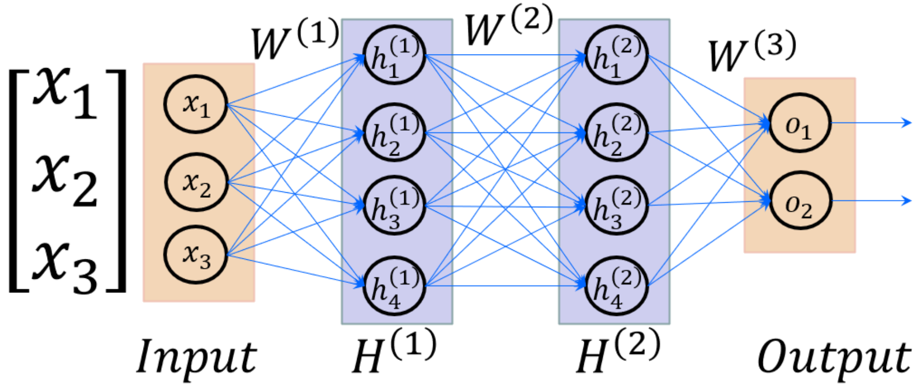
$$- H^{(2)} = \sigma\left(a^{(2)}\right) = \frac{1}{1+e^{-a^{(2)}}}$$

$$- H^{(2)} = \begin{bmatrix} h_1^{(2)} \\ h_2^{(2)} \\ h_3^{(2)} \\ h_4^{(2)} \end{bmatrix} = \begin{bmatrix} h_1^{(2)} \\ h_2^{(2)} \\ h_3^{(2)} \\ h_4^{(2)} \end{bmatrix}$$

Forward propagation

– Input của lớp output

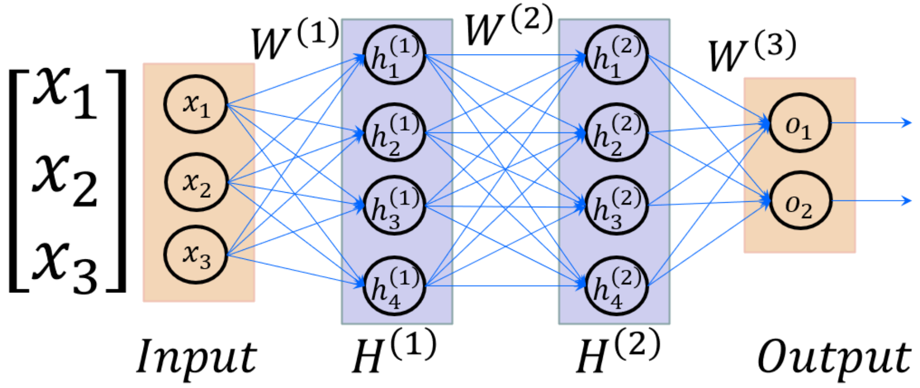
$$H^{(2)} = \begin{bmatrix} h_1^{(2)} \\ h_2^{(2)} \\ h_3^{(2)} \\ h_4^{(2)} \end{bmatrix}$$



Forward propagation

– Ma trận trọng số $W^{(3)}$

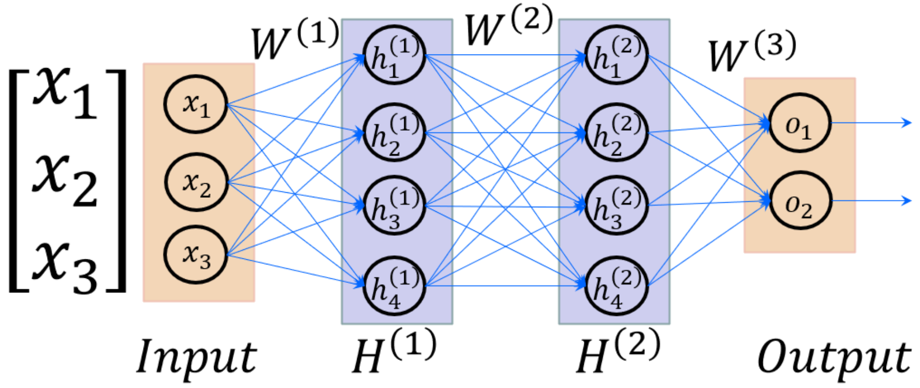
$$W^{(3)} = \begin{bmatrix} w_{11}^{(3)} & w_{12}^{(3)} & w_{13}^{(3)} & w_{14}^{(3)} \\ w_{21}^{(3)} & w_{22}^{(3)} & w_{23}^{(3)} & w_{24}^{(3)} \end{bmatrix}$$



Forward propagation

– Hệ số bias của lớp output

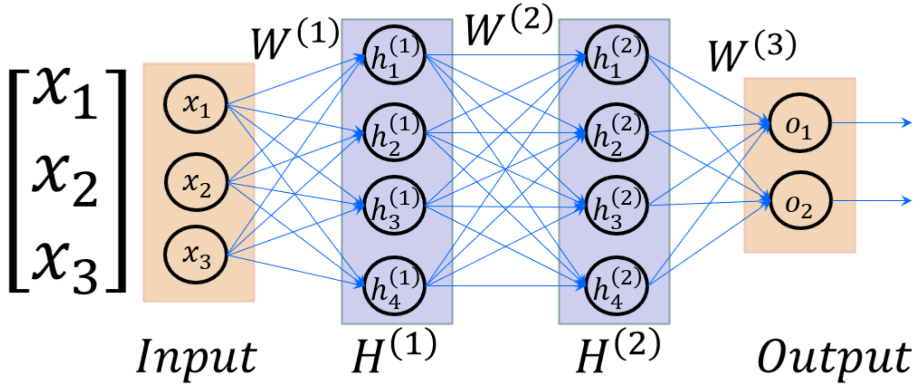
$$b^{(3)} = \begin{bmatrix} b_1^{(3)} \\ b_2^{(3)} \end{bmatrix}$$



Forward propagation

– Lớp output

$$H^{(3)} = \begin{bmatrix} h_1^{(3)} \\ h_1^{(3)} \end{bmatrix} = \begin{bmatrix} o_1 \\ o_2 \end{bmatrix} = [output]$$



Forward propagation

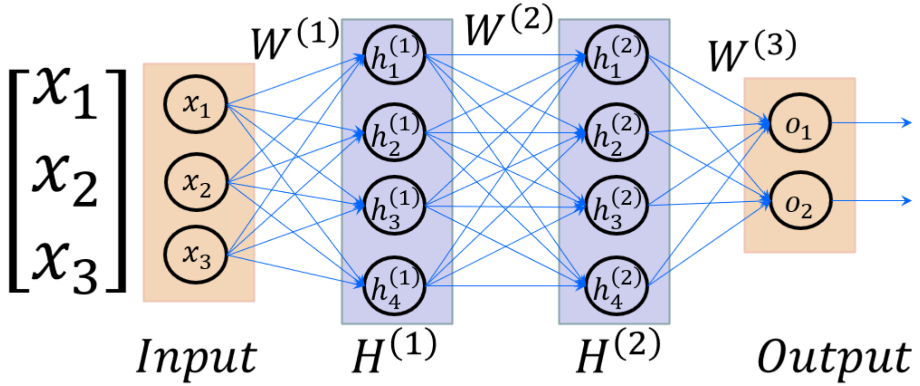
– Lớp output được tính toán như sau:

$$a^{(3)} = W^{(3)}H^{(2)} + b^{(3)}$$

$$O = \sigma\left(a^{(3)}\right)$$

$$O = \sigma\left(\begin{bmatrix} w_{11}^{(3)} & w_{12}^{(3)} & w_{13}^{(3)} & w_{14}^{(3)} \\ w_{21}^{(3)} & w_{22}^{(3)} & w_{23}^{(3)} & w_{24}^{(3)} \end{bmatrix}\begin{bmatrix} h_1^{(2)} \\ h_2^{(2)} \\ h_3^{(2)} \\ h_4^{(2)} \end{bmatrix} + \begin{bmatrix} b_1^{(3)} \\ b_2^{(3)} \end{bmatrix}\right)$$

$$o_1 = \frac{1}{1+e^{-a_1^{(3)}}} = \frac{1}{1+e^{-\left(w_{11}^{(3)}h_1^{(2)}+w_{12}^{(3)}h_2^{(2)}+w_{13}^{(3)}h_3^{(2)}+w_{14}^{(3)}h_4^{(2)}+b_1^{(3)}\right)}}$$



UIT VNUHCM University of Information Technology

Forward propagation

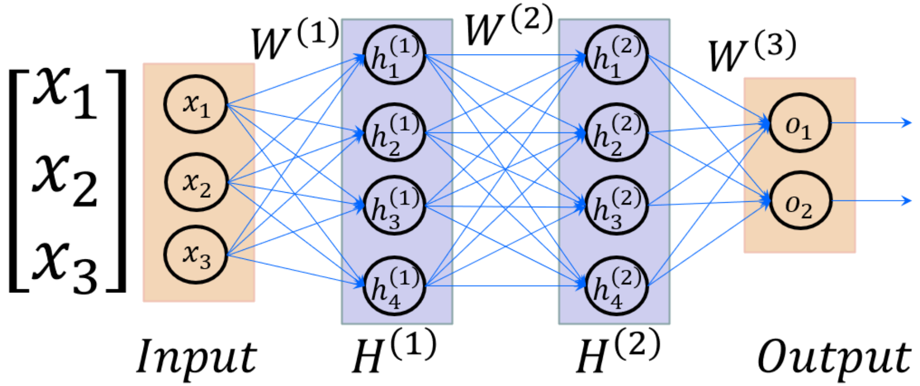
– Lớp output được tính toán như sau:

$$a^{(3)} = W^{(3)}H^{(2)} + b^{(3)}$$

$$O = \sigma\left(a^{(3)}\right)$$

$$O = \sigma\left(\begin{bmatrix} w_{11}^{(3)} & w_{12}^{(3)} & w_{13}^{(3)} & w_{14}^{(3)} \\ w_{21}^{(3)} & w_{22}^{(3)} & w_{23}^{(3)} & w_{24}^{(3)} \end{bmatrix}\begin{bmatrix} h_1^{(2)} \\ h_2^{(2)} \\ h_3^{(2)} \\ h_4^{(2)} \end{bmatrix} + \begin{bmatrix} b_1^{(3)} \\ b_2^{(3)} \end{bmatrix}\right)$$

$$o_2 = \frac{1}{1+e^{-a_2^{(3)}}} = \frac{1}{1+e^{-\left(w_{21}^{(3)}h_1^{(2)}+w_{22}^{(3)}h_2^{(2)}+w_{23}^{(3)}h_3^{(2)}+w_{24}^{(3)}h_4^{(2)}+b_2^{(3)}\right)}}$$



Neural networks

LEARNING PROCESS

Learning process

- Bước 01: Initialize network weights (often small random values)
- Bước 02: For each datapoint i in training set:
  + Bước 2.1: Forward propagation.
  + Bước 2.2: Backward propagation.
  + Bước 2.3: Weight update.
- Bước 3: Return the network.

Chúc các bạn học tốt

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN TP.HCM

Nhóm UIT-Together

TS. Nguyễn Tấn Trần Minh Khang