

الگوریتم پس انتشار خطا و شبکه‌های عصبی

سید ناصر رضوی www.snrazavi.ir

۱۳۹۷

یادآوری: محاسبه گرادیان به صورت تحلیلی

۲

□ تابع هزینه یک تابع از پارامترهای W است.

$$s = f(x; W) = Wx$$

$$L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)$$

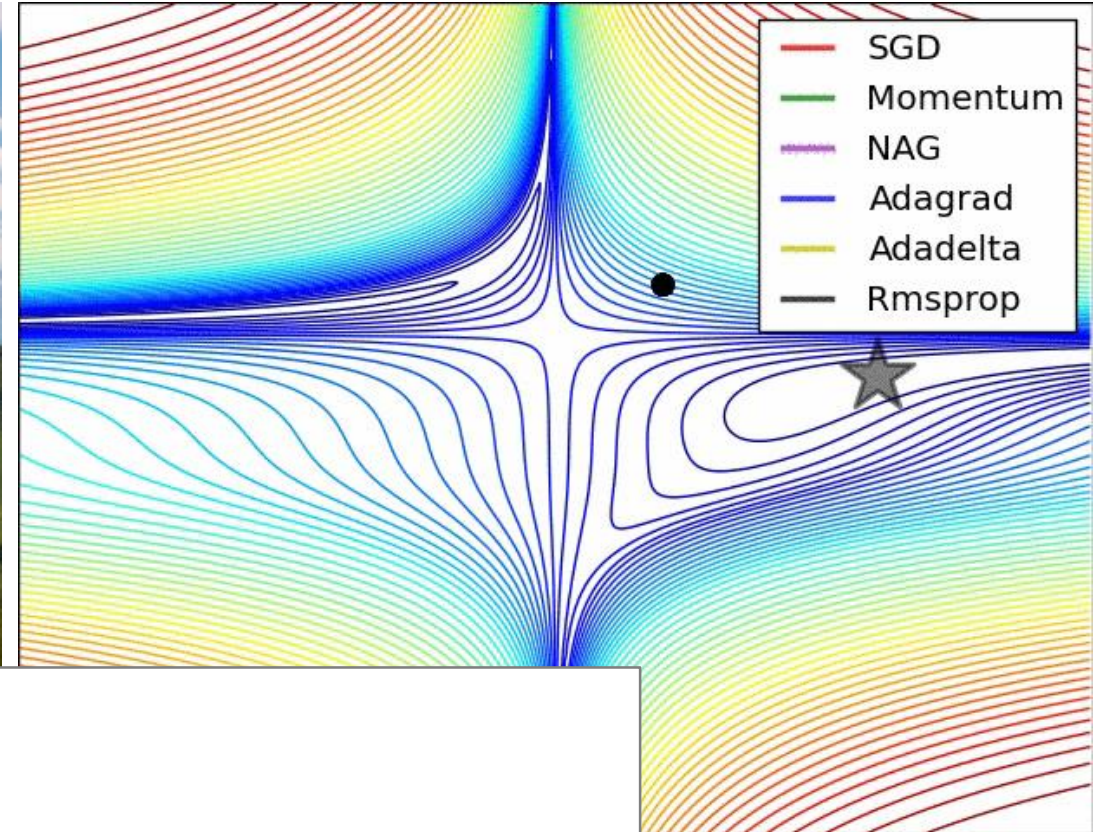
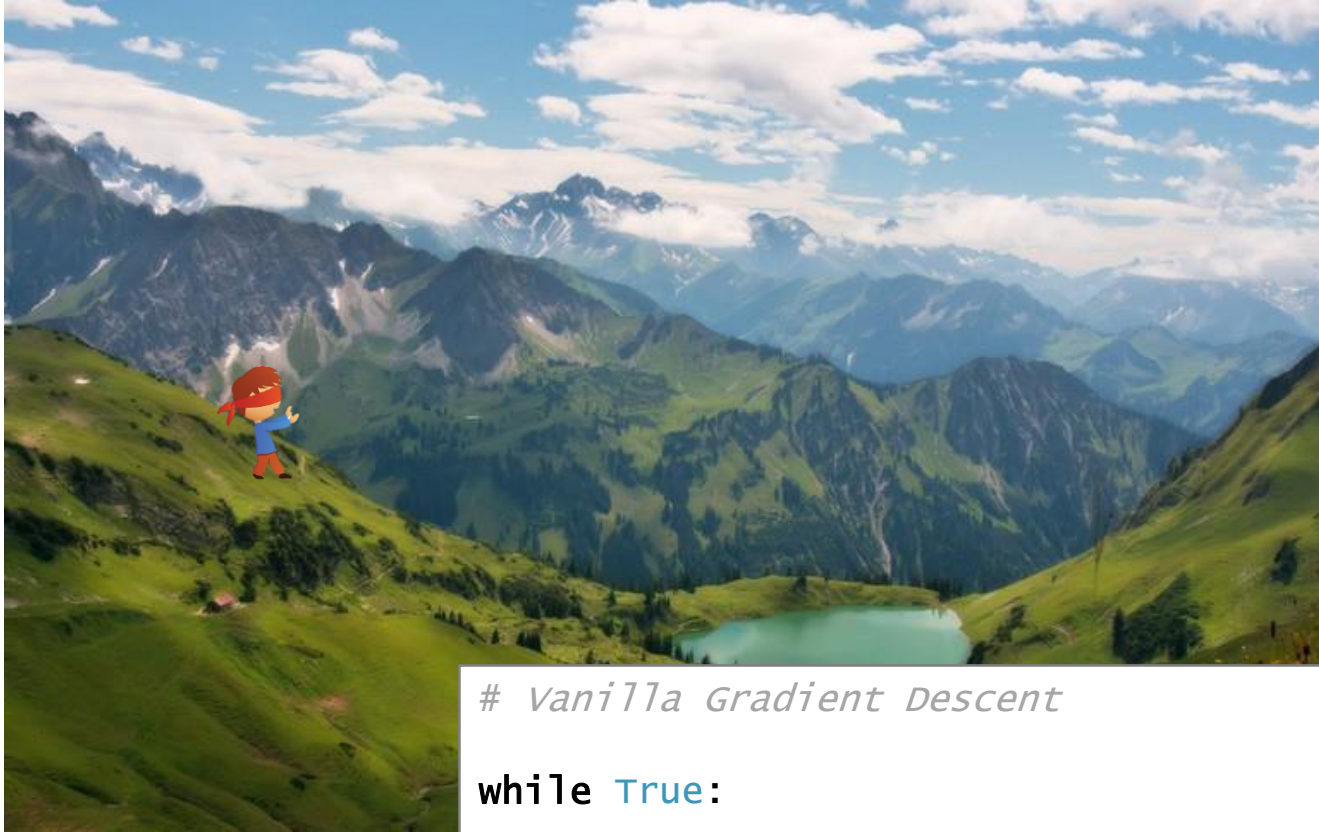
$$L = \frac{1}{N} \sum_{i=1}^N L_i + \lambda \sum_k W_k^2$$

$$\nabla_W L = \dots$$



یادآوری: بهینه‌سازی

۳



```
# Vanilla Gradient Descent
```

```
while True:
```

```
    gradient = evaluate_gradient(loss_fun, data, weights)
```

```
    weights += -step_size * gradient # weight update
```

یادآوری: بررسی گرادیان

۴

□ به طور خلاصه.

□ گرادیان عددی: تقریبی، زمان بر، پیاده سازی آسان!

□ گرادیان تحلیلی: دقیق، سریع، امکان بروز خطا در پیاده سازی!

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

□ در عمل.

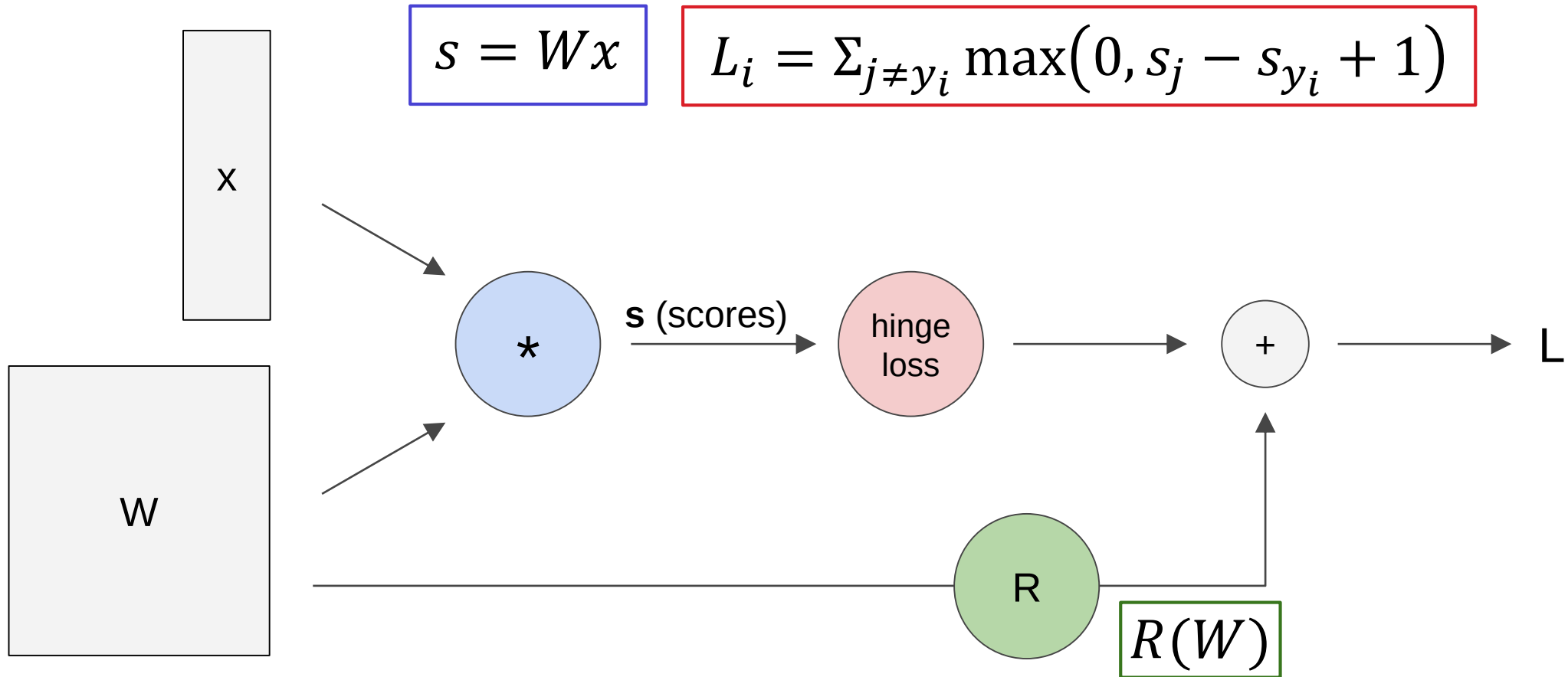
□ همیشه از گرادیان تحلیلی استفاده می کنیم.

□ اما به منظور اطمینان از صحت پیاده سازی، گرادیان تحلیلی را با گرادیان عددی مقایسه می کنیم.

بررسی گرادیان

گراف محاسباتی

۵

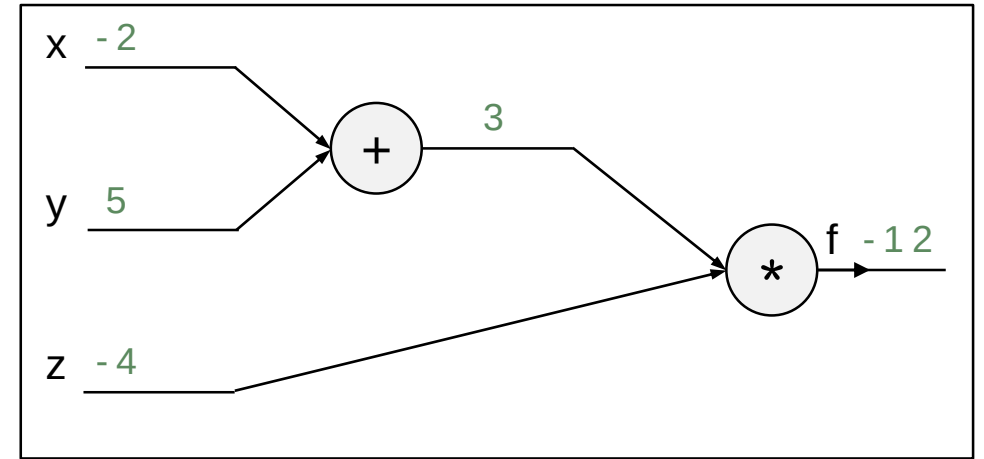


گراف محاسباتی

۶

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

$$x = -2, y = 5, z = -4$$



گراف محاسباتی

۷

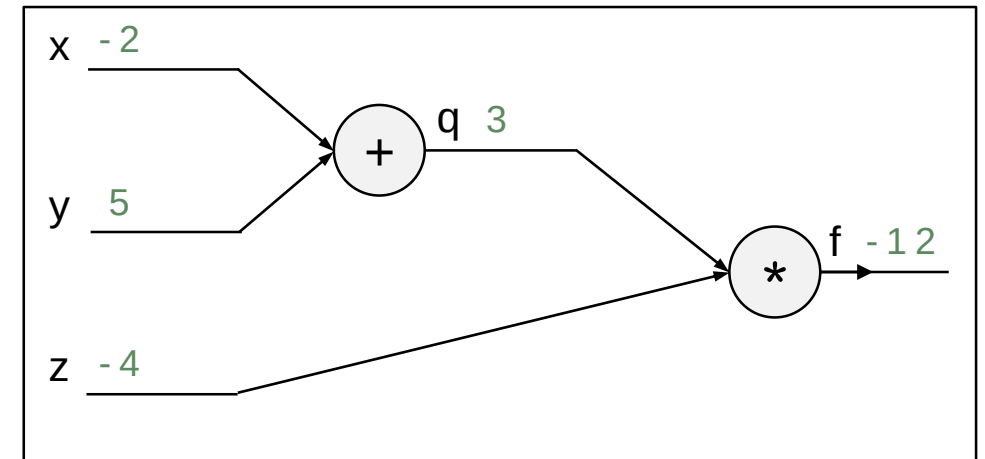
$$f(x, y, z) = (x + y) \cdot z$$

$$x = -2, y = 5, z = -4$$

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

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

$$\frac{\partial f}{\partial x}, \quad \frac{\partial f}{\partial y}, \quad \frac{\partial f}{\partial z} \quad \text{مقادیر مورد نیاز}$$



گراف محاسباتی

۸

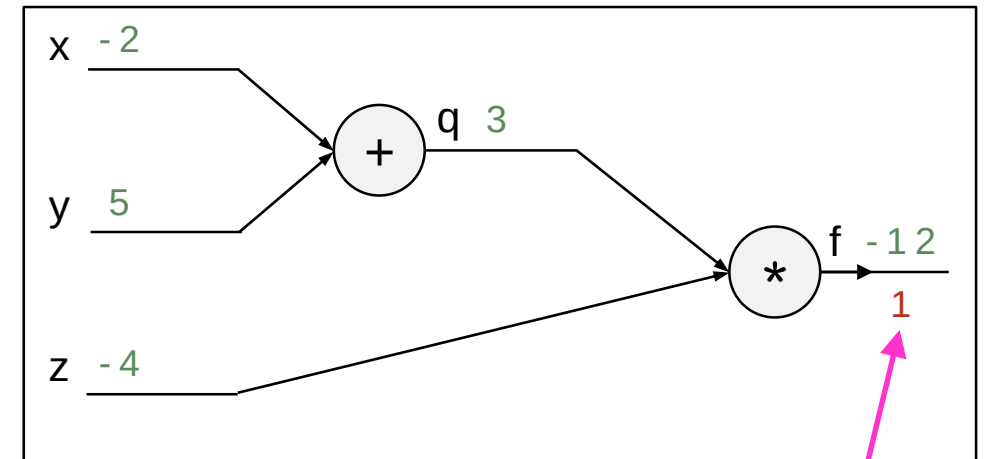
$$f(x, y, z) = (x + y) \cdot z$$

$$x = -2, y = 5, z = -4$$

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

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

$$\frac{\partial f}{\partial x}, \quad \frac{\partial f}{\partial y}, \quad \frac{\partial f}{\partial z} \quad \text{مقادیر مورد نیاز}$$



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

گراف محاسباتی

۹

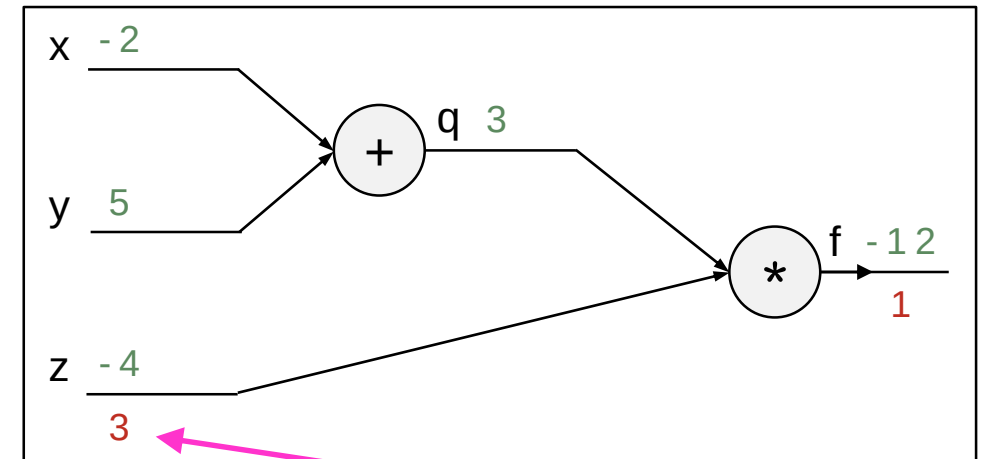
$$f(x, y, z) = (x + y) \cdot z$$

$$x = -2, y = 5, z = -4$$

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

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

$$\frac{\partial f}{\partial x}, \quad \frac{\partial f}{\partial y}, \quad \frac{\partial f}{\partial z} \quad \text{مقادیر مورد نیاز}$$



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

گراف محاسباتی

۱۰

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

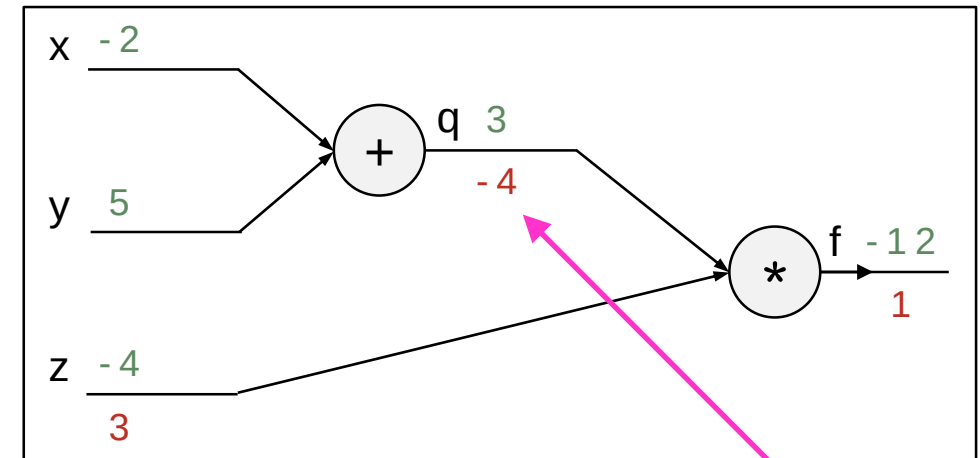
$$x = -2, y = 5, z = -4$$

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

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

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

مقادیر مورد نیاز



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

گراف محاسباتی

۱۱

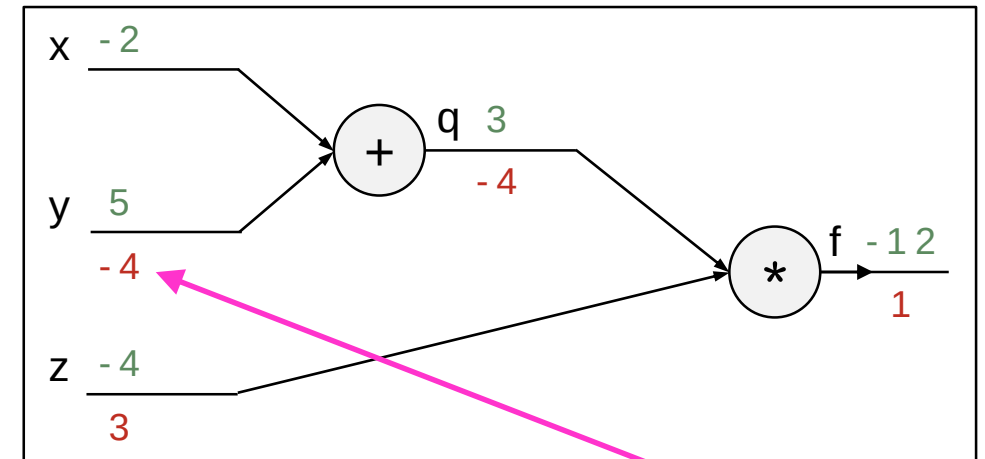
$$f(x, y, z) = (x + y) \cdot z$$

$$x = -2, y = 5, z = -4$$

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

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

$$\frac{\partial f}{\partial x}, \quad \frac{\partial f}{\partial y}, \quad \frac{\partial f}{\partial z} \quad \text{مقادیر مورد نیاز}$$



قاعده زنجیری

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

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

گراف محاسباتی

۱۲

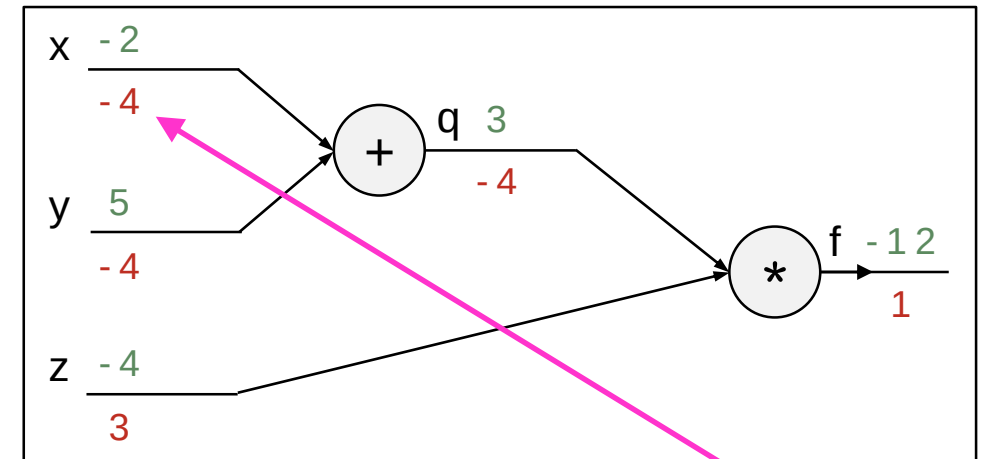
$$f(x, y, z) = (x + y) \cdot z$$

$$x = -2, y = 5, z = -4$$

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

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

$$\frac{\partial f}{\partial x}, \quad \frac{\partial f}{\partial y}, \quad \frac{\partial f}{\partial z} \quad \text{مقادیر مورد نیاز}$$



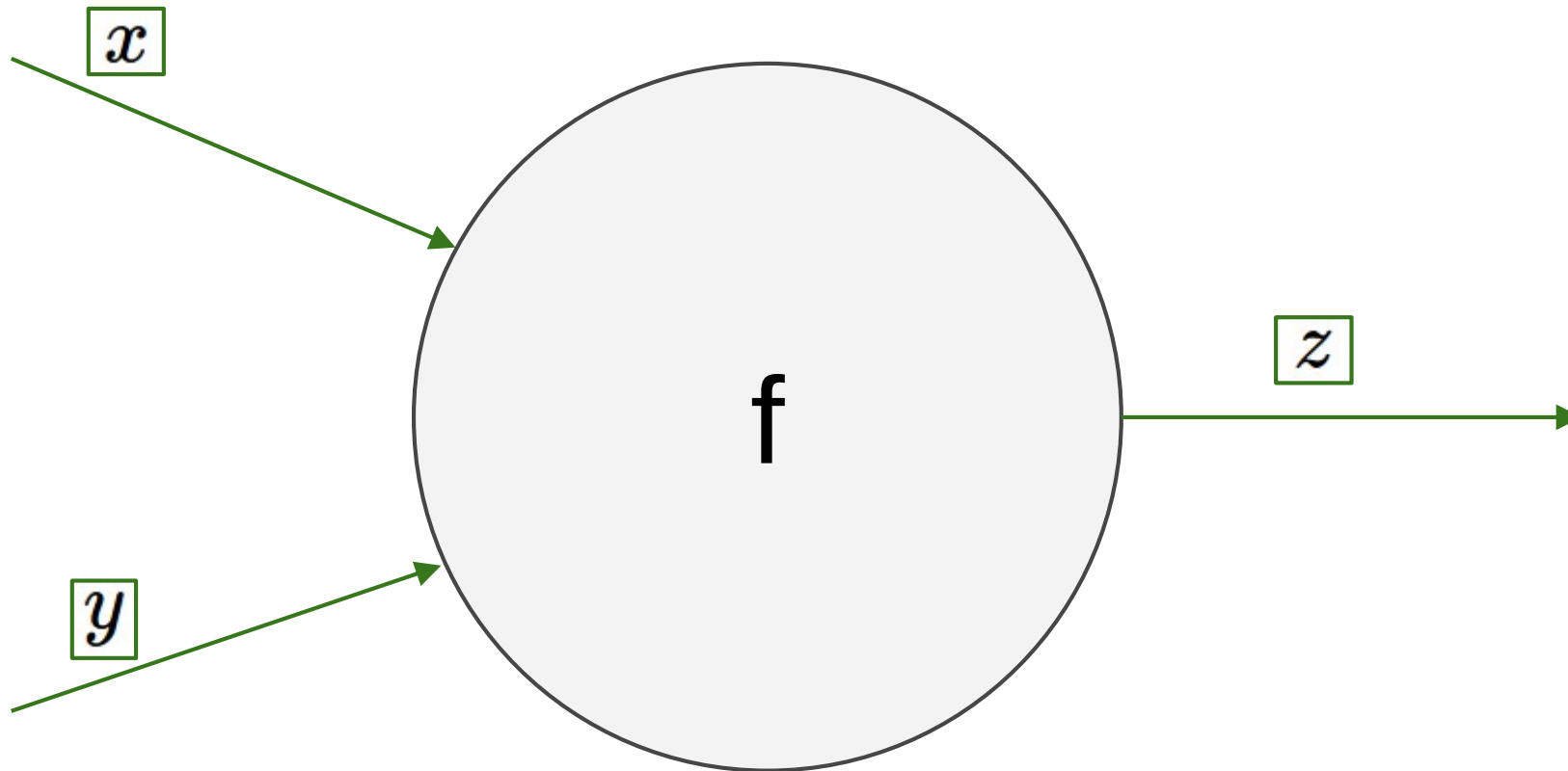
$$\frac{\partial f}{\partial x}$$

قاعده زنجیری

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

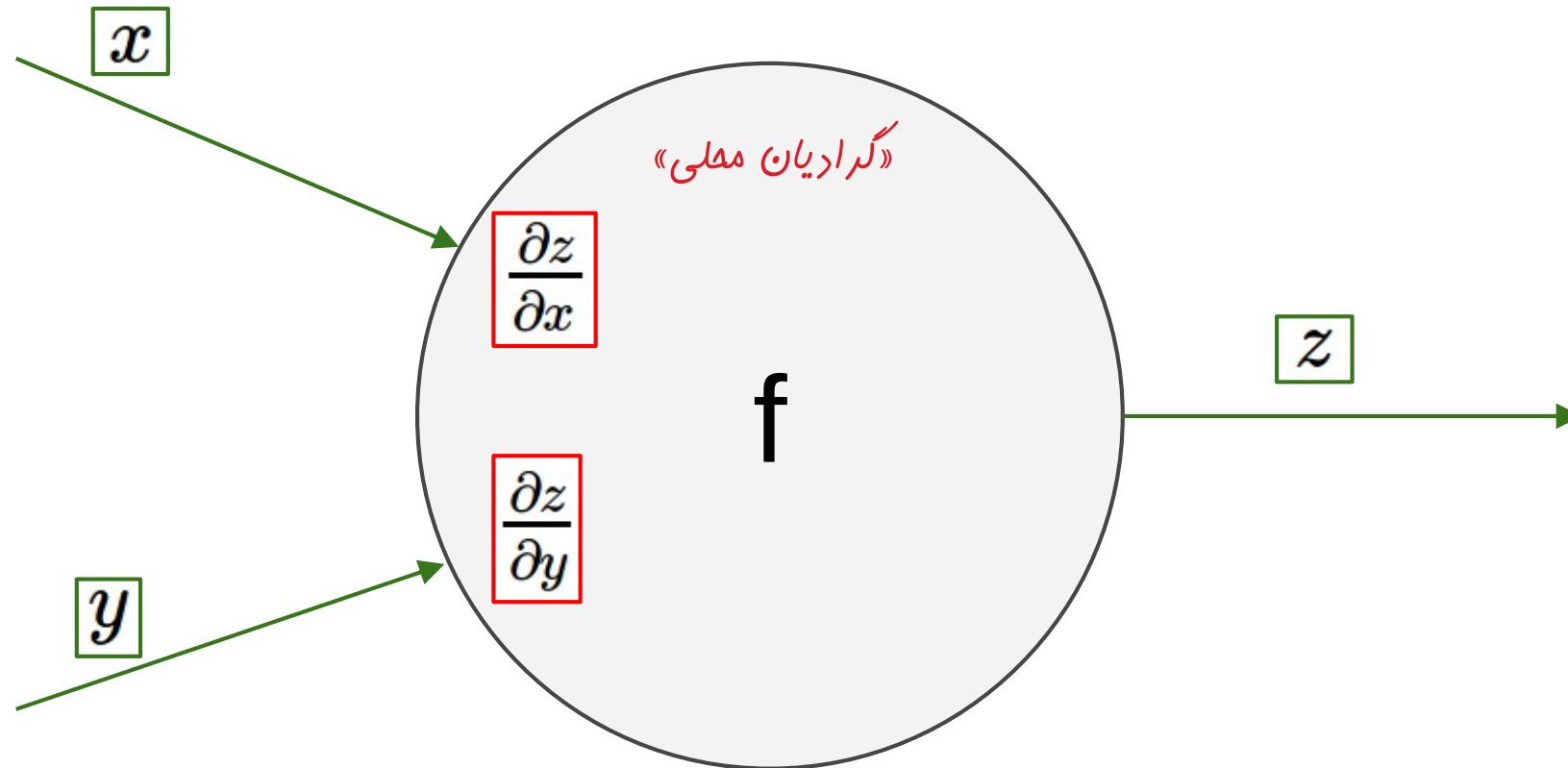
گراف محاسباتی

۱۳



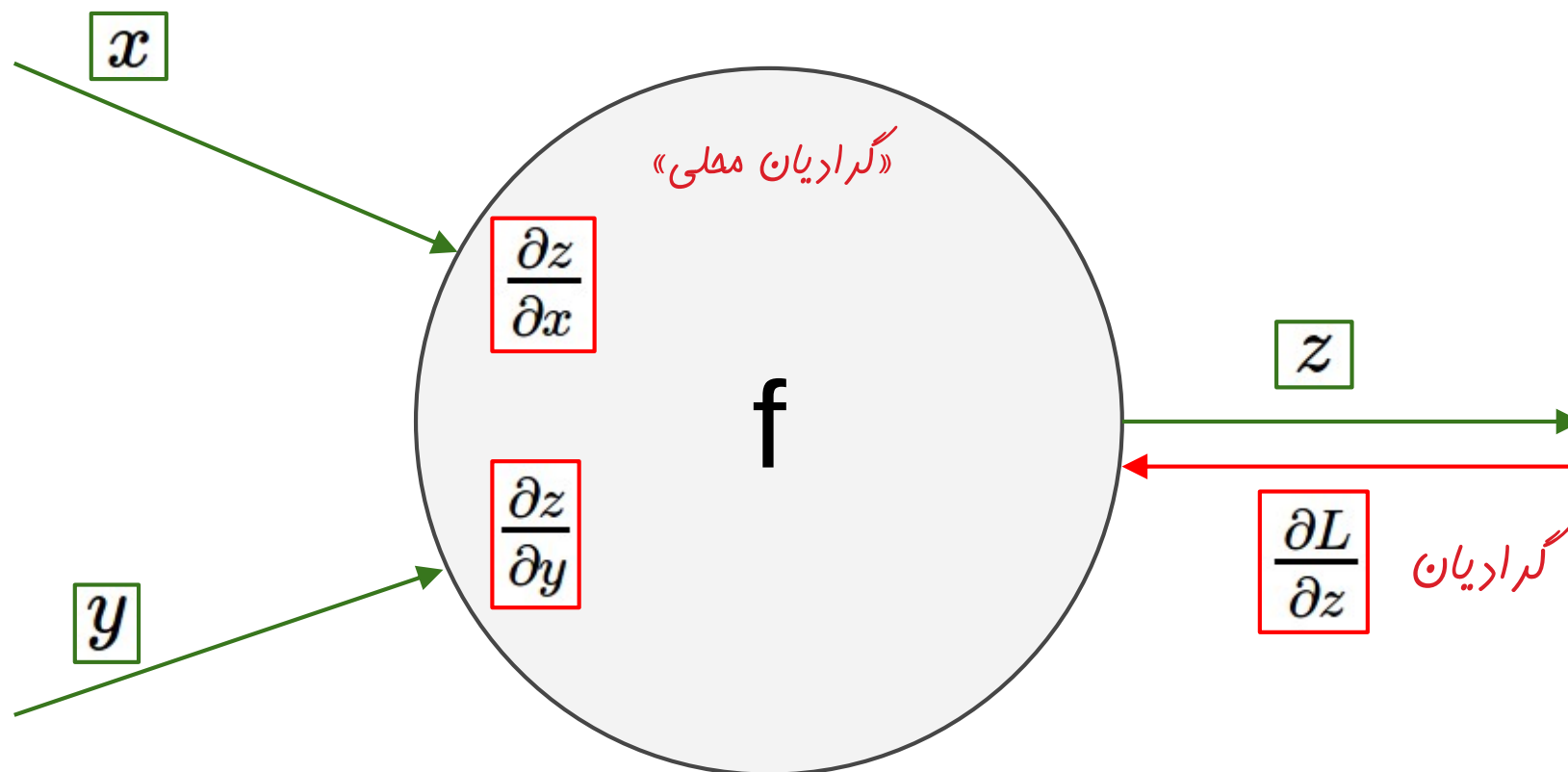
گراف محاسباتی

۱۴



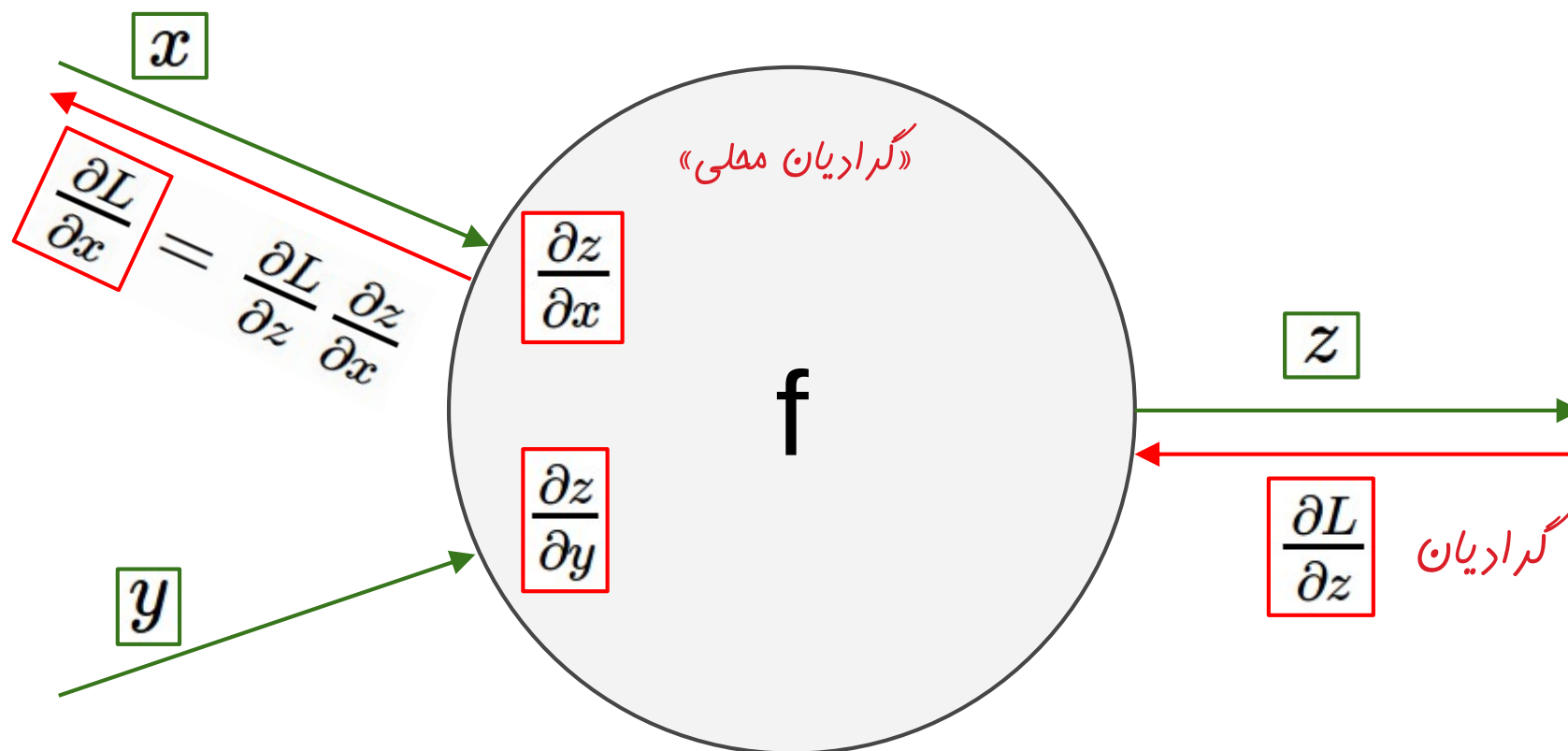
گراف محاسباتی

۱۵



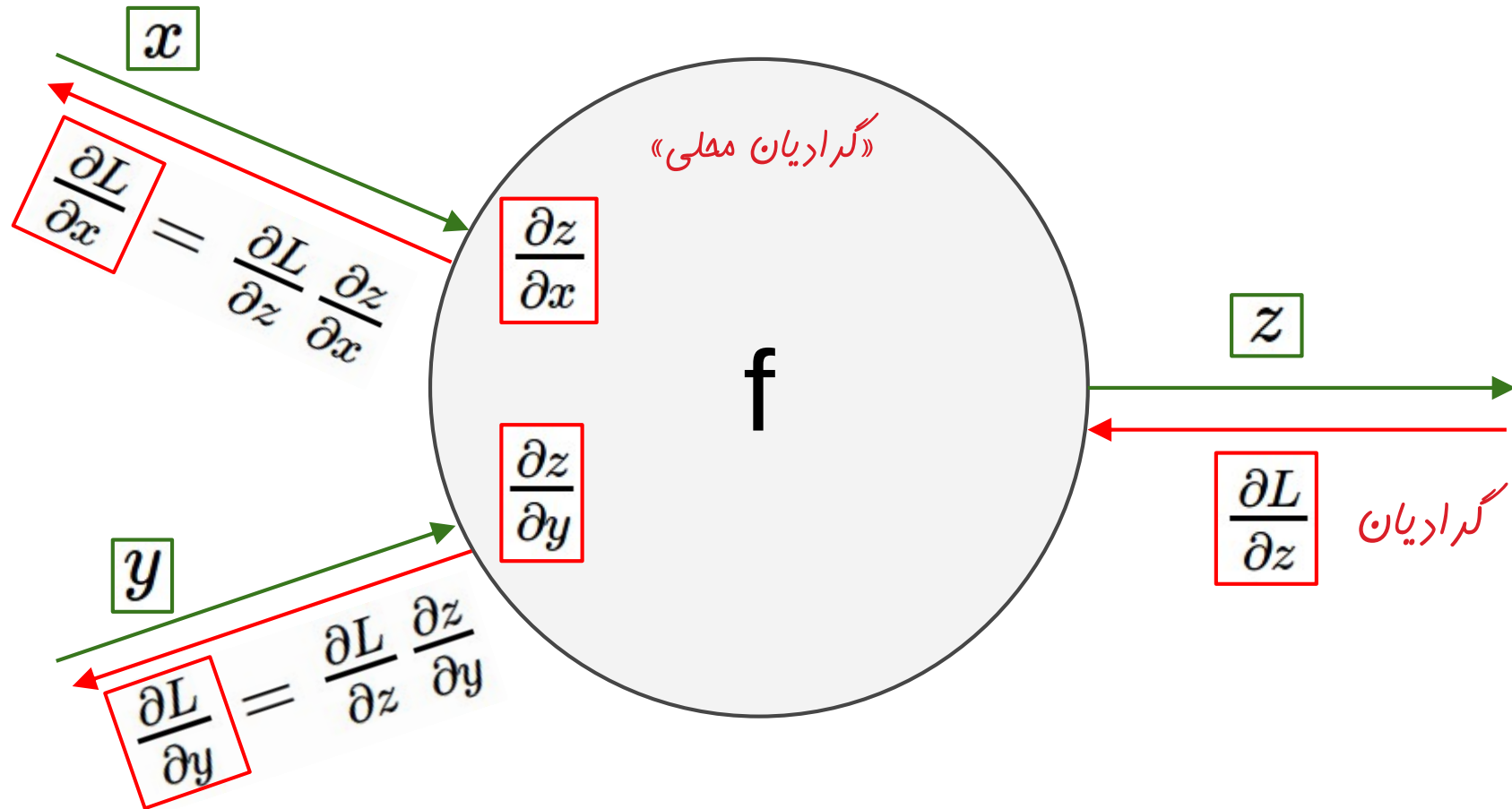
گراف محاسباتی

۱۶



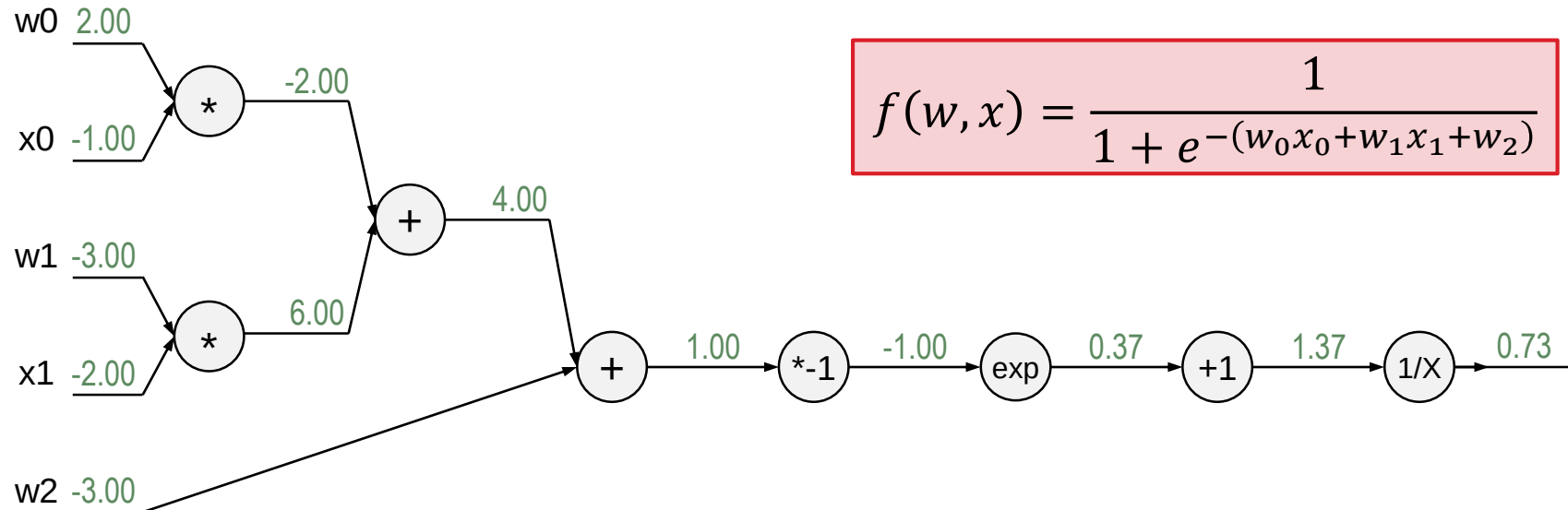
گراف محاسباتی

۱۷



گراف محاسباتی: یک مثال دیگر

۱۸



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

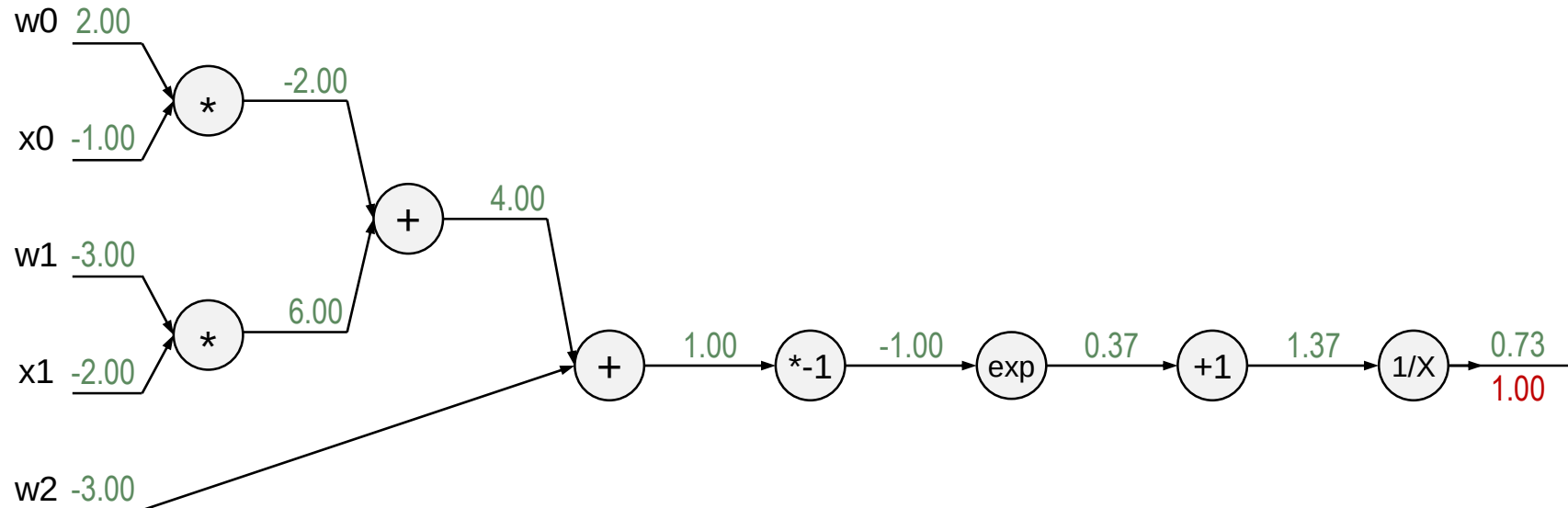
$$f_a(x) = ax \rightarrow \frac{df}{dx} = a$$

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

$$f_c(x) = x + c \rightarrow \frac{df}{dx} = 1$$

گراف محاسباتی: یک مثال دیگر

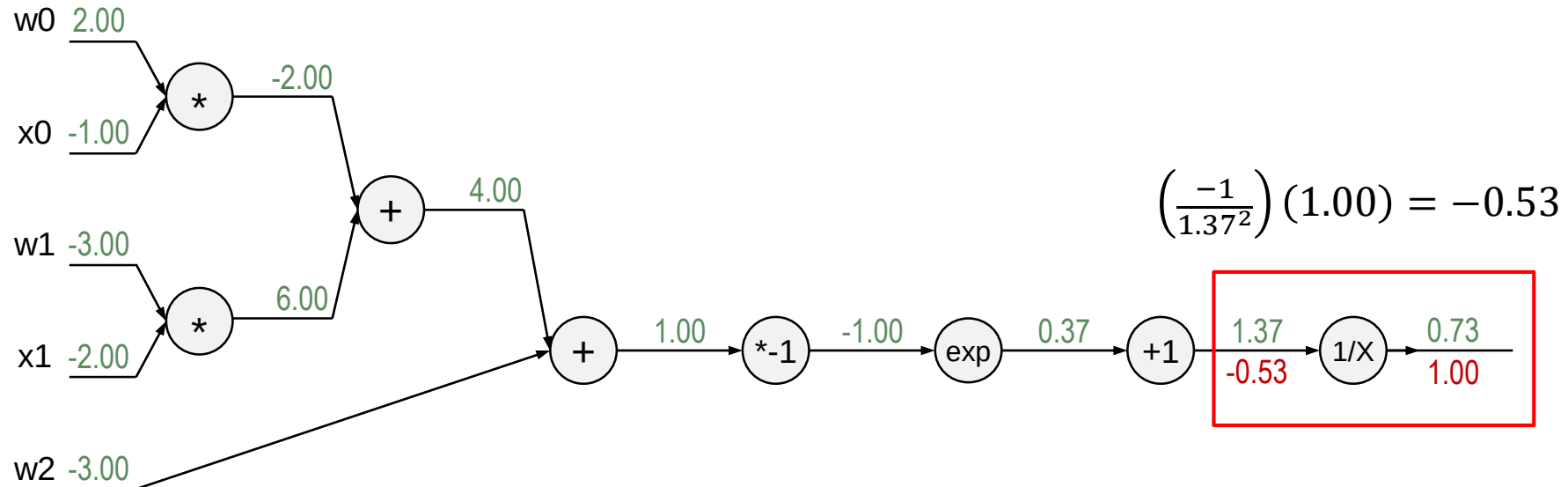
۱۹



$f(x) = e^x$	$\rightarrow$	$\frac{df}{dx} = e^x$		$f(x) = \frac{1}{x}$	$\rightarrow$	$\frac{df}{dx} = -1/x^2$
$f_a(x) = ax$	$\rightarrow$	$\frac{df}{dx} = a$		$f_c(x) = x + c$	$\rightarrow$	$\frac{df}{dx} = 1$

گراف محاسباتی: یک مثال دیگر

۲۰



$$f(x) = e^x \quad \rightarrow \quad \frac{df}{dx} = e^x$$

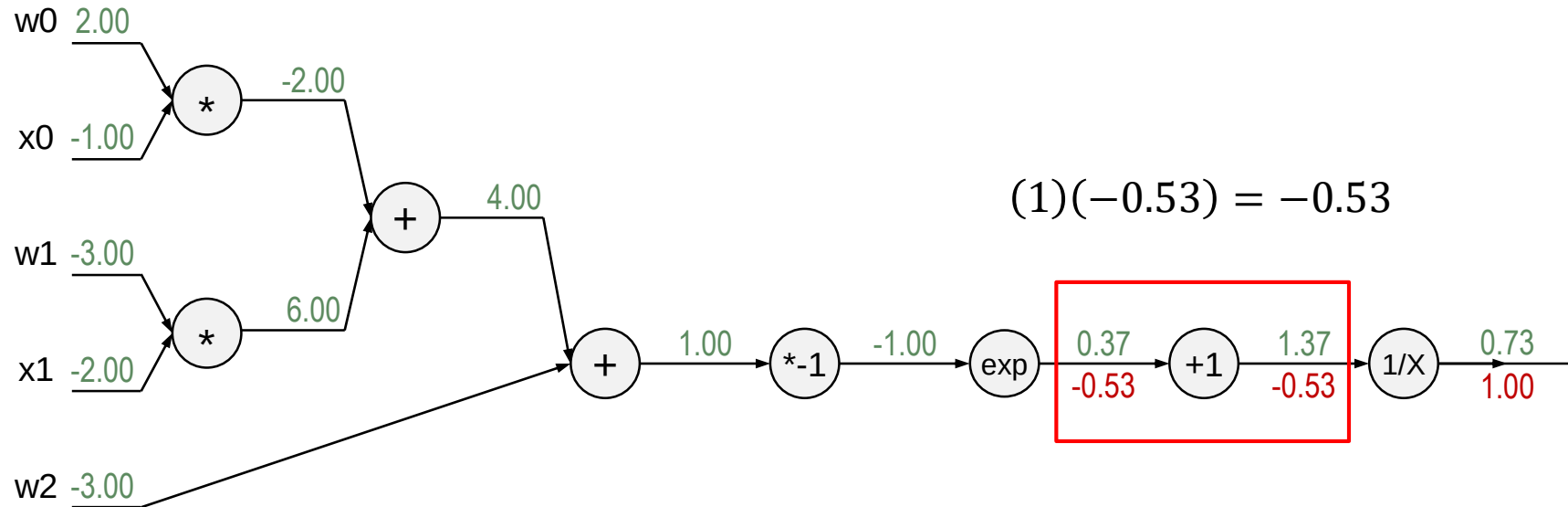
$$f_a(x) = ax \quad \rightarrow \quad \frac{df}{dx} = a$$

$$f(x) = \frac{1}{x} \quad \rightarrow \quad \frac{df}{dx} = -1/x^2$$

$$f_c(x) = x + c \quad \rightarrow \quad \frac{df}{dx} = 1$$

گراف محاسباتی: یک مثال دیگر

۲۱



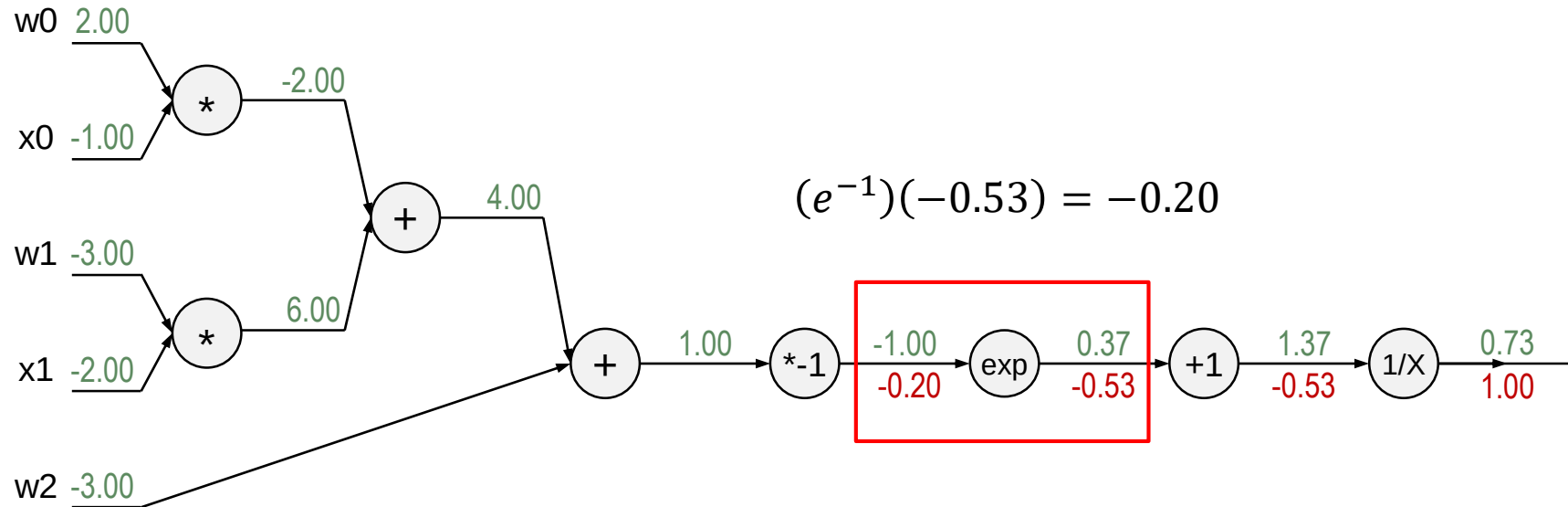
$$f(x) = e^x \rightarrow \frac{df}{dx} = e^x$$

$$f_a(x) = ax \rightarrow \frac{df}{dx} = a$$

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

$$f_c(x) = x + c \rightarrow \frac{df}{dx} = 1$$

گراف محاسباتی: یک مثال دیگر



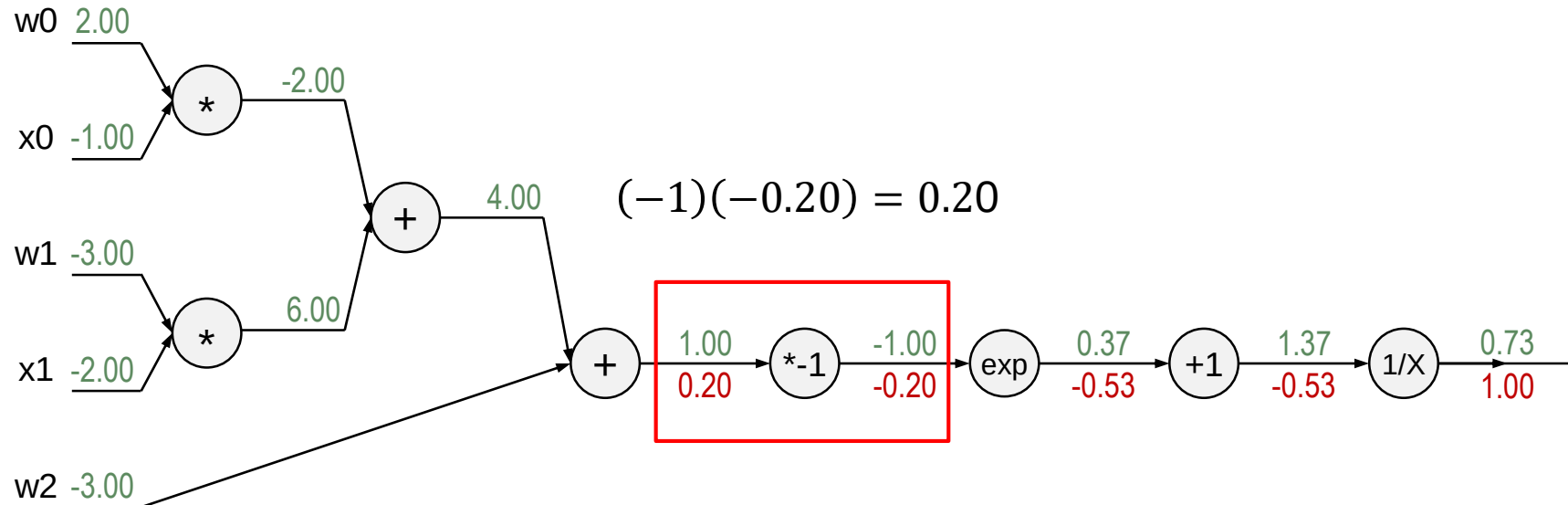
$$f(x) = e^x \rightarrow \frac{df}{dx} = e^x$$

$$f_a(x) = ax \rightarrow \frac{df}{dx} = a$$

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

$$f_c(x) = x + c \rightarrow \frac{df}{dx} = 1$$

گراف محاسباتی: یک مثال دیگر



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

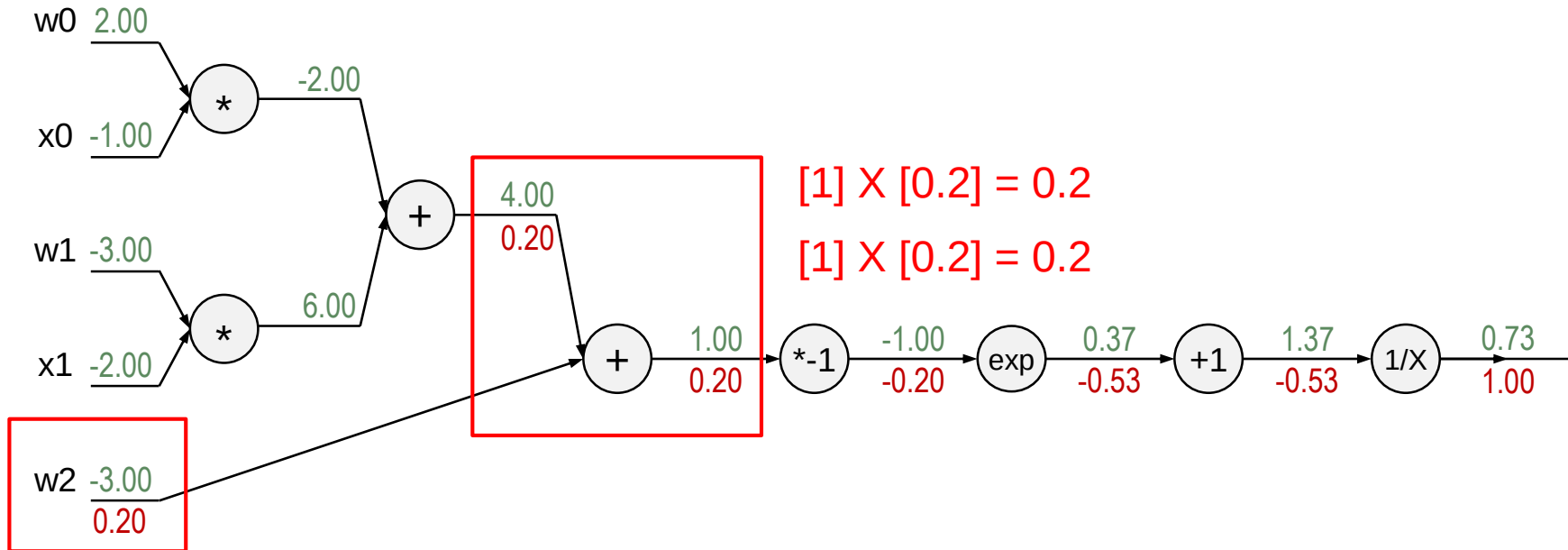
$$f_a(x) = ax \rightarrow \frac{df}{dx} = a$$

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

$$f_c(x) = x + c \rightarrow \frac{df}{dx} = 1$$

گراف محاسباتی: یک مثال دیگر

۲۴



$$[1] \times [0.2] = 0.2$$

$$[1] \times [0.2] = 0.2$$

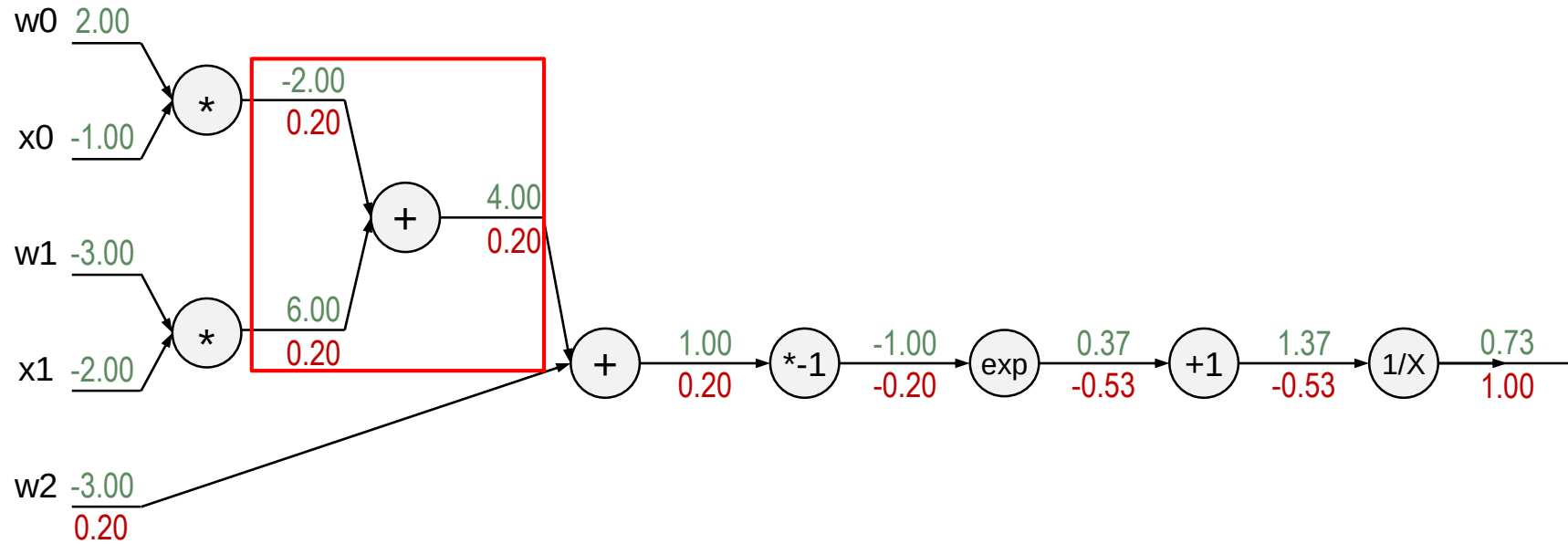
$$f(x) = e^x \rightarrow \frac{df}{dx} = e^x$$

$$f_a(x) = ax \rightarrow \frac{df}{dx} = a$$

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

$$f_c(x) = x + c \rightarrow \frac{df}{dx} = 1$$

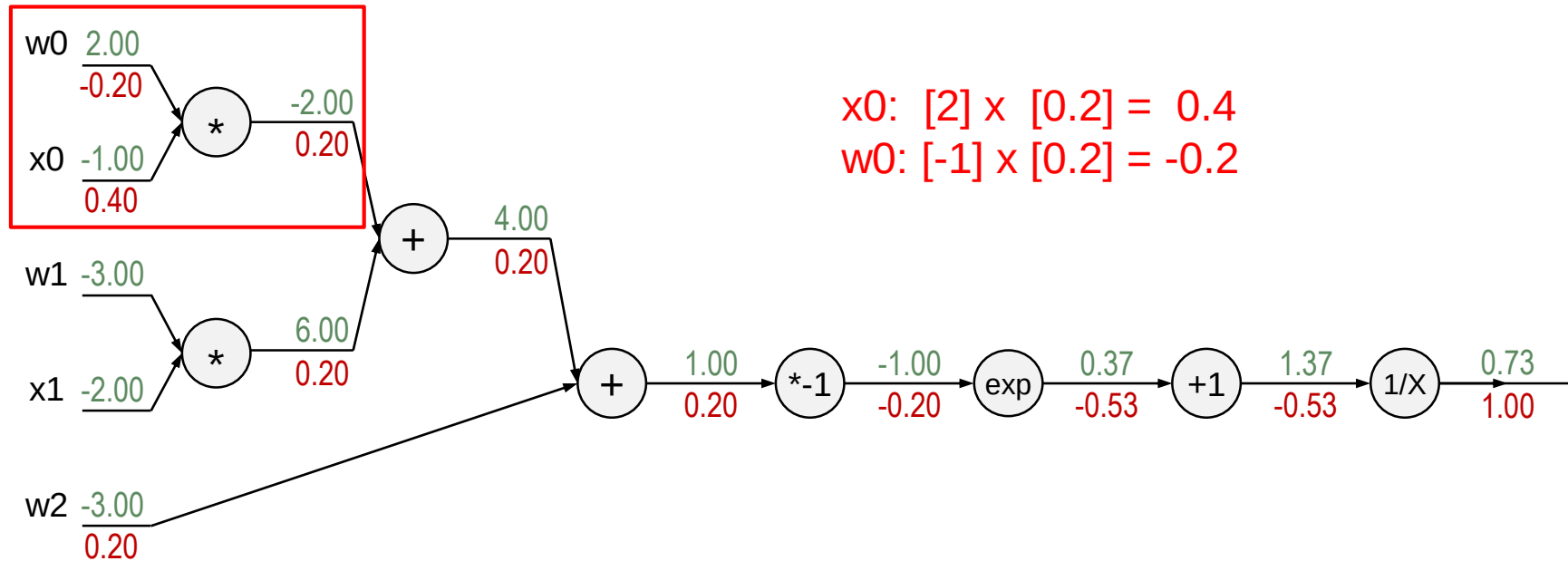
گراف محاسباتی: یک مثال دیگر



$f(x) = e^x$	$\rightarrow$	$\frac{df}{dx} = e^x$	$f(x) = \frac{1}{x}$	$\rightarrow$	$\frac{df}{dx} = -1/x^2$
$f_a(x) = ax$	$\rightarrow$	$\frac{df}{dx} = a$	$f_c(x) = x + c$	$\rightarrow$	$\frac{df}{dx} = 1$

گراف محاسباتی: یک مثال دیگر

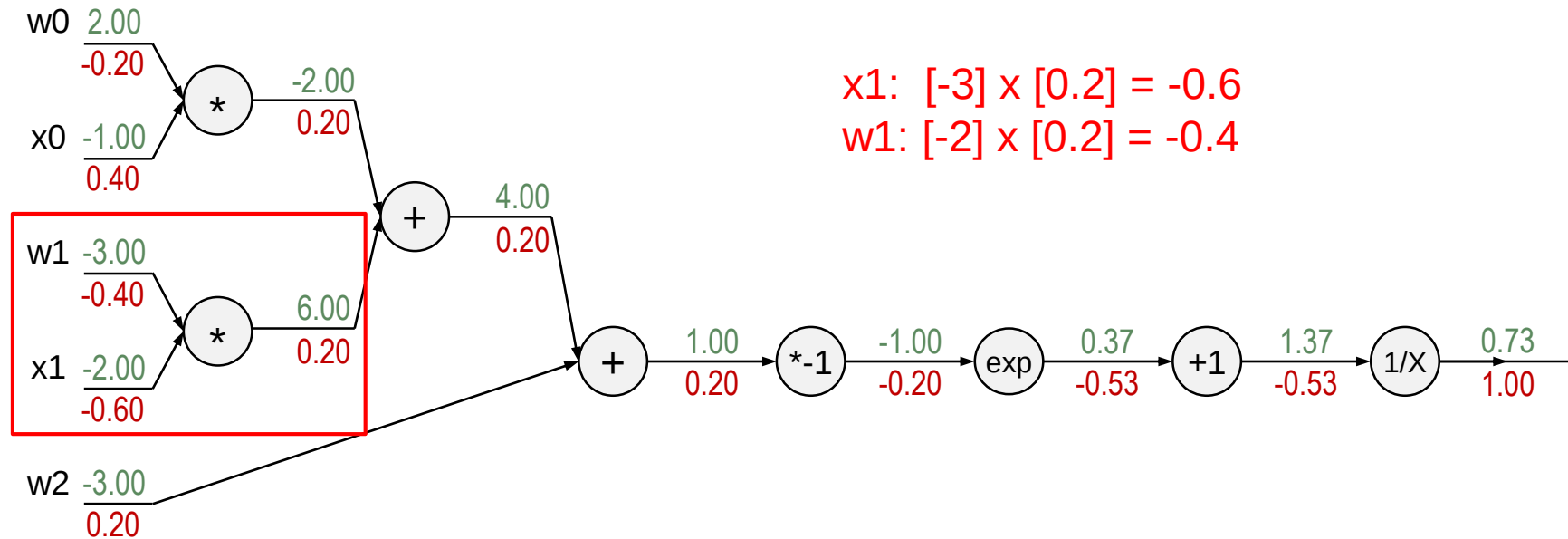
۲۶



$f(x) = e^x$	$\rightarrow$	$\frac{df}{dx} = e^x$		$f(x) = \frac{1}{x}$	$\rightarrow$	$\frac{df}{dx} = -1/x^2$
$f_a(x) = ax$	$\rightarrow$	$\frac{df}{dx} = a$		$f_c(x) = x + c$	$\rightarrow$	$\frac{df}{dx} = 1$

گراف محاسباتی: یک مثال دیگر

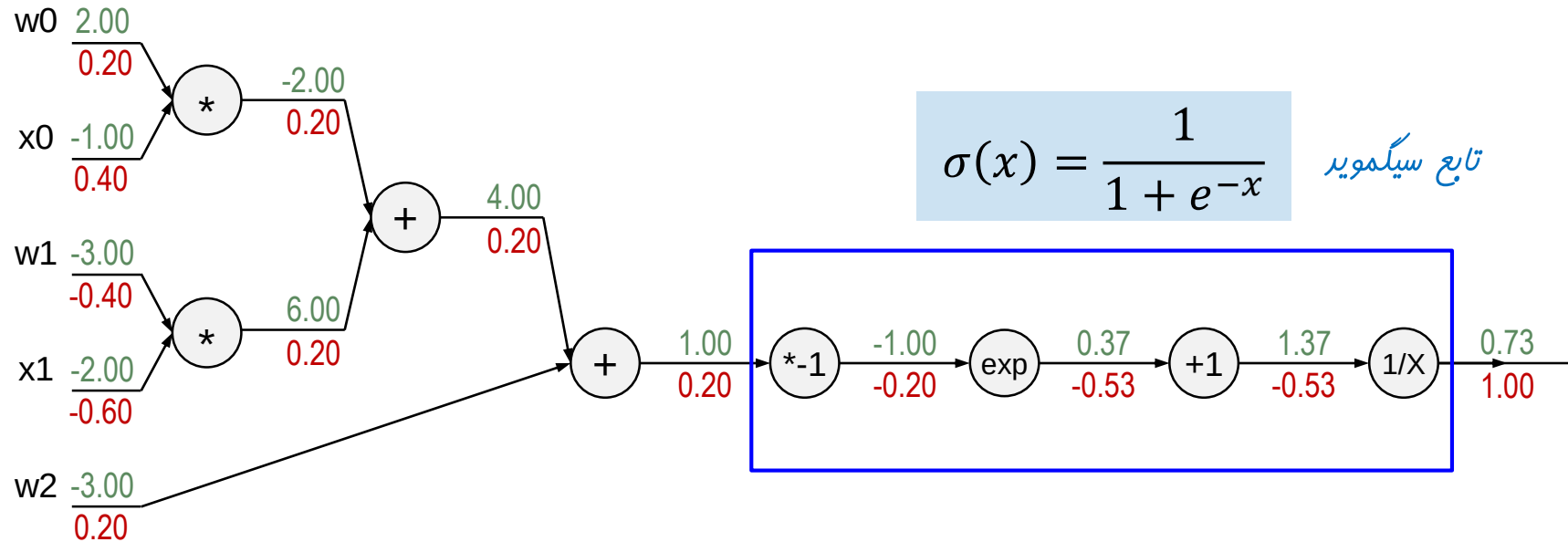
۲۷



$f(x) = e^x$	$\rightarrow$	$\frac{df}{dx} = e^x$		$f(x) = \frac{1}{x}$	$\rightarrow$	$\frac{df}{dx} = -1/x^2$
$f_a(x) = ax$	$\rightarrow$	$\frac{df}{dx} = a$		$f_c(x) = x + c$	$\rightarrow$	$\frac{df}{dx} = 1$

گراف محاسباتی: یک مثال دیگر

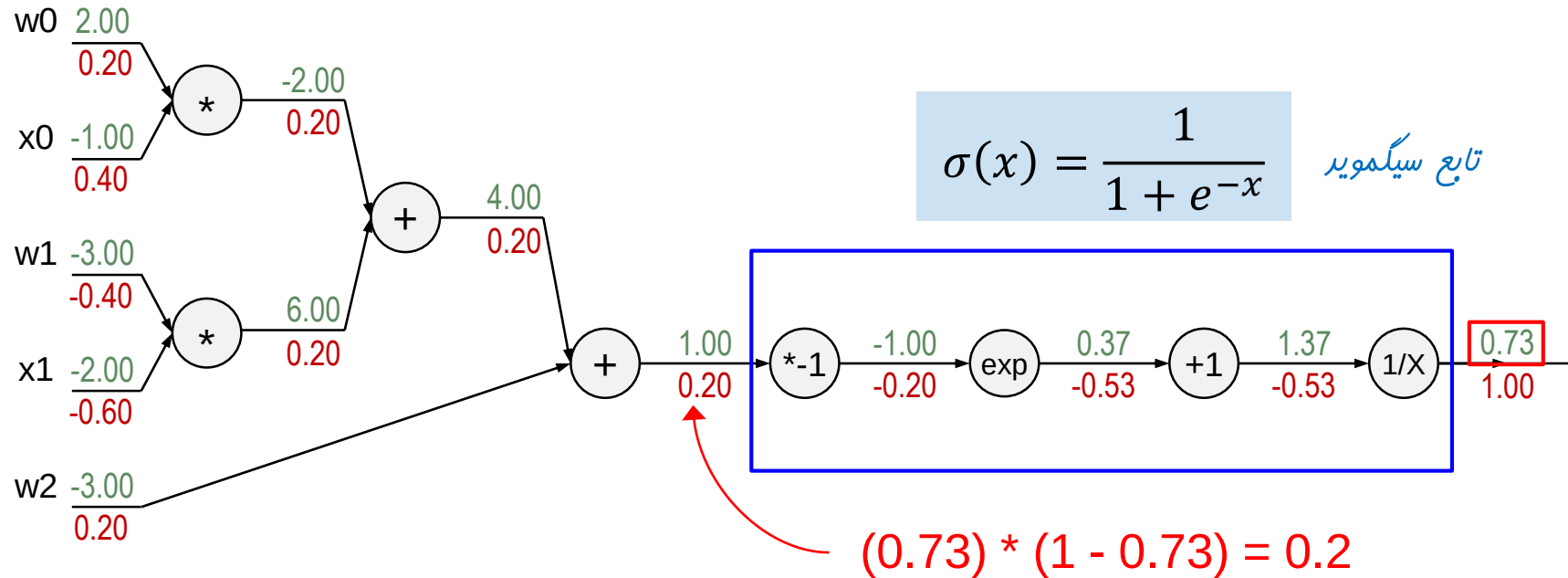
۲۸



$$\frac{d\sigma(x)}{dx} = \frac{e^{-x}}{(1 + e^{-x})^2} = \left(\frac{1}{1 + e^{-x}} \right) \left(\frac{1 + e^{-x} - 1}{1 + e^{-x}} \right) = \sigma(x)(1 - \sigma(x))$$

گراف محاسباتی: یک مثال دیگر

۲۹

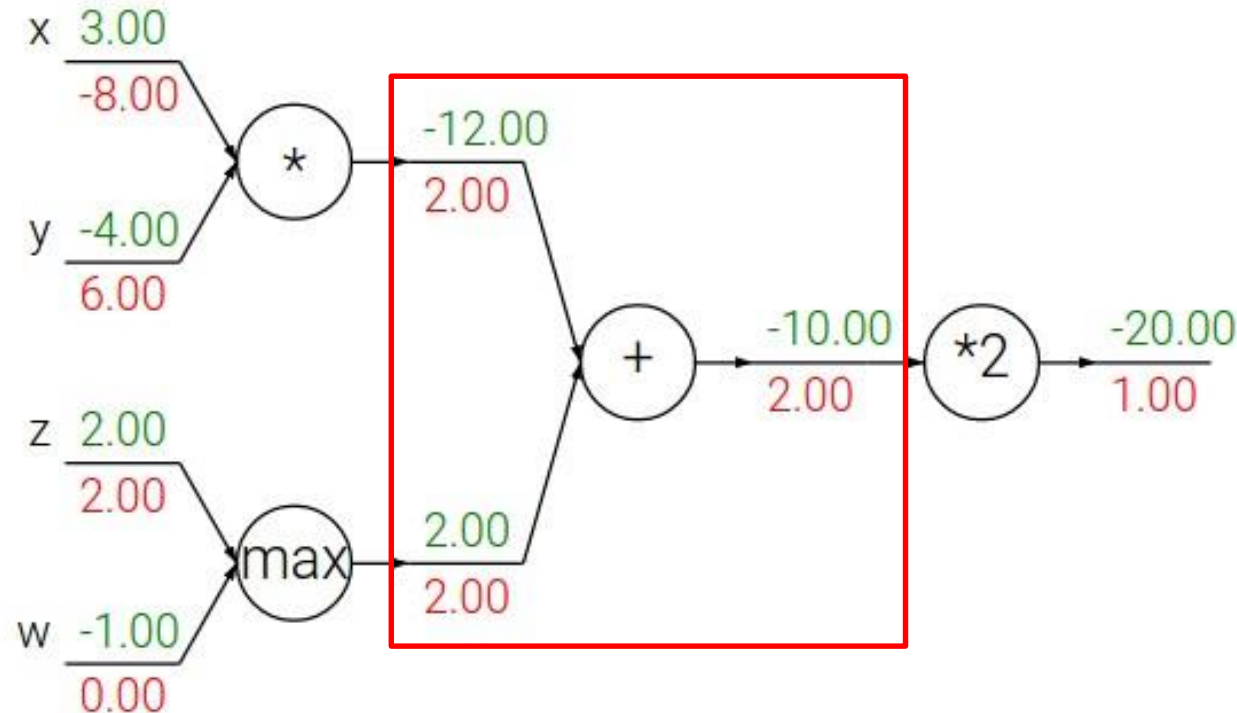


$$\frac{d\sigma(x)}{dx} = \frac{e^{-x}}{(1 + e^{-x})^2} = \left(\frac{1}{1 + e^{-x}} \right) \left(\frac{1 + e^{-x} - 1}{1 + e^{-x}} \right) = \sigma(x)(1 - \sigma(x))$$

چند الگو در محاسبات رو به عقب

۳۰

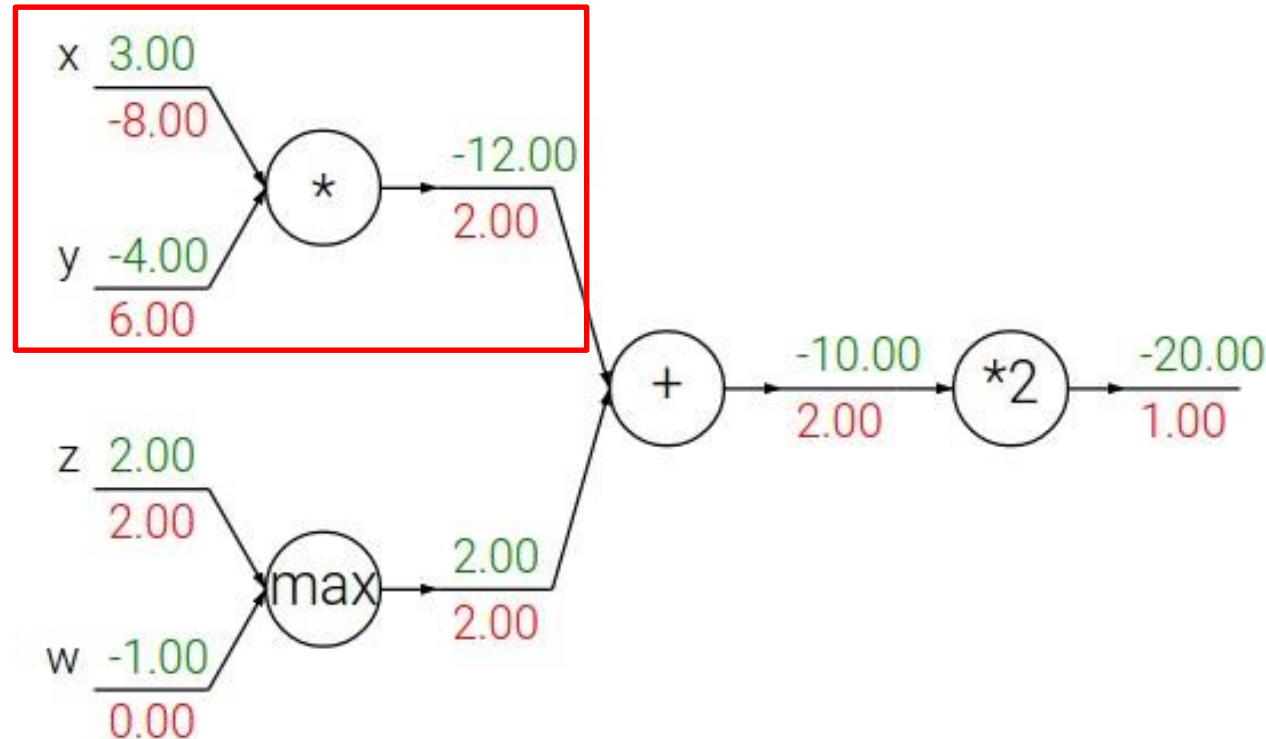
□ عمل جمع. پخش کننده گرادینان!



چند الگو در محاسبات رو به عقب

۳۱

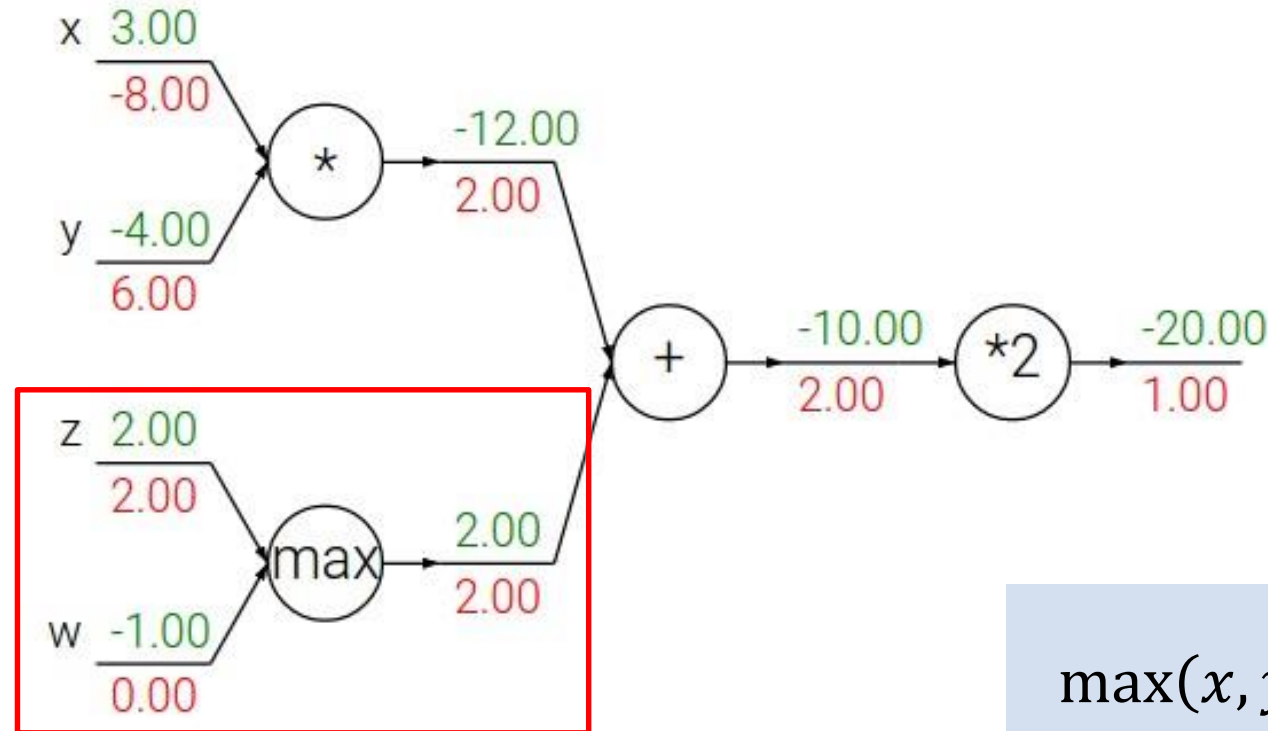
□ عمل ضرب. جابه‌جا کننده گرادینان!



چند الگو در محاسبات رو به عقب

۳۲

□ عمل بیشینه‌یابی. مسیریابی گرادیان!

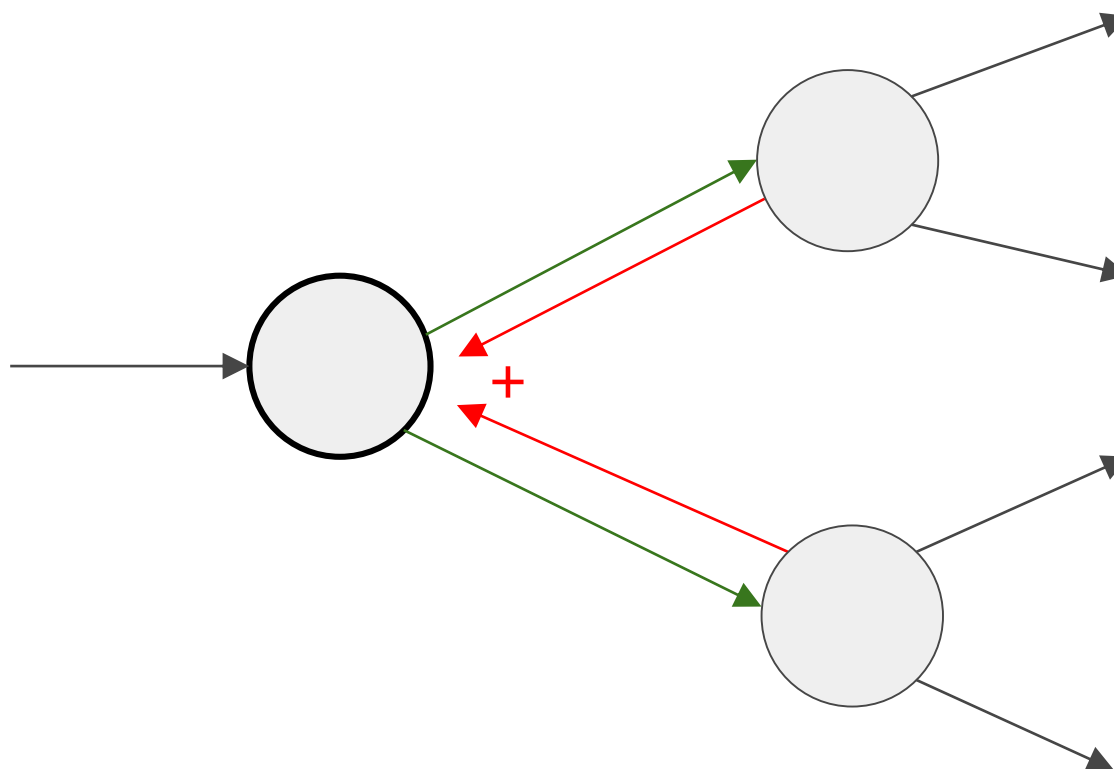


$$\max(x, y) = \begin{cases} x, & x \geq y \\ y, & y > x \end{cases}$$

چند الگو در محاسبات رو به عقب

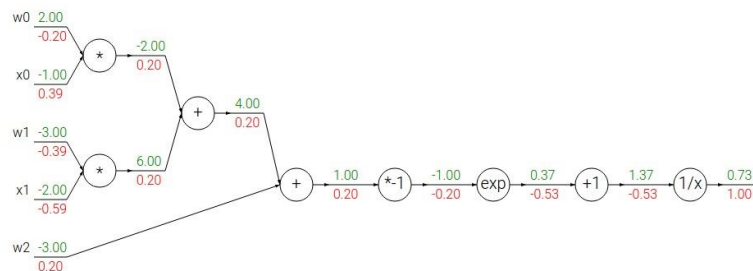
۳۳

□ جمع شدن گرادیان‌ها در انشعاب‌ها.



پیاده‌سازی: محاسبات رو به جلو و رو به عقب

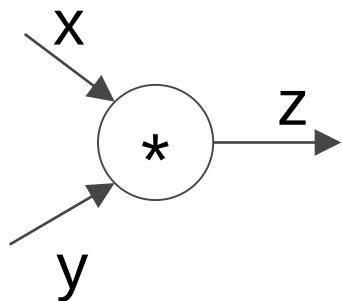
۳۴



```
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
```

پیاده‌سازی: محاسبات رو به جلو و رو به عقب

۳۵



```
class MultiplyGate(object):
```

```
    def forward(x,y):
```

```
        z = x*y
```

```
        return z
```

```
    def backward(dz):
```

```
        # dx = ... #todo
```

```
        # dy = ... #todo
```

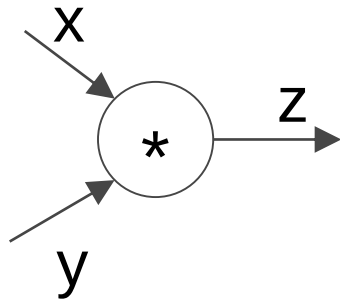
```
        return [dx, dy]
```

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

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

پیاده‌سازی: محاسبات رو به جلو و رو به عقب

۳۶

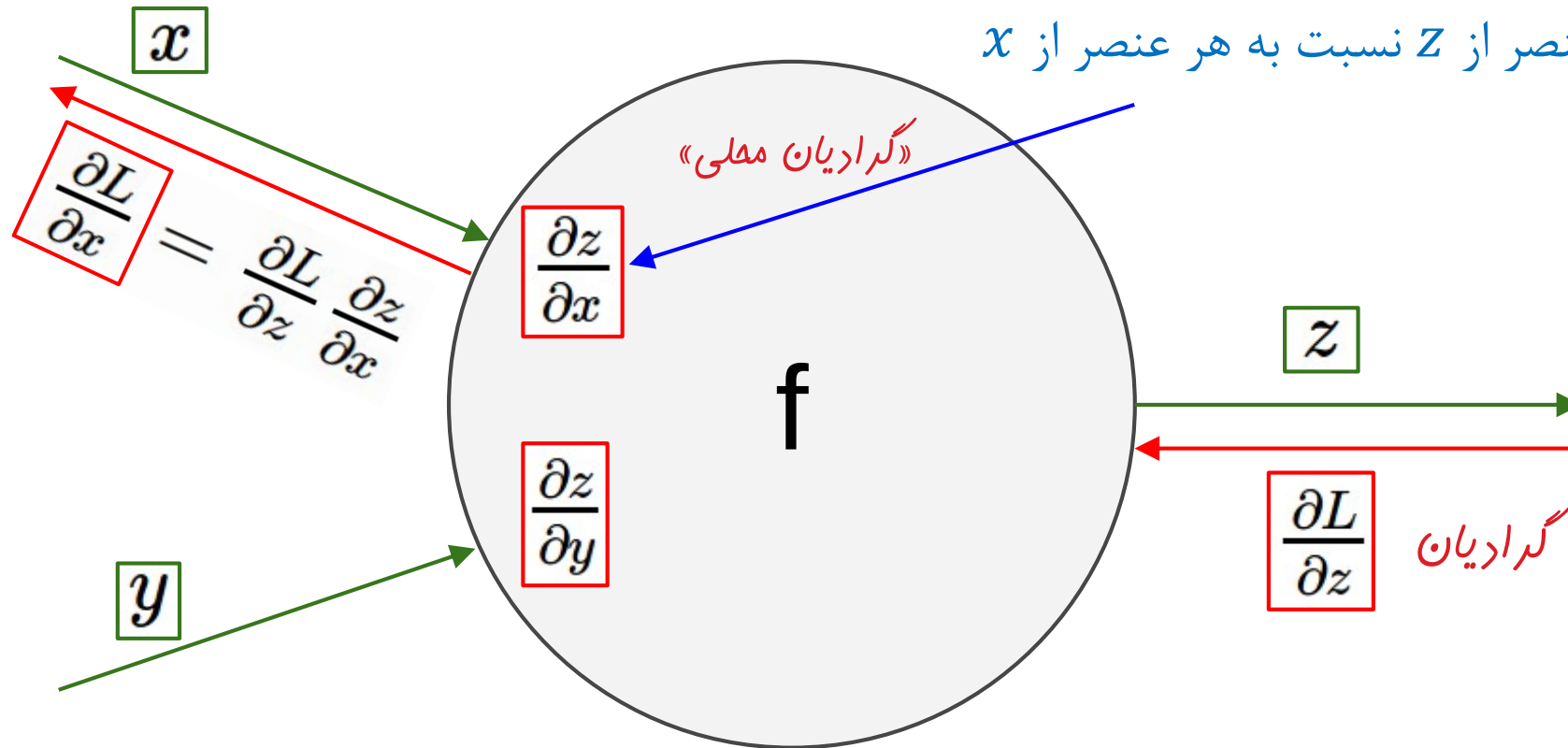


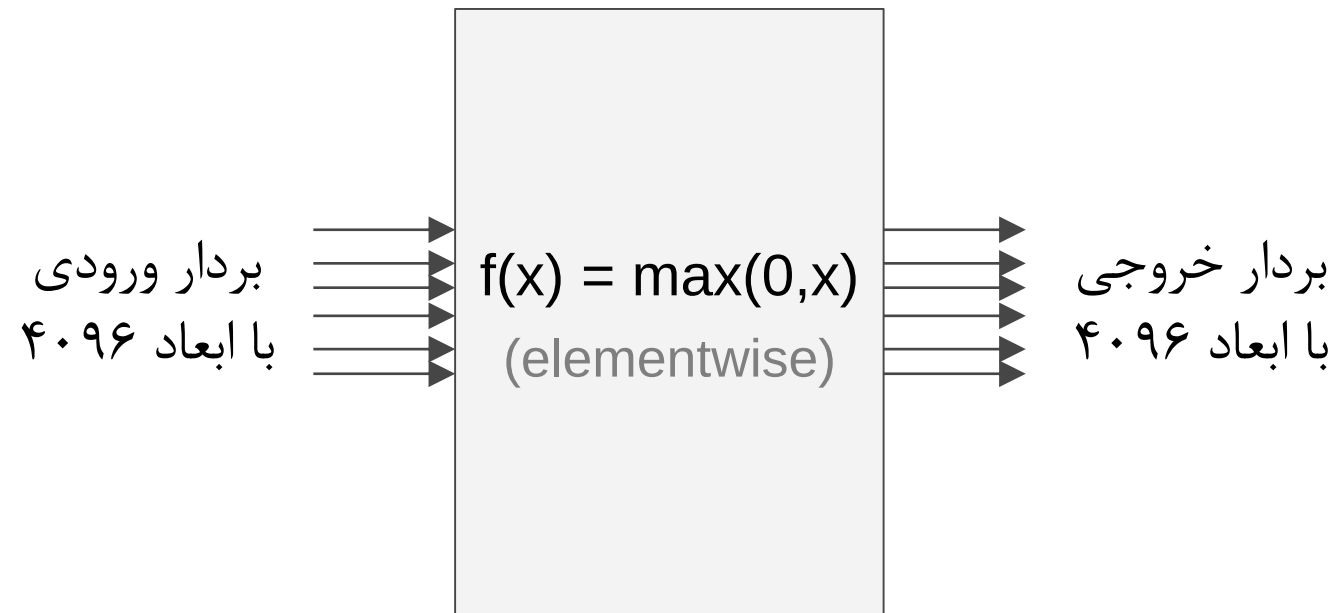
```
class MultiplyGate(object):  
    def forward(x,y):  
        z = x*y  
        self.x = x # must keep these around!  
        self.y = y  
        return z  
    def backward(dz):  
        dx = self.y * dz # [dz/dx * dL/dz]  
        dy = self.x * dz # [dz/dy * dL/dz]  
        return [dx, dy]
```

گرادیان برای بردارها

۳۷

اکنون این گرادیان محلی یک ماتریس ژاکوبی است.
یعنی مشتق هر عنصر از z نسبت به هر عنصر از x



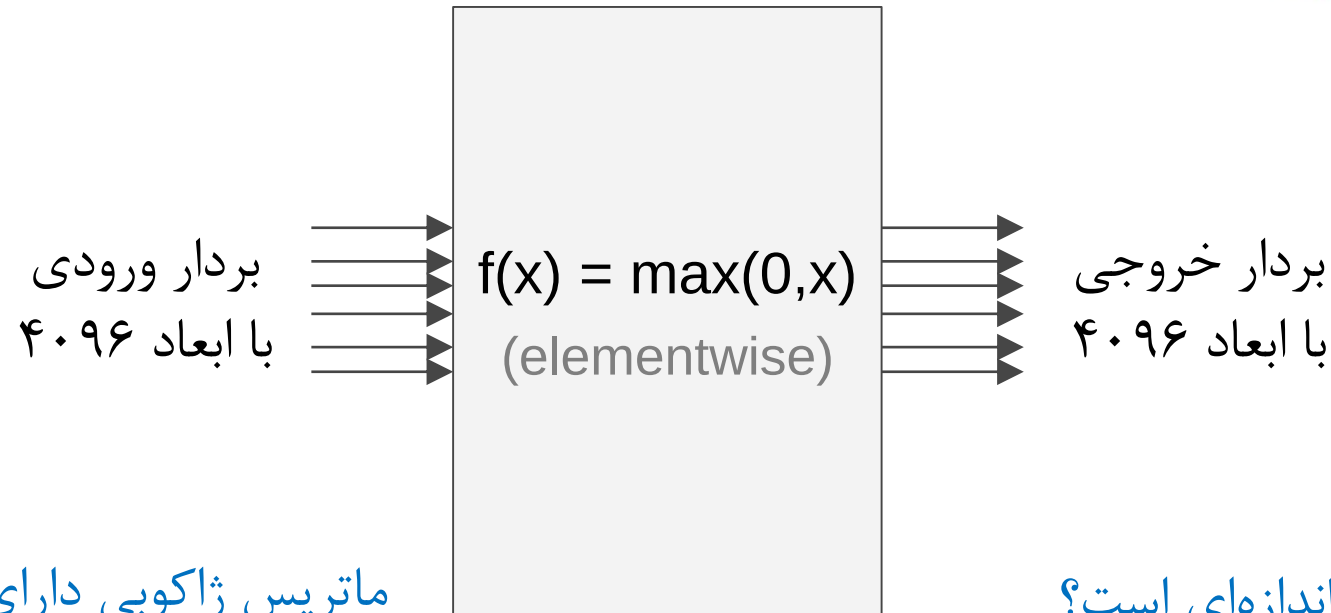


عملیات برداری

۳۹

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

ماتریس ژاکوبی



ماتریس ژاکوبی دارای چه شکلی است؟

یک ماتریس قطری که برخی از عناصر قطر اصلی دارای مقدار یک و برخی دیگر دارای مقدار صفر هستند.

ماتریس ژاکوبی دارای چه اندازه‌ای است؟

(4096×4096)

تمرین: پیاده‌سازی توابع هزینه

۴۰

```
# receive W (weights), X (data)
```

```
# forward pass (we have 8 lines)
```

```
scores = #...
```

```
margins = #...
```

```
data_loss = #...
```

```
reg_loss = #...
```

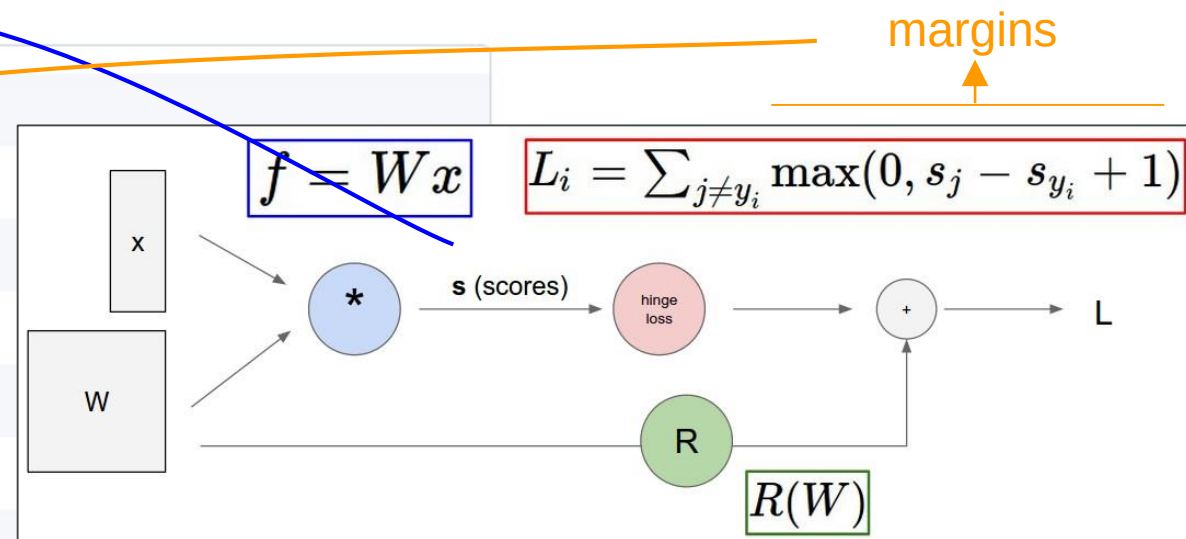
```
loss = data_loss + reg_loss
```

```
# backward pass (we have 5 lines)
```

```
dmargins = # ... (optionally, we go direct to dscores)
```

```
dscores = #...
```

```
dW = #...
```



□ تعداد پارامترها در یک شبکه عصبی می‌تواند بسیار زیاد باشد:

□ نوشتن رابطه مربوط به گرادیان تمام پارامترها به صورت دستی غیر ممکن است!

□ پس‌انتشار. به کار بردن **قاعده زنجیری** به صورت بازگشتی در طول یک گراف محاسباتی به منظور محاسبه گرادیان تابع هزینه نسبت به پارامترها، ورودی‌ها و مقادیر میانی.

□ **گراف محاسباتی**. یک ساختار گرافی که هر گره آن **محاسبات رو به جلو** و **محاسبات رو به عقب** را پیاده‌سازی می‌کند.

□ **محاسبات رو به جلو**. محاسبه نتیجه یک عمل و ذخیره مقادیر میانی مورد نیاز برای محاسبه گرادیان.

□ **محاسبات رو به عقب**. استفاده از قاعده زنجیری به منظور محاسبه گرادیان تابع هزینه نسبت به ورودی‌ها.