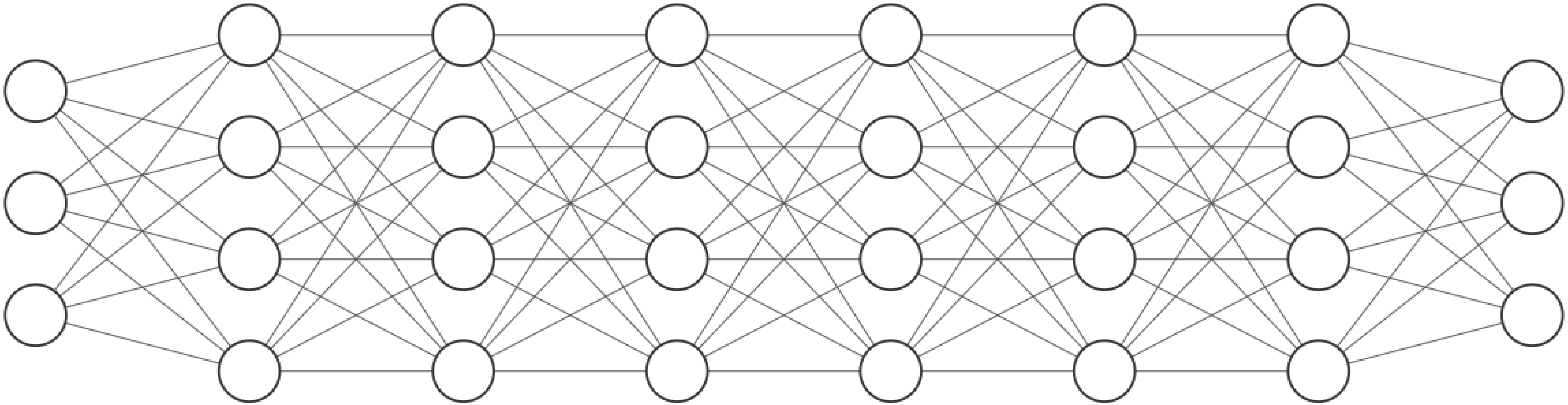
Grafos Computacionais e Diferenciação Automática

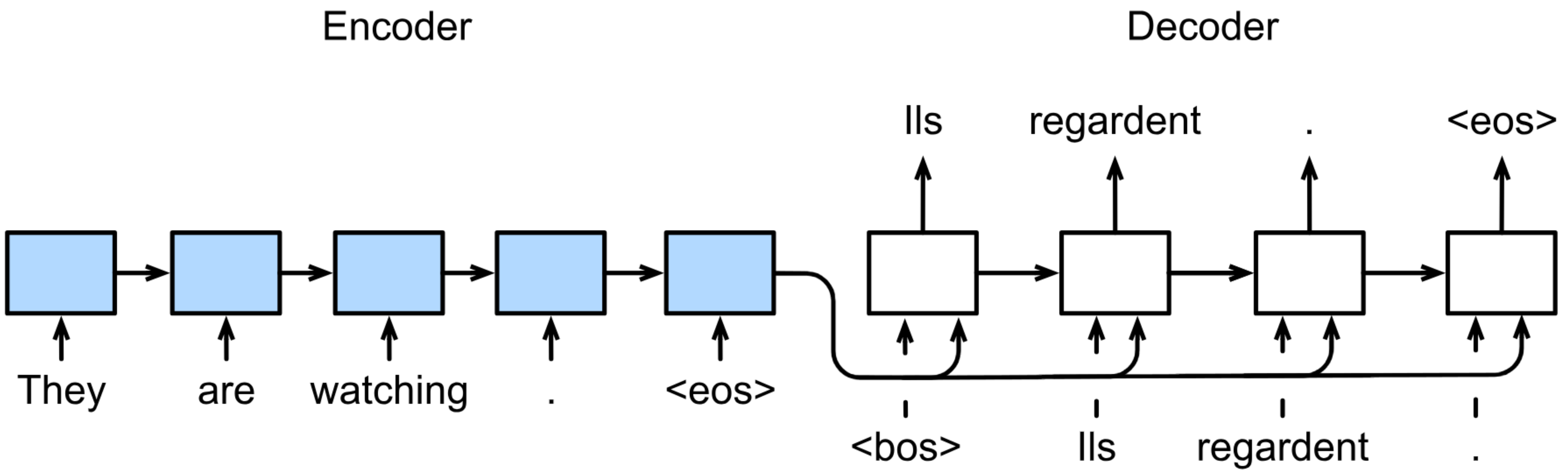
Introdução

Computar gradientes manualmente é desafiador para cadeias longas de composição de funções



Introdução

Alguns métodos realizam números variados de operações dependendo do tamanho da entrada



Introdução

Seria bom se só precisássemos calcular derivadas uma vez e depois pudéssemos reusar as funções (e derivadas) como componentes.

arcocotangente hiperbólica

$$f(x) = \operatorname{arccoth}(x) \Longrightarrow f'(x) = \frac{1}{1 - x^2}$$

Diferenciação Automática

- Considere valores calculados pela aplicação sucessiva de operações matemáticas (e.g., +, -, *, cos, exp, etc.).
- Métodos de diferenciação automática (*automatic differentiation*) armazenam operações realizadas e seus resultados de forma que seja possível calcular as derivadas parciais dos argumentos automaticamente.

Grafos Computacionais

- São as representações mais comuns de sequências de operações em sistemas modernos de diferenciação automática.
- Valores e operações (junto com seus resultados) são armazenados como nós em um grafo.

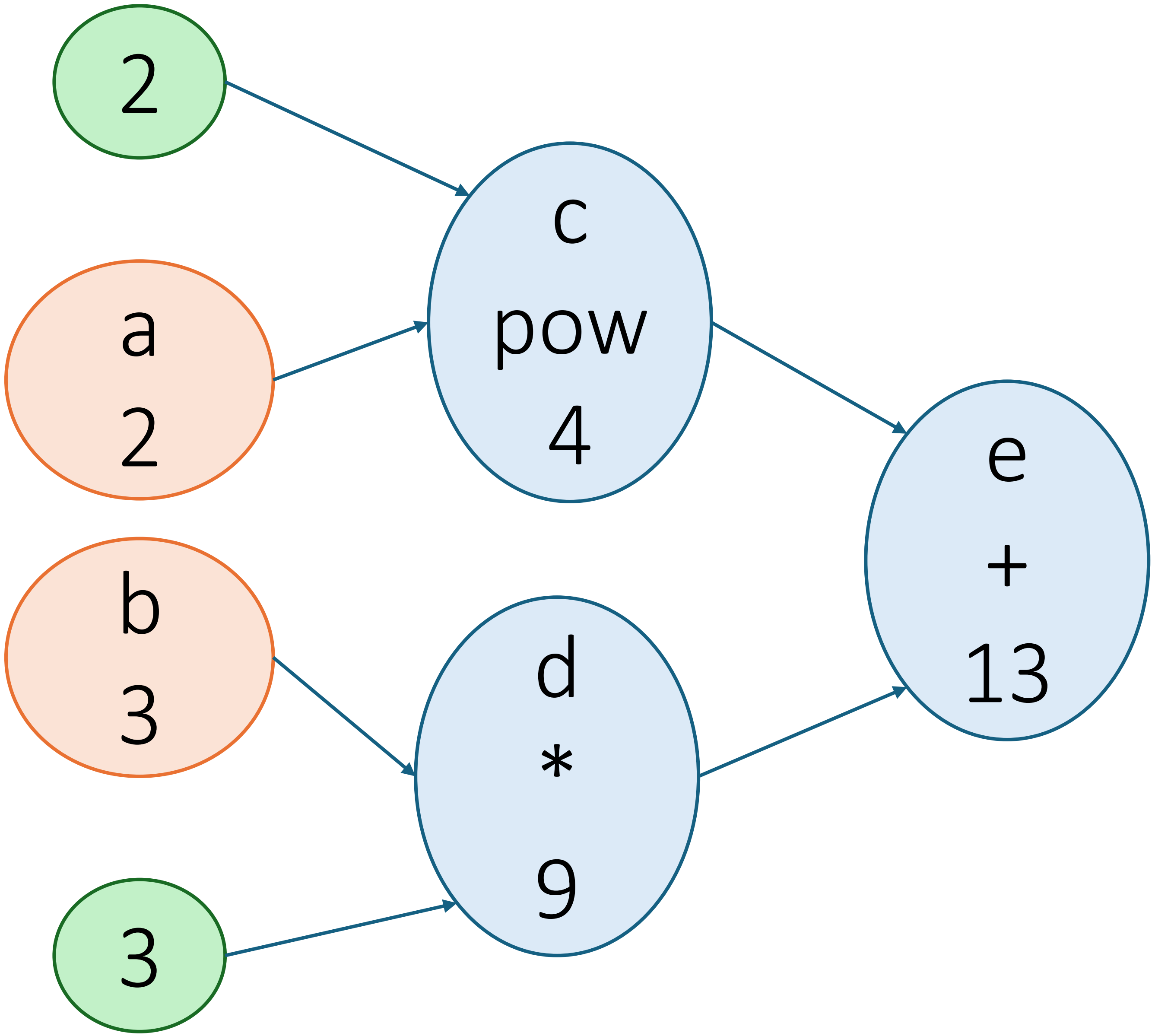
$a = 2$

$b = 3$

$c = pow(a, 2)$

$d = 3 * b$

$e = c + d$



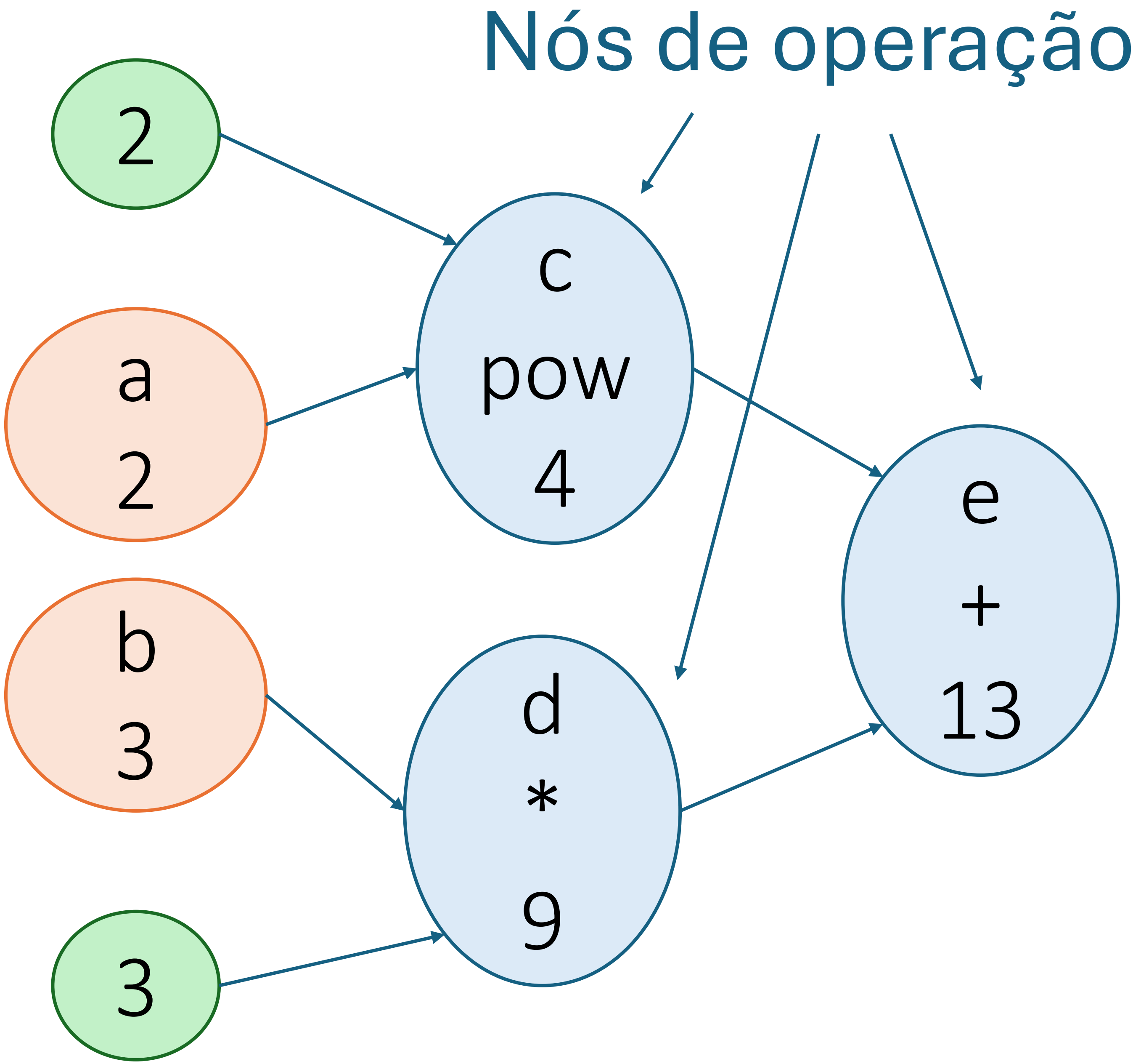
$a = 2$

$b = 3$

$c = a ** 2$

$d = 3 * b$

$e = c + d$



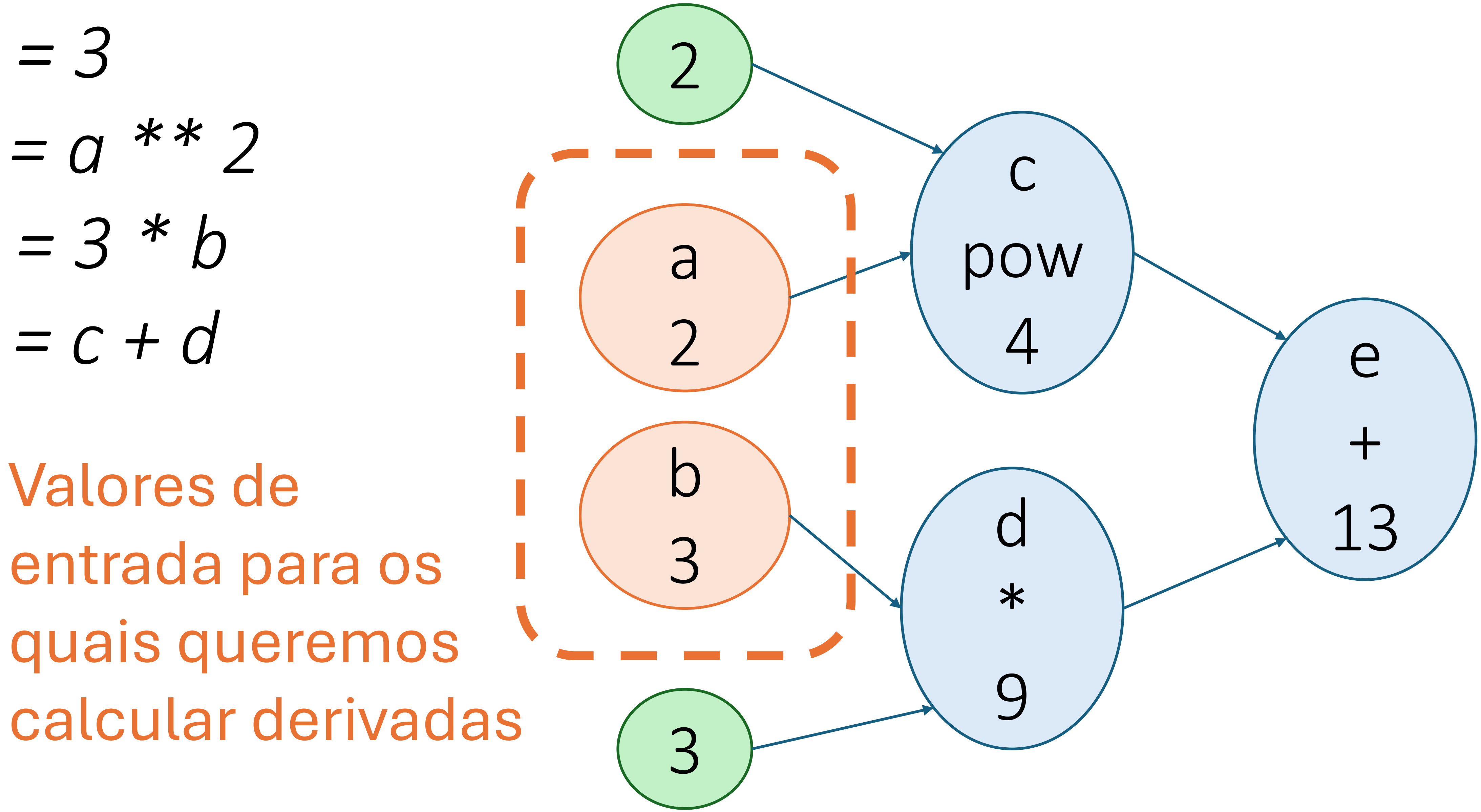
a = 2

b = 3

c = *a* ** 2

d = 3 * *b*

e = *c* + *d*



$a = 2$

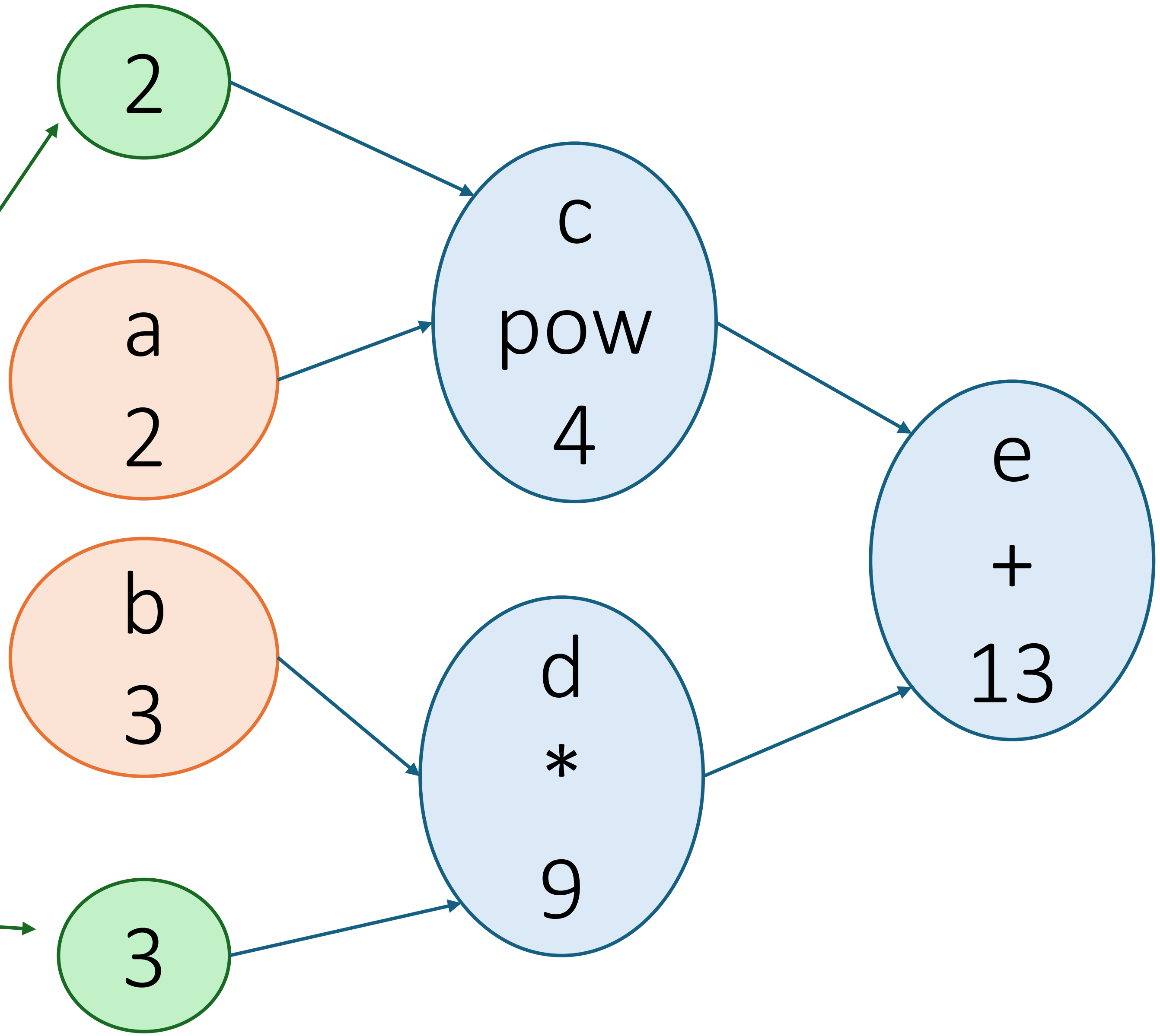
$b = 3$

$c = a ** 2$

$d = 3 * b$

$e = c + d$

Constantes para as quais não precisamos calcular derivadas



Diferenciação Automática com Grafos Computacionais

- Cada nó de operação armazena seus pais, isto é, os argumentos usados como entrada.
- Podemos caminhar do final para o começo usando regras de diferenciação (soma, produto, cadeia, etc.) para calcular derivadas parciais em relação às entradas.

Diferenciação Automática com Grafos Computacionais

- Cada nó de operação armazena seus pais, isto é, os argumentos usados como entrada.
- Cada operação* define c
- Com isto, podemos caminhar para trás pelas operações até chegar nos nós iniciais.

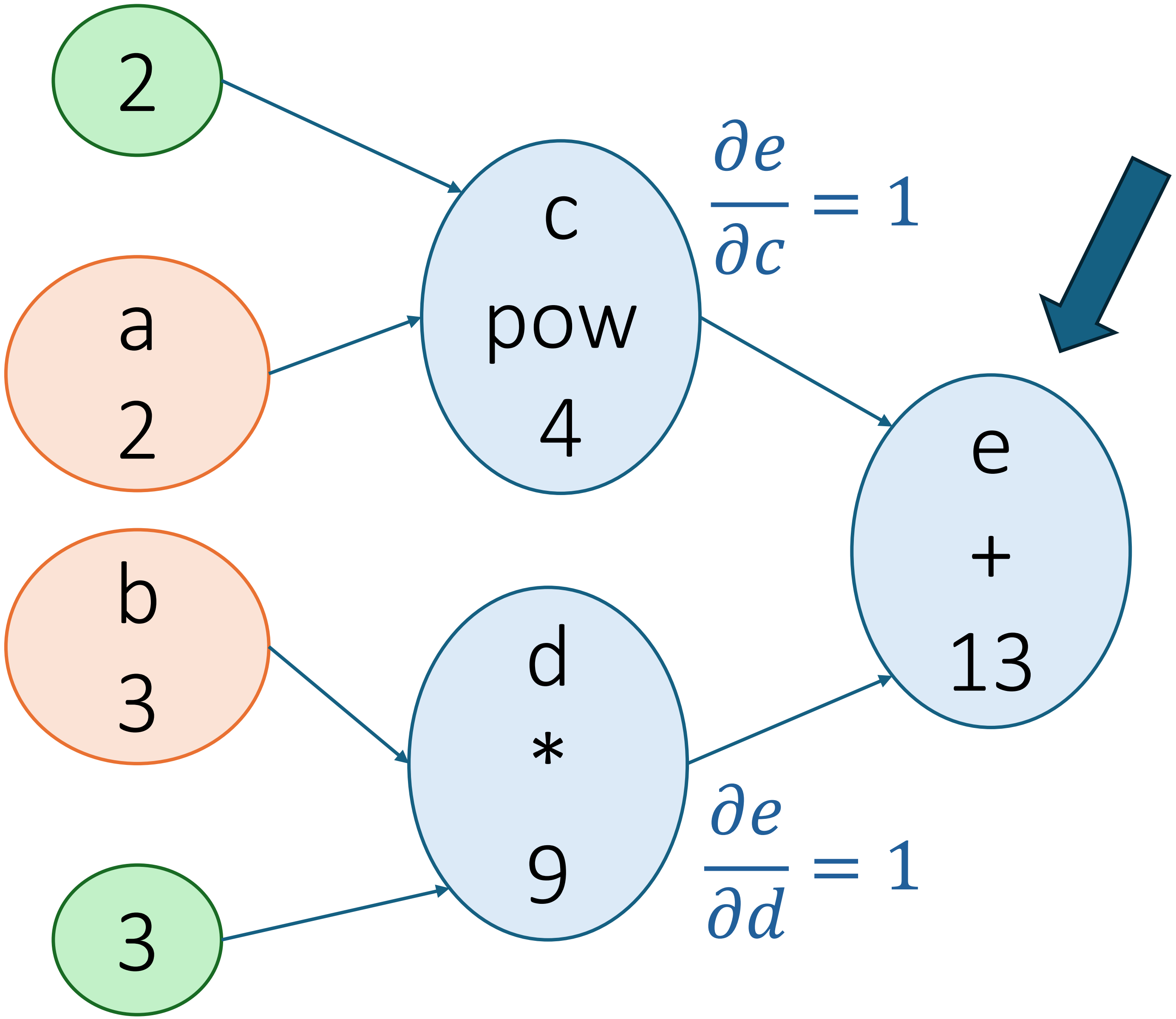
$a = 2$

$b = 3$

$c = a ** 2$

$d = 3 * b$

$e = c + d$



$a = 2$

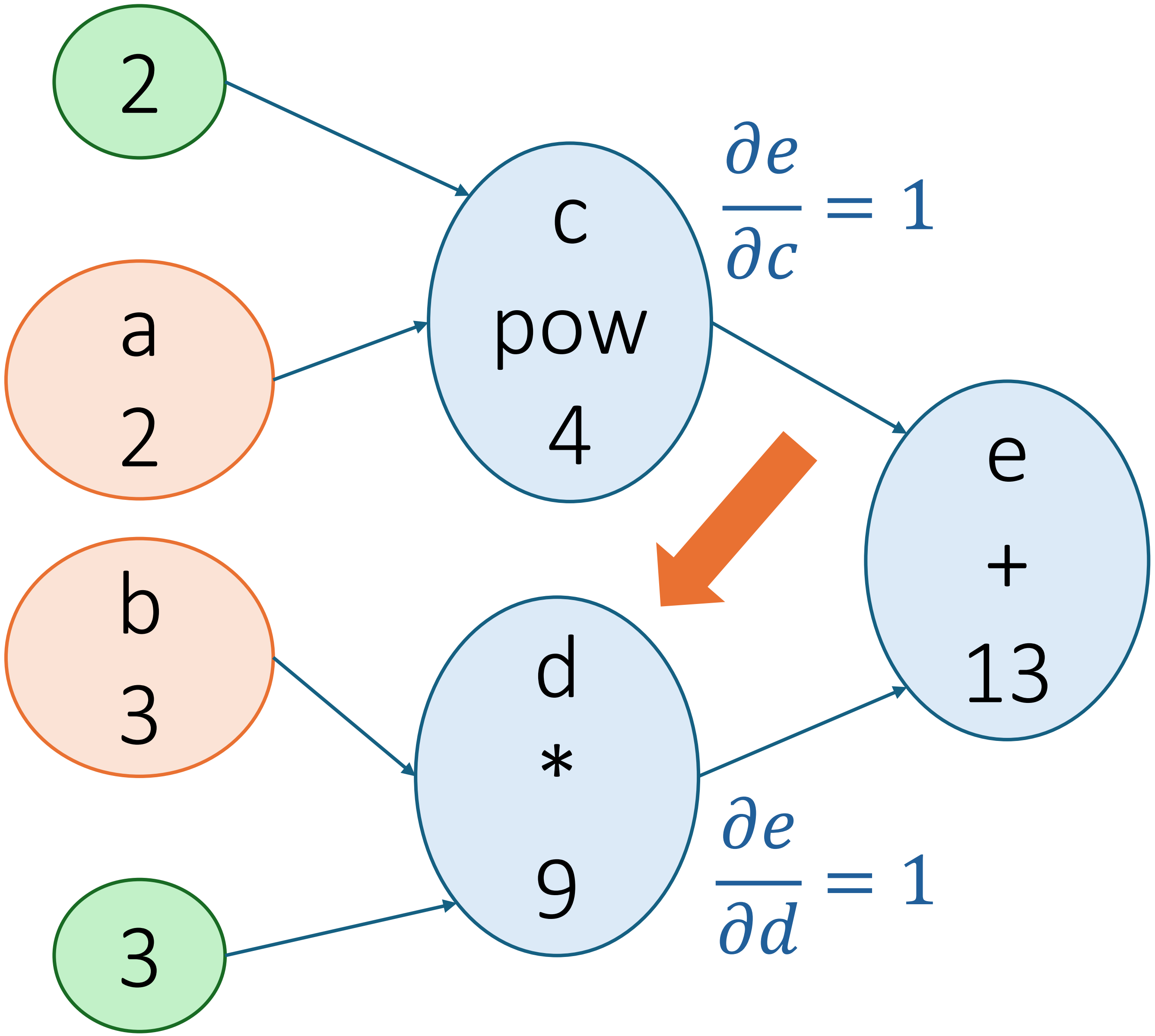
$b = 3$

$c = a ** 2$

$d = 3 * b$

$e = c + d$

$$\frac{\partial e}{\partial b} = \frac{\partial e}{\partial d}\frac{\partial d}{\partial b}$$



$a = 2$

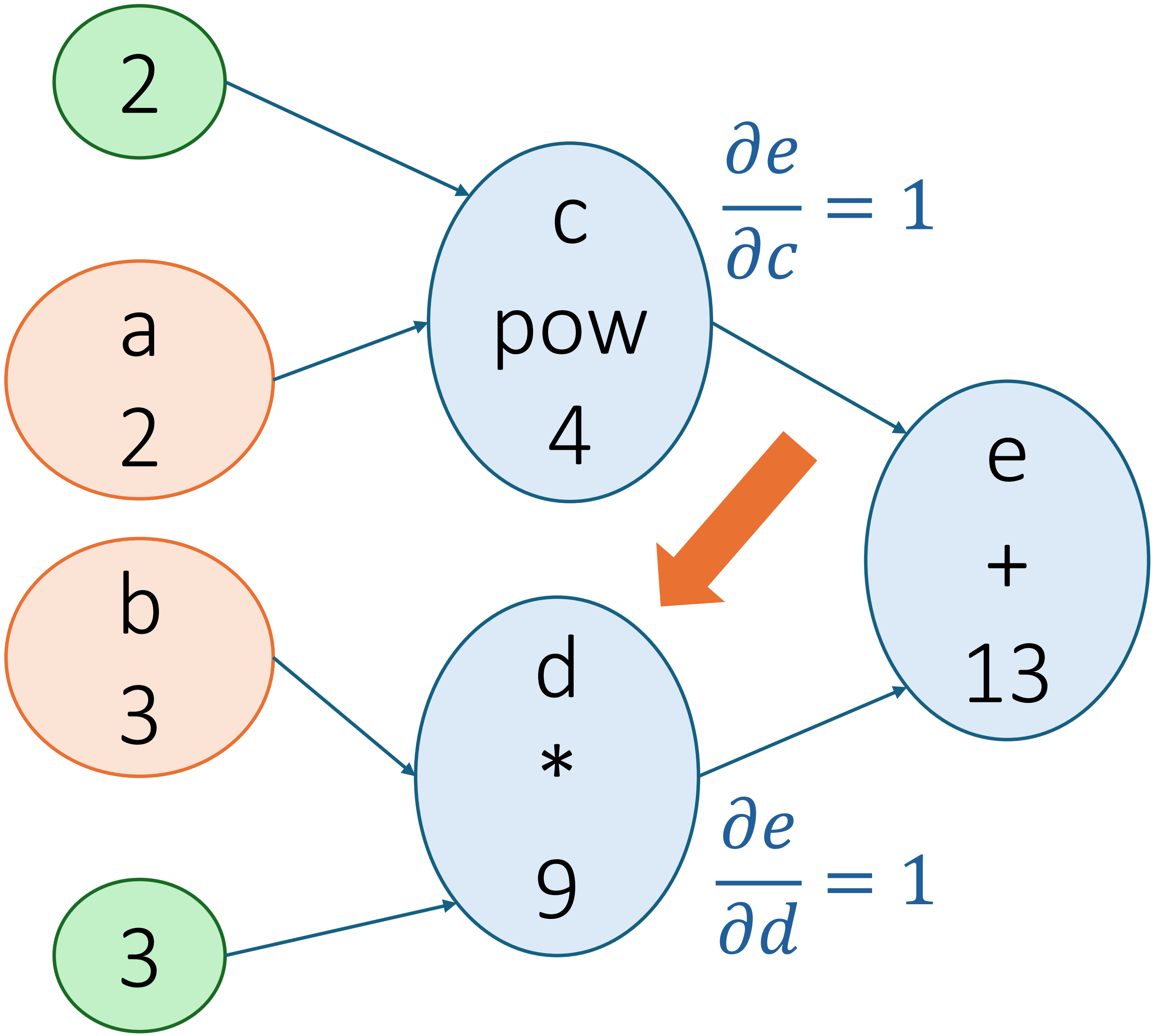
$b = 3$

$c = a ** 2$

$d = 3 * b$

$e = c + d$

$$\frac{\partial e}{\partial b} = 1 \frac{\partial d}{\partial b}$$



$a = 2$

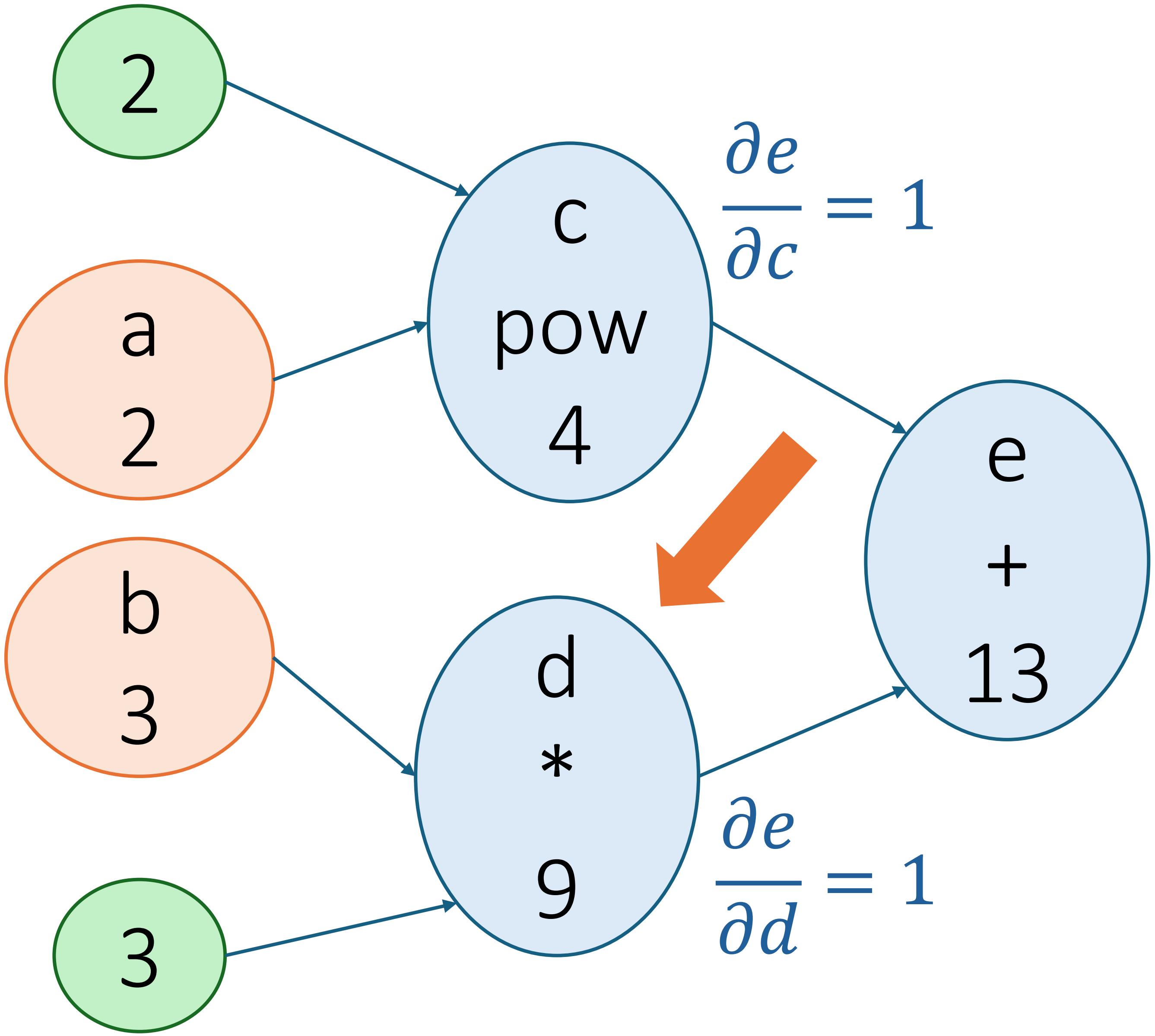
$b = 3$

$c = a ** 2$

$d = 3 * b$

$e = c + d$

$$\frac{\partial e}{\partial b} = 1 \times 3$$



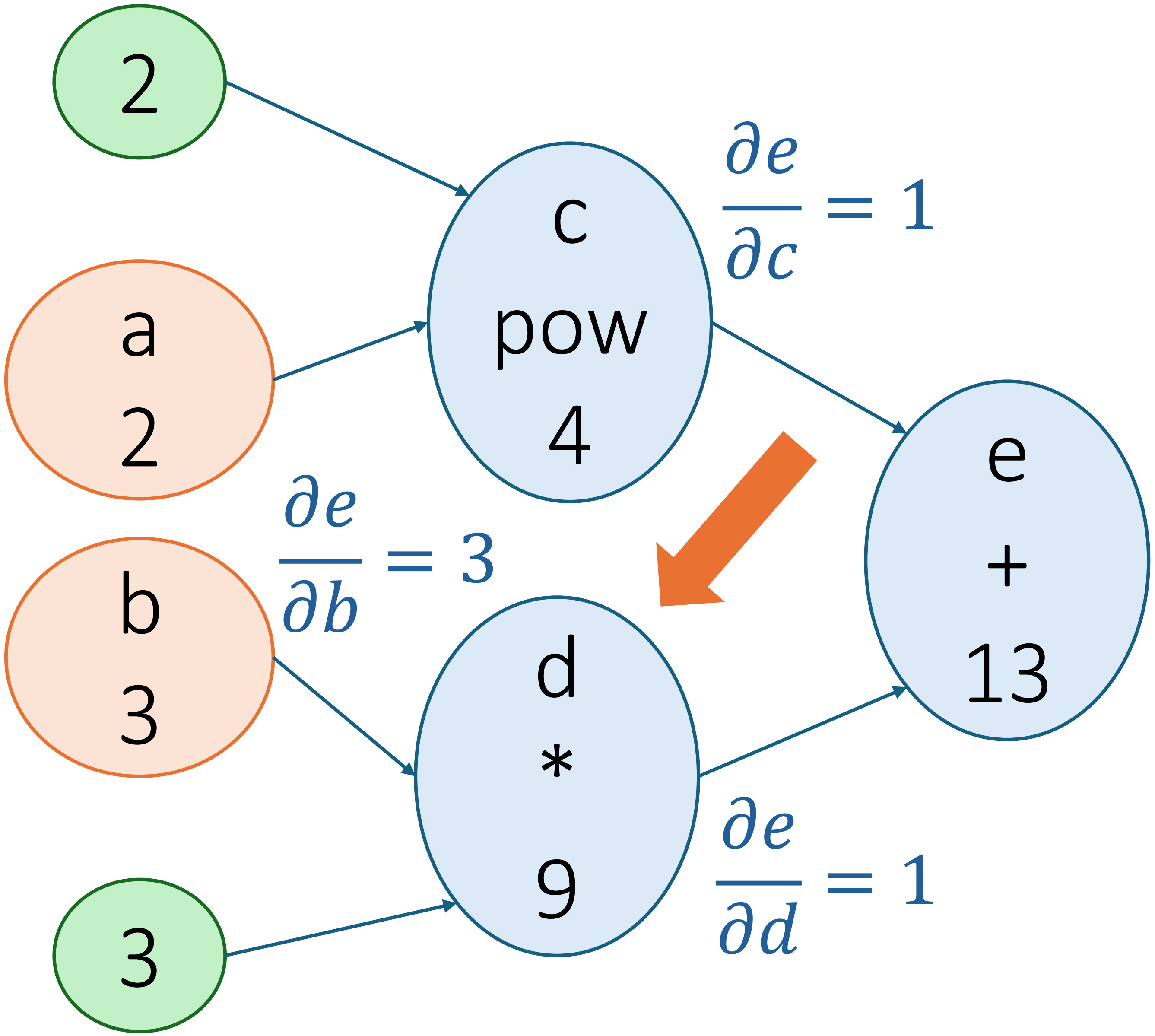
$a = 2$

$b = 3$

$c = a ** 2$

$d = 3 * b$

$e = c + d$



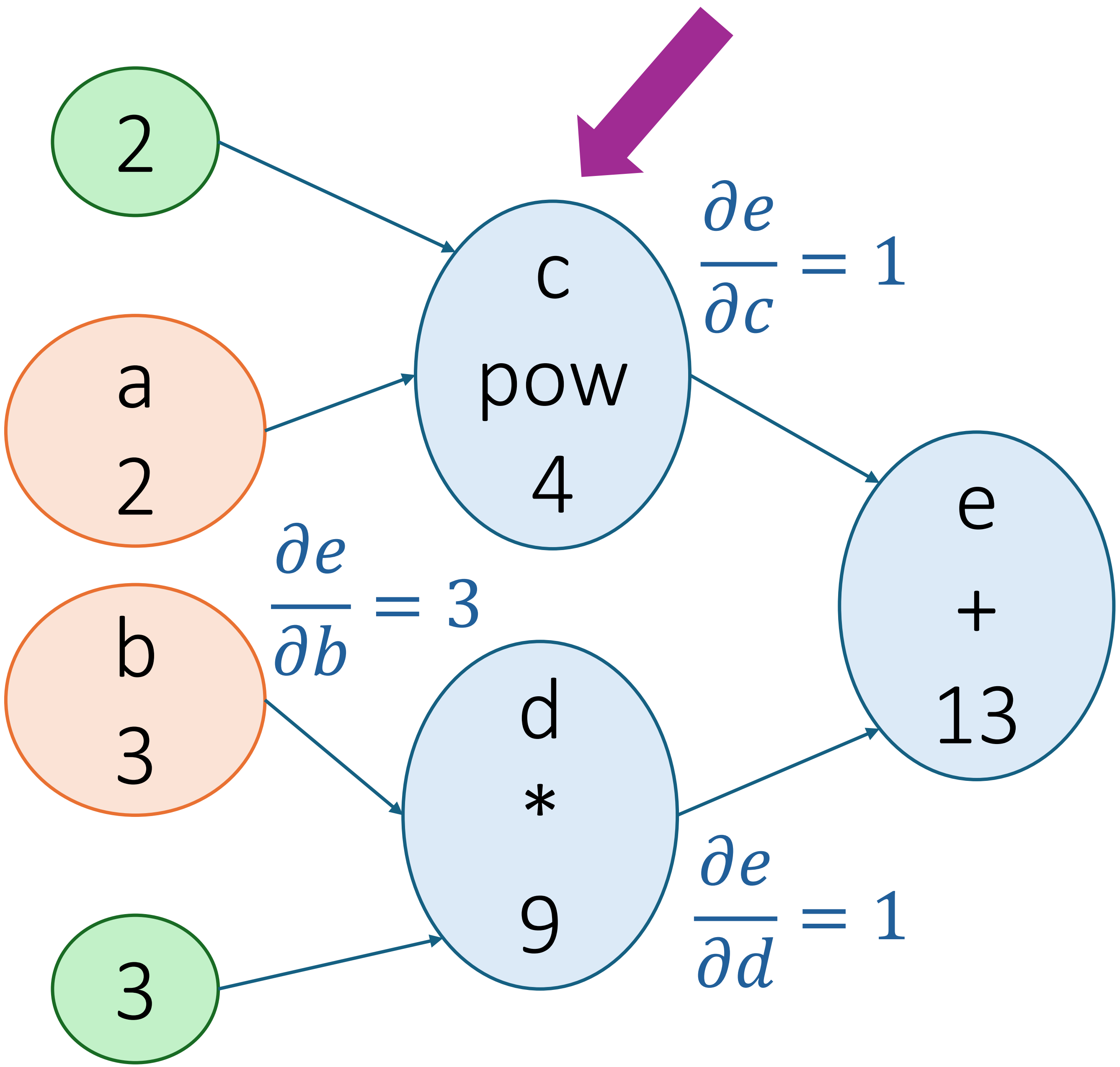
$a = 2$

$b = 3$

$c = a ** 2$

$d = 3 * b$

$e = c + d$



$$\frac{\partial e}{\partial a} = \frac{\partial e}{\partial c}\frac{\partial c}{\partial a}$$

$a = 2$

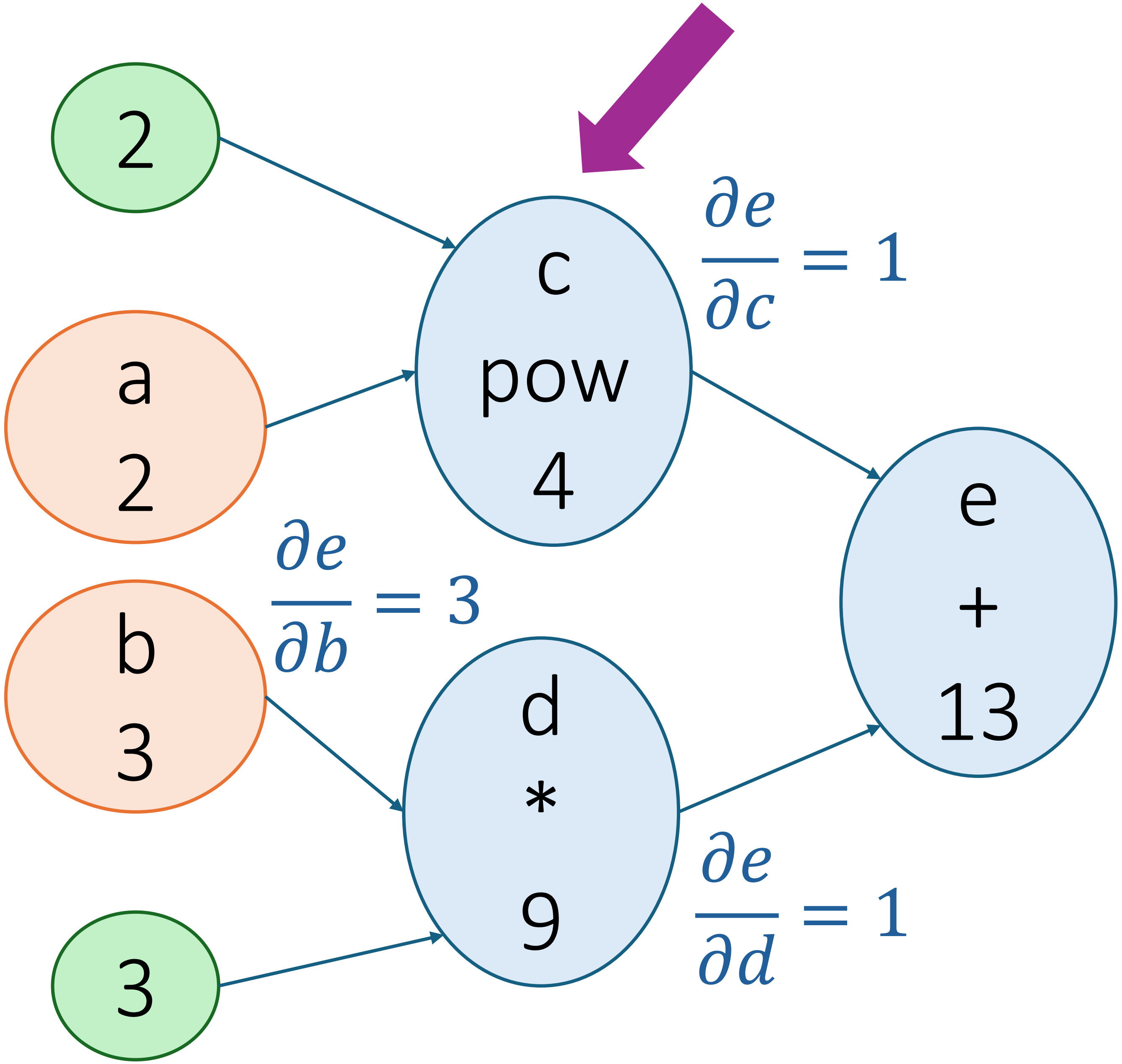
$b = 3$

$c = a ** 2$

$d = 3 * b$

$e = c + d$

$$\frac{\partial e}{\partial a} = 1 \times \frac{\partial c}{\partial a}$$



$a = 2$

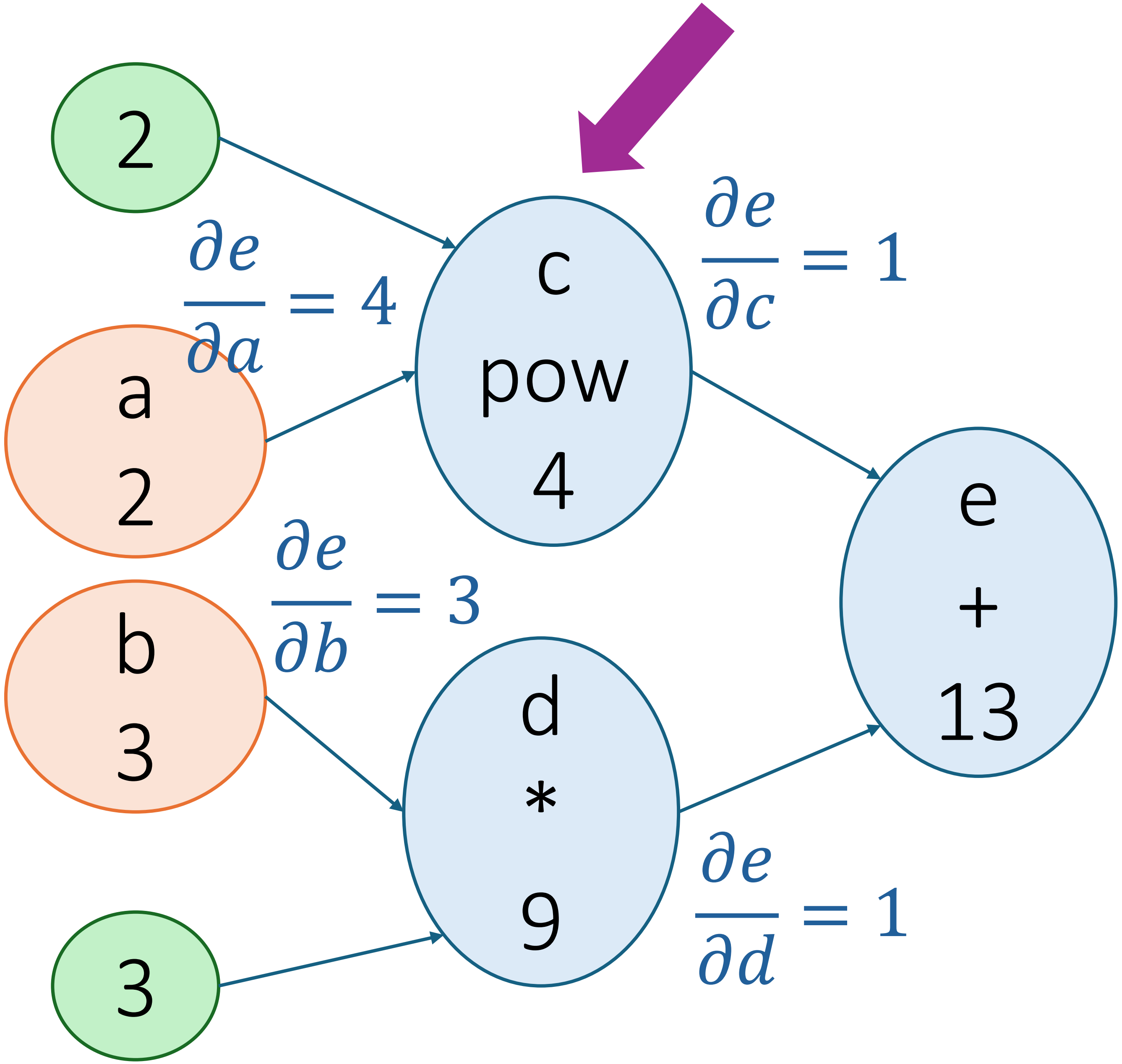
$b = 3$

$c = a ** 2$

$d = 3 * b$

$e = c + d$

$$\frac{\partial e}{\partial a} = 1 \times 2 \times a = 4$$



vamos verificar:

$$e = a^2 + 3b$$

$$\frac{\partial e}{\partial a} = 2a \;\rightarrow \frac{\partial e}{\partial a}(a = 2) = 4$$

$$\frac{\partial e}{\partial b} = 3$$

```
a = 2
b = 3
c = a ** 2
d = 3 * b
e = c + d
```

😊 👍

$x = 5$

$y = 3$

$a = square(x)$

$b = a + y$

$c = sin(b)$

Atividade: Construa o grafo computacional. Calcule as derivadas parciais em relação a x e y caminhando para trás no grafo.

$x = 5$

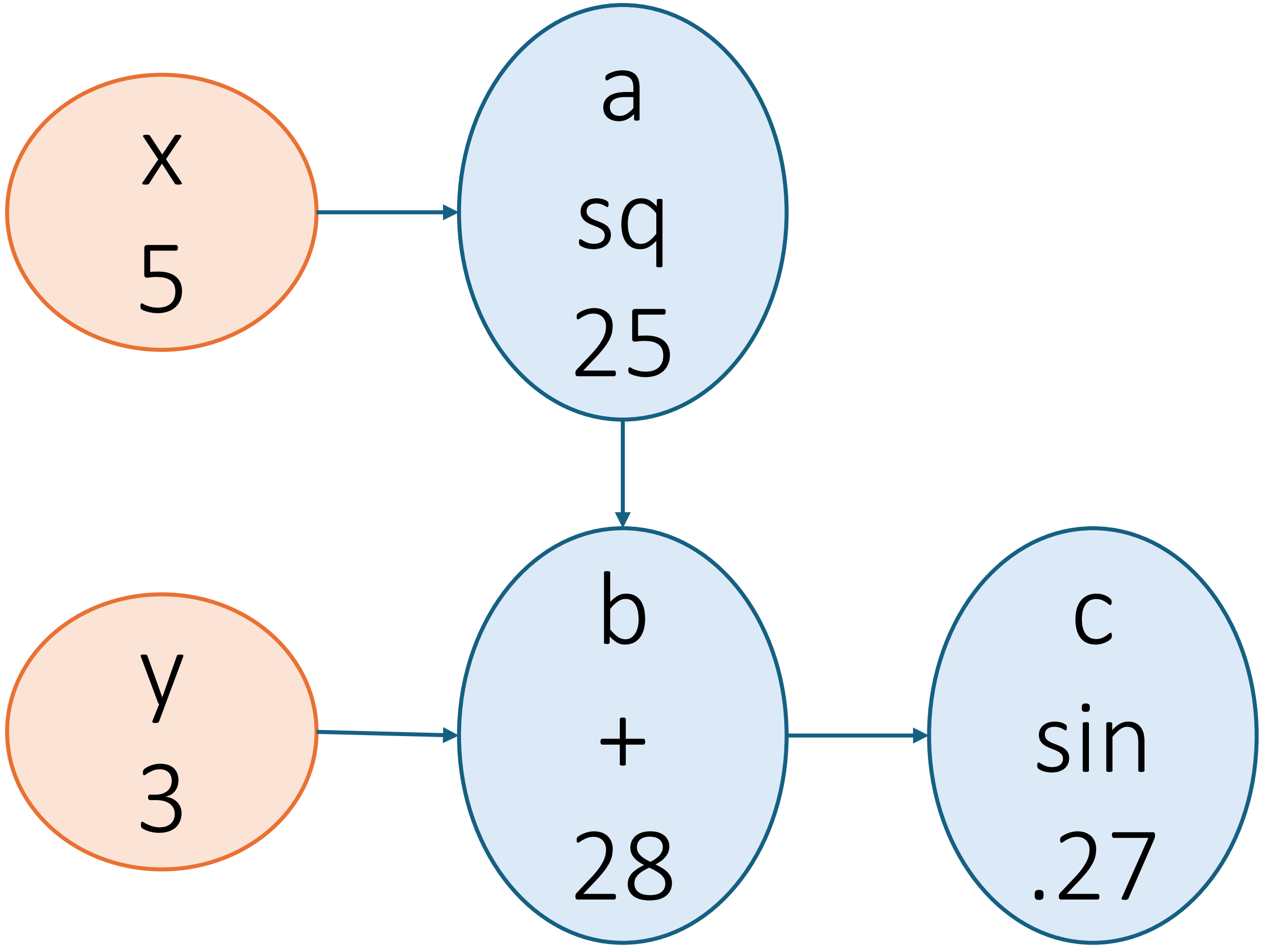
$y = 3$

$a = square(x)$

$b = a + y$

$c = sin(b)$

Atividade: Construa o grafo computacional. Calcule as derivadas parciais em relação a x e y caminhando para trás no grafo.



$x = 5$

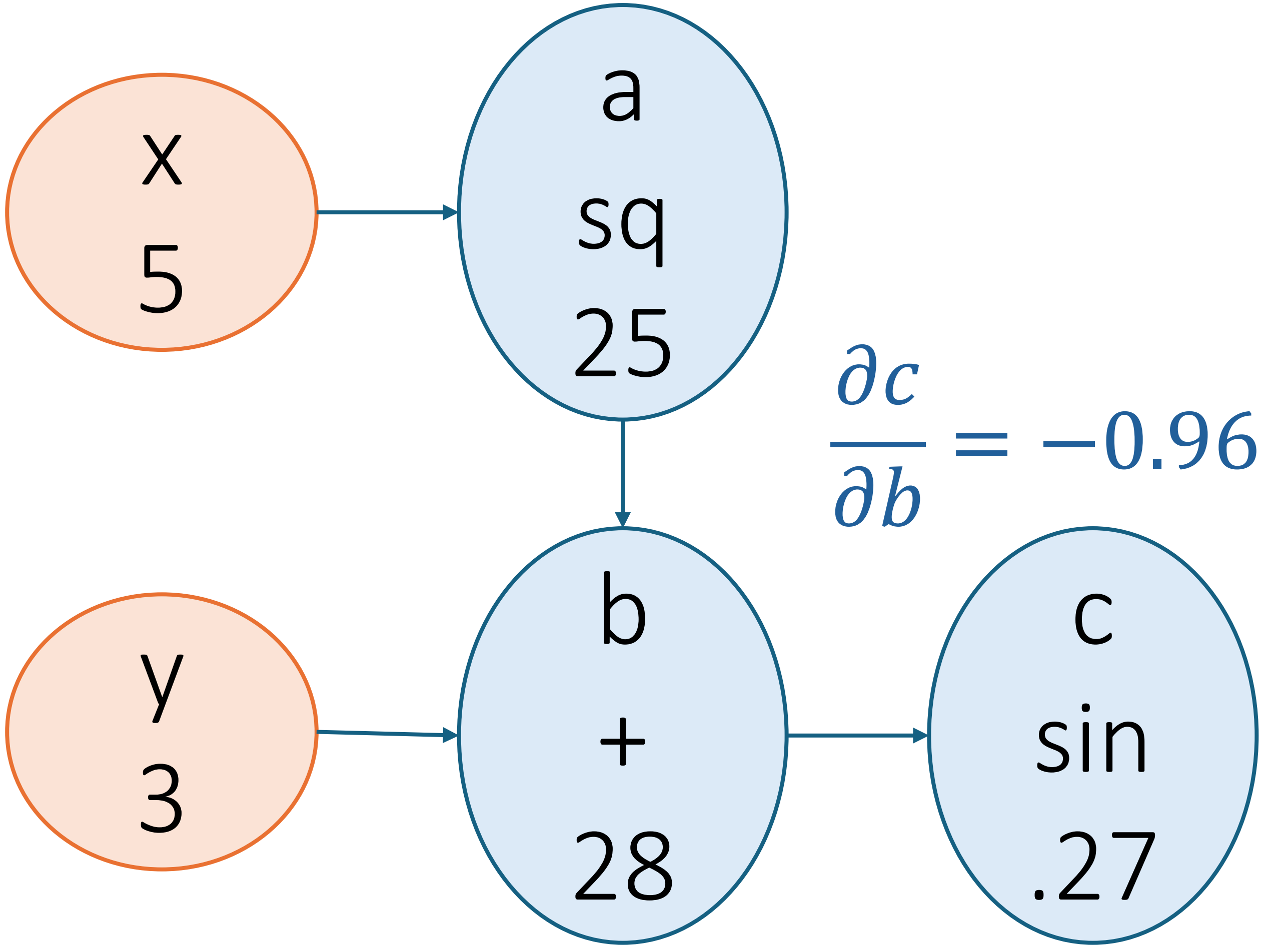
$y = 3$

$a = square(x)$

$b = a + y$

$c = sin(b)$

Atividade: Construa o grafo computacional. Calcule as derivadas parciais em relação a x e y caminhando para trás no grafo.



$$\frac{\partial c}{\partial b} = -0.96$$

$$\frac{\partial c}{\partial b} = \cos(b) = -0.96$$

$x = 5$

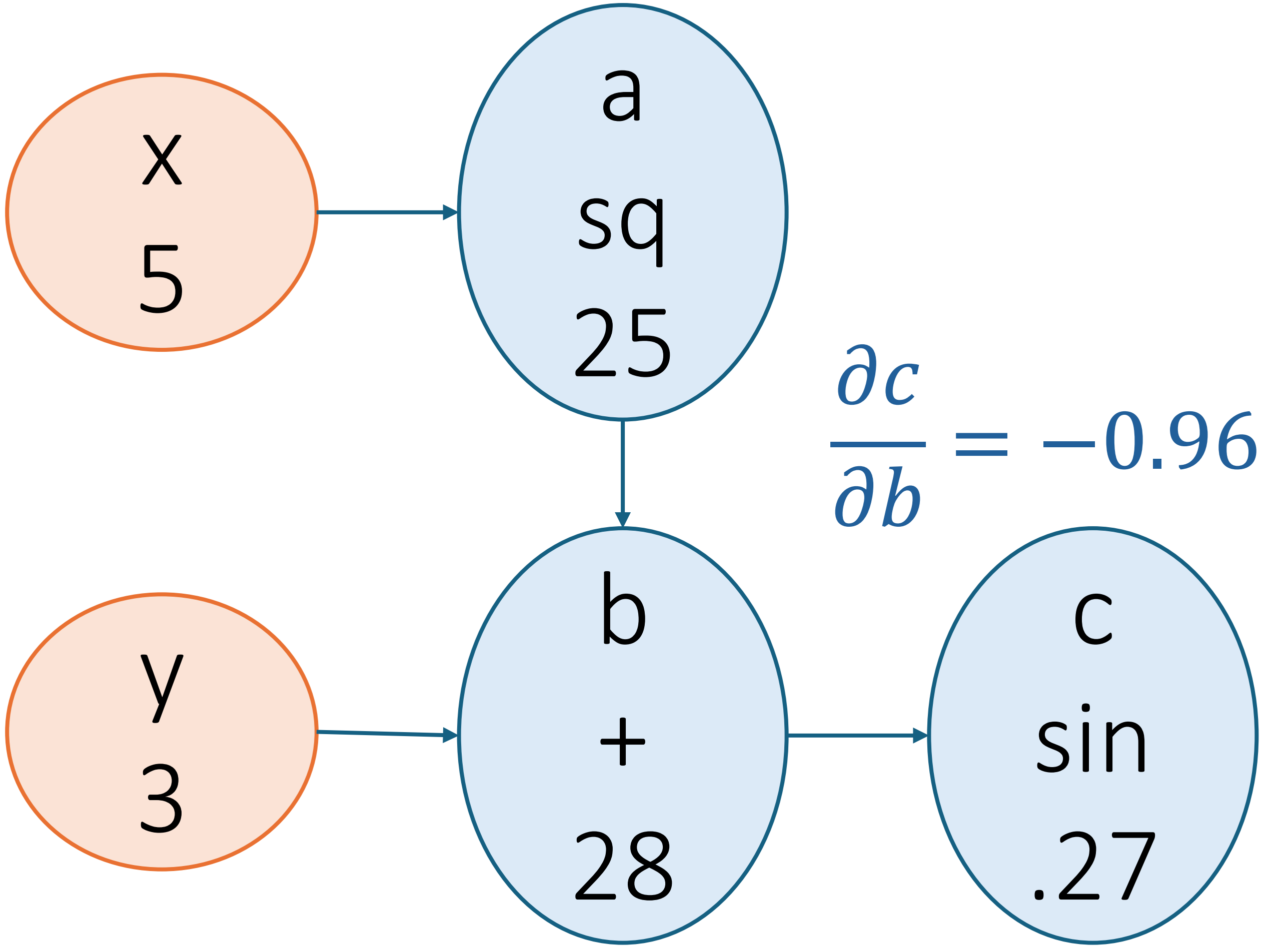
$y = 3$

$a = square(x)$

$b = a + y$

$c = sin(b)$

Atividade: Construa o grafo computacional. Calcule as derivadas parciais em relação a x e y caminhando para trás no grafo.



$$\frac{\partial c}{\partial b} = -0.96$$

$$\frac{\partial c}{\partial a} = \frac{\partial c}{\partial b}\frac{\partial b}{\partial a} = -0.96\frac{\partial b}{\partial a}$$

$$\frac{\partial c}{\partial y} = \frac{\partial c}{\partial b}\frac{\partial b}{\partial y} = -0.96\frac{\partial b}{\partial y}$$

$x = 5$

$y = 3$

$a = square(x)$

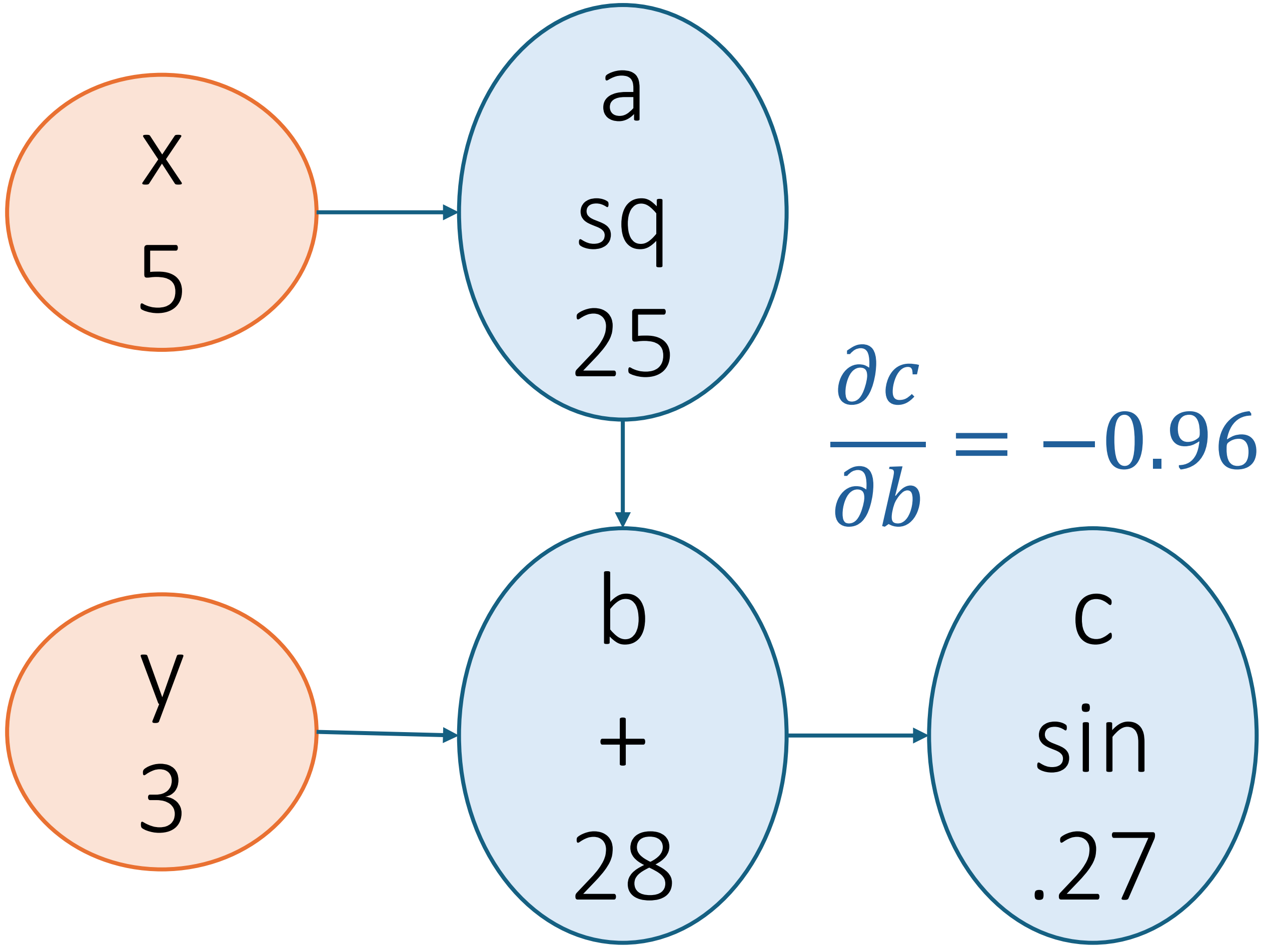
$b = a + y$

$c = sin(b)$

Atividade: Construa o grafo computacional. Calcule as derivadas parciais em relação a x e y caminhando para trás no grafo.

$$\frac{\partial c}{\partial a} = \frac{\partial c}{\partial b}\frac{\partial b}{\partial a} = -0.96 \times 1$$

$$\frac{\partial c}{\partial y} = \frac{\partial c}{\partial b}\frac{\partial b}{\partial y} = -0.96 \times 1$$



$$\frac{\partial c}{\partial b} = -0.96$$

$x = 5$

$y = 3$

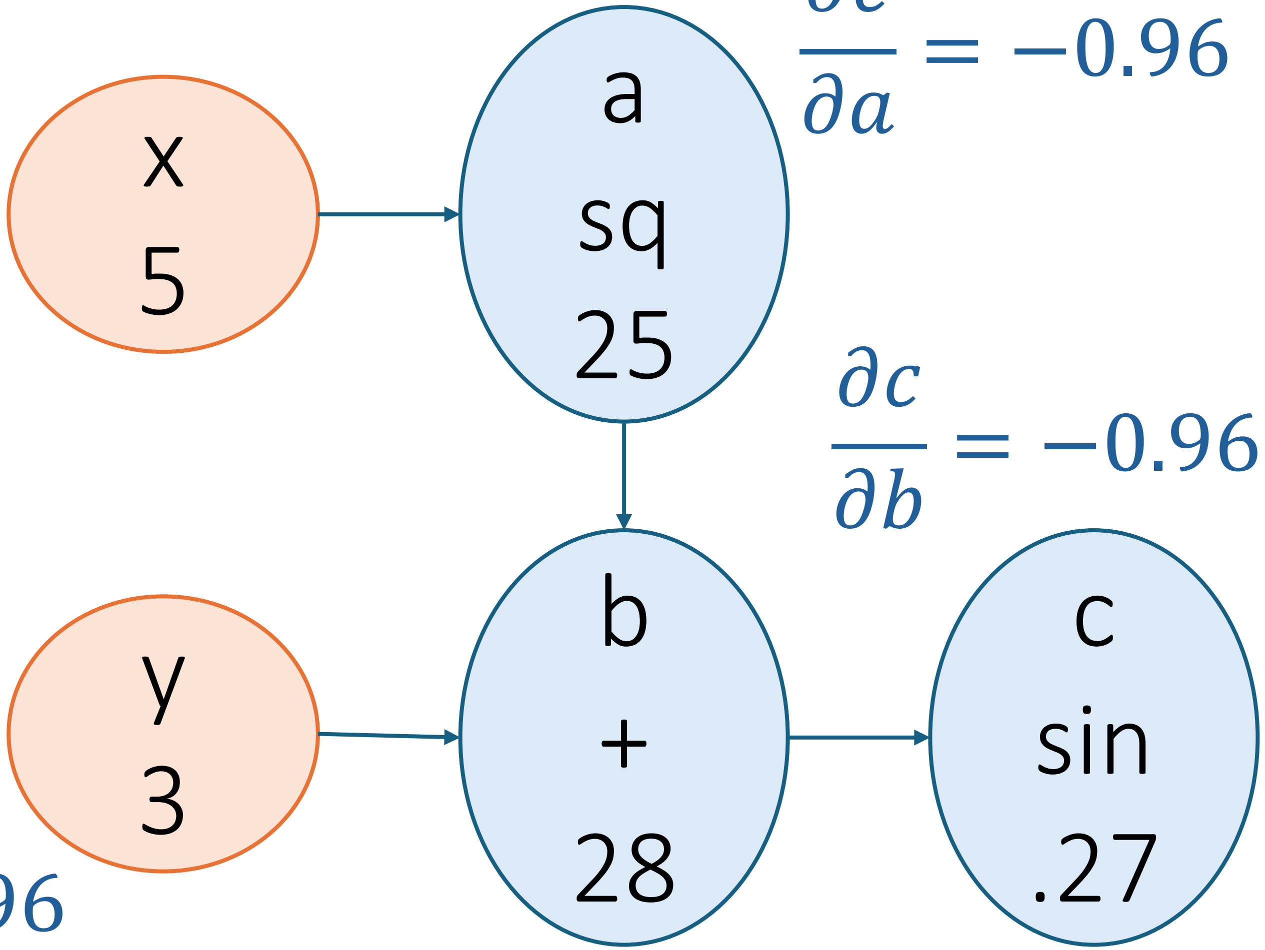
$a = square(x)$

$b = a + y$

$c = sin(b)$

Atividade: Construa o grafo computacional. Calcule as derivadas parciais em relação a x e y caminhando para trás no grafo.

$$\frac{\partial c}{\partial a} = -0.96$$



$$\frac{\partial c}{\partial b} = -0.96$$

$$\frac{\partial c}{\partial a} = \frac{\partial c}{\partial b}\frac{\partial b}{\partial a} = -0.96 \times 1$$

$$\frac{\partial c}{\partial y} = \frac{\partial c}{\partial b}\frac{\partial b}{\partial y} = -0.96 \times 1$$

$$\frac{\partial c}{\partial y} = -0.96$$

$x = 5$

$y = 3$

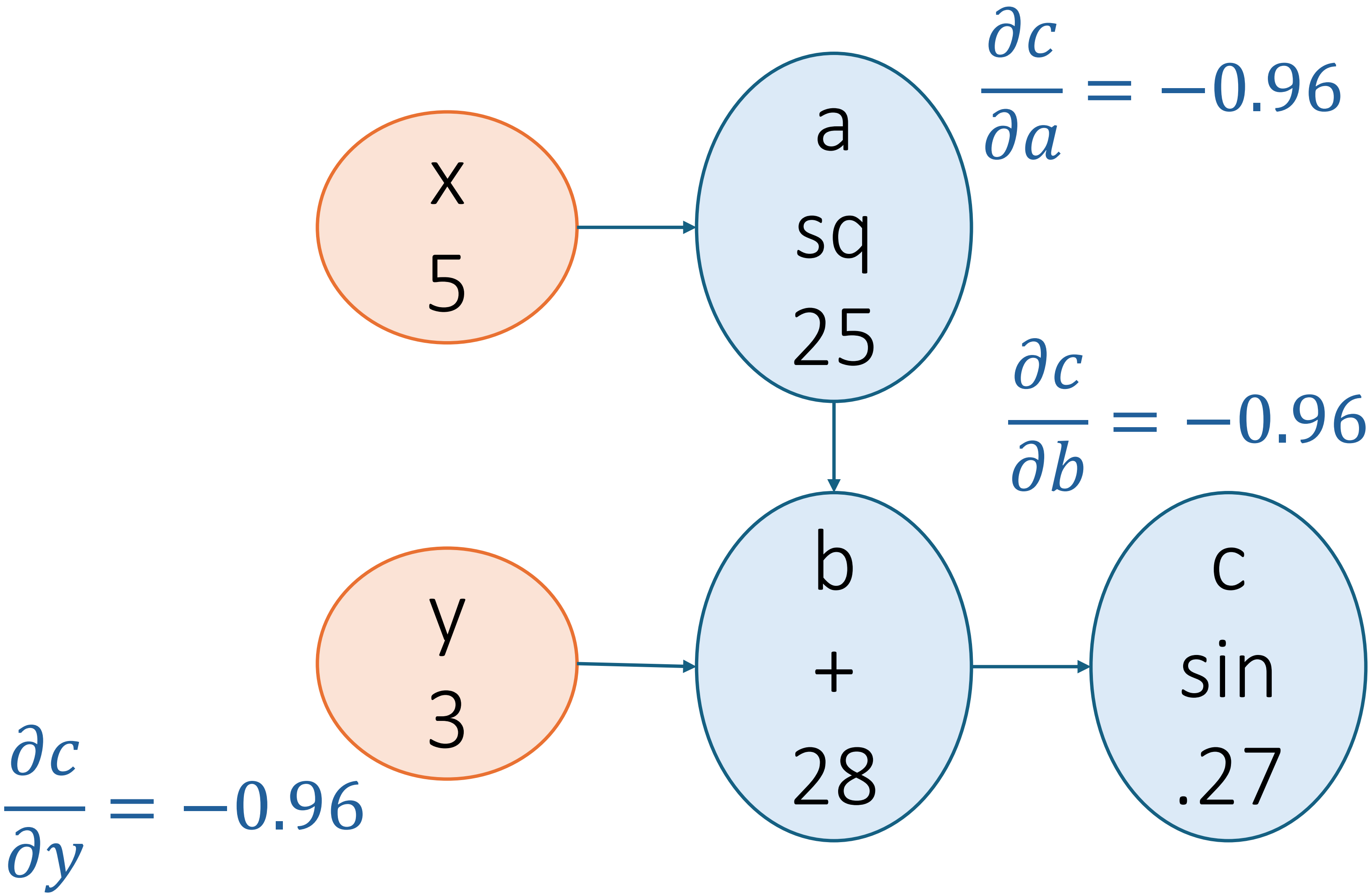
$a = square(x)$

$b = a + y$

$c = sin(b)$

Atividade: Construa o grafo computacional. Calcule as derivadas parciais em relação a x e y caminhando para trás no grafo.

$$\frac{\partial c}{\partial x} = \frac{\partial c}{\partial b}\frac{\partial b}{\partial a}\frac{\partial a}{\partial x}$$



$x = 5$

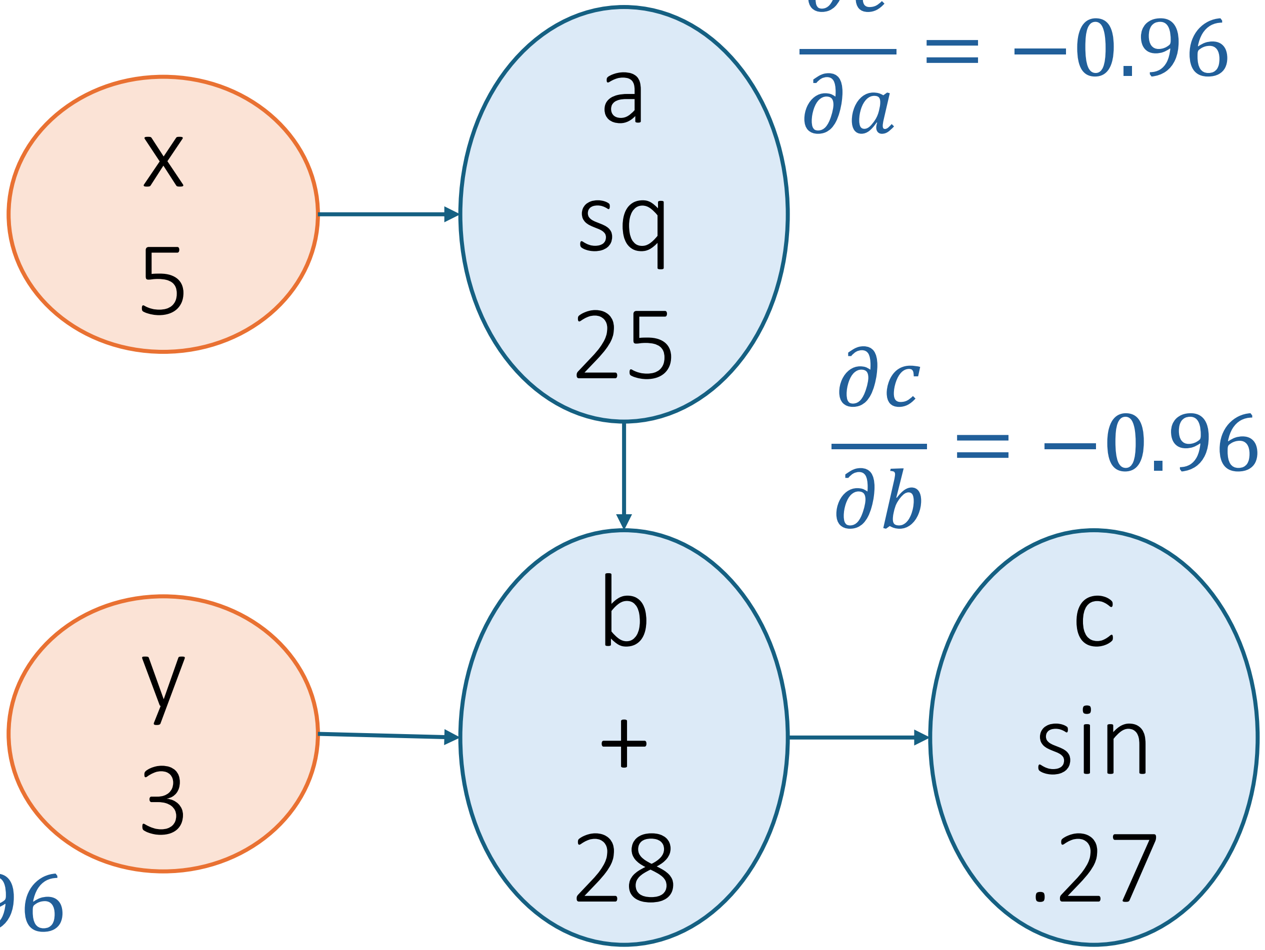
$y = 3$

$a = square(x)$

$b = a + y$

$c = sin(b)$

Atividade: Construa o grafo computacional. Calcule as derivadas parciais em relação a x e y caminhando para trás no grafo.



$$\frac{\partial c}{\partial a} = -0.96$$

$$\frac{\partial c}{\partial b} = -0.96$$

$$\frac{\partial c}{\partial x} = \underbrace{\frac{\partial c}{\partial b}\frac{\partial b}{\partial a}}_{}\frac{\partial a}{\partial x}$$

$$= -0.96 \times (2 \times x)$$

Já calculamos!

$$\frac{\partial c}{\partial y} = -0.96$$

x = 5

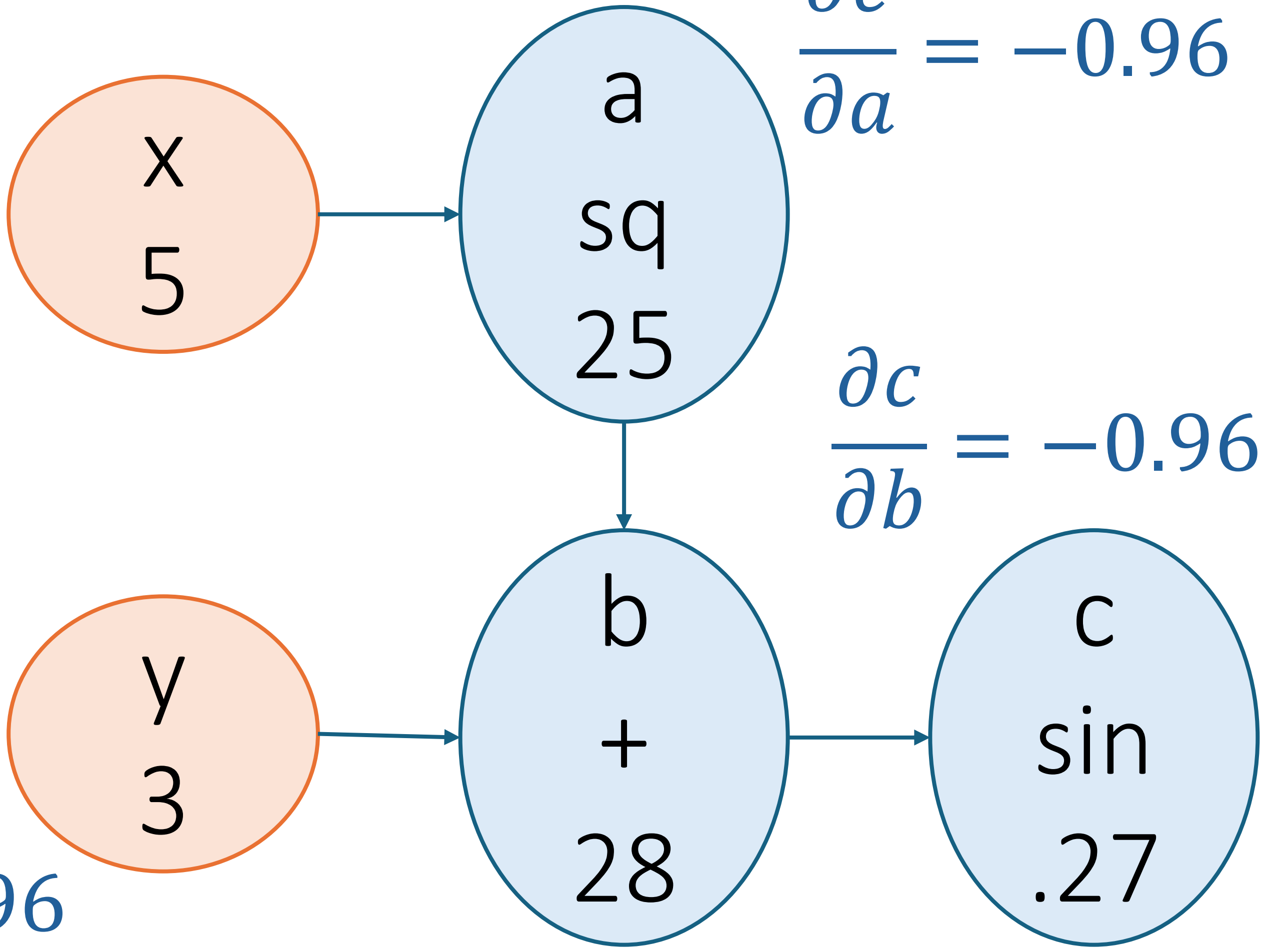
y = 3

a = square(x)

b = a + y

c = sin(b)

<u>Atividade</u>: Construa o grafo computacional. Calcule as derivadas parciais em relação a x e y caminhando para trás no grafo.



$$\frac{\partial c}{\partial a} = -0.96$$

$$\frac{\partial c}{\partial b} = -0.96$$

$$\frac{\partial c}{\partial x} = \frac{\partial c}{\partial b}\frac{\partial b}{\partial a}\frac{\partial a}{\partial x}$$

$$= -0.96 \times (2 \times 5)$$

$$\frac{\partial c}{\partial y} = -0.96$$

$x = 5$

$y = 3$

$a = square(x)$

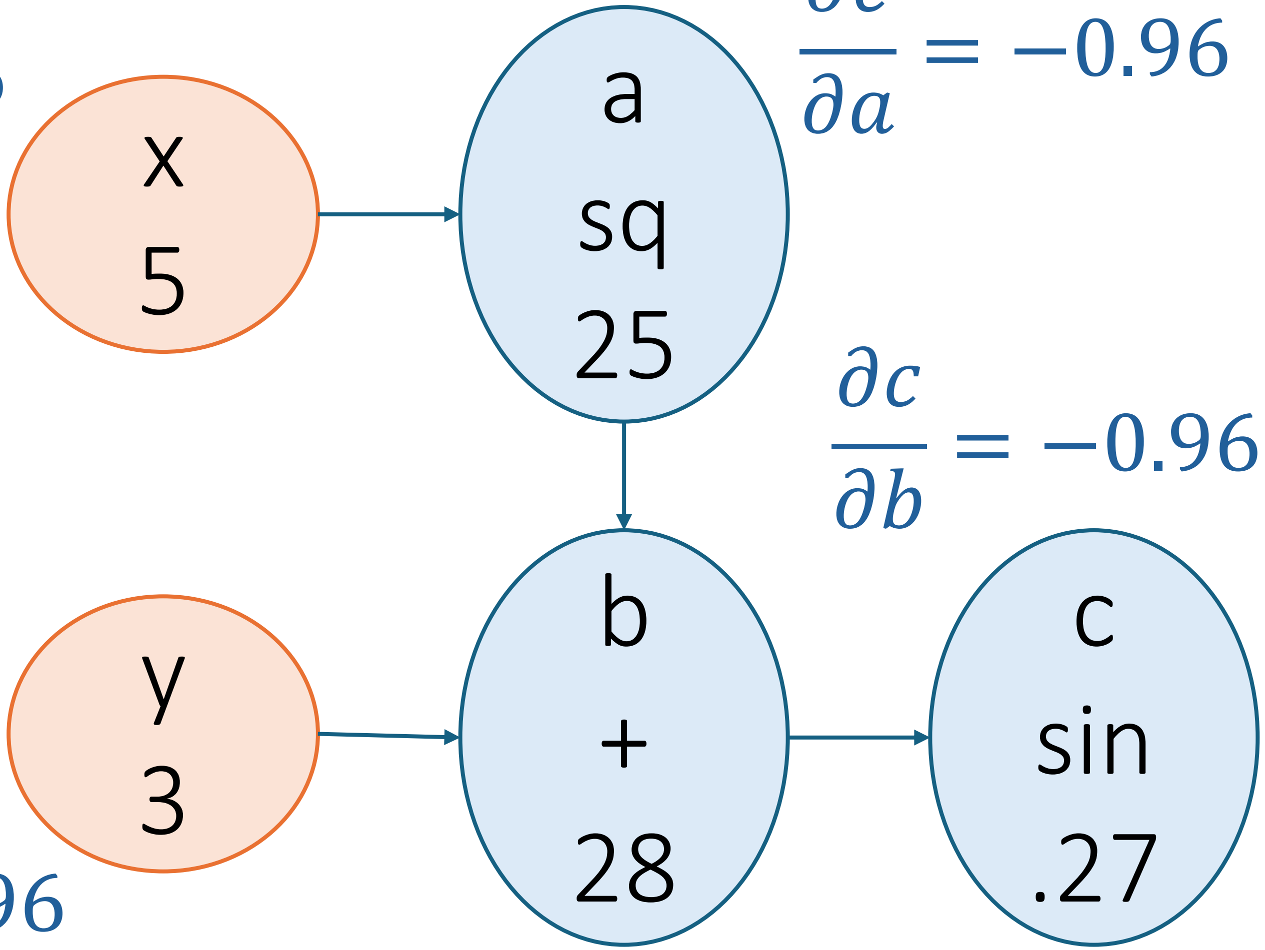
$b = a + y$

$c = sin(b)$

Atividade: Construa o grafo computacional. Calcule as derivadas parciais em relação a x e y caminhando para trás no grafo.

$$\frac{\partial c}{\partial x} = -9.6$$

$$\frac{\partial c}{\partial a} = -0.96$$



$$\frac{\partial c}{\partial b} = -0.96$$

$$\frac{\partial c}{\partial y} = -0.96$$

$$\frac{\partial c}{\partial x} = \frac{\partial c}{\partial b}\frac{\partial b}{\partial a}\frac{\partial a}{\partial x}$$

$$= -9.6$$

Observações

- Cada nó calcula sua “contribuição” para a derivada parcial.
- Derivadas parciais mais para frente no grafo não precisam ser recalculadas e podem ser “reutilizadas”.

IMPORTANTE

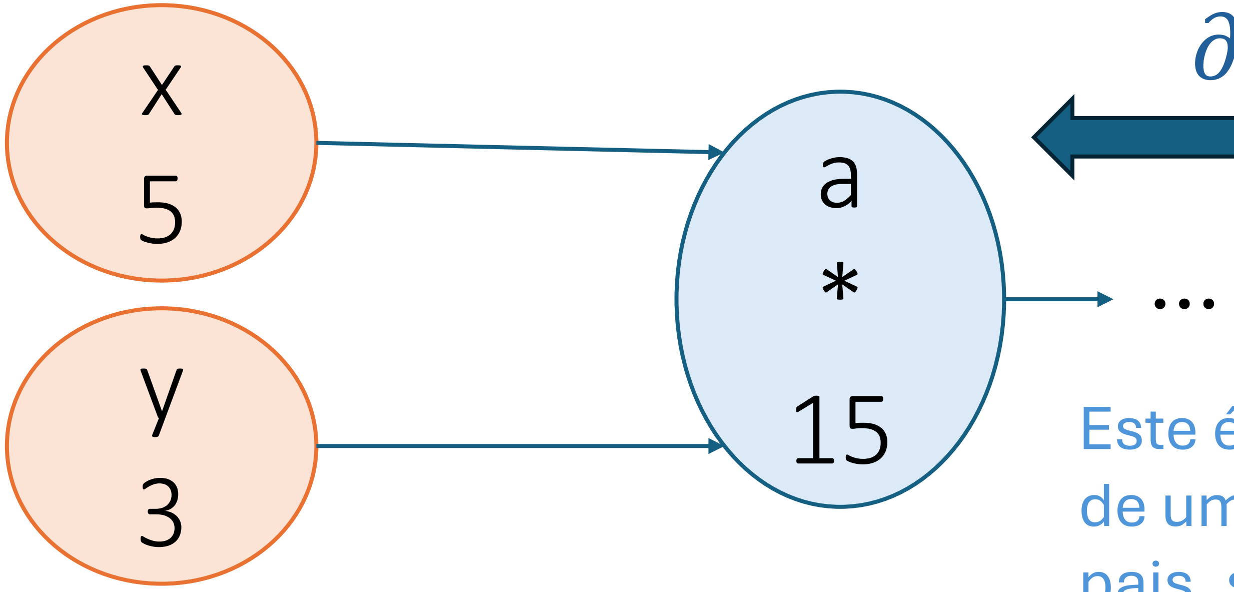
1) Vamos modelar a solução de forma que cada operação saiba como calcular sua derivada “local”.

- f=sin(x) -> df/dx = cos(x)
- f=cos(x) -> df/dx = -sin(x)
- f=a+b -> df/da = df/db = 1
- f=a*b -> df/da = b; df/db = a

Bem mais simples que pensar em derivadas de expressões longas e complexas!

2) Usando isto, cada operação vai calcular as derivadas parciais dos pais.

$a = xy$



$$\frac{\partial \, out}{\partial a}$$

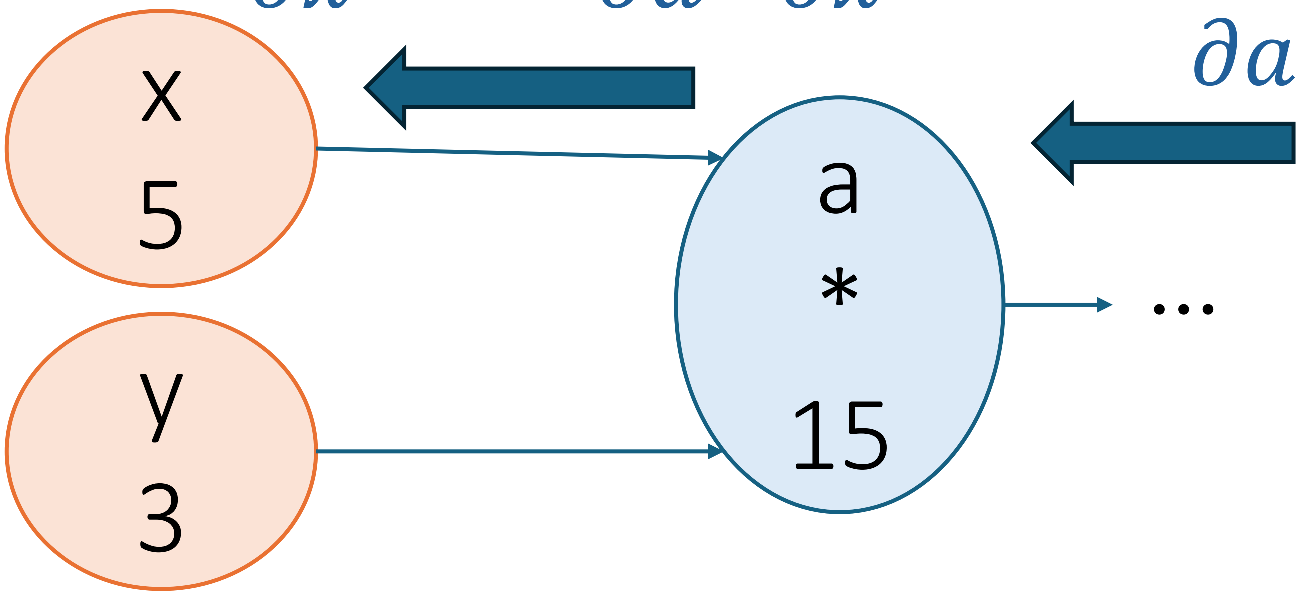
Este é o campo de visão de um nó do grafo: seus pais, sua operação e a derivada recebida de seus filhos

2) Usando isto, cada operação vai calcular as derivadas parciais dos pais.

$$\frac{\partial\, out}{\partial x} = \frac{\partial\, out}{\partial a}\frac{\partial a}{\partial x}$$

$$\frac{\partial\, out}{\partial a}$$

$a = xy$

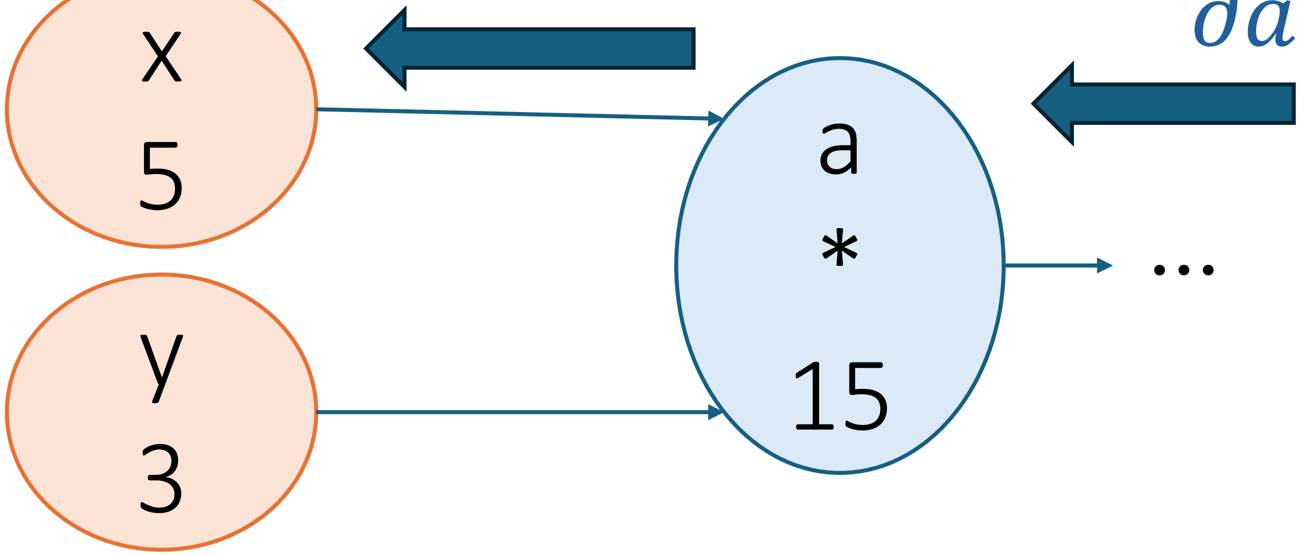


2) Usando isto, cada operação vai calcular as derivadas parciais dos pais.

$$\frac{\partial\, out}{\partial x} = \frac{\partial\, out}{\partial a} y$$

$$\frac{\partial\, out}{\partial a}$$

$$a = xy$$

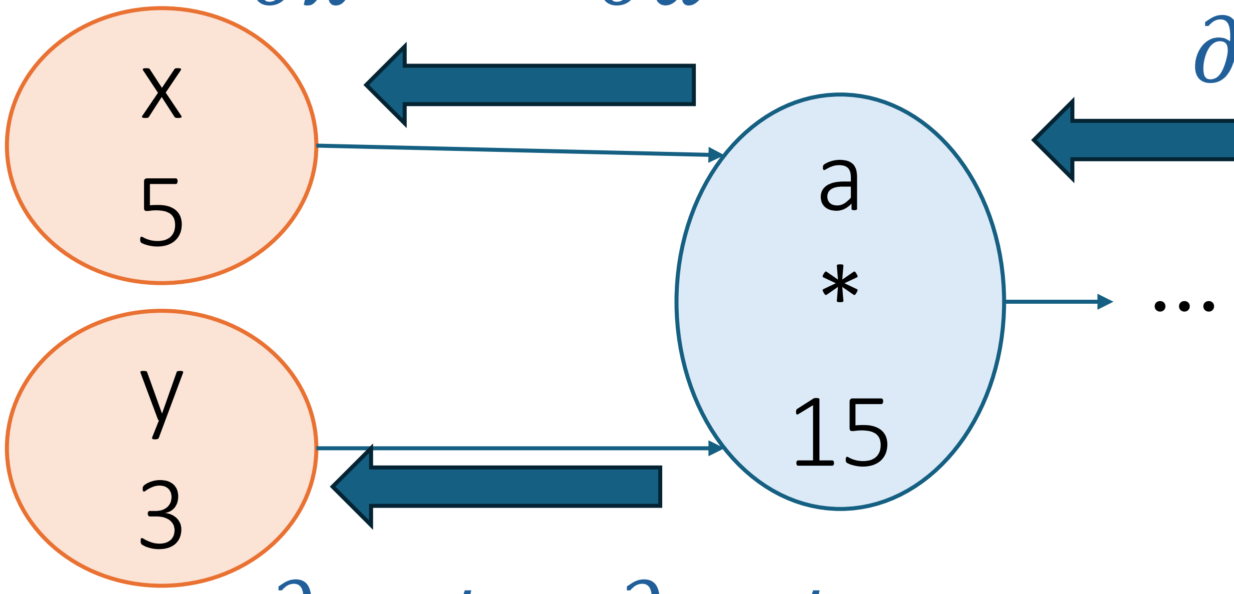


2) Usando isto, cada operação vai calcular as derivadas parciais dos pais.

$$\frac{\partial\, out}{\partial x} = \frac{\partial\, out}{\partial a} y$$

$$\frac{\partial\, out}{\partial a}$$

$$a = xy$$



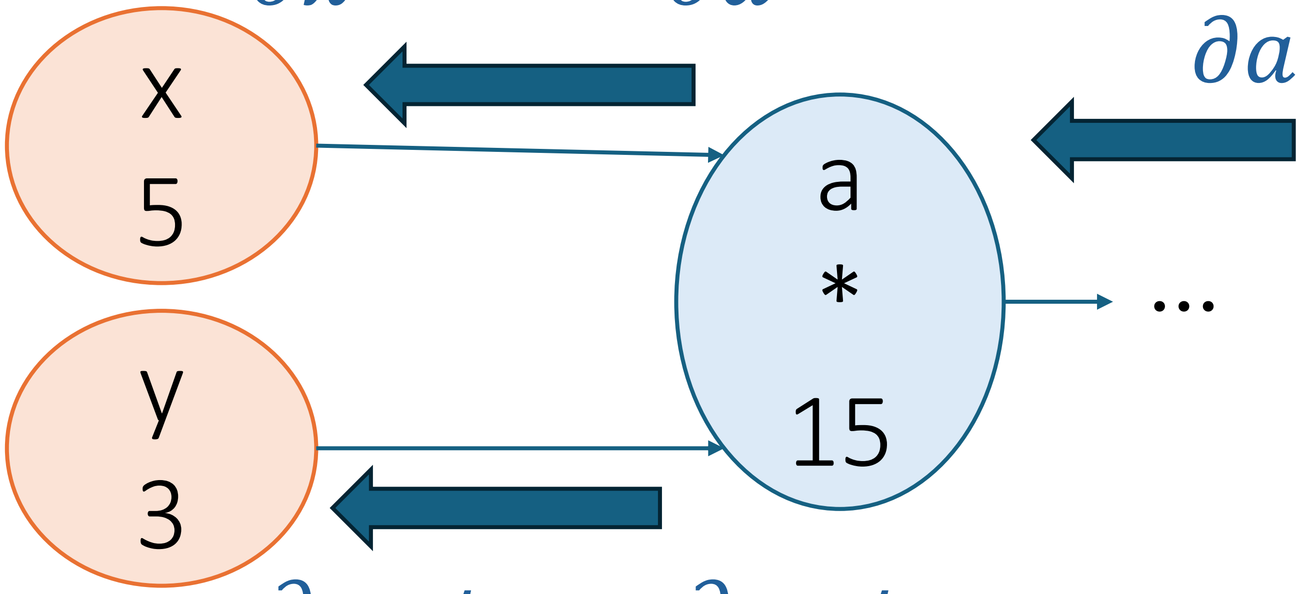
$$\frac{\partial\, out}{\partial y} = \frac{\partial\, out}{\partial a} x$$

2) Usando isto, cada operação vai calcular as derivadas parciais dos pais.

$$\frac{\partial\, out}{\partial x} = 3\,\frac{\partial\, out}{\partial a}$$

$$\frac{\partial\, out}{\partial a}$$

$$a = xy$$

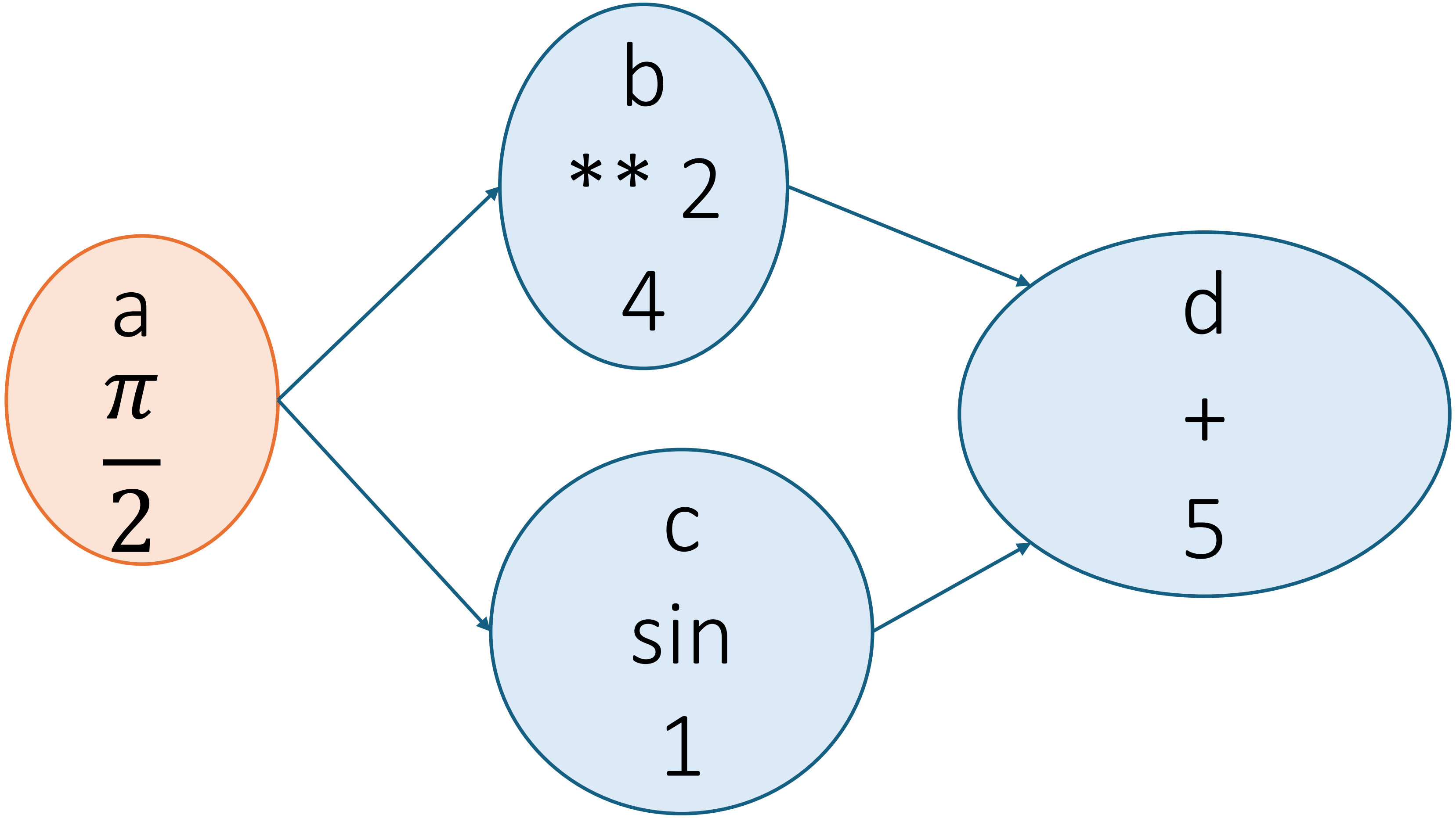


$$\frac{\partial\, out}{\partial y} = 5\,\frac{\partial\, out}{\partial a}$$

Fluxo de Gradientes em Bifurcações e Junções

$$a = \frac{\pi}{2}$$
$$b = a**2$$
$$c = \sin(a)$$
$$d = b + c$$

$$d = a**2 + \sin(a)$$



$$a = \frac{\pi}{2}$$

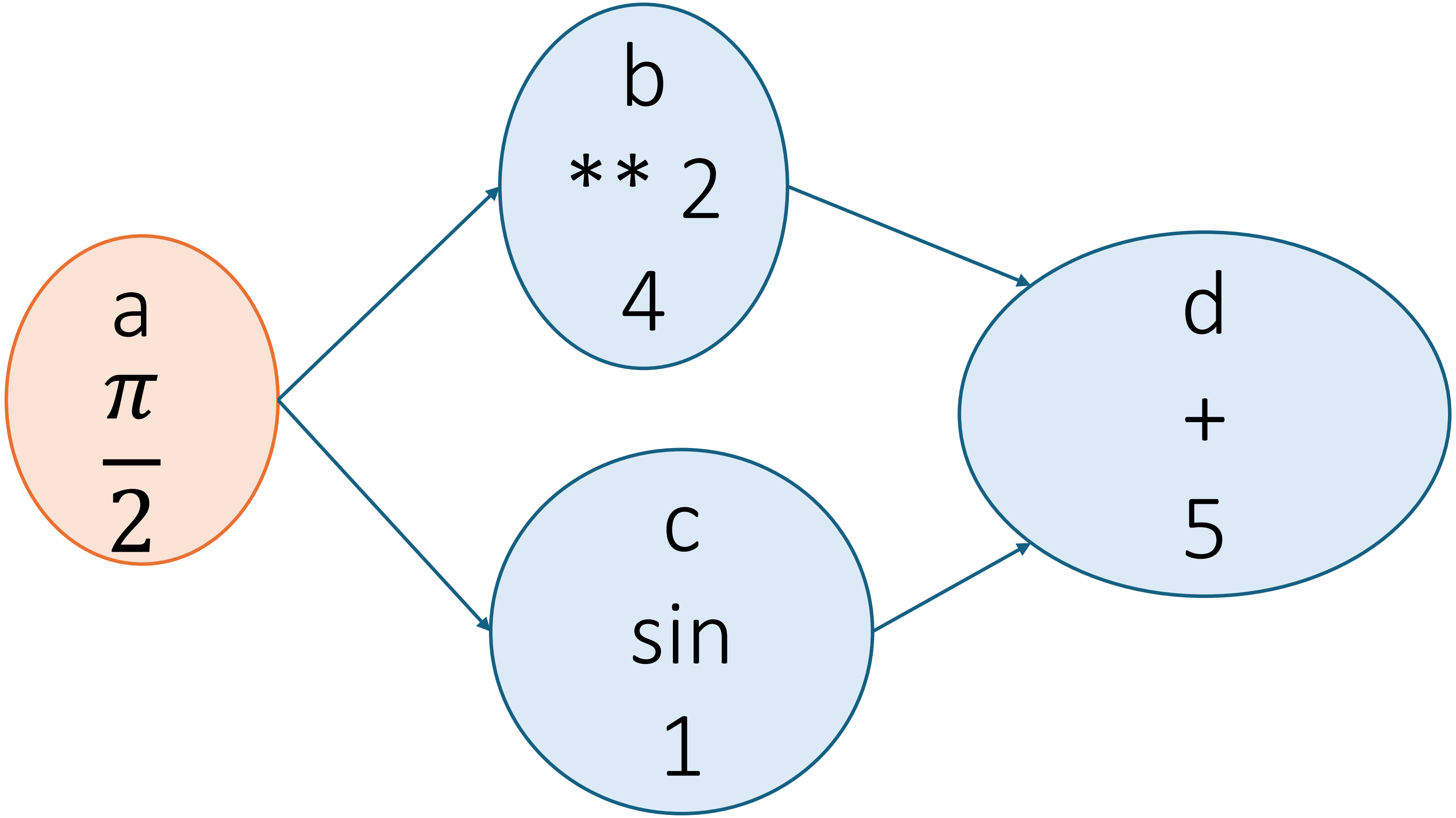
$$b = a**2$$

$$c = sin(a)$$

$$d = b + c$$

$$d = a**2 + sin(a)$$

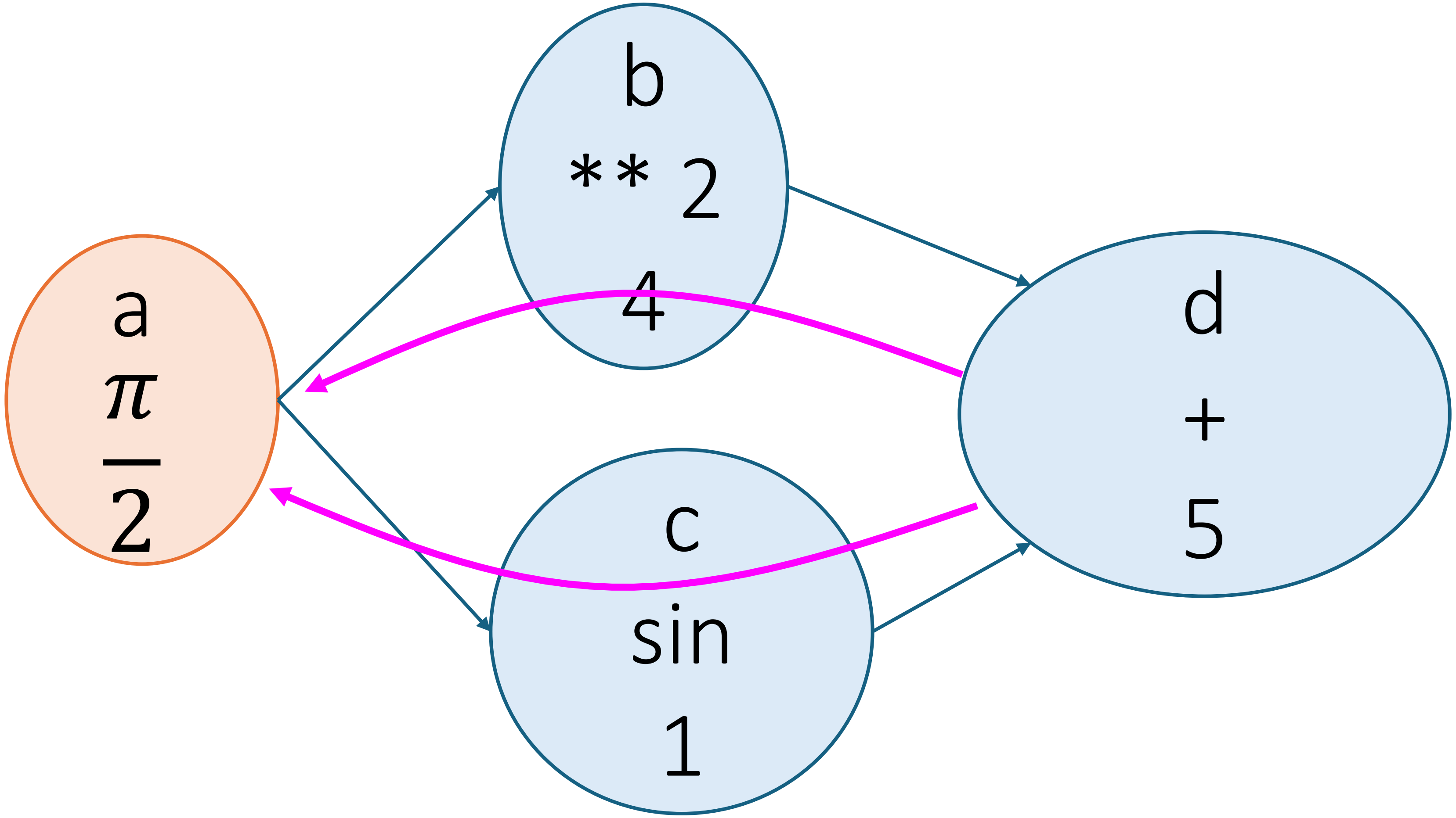
Olhando a fórmula de d, concluímos que a derivada verdadeira é dada por: $\frac{\partial d}{\partial a} = 2a + \cos(a)$



$$a = \frac{\pi}{2}$$
$$b = a**2$$
$$c = sin(a)$$
$$d = b + c$$

$$d = a**2 + sin(a)$$

Olhando a fórmula de d, concluímos que a derivada verdadeira é dada por: $\frac{\partial d}{\partial a} = 2a + \cos(a)$



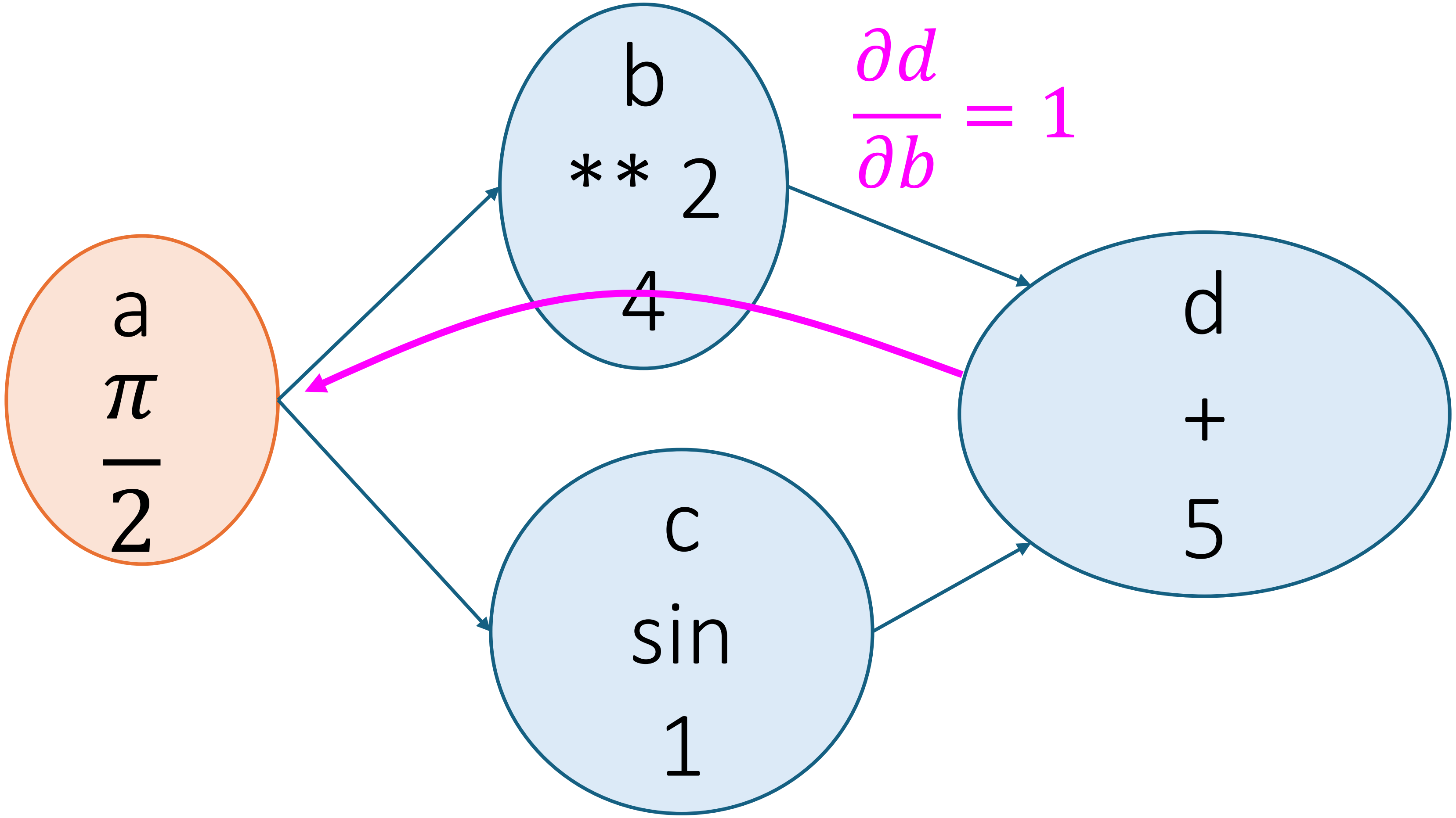
Em caso de bifurcações, chegaremos em *a* via dois caminhos ao executar backward() e as derivadas dos dois caminhos devem ser **somadas**.

Intuição: *estamos somando a contribuição de a nos dois caminhos.*

$$a = \frac{\pi}{2}$$
$$b = a**2$$
$$c = sin(a)$$
$$d = b + c$$

$$d = a**2 + sin(a)$$

Olhando a fórmula de d, concluímos que a derivada verdadeira é dada por: $\frac{\partial d}{\partial a} = 2a + \cos(a)$



$$a = \frac{\pi}{2}$$
$$b = a**2$$
$$c = sin(a)$$
$$d = b + c$$

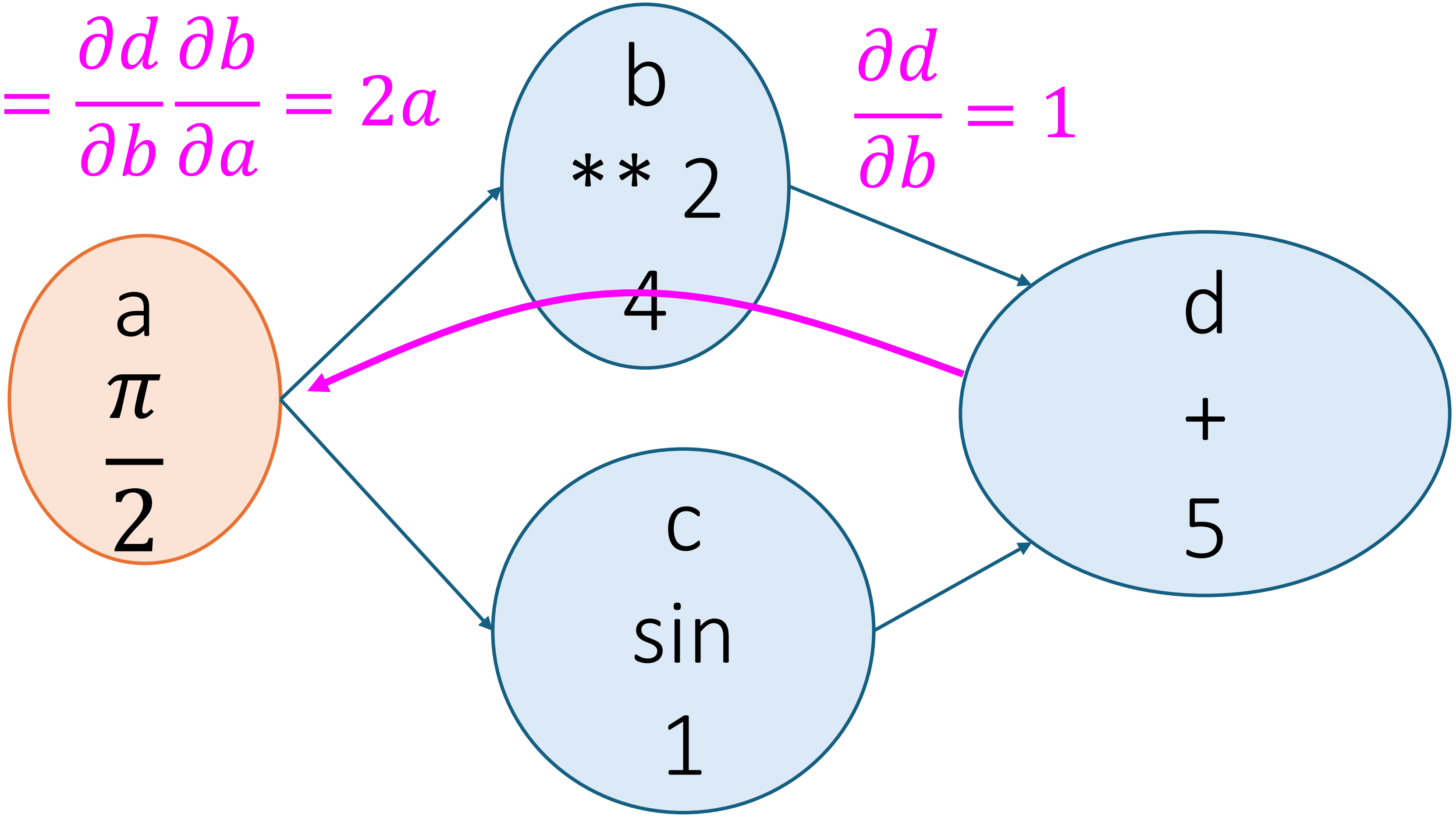
$$\Rightarrow \quad d = a**2 + sin(a)$$

Olhando a fórmula de d, concluímos que a derivada verdadeira é dada por: $\dfrac{\partial d}{\partial a} = 2a + \cos(a)$

caminho 1

$$\frac{\partial d}{\partial a} = \frac{\partial d}{\partial b}\frac{\partial b}{\partial a} = 2a$$

$$\frac{\partial d}{\partial b} = 1$$



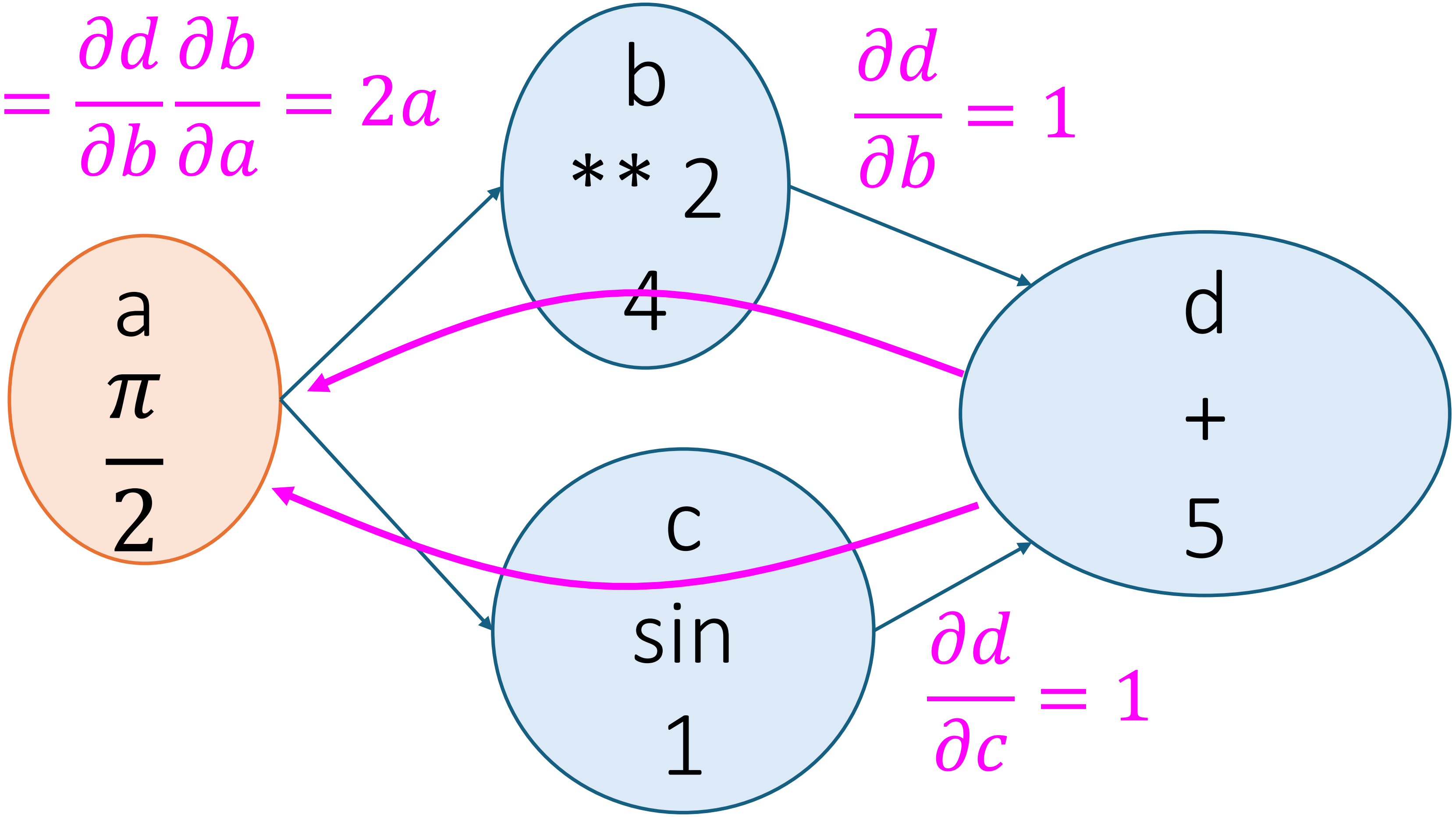
$$a = \frac{\pi}{2}$$
$$b = a**2$$
$$c = sin(a)$$
$$d = b + c$$

$$d = a**2 + sin(a)$$

Olhando a fórmula de d, concluímos que a derivada verdadeira é dada por: $\frac{\partial d}{\partial a} = 2a + \cos(a)$

caminho 1

$$\frac{\partial d}{\partial a} = \frac{\partial d}{\partial b}\frac{\partial b}{\partial a} = 2a$$



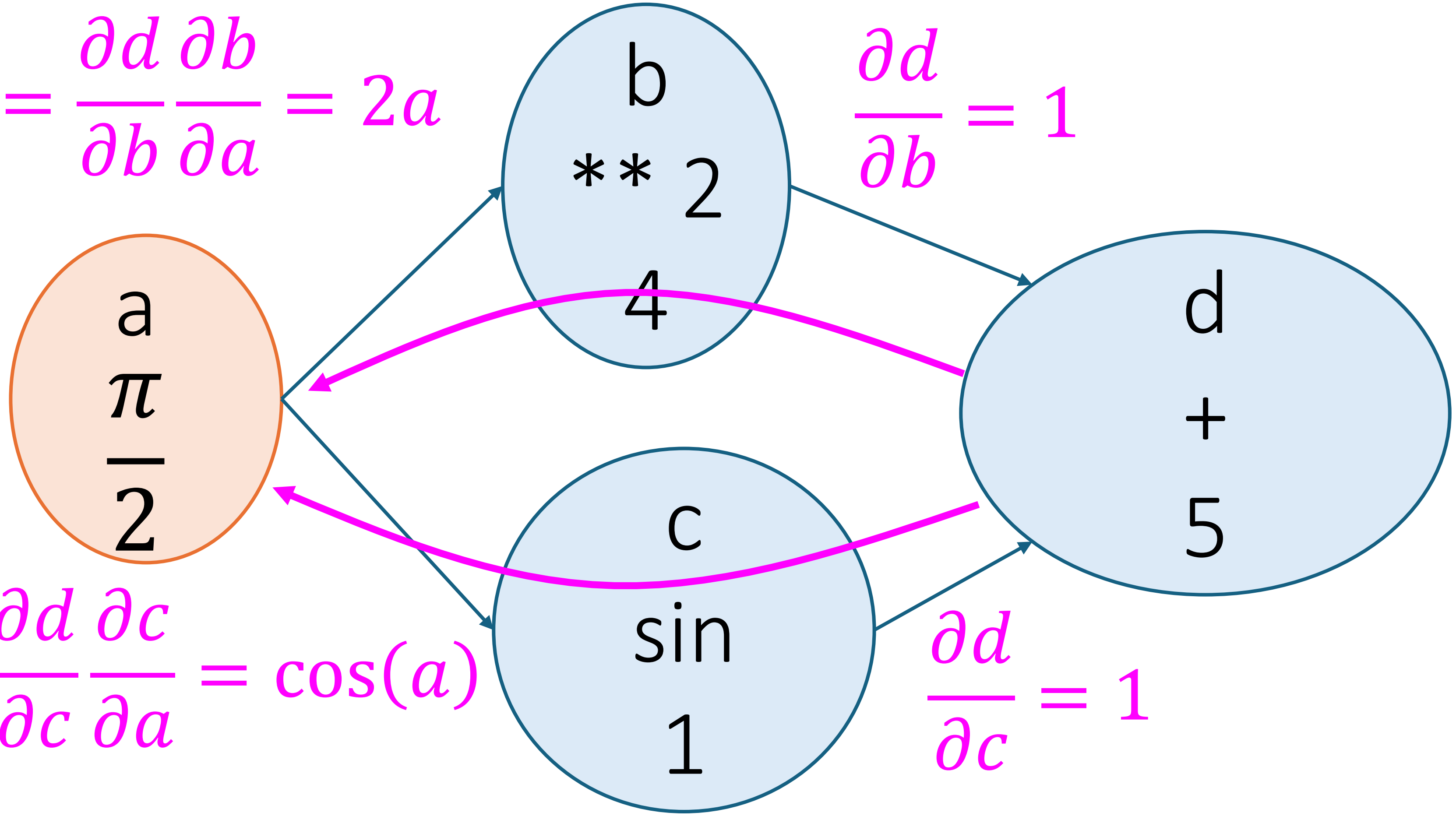
$$a = \frac{\pi}{2}$$
$$b = a**2$$
$$c = sin(a)$$
$$d = b + c$$

$$d = a**2 + sin(a)$$

Olhando a fórmula de d, concluímos que a derivada verdadeira é dada por: $\frac{\partial d}{\partial a} = 2a + \cos(a)$

caminho 1

$$\frac{\partial d}{\partial a} = \frac{\partial d}{\partial b}\frac{\partial b}{\partial a} = 2a$$



$$\frac{\partial d}{\partial b} = 1$$

$$\frac{\partial d}{\partial c} = 1$$

caminho 2

$$\frac{\partial d}{\partial a} = \frac{\partial d}{\partial c}\frac{\partial c}{\partial a} = \cos(a)$$

$$a = \frac{\pi}{2}$$
$$b = a**2$$
$$c = sin(a)$$
$$d = b + c$$

$$d = a**2 + sin(a)$$

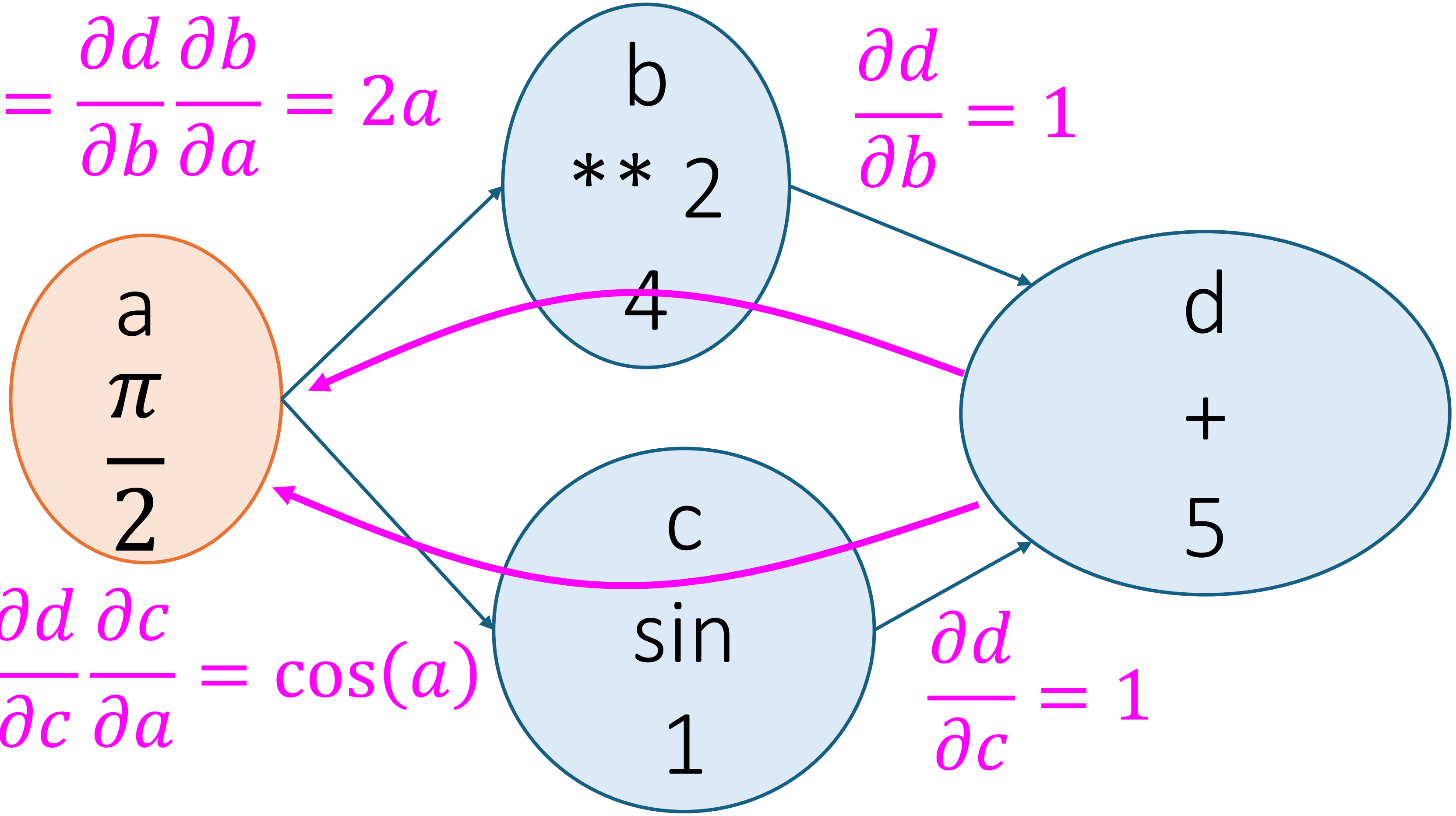
Olhando a fórmula de d, concluímos que a derivada verdadeira é dada por: $\dfrac{\partial d}{\partial a} = 2a + \cos(a)$

caminho 1

$$\frac{\partial d}{\partial a} = \frac{\partial d}{\partial b}\frac{\partial b}{\partial a} = 2a$$

derivada “total”

$$\frac{\partial d}{\partial a} = 2a + \cos(a)$$



caminho 2

$$\frac{\partial d}{\partial a} = \frac{\partial d}{\partial c}\frac{\partial c}{\partial a} = \cos(a)$$

Derivadas de Operações usadas em Redes Neurais

Multiplicação de Matrizes

$$y = Wx + b$$

$$l = loss(y, \hat{y})$$

$$\frac{\partial l}{\partial W} = \frac{\partial l}{\partial y}\frac{\partial y}{\partial W} = \frac{\partial l}{\partial y}x^T \qquad \frac{\partial l}{\partial b} = \frac{\partial l}{\partial y}\frac{\partial y}{\partial b} = \frac{\partial l}{\partial y}$$

$$\frac{\partial l}{\partial x} = \frac{\partial l}{\partial y}\frac{\partial y}{\partial x} = W^T\frac{\partial l}{\partial y}$$

ReLU

$$y = relu(z) = \begin{cases} z, se\ z > 0 \\ 0, c.c. \end{cases}$$

$$l = loss(y, \hat{y})$$

$$\frac{\partial l}{\partial z} = \frac{\partial l}{\partial y}\frac{\partial y}{\partial z} = \frac{\partial l}{\partial y} * \begin{bmatrix} dz_1 \\ \vdots \\ dz_n \end{bmatrix} \qquad dz_i = \begin{cases} 1, se\ z > 0 \\ 0, c.c. \end{cases}$$

Sigmoide

$$y = sigmoid(z) = \frac{1}{1 + e^{-x}}$$

$$l = loss(y, \hat{y})$$

$$\frac{\partial l}{\partial z} = \frac{\partial l}{\partial y}\frac{\partial y}{\partial z} = \frac{\partial l}{\partial y} * \begin{bmatrix} dz_1 \\ \vdots \\ dz_n \end{bmatrix} \qquad dz_i = sigm(x) * (1 - sigm(x))$$

Tanh

$$y = \tanh(z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}$$

$$l = loss(y, \hat{y})$$

$$\frac{\partial l}{\partial z} = \frac{\partial l}{\partial y}\frac{\partial y}{\partial z} = \frac{\partial l}{\partial y} * \begin{bmatrix} dz_1 \\ \vdots \\ dz_n \end{bmatrix} \qquad dz_i = 1 - \tanh(z)^2$$

Softmax

A softmax é uma função $\mathbb{R}^n \rightarrow \mathbb{R}^n$ e diferente das anteriores, a operação não é feita elemento a elemento. Todas as entradas afetam todas as saídas!

$$y = \mathrm{softmax}(z_1, \dots, z_n) = \left[\frac{e^{z_1}}{\sum_{j=1}^{n} e^{z_j}}, \dots, \frac{e^{z_n}}{\sum_{j=1}^{n} e^{z_j}}\right]$$

$$l = loss(y, \hat{y})$$

$$\frac{\partial l}{\partial z} = \frac{\partial l}{\partial y}\frac{\partial y}{\partial z}$$

Softmax

A softmax é uma função $\mathbb{R}^n \to \mathbb{R}^n$ e diferente das anteriores, a operação não é feita elemento a elemento. Todas as entradas afetam todas as saídas!

$$y = \text{softmax}(z_1, \dots, z_n) = \left[\frac{e^{z_1}}{\sum_{j=1}^{n} e^{z_j}}, \dots, \frac{e^{z_n}}{\sum_{j=1}^{n} e^{z_j}}\right]$$

$$l = loss(y, \hat{y})$$

Note que z_1 aparece no numerador do primeiro termo e em todos os denominadores!

$$\frac{\partial l}{\partial z} = \frac{\partial l}{\partial y}\frac{\partial y}{\partial z}$$

Softmax

A softmax é uma função $\mathbb{R}^n \rightarrow \mathbb{R}^n$ e diferente das anteriores, a operação não é feita elemento a elemento. Todas as entradas afetam todas as saídas!

$$y = \text{softmax}(z_1, \dots, z_n) = \left[\frac{e^{z_1}}{\sum_{j=1}^{n} e^{z_j}}, \dots, \frac{e^{z_n}}{\sum_{j=1}^{n} e^{z_j}}\right]$$

$$l = loss(y, \hat{y})$$

$$\frac{\partial l}{\partial z} = \mathrm{J}^T \frac{\partial l}{\partial y}$$

Nestes casos, precisamos computar a matriz **jacobiana** para propagar o gradiente do erro em relação a z.

$$
\mathrm{J} = \begin{pmatrix} \frac{\partial y_1}{\partial z_1} & \cdots & \frac{\partial y_1}{\partial z_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial z_1} & \cdots & \frac{\partial y_n}{\partial z_n} \end{pmatrix}
$$

O **jacobiano** contém as derivadas parciais de cada saída em relação à cada entrada

caso 1: diagonal (derivada da saída i em relação à entrada i).

$$
\frac{\partial y_i}{\partial z_i} = \frac{\partial}{\partial z_i} \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} = \frac{\partial}{\partial z_i} e^{z_i} \left[\frac{1}{\sum_{j=1}^{n} e^{z_j}} \right] + e^{z_i} \frac{\partial}{\partial z_i} \frac{1}{\sum_{j=1}^{n} e^{z_j}}
$$

$$
= \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} + e^{z_i} \frac{\partial}{\partial z_i} \left(\sum_{j=1}^{n} e^{z_j} \right)^{-1}
$$

$$J = \begin{pmatrix} \frac{\partial y_1}{\partial z_1} & \cdots & \frac{\partial y_1}{\partial z_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial z_1} & \cdots & \frac{\partial y_n}{\partial z_n} \end{pmatrix}$$

O **jacobiano** contém as derivadas parciais de cada saída em relação à cada entrada

caso 1: diagonal (derivada da saída i em relação à entrada i).

$$\frac{\partial y_i}{\partial z_i} = \frac{\partial}{\partial z_i} \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} = \frac{\partial}{\partial z_i} e^{z_i} \left[\frac{1}{\sum_{j=1}^{n} e^{z_j}} \right] + e^{z_i} \frac{\partial}{\partial z_i} \frac{1}{\sum_{j=1}^{n} e^{z_j}}$$

$$= \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} + e^{z_i}(-1) \left(\sum_{j=1}^{n} e^{z_j} \right)^{-2} \frac{\partial}{\partial z_i} \sum_{j=1}^{n} e^{z_j}$$

$$J = \begin{pmatrix} \frac{\partial y_1}{\partial z_1} & \cdots & \frac{\partial y_1}{\partial z_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial z_1} & \cdots & \frac{\partial y_n}{\partial z_n} \end{pmatrix}$$

O **jacobiano** contém as derivadas parciais de cada saída em relação à cada entrada

caso 1: diagonal (derivada da saída i em relação à entrada i).

$$\frac{\partial y_i}{\partial z_i} = \frac{\partial}{\partial z_i} \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} = \frac{\partial}{\partial z_i} e^{z_i} \left[\frac{1}{\sum_{j=1}^{n} e^{z_j}} \right] + e^{z_i} \frac{\partial}{\partial z_i} \frac{1}{\sum_{j=1}^{n} e^{z_j}}$$

$$= \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} - \left(\frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} \right)^2$$

$$J = \begin{pmatrix} \frac{\partial y_1}{\partial z_1} & \cdots & \frac{\partial y_1}{\partial z_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial z_1} & \cdots & \frac{\partial y_n}{\partial z_n} \end{pmatrix}$$

O **jacobiano** contém as derivadas parciais de cada saída em relação à cada entrada

caso 1: diagonal (derivada da saída i em relação à entrada i).

$$\frac{\partial y_i}{\partial z_i} = \frac{\partial}{\partial z_i} \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} = \frac{\partial}{\partial z_i} e^{z_i} \left[\frac{1}{\sum_{j=1}^{n} e^{z_j}}\right] + e^{z_i} \frac{\partial}{\partial z_i} \frac{1}{\sum_{j=1}^{n} e^{z_j}}$$

$$= y_i - y_i y_i$$

$$= \boxed{y_i(1 - y_i)}$$

$$J = \begin{pmatrix} \frac{\partial y_1}{\partial z_1} & \cdots & \frac{\partial y_1}{\partial z_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial z_1} & \cdots & \frac{\partial y_n}{\partial z_n} \end{pmatrix}$$

O **jacobiano** contém as derivadas parciais de cada saída em relação à cada entrada

caso 2: fora da diagonal (derivada da saída k em relação à entrada i, com k != i).

$$\frac{\partial y_k}{\partial z_i} = \frac{\partial}{\partial z_i} \frac{e^{z_k}}{\sum_{j=1}^{n} e^{z_j}} = \frac{\partial}{\partial z_i} e^{z_k} \left[\frac{1}{\sum_{j=1}^{n} e^{z_j}} \right] + e^{z_k} \frac{\partial}{\partial z_i} \frac{1}{\sum_{j=1}^{n} e^{z_j}}$$

$$J = \begin{pmatrix} \frac{\partial y_1}{\partial z_1} & \cdots & \frac{\partial y_1}{\partial z_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial z_1} & \cdots & \frac{\partial y_n}{\partial z_n} \end{pmatrix}$$

O **jacobiano** contém as derivadas parciais de cada saída em relação à cada entrada

caso 2: fora da diagonal (derivada da saída k em relação à entrada i, com k != i).

$$\frac{\partial y_k}{\partial z_i} = \frac{\partial}{\partial z_i} \frac{e^{z_k}}{\sum_{j=1}^{n} e^{z_j}} = \cancel{\frac{\partial}{\partial z_i} e^{z_k} \left[\frac{1}{\sum_{j=1}^{n} e^{z_j}} \right]} + e^{z_k} \frac{\partial}{\partial z_i} \frac{1}{\sum_{j=1}^{n} e^{z_j}}$$

$$J = \begin{pmatrix} \frac{\partial y_1}{\partial z_1} & \cdots & \frac{\partial y_1}{\partial z_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial z_1} & \cdots & \frac{\partial y_n}{\partial z_n} \end{pmatrix}$$

O **jacobiano** contém as derivadas parciais de cada saída em relação à cada entrada

caso 2: fora da diagonal (derivada da saída k em relação à entrada i, com k != i).

$$\frac{\partial y_k}{\partial z_i} = \frac{\partial}{\partial z_i} \frac{e^{z_k}}{\sum_{j=1}^{n} e^{z_j}} = \cancel{\frac{\partial}{\partial z_i} e^{z_k} \left[\frac{1}{\sum_{j=1}^{n} e^{z_j}}\right]} + e^{z_k} \frac{\partial}{\partial z_i} \frac{1}{\sum_{j=1}^{n} e^{z_j}}$$

$$= e^{z_k} \frac{\partial}{\partial z_i} \left(\sum_{j=1}^{n} e^{z_j}\right)^{-1}$$

$$J = \begin{pmatrix} \frac{\partial y_1}{\partial z_1} & \cdots & \frac{\partial y_1}{\partial z_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial z_1} & \cdots & \frac{\partial y_n}{\partial z_n} \end{pmatrix}$$

O **jacobiano** contém as derivadas parciais de cada saída em relação à cada entrada

caso 2: fora da diagonal (derivada da saída k em relação à entrada i, com k != i).

$$\frac{\partial y_k}{\partial z_i} = \frac{\partial}{\partial z_i} \frac{e^{z_k}}{\sum_{j=1}^{n} e^{z_j}} = \cancel{\frac{\partial}{\partial z_i} e^{z_k} \left[\frac{1}{\sum_{j=1}^{n} e^{z_j}} \right]} + e^{z_k} \frac{\partial}{\partial z_i} \frac{1}{\sum_{j=1}^{n} e^{z_j}}$$

$$= e^{z_k}(-1) \left(\sum_{j=1}^{n} e^{z_j} \right)^{-2} \frac{\partial}{\partial z_i} \sum_{j=1}^{n} e^{z_j}$$

$$J = \begin{pmatrix} \frac{\partial y_1}{\partial z_1} & \cdots & \frac{\partial y_1}{\partial z_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial z_1} & \cdots & \frac{\partial y_n}{\partial z_n} \end{pmatrix}$$

O **jacobiano** contém as derivadas parciais de cada saída em relação à cada entrada

caso 2: fora da diagonal (derivada da saída k em relação à entrada i, com k != i).

$$\frac{\partial y_k}{\partial z_i} = \frac{\partial}{\partial z_i} \frac{e^{z_k}}{\sum_{j=1}^{n} e^{z_j}} = \cancel{\frac{\partial}{\partial z_i} e^{z_k} \left[\frac{1}{\sum_{j=1}^{n} e^{z_j}} \right]} + e^{z_k} \frac{\partial}{\partial z_i} \frac{1}{\sum_{j=1}^{n} e^{z_j}}$$

$$= -\frac{e^{z_k} e^{z_i}}{\left(\sum_{j=1}^{n} e^{z_j}\right)^2}$$

$$J = \begin{pmatrix} \frac{\partial y_1}{\partial z_1} & \cdots & \frac{\partial y_1}{\partial z_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial z_1} & \cdots & \frac{\partial y_n}{\partial z_n} \end{pmatrix}$$

O **jacobiano** contém as derivadas parciais de cada saída em relação à cada entrada

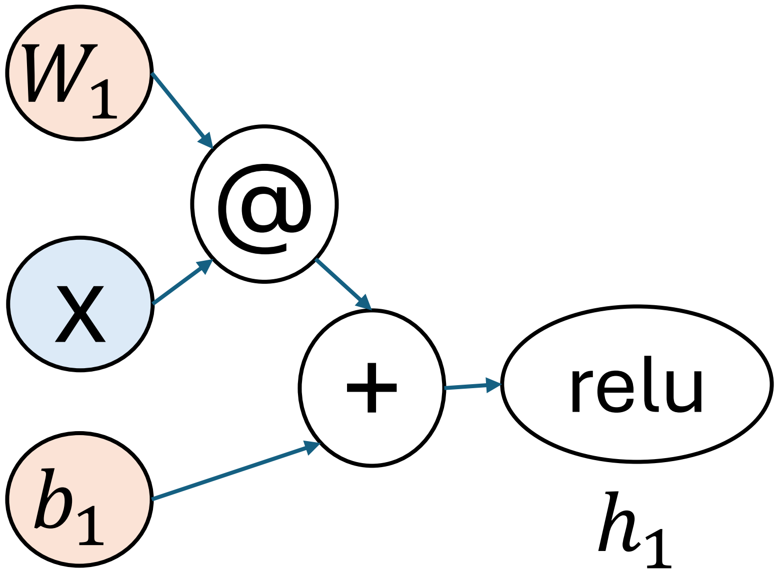
caso 2: fora da diagonal (derivada da saída k em relação à entrada i, com k != i).

$$\frac{\partial y_k}{\partial z_i} = \frac{\partial}{\partial z_i} \frac{e^{z_k}}{\sum_{j=1}^{n} e^{z_j}} = \cancel{\frac{\partial}{\partial z_i} e^{z_k} \left[\frac{1}{\sum_{j=1}^{n} e^{z_j}}\right]} + e^{z_k} \frac{\partial}{\partial z_i} \frac{1}{\sum_{j=1}^{n} e^{z_j}}$$

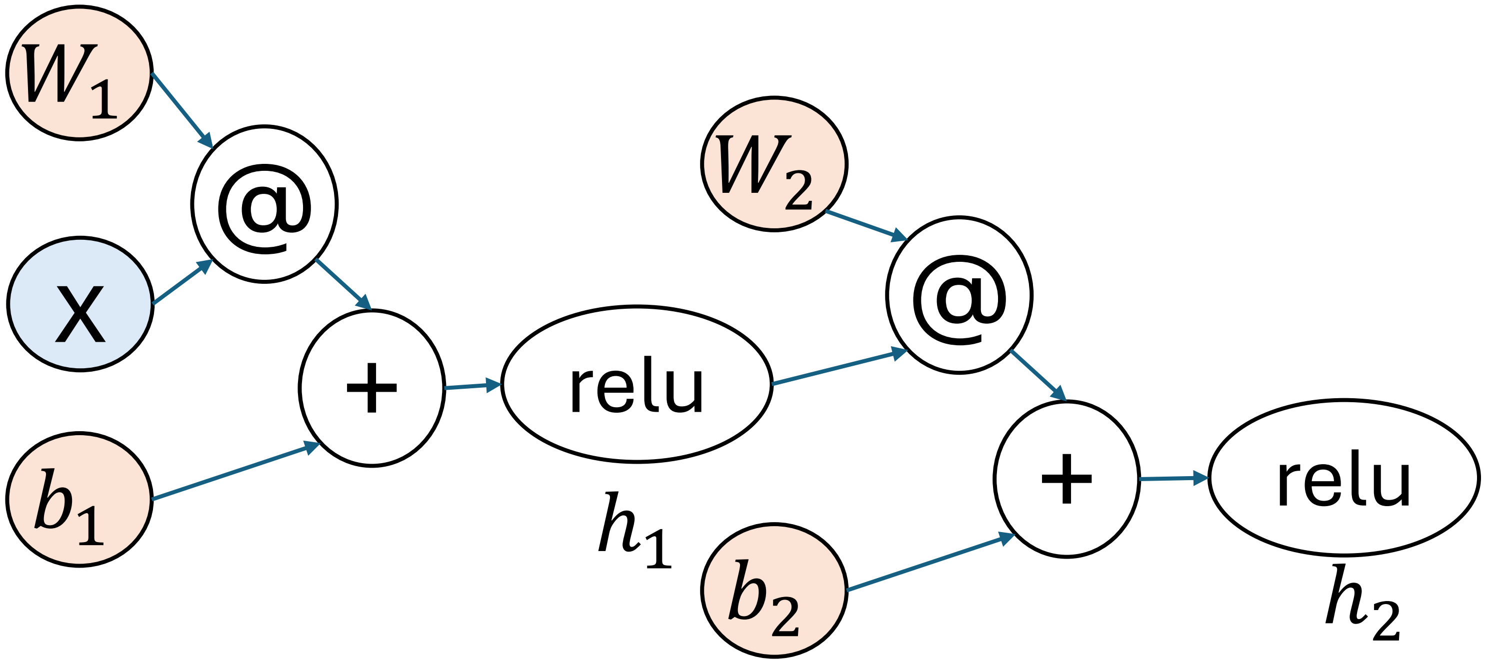
$$= \boxed{-y_k y_i}$$

Redes Neurais com Grafos Computacionais

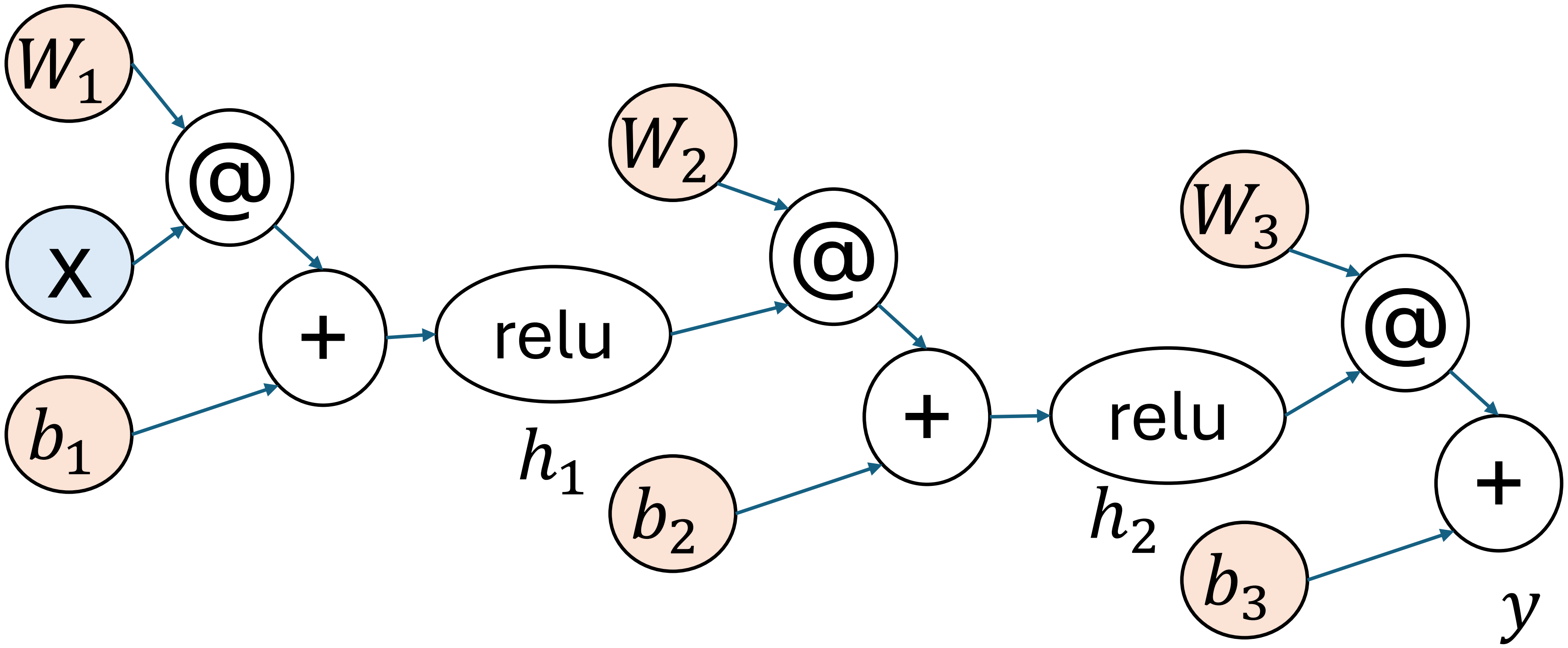
$$h_1 = relu(W_1 x + b_1)$$
$$h_2 = relu(W_2 h_1 + b_2)$$
$$y = W_3 h_2 + b_3$$
$$l = loss(y, \hat{y})$$



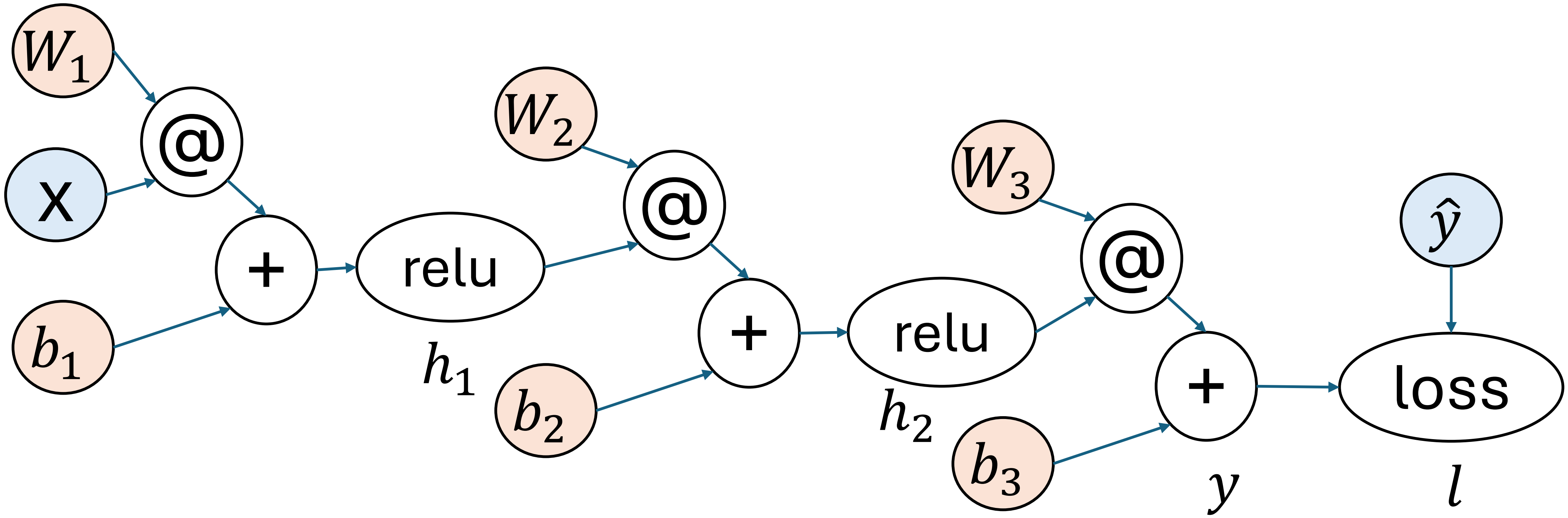
$$h_1 = relu(W_1 x + b_1)$$
$$h_2 = relu(W_2 h_1 + b_2)$$
$$y = W_3 h_2 + b_3$$
$$l = loss(y, \hat{y})$$



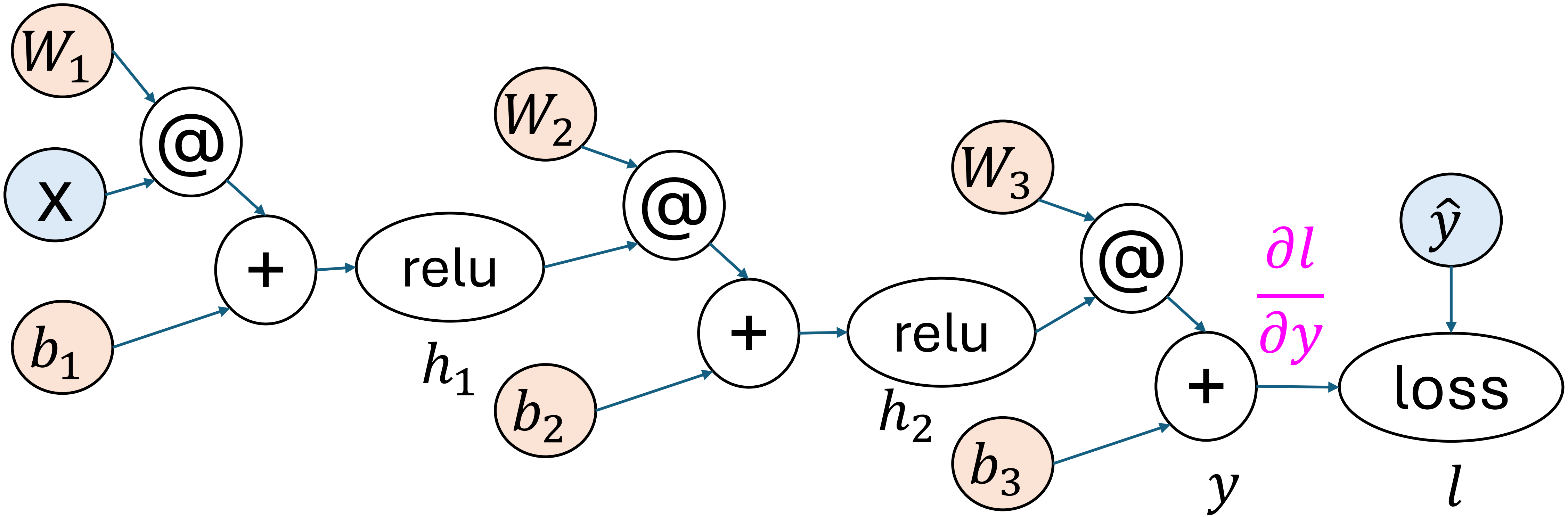
$$h_1 = relu(W_1 x + b_1)$$
$$h_2 = relu(W_2 h_1 + b_2)$$
$$y = W_3 h_2 + b_3$$
$$l = loss(y, \hat{y})$$



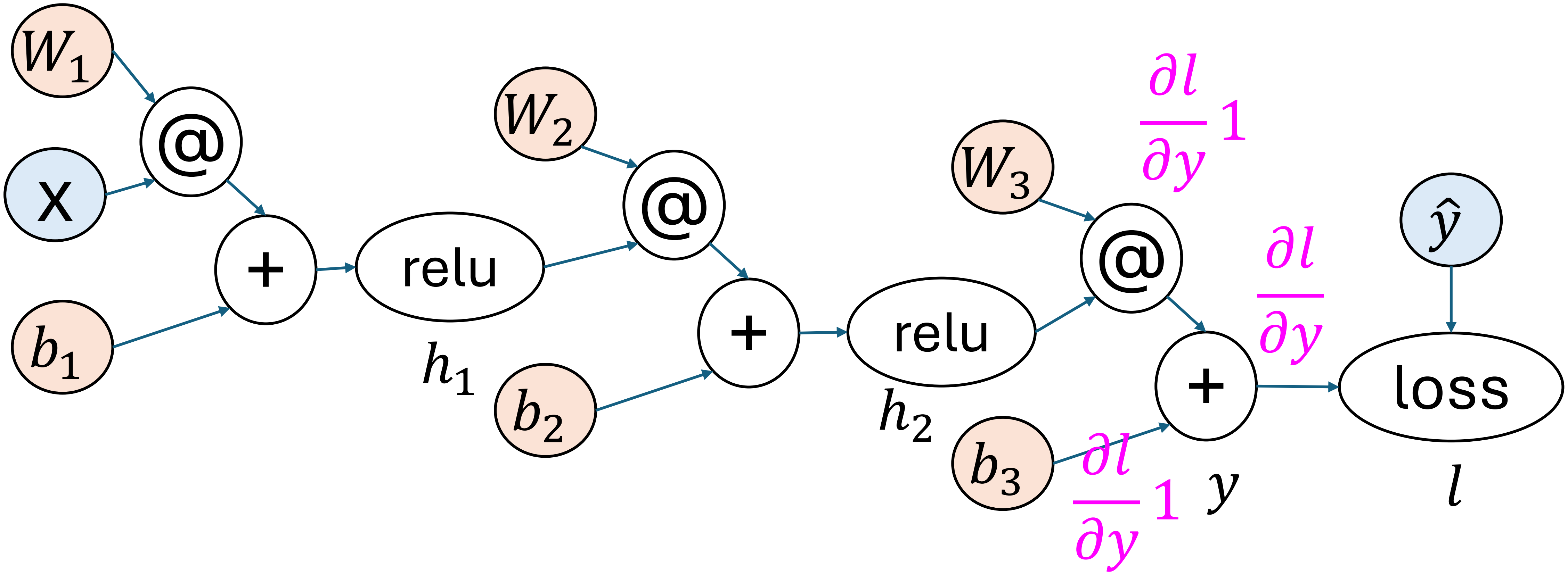
$$h_1 = relu(W_1 x + b_1)$$
$$h_2 = relu(W_2 h_1 + b_2)$$
$$y = W_3 h_2 + b_3$$
$$l = loss(y, \hat{y})$$



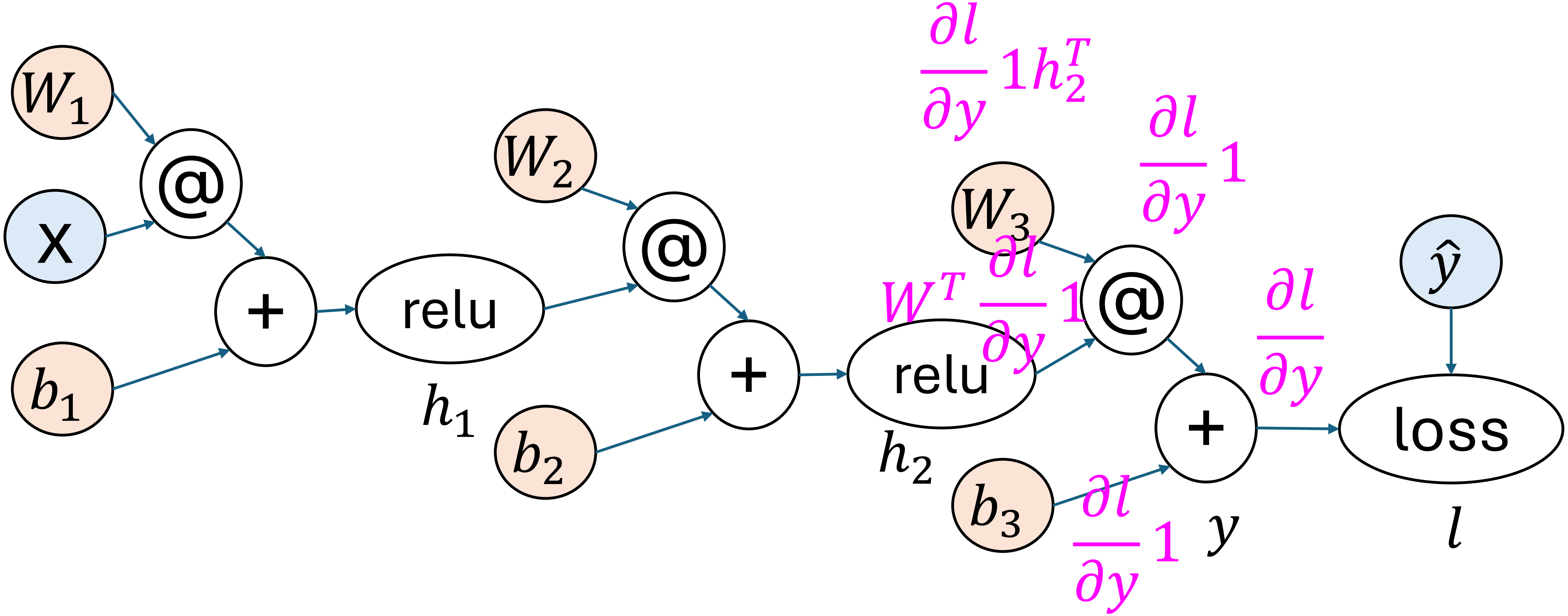
$$h_1 = relu(W_1 x + b_1)$$
$$h_2 = relu(W_2 h_1 + b_2)$$
$$y = W_3 h_2 + b_3$$
$$l = loss(y, \hat{y})$$



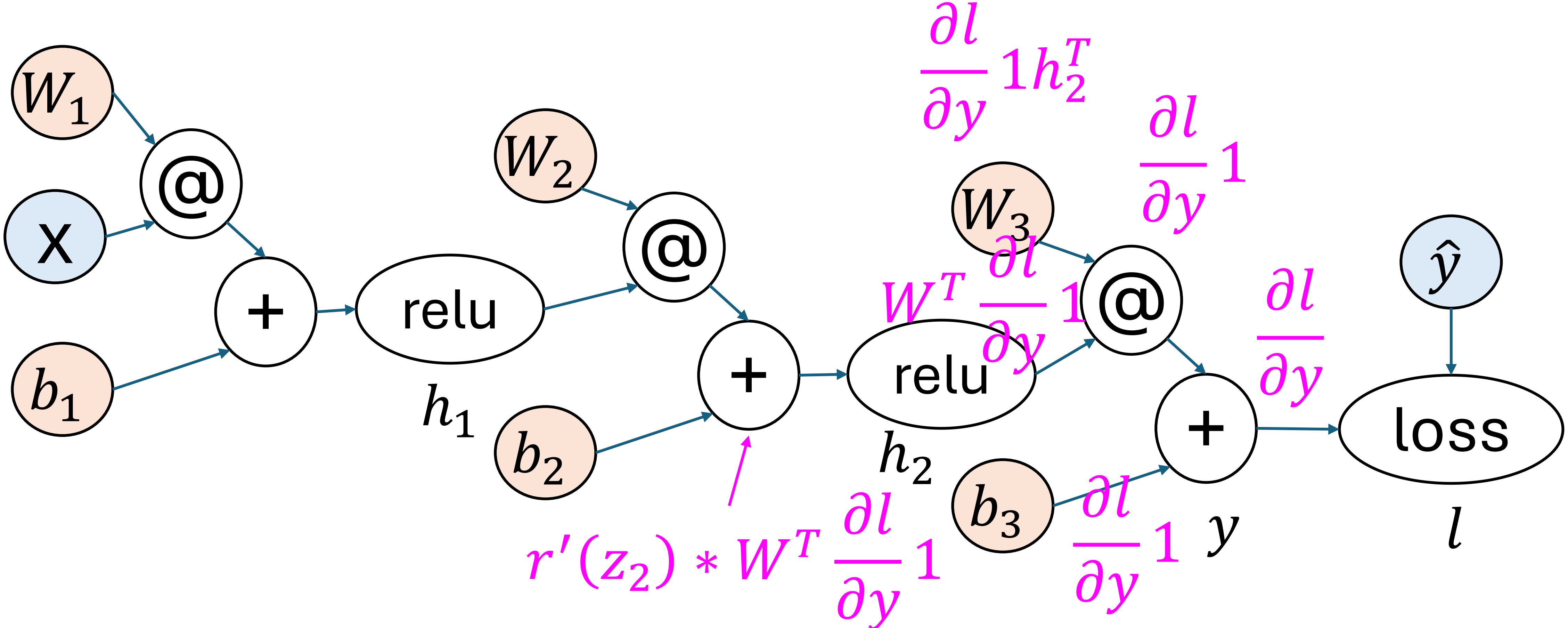
$$h_1 = relu(W_1 x + b_1)$$
$$h_2 = relu(W_2 h_1 + b_2)$$
$$y = W_3 h_2 + b_3$$
$$l = loss(y, \hat{y})$$



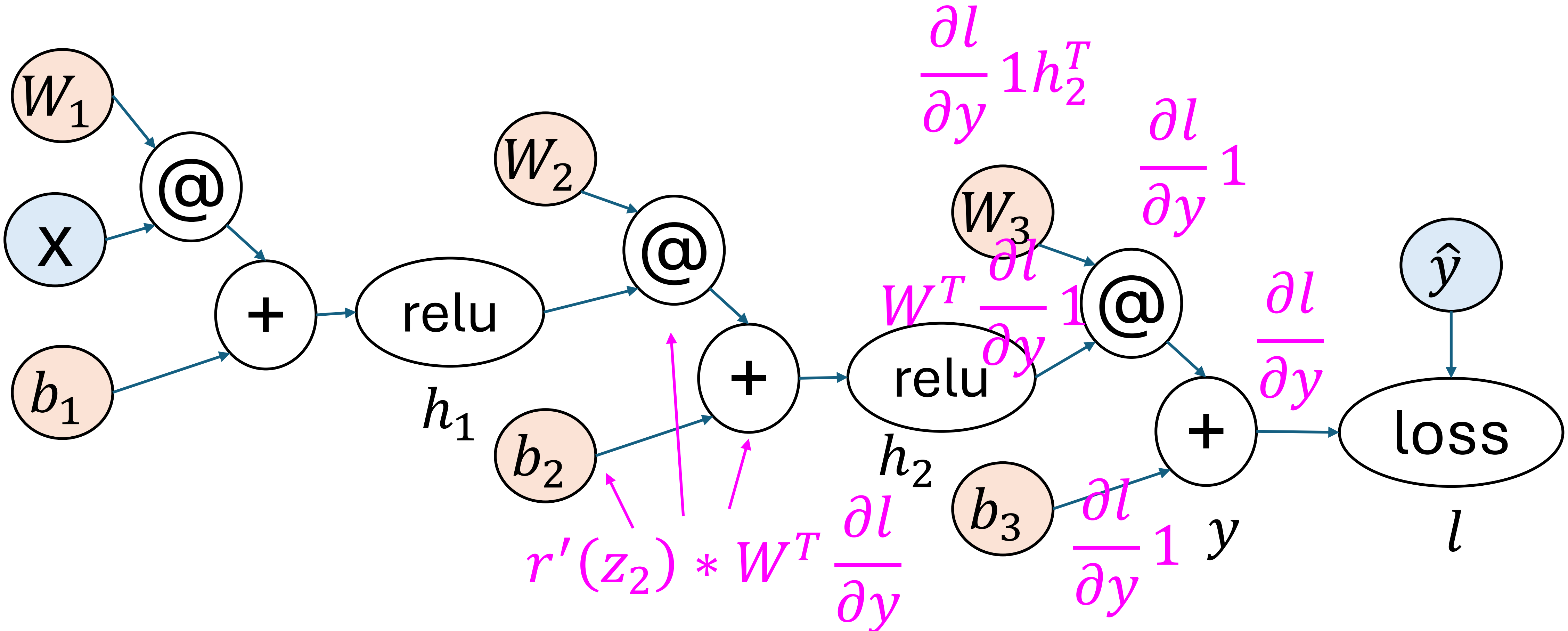
$h_1 = relu(W_1 x + b_1)$
$h_2 = relu(W_2 h_1 + b_2)$
$y = W_3 h_2 + b_3$
$l = loss(y, \hat{y})$



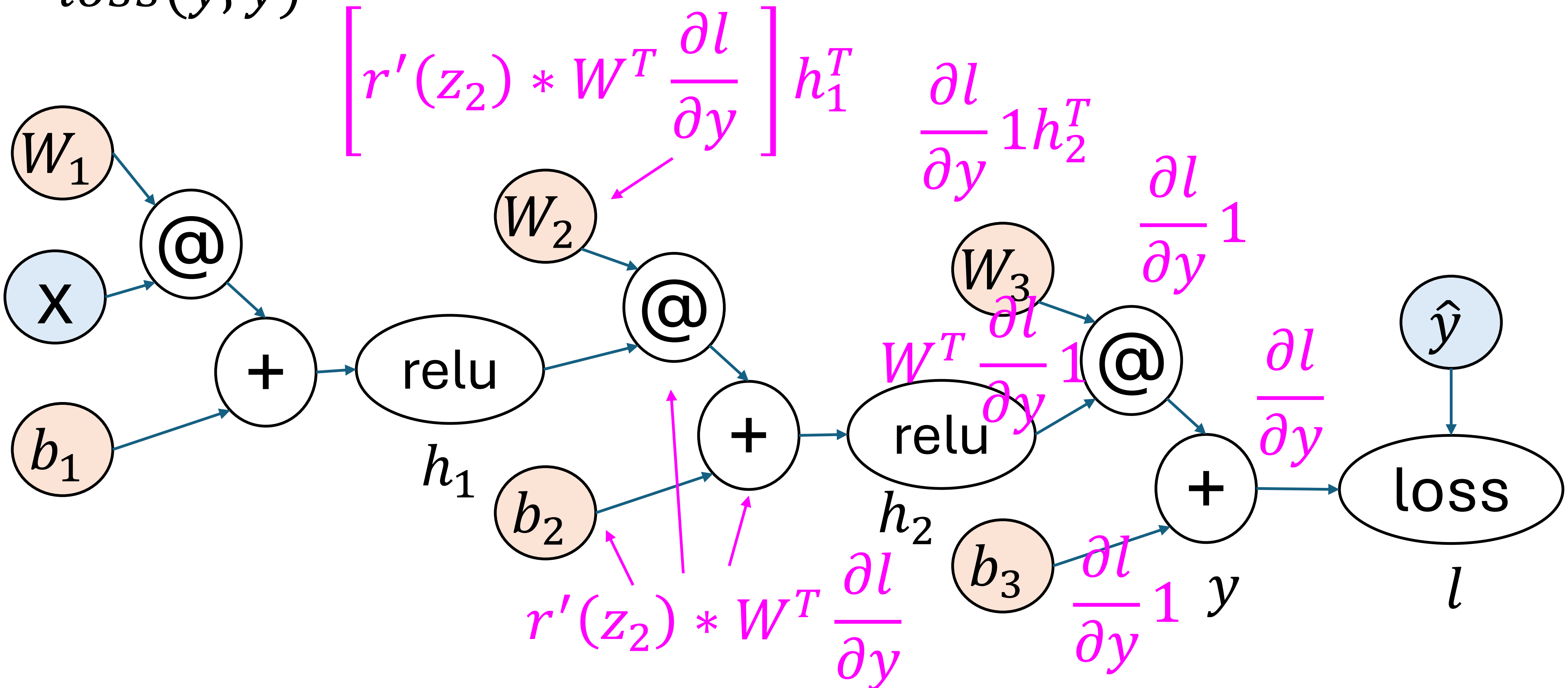
$$h_1 = relu(W_1 x + b_1)$$
$$h_2 = relu(W_2 h_1 + b_2)$$
$$y = W_3 h_2 + b_3$$
$$l = loss(y, \hat{y})$$



$$h_1 = relu(W_1 x + b_1)$$
$$h_2 = relu(W_2 h_1 + b_2)$$
$$y = W_3 h_2 + b_3$$
$$l = loss(y, \hat{y})$$



$$h_1 = relu(W_1 x + b_1)$$
$$h_2 = relu(W_2 h_1 + b_2)$$
$$y = W_3 h_2 + b_3$$
$$l = loss(y, \hat{y})$$



$$h_1 = relu(W_1 x + b_1)$$
$$h_2 = relu(W_2 h_1 + b_2)$$
$$y = W_3 h_2 + b_3$$
$$l = loss(y, \hat{y})$$