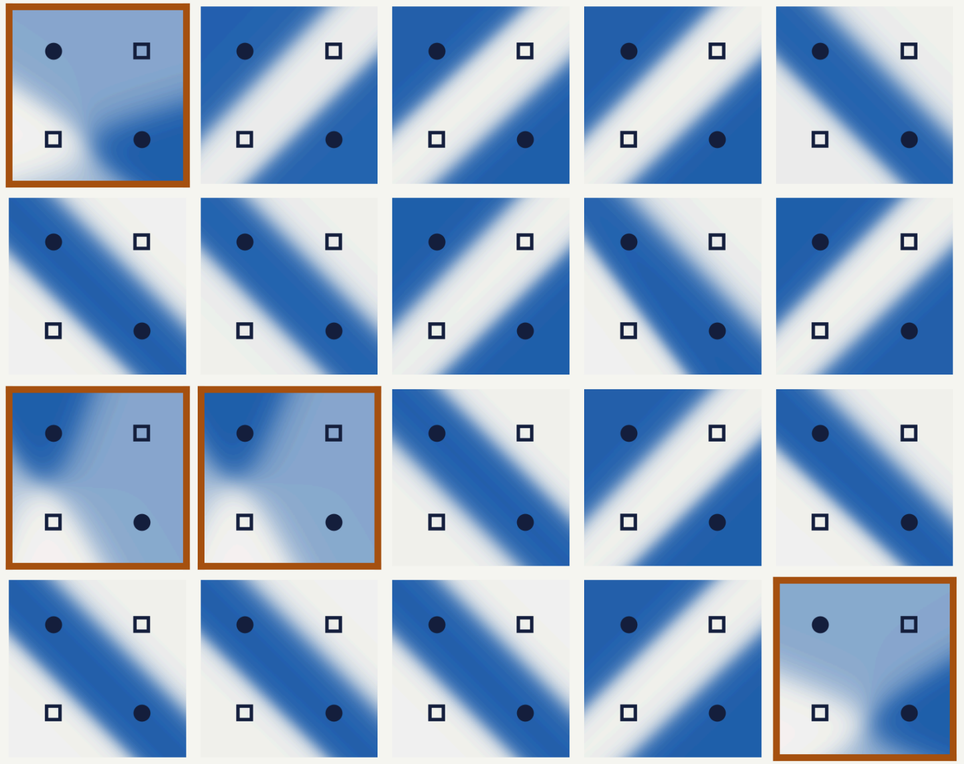
SIENA, NIPS-CERN, UFJF

Limiar

Redes neurais, do neurônio ao fractal

Aula 5: retropropagação

13/11/2026. C∴ Chrysthofer A. A. Afonso



O XOR como 20 redes iguais o aprenderam, uma por semente: azul onde a rede responde 1. Nas 4 de moldura laranja, o treino parou no patamar.

1 · REVISÃO

O que ficou das aulas 3 e 4, de memória

1. O que uma camada calcula, escrita como matriz?

2. Por que a camada oculta precisa de uma ativação que não seja linear?

3. O que diz a regra da cadeia?

4. Que pesos à mão calculam o XOR?

1 · REVISÃO

O fio do curso, da aula 1 à 10

1	**O neurônio e o perceptron**	um neurônio desenha uma reta, e nenhuma reta separa o XOR	25/09, presencial
2	**Derivada e gradiente**	o passo contra a derivada; a taxa grande demais faz o treino divergir	02/10, presencial
3	**Matrizes e camadas**	duas camadas separam o XOR, com pesos escolhidos à mão	13/11, on-line
4	**O Jacobiano**	a regra da cadeia com vetores, uma camada de cada vez	13/11, on-line
5	**Retropropagação**	a regra da cadeia, de trás para a frente, acha os pesos do XOR	13/11, on-line
6	**O MNIST em NumPy**	um MLP em 70 mil dígitos escritos à mão, com treino, validação e teste	27/11, on-line
7	**PyTorch**	a biblioteca faz a retropropagação; o erro final contra a taxa	27/11, on-line
8	**Os dados do CERN**	um MLP estima a massa dos dois múons, conferida pela fórmula	04/12, on-line
9	**O fractal e o projeto**	a fronteira entre a taxa que converge e a que diverge	04/12, on-line
10	**A apresentação**	a apresentação do projeto e a revisão do curso inteiro	11/12, on-line

2 · A regra da cadeia na rede

Como achar os nove números da rede sem escolher à mão?

A pergunta que fechou a aula 3.

A regra da cadeia: as taxas se multiplicam

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial z} \cdot \frac{\partial z}{\partial w}$$

UM EXEMPLO

$$z = 3w$$

$$L = z^2$$

$$\frac{\partial L}{\partial w} = 2z \cdot 3 = 18w$$

conferido sem a regra: $L = 9w^2$, e a derivada é $18w$

UM NEURÔNIO COM A SIGMOIDE, UM EXEMPLO

$$L = (y - t)^2$$

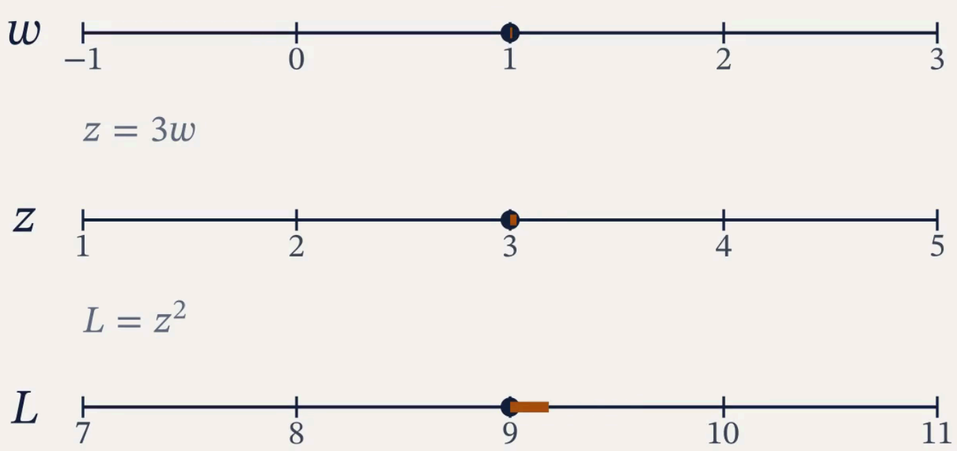
$$y = \sigma(z)$$

$$z = w\,x + b$$

$$\frac{\partial L}{\partial w} = \underbrace{2\,(y - t)}_{\partial L/\partial y}\ \underbrace{\sigma'(z)}_{\partial y/\partial z}\ \underbrace{x}_{\partial z/\partial w}$$

As taxas se multiplicam

A regra da cadeia: $\partial L/\partial w = 2z \cdot 3 = 18w$



$\Delta w = 0{,}01$

$\Delta z = 0{,}03$

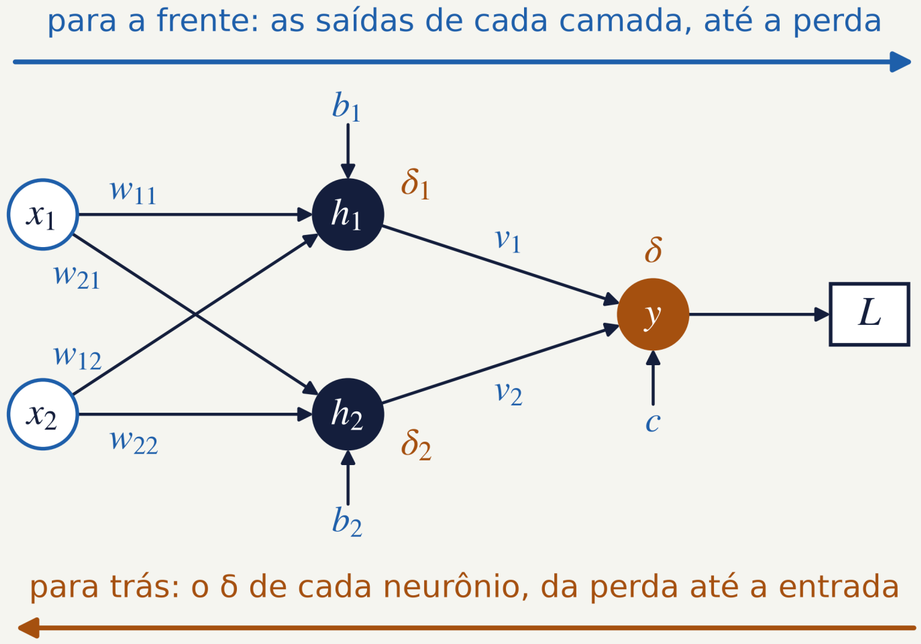
$\Delta z/\Delta w = 3{,}00$

$\Delta L = 0{,}1809$

$\Delta L/\Delta z = 6{,}03$

$\Delta L/\Delta w = 18{,}09$

Para a frente, os valores; para trás, as derivadas



O DELTA: A DERIVADA DA PERDA PELA SOMA DO NEURÔNIO

$$\delta = 2\,(y - t)\,\sigma'(z)$$

δ, delta; o da saída, para um exemplo

$$\delta_k = \delta\, v_k\, \sigma'(z_k)$$

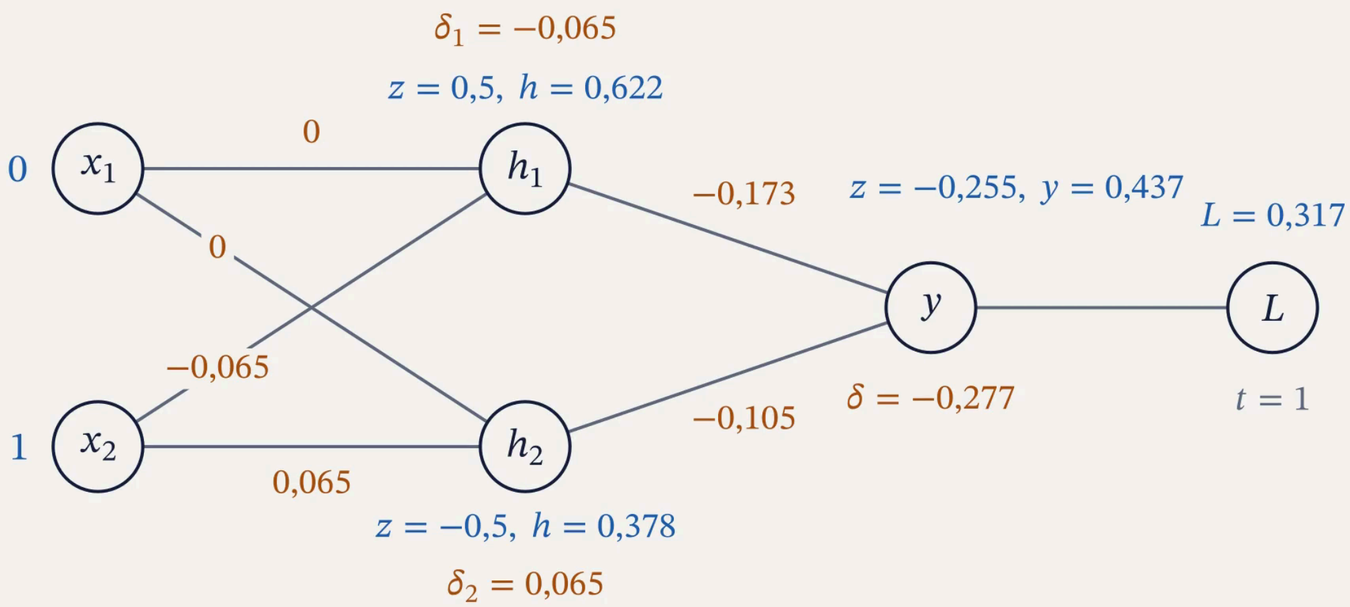
$$\frac{\partial L}{\partial v_k} = \delta\, h_k$$

$$\frac{\partial L}{\partial w_{kj}} = \delta_k\, x_j$$

cada peso: o δ do neurônio aonde ele chega vezes a entrada que ele multiplica

A retropropagação, com números

As nove derivadas, numa passada para trás



$$\partial L/\partial w_{12} = \delta_1\, x_2 = -0{,}065 \cdot 1 = -0{,}065$$

Um passo com um exemplo só

	à mão	derivada	depois
w_{11}	1	0	1,000
w_{12}	1	−0,065	1,130
w_{21}	1	0	1,000
w_{22}	1	0,065	0,870
b_1	−0,5	−0,065	−0,370
b_2	−1,5	0,065	−1,630
v_1	1	−0,173	1,345
v_2	−1	−0,105	−0,791
c	−0,5	−0,277	0,054

O exemplo (0; 1), sozinho; o passo é novo = velho − η · derivada, com η = 2.

entrada	certa	antes	depois
(0; 0)	0	0,424	**0,616**
(0; 1)	1	0,437	**0,672**
(1; 0)	1	0,437	**0,659**
(1; 1)	0	0,424	**0,681**

A saída sobe nas quatro entradas, também onde a certa é 0. A perda dos quatro piora: de 0,249 para 0,267.

Os quatro exemplos de uma vez

$$L = \frac{1}{4}\sum_{i=1}^{4}(y_i - t_i)^2$$

a perda dos quatro é a média; a derivada também

	(0; 0)	(0; 1)	(1; 0)	(1; 1)	perda
resposta certa	0	1	1	0	
antes	0,424	0,437	0,437	0,424	0,2488
um passo, só (0; 1)	0,616	0,672	0,659	0,681	0,2670
um passo, os quatro	0,449	0,468	0,468	0,459	0,2448
200 passos, os quatro	**0,188**	**0,816**	**0,816**	**0,202**	0,0359

Os pesos da aula 3, que erravam com a sigmoide, acertam o XOR em 200 passos com os quatro exemplos.

3 · CONFERIR E TREINAR

A retropropagação confere com a derivada numérica

derivada	retropropagação	numérica
$\partial L/\partial w_{11}$	0,001178	0,001178
$\partial L/\partial w_{12}$	0,000624	0,000624
$\partial L/\partial w_{21}$	0,002585	0,002585
$\partial L/\partial w_{22}$	0,001618	0,001618
$\partial L/\partial b_1$	−0,002508	−0,002508
$\partial L/\partial b_2$	0,005113	0,005113
$\partial L/\partial v_1$	0,018787	0,018787
$\partial L/\partial v_2$	0,015281	0,015281
$\partial L/\partial c$	0,026299	0,026299

semente 1, nos pesos iniciais

$$\text{maior diferença} = 2{,}8 \cdot 10^{-11}$$

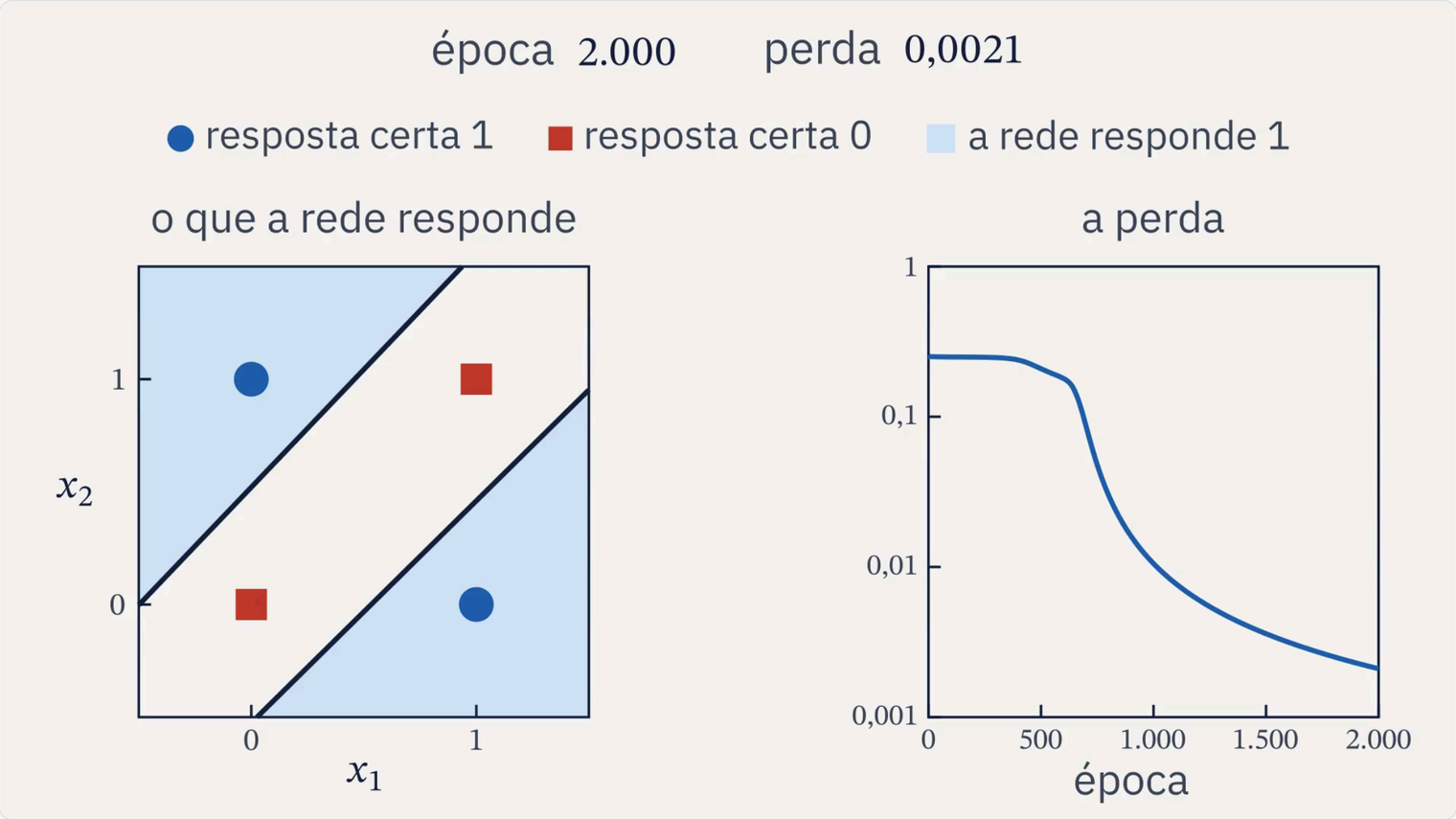
DERIVADA NUMÉRICA

duas passadas para a frente por parâmetro: 18 nesta rede

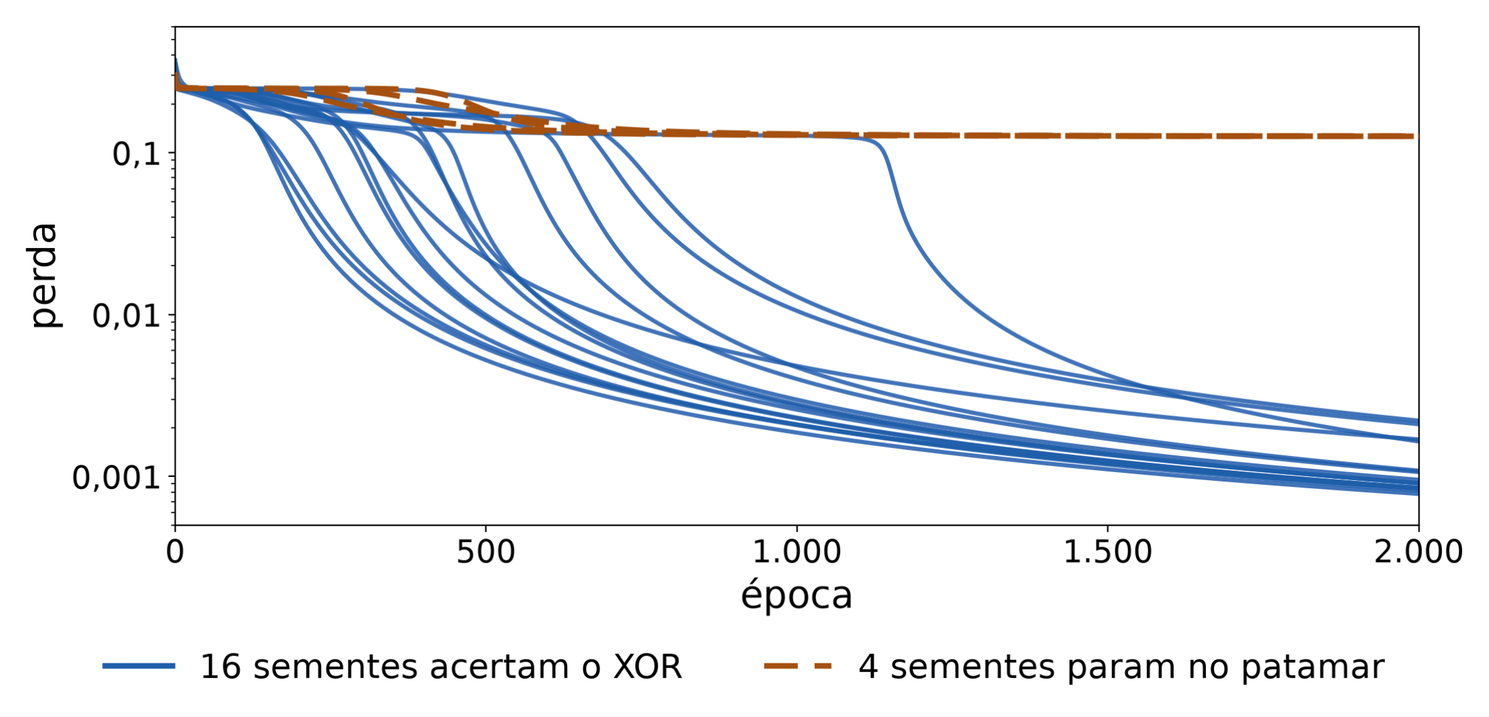
RETROPROPAGAÇÃO

uma passada para a frente e uma para trás, para os nove de uma vez

A retropropagação acha os pesos do XOR



Nem toda semente chega



Vinte sementes, vinte sorteios dos pesos iniciais, com $\eta = 2$ e 2.000 épocas: 16 acertam o XOR, e 4 param com a perda perto de 0,126, respondendo 0,5 em duas entradas.

1986: Rumelhart, Hinton e Williams

> *“Em consequência dos ajustes dos pesos, unidades internas, ‘ocultas’, que não fazem parte da entrada nem da saída, passam a representar características importantes do domínio da tarefa.”*

Rumelhart, Hinton e Williams, Nature, 9 de outubro de 1986, tradução livre

- a retropropagação, em quatro páginas da Nature
- a conta de trás para a frente é mais antiga: Linnainmaa, 1970; Werbos, 1974
- Hinton dividiu o Nobel de Física de 2024 com John Hopfield



Da esquerda para a direita: Paul Werbos e David Rumelhart, em 1991, fotos de Rolf Kickuth, CC BY-SA 4.0; Geoffrey Hinton, em 2013, foto de Eviatar Bach, CC BY-SA 3.0. Wikimedia Commons.

5 · O CÓDIGO

A retropropagação em NumPy, passo a passo

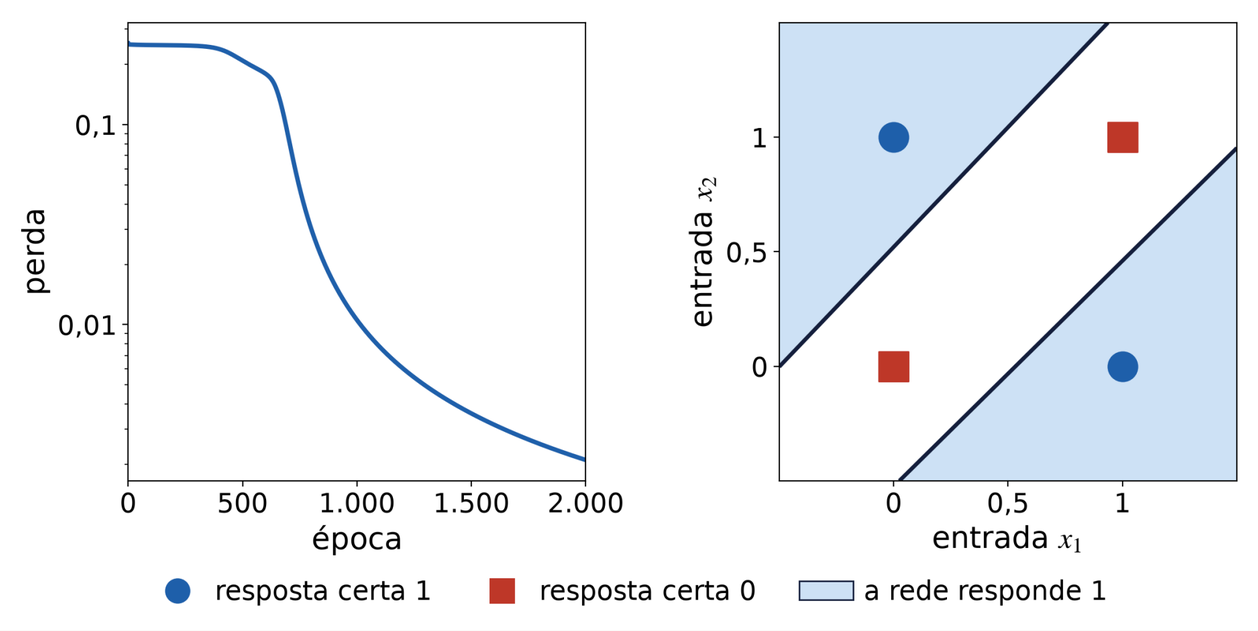
1. Os quatro exemplos do XOR e a semente que sorteia os pesos
2. A passada para a frente com a sigmoide, guardando a saída de cada camada
3. A retropropagação: o δ da saída, o δ de cada neurônio oculto, as nove derivadas
4. Cada derivada conferida pela derivada numérica, antes de treinar
5. O laço do gradiente descendente, guardando a perda de cada época

O NUMPY QUE RESOLVE

```
X @ W.T + b
1 / (1 + np.exp(-z))
s * (1 - s)
H.T @ delta
```

com X, os quatro exemplos de uma vez, um por linha

O que a entrega pede



O XOR treinado com a própria retropropagação, em NumPy, com as nove derivadas conferidas pela derivada numérica. A semente, a taxa e as épocas vão junto com o resultado.

6 · FECHAMENTO

Para estudar e entregar

Estudo

- 3Blue1Brown, redes neurais, capítulo 3
- 3Blue1Brown, redes neurais, capítulo 4
- PyTorch, “Learn the Basics”

Entrega

- o XOR treinado com a própria retropropagação, em NumPy
- o gradiente conferido pela derivada numérica
- uma pergunta sobre o estudo

A seguir

- aula 6: o MNIST em NumPy, com treino, validação e teste
- aula 7: o PyTorch faz a retropropagação sozinho