



Insper

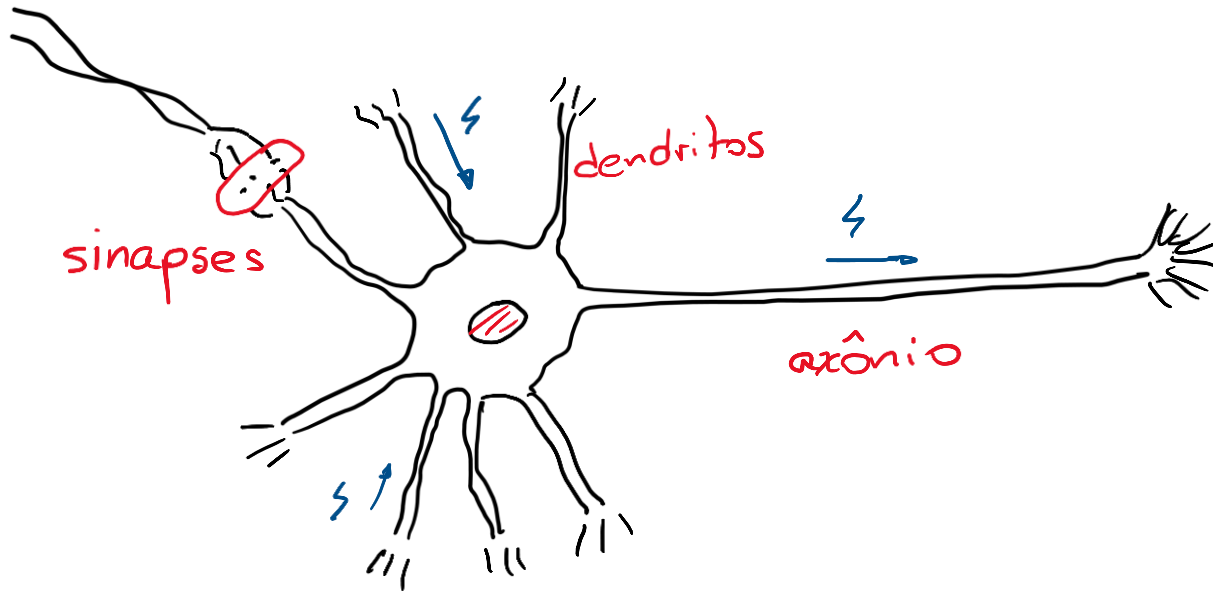
# Machine Learning

Introdução à redes neurais

Fábio Ayres <[fabioja@insper.edu.br](mailto:fabioja@insper.edu.br)>

# Neurônio biológico

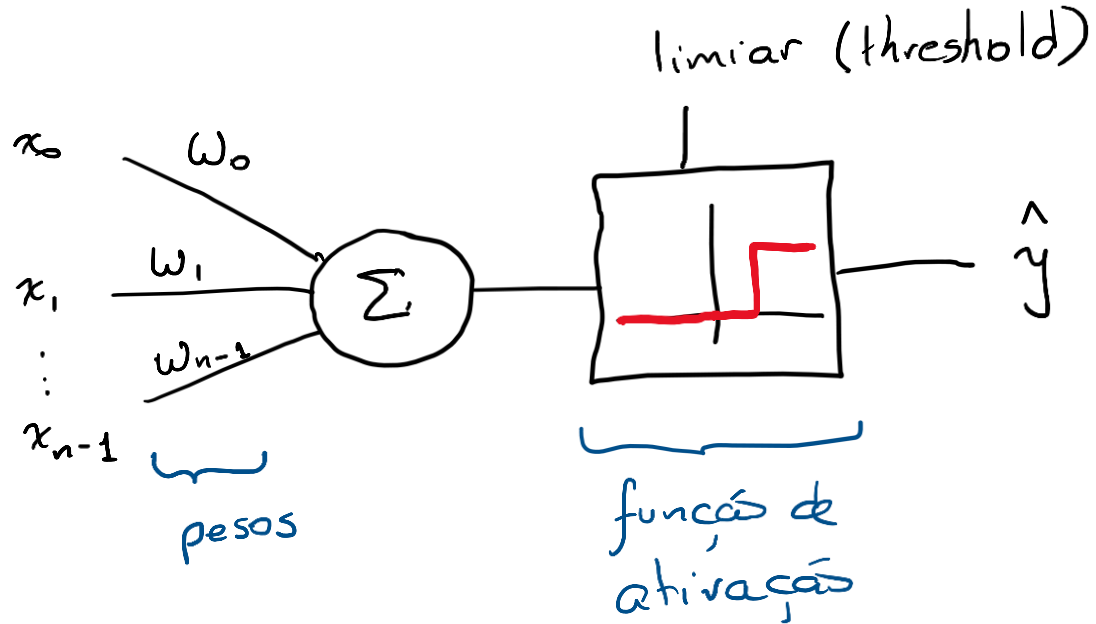
2



$$\sum_{\text{dendritos entrada}} (\text{impulso na entrada}) \cdot (\text{peso da sinapse}) > \text{limiar} \Rightarrow \text{pulso elétrico na saída}$$

# Neurônios artificiais

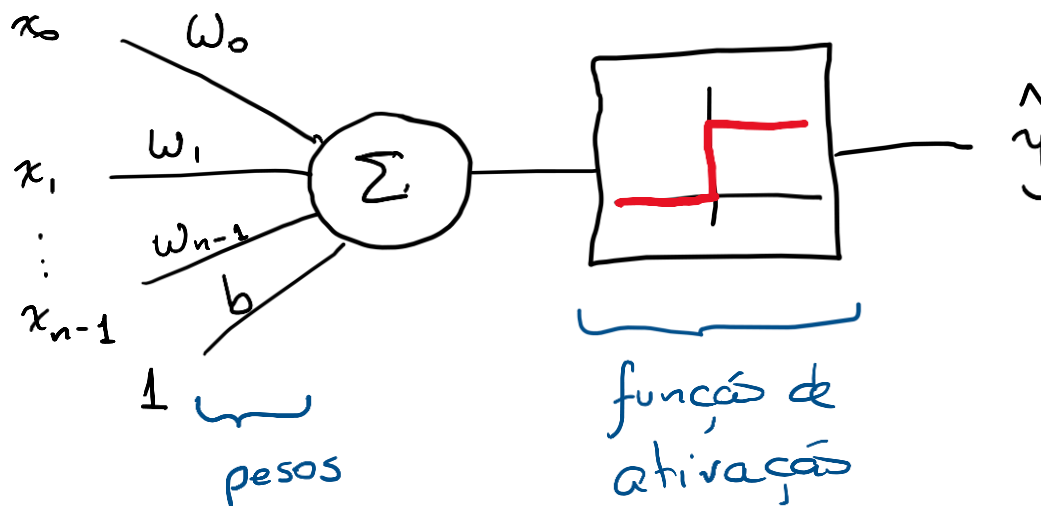
3

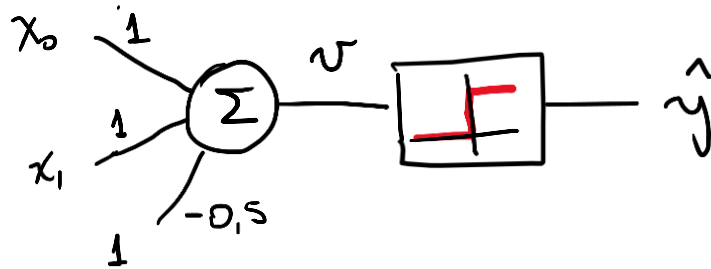
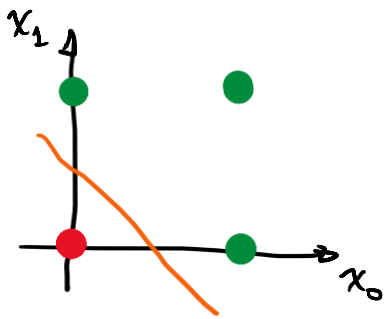


Perceptron  
(Rosenblatt)

Como ele  
aprende?

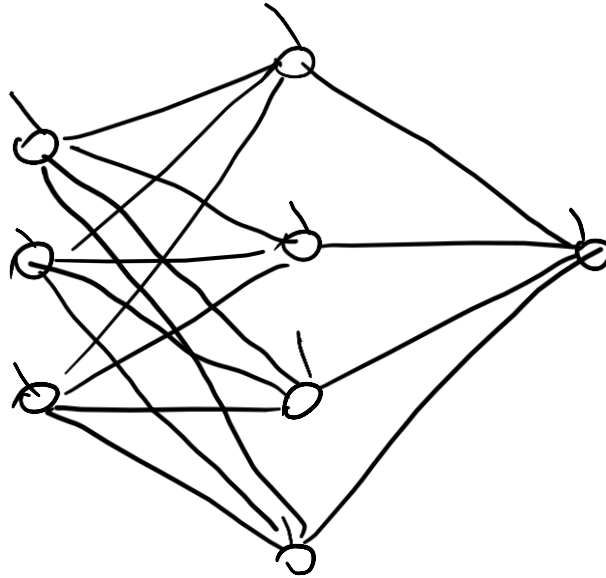
## Neurônios artificiais





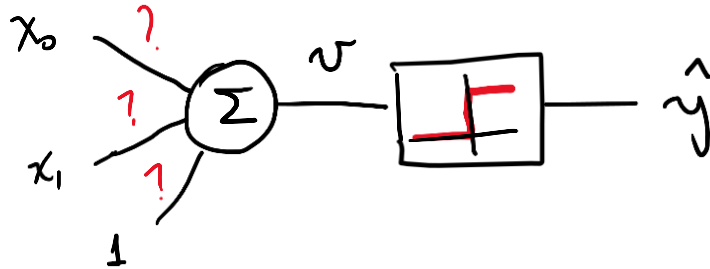
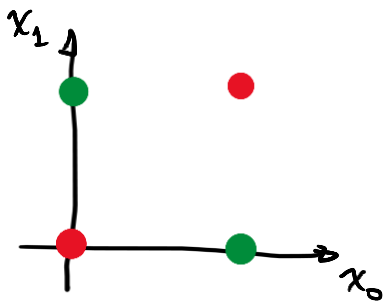
$x_0$	$x_1$	$\hat{y}$
0	0	0
0	1	1
1	0	1
1	1	1

Porta OR



Hebb learning  $\rightarrow$  não funciona p/ redes maiores!

1º inverno das redes neurais

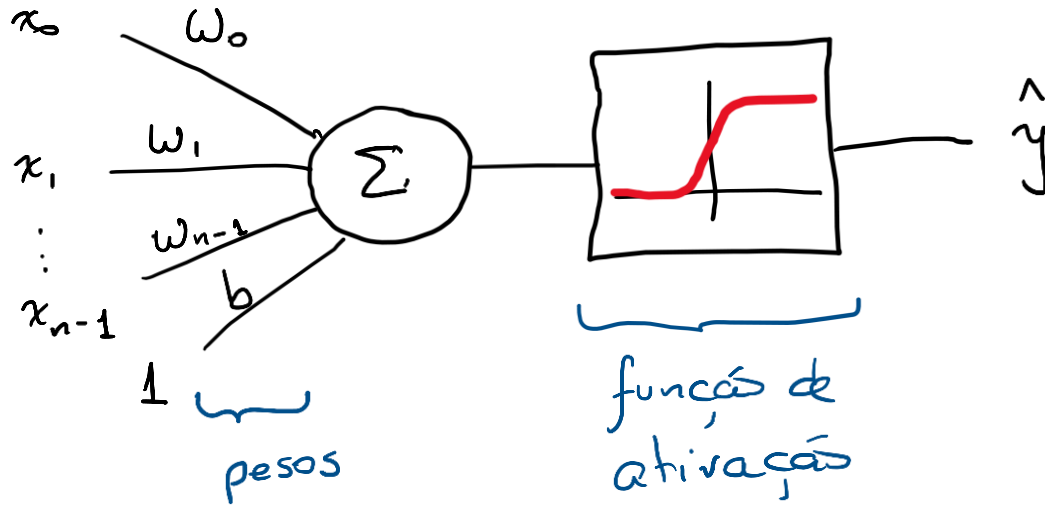


$x_0$	$x_1$	$\hat{y}$
0	0	0
0	1	1
1	0	1
1	1	0

Porta XOR

# Multilayer Perceptron e Backpropagation

8



$$\hat{y} = \sigma(\theta_0 + \theta_1 x_1 + \dots + \theta_n x_n)$$

Regressão logística

$\Rightarrow$  Gradient Descent



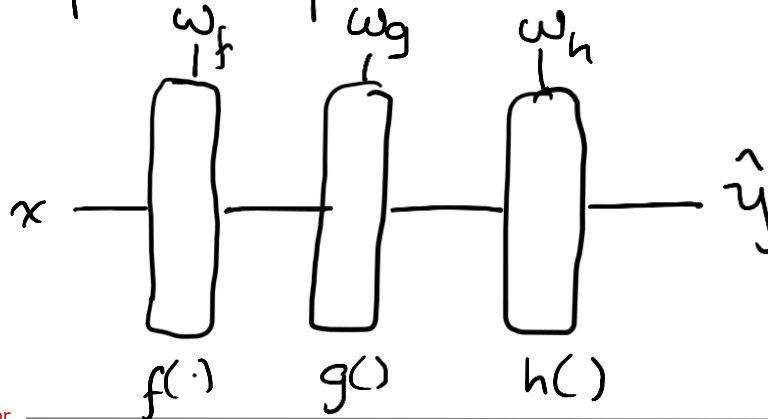
# Regra da cadeia

9

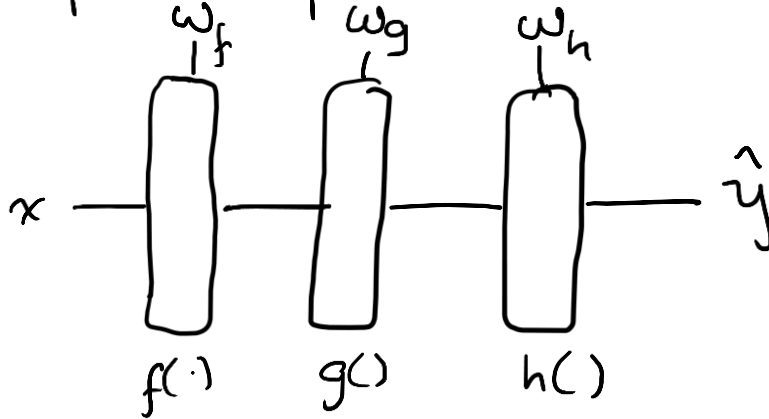
$$y = h(g(f(x)))$$

$$\Rightarrow \frac{dy}{dx} = \frac{dh}{dg} \cdot \frac{dg}{df} \cdot \frac{df}{dx}$$

Aplicando p/ redes neurais



Aplicando p/ redes neurais



$$\mathcal{E} = \mathcal{E}(w_f, w_g, w_h)$$

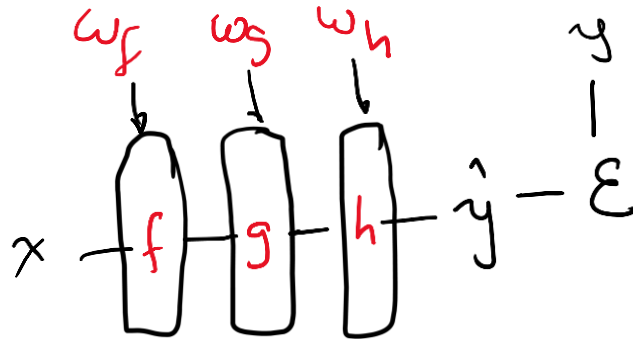
$$\mathcal{E} = \frac{1}{m} \sum_{i=1}^m (\hat{y}_i - y_i)^2 \quad \text{MSE} \quad \text{regressões}$$

$$\mathcal{E} = -\frac{1}{m} \sum_{i=1}^m \sum_{k=1}^K y_{ik} \log \hat{y}_{ik} \quad \text{entropia cruzada} \\ \text{(classificação)}$$

Gradient descent

$$\Theta_{i+1} = \Theta_i - \eta \nabla_{\Theta} \mathcal{E}$$

11



$$\frac{\partial \mathcal{E}}{\partial w_f} = \frac{\partial \mathcal{E}}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial g} \cdot \frac{\partial g}{\partial f} \cdot \frac{\partial f}{\partial w_f}$$

Diagram illustrating the backpropagation of error through the layers. Blue arrows labeled 'erro' (error) point from the output layer back to the hidden layer, and from the hidden layer back to the input layer.

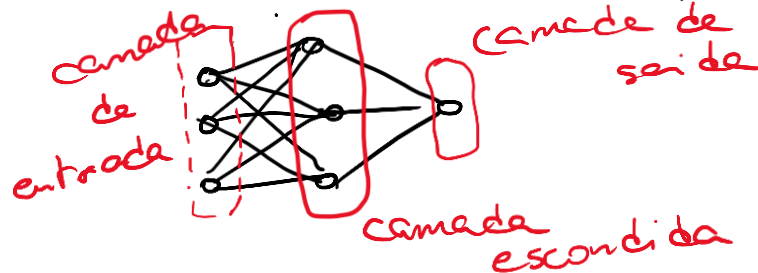
backpropagation  
2ª era das  
reds  
neurais

No início dos anos 90...

- Problema do treinamento
  - O treinamento das redes não converge ou
  - converge p/ mínimos locais ruins

- Teorema da universalidade

Qualquer função (quase) pode ser reproduzida com uma rede neural de apenas ~~2 camadas~~ 1 camada escondida



2<sup>o</sup> inverno das redes neurais

SVM, R.F., etc

2000 - 2010

Hinton, LeCun, Bengio,  
Schmidhuber, etc

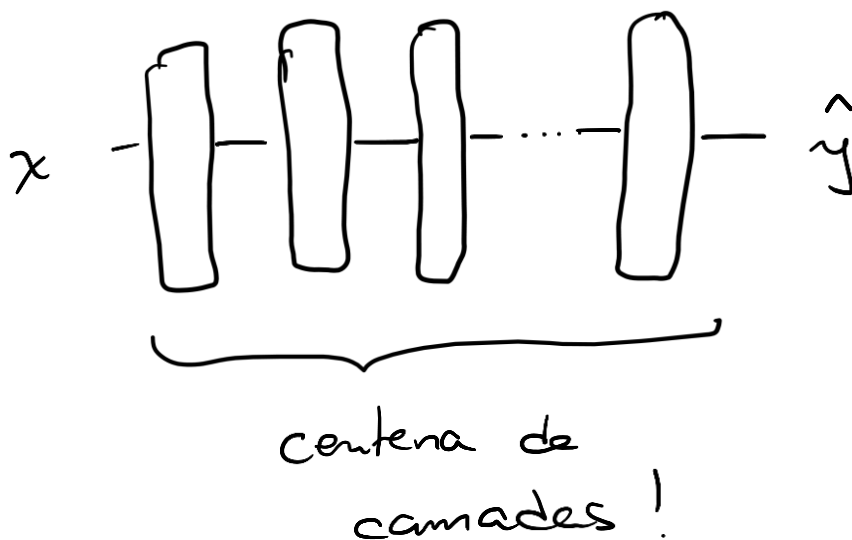
13

- Na prática, aumentar o número de camadas ajuda  $\Rightarrow$  redes neurais profundas  $\rightarrow$  deep learning
- Topologias especiais resolvem melhor certos problemas
  - Visão Computacional — Redes convolucionais
  - Sequências (e.g. texto) — Redes recorrentes
  - etc...

3<sup>a</sup> era das redes neurais

# Redes neurais profundas

14



$$z = x \cdot y \quad \Rightarrow \quad \frac{\partial z}{\partial x} = y$$

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

The background of the slide is composed of several concentric, partial circular arcs in red and grey, creating a dynamic, layered effect. The arcs are of varying thicknesses and are positioned at different radii from the center, which is where the text is located.

# Insper