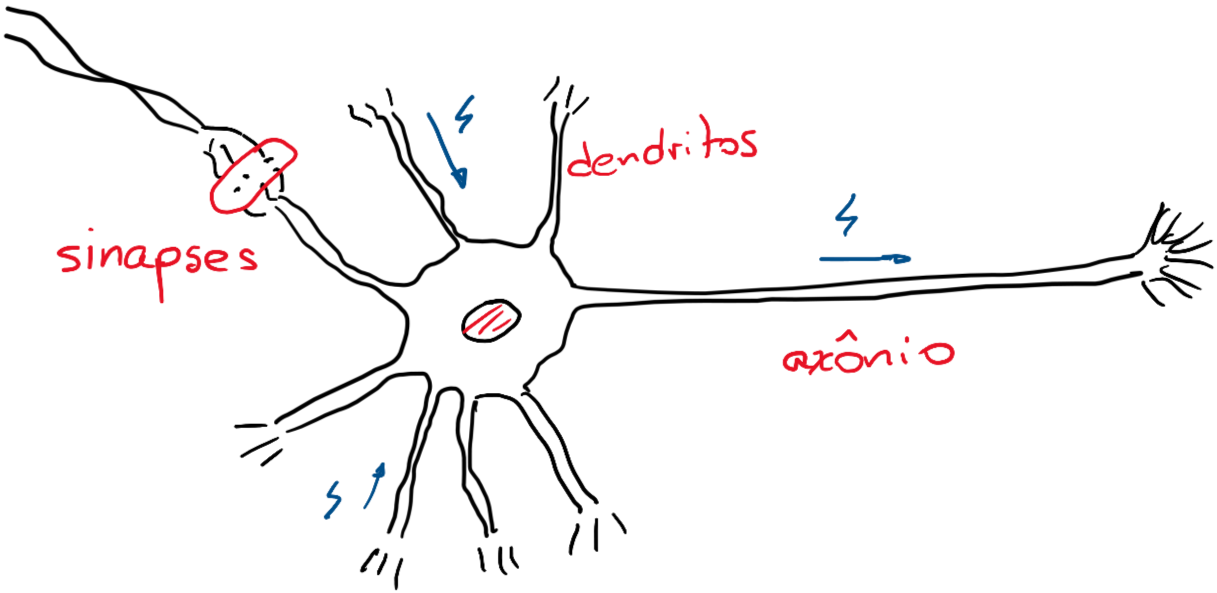
Insper

Machine Learning

Introdução à redes neurais

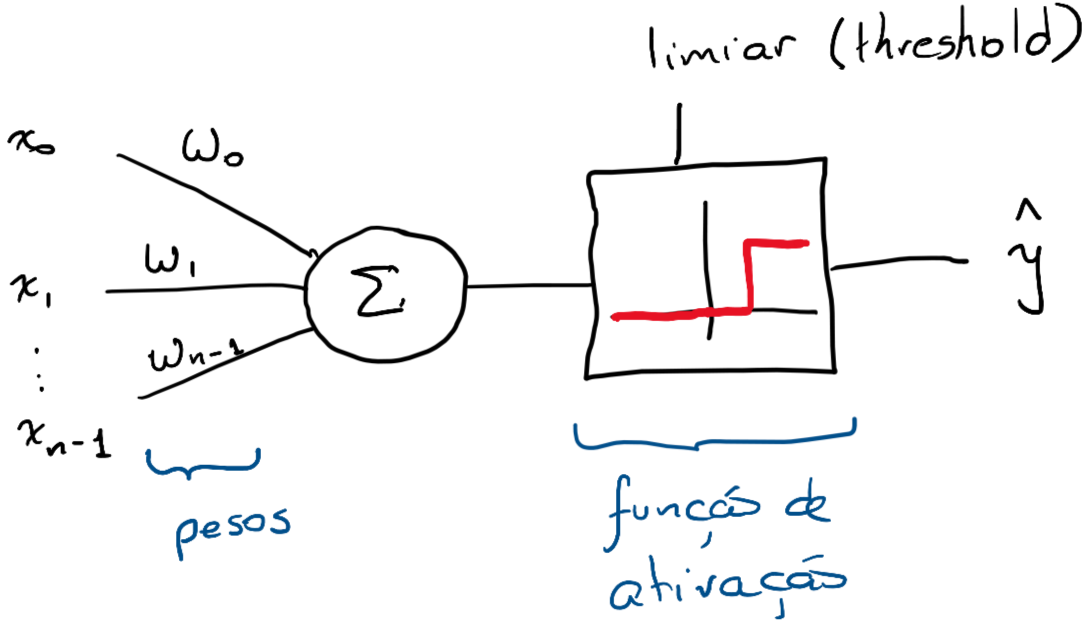
Fábio Ayres <fabioja@insper.edu.br>

Neuronio biológico



$$\sum_{\text{dendritos}} \begin{pmatrix}\text{impulso}\\ \text{na}\\ \text{entrada}\end{pmatrix} \cdot \begin{pmatrix}\text{peso}\\ \text{da}\\ \text{sinapse}\end{pmatrix} > \text{limiar} \Rightarrow \begin{matrix}\text{pulso}\\ \text{eletrico}\\ \text{na}\\ \text{saida}\end{matrix}$$

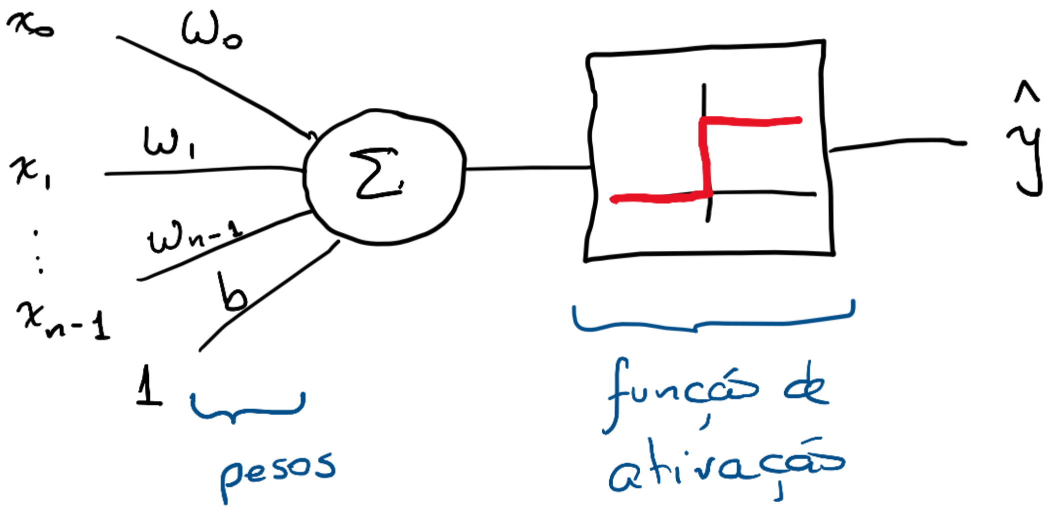
Neuronios artificiais

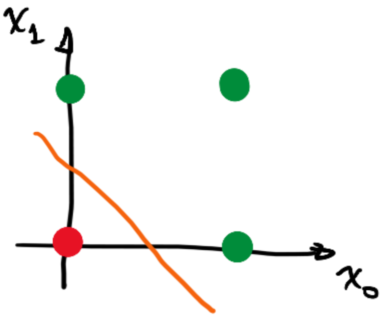


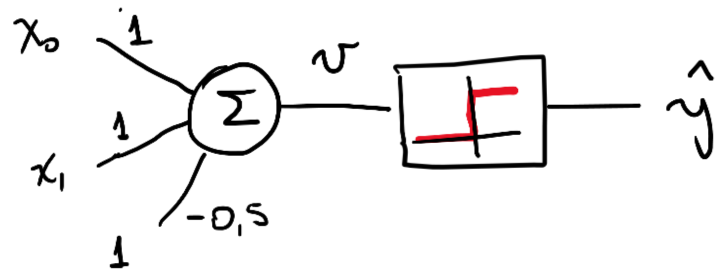
Perceptron
(Rosenblatt)

Como ele aprende?

Neuronios artificiais

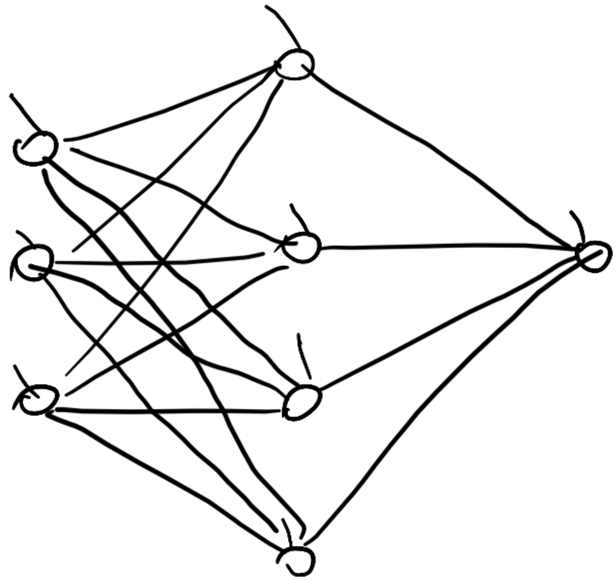






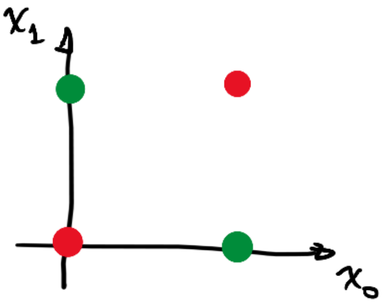
x_0	x_1	$\hat{y}$
0	0	0
0	1	1
1	0	1
1	1	1

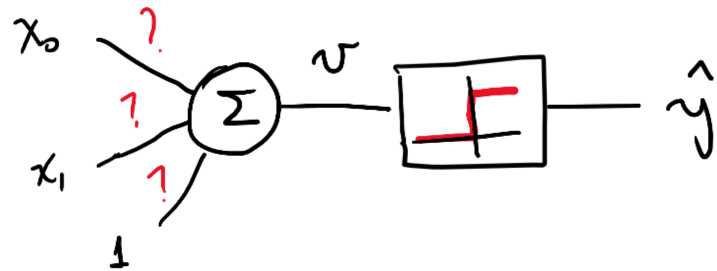
<u>Porta OR</u>



Hebb learning → não funciona p/ redes maiores!

1º inverno das redes neurais

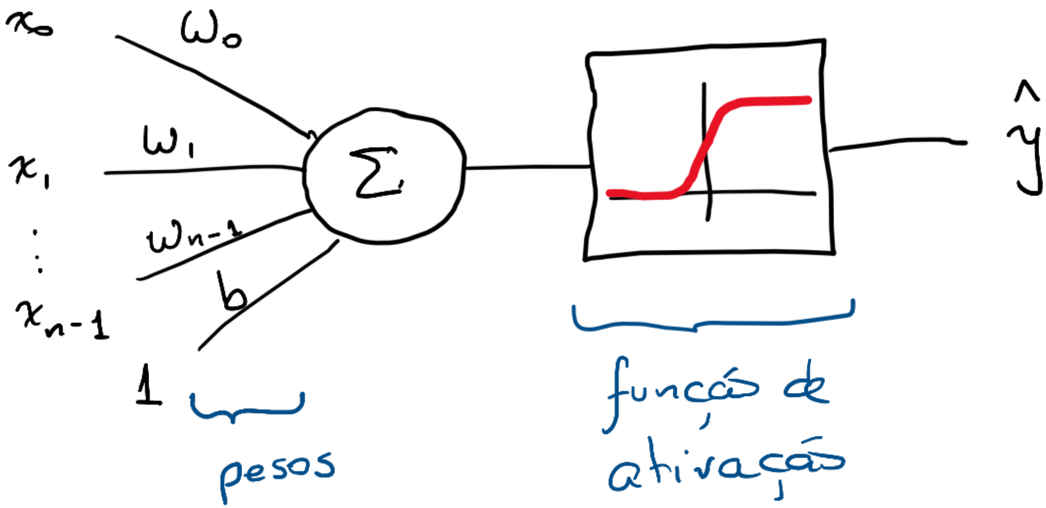




x_0	x_1	$\hat{y}$
0	0	0
0	1	1
1	0	1
1	1	0

Porta XOR

Multilayer Perceptron e Backpropagation



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

Regressão logística

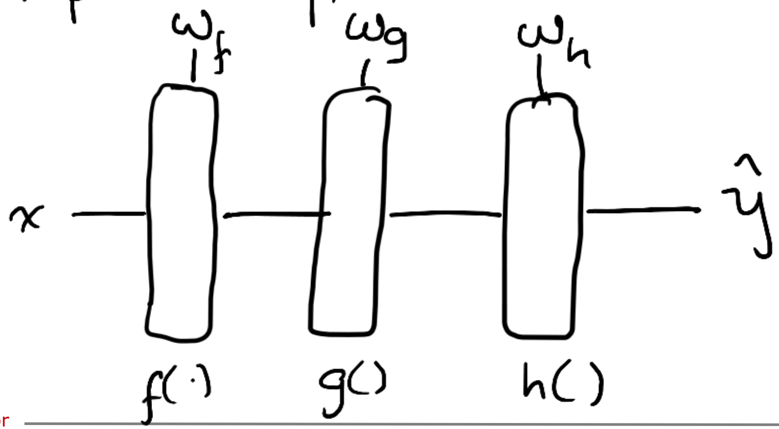
$\Rightarrow$ Gradient Descent

Regra da cadeia

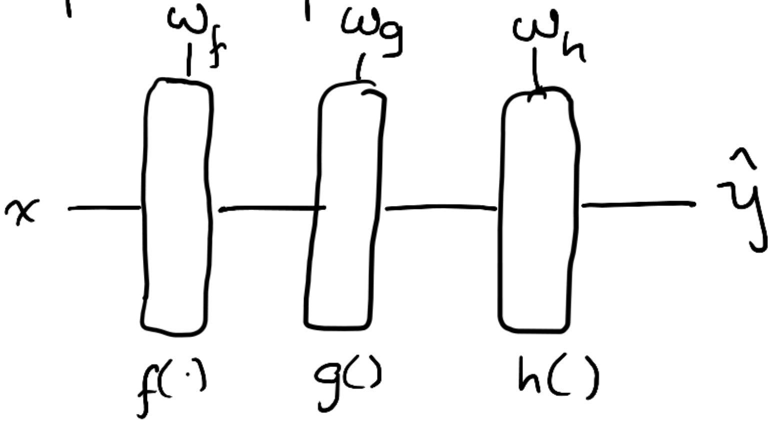
$$y = h(g(f(x)))$$

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

Aplicando p/ redes neurais



Aplicando p/ redes neurais

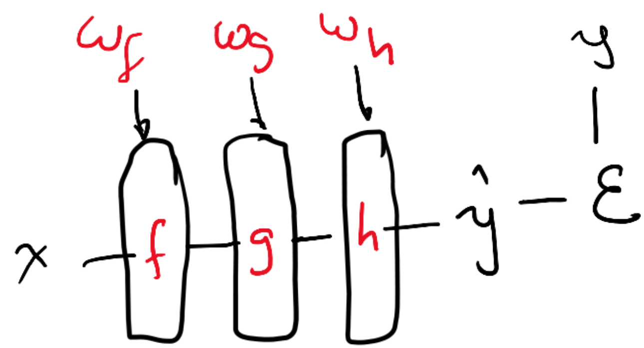


$$\mathcal{E} = \mathcal{E}(w_f, w_g, w_h)$$

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

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

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



erro erro erro

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

backpropagation

2ª era das redes neurais

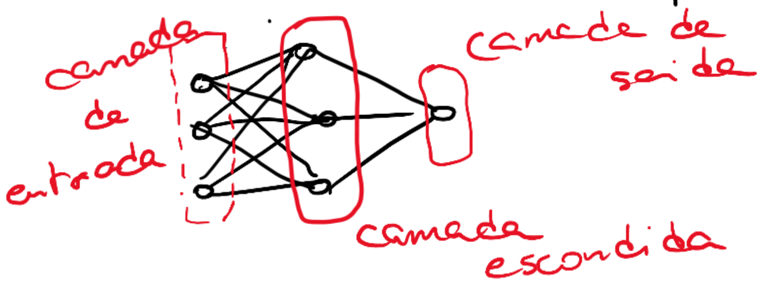
No início dos anos 90...

- 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



2º inverno das redes neurais

SVM, R.F., etc

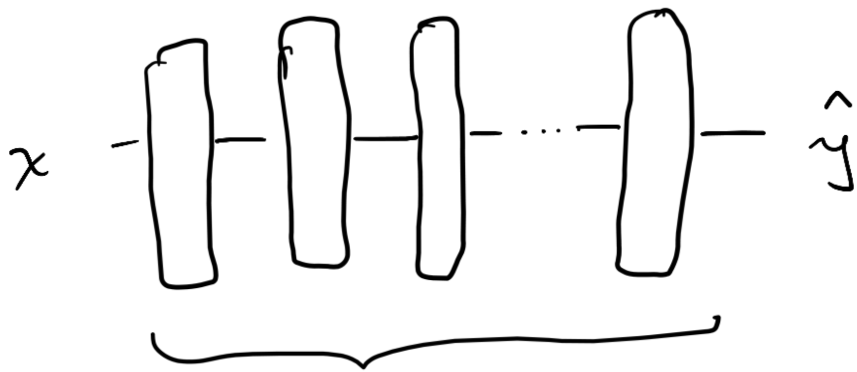
2000 - 2010

Hinton, LeCun, Bengio,
SchmidtHuber, etc

- Na pratica, aumentar o numero de camadas ajuda
  ⇒ redes neurais profundas ⟶ deep learning
- Topologias especiais resolvem melhor certos problemas
  - Visão Computacional — Redes convolucionais
  - Sequencias (e.g. texto) — Redes recorrentes
  - etc ...

3ª era das redes neurais

Redes neurais profundas



centena de camadas !

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

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

Insper