



Universidade Federal da Paraíba

Coordenação do Curso de Ciência de Dados e
Inteligência Artificial



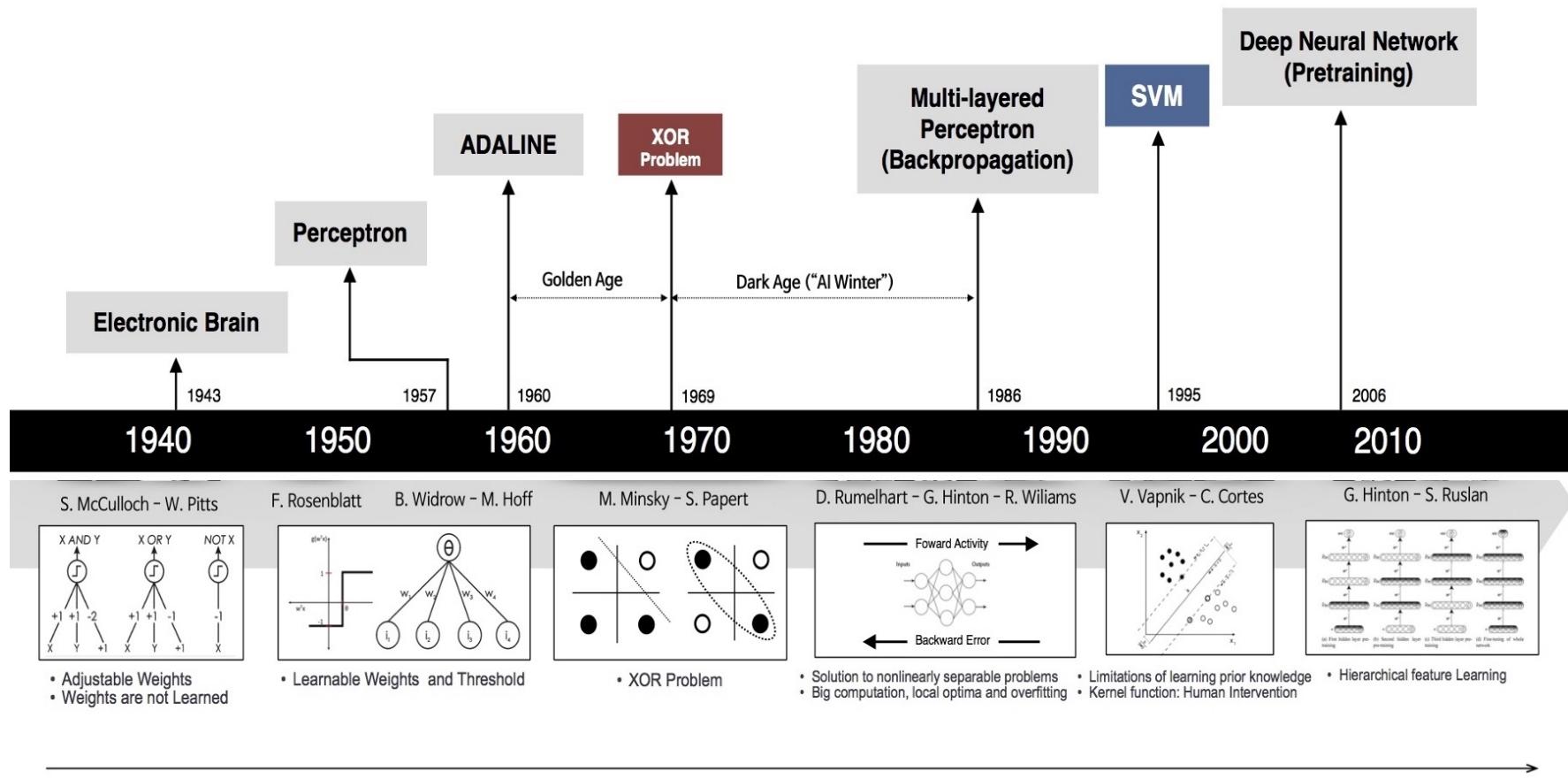
O significado de aprendizagem

Prof. Dr. Bruno Pessoa

Roteiro

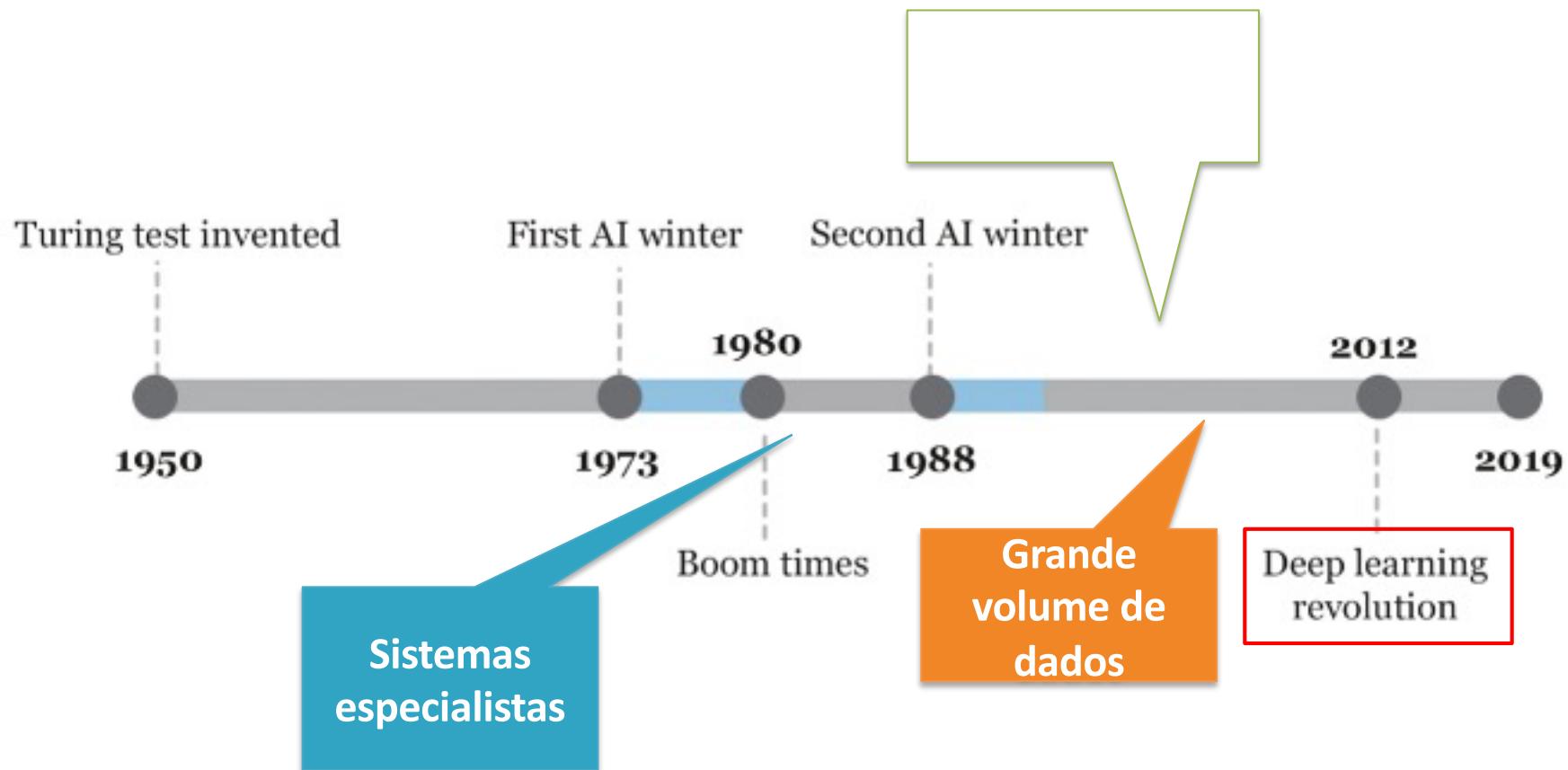
- Breve histórico da IA
- Panorama atual e futuro da IA
- Inteligência Artificial e Aprendizagem de Máquina
- Hierarquia de aprendizagem
- Componentes de aprendizagem
- Perceptron
- PLA

Breve histórico



Brief history of neural network

Breve histórico



Breve histórico

ANO	FATO RELEVANTE
1950	O Teste de Turing é criado (“Can machines think?”).
1955	Arthur Samuel implementou um programa para jogar Damas baseado em heurísticas e em pesos auto-ajustáveis que aprendia a partir de suas jogadas.
1956	John McCarthy define o termo “Inteligência Artificial”.
1957	RosenBlatt cria o modelo Perceptron.
1959	Arthur Samuel usa o termo “machine learning” pela primeira vez.
1986	O algoritmo backpropagation é apresentado.

Breve histórico

ANO	FATO RELEVANTE
1997	O modelo SVM é publicado.
2006	Geoffrey Hinton publica um dos mais importantes artigos na área de Redes Neurais Profundas.
2012	Hinton demonstrou o poder das Redes Neurais Profundas ao vencer a competição ImageNet de 2012.
2016	O programa AlphaGo vence o campeão mundial de GO.
2022	Lançamento do ChatGPT.

Grandes nomes da IA



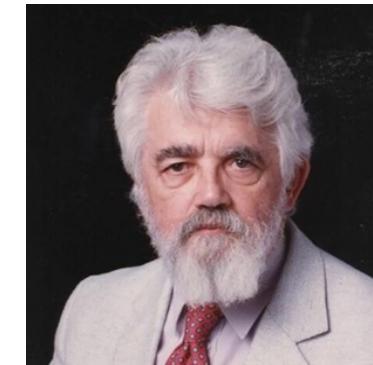
Alan Turing
(1912-1954)



Arthur Samuel
(1901-1990)



Marvin Minsky
(1927-2016)



John McCarthy
(1927-2011)

Grandes nomes da IA



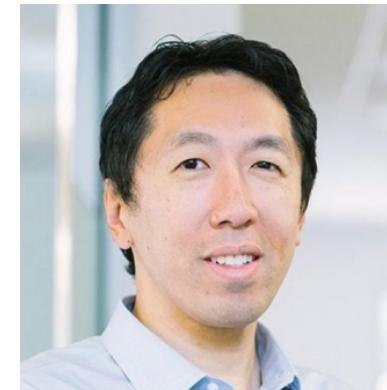
**Geoffrey
Hinton**



**Yann
LeCun**



**Yoshua
Bengio**



**Andrew
Ng**

Trabalhos revolucionários

The screenshot shows the top navigation bar of the Nature journal website. The 'nature' logo is circled in red. Below it, the menu includes 'Explore content', 'About the journal', 'Publish with us', and 'Subscribe'. A horizontal line separates this from the main content area.

[nature](#) > [letters](#) > [article](#)

[Published: 09 October 1986](#)

Learning representations by back-propagating errors

[David E. Rumelhart](#), [Geoffrey E. Hinton](#) & [Ronald J. Williams](#)

[Nature](#) 323, 533–536 (1986) | [Cite this article](#)

112k Accesses | 14978 Citations | 386 Altmetric | [Metrics](#)

Abstract

We describe a new learning procedure, back-propagation, for networks of neurone-like units. The procedure repeatedly adjusts the weights of the connections in the network so as to minimize a measure of the difference between the actual output vector of the net and the desired output vector. As a result of the weight adjustments, internal 'hidden' units which are not part of the input or output come to represent important features of the task domain, and the regularities in the task are captured by the interactions of these units. The ability to create useful new features distinguishes back-propagation from earlier, simpler methods such as the perceptron-convergence procedure¹.

Trabalhos revolucionários

> [Neural Comput.](#) 2006 Jul;18(7):1527-54. doi: 10.1162/neco.2006.18.7.1527.

A fast learning algorithm for deep belief nets

[Geoffrey E Hinton](#)¹, [Simon Osindero](#), [Yee-Whye Teh](#)

Affiliations + expand

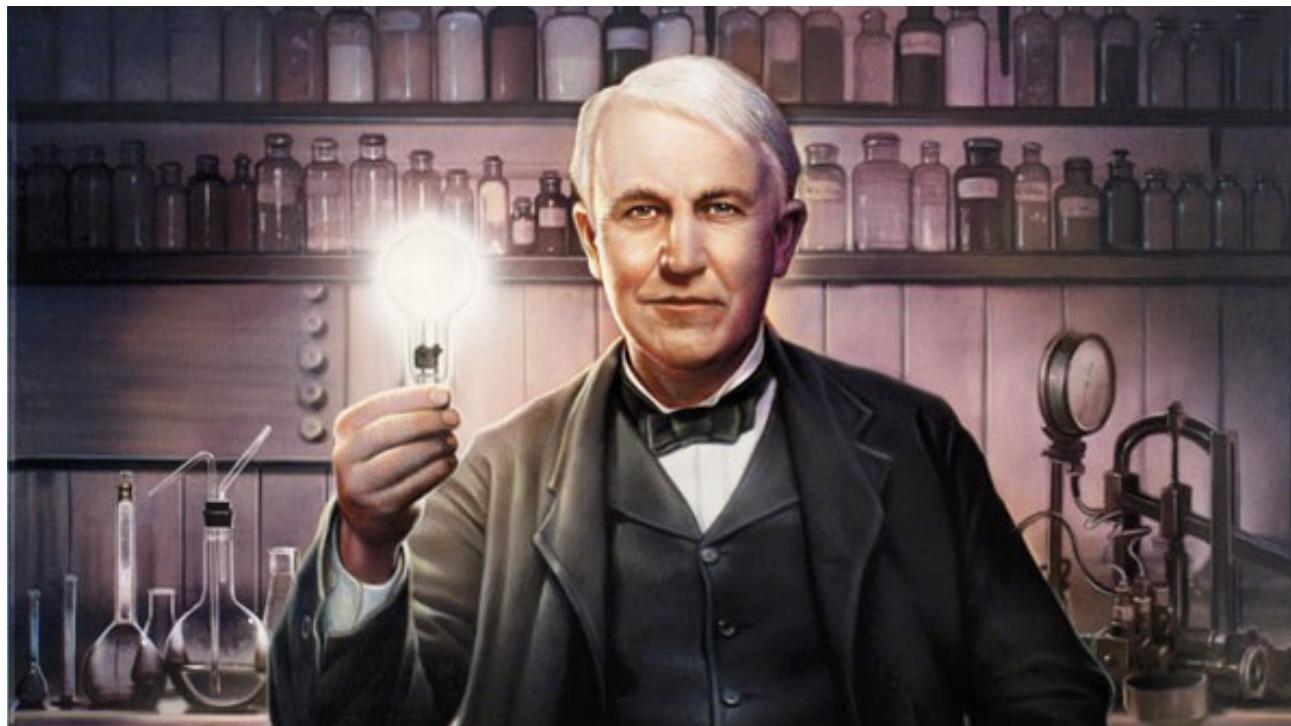
PMID: 16764513 DOI: [10.1162/neco.2006.18.7.1527](#)

Abstract

We show how to use "complementary priors" to eliminate the explaining-away effects that make inference difficult in densely connected belief nets that have many hidden layers. Using complementary priors, we derive a fast, greedy algorithm that can learn deep, directed belief networks one layer at a time, provided the top two layers form an undirected associative memory. The fast, greedy algorithm is used to initialize a slower learning procedure that fine-tunes the weights using a contrastive version of the wake-sleep algorithm. After fine-tuning, a network with three hidden layers forms a very good generative model of the joint distribution of handwritten digit images and their labels. This generative model gives better digit classification than the best discriminative learning algorithms. The low-dimensional manifolds on which the digits lie are modeled by long ravines in the free-energy landscape of the top-level associative memory, and it is easy to explore these ravines by using the directed connections to display what the associative memory has in mind.

Panorama atual

- Era da implementação



As quatro ondas da IA



As quatro ondas da IA

- IA da Internet

SISTEMAS DE RECOMENDAÇÃO



NETFLIX

eBay

amazon®

twitter

You Tube

movielens
helping you find the right movies

YAHOO!

PHOTOREE

Google

last.fm

facebook.

Submarino

As quatro ondas da IA

- IA de Negócios



As quatro ondas da IA

- IA da Percepção



As quatro ondas da IA

- IA Autônoma



IA Específica x IA Geral

- IA específica, restrita ou fraca (do inglês, “**Narrow AI**”) é aquela voltada para realização de atividades específicas que requerem inteligência humana.
- IA geral ou forte (do inglês, “**Artificial General Intelligence**”)
- AGI está associada ao fenômeno da **singularidade tecnológica**.

Perspectivas sobre o futuro

- Estima-se que 27% dos empregos de países da OCDE são de profissões com alto risco de automatização pela inteligência artificial (OCDE, 2023).
- Há um consenso entre especialistas de que a inteligência artificial superará a humana.

Perspectiva sobre o futuro

- Propostas de soluções
 - Reciclar trabalhadores
 - Reduzir jornada de trabalho
 - Redistribution de renda



Perspectiva sobre o futuro

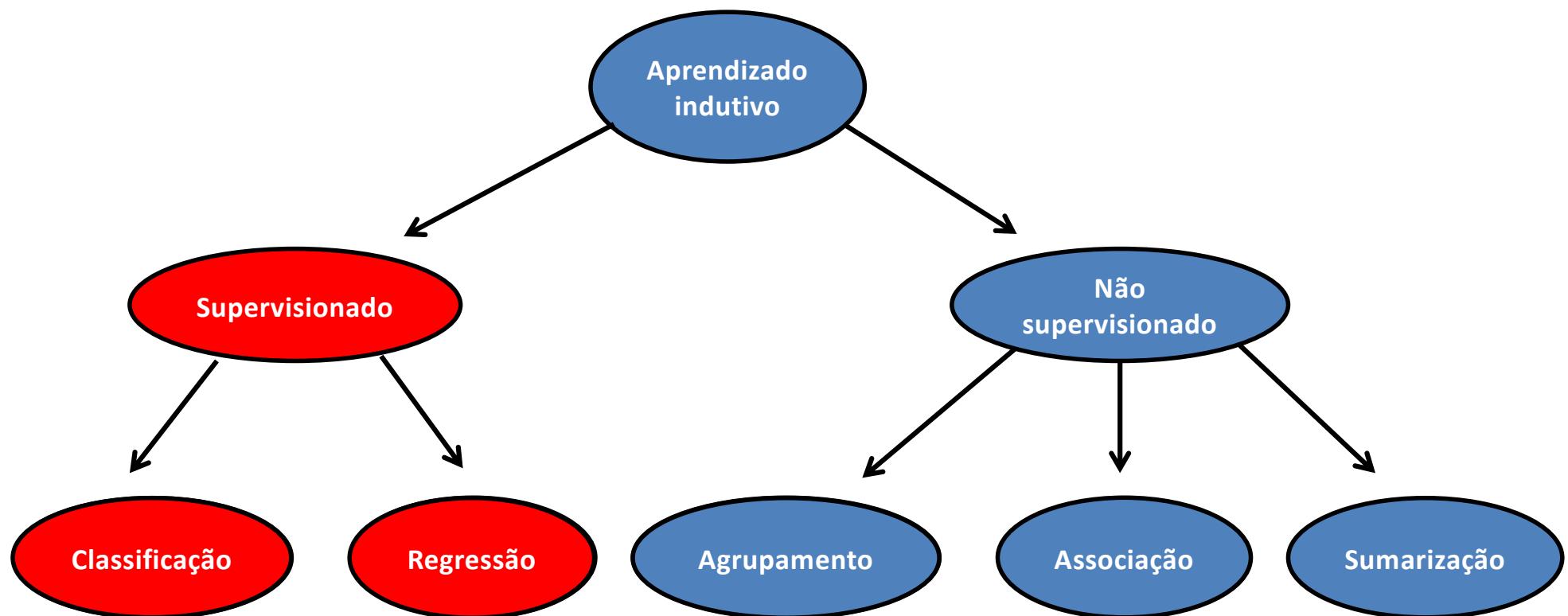
- Propostas alternativas



Inteligência Artificial e Aprendizagem de Máquina

- A aprendizagem de máquina é uma **subárea** da Inteligência Artificial.
- Pode ser definida como o processo de **indução de hipóteses** a partir de um **conjunto de dados** (experiência passada).
- Pretende-se então generalizar o conhecimento adquirido de amostras de dados.

Hierarquia de aprendizado



Componentes de aprendizagem

- **Condições necessárias para a aprendizagem de máquina**
 - Existência de um padrão.
 - Não há método matemático capaz de resolver o problema diretamente.
 - Existência de dados.

Componentes de aprendizagem

- **Exemplo: Aprovação de crédito**

Variável	Valor
Idade	25
Sexo	Masculino
Salário	R\$ 3000,00
Anos na mesma residência	2
Anos no mesmo trabalho	5
Possui outras dívidas	Sim
(...)	

Decisão a ser tomada: Aprovar ou não o crédito solicitado?

Componentes de aprendizagem

- **Formalização:**

Entrada: x (características do requerente)

Saída: y (decisão sobre a concessão de crédito)

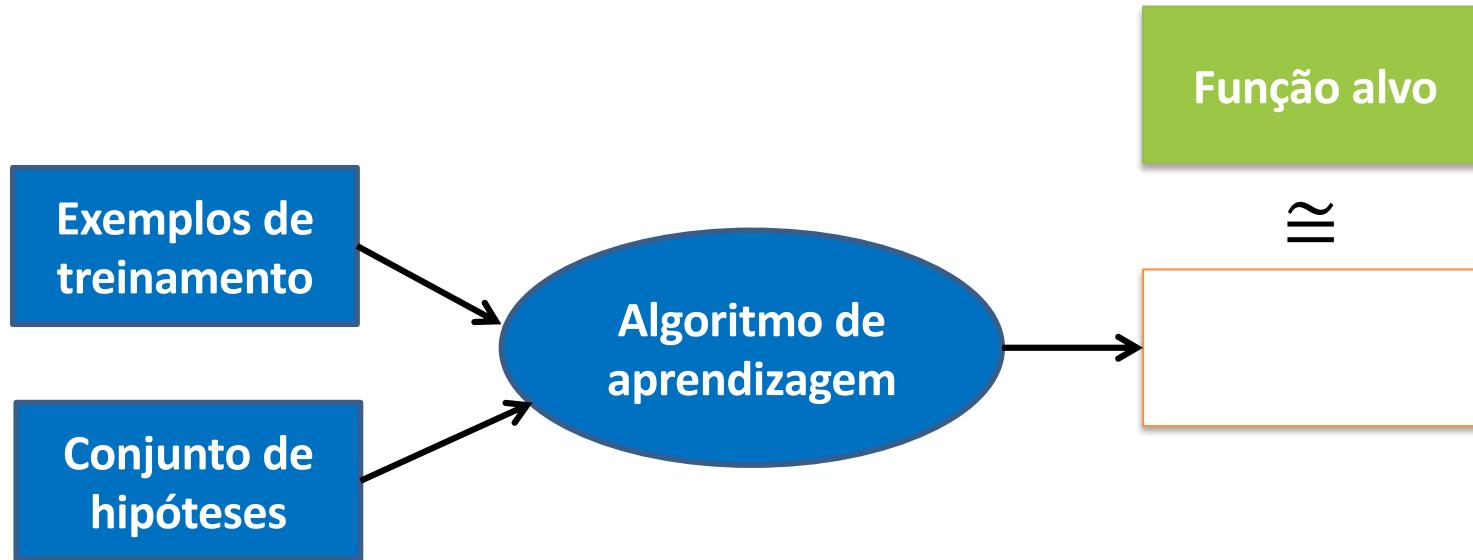
Função alvo: $f: X \rightarrow Y$ (função que define o padrão)

Dados: $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$ (dados de clientes)

Hipótese: $g: X \rightarrow Y$, (resposta do modelo de AM)

$g \in H$ (conjunto de hipóteses)

Componentes de aprendizagem



Entrada: Exemplos de treinamento

Processamento: Algoritmo de Aprendizagem + Conjunto de Hipóteses

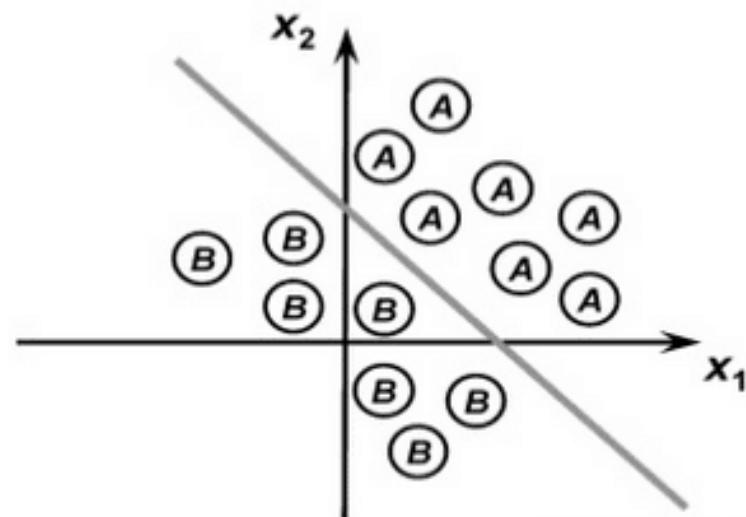
Saída: Hipótese solução

Perceptron

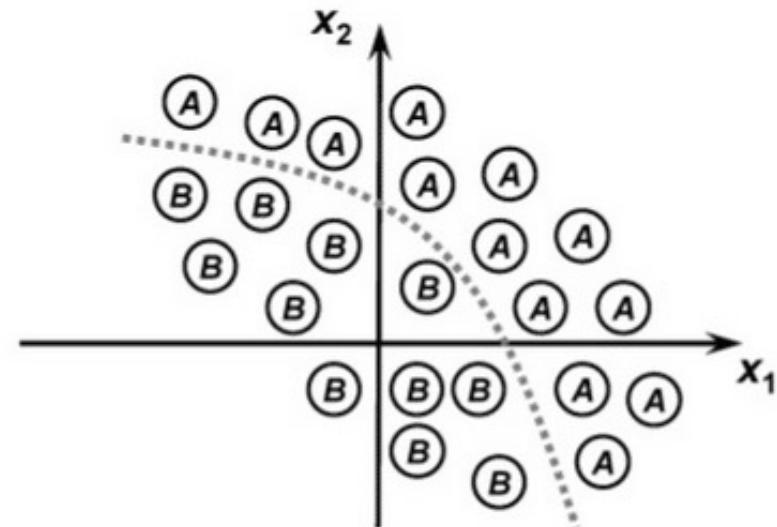
- **Modelo linear** de aprendizagem de máquina.
- Unidade básica de uma **rede neural artificial**.
- Aplicável a dados **linearmente separáveis**.

Perceptron

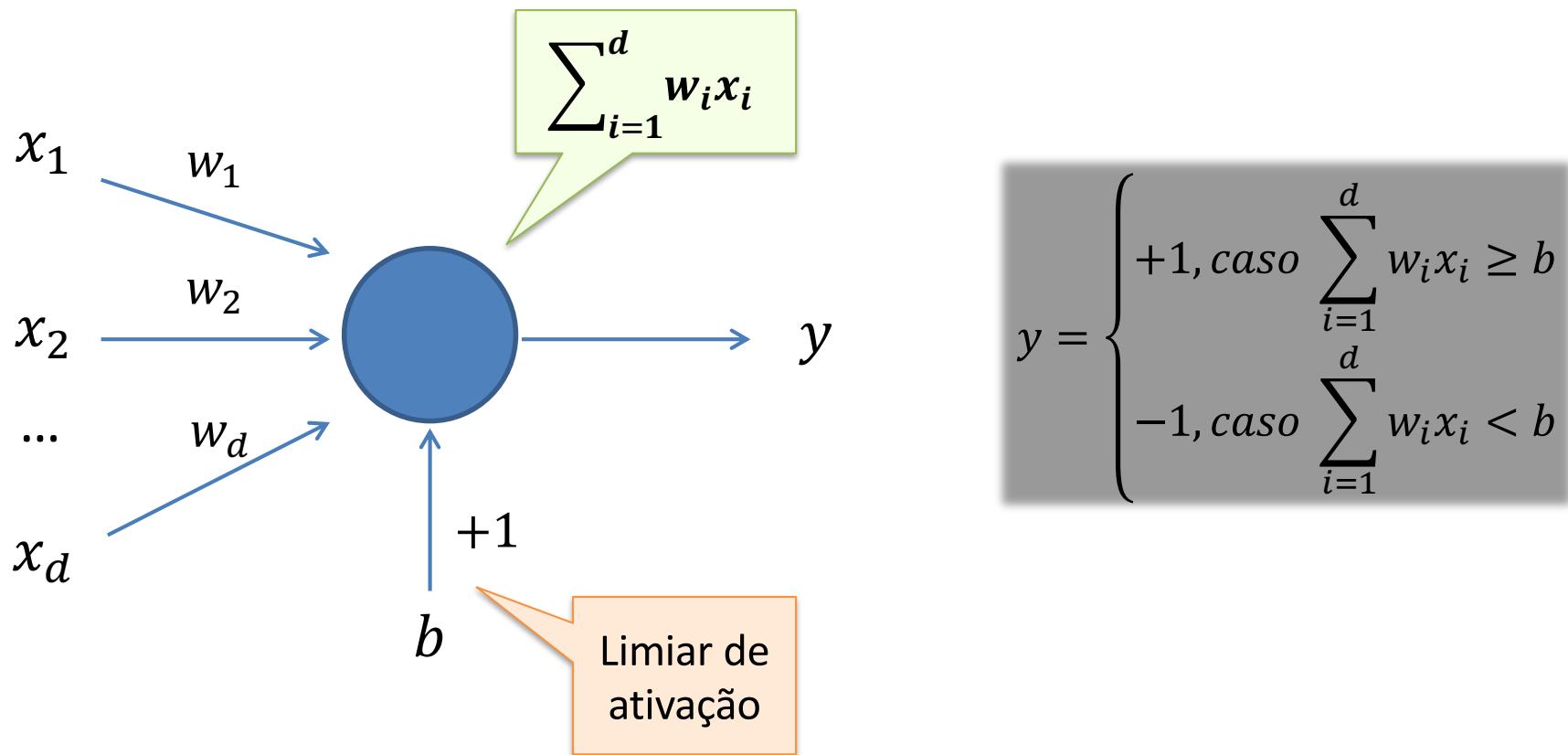
Dados linearmente
separáveis



Dados NÃO linearmente
separáveis



Perceptron



Interpretação geométrica

Considere $x \in \mathbb{R}^2$ e seja $\mathbf{h}(x) = \text{sign}(\sum_{i=1}^d w_i x_i - b)$ a função que define o conjunto de hipóteses.

Assim, $\mathbf{h}(x_1, x_2) = w_1 x_1 + w_2 x_2 - b$, é a equação de uma reta em \mathbb{R}^2 .

Podemos concluir então que as **hipóteses** no modelo Perceptron em \mathbb{R}^2 consistem em **retas**, onde **b** é o **coeficiente linear**.

Forma vetorial

Para $\mathbf{x} = (x_1, \dots, x_d)$ e $h \in H$, temos que:

$$h(\mathbf{x}) = \text{sign}\left(\left(\sum_{i=1}^d w_i x_i\right) - b\right)$$

Fazendo $w_0 = -b$ e $x_0 = 1$, obtemos:

$$h(\mathbf{x}) = \text{sign}\left(\sum_{i=0}^d w_i x_i\right)$$

Na forma vetorial:

$$h(\mathbf{x}) = \text{sign}(\mathbf{w}^T \mathbf{x})$$

Algoritmo de aprendizagem (PLA)

```
# X = {(x1, y1), (x2, y2), ..., (xN, yN)}

Algoritmo PLA(X) {
    w = 0

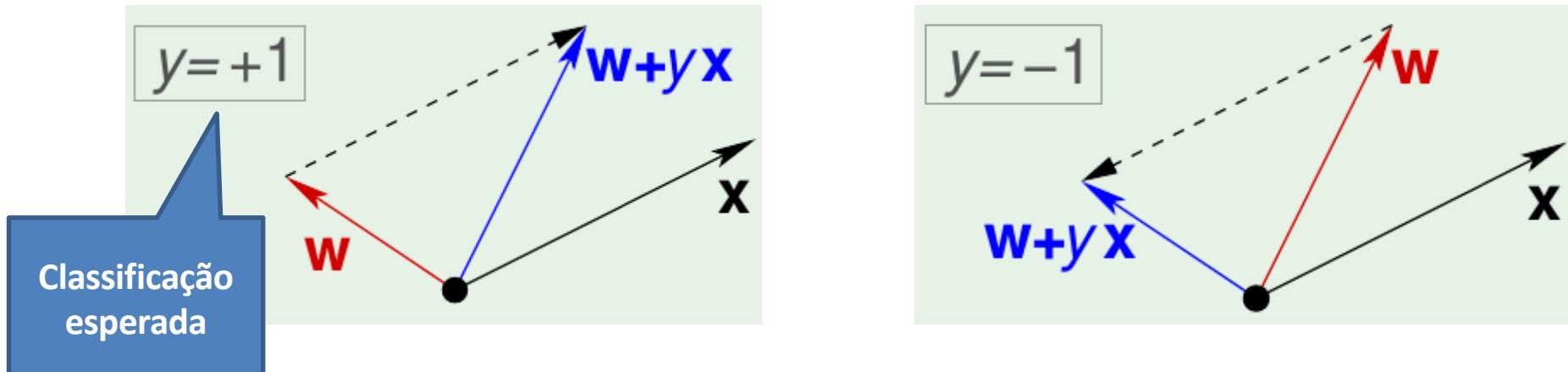
    # Lista de pontos classificados incorretamente
    listaPCI = X
    enquanto (len(listaPCI) > 0):
        escolha aleatoriamente um ponto  $x_i$  pertencente à lista

        # Atualiza o vetor de pesos ao corrigir a classificação de x
        w = w + yixi
        listaPCI = constroiListaPCI(X, w)

    return w
}
```

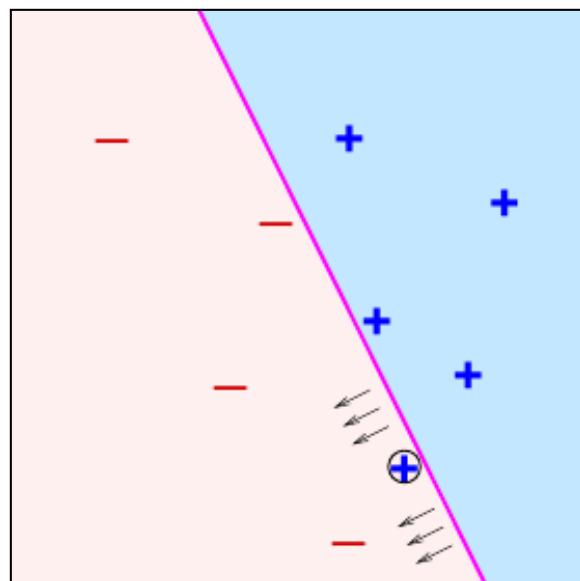
Algoritmo de aprendizagem (PLA)

- Atualização dos pesos ($w = w + y_i x_i$)
 - O produto interno de dois vetores é **negativo** quando os vetores formam um **ângulo obtuso** e **positivo** quando formam um **ângulo agudo**.



Algoritmo de aprendizagem (PLA)

- **Atualização dos pesos**
 - A correção de um ponto pode levar a um erro na classificação do outro.



O algoritmo tem fim?

Referências bibliográficas

- Abu-Moustafa, Y.S.; Magdon-Ismail, M.; Lin, H-S.
“Learning from data”. AMLBook, 2012.
- Faceli, K.; Lorena, A.C.; Gama, J.; Carvalho, A.C.P.L.F.
“Inteligência Artificial Uma Abordagem de Aprendizado de Máquina”. LTC, 2011.
- Notas de aula do prof. Abu-Moustafa.