PPGEE
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA
UNIVERSIDADE FEDERAL DO CEARÁ

INSTITUTO FEDERAL
Ceará
Campus Fortaleza

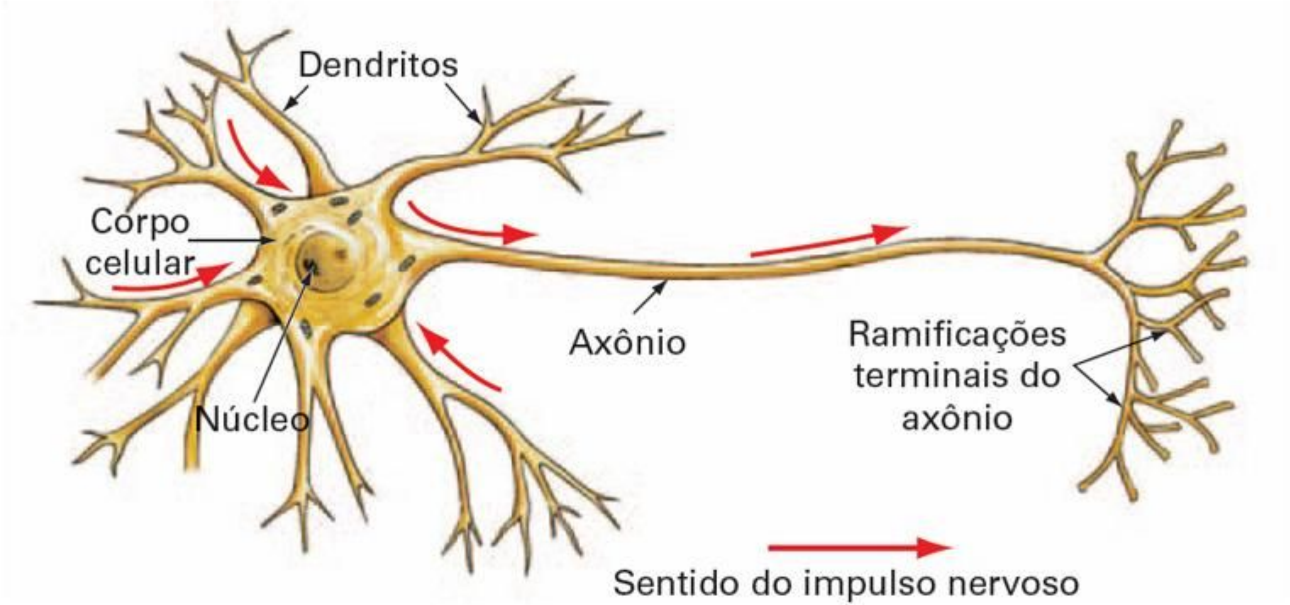
LAPISCO

Redes Neurais

Prof. Matheus Santos
matheus.santos@lapisco.ifce.edu.br

Introdução

O Neurônio Biológico



Introdução

O Neurônio Biológico

- **Dentritos:** Ramificações correspondentes aos canais de entrada de informação (sinais elétricos, mV).
- **Corpo celular:** Local onde é feito o balanço energético da célula nervosa (soma das contribuições de energia).
- **Axônio:** Canal de saída do neurônio, ou seja, caminho de propagação dos impulsos nervosos em direção a outros neurônios ou músculos.
- **Sinapses:** Pontos de contato entre neurônios onde há passagem de neurotransmissores do axônio de um neurônio para os dendritos de outro neurônio.

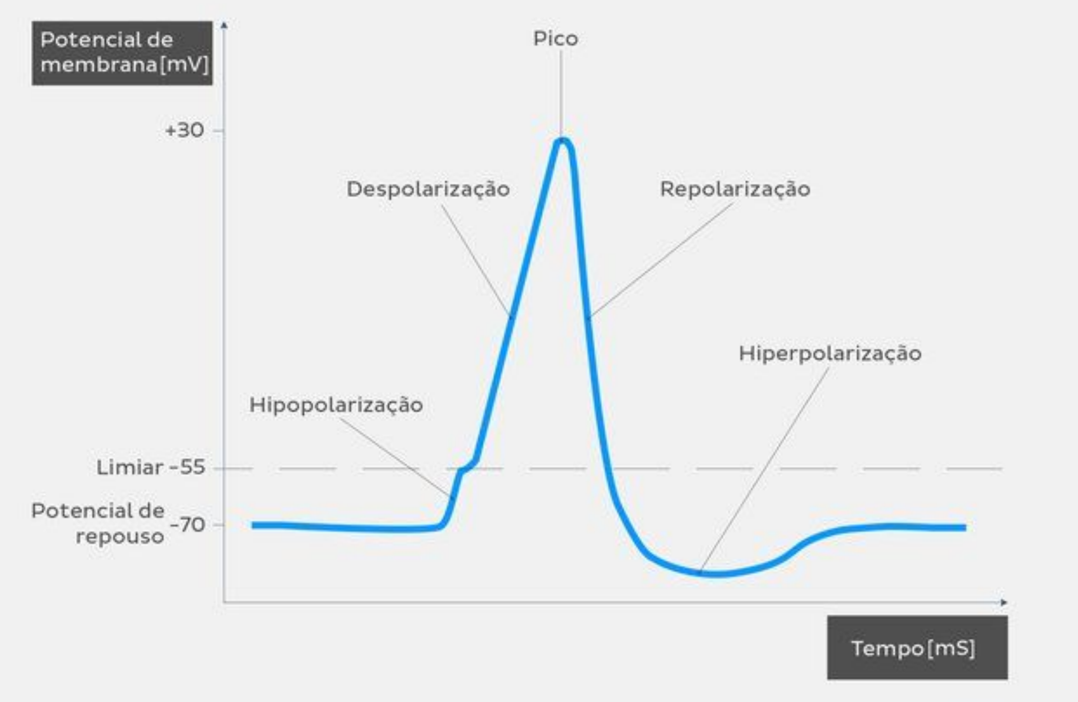
Introdução

O Neurônio Biológico

Impulso elétrico do axônio apenas se o balanço energético realizado no corpo celular for maior que um certo limiar. Assim, o neurônio **dispara** ou está **ativa**.

Introdução

O Neurônio Biológico



Introdução

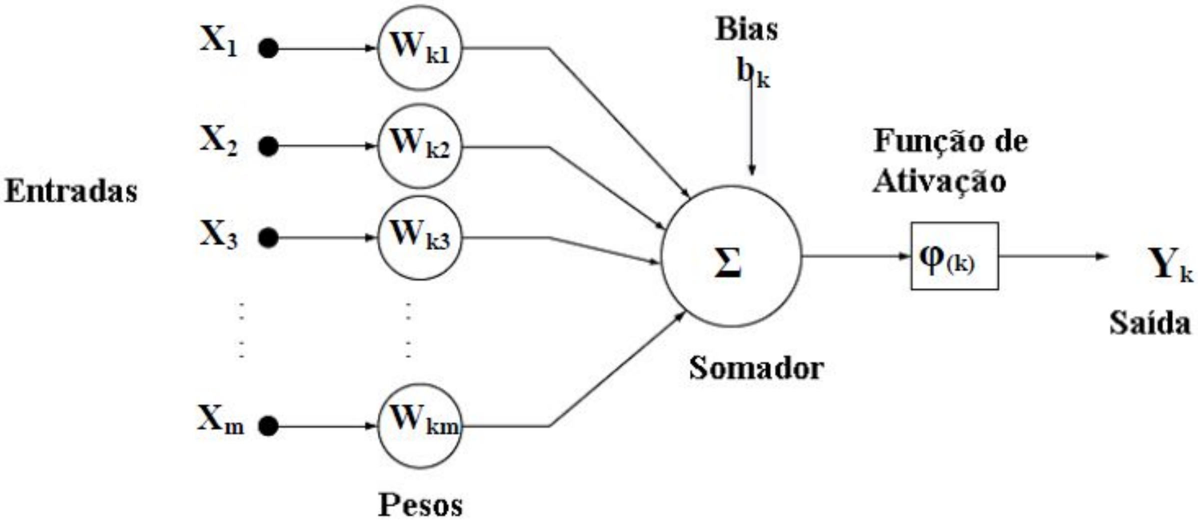
O Neurônio Biológico: Curiosidades

- Cerca de 80 a 100 bilhões deles no cérebro humano.
- Cada neurônio tem cerca de 10.000 sinapses com outros neurônios.
- O cérebro produz sim novos neurônios (ex.: hipocampo).
- Logo, a freqüência de disparo de um neurônio é da ordem de kHz.
- O peso do cérebro é aprox. 2% do peso de uma pessoa. Mesmo em repouso, o cérebro consome 20% de sua energia.
- O cérebro consome 10 vezes mais energia que o resto do corpo por grama de tecido.

Neurônio Artificial

Neurônio Artificial de McCulloch-Pitts (1943)

W. S. McCullogh and W. Pitts (1943). "A logical calculus of the ideas immanent in nervous activity", *Bulletin of Mathematical Biophysics*, vol. 5, p. 115-133.



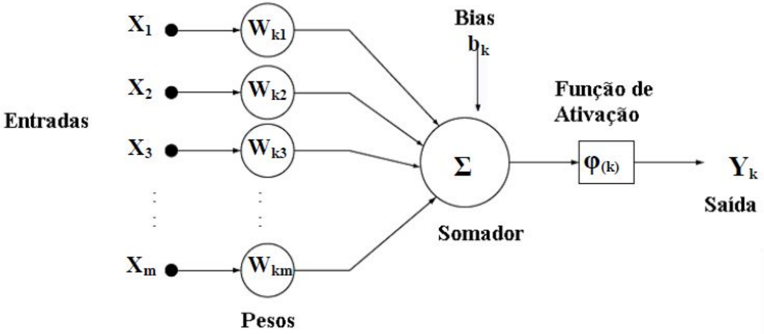
Neurônio Artificial

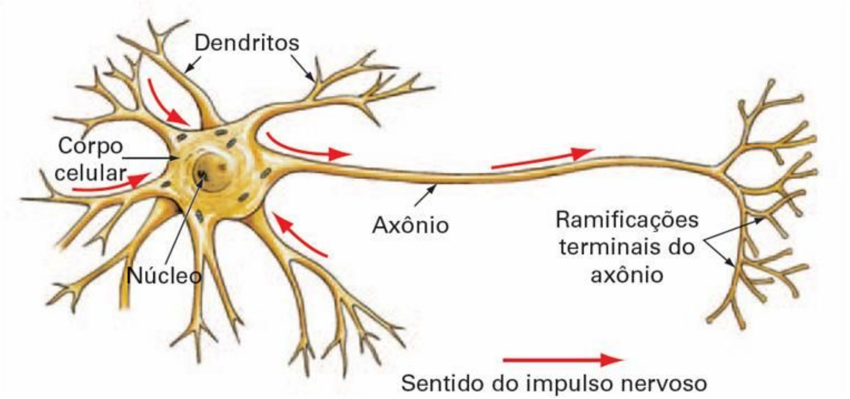
Neurônio Artificial de McCulloch-Pitts (1943)

- Cada dendrito é modelado como um canal de transmissão por onde flui a informação de entrada.
- A força (ou eficiência) das conexões sinápticas é modelada como um fator (peso sináptico), cujo papel é modular o fluxo de informações de entrada.
- A função do corpo celular é realizar o acúmulo energético (somatório) sobre as entradas moduladas pelos pesos sinápticos.
- O axônio é modelado como uma chave ON-OFF, que indica se o neurônio respondeu ao estímulo. Em outras palavras, atingiu ou não o potencial de ação.

Neurônio Artificial

Neurônio Artificial de McCulloch-Pitts (1943)

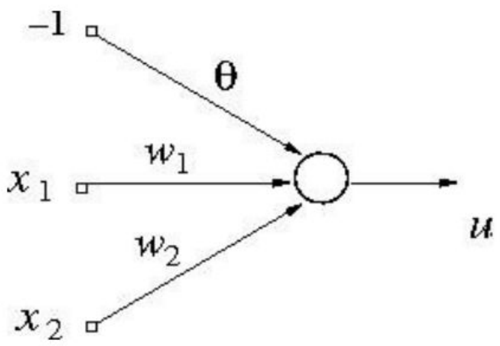




Neurônio Artificial

Neurônio Artificial de McCulloch-Pitts (1943): Análise Geométrica

Dado o neurônio:



x_1, x_2: entradas

w_1, w_2 : pesos sinápticos

θ : limiar (*bias*)

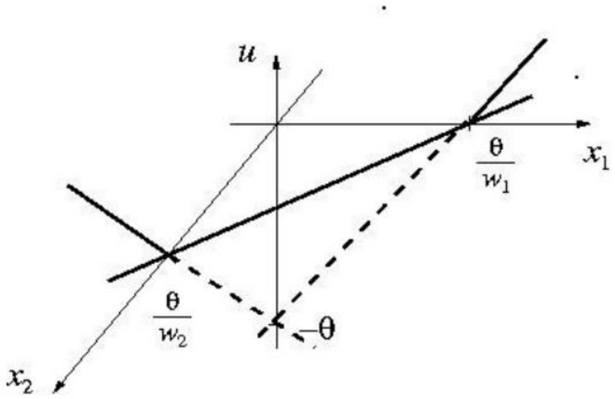
u : ativação

Sua ativação é dada por: $u = w_1 x_1 + w_2 x_2 - \theta$

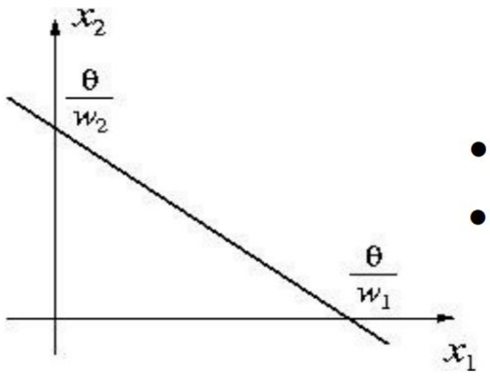
Neurônio Artificial

Neurônio Artificial de McCulloch-Pitts (1943): Análise Geométrica

A Eq. (1) define um plano em (x_1, x_2, u).



A Eq. (2) define a seguinte reta em (x_1, x_2).

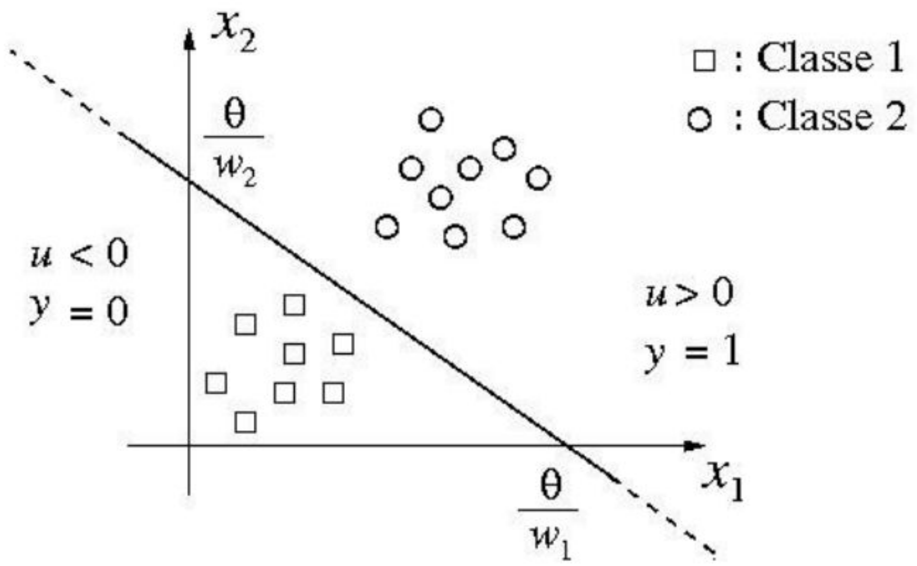


$$x_2 = -(w_1/w_2)x_1 + \theta/w_2$$

- Corta eixo x_1 no ponto $(\theta/w_1, 0)$
- Corta eixo x_2 no ponto $(0, \theta/w_2)$

Neurônio Artificial

Neurônio Artificial de McCulloch-Pitts (1943): Análise Geométrica



Redes Neurais

Contribuições após o neurônio MCP

➢ Teoria do aprendizado de Hebb (1949)

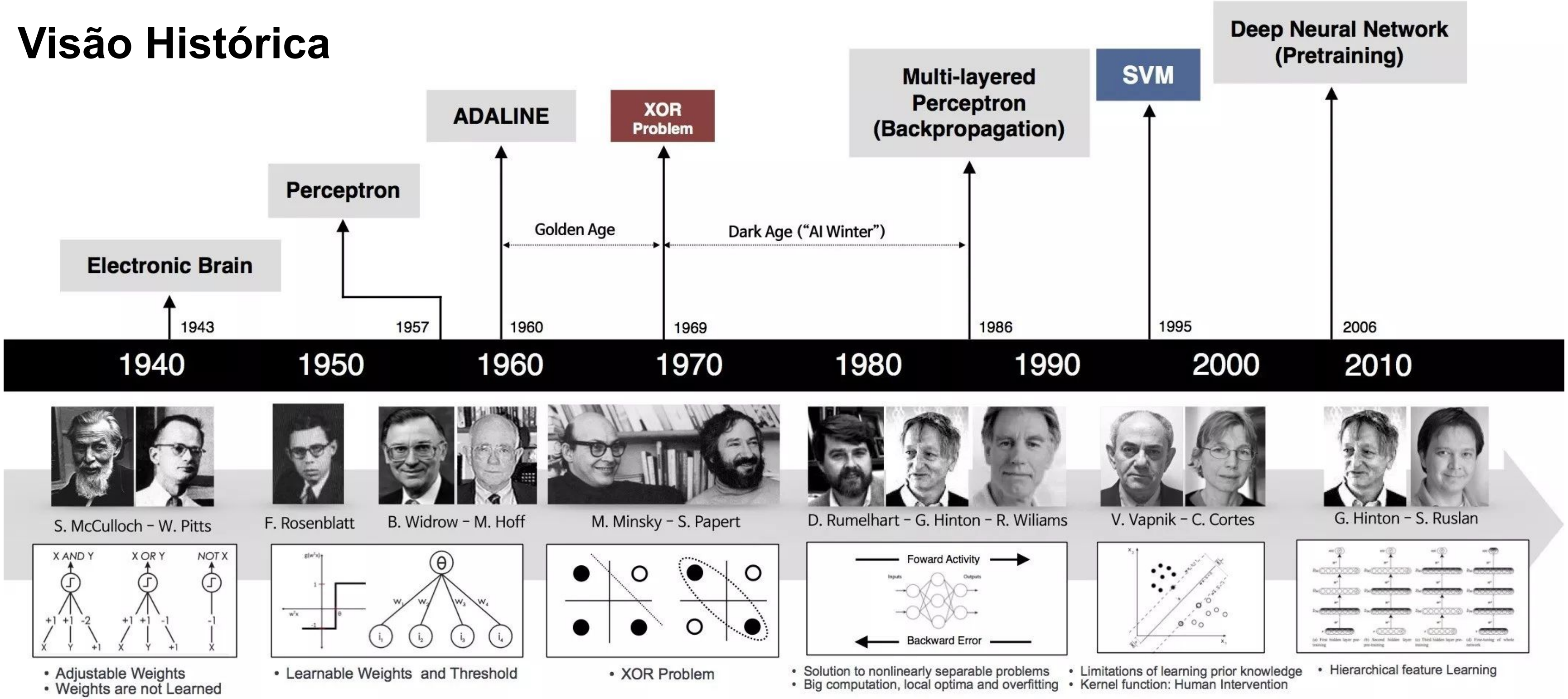
Hebb, D.O. (1949) The Organization of Behavior. A Neuropsychological Theory. John Wiley and Sons, Inc., New York.

➢ Perceptron de Rosenblatt (1958)

F. Rosenblatt (1958). "The Perceptron: A probabilistic model for information storage and organization in the brain", *Psychological Review*, vol. 65, p. 386-408.

Redes Neurais

Visão Histórica



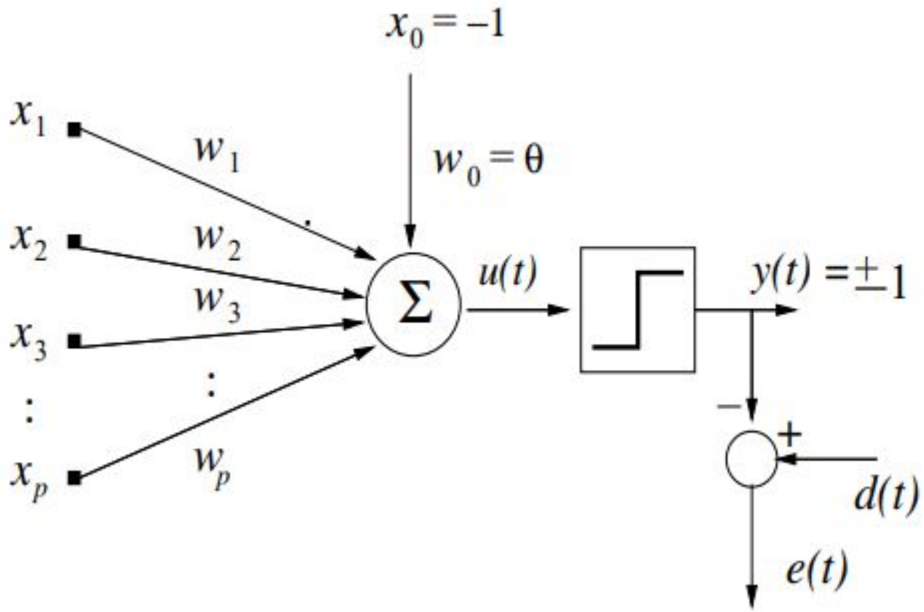
Perceptron Simples

Rede Neural Perceptron Simples:

- ➢ Neurônio de MCP + Regra de Aprendizagem
- ➢ A regra de aprendizagem torna o Perceptron inteligente
- ➢ Aprendizagem Supervisionada
- ➢ Função-Objetivo
- ➢ Taxa de Aprendizagem

Perceptron Simples

Rede Neural Perceptron Simples:



Perceptron Simples

Rede Neural Perceptron Simples:

Variáveis e pesos como vetores daqui em diante

$$\mathbf{x} = \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_p \end{pmatrix} = \begin{pmatrix} -1 \\ x_1 \\ \vdots \\ x_p \end{pmatrix} \qquad \mathbf{w} = \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_p \end{pmatrix} = \begin{pmatrix} \theta \\ w_1 \\ \vdots \\ w_p \end{pmatrix}$$

Perceptron Simples

Rede Neural Perceptron Simples:

Ativação = Produto Escalar (medida de similaridade entre vetores)

- Definição 1:

$$u = \mathbf{w}^T\mathbf{x} = \mathbf{x}^T\mathbf{w} = w_0x_0 + w_1x_1 + \dots + w_px_p$$
$$= w_1x_1 + w_2x_2 + \dots + w_px_p - \theta$$

- Definição 2:

$$u = ||\mathbf{w}||.||\mathbf{x}||.\cos(\alpha)$$

Regra de Aprendizagem

- A forma vetorial da ativação (u) nos ajudará no processo de obtenção de uma regra de aprendizagem.
- O processo de aprendizagem consiste na modificação dos pesos e do limiar até que ele resolva o problema de interesse ou que o período de aprendizagem tenha finalizado.
- Função a partir de dois fatores:
  - Erro entre a saída desejada (*d*) e a saída da rede (*y*): $e = d\text{-}y$
  - Vetor de entrada (*x*)

Regra de Aprendizagem

Em geral, uma regra de aprendizagem tem a seguinte forma:

$$\mathbf{w}(t+1) = \mathbf{w}(t) + \Delta\mathbf{w}(t)$$

$\mathbf{w}(t)$ = memória (conhecimento atual).

$\Delta\mathbf{w}(t)$ = incremento na memória (informação adquirida)

$\mathbf{w}(t+1)$ = memória modificada com acréscimo de nova informação.

Sendo: $\Delta\mathbf{w}(t) = \mathbf{F}\big(e(t), \mathbf{x}(t)\big)$

Regra de Aprendizagem

Com base na função do incremento de *w*, temos:

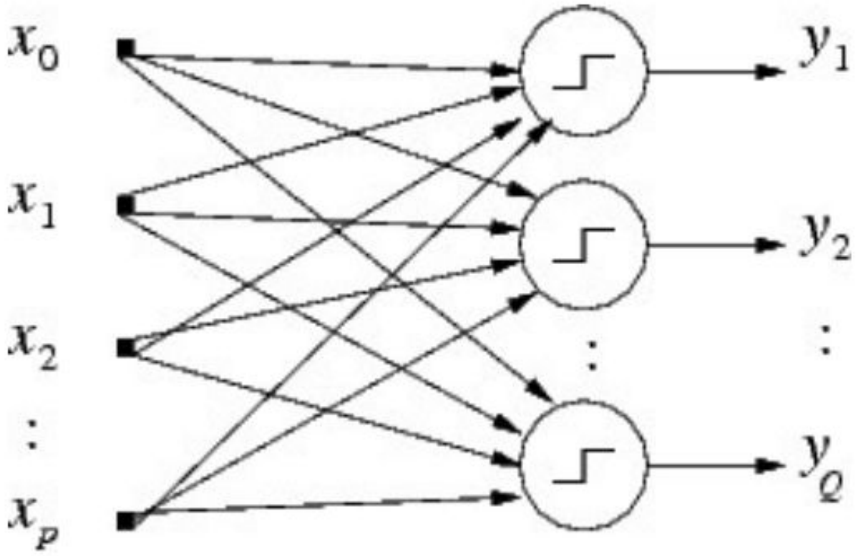
$$\mathbf{w}(t+1) = \mathbf{w}(t) + e(t)\mathbf{x}(t)$$

A fim de tornar o ajuste do vetor dos pesos mais estável, utiliza-se um fator chamado **taxa de aprendizagem** (η), frequentemente variado de 0 a 1.

$$\mathbf{w}(t+1) = \mathbf{w}(t) + \eta e(t)\mathbf{x}(t)$$

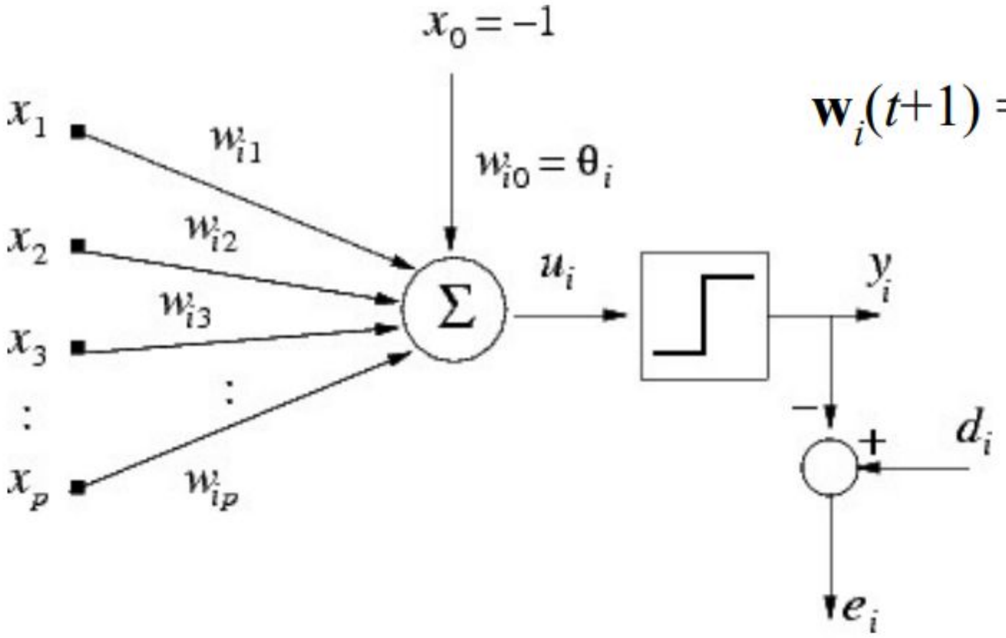
Regra de Aprendizagem

Lembrando que um neurônio categoriza apenas duas classes de dados. Ao tratar múltiplas classes, deve-se usar neurônios em paralelo.



Regra de Aprendizagem

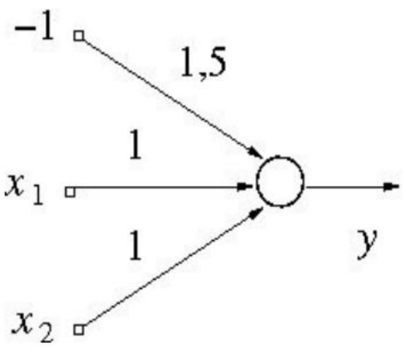
Em resumo, o i-ésimo neurônio de uma rede Perceptron é representado da seguinte forma:

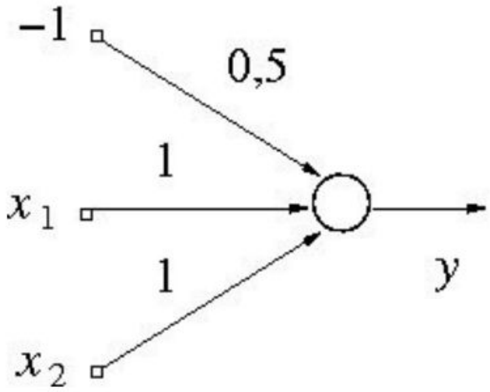


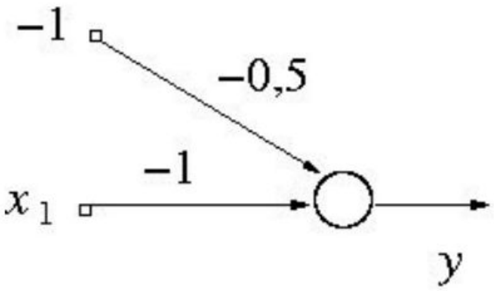
$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + \eta e_i(t)\mathbf{x}(t)$$

Aplicações

- ➢ Implementação para dados (Iris)
- ➢ Implementação para imagens (MNIST)
- ➢ Portas lógicas







Aplicações

➢ **Exemplo:** Porta OR

Porta OR

x_1	x_2	d
0	0	0
0	1	1
1	0	1
1	1	1

t =0: Iniciar com zeros os pesos e o limiar.

$$w_1(0) = w_2(0) = 0 \quad \text{e} \quad \theta = 0$$

Logo:

$$w(0) = \begin{bmatrix} \theta(0) \\ w_1(0) \\ w_2(0) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \qquad x(t) = \begin{bmatrix} -1 \\ x_1(t) \\ x_2(t) \end{bmatrix}$$

Passo de aprendizagem escolhido: $\eta = 0{,}5$;

Aplicações

t=1: Calcular saída para $\mathbf{w}(1) = [0\ \ 0\ \ 0]^T$ e $\mathbf{x}(1) = [-1\ \ 0\ \ 0]^T$.

$$u(1) = (0)(-1) + (0)(0) + (0)(0) = 0 \Rightarrow y(1) = 1,\ e(1) = -1.$$

$$\mathbf{w}(2) = \mathbf{w}(1) + (0{,}5)(-1)\mathbf{x}(1) = \begin{bmatrix}0\\0\\0\end{bmatrix} - 0{,}5\begin{bmatrix}-1\\0\\0\end{bmatrix} = \begin{bmatrix}0{,}5\\0\\0\end{bmatrix}$$

t=2: Calcular saída para $\mathbf{w}(2) = [0{,}5\ \ 0\ \ 0]^T$ e $\mathbf{x}(2) = [-1\ \ 0\ \ 1]^T$.

$$u(2) = (0{,}5)(-1) + (0)(0) + (0)(1) = -0{,}5 \Rightarrow y(2) = 0,\ e(2) = 1$$

$$\mathbf{w}(3) = \mathbf{w}(2) + (0{,}5)(1)\mathbf{x}(2) = \begin{bmatrix}0{,}5\\0\\0\end{bmatrix} + 0{,}5\begin{bmatrix}-1\\0\\1\end{bmatrix} = \begin{bmatrix}0\\0\\0{,}5\end{bmatrix}$$

Aplicações

t=3: Calcular saída para $\mathbf{w}(3) = [0\ \ 0\ \ 0{,}5]^T$ e $\mathbf{x}(3) = [-1\ \ 1\ \ 0]^T$.

$u(3)=(0)(-1)+(0)(1) + (0{,}5)(0) = 0 \Rightarrow y(3) = 1,\ e(3)=0.$

$\mathbf{w}(4) = \mathbf{w}(3)$

t=4: Calcular saída para $\mathbf{w}(4) = [0\ \ 0\ \ 0{,}5]^T$ e $\mathbf{x}(4) = [-1\ \ 1\ \ 1]^T$.

$u(4) = (0)(-1) + (0)(1) + (0{,}5)(1) = 0{,}5 \Rightarrow y(4) = 1,\ e(4)=0.$

$\mathbf{w}(5) = \mathbf{w}(4)$

t=5: Calcular saída para $\mathbf{w}(5) = [0\ \ 0\ \ 0{,}5]^T$ e $\mathbf{x}(5) = \mathbf{x}(1) = [-1\ \ 0\ \ 0]^T$.

$u(5) = (0)(-1) + (0)(0) + (0{,}5)(0) = 0 \Rightarrow y(5) = 1,\ e(5)=-1.$

$$\mathbf{w}(5) = \mathbf{w}(4)+(0{,}5)(-1)\mathbf{x}(5) = \begin{bmatrix} 0 \\ 0 \\ 0{,}5 \end{bmatrix} - 0{,}5 \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0{,}5 \\ 0 \\ 0{,}5 \end{bmatrix}$$

Aplicações

t=6: Calcular saída para $\mathbf{w}(6) = [0,5 \;\; 0 \;\; 0,5]^T$ e $\mathbf{x}(6) = \mathbf{x}(2) = [-1 \;\; 0 \;\; 1]^T$.

$u(6)=(0,5)(-1)+(0)(0) + (0,5)(1) = 0 \Rightarrow y(6)= 1, e(6)=0.$

$\mathbf{w}(7) = \mathbf{w}(6)$

t=7: Calcular saída para $\mathbf{w}(7) = [0,5 \;\; 0 \;\; 0,5]^T$ e $\mathbf{x}(7) = \mathbf{x}(3) = [-1 \;\; 1 \;\; 0]^T$.

$u(7) = (0,5)(-1) + (0)(1) + (0,5)(0) = -0,5 \Rightarrow y(7) = 0, e(7)=1.$

$$\mathbf{w}(8) = \mathbf{w}(7)+(0,5)(1)\mathbf{x}(7) = \begin{bmatrix} 0,5 \\ 0 \\ 0,5 \end{bmatrix} + 0,5 \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0,5 \\ 0,5 \end{bmatrix}$$

t=8: Calcular saída para $\mathbf{w}(8) = [0 \;\; 0,5 \;\; 0,5]^T$ e $\mathbf{x}(8) = \mathbf{x}(4) = [-1 \;\; 1 \;\; 1]^T$.

$u(8) = (0)(-1)+(0,5)(1)+(0,5)(1) = 1 \Rightarrow y(8) = 1, e(8)=0.$

$\mathbf{w}(9) = \mathbf{w}(8)$

Aplicações

t=9: Calcular saída para $\mathbf{w}(9) = [0\ 0{,}5\ \ 0{,}5]^T$ e $\mathbf{x}(9)=\mathbf{x}(1)=[-1\ \ 0\ \ 0]^T$.

$u(9)=(0)(-1)+(0{,}5)(0) + (0{,}5)(0) = 0 \Rightarrow y\,(9) = 1,\ e(9)=-1.$

$$\mathbf{w}(10) = \mathbf{w}(9)+(0{,}5)(-1)\mathbf{x}(9) = \begin{bmatrix} 0 \\ 0{,}5 \\ 0{,}5 \end{bmatrix} - 0{,}5 \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0{,}5 \\ 0{,}5 \\ 0{,}5 \end{bmatrix}$$

t=10: Calcular saída para $\mathbf{w}(10) = [0{,}5\ 0{,}5\ 0{,}5]^T$ e $\mathbf{x}(10)=\mathbf{x}(2)=[-1\ \ 0\ \ 1]^T$.

$u(10) = (0{,}5)(-1)+(0{,}5)(0)+(0{,}5)(1) = 0 \Rightarrow y(10) = 1,\ e(10) = 0.$

$\mathbf{w}(11) = \mathbf{w}(10)$

t=11: Calcular saída para $\mathbf{w}(11) = [0{,}5\ \ 0{,}5\ \ 0{,}5]^T$ e $\mathbf{x}(11)=\mathbf{x}(3)=[-1\ \ 1\ \ 0]^T$.

$u(11) = (0{,}5)(-1)+(0{,}5)(1)+(0{,}5)(0) = 0 \Rightarrow y(11) = 1,\ e(11)=0.$

$\mathbf{w}(12) = \mathbf{w}(11)$

Aplicações

t=12: Calcular saída para $\mathbf{w}(12) = [0{,}5 \;\; 0{,}5 \;\; 0{,}5]^T$ e $\mathbf{x}(12)=\mathrm{x}(4)=[-1 \;\; 1 \;\; 1]^T$.

$u(12)=(0{,}5)(-1)+(0{,}5)(1) + (0{,}5)(1) = 0{,}5 \Rightarrow y\,(12)= 1,\; e(12)=0.$

$\mathbf{w}(13) = \mathbf{w}(12)$

t=13: Calcular saída para $\mathbf{w}(13) = [0{,}5 \;\; 0{,}5 \;\; 0{,}5]$ e $\mathbf{x}(13) =[-1 \;\; 0 \;\; 0]$.

$u(13) = (0{,}5)(-1) + (0{,}5)(0) + (0{,}5)(0) = -0{,}5 \Rightarrow y(13) = 0,\; e(13)=0.$

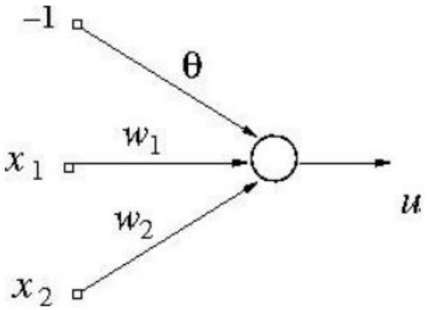
$\mathbf{w}(14) = \mathbf{w}(13)$

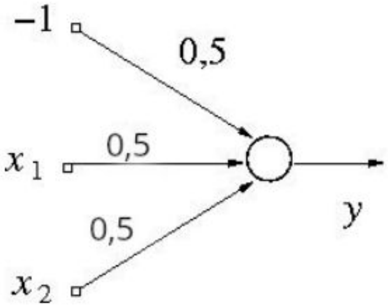
FIM do treinamento!

Aplicações

Solução Encontrada:

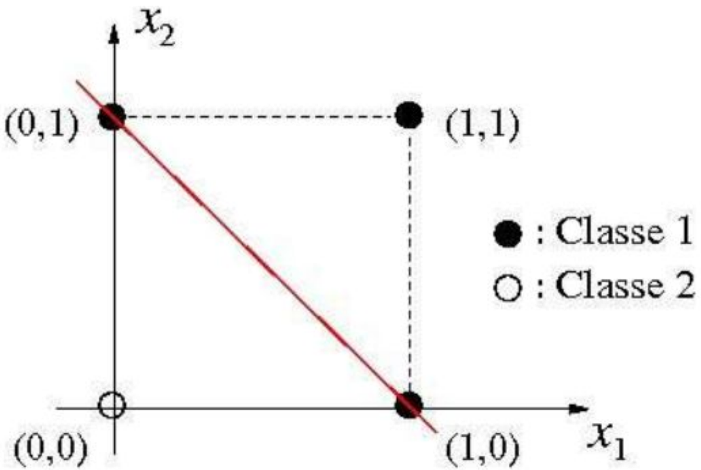
$$\mathbf{w} = \begin{bmatrix} \theta \\ w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 0{,}5 \\ 0{,}5 \\ 0{,}5 \end{bmatrix}$$





Aplicações

Solução Encontrada:



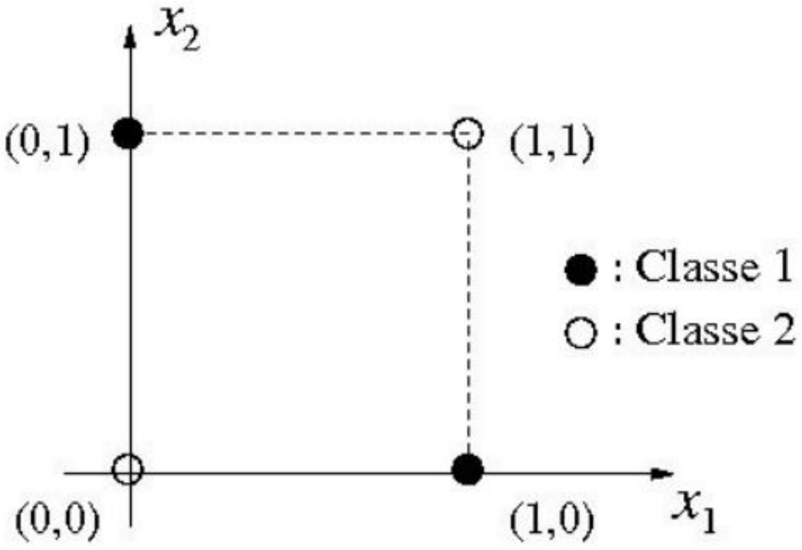
Note que esta não é a melhor das soluções, porque a reta passa bem em cima dos pontos (1,0) e (0,0). Se os pesos tivessem sido iniciados aleatoriamente, dificilmente uma situação como essa ocorreria.

Aplicações

➢ **Problema:** Porta XOR (?)

Porta XOR

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0



Referências

1. A. Webb. Statistical Pattern Recognition. John Wiley & Sons, 2nd edition, 2002.
2. HAYKIN, Simon. Redes Neurais: Princípios e prática. Bookman, 2nd edition, 2001.
3. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. "Gradient-based learning applied to document recognition." Proceedings of the IEEE, 86(11):2278-2324, November 1998.
4. M.-L. Zhang and Z.-H. Zhou, "A review on multi-label learning algorithms", IEEE transactions on knowledge and data engineering, vol. 26, no. 8, pp. 1819–1837, 2013.
5. Hastie, T., Tibshirani, R., Friedman, J., The elements of statistical learning: Data mining, inference, and prediction. Springer, 2009.
6. FISHER, Ronald A.; MARSHALL, Michael. Iris data set. RA Fisher, UC Irvine Machine Learning Repository, v. 440, p. 87, 1936.

Obrigado pela atenção!

Dúvidas?

Prof. Matheus Santos
matheus.santos@lapisco.ifce.edu.br