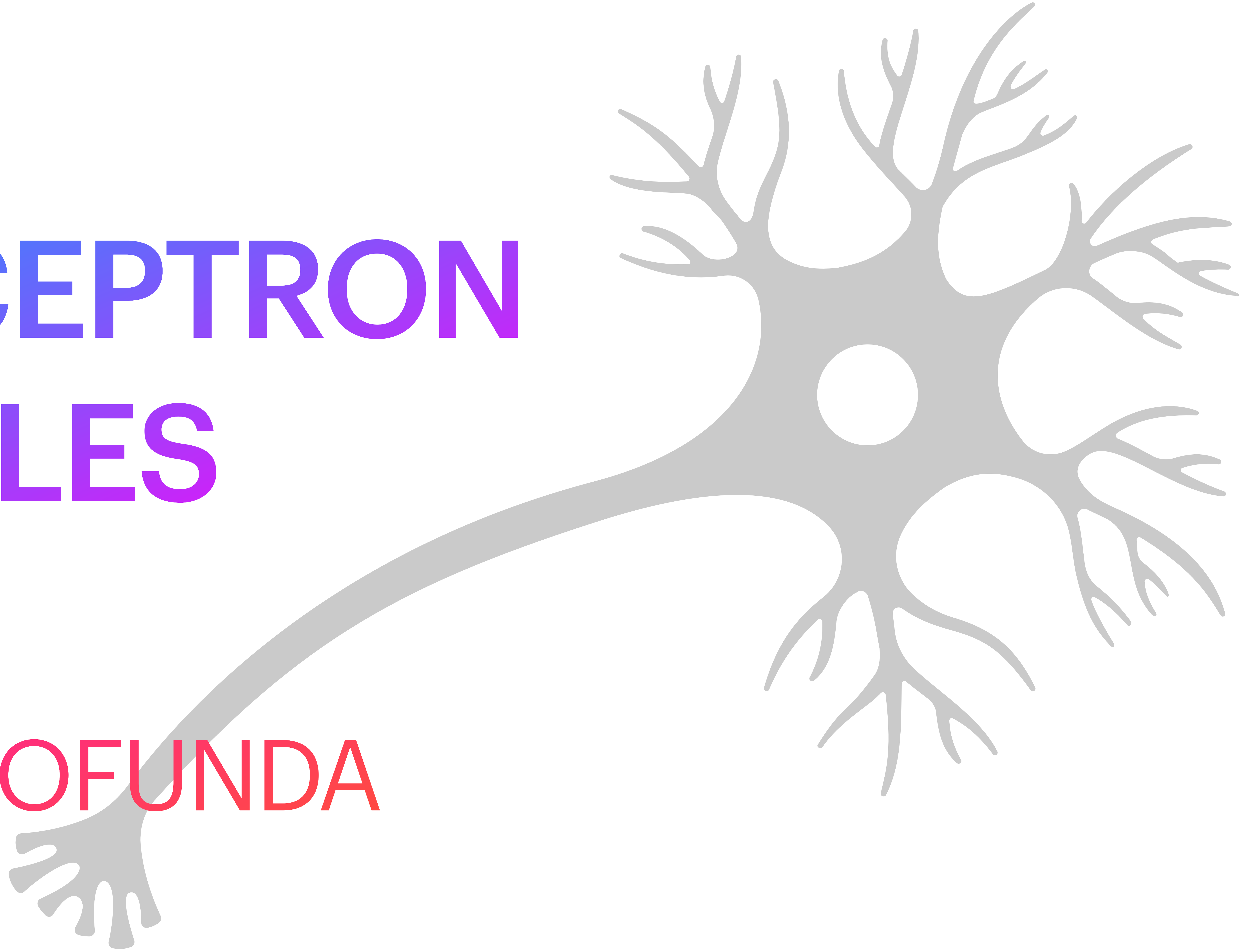
INSTITUTO FEDERAL
Ceará
Programa de Pós-Graduação
em Ciência da Computação



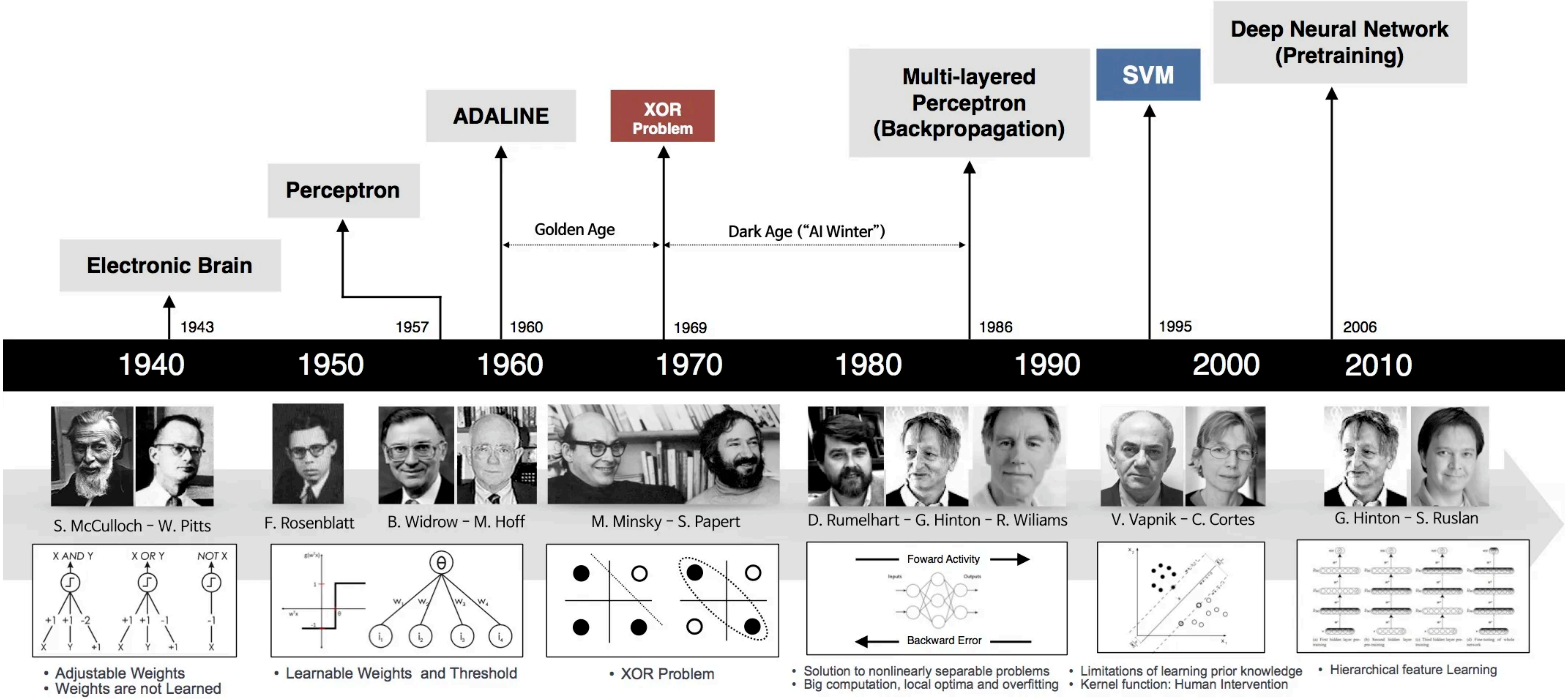
02 PERCEPTRON SIMPLES

APREDIZAGEM PROFUNDA

PPGCC – 2023.1

Prof. Saulo Oliveira <saulo.oliveira@ifce.edu.br>

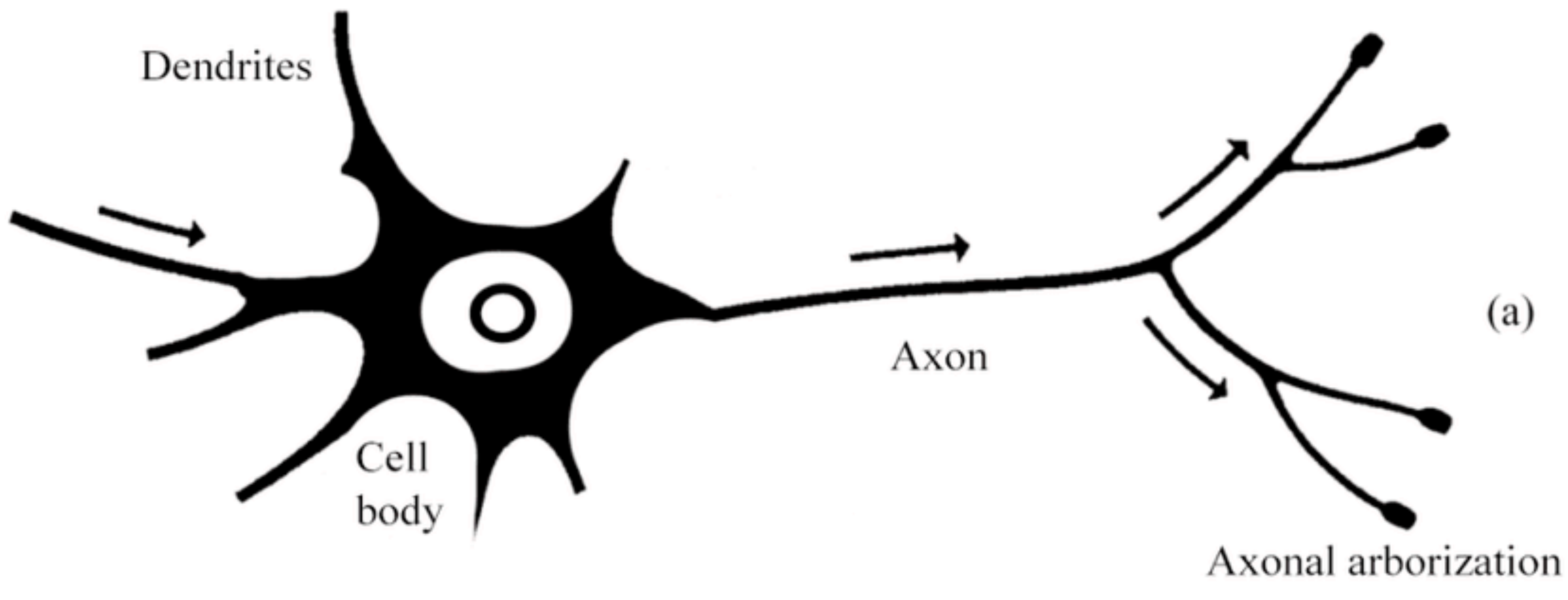
Evolução das Redes Neurais Artificiais



Fonte: https://sefiks.com/2017/10/14/evolution-of-neural-networks/

Neurônios

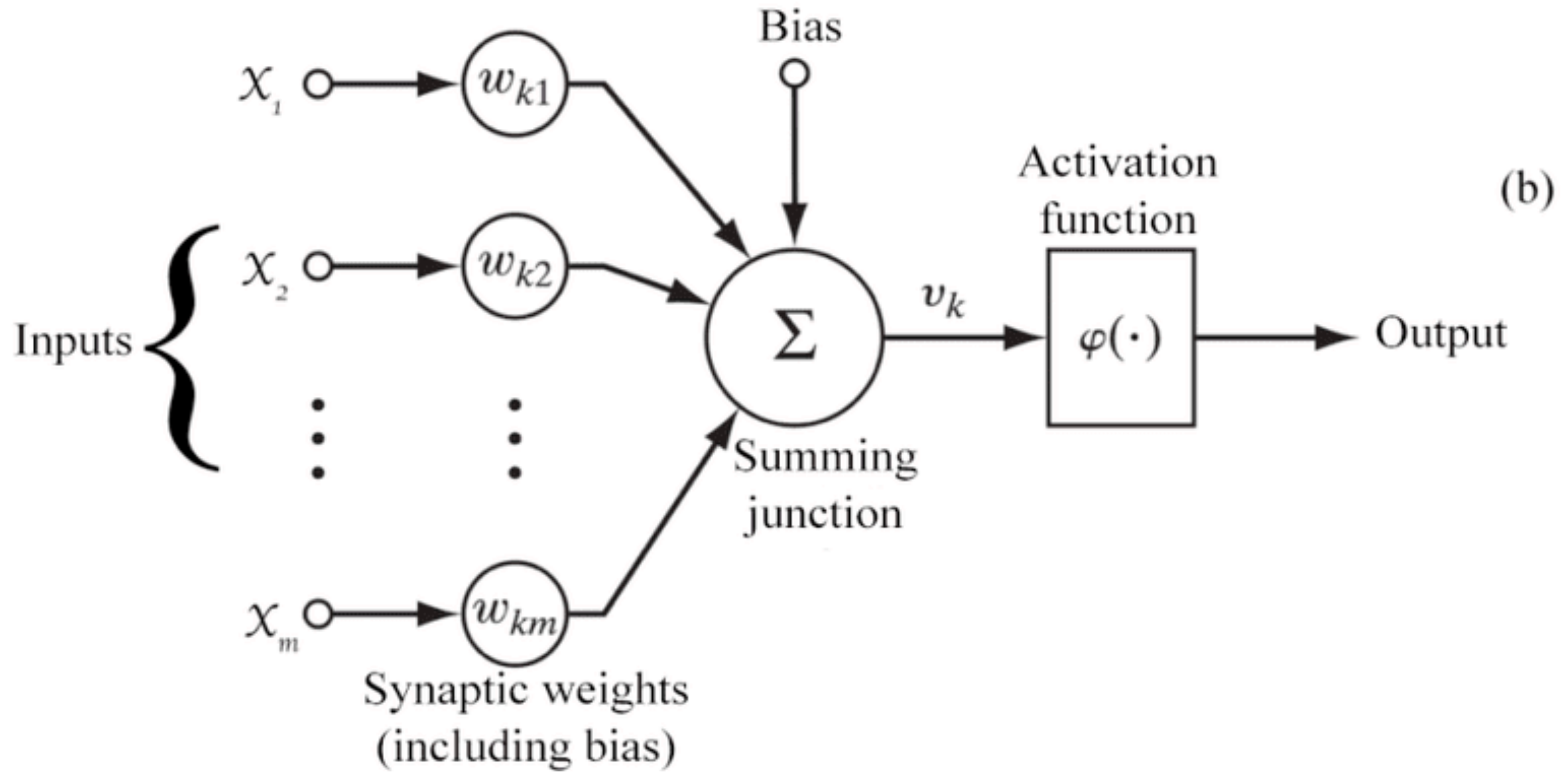
Neurônio biológico



- Dendritos: Recebem sinais de outros neurônios;
- Corpo celular: Processa a informação;
- Axônio: Transmite a saída do neurônio em questão;
- Sinapse: Ponto de conexão para outros neurônios.

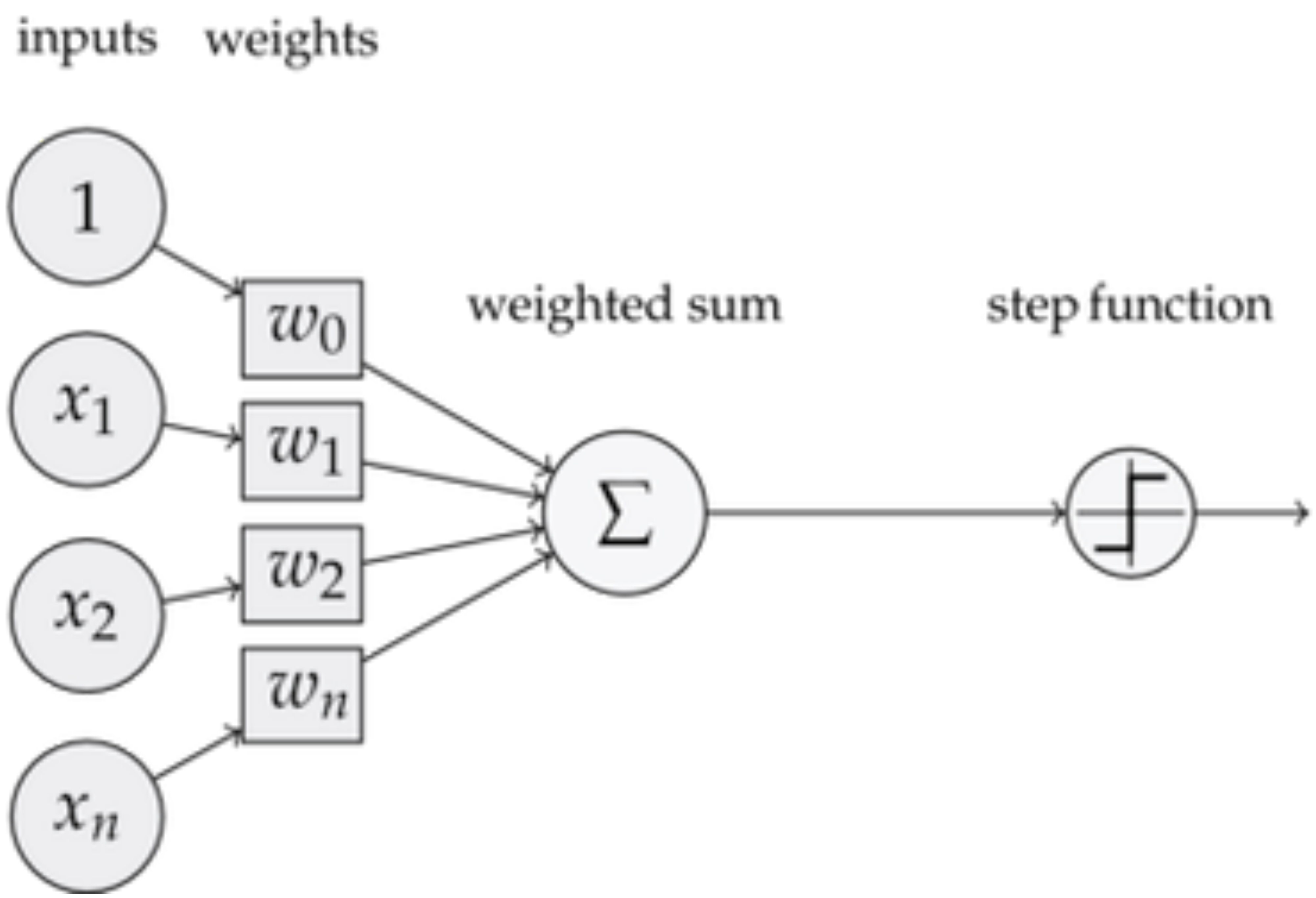
Fonte: https://dominicm73.blogspot.com/2020/08/modeling-threshold-logic-neurons-and.html

Neurônio artificial



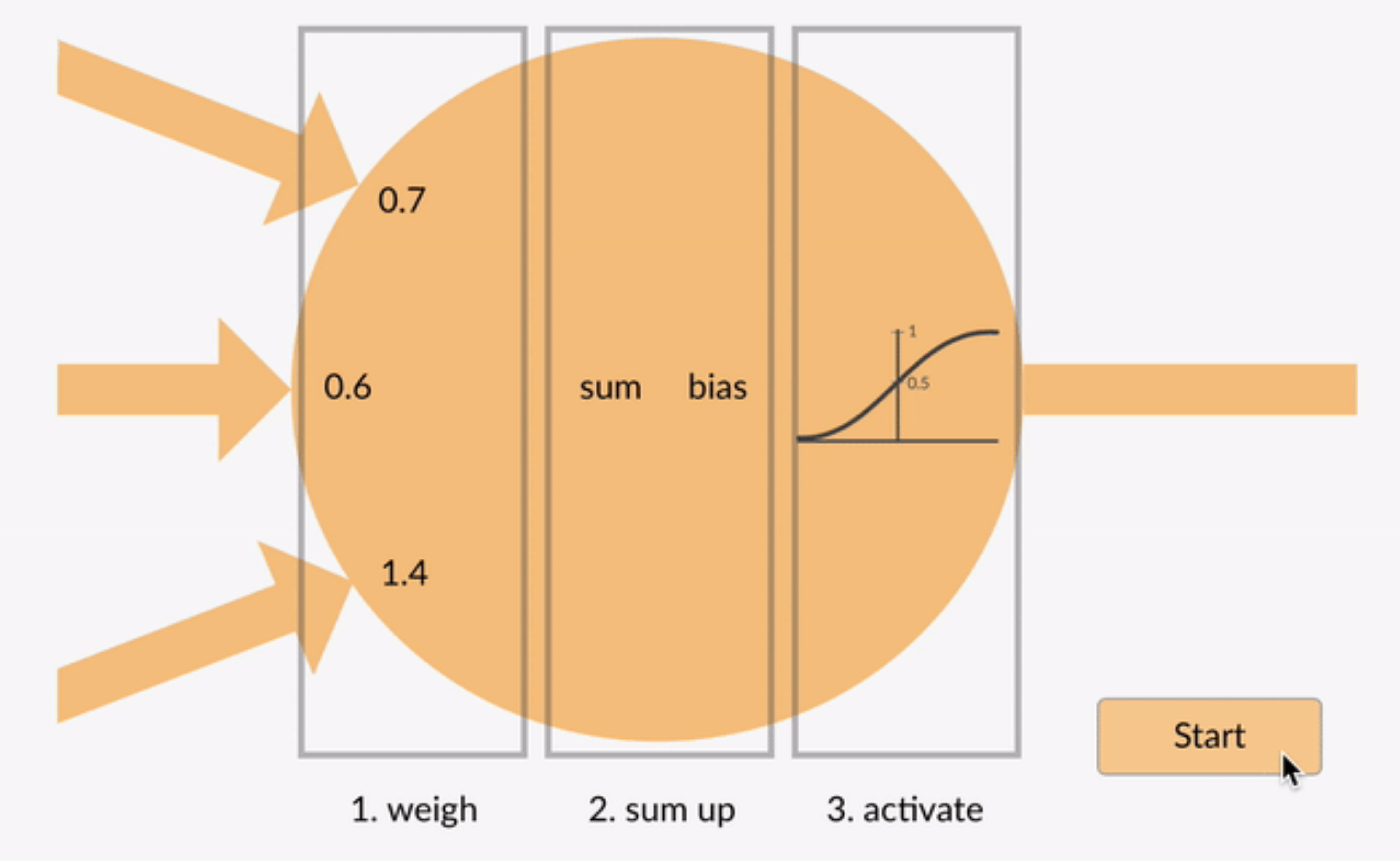
- Entrada: Recebem as informações de entrada;
- Pesos sinápticos: Ponderam as informações de entrada;
- Junção aditiva: Combina (soma) as informações ponderadas;
- Função de ativação: Despenha o papel de excitação/inibição da informação processada.
- Saída: Ponto de conexão para outros neurônios.

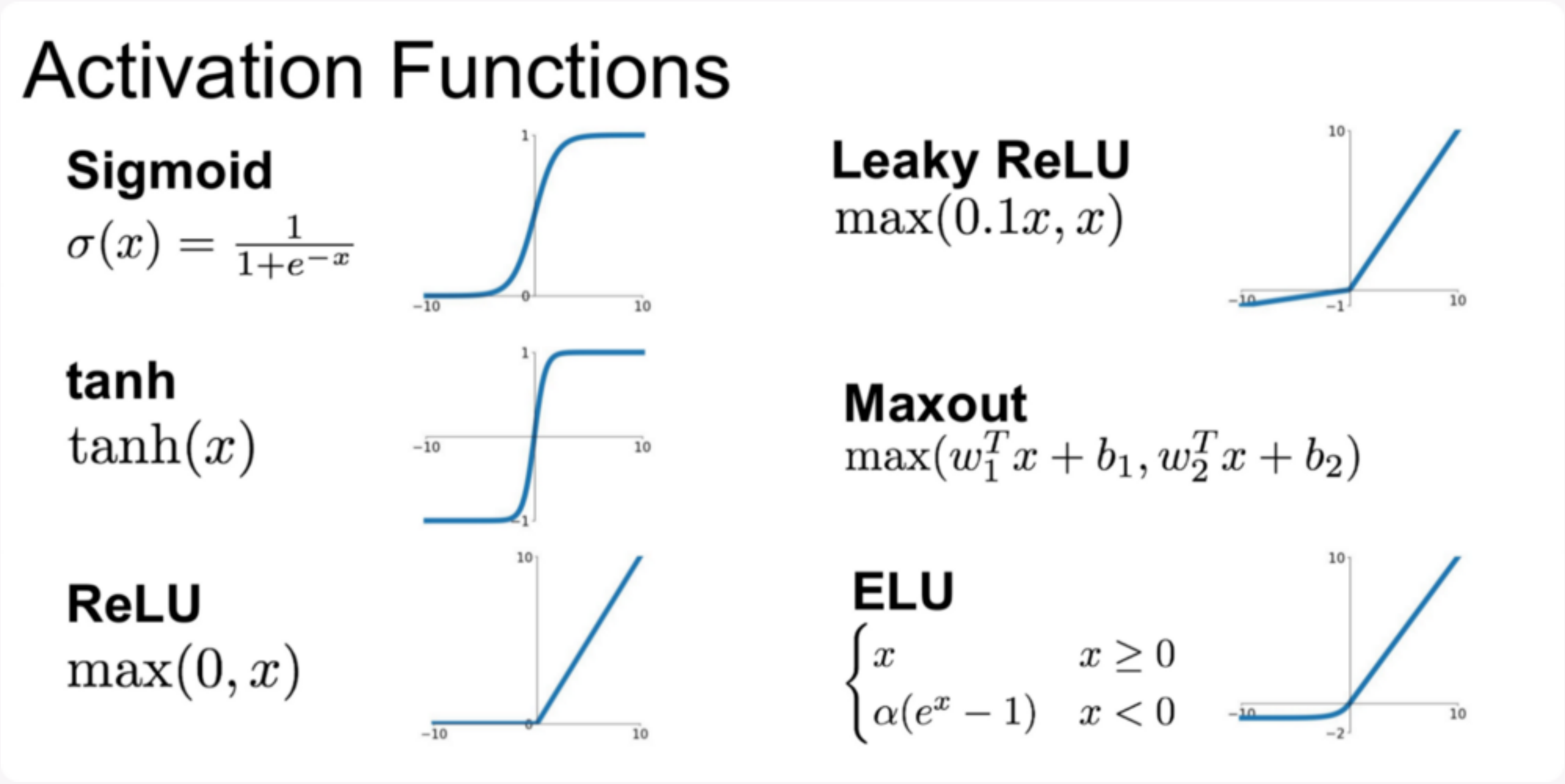
Neurônio artificial



- Entrada: Recebem as informações de entrada;
- Pesos sinápticos: Ponderam as informações de entrada;
- Junção aditiva: Combina (soma) as informações ponderadas;
- Função de ativação: Despenha o papel de excitação/inibição da informação processada.
- Saída: Ponto de conexão para outros neurônios.

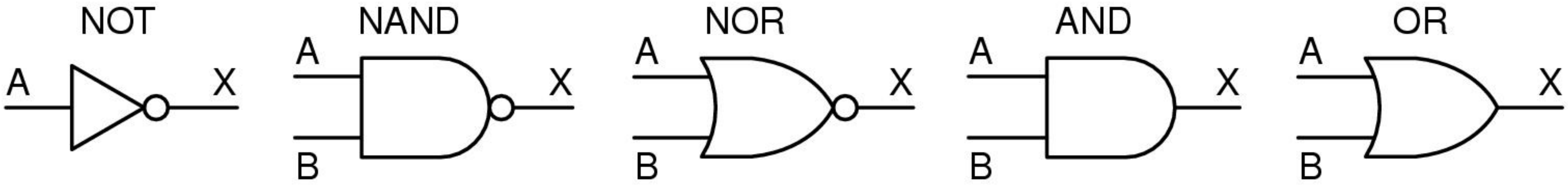
Visão geral de um neurônio artificial





Fonte: https://laptrinhx.com/introduction-to-deep-learning-feed-forward-neural-networks-ffnns-a-k-a-1737445691/

Portas lógicas



A	X
0	1
1	0

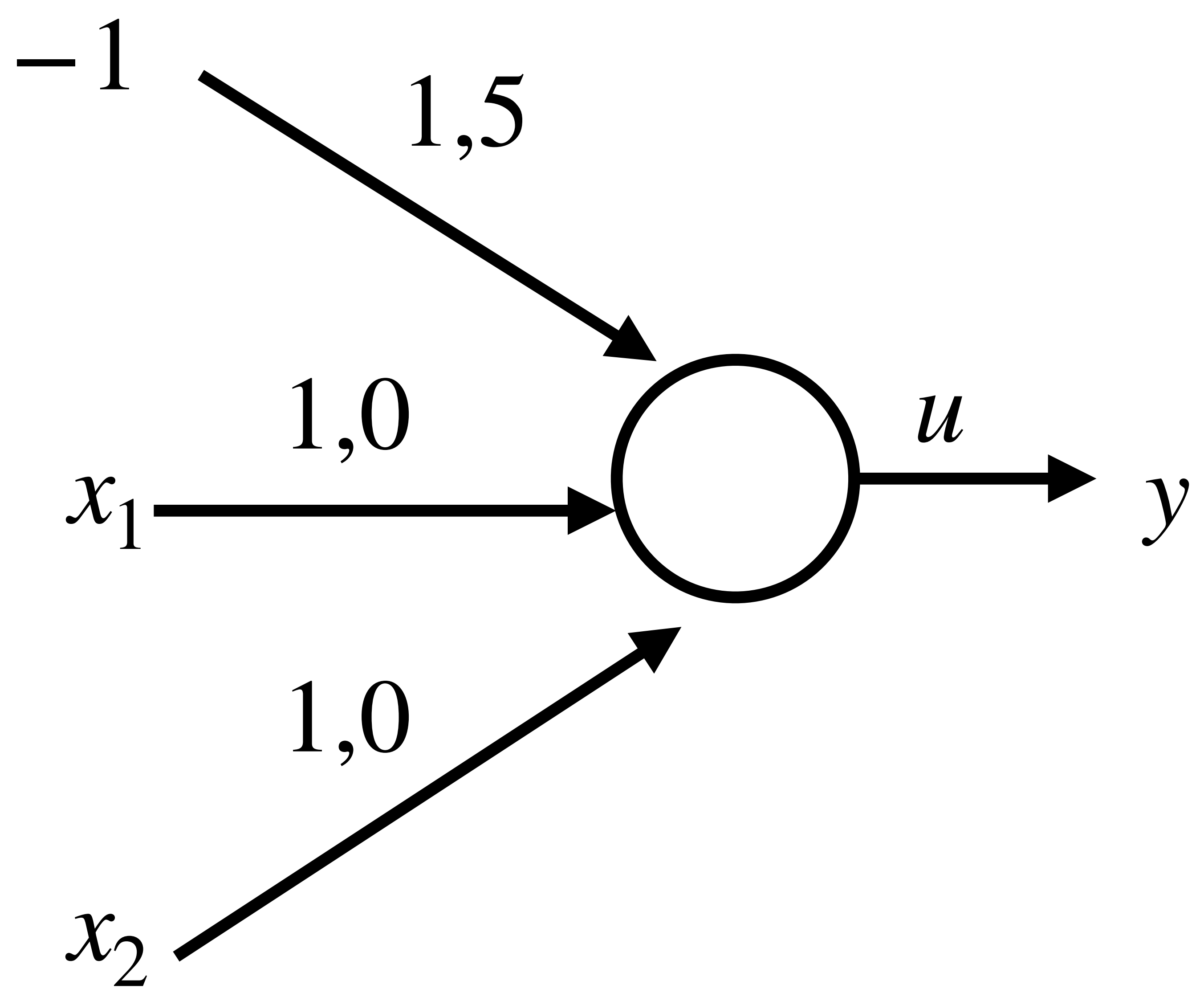
A	B	X
0	0	1
0	1	1
1	0	1
1	1	0

A	B	X
0	0	1
0	1	0
1	0	0
1	1	0

A	B	X
0	0	0
0	1	0
1	0	0
1	1	1

A	B	X
0	0	0
0	1	1
1	0	1
1	1	1

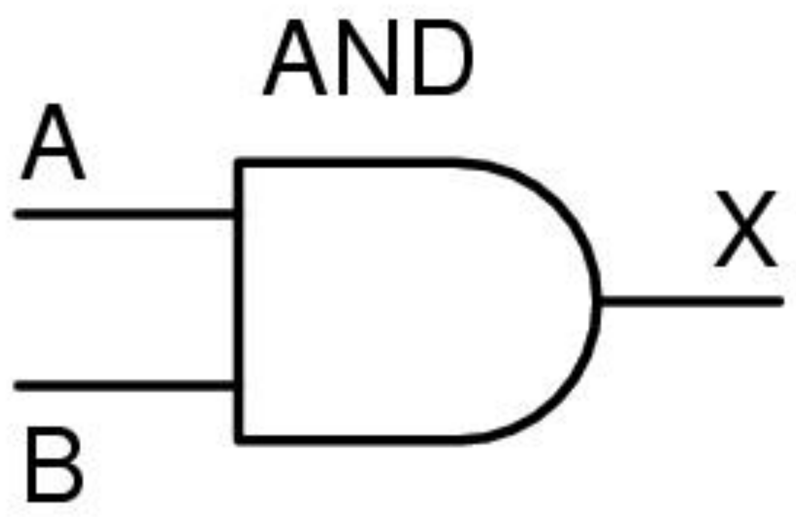
Portas lógicas: AND



$$w_1 = w_2 = 1 \text{ e } \theta = 1{,}5$$

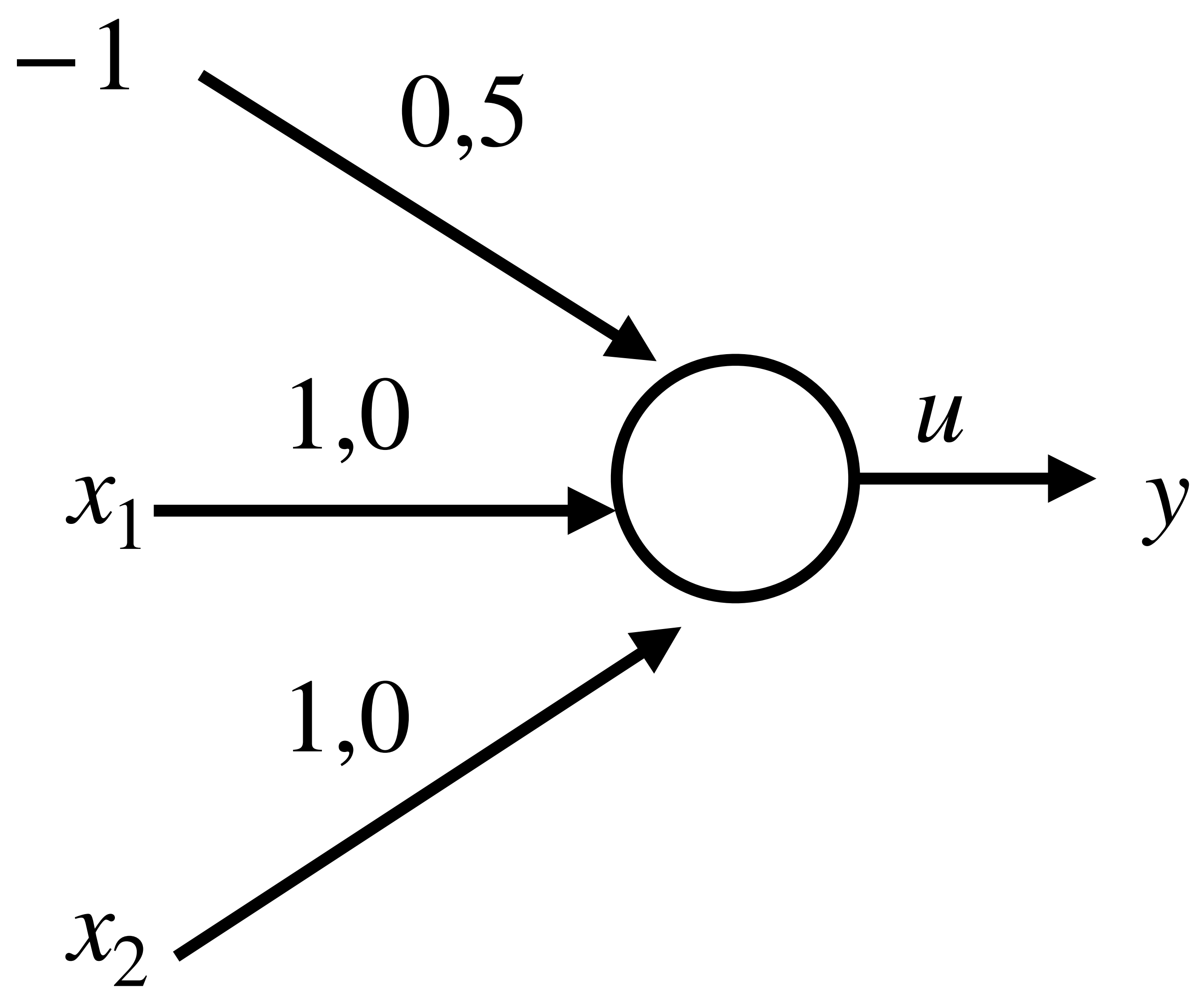
$$y = 1 \text{ se } u \geq 0.$$

$$y = 0 \text{ se } u < 0.$$



A	B	X
0	0	0
0	1	0
1	0	0
1	1	1

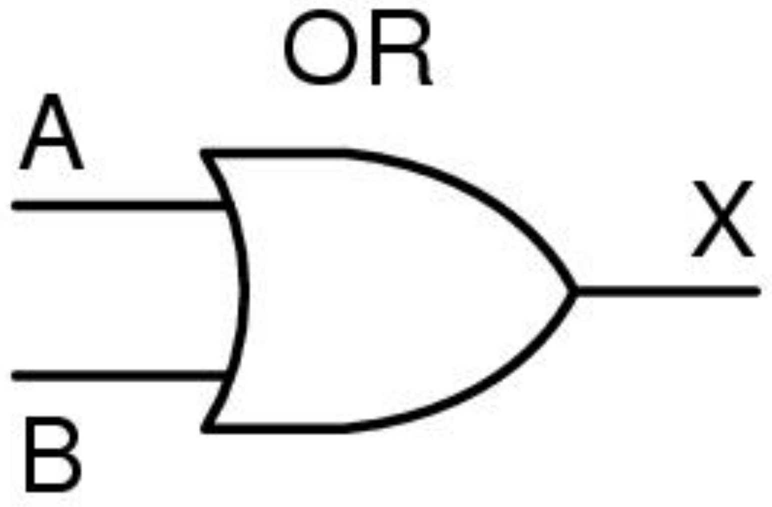
Portas lógicas: OR



$$w_1 = w_2 = 1 \text{ e } \theta = 0{,}5$$

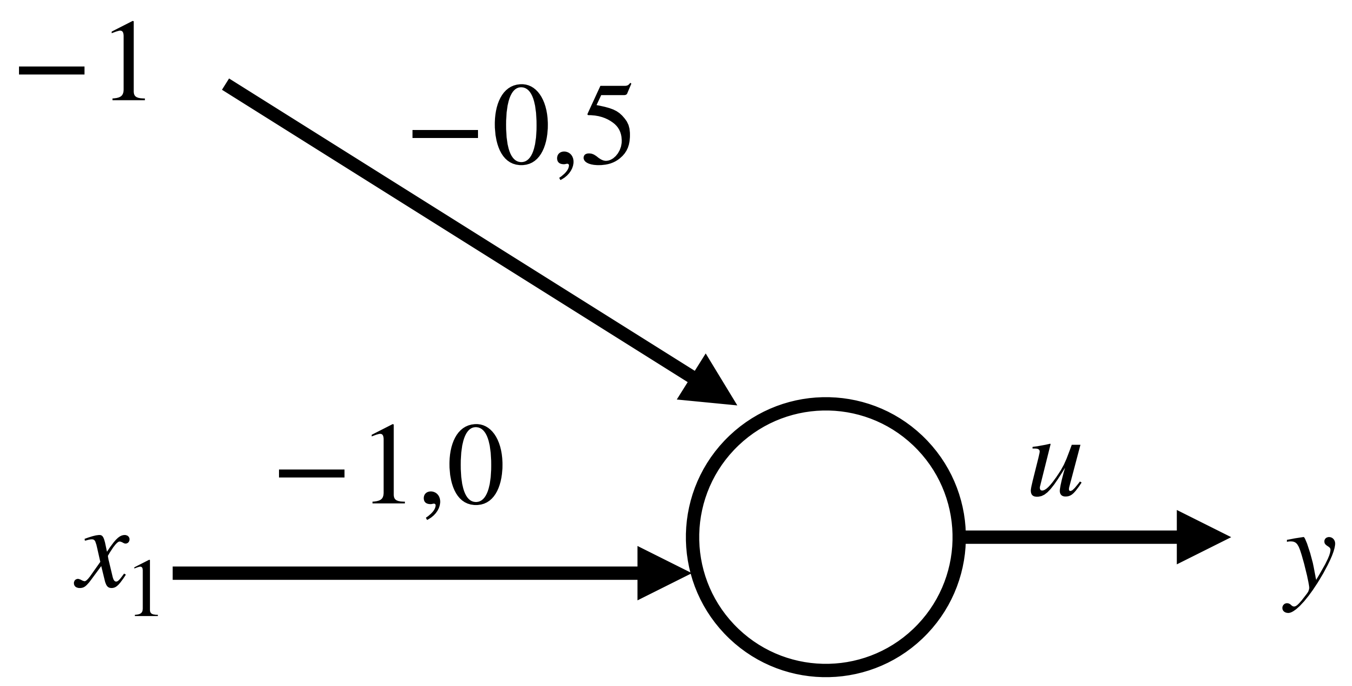
$$y = 1 \text{ se } u \geq 0.$$

$$y = 0 \text{ se } u < 0.$$



A	B	X
0	0	0
0	1	1
1	0	1
1	1	1

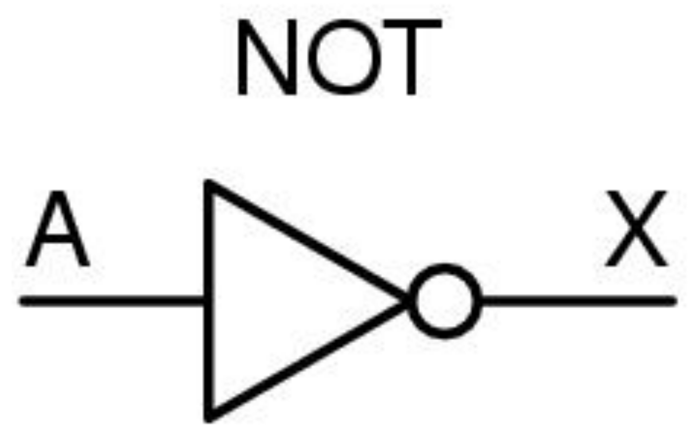
Portas lógicas: NOT



$$w_1 = -1 \text{ e } \theta = -0,5$$

$$y = 1 \text{ se } u \geq 0.$$

$$y = 0 \text{ se } u < 0.$$



A	X
0	1
1	0

Como poderia se determinar cada um dos valores para os pesos sinápticos?

Como poderia se determinar cada um dos valores para os pesos sinápticos?

R: **REGRA DE APRENDIZADO**

Regra de aprendizagem

- O processo de aprendizagem consiste basicamente na modificação dos pesos e do limiar do neurônio de M-P até que ele resolva o problema de interesse ou que o período de aprendizagem tenha finalizado.
- A ideia é pensar que deve haver uma variação dos valores contidos nos pesos durante o aprendizado, ou seja,

$$\Delta \mathbf{w} = \mathbf{w}^{(t+1)} - \mathbf{w}^{(t)}\,.$$
$$\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} + \Delta \mathbf{w}\,.$$

em que:

- $\Delta \mathbf{w}$ é o incremento na memória necessário para o ajuste em relação a um vetor de entrada $\mathbf{x}$;
- $\mathbf{w}^{(t)}$ representa o conhecimento no tempo t salvo na memória; e
- $\mathbf{w}^{(t+1)}$ representa o conhecimento no tempo $t+1$ salvo na memória.

Produto escalar

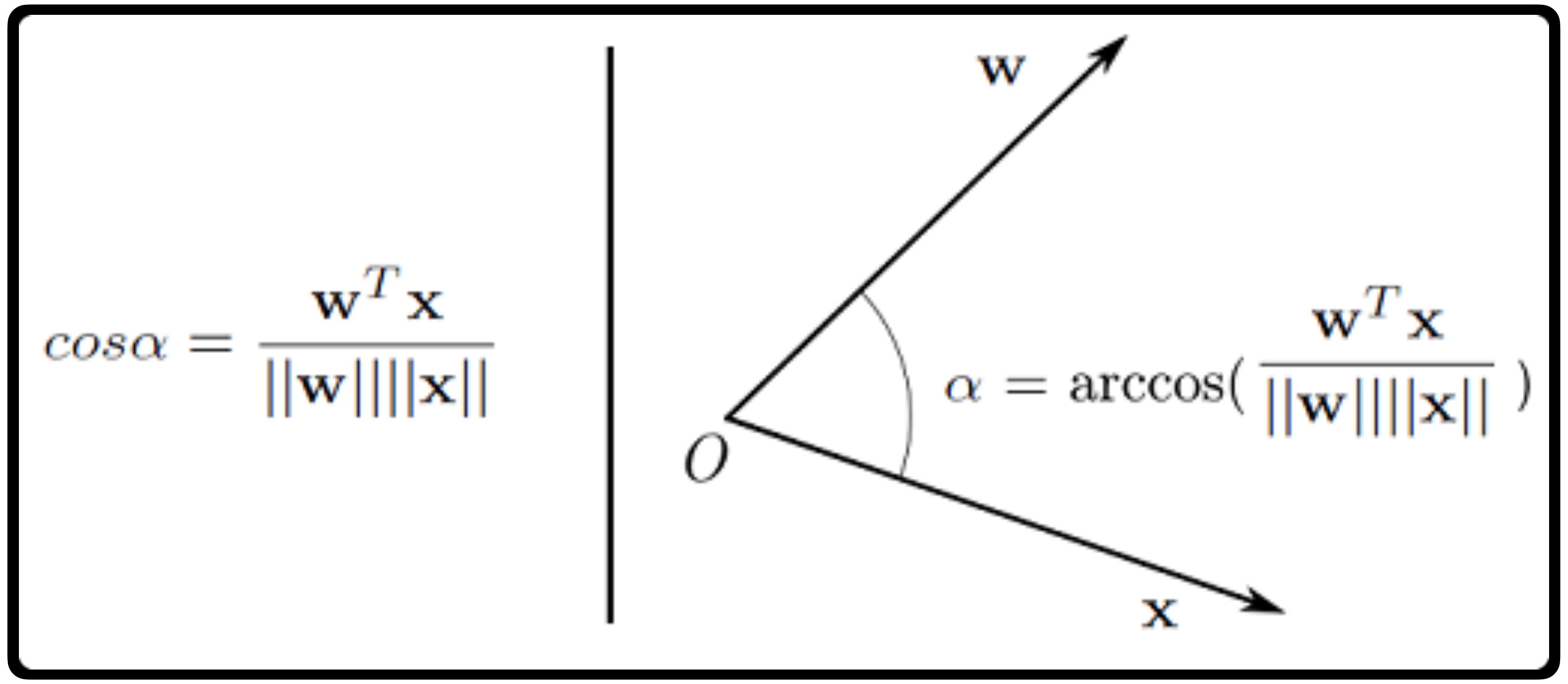
- O produto escalar é definido como o produto de um vetor linha por um vetor coluna, o que equivale a multiplicar cada componente de um vetor pelo seu correspondente no outro vetor e depois somar cada produto:

$$u = \mathbf{w}^{\mathsf{T}}\mathbf{x} = \mathbf{x}^{\mathsf{T}}\mathbf{w} = \sum_i x_i w_i.$$

- Alternativamente, o produto escalar pode ser definido como o produto dos comprimentos dos vetores com o cosseno do menor ângulo entre eles:

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

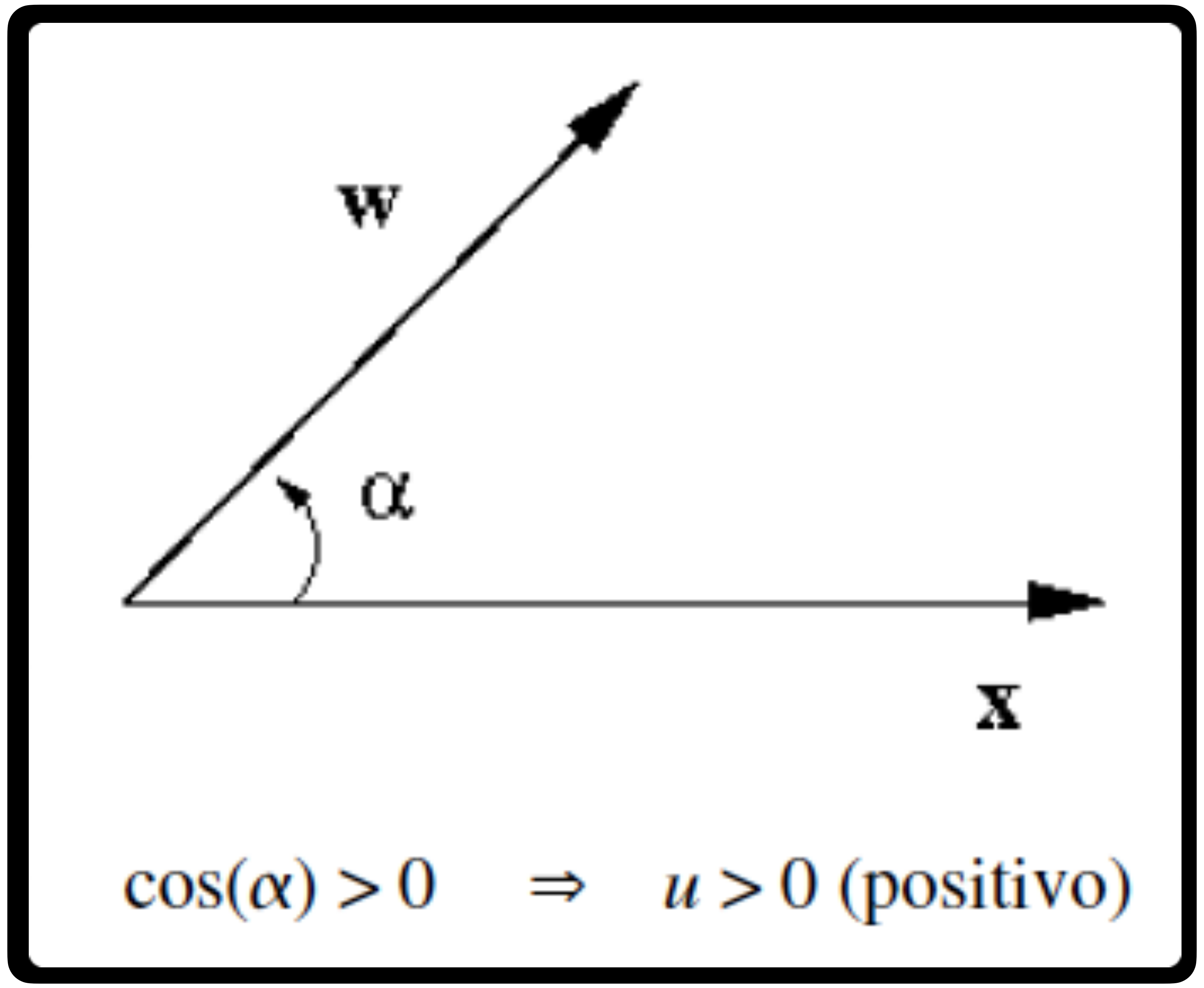
- O produto escalar é uma medida de similaridade entre vetores.
- Para vetores de comprimento fixo, quanto menor o ângulo entre eles, maior é o valor resultante do produto escalar.

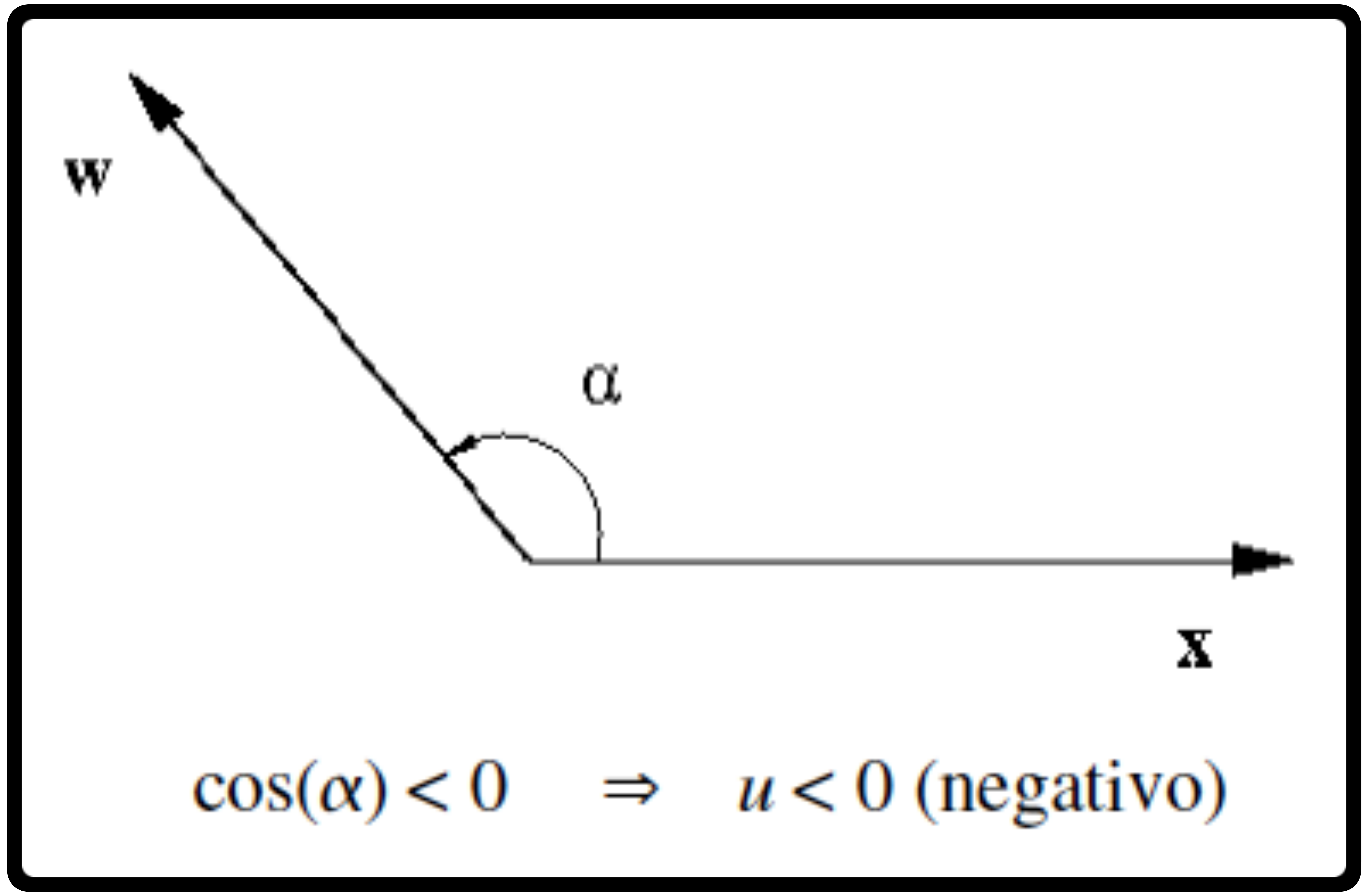


Produto escalar: $u = \| \mathbf{x} \| \| \mathbf{w} \| \cos\alpha$

Considerando a definição $u = \| \mathbf{x} \| \| \mathbf{w} \| \cos\alpha$.

- **CENÁRIO A:** $\cos\alpha > 0,\ u > 0$ **(positivo)**
- CENÁRIO B: $\cos\alpha < 0,\ u < 0$ **(negativo)**





Regra de aprendizagem: definindo $\Delta \mathbf{w}$

Consideremos os argumentos geométricos para esta definição. Assim, 03 casos são derivados para a variável de erro $e^{(t)}$:

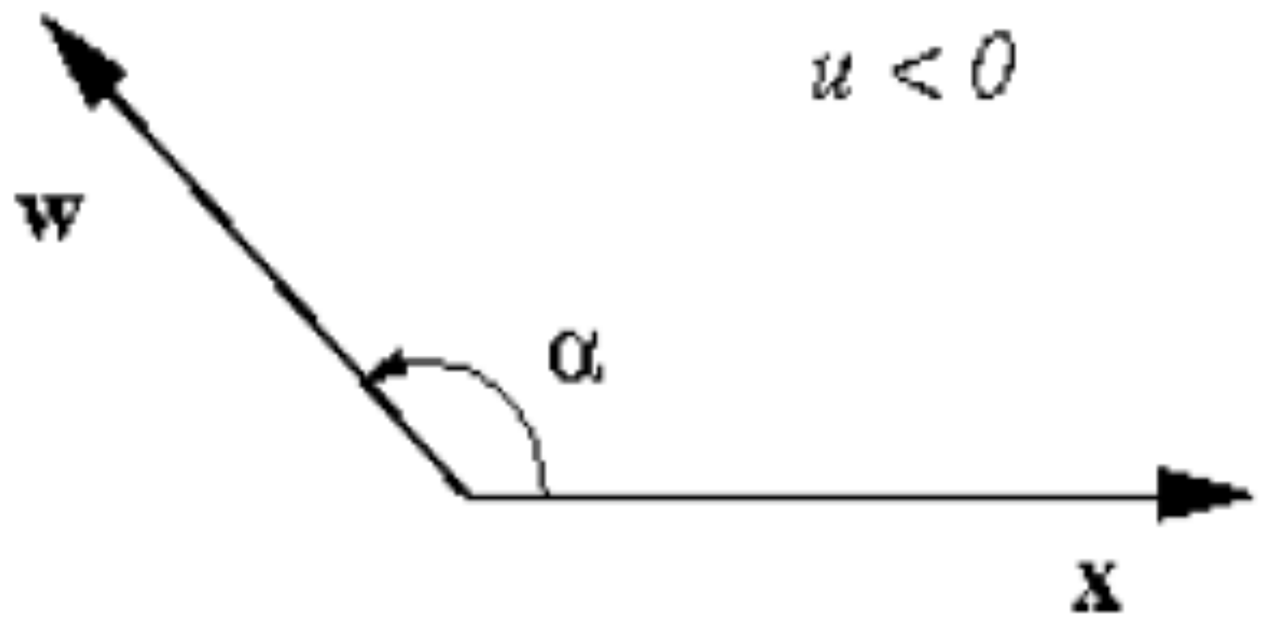
- CASO 1: $e^{(t)} = d^{(t)} - y^{(t)} = +1$, para $d^{(t)} = 1,\ y^{(t)} = 0$;
- CASO 2: $e^{(t)} = d^{(t)} - y^{(t)} = -1$, para $d^{(t)} = 0,\ y^{(t)} = 1$;
- CASO 3a: $e^{(t)} = d^{(t)} - y^{(t)} = 0$, para $d^{(t)} = 0,\ y^{(t)} = 0$.
- CASO 3b: $e^{(t)} = d^{(t)} - y^{(t)} = 0$, para $d^{(t)} = 1,\ y^{(t)} = 1$.

Nos CASO 1 e no CASO 2 houve erro. No CASO 3 houve acerto.

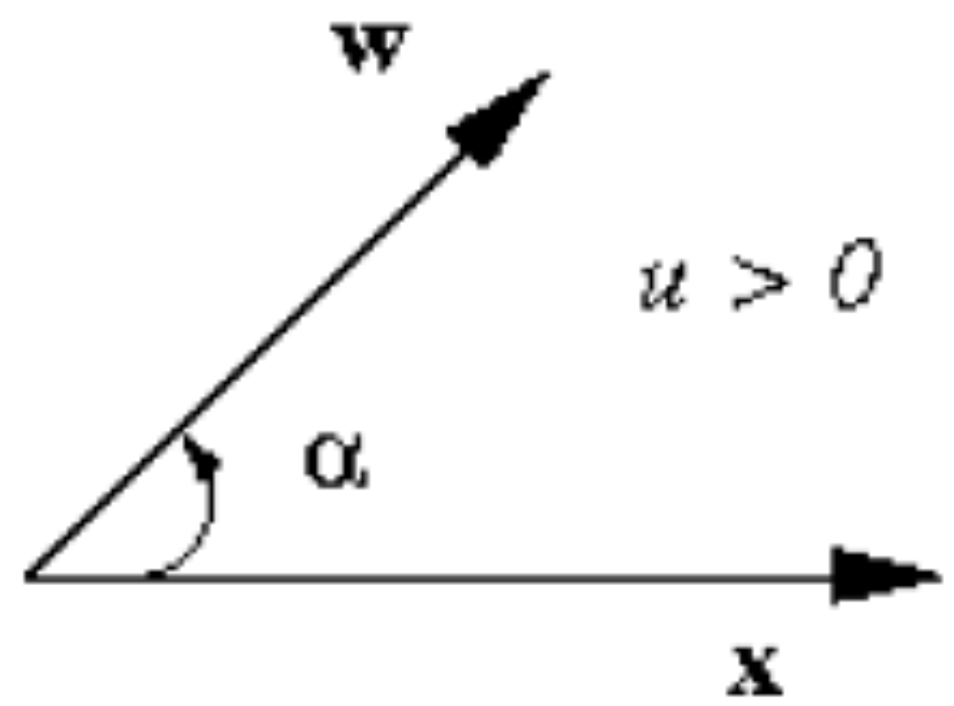
Representação gráfica do CASO 1

Caso 1: $e = d - y = +1$ (d=+1 e y=0)

Situação ocorrida (u<0, y=0):



Situação desejada (u>0, y=1):



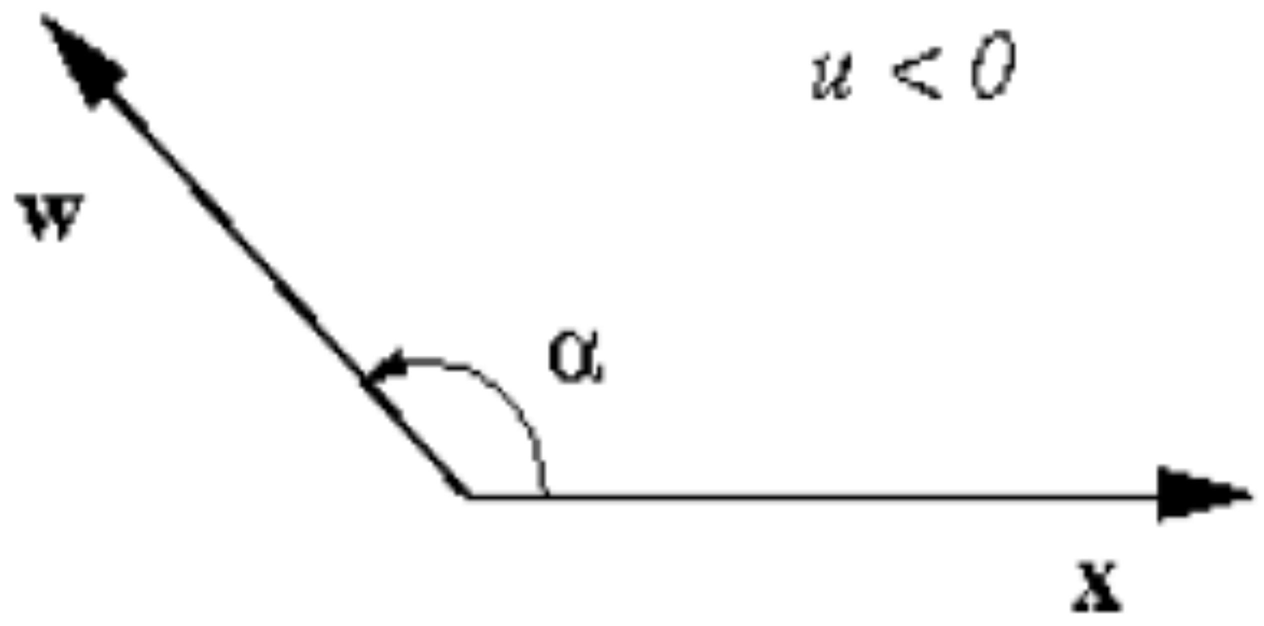
$y = 1$, se $u \geqslant 0$.

$y = 0$, se $u < 0$.

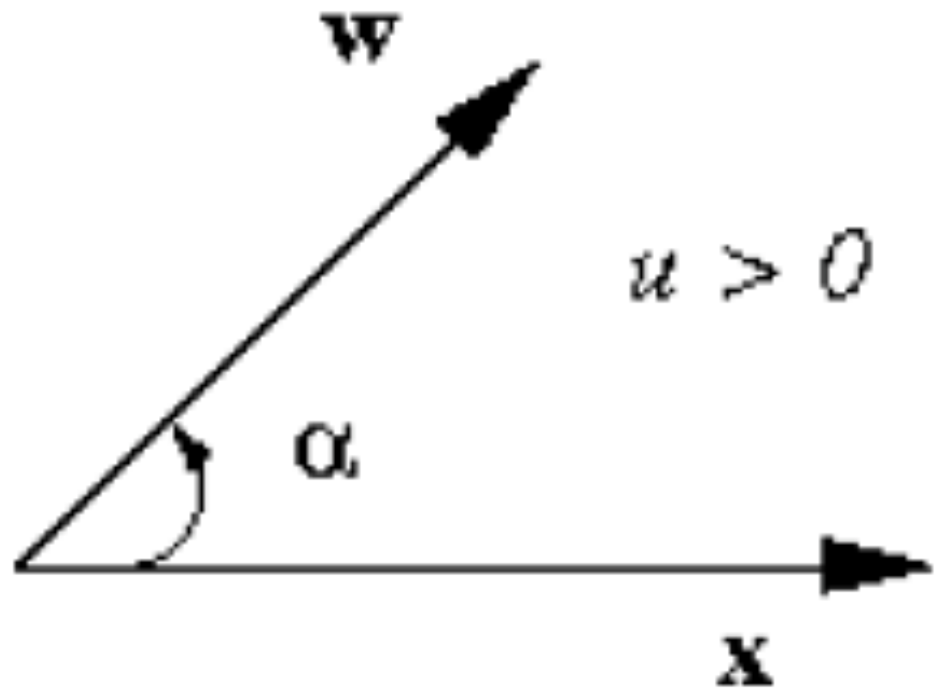
Representação gráfica do CASO 1

Caso 1: $e = d - y = +1$ (d=+1 e y=0)

Situação ocorrida (u<0, y=0):



Situação desejada (u>0, y=1):



Lembre-se que:

$y = 1$, se $u \geqslant 0$.

$y = 0$, se $u < 0$.

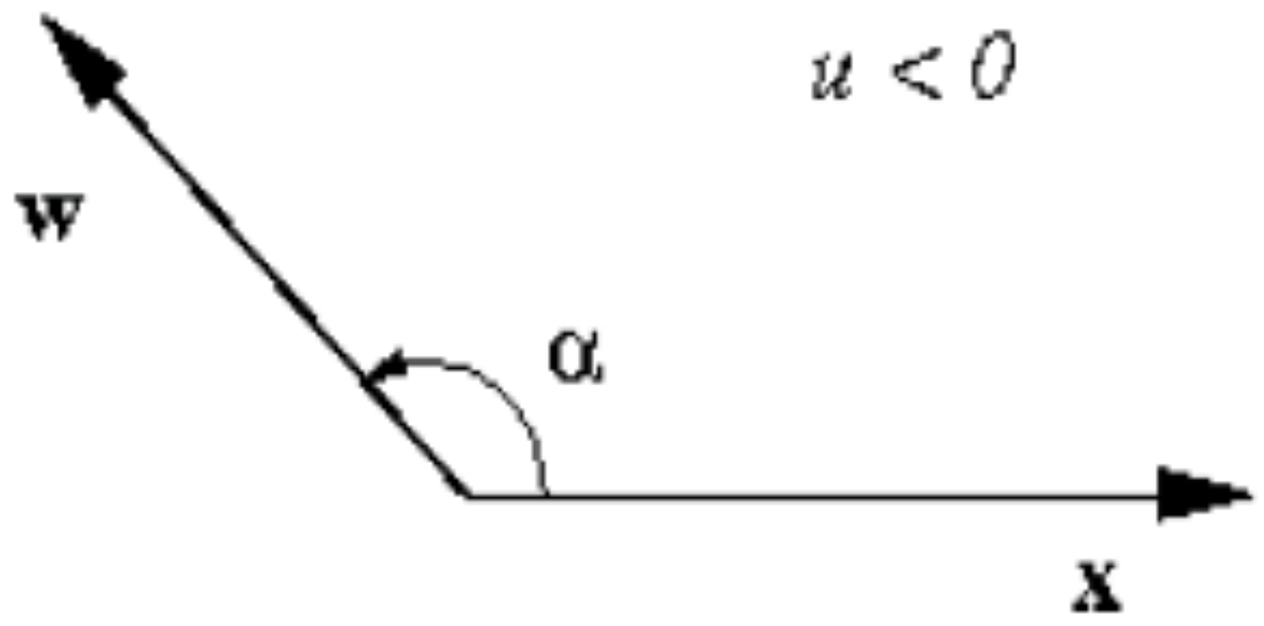
O vetor $\mathbf{w}$ deve ser ajustado para se aproximar de $\mathbf{x}$. Assim, neste caso, a regra de atualização é descrita como:

$$\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} + \mathbf{x}\,.$$

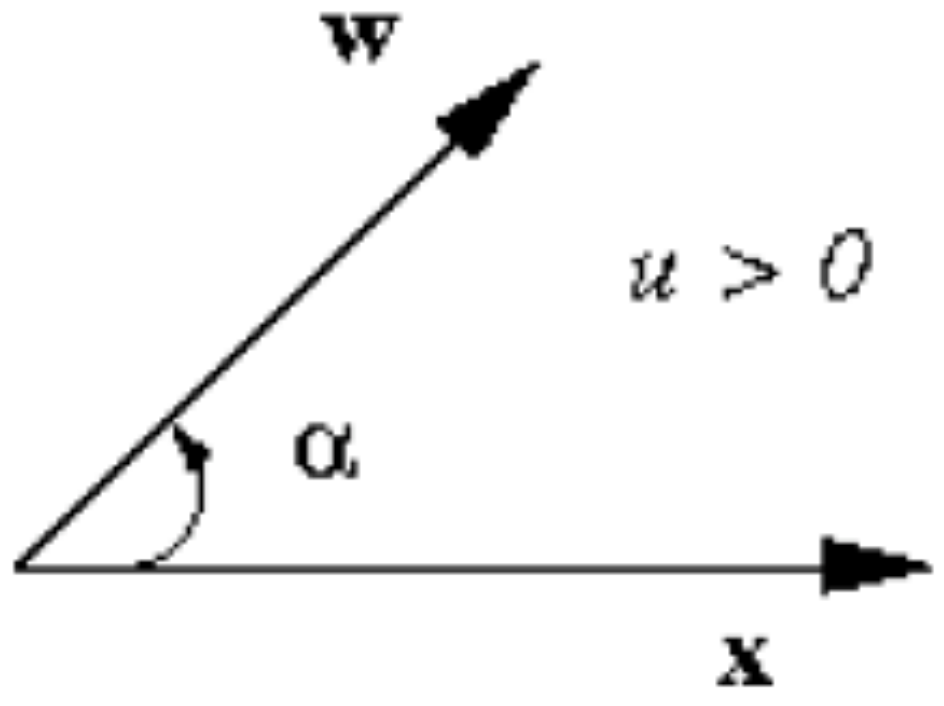
Representação gráfica do CASO 1

Caso 1: $e = d - y = +1 \quad (d{=}+1 \text{ e } y{=}0)$

Situação ocorrida (u<0, y=0):



Situação desejada (u>0, y=1):



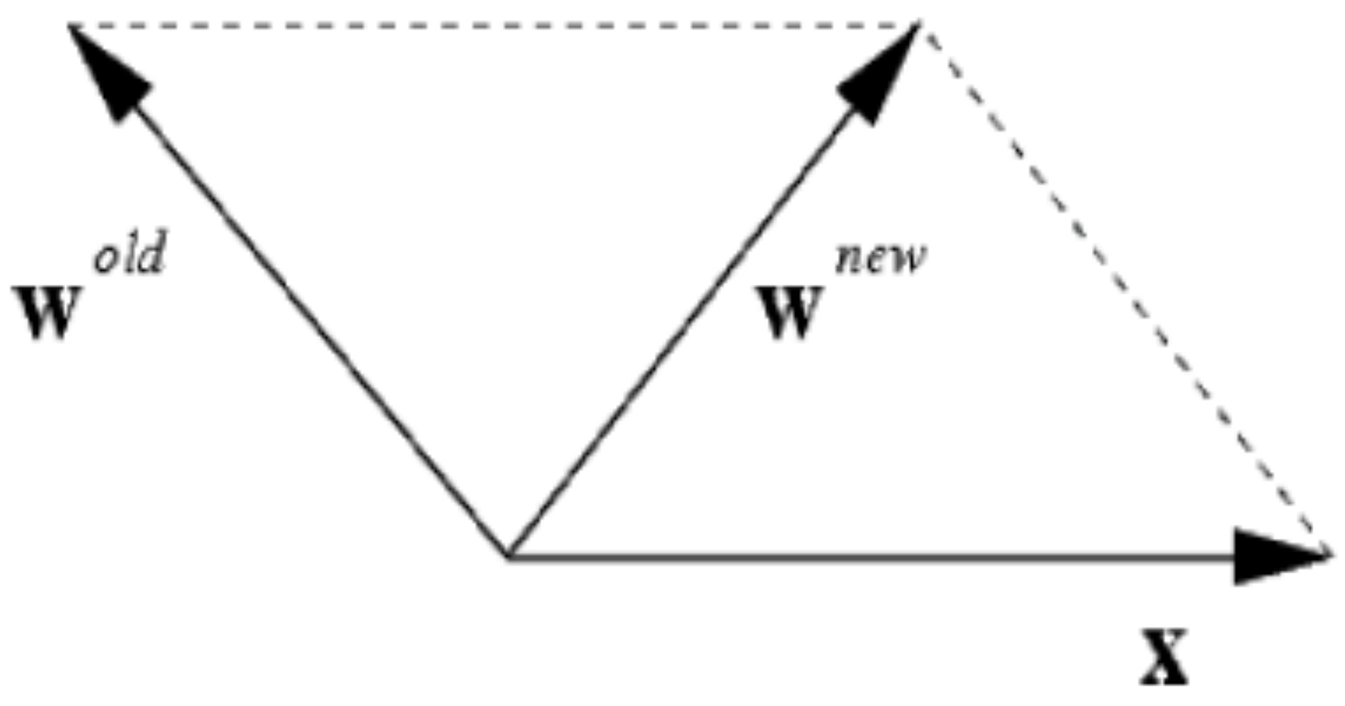
Lembre-se que:

$y = 1$, se $u \geqslant 0$.

$y = 0$, se $u < 0$.

O vetor $\mathbf{w}$ deve ser ajustado para se aproximar de $\mathbf{x}$. Assim, neste caso, a regra de atualização é descrita como:

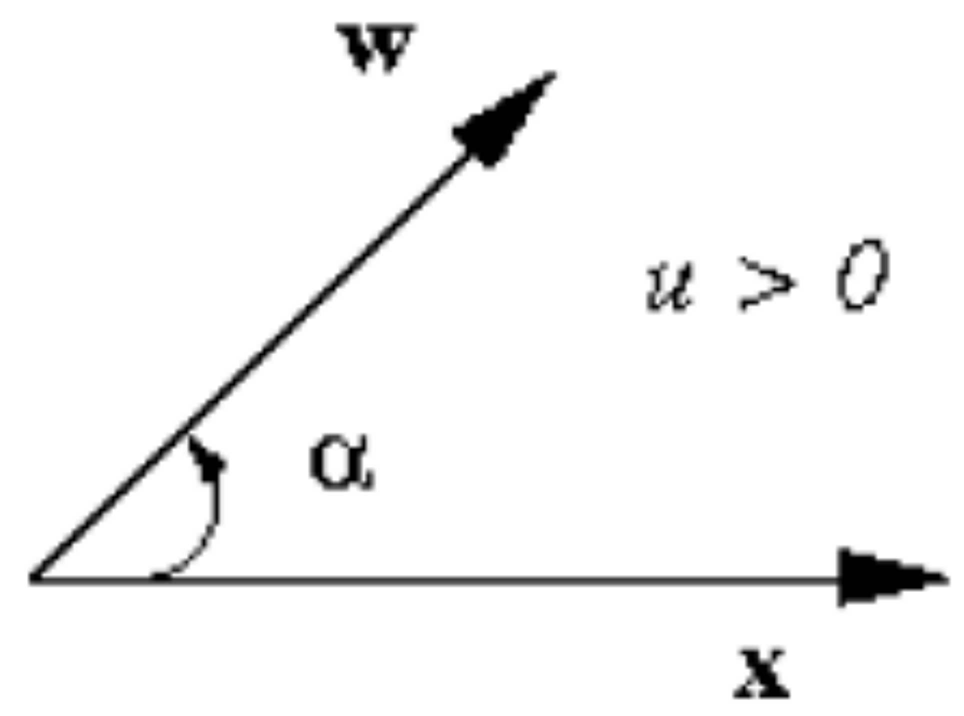
$$\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} + \mathbf{x}\,.$$



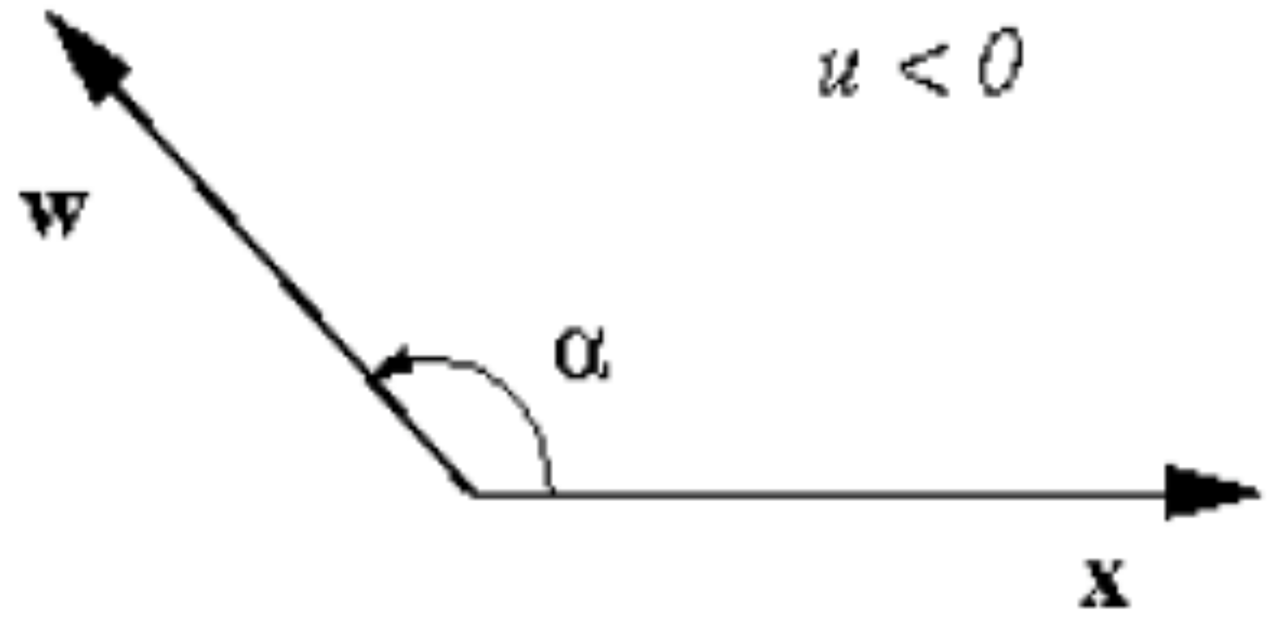
Representação gráfica do CASO 2

Caso 2: $e = d - y = -1$ (d=0 e y=+1)

Situação ocorrida (u>0, y=+1):



Situação desejada (u<0, y=0):



Lembre-se que:

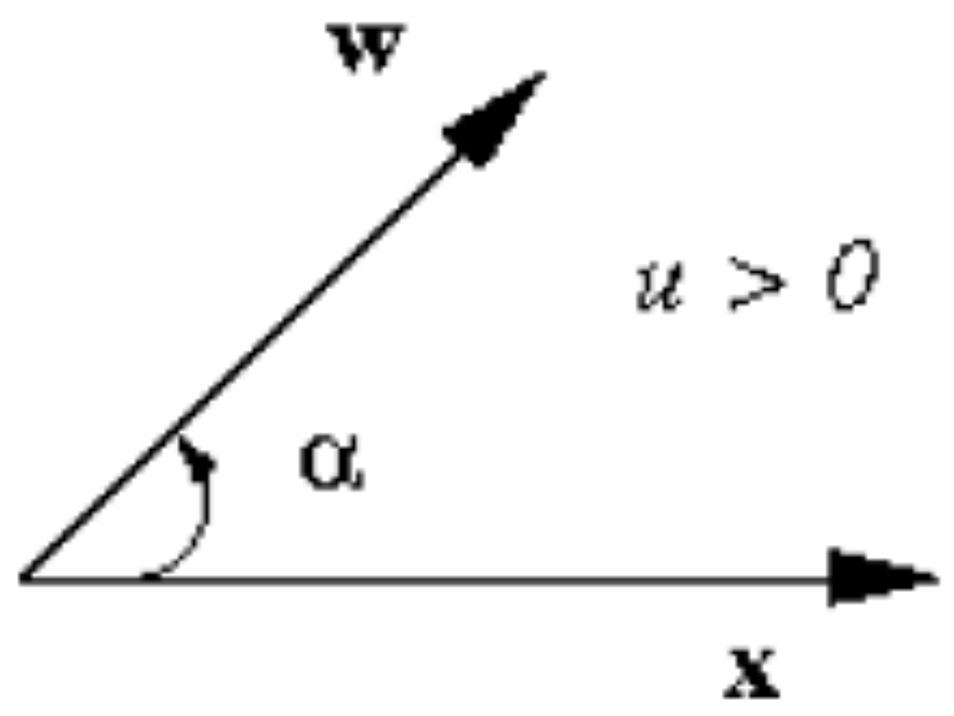
$y = 1$, se $u \geqslant 0$.

$y = 0$, se $u < 0$.

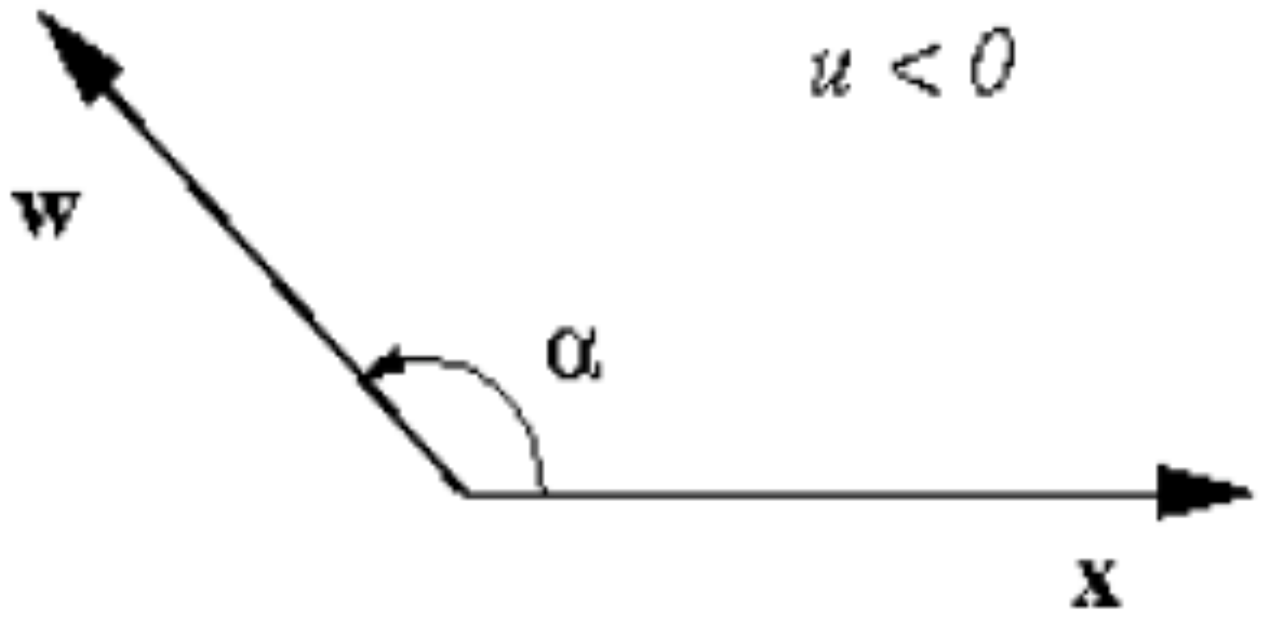
Representação gráfica do CASO 2

Caso 2: $e = d - y = -1$ (d=0 e y=+1)

Situação ocorrida (u>0, y=+1):



Situação desejada (u<0, y=0):



Lembre-se que:

$y = 1$, se $u \geqslant 0$.

$y = 0$, se $u < 0$.

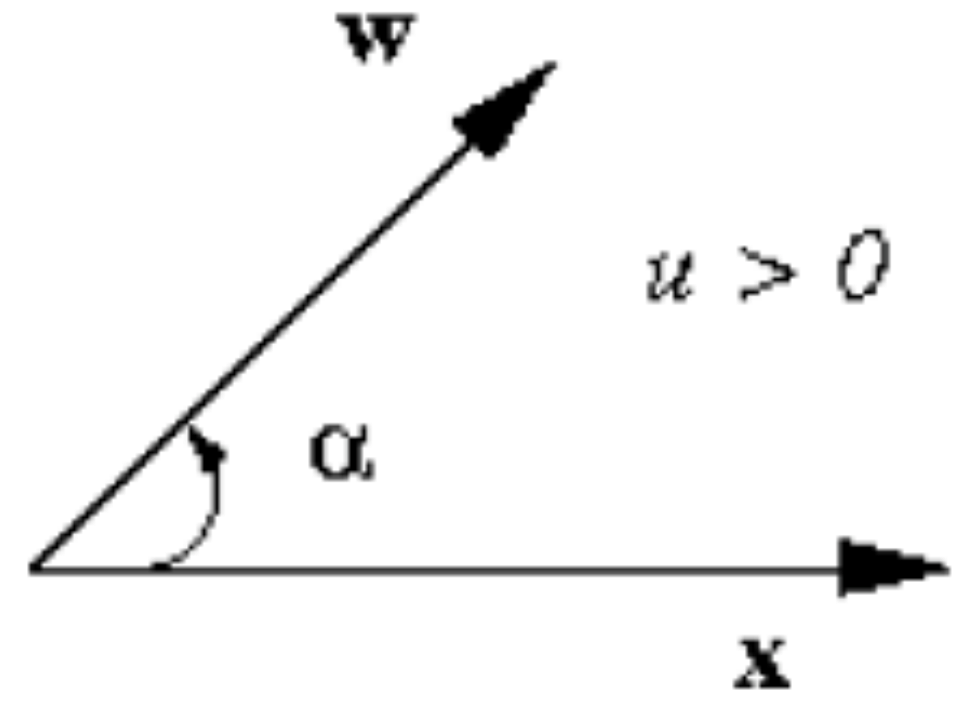
O vetor $\mathbf{w}$ deve ser ajustado para se afastar de $\mathbf{x}$. Assim, neste caso, a regra de atualização é descrita como:

$$\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} - \mathbf{x}\,.$$

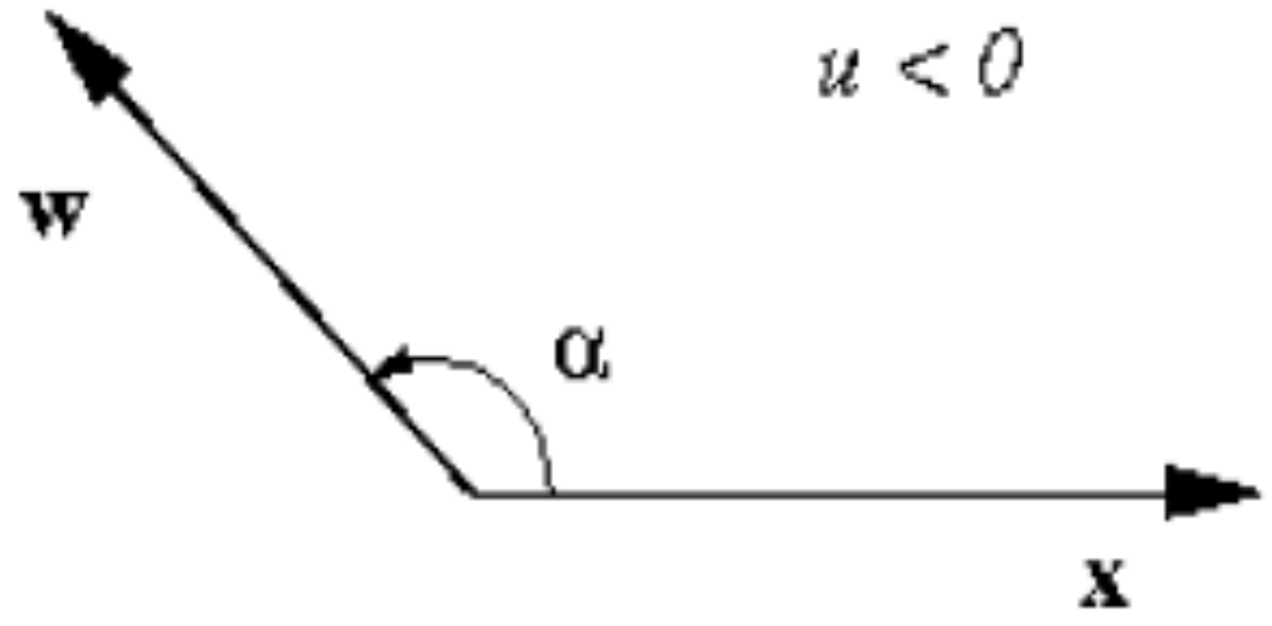
Representação gráfica do CASO 2

Caso 2: $e = d - y = -1$ (d=0 e y=+1)

Situação ocorrida (u>0, y=+1):



Situação desejada (u<0, y=0):



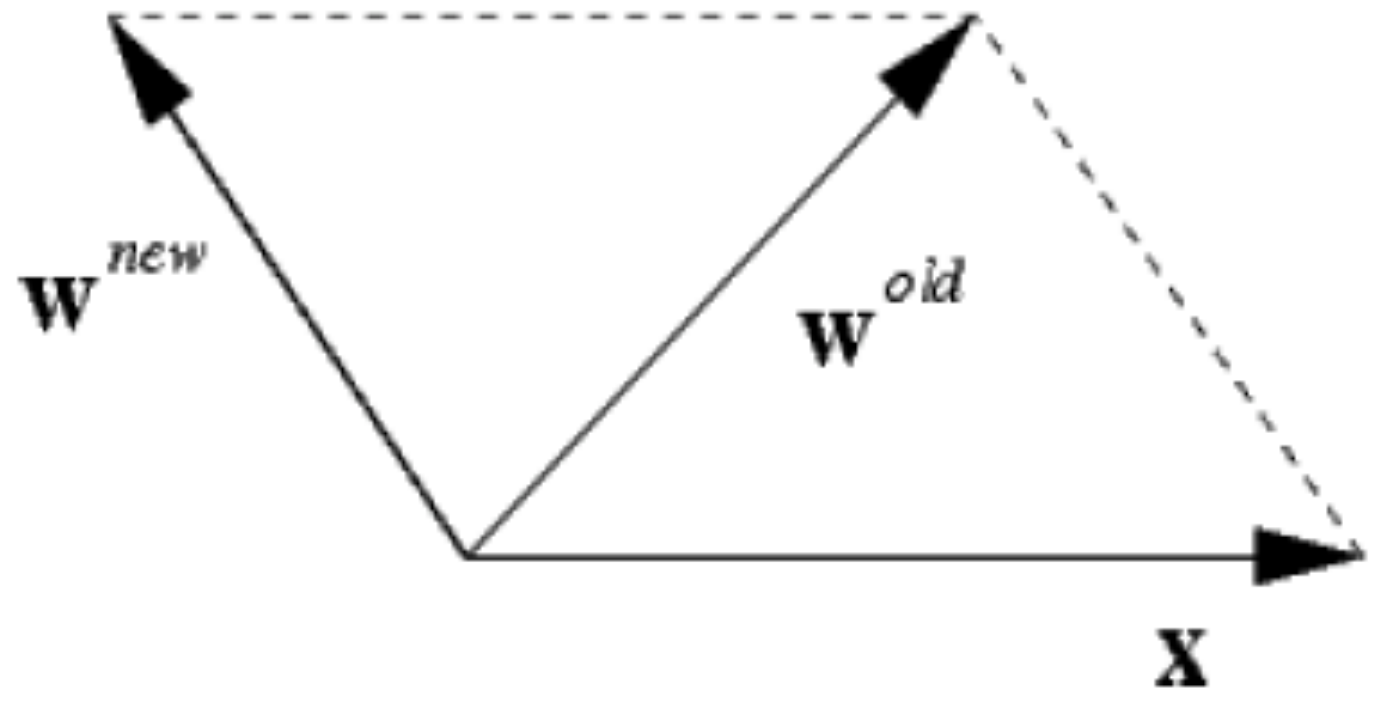
Lembre-se que:

$y = 1$, se $u \geqslant 0$.

$y = 0$, se $u < 0$.

O vetor $\mathbf{w}$ deve ser ajustado para se afastar de $\mathbf{x}$. Assim, neste caso, a regra de atualização é descrita como:

$$\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} - \mathbf{x}\,.$$



E o CASO 3?

Regra de aprendizagem

Consideremos os argumentos geométricos para esta definição. Assim, 03 casos são derivados para a variável de erro $e^{(t)}$:

- CASO 1: $e^{(t)} = d^{(t)} - y^{(t)} = +1,\ \mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} + \mathbf{x}$;
- CASO 2: $e^{(t)} = d^{(t)} - y^{(t)} = -1,\ \mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} - \mathbf{x}$;
- CASO 3: $e^{(t)} = d^{(t)} - y^{(t)} = 0$, não faz nada.

Regra de aprendizagem

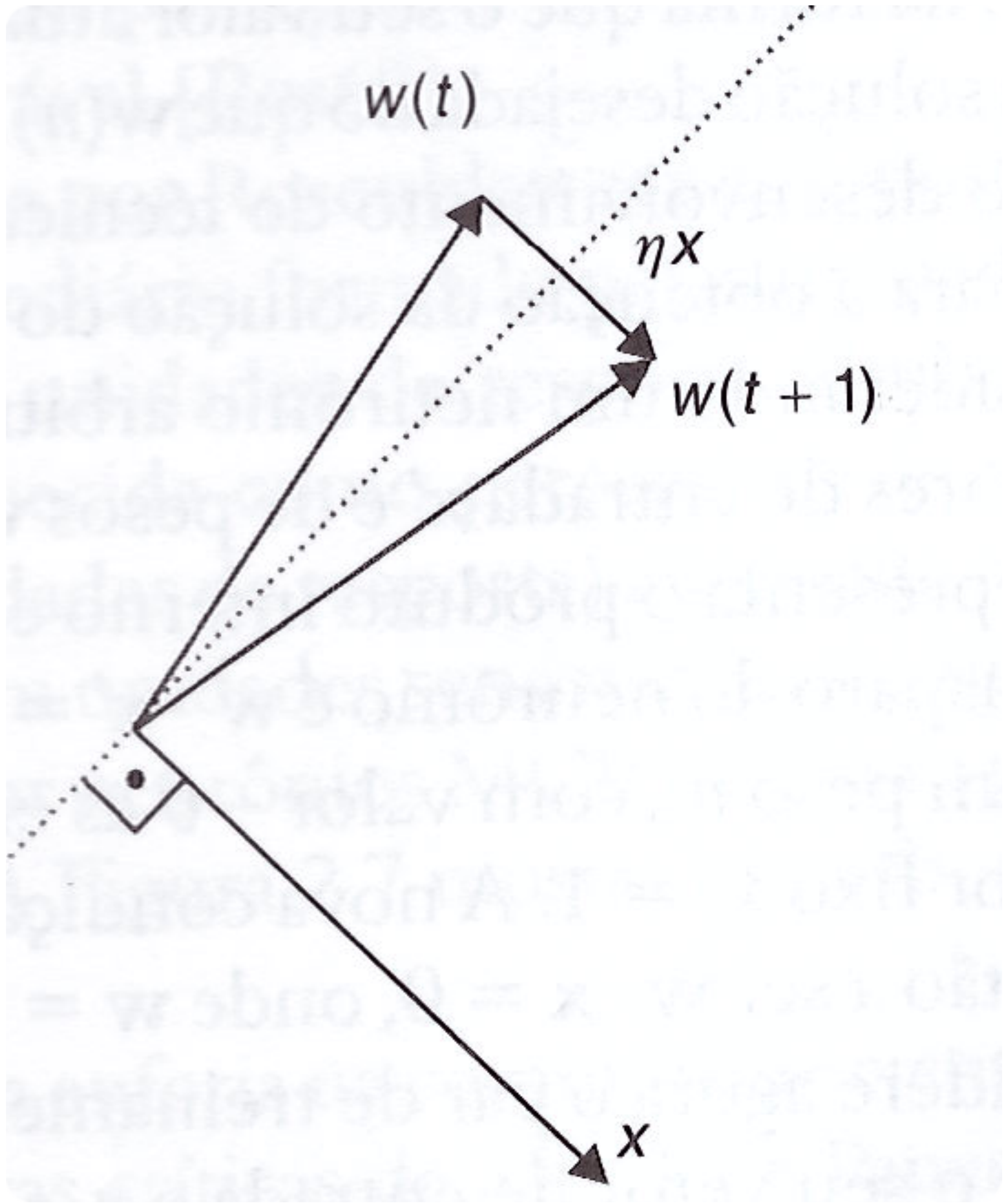
Consideremos os argumentos geométricos para esta definição. Assim, 03 casos são derivados para a variável de erro $e^{(t)}$:

- CASO 1: $e^{(t)} = d^{(t)} - y^{(t)} = +1$, $\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} + \mathbf{x}$;
- CASO 2: $e^{(t)} = d^{(t)} - y^{(t)} = -1$, $\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} - \mathbf{x}$;

$$\left\{ \mathbf{w}^{(t)} + e^{(t)}\mathbf{x}^{(t)} \right.$$

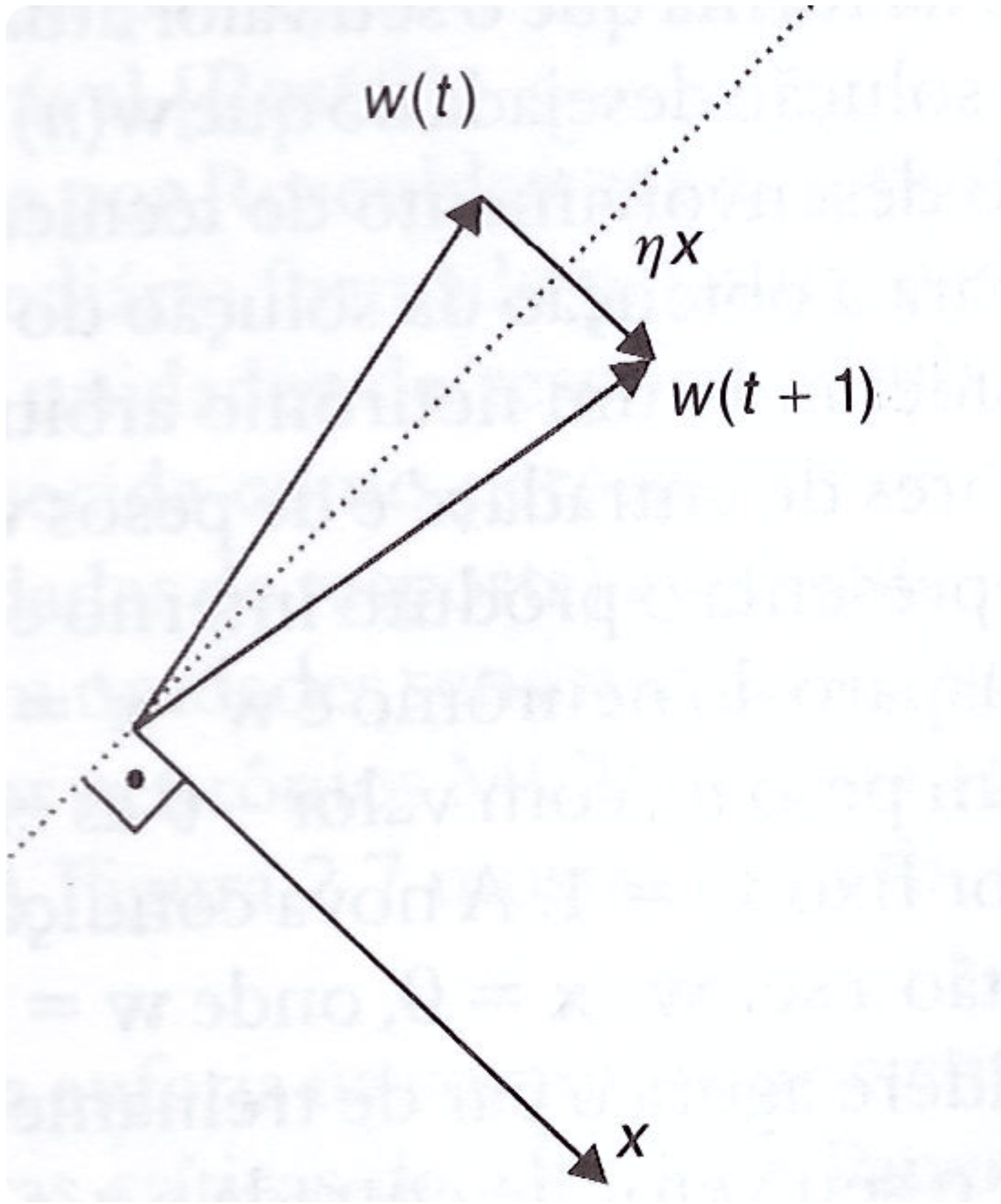
- CASO 3: $e^{(t)} = d^{(t)} - y^{(t)} = 0$, não faz nada.

Regra de aprendizagem



Em geral utiliza-se um valor para η, denominado fator ou taxa de aprendizagem, pequeno $(0 < \eta \leq 1)$. Assim, que como $\eta \mathbf{x}^{(t)}$ representa o vetor a ser somado para aproximação ou afastamento de $\mathbf{w}^{(t)}$.

Regra de aprendizagem



Em geral utiliza-se um valor para η, denominado fator ou taxa de aprendizagem, pequeno $(0 < \eta \leq 1)$. Assim, que como $\eta\mathbf{x}^{(t)}$ representa o vetor a ser somado para aproximação ou afastamento de $\mathbf{w}^{(t)}$.

REGRA GERAL DE APRENDIZAGEM PERCEPTRON SIMPLES

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

Como poderia se determinar cada um dos valores para os pesos sinápticos?

Como poderia se determinar cada um dos valores para os pesos sinápticos?

REGRA DE APRENDIZADO

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

Resumo do algoritmo de treinamento do Perceptron

1. INICIO $(t = 0)$
   1.1. Definir o valor de $\eta,\ 0 < \eta \leq 1$;
   1.2. Iniciar $\mathbf{w}^{(0)}$ com valores aleatórios;
2. FUNCIONAMENTO
   2.1. Selecionar um valor de entrada $\mathbf{x}^{(t)}$.
   2.2. Calcular a saída do neurônio $y^{(t)}$.
3. TREINAMENTO
   3.1. Calcular o erro: $e^{(t)} = d^{(t)} - y^{(t)}$.
   3.2. Ajustar o peso via REGRA DE APRENDIZAGEM.
   3.3. Verificar o critério de parada.
      3.3.1. Se o critério for atendido, finalizar treinamento.
      3.3.2. Caso contrário, fazer $t = t + 1$ e voltar ao PASSO 2. FUNCIONAMENTO.

Referências

- Jardel Rodrigues. **Aula 03**: Perceptron Simples. IFCE *campus* Jaguaribe, 2021.
- PÁDUA BRAGA, Antônio; DE LEON FERREIRA, André Carlos Ponce; LUDERMIR, Teresa Bernarda. **Redes neurais artificiais**: teoria e aplicações. LTC editora, 2007.
- Richard O. Duda, Peter E. Hart, David G. Stork. **Pattern Classification**. John Wiley & Sons, 2012.