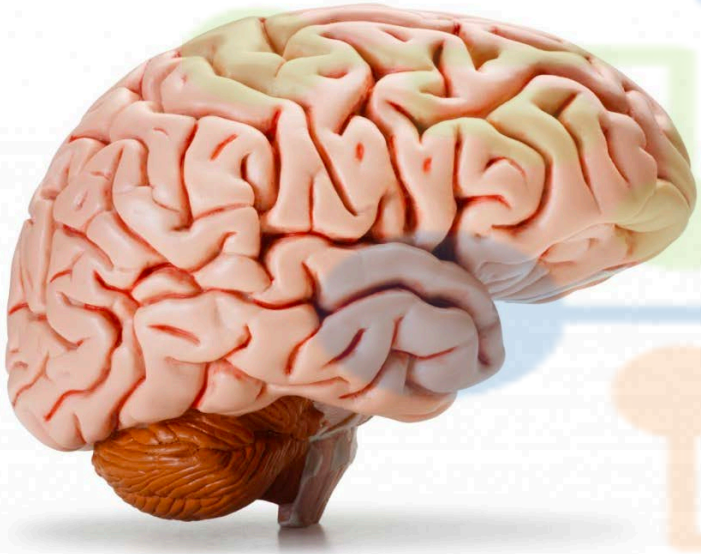
Data Science
Academy

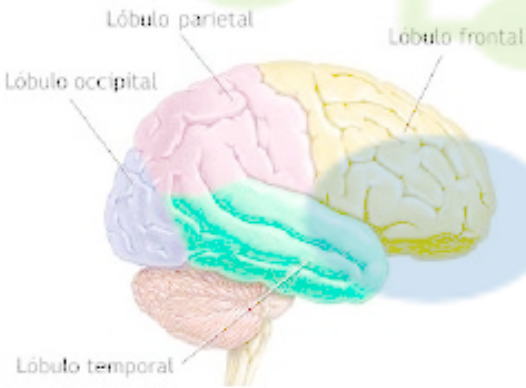
Seja muito bem-vindo(a)!

Data Science Academy

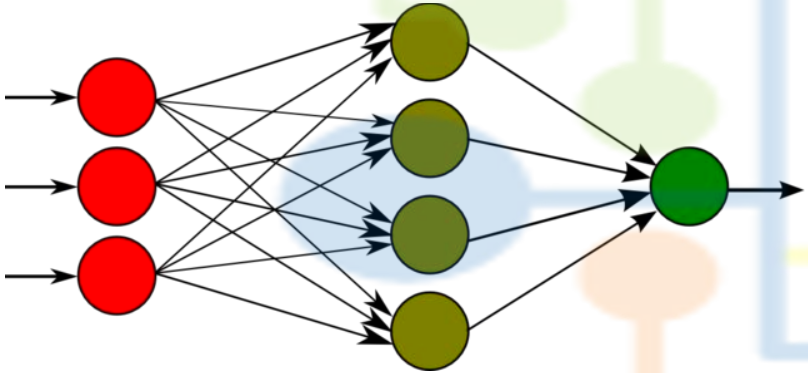
Redes Neurais



Este é o dispositivo mais incrível da história humana e que os cientistas estão tentando reproduzir em computadores!



O **cérebro humano** tem sido extensamente estudado, mas ainda não somos capazes de **entender completamente** o seu funcionamento.



Redes Neurais Artificiais podem ser consideradas um paradigma diferente de computação.

Redes Neurais Artificiais consistem em um modo de abordar a solução de problemas de **Inteligência Artificial**

Assim como um cérebro usa uma rede de células interconectadas chamadas **neurônios** para criar um processador paralelo maciço, a rede neural usa uma **rede de neurônios artificiais** para resolver problemas de aprendizagem

- Cérebro humano – 85 bilhões de neurônios
- Cérebro de um gato – 1 bilhão de neurônios
- Cérebro de um rato – 75 milhões de neurônios
- Cérebro de uma barata – 1 milhão de neurônios

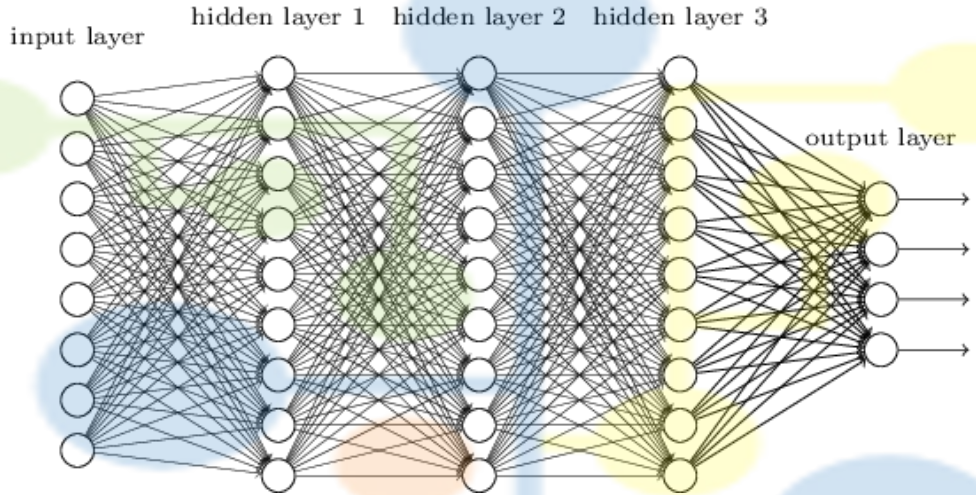
Agora você entende porque a computação paralela em GPU's está acelerando o desenvolvimento de sistemas inteligentes, pois somos capazes de processar cada vez mais dados em redes neurais artificiais com cada vez mais neurônios

Programas de reconhecimento de voz e escrita

Automação de dispositivos inteligentes

Modelos sofisticados de padrões climáticos

As Redes Neurais Artificiais são modelos versáteis que podem ser aplicadas a quase todas as tarefas de aprendizagem: classificação, previsão numérica e mesmo reconhecimento não supervisionado de padrões



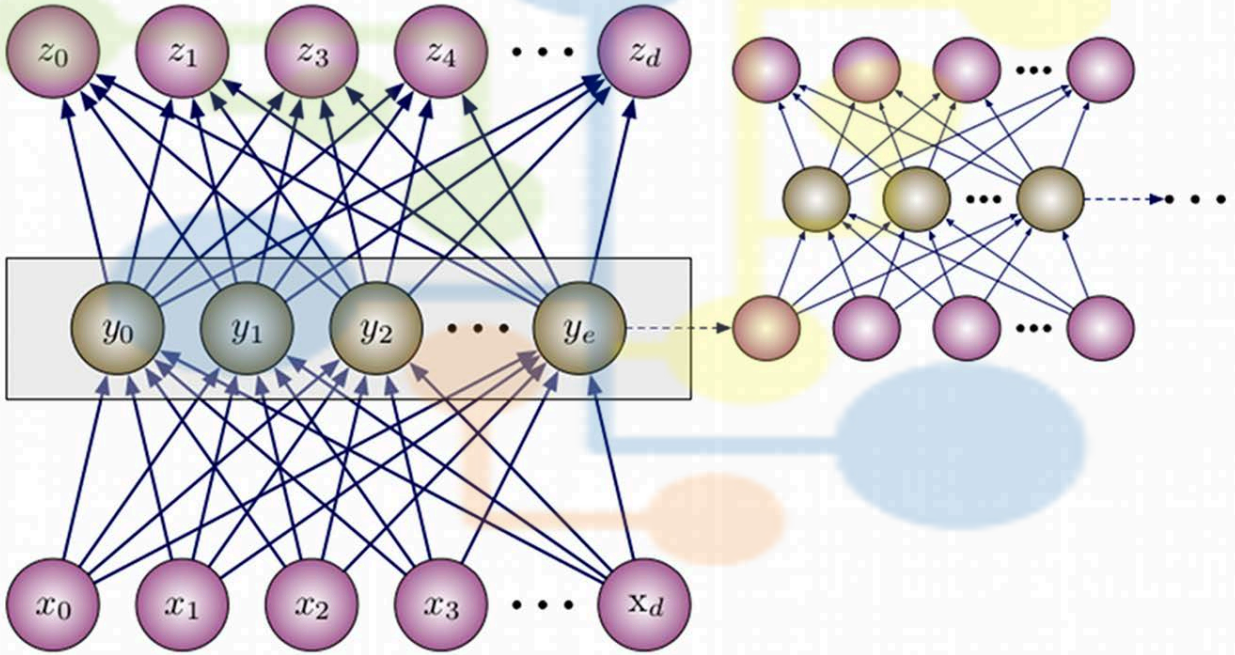
As redes neurais artificiais são melhor aplicadas a problemas onde os dados de entrada e os dados de saída são bem definidos ou, pelo menos, bastante simples, mas o processo que relaciona a entrada com a saída é extremamente complexo

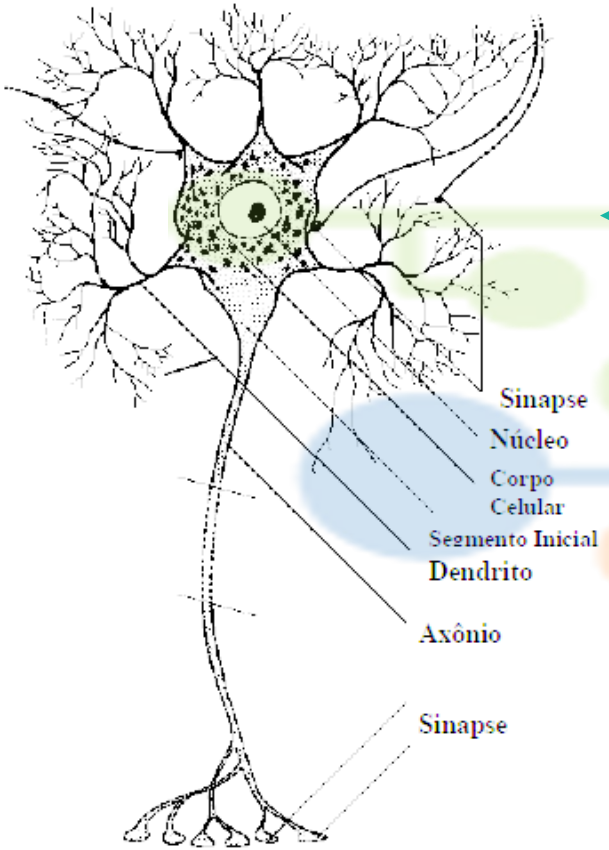
Data Science Academy

O que são Redes Neurais Artificiais?

Warren McCulloch

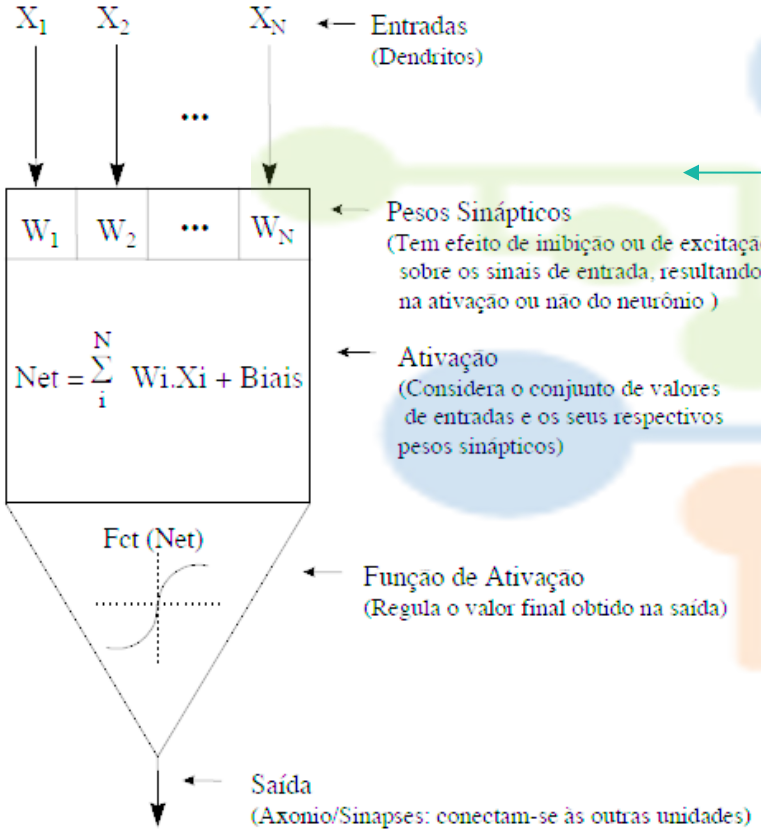
Walter Pitts


z_0 z_1 z_3 z_4 z_d
y_0 y_1 y_2 y_e
x_0 x_1 x_2 x_3 x_d



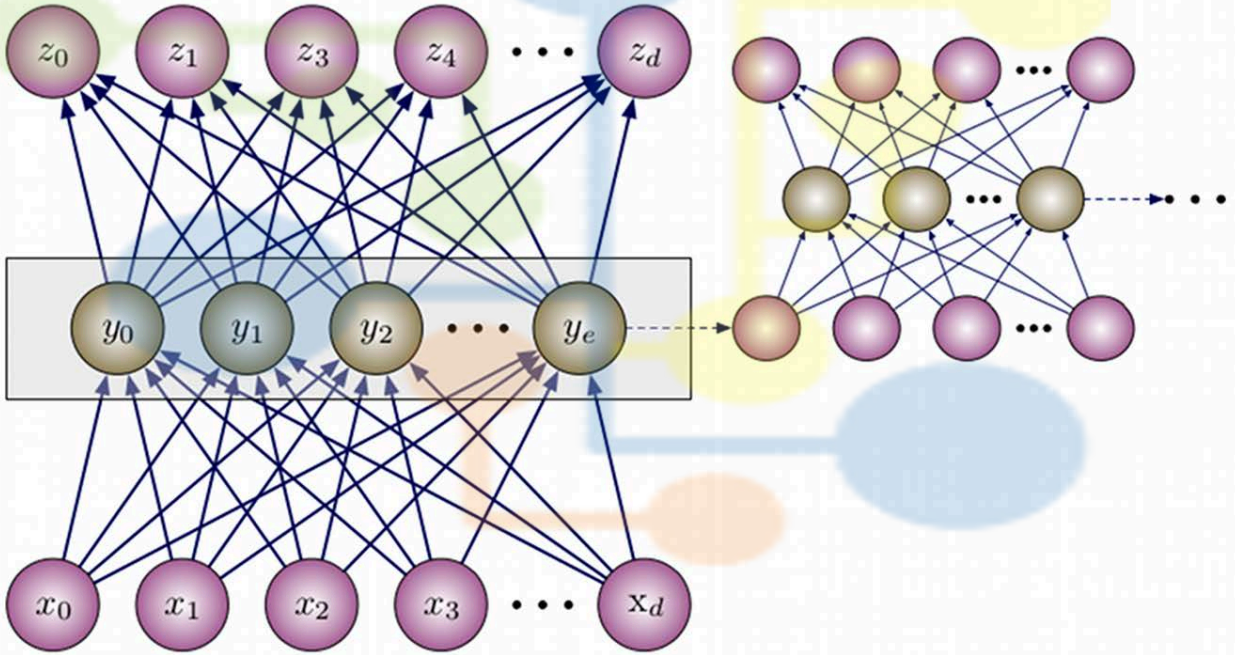
Neurônio Biológico

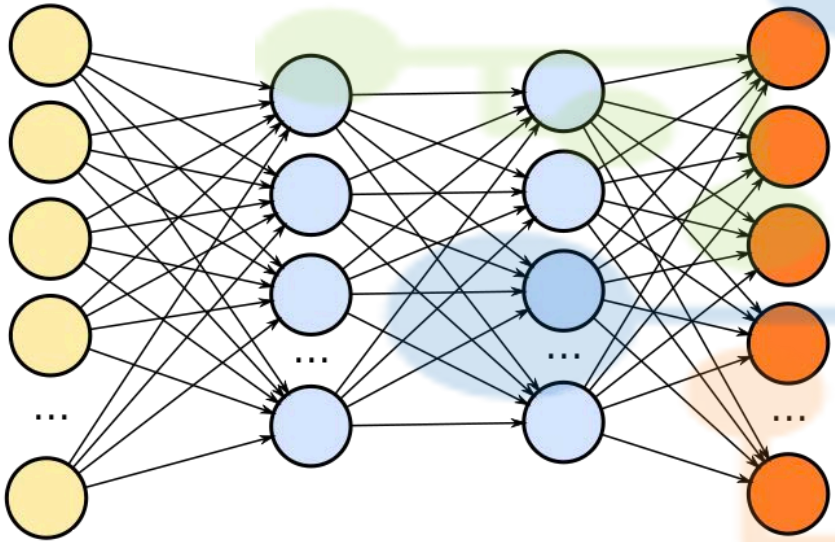
O conhecimento de uma Rede Neural Artificial (RNA) está codificado na estrutura da rede, onde se destacam as conexões (sinapses) entre as unidades (neurônios) que a compõe



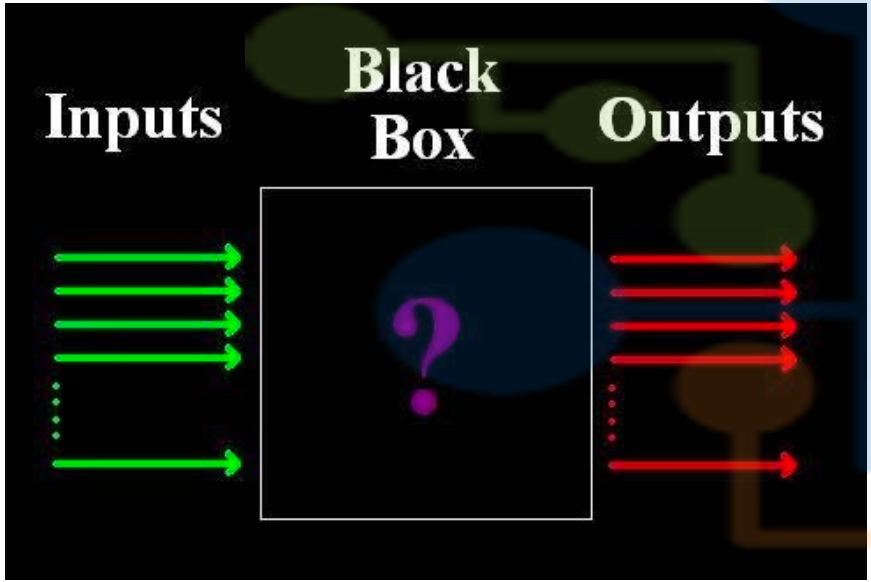
Neurônio Matemático

O conhecimento de uma Rede Neural Artificial (RNA) está codificado na estrutura da rede, onde se destacam as conexões (sinapses) entre as unidades (neurônios) que a compõe

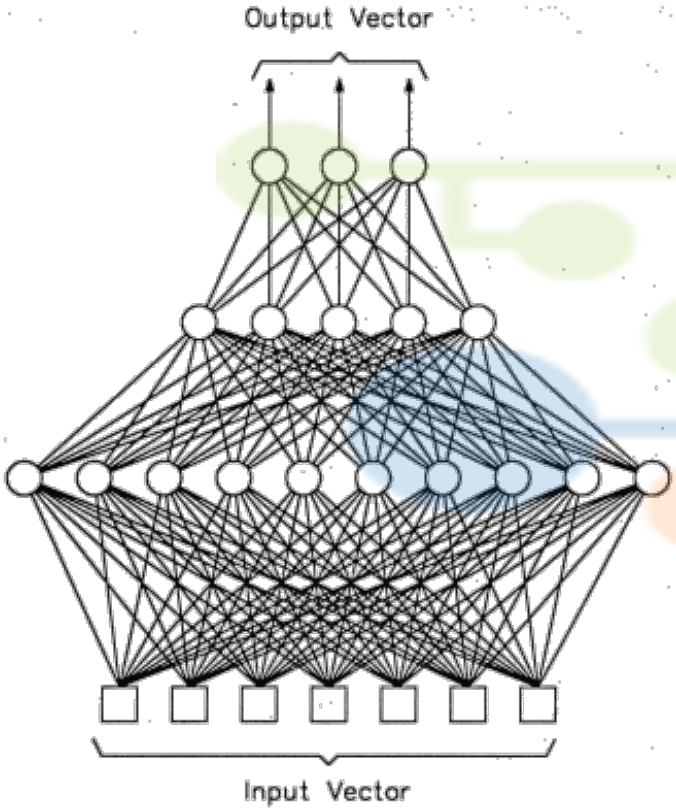

z_0 z_1 z_3 z_4 z_d
y_0 y_1 y_2 y_e
x_0 x_1 x_2 x_3 x_d



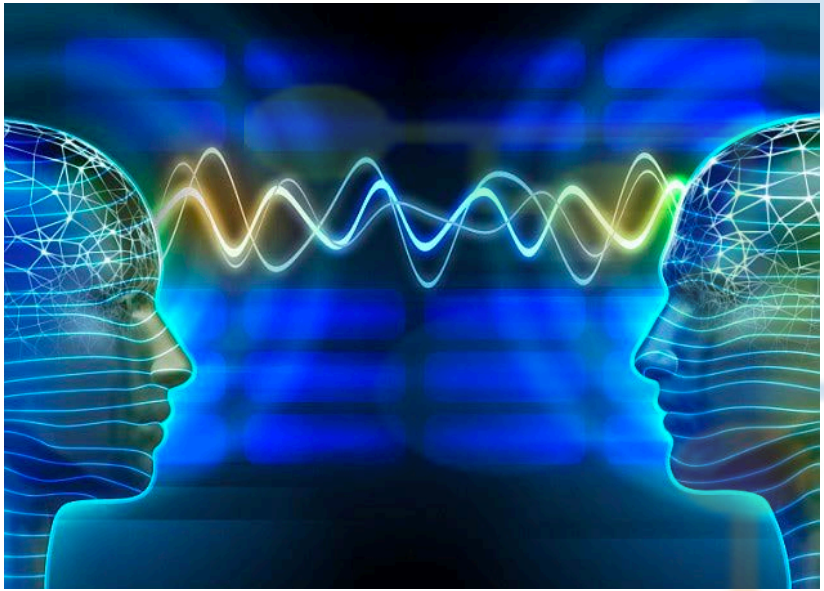
Um dos benefícios das redes diz respeito ao tratamento de um problema clássico da Inteligência Artificial, que é a representação de um universo não-estacionário (onde as estatísticas mudam com o tempo)



Uma desvantagem das redes neurais é o fato delas , normalmente, serem uma ”caixa preta”

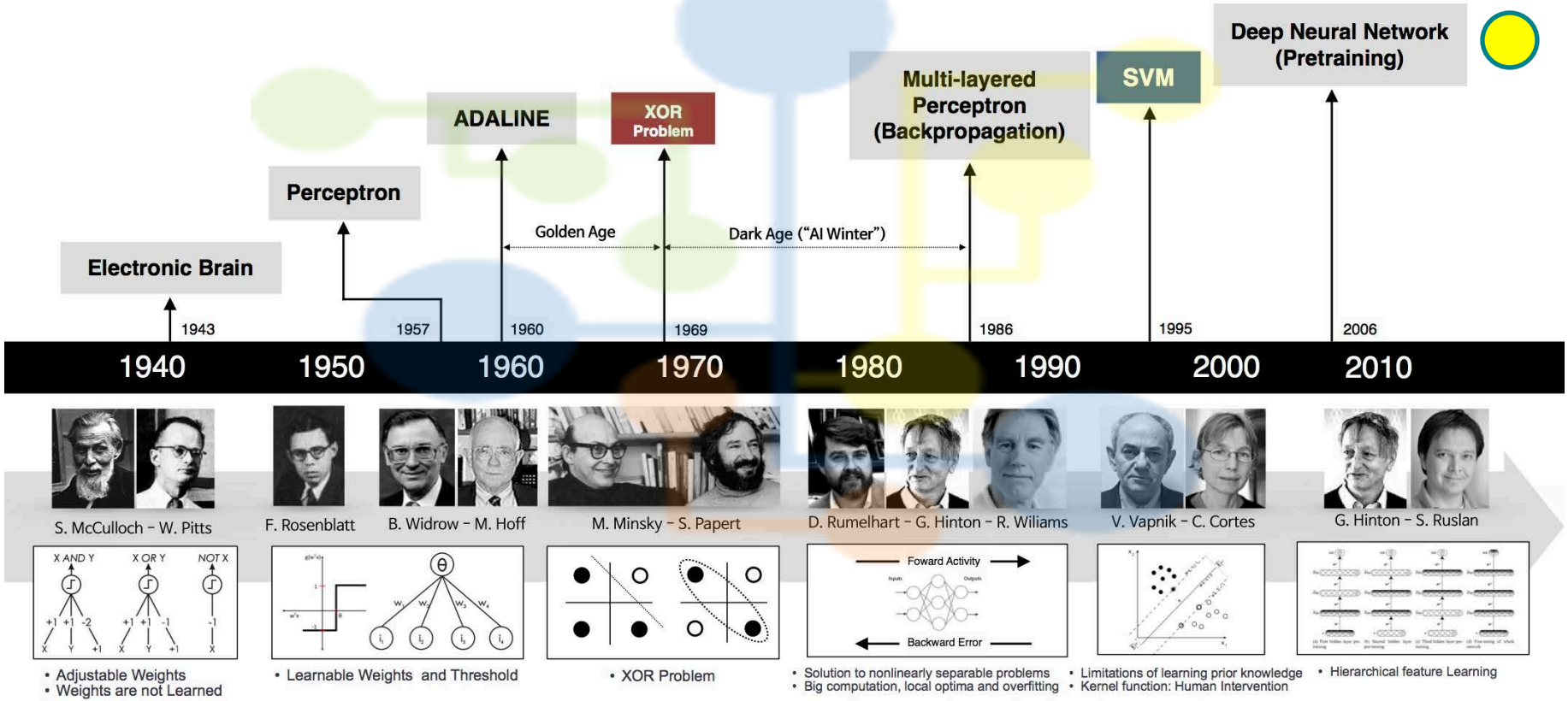


A solução de problemas através das RNAs é bastante atrativa, pois o paralelismo constitui-se na característica principal das RNAs, onde esta cria a possibilidade de um desempenho superior em relação a solução de problemas baseados nos modelos convencionais.



A generalização está associada à capacidade da rede em aprender através de um conjunto reduzido de exemplos, e posteriormente, dar respostas coerentes a dados não apresentados a rede.

Redes Neurais Artificiais
Origem e Evolução


Deep Neural Network (Pretraining)
SVM
Multi-layered Perceptron (Backpropagation)
XOR Problem
ADALINE
Perceptron
Electronic Brain
Golden Age
Dark Age ("AI Winter")
1943
1957
1960
1969
1986
1995
2006
1940
1950
1960
1970
1980
1990
2000
2010
S. McCulloch - W. Pitts
F. Rosenblatt
B. Widrow - M. Hoff
M. Minsky - S. Papert
D. Rumelhart - G. Hinton - R. Wiliams
V. Vapnik - C. Cortes
G. Hinton - S. Ruslan
X AND Y
X OR Y
NOT X
Foward Activity
Backward Error
• Adjustable Weights
• Weights are not Learned
• Learnable Weights and Threshold
• XOR Problem
• Solution to nonlinearly separable problems
• Big computation, local optima and overfitting
• Limitations of learning prior knowledge
• Kernel function: Human Intervention
• Hierarchical feature Learning

Para compreender a lógica de funcionamento das redes neurais, alguns conceitos básicos referentes ao funcionamento do cérebro humano e seus componentes, os neurônios, são de fundamental importância

Data Science Academy

O Neurônio Biológico

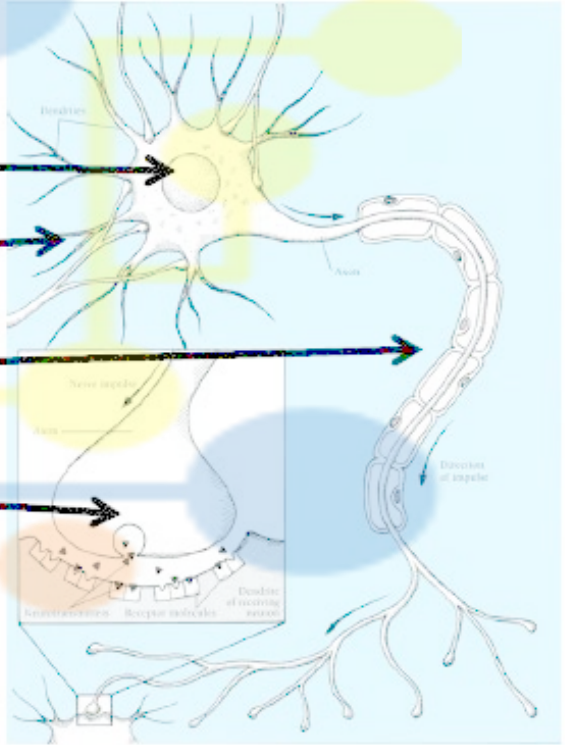
Estrutura de um Neuronio:

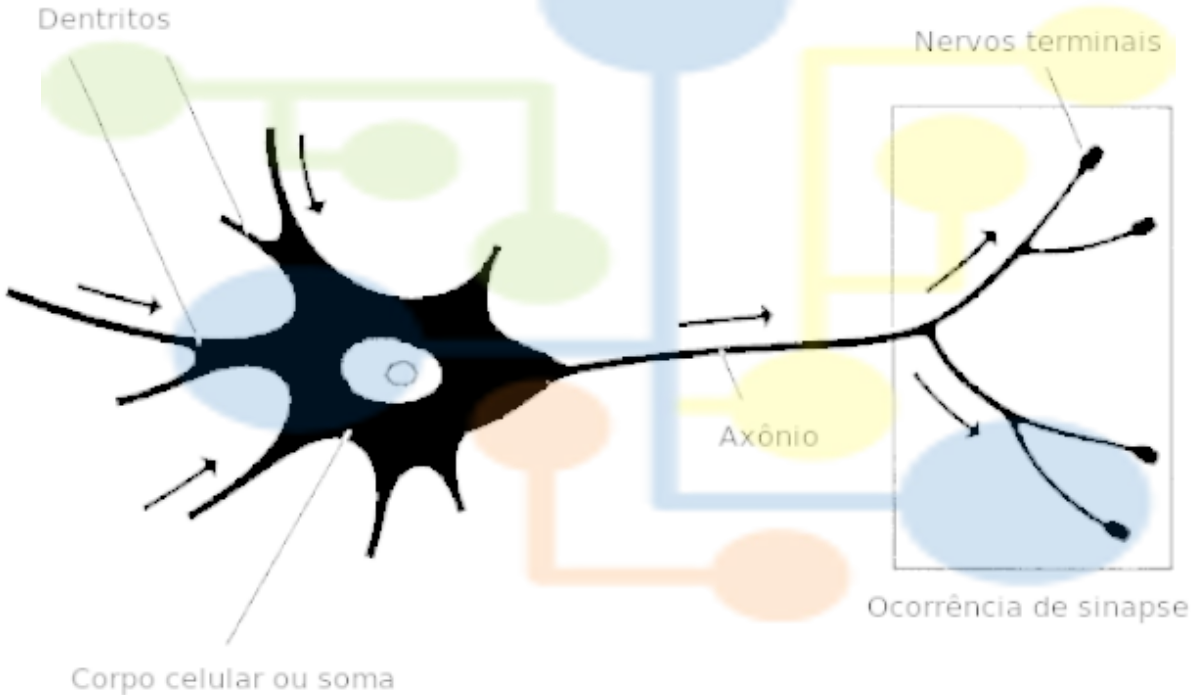
Corpo celular

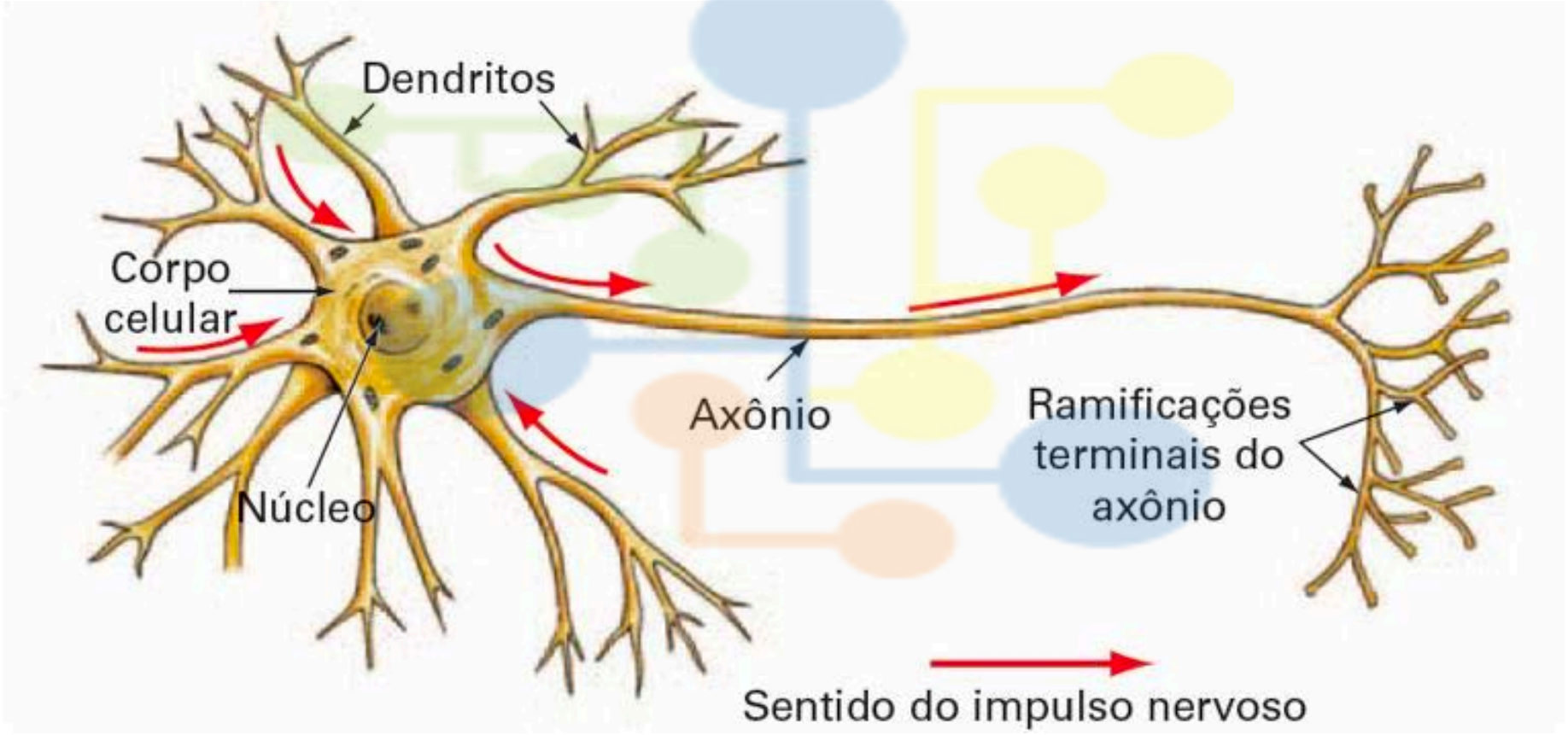
Dendritos

Axônio

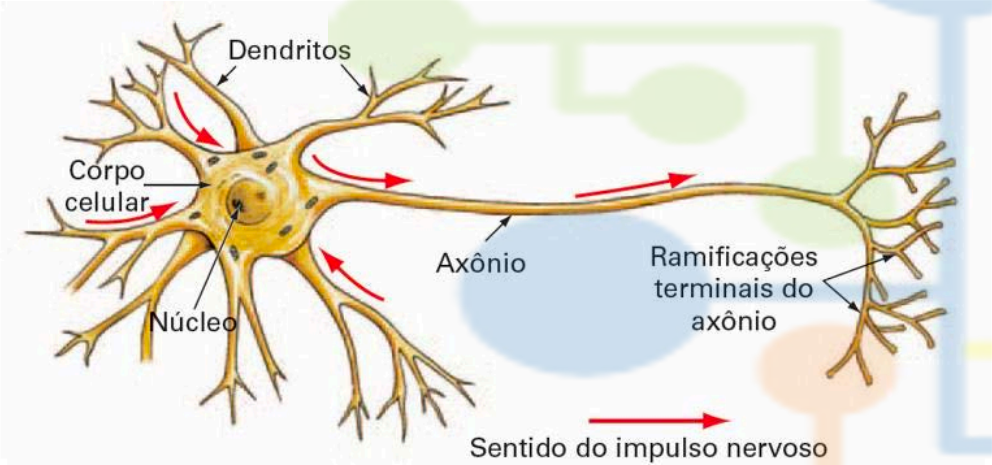
Terminais sinápticos




Dentritos
Nervos terminais
Axônio
Ocorrência de sinapse
Corpo celular ou soma

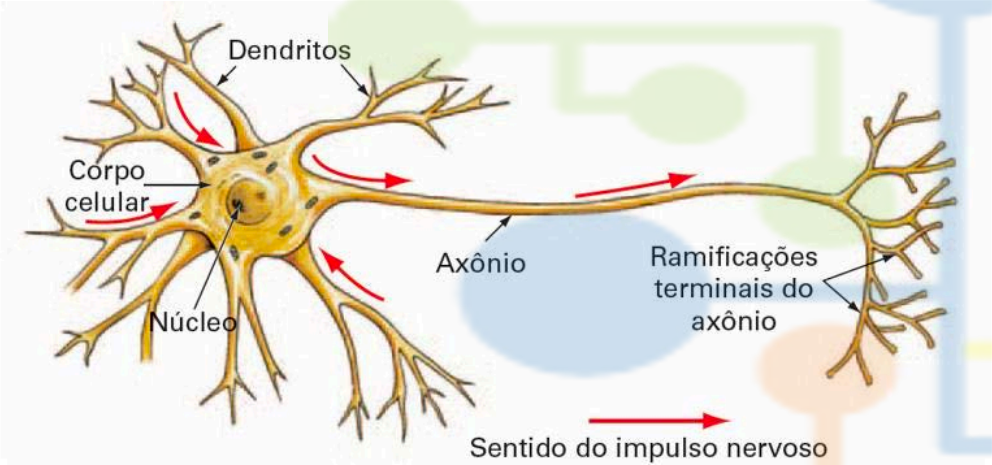

Dendritos
Corpo celular
Núcleo
Axônio
Ramificações terminais do axônio
Sentido do impulso nervoso

O ponto de contato entre a terminação axônica de um neurônio e o dentrito de outro é chamado ***sinapse***



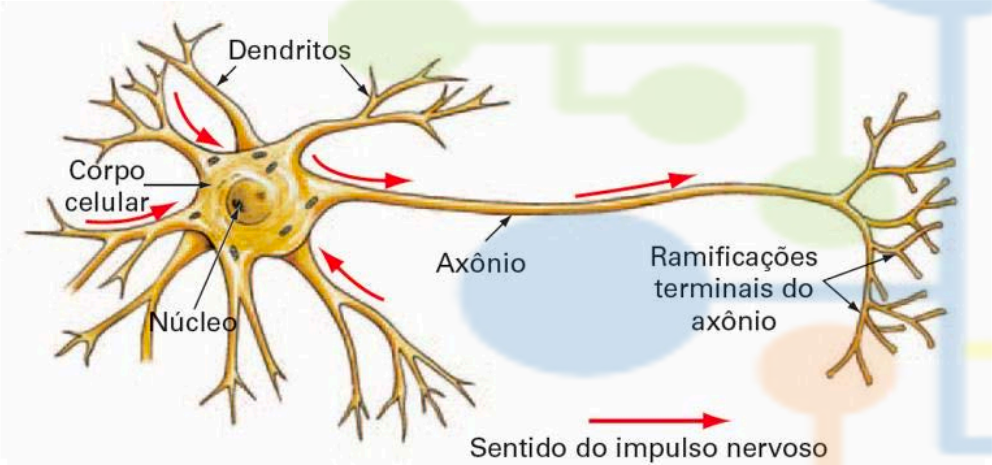
Se esses sinais forem superiores a aproximadamente 50mV (limiar do disparo), seguem pelo axônio. Caso contrário, são bloqueados e não preosseguem (são considerados irrelevantes).

Se o sinal for superior a certo limite (***threshold***), vai em frente; caso contrário é bloqueado e não segue.



Um neurônio recebe sinais através de inúmeros dendritos, os quais são ***ponderados*** e enviados para o axônio, podendo ou não seguir adiante (***threshold***)

Cada condutor, está associado um ***peso*** pelo qual o sinal é multiplicado.
A memória são os pesos.

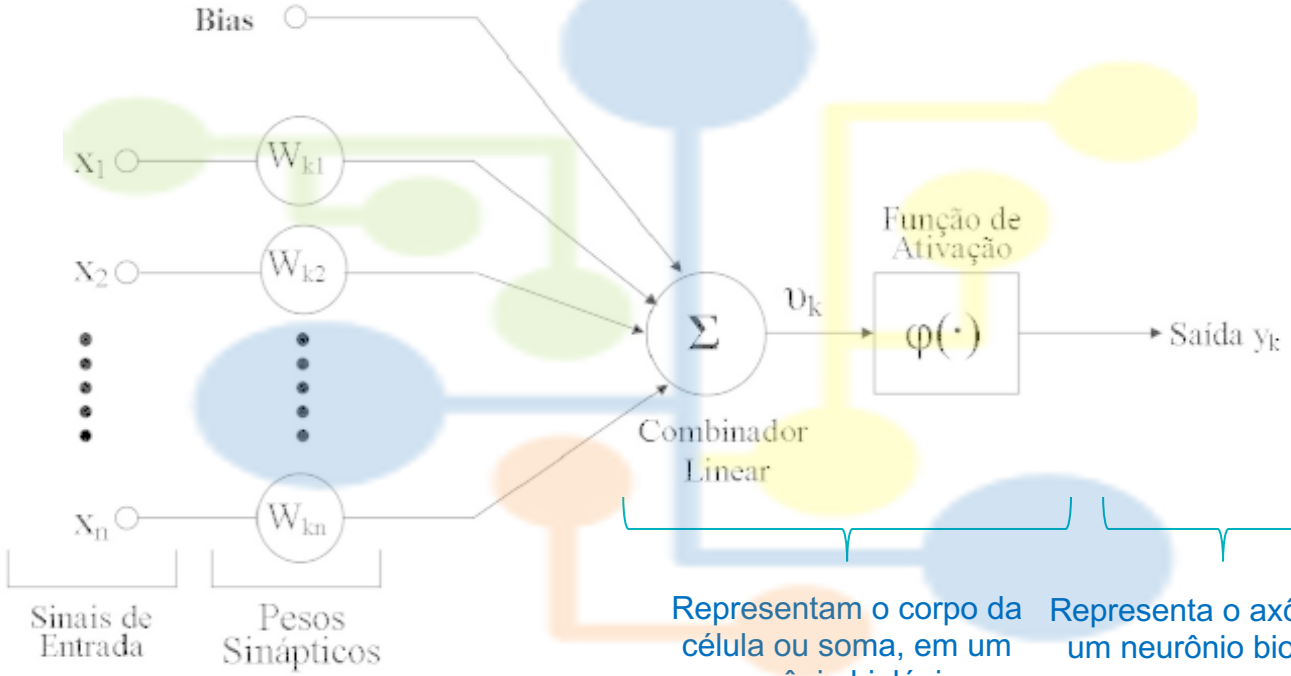


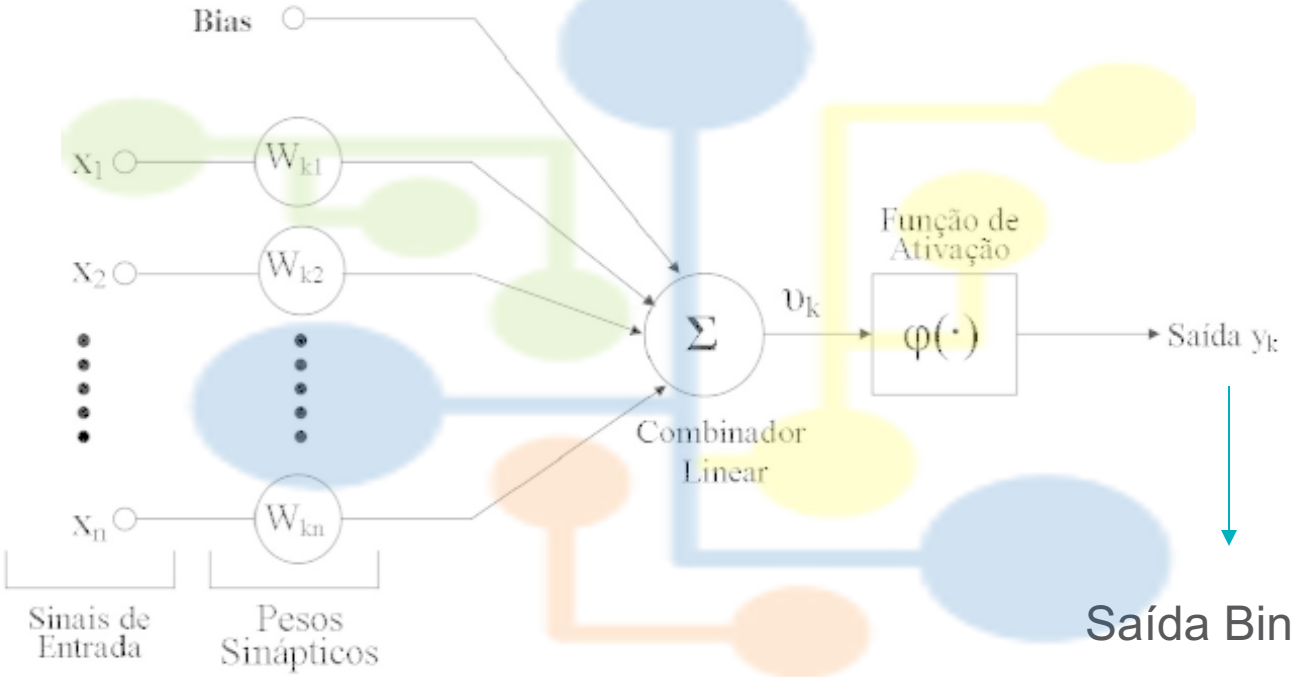
Cada região do cérebro possui uma arquitetura de rede diferente: varia o número de neurônios, de sinapses por neurônio, valor dos thresholds e dos pesos, etc…

Os valores dos pesos são estabelecidos por meio de treinamento recebido pelo cérebro durante a vida útil.
É a memorização.

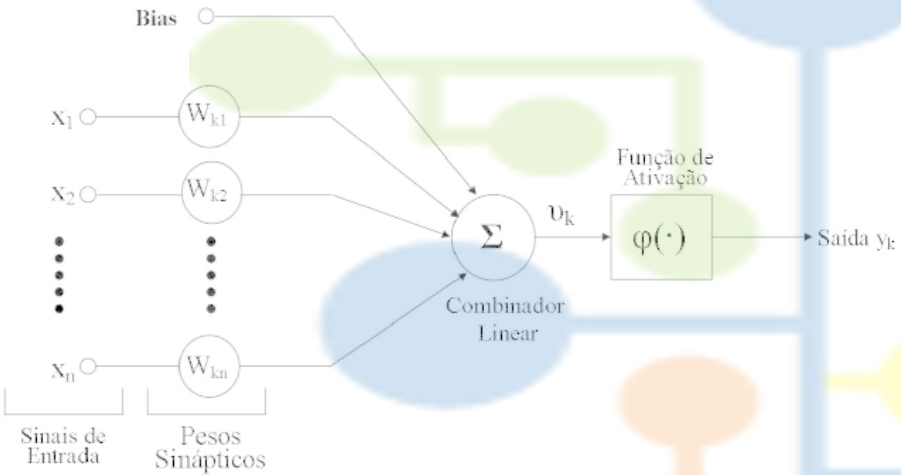
Data Science Academy

O Neurônio Matemático


Bias
x_1
x_2
x_n
W_{k1}
W_{k2}
W_{kn}
Sinais de Entrada
Pesos Sinápticos
Σ
υ_k
Função de Ativação
$\varphi(\cdot)$
Saída y_k
Combinador Linear
Representam o corpo da célula ou soma, em um neurônio biológico
Representa o axônio em um neurônio biológico



Saída Binária



$X_1W_1, X_2W_2, ...X_nW_n$

$$v = \sum_{i=0}^{n} w_i x_i$$

Um neurônio dispara quando a soma dos impulsos que ele recebe ultrapassa o seu limiar de excitação chamado de ***threshold***

Note que este modelo matemático simplificado de um neurônio é estático, ou seja, não considera a dinâmica do neurônio natural. No neurônio biológico, os sinais são enviados em pulsos e alguns componentes dos neurônios biológicos, a exemplo do axônio, funcionam como filtros de frequência.

$$u_k = \sum_{j=1}^{m} w_{kj} \cdot x_j$$

Fórmula do Neurônio Artificial

$$y_k = \varphi(u_k)$$

Fórmula da Função de Ativação

$$\varphi(u) = \begin{cases} 1 & , \text{se } u \geq 0 \\ 0 & , \text{se } u < 0 \end{cases}$$

Dentre as funções de ativação utilizadas, podemos destacar:

$$\varphi(v) = \frac{1}{1 + e^{-v}}$$

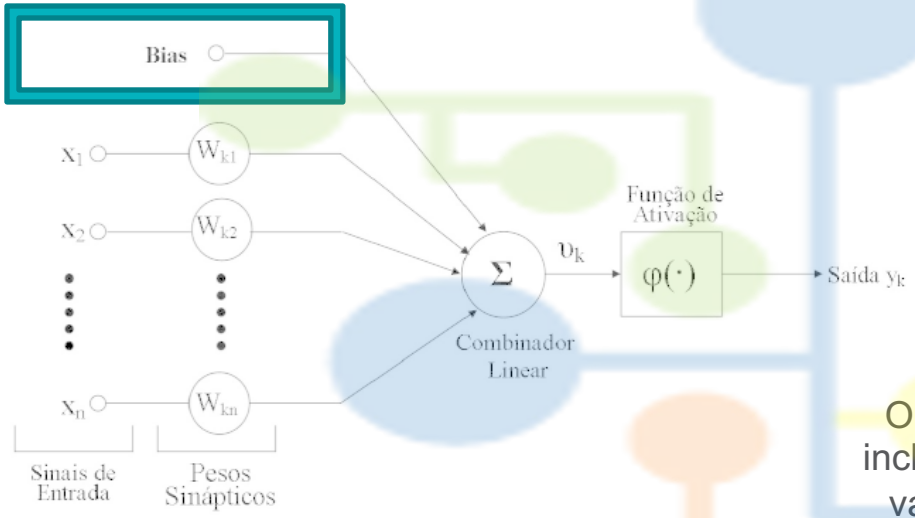
Função Sigmóide

$$\varphi(v) = e^{-v^2}$$

Função Gaussiana

$$\varphi(v) = \tanh(v)$$

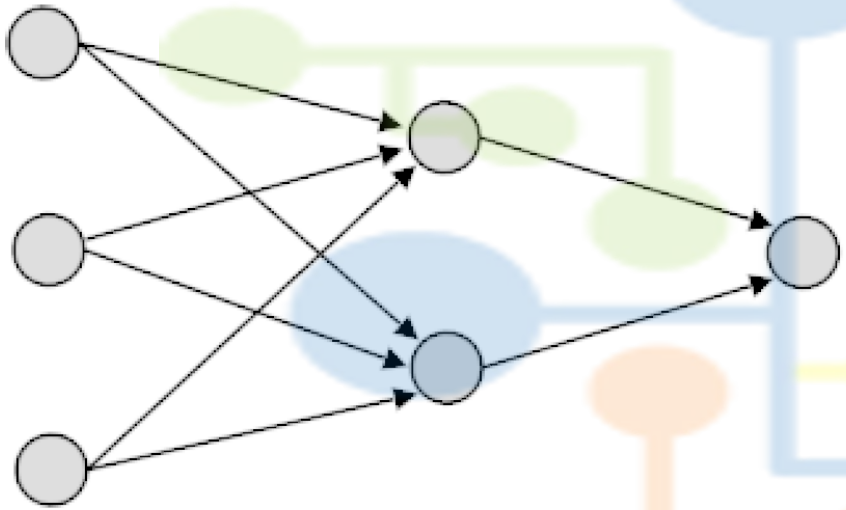
Função Tangente Hiperbólica



O modelo neuronal matemático também pode incluir uma polarização ou ***bias*** de entrada. Esta variável é incluída ao somatório da função de ativação, com o intuito de aumentar o grau de liberdade desta função e, consequentemente, a capacidade de aproximação da rede

Data Science Academy

A Arquitetura de Redes Neurais Artificiais



Vértices são os neurônios e as arestas as sinapses

Arquitetura das Redes Neurais Artificiais

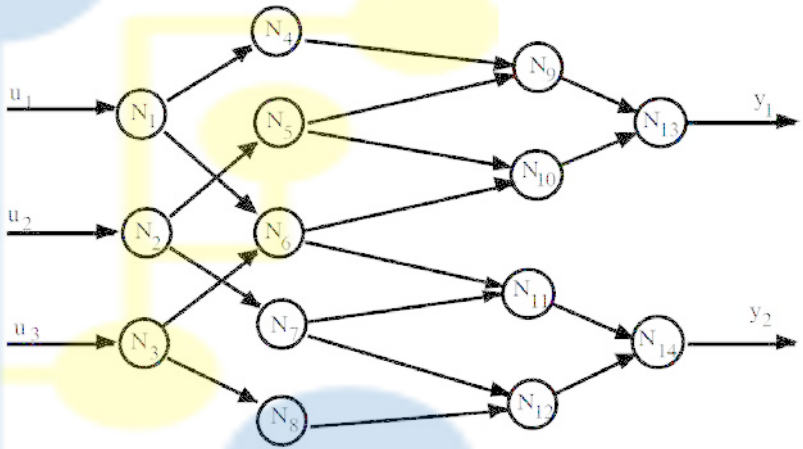
Feed Forward Networks
(Redes Diretas)

Feed Backward Networks
(Redes Recorrentes)

Redes Competitivas

Feed Forward Networks
(Redes Diretas)

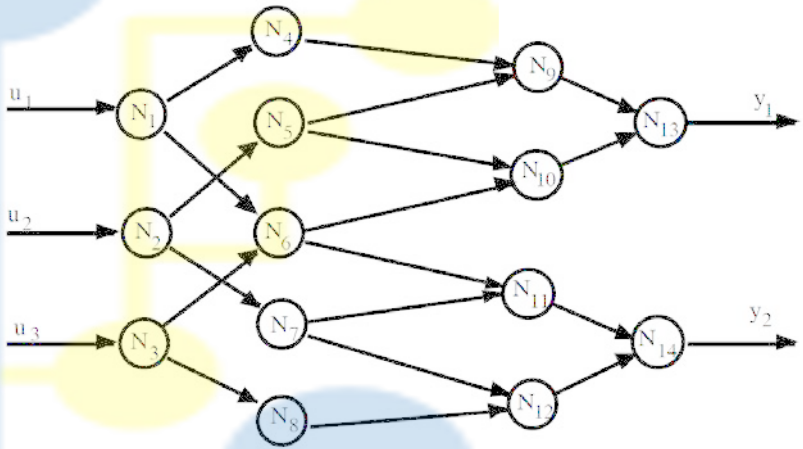
Os neurônios são arranjados em camadas, sendo que a camada inicial recebe os sinais de entrada enquanto a camada final obtém as saídas. As camadas intermediárias são chamadas de camadas ocultas.



Fluxo da Informação

Feed Forward Networks
(Redes Diretas)

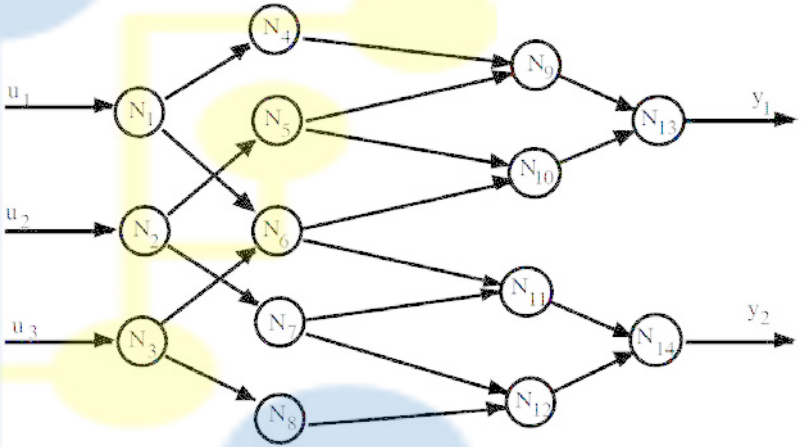
Cada neurônio de uma camada é conectado com todos os neurônios da camada seguinte.



Fluxo da Informação

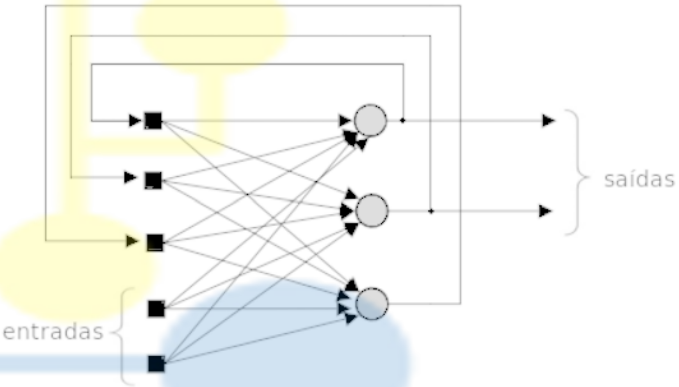
Feed Forward Networks
(Redes Diretas)

Não há conexões entre neurônios de uma mesma camada.

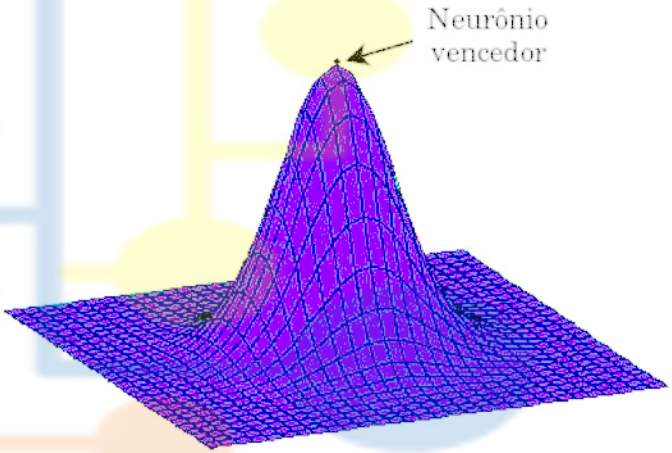


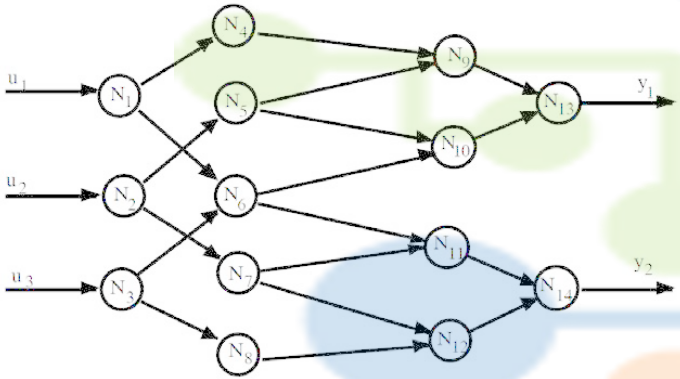
Fluxo da Informação

Feed Backward Networks
(Redes Recorrentes)

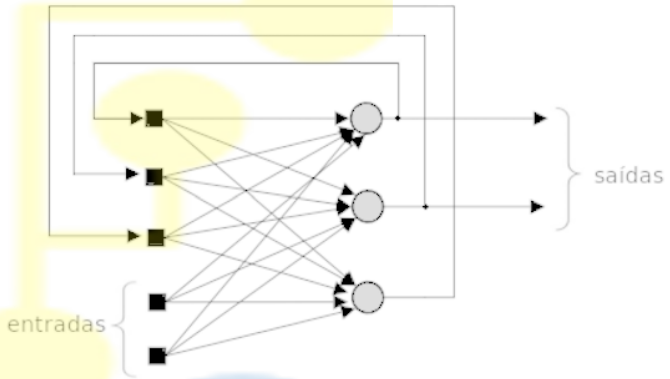


Rede Competitiva





Rede Direta

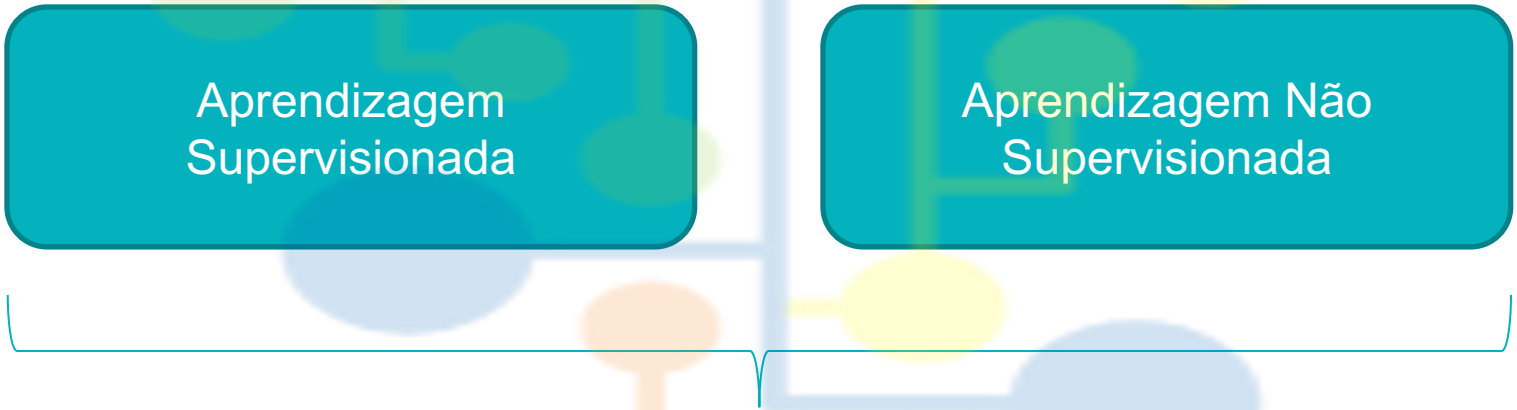


Rede Recorrente

VC Dimension

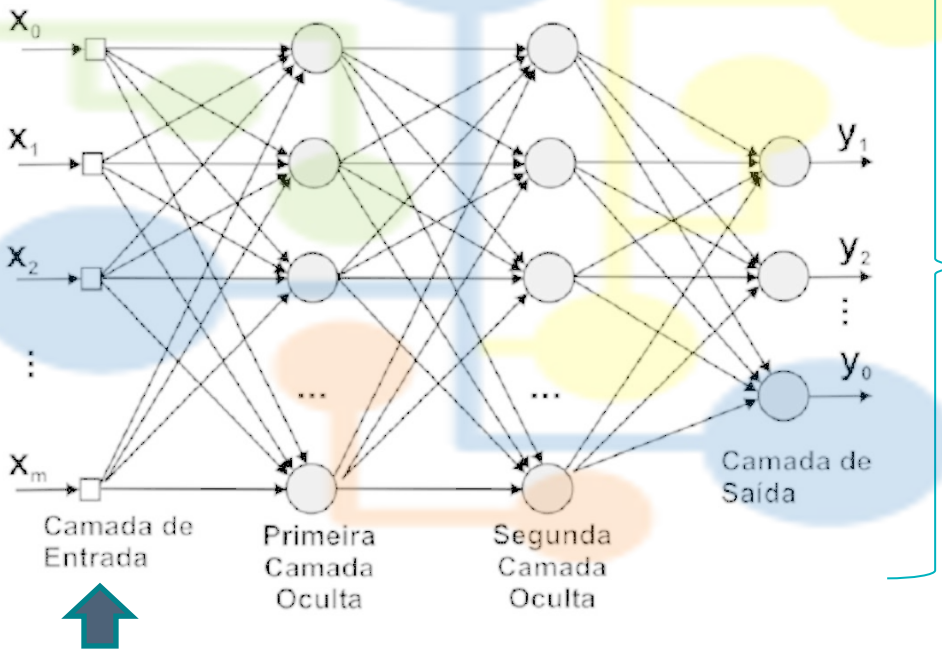
Data Science Academy

Processo de Aprendizagem



Processo de Aprendizagem de Redes Neurais

Sinal de Entrada



Sinal de Saída

Tipo de Regra de Aprendizagem	Descrição
Aprendizagem por Correção de Erro (Regra Delta)	Utilizado em treinamento supervisionado, esta técnica ajusta os pesos sinapticos por meio do erro, que é obtido através da diferença entre o valor de saída da rede e o valor esperado em um ciclo de treinamento. Com isso gradualmente vai diminuindo o erro geral da rede.
Aprendizagem Hebbiana	Baseado no postulado de aprendizagem de Hebb, que afirma: “se dois neurônios em ambos os lados de uma sinapse são ativados sincronamente e simultaneamente, então a força daquela sinapse é seletivamente aumentada". Este processo de treinamento é feito localmente, ajustando o peso das conexões baseado nas atividades dos neurônios.
Aprendizagem de Boltzmann	Método de aprendizagem estocástico derivado de conceitos da Estatística. Neste modelo os neurônios são estocásticos, podendo residir em dois estados possíveis, ligado (+1) e desligado (-1), e ainda são divididos em dois grupos funcionais, presos e livres, sendo responsáveis pela interação com o ambiente e pela explicação das restrições subjacentes dos padrões de entrada do ambiente, respectivamente. Um ponto importante deste modelo de aprendizagem é que os neurônios possuem conexões bidirecionais.
Aprendizagem Competitiva	Neste modelo de aprendizagem os neurônios são forçados a competir entre si e somente um será ativo, em uma dada iteração, o vencedor, ou seja, o que tiver maior similaridade com o padrão de entrada. Todos os pesos dos neurônios próximos ao neurônio vencedor terão seus valores ajustados.

Aprendizagem Hebbiana

”Quando um axônio da célula A está perto o suficiente para excitar uma célula B e participa do seu disparo repetidamente, então algum processo de crescimento ou modificação metabólica acontece em uma das células ou em ambas, de tal forma que a eficiência de A como uma das células que dispara B é aumentada.”

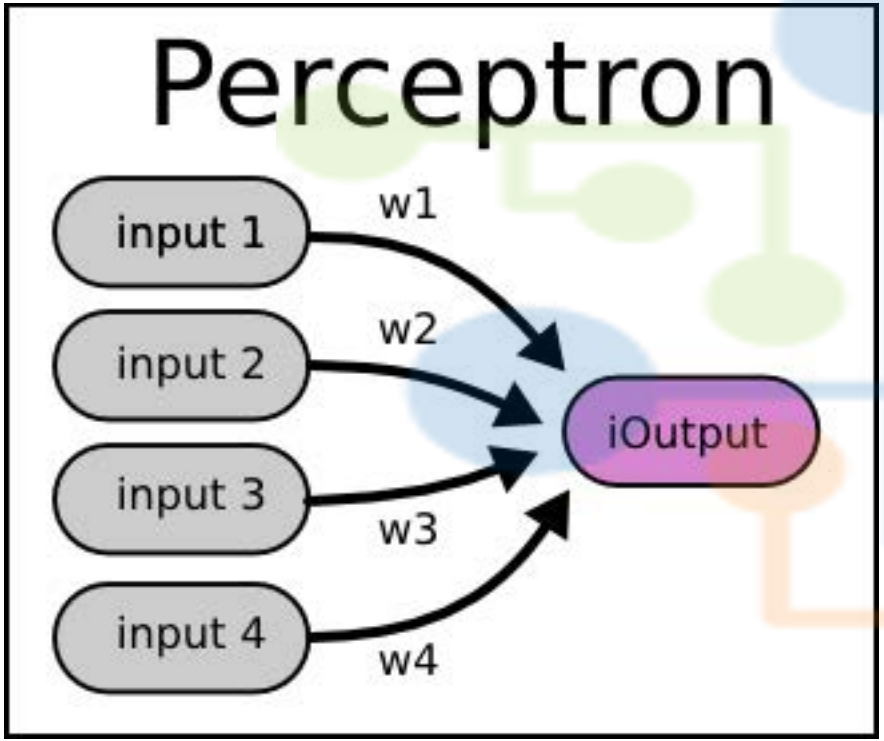
Aprendizagem Hebbiana

Se dois neurônios em ambos os lados de uma sinapse são ativados ***sincronamente***, então a força desta sinapse é aumentada

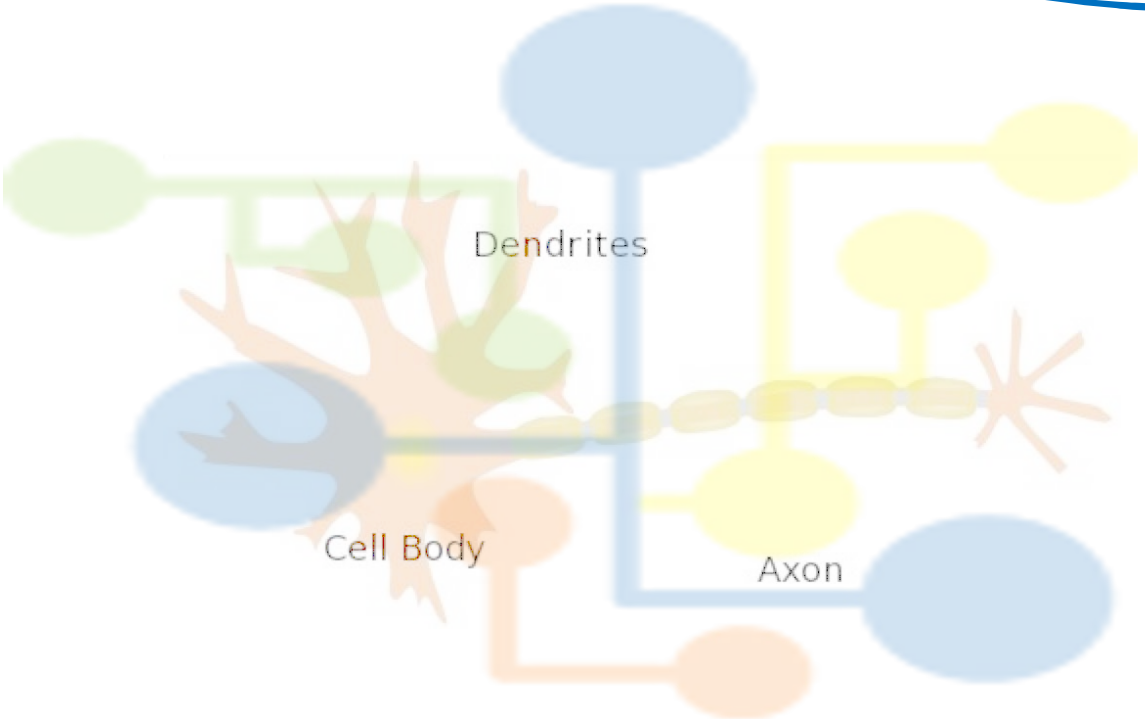
Se dois neurônios em ambos os lados de uma sinapse são ativados ***assincronamente***, então a força desta sinapse é enfraquecida

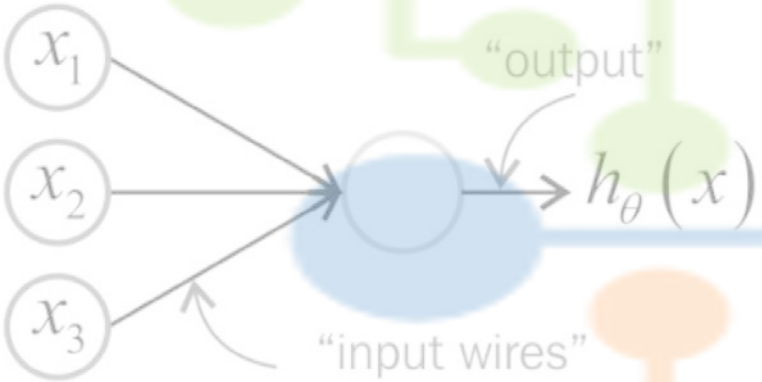
Data Science Academy

Perceptron

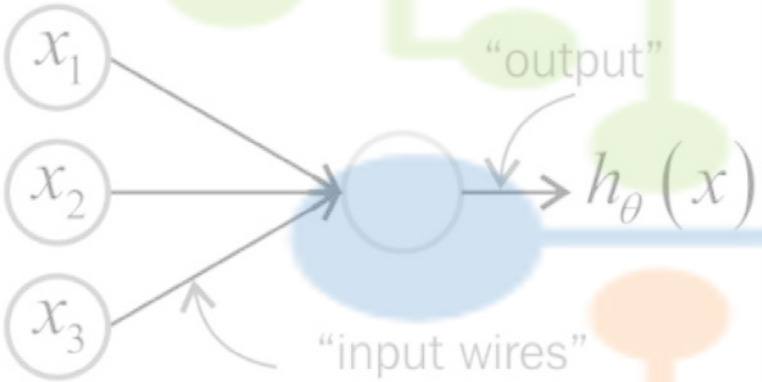


O Perceptron e suas limitações inspiraram os modelos mais avançados de redes neurais

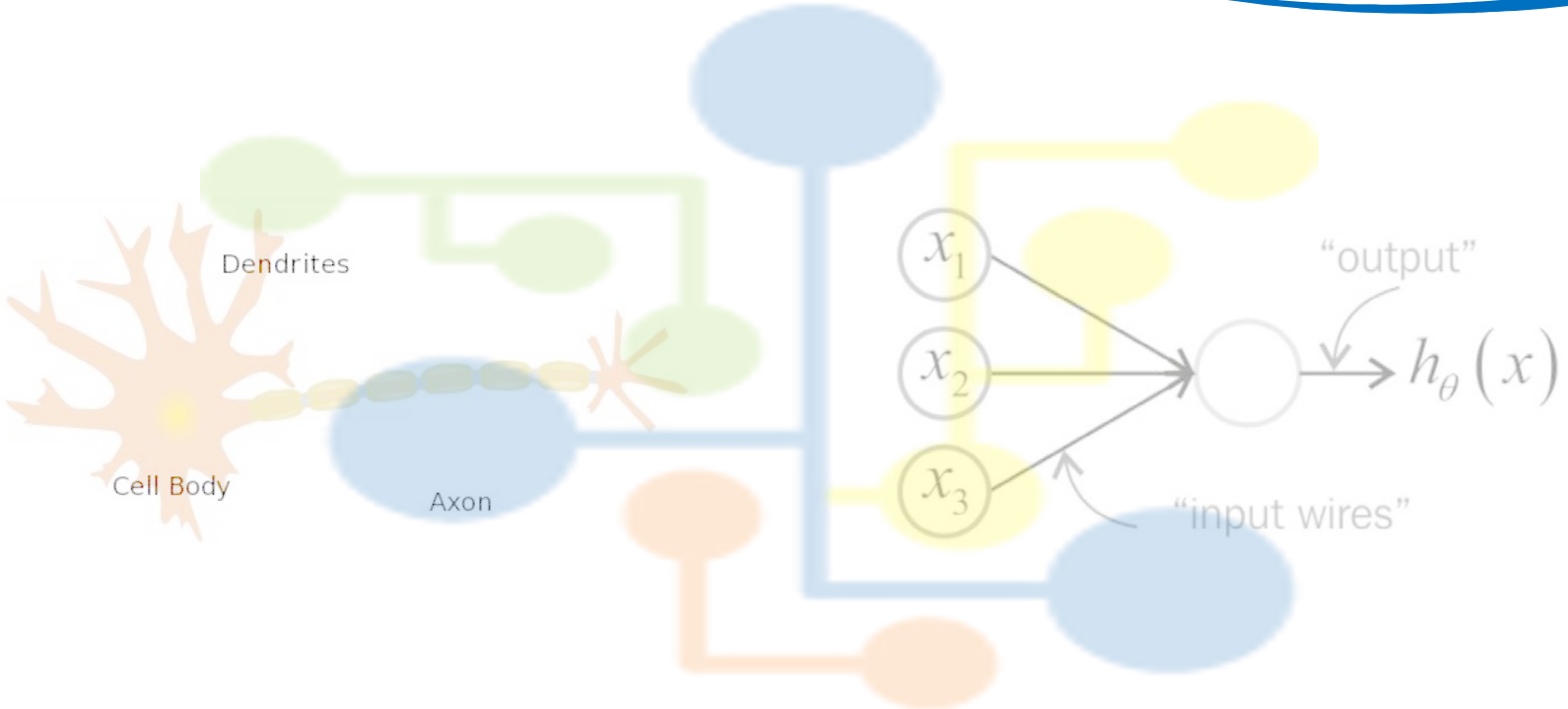

Dendrites
Cell Body
Axon

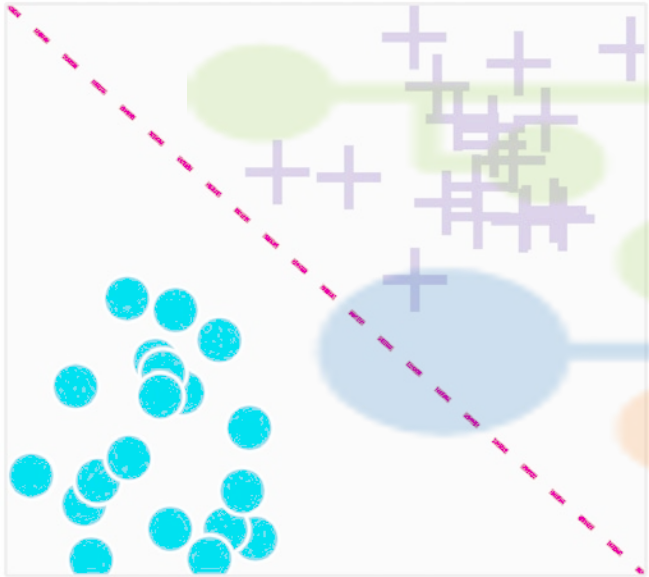


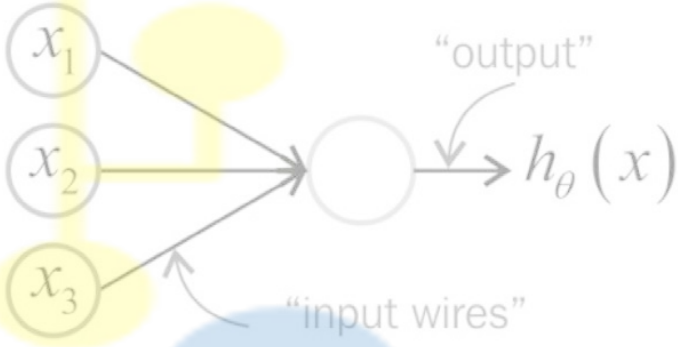
Um perceptron funciona analogamente a um neurônio

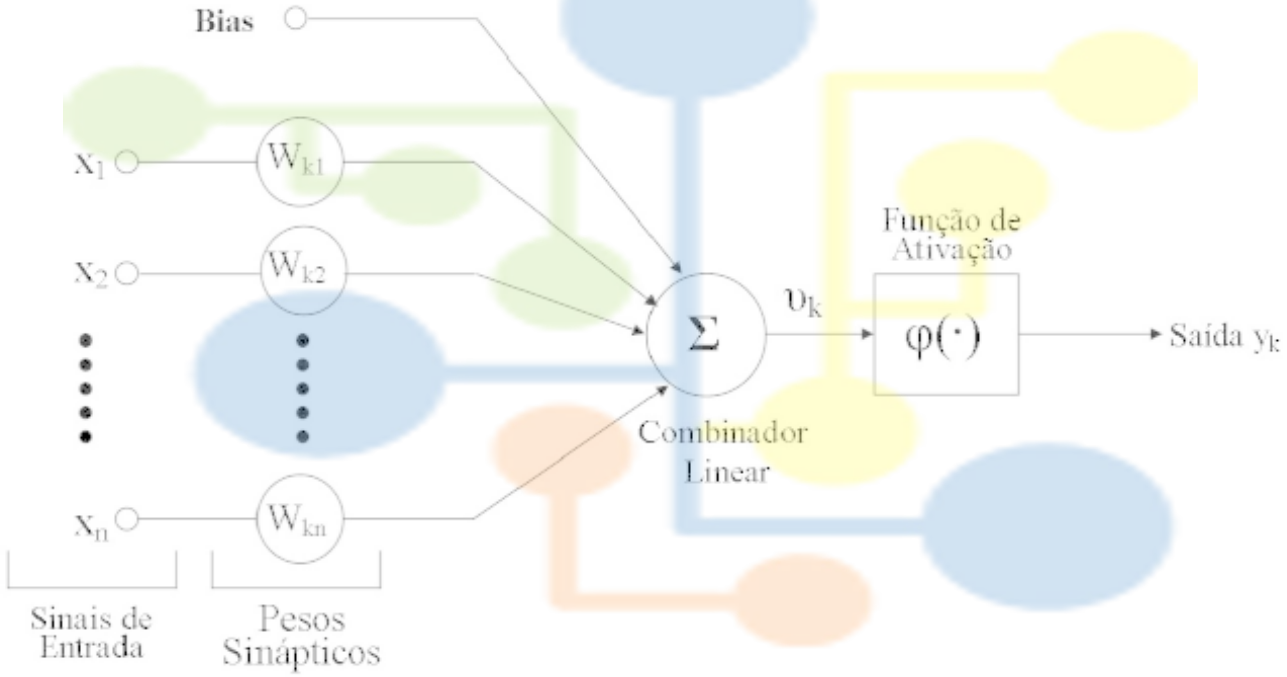


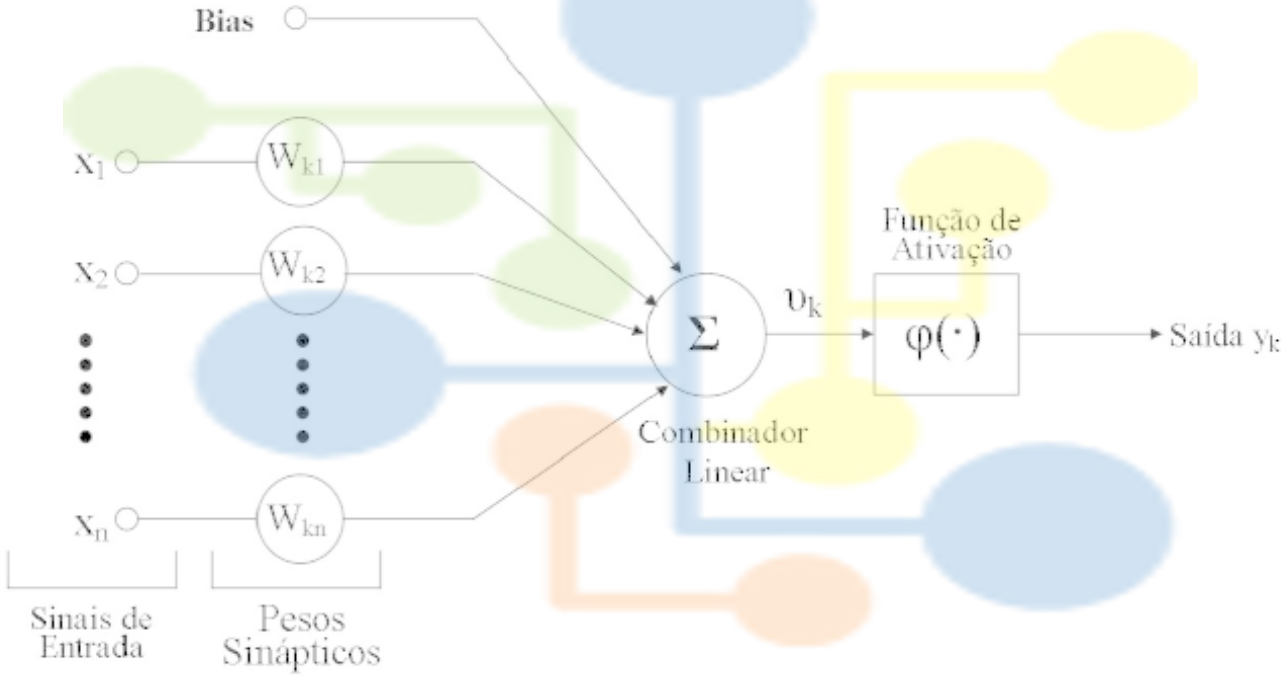
Os perceptrons são capazes de aprender online, com o erro


Dendrites
Cell Body
Axon
x_1
x_2
x_3
"output"
$h_\theta(x)$
"input wires"




x_1
x_2
x_3
"output"
$h_\theta(x)$
"input wires"


Bias
x_1
x_2
x_n
W_{k1}
W_{k2}
W_{kn}
Sinais de Entrada
Pesos Sinápticos
Σ
Combinador Linear
v_k
Função de Ativação
φ(·)
Saída y_k

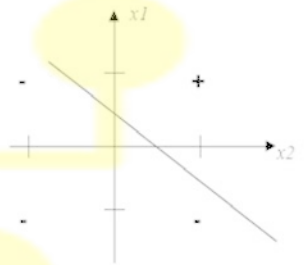

Bias
x_1
x_2
x_n
W_{k1}
W_{k2}
W_{kn}
Sinais de Entrada
Pesos Sinápticos
Σ
Combinador Linear
v_k
Função de Ativação
φ(·)
Saída y_k

Com o Perceptron, o objetivo é aprender **pesos sinápticos** de tal forma que a unidade de saída produza a saída correta para cada exemplo.

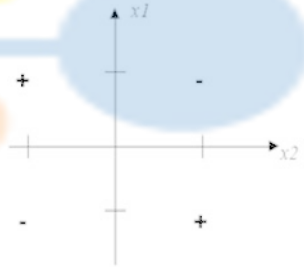
O algoritmo faz atualizações iterativamente até chegar aos pesos corretos.

Limitações do Perceptron

Um único Perceptron consegue resolver somente funções linearmente separáveis.



Em funções não linearmente separáveis o perceptron não consegue gerar um hiperplano para separar os dados.



Data Science Academy

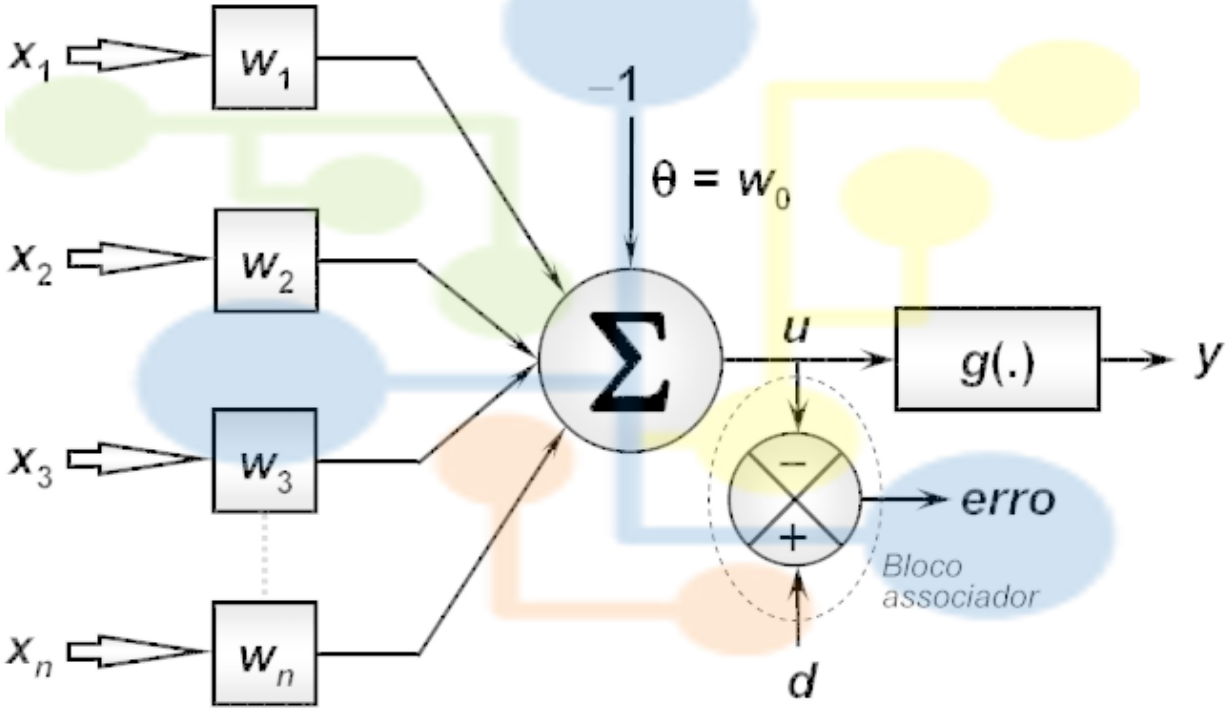
Adaline (Adaptive Linear Element)

Adaline

ADAptive LINear Element

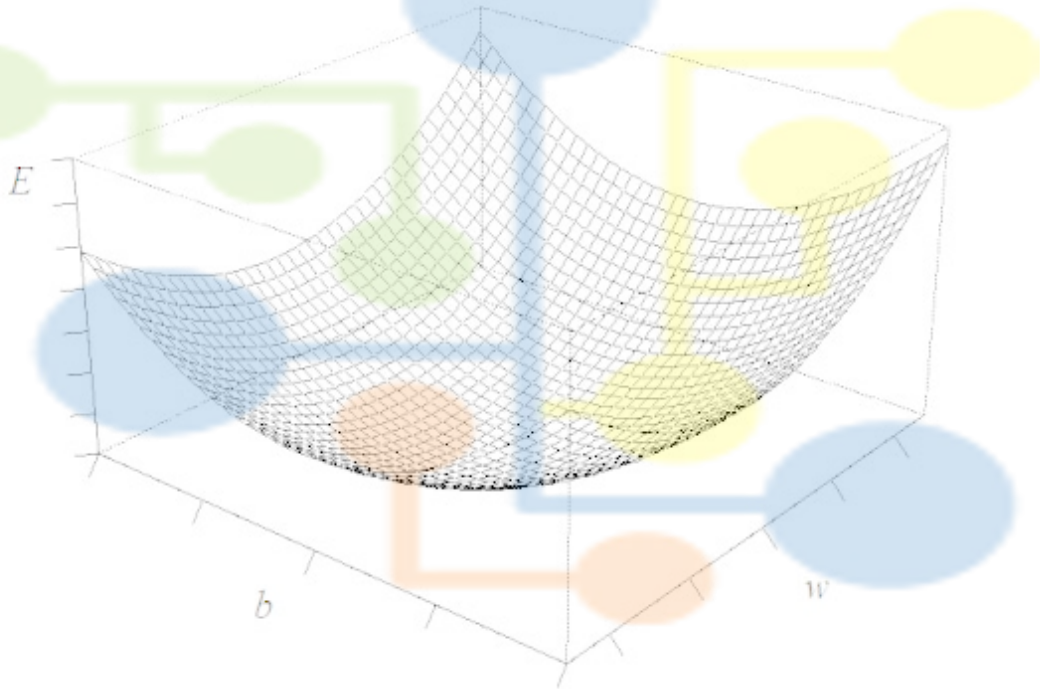
ou ADAptive LInear NEuron

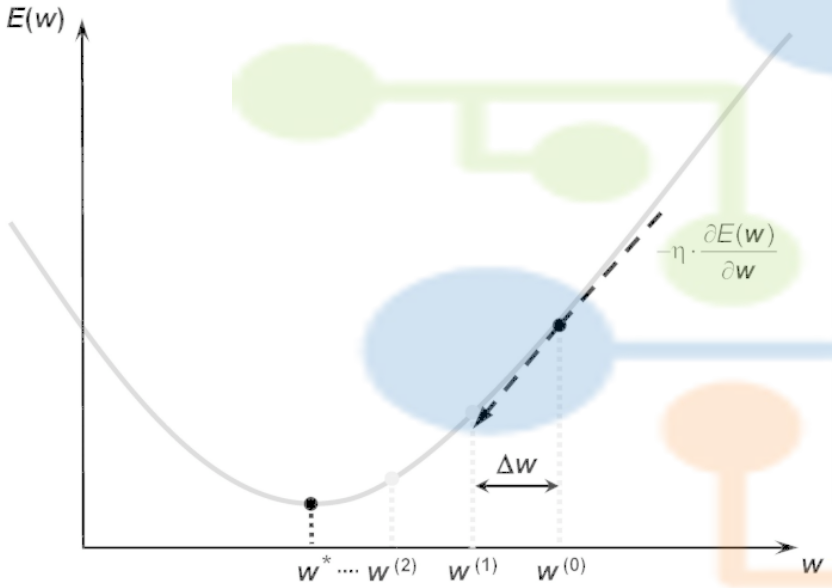
Bernard Widrow e Marcian Hoff


x1
x2
x3
xn
w1
w2
w3
wn
−1
θ = w0
Σ
u
g(.)
y
−
+
erro
Bloco
associador
d

O Adaline é similar ao Perceptron, com diferença apenas pelo seu algoritmo de treinamento. Enquanto o Perceptron ajusta os pesos somente quando um padrão é classificado incorretamente, o Adaline utiliza a Regra Delta para minimizar o erro médio (MSE) após cada padrão ser apresentado, ajustando os pesos proporcionalmente ao erro

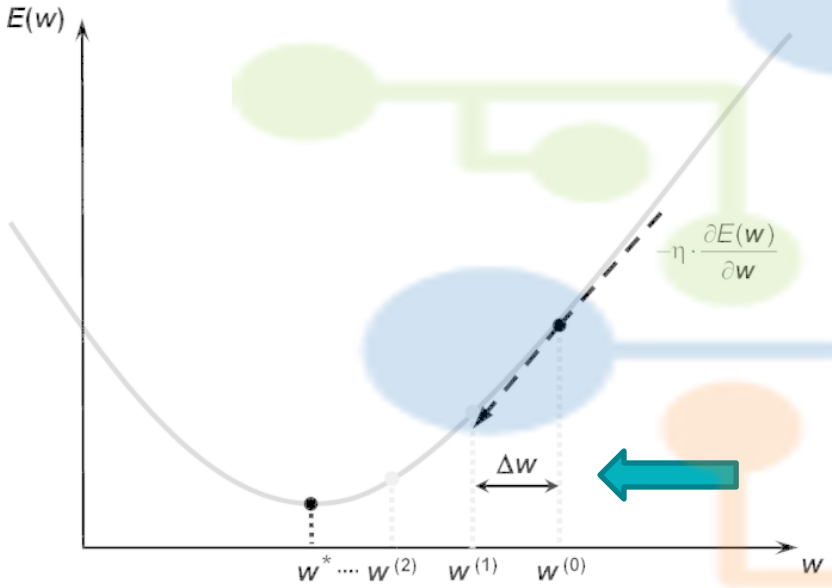
$$E = \frac{1}{2M} \sum_{p=1}^{M} \left(d^{\,p} - S^{\,p} \right)^2 = \frac{1}{2M} \sum_{p=1}^{M} \left[d^{\,p} - \sum_{i=0}^{N} \omega_i x_i^p \right]^2$$


E
b
w



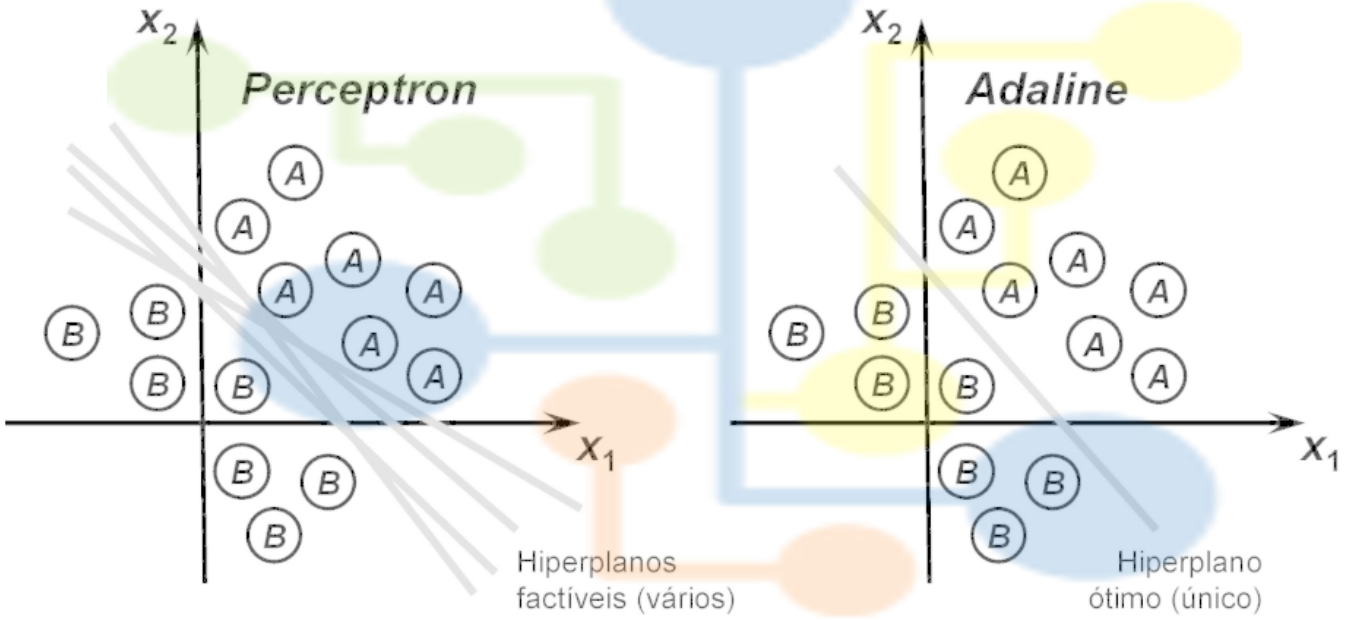
Regra Delta

Inicialmente, atribui-se aos pesos valores aleatórios e, com eles, apresenta-se um conjunto de sinais de entrada e calcula-se a resposta da rede. Então, comparam-se os valores calculados com os valores desejados (treinamento supervisionado).



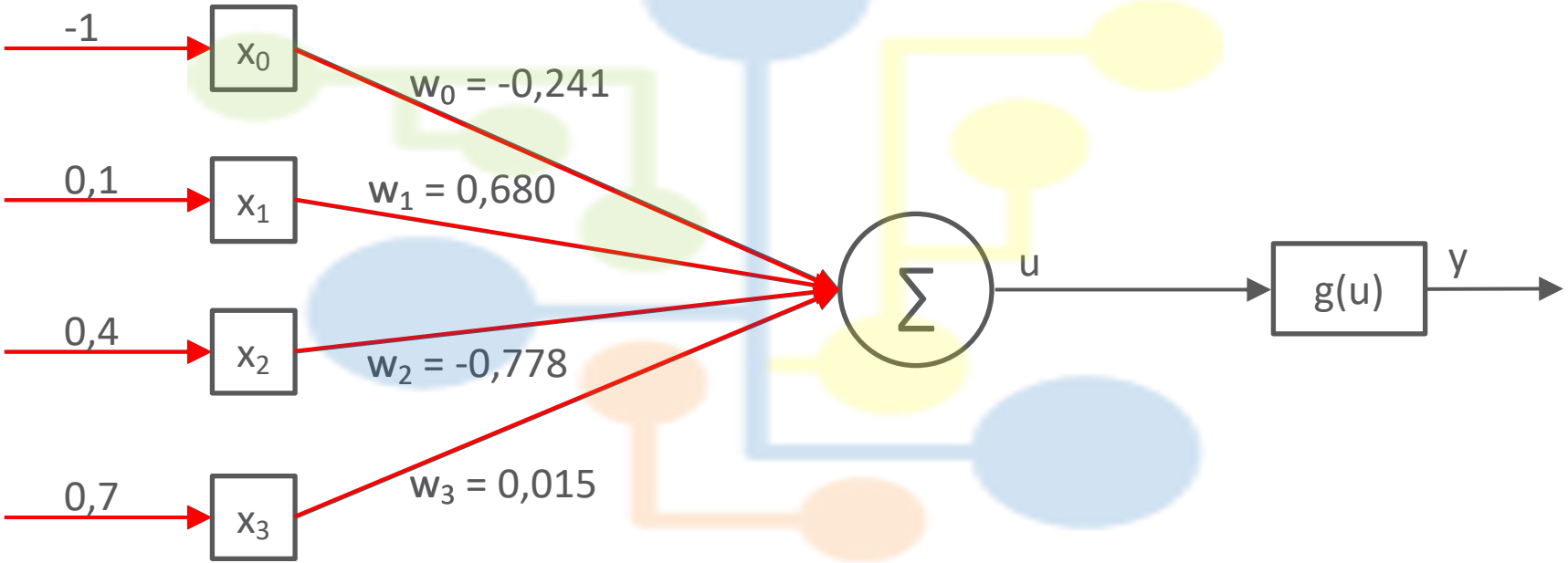
Regra Delta

Esse algoritmo, conhecido como regra delta, deu origem, anos mais tarde, ao primeiro algoritmo de treinameto de redes Perceptron de múltiplas camadas, o Backpropagation.

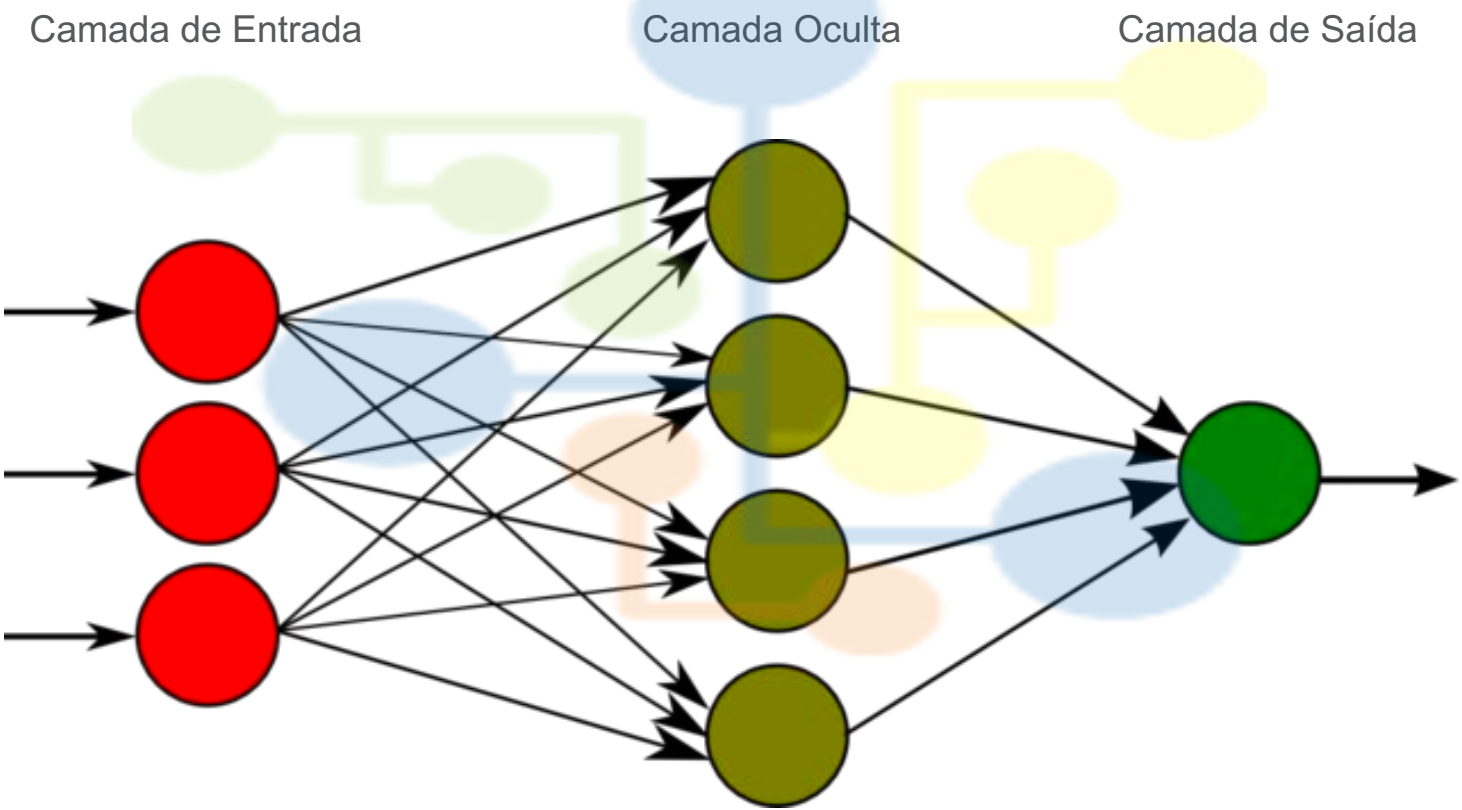

Perceptron
Adaline
x_2
x_1
A
B
Hiperplanos factíveis (vários)
Hiperplano ótimo (único)

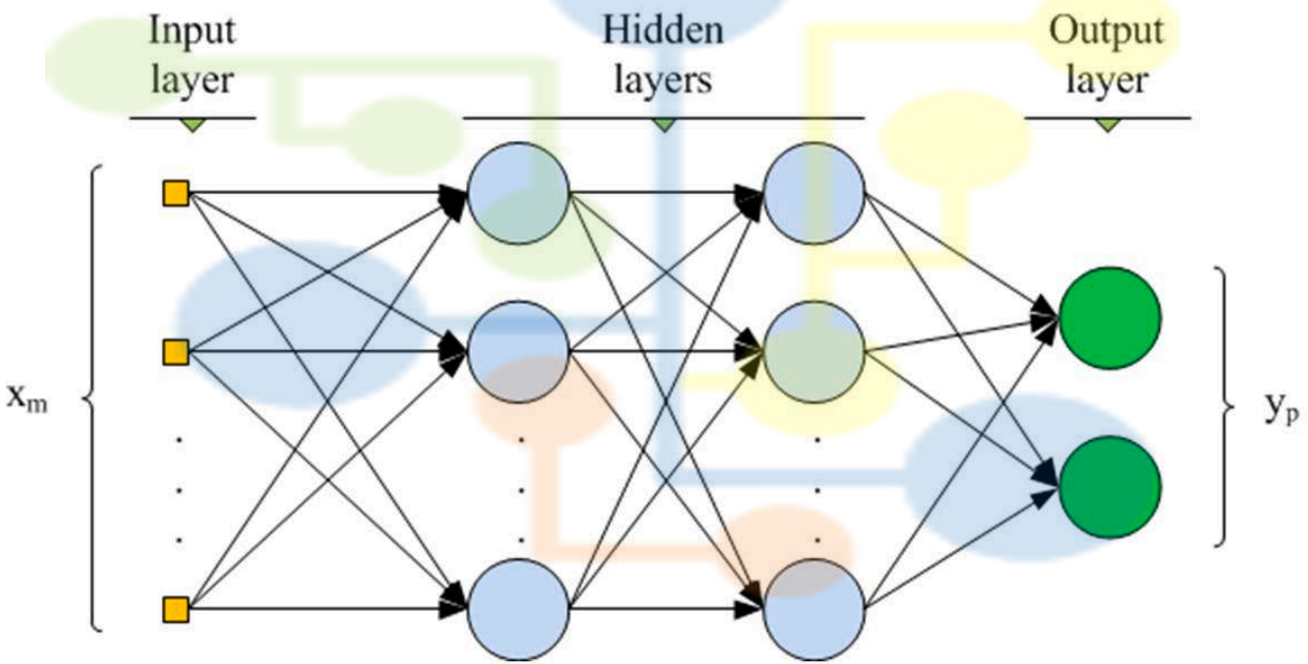
Data Science Academy

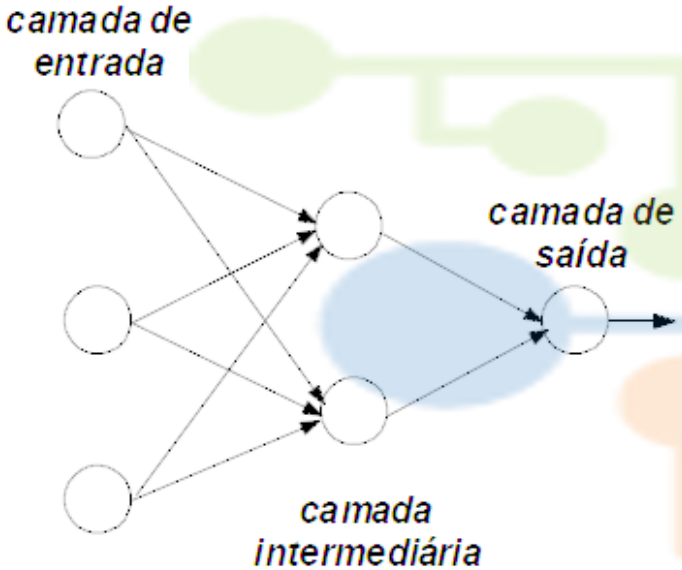
Perceptrons de Múltiplas Camadas


-1
x_0
w_0 = -0,241
0,1
x_1
w_1 = 0,680
0,4
x_2
w_2 = -0,778
0,7
x_3
w_3 = 0,015
Σ
u
g(u)
y

Redes Multilayer Perceptron são redes diretas (feed forward) que possuem uma ou mais camadas de neurônios entre as camadas de entrada e saída, chamada de camada oculta. Esta camada adiciona um poder maior em relação às redes Perceptron de camada única, que classifica apenas padrões linearmente separáveis, sendo os neurônios ocultos responsáveis por capturar a não-linearidade dos dados.

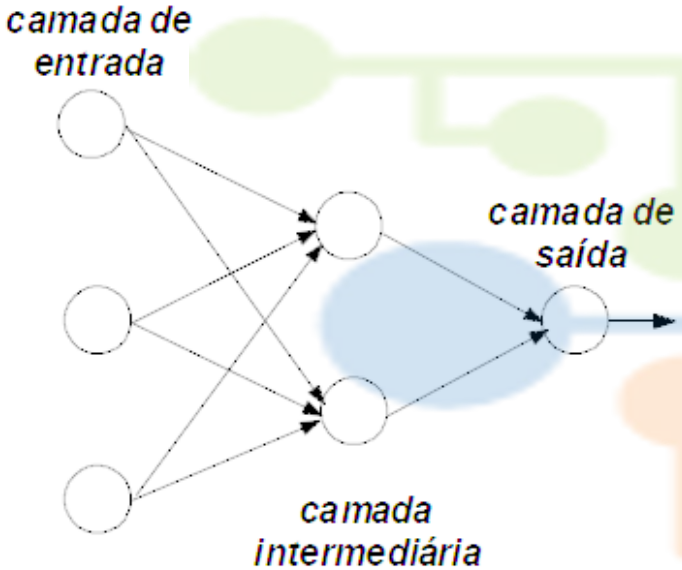

Camada de Entrada
Camada Oculta
Camada de Saída


Input layer
Hidden layers
Output layer
x_m
y_p



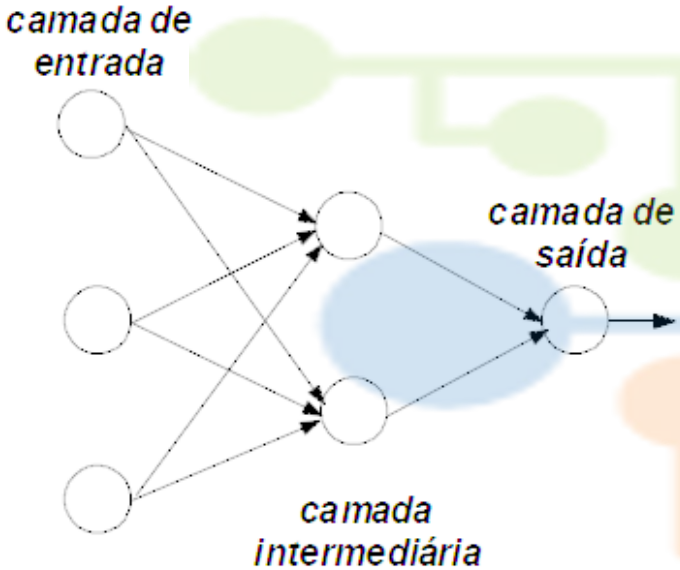
Treinamento da Rede – Opção 1

Dividir a rede neural em sub-redes



Treinamento da Rede – Opção 2

Treinar toda a rede como uma única unidade

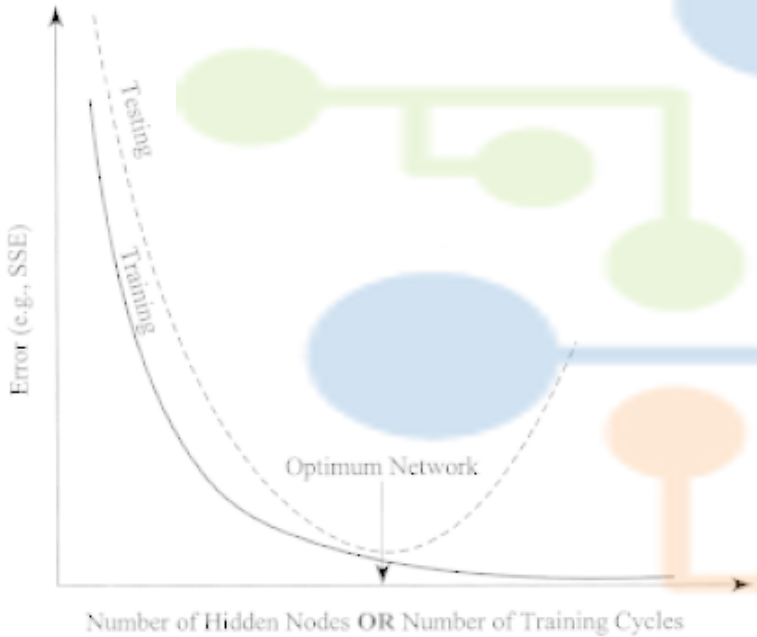


Treinamento da Rede – Opção 3

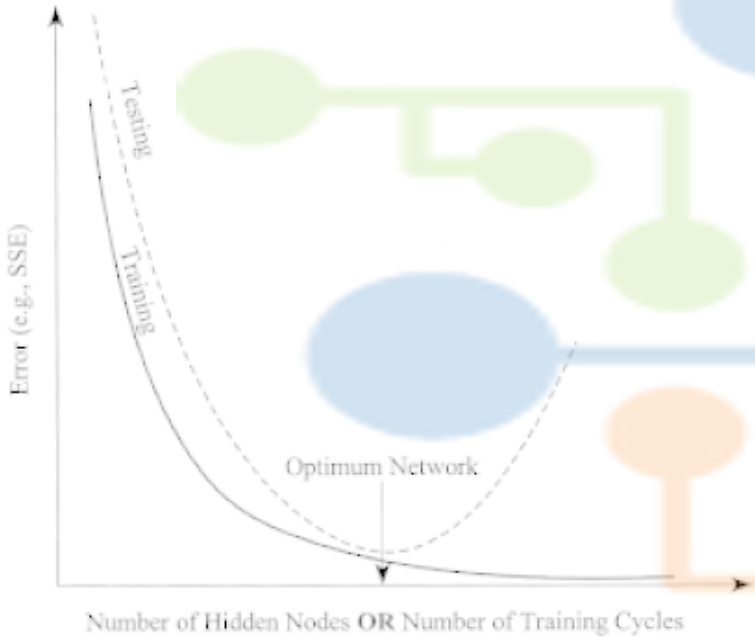
Treinamento baseado no Gradiente Descendente

Algoritmo 1: Algoritmo de treinamento *backpropagation*

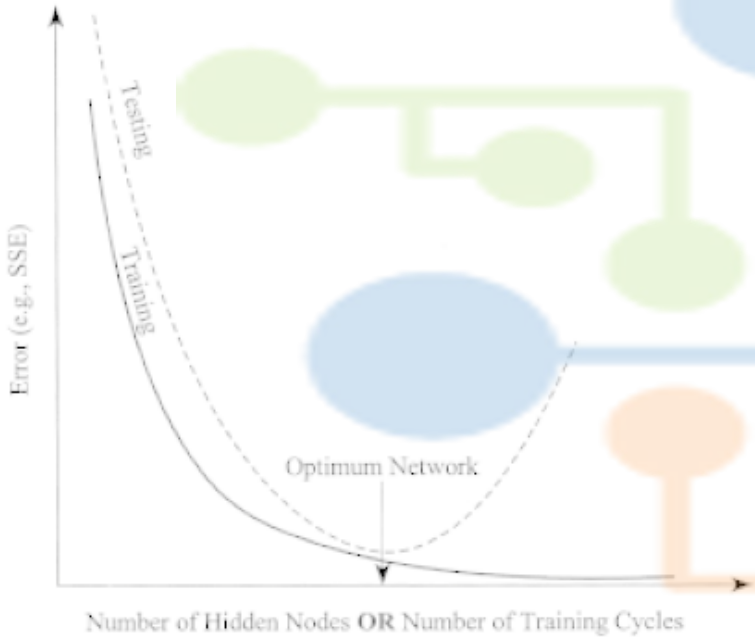
```
1  begin
2      Atribuição de valores iniciais aos pesos sinápticos;
3      repeat
4          Apresentação à rede dos padrões de entrada e as saídas desejadas;
5          Cálculo dos valores de saída dos neurônios ocultos;
6          Cálculo dos valores de saída dos neurônios de saída (resposta real da rede);
7          Cálculo do erro (diferença entre resposta da rede e valor esperado);
8          Ajuste dos pesos sinápticos;
9      until Condição de parada não satisfeita ;
10 end
```



Alguns parâmetros são determinados por tentativa e erro, ou seja, são atribuídos vários valores distintos aos parâmetros e analisando os resultados obtidos, a melhor conguração é escolhida!



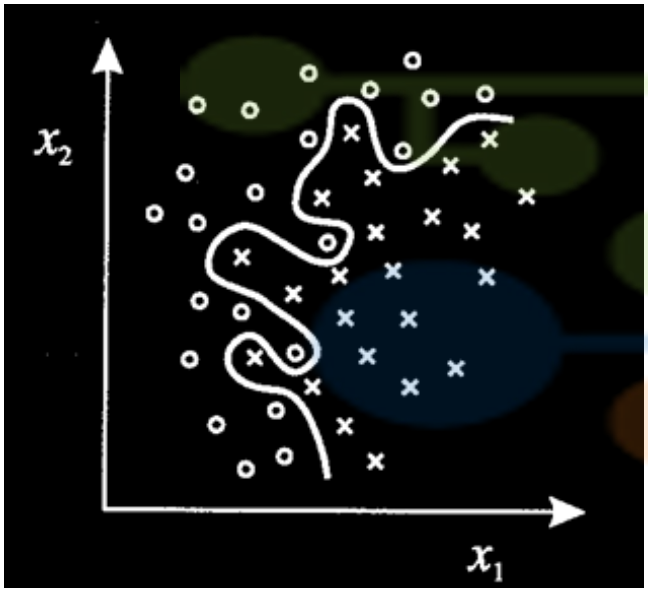
Outra diculdade é a determinação do número ideal de ciclos de treinamento da rede, que é determinado por tentativa e erro



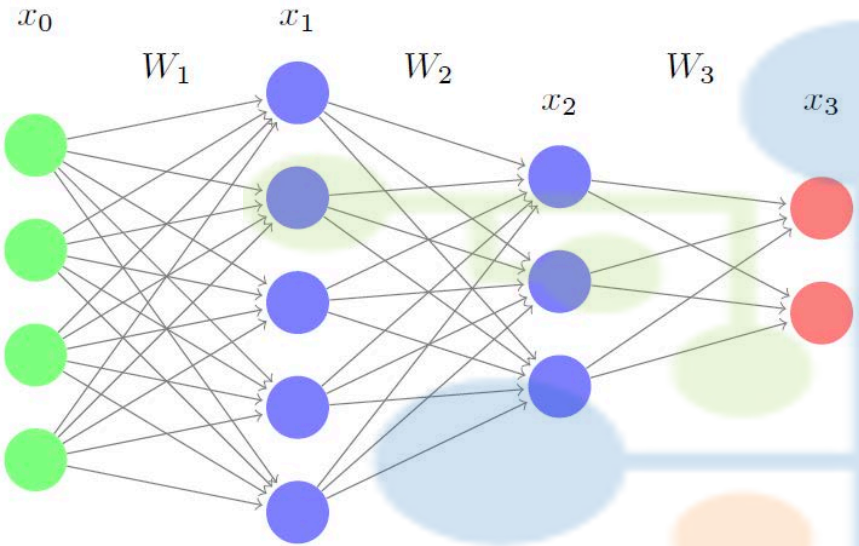
O overfitting é identicado quando o erro de teste, obtido pela validação cruzada, começa a aumentar depois de ter diminuído

Data Science Academy

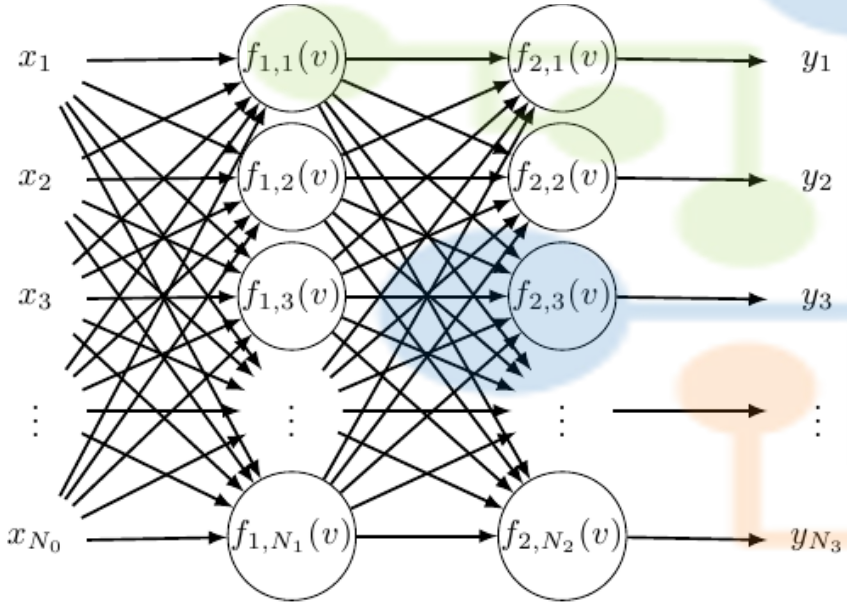
O Algoritmo Backpropagation



Classes não-linearmente separáveis



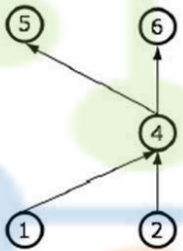
O Backpropagation é multicamada, pois tem no mínimo 3 camadas (entrada, intermediária, saída)



Os pesos das conexões das unidades das camadas internas vão sendo modificados conforme o erro é retropropagado.

Generalized Delta Rule

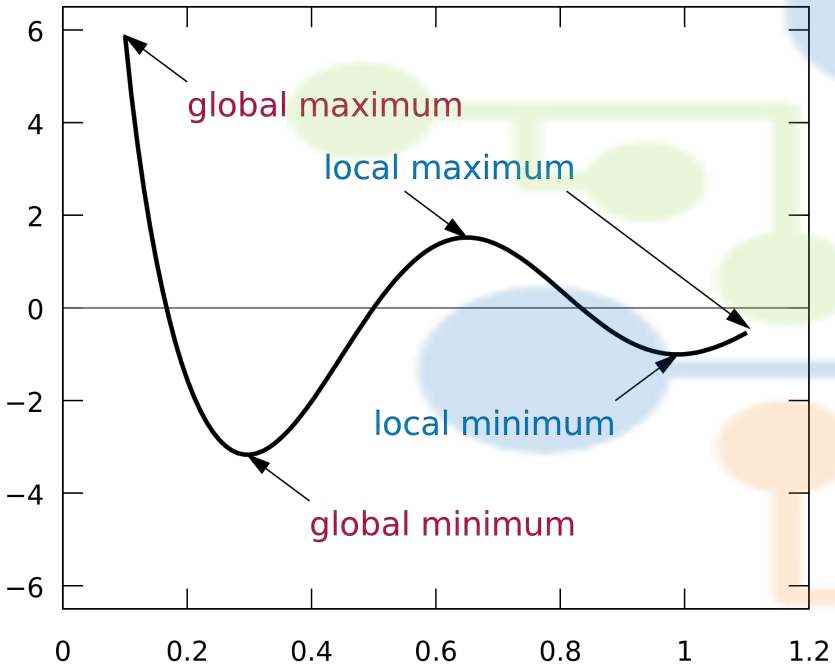
$$\delta_4 = y_4(1-y_4)(\delta_5 w_{4\to5} + \delta_6 w_{4\to6})$$
$$\Delta w_{1\to4} = -\eta\delta_4 y_1$$
$$\Delta w_{2\to4} = -\eta\delta_4 y_2$$



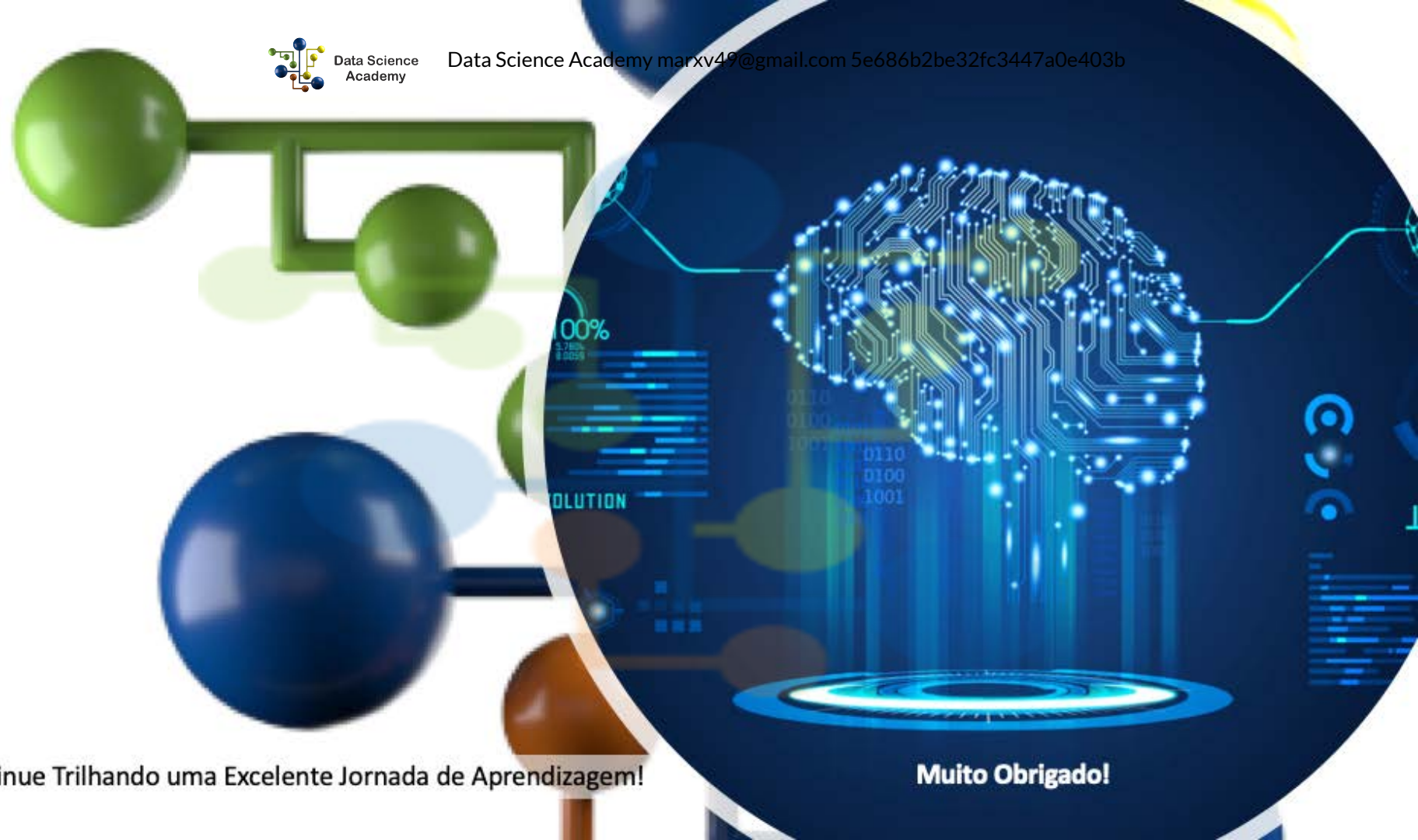
In general, for a hidden unit *j* we have

$$\delta_j = y_j(1-y_j)\sum_{k\in downstream(j)} \delta_k w_{j\to k}$$
$$w_{i\to j} = w_{i\to j} - \eta\delta_j y_i$$

As redes que utilizam backpropagation trabalham com uma variação da regra delta, apropriada para redes multi-camadas: a regra delta generalizada



O treinamento das redes MLP com backpropagation pode demandar muitos passos no conjunto de treinamento, resultando um tempo de treinamento considerávelmente longo



Continue Trilhando uma Excelente Jornada de Aprendizagem!

Muito Obrigado!