



## Objectivos

- Adquirir a noção de rede neuronal artificial
- Descrever a arquitectura de uma rede neuronal multicapa com conexões para diante
- Descrever o algoritmo de retropropagação do erro para a aprendizagem de redes neuronais



- Sumário:

- Redes Neurais Artificiais
  - Introdução
  - Algoritmo de retropropagação do erro

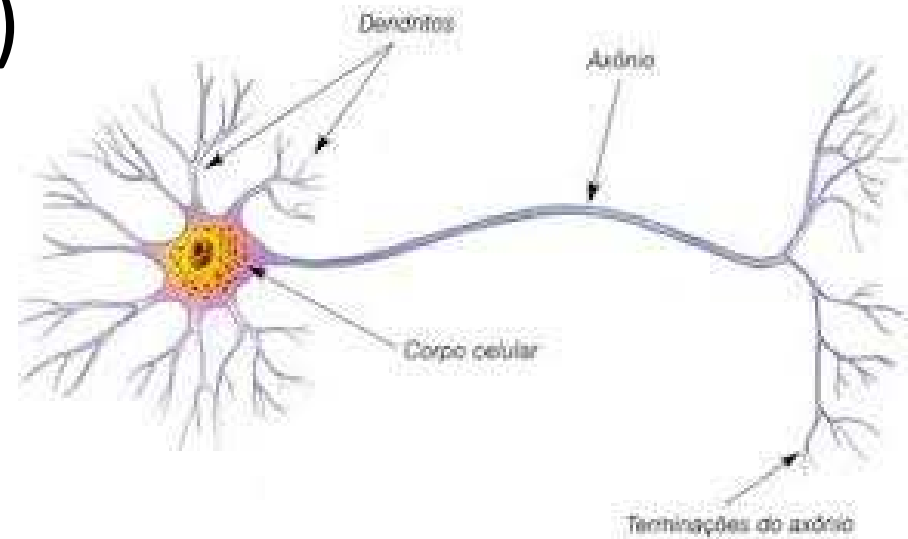


# Redes Neurais Artificiais (ANN)

- Constituem um enfoque de aprendizagem robusto para a aproximação de diferentes tipos de funções
- Se encontram entre os métodos de aprendizagem mais efectivos ao abordar certos tipos de problemas, tais como a interpretação de dados complexos captados por sensores
- Têm sido aplicadas com sucesso em problemas como a interpretação visual de cenas, reconhecimento de fala, a aprendizagem de estratégias de controlo de robots...

## Inspiração biológica (1/3)

- O seu estudo foi em parte inspirado na observação de que os sistemas de aprendizagem biológicos são constituídos por complexas redes de neurónios interconectados



## Inspiração biológica (2/3)

- Em analogia, as Redes Neurais Artificiais são compostas por um conjunto de unidades simples densamente interconectadas (designadas por isso *modelos connexionistas*)





## Inspiração biológica (3/3)

- Historicamente, a investigação na área foi desenvolvida por dois grupos de investigadores:
  - Um grupo motivado pela utilização das ANN para estudar e modelar processos de aprendizagem biológicos
  - Outro motivado pelo objectivo da obtenção de algoritmos de aprendizagem automática efectivos

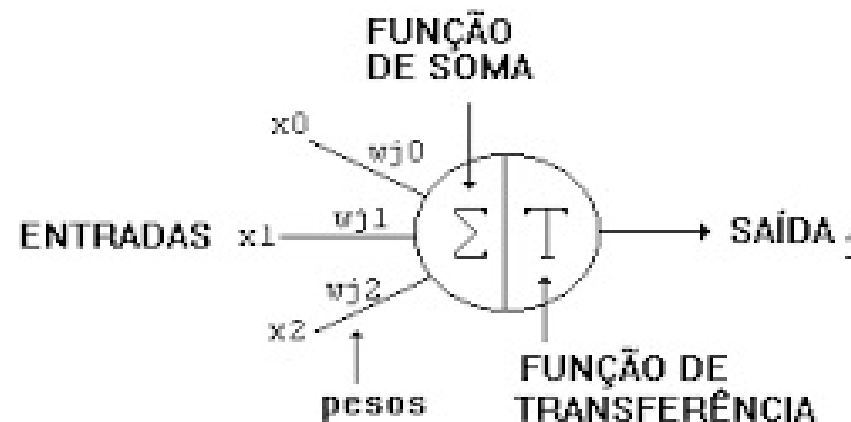


## Problemas apropriados

- Se adequam a problemas com as seguintes características
  - As instâncias são representadas por vectores de atributos
  - A função objectivo pode ser discreta, real ou um vector de vários atributos discretos ou reais
  - Os exemplos de treino podem conter erros
  - Pode ser requerida uma avaliação rápida da função objectivo
  - São aceitáveis longos tempos de treino
  - Não é importante a interpretação da função objectivo aprendida por parte dos seres humanos

## Estrutura básica (1/4)

- Existem diferentes modelos (diferentes filosofias de desenho, regras de aprendizagem, funções de construção das respostas...)
- Elemento básico -> unidade de processo
  - Também designadas *células* ou *neurónios*





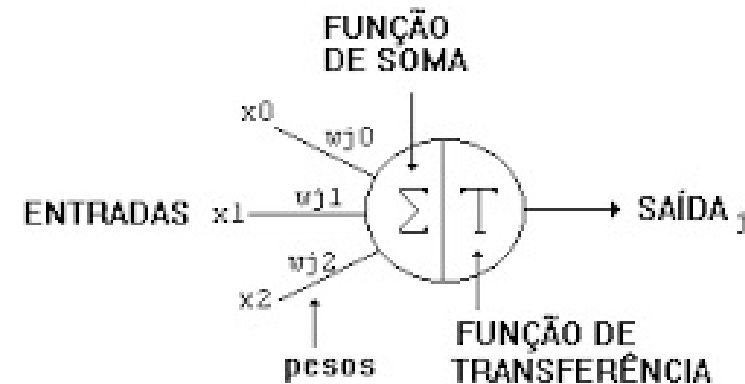
## Estrutura básica (2/4)

- Entradas
  - Podem ser provenientes de uma das entradas da rede ou
  - Das saídas de outras células -> conexão
    - Cada conexão tem um peso associado
  - O valor das entradas a uma célula é dado pela saída de cada célula com a qual está conectada, ponderada pelos respectivos pesos



## Estrutura básica (3/4)

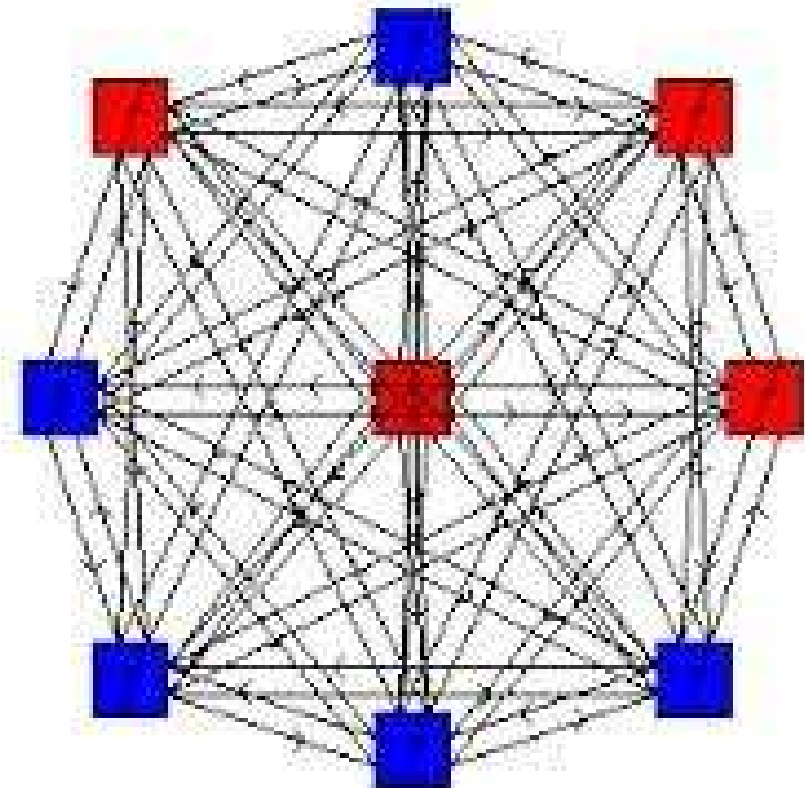
- Cada célula produz uma única saída que pode servir de entrada para outras células
- Para calcular a saída de uma célula
  - Se realiza a soma de todas as entradas que esta recebe e
  - Se aplica uma função não linear,  $F$ , que pode ter diferentes formas



$$s_j = F\left(\sum_{i=0}^n s_i w_{ij}\right)$$

## Estrutura básica (4/4)

- Uma ANN é formada por um conjunto de células interconectadas
- A estrutura de conexão é designada *arquitectura da rede*





## Objectivo da aprendizagem (1/2)

- As ANN são sistemas computacionais capazes de determinar, a partir de um conjunto de exemplos de treino, uma função implicitamente determinada



## Objectivo da aprendizagem (2/2)

- Dada:
  - A arquitectura da ANN (número de neurónios e padrão de conectividade entre eles)
  - Conjunto de parâmetros da rede (valores iniciais dos pesos, regras de actualização dos mesmos e função de activação)
  - Conjunto de exemplos de treino
- Determinar
  - Conjunto de valores dos pesos das conexões da ANN que caracterizam a todos os exemplos do conjunto de treino

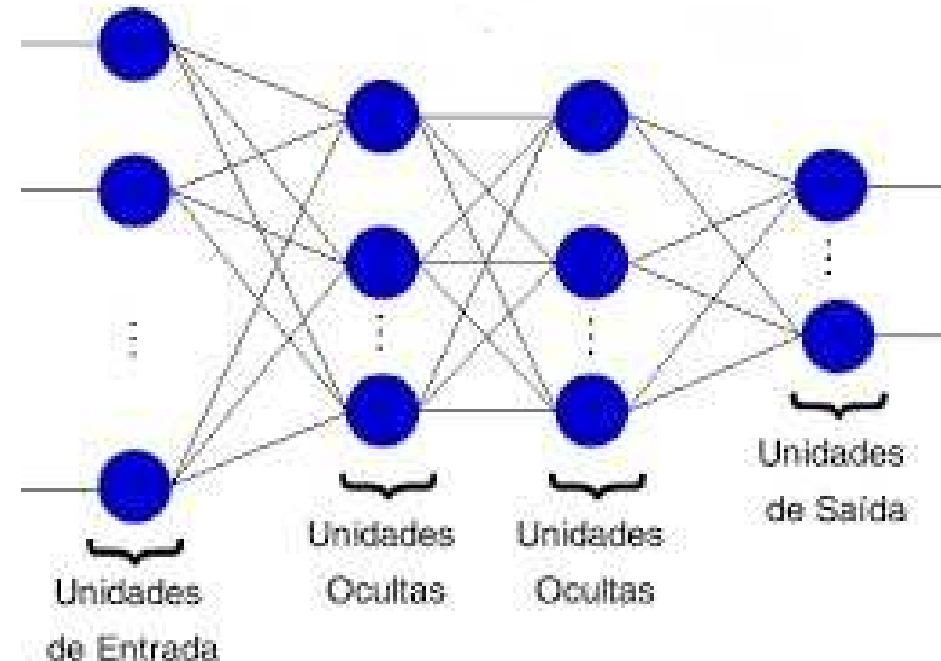


## Redes neuronais multicapa (1/4)

- Arquitectura típica -> rede neuronal multicapa com conexões para diante (*feed-forward neural network*)
- Composta por:
  - Um primeiro nível (células de entrada)
  - Uma ou várias capas intermédias (camadas ocultas)
  - Um último nível de saída

## Redes neuronais multicapa (2/4)

- As saídas das células de uma camada determinada podem ser conectadas somente às entradas de células da camada imediatamente a seguir



## Redes neuronais multicapa (3/4)

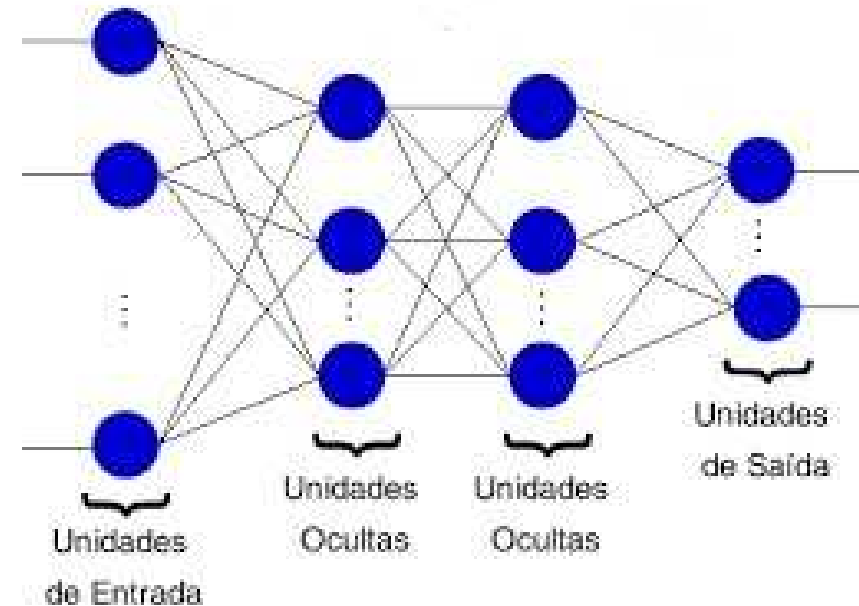
- A entrada à  $i$ -ésima célula de uma camada  $k$

$$E_i = \sum_{j=0}^{n_{k-1}} s_j w_{ji}$$

- A saída da célula  $i$ -ésima

$$s_i = F(E_i)$$

- Função de activação não linear, contínua e derivável

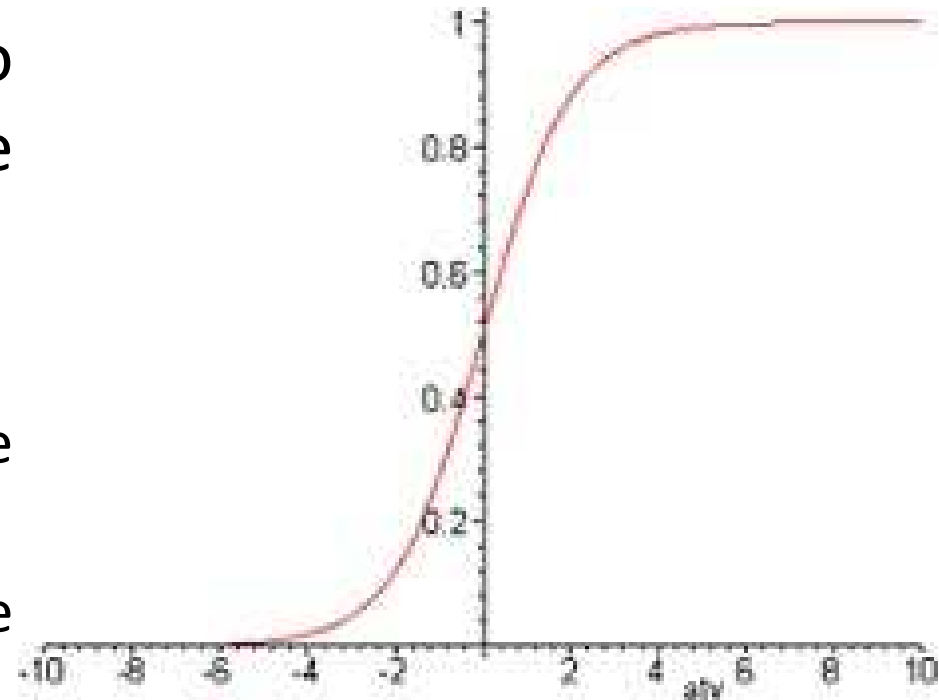




## Redes neuronais multicapa (4/4)

- Função de activação não linear, contínua e derivável
- Função sigmoide
  - Contínua e facilmente derivável
  - Saída normalizada entre 0 e 1

$$F(x) = \frac{1}{1 + e^{-x}}$$





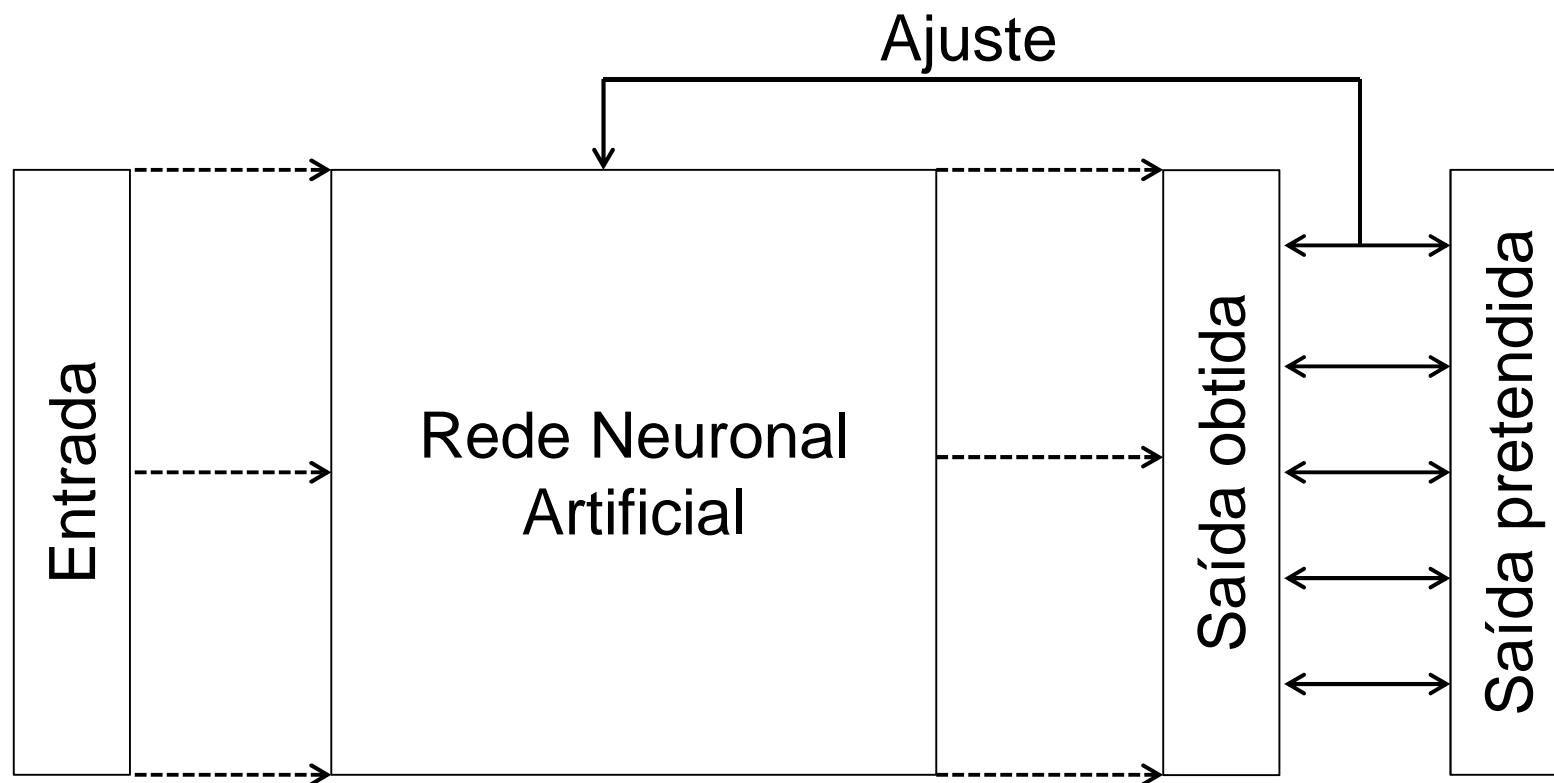
## Algoritmo de retropropagação do erro (1/3)

- Mecanismo de ajuste dos pesos da rede designado *método de retropropagação do erro (backpropagation)*
- Trata de adaptar os valores dos pesos das conexões de forma tal que seja minimizado o erro quadrático médio para o conjunto de exemplos de treino

$$\mathcal{E} = \sum_{j=1}^N \mathcal{E}_j$$

$$\mathcal{E}_j = \sum_{i=1}^n \frac{1}{2} (s_i - y_i)^2$$

## Algoritmo de retropropagação do erro (2/3)



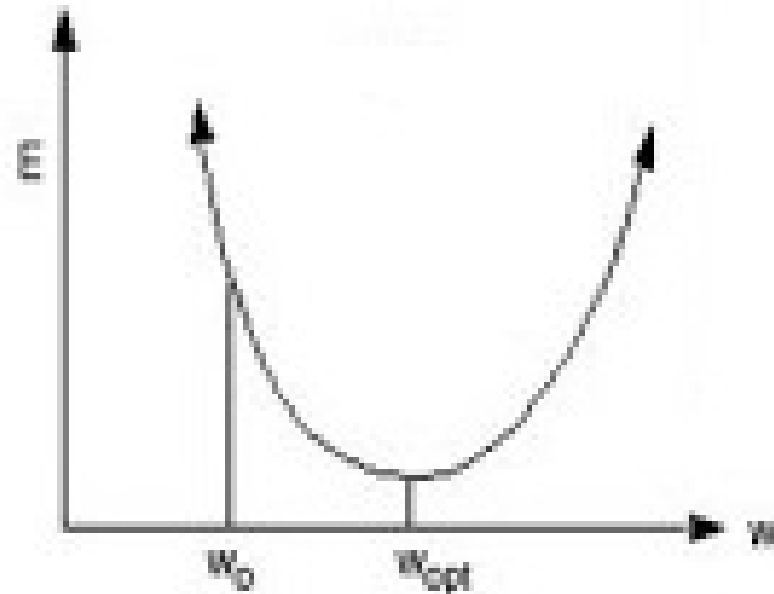


# Algoritmo de retropropagação do erro (3/3)

- Função Retropropagação
  - Entradas:
    - $C$  : matriz de conexões da rede
    - $E$  : exemplos de treino
  - Inicializar os pesos da rede com valores pequenos (ex. entre -0,05 e 0,05)
  - Enquanto não se alcança a condição de finalização
    - $\forall X = (\vec{x}, \vec{y}) \in E$ 
      - $\vec{s} = Saída(\vec{x})$
      - $\varepsilon = Erro(\vec{s}, \vec{y})$
      - $C = RegraAprendizagem(C, \varepsilon)$
  - Retornar  $C$

## Regra de aprendizagem (1/2)

- Para ajuste dos pesos se utiliza o método do gradiente descendente
- Se baseia no facto de que no ponto mínimo a derivada do erro com respeito aos pesos é igual a zero





## Regra de aprendizagem (2/2)

- Adaptação se faz de forma recursiva começando pela camada de saída, propagando o erro desde a saída até à entrada ajustando ao mesmo tempo os pesos

$$w_{ij} = w_{ij} + \mu \delta_j s_i$$

- $w_{ij}$  : peso da conexão entre a  $i$ -ésima célula da camada  $k - 1$  e a  $j$ -ésima célula da camada  $k$
- $\mu$  : taxa de aprendizagem (ou ganho)
- $s_i$  : saída produzida pela  $i$ -ésima célula da camada anterior
- $\delta_j$  : valor de incremento que depende da derivada do erro associado à  $j$ -ésima célula



## Término da aprendizagem

- O ciclo de actualização dos pesos pode ser repetido centenas de vezes numa aplicação típica
- Diferentes condições de finalização podem ser utilizadas
  - Finalização depois de uma quantidade fixa de iterações
  - Quando o erro sobre o conjunto de treino se encontra por baixo de certo nível
  - Quando o erro sobre um conjunto separado de exemplos de validação cumpre determinadas condições



## Bibliografia

- Mitchell, pg. 81 – 100
- Witten 232 – 241
- Borrajo Millán