3 REDES NEURAIS ARTIFICIAIS

Neste capítulo será apresentado um breve histórico das redes neurais artificiais de modo a situar o leitor, descrevendo-se suas aplicações, teorias e finalmente detalhando-se a rede *multi-layer perceptron (MLP)*, suas fórmulas de ativação e processo de aprendizagem.

As redes neurais artificiais são modelos computacionais fundamentados na estrutura neural biológica do cérebro humano, onde os neurônios (Figura 6) estão conectados uns aos outros através de sinapses, formando assim uma grande rede de processadores chamada rede neural.

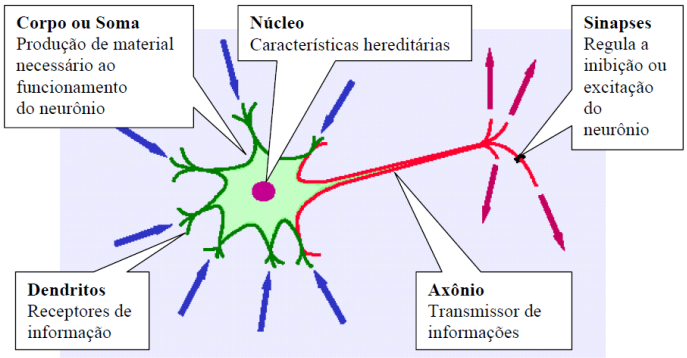


Figura 6 - Esquema de uma Célula Neural

No corpo humano, as sinapses transmitem estímulos por meio de diferentes concentrações de íons de Na+ (Sódio) e K+ (Potássio), adquirindo e compartilhando o processamento de grupos de informações ou sinais. Esta grande rede de informações proporciona a capacidade de processamento e armazenamento de informação e o resultado disto se estende pelo corpo humano.

O sistema nervoso é formado por um conjunto extremamente complexo de neurônios no qual a comunicação é realizada por meio de pulsos. Quando um pulso é recebido, o neurônio processa o sinal e, após um curto período, dispara um segundo pulso que produz uma substância neurotransmissora que flui do corpo celular para o axônio, que pode ou não estar conectado a um dendrito de outra célula. O neurônio que transmite o pulso pode controlar a frequência de

pulsos, aumentando ou diminuindo a polaridade na membrana pós-sináptica, o que determina o funcionamento, comportamento e raciocínio humano. Outra característica das redes neurais naturais é que elas não transmitem sinais negativos, sendo a sua ativação medida pela frequência com que emite pulsos contínuos e sempre positivos, ou seja, seus pulsos não são síncronos ou assíncronos, devido ao fato de não serem contínuos, o que a difere de redes neurais artificiais.

Pesquisas desenvolvidas baseadas nas principais características das redes neurais naturais (sendo possível citar: não linearidade, mapeamento entrada-saída, adaptabilidade, respostas a evidências, informação contextual, tolerância a falhas, implementação em VLSI, uniformidade de análise e projeto e analogia neurobiológica) permitiram o desenvolvimento de modelos matemáticos que definem a base estrutural das redes neurais artificiais.

É evidente que uma rede neural extrai seu poder computacional, primeiro, de sua estrutura maciçamente paralelamente distribuída e, segundo, de sua habilidade de aprender e, portanto, de generalizar. A generalização se refere ao fato de a rede neural produzir saídas adequadas para entradas que não estavam presentes durante o treinamento (aprendizado) (Haykin, 2001).

3.1 Histórico das Redes Neurais Artificiais

Os primeiros estudos de que se tem conhecimento datam dos anos 1940, quando o neurofisiologista, filósofo e poeta americano Warren McCulloch (1960) e o lógico Walter Pitts (1960) desenvolveram o primeiro modelo matemático de um neurônio (McCulloch, 1960), mostrado na equação A.1.:

$$y = \Phi\left(\sum_{i=1}^{n} w_i \cdot x_i - b\right) \qquad \text{(A.1)}$$

A função (Φ) é denominada de função de ativação e, neste modelo, é uma função limiar simples. As entradas (x_i) chegam ao neurônio através dos pesos das conexões (w_i). A função de ativação também leva em consideração um termo de polarização ou *bias* (b), valor abaixo do qual a saída é nula.

A rede neural artificial utiliza processadores paralelamente distribuídos e possuem características congruentes com os neurônios biológicos, evidenciando:

- Múltiplas entradas e apenas uma única saída;

- Ação excitatória ou inibitória dos sinais de entrada refletidos no sinal de saída;
- Limiar de ação dos sinais de entrada refletidos no sinal de saída.

A figura 7 ilustra esquematicamente o modelo desenvolvido por McCulloch e Pitts (1960).

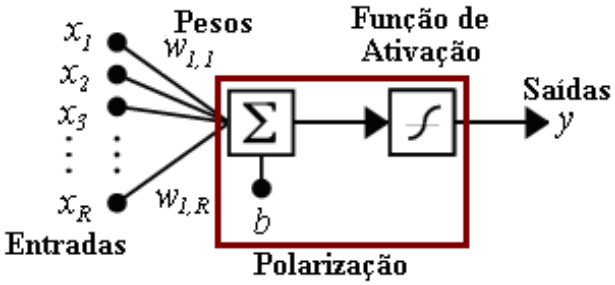


Figura 7 - Ilustra o Modelo Não Linear de um Neurônio (Haykin, 2001)

Em 1958 Rosenblatt (1969) realizou pesquisas sobre redes neurais artificiais e seus resultados apresentaram a criação da rede *perceptron*, um modelo cognitivo que consistia de unidades sensoriais conectadas a uma única camada dos neurônios de McCulloch e Pitts. Essa nova rede teve como característica a capacidade de aprender tudo o que pudesse representar. Rosenblatt demonstrou que acrescentando sinapses ajustáveis, as redes neurais de McCulloch e Pitts poderiam ser treinadas para classificar padrões em classes linearmente separáveis, convergindo em um número limitado de passos (Másson & Wang1990).

- A representação significa a habilidade do sistema neural em simular uma função específica;
- O aprendizado refere-se à existência de procedimentos sistemáticos de aquisição de conhecimento, segundo os quais os pesos são ajustados de forma a produzir a função desejada.

Em 1960 Minsky & Papert (1988) fizeram novas descobertas observando que as limitações das redes perceptrons desenvolvida por Rosenblatt (1969), suas capacidade de representação, que se referem ao OU-Exclusivo, provando que tais redes não são capazes de resolver uma ampla classe de problemas, uma vez que a rede perceptron divide o plano de trabalho em dois hemisférios. Isso causou um desinteresse em novas pesquisas sobre Redes Neurais.

Nos anos 1980 Boltzmann (1974) observou o modelo de aprendizado das redes perceptrons, redes de multicamadas e teorias adaptativas, e com isso

demonstrou a diferença entre as redes neurais perceptrons e os modelos ADALINE.

Hinton e Seynowsky, em 1983, estenderam o modelo desenvolvido por Boltzmann e incorporaram a dinâmica estocástica. Este modelo de rede neural passou a ser conhecido como Máquina de Boltzmann (Haykin, 1994) e as redes com várias camadas.

Anos mais tarde, Rumelhart, Hinton e Williams aperfeiçoaram a idéia da rede perceptron e desenvolveram o algoritmo *Backpropagation* (Rumelhart,1986), retomando o interesse nos estudos das Redes Neurais. Este algoritmo foi aplicado em grande variedade de problemas, como na identificação da estrutura de proteínas, hifenização de palavras em inglês, reconhecimento da fala, compressão de imagens e previsão de séries temporais (Másson & Wang1990). O sucesso deste algoritmo estimulou o desenvolvimento de muitas pesquisas em redes neurais artificiais e de uma variedade de modelos cognitivos.

O último modelo de destaque neste período foi o ART (*Adaptive Ressonance Theroy*) criado por Gail Carpenter e Stephen Grossberg (Carpenter, 1993). Este modelo possui aprendizado não supervisionado, criando *clusters* dos padrões aprendidos. O modelo ART teve diversas versões posteriores, entre elas versões do tipo semi-supervisionado e com uso de conceitos da lógica nebulosa Fuzzy-ART.

Os estudos sobre as redes neurais sofreram uma grande revolução a partir dos anos 1980, se destacando com promissoras características apresentadas pelos modelos de redes neurais.

3.2 O Neurônio Artificial

Para Haykin (2001), uma rede neural assemelha-se ao cérebro humano em dois aspectos:

1. O conhecimento é adquirido pela rede através de um processo de aprendizagem;
2. Forças de conexão entre neurônios (os pesos sinápticos) são utilizadas para armazenar o conhecimento adquirido.

Assim, as redes neurais artificiais são constituídas por conjuntos de unidades de processamento conectadas entre si, denominadas neurônios artificiais e ilustradas na Figura 8.

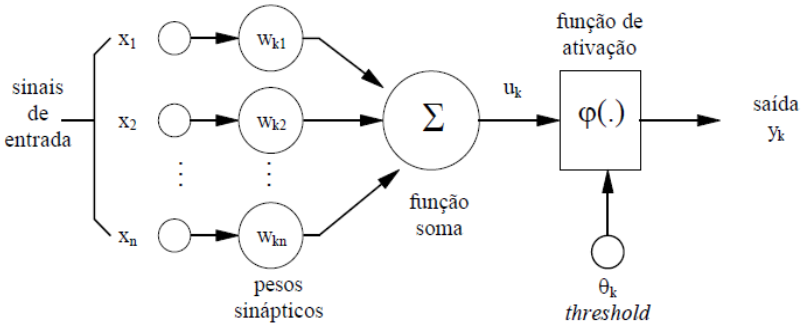


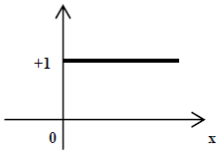
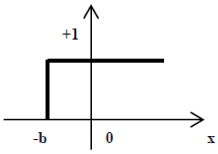
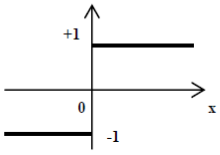
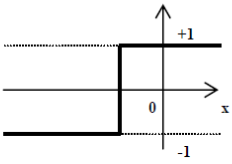
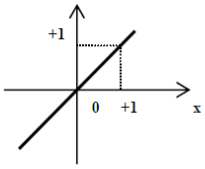
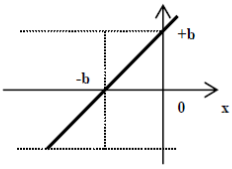
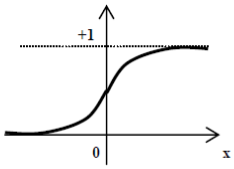
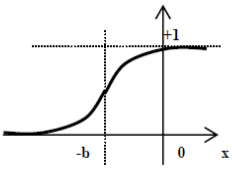
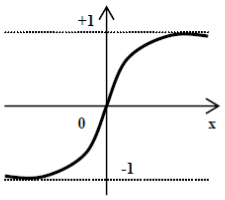
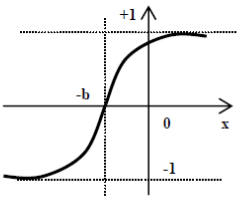
Figura 8 - Modelo Não Linear de um Neurônio (Haykin, 1994)

Cada neurônio artificial calcula a soma das entradas (no exemplo da Figura 3, x_1, x_2, x_3) ponderadas pelos pesos sinápticos (w_{k1}, w_{k2}, w_{k3}), além de um valor constante denominado *bias ou threshold* (θ_k). O valor encontrado é fornecido a uma função não-linear denominada função de ativação ($\varphi(.)$), que por sua vez gera o valor de saída y_k do neurônio, que será fornecido a outros neurônios da rede neural artificial.

Com objetivo de restringir a amplitude da saída do neurônio é empregada uma função restritiva, que restringe o intervalo permissível de amplitude do sinal de saída a um valor finito. Normalmente o intervalo normalizado da amplitude da saída de um neurônio deve estar contido no intervalo unitário fechado em [0,1] ou em [-1,1].

Há vários tipos de funções de ativação sendo suas características definidas conforme a Tabela 8, que define as funções mais usuais com as respectivas expressões.

Tabela 8 - Funções de Ativação

Função	Equação (com polarização)	Gráfico (sem polarização)	Gráfico (com polarização)
Degrau	$y = \begin{cases} 1 , x > -b \\ 0 , x < -b \end{cases}$		
Degrau Simétrico	$y = \begin{cases} 1 , x > -b \\ -1 , x < -b \end{cases}$		
Linear	$y = x + b$		
Logística Sigmoidal	$y = \frac{1}{1+e^{-(n+b)}}$		
Tangente Sigmoidal	$y = \frac{e^{(x+b)} - e^{-(x+b)}}{e^{(x+b)} + e^{-(x+b)}}$		

3.3 Topologia da Rede Perceptrons de Múltiplas Camadas – MLP

Como descrito anteriormente, a capacidade limitada de representação das Redes Neurais Artificiais com uma única camada (Minsk & Papert, 1988) foi superada na década dos anos 80, a partir do perfeito entendimento do problema da separabilidade linear (Rumelhart,1986).

A organização dos neurônios é definida como a topologia da rede. A topologia afeta o desempenho e as aplicações da rede. Sua estrutura está intimamente ligada ao algoritmo de aprendizado usado na fase de treinamento. Algumas redes permitem conexões tanto no sentido entrada-saída quanto no sentido saída entrada. Outras permitem que os neurônios da mesma camada estejam conectados entre si. Ainda há redes que permitem que o neurônio envie sinais de volta para ele mesmo (Tubb, 1993).

O desenvolvimento de novas redes com o acréscimo de novas camadas foi denominado de redes MLP (*Multilayer Perceptron*) formada por uma camada de entrada, uma ou mais camadas intermediárias e uma camada de saída, sendo esta normalmente a arquitetura mais encontrada.

As MLP (*Multilayer Perceptron*) possuem diferentes camadas (*layers*) intermediárias, trabalhando em uma forma acíclica (*feedforward*). A presença das camadas intermediárias nos modelos de redes MLP permite a solução de problemas mais complexos, sabendo-se que uma camada intermediária é suficiente para aproximar qualquer função contínua e que duas camadas intermediárias são suficientes para aproximar qualquer função matemática, sendo que outra característica das MLP é capacidade de prover soluções para problemas não lineares.

Assim, uma rede neural artificial é um sistema de neurônios artificiais ligados por conexões sinápticas e divididos em neurônios de entrada, que recebem as variáveis, neurônios internos ou *hidden* (ocultos) e neurônios de saída, que fornecem o resultado da rede. Esta forma de arranjo dos neurônios em camadas é denominado *Multilayer Perceptron*. O *multilayer perceptron* foi desenvolvido com o objetivo de solucionar os problemas mais complexos, até então não resolvidos pelo modelo de neurônio básico, como falta de linearidade.

Os neurônios internos são de suma importância na rede neural, pois provou-se que sem estes torna-se impossível a resolução de problemas linearmente não separáveis.

3.4 Aprendizado

Assim como as características humanas, as redes neurais artificiais têm a habilidade de aprender a partir de seu próprio ambiente e com isso melhorar seu desempenho. Isto ocorre por meio de um processo iterativo de ajuste aplicado aos pesos de entrada da rede e o treinamento. O aprendizado ocorre quando a rede neural atinge uma solução generalizada para uma classe de problemas.

O processo de aprendizado nas redes neurais acontece, internamente, por meio do ajuste dos pesos sinápticos das conexões durante a exposição dos exemplos, em resposta à quantidade de erros gerados pela rede. Ou seja, a rede neural é capaz de modificar-se em função da necessidade de aprender a informação que lhe foi apresentada (Tafner, 1998).

Conforme Mello (2004), os passos para o treinamento da rede, listados, são demonstrados no fluxograma da Figura 9.

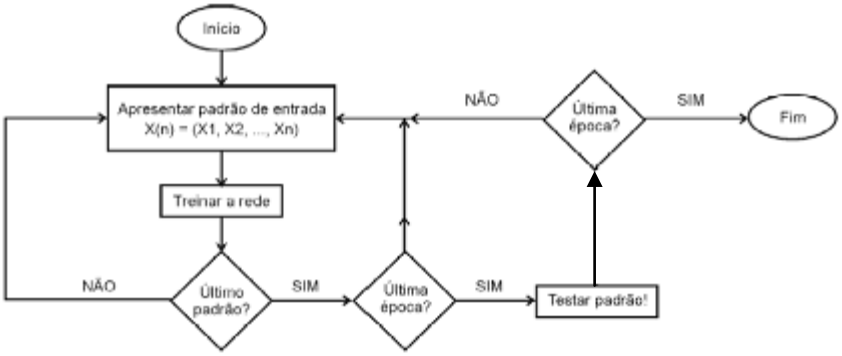


Figura 9– Fluxograma do Processo de Treinamento da Rede

(Mello, 2004)

As redes neurais artificiais possuem vários algoritmos de aprendizado, dentre eles: Regra de Hebb, Perceptron, Delta Rule, Backpropagation, Competitive Learning. Esta dissertação empregou rede neurais do tipo *multi-layer perceptron*, implementadas em Matlab, treinadas por um algoritmo derivado do *backpropagation* e denominado Levenberg-Marquardt (1981).

Assim, o algoritmo de aprendizado se dá por meio dos passos abaixo citados:

1. Iniciar os pesos e bias com valores aleatórios entre -0,5 e 0,5 ou entre -1 e 1, evitando que os pesos iniciais sejam muito grandes e que as derivadas das funções de ativação sejam muito próximas a zero, o que tornaria o aprendizado muito lento;
2. Aplicar um padrão de entrada, com seu respectivo valor desejado de saída (tj) e verificar a saída da rede (sj);
3. Calcular o erro na saída, dado geralmente por $ej = tj - sj$;
4. Definir modelo de decisão onde $ej = 0$ voltar ao passo 2, se $ej \neq 0$;
5. Atualizar os pesos sinápticos, por meio de algum algoritmo que calcule a variação do erro da rede em relação aos valores dos pesos;
6. Retornar ao passo 2 até que a rede esteja treinada.

Para a compreensão do processo de aprendizado, dois conceitos são importantes: o número de ciclos ou épocas e a taxa de aprendizado. O número de ciclos refere-se ao número de vezes que os padrões de treinamento serão

apresentados às redes neurais, a fim de que se faça a atualização dos pesos. A taxa de aprendizado controla a intensidade das alterações dos pesos.

O aprendizado, por sua vez, define a maneira como a rede se relaciona com o ambiente e se dividem em três grupos principais:

1. Supervisionado - a rede apresenta, na fase de treinamento, um conjunto de entradas acompanhadas de suas respectivas saídas. O objetivo é minimizar o sinal de erro, que é uma função da diferença entre a saída desejada e aquela fornecida pela rede. Esta minimização se dá pelo ajuste dos pesos da rede. Um exemplo deste método de aprendizado é o *backpropagation*;
2. Não supervisionado - a rede aprende sozinha, sem uma supervisão externa. É necessário que entradas parecidas sejam apresentadas à rede, para que esta possa extrair características estatisticamente relevantes e criar classes de maneira automática;
3. Híbrido - consiste de uma combinação dos aprendizados supervisionado e não supervisionado. Um exemplo é o aprendizado por reforço, onde a rede aprende de seu próprio ambiente, a partir dos dados de entrada.

3.5 Função Analítica de uma Rede Neural

Como descrito anteriormente, as redes Multilayer Perceptron são cíclicas (*feedforward*), com uma ou mais camadas intermediárias, e constituem os modelos de redes neurais artificiais mais utilizados. Tipicamente, a arquitetura do tipo *perceptron* de múltiplas camadas é composta por um conjunto de unidades sensoriais que formam uma camada de entrada, uma ou mais camadas intermediárias ou escondidas de unidades computacionais e uma camada de saída. Os sinais de entrada são propagados camada a camada pela rede em uma direção positiva, ou seja, da entrada para a saída, conforme demonstrado na figura 10 para uma rede neural com duas variáveis de entrada, uma camada escondida com dois neurônios e uma saída.

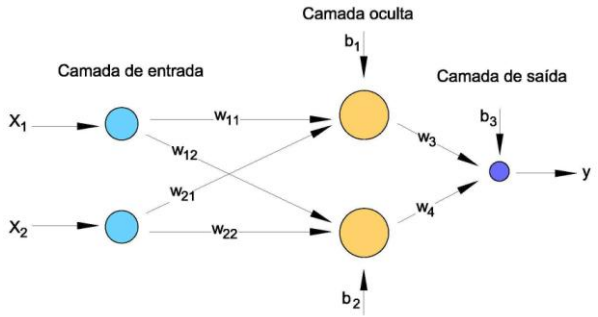


Figura 10 – Rede Neural Modelo Multi-Layer Perceptron

Caracterizando a Rede, tem-se que:

1. As sinapses caracterizadas por um peso w representam a sua intensidade. O papel do peso w_{kj} é multiplicar o sinal X_j na entrada da sinapse *j*, conectada a um neurônio *k*. O peso w_{kj} é positivo se a sinapse associada é excitatória e negativo se a sinapse associada é inibitória;
2. Um somatório adiciona as variáveis de entradas ponderadas pelos seus pesos respectivos conforme A.2, ou seja:

$$u_k = \sum_{i=1}^{n} w_{ij} \, . \, X_j \qquad \text{(A.2)}$$

3. Um limiar (threshold) b_n, que é subtraído do somatório acima;
4. Uma função de ativação, que limita a amplitude da saída do neurônio, ou seja, a entrada é normalizada dentro de um intervalo fechado [0,1] ou [-1,1];
5. A saída do neurônio y_k , onde:

$$y_k = \varphi(u_k - b_n) \qquad \text{(A.3)}$$

onde φ é a função de ativação.

Em geral, o valor do *threshold* é aplicado com a inclusão de uma entrada x0 igual a -1 e um peso wk0 igual ao valor de b_n. Portanto, a nova entrada da função de ativação, já incluindo o limiar (A.4), é dada por:

$$X_k = \sum_{j=0}^{n} w_{kj} \, . x_j - b_n \qquad \text{(A.4)}$$

Supondo agora uma rede neural conforme a figura 10, com duas entradas e dois neurônios na camada escondida, pode-se escrever que esta rede como um todo implementa uma função.

$$y = f\,(x_1\, x_2) \tag{A.5}$$

$$y = f\left(x_1, x_2\right) = \varphi\{b_3 + w_3 \cdot \varphi\left[b_1 + w_{11} \cdot x_1 + w_{21} \cdot x_2\right] + w_4 \cdot \varphi\left[b_2 + w_{12} \cdot x_1 + w_{22} \cdot x_2\right]\} \tag{A.6}$$