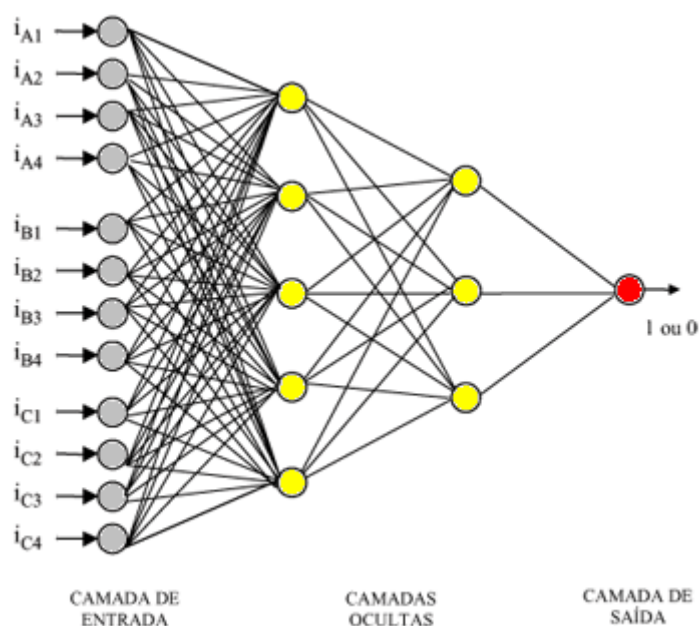
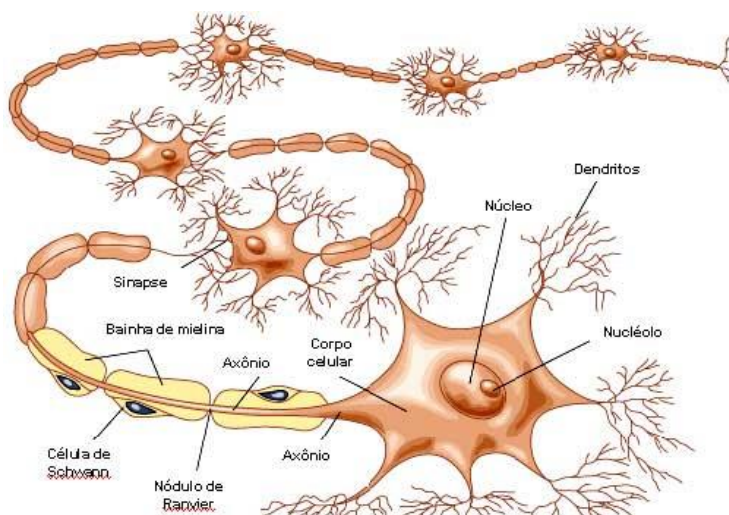


UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ  
CAMPUS CORNÉLIO PROCÓPIO  
COORDENAÇÃO DE ELETROTÉCNICA - ENGENHARIA ELÉTRICA  
LABORATORIO DE SEGURANÇA ILUMINAÇÃO E EFICIÊNCIA ENERGÉTICA



## NOÇÕES DE REDES NEURAIS ARTIFICIAIS

Prof. Me. Marco Antonio Ferreira Finocchio

Julho de 2014

## NOTA DO PROFESSOR

Esta apostila é de apoio didático utilizada para fornecer informações básicas de Redes Neurais Artificiais.

Este material não tem a pretensão de esgotar, tão pouco inovar o tratamento do conteúdo aqui abordado, mas, simplesmente, facilitar a dinâmica do estudo do tema, com expressivo ganho de tempo e de compreensão do assunto por parte dos interessados. A complementação do tema ocorrerá através de exemplificações, notas de aula, trabalhos e discussões.

Este trabalho se baseia nas referências, citadas na bibliografia, nos apontamentos de aula e na experiência do autor na abordagem do assunto. Sendo esta experiência baseada na atuação do autor aplicada aos transformadores de energia elétrica. Por se tratar de um material didático elaborado em uma Instituição Pública de Ensino, é permitida a reprodução do texto, desde que devidamente citada a fonte.

Quaisquer contribuições e críticas construtivas a este trabalho serão bem-vindas.

*“Lauda parce et vitupera parcius”.*

*Louva com moderação e censura com mais moderação ainda.*

*“In nomine XPI vicas semper”.*

Em nome de Cristo vencerás sempre.

**Prof. Marco Antonio Ferreira Finocchio**

**mafinocchio@utfpr.edu.br**

**ÍNDICE:**

<b>01. INTRODUÇÃO</b>	<b>04</b>
<b>02. TOPOLOGIA DE REDES NEURAIS ARTIFICIAIS</b>	<b>17</b>
<b>03. DETERMINAÇÃO DO NÚMERO DE NEURÔNIOS</b>	<b>27</b>
<b>REFERÊNCIAS</b>	<b>29</b>

## 1. INTRODUÇÃO

Redes neurais artificiais (RNA's) são sistemas que mimetizam o comportamento e a estrutura do cérebro humano, todavia, possuindo um conjunto muito limitado de neurônios. Esses neurônios, por sua vez, processam paralelamente os dados e os propagam através de uma complexa malha de interconexão. Analogamente ao cérebro humano, as RNA's têm a capacidade de interagir com o meio externo e adaptar-se a ele. Essas características conferem as RNA's uma importância multidisciplinar, razão pela qual essa ferramenta vem ganhando destaque em diferentes áreas do conhecimento, tais como engenharia, matemática, física, informática, etc. Particularmente, na área de transformadores, as RNA's vem sendo aplicadas na previsão de manutenção, no controle e monitoramento da operação e, como nos casos desse trabalho e outros, na modelagem.

De acordo com diversas estruturas neurais e algoritmos de aprendizagem propostos por vários pesquisadores, as redes neurais possuem certas características exclusivas de sistemas biológicos. Sistemas de computação baseados em redes neurais têm a capacidade de receber ao mesmo tempo várias entradas e distribuí-las de maneira organizada. Geralmente, as informações armazenadas por uma rede neural são compartilhadas por todas as suas unidades de processamento. Essa característica contrasta com os atuais esquemas de memória dos computadores convencionais, onde a informação fica confinada em um determinado endereço.

Esse capítulo traz alguns fundamentos e propriedades das RNA's, importantes no desenvolvimento de seus trabalhos, começando por um breve histórico. Não existe a preocupação de ir além do estritamente necessário para a compreensão da aplicação desta ferramenta em Engenharia.

### 1.1 HISTÓRICO

Nesta seção é apresentado um histórico da evolução do estudo de RNA's. O tema é abordado considerando-se uma sequência cronológica das atividades científicas de relevância no processo de evolutivo da técnica.

Esse processo é marcado por um longo período de estagnação e seguido por um recente e crescente interesse científico, devido às inovações tecnológicas e também pela característica multidisciplinar das RNA's.

Os primeiros documentos mencionados sobre redes neurais ou neurocomputação datam de 1943 com McCulloch e Pitts que projetaram a estrutura que é conhecida como a primeira rede neural. Estes pesquisadores propuseram um modelo de neurônio como uma unidade de processamento binária e provaram que esta unidade é capaz de executar muitas operações lógicas. Este modelo, apesar de muito simples, trouxe uma grande contribuição para as discussões sobre a construção dos primeiros computadores digitais, permitindo a criação dos primeiros modelos matemáticos de dispositivos artificiais que buscavam analogias biológicas (uma máquina baseada ou inspirada no cérebro humano). Muitos outros artigos e livros surgiram desde então, porém, pouco resultado foi obtido. Até que, em 1948, N. Wiener criou a palavra cibernética para descrever, de forma unificada, controle e comunicação nos organismos vivos e nas máquinas. Em 1949, Donald O. Hebb apresentou uma hipótese a respeito da maneira com que a força das sinapses no cérebro se altera em resposta à experiência. Em particular, ele sugeriu que as conexões entre células que são ativadas ao mesmo tempo tendem a se fortalecer, enquanto que as outras conexões tendem a se enfraquecer. Esta hipótese passou a influir decisivamente na evolução da teoria de aprendizagem em RNAs. Ele escreveu o livro intitulado "*The Organization of Behavior*" que perseguia a idéia de que o condicionamento psicológico clássico está presente em qualquer parte dos animais pelo fato de que isto é uma propriedade de neurônios individuais. Suas idéias não eram completamente novas, mas Hebb foi o primeiro a

propor uma lei de aprendizagem específica para as sinapses dos neurônios. Este primeiro passo serviu de inspiração para que muitos outros pesquisadores perseguissem a mesma idéia.

Mark I Perceptron foi o primeiro neurocomputador a obter sucesso e surgiu por volta de 1957. Ele foi criado por Frank Rosenblatt, Charles Wightman e outros. Rosenblatt introduziu uma nova abordagem para o problema de reconhecimento de padrões com o desenvolvimento do Perceptron.

Ele também propôs para o Perceptron um algoritmo para o ajuste dos pesos.

Devido à profundidade de seus estudos, suas contribuições técnicas e sua maneira moderna de pensar, muitos o vêem como o fundador da neurocomputação na forma em que encontra-se.

Por volta do mesmo período, Bernard Widrow e alguns colaboradores, desenvolveram um novo tipo de elemento de processamento de redes neurais chamado de Adaline (abreviação de Adaptive Linear Element), e mais tarde a sua generalização multidimensional, o Madaline (múltipla Adaline). Esta rede era equipada com uma nova lei de aprendizado, conhecida como "Regra Delta", que depois foi generalizada para redes com modelos neurais mais elaborados.

Infelizmente, apesar do sucesso do Perceptron e do Adaline, a pesquisa em redes neurais passou gradualmente a conviver com dois problemas fundamentais. Devido ao fato de a maior parte das pesquisas até então desenvolvidas ser de natureza heurística, o primeiro problema estava vinculado à carência de resultados teóricos que justificassem a manutenção do interesse científico pela área, o que ocasionou a redução na produção de novas idéias.

O segundo problema, e talvez o de maior significado histórico, foi a expectativa exagerada criada pelos próprios pesquisadores desta área, não acompanhada de resultados à altura, o que agravou a queda de financiamentos para pesquisa. Em 1969, L. M. Minsky e S. A. Papert publicaram o livro *"Perceptrons - An Introduction to Computational Geometry"*, onde os autores argumentaram enfaticamente quanto às limitações básicas das redes perceptrons isoladas, a começar pela impossibilidade de se implementar regras lógicas tão simples, como o caso do ou-exclusivo. Neste livro conceitos de matemática moderna, tais como, topologia e teoria de grupo, são aplicados com o objetivo de analisar as capacidades adaptativas e computacionais de neurônios. Os autores demonstraram que a rede perceptron, apesar de ser capaz de executar as operações booleanas AND e OR, não é capaz de executar outras operações elementares, como XOR (ou-exclusivo). Além disso, esses autores não acreditavam que uma arquitetura adequada em conjunto com um algoritmo de ajuste dos pesos pudesse ser desenvolvida de forma a superar esta limitação.

Historicamente, a publicação deste livro paralisou, a partir de 1970, as pesquisas na área de redes neurais. Os pesquisadores passaram a buscar por alternativas dentro do campo da engenharia e, principalmente, da lógica matemática, que, na época, encontrava-se em franca expansão, devido às grandes conquistas realizadas na área de computação.

Apesar do êxodo generalizado, um período de pesquisa silenciosa seguiu-se durante período de 1969 a 1982, quando poucos desenvolvimentos foram publicados devido aos fatos ocorridos anteriormente. Entretanto, nomes como Teuvo Kohonen, Stephen Grossberg, B. Widrow, James Anderson, Edoardo Caianiello, Kunuhito Fukushima, Igor Aleksander, entre outros, conseguiram novamente estabelecer um campo concreto para o renascimento da área. Eles passaram a publicar diversas propostas para o desenvolvimento e para as aplicações de redes neurais. Esta retomada do interesse pela exploração das RNA's deu-se por vários fatores, entre os quais ressaltam-se:

- os melhores conhecimentos da estrutura real do cérebro;
- a disponibilidade de computadores com maior capacidade de cálculo;
- o desenvolvimento de novos algoritmos de aprendizado.

Nos anos 80, muitos fatores contribuíram para o ressurgimento definitivo das pesquisas em redes neurais e, conseqüentemente, permitindo a sua aplicação em sistemas reais. Os mais importantes são:

- neurofisiologistas foram adquirindo um maior conhecimento sobre o processamento de informações nos organismos vivos;
- avanços tecnológicos tornaram disponível um maior potencial computacional a baixo custo, viabilizando ou facilitando simulações e testes com modelos neurais;
- novas teorias para a implementação de algoritmos foram desenvolvidas.

Neste período, John Hopfield, renomado físico de reputação mundial e ganhador do Prêmio Nobel, se interessou pela neurocomputação. O trabalho dele conquistou centenas de cientistas, matemáticos e engenheiros altamente qualificados para a pesquisa nesta área. Contudo, o fato que efetivamente colocou a área de redes neurais como uma das prioritárias na obtenção de recursos foi o desenvolvimento de um método para ajuste de parâmetros de redes não-recorrentes de múltiplas camadas. Este método, baseado em um algoritmo denominado *backpropagation*, tornou-se largamente conhecido após a publicação, em 1986, do livro “*Parallel Distributed Processing*”, por J. L. McClelland e D. E. Rumelhart, fazendo com que pesquisadores das mais diferentes áreas passassem a visualizar interessantes aplicações para redes neurais artificiais.

Desde então, várias aplicações tem sido mapeadas através de redes neurais artificiais, tais como: reconhecimento de padrões, processamento de imagens, sistemas de controle, robótica e identificação de sistemas. Na área específica de transformadores de potência, encontram-se vários trabalhos com aplicação de redes neurais, que mostram as diversas pesquisas que estão em desenvolvimento nesta área.

## 1.2 CÉREBRO HUMANO

O cérebro humano é responsável pelo controle das funções e dos movimentos do organismo. Por esta razão, ele é considerado o mais fascinante processador baseado em carbono existente. Ele é composto por aproximadamente 100 bilhões de neurônios. Cada neurônio está conectado a aproximadamente 100 outros através de sinapses, formando, juntos, uma grande rede, chamada rede neural biológica. Esta grande rede proporciona uma fabulosa capacidade de processamento e armazenamento de informação. As mais de 10 trilhões de sinapses transmitem estímulos através de diferentes concentrações de Sódio e Potássio, e os resultados destas podem ser estendidos por todo o corpo humano.

O sistema nervoso humano pode ser visto como um sistema com três estágios. São eles:

- o centro do sistema (cérebro), representado pela rede neural que recebe informações e toma decisões;
- os receptores, que convertem os estímulos do corpo ou do ambiente em impulsos elétricos que transmitem informação para a rede neural;
- os atuadores, que, por sua vez, convertem impulsos elétricos em respostas para a saída do sistema.

A rede neural é formada por um conjunto extremamente complexo de neurônios e, entre eles, a comunicação é realizada através de impulsos.

Quando um impulso é recebido, o neurônio o processa, e se o resultado do processamento for maior que um determinado limite de ação, o mesmo dispara um segundo impulso, que produz uma substância neurotransmissora e, este impulso flui do corpo celular para o axônio (que, por sua vez, 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. Os neurônios têm, portanto, um papel essencial no funcionamento, no comportamento e no raciocínio do ser humano.

As redes neurais naturais não transmitem sinais negativos, ao contrário das redes neurais artificiais, sua ativação é medida pela frequência com que os pulsos são emitidos. As redes naturais não são uniformes como as redes artificiais, e apresentam certa uniformidade apenas em alguns pontos do organismo. A figura 1 ilustra a aparência de uma célula nervosa (neurônio).

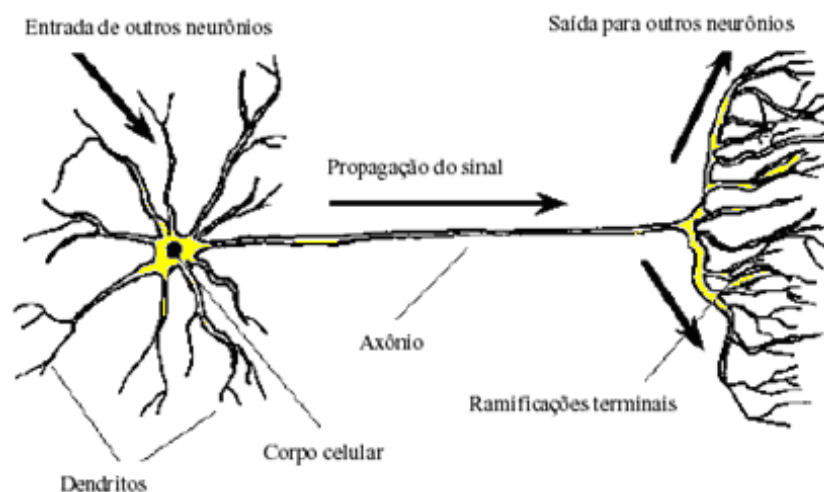


Figura 1: Principais componentes do neurônio

Os principais constituintes de um neurônio são:

- dendritos - responsáveis por receber os estímulos transmitidos pelos outros neurônios;
- corpo celular - responsável por coletar e combinar informações vindas de outros neurônios;
- axônio - responsável por transmitir os estímulos para outras células (constituído de uma fibra tubular que pode alcançar até alguns metros).

É um recurso muito aplicado em ciências exatas baseia-se no ajuste de modelos matemáticos a dados experimentais. O caso mais comum é o de regressão seja ela linear ou não, para uma única ou várias variáveis independentes.

Este último tipo de ajuste é de natureza estatística e resulta em expressão matemática (modelo) que permite fazer previsão para casos não experimentados na prática.

As redes neurais artificiais estão relacionadas a esse tipo de técnica, com aplicação a grande variedade de problemas práticos. As RNA's são expressões matemáticas com uma estrutura conveniente. Elas se ajustam melhor a dados experimentais e têm mais capacidade de reproduzi-los (previsão).

As células nervosas naturais possuem terminações (axônios e dendritos), sendo que substâncias denominadas neurotransmissores participam da transferência de sinais de natureza elétrica entre células.

Um sinal elétrico chega à célula nervosa percorrendo suas terminações conforme Figura 2. Estas podem atenuar ou amplificar o sinal. A fisiologia da célula também altera ainda mais o sinal. O sinal de saída modificado vai para outra célula.



Figura 2: Direção da transmissão de sinal no neurônio

Todos os estímulos do nosso ambiente causam, nos seres humanos, sensações como dor e calor. Os sentimentos, pensamentos, programação de respostas emocionais e motoras, causas de distúrbios mentais, e qualquer outra ação ou sensação do ser humano, não podem ser entendidas sem o conhecimento do processo de comunicação entre os neurônios.

Os neurônios são células especializadas, que junto com as células gliais formam o sistema nervoso.

Os neurônios são constituídos fundamentalmente por:

- Corpo Celular, ou também conhecido como soma, é a região onde está o núcleo e a maioria das estruturas citoplasmáticas;
- Dendritos (do grego *dendron*, que significa árvore) são prolongamentos geralmente muito ramificados. Eles são responsáveis pelo recebimento dos sinais;
- Axônio (do grego *axoon*, que significa eixo) só possui ramificações na extremidade. Ele é responsável por conduzir os sinais;
- Contatos Sinápticos, ou sinapses, são responsáveis por transmitir as informações de uma célula para outra célula.

A conexão entre neurônios é estabelecida pelas sinapses Figura 3.

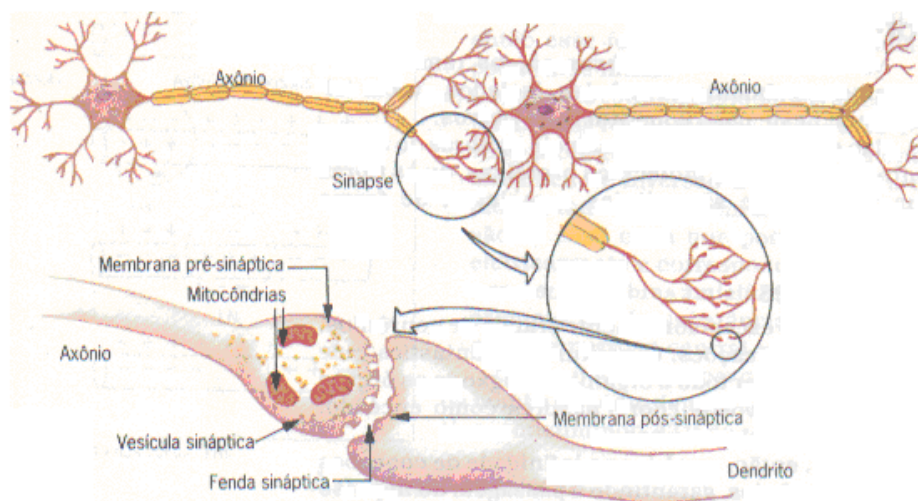


Figura 3: Transmissão de sinal entre neurônios

Além destas estruturas os neurônios possuem outras, devido o seu nível de especialização. Dentre elas podemos citar a bainha de mielina.

Alguns neurônios têm seus axônios envolvidos por um tipo celular denominado *célula de Schwann*. As células de Schwann determinam a formação da *bainha de mielina* - invólucro lipídico que atua como isolante térmico e facilita a transmissão do impulso nervoso. Entre uma célula de Schwann e outra existe uma região de descontinuidade da bainha de mielina,



denominada *nódulo de Ranvier*. A parte celular da bainha de mielina, onde estão o citoplasma e o núcleo da célula de Schwann, constitui o neurilema.

O percurso do impulso nervoso no neurônio é sempre no sentido

**Dendrito → Corpo Celular → Axônio**



Um impulso é transmitido de uma célula a outra através das sinapses (do grego *synapsis*, ação de juntar). A sinapse é uma região de contato muito próximo entre a extremidade do axônio de um neurônio e a superfície de outras células. Estas células podem ser tanto outros neurônios como células sensoriais, musculares ou glandulares.

As terminações de um axônio podem estabelecer muitas sinapses simultâneas. Na maioria das sinapses nervosas, as membranas das células que fazem sinapses estão muito próximas, mas não se tocam. Há um pequeno espaço entre as membranas celulares (o espaço sináptico ou fenda sináptica).

Quando os impulsos nervosos atingem as extremidades do axônio da célula pré-sináptica, ocorre liberação, nos espaços sinápticos, de substâncias químicas denominadas neurotransmissores ou mediadores químicos, que tem a capacidade de se combinar com receptores presentes na membrana das células pós-sináptica, desencadeando o impulso nervoso. Esse tipo de sinapse, por envolver a participação de mediadores químicos, é chamado *sinapse química*.

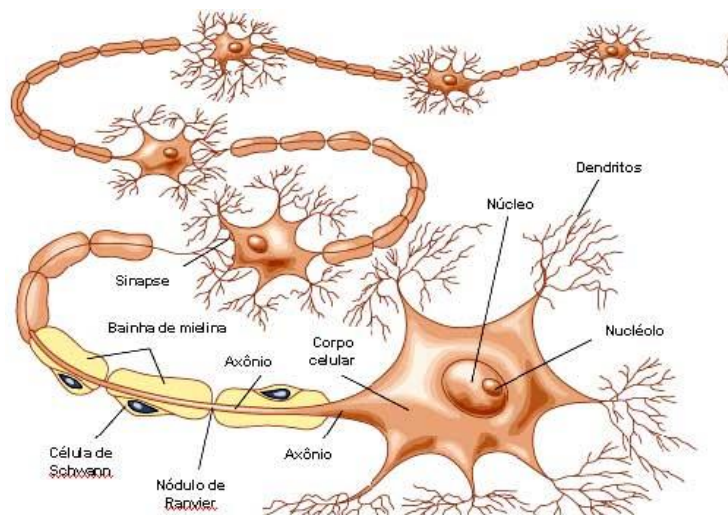


Figura 4: Transmissão de sinal entre neurônios

Os cientistas já identificaram mais de dez substâncias que atuam como neurotransmissores, como a acetilcolina, a adrenalina, a noradrenalina, a dopamina e a serotonina.

Existe também um outro tipo de sinapse, a *sinapse elétrica*, onde íons e pequenas moléculas passam por eles, conectando então canais de uma célula a próxima, de forma que alterações elétricas em uma célula são transmitidas quase instantaneamente à próxima.

As diferentes terminações no neurônio artificial modulam o sinal de entrada de modo diferente, conforme seu comprimento e natureza. Diferentes terminações pesam diferentemente a transmissão do sinal.

A bioquímica celular também altera a transmissão do sinal. Há padrões gerais definidos para transferência, em função de características gerais válidas para classes de células. Para uma mesma função de transferência, a alteração depende de particularidades da célula, ou seja, depende da função.

Expressando tais idéias em linguagem matemática, a intensidade do sinal de entrada ( $p$ ) é modulada com pesos ( $w$ -weight) e com uma constante ( $b$ -bias) nas terminações celulares:  $w \cdot p + b$ . Esse sinal transformado é, em seguida, alterado com uma função que depende da atividade típica da célula.

O resultado é um sinal de saída  $a$  que pode ser expresso por:

$$a = F(w \times p + b)$$

É uma lei de transmissão dos impulsos através de um neurônio.

Os sinais de um neurônio vão para um ou mais neurônios seguintes. Nestes, novamente há alteração do sinal com outros pesos, bias e uma função que dependem da natureza da célula.

Podem-se representar os neurônios e a transmissão de sinais elétricos através de um diagrama de blocos Figura 5.

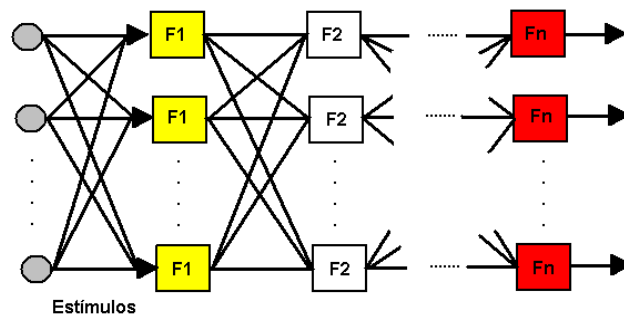


Figura 5: Representação em blocos de neurônios

Os estímulos de entrada vão sendo modificados de maneira a gerar finais que dependem da natureza do tecido nervoso.

Há dois contextos diferentes em que ocorre a transmissão de sinais entre neurônios, correspondendo a atividades distintas do tecido nervoso:

- Aprendizagem
- Utilização do conhecimento

## Aprendizagem

Os estímulos (dados de entrada) e sinais de saída (dados de saída) são simultaneamente submetidos aos neurônios.

Neste caso, o caminho da transmissão do sinal é ajustado: seleção das células participantes da transmissão do sinal. Isso corresponde a uma seleção de pesos, funções de transferência, quantidade e sequência de neurônios que melhor reproduza os dados de saída frente aos dados de entrada.

Esse é um caso de aprendizagem: a transmissão de sinal é cuidadosamente modulada de maneira a tornar possível a reprodução de percepções dos dados do mundo exterior. O tecido nervoso é treinado, sendo feito ajuste do caminho da transmissão de sinal de modo a se permitir reprodução dos dados da experiência diária (aprendizado por associação).

## Utilização do conhecimento

Nesse caso, um conjunto de neurônios que foi submetido a um aprendizado anterior é exposto a novos estímulos.

Tais estímulos de entrada, novos e diferentes dos utilizados para aprendizado, serão tratados pelos neurônios treinados. Resultam dados de saída correspondentes aos novos estímulos. Em última análise, os dados de saída são consequência do aprendizado anteriores.

Trata-se da utilização do tecido nervoso para fazer uma previsão de situações novas (dados de saída) correspondentes a novas questões (dados de entrada novos) com base em experiência adquirida.

Com as RNA's ocorrem algo semelhante:

- **Dados de ensino são submetidos às redes:**

As redes são treinadas com dados experimentais para aprendizagem. A qualidade dos dados é importante para o aprendizado, exercendo forte influência no desempenho das redes. O aprendizado é uma etapa prévia e consiste em ajustes de pesos e bias de redes cujas funções de transferência e estrutura de neurônios tenha sido pré-definidas.

Também no procedimento de ajuste de modelos matemáticos (regressão estatística) a dados experimentais há algo semelhante. Escolhe-se a forma da expressão matemática (reta, polinômio, ...) e prossegue com um ajuste de parâmetros. O critério de ajuste é o de minimizar a soma dos quadrados dos erros entre os dados experimentais e os calculados para variáveis de saída.

- **Uso da rede treinada:**

Uma vez treinada, a rede é utilizada para reprodução (previsão) de dados de saída correspondentes a novos dados de entrada.

Essa etapa é a do uso efetivo da correlação matemática para estimar novos dados de saída correspondentes a novas entradas.

- **Reaprendizado:**

Sempre que novos dados estejam disponíveis, o sistema nervoso pode ser submetido a um reaprendizado. Isso corresponde a um reajuste dos caminhos da propagação dos sinais através dos neurônios.

Em regressão estatística, essa etapa corresponde a reestimação de parâmetros para nova massa de dados.

Com redes neurais, o reaprendizado corresponde a novos treinamentos da rede, incluindo-se os novos dados experimentais.

No restante destas notas, detalham-se as estruturas que permitem a um ajuste matemático, seguir o padrão de funcionamento do sistema nervoso. Detalham-se as estruturas de algumas redes neurais.

## Estrutura dos Neurônios

Todas as redes neurais são compostas por elementos matemáticos denominados neurônios. O neurônio pode ser entendido a partir da Figura 6:

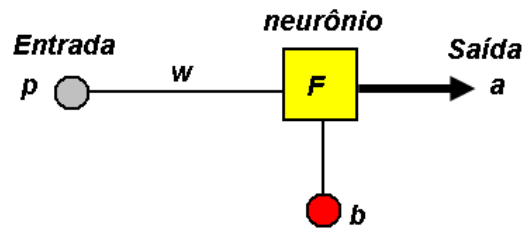


Figura 6: Esquema do neurônio

onde:

$p$  – entradas de sinal

$w$  – peso

$b$  – um termo linear (bias)

$F$  – as entradas são somadas e tratadas por uma função do neurônio, que gera a saída

$a$  – saída de sinal (número)

Em termos matemáticos:

$$a = F(w \cdot p + b)$$

O peso  $w$  e o bias são parâmetros ajustáveis (livres, variáveis) do neurônio e da rede neural.

### Função de Transferência do Neurônio

Uma característica do neurônio é a sua função  $F$ , que transforma os sinais combinados de entrada no sinal de saída. Há algumas funções mais comuns utilizadas:

- hard limit (degrau): se a saída vale 1 para entrada positiva e 0 ou -1 (simétrica) para entrada negativa;
- linear: saída proporcional ao valor de entrada (puramente linear);
- sigmoidal: quando a saída é uma sigmóide logarítmica ou tangencial.

Estas funções de transferência são também chamadas na literatura de função de ativação que serão observadas com detalhe na seção 2.1.

## 1.3 CAMADAS DE NEURÔNIOS COM VÁRIAS ENTRADAS

O modelo descrito para um neurônio pode receber mais de um sinal de entrada. Cada sinal  $p_i$  é combinado conforme mostra a figura 7.

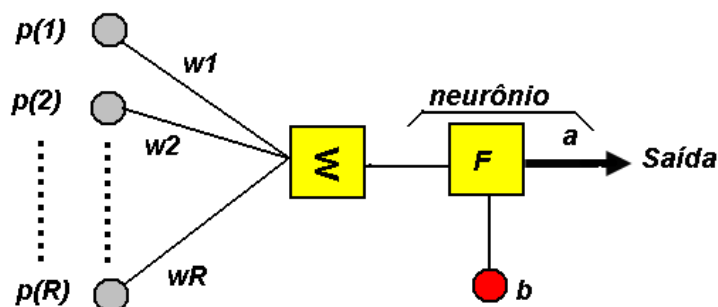


Figura 7: Neurônio recebendo R sinais de entrada mais bias

A descrição matemática para esse caso pode fazer uso de notação com vetores (matrizes), uma vez que:

$$a = F(w_1 \cdot p_1 + w_2 \cdot p_2 + \dots + w_g \cdot p_g + b) = F(\vec{w} \cdot \vec{p} + b)$$

onde:

$$\vec{w} = [w_1 \dots w_2 \dots \dots w_R]$$

$$\vec{p} = [p_1 \dots p_2 \dots \dots p_R]^T$$

Vários neurônios podem formar uma camada, cada um deles recebendo todas as mesmas R entradas descritas. Uma representação da camada de neurônios pode ser a seguinte:

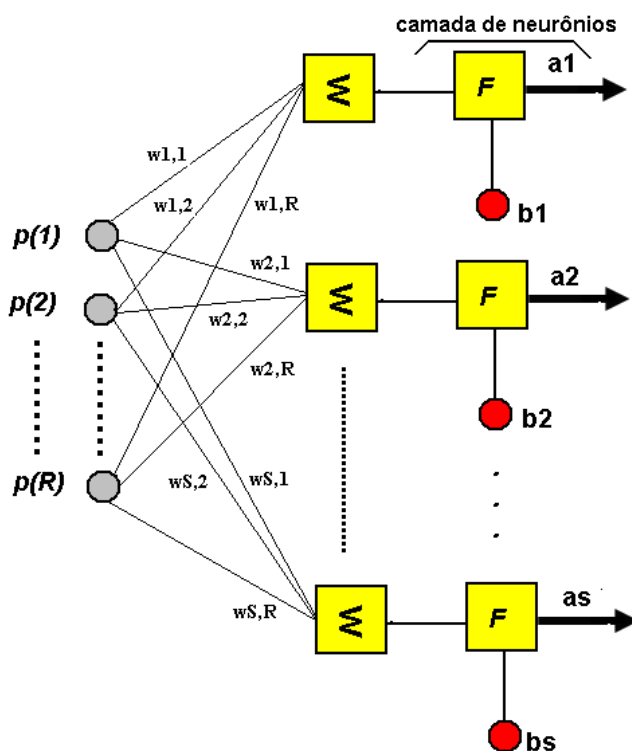


Figura 8: Camada de neurônio recebendo R sinais de entrada mais bias

Nesse caso, cada sinal de entrada está ligado a cada neurônio através de um peso. Sendo  $R$  sinais de entrada e  $S$  neurônios, serão  $R \times S$  pesos e  $S$  bias, que se agrupam em matriz de pesos e vetores de bias. Seu relacionamento com dados de entrada e saída (igualmente organizados em vetores) é expresso por:

$$\vec{a} = F(\underline{\underline{w}} \cdot \vec{p} + \vec{b})$$

$\underline{\underline{w}}$  matriz de pesos  
 $\vec{p}$ ,  $\vec{a}$ ,  $\vec{b}$  vetores: entradas, saídas e bias

com:

$$\underline{\underline{W}} = \begin{bmatrix} w_{1,1} & \dots & w_{1,2} & \dots & w_{1,R} \\ w_{2,1} & \dots & w_{2,2} & \dots & w_{2,R} \\ \vdots & & & & \\ w_{S,1} & \dots & w_{S,2} & \dots & w_{S,R} \end{bmatrix} \quad \vec{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_S \end{bmatrix} \quad \vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_S \end{bmatrix}$$

## 1.4 ESTRUTURA DA REDE NEURAL

As redes neurais são compostas por várias dessas camadas de neurônios, disposta em sequência como mostra a figura 9.

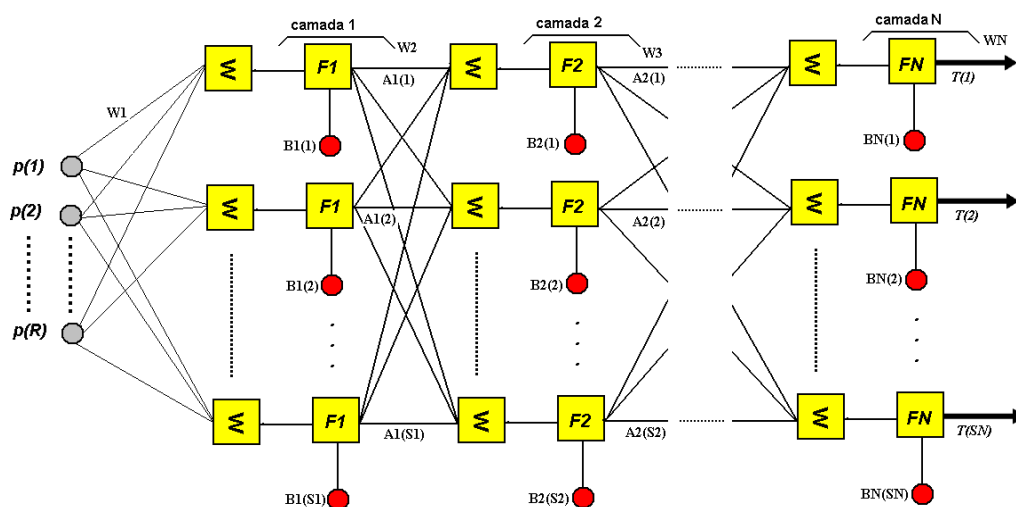


Figura 9: Rede neural com  $N$  camadas de  $S_i$  neurônios

Essa é uma rede neural com:

- $N$  camadas. Há  $S_i$  neurônios na camada  $i$  ( $i=1, 2, \dots, N$ );
- Diferentes funções de transferência em cada camada;
- $R$  entradas organizadas no vetor  $\vec{P}$ ;
- Vetores de saída  $A_i$  com dimensão  $S_i$  em cada camada. As saídas de uma camada são entradas para a camada seguinte; a saída calculada  $A_N$  da última camada é saída da rede, desejando-se que seja muito próxima a dados disponíveis  $\vec{T}$  (target);
- Matrizes de pesos  $\underline{\underline{W}}$  e vetores de vieses  $\vec{B}_i$  para cada camada.

Uma representação simplificada da rede pode ser a seguinte:

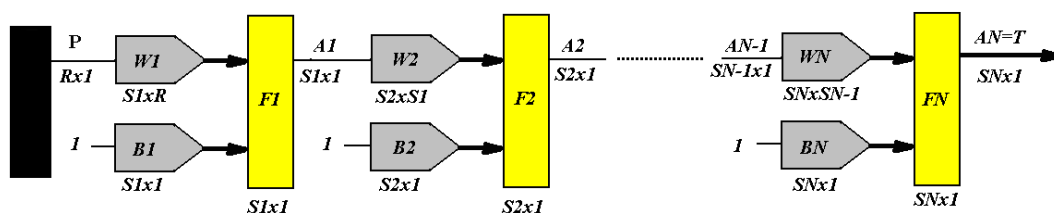


Figura 10: Representação simplificada de rede neural

Essa é uma estrutura típica de rede neural. A rede envolve dados de entrada que se propagam até uma saída, através de uma sucessão de aplicações de pesos, bias e funções de transferência.

Para uso prático de redes neurais, basta interpreta-las como estruturas matemáticas de matrizes. As redes neurais são seqüências de multiplicações e somas de matrizes e aplicações de certas funções de transferência.

A rede neural como estrutura matemática de matrizes:

$$\begin{aligned}\overline{A}_1 &= F(\overrightarrow{W}_1 \cdot \overrightarrow{P} + B_1) \\ \overline{A}_2 &= F(\overrightarrow{W}_2 \cdot \overline{A}_1 + B_2) \\ &\vdots \\ \overline{T} &= F(\overrightarrow{W}_N \cdot \overline{A}_{N-1} + B_N)\end{aligned}$$

## 1.5 TREINAMENTO DAS REDES E REPETIÇÃO DOS DADOS DE ENTRADA

Para a aplicação mais comum de redes neurais, geralmente se conhecem os dados de entrada e os dados de saída. Na notação matricial, isso significa conhecer os vetores  $\mathbf{P}$  e  $\mathbf{T}$ . Os pesos e os bias para os diversos neurônios das diferentes camadas devem ser estimados da melhor maneira possível, de modo que a rede seja capaz de reproduzir os dados de saída com os de entrada (associação).

Observa-se, portanto, que o uso de redes neurais se assemelha bastante ao problema de ajuste estatístico de expressões matemáticas a dados experimentais (levantamento de correlações). Ou seja, um método matemático deve ser usado para estimar os parâmetros de uma correlação cuja forma é determinada pela estrutura da rede.

Antes de discutir como são esses métodos matemáticos, cabe lembrar um último aspecto da estrutura de redes neurais. Como use de redes está muito associado a dados experimentais, é comum trabalhar com várias amostras.

Na coleta de dados para ajuste, o vetor  $\mathbf{P}$  pode ser amostrado diversas vezes. Sendo  $Q$  o número de amostras, resulta a matriz  $\mathbf{P}$  de dimensões  $R \times Q$ . Cada uma das  $Q$  colunas da matriz  $\mathbf{P}$  correspondem a uma amostragem dos dados. Com isso, também a cada amostra.

Para uma rede neural com uma única camada, por exemplo, o efeito de várias amostragens das variáveis de entrada e de saída da rede pode ser expresso através da representação simplificada indicada na figura 11.

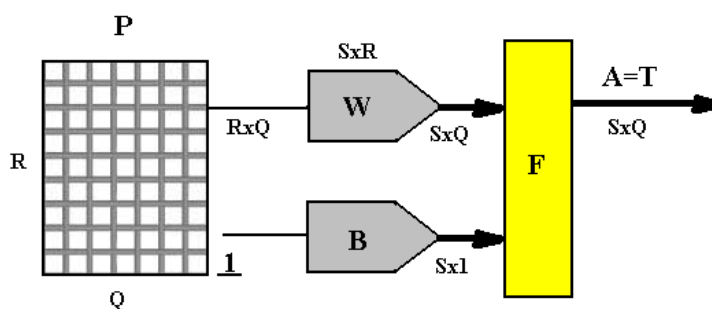


Figura 11: Representação simplificada de rede neural

O processo de ajuste da rede neural aos dados é denominado **treinamento da rede neural**.

Definida uma arquitetura para rede (número de camadas de neurônios, número neurônios em cada camada e funções de transferência de cada neurônio), seu treinamento começa com uma atribuição de estimativas iniciais para pesos e bias.

Seguem, então, iterações que incluem:

- Propagação dos dados de entrada através da rede (cálculos matriciais) até calculada;
- Comparação da saída calculada com a saída real (erro na saída);
- Aplicação de método matemático para atualizar os bias e pesos conforme o erro calculado.

Ajustar a rede significa minimizar o erro entre saída calculada e saída real. O erro mínimo alcançado através de reajustes das variáveis livres (pesos e bias dos neurônios). Cada execução do laço, cada iteração ao longo da pesquisa rumo ao mínimo (convergência) é denominada **época** de treinamento.

Os critérios matemáticos que determinam as modificações necessárias nos pesos e bias em função do erro de saída da rede neural são chamados **regras de aprendizado** da rede neural (*learning rule*). Verifica-se que eles são essencialmente, métodos de pesquisa do mínimo de uma função de varias variáveis.

A sequência para aplicação de redes neurais é a seguinte:

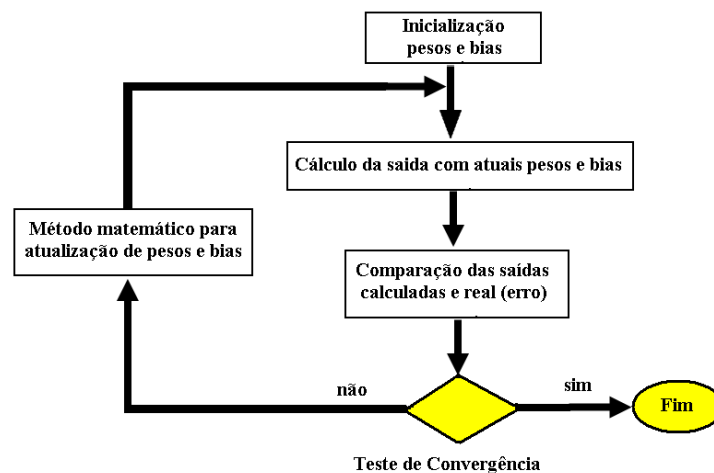


Figura 12: Fluxograma com a lógica de treinamento de rede neural

- Especificação do problema a ser resolvido e escolha da arquitetura de rede mais apropriada. No problema de ajuste, normalmente são empregadas redes com duas ou três camadas de neurônios, as primeiras com funções de transferência sigmoidais e a última com função de transferência linear. O número de neurônios em cada camada depende fortemente do problema a ser resolvido;
- Especificação do critério de convergência (precisão para a convergência). A rigidez desse critério depende da análise cuidadosa do problema e dos dados;
- A aplicação de método de pesquisa de mínimo erro na saída da rede. Normalmente, usa-se o método denominado **retropropagação** (*backpropagation*).

Existem várias arquiteturas, porém as mais utilizadas na análise de dados numéricos são:

- Perceptron
- Rede Neural não linear (*backpropagation*).



## 2. TOPOLOGIA DE REDES NEURAIS ARTIFICIAIS

As RNA's são técnicas computacionais que têm capacidade para solucionar problemas por intermédio de circuitos simples que simulam o funcionamento e o comportamento do cérebro humano. Elas apresentam um modelo inspirado na estrutura neural de organismos inteligentes, que adquirem conhecimento através da experiência, ou seja, aprendendo, errando e fazendo descobertas. Uma rede neural artificial pode ter centenas ou até milhares de unidades de processamento, enquanto o cérebro de um mamífero pode conter muitos bilhões de neurônios.

A operação de uma célula (neurônio) em uma rede neural, geralmente, pode ser descrita da seguinte forma:

- os sinais são apresentados à entrada;
- cada sinal é multiplicado por um peso, o qual indica sua influência na saída da célula;
- executa-se a soma ponderada dos sinais, o que produz um nível de atividade;
- quando este nível excede um limite (*threshold*), a unidade produz uma saída.

Assim como o sistema nervoso é composto por bilhões de células nervosas, a rede neural artificial também seria formada por unidades que nada mais são que pequenos módulos que simulam o funcionamento de um neurônio.

Estes módulos devem funcionar de acordo com os elementos em que foram inspirados, recebendo, processando e retransmitindo informações.

O modelo de neurônio mais simples, e que engloba as principais características de uma rede neural biológica, paralelismo e alta conectividade, foi proposto por McCulloch e Pitts. A Figura 13 mostra o modelo geral do neurônio artificial, onde:  $x_1, x_2, \dots, x_n$  representam os sinais de entrada;  $w_1, w_2, \dots, w_n$  são os pesos ou conexões sinápticas; o *bias* representa o limiar de ativação do neurônio;  $u$  é a saída do combinador linear;  $g(u)$  é a função de ativação (limita a saída do neurônio);  $y$  é o sinal de saída do neurônio.

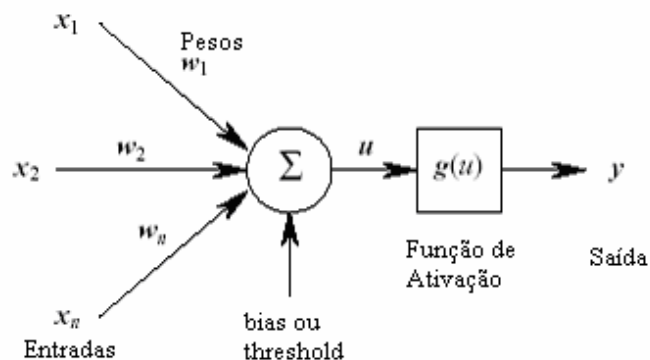


Figura 13: Modelo geral do neurônio artificial

As equações (2.1) e (2.2) descrevem o comportamento do neurônio artificial, onde  $x_i$  representa a  $i$ -ésima entrada do neurônio,  $w_i$  é o peso associado à entrada  $x_i$ ,  $n$  é o número de entradas,  $g(\ )$  é a função de ativação do neurônio e  $y$  é a saída dele.

$$u = \sum_{i=1}^n W_i X_i \quad (2.1)$$

$$y = g(u) \quad (2.2)$$

A maioria dos modelos de redes neurais possui alguma regra de treinamento, onde os pesos de suas conexões são ajustados de acordo com os padrões apresentados. Em outras palavras, elas aprendem através de exemplos (padrões).

## 2.1 FUNÇÕES DE ATIVAÇÃO

Na equação (2.2), a função de ativação processa o conjunto de entradas recebidas e o transforma em estado de ativação. Normalmente, o estado de ativação dos neurônios pode assumir:

- valores binários (0 e 1);
- valores bipolares (-1 e 1);
- valores reais.

As funções de ativação são escolhidas em função da necessidade do problema em que a rede esteja trabalhando. As mais típicas são:

### 2.1.1 Função linear

A função de ativação linear é mostrada na Figura 14 e representada pela equação (2.3).

$$g(u) = u \quad (2.3)$$

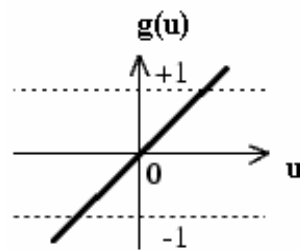


Figura 14: Função linear

### 2.1.2 Função degrau

Na função de ativação degrau (hard limit), mostrada na Figura 15 e representada pela equação (2.4), a saída do neurônio assumirá o valor 1 se o nível de atividade interna total do neurônio for não-negativo, caso contrário, a saída do neurônio assumirá o valor 0.

$$g(u) = \begin{cases} 1 & \text{se } u \geq 0 \\ 0 & \text{se } u < 0 \end{cases} \quad (2.4)$$

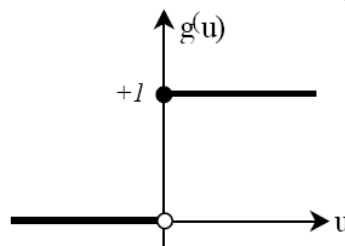


Figura 15: Função degrau

### 2.1.3 Função degrau bipolar

Na função de ativação degrau bipolar (hard limit-simetric), apresentada na Figura 16 e representada através da expressão (2.5), a saída do neurônio assumirá o valor 1 se o nível de atividade interna total do neurônio for não-negativo; caso contrário, a saída do neurônio assumirá o valor -1.

$$g(u) = \begin{cases} 1 & \text{se } u \geq 0 \\ -1 & \text{se } u < 0 \end{cases} \quad (2.5)$$

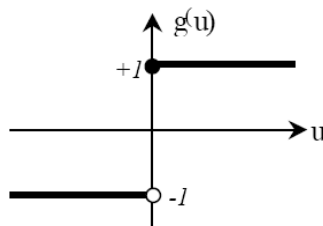


Figura 16: Função degrau bipolar

### 2.1.4 Função rampa

Nesse tipo de ativação (satlin), mostrada pela Figura 17 e expressa pela equação (2.6), a saída do neurônio pode assumir valores positivos e negativos no domínio de -1 a 1, e no intervalo  $\{-a, a\}$ , a saída assume o valor da função  $g(u)=u$ .

$$g(u) = \begin{cases} 1 & \text{se } u \geq a \\ u & \text{se } -a < u < a \\ -1 & \text{se } u \leq -a \end{cases} \quad (2.6)$$

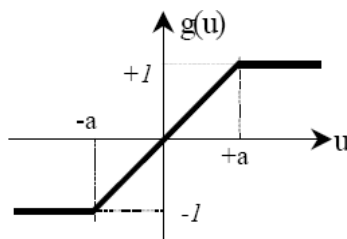


Figura 17: Função rampa

### 2.1.5 Função sigmóide

Para a função sigmóide (logsig) ilustrada na Figura 18 e apresentada pela expressão (2.7),  $\beta$  é o parâmetro que define a inclinação (ganho) da função sigmóide. Nesse tipo de função, a saída do neurônio assumirá valores reais entre 0 e 1.

$$g(u) = \frac{1}{1 + e^{(-\beta u)}} \quad (2.7)$$

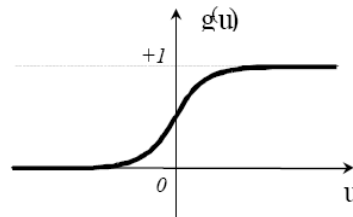


Figura 18 Função sigmóide

### 2.1.6 Função tangente hiperbólica

Na função de ativação do tipo tangente hiperbólica (tansig), apresentada na Figura 19 e representada pela equação (2.8), a saída do neurônio pode assumir valores reais negativos e positivos no domínio de -1 a 1.

$$g(u) = \tanh(u) = \frac{1 - e^{-u}}{1 + e^{-u}} \quad (2.8)$$

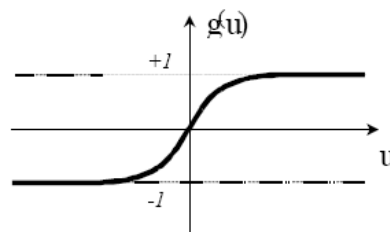


Figura 19 Função tangente hiperbólica

## 2.2 CAMADAS

Arquiteturas neurais são tipicamente organizadas em camadas, com unidades que podem estar conectadas às unidades da camada posterior.

Conforme apresentado na Figura 20, as camadas de uma rede neural são usualmente classificadas em três grupos:

- camada de entrada;
- camadas intermediárias ou ocultas;
- camada de saída.

A camada de entrada é a camada por onde os padrões são introduzidos à rede. As camadas intermediárias, escondidas ou ocultas, como também são conhecidas, são responsáveis pela maior parte do processamento e através das conexões ponderadas, elas podem ser consideradas como extratoras de características. A camada de saída é responsável pela apresentação dos resultados finais.

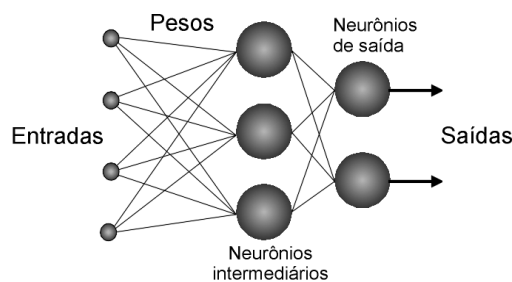


Figura 20: Organização da rede em camadas

## 2.3 ARQUITETURA

Do ponto de vista estrutural, a arquitetura de redes neurais pode ser classificada como estática, dinâmica ou *fuzzy*. Ela pode ser constituída ainda de única camada ou múltiplas camadas. Além disso, algumas diferenças computacionais surgem quando se trata da maneira com que são feitas as conexões existentes entres os neurônios. Estas conexões podem ser estritamente no sentido de ida, no sentido de ida e volta, lateralmente conectadas, topologicamente ordenadas ou híbridas.

As redes neurais são também classificadas de acordo com a arquitetura em que foram implementadas, topologias, características de seus nós, regras de treinamento e tipos de modelos de neurônio empregado. A seguir serão descritas as mais importantes.

### 2.3.1 Redes feedforward (camada única)

São redes neurais contendo apenas uma camada de entrada e uma camada de neurônios que é a própria camada de saída, como mostra a Figura 21. Suas principais aplicações são em memória associativa e no reconhecimento de padrões. O Perceptron e o Adaline são exemplos desse tipo de rede.

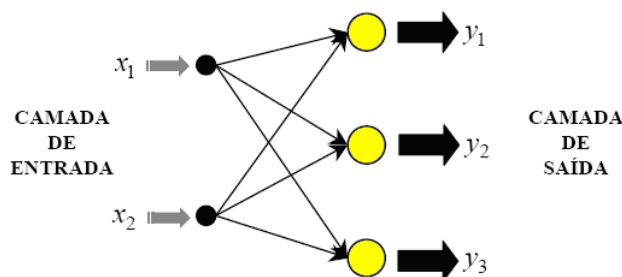


Figura 21: Exemplo de rede *feedforward* (camada única)

### 2.3.2 Redes feedforward (multicamadas)

Este tipo de rede distingue-se da anterior pela presença de uma ou mais camadas escondidas de neurônios. A Figura 22 mostra um exemplo de rede *feedforward* (multicamadas). Nesta figura:

- os neurônios que recebem sinais de excitação do meio externo estão na camada de entrada;
- os neurônios que estão na saída representam a camada de saída;
- os neurônios intermediários estão nas camadas escondidas.

Suas principais aplicações são em reconhecimento de padrões, aproximador universal de funções e em controle. O Madaline, o Perceptron Multicamadas e o de Função Base Radial são exemplos deste tipo de rede.

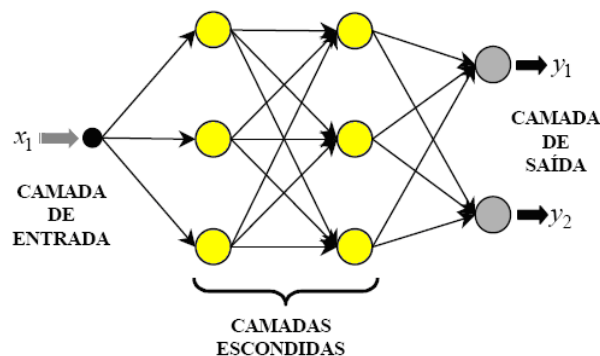


Figura 22: Exemplo de rede feedforward (multicamadas)

### 2.3.3 Redes recorrentes

A Figura 23 apresenta um exemplo de rede recorrente. Este tipo de rede contém realimentação entre neurônios de camadas diferentes. Suas principais aplicações são em sistemas dinâmicos, memória associativa, previsão e estimação, otimização e em controle. O modelo de Hopfield e o Perceptron com realimentação são exemplos desse tipo de rede.

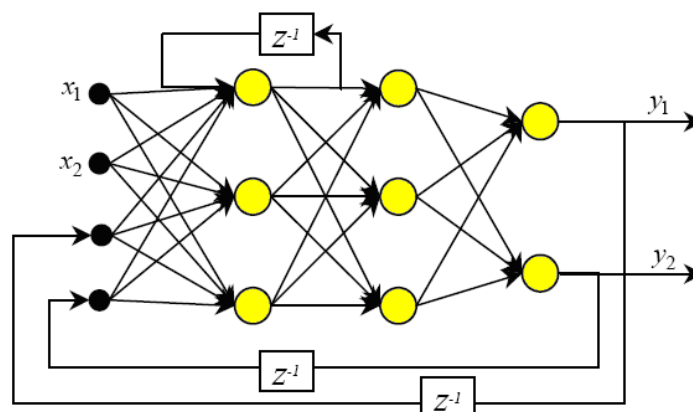


Figura 23: Exemplo de rede recorrente

## 2.4 TREINAMENTO DE REDES NEURAIS ARTIFICIAIS

O objetivo principal da pesquisa sobre redes neurais na área de computação é desenvolver morfologias neurais, baseadas em modelos matemáticos, que podem realizar funções diversas [22]. Na maior parte dos casos, modelos neurais são compostos de muitos elementos não lineares que operam em paralelo e que são classificados de acordo com padrões ligados à biologia.

A rede neural passa por um processo de treinamento com fundamentação nos casos reais conhecidos, adquirindo, a partir daí, a sistemática necessária para executar adequadamente o processamento desejado dos dados fornecidos.

Ela é, portanto, capaz de extrair regras básicas (conjunto de pesos) em consequência dos dados reais, diferindo, assim, da computação convencional, onde são necessários um conjunto de regras rígidas pré-fixadas e algoritmos.

A propriedade mais importante das redes neurais é a habilidade de aprender por intermédio de seu ambiente e com isso melhorar seu desempenho.

Isso é feito através de um processo iterativo de ajustes (treinamento) aplicado a seus pesos. O processo de aprendizado encerra-se quando a rede neural consegue generalizar soluções para uma classe de problemas.

Todo o conhecimento de uma rede neural está armazenado nas sinapses, ou seja, nos pesos atribuídos às conexões entre os neurônios. Cerca de 60 a 90% do total de dados devem ser

separados para o treinamento da rede neural, dados estes escolhidos aleatoriamente, a fim de que a rede "aprenda" as regras associadas ao processo. O restante dos dados só é apresentado à rede neural na fase de testes para verificar se as regras produzem saídas adequadas para os dados não utilizados no treinamento (testar o grau de generalização).

### **2.4.1 Aprendizado**

Um algoritmo de aprendizado é composto por um conjunto de regras bem definidas visando a solução de um problema de aprendizado. Existem muitos tipos de algoritmos de aprendizado específicos para determinados modelos de redes neurais. Estes algoritmos diferem entre si principalmente pelo modo como os pesos são modificados. As principais formas de aprendizado, que estão associadas aos processos de ajuste de pesos da rede, podem ser divididas em:

#### **2.4.1.1 Aprendizado supervisionado**

O aprendizado supervisionado utiliza um agente externo que indica à rede um comportamento adequado ou inadequado em conformidade com o padrão de entrada. A rede é treinada para fornecer a saída desejada em relação a um estímulo de entrada específico. Quando um vetor de entrada é aplicado, a saída da rede é calculada e comparada com o respectivo padrão de saída. A diferença ou erro é então propagada da saída para a entrada (em sentido inverso ao fluxo de informações da rede) e os pesos são alterados em conformidade com algum algoritmo que tende a minimizar esse erro ou diferença. Os vetores de entrada e saída do conjunto de treinamento são sequencialmente aplicados, as diferenças ou erros são calculados e, para cada vetor, os pesos são ajustados até o erro atingir um nível aceitável para o conjunto de treinamento.

#### **2.4.1.2 Aprendizado não supervisionado**

Essa forma de aprendizado dispensa um agente externo indicando a resposta desejada para os padrões de entrada, utiliza-se, entretanto, de exemplos de coisas análogas para que a rede responda de maneira similar. A rede se autoorganiza em relação a alguns subconjuntos de entrada cujos elementos possuem características similares. Os vetores do conjunto de treinamento consistem, unicamente, de vetores de entrada. O algoritmo de treinamento modifica os pesos da rede para produzir vetores de saída que são consistentes, isto é, vetores do conjunto de treinamento que são similares entre si produzirão o mesmo padrão de saída. Nesse tipo de aprendizagem, espera-se que o sistema deva, estatisticamente, descobrir características e particularidades marcantes da população de entrada. Ao contrário do aprendizado supervisionado, não existe um conjunto à priori de categorias dentro do qual os padrões irão ser classificados, mas sim o sistema é quem deve desenvolver sua própria representação do estímulo de entrada.

Ciclo de aprendizado (época) é a apresentação de todos os pares (entrada e saída) do conjunto de treinamento no processo de aprendizado. A correção dos pesos num ciclo pode ser executada de dois modos (métodos para apresentação dos dados para o treinamento):

- Modo Padrão, onde a correção dos pesos acontece a cada apresentação à rede de um exemplo do conjunto de treinamento. Cada correção de pesos baseia-se somente no erro do exemplo apresentado naquela iteração;
- Modo Batch, onde apenas uma correção é feita por ciclo. Todos os exemplos do conjunto de treinamento são apresentados à rede, o erro médio é calculado e, a partir deste erro, fazem-se correções nos pesos.

## 2.5 REDES PERCEPTRONS MULTICAMADAS (MLP)

Quando RNAs de uma só camada são utilizadas, os padrões de treinamento apresentados à entrada são mapeados diretamente em um conjunto de padrões de saída da rede, ou seja, não é possível a formação de uma representação interna. Neste caso, a codificação proveniente do mundo exterior deve ser suficiente para implementar esse mapeamento.

Tal restrição implica que padrões de entrada similares resultem em padrões de saída similares, o que leva o sistema à incapacidade de aprender importantes mapeamentos. Como resultado, padrões de entrada com estruturas similares, fornecidos pelo mundo externo, que levem à saídas diferentes, não são possíveis de serem mapeados por redes sem representações internas, isto é, sem camadas intermediárias. Um exemplo clássico deste caso é a função ou exclusivo (XOR). Redes de uma camada não são capazes de solucionar problemas que não sejam linearmente separáveis.

O desenvolvimento do algoritmo de treinamento *backpropagation* em 1986, mostrou que é possível treinar eficientemente redes com camadas intermediárias, resultando no modelo de RNAs mais utilizado atualmente, as redes Perceptron Multicamadas (MLP), treinadas com o algoritmo

*backpropagation*. Nessas redes, cada camada tem uma função específica. A camada de saída recebe os estímulos da camada intermediária e constrói o padrão que será a resposta. As camadas intermediárias funcionam como extratoras de características e nelas os pesos representam uma codificação das características apresentadas nos padrões de entrada. Este fato permite que a rede crie sua própria representação do problema.

São necessárias, no máximo, duas camadas intermediárias, com um número suficiente de unidades por camada, para se produzir quaisquer mapeamentos e é suficiente apenas uma camada intermediária para aproximar qualquer função contínua.

Neste trabalho, as redes MLPs, que constituem o sistema de projeto, têm a arquitetura “*multilayer feedforward*”, ou seja, o fluxo de informações é executado numa única direção, não possuindo retro-alimentação entre os neurônios.

## 2.6 PROCESSOS DE APRENDIZAGEM DE REDES PERCEPTRONS

Esta seção traz detalhes dos algoritmos de treinamento referentes aos processos de aprendizagem supervisionada das redes neurais.

### 2.6.1 Backpropagation

A estratégia de treinamento da rede, utilizando o algoritmo *backpropagation* (regra delta generalizada), é descrita através de uma sequência de ações:

- um padrão é apresentado à camada de entrada da rede;
- a atividade resultante é propagada pela rede, camada por camada, até que a resposta seja produzida pela camada de saída;
- a saída obtida é comparada à saída desejada para esse padrão particular;
- se a saída não estiver correta, o erro é calculado;
- o erro é então propagado a partir da camada de saída até a camada de entrada;
- os pesos das conexões das unidades das camadas internas vão sendo modificados conforme o erro é retro-propagado;



- o processo é repetido para todos os vetores de entrada da rede até que o erro quadrático médio das saídas da rede esteja num valor aceitável.

## 2.6.2 Método de Levenberg-Marquardt

Como descrito anteriormente o algoritmo *backpropagation* ajusta os valores das matrizes de pesos em função erro quadrático médio. Entretanto, a utilização desse algoritmo na prática tende a convergir muito lentamente, exigindo assim um elevado esforço computacional. Para contornar este problema várias técnicas de otimização têm sido incorporadas ao algoritmo *backpropagation* a fim de reduzir o seu tempo de convergência e diminuir o esforço computacional exigido pelo mesmo. Dentre as técnicas de otimização mais utilizadas para este propósito destaca-se o algoritmo de Levenberg-Marquardt.

O algoritmo de Levenberg-Marquardt (LM) é uma técnica que pode ser incorporada ao algoritmo *backpropagation* a fim de aumentar a eficiência do processo de treinamento. Este método é bastante eficiente quando estamos tratando de redes que não possuem mais do que algumas centenas de conexões a serem ajustadas. Isto se deve, principalmente, ao fato de que estes algoritmos necessitam armazenar uma matriz quadrada cuja dimensão é da ordem do número de conexões da rede. Comprova-se que, em determinados problemas, o algoritmo LM torna-se de 10 a 100 vezes mais rápido que o algoritmo *backpropagation* convencional.

## 2.7 APLICAÇÃO DO PROCESSO DE DESENVOLVIMENTO DE RNA'S

Nesta seção, uma seqüência de ações é apresentada com a finalidade de sistematizar o processo de desenvolvimento das RNA's. Essas ações propiciam a utilização de forma adequada da técnica, prevenindo e evitando a ocorrência de problemas durante os treinamentos.

### 2.7.1 Coleta de dados

A primeira ação no processo de desenvolvimento de redes neurais artificiais é a coleta de dados relativos ao problema em questão. Esta tarefa requer uma cuidadosa análise do problema, usando técnicas de amostragem de forma a minimizar ambigüidades e erros nos dados. Além disso, os dados coletados devem ser significativos e cobrir amplamente o domínio do problema. Eles não devem cobrir apenas as operações normais ou rotineiras, mas também as exceções e condições pertencentes aos limites do domínio do problema.

### 2.7.2 Separação em conjuntos

Normalmente, os dados coletados devem ser separados em duas categorias:

- dados de treinamento - utilizados para o treinamento da rede;
- dados de teste - utilizados para verificar o desempenho no referente às condições reais de utilização e a capacidade de generalização da rede.

Depois de determinados estes conjuntos, eles são geralmente colocados em ordem aleatória para prevenção de tendências associadas à ordem de apresentação dos dados. Além disso, pode ser necessário pré-processar estes dados, através de normalizações, escalonamentos e conversões de formato para torná-los mais apropriados à sua utilização na rede.

### 2.7.3 Configuração da rede

A terceira ação é a especificação da configuração da rede, que pode ser dividida em três etapas:

- a seleção do paradigma neural apropriado à aplicação;
- a determinação da topologia da rede a ser utilizada: o número de camadas, o número de unidades em cada camada, etc;
- a determinação de parâmetros do algoritmo de treinamento e das funções de ativação dos neurônios. Esta etapa tem um grande impacto no desempenho do sistema resultante.

Existem algumas metodologias na condução destas tarefas. Porém, normalmente, parte dessas escolhas é feita de forma empírica. A definição da configuração de redes neurais é ainda considerada uma arte que requer grande experiência dos projetistas.

### 2.7.4 Treinamento

A quarta ação é o treinamento da rede. Nesta fase, seguindo o algoritmo de treinamento escolhido, serão ajustados os pesos das conexões. É importante considerar, nesta fase, alguns aspectos tais como a inicialização da rede, o modo e o tempo de treinamento.

Uma boa escolha dos valores iniciais dos pesos da rede pode diminuir o tempo necessário para o treinamento. Normalmente, os valores iniciais dos pesos da rede são números aleatórios pequenos, uniformemente distribuídos em um intervalo definido.

Quanto ao tempo de treinamento, vários fatores podem influenciar a sua duração, porém sempre será necessário utilizar algum critério de parada. Os principais critérios de parada do algoritmo de aprendizagem são o número máximo de ciclos (épocas) e o erro quadrático médio por ciclo. Pode ocorrer que, em um determinado instante do treinamento, a generalização comece a degenerar, causando o problema denominado *over-training*, ou seja a rede se especializa no conjunto de dados do treinamento e perde a capacidade de generalização.

O treinamento deve ser encerrado quando a rede apresentar uma boa capacidade de generalização e quando a taxa de erro for suficientemente pequena, ou seja, menor que um nível admissível previamente estabelecido.

### 2.7.5 Teste e integração

Durante esta fase, o conjunto de teste é utilizado para determinar o desempenho da rede com dados que não foram previamente utilizados.

Finalmente, com a rede treinada e devidamente avaliada, ela pode ser integrada em um sistema do ambiente de operação da aplicação.

### 3. DETERMINAÇÃO DO NÚMERO DE NEURÔNIOS

O ponto mais importante das RNA's está na sua capacidade de aprender a partir da apresentação de amostras (padrões) que exprimem o comportamento do sistema, sendo que, em seguida, após a rede ter aprendido o relacionamento entre as entradas e saídas, esta é capaz de generalizar soluções. A rede será então capaz de produzir uma saída próxima daquela esperada (desejada) a partir de quaisquer sinais inseridos em suas entradas.

Assim, a etapa de treinamento de uma rede neural consiste da aplicação de passos ordenados que sejam necessários para sintonização dos pesos sinápticos e limiares de seus neurônios. Tendo como objetivo final a generalização de soluções a serem produzidas pelas suas saídas, cujas respostas são representativas do sistema físico em que estas estão mapeando.

O conjunto desses passos ordenados visando o treinamento da rede é denominado de **algoritmo de aprendizagem**. Ao longo de sua aplicação, a rede será então capaz de extrair características discriminantes do sistema a ser mapeado por intermédio de amostras que foram retiradas de seu contexto.

Geralmente, o conjunto total das amostras disponíveis sobre o comportamento do sistema é dividido em dois subconjuntos, um subconjunto de **treinamento** e outro de **teste**. O subconjunto de treinamento, é composto aleatoriamente com cerca de **60% a 90%** das amostras do conjunto total, será usado essencialmente no processo de aprendizado da rede. Já o subconjunto de teste, cuja composição está entre **10% e 40%** do conjunto total de amostras, será utilizado para verificar se os aspectos referentes à generalização de soluções por parte da rede já estão em níveis aceitáveis, possibilitando assim a validação da topologia assumida. Contudo, o dimensionamento desses conjuntos deve também levar em consideração a caracterização estatística dos dados.

Durante o processo de treinamento de RNA's, cada apresentação completa das amostras pertencentes ao subconjunto de treinamento, visando, sobretudo, o ajuste dos pesos sinápticos e liminares de seus neurônios, será denominada de **época de treinamento**.

#### 3.1 DETERMINAÇÃO DOS NÚMEROS DE NEURÔNIOS

Neste trabalho será comentado apenas três dos diversos métodos para determinação (estimativa) do número de neurônios da camada. Os dois métodos são:

- Método de Kolmogorov
- Método de Fletcher-Gloss
- Método de Weka

A finalidade de tais métodos é indicar um ponto de partida para construção da rede.

**a) Método de Kolmogorov:** para determinar o número de neurônios ( $n$ ) da camada escondida é estimado pela equação (3.1):

$$n = 2 \cdot n_1 + 1 \quad (3.1)$$

onde:

$n_1$  – representa o número de variáveis de entradas da rede.

**b) Método de Fletcher-Gloss:** o número de neurônios da camada escondida será dado pela equação (3.2):

$$2 \cdot \sqrt{n_1} + 1 \leq n \leq 2 \cdot n_2 + 1 \quad (3.2)$$

onde:

$n_1$  – representa o número de variáveis de entradas da rede.

$n_2$  – representa o número de neurônios de saída da rede.

$n$  – quantidade de neurônios

**b) Método de Weka:** para classificação de padrões de classes. O número de neurônios será fornecido pela equação (3.3)

$$n_1 = \frac{n + n_c}{2} \quad (3.3)$$

$n_c$  – número de classes

Embora muito utilizados tais métodos, são recomendados apenas para alguns tipos de problemas bem comportados. Porque desconsideram atributos que possam ser relevantes para especificação da topologia de redes PMC. Como a quantidade de dados, a complexidade do problema, a qualidade dos dados e suas disposições no espaço amostral.

## REFERENCIAS

- [01] A. P Braga, A. P. L. F. de Carvalho e T. B. Ludemir. Redes Neurais Artificiais Teoria e Aplicações. Rio de Janeiro: LTC, 2000.
- [02] Barreto, Guilherme de Alencar. Perceptron multicamadas e o algoritmo de retropropagação do erro. Publicação interna: Programa de pós-graduação em engenharia de teleinformática. Fortaleza: UFC, 2007.
- [03] Caio C. O. Ramos, André N. de Souza, Lucas I. Pereira, Danilo S. Gastaldello, Maria G. Zago e João P. Papa. “Técnicas Inteligentes Aplicadas na Identificação de Consumidores Industriais Fraudadores de Energia Elétrica”. The 8th Latin-American Congress on Electricity Generation and Transmission. CLAGTEE, 2009.
- [04] Fernando José Von Zuben. “RNAs - Notas de aula”. Campinas: Unicamp, 2001.
- [05] Finocchio, Marco Antonio Ferreira. Determinação da temperatura de enrolamentos de transformadores a seco e de suas perdas totais baseado em redes neurais artificiais. Dissertação de Mestrado em Engenharia Elétrica – Universidade Estadual de Londrina, 2010.
- [06] G. Lupi Filho, R. A. S. Fernandes, A. A. Vallada, I. N. da Silva e R. A. C. Altafim. “Um Estudo Comparativo entre Abordagens Convencionais e Redes Neurais Artificiais para Diagnóstico de Transformadores de Potência”. VIII CBQEE Conferência Brasileira de Qualidade de Energia Elétrica. Blumenau, 2009
- [07] M. Hagan e M. Menhaj; “Training feedforward networks with the Marquardt algorithm”; IEEE Transaction on neural networks; 989-993; Novembro; 1994.
- [08] Silva, Ivan Nunes da; Spatti, Danilo Hernane; Flauzino, Rogério Andrade. Redes neurais artificiais: para engenharia e ciências aplicadas. São Paulo: Artliber, 2010.
- [09] Simon Haykin. Neural Networks and Learning Machines. 3ª ed.. New York: Prentice Hall, 2008.
- [10] Spandri, Renato. Introdução a redes neurais. *Controle & Instrumentação*, São Paulo: Editora Técnica Comercial Ltda, nº 43, p. 68-73, fev. 2000.