

[Imprimir](#)[Fechar](#)*Referência Completa*

Tipo da Referência	Conference Proceedings
Chave Secundária	INPE-11558-PRE/6946
Chave de Citação	Matuck:2004:EsMoRe
Autor(es)	Matuck, Gustavo Ravanhani
Título	Estudos de modelos de redes neurais artificiais para aprendizagem on-line
Nome do Evento	Seminário de Iniciação Científica do INPE (SICINPE)
Localização do Evento	São José dos Campos
Data	03 - 04 ago.
Editora (Publisher)	INPE
Ano	2004
Páginas	1-78
Título do Livro	Anais
Endereço da Editora	São José dos Campos
Organização	Instituto Nacional de Pesquisas Espaciais
Repositório do Metadado	sid.inpe.br/marciana/2004/11.19.10.28.37
Palavras-Chave	COMPUTAÇÃO APLICADA, Redes neurais, Comportamento humano, Testes psicológicos, Aprendizagem, COMPUTER SCIENCE, Neural nets, Human behavior, Psychological tests, Learning.
Resumo	<p>O ser humano quando submetido a situações em regime de confinamento por longos períodos pode ser influenciado pelo ambiente restrito e sofrer alterações comportamentais e fisiológicas. Essa é uma preocupação existente em missões em que grupos reduzidos são requisitados para realizar tarefas em ambientes como submarinos, missões espaciais e plataformas petrolíferas, por exemplo. O acompanhamento psicológico é médico é essencial para evitar o agravamento de possíveis problemas. Entretanto, esse acompanhamento é, na maioria das vezes, realizado após a volta das equipes por não haver meios para avaliação direta no sentido médico-paciente. Uma estratégia para iniciar antecipadamente o diagnóstico, seria o monitoramento das equipes por psicólogos através da observação remota ou auxiliado por sistemas de monitoramento autônomo que permitam pré-diagnósticos parciais a partir de características que possam ser captadas ou medidas remotamente como, expressões faciais, frequência e tom da voz, e postura, etc. Dentro deste contexto, este trabalho utiliza as redes neurais artificiais como ferramentas para aprender a reconhecer padrões de comportamento pela análise de testes psicológicos utilizando mecanismos como o questionário VARK, que verifica a forma como o observado trata a informação (V - visual; A - Auditiva; R - leitura/escrita; K - cinestésica); o teste MCPL - Mooney Problem Check List (lista de problemas de Mooney) que ajuda o indivíduo a expressar seus problemas pessoais; e a análise de estado emocional do indivíduo pela análise</p>

das características da voz. Foram utilizados diferentes modelos de redes neurais como tarefas de aprendizagem das técnicas, e alguns resultados promissores foram conseguidos utilizando redes com aprendizagem supervisionada.

Ultima Atualização do Metadado 2005:06.10.18.15.37 sid.inpe.br/banon/2003/08.15.17.40 sergio

Site <iris.sid.inpe.br:1916>

Idioma Portuguesse

Detentor da Cópia SID/SCD

Tipo Secundário PRE CN

Divulgação <E>

Formato Papel

Area COMP

Grupo(s) LAC

Afiliação Instituto Tecnológico Aeroespacial (ITA)
Instituto Nacional de Pesquisas Espaciais, Laboratório Associado de Computação e Matemática Aplicada (INPE.LAC)

Projeto Redes neurais

Usuário Avançado sergio

Grupo de Usuários sergio

Visibilidade shown

Data de Acesso 10 jun. 2005

[atualizar](#)

[Fechar](#)

Estudos de Modelos de Redes Neurais Artificiais Para Aprendizagem On-Line

Gustavo Ravanhani Matuck¹ (UNIVAP, Bolsista PIBIC/CNPq)
Dr. José Demísio Simões da Silva² (LAC/INPE)

RESUMO

O ser humano quando submetido a situações em regime de confinamento por longos períodos pode ser influenciado pelo ambiente restrito e sofrer alterações comportamentais e fisiológicas. Essa é uma preocupação existente em missões em que grupos reduzidos são requisitados para realizar tarefas em ambientes como submarinos, missões espaciais e plataformas petrolíferas, por exemplo. O acompanhamento psicológico e médico é essencial para evitar o agravamento de possíveis problemas. Entretanto, esse acompanhamento é, na maioria das vezes, realizado após a volta das equipes por não haver meios para avaliação direta no sentido médico-paciente. Uma estratégia para iniciar antecipadamente o diagnóstico, seria o monitoramento das equipes por psicólogos através da observação remota ou auxiliado por sistemas de monitoramento autônomo que permitam pré-diagnósticos parciais a partir de características que possam ser captadas ou medidas remotamente como, expressões faciais, frequência e tom da voz, e postura, etc. Dentro deste contexto, este trabalho utiliza as redes neurais artificiais como ferramentas para aprender a reconhecer padrões de comportamento pela análise de testes psicológicos utilizando mecanismos como o questionário VARK, que verifica a forma como o observado trata a informação (V - visual; A - Auditiva; R - leitura/escrita; K - cinestésica); o teste MCPL - Mooney Problem Check List (lista de problemas de Mooney) que ajuda o indivíduo a expressar seus problemas pessoais; e a análise de estado emocional do indivíduo pela análise das características da voz. Foram utilizados diferentes modelos de redes neurais como tarefas de aprendizagem das técnicas, e alguns resultados promissores foram conseguidos utilizando redes com aprendizagem supervisionada.

BANCO DE FONTES DE CONSULTA EM CIÊNCIA E TECNOLOGIA

¹ Aluno do Curso de Ciência da Computação, UNIVAP. E-mail: gmatuck@hotmail.com

² Pesquisador do Laboratório Associado de Computação e Matemática Aplicada - LAC. E-mail: demisio@lac.inpe.br



MINISTÉRIO DA AERONÁUTICA
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

AUTORIZAÇÃO PARA PUBLICAÇÃO

Número

PIBIC-029/04

Título

MM3

Estudo de Modelos de Redes Neurais Artificiais para Aprendizagem On-Line

Autor

Gustavo Ravanhani Matuck

Tradutor

Não há

Editor

Origem

Projeto

Série

No. de Páginas

No. de Fotos

No. de Mapas

78

0

0

Tipo

RPQ PRE NTC PRP MAN PUD TAE

Divulgação

Externa Interna Reservada Lista de Distribuição Anexa

Periódico / Evento

Seminário de Iniciação Científica do INPE - SICINPE 2004

Convênio

Autorização Preliminar

___/___/___
Data

Coordenador do PIBIC
Programa Institucional de Bolsas
de Iniciação Científica do INPE

Revisão Técnica

Solicitada

Dispensada

Recebida ___/___/___

Devolvida ___/___/___

Titular de Nível "A"

Assinatura do Revisor

Revisão de Linguagem

Solicitada

Dispensada

Recebida ___/___/___

Devolvida ___/___/___

Titular de Nível "A"

Assinatura do Revisor

Autorização Final

___/___/___
Data

Marcos Dias da Silva
Coordenador de Ensino, Documentação
e Projetos de Pesquisa - PIBIC

Palavras Chave

Redes Neurais - On-Line - Artificiais

(continua na folha 2/2)

INPE - 106.2



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

INPE-11558-PRE/6946

**ESTUDO DE MODELOS DE REDES NEURAIIS ARTIFICIAIS
PARA APRENDIZAGEM ON-LINE**

Gustavo Ravanhani Matuck

Relatório Final de Projeto de Iniciação Científica
(PIBIC/CNPq/INPE)

INPE
São José dos Campos
2004



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

**RELATÓRIO FINAL DE PROJETO DE INICIAÇÃO CIENTÍFICA
(PIBIC/CNPq/INPE)**

**ESTUDO DE MODELOS DE REDES NEURAIIS ARTIFICIAIS PARA
APRENDIZAGEM ON-LINE**

**Gustavo Ravanhani Matuck (Bolsista PIBIC/CNPq)
E-mail: gmatuck@hotmail.com**

**Dr. José Demisio Simões da Silva (LAC/INPE, Orientador)
E-mail: demisio@lac.inpe.br**

Junho de 2004

SUMÁRIO

1 - INTRODUÇÃO	1
1.1 - Apresentação do tema	1
1.2 - Sistemas Conexionistas	2
1.3 - Objetivos.....	3
1.3.1 - Objetivos Gerais.....	3
1.3.2 - Objetivos Específicos.....	3
1.4 - Organização do trabalho	3
2 - O SISTEMA NERVOSO HUMANO	4
2.1 - Introdução	4
2.2 - Níveis de Organização Estrutural de um Cérebro	4
2.3 - Plasticidade Neuronal	6
2.4 - O Neurônio Biológico	6
3 - REDES NEURAIS ARTIFICIAIS	8
3.1 - Introdução	8
3.2 - Histórico	11
3.3 - O Neurônio Artificial	13
3.4 - Tipos de Função de Ativação.....	14
3.5 - Arquiteturas	16
3.6 - Aprendizado	21
3.6.1 - Aprendizado Supervisionado	21
3.6.2 - Aprendizado Não-Supervisionado	22
3.6.3 - Velocidade de Aprendizado	23
3.6.4 - Aprendizado na Correção de Erros	24
3.6.5 - Aprendizado Hebbiano	25
3.6.6 - Aprendizado por Competição	26
3.6.7 - Aprendizado por Reforço.....	27
3.6.8 - A Regra Delta de Widrow.....	28
3.7 - Rede Perceptron	29
3.7.1 - Limitações: O problema do OU-EXCLUSIVO	32
3.8 - Rede Adaline	33
3.9 - Rede Madaline	34
3.10 - Rede Perceptron de Múltiplas Camadas	35
3.11 - Redes SOM – Kohonen	40
3.12 - Implementações de Redes Neurais Artificiais.....	43
3.12.1 - Problema XOR.....	43
3.12.2 - Reconhecimento de Caracteres	45
4 - ANÁLISES DOS TESTES PSICOLÓGICOS.....	50
4.1 - Teste Psicológico VARK	50
4.1.1 - Introdução	50
4.1.2 - Aplicação	52

4.1.3 - Rede Adaline.....	55
4.1.4 - Rede Kohonen (SOM)	56
4.1.5 - Rede Perceptron de Múltiplas Camadas (MLP)	57
4.1.6 - Rede Perceptron de Camada Simples (SLP).....	59
4.1.7 - Conclusão.....	59
4.2 - Teste Psicológico MPCL	60
4.2.1 - Introdução	60
4.2.2 - Aplicação	61
4.2.3 - Rede Adaline.....	63
4.2.4 - Rede Kohonen (SOM)	65
4.2.5 - Rede Perceptron de Múltiplas Camadas (MLP)	66
4.2.6 - Rede Perceptron de Camada Simples (SLP).....	67
4.2.7 - Conclusão.....	68
5 - RECONHECIMENTO DE VOZ.....	69
5.1 - Aplicação	69
5.2 - Conclusão	74
6 - CONCLUSÕES E RECOMENDAÇÕES FINAIS	76
7 - REFERÊNCIAS BIBLIOGRÁFICAS	77

Lista de Figuras

Figura 2.1 - Representação do diagrama de bloco do sistema nervoso.....	4
Figura 2.2 - Níveis de organização estrutural de um cérebro.....	5
Figura 2.3 - A Estrutura do Neurônio.....	7
Figura 3.1 - Modelo do neurônio artificial de McCulloch-Pitts.....	13
Figura 3.2 - Algumas funções de ativação.....	15
Figura 3.3 - Rede Neural Artificial.....	16
Figura 3.4 - Rede Neural Artificial de uma única camada.....	17
Figura 3.5 - Rede Neural Artificial de múltiplas camadas.....	18
Figura 3.6 - Redes Neurais Artificiais diretas (feedforward ou acíclicas).....	19
Figura 3.7 - Redes Neurais Artificiais feedback ou cíclica.....	20
Figura 3.8 - Tipos de conectividade de uma rede neural artificial.....	20
Figura 3.9 - Diagrama de bloco do aprendizado supervisionado.....	21
Figura 3.10 - Diagrama de bloco do aprendizado não-supervisionado.....	23
Figura 3.11 - Arquitetura do aprendizado competitivo com conexões feedforward (excitatória) dos nodos de entrada para os neurônios, e conexões laterais (inibitória) entre os neurônios; as correções laterais são identificadas por setas pontilhadas [HAY-94].....	27
Figura 3.12 - Aprendizado por Reforço.....	28
Figura 3.13 - Topologia de um Perceptron simples com uma única camada.....	29
Figura 3.14 - Regiões de decisão de um Perceptron.....	30
Figura 3.15 - Unidade de processamento do Perceptron.....	31
Figura 3.16 - Passos do algoritmo Perceptron simples.....	32
Figura 3.17 - Representação geométrica do espaço de entrada.....	33
Figura 3.18 - Diagrama esquemático de um nodo Adaline.....	34
Figura 3.19 - Regiões de decisão da Madaline.....	35
Figura 3.20 - Esquema de uma rede do tipo MLP.....	36
Figura 3.21 - Regiões definidas pelo processamento das redes MLP.....	37
Figura 3.22 - Rede MLP com o algoritmo backpropagation.....	39
Figura 3.23 - Passos do algoritmo Backpropagation.....	39
Figura 3.24 - Rede SOM típica.....	41
Figura 3.25 - Formatos para a região vencedora utilizados pelas redes SOM.....	42
Figura 3.26 - Passos do algoritmo para as redes SOM.....	43
Figura 3.27 - Erro quadrático médio da rede MLP no problema XOR.....	44
Figura 3.28 - Erro quadrático médio encontrado na rede MLP.....	47
Figura 3.29 - Erro encontrado na rede Adaline no reconhecimento de caracteres.....	49
Figura 4.1 - Exemplo da correção de um questionário VARK.....	51
Figura 4.2 - Erro encontrado na rede Adaline no teste VARK.....	55
Figura 4.3 - Soma dos erros quadrático encontrado no Backpropagation.....	58
Figura 4.4 - Erro encontrado na rede Adaline no reconhecimento dos testes.....	64
Figura 4.5 - Erro encontrado na rede Adaline no reconhecimento dos testes.....	66
Figura 5.1 - Ilustração das vozes gravadas.....	69
Figura 5.2 - Ilustração das Etapas utilizadas no treinamento da rede MLP.....	71
Figura 5.3 - Erro encontrado no treinamento da rede MLP.....	72
Figura 5.4 - Instrumento utilizado na gravação de voz.....	74

Lista de Tabelas

Tabela 3.1 - Correspondência entre as portas lógicas	33
Tabela 3.2 - Resultado da rede SingleLayer Perceptron	43
Tabela 3.3 - Resultado da rede MLP no problema XOR.....	44
Tabela 3.4 - Resultado obtido da rede Adaline	45
Tabela 3.5 - Resultado da rede Kohonen.....	45
Tabela 3.6 - Padrões de entrada para exemplos de reconhecimento de caracteres	46
Tabela 3.7 - Resultado da rede SOM no reconhecimento de caracteres	47
Tabela 3.8 - Resultado da rede MLP no reconhecimento de caracteres.....	48
Tabela 3.9 - Resultado da rede SLP no reconhecimento de caracteres	48
Tabela 3.10 - Resultado obtido da rede Adaline	49
Tabela 4.1 - Distância do passo de acordo com a soma do escores VARK	51
Tabela 4.2 - Encontrando o diagnóstico	52
Tabela 4.3 - Diagnóstico de cada Teste.....	53
Tabela 4.4 - Disposição dos Testes utilizados na Rede Neural Artificial	54
Tabela 4.5 - Diagnósticos dos testes codificados	54
Tabela 4.6 - Resultado do teste VARK na rede Adaline	56
Tabela 4.7 - Resultado da rede Kohonen no teste psicológico	57
Tabela 4.8 - Resultado da rede MLP no teste VARK	58
Tabela 4.9 - Resultado da rede Perceptron de Camada Simples	59
Tabela 4.10 - Resultados do treinamento das redes.....	60
Tabela 4.11 - Disposição dos blocos dos diagnósticos no teste	61
Tabela 4.12 - Diagnóstico de cada Teste.....	62
Tabela 4.13 - Disposição dos Testes utilizados na Rede Neural Artificial	62
Tabela 4.14 - Diagnósticos dos testes codificados	63
Tabela 4.15 - Resultado obtido da rede Adaline no reconhecimento dos testes	64
Tabela 4.16 - Resultado da rede Kohonen no teste MPCL	65
Tabela 4.17 - Resultado obtido da rede MLP no reconhecimento dos testes.....	67
Tabela 4.18 - Resultado da rede Perceptron de Camada Simples	68
Tabela 4.19 - Resultados do treinamento das redes.....	68
Tabela 5.1 - Disposição das vozes utilizadas na rede MLP.....	70
Tabela 5.2 - Codificação dos diagnósticos de voz	70
Tabela 5.3 - Porcentagem de acerto das vozes <i>treinadas</i> na rede MLP	73
Tabela 5.4 - Porcentagem de acerto das vozes <i>generalizadas</i> na rede MLP.....	74

1 - INTRODUÇÃO

1.1 - Apresentação do tema

Nos dias atuais, a tecnologia da informação está cada vez mais presente em nossas vidas. Esse conceito está cada vez mais propiciando uma solução para uma qualidade de vida melhor das pessoas. Uma questão ainda preocupante hoje em dia são as conseqüências trazidas pelo isolamento social onde não haja possibilidades de se fazer uma avaliação à distância na mudança desse comportamento.

As conseqüências do isolamento do ser humano nas suas mais diversas áreas de atuação vêm trazendo problemas psicológicos dos mais variados níveis de complexidade. Como exemplos de isolamento social podemos citar o confinamento das pessoas em submarinos, plataformas de petróleo, presídios, na Base de Pesquisa da Antártida, viagens espaciais e até para um futuro problema que poderá ocorrer na viagem da NASA programada para 2020 para Marte no qual os tripulantes estarão totalmente isolados por um período de aproximadamente 2 anos.

Para problema em questão foi abordada uma solução para a construção de uma sistemática computacional de confiabilidade que possibilita fazer a avaliação a distancia de qualquer mudança do estado emocional de uma pessoa. Esse estudo tem como base a avaliação do comportamento humano em isolamento social nas mais diversas situações, sendo executadas por sistemas inteligentes autônomos.

Para a realização de tal diagnóstico, três áreas de atuação se reuniram para concretização dessa linha de pesquisa, a Informática, a Psiquiatria e a Psicologia. A princípio, tem-se como idéia utilizar diversos mecanismos como medição na avaliação do comportamento humano, tantos computacionais como psicológicos. Tais funcionalidades têm o objetivo de proporcionar a elaboração de um protótipo computacional cuja finalidade é à sistemática de avaliação da mudança do comportamento humano.

Para os mecanismos computacionais tem-se como interesse tanto um estudo como a aplicação relacionada à área de inteligência artificial. Acredita-se no fato de que, caso for verificado a existência de padrões na identificação da mudança do comportamento humano, este conceito computacional poderá formular um banco de dados relacionados a esses padrões e especificar qualquer mudança relacionado a esse comportamento. Para isso está sendo feita uma abordagem as Redes Neurais Artificiais (RNAs), pois acredita-se na sua capacidade de aprendizagem e propagação na identificação desses padrões. No estudo dessa abordagem foram utilizados procedimentos conceituais bibliográficos no intuito de propiciar o seu entendimento e aplicação dessa metodologia nas mais diversas finalidades.

Para os procedimentos psicológicos, será dada importância à identificação de padrões para abordagens como a expressão facial, pulsação, temperatura, voz, transpiração entre possíveis outros.

Foi utilizado também como base para esses estudos dois testes psicológicos o qual auxiliam na identificação de características do ser humano.

No primeiro teste psicológico o questionário LPM (Lista de Problemas Pessoais) tem como base ajudar as pessoas a expressar seus problemas pessoais. Esses problemas são relacionados com saúde, trabalho, família, religião, temperamento, financeiro, etc.

Já no segundo teste psicológico foi utilizado o questionário VARK. Esse questionário tem por finalidade investigar as preferências por modos pelos quais as pessoas trabalham com informação. Será adquirida com a aplicação do teste VARK um estilo de aprendizagem preferido e que, uma parte desse estilo será a preferência pelo recebimento e pela transmissão de idéias e de informações.

Pouco se sabe sobre o sucesso na aplicação dessa metodologia abordada até o momento em sistemas computacionais. No ramo da psicologia e psiquiatria há vários métodos sistemáticos na avaliação do estado emocional de uma pessoa que são utilizados produzindo resultados de grande confiança. A questão da confiabilidade de um diagnóstico pode ser representada na base da “tentativa e erro”, onde a teoria e a prática caminham lado a lado para o assunto em questão.

Pode se perceber então que dentre os diversos ramos de pesquisa como Ciência da Computação, Linguística, Matemática, Neurofisiologia, Psicologia, Psiquiatria entre outras que buscam em conjunto a representação do conhecimento inspirados na organização física do cérebro humano, suas potencialidades e limitações, é encontrado o Conexionismo.

1.2 - Sistemas Conexionistas

O Conexionismo revela uma forma particular de processamento da informação, sendo já há algum tempo, uma área de pesquisa que envolve indistintamente várias disciplinas do conhecimento. Os sistemas conexionistas são modelos computacionais inspirados na estrutura e funcionamento do cérebro humano e caracterizam-se pela reunião de uma grande quantidade de células (elementos ou unidades) de processamento interligadas por um grande número de conexões (links), que processam a informação de forma paralela.

Este tipo de processamento da informação é também chamado de processamento paralelo distribuído (PPD), ou sistemas neuronais, ou ainda, redes neurais, em alusão à tentativa de imitação das células cerebrais, suas características básicas e organização. Contudo, conforme citam diversos autores a denominação mais adequada seria redes neurais artificiais (RNA), pois, o conhecimento do funcionamento do cérebro, bem como dos neurônios ainda é bastante incompleto, portanto, tais modelos computacionais nada mais são que grosseiras imitações parciais das redes de neurônios do cérebro.

1.3 - Objetivos

1.3.1 - Objetivos Gerais

O principal objetivo desse trabalho é a realização de um estudo sobre redes neurais artificiais bem como sua implementação na solução de problemas específicos.

1.3.2 - Objetivos Específicos

- Adquirir conhecimentos sobre redes neurais artificiais(neural networks).
- Implementar modelos de redes neurais em linguagem C.
- Desenvolver um ambiente para aprendizagem on-line de padrões de uso de sistemas interativos como um teclado, por exemplo.
- Aplicação de redes neurais artificiais no reconhecimento de voz.

1.4 - Organização do trabalho

Este trabalho está disposto em 6 capítulos. O capítulo 1 trata dos aspectos introdutórios do trabalho, como a apresentação do tema e uma breve introdução aos sistemas conexionistas. O capítulo 2 faz uma abordagem geral referente ao sistema nervoso, demonstrando a estrutura do neurônio e sua comunicação.

Já no capítulo 3 é feita uma introdução sobre redes neurais artificiais, como o seu surgimento, arquitetura e disposição dos elementos formadores destas redes (os neurônios artificiais), os conceitos referentes aos métodos e paradigmas de treinamento e algumas implementações práticas.

No capítulo 4 é feita a análise dos testes psicológicos VARK e MPCL para sua aplicação em Redes Neurais Artificiais. Foram testados quatro tipos de redes para a análise dos testes: Perceptron de camada simples (SLP), Perceptron de Múltiplas Camadas (backpropagation - MLP), rede Kohonen (SOM) e Adaline.

No capítulo 5 é realizado um estudo sobre o comportamento da rede Perceptron de Múltiplas Camadas no reconhecimento de vozes calmas e nervosas. Finalmente no capítulo 6 são feitas as considerações finais e conclusões do trabalho.

2 - O SISTEMA NERVOSO HUMANO

2.1 - Introdução

O cérebro humano é considerado o mais fascinante processador baseado em carbono existente, sendo composto por aproximadamente 10 bilhões de neurônios. Todas as funções e movimentos do organismo estão relacionados ao funcionamento destas pequenas células. Os neurônios estão conectados uns aos outros através de sinapses, e juntos formam uma grande rede, chamada **rede neural**. Esta grande rede proporciona uma fabulosa capacidade de processamento e armazenamento de informação.

Um dos principais objetivos da neurociência é a compreensão de como são realizadas as atividades do encéfalo, explicando como atuam as milhões de células nervosas no encéfalo que produzem uma acentuada individualidade da ação humana, isto é, como é conduzido o relacionamento das moléculas responsáveis pelas atividades das células nervosas resultando nos complexos processos mentais.

Todo ser vivo dotado de um sistema nervoso é capaz de modificar o seu comportamento em função de experiências passadas. Essa modificação comportamental é chamada de aprendizado, e ocorre no sistema nervoso através da propriedade chamada **plasticidade neuronal**.

2.2 - Níveis de Organização Estrutural de um Cérebro

O sistema nervoso humano pode ser visto num sistema de três estágios, como descrito na figura 2.1. O sistema central do cérebro representado por “Rede neural (neural net)”, que recebe informação continuamente, percebe isto, e faz a decisão apropriada. As setas que apontam da esquerda para a direita indica a transmissão de sinais de informação pelo sistema. Na outra mão, as setas que apontam da direita para a esquerda significa a presença de avaliação no sistema. Os receptores na figura 2.1 convertem estímulos do corpo humano ou do ambiente externo em impulsos elétricos que carrega a informação para a rede neural (cérebro). Os Atuadores, na outra mão, convertem impulsos elétricos gerados pela rede neural em respostas discerníveis como saídas do sistema.

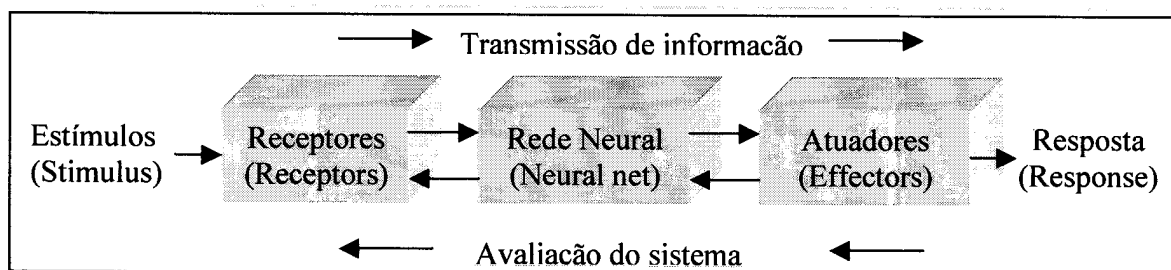


Figura 2.1 – Representação do diagrama de bloco do sistema nervoso

No cérebro há uma pequena e uma grande escala de organizações anatômicas, e funções diferentes que levam a níveis altos e baixos. A figura 2.2 mostra a hierarquia dos

níveis de organização que emergiu do trabalho extensivo feito da análise de regiões locais do cérebro.

Nós temos o mais fundamental nível que são as **moléculas** e íons que em ação, propiciam funcionalidade às sinapses. O **microcircuito neural** refere-se a uma assembléia de sinapses organizada em padrões de conectividade para produzir uma operação funcional de interesse. Os microcircuitos neurais são agrupados formando subunidades de dendritos com a **árvore de dendritos** de neurônios individuais. Todo o **neurônio**, com um micrômetro de tamanho, contém várias subunidades de dendritos.

No próximo nível de complexibilidade, nós temos os **circuitos locais** (com um milímetro de tamanho) formado de neurônio com propriedades similares ou diferentes; esses conjuntos neurais executam operações características de uma região localizada no cérebro. Isto é seguido pelo **circuito inter-regional** formado de caminhos, colunas, e mapas topográficos, que envolvem múltiplas regiões localizadas em diferentes partes do cérebro.

Finalmente, os mapas topográficos, e outros circuitos inter-regionais medem tipos específicos de comportamentos dentro do **sistema nervoso central**. É importante reconhecer que os níveis de estrutura de organização descrito nisto são a única característica de um cérebro. Eles não serão achados em nenhuma parte em um computador digital, e não estão em nenhuma parte perto de percebê-los com redes neurais artificiais.

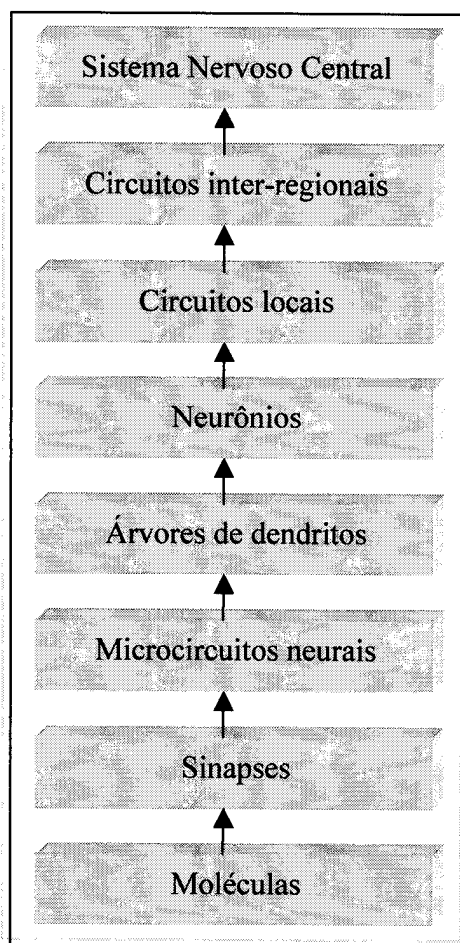


Figura 2.2 – Níveis de organização estrutural de um cérebro

2.3 - Plasticidade Neuronal

Pesquisas em Neurobiologia têm comprovado que a plasticidade do sistema nervoso é uma característica única em relação a todos os outros sistemas orgânicos.

Por modelagem de aprendizado ou plasticidade entende-se as especificações do conjunto de leis que determinam como os pesos das conexões¹ serão inicializados e alterados durante o funcionamento da rede. Conforme deGroot, "a plasticidade neural é a propriedade do sistema nervoso que permite o desenvolvimento de alterações estruturais em resposta à experiência, e como adaptação a condições mutantes e a estímulos repetidos".

As regras de aprendizagem quanto ao tipo de treinamento podem ser divididas em dois grandes grupos: treinamento com supervisão ou sem supervisão. O treinamento supervisionado consiste em apresentar à rede um padrão a ser reconhecido, juntamente com a resposta que a rede deve fornecer ao reconhecer novamente este mesmo padrão. O treinamento sem supervisão consiste em apresentar apenas os padrões para a rede e esta se encarrega de agrupar aqueles que possuem propriedades similares (*clusterization*).

Outro aspecto a ser avaliado no aprendizado é a capacidade de treino adaptativo. Ou seja, se é possível modificar os pesos das conexões durante o funcionamento da rede. No grupo de algoritmos adaptativos encontramos aqueles modelos cujos pesos das conexões podem ser modificados durante o seu funcionamento e no outro grupo temos os modelos como o de Hopfield e Hamming, que necessitam de uma etapa de treinamento separada em que os pesos são inicializados. Após esta inicialização e durante o uso, os pesos permanecem inalterados neste tipo de modelo.

É cientificamente provado que a "aprendizagem pode levar a alterações estruturais no cérebro". A cada nova experiência adquirida pelo indivíduo, as redes de neurônios são rearranjadas, outras tantas sinapses são reforçadas e múltiplas possibilidades de respostas ao ambiente tornam-se possíveis. Portanto, o mapa cortical de um adulto está sujeito a constantes modificações com base no uso ou atividade de seus caminhos sensoriais periféricos.

¹Os pesos das conexões devem ser inicializados ou modificados de forma que a rede responda adequadamente aos estímulos do ambiente externo.

2.4 - O Neurônio Biológico

Todos os estímulos do nosso ambiente causando sensações como dor e calor, todos os sentimentos, pensamentos, programação de respostas emocionais e motoras, bases neurais da aprendizagem e memória, ação de drogas psicoativas, 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.

A célula nervosa, ou, simplesmente, neurônio, é o principal componente do sistema nervoso, estima-se que no cérebro humano existam aproximadamente 15 bilhões destas células, que são unidades de processamento de informações responsáveis por todas as funções de uma rede neural. Neurônios são células especializadas. Eles são feitos para

receber certas conexões específicas, executar funções apropriadas e passar suas decisões a um evento particular a outros neurônios que estão relacionados com aqueles eventos.

Na figura 2.3 é mostrada a estrutura do neurônio. Neurônios recebem sinais nervosos de axônios de outros neurônios. A maioria dos sinais é liberada aos **dendritos (1)**. Os sinais gerados por um neurônio são enviados através do **corpo celular (2)**, que contém o **núcleo (2a)**, o "armazém" de informações genéticas. **Axônios (3)** são as principais unidades condutoras do neurônio. O **cone axonal (2b)** é a região na qual os sinais das células são iniciados. **Células de Schwann (6)**, as quais não são partes da célula nervosa, mas um dos tipos das células gliais, exercem a importante função de isolar neurônios por envolver seus processos membranosos ao redor do axônio formando a **bainha de mielina (7)**, uma substância gordurosa que ajuda os axônios a transmitirem mensagens mais rapidamente do que as não mielinizadas. A mielina é quebrada em vários pontos pelos **nodos of Ranvier (4)**, de forma que em uma secção transversal o neurônio se parece como um cordão de salsichas. Ramos do axônio de um neurônio (o neurônio pré-sináptico) transmitem sinais a outro neurônio (o neurônio pós-sináptico) em um local chamado **sinapse (5)**. Os ramos de um único axônio podem formar sinapses com até 1000 outros neurônios.

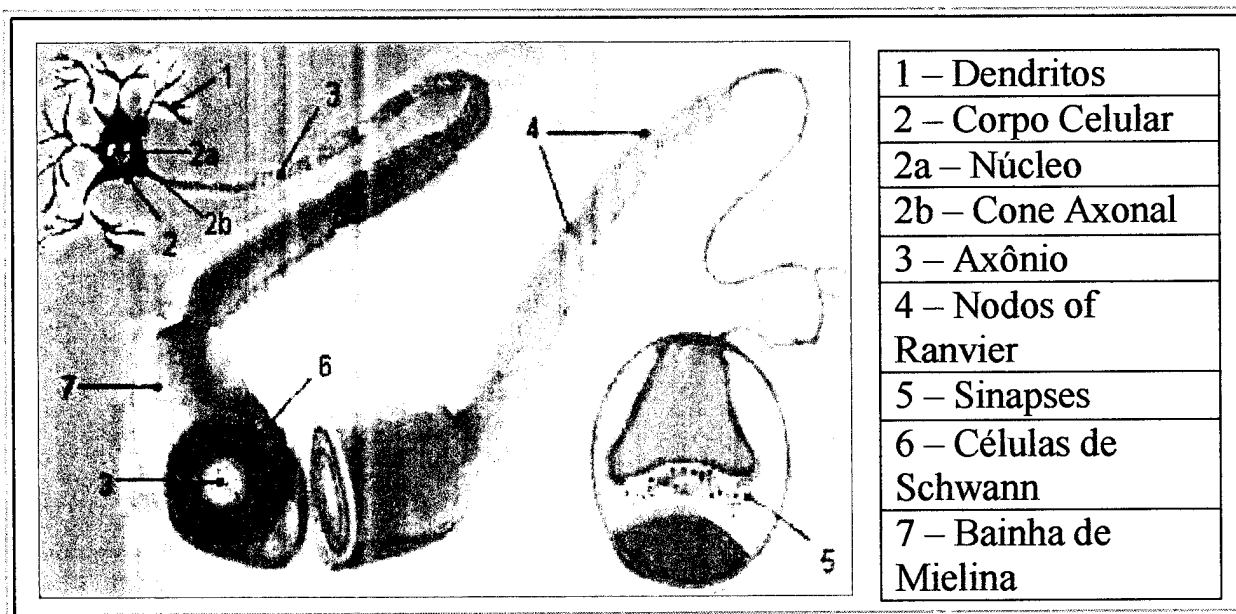


Figura 2.3 – A Estrutura do Neurônio

Nos neurônios a comunicação é realizada através de impulsos, quando um impulso é recebido, o neurônio o processa, e ultrapassado um limite de ação, dispara um segundo impulso, que produz uma substância neurotransmissora, o qual flui do corpo celular para o axônio (que por sua vez pode ou não estar conectado a um dendrito de outra célula). Eles têm um papel essencial na determinação do funcionamento, comportamento e do raciocínio do ser humano.

Ao contrário das redes neurais artificiais, redes neurais naturais não transmitem sinais negativos, sua ativação é medida pela frequência com que emite pulsos, frequência esta de pulsos contínuos e positivos. As redes naturais não são uniformes como as redes artificiais, e apresentam uniformidade apenas em alguns pontos do organismo.

3 - REDES NEURAIS ARTIFICIAIS

3.1 - Introdução

Já é um sonho antigo do homem criar máquinas capazes de gerar pensamento inteligente. Este sonho, barrado pelo desempenho insatisfatório das técnicas de computação convencionais, levou cientistas e pesquisadores a buscar formas alternativas de computação. Entre estas formas estão as Redes Neurais Artificiais.

Ao contrário dos sistemas heurísticos, assim chamados porque procuram obter sistemas inteligentes baseados em lógica e em processamento simbólico (por exemplo, os sistemas especialistas); as redes neurais artificiais se inspiram em um modelo biológico para a inteligência, isto é, na maneira como o cérebro é organizado em sua arquitetura elementar, e em como a mesma é capaz de executar tarefas computacionais.

A tecnologia das Redes Neurais Artificiais (RNAs) visa solucionar problemas de reconhecimento de padrões que geralmente são baseados em um conjunto de informações previamente conhecidos. Geralmente os conjuntos de dados são divididos em conjunto de treinamento e conjunto de teste. Atualmente pesquisadores em RNAs estão buscando uma compreensão das capacidades da natureza humana, as quais possibilitam que as pessoas construam soluções para problemas que não sejam resolvidos através de métodos tradicionais.

As redes neurais artificiais representam um novo paradigma metodológico no campo da Inteligência Artificial, onde acredita-se que construindo um sistema que simule a estrutura cerebral humana, este apresentará inteligência e será capaz de aprender, assimilar, errar e, com esta experiência, adquirir novos conhecimentos. Desta forma, as Redes Neurais tendem a representar uma estrutura interconectada similar a estrutura do cérebro humano e com isto resolver situações não facilmente resolvidas na computação convencional, onde a informática é transferida linear e sincronizadamente.

Segundo o pesquisador Teuvo Kohonen da Universidade de Helsinki, uma rede neural artificial possui a seguinte definição: “uma rede massivamente paralela de elementos interconectados e suas organizações hierárquicas que estão preparadas para iterar com objetos do mundo real de um mesmo modo que um sistema nervoso biológico faz”.

Da mesma maneira que no cérebro, as redes neurais artificiais **são organizadas** na forma de um número de elementos individuais simples (os neurônios), que se interconectam uns aos outros, formando redes capazes de armazenar e transmitir informação provinda do exterior. Outra capacidade importante das redes neurais artificiais é a auto-organização, ou **plasticidade**, ou seja, através de um processo de aprendizado, é possível alterar-se os padrões de interconexão entre seus elementos. Por este motivo, as redes neurais artificiais são um tipo de sistema conexionista, no qual as propriedades computacionais são resultado dos padrões de interconexão da rede, como acontece também no sistema nervoso biológico.

Principais características das Redes Neurais Artificiais:

- Capacidade de “aprender” através de exemplos e de generalizar este aprendizado de forma a reconhecer elementos similares, que não foram apresentados no conjunto de exemplos (treinamento);
- Não linearidade: O neurônio é basicamente um dispositivo não linear. Conseqüentemente, a rede neural compõe-se de uma interconexão de neurônios, que não são lineares. Não linearidade é uma propriedade altamente importante, particularmente se o mecanismo físico subjacente responsável pela geração de um sinal de entrada (sinal de contribuição) for inerentemente não linear;
- Mapeamento de entradas-saídas: O paradigma popular de aprendizado chamado aprendizagem supervisionada envolve a modificação dos pesos das sinapses da rede neural aplicando um conjunto de treinamento chamado “exemplos de treinamento” ou “exemplos de tarefa”. Cada exemplo consiste de um único sinal de entrada e a resposta desejada correspondente. O treinamento da rede é repetido para muitos exemplos e a rede alcança o estado fixo, onde não há nenhuma mudança significativa nos pesos das sinapses; o treinamento de exemplo previamente aplicados pode ser repetidos durante a seção de treinamento, mas em diferente ordem. Assim a rede aprende com os exemplos construindo uma entrada e saída mapeada para o problema à mão;
- Adaptativo: Redes Neurais tem uma capacidade embutida de adaptar os pesos de suas sinapses para mudar o ambiente fechado. Em particular, a rede neural treinada para operar num ambiente específico pode ser novamente treinada facilmente para lidar com umas mudanças secundárias dentro de condições ambientais de operação. Além disso, quando se está operando em um ambiente não estacionário, a rede neural pode ser designada para mudar seus pesos sinápticos em tempo real. A arquitetura natural de uma rede neural para uma classificação padrão, processamento de sinal, e aplicações de controle, junto com a capacidade adaptativa da rede neural, forma uma ferramenta ideal para a classificação padrão de adaptação, processamento de sinal adaptativo, e controle adaptativo. Para realizar todos os benefícios de adaptação, o tempo constante principal de uma sistema deve ser longo o bastante para o sistema ignorar essas perturbações espúrias e ainda curto o bastante para responder a uma significativa mudança no ambiente;
- Resposta Evidencial: No contexto de uma classificação padrão, a rede neural pode ser designada para prover informação não somente sobre um padrão particular escolhido, mas também sobre a confiança na decisão feita. Essa informação posterior pode ser usada para rejeitar padrões ambíguos, que devem surgir, e assim melhorar o desempenho de classificação da rede;
- Informação Contextual: O conhecimento é representado pela mesma estrutura e estado de ativação da rede neural. Todo neurônio na rede é potencialmente afetado pela atividade global de todos os outros neurônios da rede. Conseqüentemente, informação contextual é tratada com naturalidade pela rede neural;
- Tolerância a Falhas: A rede neural, implementada na forma de hardware, tem o potencial para ser inerentemente tolerante a falhas no senso que o desempenho é

degradado debaixo de condições operacionais adversas. Por exemplo, se um neurônio ou seu link de conexão é danificado, é prejudicada em qualidade. Porém, devido à natureza distribuída de informação da rede, o dano tem que ser extenso antes que resposta global da rede seja degradada seriamente. Assim, em princípio, a rede neural exhibe uma graciosa degradação em desempenho em lugar de fracasso catastrófico;

- Implementabilidade da VLSI: A natureza maciçamente paralela de uma rede neural faz aumentar rapidamente o potencial da computação em certas tarefas. A mesma característica faz a rede neural estar pronto para a implementação usando a tecnologia VLSI (Very Large Scale Integrated). A virtude particular do VLSI é que esta provê os meios de captação do comportamento verdadeiramente complexo, o qual faz isso ser possível usando redes neurais como uma ferramenta para aplicações em tempo real envolvendo reconhecimentos de padrões, sinais de processamento, e controle;
- Uniformidade de Análises e Projeto: Basicamente, redes neurais desfruta a generalidade como processador de informação. Isso é dito no senso que a mesma notação é utilizada em todos os domínios envolvendo aplicações de redes neurais. Esta característica manifesta a si mesmo em diferentes caminhos:
 - Neurônios, em uma forma ou outra, representa o ingrediente comum para todas as redes neurais;
 - Isto faz ser possível compartilhar teorias e aprender algoritmos em diferentes aplicações de redes neurais;
 - Redes modulares podem ser construídas por módulos de integração sem emenda.
- Analogia Neurobiológica: O escopo de uma rede neural é motivado pela analogia com um cérebro, o qual é uma prova viva de que tolerância a falhas no processamento paralelo não é somente fisicamente possível, mas também rápido e poderoso. Neurobiológicos olham para redes neurais como uma ferramenta de descoberta para interpretação de fenômenos neurobiológicos. As analogias neurobiológicas são também úteis em outro importante caminho: Essas analogias provêm uma esperança e convicção que o entendimento físico de estruturas neurobiológicas podem realmente influenciar a arte da eletrônica e do VLSI.
- Bom desempenho em tarefas pouco ou mal definidas, onde falta o conhecimento explícito de como resolvê-las, o aprendizado se dá através de exemplos;
- Robustez a presença de informações falsas ou ausentes, escolha dos elementos no próprio conjunto de treinamento (integridade do conjunto de treinamento);
- No contexto de classificações de padrões, uma rede neural pode fornecer informações sobre quais padrões selecionar em função do grau de confiança apresentado (confiabilidade no conjunto de treinamento).

Enquanto hoje os computadores funcionam de modo seqüencial, proporcionando maior eficiência na resolução de tarefas nas quais devem ser seguido determinadas etapas, o cérebro humano funciona com processamento paralelo, e sendo extremamente conectado (massivamente paralelo) é mais eficiente na resolução de tarefas que exigem um conjunto muito grande de variáveis.

Redes neurais artificiais é um processador distribuído maciçamente paralelo que tem a tendência natural de armazenar conhecimento experimental e fazer a avaliação de sua utilização. Isso se assemelha ao cérebro nos aspectos de que o conhecimento é adquirido pela rede pelo processo de aprendizagem e que forças de conexões entre neurônios conhecidas como pesos de sinapses são utilizadas para armazenar o conhecimento.[HAY-94].

3.2 - Histórico

As Redes Neurais Artificiais surgiram na década de 40, mais precisamente em 1943, quando o psiquiatra e neuroanatomista Warren McCulloch e o matemático Walter Pitts, da Universidade de Illinois, fizeram uma analogia entre as células nervosas e o processo eletrônico onde sugeriam a construção de uma máquina baseada ou inspirada no cérebro humano. O artigo foi publicado no *Bulletin of Mathematical Biophysics* com o título: “*A Logical Calculus of the Ideas Immanent in Nervous Activity*”. Muitos outros artigos e livros surgiram desde então, porém, por um longo período de tempo, pouco resultado foi obtido.

Em 1949, o biólogo e psicólogo Donald Hebb, que estudava o comportamento dos animais, escreveu um livro chamado “*The Organization of Behavior*”, que reforçava as teorias de que o condicionamento psicológico estava presente em qualquer parte dos animais, pelo fato de que esta é uma propriedade de neurônios individuais. As idéias de Hebb não eram pioneiras, mas ele propôs um princípio de aprendizado em sistemas nervosos complexos, ou seja, uma lei que descreve o funcionamento quantitativo da sinapse e do processo de treinamento humano.

Este primeiro e corajoso passo serviram de inspiração para que muitos outros pesquisadores perseguissem a mesma idéia. E embora muito tenha sido estudado e publicado nos anos que seguiram (1940-1950), estes serviram mais como base para desenvolvimento posterior que para o próprio desenvolvimento.

Em 1951, Marvin Minsky, co-fundador do Laboratório de Inteligência Artificial do MIT, construiu o SNARK, o primeiro neurocomputador de simulação de cadeia neural. O SNARK trabalhava com êxito a partir de um ponto de partida técnico e podia ajustar seus pesos sinápticos automaticamente. Entretanto, ele nunca chegou a executar qualquer função de processamento de informação interessante, mas serviu de inspiração para as idéias de estruturas que o sucederam.

Em 1956, na Primeira Conferência de Inteligência Artificial, foi apresentado um modelo de rede neural artificial pelo pesquisador da IBM Nathaniel Rochester. Seu modelo consistia numa simulação de centenas de neurônios interconectados através de um sistema que verificaria como a rede responderia aos estímulos ambientais.

Também em 1956 no Dartmouth College nasceram os dois paradigmas da Inteligência Artificial, a **simbólica** e o **conexionista**. A Inteligência Artificial Simbólica tenta simular o comportamento inteligente humano desconsiderando os mecanismos responsáveis por tal. Já a Inteligência Artificial Conexionista acredita que construindo um sistema que simule a estrutura do cérebro, este sistema apresentará inteligência, ou seja, será capaz de aprender, assimilar, errar e aprender com seus erros.

O primeiro neurocomputador a obter sucesso surgiu em 1957 e 1958, criado por Frank Rosenblatt da Universidade de Cornell. Rosenblatt criou uma rede de múltiplos neurônios do tipo discriminadores lineares e a batizou de rede Perceptron, baseando-se nas linhas de

pensamento de McCulloch para desenvolver o seu modelo matemático da sinapse humana. Devido à profundidade de seus estudos, suas contribuições técnicas e de sua maneira moderna de pensar, muitos o vêem como o fundador da neurocomputação na forma em que a temos hoje. Seu interesse inicial para a criação do Perceptron era o reconhecimento de padrões.

No final da década de 50, Minsky e Seymour Papert lançaram em uma obra chamada Perceptron, a qual mostrava que o modelo apresentado por Rosenblatt não era muito promissor, devido ao uso de técnicas empíricas, das grandes dificuldades da matemática envolvida e dos poucos recursos computacionais disponíveis na época. A publicação de Minsky e Papert acabou esfriando as pesquisas e praticamente todo o investimento financeiro nesta época foi cancelado.

Enquanto Rosenblatt trabalha no Perceptron, Bernard Widrow da Universidade de Stanford com a ajuda de alguns estudantes desenvolveu um novo modelo de processamento de redes neurais artificiais chamado de Adaline (ADaptive LINear Element), equipado com uma poderosa lei de aprendizado. O princípio de treinamento para as redes Adalines ficou conhecido como a Regra Delta, que foi mais tarde generalizada para redes com modelos neurais mais sofisticados. Mais tarde Widrow criou a Madaline, que era uma generalização multidimensional do adaline. Widrow também fundou a primeira companhia de hardware de neurocomputadores e componentes.

Infelizmente, os anos seguintes foram marcados por um entusiasmo exagerado de muitos pesquisadores, que passaram a publicar mais e mais artigos e livros que fazia uma previsão pouco confiável para a época, sobre máquinas tão poderosas quanto o cérebro humano que surgiriam em um curto espaço de tempo. Isto tirou quase toda a credibilidade dos estudos desta área e causou grandes aborrecimentos aos técnicos de outras áreas.

Um período de pesquisa silenciosa seguiu-se durante 1967 a 1982, quando poucas pesquisas foram publicadas devido aos fatos ocorridos anteriormente. Entretanto, aqueles que pesquisavam nesta época, e todos os que se seguiram no decorrer de treze anos conseguiram novamente estabelecer um campo concreto para o renascimento da área.

No início da década de 80, muitos pesquisadores publicaram inúmeras propostas para a elaboração de desenvolvimento e pesquisa em redes neurais. Foi quando o administrador de programas da DARPA (*Defense Advanced Research Projects Agency*) Ira Skurnick resolveu ouvir os argumentos da neurocomputação e seus projetistas, e divergindo dos caminhos tradicionais dos conhecimentos convencionais, fundou em 1983 pesquisas em neurocomputação da DARPA. Este ato não só abriu as portas para a neurocomputação, como também deu a DARPA o status de uma das líderes mundiais em se tratando de "moda" tecnológica.

O físico e biólogo de reputação mundial John Hopfield também se interessou pela neurocomputação e escreveu vários artigos em 1982 que levaram vários cientistas, matemáticos e tecnólogos altamente qualificados a se unirem nesta nova área emergente. Hopfield reascendeu as pesquisas em neurocomputação, criticando fortemente as teorias apresentadas por Minsky e Papert na década de 50.

Este campo de pesquisa explodiu mesmo em 1986, quando o professor de psicologia da Universidade de Stanford, David E. Rumelhart e seu colega James L. McClelland, professor de psicologia da Universidade de Carnegie Mellon, publicaram o

livro “*Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (vol. 1: Foundations, vol. 2: Psychological and Biological Models)”. Nesse livro eles apresentam um modelo matemático e computacional que propicia o treinamento supervisionado dos neurônios artificiais. Surgiu então o algoritmo *backpropagation*, um algoritmo de otimização global sem restrições.

Em 1987 ocorreu em São Francisco a primeira conferência de redes neurais em tempos modernos. Também foi formada a Sociedade Internacional de Redes Neurais (*International Neural Networks Society - INNS*) juntamente com o *INNS Journal* em 1989, do *Neural Computation* e do *IEEE Transactions on Neural Networks* em 1990.

A partir destes acontecimentos muitas universidades formaram institutos de pesquisa e programas de educação em neurocomputação.

3.3 - O Neurônio Artificial

Em 1943 foi proposto por Warren McCulloch e Walter Pitts o primeiro modelo matemático para uma rede neural, que era bem simples quando comparado com o neurônio biológico que possui uma imensa estrutura complexa e um grande número de detalhes. Porém, o modelo integrador no neurônio de McCulloch-Pitts, é uma boa aproximação que satisfaz alguns dos modelos encontrados na neurofisiologia e que, abstraídos das propriedades fisiológicas dos neurônios biológicos e de suas conexões, é considerado como referência número um para a teoria de Redes Neurais Artificiais. Na figura 3.1 é mostrada a estrutura do neurônio artificial.

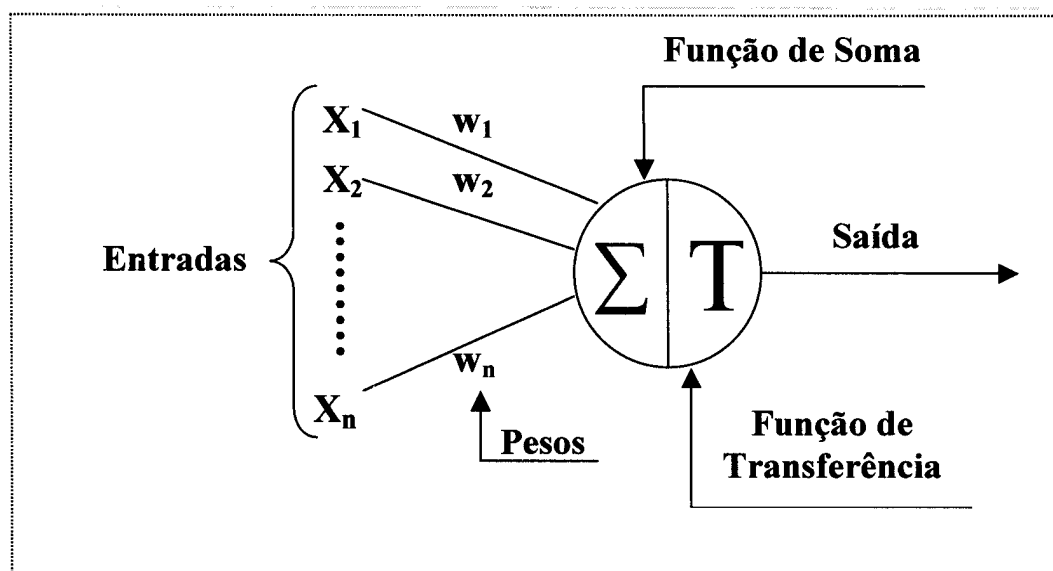


Figura 3.1 – Modelo do neurônio artificial de McCulloch-Pitts

Como o neurônio biológico, o neurônio artificial é um modelo com n terminais de entrada x_1, x_2, \dots, x_n (que representam os dendritos) e apenas um terminal de saída y (representando o axônio). Para emular o comportamento das sinapses, os terminais de entrada do neurônio têm pesos acoplados w_1, w_2, \dots, w_n cujos valores podem ser positivos ou negativos, dependendo de as sinapses correspondentes serem inibitórias ou excitatórias, determinando assim “em que grau” o neurônio deve considerar sinais de disparo que ocorrem naquela conexão.

Um neurônio biológico dispara quando a soma dos impulsos que ele recebe ultrapassa o seu limiar de excitação (*threshold*). O corpo do neurônio, por sua vez, é formado por um mecanismo simples que faz a soma dos valores $x_i w_i$ recebidos pelo neurônio (soma ponderada) e decide se o neurônio deve ou não disparar (saída igual a 1 ou a 0) comparando a soma obtida ao limiar ou *threshold* do neurônio. No modelo de McCulloch-Pitts a ativação do neurônio é obtida através da aplicação de uma “função de ativação”, que ativa ou não a saída, dependendo do valor da soma ponderada das suas entradas. Nesse modelo a função de ativação é dada pela função de limiar descrita na equação 3.1 que terá sua saída ativa quando:

$$\sum_{i=1}^n x_i w_i \geq \theta \quad (3.1)$$

onde n é o número de entradas do neurônio, w_i é o peso associada a entrada x_i e θ é o limiar (*threshold*) do neurônio.

Finalmente, pode-se levantar algumas limitações na descrição desse modelo de neurônio:

1. redes com apenas uma camada só conseguem implementar funções linearmente separáveis;
2. pesos negativos são mais adequados para representar disparos inibitórios;
3. o modelo foi proposto com pesos fixos, não-ajustáveis.

3.4 - Tipos de Função de Ativação

A partir do modelo proposto por McCulloch e Pitts, foram derivados vários outros modelos que permitem a produção de uma saída qualquer, não necessariamente zero ou um, e com diferentes funções de ativação. A figura 3.2 ilustra graficamente quatro funções de ativação diferente: a função linear, a função rampa, a função degrau (*step*) e a função sigmoideal.

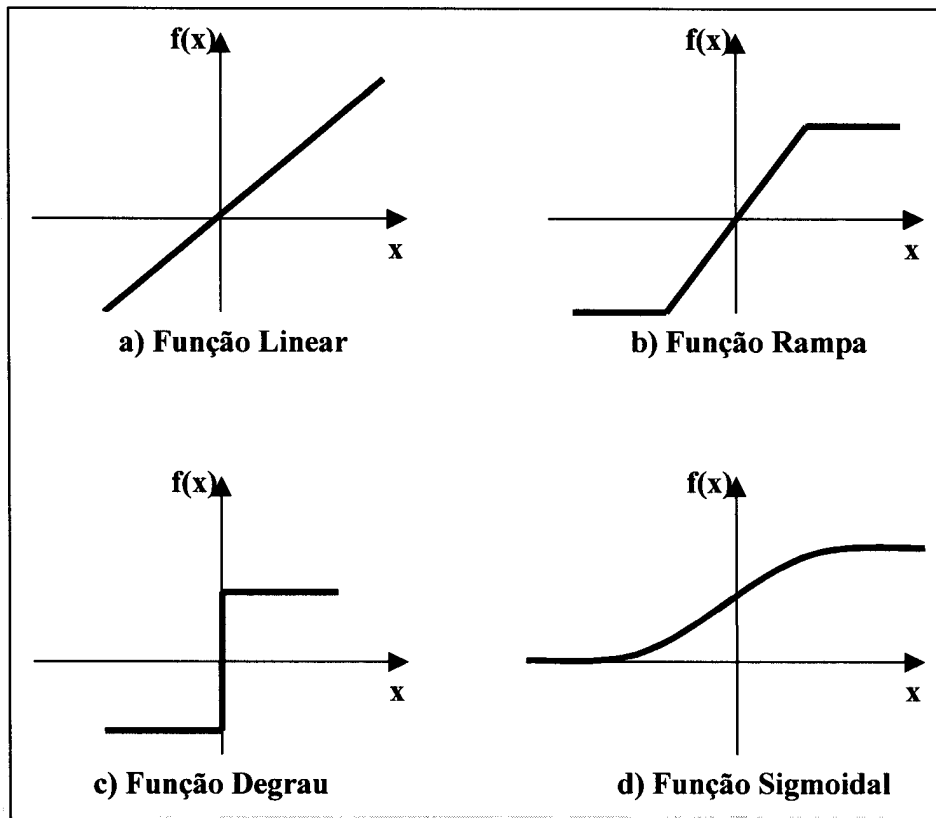


Figura 3.2 – Algumas funções de ativação

A função de ativação linear mostrada na Figura 3.2a é definida pela equação 3.2, onde α é um número real que define a saída linear para os valores de entrada, y é a saída e x é a entrada.

$$y = \alpha x \quad (3.2)$$

Os valores máximo e mínimo da saída são representados por $+\gamma$ e $-\gamma$, respectivamente. A função linear pode ser limitada para produzir valores constantes em uma faixa $[-\gamma, +\gamma]$, e neste caso a função passa a ser a função rampa como mostra graficamente a Figura 3.2b e a equação 3.3. A função rampa é geralmente usada como função não-linear simplificada.

$$y = \begin{cases} +\gamma & \text{se } x \geq +\gamma \\ x & \text{se } |x| < +\gamma \\ -\gamma & \text{se } x \leq -\gamma \end{cases} \quad (3.3)$$

A função de ativação degrau é definida matematicamente abaixo pela equação 3.4, onde a função produz uma saída $+\gamma$ para os valores de x maiores que zero, caso contrário a função produz o valor $-\gamma$.

$$y = \begin{cases} +\gamma & \text{se } x > 0 \\ -\gamma & \text{se } x \leq 0 \end{cases} \quad (3.4)$$

A função de ativação sigmoide é a forma mais comum de função de ativação usada na construção de redes neurais artificiais. Ela é definida como uma função de aumento restrito que exibe suavidade e propriedades assintóticas. Uma das funções sigmoideais mais importantes é a função logística, definida matematicamente na equação 3.5:

$$y = \frac{1}{1 + \exp^{-x/T}} \quad (3.5)$$

onde o parâmetro T determina a suavidade do formato da curva.

Assim como as redes neurais biológicas, o conjunto de vários neurônios artificiais interconectados, formam as redes neurais artificiais como exibida na figura 3.3.

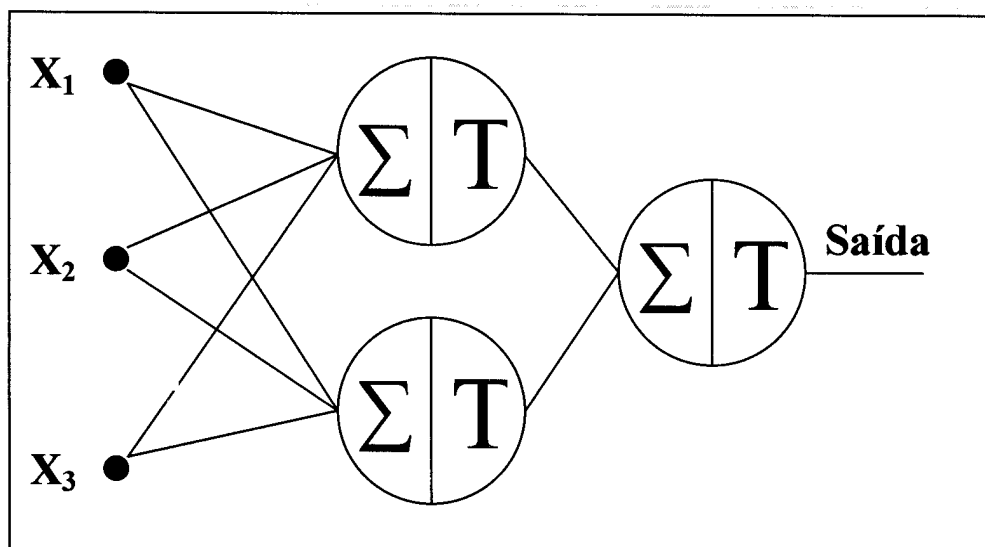


Figura 3.3 – Rede Neural Artificial

3.5 - Arquiteturas

Um outro detalhe importante a ser considerado é a maneira com que os neurônios artificiais podem ser agrupados. Este agrupamento se sucede no cérebro humano de maneira que as informações possam ser processadas de forma dinâmica ou interativa. Biologicamente, as redes neurais são organizadas e construídas de forma tridimensional por componentes microscópicos.

A arquitetura de uma rede é um parâmetro importante na sua concepção, pois restringe o tipo de problema que pode ser tratado pela rede. Redes com uma única camada de neurônios são as redes que possuem um nó entre uma entrada e uma saída da rede. Esse tipo de rede é indicado para a solução de problemas linearmente separáveis, não resolvendo problemas simples como o problema XOR (função lógica “OU” exclusivo). A figura 3.4 abaixo mostra a arquitetura de uma rede com uma única camada de neurônios.

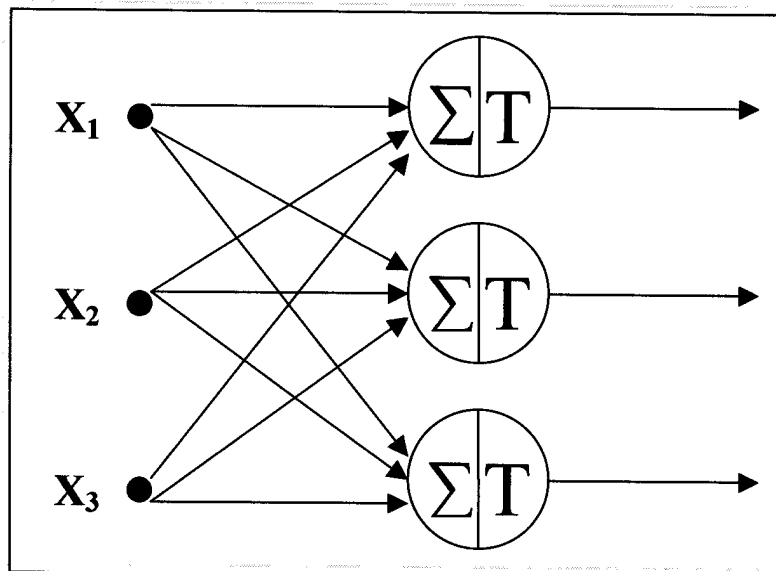


Figura 3.4 – Rede Neural Artificial de uma única camada

As redes neurais artificiais de múltiplas camadas possuem as chamadas camadas escondidas (hidden), que também são chamadas de intermediárias ou ocultas. Esse número de camadas pode ser indeterminado, e estão situadas entre a camada de entrada e a camada de saída da rede neural.

As camadas ocultas são constituídas por neurônios artificiais, da mesma forma com que as camadas externas (entrada e saída) são compostas, e tendo como característica diferenciada o não contato com o mundo externo. Os sinais são passados para os outros neurônios obedecendo às funções de transferência que cada neurônio possui. A figura 3.5 mostra a arquitetura de uma rede neural de múltiplas camadas.

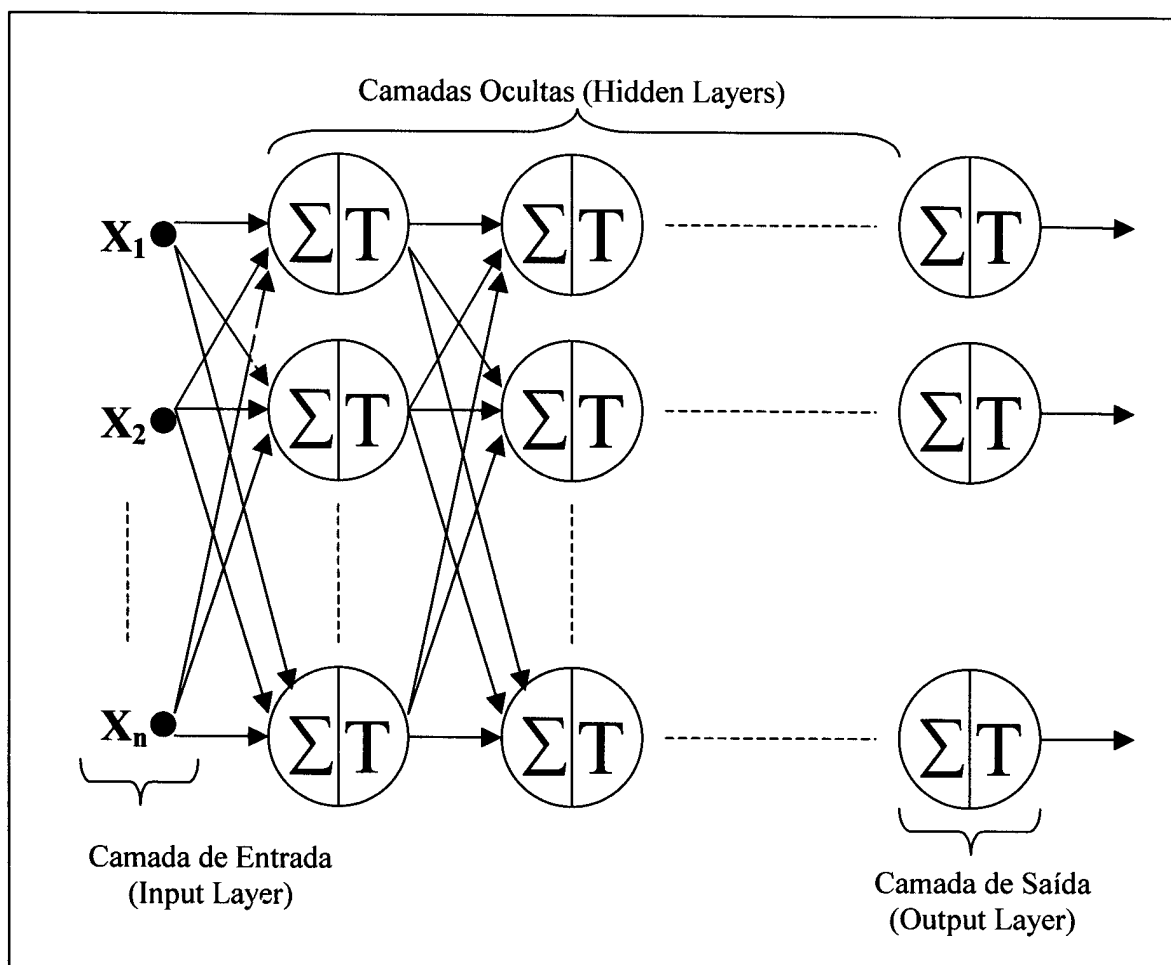


Figura 3.5 – Rede Neural Artificial de múltiplas camadas

Os nodos que compõem a rede neural artificial podem ter conexões do tipo:

- **Redes Diretas (Feedforward ou Acíclicas)** (Figura 3.6): A saída do neurônio na i -ésima camada da rede não pode ser usada como entrada de nodos em camadas de índice menor ou igual a i . Uma aplicação típica para as redes neurais artificiais **feedforward** é de desenvolver modelos não-lineares que também são usados para o reconhecimento e classificação de padrões. Estas redes só possuem uma memória de curtíssimo prazo, sendo as únicas informações armazenadas relativas ao processamento imediatamente anterior de cada neurônio. Seus sinais se propagam somente num sentido e as saídas só dependem dos sinais que estão chegando dos outros neurônios. Uma vez treinadas, estas redes produzem sempre o mesmo resultado de saída, para uma mesma entrada.

Esta rede geralmente é treinada de modo supervisionado e, superficialmente falando, pode ser descrito como: dado um vetor com o padrão a ser reconhecido. Insere-se esse vetor na camada de entrada da rede, que repassará os valores para as camadas intermediárias que a seguem e finalmente para a camada de saída através dos pesos das conexões, obtendo-se dessa forma o resultado. Cada neurônio da rede opera realizando a soma dos pesos de entrada e repassando o resultado para a próxima camada, ou para a saída da rede, através de uma função de ativação não-linear.

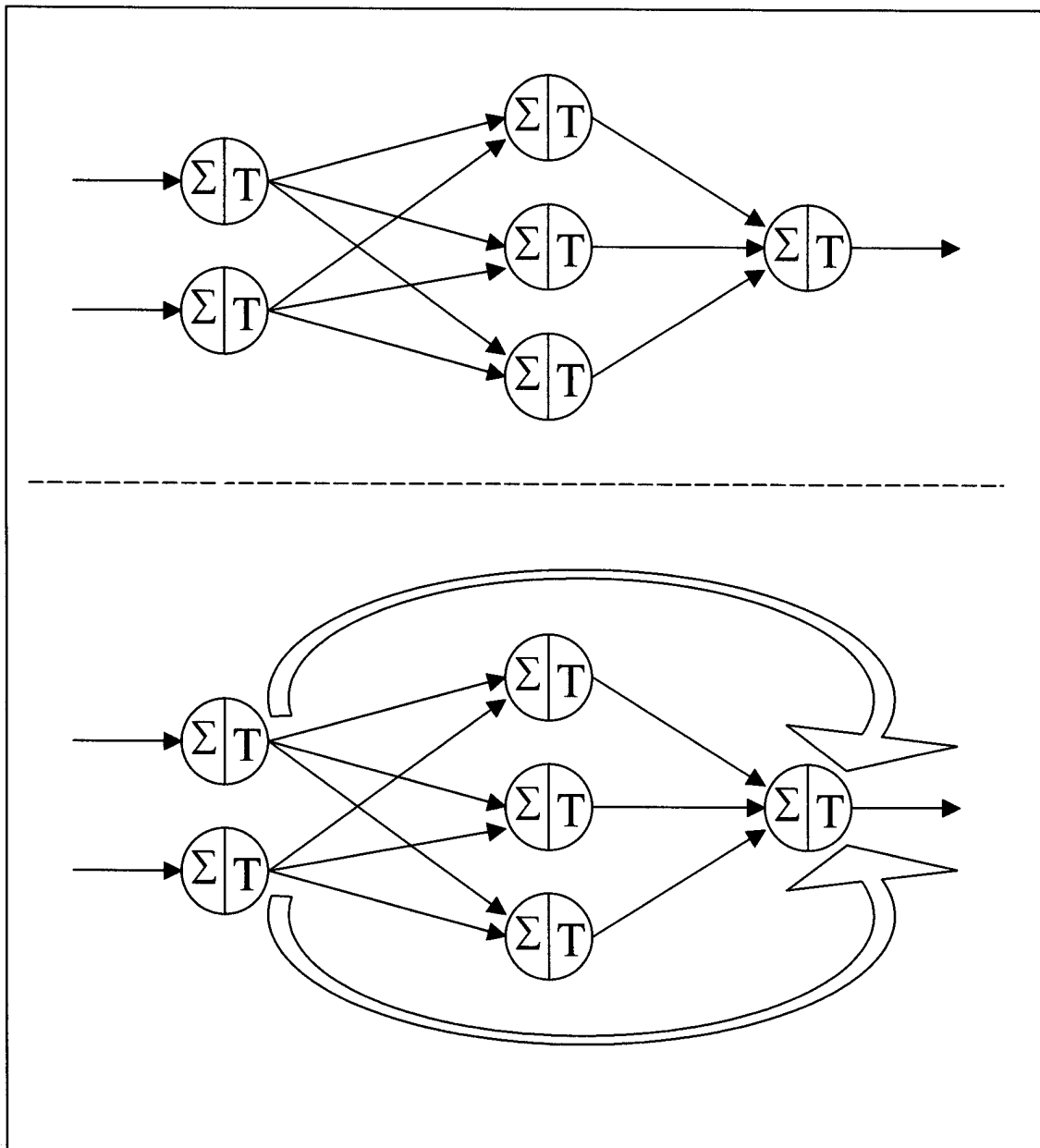


Figura 3.6 – Redes Neurais Artificiais diretas (feedforward ou acíclicas)

- **Redes com Realimentação (Feedback ou Cíclica)** (Figura 3.7): A saída de algum neurônio na *i*-ésima camada da rede é usada como entrada de nodos em camadas de índice menor ou igual a *i*. Se todas as ligações entre os neurônios forem cíclicas, a rede é chamada auto-associativa, estas redes associam um padrão de entrada com ele mesmo, e são particularmente úteis para recuperação ou “regeneração” de um padrão de entrada.

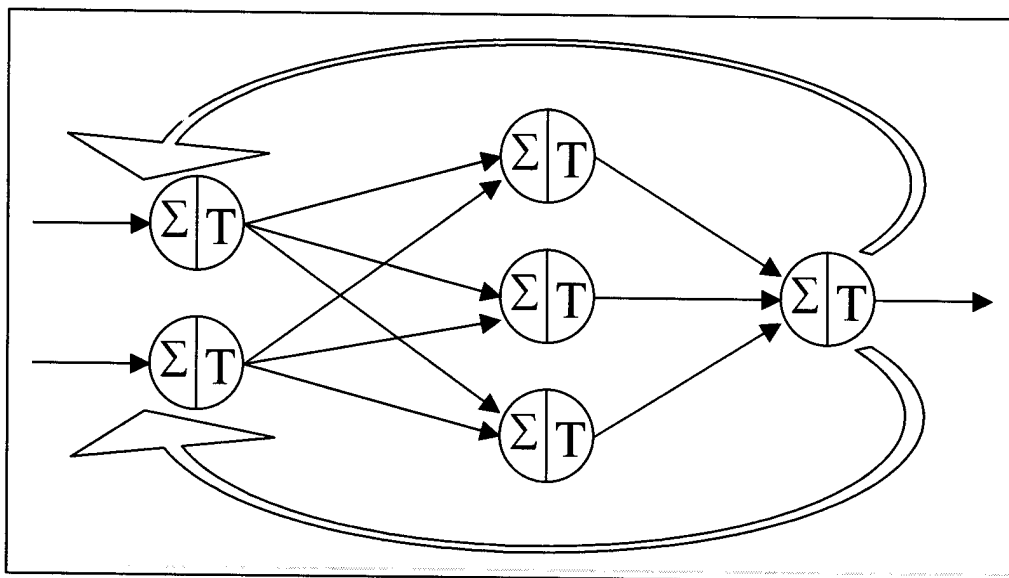


Figura 3.7 – Redes Neurais Artificiais feedback ou cíclica

Finalmente, as redes neurais artificiais podem também serem classificadas quanto a sua conectividade, como exibido na figura 3.8: rede fracamente (parcialmente) conectada – Figura 3.8-a; rede completamente conectada – Figura 3.8-b.

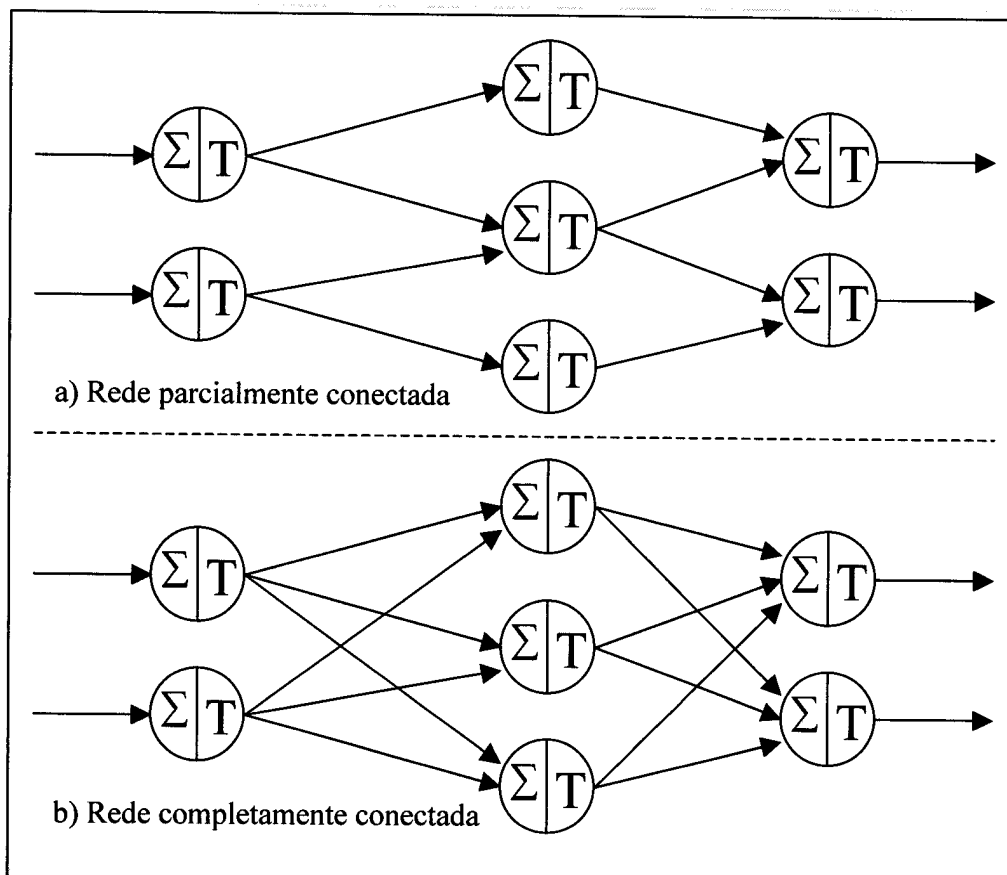


Figura 3.8 – Tipos de conectividade de uma rede neural artificial

3.6 - Aprendizado

As Redes Neurais Artificiais possuem a capacidade de aprender por exemplos e fazer interpolações e extrapolações do que aprenderam. No aprendizado conexionista, procura-se encontrar a intensidade de conexões entre neurônios, sendo que o conjunto de procedimentos bem definidos para adaptar os parâmetros de uma rede neural artificial para que a mesma possa aprender uma determinada função e chamado de *algoritmo de aprendizado*.

Não há muitas noções associadas com "aprendizagem" para justificar a definição do termo de uma maneira precisa. Reconhecendo que nosso interesse particular é nas redes neurais, vamos usar a definição de aprendizagem que é adaptado por Mendel e McClaren(1970):

“Aprendizagem é um processo o qual parâmetros livres de uma rede neural são adaptados por um processo contínuo de simulação através do ambiente no qual a rede é embutida. O tipo de aprendizagem é determinado pela maneira o qual ocorre mudança no parâmetro” [HAY-94].

A definição do processo de aprendizagem implica numa seqüência de eventos:

- 1- A rede neural é estimulada pelo ambiente.
- 2- A rede neural sofre mudança como o resultado dessa estimulação.
- 3- A rede neural responde em um novo caminho para o ambiente, por causa das mudanças ocorridas em sua estrutura interna.

Diversos métodos para treinamento de redes neurais foram desenvolvidos, podendo ser agrupados em dois paradigmas principais: aprendizado supervisionado e aprendizado não-supervisionado.

3.6.1 - Aprendizado Supervisionado

O método de aprendizado supervisionado é o mais comum no treinamento das redes neurais artificiais. É chamado de *aprendizado supervisionado* porque a entrada e saída desejadas para a rede são fornecidas por um supervisor (professor) externo. O objetivo é justar os parâmetros da rede, de forma a encontrar uma ligação entre os pares de entrada e saída fornecidos. A figura 3.9 ilustra o mecanismo do aprendizado supervisionado.

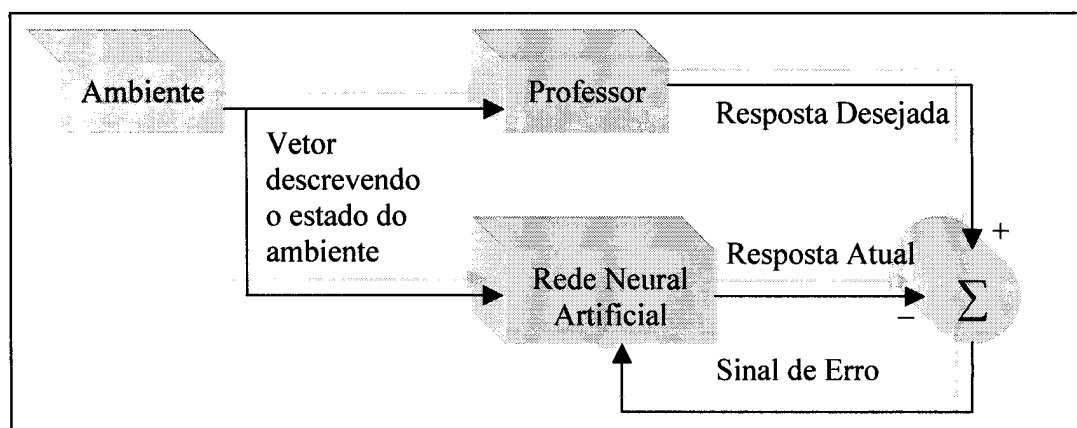


Figura 3.9 – Diagrama de bloco do aprendizado supervisionado

Em termos conceituais, nós podemos pensar no professor como tendo conhecimento de um ambiente que é representado pela configuração de exemplos de entrada e saída. O ambiente é, porém, desconhecido para a rede neural de interesse. Por isso, são expostos pelo ambiente o vetor de treinamento(exemplos) para a rede neural e o professor. O professor é habilitado a prover a rede neural com a resposta desejada para aquele vetor de treinamento. Realmente, a resposta desejada representa a ação para ser executada pela rede neural.

A rede tem sua saída corrente (calculada) comparada com a saída desejada, recebendo informações do supervisor sobre o erro da resposta atual. A cada padrão de entrada submetido à rede compara-se a resposta desejada (que representa uma ação ótima a ser realizada pela rede) com a resposta calculada, ajustando-se os pesos das conexões para minimizar o erro. A soma dos erros quadráticos de todas as saídas é normalmente utilizada como medida de desempenho da rede e também como função de custo a ser minimizada pelo algoritmo de treinamento.

O conjunto de treinamento precisa ser suficientemente grande para conter as informações necessárias para que a rede aprenda os moldes e as relações importantes. Se a rede é treinada somente com um exemplo em um determinado tempo, todos os pesos serão alterados meticulosamente para este fato, os quais poderiam sofrer alterações drásticas no aprendizado de um próximo fato. Conforme um resultado, o sistema precisa aprender com todos os fatos em conjunto, provendo posteriormente o melhor ajuste dos pesos para todo o conjunto de fatos.

Normalmente, as redes neurais artificiais somente manipulam, ou trabalham, com dados numéricos como entrada. Por este motivo, os dados do mundo exterior, devem ser tratados e convertidos para que se possa alimentar a rede. Esta captura de estímulos do mundo real pode ser realizada através de vários tipos de dispositivos, tais como: câmeras de vídeo, diversos tipos de sensores, microfones, entre outros.

Após o treinamento supervisionado, é importante analisar o que a rede pode realizar com os dados que ainda não foram apresentados à mesma. Se o resultado de saída do sistema não for razoável para este novo conjunto de dados (chamado conjunto de teste), presume-se que o treinamento da rede ainda não foi suficiente ou que o número de exemplos de treinamento (entradas) foram insuficientes para que a rede generalize de forma promissora.

Esta avaliação é crítica para assegurar que a rede simplesmente não memorizou um dado conjunto de dados, mas sim aprendeu modelos (padrões) gerais envolvidos na aplicação (generalização). É importante ressaltar que às vezes o problema da generalização é devido à má qualidade dos dados usados para o treinamento e não um problema da rede.

3.6.2 - Aprendizado Não-Supervisionado

O aprendizado não supervisionado é a grande promessa para o futuro, visto que implica que a rede aprenda sem a necessidade de um conjunto de treinamento. Nesse aprendizado, como o próprio nome sugere, não há um professor ou supervisor para acompanhar o processo de aprendizado. Este método está ilustrado na figura 3.10.



Figura 3.10 – Diagrama de bloco do aprendizado não-supervisionado

Para este aprendizado, somente os padrões de entrada estão disponíveis para a rede, ao contrário do aprendizado supervisionado, cujo conjunto de treinamento possui pares de entrada e saída. A partir do momento em que a rede estabelece uma harmonia com as regularidades estatísticas da entrada de dados, desenvolve-se nela uma habilidade de formar representações internas para codificar características de entrada e criar novas classes ou grupos automaticamente. Este tipo de aprendizado só se torna possível quando existe redundância nos dados de entrada. Sem redundância seria impossível encontrar quaisquer padrões ou características dos dados de entrada.

Uma rede com o algoritmo de aprendizado não supervisionado deve ter enfatizado a cooperação entre as camadas de unidades de processamento. A competição entre estas unidades é à base de aprendizado da rede. Normalmente, quando a competição pelo aprendizado ocorre de fato, somente os pesos pertencentes a unidades de processamento vencedora são ajustados.

3.6.3 - Velocidade de Aprendizado

A velocidade com que uma rede neural artificial aprende, depende de vários fatores. A baixa taxa de aprendizado resulta num tempo relativamente grande para a realização do aprendizado da rede, a fim de se produzir um resultado adequado para o sistema em questão. Já com um alta taxa de aprendizado, sua velocidade no aprendizado é mais rápida, mas a rede pode não ser capaz de fazer uma possível discriminação fina em relação aos sistemas que aprendem de forma mais lenta.

Geralmente, vários fatores, além do tempo de aprendizado precisam ser considerados quando se aborda a tarefa de aprendizado em treinamento de rede. Abaixo estão alguns fatores que estão fortemente acoplados ao processo de aprendizado de uma rede neural artificial:

- a complexidade da rede;
- o número de camadas (tamanho);
- o paradigma da seleção;
- a arquitetura adotada;
- o algoritmo de aprendizado;
- as regras empregadas;
- a precisão desejada.

Todos esses fatores contribuem e alteram o tempo de treinamento da rede. A mudança de qualquer um destes fatores pode estender o tempo de treinamento para uma razão não muita significativa, ou resultando em uma precisão não satisfatória.

A maioria dos algoritmos de aprendizado possui alguma provisão para a taxa de aprendizado ou em muitos casos, uma constante de tempo. Normalmente, este valor está compreendido entre zero e um $[0,1]$. Se a taxa de aprendizado exceder o valor máximo, o algoritmo de aprendizado irá corrigir os pesos da rede.

Pequenos valores para a taxa de aprendizado não retificarão os erros tão rapidamente, mas se pequenos passos são tomados na correção de erros, há uma grande possibilidade de se alcançar uma boa convergência de aprendizado da rede.

3.6.4 - Aprendizado na Correção de Erros

O aprendizado na correção de erros procura diminuir a diferença entre a soma ponderada das entradas pelos pesos (saída calculada pela rede) e a saída desejada, ou seja, o erro encontrado na saída da resposta atual da rede. A correção de erros pode ser vista a equação 3.6, onde $d(t)$ é a resposta desejada ou resposta alvo para o neurônio no instante de tempo t , $y(t)$ é a resposta atual (calculada) e $e(t)$ é a resposta desejada.

$$e(t) = d(t) - y(t) \quad (3.6)$$

A equação 3.7 mostra a forma genérica para alteração dos pesos para a correção de erros, onde η é o parâmetro da taxa de aprendizado e $x_i(t)$ é a entrada para o neurônio i no tempo t . Segundo essa equação, o ajuste dos pesos deve ser proporcional ao produto do erro pelo valor de entrada da sinapse naquele instante de tempo.

$$w_i(t+1) = w_i(t) + \eta e(t) x_i(t) \quad (3.7)$$

A ultima proposta para o aprendizado na correção de erros é a *minimização da função de custo* baseado no sinal de erro $e(t)$, tal que a resposta atual de cada saída do neurônio na rede aproxima-se da resposta desejada para aquele neurônio. Um critério usualmente utilizado para a função de custo é o critério do erro quadrado médio, definido pela *soma dos erros quadráticos* na equação 3.8. O número de nodos na saída da rede é dado por k , d_i é a saída desejada para o nodo i , e y_i é a saída corrente da rede.

$$F(w) = \frac{1}{2} \sum_{i=0}^k (d_i - y_i(w))^2 \quad (3.8)$$

Com relação à superfície de erro obtida através da equação 3.8, dependendo do tipo de unidade de processamento utilizada para construir a rede, podem-se identificar duas situações diferentes:

- A rede ser formada inteiramente por unidades de processamento lineares, onde neste caso a superfície de erro é dada exatamente pela função quadrática dos pesos da rede, isto e, dizendo que a rede possui um único mínimo global;

- A rede ser formada por unidades de processamento não-lineares. Neste caso, a superfície de erro poderá ter, além do mínimo global, um ou mais mínimos locais.

Em ambas situações, o objetivo deste método de aprendizado é partir de um ponto arbitrário da superfície movendo-se até o mínimo global. Na primeira situação só existe um mínimo global, tratando-se de uma superfície de erro quadrática. Na segunda situação o mínimo global nem sempre é alcançado, já que as saídas não-lineares geram superfícies de erro irregulares, podendo deixar a rede se estabilizar em um mínimo local indesejável.

3.6.5 - Aprendizado Hebbiano

O princípio do aprendizado de Hebb é a primeira e indiscutivelmente a mais famosa regra de aprendizado. Foi postulado pelo biólogo e psicólogo Donald Hebb em 1949 em seu famoso livro *The Organization of Behavior*. Retiramos do livro de Hebb o seguinte conceito para esta regra de aprendizado:

Quando o axônio da célula A está perto o bastante para excitar a célula B e repetidamente ou persistentemente participa de sua ativação, algum processo de crescimento ou mudança metabólica acontecem em uma ou ambas as células tal que, a eficiência de A como uma das células que dispara em B, é aumentada.

Esta regra de aprendizado propõe que o peso de uma conexão sináptica deve ser ajustado se houver sincronismo entre os “níveis de atividade” das entradas e saídas. Em outras palavras, se um neurônio recebe uma entrada, proveniente de um outro neurônio, e ambos estão ativos (possuem o mesmo sinal), os pesos entre os neurônios precisam ser excitados (fortalecimento desta sinapse). Se a ativação é assíncrona, então a sinapse deve ser enfraquecida ou eliminada.

A sinapse hebbiana possui quatro características principais:

- *Mecanismo Interativo* → Aqui nós notamos que a ocorrência da mudança da sinapse hebbiana depende (estatisticamente) do nível de atividade de ambos os lados da sinapse (pré e pós-sináptica);
- *Mecanismo Local* → Uma sinapse hebbiana é uma transmissão com sinais contínuos que produz modificações sinápticas locais que são entradas específicas. É a partir do mecanismo local que sinapses hebbianas efetuam o aprendizado não-supervisionado;
- *Mecanismo dependente do tempo* → Este mecanismo refere-se ao fato de que modificações em uma sinapse hebbiana dependem do momento exato de ocorrência das atividades pré e pós-sinápticas;
- *Mecanismo correlacional ou conjuncional* → A sinapse hebbiana pode ser chamada de sinapse conjuncional pelo fato de a ocorrência conjunta de atividades pré e pós-sinápticas ser suficiente para que haja uma modificação. Além disso, pode também ser chamada de sinapse correlacional porque uma correlação entre estas mesmas atividades também é suficiente para gerar mudanças.

Matematicamente, o postulado de Hebb consiste em uma regra para a mudança do peso sináptico w_{ij} (regra do produto de atividade) como mostrada na equação 3.9, onde η é uma constante positiva que determina a taxa de aprendizado.

$$\Delta w_{ij}(t) = \eta y_i(t) x_j(t) \quad (3.9)$$

Apesar de os vetores de entrada e saída serem fornecidas a rede (como no aprendizado supervisionado), a regra de Hebb é classificada como aprendizado não-supervisionado, já que não existe supervisor externo para verificar a qualidade da resposta da rede para prover um ajuste de pesos. Neste caso, o treinamento da rede é feito independentemente da resposta atual da rede, através de um mecanismo local à sinapse.

3.6.6 - Aprendizado por Competição

Numa aprendizagem competitiva, como o nome insinua, a saída dos neurônios da rede neural competem entre si para que apenas um seja ativado, sendo assim um caso particular de aprendizado não-supervisionado. Existe, portanto, uma competição entre as unidades de saída para decidir qual delas será a vencedora e, conseqüentemente, terá sua saída ativada e seus pesos atualizados no treinamento.

Existem três elementos básicos para a regra de aprendizado competitiva (Rumelhart e Zipser, 1985):

- Um conjunto de neurônios que são todos parecidos exceto por alguns pesos sinápticos distribuídos aleatoriamente, e que então responde diferentemente para um determinado conjunto de padrões de entrada.
- O **limite** é imposto na “força” de cada neurônio.
- O mecanismo que permite os neurônios competirem pelo direito de responder para dado subconjunto de entradas, tal que somente uma saída de neurônio, ou somente um neurônio por grupo é ativado de cada vez. O neurônio que vence a competição é chamado "*winner takes all*".

Adequadamente, os neurônios individuais da rede neural aprende a se especializar nos conjuntos de padrões semelhantes, e assim tornam-se *detectores de característica*. Numa forma mais simples de aprendizagem competitiva, a rede neural possui uma camada simples de saída de neurônios, cada qual é completamente conectada aos nós de entrada. A rede pode incluir *conexões laterais* entre os neurônios, como indicado na figura 3.6.6. Na arquitetura da rede descrita aqui, as conexões laterais executam uma **inibição lateral**, com cada neurônio tendendo a inibir o neurônio o qual é conectada lateralmente. O resto das conexões sinápticas na rede da figura 3.11 são excitatórias.

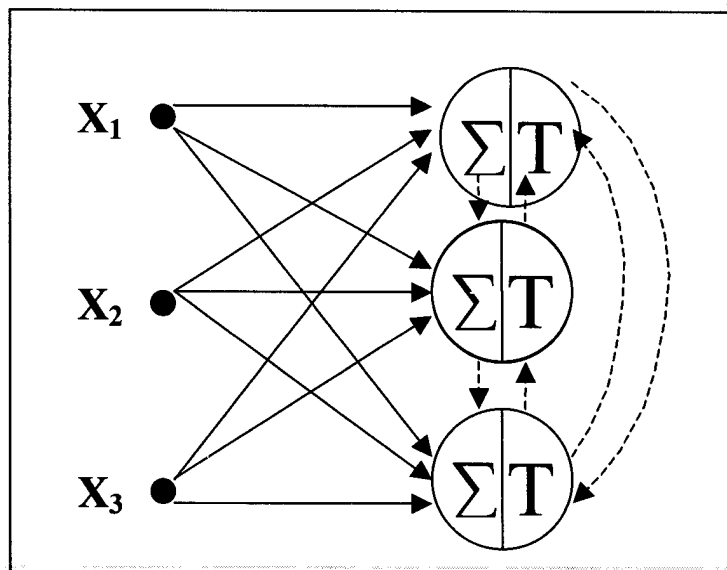


Figura 3.11 – Arquitetura do aprendizado competitivo com conexões feedforward (excitatória) dos nodos de entrada para os neurônios, e conexões laterais (inibitória) entre os neurônios; as correções laterais são identificadas por setas pontilhadas [HAY-94].

Um problema neste algoritmo é que a unidade de saída pode se tornar dominante e ficar ativa o tempo todo podendo captar para si todo o espaço de entradas. A solução para tal problema é racionar os pesos, de tal forma que a soma dos pesos sobre as linhas de entrada de uma unidade seja limitada a 1. Para aumentar o peso de uma conexão, é preciso diminuir o peso de alguma outra, conforme descrito na equação 3.10:

$$\Delta w_j = \frac{\eta x_j}{M - \eta w_i} \quad (3.10)$$

Onde para todo $j = 1, 2, \dots, n$, o w_j é o peso da conexão da unidade de entrada j com a unidade ativa, x_j é o valor do j -ésimo bit de entrada, M é o número de unidades de entrada ativas no vetor de entrada e η é o parâmetro da taxa de aprendizado.

3.6.7 - Aprendizado por Reforço

O aprendizado por reforço, ilustrado na figura 3.12, pode ser visto como um caso particular de aprendizado supervisionado. A principal diferença é porque a medida de desempenho não é baseada no conjunto de respostas desejadas usando um critério de erro conhecido (aprendizado supervisionado), mas seu desempenho baseia-se em qualquer medida que possa ser fornecida ao sistema. No aprendizado por reforço, a única informação de realimentação fornecida à rede é se uma saída está correta ou não, isto é, não é fornecida à rede a resposta correta para o padrão de entrada.

O aprendizado por reforço é uma forma de aprendizado on-line obtido por um mapeamento de entrada-saída através de um processo de triagem e erro desenvolvido para

maximizar o índice de desempenho escalar chamado *signal de reforço*. Em outras palavras, o conceito de aprendizado por reforço pode ser dado pela seguinte definição:

Se uma ação tomada por um sistema de aprendizagem é seguida de estados satisfatórios, então a tendência do sistema de produzir esta ação particular é reforçada. Se não, a tendência do sistema de produzir esta ação é enfraquecida [HAY-94].

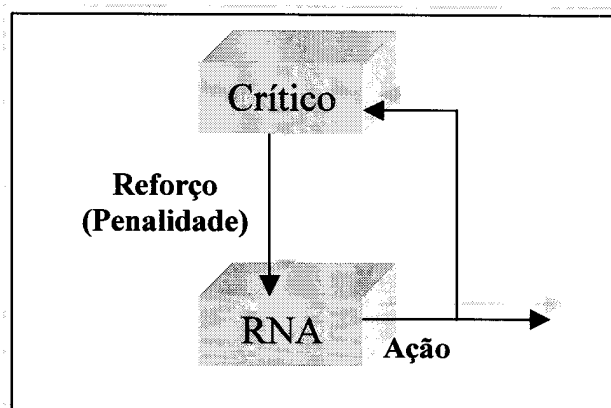


Figura 3.12 – Aprendizado por Reforço

O paradigma do aprendizado por reforço pode ser:

- *Aprendizagem associativa*: o meio fornece outras informações além do reforço, e um mapeamento, na forma estímulo-ação, deve ser aprendido;
- *Aprendizagem não-associativa*: o sinal de reforço é a única entrada que o sistema recebe do meio. O sistema seleciona uma única ação ótima, em vez de associar diferentes ações com diferentes estímulos.

3.6.8 - A Regra Delta de Widrow

A regra delta ou Regra de Widrow Hoff, por ter sido desenvolvida por Bernard Widrow e Ted Hoff, pode ser considerada uma variação da regra de Hebb. Esta regra baseia na simples idéia de contínua modificação da intensidade e a importância das conexões de entrada, reduzindo consideravelmente a diferença entre o valor de saída desejado e o atual valor de saída da unidade de processamento. Esta regra pode ser expressa matematicamente na equação 3.11, onde $w(t+1)$ é o ajuste dos pesos da conexão, η é a taxa de aprendizagem, e e é o erro encontrado na rede ($e(t) = d(t) - y(t)$) e $x(t)$ é o vetor de entrada.

$$w(t+1) = w(t) + \eta e x(t) \quad (3.11)$$

A regra delta altera os pesos sinápticos de modo que minimize o erro quadrático da rede, trabalhando da seguinte forma: o erro calculado na saída é transformado pela derivação da função de transferência e conseqüentemente usado para ajustar os pesos de entrada da camada prévia da rede, ou seja, o erro é retro-propagado as camadas anteriores,

sendo uma de cada vez. O processo de retro-propagação dos erros da rede continua até que a primeira camada da rede seja alcançada.

Quando se utiliza a regra delta, é importante assegurar que o conjunto de dados de entrada está disposto de forma aleatória ou gerado de forma randômica. Uma vez mal organizado, este conjunto de treinamento pode conduzir a não convergência da precisão desejada, impossibilitando o aprendizado do problema em questão.

3.7 - Rede Perceptron

As redes neurais artificiais com função de ativação foram inicialmente estudadas por Frank Rosenblatt em 1985, as quais foram chamadas por ele de Perceptrons. Este modelo proposto por Rosenblatt era composto por uma estrutura de rede, tendo como unidades básicas nodos de MCP (neurônio artificial de McCulloch e Pitts), e por uma regra de aprendizado. Alguns anos mais tarde, Rosenblatt demonstrou o teorema de convergência do Perceptron, que mostra que um nodo MCP treinado com o algoritmo de aprendizado do Perceptron sempre converge caso o problema em questão seja linearmente separável.

A topológica original descrita por Rosenblatt era composta por unidades de entrada (retina), por um nível intermediário formado pelas unidades de associação e por um nível de saída formado pelas unidades de resposta. Embora esta topologia original possua três níveis ela é conhecida como Perceptron de uma única camada, já que somente o nível de saída (unidades de resposta) possui propriedades adaptativas. A figura 3.13 mostra um esboço da topologia do Perceptron.

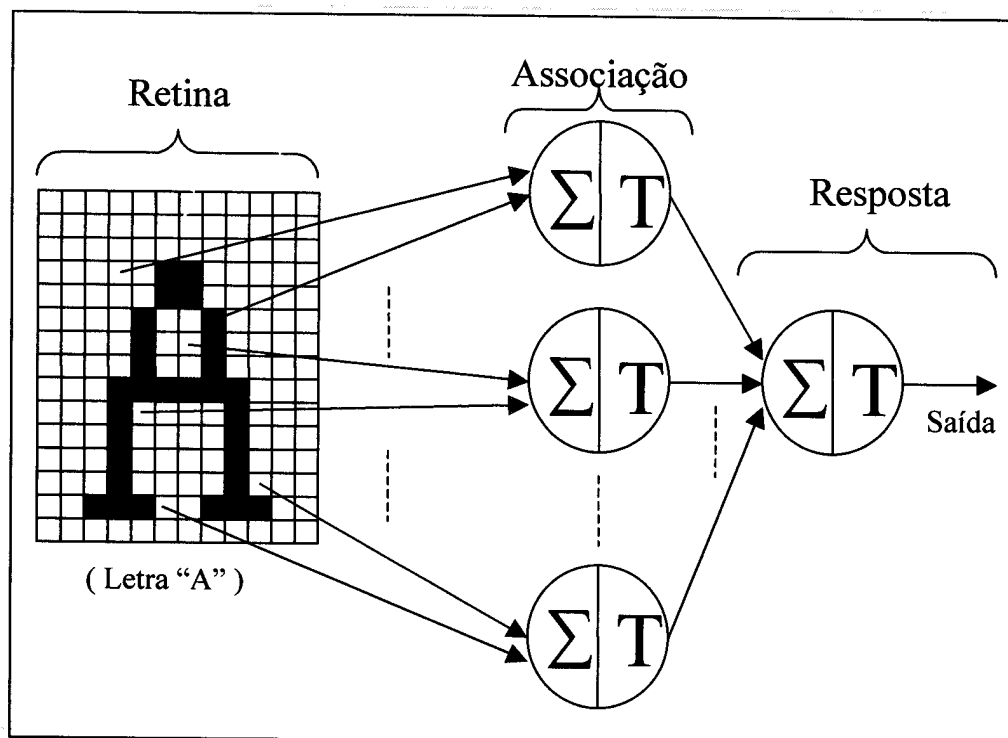


Figura 3.13 – Topologia de um Perceptron simples com uma única camada

A figura 3.14 mostra a distribuição das entradas possíveis em um plano cartesiano. As entradas estão divididas em duas classes A e B, e representam os grupos de entradas que devem ser reconhecidos (distinguidos) pela rede neural artificial. A reta é o resultado da combinação de entradas e pesos (dos neurônios) que resultam em uma equação da reta, desta forma a rede neural artificial gera uma saída igual a *zero* se a entrada estiver sobre a reta, por outro lado gera um uma saída *maior que zero* se a entrada estiver em um dos lados desta reta, e gera uma saída *menor que zero* se a entrada estiver no outro lado desta mesma reta. Em resumo o gráfico mostra como podemos encontrar um conjunto de pesos para cada entrada de forma a colocar a reta em uma posição tal que os exemplos possam ser separados nas duas classes.

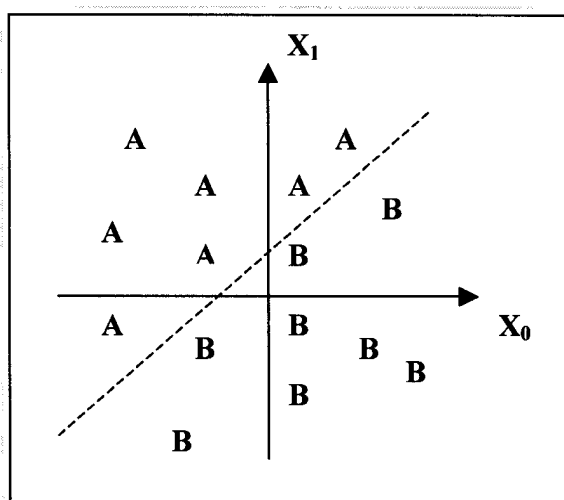


Figura 3.14 – Regiões de decisão de um Perceptron

Cabe ainda uma ressalva, a figura 3.14 é um exemplo simplificado, que separa apenas dois valores de entrada, e para tanto só é necessária uma reta, já para 3 entradas seria necessário usar um plano, e para N valores de entrada é necessário um hiperplano de N dimensões.

Um conceito que pode ser melhor entendido através da figura 3.14 é o deslocamento (*Bias ou Viés*), que é adicionado ao resultado da função somatória. Este deslocamento é necessário porque a equação da reta gerada pela rede é do tipo " $ax + by = 0$ ", este tipo de equação sempre passa pela origem $(0,0)$, porém, no entanto, muitas vezes não é adequado (como mostra o exemplo). Para permitir que a reta seja deslocada de forma que não precise passar exatamente sobre a origem usa-se o *viés*.

Inspirado também pelas idéias de McCulloch, Rosenblatt compôs a rede Perceptron por uma camada de entrada, onde cada elemento pertencente à camada de entrada fazia a distribuição do sinal que ele recebia para todas as unidades de processamento. Os neurônios eram essencialmente compostos por unidades sigma e de funções de transferência, sendo que estas, eram responsáveis pelas soma ponderada dos sinais das conexões com os dados de entrada. Foi adicionada a camada de entrada um elemento especial chamado *viés*, o qual possui sinal de valor sempre um. A conexão entre o viés e a unidade sigma tem peso w_0 , que por sua vez é ajustado da mesma maneira como os demais pesos o são. O modelo de Rosenblatt pode ser visto na figura 3.15.

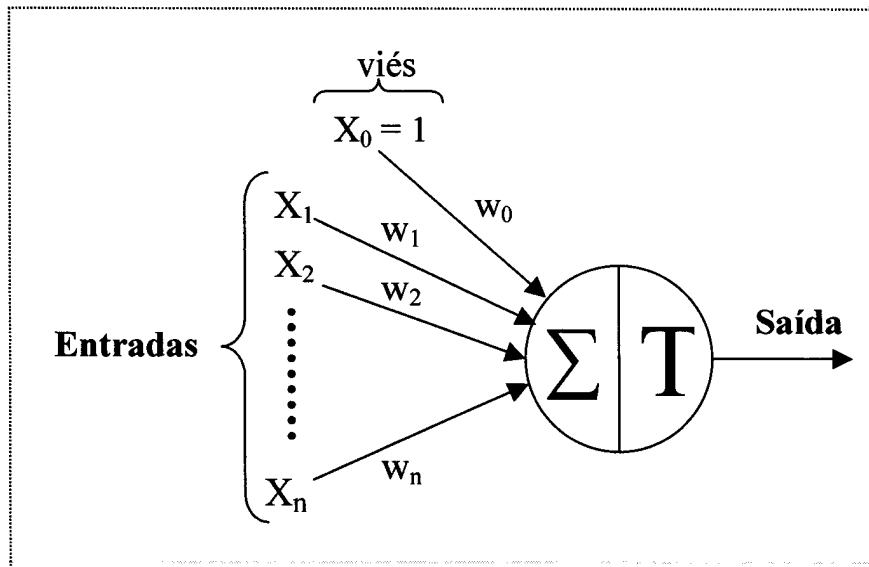


Figura 3.15 – Unidade de processamento do Perceptron

O nível de ativação de uma rede Perceptron é dado pela soma ponderada dos pesos sinápticos com os valores de entrada, $\sum x_i \cdot w_i$. Estas redes usam uma função de transferência do tipo *hard-limiter* (limite ríspido), onde a ativação do limiar resulta num valor de saída **1**, ou **-1** caso contrário. Dados os valores de entrada x_i , os pesos w_i , e um limiar t , o Perceptron computa os valores de saída da seguinte maneira:

$$\begin{cases} 1 & \text{se } \sum x_i w_i \geq t \\ -1 & \text{se } \sum x_i w_i < t \end{cases} \quad (3.12)$$

As redes Perceptron usam como configuração, o treinamento supervisionado. O Perceptron altera os seus pesos, visando reduzir o erro. Para fazer os ajustes dos pesos sinápticos da rede é utilizado a equação geral 3.13, que leva sempre a uma solução caso as classes em questão seja linearmente separáveis.

$$\mathbf{w}(t+1) = \mathbf{w}(t) + \eta e \mathbf{x}(t) \quad (3.13)$$

onde η é o parâmetro da taxa de aprendizado, e é o erro encontrado ($e = d - y$) e $\mathbf{x}(t)$ é o vetor de entrada do nodo.

O Perceptron simples sempre chega numa solução para o problema de separação de duas classes linearmente separáveis em um tempo finito. De maneira geral, o algoritmo de treinamento para um nodo do perceptron simples pode ser descrito da seguinte forma, como ilustra a figura 3.16.

1. Iniciar o η e o vetor de pesos w .
2. Repetir;
3. Para cada par do conjunto de treinamento X
 - 3.1. Atualizar o vetor de pesos cada um dos nodos da rede segundo a regra da equação 3.13;
4. Até $e = 0$ para todos os p elementos do conjunto de treinamento em todos os nodos da rede.

Figura 3.16 – Passos do algoritmo Perceptron simples

3.7.1 - Limitações: O problema do OU-EXCLUSIVO

Os modelos de um único nível baseados no Perceptron sofreram um grande abalo após a publicação de Minsky que demonstrou uma grande limitação na sua capacidade de aprendizado. A pergunta crítica feita por Minsky foi: como uma rede de neurônios de um único nível, pode aprender um padrão de uma simples função lógica “ou exclusivo” (XOR).

O problema das redes de um nível baseadas em Perceptron é que elas dividem o espaço de soluções em duas regiões distintas, mas em uma função lógica XOR não se consegue abranger as duas respostas em uma só região. Na figura 3.16 é mostrada uma representação geométrica do domínio de entrada das funções AND, OR e XOR. Para auxiliar na compreensão da figura 3.16, a tabela 3.1 mostra a correspondência entre os gráficos e estas funções. Quando a entrada $E1$ é igual a zero usamos o valor -1 em $X1$ e quando a entrada $E1$ é igual a 1 usamos o valor $+1$ em $X1$, e a mesma coisa vale para $E2$ e $X2$, além disso quando a saída é igual a 1 esta é representada no gráfico pela letra “A” e quando a saída é igual a 0 esta é representada pela letra “B”.

Como pode ser visto na figura 3.17, o problema fica claro quando se verifica que no gráfico XOR as duas letras “A” não podem ser separadas das letras “B” por apenas uma reta, como ocorre nas portas lógicas AND e OR.

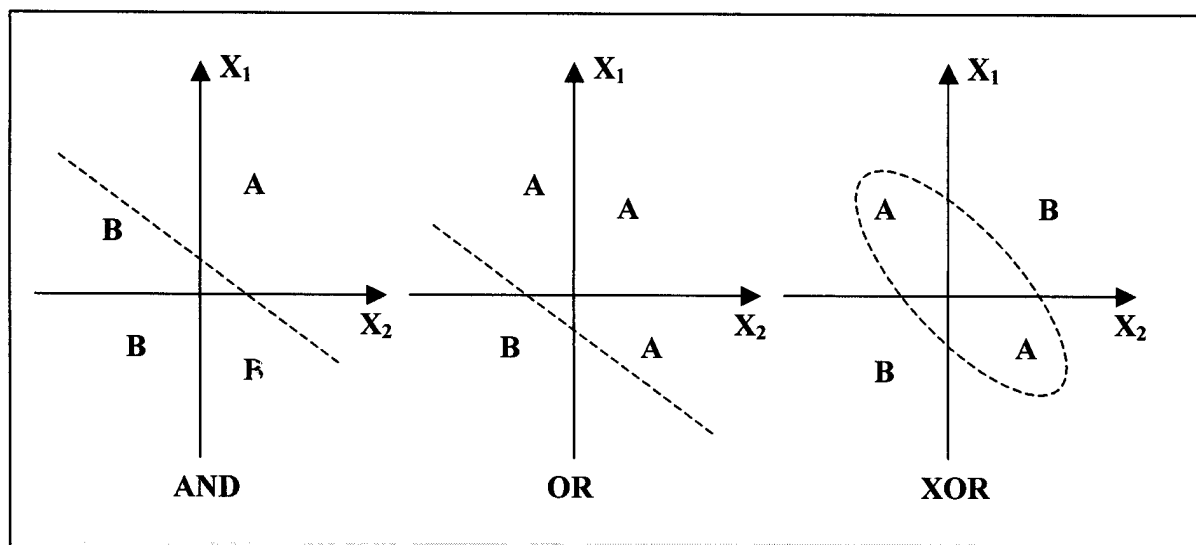


Figura 3.17 – Representação geométrica do espaço de entrada

E1	X1	E2	X2
0	-1	0	-1
0	-1	1	+1
1	+1	0	-1
1	+1	1	+1

AND	OR	XOR
0	0	0 (B)
0	1	1 (A)
0	1	1 (A)
1	1	0 (B)

Tabela 3.1 – Correspondência entre as portas lógicas

3.8 - Rede Adaline

Enquanto Rosenblatt estava desenvolvendo o Perceptron, Widrow e Hoff, estavam trabalhando em uma linha de pesquisa similar a de Rosenblatt, mais conhecida como ADALINE (chamado inicialmente por ADaptive LINEar Element, e posteriormente por ADaptive LINEar NEuron), o qual refere-se a uma única unidade de processamento com um limiar não linear.

Este modelo utiliza o algoritmo de aprendizado *LMS* (*Least Mean Square*), também conhecido como regra delta de Widrow, como abordado anteriormente. O algoritmo LMS procura minimizar o erro médio quadrático entre a saída desejada e a saída apresentada por uma rede semelhante ao Perceptron. Na figura 3.18 temos o esquema de um neurônio tipo Adaline.

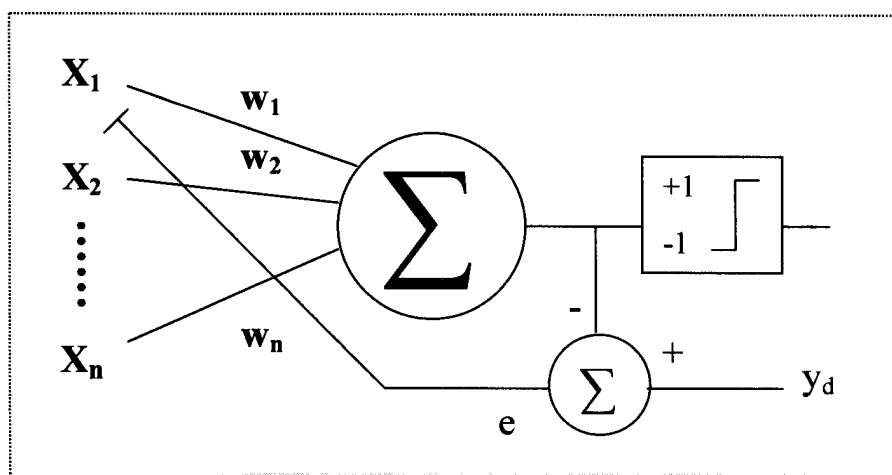


Figura 3.18 – Diagrama esquemático de um nodo Adaline

As características do Adaline são muito semelhantes às características do Perceptron. O Adaline, assim como o Perceptron, cria também uma reta limite entre duas regiões de decisão. A diferença está no algoritmo de aprendizado e nas estruturas que devem ser utilizadas para implementá-lo. A arquitetura é de uma rede de uma única camada disposta linearmente com cada neurônio possuindo uma retro-alimentação do erro na sua própria saída, como mostrado na figura 3.17.

Segundo P. Simpson, A única diferença entre os dois é que o Adaline não usa uma função de transferência em seus neurônios quando o erro é calculado, fazendo apenas a soma ponderada. A função de transferência (*hard limiter*) só era empregada para obter a saída final do neurônio.

A rede Adaline utiliza o mesmo algoritmo que o Perceptron para a atualização dos seus pesos como mostrado pelas equações 3.14 e 3.13. Porém, cabe reforçar que estas equações foram obtidas para nodos com estruturas diferentes e para condições diferentes de ajuste. A equação de ajuste para o nodo Adaline foi obtida para a saída linear do nodo, enquanto a equação de ajuste do Perceptron foi obtida para a saída do nodo após a aplicação da função de ativação.

$$\mathbf{w}(t+1) = \mathbf{w}(t) + \eta \mathbf{e} \mathbf{x}(t) \quad (3.14)$$

A solução para problemas como a porta lógica XOR, onde encontra-se a necessidade na delimitação de ambientes não lineares, são as redes de múltiplos níveis como o Madaline que é mostrado a seguir.

3.9 - Rede Madaline

O Madaline é uma extensão proposta por Widrow ao modelo Adaline. Como o Adaline só é capaz de aprender a reconhecer um padrão simples, a partir dele, surgiu a idéia de criar um elemento composto, chamado Madaline (*Many Adaline*). O Madaline

apenas agrupa a resposta de dois ou mais Adalines, realizando um “e lógico”(AND), “ou lógico” (OR) ou pela obtenção de maioria dos valores obtidos das saídas de um conjunto de neurônios do tipo Adaline. Com isso, é possível fazer com que uma saída da rede seja ativa para mais de um tipo de padrão de entrada. Para isso, basta treinar cada elemento interno do Madaline para reconhecer o padrão em uma dada posição ou orientação e fazer o “ou” das saídas destes elementos.

Na figura 3.19 pode-se ver a relação entre número de camadas(níveis) de diferentes modelos de redes e a formação das regiões de decisão da Madaline.

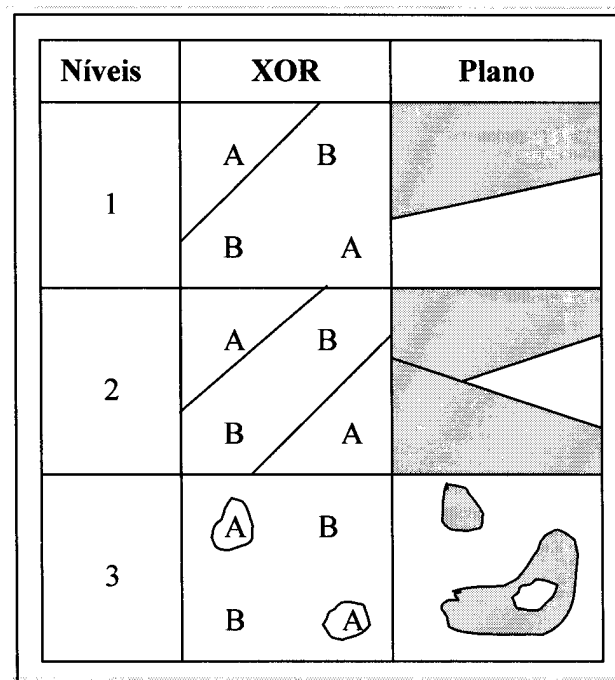


Figura 3.19 – Regiões de decisão da Madaline

3.10 - Rede Perceptron de Múltiplas Camadas

As redes Perceptrons de Múltiplas Camadas (MultiLayer Perceptron -MLP) apresentam um poder computacional muito maior do que as redes sem camadas intermediárias. Ao contrário destas redes, MLPs podem tratar dados que não são linearmente separáveis. A figura 3.20 apresenta uma rede típica de múltiplas camadas.

Tipicamente, a rede consiste de um conjunto de unidades sensoriais (nodos de fonte) que constituem a *camada de entrada* (input layer), uma ou mais *camadas intermediárias ou escondidas* (hidden layers) e a *camada de saída* (output layer).

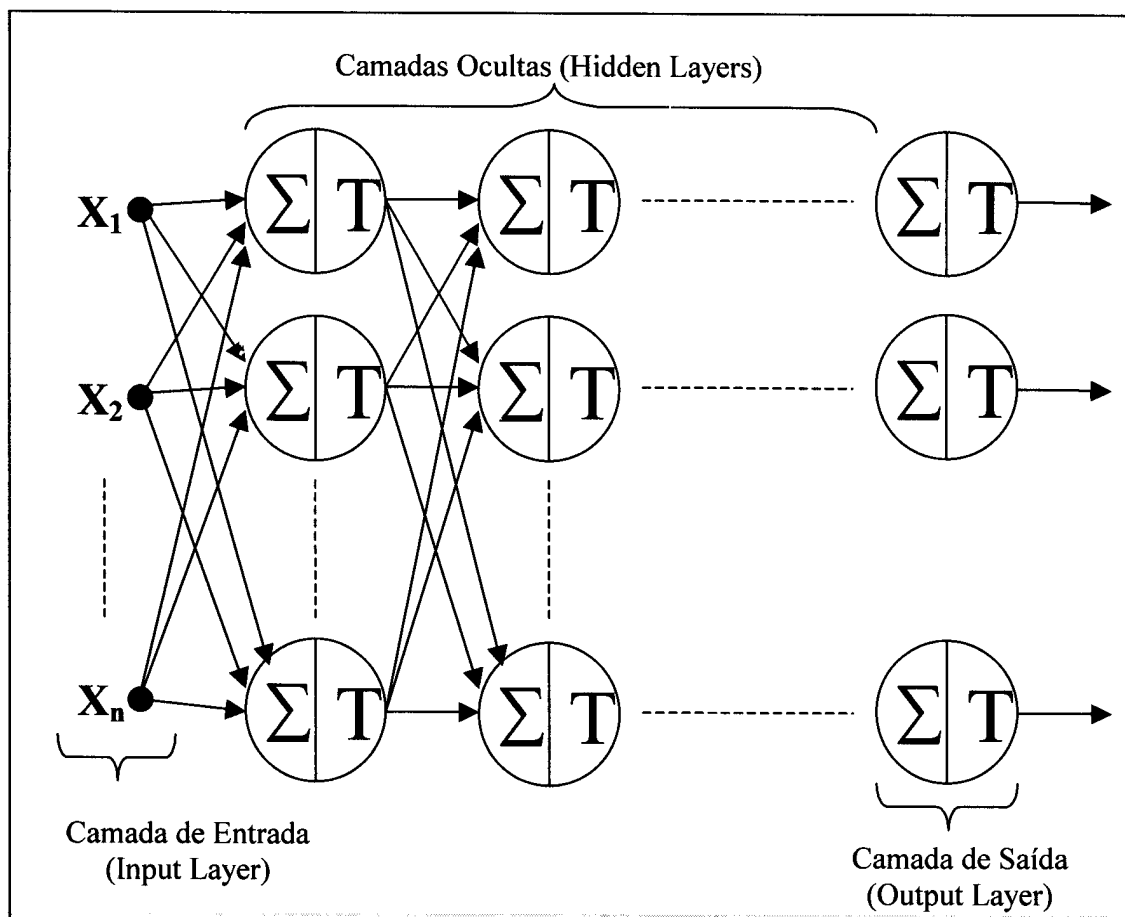


Figura 3.20 – Esquema de uma rede do tipo MLP

O Perceptron de Múltiplas Camadas possui três características distintas [HAY-94]:

1. O modelo de cada neurônio inclui uma não linearidade no final da saída. Um importante ponto a enfatizar aqui é que a não linearidade é SMOOTH (diferente em qualquer lugar), ao contrário do limite ríspido usado no Perceptron de Rosenblatt. Uma forma usualmente utilizada de não linearidade que satisfaz esse requerimento e a *não linearidade sigmoideal* definido pela função logística da equação 3.15.

$$y = \frac{1}{1 + \exp(-v)} \quad (3.15)$$

onde y é a saída do neurônio, e v é o nível de atividade interna da rede definido por $v = w(t) * x(t) - \text{bias}$. A presença da não linearidade é importante porque, por outro lado, a relação de entrada-saída da rede pode ser reduzida para uma única camada do perceptron;

2. A rede contém uma ou mais camadas de *neurônios escondidos* que não fazem parte da entrada ou saída da rede. Esses neurônios escondidos habilitam a rede a aprender tarefas complexas pela extração progressivamente de características significantes dos padrões de entrada (vetores);

3. A rede exibe um alto grau de *conectividade*, determinada pelas sinapses da rede. A mudança da conectividade da rede requer a mudança da população das conexões sinápticas ou de seus pesos;
4. Uma rede com uma única camada intermediária é suficiente para aproximar qualquer função contínua e duas camadas intermediárias são suficientes para aproximar qualquer função matemática.

Em uma rede múltiplas camadas, o processamento realizado por cada nodo é definido pela combinação dos procedimentos realizados pelos nodos de cada camada anterior que estão conectados a ele. Quando se segue da primeira camada intermediária em direção à camada de saída, as funções implementadas se tornam cada vez mais complexas. Estas funções definem como é realizada a divisão do espaço de decisão. Para uma rede com pelo menos duas camadas escondidas, pode-se dizer que o seguinte processamento ocorre em cada uma das camadas:

- *Primeira camada intermediária*: cada nodo traça retas no espaço de padrões de treinamento;
- *Segunda camada intermediária*: cada nodo combina as retas traçadas pelos neurônios de cada camada anterior conectados a ele, formando regiões convexas, onde o numero de lados é definido pelo numero de unidades a eles conectadas. A figura 3.21a ilustra possíveis regiões convexas definidas pelo nodos da segunda camada intermediária;
- *Camada de saída*: cada nodo forma regiões que são combinações das regiões convexas definidas pelos nodos a ele conectados da camada anterior. Os nodos definem, desta maneira, regiões com formatos abstratos. Possíveis regiões formadas pelos nodos da camada de saída podem ser vistas na figura 3.21b.

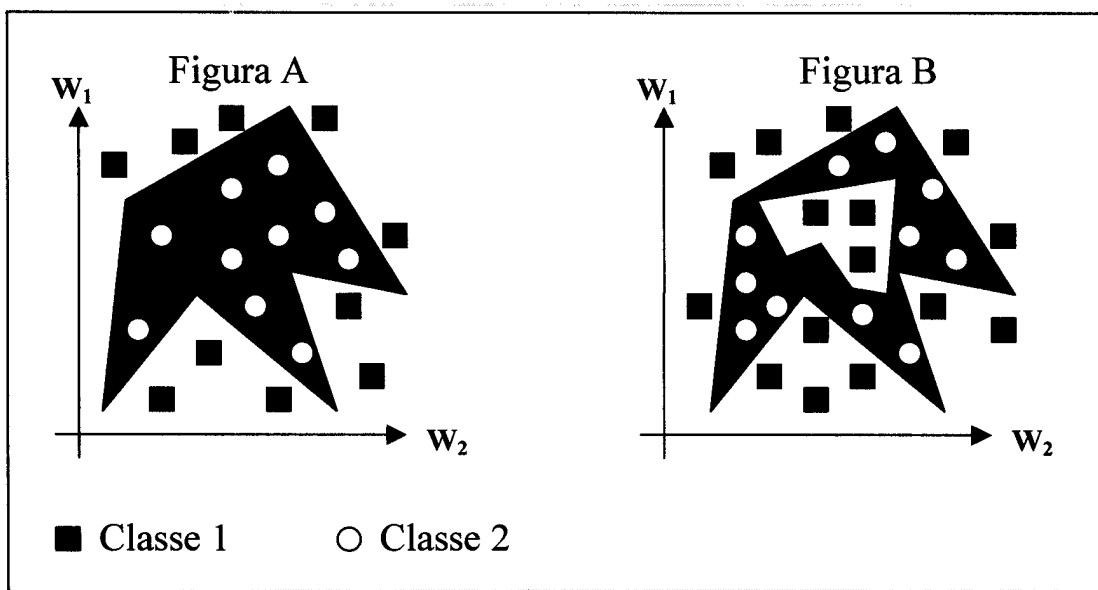


Figura 3.21 – Regiões definidas pelo processamento das redes MLP

As camadas intermediárias das redes MLP funcionam como detectores de características, gerando uma codificação interna dos padrões de entrada que são utilizadas para a definição da saída da rede. Se o número de camadas ocultas for suficientemente grande, é possível formar representações internas para qualquer conjunto de padrões de entrada.

O algoritmo de aprendizado mais conhecido para o treinamento das redes de múltiplas camadas é o **backpropagation**. Este algoritmo provê um aprendizado supervisionado, isto é, ele procura achar iterativamente a mínima diferença entre as saídas desejadas e as saídas obtidas pela rede neural, segundo um erro mínimo. Desta forma, ajustando os pesos entre as camadas através da retro-propagação do erro encontrado em cada iteração.

Durante a fase de treinamento, o algoritmo backpropagation é dividido em duas fases sequenciais, a fase *forward* e *backward*. Na primeira fase os sinais (padrões) de entrada são apresentados à rede artificial em uma determinada ordem. Cada padrão de treinamento é propagado adiante, camada após camada, até a produção do sinal (padrão) de saída. A saída computada pela rede é então comparada como a saída desejada. Esta comparação irá gerar um valor que determinará o erro. Já na fase backward este erro será utilizado como realimentação para as conexões, que resultará no ajuste dos pesos sinápticos de cada camada num sentido oposto à propagação dos sinais de treinamento.

O processo é repetido por um número determinado de vezes para cada padrão de dados durante o treinamento até que o erro total da saída tenha convergido a um valor mínimo, ou até que algum limite pré-determinado de interações tenha sido completado. Na figura 3.22 a fase *forward* está ilustrada pelas setas cheias (sentido entrada → saída) e a fase backward é representada pelas setas pontilhadas (sentido saída → entrada).

O algoritmo backpropagation é baseado na regra delta proposta por Widrow e Hoff, sendo por isto também chamada de regra delta generalizada. A atualização dos pesos se dá através da equação 3.16, onde δ é o erro entre a saída apresentada e a desejada do neurônio j .

$$\begin{bmatrix} \text{Correção} \\ \text{do peso} \\ \Delta w_{ji}(t) \end{bmatrix} = \begin{bmatrix} \text{Taxa de} \\ \text{aprendizado} \\ \eta \end{bmatrix} * \begin{bmatrix} \text{Gradiente} \\ \text{local} \\ \delta_j(t) \end{bmatrix} * \begin{bmatrix} \text{Sinal de entrada} \\ \text{do neurônio } j \\ y_j(t) \end{bmatrix} \quad (3.16)$$

A computação do gradiente local δ_j depende se o neurônio j é um nodo de saída ou um nodo escondido, o qual possuirá uma equação específica.

Na derivação deste algoritmo, a soma dos erros quadráticos é a função custo a ser minimizada, representada pela equação 3.17. Nesta equação, E é a medida do erro total, p é o número de padrões, k é o número de unidades de saída, d_i é a i -ésima saída desejada e y_i é a i -ésima saída gerada pela rede.

$$E = \frac{1}{2} \sum_p \sum_{i=1}^k (d_i^p - y_i^p)^2 \quad (3.17)$$

Mesmo com o aparente sucesso do algoritmo backpropagation, existem alguns aspectos que fazem com que este não possa ser utilizado em todo o tipo de aplicações. O maior problema é o longo processo de treinamento. Ele pode também resultar em uma taxa de sucesso do aprendizado e uma taxa de generalização ruim, pois não há garantias de que os mínimos globais de erro sejam encontrados. Em outras palavras, algumas vezes, dependendo dos pesos iniciais, da curva de aprendizado e dos dados de entrada, a rede neural artificial pode até não aprender. Vários algoritmos avançados baseados em backpropagation otimizaram e aprimoraram o processo de treinamento, mas as falhas no aprendizado ainda ocorrem, e normalmente tem duas fontes: a paralisação da rede e os mínimos locais.

3.11 - Redes SOM – Kohonen

Existem redes neurais artificiais que possuem a propriedade de auto-organização, sendo por isso chamado de redes *self-organizing* ou auto-organizativas. Essas redes têm como idéia o paradigma do aprendizado não supervisionado, a qual utilizam um conjunto de regras da natureza local para o ajuste dos pesos. Neste tópico será abordado as redes mapas *self-organizing* (Self-Organizing Maps – SOM), desenvolvidas por Teuvo Kohonen na década de 80.

As principais características das RNAs é a sua capacidade de aprender através de exemplos, utilizando o aprendizado supervisionado no qual fornece as respostas desejadas para os padrões de entrada, e assim limitando o seu campo de aplicação àqueles onde estes dados são conhecidos. A idéia das redes *self-organizing* é fazer o treinamento de uma rede

sem o auxílio do professor externo, onde a única informação fornecida está no conjunto de padrões de entradas.

As redes *self-organizing* são mais semelhantes às estruturas neurobiológicas que as redes supervisionadas. A estrutura básica de modelos *self-organizing* é formada por uma camada de entrada (nodos de fonte) e uma camada de saída. Existem também estruturas com varias camadas, em que o processo *self-organizing* segue de camada em camada. RNAs *self-organizing* possuem um amplo leque de aplicações potenciais, principalmente em problemas de reconhecimento de padrões e categorização de dados em que as classes não são inicialmente conhecidas.

As redes SOM são baseadas no mapa topológico presente no córtex cerebral, funcionando da seguinte maneira: quando um padrão de entrada X é apresentado, a rede procura a unidade mais parecida com X . Durante o seu treinamento, a rede aumenta a semelhança (ajuste dos pesos) do nodo escolhido e de seus vizinhos a X . Desta forma, a rede constrói um mapa topológico onde nodos que estão topologicamente próximos respondem de forma semelhante a padrões de entrada semelhantes. Assim como no córtex cerebral, nodos ativos apresentam conexões mais fortes com nodos que estão fisicamente mais próximos. A rede SOM de Kohonen pode ser vista na figura 3.24.

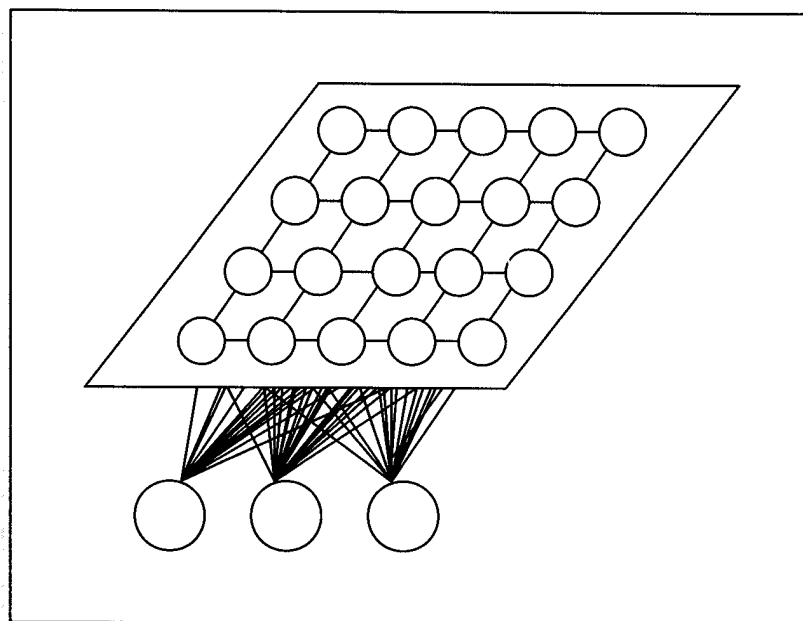


Figura 3.24 – Rede SOM típica

A rede SOM utiliza um algoritmo de aprendizado competitivo. Como mostra a equação 3.18, o estado de ativação de um nodo é determinado pela medida da distancia euclidiana D_j entre seu peso e o vetor de entrada, onde x é o vetor de entrada e w é o vetor peso e n é o número de entradas.

$$D_j = \sum_{i=1}^n \| x_i - w_{ji} \| \quad (3.18)$$

Durante o processo *self-organizing*, a unidade de agrupamento cujo vetor peso emparelha a entrada padrão mais próxima (tipicamente, o quadrado da distância euclidiana mínima) é escolhida como vencedora. A unidade vencedora e as unidades vizinhas (em termos da topologia das unidades de agrupamento) atualizam seus pesos. Durante o processo, a taxa de aprendizado e o raio da vizinhança topológica são continuamente decrementados. A equação 3.19 ilustra como são atualizados os pesos da unidade vencedora e daquelas situadas na sua vizinhança.

$$\mathbf{w}_{ji}(\mathbf{t}+1) = \begin{cases} \mathbf{w}_{ji}(\mathbf{t}) + \eta(\mathbf{t})(\mathbf{x}_i(\mathbf{t}) - \mathbf{w}_{ji}(\mathbf{t})), & \text{se } \mathbf{j} \in \text{a vizinhança de } \mathbf{t} \\ \mathbf{w}_{ji}(\mathbf{t}), & \text{caso contrário} \end{cases} \quad (3.19)$$

onde $\mathbf{w}_{ji}(\mathbf{t})$ é o peso da conexão entre o elemento de entrada $\mathbf{x}_i(\mathbf{t})$ e o nodo \mathbf{j} , η é a taxa de aprendizado, tudo no instante \mathbf{t} .

Em uma rede SOM, a vizinhança pode assumir vários formatos diferentes. Embora o formato quadrado seja o mais comum, a região de vizinhança pode apresentar também, conforme mostra a figura 3.25 a forma de um hexágono ou um círculo (gaussiana). O formato dessa vizinhança geralmente é definido por tentativa e erro, podendo ter uma topologia arbitrária.

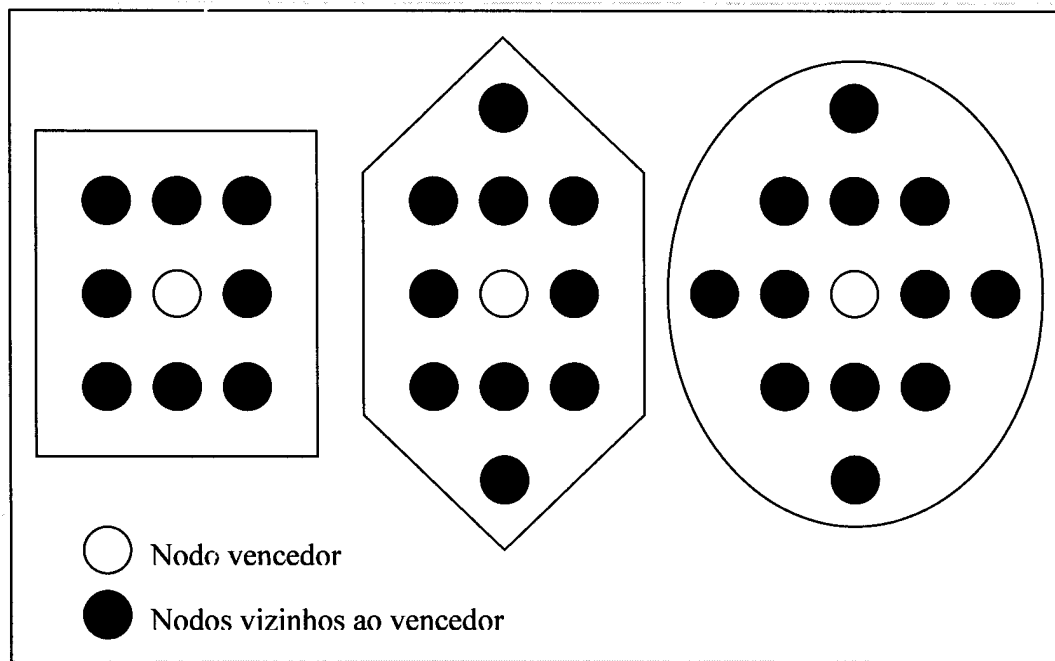


Figura 3.25 – Formatos para a região vencedora utilizados pelas redes SOM

O algoritmo de treinamento para as redes do tipo SOM é demonstrado através da ilustração 3.26.

1. Iniciar os pesos e parâmetros.
2. Repetir
 - 2.1. Para cada padrão de treinamento **X** faça
 - 2.1.1. Definir nodo vencedor
 - 2.1.2. Atualizar os pesos deste nodo e de seus vizinhos.
 - 2.1.3. Se o numero de ciclo (época) for múltiplo de N
Então reduzir a taxa de aprendizado e área de vizinhança
3. Até o mapa de características não mudar.

Figura 3.26 – Passos do algoritmo para as redes SOM

3.12 - Implementações de Redes Neurais Artificiais

3.12.1 - Problema XOR

A implementação da porta lógica não linear “OU” exclusivo (XOR) foi realizada em quatro tipos de redes, Perceptron de uma única camada, Perceptron de múltiplas camadas, rede Adaline e rede Kohonen.

Na rede **Perceptron de camada simples** (SingleLayer Perceptron - SLP), foi utilizada uma taxa de aprendizado de **0,5** e um treinamento de **30.000** épocas. Após o treinamento foi feita a ativação dos dados apresentados à rede, obtendo um acerto de **50%**. A tabela 3.2 mostra os resultados obtidos.

Entrada	Saída desejada	Saída encontrada	
0 0	0	1	Errado
0 1	1	1	Correto
1 0	1	0	Errado
1 1	0	0	Correto

Tabela 3.2 – Resultado da rede SingleLayer Perceptron

No treinamento da rede **Perceptron de múltiplas camadas** (MLP) foi utilizada uma taxa de aprendizado de **0,2** o qual precisou de **8051** épocas para encontrar um erro

quadrático médio menor que **0,01** (erro alvo). Nesta rede utilizou-se uma função de ativação sigmoial e apenas uma camada escondida com **2** neurônios ocultos. O resultado obtido foi um acerto de **100%** que pode ser visto na tabela 3.3.

Entrada	Saída desejada	Saída encontrada	Saída padronizada	
0 0	0	0,0503	0	Correto
0 1	1	0,9412	1	Correto
1 0	1	0,9549	1	Correto
1 1	0	0,0443	0	Correto

Tabela 3.3 – Resultado da rede MLP no problema XOR

Na figura 3.27 está ilustrado o erro quadrático médio encontrado ao longo do treinamento da porta lógica XOR em função do número de épocas utilizadas no seu treinamento.

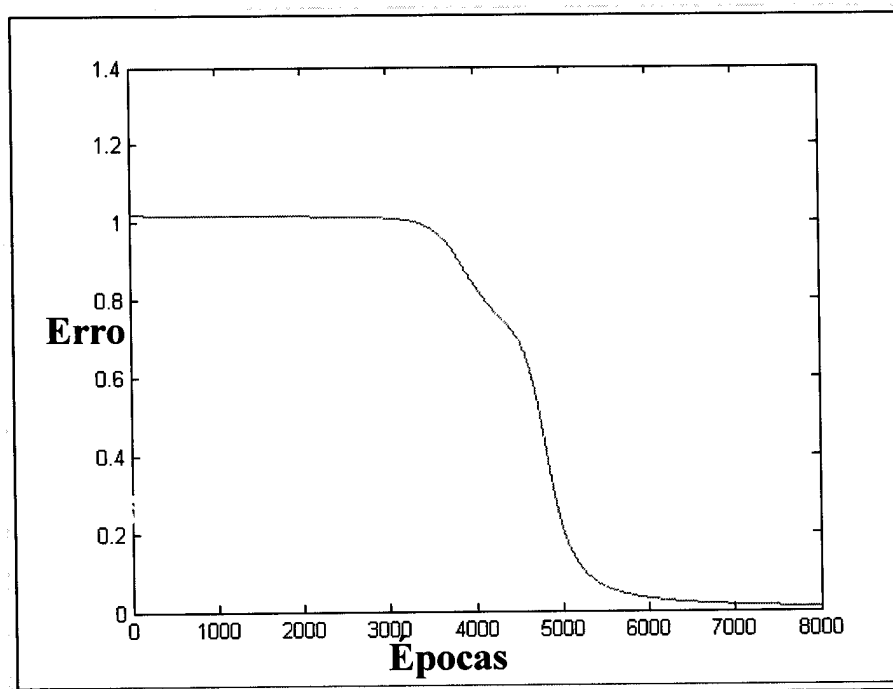


Figura 3.27 – Erro quadrático médio da rede MLP no problema XOR

Já a rede **Adaline** foi treinada por **30.000** épocas gerando uma medida de desempenho igual a **3,125**. Com uma taxa de aprendizado de **0,2** foi obtido um resultado de **25%** de acerto, conforme mostra a tabela 3.4.

Entrada	Saída desejada	Saída encontrada	Saída padronizada	
0 0	0	0,625	1	Errado
0 1	1	0,5	0	Errado
1 0	1	0,375	0	Errado
1 1	0	0,25	0	Correto

Tabela 3.4 – Resultado obtido da rede Adaline

Na rede **Kohonen** foi utilizada uma taxa de aprendizado igual a **0,6**, com a possibilidade de formar apenas **2** unidades de agrupamento (0 e 1). A duração do treinamento foi de **10.000** épocas gerando o resultado de **25%** de acerto. A tabela 3.5 mostra o resultado obtidos na resolução do problema XOR.

Entrada	Saída desejada	Saída encontrada	
0 0	0	0	Correto
0 1	1	0	Errado
1 0	1	1	Errado
1 1	0	1	Errado

Tabela 3.5 – Resultado da rede Kohonen

Com os resultados obtidos pelas quatro redes pode-se concluir que o Perceptron de múltiplas camadas é a arquitetura que melhor demonstra o funcionamento da porta lógica XOR (100% de acerto) e que redes como Adaline e Perceptron de apenas uma camada não são redes capazes de resolver problemas não lineares.

3.12.2 - Reconhecimento de Caracteres

Para o aprendizado no reconhecimento de caracteres foi utilizado um banco de dados para o vetor de entrada conforme ilustrada na tabela 3.6. Esse conjunto de entradas foi utilizado no treinamento das redes Perceptron de uma única camada, Perceptron de

múltiplas camadas, rede Adaline e rede Kohonen. Em todas as redes foram utilizadas as fontes 1 e 2 para o treinamento, e a fonte 3 para a generalização dos dados aprendidos pela rede.

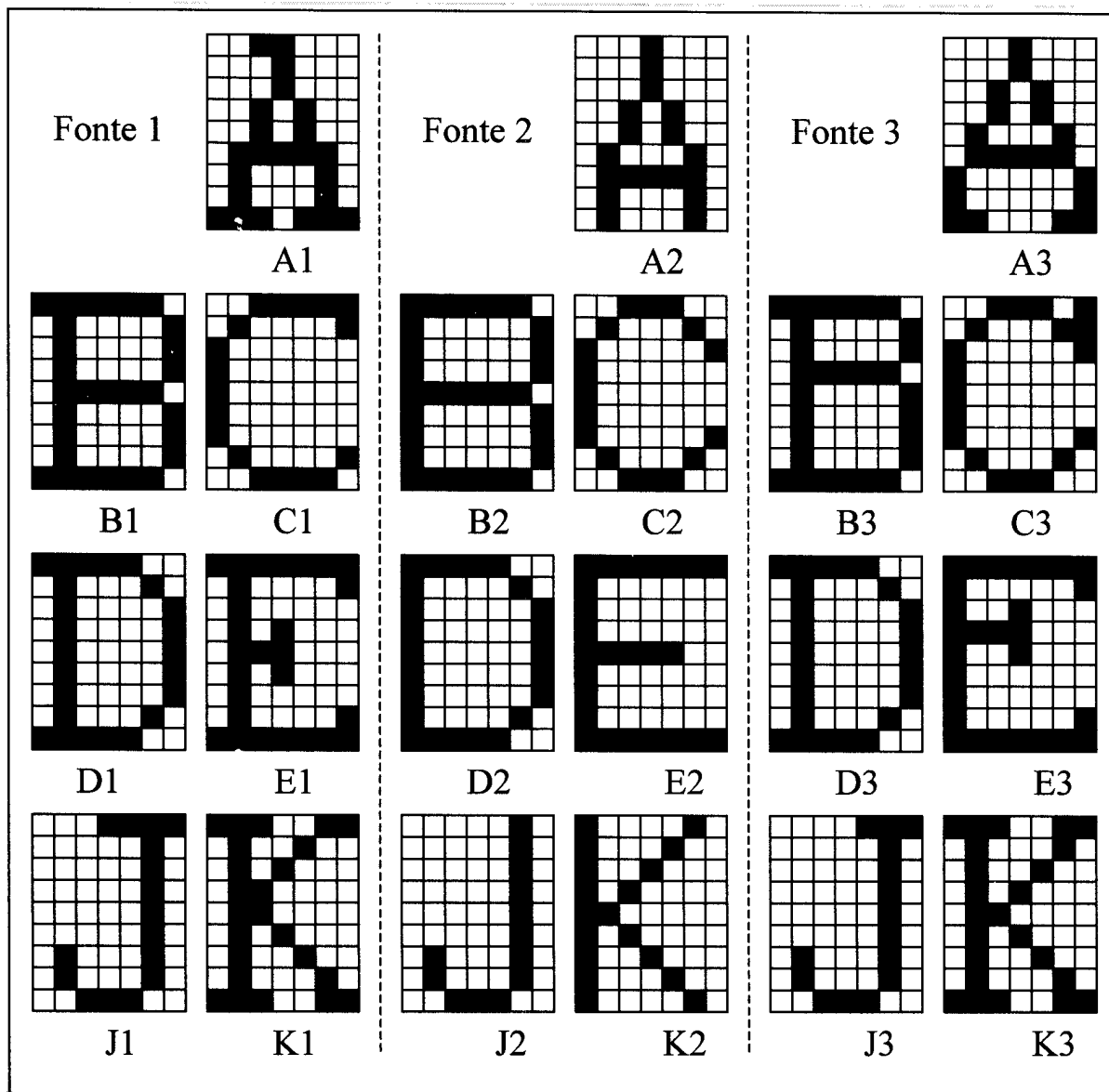


Tabela 3.6 – Padrões de entrada para exemplos de reconhecimento de caracteres

Na rede **Kohonen**, foi utilizado uma taxa de aprendizado igual a **0,8**, um ciclo de **10.000** épocas para o treinamento das fontes 1 e 2, possibilitando a formação de até **25** unidades de agrupamento (letras do alfabeto). Após o treinamento, a rede SOM encontrou **6** grupos diferentes para os padrões apresentados e obteve um acerto de **42,9%** na generalização da fonte 3, reconhecendo a letra **A, C e J** (cor cinza na tabela). O resultado pode ser visto na tabela 3.7 ilustrada abaixo, onde os dados com o fundo na cor cinza são as letras encontradas corretamente pela rede.

Fonte 1	Grupo	Fonte 2	Grupo	Fonte 3	Grupo
A1	20	A2	20	A3	20
B1	5	B2	16	B3	5
C1	12	C2	12	C3	12
D1	5	D2	16	D3	5
E1	5	E2	16	E3	5
J1	3	J2	3	J3	3
K1	5	K2	19	K3	5

Tabela 3.7 – Resultado da rede SOM no reconhecimento de caracteres

Na rede **Perceptron de múltiplas camadas (MLP)** o treinamento das fontes 1 e 2 foi conduzido com uma taxa de aprendizado igual a **0,5** com apenas uma camada escondida com **20** neurônios, utilizando uma função de ativação sigmoial. A duração do treinamento foi de **11676** épocas até atingir o erro quadrático médio menor que **0,001** (erro alvo), conforme mostra a figura 3.28.

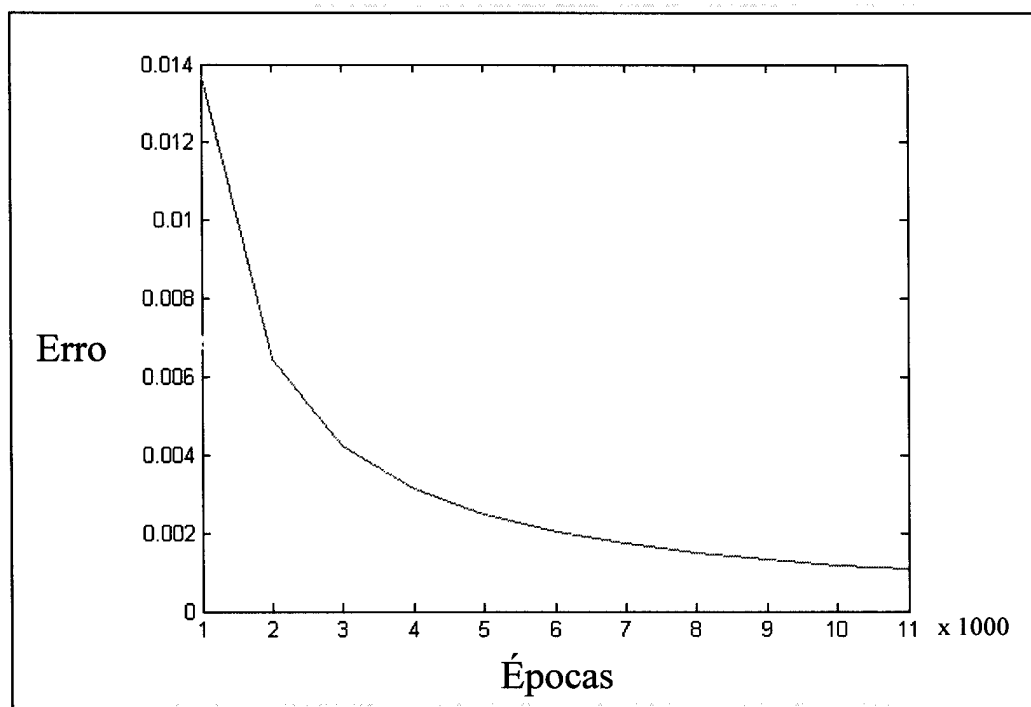


Figura 3.28 – Erro quadrático médio encontrado na rede MLP

Na generalização da rede MLP com a fonte 3 a rede obteve um acerto de **100%** aos dados apresentados, conforme mostra a tabela 3.8.

Fonte 3	Saída desejada	Saída encontrada	Saída padronizada	
A3	0 0 0	0,0071 0,0065 0,0334	0 0 0	Correto
B3	0 0 1	0,0025 0,3171 0,9920	0 0 1	Correto
C3	0 1 0	0,0020 0,9950 0,0067	0 1 0	Correto
D3	0 1 1	0,0039 0,9960 0,9933	0 1 1	Correto
E3	1 0 0	0,8921 0,0277 0,0043	1 0 0	Correto
J3	1 0 1	0,9957 0,0048 0,9933	1 0 1	Correto
K3	1 1 0	0,9989 0,9876 0,0228	1 1 0	Correto

Tabela 3.8 – Resultado da rede MLP no reconhecimento de caracteres

Já na rede **Perceptron de camada simples (SLP)** o treinamento das fontes 1 e 2 foi realizado com uma taxa de aprendizado igual a **0,9** com ciclos de **20.000** épocas. Na generalização dos dados treinados com a fonte 3, conforme mostra na tabela 3.9, foi obtido um acerto de **100%**.

Fonte 3	Saída desejada	Saída encontrada	
A3	0 0 0	0 0 0	Correto
B3	0 0 1	0 0 1	Correto
C3	0 1 0	0 1 0	Correto
D3	0 1 1	0 1 1	Correto
E3	1 0 0	1 0 0	Correto
J3	1 0 1	1 0 1	Correto
K3	1 1 0	1 1 0	Correto

Tabela 3.9 – Resultado da rede SLP no reconhecimento de caracteres

Na rede Adaline foi utilizado uma taxa de aprendizado de **0,05** e necessitou de **125** épocas para atingir um erro alvo menor que **0,0001**. A figura 3.29 mostra o erro quadrático médio em função das épocas do treinamento.

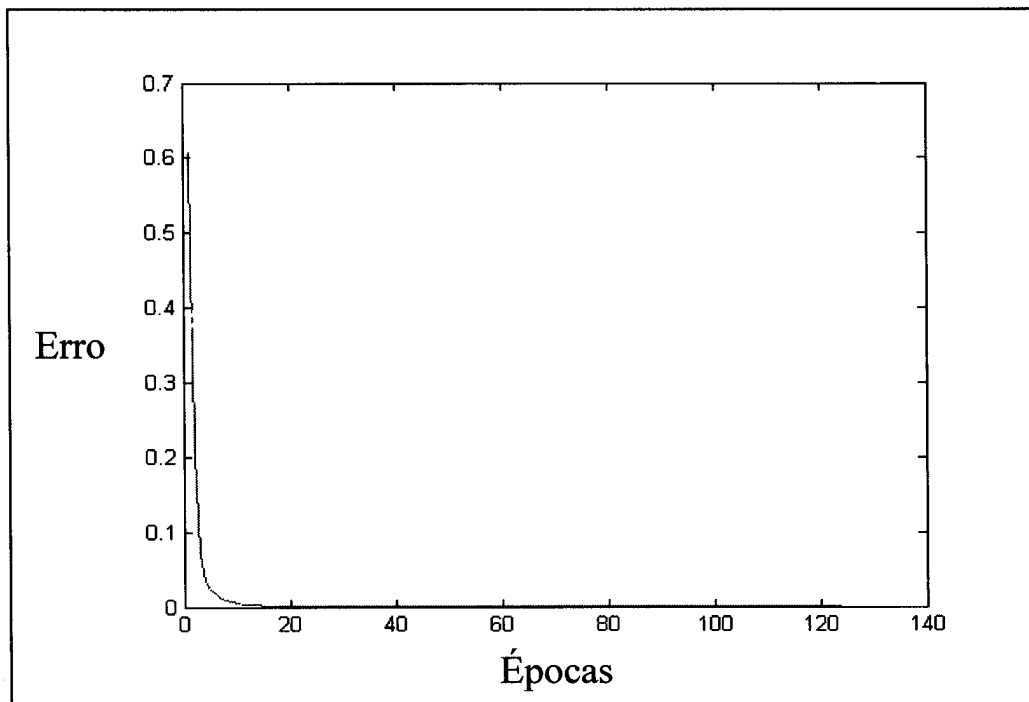


Figura 3.29 – Erro encontrado na rede Adaline no reconhecimento de caracteres

Após o treinamento da rede foi verificada a porcentagem de aprendizado nos dados não visto no treinamento, obtendo um acerto de **85,71%**. O resultado pode ser visto na tabela 3.10 abaixo, onde mostra que a rede encontrou dificuldade no reconhecimento da fonte B3.

Fonte 3	Saída desejada	Saída encontrada	Saída padronizada	
A3	0 0 0	-0,1197 -0,1104 0,1631	0 0 0	Correto
B3	0 0 1	0,0746 0,7714 0,8662	0 1 1	Errado
C3	0 1 0	-0,0437 1,0560 0,0602	0 1 0	Correto
D3	0 1 1	-0,0162 1,0113 1,0067	0 1 1	Correto
E3	1 0 0	0,5447 0,4205 0,1796	1 0 0	Correto
J3	1 0 1	0,9842 0,0609 0,9411	1 0 1	Correto
K3	1 1 0	1,1908 0,5983 -0,0254	1 1 0	Correto

Tabela 3.10 - Resultado obtido da rede Adaline

4 - ANÁLISES DOS TESTES PSICOLÓGICOS

4.1 - Teste Psicológico VARK

4.1.1 - Introdução

Esse questionário tem por finalidade investigar as preferências por modos pelos quais as pessoas trabalham com informação. Foi adquirida com a aplicação do teste VARK um estilo de aprendizagem preferido e que, uma parte desse estilo foi a preferência pelo recebimento e pela transmissão de idéias e de informações.

O acrônimo VARK representa as iniciais das palavras Visual, Aural, Read/write. E Kinesthetic, respectivamente em português Visual, Aural, Leitura/Escrita e Cinestésico, todas as modalidades sensoriais que são usadas para aprender informação. Fleming e Mills (1992) propuseram quatro categorias que são apropriadas para refletir as experiências dos estudantes. Embora haja alguma sobreposição entre as categorias, para os propósitos de nossa discussão, elas são definidas como segue:

- **Visual (V):** Esta preferência inclui a representação da informação em esquemas, gráficos, fluxogramas e todos os tipos de setas, círculos, hierarquias e outros artifícios que os instrutores usam para representar o que poderia ser também representado em palavras.
- **Aural (A):** Esta modalidade perceptual descreve uma preferência pela informação que é “ouvida”. Os estudantes com essa modalidade contam que aprendem melhor com aulas expositivas, palestras, fitas e conversando com outros estudantes.
- **Read/write (R):** É a preferência pela informação mostrada em palavras. Não surpreende que muitos professores universitários tenham forte preferência por essa modalidade.
- **Kinesthetic (K):** Por definição, essa modalidade refere-se a “preferência perceptual relacionada ao uso da experiência e da prática (simulada ou real)”. Embora uma experiência possa envolver outras modalidades, o essencial é que o estudante está ligado à realidade, “através da experiência, do exemplo, da prática ou da simulação” [Fleming & Mills, 1992].

O questionário é composto por 13 questões no qual cada alternativa representa uma letra da palavra VARK sendo que pode ser assinalado mais de uma alternativa por questão. A figura 4.1 abaixo mostra como é feito diagnóstico de um teste.

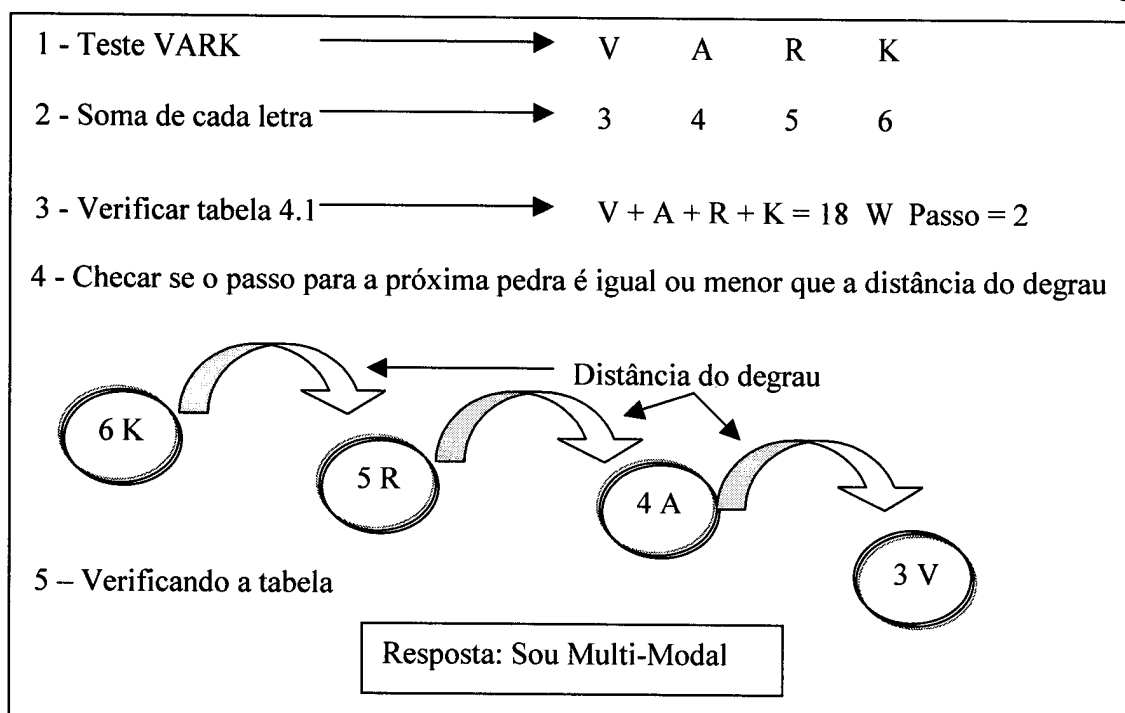


Figura 4.1 – Exemplo da correção de um questionário VARK

Após a pessoa responder os testes é feita a soma de cada letra assinalada e a seguir é feita a soma de todas as letras e comparada com a tabela 4.1 a seguir:

Soma dos quatro escores VARK é:	A distância do meu passo é:
10 – 16	1
17 – 22	2
23 – 26	3
Mais que 26	4

Tabela 4.1 – Distância do passo de acordo com a soma dos escores VARK

Após esse procedimento, a letra com o maior valor vai para a primeira pedra. Se a distância do degrau da primeira pedra para a segunda for menor ou igual à distância do passo então é colocada a letra com o segundo maior valor na segunda pedra e assim por diante. Se a pessoa não alcançar a próxima pedra então foi concluída a definição do conjunto de preferências o teste em questão.

O próximo passo é calcular a diferença entre o valor colocado na primeira pedra e a distância do degrau para a próxima pedra. Essa diferença será conferido na tabela 4.2 e assim será encontrado o diagnóstico o teste.

Coluna 1	A diferença entre meus dois maiores escores foi:						
Meu total número de respostas é:	Zero. Eles são iguais.	1	2	3	4	5	6 ou mais
Menos que 17	Multi-Modal	Multi-Modal	Mild	Strong	Very Strong	Very Strong	Very Strong
Entre 17 e 22	Multi-Modal	Multi-Modal	Multi-Modal	Mild	Strong	Very Strong	Very Strong
Entre 23 e 26	Multi-Modal	Multi-Modal	Multi-Modal	Multi-Modal	Mild	Strong	Very Strong
Mais que 26	Multi-Modal	Multi-Modal	Multi-Modal	Multi-Modal	Multi-Modal	Mild	Strong

Tabela 4.2 – Encontrando o diagnóstico

4.1.2 - Aplicação

Para a aplicação do teste VARK, foram utilizados 18 questionários e seu treinamento conduzido em quatro tipos de redes neurais artificiais, Adaline, SingleLayer Perceptron, MultiLayer Perceptron (Backpropagation) e Kohonen. Desses 18 questionário, foram identificados 13 testes para o diagnóstico "Multi-Modal", 1 para "Very Strong" e 4 para "Mild". A tabela 4.3 mostra o diagnóstico de cada questionário.

Classificação dos Testes			
Questionário	Diagnóstico	Questionário	Diagnóstico
1	Multi-Modal	10	Mild
2	Multi-Modal	11	Multi-Modal
3	Multi-Modal	12	Multi-Modal
4	Very Strong	13	Multi-Modal
5	Multi-Modal	14	Multi-Modal
6	Multi-Modal	15	Multi-Modal
7	Mild	16	Multi-Modal
8	Mild	17	Mild
9	Multi-Modal	18	Multi-Modal

Tabela 4.3 – Diagnóstico de cada Teste

Já a tabela 4.4 exhibe a disposição dos testes para o treinamento e generalização da rede. Sempre para o teste de ativação das redes (generalização) faremos sempre o uso dos questionários que não foram utilizados nos seus treinamentos, verificando assim a porcentagem de aprendizagem de dados nunca vistos anteriormente. Os resultados das redes foram padronizados, isto é, se a saída for maior que **0,5**, foi utilizado como resultado o valor **1**, e o valor **0** caso contrário.

Disposição dos Testes utilizados na Rede Neural Artificial		
Questionário	Treinamento	Generalização
1	X	
2		X
3	X	
5	X	
6	X	
9	X	
11		X
12	X	
13	X	
14	X	
15	X	
16		X
18	X	
4	X	X
7	X	
8		X
10	X	
17		X

Tabela 4.4 – Disposição dos Testes utilizados na Rede Neural Artificial

OBS: Nesse teste psicológico realizado, foi encontrado apenas um diagnóstico “Very Strong”, então por conveniência este teste será utilizado tanto para os treinamentos quanto para as generalizações das redes.

Os três tipos de diagnósticos encontrados nos testes foram codificados em números binários conforme mostra a tabela 4.5.

Diagnóstico	Codificação
Multi-Modal	0 1
Very Strong	1 1
Mild	1 0

Tabela 4.5 – Diagnósticos dos testes codificados

4.1.3 - Rede Adaline

Na rede Adaline foi utilizado uma taxa de aprendizagem (**K**) igual a **0,2** e para o seu treinamento foram necessário **1.323** épocas encontrando um erro menor que **0,05** . A figura 4.2 mostra o erro encontrado no treinamento da rede Adaline.

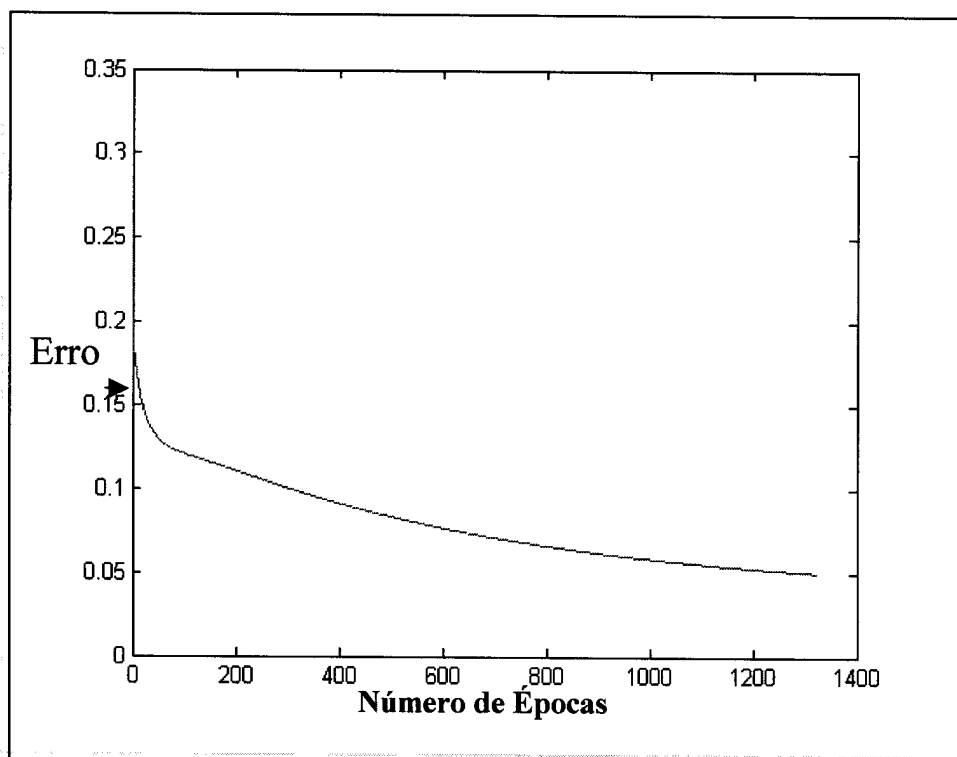


Figura 4.2 – Erro encontrado na rede Adaline no teste VARK

Na utilização dos testes selecionados para a generalização (testes nunca vistos anteriormente pela rede) obteve uma aprendizagem de **50%** de acerto. A tabela 4.6 mostra o resultado da rede Adaline.

Teste	Saída desejada	Saída encontrada	Saída padronizada	
2	0 1	-0,3093 1,1535	0 1	Correto
4	1 1	0,8405 0,5640	1 1	Correto
8	1 0	0,6610 0,4472	1 0	Correto
11	0 1	0,7951 0,5993	1 1	Errado
16	0 1	0,6649 0,4801	1 0	Errado
17	1 0	0,5479 0,6705	1 1	Errado

Tabela 4.6 – Resultado do teste VARK na rede Adaline

4.1.4 - Rede Kohonen (SOM)

A rede Kohonen, com uma taxa de aprendizagem igual a **0,5**, foi conduzida até **20,000** épocas para realizar o seu treinamento. Na utilização dos testes seleccionados para a generalização nunca vistos anteriormente pela rede, esta obteve uma aprendizagem de **50%** de acerto. A tabela 4.7 ilustra o resultado da rede Kohonen. Os quadros que estão com preenchimento cinza são os testes que encontraram o mesmo grupo tanto no treinamento quanto na generalização dos dados.

MULTI-MODAL				V E R Y	Testes treinados	Grupo
Testes Treinados	Grupo	Testes Generalizados	Grupo		4	3
1	1	2	1		Testes generalizados	Grupo
3	1					
5	3					
6	3	11	3	4	3	
9	2					
12	3					
13	1	16	3	S T R O N G	4	3
14	3					
15	1					
18	1					

MILD			
Testes Treinados	Grupo	Testes Generalizados	Grupo
7	2	8	2
10	1	17	3

Tabela 4.7 – Resultado da rede Kohonen no teste psicológico

Para os testes treinados do diagnóstico Multi-Modal, a maioria dos testes encontrou o **grupo 1** (testes 1,3, 13, 15, 18) e os testes generalizados para esse mesmo diagnóstico encontrou o teste 2. O diagnóstico Mild encontrou o **grupo 2** (teste 7) no treinamento e generalizou para o mesmo grupo o teste 8. Já o teste 4 ficou sozinho no **grupo 3**.

4.1.5 - Rede Perceptron de Múltiplas Camadas (MLP)

Na rede Perceptron de Múltiplas Camadas (MLP) foi utilizado uma taxa de aprendizagem (**K**) igual a **0,5** e necessitou de **6169** épocas para encontrar a soma dos erros quadrático médio menor que **0,01**. Foi utilizada apenas uma camada escondida com 10 neurônios e uma função de ativação sigmoidal. Na figura 4.3 está ilustrada a soma dos erros quadráticos de todas as épocas do treinamento.

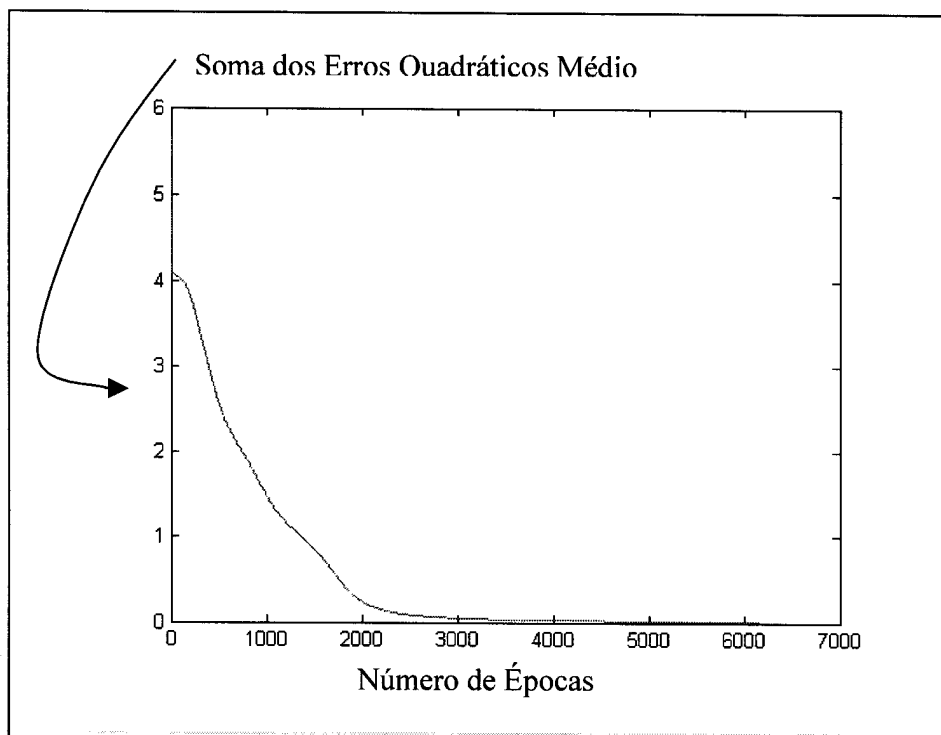


Figura 4.3 – Soma dos erros quadrático encontrado no Backpropagation

Após o treinamento foi verificado o aprendizado da rede nos testes nunca vistos anteriormente, no qual obteve na generalização um acerto de **50%**. A tabela 4.8 mostra o resultado da rede MLP.

Teste	Saída desejada	Saída encontrada	Saída padronizada	
2	0 1	0,0144 0,4246	0 0	Errado
4	1 1	0,9580 0,9979	1 1	Correto
8	1 0	0,9680 0,0075	1 0	Correto
11	0 1	1,0000 0,0104	1 0	Errado
16	0 1	0,0000 1,0000	0 1	Correto
17	1 0	0,0011 1,0000	0 1	Errado

Tabela 4.8 – Resultado da rede MLP no teste VARK

A rede Perceptron de Múltiplas Camadas generalizou corretamente o teste 4 (Very Strong), o teste 8 (Mild) e o teste 16 (Multi-Modal), mas não conseguiu reconhecer o teste 2 e o teste 11 (Multi-Modal) nem o teste 17 (Mild).

4.1.6 - Rede Perceptron de Camada Simples (SLP)

Na rede Perceptron de camada simples foi utilizado uma taxa de aprendizagem igual a **0,5** e um treinamento com duração de **20.000** épocas. Após o treinamento foi apresentado os testes nunca vistos anteriormente pela rede SLP, gerando um acerto de **66,67%**. O resultado pode ser visto na tabela 4.9.

Teste	Saída desejada	Erro	Saída encontrada	
2	0 1	0 0	0 1	Correto
4	1 1	0 0	1 1	Correto
8	1 0	0 0	1 0	Correto
11	0 1	-1 1	1 0	Errado
16	0 1	-1 0	1 1	Errado
17	1 0	0 0	1 0	Correto

Tabela 4.9 – Resultado da rede Perceptron de Camada Simples

Através da tabela pode-se perceber que a rede conseguiu aprender os testes nunca vistos anteriormente como os testes 2 e 11 (Multi-Modal) e os testes 8 e 17 (Mild). Mas a rede encontrou dificuldade para aprender os testes 11 e 16 (Multi-Modal).

4.1.7 - Conclusão

Pode-se concluir de acordo com a tabela 4.10 abaixo que para o teste em questão a rede Perceptron de Camada Simples obteve um desempenho melhor do que as demais redes. Mas deve se levar em conta que o número de questionários (padrões) aplicados em cada rede é insuficiente para a obtenção de um resultado promissor na aprendizagem das redes neurais artificiais. Porém, percebe-se que mesmo assim as redes neurais artificiais possibilita a aprendizagem de padrões de testes psicológicos e sua posterior aplicação.

Rede Neural Artificial	Porcentagem de acerto
Adaline	50 %
Kohonen	50 %
Perceptron de Múltiplas Camadas	50 %
Perceptron simples	66,67 %

Tabela 4.10 – Resultados do treinamento das redes

4.2 - Teste Psicológico MPCL

4.2.1 - Introdução

O teste psicológico MPLC (*Mooney Problem Check List*) foi desenvolvido por Leonard V. Gordon e Roos L. Mooney em 1950. Este teste foi projetado para ajudar indivíduos a expressarem seus problemas pessoais e são úteis para aumentar a compreensão do professor sobre seus estudantes. Os questionários são separados por faixa etária e abrangem áreas como: saúde e desenvolvimento, repouso e família, morais e religião, sexo e casamento.

O teste é composto por 288 questões divididos em 48 blocos com 6 perguntas em cada bloco. Cada conjunto pré-determinado de blocos condiciona a um diagnóstico específico e são encontrados aleatoriamente no teste em questão. Após a pessoa fazer o teste, é feita a contagem do número de testes assinalados por blocos e que, o bloco que conteve o maior número de respostas assinaladas é o maior problema encontrado por esta pessoa nesse momento como resultado do teste, podendo haver mais de um diagnóstico por pessoa. A disposição dos blocos é mostrada na tabela 4.11.

Relação Bloco-Diagnóstico					
Diagnóstico	Saúde	Sócio Econômico	Lazer	Família	Aptidão
Questões	1 – 6	7 – 12		31 – 42	
	49 – 54	55 – 60	13 – 18	79 – 90	19 – 24
	97 – 102	103 – 108	61 – 66	127 – 138	67 – 72
	145 – 150	151 – 156	205 – 110	175 – 186	
	193 – 246		253 – 258	223 – 228	
				271 – 276	
Diagnóstico	Negócio	Personalidade	Religião	Ocupação	Sexual
Questões		25 – 30			
		73 – 78			
	199 – 204	109 – 126	43 – 48	187 – 192	229 – 234
	247 – 252	157 – 174	91 – 96	235 – 240	277 – 282
		211 – 222	139 – 144	283 – 288	
		259 – 270			

Tabela 4.11 – Disposição dos blocos dos diagnósticos no teste

4.2.2 - Aplicação

O teste psicológico foi aplicado em 17 pessoas e o seu treinamento em Redes Neurais Artificiais foi conduzido em quatro tipos de redes: Adaline, Perceptron de Camada Simples, Perceptron de Múltiplas Camadas (Backpropagation) e rede Kohonen. Desses 17 testes, 11 questionários apresentaram o diagnóstico “Personalidade”, 4 testes para “Família/Lazer” 1 para Família e 1 para “Saúde”. A tabela 4.12 mostra os resultados de cada teste.

Classificação dos Testes			
Questionário	Diagnóstico	Questionário	Diagnóstico
1	Personalidade	10	Personalidade
2	Família/Lazer	11	Família/Lazer
3	Família/Lazer	12	Personalidade
4	Família/Lazer	13	Saúde
5	Personalidade	14	Personalidade
6	Personalidade	15	Personalidade

Tabela 4.12 – Diagnóstico de cada Teste

A tabela 4.13 exibe os testes utilizados para o treinamento e generalização da rede.

Diagnóstico	Questionário	Treinamento	Generalização
Personalidade	1	X	
	5		X
	6	X	
	8		X
	9	X	
	10		X
	12	X	
	14		X
	15	X	
	16		X
Saúde	13	X	X
Família	7	X	X
Família/Lazer	2	X	
	3		X
	4	X	
	11		X

Tabela 4.13 – Disposição dos Testes utilizados na Rede Neural Artificial

OBS: Nesse teste psicológico realizado, foi encontrado apenas um diagnóstico para “Saúde” e para “Família”, então por conveniência estes dois testes serão utilizados tanto para os treinamentos quanto para as generalizações das redes.

O vetor de entrada para o treinamento das redes foram codificados com **1** para as alternativas assinaladas e **0** para as não assinaladas. Os resultados das redes foram padronizados, isto é, se a saída for maior que **0,5**, utilizou-se valor **1**, caso contrário foi usado o valor **0**. Os quatro tipos de diagnósticos encontrados nos testes foram codificados em números binários conforme mostra a tabela 4.14.

Diagnóstico	Codificação
Personalidade	0 0 0 1
Saúde	0 0 1 0
Família	0 1 0 0
Família/Lazer	1 0 0 1

Tabela 4.14 – Diagnósticos dos testes codificados

4.2.3 - Rede Adaline

Na rede Adaline foi utilizada uma taxa de aprendizado igual a **0,01** e seu treinamento necessitou de **62** épocas para encontrar um erro menor que **0,00001**. A figura 4.4 exibe o erro encontrado em função do número de épocas utilizados no treinamento.

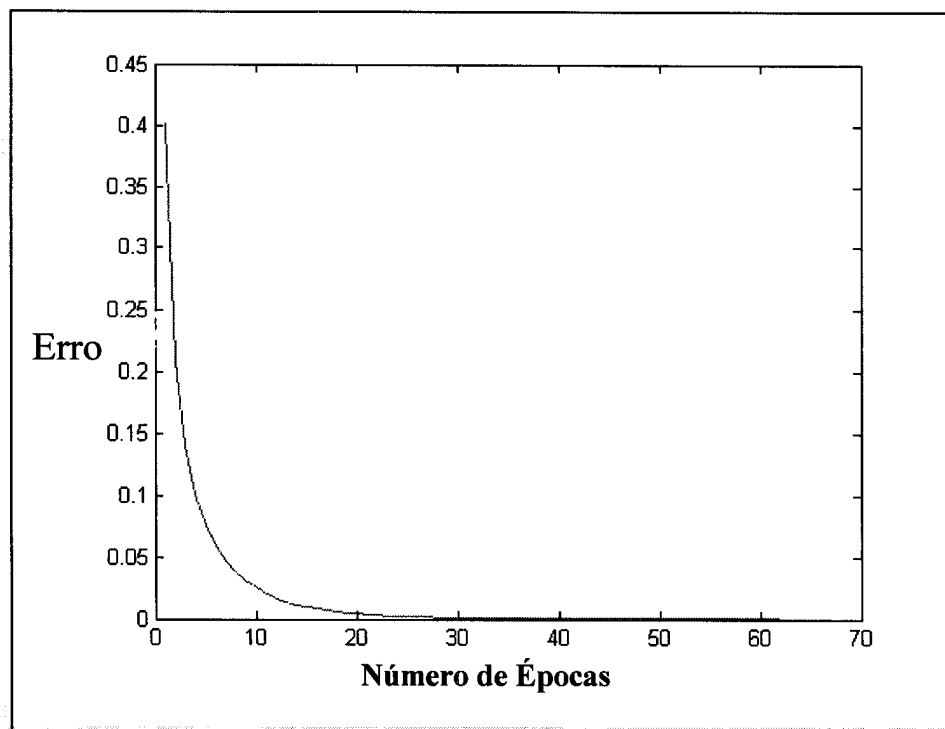


Figura 4.4 – Erro encontrado na rede Adaline no reconhecimento dos testes

Após o treinamento da rede foi verificada a porcentagem de aprendizado nos dados não visto no treinamento, obtendo um acerto de **55,56%**. O resultado pode ser visto na tabela 4.15 abaixo, onde mostra que a rede encontrou dificuldade no reconhecimento do teste 3 (Família\Lazer), teste 11 (Família\Lazer), teste 14 (Personalidade) e teste 16 (Personalidade).

Generalização dos testes psicológicos						
Teste	Saída desejada	Saída encontrada				Saída padronizada
3	0 0 1 0	0,0737	-0,0852	0,0616	0,5217	0 0 0 1 Errado
5	0 0 0 1	0,2497	0,1887	0,1205	0,7802	0 0 0 1 Correto
7	0 1 0 0	-0,0030	0,9963	0,0002	0,0039	0 1 0 0 Correto
8	0 0 0 1	0,0607	-0,0477	0,1333	0,5942	0 0 0 1 Correto
10	0 0 0 1	-0,1116	-0,2029	0,1576	0,8783	0 0 0 1 Correto
11	0 0 1 0	-0,1688	-0,0504	0,0806	0,5194	0 0 0 1 Errado
13	1 0 0 1	0,9954	-0,0035	0,0003	1,0033	1 0 0 1 Correto
14	0 0 0 1	0,0281	-0,0140	0,2327	0,3960	0 0 0 0 Errado
16	0 0 0 1	0,0544	0,2244	0,1423	0,4518	0 0 0 0 Errado

Tabela 4.15 – Resultado obtido da rede Adaline no reconhecimento dos testes

4.2.4 - Rede Kohonen (SOM)

Os testes para treinamento foram aplicados na rede Kohonen com uma taxa de aprendizado de **0,05** e treinados com uma duração de **20.000** épocas. Após o treinamento foi feita a generalização com os testes nunca vistos anteriormente pela rede o qual obteve um acerto de **55,56%**. A tabela 4.16 mostra o resultado da rede Kohonen. Nesta tabela, os quadros com preenchimento cinza são os grupos iguais encontrados na generalização da rede após o treinamento.

Como havia apenas um teste tanto para o diagnóstico Saúde como para Família, e como eles acabaram encontrando o mesmo grupo (grupo 1), por conveniência foi escolhido o diagnóstico Saúde como pertencente ao grupo 1 (vencedor). Já o teste 7 (Família) foi considerado como não obtendo aprendizagem alguma, pois os diagnósticos Família e Saúde têm que ser considerados como grupos diferentes.

Personalidade			
Testes Treinados	Grupo	Testes Generalizados	Grupo
1	4	5	2
6	4	8	4
9	3	10	4
12	4	14	1
15	1	16	3
17	4		

Saúde	Testes treinados	Grupo
	13	1
	Testes generalizados	Grupo
	13	1

Família	Testes treinados	Grupo
	7	1
	Testes generalizados	Grupo
	7	1

Família/Lazer			
Testes Treinados	Grupo	Testes Generalizados	Grupo
2	1	3	2
4	2	11	2

Tabela 4.16 – Resultado da rede Kohonen no teste MPCL

4.2.5 - Rede Perceptron de Múltiplas Camadas (MLP)

Na rede MLP foi utilizado uma taxa de aprendizado igual a **0,2**, apenas uma camada escondida com **10** neurônios na camada e uma função de ativação sigmoideal. O treinamento foi conduzido por **3685** épocas até atingir um erro quadrático médio menor que **0,01** (erro alvo). A figura 4.5 ilustra o erro encontrado em função do número de épocas necessárias no treinamento.

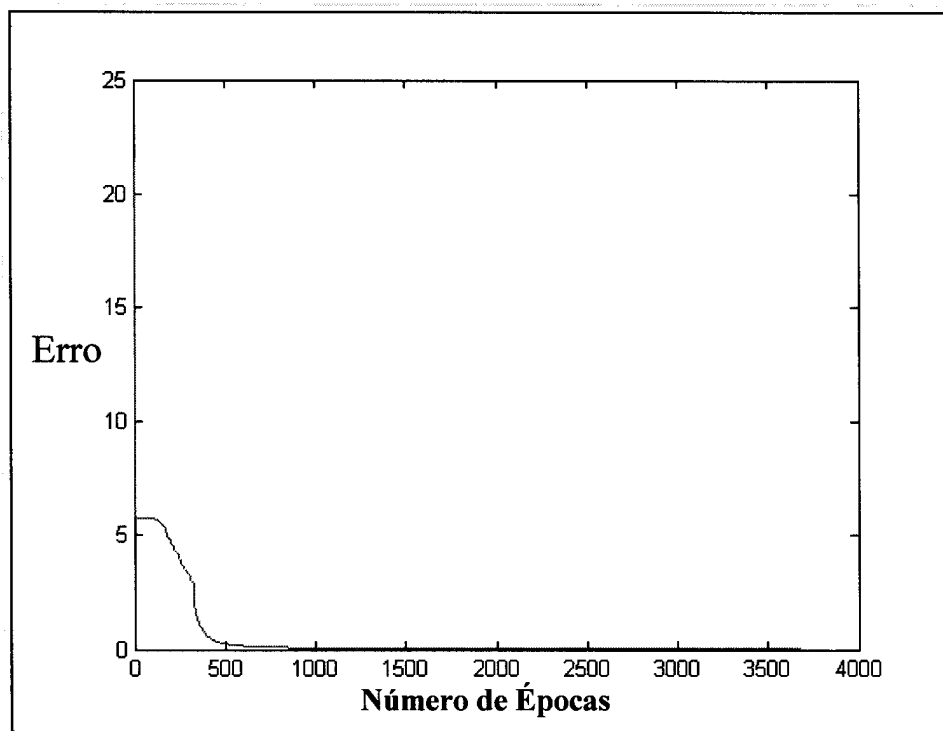


Figura 4.5 – Erro encontrado na rede Adaline no reconhecimento dos testes

Após o treinamento, o próximo experimento conduzido foi à generalização dos testes não vistos pela rede durante o treinamento, o qual obteve um acerto de **77,78%** aos dados apresentados. Na ativação do teste pelo MLP, a rede encontrou dificuldade na caracterização do teste 3 (Família/Lazer) e no teste 11 (Família/Lazer). Na tabela 4.17 está exibido o resultado da rede MLP.

Generalização dos testes psicológicos						
Teste	Saída desejada	Saída encontrada		Saída padronizada		
3	0 0 1 0	0,0105	0,0012	0,0188 0,9833	0 0 0 1	Errado
5	0 0 0 1	0,0080	0,0012	0,0181 0,9832	0 0 0 1	Correto
7	0 1 0 0	0,0005	0,9739	0,0249 0,0155	0 1 0 0	Correto
8	0 0 0 1	0,0080	0,0012	0,0181 0,9832	0 0 0 1	Correto
10	0 0 0 1	0,0080	0,0012	0,0181 0,9832	0 0 0 1	Correto
11	0 0 1 0	0,0080	0,0012	0,0181 0,9832	0 0 0 1	Errado
13	1 0 0 1	0,9435	0,0002	0,0121 0,9989	1 0 0 1	Correto
14	0 0 0 1	0,0080	0,0018	0,0496 0,9520	0 0 0 1	Correto
16	0 0 0 1	0,0080	0,0013	0,0195 0,9815	0 0 0 1	Correto

Tabela 4.17 – Resultado obtido da rede MLP no reconhecimento dos testes

4.2.6 - Rede Perceptron de Camada Simples (SLP)

Na rede Perceptron de Camada Simples (SLP) foi utilizado uma taxa de aprendizado igual a 0,5 com ciclos de 20.000 épocas para o seu treinamento. Após este processo, foi feita a generalização da rede que obteve 66,67% de acerto aos dados nunca vistos anteriormente. Na generalização da rede, foi encontrada uma dificuldade no aprendizado do teste 3 (Família/Lazer), no teste 11 (Família/Lazer) e no teste 14 (Personalidade). A tabela 4.18 mostra o resultado da rede perceptron de camada simples.

Teste	Saída desejada	Saída encontrada	
3	0 0 1 0	0 0 0 1	Errado
5	0 0 0 1	0 0 0 1	Correto
7	0 1 0 0	0 1 0 0	Correto
8	0 0 0 1	0 0 0 1	Correto
10	0 0 0 1	0 0 0 1	Correto
11	0 0 1 0	0 0 0 1	Errado
13	1 0 0 1	1 0 0 1	Correto
14	0 0 0 1	0 0 0 0	Errado
16	0 0 0 1	0 0 0 1	Correto

Tabela 4.18 – Resultado da rede Perceptron de Camada Simples

4.2.7 - Conclusão

Através da tabela 4.19, pode se observar que a rede Perceptron de Múltiplas Camadas (MLP) obteve um resultado melhor do que as demais redes neurais artificiais. Mas a mesma observação feita no teste VARK vale para esse teste MPCL, pois o número de padrões (questionários) são insuficientes para um resultado realmente promissor (100%) no aprendizado de Redes Neurais Artificiais. Por outro lado, pode-se perceber que pelos dois testes aplicados é possível obter um aprendizado de padrões de testes psicológicos usando como ferramenta as RNAs.

Rede Neural Artificial	Porcentagem de acerto
Adaline	55,56 %
Kohonen	55,56 %
Perceptron de Múltiplas Camadas	77,78 %
Perceptron simples	66,67 %

Tabela 4.19 – Resultados do treinamento das redes

5 - RECONHECIMENTO DE VOZ

5.1 - Aplicação

O objetivo na aplicação do reconhecimento de voz em Redes Neurais Artificiais é a identificação de níveis de stress na voz humana. Como resultado dessa aplicação, deseja-se saber se as RNAs conseguem distinguir se determinada voz de uma pessoa pode ser diagnosticada como “nervosa” (estressada) ou “calma”.

Para a concretização de tal objetivo, foi proposto a formulação de um banco de voz contendo falas de diversas pessoas. A gravação de todas as vozes se fez com a utilização de um gravador portátil da marca “Panasonic” que possui o nome de identificação “FP-Fast Playback / 2-Speed”. Todas as vozes foram gravadas em formato “WAV” a partir de um texto pré-determinado o qual possui uma duração média em torno de trinta segundos.

Após a gravação das vozes, estas foram manipuladas no programa “Sound Forge (versão 4.5)”, que é um sistema profissional de editor de som. Com esse programa foi obtida a gravação correta de todas as vozes “calmas” e “nervosas”. Na figura 5.1 está ilustrado duas das quatorze vozes gravadas utilizadas no reconhecimento de voz pelas RNA.

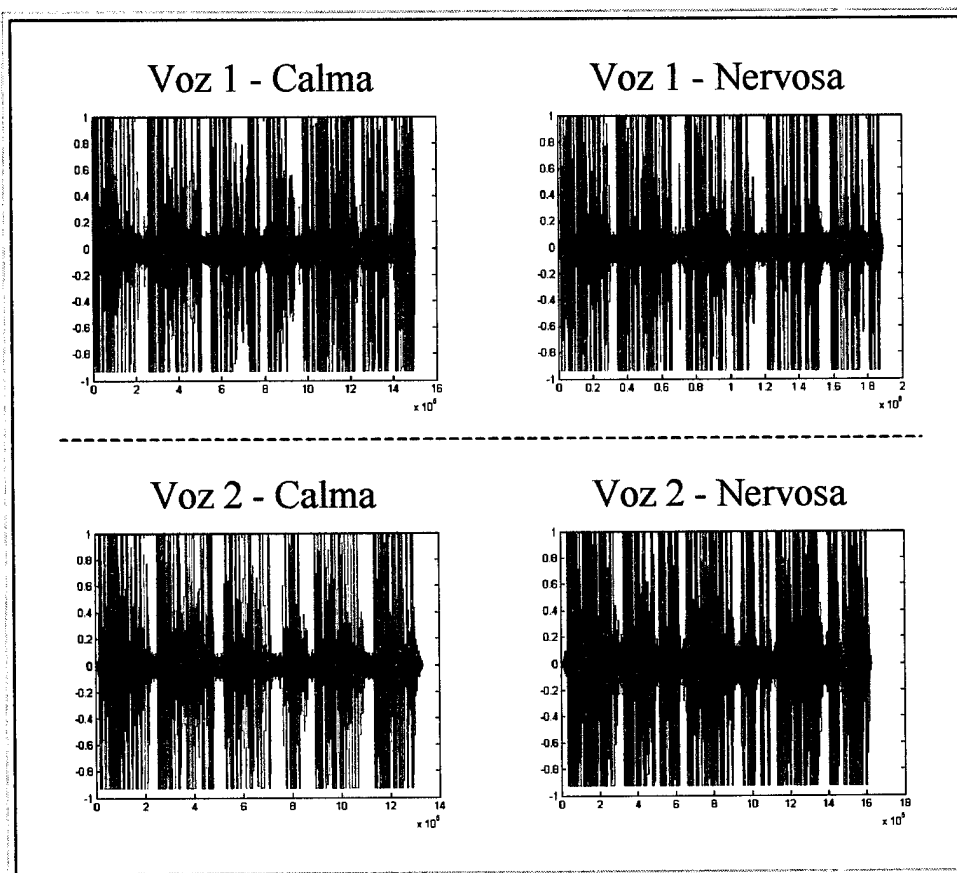


Figura 5.1 – Ilustração das vozes gravadas

As quatorze vozes gravadas foram utilizadas no treinamento e generalização da rede Perceptron de Múltiplas Camadas (Backpropagation) conforme mostra a tabela 5.1.

Voz	Diagnóstico	Treinamento	Generalização
1	Calmo	X	
	Nervoso	X	
2	Calmo	X	
	Nervoso	X	
3	Calmo	X	
	Nervoso	X	
4	Calmo	X	
	Nervoso	X	
5	Calmo	X	
	Nervoso	X	
6	Calmo	X	
	Nervoso	X	
7	Calmo	X	
	Nervoso	X	
8	Calmo	X	
	Nervoso	X	
9	Calmo	X	
	Nervoso	X	
10	Calmo	X	
	Nervoso	X	
11	Calmo		X
	Nervoso		X
12	Calmo		X
	Nervoso		X
13	Calmo		X
	Nervoso		X
14	Calmo		X
	Nervoso		X

Tabela 5.1 – Disposição das vozes utilizadas na rede MLP

A saída desejada para essas vozes pode ser vista codificada na tabela 5.2 onde a voz só pode ser diagnosticada como “calmo” ou “nervoso”.

Diagnóstico	Codificação
Voz calma	0 1
Voz nervosa	1 0

Tabela 5.2 – Codificação dos diagnósticos de voz

Como cada vetor de voz possui uma dimensão muito grande (441000 x 1), este foi dividido em blocos mil (441 blocos com o tamanho: 1000 x 1) e assim colocados seqüencialmente cada bloco na camada de entrada da rede MLP. A figura 5.2 ilustra esse processo.

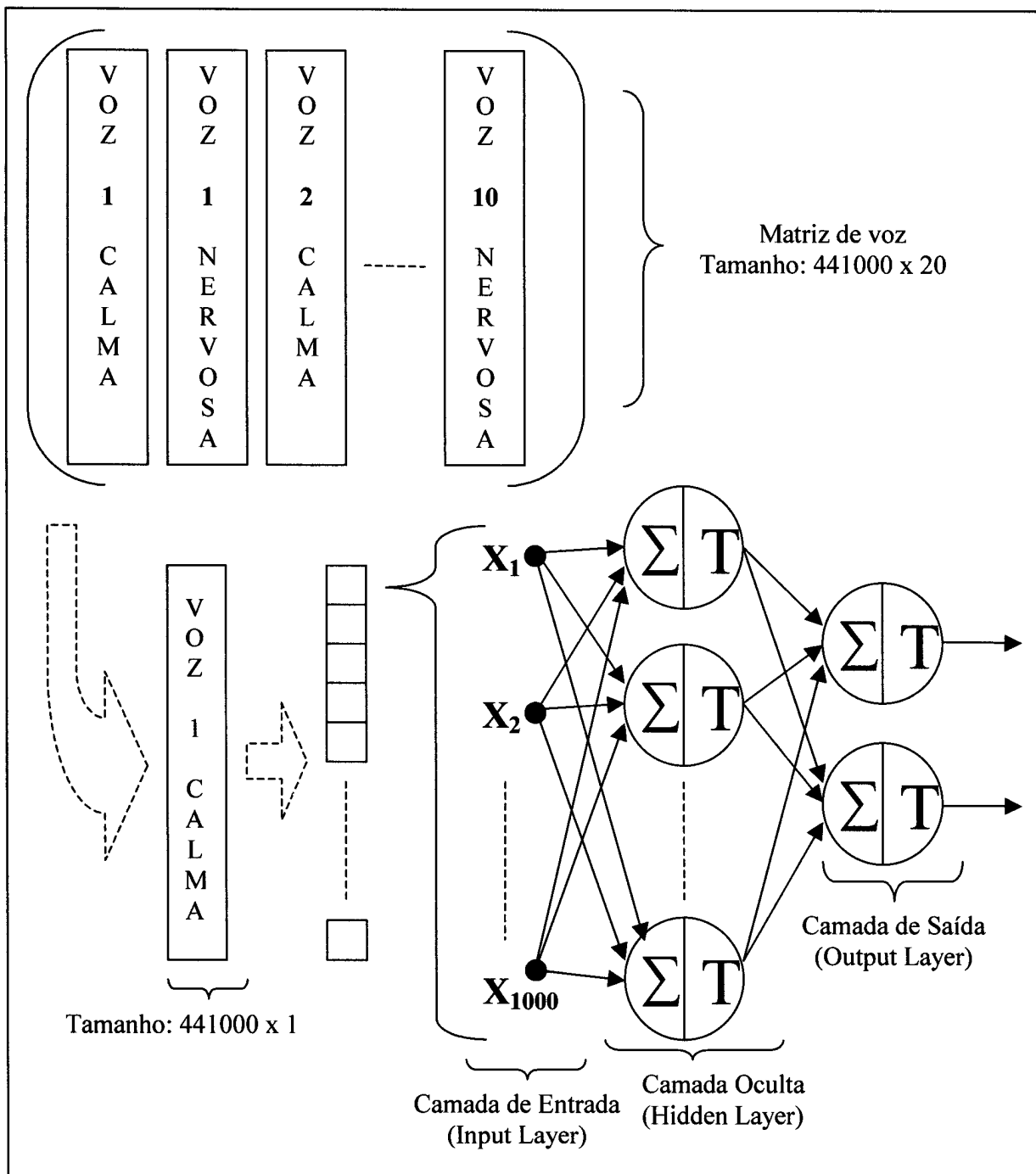


Figura 5.2 – Ilustração das Etapas utilizadas no treinamento da rede MLP

No treinamento, foi utilizada uma taxa de aprendizado de 0,6, uma função de ativação sigmoideal e apenas uma camada oculta (hidden layer) com 25 neurônios nesta camada. A rede MLP foi processada por um período de 16.007 épocas atingindo um erro

quadrático médio igual a **9.5703**. A figura 5.3 mostra o erro encontrado durante o treinamento.

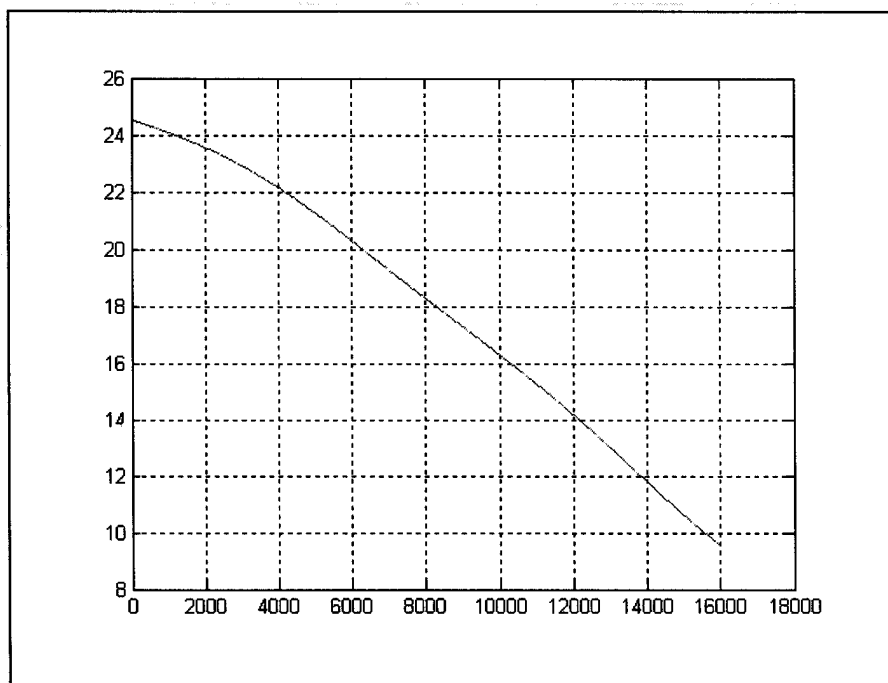


Figura 5.3 – Erro encontrado no treinamento da rede MLP

O próximo passo foi fazer a generalização das dez vozes que foram utilizadas no treinamento, para saber se a rede Perceptron de Múltiplas Camadas aprendeu os padrões o qual treinou. Na ativação da rede com as vozes que foram treinadas, observou um acerto de **97,78%**. Cada voz utilizada no treinamento conseguiu determinada taxa de acerto conforme mostra a tabela 5.3

Voz	Diagnóstico	Porcentagem de acerto
1	Calmo	99,32 %
	Nervoso	99,09 %
2	Calmo	99,32 %
	Nervoso	99,09 %
3	Calmo	96,83 %
	Nervoso	99,55 %
4	Calmo	99,09 %
	Nervoso	98,19 %
5	Calmo	96,83 %
	Nervoso	98,87 %
6	Calmo	95,01 %
	Nervoso	99,09 %
7	Calmo	93,88 %
	Nervoso	98,87 %
8	Calmo	99,09 %
	Nervoso	99,32 %
9	Calmo	94,10 %
	Nervoso	97,96 %
10	Calmo	92,06 %
	Nervoso	99,77 %

Tabela 5.3 – Porcentagem de acerto das vozes *treinadas* na rede MLP

Num segundo experimento, foram mostradas à rede as quatro vozes que não foram utilizadas no treinamento, com o intuito de verificar a porcentagem de acerto com padrões de vozes nunca vistos anteriormente pela rede MLP. Verificou-se que na ativação a rede conseguiu obter um acerto de **50,2%** e que a porcentagem de acerto dos diagnósticos de cada voz pode ser visto na tabela 5.4 abaixo.

Voz	Diagnóstico	Porcentagem de acerto
1	Calmo	45,12 %
	Nervoso	60,32 %
2	Calmo	43,99 %
	Nervoso	56,46 %
3	Calmo	55,78 %
	Nervoso	44,44 %
4	Calmo	58,73 %
	Nervoso	32,65 %

Tabela 5.4 – Porcentagem de acerto das vozes *generalizadas* na rede MLP

5.2 - Conclusão

Através dessa aplicação de reconhecimento de voz em Redes Neurais Artificiais, conclui-se que esta ferramenta propicia o aprendizado de padrões na mudança no comportamento humano (50,2% de sucesso). Cabe ressaltar que alguns métodos utilizados impossibilitam que as RNAs aprendam de modo eficiente os dados apresentados, como por exemplo:

- Utilização de arquivos de voz muito grande (441000 dados cada voz), fazendo com que o treinamento demande muito tempo de processamento;
- Utilização de um gravador (Panasonic) e uma fita cassete (miniatura) que impossibilita a gravação de voz de alta qualidade, fazendo com que fatores sonoros externos influenciem nos dados de voz. A figura 5.4 ilustra as ferramentas utilizadas na gravação de voz;

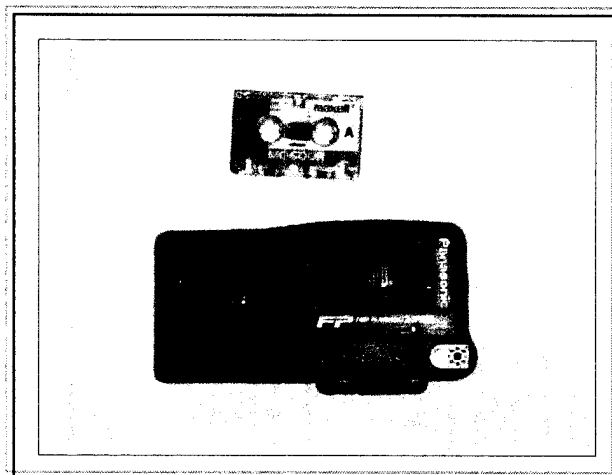


Figura 5.4 – Instrumento utilizado na gravação de voz

- Pode se dizer que as pessoas que forneceram suas vozes para a gravação foram forçadas a gravar a voz em tom nervoso, o que pode diferir de uma voz com um estado emocional realmente nervoso.
- A utilização de um banco de dados de vozes pequeno (14 vozes), o que faz com que a rede neural artificial tenha poucas variedades de padrões para o aprendizado.

Acredita-se que usando um banco de dados com um número maior de padrões, e sendo essas vozes refinadas (alta qualidade na gravação da voz da pessoa), a possibilidade de uma rede como o Perceptron de Múltiplas Camadas (MLP) de aprender dados nunca vistos no treinamento é bem maior (nível mais aceitável – acima de 70% de acerto). Por outro lado, percebe-se que a potencialidade das Redes Neurais Artificiais é grande e promissor no reconhecimento das mudanças de tonalidades das vozes humanas.

6 - CONCLUSÕES E RECOMENDAÇÕES FINAIS

O principal objetivo deste trabalho foi adquirir conhecimento sobre Redes Neurais Artificiais, sua aplicação em diversos propósitos, bem como a verificação do aprendizado desta técnica no reconhecimento de mudanças no comportamento humano (nível de stress) condicionados pelo isolamento social.

O modelo proposto no reconhecimento de vozes calmas e nervosas (capítulo 5) demonstra que esse objetivo pode ser alcançado. O desempenho do sistema utilizado pode ser melhorado. Para tanto, podem ser aplicadas técnicas, além da ordenação e supressão do silêncio, capazes de realizar um refinamento mais detalhado das características ocultas da voz a ser processada. Citamos por exemplo, a FFT (*Fast Fourier Transform*), *Wavelet*, conjuntos difusos, no entanto, a aplicação dessas técnicas fica como sugestão para novos trabalhos.

A aplicação dos dois testes psicológicos demonstra que Redes Neurais Artificiais podem ser capazes de produzir um resultado confiável na realização de diagnósticos psicológicos e que quanto maior o número de padrões utilizados no seu treinamento, maior será a capacidade de aprendizado da rede e sua posterior generalização.

Para a continuação desse trabalho, sugere-se a utilização de técnicas em conjunto na identificação do nível de stress de uma pessoa que vive em confinamento, como o batimento cardíaco, sudorese, testes psicológicos específicos, correntes elétricas emitidas pelo cérebro na mudança de comportamento, expressão facial, a pressão com que o usuário responde um questionário através de um teclado, entre outros métodos.

7 - REFERÊNCIAS BIBLIOGRÁFICAS

- [HAY-94] HAYKIN, Simon. **Neural Networks: A Comprehensive Foundation**.
New York: Macmillan College Publish Company, 1994.
- [FAU-94] FAUSET, Laurene. **Fundamentals of Neural Networks: Architectures, Algorithms, and Applications**.
- [NEU-04] Neuroquímica. Tudo sobre a química do cérebro. Disponível por WWW em <http://quark.qmc.ufsc.br/qmcweb/artigos/neuroquimica.html>
- [MEC-04] O cérebro, conceitos a explorar. Disponível por WWW em <http://www.mec.gov.br/semtec/ensmed/ftp/cerebro.pdf>.
- [HEL-04] HELENA, Silvia Cardoso. Guia Ilustrado da Anatomia e Funcionamento do Corpo Humano. Disponível por WWW em <http://www.nib.unicamp.br/svol/giovanna.html>.
- [SNC-04] Sistema Nervoso Central em Vertebrados. Disponível por WWW em <http://curlygirl3.no.sapo.pt/nervoso.htm#evolucao>.
- [BRA-04] BRAGA, Antônio de Pádua. **Redes Neurais Artificiais: Teoria e aplicações**. Disponível por WWW em <http://www.cpdee.ufmg.br/~apbraga/discpos.html>.
- [TPC-04] The Psychological Corporation: Meeting your Assessment Needs. Disponível por WWW em <http://www.harcourt-uk.com>.
- [MPC-50] The Mooney Problem Check List. Disponível por WWW em http://www.yorku.ca/psycentr/tests/per_test.html#67.
- [TAF-04] TAFNER, Malcon Anderson. **As Redes Neurais Artificiais: Aprendizado e Plasticidade**. Disponível por WWW em <http://www.epub.org.br/cm/n05/tecnologia/plasticidade2.html>.
- [DES-04] DESCHAMPS, Fernando. **Redes Neurais Artificiais: Fundamentos**. Disponível por WWW em http://s2i.das.ufsc.br/downloads/Apresentacao_RNA.pdf.
- [OSO-04] OSÓRIO, Fernando Santos. **Redes Neurais Artificiais**. Universidade do Vale do Rio dos Sinos - Centro de Ciências Exatas e Tecnológicas. Disponível por WWW em <http://www.inf.unisinos.br/~osorio/neural.html>.
- [INF-04] INFO WESTER. **Redes Neurais Artificiais. Seção software**. Disponível por WWW em <http://www.inf.unisinos.br/~osorio/neural.html>.
- [FAB-04] FABRI, José Augusto. **Redes Neurais Artificiais**. Disponível por WWW em <http://users.femanet.com.br/~fabri/RNAs.htm>.

- [SOU-04] SOUTO, Marcílio. Redes Neurais Artificiais. Centro de Informática – UFPE. Disponível por WWW em <http://users.femanet.com.br/~fabri/RNAs.htm>.
- [DAC-92] Data and Analysis Center for Software. Artificial Neural Networks. Disponível por WWW em <http://www.dacs.dtic.mil>.
- [VEL-04] VELOSO, Renê Rodrigues. Sistemas baseados em aprendizagem – Redes Neurais Artificiais. Algoritmo de Kohonen. Disponível por WWW em www.inf.furb.rct-sc.br/~alx/ia.html.
- [RNA-04] Redes Neurais Artificiais: Do Aprendizado Natural ao Aprendizado Artificial. Disponível por WWW em <http://www.inf.unisinos.br/~osorio/IForumIA/fia99-ppt.pdf>.
- [NEP-04] NEPOMOCENO, Marcelo Kapp. Reconhecimento de palavras manuscritas utilizando Redes Neurais Artificiais. Disponível por WWW em <http://www.livia.etsmtl.ca/publications/2004/MNK-Dissertation-final.pdf>.
- [ANN-04] Artificial Neural Networks. Tutorial introdutório sobre Redes Neurais Artificiais, em especial sobre as Redes Multi Layer Perceptron treinadas com BackPropagation. Disponível por WWW em <http://7mares.terravista.pt/ivan/neurais.html>.