

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE MATEMÁTICA

DEPARTAMENTO DE ESTATÍSTICA

**REDES NEURAS ARTIFICIAIS NO CONTEXTO
ESTATÍSTICO**

LISIANE PRISCILA ROLDÃO SELAU

ORIENTADOR: JOÃO RIBOLDI

Monografia apresentada para obtenção do
grau de Bacharel em Estatística

Porto Alegre, Dezembro de 2000

AGRADECIMENTOS

Para que este trabalho fosse realizado contei com a colaboração de várias pessoas, as quais gostaria, neste momento, de expressar meus mais sinceros agradecimentos:

Ao meu orientador, Prof. Dr. João Riboldi; que nunca desistiu de me orientar; pelo seu interesse, sugestões, dedicação, paciência e total apoio.

Ao meu namorado, Everaldo, que sempre me incentivou a prosseguir no desenvolvimento deste trabalho, não me deixando desistir em momento algum, transmitindo-me confiança e incentivo, que demonstrou enorme calma nos meus longos períodos de ausência.

Às minhas amigas, Dierê, Juliane e Mariana, colegas de graduação, que sempre me acompanharam nas horas de estudo, que me distraíram nas horas difíceis e que, acima de tudo, sempre me apoiaram em todas minhas decisões.

À minha irmã, Lívian, pelo companheirismo em vários momentos da minha graduação.

Ao professor da Universidade Federal do Rio Grande do Norte, Paulo Roberto M. de Azevedo, pela disponibilização de vários materiais usados como referências neste trabalho e pela força transmitida em várias ajudas dadas via e-mail.

Agradeço, ainda, a todos que de alguma forma contribuíram, mesmo de maneira indireta, para a elaboração deste trabalho.

GLOSSÁRIO DE PALAVRAS

Algoritmos genéticos: Modelos de aprendizado desenvolvidos baseados nos princípios de evolução. Dentre soluções parciais de um problema a melhor solução é selecionada e combinada para dar a base para a futura solução do problema.

Amostra de treinamento: Observações utilizadas na calibração da rede neural. Deve conter valores para os nós de saída para que os erros nos valores finais preditos possam ser determinados e usados para o processo de aprendizado.

Aprendizado/Treinamento: Processo seqüencial de grandes amostras de observações, conhecidas como amostras de treinamento, em que os erros de predição ou classificação são usados para recalibrar os pesos para melhorar a predição.

Retropagação: Processo de aprendizado mais comum de redes neurais, em que erros de estimação nos nós de saída são retornados através do sistema às camadas de entrada e são usados como indicadores para recalibrar os pesos para cada nó.

Camadas escondidas ou intermediárias: Camadas de nós em um perceptron multi-camada que estão entre os nós de entrada e saída. O pesquisador pode ou não controlar o número de nós bem como o número de camadas escondidas. Também conhecidas como camadas ocultas.

Data mining: Processo de busca de informações que analisa grande quantidade de dados de forma a descobrir relacionamentos entre eles, procurando identificar padrões e tendências.

Função de ativação: Função matemática dentro do nó que converte a soma de ativação das entradas em ativação de saída. Existem vários tipos de funções de ativação, a forma mais comum é a função sigmóide.

Função sigmóide: Função de ativação não linear com uma distribuição em forma de S. Um exemplo mais comum é a função logística.

Função somatório: Combinador linear que soma os sinais de entrada ponderando-os pelos respectivos pesos sinápticos do neurônio.

Inteligência artificial (IA): Área da ciência da computação habituada com a criação de programas de computadores que tenta simular as propriedades de aprendizado do cérebro humano. IA está envolvida em diversas áreas especializadas, incluindo redes neurais e algoritmos genéticos.

Nó ou neurônio: Unidade de processamento mais básica de uma rede que pode atuar como entrada, saída ou como função de processamento e análise. Análogo ao neurônio de um cérebro humano.

Perceptron multi-camada (MLP): A mais conhecida e utilizada rede neural que contém pelo menos uma camada de nós escondidas entre as camadas de entrada e saída.

Pesos: Constantes que são usadas em um modelo para informar a influência de cada nó na saída da unidade.

Rede Neural Artificial (RNA): Também chamada, de modo simplificado, de Rede Neural é um modelo preditivo não linear que adquire conhecimento através do treinamento com exemplos. Se assemelha a estrutura do sistema neural biológico.

Sinapse: Mecanismo transmissor de informação entre os neurônios. A informação recebida pelo neurônio é processada e produz uma saída que pode excitar ou inibir outros neurônios. O tipo mais comum de sinapse é a sinapse química, que converte um sinal químico e retorna um sinal elétrico e vice-versa.

Supervisionado: Processo de aprendizado que utiliza uma amostra de treinamento e dá retorno para a rede neural a respeito dos erros dos nós de saída.

Topologia: É a arquitetura de uma Rede Neural com suas camadas de entrada e saída e quando existirem suas camadas intermediárias ou escondidas.

SUMÁRIO

1	- INTRODUÇÃO	7
	1.1 - OBJETIVOS	8
	1.2 - ESTRUTURA DA MONOGRAFIA	9
2	- REFERENCIAL HISTÓRICO	11
3	- REDES NEURAIIS NATURAIS	15
4	- REDES NEURAIIS ARTIFICIAIS	19
	4.1 - ARQUITEURA DE UMA REDE	26
	4.2 - APRENDIZADO DA REDE	29
	4.2.1- TEORIA ESTATÍSTICA DO PROCESSO DE APRENDIZAGEM	33
	4.3 - INTELIGÊNCIA ARTIFICIAL E NEUROCOMPUTAÇÃO	40
	4.4 - REDES NATURAIS x REDES ARTIFICIAIS	42
	4.5 - ALGUMAS REDES MAIS UTILIZADAS	44
	4.5.1- REDE PERCEPTRON.....	44
	4.5.2- REDE HOPFIELD.....	47
	4.5.3- REDES ADALINE E MADALINE	48
	4.5.4- REDE DE RESSONÂNCIA ADAPTATIVA.....	50
	4.5.5- REDE KOHONEN	51
	4.6 - RESUMO DA APLICAÇÃO DE REDES NEURAIIS	52
	4.6.1- COLETA DE DADOS E SEPARAÇÃO EM CONJUNTOS	53
	4.6.2- CONFIGURAÇÃO DA REDE	53
	4.6.3- TREINAMENTO	54

4.6.4-	TESTE	55
4.6.5-	INTEGRAÇÃO	55
5	- MODELAGEM ESTATÍSTICA	57
5.1	- ANÁLISE DISCRIMINANTE	58
5.2	- ANÁLISE DE REGRESSÃO LINEAR	61
5.3	- ANÁLISE DE REGRESSÃO LOGÍSTICA	63
5.4	- ANÁLISE DE SÉRIES TEMPORAIS	64
6	- APLICAÇÕES	68
6.1	- ANÁLISE DE CRÉDITO	68
6.2	- FINANÇAS	70
6.3	- MARKETING	71
6.4	- MORTALIDADE INFANTIL	73
6.5	- SAÚDE	74
6.6	- MEDICINA	76
7	- EXEMPLO	79
7.1	- CARACTERIZAÇÃO E PREPARAÇÃO DOS DADOS	81
7.2	- ANÁLISE E RESULTADOS	88
7.3	- INTEGRAÇÃO DA REDE	96
8	- CONCLUSÃO	100
9	- REFERÊNCIAS BIBLIOGRÁFICAS	102

1. INTRODUÇÃO

A partir do momento em que as máquinas começaram a evoluir, um grande desejo do homem tem sido a criação de uma máquina que possa operar independentemente do controle humano. Uma máquina cuja independência seja desenvolvida de acordo com seu próprio aprendizado e que tenha a capacidade de interagir com ambientes incertos, que possa ser chamada de autônoma, inteligente ou cognitiva. Organismos humanos são uma fonte de motivação para o desenvolvimento destas máquinas, e proporcionam diversas opções para o desenvolvimento de algoritmos de aprendizado e adaptação.

Redes Neurais são técnicas computacionais que apresentam um modelo matemático inspirado na estrutura neural de organismos inteligentes e que adquirem conhecimento através da experiência.

O sistema é composto de um grande número de elementos de processamento, interconectados que se assemelham com neurônios e são ligados com conexões de pesos que se assemelham com sinapses. O nome Redes Neurais decorre desta semelhança com o cérebro humano.

Redes neurais é considerada uma técnica emergente na análise multivariada e é utilizada em conjunto com Data Mining para, dentre outras aplicações, reconhecer padrões de perfis. O Data Mining é uma metodologia de identificação de padrões e tendências de dados que usa técnicas estatísticas para examiná-los, detectando tendências e associações escondidas. Dentre estas várias técnicas está a metodologia de Redes Neurais que, utilizando inteligência artificial, auxilia no conhecimento da base de dados. Um melhor entendimento do que é o Data Mining pode ser obtido através da visualização do esquema da Figura 1.1.

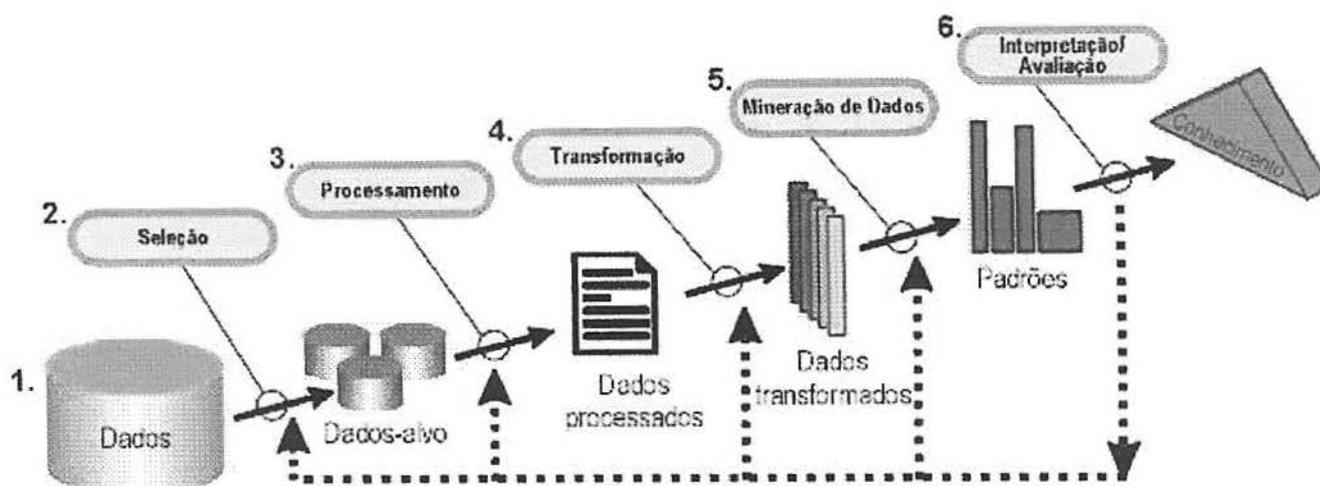


FIGURA 1.1: Visão geral das etapas que compõe o processo de Data Mining.

O estudo de Redes Neurais é um dos ramos da Inteligência Artificial que mais tem se desenvolvido, atraindo o interesse de várias áreas do conhecimento. Aplica-se em diversas áreas do conhecimento como finanças, varejo e logística, medicina e saúde, marketing, manufaturamento e controle industrial, energia, dentre outras.

Dada a característica multidisciplinar de Redes Neurais, o papel do estatístico é essencial juntamente com profissionais de outras áreas, especialmente da computação. Justifica-se a participação do Estatístico pelos seus conhecimentos avançados em técnicas estatísticas e sua sólida formação matemática e probabilística, e por ser capaz de auxiliar no desenvolvimento computacional da rede no que diz respeito, principalmente, aos modelos matemáticos de uma rede neural.

1.1. OBJETIVOS

O problema de interesse desta monografia é mostrar a real importância do Estatístico no desenvolvimento de Redes Neurais, ressaltando-se a interface potencial com outras áreas do conhecimento.

Quer-se mostrar, com este estudo, a superioridade da técnica em comparação com soluções triviais utilizadas freqüentemente em diversas situações. Esta superioridade pode ser de certa forma explicada pelo fato de que essas novas gerações de algoritmos podem aprender com o passado e prever o futuro e, também, extrair regras para trabalhar em ambientes complexos.

Pretende-se detalhar e exemplificar algumas aplicações da técnica nas diversas áreas do conhecimento, bem como ilustrar a análise correta dos resultados, mostrando, assim, como é indispensável a presença do Estatístico desde a modelagem da rede até a etapa final da interpretação dos resultados, juntamente com os especialistas da área.

Com o intuito de explorar o tema em questão, procurar-se-á entender redes neurais como uma ferramenta estatística, proceder-se-á a revisão bibliográfica, para detalhar as abordagens e diferentes enfoques existentes e estabelecer-se-á a fundamentação teórica da técnica, explorando, principalmente, a sua interface com a Estatística.

Pretende-se, também mostrar que uma rede neural não pode ser encarada simplesmente como uma caixa preta, que existem metodologias que lhe dão sustentabilidade, como por exemplo, para a obtenção dos pesos da rede.

1.2. ESTRUTURA DA MONOGRAFIA

Neste trabalho serão apresentados enfoques da técnica de Redes Neurais para análise de bases de dados, visando, dentre outros aspectos, a classificação e o reconhecimento de padrões.

O capítulo 2 apresenta um referencial histórico do que tem sido referido ao longo dos últimos anos a respeito da técnica. Os fatos ocorridos são mencionados em ordem cronológica mostrando a evolução da técnica, dando especial ênfase ao aspecto de entender redes neurais como ferramenta estatística.

No capítulo 3 apresenta-se a rede neural natural com a fisiologia de um neurônio biológico como ponto de partida para a apresentação de uma Rede Neural Artificial.

O capítulo 4 caracteriza de forma matemática e, também sob o ponto de vista estatístico, a Rede Neural Artificial, detalhando seu funcionamento e componentes. Em seguida é apresentada as diferentes formas de subdivisão e algumas das redes mais conhecidas e utilizadas.

O capítulo 5 é o mais importante do ponto de vista estatístico por detalhar algumas aplicações e relações com algumas conhecidas técnicas de Inferência Estatística com Redes Neurais. Entendendo-se Redes Neurais Artificiais como ferramenta estatística compara-se com algumas das técnicas estatísticas clássicas.

No capítulo 6 são descritas algumas aplicações de Redes Neurais Artificiais nas diversas áreas do conhecimento como marketing, medicina, finanças e lingüística.

No Capítulo 7 explora-se um exemplo de uma das aplicações de Redes Neurais e é ilustrado o uso de um software específico para a técnica. O exemplo estudado foi o de avaliação de crédito na área de finanças, utilizando-se o Módulo Enterprise Miner especializado em Data Mining componente do Pacote Estatístico SAS (Statistical Analysis Systems). A técnica de Redes Neurais é uma das diversas ferramentas disponíveis no Enterprise Miner para a solução de problemas utilizando Data Mining.

Aos leitores mais interessados na aplicação dos procedimentos de solução e nos resultados que podem ser obtidos sugere-se que baseiem sua leitura no capítulo 7. Aos demais leitores aconselha-se a leitura completa na ordem apresentada.

2. REFERENCIAL HISTÓRICO

As primeiras informações sobre a neurocomputação aparecem em McCULLOCH & PITTS (1943) que sugeriram a construção de uma máquina baseada ou inspirada no cérebro humano, definindo os primeiros conceitos para o estudo de redes neurais artificiais e apresentando, assim, o modelo inicial de neurônio artificial denominado Psychon. Segundo KOVACS (1996) a essência da proposta de McCULLOCH & PITTS (1943) foi a seguinte: *'A inteligência é equivalente a cálculo proposicional que por sua vez pode ser implementado por funções booleanas. Por outro lado, o sistema nervoso é composto de redes de neurônios, que com as devidas simplificações, tem a capacidade básica de implementar estas funções booleanas. Conclusão: a ligação entre inteligência e atividade nervosa fica estabelecida de forma científica.'*

HEBB (1944) perseguiu a idéia de que o condicionamento psicológico clássico está presente em qualquer parte dos animais pelo fato de que esta é uma propriedade de neurônios individuais. Suas idéias não eram completamente novas, ele foi o primeiro a propor uma lei de aprendizagem específica para as sinapses dos neurônios. Este primeiro e corajoso passo serviu de inspiração para que muitos outros pesquisadores perseguissem a mesma idéia.

O primeiro neurocomputador, denominado Snark foi construído por Mavin Minsky em 1951. O Snark operava com sucesso a partir de um ponto de partida técnico, ajustando seus pesos automaticamente, entretanto, ele nunca executou qualquer função de processamento de informação interessante, mas serviu de inspiração para as idéias de estruturas que o sucederam.

Em 1956 no "Darthmouth 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 isso. Já

a Inteligência Artificial Conexionista acredita que construindo-se 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.

ROSENBLATT (1962) idealizou o modelo básico do Perceptron que despertou o interesse de muitos pela sua capacidade de reconhecer padrões simples.

WIDROW & HOFF (1960) desenvolveram um novo tipo de elemento de processamento de redes neurais chamado de Adaline, equipado com uma poderosa lei de aprendizado, que diferente do Perceptron utiliza a Regra Delta de aprendizado, ou como também é chamado, algoritmo de aprendizado LMS (Least Mean Square). Widrow também fundou a primeira companhia de hardware de neurocomputadores e componentes. Estes modelos foram muito criticados por MINSKY & PAPERT (1969) que apontavam a impossibilidade de uma rede de um único nível como o Perceptron e o Adaline ser capaz de solucionar problemas que não sejam linearmente separáveis.

Os anos seguintes foram marcados por um entusiasmo exagerado de muitos pesquisadores, que passaram a publicar muitos artigos e livros que faziam uma previsão pouco confiável para a época, sobre máquinas que seriam tão poderosas quanto o cérebro humano, que surgiriam em um curto espaço de tempo. Isto reduziu a credibilidade dos estudos desta área e provocou grandes aborrecimentos aos técnicos de outras áreas.

Somente com RUMELHART, et al (1986), que propuseram um algoritmo que permitia ajustar os pesos em uma rede com mais de um nível, o problema veio a se resolver. O algoritmo Retropropagação era capaz de propagar o erro através dos níveis numa fração proporcional à influência de cada neurônio na resposta incorreta. Este estudo resultou no modelo de Redes Neurais Artificiais, mais utilizado atualmente, o das redes Perceptron Multi-Camadas (MLP), treinadas com o algoritmo de Retropropagação.

Em 1987 ocorreu em São Francisco a primeira conferência de redes neurais em tempos modernos, a IEEE International Conference on Neural Networks, e também foi

formada a International Neural Networks Society (INNS). A partir destes acontecimentos seguiu-se a criação do INNS journal em 1989 e do Neural Computation e do IEEE Transactions on Neural Networks em 1990.

Desde 1987, muitas universidades no país e no exterior têm estruturado institutos de pesquisa e têm desenvolvido programas de educação em neurocomputação.

Redes Neurais têm sido utilizadas em uma grande quantidade de aplicações onde os métodos estatísticos são tradicionalmente empregados. Elas têm sido utilizadas em problemas de classificação como identificação de contatos de sonáres aquáticos (GORMAN & SEJNOWSKI, 1998), e problemas de predição de problemas cardíacos em pacientes (BAXT, 1990, 1991; FUJITA et al, 1992). Redes Neurais também são usadas em diversas áreas como diagnóstico de hipertensão (POLI et al, 1991), jogo de gamão (TESAURO, 1990) e reconhecimento de fala (LIPPMANN, 1989). Em aplicações de séries temporais as redes neurais tem sido utilizadas no melhoramento da performance de estoque de supermercados (HUTCHINSON, 1994). Rede Neural é atualmente a ferramenta preferida na predição de estruturas secundárias protéicas (QIAN & SEJNOWSKI, 1988). Pode-se resolver estes tipos de problemas através dos modelos clássicos de estatística como análise discriminante (FLURY & RIEDWYL, 1990), regressão logística (STUDENMUND, 1992), Bayesiana e outros tipos de métodos classificatórios (DUDA & HART, 1973), regressão múltipla (NETER et al, 1990), e modelos de séries temporais como ARIMA e outros métodos de previsão (STUDENMUND, 1992). É então hora de reconhecer Redes Neurais como ferramenta de análise de dados.

Vários autores têm feito estudos comparativos entre métodos estatísticos e redes neurais (HRUSCHKA, 1993; WU & YEN, 1992). Há um grande número de artigos introdutórios em redes neurais, por exemplo, LIPPMANN (1987) dá uma excelente visão de redes neurais para o processamento de sinais. Há também um grande número de bons livros introdutórios em redes neurais, como HERTZ et al (1991) reproduz uma boa descrição

matemática, SMITH (1993) explora o algoritmo de retropropagação em um ambiente aplicado, e FREEMAN (1994) usando exemplos para explicar redes neurais. Há também jornais falando a respeito de redes neurais e métodos estatísticos (BUNTINE & WEIGEND, 1991; RIPLEY, 1994; SARLE, 1994; WERBOS, 1991). Um dos melhores para uma visão geral sobre o assunto é RIPLEY (1993).

3. REDES NEURAIIS NATURAIS

Composto por aproximadamente 10 bilhões de neurônios, o cérebro humano é considerado o mais fascinante processador existente baseado em carbono. 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. As sinapses transmitem estímulos através de diferentes concentrações de sódio e potássio, e o resultado disto pode ser estendido por todo o corpo humano. Esta grande rede proporciona uma enorme capacidade de processamento e armazenamento de informação.

O sistema nervoso é formado por um conjunto extremamente complexo de neurônios. Nos neurônios a comunicação é realizada através de impulsos. Quando um impulso é recebido, o neurônio o processa e, passado um limite de ação, dispara um segundo impulso que produz uma substância neurotransmissora, fluindo do corpo celular para o axônio, que por sua vez pode ou não estar conectado a um dendrito de outra célula. O neurônio que transmite o pulso pode controlar a frequência de pulsos aumentando ou diminuindo a polaridade na membrana pós sináptica. Eles tem um papel essencial na determinação do funcionamento, comportamento e do raciocínio do ser humano.

O sistema nervoso humano pode ser visto como um sistema de três estágios, como mostrado no diagrama em bloco na Figura 3.1 (ARBIB, 1987).

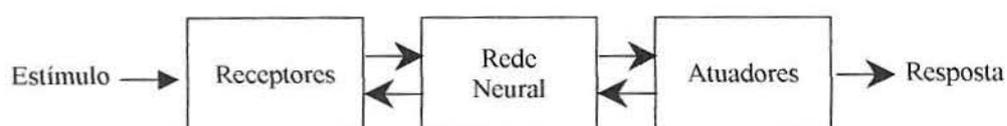


FIGURA 3.1: Representação em diagrama em blocos do sistema nervoso.

O centro do sistema é o cérebro, representado pela rede neural (natural), que recebe continuamente informação, percebe-a e toma decisões apropriadas. Dois conjuntos de setas

são mostrados na figura. Aquelas que apontam da esquerda para direita indicam a transmissão para frente do sinal portador de informação, através do sistema. As setas apontando da direita para a esquerda indicam a presença de realimentação no sistema. Os receptores convertem estímulos do corpo humano ou do ambiente externo em impulsos elétricos que transmitem informação para a rede neural natural (cérebro). Os atuadores convertem impulsos elétricos gerados pela rede neural natural em respostas discerníveis como saída do sistema.

A origem da teoria de Redes Neurais remonta aos modelos matemáticos de neurônios biológicos. A célula nervosa, ou neurônio, foi identificado anatomicamente e descrito com notável detalhe, pelo neurologista espanhol Ramón y Cajal no século 19 (KOVÁCS, 1996).

A Figura 3.2 detalha os principais componentes dos neurônios que são:

- Os dentritos, que tem por função, receber os estímulos transmitidos pelos outros neurônios;
- O corpo de neurônio, também chamado de soma; composto do núcleo, do citoplasma e da membrana celular; que é o responsável por coletar e combinar informações vindas de outros neurônios;
- E finalmente o axônio, que é constituído de uma fibra tubular que pode alcançar até alguns metros, e é responsável por transmitir os estímulos para outras células.

O tecido nervoso forma o cérebro, a medula espinhal, os nervos e os gânglios nervosos. O cérebro e a medula espinhal constituem o sistema nervoso central; os nervos e gânglios nervosos, por sua vez, constituem o sistema nervoso periférico. O sistema nervoso forma-se a partir do tubo nervoso embrionário. A unidade do tecido nervoso é o neurônio. O neurônio é uma célula altamente especializada na percepção e transmissão de estímulos. Ao receber um estímulo qualquer, seja de natureza química, elétrica ou mecânica, os dentritos geram um impulso nervoso, o qual se transmite ao corpo celular, de onde se propaga até as extremidades do axônio. O sentido de propagação do impulso em um neurônio segue sempre a seqüência: dentrito \Rightarrow corpo celular \Rightarrow axônio.

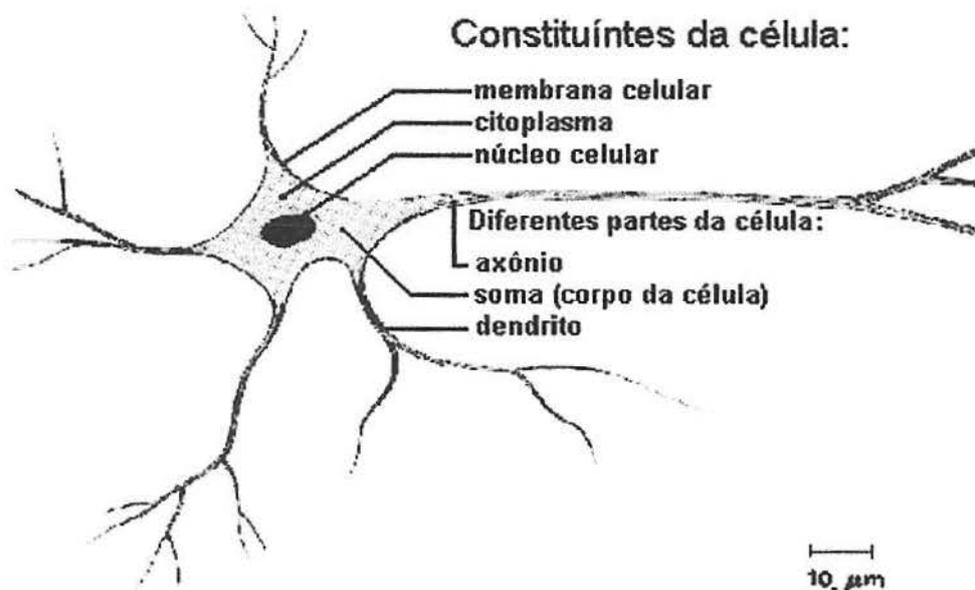


FIGURA 3.2: Os Constituintes de um Neurônio Natural

O impulso se propaga de um neurônio a outro através das sinapses nervosas. Entende-se por sinapse como a ligação estabelecida entre as terminações axônicas de um neurônio com os dendritos de outro. A análise de neurônios ao microscópio eletrônico mostra que, apesar de muito próximos, eles não se tocam na região da sinapse.

Pode-se dizer que as sinapses são regiões eletronicamente ativas, compreendidas entre duas membranas celulares: a membrana pré-sináptica, por onde chega um estímulo proveniente de outra célula, e a membrana pós-sináptica, que é a do dendrito. Nesta região inter-sináptica, o estímulo nervoso que chega a sinapse é transferido à membrana dendrital através de substâncias conhecidas como neurotransmissores. O resultado desta transferência é uma alteração do potencial elétrico da membrana pós-sináptica. Dependendo do tipo de neurotransmissor, a conexão sináptica será excitatória ou inibitória. Uma conexão excitatória provoca uma alteração no potencial da membrana que contribui para a formação de um impulso nervoso no axônio de saída enquanto que uma conexão inibitória age no sentido oposto (KOVÁCS, 1996).

Quando estimulado, um neurônio propaga o impulso nervoso, de natureza elétrica, ao longo do seu axônio. Ao chegar nas extremidades axônicas, o impulso determina a liberação de substâncias químicas chamadas mediadores sinápticos. As substâncias mediadoras, como a acetilcolina e a adrenalina, estimulam os dendritos do neurônio pós-sináptico de tal maneira que gerem nele o impulso nervoso. As terminações axônicas de muitos neurônios localizam-se nos músculos, onde também são liberados mediadores químicos que o estimulam, desencadeando a contração. Terminações nervosas de outros neurônios atingem certas glândulas, estimulando-as a liberarem sua secreção.

O neurônio biológico pode ser visto como o dispositivo computacional elementar do sistema nervoso, composto de muitas entradas e uma saída. As entradas são formadas através das conexões sinápticas que conectam os dendritos aos axônios de outras células nervosas. Os sinais que chegam por estes axônios são pulsos elétricos conhecidos como impulsos nervosos ou potenciais de ação e constituem a informação que o neurônio processa para produzir como saída um impulso nervoso no seu axônio (KOVÁCS, 1996).

A partir do conhecimento da estrutura e do comportamento dos neurônios naturais foram identificadas suas características fundamentais, utilizadas na criação de modelos de neurônios artificiais que simulam os reais. Estes neurônios artificiais são utilizados na formação das Redes Neurais Artificiais (OSÓRIO, 1991).

4. REDES NEURAI ARTIFICIAIS

Inteligência Artificial, Redes Neurais Artificiais, Neurocomputação são termos que, há pouco tempo, soavam nos meios acadêmicos como utopia.

O processo mais notável deu-se, em particular, na área de Redes Neurais Artificiais, que têm seus princípios fundamentais assentados nos trabalhos de modelagem biológica de processos neurofisiológicos, cognitivos e comportamentais.

As redes neurais artificiais consistem em um método de solucionar problemas de inteligência artificial, construindo um sistema que tenha circuitos que simulem o cérebro humano, inclusive seu comportamento, ou seja, aprendendo, errando e fazendo descobertas. São mais que isso, são técnicas computacionais que apresentam um modelo inspirado na estrutura neural de organismos inteligentes e que adquirem conhecimento através da experiência. Uma grande rede neural artificial pode ter centenas ou milhares de unidades de processamento, enquanto que o cérebro de um mamífero pode ter muitos bilhões de neurônios.

O campo de estudos com aplicação das Redes Neurais Artificiais mantém dependências disciplinares com as chamadas ciências cognitivas, a psicologia, a neurofisiologia, a matemática, a estatística e a física, em suas bases. Outros campos de conhecimento são arrolados à medida que os resultados das pesquisas se aplicam.

As Redes Neurais Artificiais foram criadas partindo de fundamentos biológicos para simular processos neurais em ambientes computacionais, de forma que estes pudessem adquirir propriedades tidas como especiais, tais como “aprender”. Porém, apesar de terem sido criadas partindo de fundamentos biológicos, não o são mais. Atualmente constituem-se de modelos matemático/estatísticos que pouco lembram os conhecidos neurônios dos livros de biologia.

A tecnologia de redes neurais assemelha-se ao processo de resolução de problemas do cérebro humano. Tal como humanos aplicam o conhecimento adquirido de experiências passadas para novos problemas ou situações, uma rede neural artificial utiliza exemplos resolvidos previamente para construir um sistema de nós que toma novas decisões, faz classificações e prognósticos.

Redes Neurais Artificiais buscam por padrões em treinamento realizado em conjuntos de dados, aprendem estes padrões, e desenvolvem a habilidade de classificar corretamente ou fazer prognósticos e predições. A rede neural supera-se em diagnóstico de problemas, tomadas de decisão, predição, e outros problemas de classificação onde o reconhecimento de padrões é importante e respostas computacionais precisas não são requeridas.

A aplicabilidade de redes neurais artificiais é extensa. Elas têm sido utilizadas na identificação de padrões e tendências nos dados, e são particularmente convenientes para previsão, incluindo, previsão de vendas, pesquisas de consumidores, validação de dados, análise de riscos, controle de processos industriais, dentre outras.

As redes neurais artificiais são constituídas por camadas de nós, que conectados, formam a rede (Figura 4.1). Estes nódulos são unidades de processamento unitárias que agem em paralelo. São análogas aos neurônios do cérebro humano, aceitam uma determinada informação de entrada e produzem uma saída (HAIR et al, 1998).

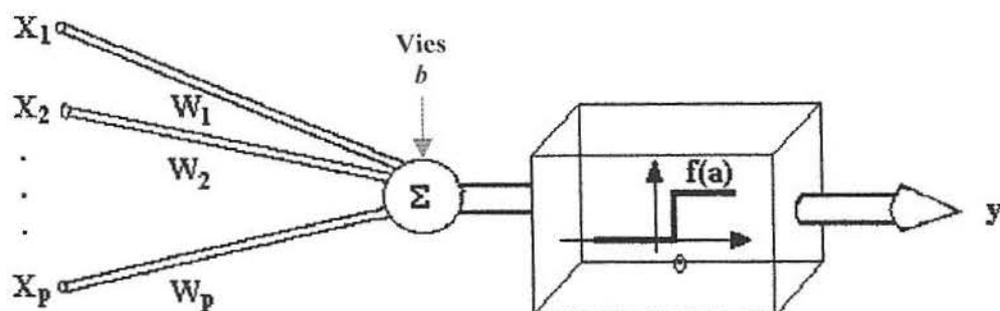


FIGURA 4.1: Representação gráfica de um nó com suas entradas e saída.

O nó aceita uma quantidade de informação de entrada x_1, x_2, x_3, \dots , pondera, respectivamente, de acordo com os pesos, w_1, w_2, w_3 . A função somatório, Σ , permite combinar todas as entradas em um único valor que representa o nível de ativação global das entradas, assim o potencial interno do nó é dado pela ponderação das entradas x_i com os pesos das respectivas conexões. A função de ativação, $f(a)$, transforma as entrada recebidas em uma informação de saída e a envia para outra unidade (nó) que a utiliza como informação de entrada.

A operação de uma unidade de processamento de uma rede neural artificial, proposta por McCULLOCH & PITTS (1943), pode ser resumida da seguinte maneira: sinais são apresentados à entrada; cada sinal é multiplicado por um peso que indica a sua influência na saída da unidade; é feita a soma ponderada dos sinais que produz um nível de atividade; e se este nível de atividade exceder um certo limite (threshold) a unidade produz uma determinada resposta de saída. O threshold é um nível determinado que as entradas ponderadas dos nós anteriores têm que exceder para que o nó seja ativado.

O peso sináptico de um neurônio artificial pode estar em um intervalo que inclui valores negativos bem como positivos.

O modelo neuronal apresentado na Figura 4.1 inclui também um vies aplicado externamente, representado por b . O vies b tem o efeito de aumentar ou diminuir a entrada líquida da função de ativação, dependendo de ser positivo ou negativo, respectivamente, e

serve para aumentar os graus de liberdade, permitindo uma melhor adaptação, por parte da rede neural, ao conhecimento à ela fornecido.

Em termos matemáticos, pode-se descrever um neurônio escrevendo-o pelas equações:

$$u = \sum_{j=1}^p w_j x_j, \quad v = f(u + b) \quad \text{e} \quad a = u + b$$

onde x_1, x_2, \dots, x_p são os sinais de entrada; w_1, w_2, \dots, w_p são os pesos sinápticos do neurônio; u é a saída do combinador linear (função somatório) devido aos sinais de entrada; b é o vies; $f(a)$ é a função de ativação; e v é o sinal de saída do neurônio.

A função de ativação $f(a)$ converte a soma de ativação das entradas de saída. Esta função que é um fator básico de diferenciação entre os nós, decide baseada no potencial interno do nó, qual o valor que deve ser enviado aos demais nós. Geralmente as funções de ativação são funções não-lineares. As duas funções de ativação mais comuns são a função threshold e a função sigmóide, apresentada respectivamente abaixo.

$$f(a) = \frac{1}{1 + e^{-a}} \quad \text{e} \quad f(a) = \tanh(a)$$

A função threshold é usada em situações onde as entradas e saídas são binárias. A função sigmóide é a função de ativação mais comum utilizada em modelagem de redes neurais.

As redes neurais artificiais são constituídas por uma camada de entrada, outra de saída, e camadas intermediárias, conhecidas como camadas de nós ocultos. A Figura 4.2 mostra a representação simplificada de uma rede neural com duas camadas de nós ocultos.

Os nós de entrada são responsáveis pela obtenção da informação inicial que será transmitida pela rede. Um nódulo de entrada representa uma única variável ou padrão. Quando as variáveis são quantitativas requerem somente um nó para cada uma delas, entretanto as variáveis qualitativas devem ser codificadas adequadamente, de tal forma que cada categoria seja representada por uma variável binária. Assim, uma variável codificada em

duas categorias seria representada por dois nós de entrada; uma categoria teria os valores (1,0), enquanto a outra (0,1).

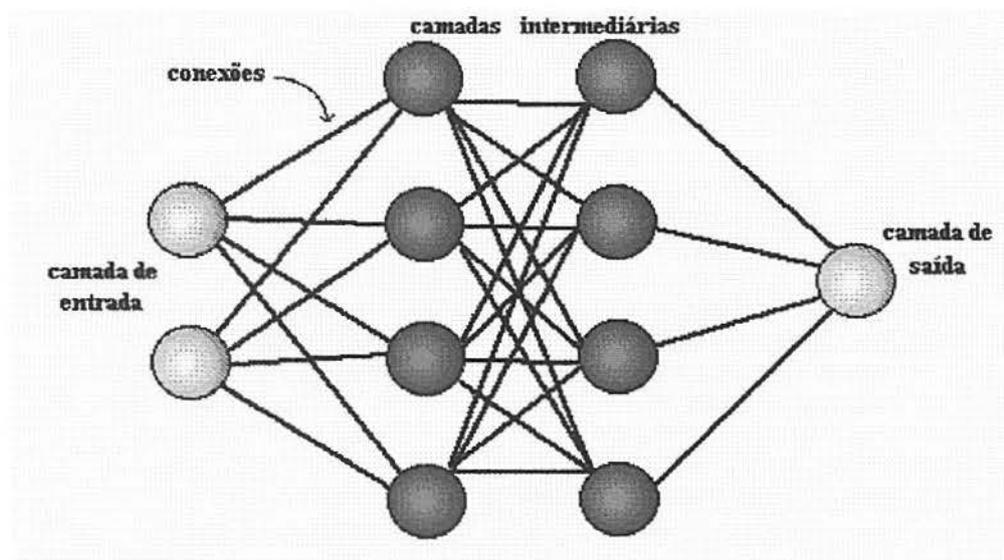


FIGURA 4.2: Representação gráfica de uma rede com duas camadas intermediárias.

Os nódulos de saída recebem entradas e produzem uma saída, mas não passam essa informação adiante para outro nódulo. Esta saída é o resultado final da rede. Se a rede é direcionada por um modelo preditivo, então esta saída é um valor predito. Se a rede é direcionada por um modelo de classificação, a saída é o valor final usado no processo de classificação.

A camada intermediária é utilizada pela rede na representação de relacionamentos mais complexos do que simplesmente relações um a um entre entrada e saída. Os nós ocultos, com suas funções de ativação, são responsáveis pela representação de relacionamento não-lineares.

A ação de cada nó em uma rede neural artificial é independente da ação de outro, mas as ações são paralelas, o que proporciona grande flexibilidade na modelagem (HAIR et al, 1998)

A generalização de uma rede se refere ao fato dela produzir saídas adequadas para entradas que não estavam presentes durante o treinamento, isto é, a aprendizagem. A capacidade de generalização e de aprendizagem possibilitam as redes neurais resolver

problemas complexos (de grande escala). Redes Neurais Artificiais apresentam as propriedades e capacidades descritas a seguir:

- Não linearidade: Uma rede composta por neurônios não-lineares é dita não-linear também. A não-linearidade é uma propriedade muito importante, particularmente se o mecanismo físico responsável pela geração do sinal de entrada for inerentemente não-linear.
- Mapeamento de Entrada-Saída: Cada exemplo da amostra de treinamento que é apresentado à rede consiste de um sinal de entrada único e de uma resposta desejada correspondente. Os exemplos de treinamento previamente aplicados podem ser reaplicados durante a sessão de treinamento, mas em ordem diferente. Assim, a rede aprende com os exemplos ao construir um mapeamento de entrada-saída para o problema considerado. Tal abordagem faz lembrar o estudo de inferência estatística não-paramétrica, que é o ramo da estatística que trata da estimação independente de modelo, ou seja, não são feitas suposições prévias sobre o modelo estatístico dos dados de entrada.
- Adaptabilidade: As redes neurais artificiais têm uma capacidade inata de adaptar seus pesos sinápticos a modificações externas. Em particular, uma rede treinada para operar em um ambiente específico pode ser facilmente retreinada para lidar com pequenas modificações operativas do ambiente. De modo geral, quanto mais adaptativo se fizer um sistema, assegurando-se de que o sistema se mantenha estável, mais robusto tenderá a ser o seu desempenho quando o sistema for exigido a operar em um ambiente não-estacionário.
- Resposta a Evidências: No contexto de classificação de padrões, uma rede neural pode ser projetada para fornecer informação não somente sobre qual padrão particular selecionar, mas também sobre a confiança ou crença na decisão tomada. Esta última informação pode ser utilizada para rejeitar padrões ambíguos e com isso melhorar o desempenho de classificação da rede.

- Informação Contextual: O conhecimento é representado pela própria estrutura e estado de ativação de uma rede neural. Cada neurônio da rede é potencialmente afetado pela atividade de todos os outros neurônios da rede. Consequentemente, a informação contextual é tratada naturalmente pela rede neural.
- Tolerância a Falhas: Uma rede neural, implementada em um hardware, tem o potencial de ser inerentemente tolerante a falhas, ou capaz de realizar computação robusta, no sentido de que seu desempenho se degrada suavemente sob condições de operações adversas. Para se assegurar que uma rede seja de fato tolerante a falhas pode ser necessário adotar-se medidas corretivas no projeto do algoritmo utilizado para treinar a rede (KERLIRZIN & VALLET, 1993).
- Implementação em VLSI: A natureza maciçamente paralela de uma rede neural a faz ser potencialmente rápida na computação de certas tarefas. Esta mesma característica torna uma rede neural adequada para implementação utilizando tecnologia de integração em escala muito ampla (very-large-scale-integration), sendo esta tecnologia que fornece um meio de capturar comportamentos realmente complexos de uma forma altamente hierarquizada. (MEAD, 1989).
- Uniformidade de Análise e Projeto: As redes neurais, basicamente, desfrutam de universalidade como processadores de informação, no sentido de que a mesma notação é utilizada em todos os domínios envolvendo a aplicação de redes neurais.
- Analogia Neurobiológica: O projeto de uma rede neural artificial é motivado pela analogia com o cérebro, que é uma prova viva de que o processamento paralelo tolerante a falhas é não somente possível fisicamente mas também rápido e poderoso.

4.1. ARQUITETURA DE UMA REDE

Um dos objetivos da pesquisa sobre redes neurais na computação é desenvolver morfologias neurais matemáticas, não necessariamente baseadas na biologia, que podem realizar funções diversas. Na maior parte dos casos, modelos neurais são compostos de muitos elementos não lineares que operam em paralelo e que são classificados de acordo com padrões ligados à biologia.

Quando um processo é criado visando utilizar aspectos de redes neurais começam com o desenvolvimento de um neurônio artificial baseado no entendimento de estruturas biológicas neurais, seguidas do aprendizado de mecanismos voltados para um determinado conjunto de aplicações. Ou em outras palavras, seguindo as três etapas:

- O desenvolvimento de modelos neurais motivado por neurônios biológicos;
- Modelos de estruturas e conexões sinápticas;
- O aprendizado das regras (método de ajuste de pesos ou forças de conexões internodais)

A maneira como os neurônios de uma rede estão estruturados está fortemente ligada com o algoritmo de aprendizagem usado para treinar a rede. Pode-se, portanto, falar de algoritmos de aprendizagem utilizados no projeto de redes neurais como sendo estruturados. Em geral, pode-se identificar classes de redes fundamentalmente diferentes:

1. Redes Alimentadas Adiante com Camada Única: Temos uma camada de entrada de nós de fonte que se projeta sobre uma camada de saída, mas não vice-versa, ou seja, esta rede é estritamente do tipo alimentada adiante ou acíclica. Ela é ilustrada na Figura 4.1.1 para o caso de quatro nós tanto na camada de entrada como na de saída. Esta rede é chamada de rede de camada única, sendo a designação “camada única” referente à camada de nós de saída. Não conta-se a camada de entrada de nós de fonte, porque lá não é realizada qualquer computação.

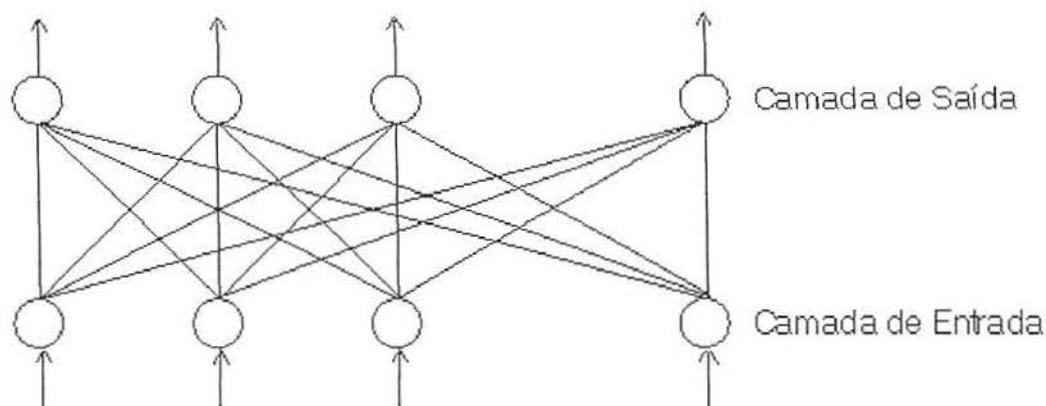


FIGURA 4.1.1: Rede alimentada adiante ou acíclica com uma única camada de nós.

2. Redes Alimentadas Diretamente com Múltiplas Camadas: Caracteriza-se pela presença de uma ou mais camadas ocultas (intermediárias), cujos nós são chamados de nós ocultos ou unidades ocultas. A função destes nós ocultos é de intervir entre a entrada externa e a saída da rede de maneira pertinente. Com a presença de uma ou mais camadas ocultas, a rede torna-se capaz de extrair estatísticas de ordem elevada. A habilidade dos nós ocultos extraírem estatísticas de ordem elevada é valiosa quando o tamanho da camada de entrada é grande. A Figura 4.1.2 ilustra a arquitetura de uma rede neural artificial de múltiplas camadas alimentada adiante para o caso de uma única camada oculta.

De acordo com RUMELHART & WEIGEND (1990), a rede neural deve possuir no mínimo duas camadas, a de entrada de dados e a da saída dos resultados. Como a rede apresenta desempenho muito limitado com somente duas camadas, a adição de uma camada intermediária faz-se necessária. Neste tipo de configuração, cada neurônio está ligado com todos os outros das camadas vizinhas, mas neurônios da mesma camada não se comunicam, além da comunicação ser unidirecional, apresentando assim um comportamento estático.

HECHT-NIELSEN (1990) afirma que com apenas uma camada oculta já é possível calcular uma função arbitrária qualquer a partir de dados fornecidos, defendendo, assim, que a camada oculta deve ter por volta de $2i+1$ neurônios, onde i é o número de variáveis de entrada.

Já LIPPMANN (1989) afirma que a segunda camada escondida deve ter o dobro de neurônios da camada de saída. No caso de apenas uma camada escondida ela deverá ter $s(i+1)$ neurônios, onde s é o número de neurônios de saída e i o número de neurônios na entrada.

Independente de cada abordagem, quanto mais camadas de neurônios, melhor é o desempenho da rede neural pois aumenta a capacidade de aprendizado, melhorando a precisão com que ela delimita regiões de decisão. Estas regiões de decisão são intervalos fixos onde a resposta pode estar.

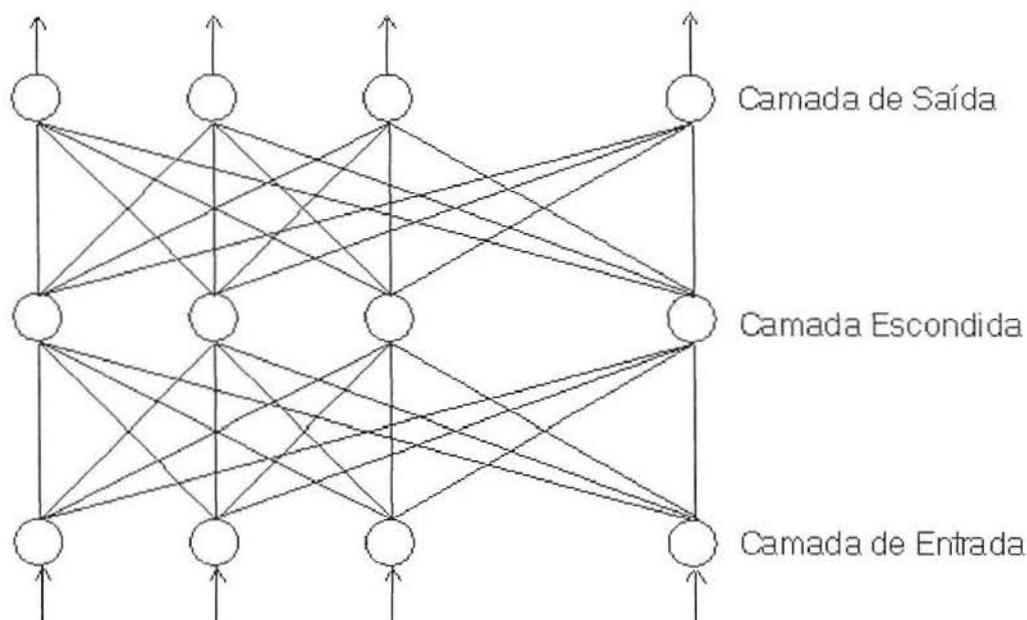


FIGURA 4.1.2: Rede alimentada adiante ou acíclica com uma camada escondida e uma camada de saída.

3. Redes Recorrentes: Distingui-se por ter pelo menos um laço de realimentação. Uma rede recorrente pode consistir, por exemplo, de uma camada de nós com cada nó alimentando seu sinal de saída de volta para as entradas de todos os outros nós, como é representado na Figura 4.1.3. As conexões de realimentação mostradas se originam dos nós ocultos bem como dos nós de saída. A presença de laços de realimentação tem um impacto profundo na capacidade de aprendizagem da rede e no seu desempenho. Este laços envolvem o uso de ramos particulares compostos de elementos de atraso unitário (representados por z^{-1}), o que

resulta em um comportamento dinâmico não-linear, admitindo-se que a rede neural contenha unidades não-lineares (HAYKIN, 2000).

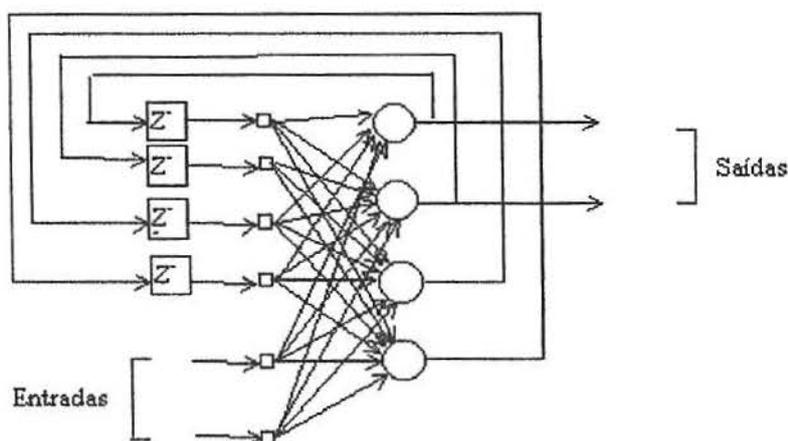


FIGURA 4.1.3: Rede recorrente com nós ocultos.

4.2. APRENDIZADO DA REDE

A propriedade mais importante das redes neurais é a habilidade de aprender de seu ambiente e com isso melhorar seu desempenho. Isso é feito através de um processo iterativo de ajustes aplicado a seus pesos, o treinamento. O aprendizado ocorre quando a rede neural atinge uma solução generalizada para uma classe de problemas.

O procedimento utilizado para realizar o processo de aprendizagem é chamado de algoritmo de aprendizagem, cuja função é modificar os pesos sinápticos e níveis de vies da rede de uma forma ordenada para alcançar um objetivo de projeto desejado. Existem muitos tipos de algoritmos de aprendizado específicos para determinados modelos de redes neurais, estes algoritmos diferem entre si principalmente pelo modo como os pesos são modificados.

Conforme MENDEL & McCLAREN (1970), aprendizagem é um processo pelo qual os parâmetros de uma rede neural são adaptados através de um processo de estimulação pelo ambiente no qual a rede está inserida. O tipo de aprendizagem é determinado pela maneira pela qual a modificação dos parâmetros ocorre.

Esta definição do processo de aprendizagem implica a seguinte seqüência de eventos:

1. A rede é estimulada por um ambiente.
2. A rede sofre modificações nos seus parâmetros livres como resultado desta estimulação.
3. A rede responde de uma maneira nova ao ambiente, devido às modificações ocorridas na sua estrutura interna.

Pode-se denominar ainda o ciclo de aprendizado da rede como sendo uma apresentação de todos os N pares (entrada e saída) do conjunto de treinamento no processo de aprendizado. A correção dos pesos num ciclo pode ser executado de dois modos:

1. Modo Padrão: A correção dos pesos acontece a cada apresentação à rede de um exemplo do conjunto de treinamento. Cada correção de pesos baseia-se somente no erro do exemplo apresentado naquela iteração. Assim, em cada ciclo ocorrem N correções.
2. Modo Batch: Apenas uma correção é feita por ciclo. Todos os exemplos do conjunto de treinamento são apresentados à rede, seu erro médio é calculado e a partir deste erro fazem-se as correções dos pesos.

Outro fator importante é a maneira pela qual uma rede neural se relaciona com o ambiente. Nesse contexto existem os seguintes paradigmas de aprendizado:

1. Aprendizagem supervisionada: utiliza um agente externo que indica à rede um comportamento bom ou ruim de acordo com o padrão de entrada. Constrõem modelos, os quais classificam padrões ou executam predições de acordo com outros padrões de "entradas" e "saídas" que eles aprenderam. Eles dão a resposta mais razoável baseada em uma variedade de padrões de aprendizado. Em uma rede supervisionada, você mostra à rede como fazer predições, classificações, ou decisões, fornecendo a ela um grande número de classificações corretas ou predições das quais ela pode aprender.

2. Aprendizagem não-supervisionada (auto-organização): não utiliza um agente externo indicando a resposta desejada para os padrões de entrada, utiliza-se entretanto, exemplos de coisas semelhantes para que a rede responda de maneira semelhante. Podem

classificar um conjunto de padrões de treinamento em um número especificado de categorias sem ser mostrado antes como categorizar. A rede faz isto agrupando os padrões. Ela agrupa-os por sua proximidade em um espaço dimensional N onde o N é o número de "entradas". O usuário diz à rede o número máximo de categorias e ela geralmente agrupa os dados em um certo número de categorias. Entretanto, ocasionalmente a rede pode não ser capaz de separar os padrões naquelas muitas categorias opcionais.

A rede neural se baseia nos dados para extrair um modelo geral. Portanto, a fase de aprendizado deve ser rigorosa e verdadeira, a fim de se evitar modelos espúrios. Todo o conhecimento de uma rede neural está armazenado nas sinapses, ou seja, nos pesos atribuídos às conexões entre os neurônios. De 50 a 90% do total de dados devem ser separados para o treinamento da rede neural, dados estes escolhidos aleatoriamente, a fim de que a rede aprenda as regras e não adicione nenhuma tendência no seu resultado final. O restante dos dados só é apresentado à rede neural na fase de testes a fim de que ela possa "deduzir" corretamente o inter-relacionamento entre os dados.

Pode-se dividir os processos de aprendizagem em cinco regras básicas de aprendizagem: aprendizagem por correção de erro, aprendizagem baseada em memória, aprendizagem hebbiana, aprendizagem competitiva e aprendizagem de boltzmann. Estas regras permitem ir muito além do alcançável por filtros adaptativos lineares, tanto em termos de capacidade como em universalidade.

No estudo de aprendizagem supervisionada, uma condição fundamental é a existência de um especialista capaz de fornecer correções exatas para as saídas da rede quando um erro ocorrer, como na aprendizagem por correção de erro; ou de fixar em relação ao ambiente as unidades de entrada e de saída livres da rede, como na aprendizagem de Boltzmann. Nenhum destes modelos é possível em organismos biológicos, que não possuem as conexões nervosas recíprocas exatas necessárias para a retropropagação das correções de erro (em uma rede de múltiplas camadas alimentada adiante), nem os meios nervosos para imposição de

comportamento pelo mundo exterior. Apesar disso, a aprendizagem supervisionada estabeleceu-se como um paradigma poderoso para o projeto de redes neurais artificiais.

Por outro lado, regras de aprendizagem não-supervisionada, tais como a aprendizagem hebbiana e a aprendizagem competitiva, são motivadas por considerações neurobiológicas. Entretanto, para aperfeiçoar o entendimento sobre a aprendizagem não-supervisionada, é necessário buscar idéias relevantes na teoria da informação de Shannon. O princípio da máxima informação mútua (Infomax) de LINSKER (1988) fornece o formalismo matemático para o processamento da informação em uma rede neural artificial não-supervisionada, de uma forma até certo ponto análoga à transmissão de informação em um canal de comunicação.

A discussão sobre os métodos de aprendizagem não seria completa sem mencionar o modelo de aprendizagem seletiva darwiniano (EDELMAN, 1987; REEKE et al., 1990). Este modelo é baseado na teoria de seleção de grupo mental, e pressupõe que o sistema nervoso opera por uma forma de seleção similar a seleção natural evolutiva, mas que isto acontece dentro do cérebro durante o período de vida de cada animal. De acordo com esta teoria, as unidades operacionais básicas não são os neurônios isolados, mas sim grupos localizados de células fortemente interligadas. A pertinência de neurônios em um grupo é modificada pela alteração dos pesos sinápticos dos neurônios. A competição local e a cooperação entre células são claramente necessárias para produzir ordenação local na rede. Uma coleção de grupos neuronais é denominada repertório. Grupos em um repertório respondem melhor a padrões de entrada superpostos mas similares, devido à natureza aleatória do crescimento neural. Um ou mais grupos neuronais em um repertório respondem a todo padrão de entrada, assegurando assim que haja uma resposta a padrões de entrada não-esperados, que podem ser importantes. A aprendizagem seletiva darwiniana é diferente dos algoritmos de aprendizagem normalmente utilizados no projeto de redes neurais artificiais, porque ela assume que, por

projeto, haja muitas subredes e que somente aqueles com resposta desejada são selecionadas durante o processo de treinamento.

Nenhum tipo de rede é garantida de sempre dar uma resposta absolutamente "correta", especialmente se os padrões são de alguma forma incompletos ou conflitantes.

Os resultados devem ser avaliados em termos da percentagem de respostas corretas que resultaram do modelo. Alguns problemas são bem apropriados para a capacidade de reconhecimento de padrões de uma rede neural e outros são melhor resolvidos com métodos estatísticos tradicionais.

4.2.1. TEORIA ESTATÍSTICA DO PROCESSO DE APRENDIZAGEM

O interesse principal aqui é expressar em termos estatísticos o desvio entre a função "alvo" $f(x)$ e a função "real" $F(\mathbf{x}, \mathbf{w})$, realizada pela rede neural, onde o vetor \mathbf{x} representa o sinal de entrada e \mathbf{w} é o vetor de pesos sinápticos.

Dado uma amostra de treinamento τ constituída de N realizações do vetor aleatório \mathbf{X} constituindo de um conjunto de variáveis independentes representadas por $\{\mathbf{x}_i\}_{i=1}^N$ e por N realizações do escalar aleatório D que representa uma variável dependente representadas por $\{d_i\}_{i=1}^N$. Através do modelo proposto por WHITE (1989), a relação funcional entre \mathbf{X} e D pode ser a descrita a seguir:

$$D = f(\mathbf{X}) + \varepsilon$$

onde $f(\cdot)$ é uma função determinística de seu argumento vetorial, e ε é um erro de predição aleatório que representa o desconhecimento sobre a dependência de D e \mathbf{X} . O modelo descrito acima é o modelo estatístico de regressão e está representado na Figura 4.2.1.1a.

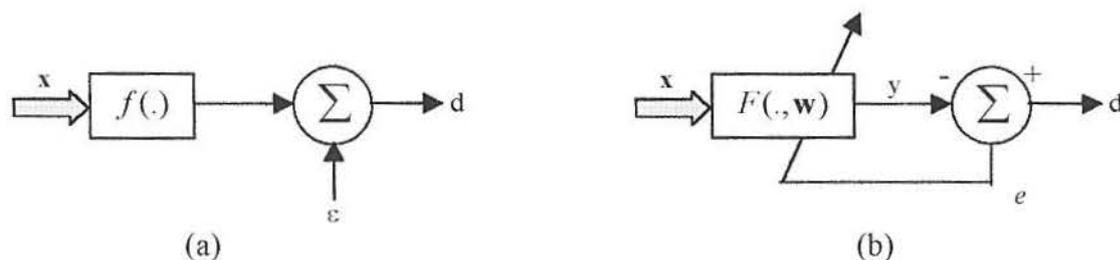


FIGURA 4.2.1.1: (a) Modelo estatístico de regressão. (b) Modelo (físico) de rede neural.

O modelo de regressão da Figura 4.2.1.1a é uma descrição estatística de um ambiente estocástico. O seu propósito é utilizar o vetor \mathbf{X} para explicar ou prever a variável dependente D . A Figura 4.2.1.1b é o modelo “físico” correspondente do ambiente. O propósito deste segundo modelo, baseado em uma rede neural, é codificar o conhecimento empírico, representado pela amostra de treinamento τ em um conjunto correspondente de vetores de pesos sinápticos \mathbf{w} . Na verdade, a rede neural fornece uma aproximação para o modelo de regressão da Figura 4.2.1.1a.

A caracterização estatística das redes neurais artificiais pode ser feita também através de uma teoria de aprendizagem que trata da questão fundamental de como controlar a habilidade de generalização de uma rede neural em termos estatísticos. A discussão é apresentada no contexto da aprendizagem supervisionada.

Um modelo de aprendizagem supervisionada consiste de três componentes interrelacionados, ilustrados na Figura 4.2.1.2 e descritos em termos estatísticos como segue (VAPNIK, 1992, 1998):

1. Ambiente: O ambiente é estacionário, fornecendo um vetor \mathbf{x} com uma função de distribuição de probabilidade cumulativa fixa, mas desconhecida $F_{\mathbf{x}}(\mathbf{x})$.
2. Especialista: O especialista fornece uma resposta desejada d para cada vetor de entrada \mathbf{x} recebido do ambiente, de acordo com uma função de distribuição cumulativa condicional $F_{\mathbf{x}}(\mathbf{x}|d)$ que é também fixa mas desconhecida. A resposta desejada d e o vetor de entrada \mathbf{x} estão relacionados por $d = f(\mathbf{x}, v)$, onde v é um termo de ruído.

3. Máquina (algoritmo) de aprendizagem: A máquina de aprendizagem (rede neural) é capaz de implementar um conjunto de funções de mapeamento de entrada-saída descritas por $y = F(\mathbf{x}, \mathbf{w})$, onde y é a resposta real produzida pela máquina de aprendizagem em resposta à entrada \mathbf{x} , e \mathbf{w} é um conjunto de parâmetros livres (pesos sinápticos).

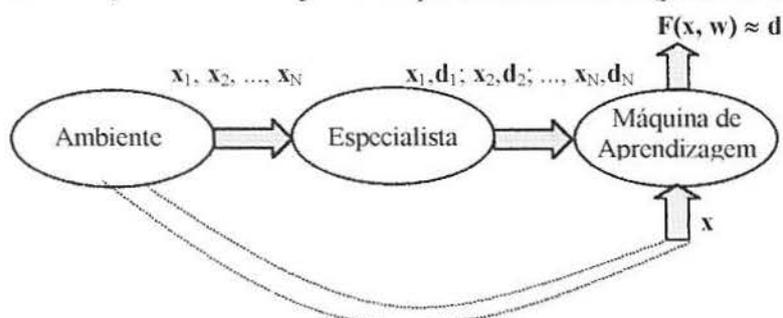


FIGURA 4.2.1.2: Modelo do processo de aprendizagem supervisionada.

O problema da aprendizagem supervisionada é selecionar a função particular $F(\mathbf{x}, \mathbf{w})$ que aproxima a resposta desejada d de uma forma ótima, sendo “ótimo” definido em um sentido estatístico. A própria seleção é baseada no conjunto dos N exemplos de treinamento independentes e identicamente distribuídos (iid).

Vale também citar o modelo de aprendizagem provavelmente aproximadamente correto (PAC) que é creditado a VALIANT (1984). Como o nome implica, o modelo PAC é uma estrutura probabilística para o estudo de aprendizagem e generalização em sistemas de classificação binária. Ele está intimamente relacionado à aprendizagem supervisionada.

Resumidamente, o modelo consiste que desde que o tamanho da amostra de treinamento seja suficientemente grande, após a rede neural ter sido treinada com o conjunto de dados, é “provável” que o mapeamento de entrada-saída calculado pela rede seja “aproximadamente correto”.

Na teoria de aprendizagem PAC, uma questão de particular interesse com implicações práticas é a questão da complexidade da amostra. O enfoque nesta questão está sobre quantos exemplos aleatórios devem ser apresentados ao algoritmo de aprendizagem para que ele adquira informação suficiente para aprender um conceito-alvo desconhecido.

Para proceder o processo de aprendizagem de uma rede neural através do algoritmo de retropropagação proposto por HERTZ et al (1991) é necessário o uso de uma função objetivo ou uma medida de performance. O objetivo é utilizar a função objetivo para otimizar os pesos. A medida de performance mais usada em redes neurais é a soma de quadrados dos resíduos dada por:

$$E = \frac{1}{2} \sum_{p=1}^n \sum_{k=1}^O (y_{pk} - \hat{y}_{pk})^2$$

onde p representa as observações num total de n observações, k é o nó de saída com um total de O nós de saída, y é a resposta observada da variável alvo, e \hat{y} é a resposta predita pelo modelo. Esta é, então, a soma dos quadrados das diferenças entre a resposta predita e a resposta observada de todas as observações. No caso da predição de um único nó de saída ($k=1$), tem-se

$$E = \frac{1}{2} \sum_{p=1}^n (y_p - \hat{y}_p)^2$$

que é a usual função para minimização por mínimos quadrados em análise de regressão.

O processo de aprendizagem inicia com os valores de entrada sendo apresentados a camada de entrada. Os nós de entrada não executam nenhuma operação com esta informação, mas simplesmente passam esta informação para os nós da camada oculta. Como já foi mostrado anteriormente, a estrutura do cálculo feito pelo j -ésimo nó da camada oculta é dado por:

$$u_{pj} = \sum_{i=1}^N w_{ji} x_{pi}$$

onde N é o número total de nós de entrada, w_{ji} é o peso do nó de entrada i para o nó oculto j , e x_{pi} é o valor do i -ésimo nó de entrada para a observação p . O j -ésimo nó escondido aplica uma função de ativação para suas entradas e saídas (neste caso a função sigmóide):

$$v_{pj} = f(u_{pj}) = \frac{1}{1 + e^{-u_{pj}}}$$

Similarmente, o nó de saída k recebe um resultado de entrada de

$$f_{pk} = \sum_{j=1}^M W_{kj} v_{pj}$$

onde M é o número de nós ocultos, e W_{kj} representa o peso do nó oculto j para o nó de saída k .

O nó de saída, então, apresenta o seguinte resultado como resposta predita.

$$\hat{y}_{pk} = f(f_{pk}) = \frac{1}{1 + e^{-f_{pk}}}$$

O valor inicial não foi considerado na equação. Isso porque o valor inicial pode ser calculado adicionando um nó extra e fixando seu valor em 1. Isto é similar a adição de uma coluna de uns na matriz de planeamento nos problemas de regressão para o cálculo do intercepto.

Vale lembrar que o objetivo é encontrar o valor de w_{ji} , os pesos das conexões dos nós de entrada aos nós ocultos, e W_{jk} , os pesos das conexões dos nós ocultos aos nós de saída que minimize a função objetivo que é a soma de quadrados dos resíduos. Assim a derivada parcial da função objetivo em relação ao peso W_{jk} representa a taxa de mudança da função objetivo em relação ao peso, isto é a inclinação da função objetivo. Movimentando os pesos na direção descendente da inclinação representará um decréscimo na função objetivo. Isto sugere um método para iterativamente encontrar valores para os pesos. Avalia-se a derivada parcial da função objetivo em relação aos pesos e então move-se os pesos em direção descendente da inclinação, continuando até que a função erro não mais decresça. Matematicamente, isto é representado como

$$\Delta W_{kj} = -\eta \frac{\partial E}{\partial W_{kj}}$$

onde o termo η é conhecido como a taxa de aprendizagem. A prática comum em redes neurais é usar um valor fixo para a taxa de aprendizagem no início do problema.

Primeiramente, deriva-se uma expressão para calcular o ajustamento dos pesos dos nós ocultos para os nós de saída W_{kj} . Fazendo as devidas substituições na função objetivo, obtém-se o seguinte resultado

$$E = \frac{1}{2} \sum_{p=1}^n \sum_{k=1}^O \left(y_{pk} - f \left(\sum_{j=1}^M W_{kj} \cdot f \left(\sum_{i=1}^N w_{ji} x_{pi} \right) \right) \right)^2$$

e, então, calculando a derivada parcial tem-se que

$$-\eta \frac{\partial E}{\partial W_{kj}} = -\eta \frac{\partial E}{\partial \hat{y}_{pk}} \frac{\partial \hat{y}_{pk}}{\partial f_{pk}} \frac{\partial f_{pk}}{\partial W_{kj}}, \quad \text{mas} \quad \frac{\partial E}{\partial W_{kj}} = -(y_{pk} - \hat{y}_{pk})$$

e $\frac{\partial \hat{y}_{pk}}{\partial f_{pk}} = f'(f_{pk}) = \hat{y}_{pk}(1 - \hat{y}_{pk})$ pela função sigmóide e $\frac{\partial f_{pk}}{\partial W_{kj}} = v_{pj}$

Substituindo estes resultados na expressão de ΔW_{kj} , tem-se que a mudança nos nós ocultos para os nós de saída é

$$\Delta W_{kj} = -\eta [(-1)(y_{pk} - \hat{y}_{pk})] \hat{y}_{pk} (1 - \hat{y}_{pk}) v_{pj}$$

Com isso chega-se a fórmula para adaptação dos pesos da rede (dos nós da camada oculta para os nós de saída) dada por

$$W_{kj}^{t+1} = W_{kj}^t + \Delta W_{kj}$$

Utiliza-se o peso atual da rede W_{kj}^t adicionado por um peso ajustado ΔW_{kj} para obter o peso estimado W_{kj}^{t+1} do passo seguinte. Para adaptar os pesos w_{ji} , utiliza-se o mesmo procedimento com as devidas modificações. O processo de adaptação dos pesos é feito até que se atinja o mínimo para a função objetivo. (WARNER & MISRA, 1996)

Dado as equações acima utiliza-se, então, os passos descritos a seguir que são necessários para calcular a mudança nos pesos da rede usando a aprendizagem de retropropagação. Este algoritmo é adaptado de HERTZ et al. (1991).

1. Inicialmente atribua os pesos aleatoriamente. Estes pesos iniciais de cada unidade podem estar por volta de 0,5.
2. Escolha um exemplo (observação) p e propague-o para frente. Isto produz os valores de v_{pj} e \hat{y}_{pk} , as saídas para as camadas escondida e de saída.
3. Compute os resíduos: $\delta_{pk} = (y_{pk} - \hat{y}_{pk})f'(f_{pk})$.
4. Compute os resíduos para a camada escondida: $\psi_{pj} = \sum_{k=1}^O \delta_{pk} W_{kj} v_{pj} (1 - v_{pj})$.
5. Calcule $\Delta W_{kj} = \eta \delta_{pk} v_{pj}$ e $\Delta w_{ji} = \eta \psi_{pj} i_{pi}$ para adaptar os pesos.
6. Repita os passos para cada exemplo (observação).

É fácil ver como isto poderia ser implementado em um programa de computador.

4.3. INTELIGÊNCIA ARTIFICIAL E NEUROCOMPUTAÇÃO

O objetivo da inteligência artificial (IA) é desenvolver paradigmas ou algoritmos que requeiram máquinas para realizar tarefas cognitivas, para as quais os humanos são altamente melhores. Um sistema de IA deve ser capaz de: (1) armazenar conhecimento, (2) aplicar o conhecimento armazenado para resolver problemas e (3) adquirir novo conhecimento através da experiência. Um sistema de IA tem três componentes fundamentais: representação, raciocínio e aprendizagem (SAGE, 1990), como é representado na Figura 4.3.1.

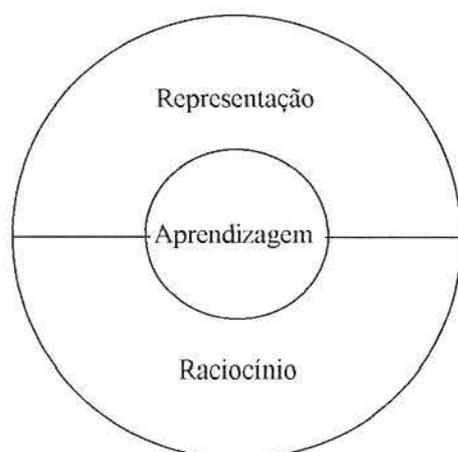


FIGURA 4.3.1: Ilustração dos três componentes principais de um sistema de IA.

1. Representação: A característica mais distintiva da IA provavelmente seja o uso difundido de uma linguagem de estruturas simbólicas para representar tanto o conhecimento genérico sobre um domínio do problema de interesse como o conhecimento específico sobre a solução do problema.
2. Raciocínio: De forma simplificada, raciocínio é a habilidade de resolver problemas. Para um sistema ser qualificado como um sistema de raciocínio ele deve ser capaz de expressar e resolver uma vasta gama de problemas, ser capaz de tornar conhecidas para ele tanto a informação explícita como a informação implícita e o sistema deve ter, também, um mecanismo de controle que determine quais operações devem ser aplicadas para um problema particular, quando uma solução para este problema foi obtida, ou quando deve ser encerrado o tratamento deste problema (FISCHLER & FIRSCHEIN, 1987).

3. Aprendizagem: Na Figura 4.3.2 é representada o modelo simples de aprendizagem de máquina. O ambiente fornece alguma informação para um elemento de aprendizagem que a utiliza para aperfeiçoar a base de conhecimento, e finalmente o elemento de desempenho utiliza a base de conhecimento para executar a sua tarefa.

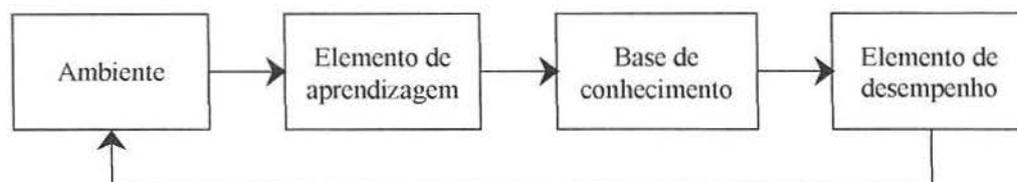


FIGURA 4.3.2: Modelo simples de aprendizagem de máquina

Uma comparação entre IA e Redes Neurais Artificiais pode ser feita sobre os aspectos de nível de explanação, estilo de processamento e estrutura representativa.

Pode-se descrever a IA como a manipulação formal de uma linguagem de algoritmos e representações de dados em uma forma de cima para baixo (top-down). Por outro lado, as redes neurais podem ser descritas como processadores distribuídos paralelamente com uma habilidade natural de aprender e que normalmente operam de uma forma de baixo para cima (bottom-up). Portanto, torna-se evidente que, para a implementação de tarefas cognitivas, melhor que procurar soluções baseadas em IA ou em redes neurais isoladamente, uma abordagem potencialmente mais vantajosa seria construir modelos conexionistas estruturados ou sistemas híbridos que interajam ambas as abordagens. Fazendo isso, obtém-se a combinação das características desejáveis de adaptabilidade, robustez e uniformidade oferecidas pelas redes neurais com a representação, inferência e universalidade, que são características inerentes da IA (FELDMAN, 1992; WALTZ, 1997).

A neurocomputação é uma ciência ligada ao desenvolvimento e estudo das Redes Neurais Artificiais, que tem como base fundamental o processamento distribuído, adaptativo e não linear. As redes neurais executam processos de forma bem diferente dos computadores convencionais, que possuem uma única unidade central de processamento resolvendo seqüencialmente todas as partes de um problema. As redes neurais são constituídas por um

significativo número de elementos de processamento extremamente simples, os quais individualmente resolvem parte de um grande problema.

Os modelos neurais, procuram aproximar o processamento dos computadores ao cérebro. As redes neurais possuem um grau de interconexão similar a estrutura do cérebro e em um computador convencional moderno a informação é transferida em tempos específicos dentro de um relacionamento com um sinal para sincronização.

Pode ser traçado um comparativo com o computador digital e com um neurocomputador como é apresentado na Tabela 4.3.1.

TABELA 4.3.1: Quadro comparativo entre computadores e neurocomputadores

Computadores	Neurocomputadores
- Executa programas	- Aprende
- Executa operações lógicas	- Executa operações não lógicas, transformações, comparações
- Depende do modelo ou do programador	- Descobre as relações ou regras dos dados e exemplos
- Testa uma hipótese por vez	- Testa todas as possibilidades em paralelo

4.4. REDES NATURAIS x REDES ARTIFICIAIS

Ao contrário das redes neurais artificiais, as 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. Seus pulsos não são síncronos ou assíncronos, devido ao fato de não serem contínuos, o que a difere de redes artificiais.

As diferenças principais entre os computadores (Redes Artificiais) e o cérebro humano (Redes Naturais) são:

Velocidade de Processamento:

Os processadores digitais tem velocidade de centenas de Mhz, levando algumas dezenas de manosegundos para executar uma instrução, enquanto que os neurônios levam pelo menos 4 milisegundos para completar um ciclo de disparo. Sendo assim, um computador pode ter componentes que são 10^5 vezes mais rápidos que um neurônio.

Modo de Processamento:

Uma possível explicação para o cérebro ter um desempenho melhor que um computador em certas tarefas tais como visão é que, ao invés de executar um programa muito grande de modo serial como os computadores fazem, o cérebro executa em paralelo um número grande de pequenos programas.

Número e Complexidade dos Processadores:

Em um computador as instruções dos programas são realizadas em poucos processadores de elevada complexidade. No cérebro humano, um número muito maior (10^{11}) de processadores relativamente simples e de baixa resolução (neurônios) trabalham com alta interconectividade (cada neurônio pode estar conectado a até cerca de 10^4 outros neurônios).

Armazenamento de Conhecimento:

Um determinado item de informação é armazenado em uma posição de memória específica em um computador. Este tipo de memória é chamado de memória localizada, pois uma unidade de memória contém um item inteiro de informação. Por outro lado, entende-se que no cérebro a informação está distribuída nas sinapses, de tal forma que nenhuma sinapse contém um item inteiro e cada sinapse pode contribuir para a representação de vários itens de informação. Este tipo de memória é chamado de memória distribuída. Além disso o computador utiliza memória endereçada por endereço, enquanto o cérebro utiliza memória

endereçada por conteúdo, ou seja, um memória é lembrada usando-se partes de seu conteúdo como as chaves de endereçamento.

Controle do Processamento:

Em um computador existe um sinal de relógio que é usado para sincronizar todos os componentes, enquanto que no cérebro não existe uma área responsável pelo controle e sincronização de todos os neurônios.

A semelhança da rede natural com a rede natural é muito grande, pois ambas possuem disparos de saídas, podendo receber várias entradas, e direcionando em uma única saída.

4.5. ALGUMAS REDES MAIS UTILIZADAS

4.5.1. REDE PERCEPTRON

Desenvolvido por ROSENBLATT (1958), em prosseguimento às idéias de McCulloch, o Modelo de Rede Neural Perceptron é uma rede genuína de múltiplos neurônios do tipo discriminadores lineares. Em sua origem era uma simulação computacional para a retina, que demonstrou como o sistema nervoso visual reconhece padrões. Um Perceptron é uma rede com uma topologia representada da Figura 4.5.1.1, com os neurônios dispostos em várias camadas. Tipicamente, a rede consiste de um conjunto de nós que constituem a camada de entrada, uma ou mais camadas ocultas (escondidas) e uma camada de saída. O sinal de entrada se propaga para frente através da rede, camada por camada. Estas redes são normalmente chamadas de perceptrons de múltiplas camadas (MLP, multilayer perceptron).

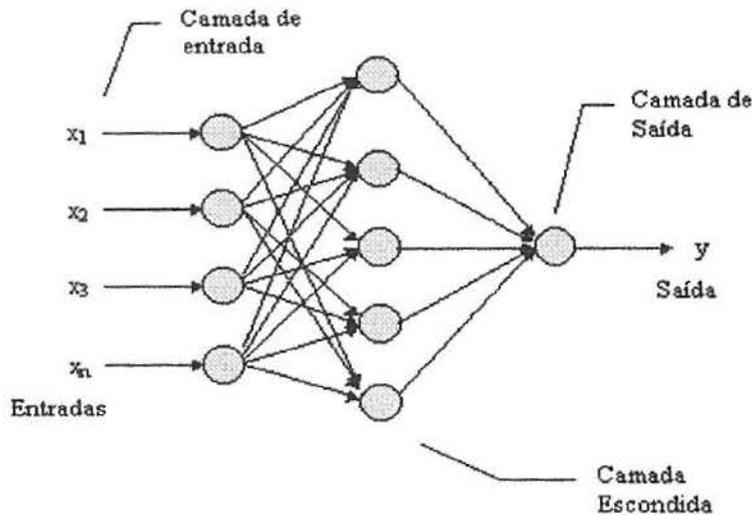


FIGURA 4.5.1.1: Arquitetura de um perceptron com uma camada escondida.

O treinamento desta rede é feito de forma supervisionada com um algoritmo conhecido como algoritmo de retropropagação. Este algoritmo é baseado na regra de aprendizagem por correção de erro. A aprendizagem por retropropagação de erro consiste de dois passos através das diferentes camadas da rede: um passo para frente, a propagação, e um passo para trás, a retropropagação. O funcionamento deste algoritmo pode ser visualizado na Figura 4.5.1.2. No passo para frente, um padrão de atividade (vetor de entrada) é aplicado aos nós da rede e seu efeito se propaga através da rede.

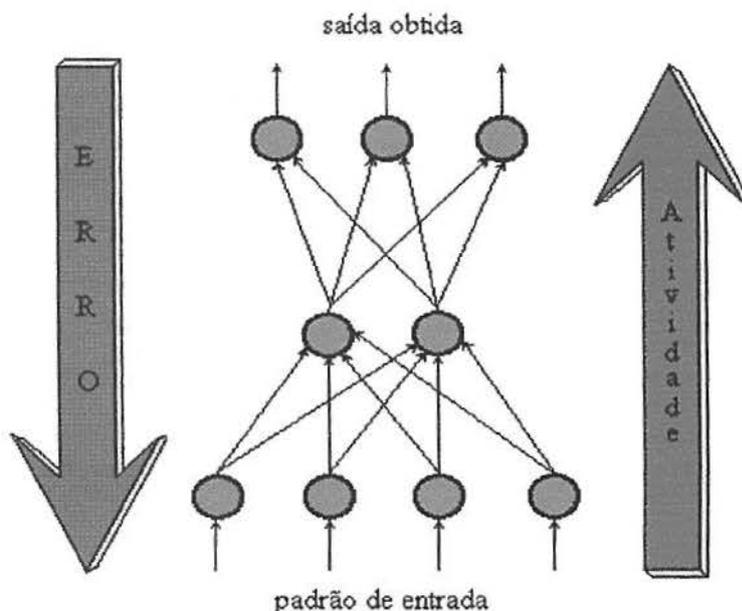


FIGURA 4.5.1.2: Esquema do algoritmo de aprendizagem de retropropagação.

Então, um conjunto de saídas é produzido como a resposta real da rede. Durante o passo de propagação, os pesos sinápticos da rede são todos fixos. Durante a retropropagação, por outro lado, os pesos sinápticos são todos ajustados de acordo com uma regra de correção de erro. Resumidamente, a resposta real da rede é subtraída da resposta alvo e produz um sinal de erro que é propagado para trás pela rede, contra a direção das conexões sinápticas. Os pesos sinápticos são ajustados de modo que a resposta real da rede se torne mais próxima da resposta desejada, em um sentido estatístico.

Em algumas redes muito complexas pode-se obter superfícies de erro com a existência de múltiplos pontos ótimos, pontos de sela e regiões planas. A Figura 4.5.1.3 ilustra um processo de otimização, onde várias soluções locais foram encontradas. (DINIZ & LOUZADA-NETO, 2000)

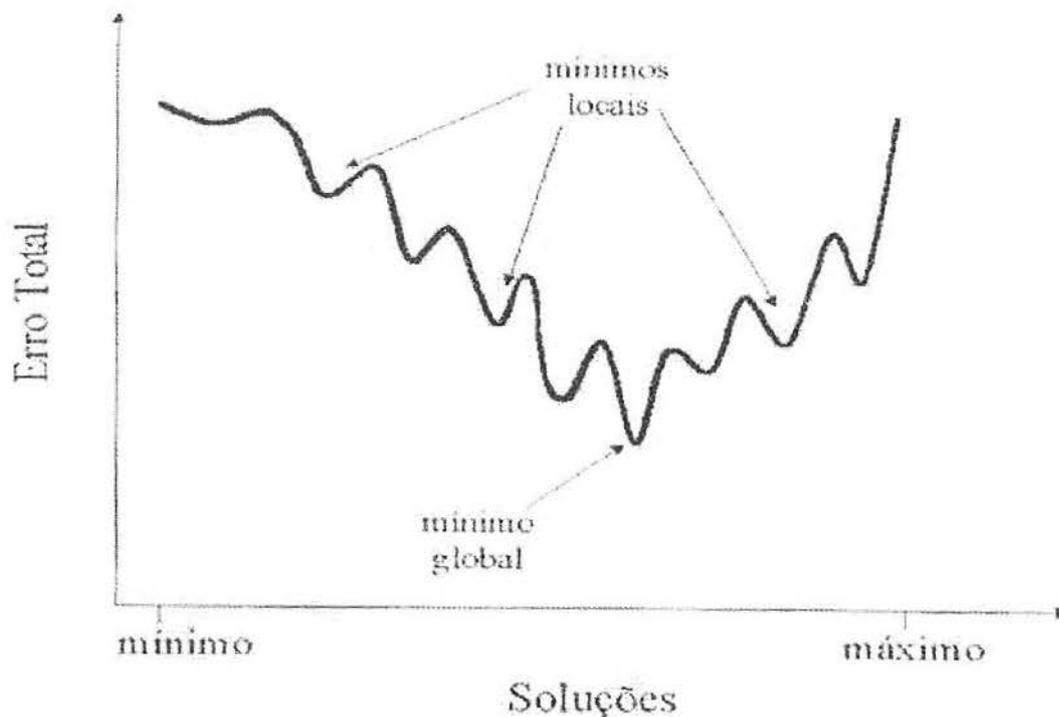


FIGURA 4.5.1.3: Representação dos resultados de um processo de otimização com várias soluções locais.

4.5.2. REDE HOPFIELD

Uma das maiores contribuições para área de redes neurais artificiais foi feita no início dos anos sessenta por John Hopfield. A contribuição de Hopfield não foi simplesmente a sugestão de um modelo adequado, mas a sua análise e estudo extensivo, o que fez seu nome ser associado com a rede. Ele desenvolveu o uso de uma função de energia, e relacionou as redes a outros sistemas físicos. A rede Hopfield consiste de um número de módulos, um conectado ao outro: é uma rede completamente conectada, isto é, há uma conexão das entradas com as saídas e por este motivo, estas redes dificilmente chegam a um estado instável, ou seja, chegará um momento em que a saída, após oscilar entre alguns valores binários, será sempre a mesma para um mesmo padrão de entrada.

É caracterizada por ser do tipo binária, de auto-aprendizado e com retro-alimentação. É baseada no conceito de energia da rede, onde esta é proporcional à troca de estados da rede. Isto é, quanto maior for a oscilação entre um estado e outro na saída, maior será a energia. Portanto, a fase de reconhecimento tem por objetivo minimizar a quantidade de energia de maneira que a rede convirja para um único estado na saída

A rede de Hopfield pode ser comparada a um modelo físico, onde a cada troca de estado da rede a energia da mesma diminui, portanto, a fase de aprendizado chegará ao fim no momento em que a rede tiver a sua energia minimizada.

4.5.3. REDES ADALINE E MADALINE

O modelo Adaline, criado por Widrow e Hoff, é composto de elementos processadores “adaline” (acrônimo de ADaptive LInear NEuron), similar ao perceptron. O Madaline (Multilayer adaline), criado pelos mesmos autores, possui uma arquitetura onde dois ou mais elementos processadores tipo “adaline” são utilizados como entrada para um outro elemento tipo “madaline”, cuja saída nada mais é do que a resposta de maior incidência entre os “adalines”, criando a primeira arquitetura de 3 camadas. (Figura 4.5.3.1)

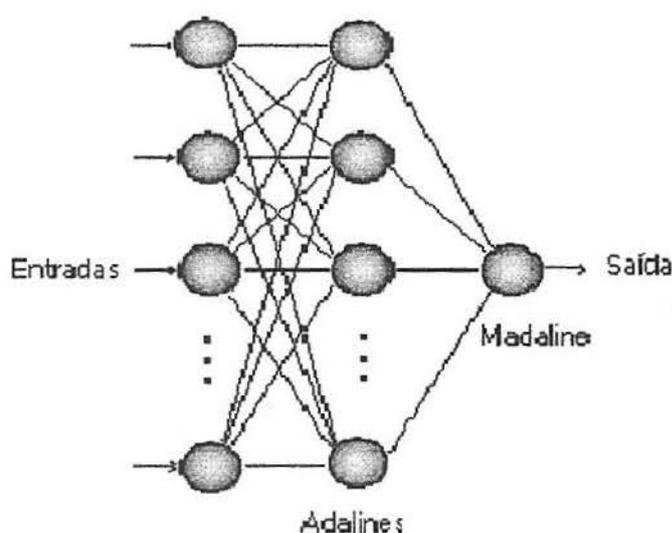


FIGURA 4.5.3.1: Arquitetura de um modelo Adaline-Madaline.

Pode-se treinar a rede Madaline para reconhecer, por exemplo, um padrão transladado ou até rotacionado, treinando cada elemento interno do Madaline para reconhecer o padrão em uma dada posição ou orientação e fazer o “ou” das saídas destes elementos.

O modelo Adaline corresponde a uma modificação do Perceptron, onde se utiliza o algoritmo de aprendizado *LMS (Least Mean Square)*, também conhecido como Regra Delta. Este algoritmo procura minimizar o erro quadrático médio entre a saída desejada e a saída real da rede. 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 retroalimentação do erro na sua própria saída. A diferença básica 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 Regra Delta pertence a um conjunto de algoritmos chamados regras lineares de correção de erros sendo baseada na descida do gradiente da curva de erro. Estas regras alteram os pesos do elemento linear adaptativo através da apresentação da resposta correta corrigindo estes pesos numa fração proporcional ao erro. Este erro é obtido pela diferença entre o valor da resposta correta desejada, e o valor obtido na saída do neurônio. Logo o erro na saída de um neurônio é determinado em função dos pesos usados em correspondência a cada entrada, ou seja, de acordo com a alteração dos pesos podemos subir ou descer a curva de erro resultante na saída do neurônio. A derivada E / P nos dará a inclinação da curva e assim permite que sigamos no sentido de descida da curva. A Figura 4.5.3.2 mostra o método de descida do gradiente da curva de erro. O problema ocorre se estamos em E2, onde podemos minimizar o erro em direção à um mínimo local, mínimo este que pode não ser a melhor solução do problema.



FIGURA 4.5.3.2: Método de descida do gradiente da curva de erro.

4.5.4. REDE DE RESSONÂNCIA ADAPTATIVA

As redes de ressonância adaptativa (ART) são uns dos mais complexos projetos de redes neurais disponíveis atualmente, contendo vários tipos de nós arranjados em diversos sub-sistemas. Uma das razões-chaves para considerar as redes ART importantes é que são elas que fornecem uma visão rápida do que é requerido para fazer sistemas verdadeiramente autônomos. Essa rede é baseada na teoria da ressonância adaptativa desenvolvida por Grossberg e Gail Carpenter. Essa rede pode processar somente dados binários de entrada.

O reconhecimento confunde-se com o aprendizado, uma vez que este é um modelo não-supervisionado que incorpora padrões desconhecidos e adapta padrões semelhantes.

O modelo ART possui a facilidade de não ter necessidade de um aprendizado prévio, podendo adaptar-se de acordo com a necessidade. Por outro lado, este modelo é muito sensível a padrões distorcidos, não podendo fazer uma boa diferenciação entre padrões semelhantes, limitando, assim, em muito suas aplicações.

Utiliza-se o ART para reconhecimento de imagens, mais comumente para reconhecimento de caracteres.

As principais características da rede ART são:

- A Teoria da Adaptação Ressonante (Adaptive Resonance Theory) resultou num classificador projetado por Carpenter e Grossberg que possui entrada binária, conexão retroalimentada e aprendizado não-supervisionado.
- A idéia de armazenamento dos padrões tem semelhanças com o modelo de Kohonen, uma vez que há a distinção de zonas, a comparação entre as distâncias vetoriais das somas ponderadas, e a escolha de um neurônio vencedor.

- aprendizado do modelo ART é bem distinto dos modelos vistos anteriormente, apenas com algumas semelhanças com o de Hopfield e o de Kohonen. Assemelha-se ao de Kohonen em sua forma de distribuição dos pesos na rede através da seleção de um neurônio vencedor, com a diferença de que no ART apenas o vencedor é atualizado, não havendo a atualização da vizinhança, que caracteriza o aprendizado competitivo. Quanto à semelhança ao Hopfield, nota-se que o padrão a aprender confunde-se com o a reconhecer, uma vez que há iterações para ambos.
- aprendizado ART não necessita a transformação da entrada binária (0 e 1) em bivalente (-1 e 1), como em outros modelos. Há somente um parâmetro a definir, o coeficiente de limiar, também chamado de *limiar de vigilância*. Como nos outros modelos, este coeficiente determina se um padrão está ou não treinado ou, neste caso, adaptado aos padrões atuais.

4.5.5. REDE KOHONEN

A rede Kohonen ou mapa de Kohonen usa aprendizado não supervisionado para modificar o estado interno da rede. A rede é formada por duas camadas: a camada de entrada onde os padrões são apresentados e a camada de Kohonen. Os neurônios da camada na rede Kohonen estão ligados a todos os neurônios da camada da rede Kohonen possui um conjunto de pesos iguais aos números de componentes formadores do padrão apresentado. Matematicamente a matriz que representa um padrão é equivalente em número de linhas e de colunas à matriz que representa um neurônio na camada da Rede Kohonen.

O esquema básico da rede desenvolvida por Kohonen é que tem a propriedade de se automodificar, ou seja, trata-se de uma rede não supervisionada e auto organizável. Resumidamente, o mecanismo de Kohonen funciona da seguinte forma: os pesos sinápticos iniciam contendo valores aleatoriamente baixos, e um sinal de entrada x (com valores que representam uma informação qualquer) é provido para a rede sem que se especifique a saída

desejada. Conforme o sinal de entrada, um neurônio de saída y deverá responder melhor para a entrada respectiva, e este será o vencedor, ou seja, sempre que essa entrada for apresentada, este neurônio será disparado. O neurônio vencedor y , e seus vizinhos, terão seus pesos sinápticos ajustados e, dessa forma, responderão melhor do que antes à entrada apresentada x .

As principais características da Rede Kohonen são:

- Possui auto-aprendizado, entrada intervalar e conexão competitiva.
- Não há propriamente um reconhecimento de padrão como em outros modelos, mas há a classificação de um padrão junto com outros que têm características semelhantes, formando classes. Estas classes são organizadas num mapa, onde se pode observar a distribuição dos padrões. Desta maneira, no instante em que um padrão é inserido na rede, esta o coloca na classe onde melhor o padrão se adequa, em função das suas características.
- Um outro aspecto importante é que o modelo de Kohonen é chamado de biologicamente plausível. No córtex auditivo, por exemplo, existem conjuntos de células que só reagem a determinados impulsos ou frequências, enquanto a outros não. No modelo ocorre o mesmo, onde um padrão ao ser reconhecido faz com que um ou somente alguns neurônios de saída sejam ativados (aqueles que mais se assemelham ao padrão inserido) enquanto outros não.
- Este tipo de rede é usado quando se deseja, por exemplo, reconhecer diversos padrões que possuam alguma relação entre si, como reconhecimento de voz, que será explicado posteriormente na seção de aplicações.

4.6. RESUMO DA APLICAÇÃO DE REDES NEURAIS

A seguir seguem os passos necessários para o desenvolvimento de aplicações utilizando redes neurais artificiais:

4.6.1. COLETA DE DADOS E SEPARAÇÃO EM CONJUNTOS

O primeiro passo do processo de desenvolvimento de redes neurais artificiais é a coleta de dados relativos ao problema e a sua separação em um conjunto de treinamento e um conjunto de testes. Esta tarefa requer uma análise cuidadosa sobre o problema para minimizar ambigüidades e erros nos dados. Além disso, os dados coletados devem ser significativos e cobrir amplamente o domínio do problema; não devem cobrir apenas as operações normais ou rotineiras, mas também as exceções e as condições nos limites do domínio do problema.

Normalmente, os dados coletados são separados em duas categorias: dados de treinamento, que serão utilizados para o treinamento da rede e dados de teste, que serão utilizados para verificar sua performance sob condições reais de utilização. Além dessa divisão, pode-se usar também uma subdivisão do conjunto de treinamento, criando um conjunto de validação, utilizado para verificar a eficiência da rede quanto a sua capacidade de generalização durante o treinamento, e podendo ser empregado como critério de parada do treinamento.

Depois de determinados estes conjuntos, eles são geralmente colocados em ordem aleatória para prevenção de tendências associadas à ordem de apresentação dos dados. Além disso, pode ser necessário pré-processar estes dados, através de padronizações, escalonamentos e conversões de formato para torná-los mais apropriados à sua utilização na rede. É importante, também, verificar a presença de “outliers” e proceder, então, a exclusão destes dados para que não afetem seriamente o treinamento da rede e os resultados obtidos.

4.6.2. CONFIGURAÇÃO DA REDE

O segundo passo é a definição da configuração da rede, que pode ser dividido em três etapas:

- 1 Seleção do paradigma neural apropriado à aplicação.
- 2 Determinação da topologia da rede a ser utilizada - o número de camadas, o número de unidades em cada camada, etc.
- 3 Determinação de parâmetros do algoritmo de treinamento e funções de ativação. Este passo tem um grande impacto na performance do sistema resultante.

Existem metodologias para a condução destas tarefas. Normalmente estas escolhas são feitas de forma empírica. A definição da configuração de redes neurais é ainda considerada uma arte, que requer grande experiência dos projetistas.

4.6.3. TREINAMENTO

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

Uma boa escolha dos valores iniciais dos pesos da rede pode diminuir o tempo necessário para o treinamento. Normalmente, os valores iniciais dos pesos da rede são números aleatórios uniformemente distribuídos, em um intervalo definido. A escolha errada destes pesos pode levar a uma saturação prematura.

Quanto ao modo de treinamento, na prática é mais utilizado o modo padrão devido ao menor armazenamento de dados, além de ser menos suscetível ao problema de mínimos locais, devido à pesquisa de natureza estocástica que realiza. Por outro lado, no modo batch se tem uma melhor estimativa do vetor gradiente, o que torna o treinamento mais estável. A eficiência relativa dos dois modos de treinamento depende do problema que está sendo tratado.

Quanto ao tempo de treinamento, vários fatores podem influenciar a sua duração, porém sempre será necessário utilizar algum critério de parada. O critério de parada do

algoritmo backpropagation não é bem definido, e geralmente é utilizado um número máximo de ciclos (iterações). Mas, devem ser considerados a taxa de erro médio por ciclo, e a capacidade de generalização da rede. Pode ocorrer que em um determinado instante do treinamento a generalização comece a degenerar, causando o problema de *over-training*, ou seja, a rede se especializa no conjunto de dados do treinamento e perde a capacidade de generalização.

O treinamento deve ser interrompido quando a rede apresentar uma boa capacidade de generalização e quando a taxa de erro for suficientemente pequena, ou seja, menor que um erro admissível. Assim, deve-se encontrar um ponto ótimo de parada com erro mínimo e capacidade de generalização máxima.

4.6.4. TESTE

O quarto passo é o teste da rede. Durante esta fase o conjunto de teste é utilizado para determinar a performance da rede com dados que não foram previamente utilizados. A performance da rede, medida nesta fase, é uma boa indicação de sua performance real.

Devem ser considerados ainda outros testes como análise do comportamento da rede utilizando entradas especiais e análise dos pesos atuais da rede, pois se existirem valores muito pequenos, as conexões associadas podem ser consideradas insignificantes e assim serem eliminadas (prunning). De modo inverso, valores substantivamente maiores que os outros poderiam indicar que houve *over-training* da rede.

4.6.5. INTEGRAÇÃO

Finalmente, com a rede treinada e avaliada, ela pode ser integrada em um sistema do ambiente operacional da aplicação. Para maior eficiência da solução, este sistema deverá

conter facilidades de utilização como interface conveniente e facilidades de aquisição de dados através de planilhas eletrônicas, interfaces com unidades de processamento de sinais, ou arquivos padronizados. Uma boa documentação do sistema e o treinamento de usuários são necessários para o sucesso do mesmo.

Além disso, o sistema deve periodicamente monitorar sua performance e fazer a manutenção da rede quando for necessário ou indicar aos projetistas a necessidade de retreinamento. Outras melhorias poderão ainda ser sugeridas quando os usuários forem se tornando mais familiares com o sistema, estas sugestões poderão ser muito úteis em novas versões ou em novos produtos.

5. MODELAGEM ESTATÍSTICA

A metodologia de redes neurais artificiais podem ajudar muito nas análises de dados estatísticos, contribuindo paralelamente às técnicas clássicas de análise estatística a obtenção de melhores resultados. O interesse principal é mostrar alguns exemplos de aplicação de redes neurais juntamente com as tradicionais técnicas estatísticas e comentar o seu inter-relacionamento, de modo geral, com a Estatística. Primeiramente, é apresentado na Tabela 5.1 um comparativo dos termos usados em Redes Neurais Artificiais com os termos equivalentes usados em Estatística.

TABELA 5.1: Quadro comparativo de termos

REDES NEURAIIS ARTIFICIAIS	ESTATÍSTICA
Neurônios de Entrada	Variáveis Independentes, Regressores
Neurônios de Saída	Valores Preditos
Valores de Treinamento, Alvos	Variáveis Dependentes
Erros	Resíduos
Treinamento, Aprendizagem, Adaptação	Estimação
Exemplo	Observações
Pesos	Estimativas de Parâmetros
Neurônios Escondidos	Interações
Topologia (Arquitetura) da Rede	Modelagem Estatística
Conexões Funcionais	Transformações
Aprendizagem Supervisionada	Regressão e Análise Discriminante
Aprendizagem não-supervisionada	Redução de Dados
Aprendizagem Competitiva	Análise de Agrupamento
Generalização	Interpolação e Extrapolação

Fonte: BISHOP (1995)

5.1. ANÁLISE DISCRIMINANTE

A Análise Discriminante, desenvolvida por FISHER (1938), é uma técnica estatística apropriada para discriminação e classificação (HAIR et al, 1998). Os objetivos da técnica envolvem a descrição, gráfica ou algébrica, das características diferenciais das observações das várias populações, além de classificar as observações em uma ou mais classes predeterminadas.

O interesse é obter uma regra que possa ser usada para classificar de forma otimizada uma nova observação à uma classe já existente. Esta é, então, a semelhança da Análise Discriminante com a técnica de Redes Neurais Artificiais que utiliza a informação de um banco de dados para criar uma regra de classificação para valores futuros.

A Análise Discriminante é adequada nas situações onde se pretende separar duas ou mais classes de objetos (pessoas, clientes, empresas, produtos, entre outros) ou alocar um novo objeto a uma das classes existentes ou, ainda, se pretende, conjuntamente, separar as classes e alocar um novo objeto. As classes poderiam ser, por exemplo, risco de crédito a clientes ruins e bons de uma instituição financeira. O vetor de variáveis classificatórias (variáveis explicativas) teria componentes tais como renda, idade, número de cartões de crédito, tamanho da família, saldo bancário, tempo como cliente. Uma vez determinada a regra de classificação, dado o conhecimento de seu vetor de variáveis medidas, um futuro cliente poderia ser classificado como mau ou bom pagador. (DINIZ & LOUZADA-NETO, 2000)

Para ilustrar a comparação entre a técnica de Redes Neurais Artificiais e a Análise Discriminante foi utilizado o exemplo de espécies de IRIS, um exemplo clássico de Análise Discriminate utilizado por FISHER (1938) para o desenvolvimento da técnica. As variáveis explicativas para este exemplo são o comprimento e a largura da sépala e o comprimento e a

largura da pétala. O exemplo consiste de 3 espécies de Iris: Setosa, Versicolor e Virginica; com 50 observações de cada uma, totalizando 150 observações.

O resultado da classificação das espécies por Análise Discriminante é apresentado na Tabela 5.1.1.

TABELA 5.1.1: Tabela de classificação das espécies por Análise Discriminante.

Predito \ Observado	Setosa	Versicolor	Virginica	Total
Setosa	50 100%	0 0%	0 0%	50 33,33%
Versicolor	0 0%	47 94%	3 6%	50 33,33%
Virginica	0 0%	2 4%	48 96%	50 33,33%
Total	50 33,33%	49 33,67%	51 34%	150 100%

Ao utilizar a Análise Discriminante para classificar as 150 observações nas 3 espécies obteve-se um erro de classificação de 3,33% (5 casos classificados incorretamente/ 150 casos totais).

Para proceder a resolução do exemplo através de Redes Neurais Artificiais, utilizou-se uma rede perceptron sem camadas ocultas e dividiu-se as observações em uma amostra de treinamento contendo 80% dos dados e uma amostra de teste com os 20% restantes dos dados. Os resultados de classificação da rede para amostras de treinamento e de teste são apresentados nas Tabelas 5.1.2 e 5.1.3, respectivamente.

TABELA 5.1.2: Tabela de classificação das espécies para a amostra de treinamento.

Predito \ Observado	Setosa	Versicolor	Virginica	Total
Setosa	40 100%	0 0%	0 0%	40 33,33%
Versicolor	0 0%	39 97,5%	1 2,5%	40 33,33%
Virginica	0 0%	1 2,5%	39 97,5%	40 33,33%
Total	40 33,33%	40 33,33%	40 33,33%	120 100%

TABELA 5.1.3: Tabela de classificação das espécies para a amostra de validação.

Predito \ Observado	Setosa	Versicolor	Virginica	Total
Setosa	10 100%	0 0%	0 0%	10 33,33%
Versicolor	0 0%	9 90%	1 10%	10 33,33%
Virginica	0 0%	0 0%	10 100%	10 33,33%
Total	10 33,33%	9 30%	11 36,67%	30 100%

Para obter o erro de classificação da rede juntou-se as duas amostras em uma única tabela. Este resultado é mostrado na Tabela 5.1.4.

TABELA 5.1.4: Tabela de classificação das espécies para a rede utilizando todos os dados.

Predito \ Observado	Setosa	Versicolor	Virginica	Total
Setosa	50 100%	0 0%	0 0%	50 33,33%
Versicolor	0 0%	48 96%	2 4%	50 33,33%
Virginica	0 0%	1 2%	49 98%	50 33,33%
Total	50 33,33%	49 32,67%	51 34%	150 100%

Utilizando-se a técnica de Redes Neurais Artificiais para classificar as 150 observações nas 3 espécies obteve-se um erro de classificação de 2% (3 casos classificados incorretamente/ 150 casos totais). Sendo assim, neste caso, a técnica de Redes Neurais mostrou-se superior a técnica de Análise Discriminante.

5.2. ANÁLISE DE REGRESSÃO LINEAR

A metodologia de regressão linear tem como objetivo prever ou classificar valores de y (variável resposta) a partir das variáveis explicativas x_i (variáveis independentes ou regressores). Para isso, obtém-se um plano de regressão para a resposta y em função de k variáveis explicativas como a que segue, onde β_i são os coeficientes para cada variável independente e ε_j a componente aleatório referente a cada observação:

$$y_j = \beta_0 + \sum_{i=1}^k \beta_i x_{ji} + \varepsilon_j$$

Este problema é equivalente a uma rede neural com uma única camada. As variáveis independentes correspondem as entradas da rede e a variável resposta y corresponde a saída da rede. Os coeficientes β_i 's correspondem aos pesos das conexões. A função de ativação é a função identidade. Para encontrar os pesos da rede utiliza-se o algoritmo de retropropagação e uma função custo similar a função objetivo da soma dos quadrados dos resíduos. A diferença entre as duas metodologias é que a análise de regressão tem uma forma de solução fechada para encontrar o valor dos coeficientes, enquanto que as redes neurais artificiais utilizam um processo iterativo para a obtenção dos pesos das conexões.

O objetivo é obter os coeficientes β_i , que minimizem $E = \sum_{j=1}^n \left(y_j - \sum_{i=1}^k \beta_i x_{ji} \right)^2$ baseado nas observações (x_j, y_j) , $j = 1, \dots, n$. Então, o problema no contexto de regressão linear é equivalente a uma rede neural com uma camada escondida. A função de ativação é a

identidade. Os pesos são obtidos por um processo iterativo ao passo que uma expressão explícita é obtida na análise clássica de regressão.

Por analogia, qualquer modelo linear generalizado é equivalente a uma rede neural com uma única camada. A função de ativação é escolhida para coincidir com o inverso da função de ligação.

Uma desvantagem na utilização de redes neurais artificiais quando comparado a utilização da análise de regressão envolve a escolha de muitos parâmetros como o número de nós escondidos, o parâmetro de aprendizagem η (taxa de aprendizagem), os pesos iniciais, a função custo e de ativação e a decisão de quando parar o treinamento.

Para ilustrar a aplicação de regressão em redes neurais artificiais, segue um exemplo descrito em (CHURCH & CURRAM, 1996) que compara previsões de despesa pessoal obtidas com rede neural e modelos de regressão (econométricos). Nenhum dos modelos foi capaz de explicar a queda no crescimento das despesas no fim dos anos 80 e início dos anos 90.

Comparando as técnicas, obteve-se que as redes neurais utilizaram exatamente as mesmas variáveis explicativas e observações usadas em cada modelo econométrico, produzindo resultados semelhantes. As redes neurais usaram 10 neurônios em uma camada escondida para todos os modelos.

Em um exercício final utilizou-se uma rede neural com entradas de todas as variáveis de todos os modelos. Esta rede foi capaz de explicar a queda de crescimento, mas resultou em uma rede com muitos parâmetros. Além de comparação das previsões, também foi realizada uma análise de sensibilidade com respeito a cada variável. Para isto, testou-se a rede com o valor médio dos dados e variou-se cada variável para verificar o grau em que cada variação afeta a previsão da variável dependente.

5.3. ANÁLISE DE REGRESSÃO LOGÍSTICA

O modelo de regressão logística, também conhecido por modelo logístico, é, em geral, utilizado para tratar problemas relacionados a dados dicotômicos em várias áreas do conhecimento. Este modelo estabelece uma relação entre a probabilidade de ocorrência dos resultados de uma variável respostas dicotômica (em geral chamada de variável dependente), que normalmente é representada pelos termos sucesso e fracasso, e variáveis explicativas categóricas ou contínuas. O interesse principal é estabelecer uma relação linear entre as variáveis explicativas e uma transformação, denominada logito (logit), da variável resposta. Este modelo é representado por

$$\log \left[\frac{P\{Y(x) = 1\}}{P\{Y(x) = 0\}} \right] = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p,$$

onde $P\{Y(x)=1\}$ representa a probabilidade de sucesso para a variável resposta, $P\{Y(x)=0\}$ representa a probabilidade de fracasso, β_0 denota o intercepto da regressão e $\mathbf{x}'=(x_1, x_2, \dots, x_p)$ é um vetor de variáveis explicativas com coeficientes $\beta_1, \beta_2, \dots, \beta_p$. (HOSMER & LEMESHOW, 1989)

Um exemplo apresentado por PEREIRA e RODRIGUES (1998) trata de medições de sonografia em 458 mulheres para diagnóstico de câncer de mama. Obteve-se que 325 destas mulheres tinham tumores benignos ($y=0$) e as outras 133 tinham tumores malignos ($y=1$).

A priori, fazendo-se uma análise de regressão logística verificou-se que três variáveis explicativas foram significativas que foram: idade da paciente, número de artérias no tumor (AT) e número de artérias na mama controlateral (AC). A comparação do modelo de Regressão Logística e da metodologia de Redes Neurais Artificiais é apresentada na Tabela 5.3.1.

TABELA 5.3.1 – Comparação dos modelos Logístico e Rede Neural.

Variável	Regressão Logística			Pesos – Rede Neural
	Coefficientes	Erro Padrão	Valor de P	
Intercepto	-8,178	0,924		$W_0=-8,108$
Idade	0,070	0,017	0,0001	$W_1=0,069$
log AT+1	5,187	0,575	0,0001	$W_2=5,162$
log AC+1	-1,074	0,437	0,0014	$W_3=-1,081$

Como pode-se observar os resultados obtidos para os coeficientes da regressão logística são muito próximos aos obtidos pela rede neural utilizando um perceptron multi-camada com algoritmo retropropagação com quatro nós de entrada ($x_0=1$, $x_1=idade$, $x_2=log AT+1$, $x_3=log AC+1$).

5.4. ANÁLISE DE SÉRIES TEMPORAIS

O problema da previsão de valores futuros de séries temporais tem se mostrado bastante apropriado à aplicação de Redes Neurais Artificiais em busca de uma solução satisfatória. O modelo de redes neurais utilizado para resolver este tipo de problema que aparece com maior frequência na literatura é o modelo dos perceptrons multi-camadas (FERNANDES et al, 1996).

O maior problema, neste contexto, está em determinar qual a melhor topologia (arquitetura) da rede que é resolvido fazendo-se uma análise estatística prévia da série temporal.

Sendo assim, pode-se detalhar as três fases para a construção da arquitetura de uma rede neural para previsão de valores futuros:

1ª) Fazer a análise dos componentes da série temporal, como a tendência e a sazonalidade. Realizar também uma análise da componente aleatória, pois quanto maior ela for, mais difícil se torna encontrar um padrão de previsão com redução dos erros.

2ª) Definir o número de nós na camada de entrada para conseguir representar as informações referentes ao comportamento da série. Desta forma, permite-se que a fase de aprendizado da rede capte as características mais importantes da série.

3ª) Determinar o número de nós na camada escondida, a partir do método de RUMELHART & WEIGEND (1990), onde a rede que melhor generaliza os dados é a menor rede apta a realizar o treinamento dos dados. Esta busca é feita de forma empírica, identificando-se a melhor arquitetura a partir dos resultados obtidos pelos experimentos realizados.

FERNANDES et al (1994) descreveram um exemplo de aplicação de Redes Neurais Artificiais, apresentando a escolha da topologia da rede utilizada, os resultados obtidos e comparações feitas com o método tradicional de Séries Temporais.

O exemplo utilizado é a série da produção industrial de 12 meses do estado do Rio Grande do Sul. Através da análise estatística observou-se que se trata de uma série estacionária e sazonal. A arquitetura proposta para o exemplo apresenta 13 nós na camada de entrada, sendo um representando a informação passada (Y_{t-1}) e os outros 12 representando os fatores sazonais, sendo que estão sob a forma de uma seqüência binária que atribui um para o mês em questão e zero para os demais. A arquitetura apresenta dois nós na camada intermediária e um nó na camada de saída que fornece o prognóstico Y_t .

Foram feitos prognósticos de valores futuros através da metodologia de Redes Neurais Artificiais (RNA) e pelo método dos modelos de componentes não-observáveis (UCM). UCM é um modelo de séries temporais que baseia-se, tanto na abordagem clássica quanto na bayesiana, no filtro de Kalman que é um algoritmo que possibilita o aprendizado através dos erros de seus prognósticos (HARVEY, 1989). Os resultados obtidos pelos dois procedimentos (UCM e RNA) são apresentados nas Tabelas 5.4.1 e 5.4.2, respectivamente.

As representações gráficas das previsões pelos dois procedimentos e os valores reais da série são apresentados nas Figuras 5.4.1 e 5.4.2.

TABELA 5.4.1 – Prognósticos dos modelos de componentes não-observáveis (UCM)

Período	Realidade	Previsão um passo à frente	Erro percentual	Previsão sete passos à frente	Erro percentual
1-93	99,70	98,69	1,01%	98,69	1,01%
2-93	105,33	104,10	1,15%	103,39	1,84%
3-93	131,95	118,04	10,54%	116,27	11,88%
4-93	127,39	129,24	-1,46%	117,91	7,44%
5-93	136,98	134,46	1,84%	123,77	9,65%
6-93	135,56	131,39	3,07%	119,63	11,75%
7-93	131,84	131,68	0,12%	117,13	11,16%

TABELA 5.4.2 – Prognósticos da rede neural artificial (RNA)

Período	Realidade	Previsão um passo à frente	Erro percentual	Previsão sete passos à frente	Erro percentual
1-93	99,70	98,76	0,95%	98,76	0,95%
2-93	105,33	100,05	5,27%	103,12	2,14%
3-93	131,95	113,73	16,02%	124,95	5,60%
4-93	127,39	126,49	0,71%	134,56	-5,33%
5-93	136,98	134,03	2,20%	137,40	-0,30%
6-93	135,56	129,74	4,48%	139,50	-2,82%
7-93	131,84	127,79	3,16%	131,67	0,13%

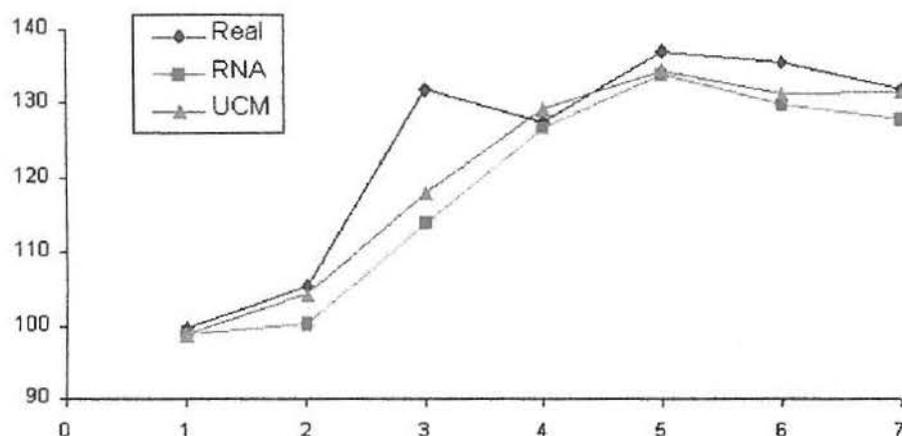


FIGURA 5.4.1 – Previsão um passo à frente

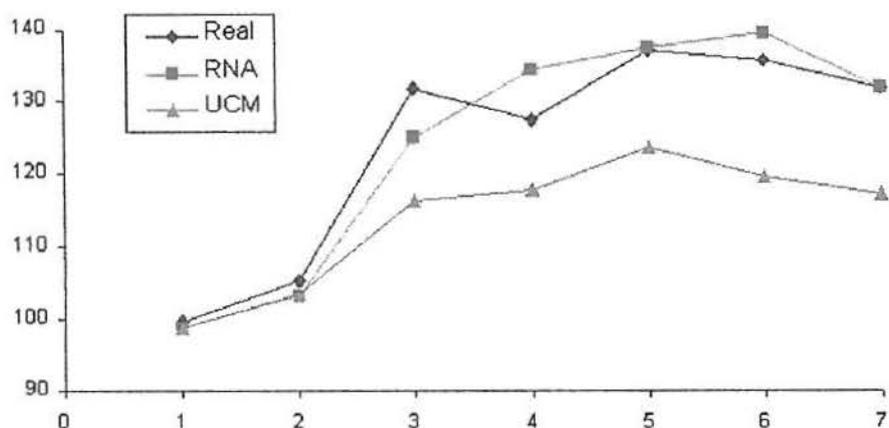


FIGURA 5.4.2 – Previsão sete passos à frente

Comparando as Figuras 5.4.1 e 5.4.2, pode-se observar uma diferenciação quanto a precisão dos dois métodos. Enquanto os UCM apresentam uma melhor performance na realização de prognósticos apenas um passo à frente, as RNA mostraram-se bem mais eficientes nas previsões de vários passos à frente. É importante ressaltar que a melhor predição foi obtida utilizando-se RNA em previsão 7 passos à frente. Percebe-se claramente isso na Tabela 5.4.3 onde são apresentados o erro absoluto médio (EAM) e o erro quadrático médio (EQM).

TABELA 5.4.3 – Comparação dos erros para cada método em função da etapa de previsão

	Previsão um passo à frente		Previsão sete passos à frente	
	UCM	RNA	UCM	RNA
EAM	3,55	5,45	10,28	3,12
EQM	5,65	7,75	11,86	4,17

6. APLICAÇÕES

O estudo de redes neurais é cada vez mais vasto tendo aplicações em várias áreas do conhecimento. Nesta seção é detalhado alguns exemplos de aplicação de redes neurais em algumas áreas do conhecimento.

6.1. ANÁLISE DE CRÉDITO

O uso de redes neurais como instrumento de apoio à avaliação de riscos de inadimplência é apresentado por ALMEIDA (1993). Neste caso as redes neurais aprenderam a identificar o risco de que uma dada empresa venha a se tornar concordatária num futuro próximo (2 anos). A seguir é detalhado como se deu o processo de aprendizado e uso da rede.

O processo é dividido em duas etapas. Uma primeira etapa é a etapa de aprendizado da rede, isto é a etapa onde a rede adquire o conhecimento sobre o problema. A segunda etapa é o uso da rede já treinada contendo o conhecimento sobre o problema.

Inicialmente escolheu-se as variáveis que permitam avaliar uma empresa. Índices financeiros são as variáveis normalmente utilizadas para a criação de ferramentas de análise de riscos de inadimplência. Um dos grupos de variáveis que foram escolhidos na pesquisa é mostrado na Tabela 6.1.1.

TABELA 6.1.1: Índices Financeiros

R1=Vendas Líquidas/Ativo total	R2= Total Exigível/Total Ativo
R3=Fluxo de Caixa/Vendas líquidas	R4=Ativo Circulante/Passivo Circulante
R5=Receita Bruta/Total de Juros Pagos	R6=log(Ativo Total)

Seleciona-se, então, um conjunto de empresas, contendo empresas inadimplentes e adimplentes e coleta-se os seis índices financeiros para cada uma das empresas selecionadas. Os dados de índices financeiros foram então utilizados para fazer com que a rede aprenda a distinguir os dois grupos de empresas. Veja na Tabela 6.1.2 como estes dados estariam organizados para serem apresentados à rede.

Tabela 6.1.2: Um exemplo de dados para as redes neurais

Empresa	Índices das Empresas						Situação da Empresa
	R1	R2	R3	R4	R5	R6	
Empresa 1	R11	R12	R13	R14	R15	R16	Inadimplente
Empresa 2	R21	R22	R23	R24	R25	R26	Inadimplente
Empresa 3	R31	R32	R33	R34	R35	R36	Inadimplente
Empresa 4	R41	R42	R43	R44	R45	R46	Adimplente
Empresa 4	R41	R42	R43	R44	R45	R46	Adimplente
Empresa 5	R51	R52	R53	R54	R55	R56	Inadimplente
Empresa 6	R61	R62	R63	R64	R65	R66	Adimplente

Cada linha é formada por um par de dados ou exemplo de uma empresa. Estes dados são apresentados à rede ainda sem conhecimento que irá aprender a distinguir as empresas boas das más. Este aprendizado se passará automaticamente e o tempo necessário será função de fatores como o número de exemplos fornecidos, estrutura da rede, e sem dúvida da velocidade do computador que se está usando.

Com o fim do aprendizado a rede passa a guardar dentro de sua estrutura o conhecimento para distinguir empresas inadimplentes de adimplentes a partir dos 6 índices financeiros usados no seu aprendizado.

Na segunda etapa, necessita-se não apenas que a rede aprenda a distinguir as empresas, dentro do grupo de empresas que lhe apresentamos, mas que a rede seja capaz de generalizar

seu conhecimento para identificar riscos de inadimplência de outras empresas que não fizeram parte de seu aprendizado.

A rede treinada na primeira etapa é capaz de fornecer o risco de inadimplência de uma empresa qualquer a partir de suas características financeiras. Começa-se, então, a usar a rede como base de conhecimento, como instrumento de apoio à decisão.

Segundo (ALMEIDA 1993), as redes foram capazes de classificar corretamente aproximadamente 70% das empresas de um grupo de empresas adimplentes e inadimplentes, dois anos antes da inadimplência, a partir de uma amostra de controle (uma amostra diferente da utilizada para o processo de aprendizado).

Outro aspecto interessante da rede neural, é que uma vez construída, ela não depende mais da base de dados que lhe deu origem. O conhecimento adquirido reside nas conexões entre os neurônios e não na base de dados. É na verdade um conhecimento implícito e conciso que é reconstituído pelo sistema a cada vez que é solicitado.

6.2. FINANÇAS

As redes neurais tem encontrado bastante interesse na área financeira. Uma vez que estas precisam de dados para serem criadas, a área financeira tem podido fornecê-los em abundância.

Um tipo de aplicação que mostra evidentemente bastante interesse é a previsão de cotações na bolsa. No entanto os resultados que se tem obtido mostram que não será fácil obter algo de interessante neste campo mesmo com redes neurais (HAWLEY et al, 1990). Em princípio em um mercado eficiente não é possível fazer previsões, o que está de acordo com os resultados obtidos pelas redes neurais. Mas também é possível que não se tenham usado os dados adequados. Tem-se por hábito acumular e tratar somente dados quantitativos. Mas o interessante das redes neurais é também a sua capacidade de tratar dados qualitativos. Pode-se

usar por exemplo dados qualitativos como setor de atividade da empresa, ou então tipo de produto vendido. Pode-se até mesmo usar variáveis menos acessíveis como características de gestão dos responsáveis pela empresa, propensão da direção a assumir riscos, ou capacidade de inovação da empresa, flexibilidade a mudanças, etc. No entanto este tipo de informação não é atualmente utilizado e os sistemas tem-se limitado a dados quantitativos. Isto ocorre muito provavelmente por uma questão de falta de disponibilidade de dados.

A área financeira oferece outras inúmeras oportunidades para o uso de redes neurais. Por exemplo gestão de portfólios. Uma instituição financeira leva em conta inúmeros tipos de papéis financeiros, ações, opções, etc. Um sistema baseado em redes neurais poderia ser usado para levar em conta as características destes papéis, considerando-se riscos, valores conjunturais, decisões do governo, flutuações do mercado, etc. As redes neurais são bastante adaptadas a problemas do tipo não estruturado, onde as relações entre as variáveis ou mesmo sua importância não é conhecida. As redes neurais permitem que se trabalhe com dados incompletos se estes não estão todos disponíveis ou mesmo com a existência de certas informações incorretas. Ela aprenderá a ignorá-las.

6.3. MARKETING

Uma área ainda pouco explorada é a área de marketing e vendas onde existem aplicações potencialmente bastante interessantes. Por exemplo as redes neurais podem ser utilizadas para fazer simulações sobre o comportamento do consumidor face a novos produtos. Uma rede poderia ser treinada a partir da característica de produtos já lançados no mercado onde se pôde medir o grau de sucesso e volume de vendas obtido. Uma vez criada a rede esta poderia ser usada para simular o resultado com novos produtos a partir de suas características. O interessante é que seria possível não só introduzir características

quantitativas e qualitativas do produto mas também características do mercado ou da conjuntura do país ou de uma região em um determinado momento.

Outra aplicação em marketing poderia ser simplesmente a simulação de vendas para um próximo período em função do resultado obtido no período anterior, das características esperadas para o mercado no futuro, etc.

Uma outra aplicação poderia ser o uso de uma rede neural para sugerir produtos mais adaptados ao perfil de cada cliente. A rede neural seria treinada a associar perfis de clientes a produtos da empresa. Poder-se-ia assim esperar por um lado a obtenção de um maior grau de satisfação do cliente, por outro lado apresentar produtos mais direcionados ao cliente em questão. Este tipo de aplicação pode ser encarado como um sistema de apoio especialista, onde um vendedor menos experimentado ou menos informado a respeito dos produtos da empresa, pode sugerir itens mais adequados ao cliente em pé de igualdade com um vendedor mais experimentado.

No caso de atividades onde há uma grande velocidade de mudança ou atualização do portfólio de produtos, este tipo de sistema permitiria à empresa fazer com que o vendedor passasse a levar em conta os novos produtos de maneira mais rápida uma vez que atualizada a base de conhecimentos, eles passariam a ser sugeridos. O sistema seria igualmente interessante no caso daqueles produtos que são freqüentemente esquecidos pelo vendedor. Caso fossem do interesse do cliente eles seriam igualmente sugeridos. Não seria menos interessante que o sistema incorporasse os interesses de venda da empresa.

Simulações, previsões ou sugestões, as redes neurais são uma tecnologia interessante na área de marketing. Há que se levar em conta também o fato de que sua atualização, isto é, a aquisição de novos conhecimentos é bastante facilitada e rápida uma vez que o aprendizado é automático. Basta entrar as novas características do novo produto ou da nova situação de mercado e refazer o processo de aprendizado.

6.4. MORTALIDADE INFANTIL

O coeficiente de mortalidade infantil tem sido considerado como um dos mais expressivos indicadores do estado de saúde, não somente de crianças menores de um ano, mas de toda a população, refletindo as condições do meio-ambiente, contexto sócio-econômico, nutrição, educação e cuidados com a saúde de uma comunidade. Tais considerações apontam para a potencialidade da modelagem da mortalidade infantil, o que permitiria, por exemplo, a avaliação do impacto de intervenções, através de simulação, e, conseqüentemente, a decisão quanto às melhores estratégias no planejamento em saúde. Com este objetivo, modelos de Regressão Linear Múltipla (RLM) e Redes Neurais Artificiais (RNA) foram desenvolvidos a partir de 43 variáveis de 59 municípios brasileiros. Para os modelos RLM, os mais comuns na literatura, o critério de seleção de variáveis independentes/variáveis explicativas baseou-se na metodologia *stepwise* ou foi orientado pela Análise Fatorial (AF). Neste último, a seleção de preditores se deu tomando-se as variáveis de maior carga fatorial em cada fator comum, e, alternativamente, segundo um *critério misto*, que consistiu em selecionar a variável de cada fator comum que também atendesse a outros requisitos, tais como completude e confiança nos dados. Para a modelagem não-linear, através de RNA, foram desenvolvidos dois ensaios; o primeiro, utilizando as 43 variáveis como entrada, e o segundo somente aquelas oito selecionadas pelo *critério misto*. Na avaliação dos modelos utilizou-se o Coeficiente de Determinação (R^2) e o Erro Quadrático Médio de Estimação (EQM), bem como replicações *bootstrap* dessas estatísticas. Os modelos RLM *stepwise* apresentaram, em geral, estimativas super-dimensionadas de R^2 e multicolinearidade, prejudicando a interpretabilidade do próprio modelo e dificultando sua utilização em estudos de simulações de intervenção. O modelo RLM *critério misto* foi o de melhor capacidade preditiva ($R^2=0,44$ e $EQM=20,44 \times 10^3$) e interpretabilidade, entre os lineares. Entretanto, a modelagem RNA *critério misto* mostrou-se muito superior ($R^2=0,80$ e $EQM=6,85 \times 10^3$) ao melhor modelo linear; resultado este

confirmado através de técnica *bootstrap*. A análise de sensibilidade do RNA *critério misto* evidenciou a elevada importância relativa das variáveis *número de pessoas alfabetizadas, pessoas ocupadas nos estabelecimentos agropecuários e número de estabelecimentos comerciais*. A exclusão de qualquer uma dessas variáveis em estimativas *bootstrap* resultou em aumento do erro de predição de 40 a 55%, bem como uma redução de 21 a 29% na capacidade do modelo prever a variância da mortalidade infantil. Tais evidências apontam para a potencialidade do uso deste modelo em estudos de simulações de intervenção, auxiliando o estabelecimento de estratégias de políticas públicas locais de saúde (GISMONDI et al, 1999).

6.5. SAÚDE

Vigilância epidemiológica é um procedimento contínuo de monitoração do comportamento de variáveis de saúde. Para que este acompanhamento seja eficaz, é necessário dispor de análise adequados aos dados epidemiológicos. Com esta finalidade diversas metodologias foram propostas, abordando o problema de algumas formas diferentes. Uma das possíveis abordagens é comparar os novos dados, a medida que chegam, com previsões feitas anteriormente, através de modelos matemáticos apropriados. Normalmente, utilizam-se os modelos estocásticos construídos a partir de séries temporais das variáveis em estudo. SÁ & NOBRE (1996) propuseram o uso de novos modelos preditivos, baseados em Redes Neurais artificiais e testaram sua aplicação a séries temporais epidemiológicas. Utilizaram redes neurais do tipo "feedforward", treinadas por Backpropagation para gerar previsões "um passo a frente", a partir de um determinado grupo de observações anteriores. No desenvolvimento de tais redes, dois importantes fatores foram estudados mais profundamente, sendo testadas um número de opções para cada um deles. O primeiro foi o tamanho e a composição do grupo de observações passadas que deve ser utilizado como

entrada da rede para cada previsão (que determina o tamanho da camada de entrada). Investiga-se quatro propostas de metodologia para definição destes grupos, baseadas em ferramentas clássicas de análise de séries temporais. O segundo foi a ordem de apresentação destes grupos de entradas e suas respectivas saídas-alvo durante o treinamento das redes. Testa-se apresentações sequenciais, aleatórias e em lote. Além disso, investiga-se também a influência da proporção dos dados usados nos conjuntos de treinamento e de teste e o comportamento do modelo para horizontes de previsão maiores. O procedimento baseado nos "lags" significativos da Função de Autocorrelação apresentam algumas vantagens do ponto de vista metodológico, mas os resultados obtidos não permitem caracterizá-lo, ou a qualquer dos outros métodos, como o mais eficiente. O treinamento por regra delta aleatória obteve, em média, resultados melhores, embora para algumas redes resultados equivalentes ou superiores foram alcançados pela regra delta sequencial e pelo treinamento em lote. A proporção na divisão dos dados entre conjunto de treinamento e teste foi de pouca influência.

PACHECO et al (1996) afirmam que os métodos mais difundidos de avaliação tecnológica em saúde, nos países desenvolvidos, têm sido orientados a avaliação de tecnologias específicas. As condições existentes nos países em desenvolvimento levam a privilegiar uma avaliação tecnológica em saúde *orientada a problemas*. Interessa identificar problemas em áreas de saúde deficitárias, de forma que a alocação de recursos resultantes da avaliação cause o maior impacto no sistema de saúde. Utilizar esta avaliação geralmente implica na avaliação de múltiplas tecnologias que atuam conjuntamente e portanto são necessários novos métodos para enfrentar a complexidade deste problema. No trabalho utilizou-se *Redes Neurais Artificiais (RNA)* de aprendizado supervisionado como um novo paradigma - baseado em dados retrospectivos - para realizar *Avaliação de Múltiplas Tecnologias de Saúde (AMTS)*. Utilizaram-se métodos estatísticos para obter o tamanho do conjunto de aprendizado e redes com capacidade de generalização. Desenhou-se um método para extrair o conhecimento de *RNA* que permite determinar o impacto das entradas nas

variáveis de saída, em estruturas hetero-associadas. Usando 523 casos (63 óbitos) coletados em duas *Unidades de Cuidados Intensivos Neonatais (UCIN)* do estado do Rio de Janeiro, aplica-se a metodologia para determinar a efetividade das tecnologias usadas e estimar a estadia. A efetividade foi obtida de um modelo de mortalidade que inclui como entradas, condições de admissão (peso ao nascer, idade gestacional e Apgar 5º min), severidade da doença (baseada em oito grupos de hipóteses diagnósticas) e 20 tecnologias. Os resultados mostram que é possível obter as efetividades relativas das tecnologias de terapia. Há uma evidência forte de que as tecnologias básicas como antibióticos e hidratação venosa são mais efetivas que tecnologias mais complexas como o exsangüíneo e ventilação mecânica. Mostra-se também que os modelos baseados em *RNA* são significativamente superiores a modelos de regressão na estimação da estadia. Considerando que a obtenção da efetividade de múltiplas tecnologias em *UCIN* constitui um problema metodologicamente complexo, é possível inferir que o paradigma proposto pode fornecer resultados adequados ao ser aplicado em outros problemas de *AMTS*.

6.6. MEDICINA

Algumas das aplicações mais frequentes de redes neurais artificiais em medicina são:

Diagnosticando Câncer de Próstata. A corporação Kaman de Ciências (Colorado Springs, CO) criou uma rede a qual pode prever o câncer de próstata. Este sistema foi apresentado na "CNN" (canal de televisão) e no "Jornal Wall Street" como um grande avanço tecnológico. Redes foram treinadas para prever resultados baseados em teste de PSA ("Prostate Specific Antigen) e para prever a recorrência de câncer depois do tratamento.

Preveno permanência de paciente. Um grupo da Escola de Medicina da Universidade Johns Hopkins desenvolveu uma rede neural para prever quais pacientes na Unidade de Tratamento Intensivo (UTI) cirúrgico do Hospital iriam permanecer sete ou mais dias. As entradas da rede

incluíam idade, o uso de droga ou outras terapias, e sinais fisiológicos tais como taxa cardíaca e pressão sanguínea, e resultados de laboratório tais como contagens de células brancas. Um estudo da Hopkins concluiu que as "três redes neurais foram substancialmente melhores modelos preditivos do que o modelo de regressão linear múltipla..." Em uma aplicação similar, um grupo do Hospital de St. Michael em Toronto, Canadá, desenvolveu uma rede neural para prever o tempo de estadia de um paciente seguido à cirurgia cardíaca.

Preveno custos de tratamento. Os médicos e administradores da Clínica Cleveland desenvolveram uma rede neural a qual prediz o custo associado com cateterização e intervenção no coração. A rede prediz se o paciente irá ser um caso normal ou de alto custo. As entradas incluem idade do paciente, sexo, e outros parâmetros médicos tal como dados estruturais sobre o coração do paciente e vasos sanguíneos.

Determinando significância de drogas. O grupo de pesquisa da "Zeneca Pharmaceuticals" em Wilmington, DE, desenvolveu uma aplicação de rede neural para examinar dados de experimentos com drogas os quais reduzem o tempo de análise em até 68%. A rede neural foi treinada para detectar "bons" picos de pressão da bexiga de centenas de picos em arquivos de dados. Estes picos são críticos para determinar a significância da droga em experimentos desenhados como "screening" de mais compostos de droga. Os pesquisadores previamente usaram o método do papel e o lápis para analisar estes picos, um método que levou 40 horas por mês. Usando redes neurais, reduziu o tempo de análise para 13 horas por mês.

Monitorando sinais fisiológicos. A Escola de Medicina da Universidade do Colorado desenvolveu um aplicação consistindo de quatro redes neurais as quais detectam anormalidades respiratórias por mal funções em equipamentos quando um paciente está sob anestesia.

Monitorando sistemas cardiovasculares. O Laboratório Nacional do Noroeste do Pacífico desenvolveu um algoritmo de treinamento que recebe dados fisiológicos tais como taxa cardíaca e respiratória de um monitor usado por um indivíduo e descreve um modelo do

sistema cardiovascular da pessoa. Este modelo pode ser usado para prever o que seria a resposta da pessoa em uma situação particular (tal como combatendo fogo ou envolvendo-se em atividade policial) ou para prever a saúde de um indivíduo ao longo do tempo.

7. EXEMPLO

O mercado financeiro é afetado por um grande número de fatores econômicos, políticos e psicológicos, altamente relacionados e que interagem entre si de uma maneira bastante complexa. A maior parte destas relações parecem ser probabilísticas e não-lineares. Portanto, estas relações dificilmente podem ser expressas através de regras determinísticas.

Análise financeira é uma das aplicações mais promissoras e mais adequadas aos sistemas de Redes Neurais Artificiais. No setor de serviços financeiros, sistemas de suporte a decisões vêm sendo amplamente utilizados em uma grande gama de tarefas, incluindo:

- projeção de autorização de crédito;
- avaliação de risco de falência;
- revisão econômica e financeira, entre outras.

Os sistemas financeiros de suporte a decisões devem transformar grandes quantidades de dados em classificações inteligíveis, reconhecendo tendências e padrões no conjunto de dados. Qualquer pequena melhoria de desempenho nestes sistemas pode gerar muitos benefícios financeiros.

Na maioria dos ambientes de decisão, os gerentes trabalham com uma enorme variedade de informações vindas das mais diversas fontes. Muitas destas informações podem ser incompletas, ambíguas, parcialmente incorretas ou de relevância dúbia. Os gerentes analisam estas informações de uma maneira muito subjetiva. Muitas vezes eles não sabem explicar os processos de tomada de decisão, mas conseguem apontar os fatores que influenciam suas decisões.

Os sistemas que utilizam Redes Neurais Artificiais estão se tornando uma técnica muito atrativa para estas aplicações, pois oferecem uma alternativa mais eficiente para tratar informações dessa natureza.

O problema de avaliação de crédito pode ocorrer na forma de solicitação de cartão de crédito, crediário, autorização de compra, empréstimos bancários e cheque especial, entre outros. Estas avaliações envolvem previsão de candidatos confiáveis e de lucratividade da empresa na transação.

Durante uma avaliação, os gerentes baseiam suas decisões em informações sobre o cliente, em sua experiência gerencial e em relatórios de departamentos administrativos ou agências de crédito. Essas decisões devem ser muito criteriosas. Decisões equivocadas podem provocar não apenas prejuízos financeiros às instituições de crédito, mas também prejuízos financeiros e morais aos clientes.

Este método de avaliação apresenta os seguintes problemas:

Baixa Eficiência. Consomem muito tempo e necessitam de pessoal altamente especializado.

Baixa Performance. A experiência de grande parte dos gerentes não é suficientemente abrangente para produzir avaliações muito meticulosas.

Pouca Flexibilidade. Uma grande experiência é necessária antes que os gerentes possam realizar esta tarefa. Este método reage muito vagarosamente a mudanças.

Inconsistência. Diferentes gerentes utilizam diferentes critérios de avaliação.

Inicialmente, técnicas estatísticas, como análise discriminante e regressão logística, foram largamente utilizadas em sistemas automatizados para auxílio de avaliações. Estes sistemas aumentaram as performances de avaliação, pois apresentam maior flexibilidade, consistência e eficiência, mas fatores que limitam o uso destas técnicas indicam que nem sempre são apropriadas para estes problemas.

Os sistemas neurais oferecem uma abordagem mais adequada para o tratamento de problemas deste tipo. Uma rede neural artificial pode ser treinada utilizando grandes

quantidades de exemplos significativos, o que contribui para a obtenção de performance elevada. Ou seja, a avaliação é modelada através de exemplos de aplicações bem ou mal sucedidas, encontrados nos históricos dos clientes e não através de modelos objetivos ou quantificáveis da operação de avaliação.

Desta forma, a rede é capaz de encontrar relações entre as diversas informações e incorporar os aspectos subjetivos do processo de avaliação da empresa. Estes sistemas são mais flexíveis, pois quando mudanças são necessárias, devido à aspectos políticos, econômicos ou regionais, pode-se alterar apenas o conjunto de exemplos de treinamento, o que não requer mudanças estruturais no sistema. Estes sistemas podem, até mesmo, detectar esta necessidade e realizar um novo treinamento. Eles também são considerados consistentes e eficientes, pois um mesmo sistema pode ser utilizado nos diferentes postos de atendimento da empresa e realizar avaliações mais rapidamente.

7.1. CARACTERIZAÇÃO E PREPARAÇÃO DOS DADOS

Com a finalidade de ilustrar a aplicação da técnica de redes neurais artificiais utilizou-se um exemplo de banco de dados do Pacote SAS (Statistical Analysis System) sobre análise de crédito. A preparação dos dados e a análise da Rede Neural é feita no Módulo Enterprise Miner do SAS.

O arquivo de dados foi obtido através do nó de fonte de dados. A base de dados constitui de 1000 pedidos de financiamentos de clientes de um banco alemão com as informações apresentadas na Tabela 7.1.1.

TABELA 7.1.1: Nome, descrição e tipo das variáveis constantes no banco de dados.

Nome	Descrição	Tipo
<i>checking</i>	Verificação da movimentação da conta	ordinal
<i>duration</i>	Duração em meses	intervalar
<i>history</i>	Histórico de crédito	ordinal
<i>purpose</i>	Motivo do empréstimo	nominal
<i>amount</i>	Valor do crédito	intervalar
<i>savings</i>	Valor em poupanças e investimentos	ordinal
<i>employed</i>	Permanência no emprego atual	ordinal
<i>installp</i>	Taxa da prestação sobre a renda disponível (em %)	intervalar
<i>marital</i>	Estado civil e sexo	nominal
<i>coapp</i>	Outro fiador	nominal
<i>resident</i>	Data de início de permanência na atual residência	intervalar
<i>property</i>	Propriedade	nominal
<i>age</i>	Idade em anos	intervalar
<i>other</i>	Outros planos de crédito	nominal
<i>housing</i>	Forma de moradia	nominal
<i>exister</i>	Número de créditos existentes neste banco	intervalar
<i>job</i>	Situação empregatícia	ordinal
<i>depends</i>	Número de dependentes	intervalar
<i>telephon</i>	Possui telefone	binária
<i>foreign</i>	Trabalhador estrangeiro	binária
<i>good_bad</i>	Bom ou mau pagador	binária

A variável binária *good_bad* foi selecionada como a variável alvo (variável resposta ou dependente). As outras 20 variáveis no banco de dados servirão como variáveis de entrada para a rede (variáveis independentes ou explicativas) para a predição de um bom ou mau pagador. (Figura 7.1.1)

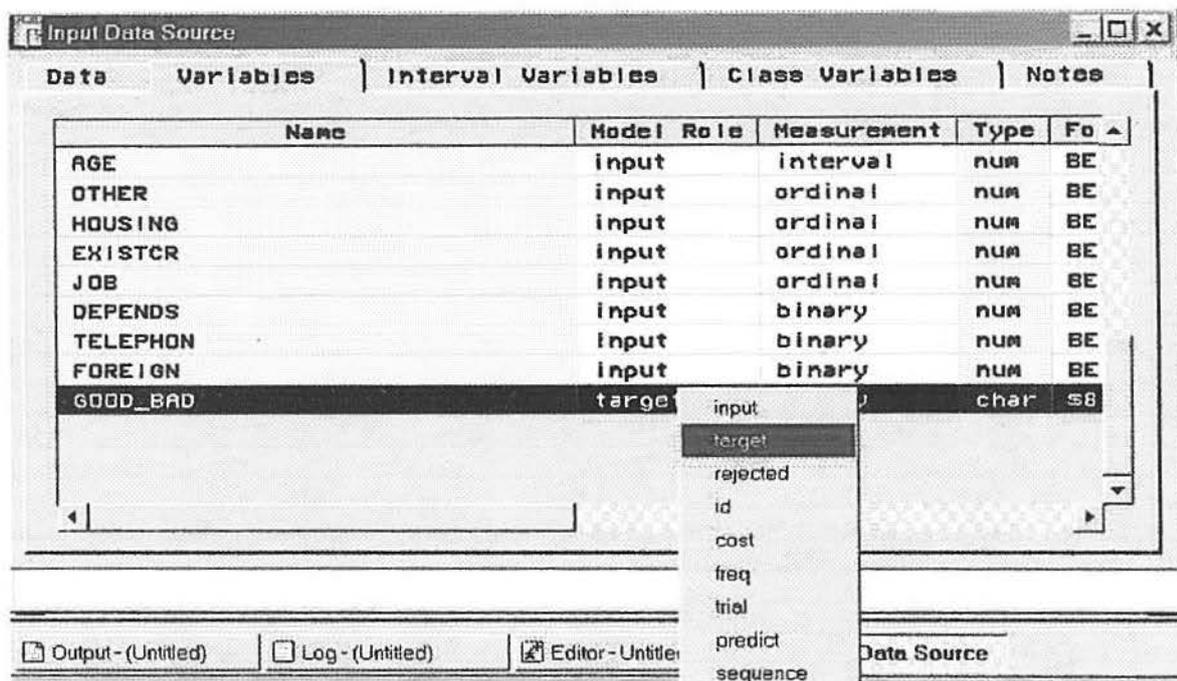


FIGURA 7.1.1: Seleção da variável *good_bad* como variável alvo.

Primeiramente, definiu-se a matriz de perda para a aceitação ou rejeição de um empréstimo. As linhas da matriz representam os valores alvos (bom ou mau pagador) e as colunas representam as decisões (aceitar ou rejeitar o empréstimo). Para esta matriz de perda foi determinado que o risco de aceitar um mau pagador é cinco vezes pior que o de rejeitar um bom pagador. Se a função perda fosse atribuída em perda monetária poderia-se dizer que ao aceitar o crédito de um mau pagador o banco perde 5 Reais e ao aceitar o crédito de um bom pagador o banco ganhará 1 Real. (Figura 7.1.2)

LEVEL	accept	reject
good	-1	0
bad	5	0

FIGURA 7.1.2: Definição da Matriz de Perda para aceitação de um empréstimo.

Cada nó da rede utilizará esta matriz de perda para calcular as perdas esperadas.

Com base no banco de dados tem-se que 70% dos clientes são bons pagadores e que 30% são maus pagadores. Porém, utilizando a informação de que esta amostra de clientes não representa a realidade dos empréstimos do banco e que na verdade estas proporções são de 90% e 10% de bons e maus pagadores, respectivamente, então modificou-se a probabilidade a priori de ser bom pagador para 0,9 e de ser mau pagador para 0,1 como mostrado na Figura 7.1.3.

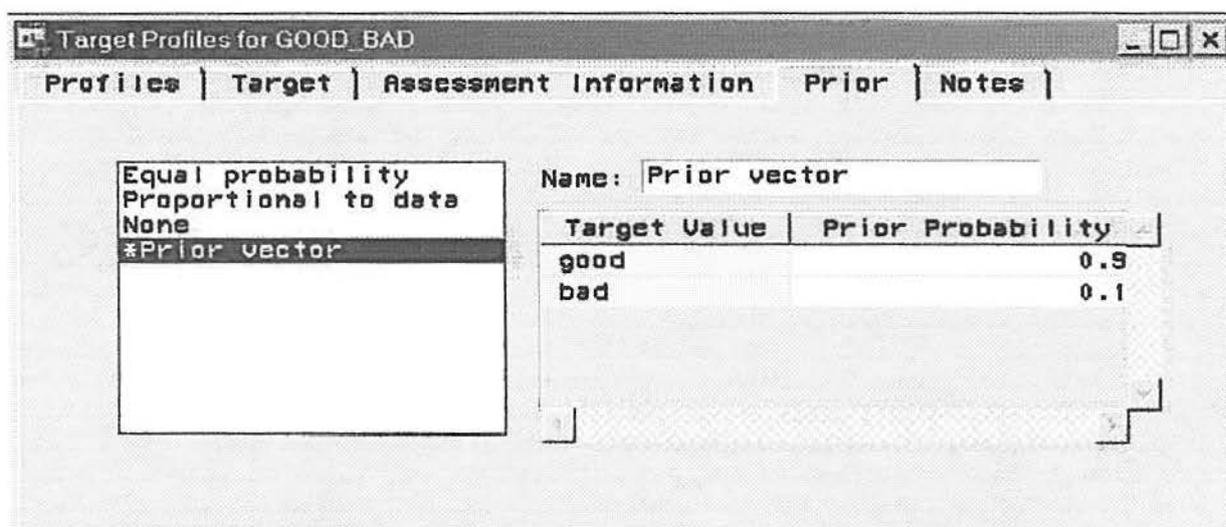


FIGURA 7.1.3: Definição da Probabilidade a Priori de ser bom e mau pagador.

As probabilidades a priori serão usadas para ajustar a contribuição relativa de cada classe ao calcular a perda média e total.

Nesta etapa da preparação dos dados, divide-se a base de dados em três amostras através do nó de partição de dados, sendo uma de treinamento, utilizada para ajustar o modelo de rede neural, outra de validação, para escolher o melhor modelo de rede para projetar novas disponibilizações de crédito baseada no modelo de perda mínima, e uma outra de teste para ajudar a verificar o poder de predição da rede criada e para fazer uma última avaliação do modelo escolhido. Neste caso, por existir apenas 1000 casos na base de dados, serão criadas apenas as amostras de treinamento e validação. Dividiu-se, então, em 60% dos dados para a amostra de treinamento e 40% para a amostra de validação. (Figura 7.1.4)

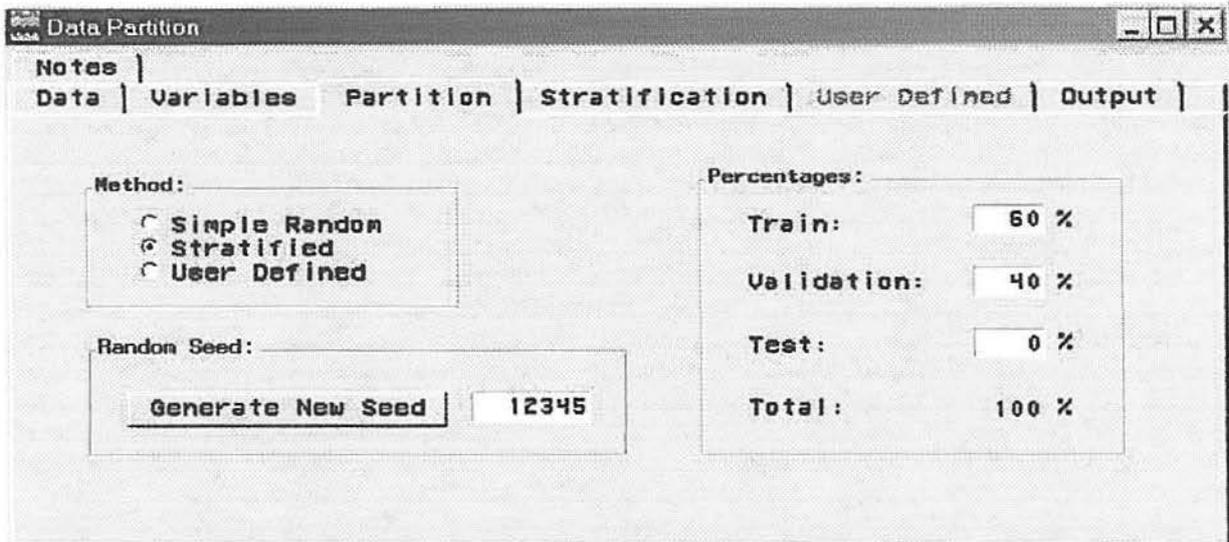


FIGURA 7.1.4: Divisão da base de dados em amostras de treinamento e validação.

O processo de amostragem para divisão da base de dados foi a amostragem estratificada, utilizando a variável alvo *good_bad* como variável de estratificação para garantir a mesma proporção de bons e maus pagadores nas duas amostras. (Figura 7.1.5)

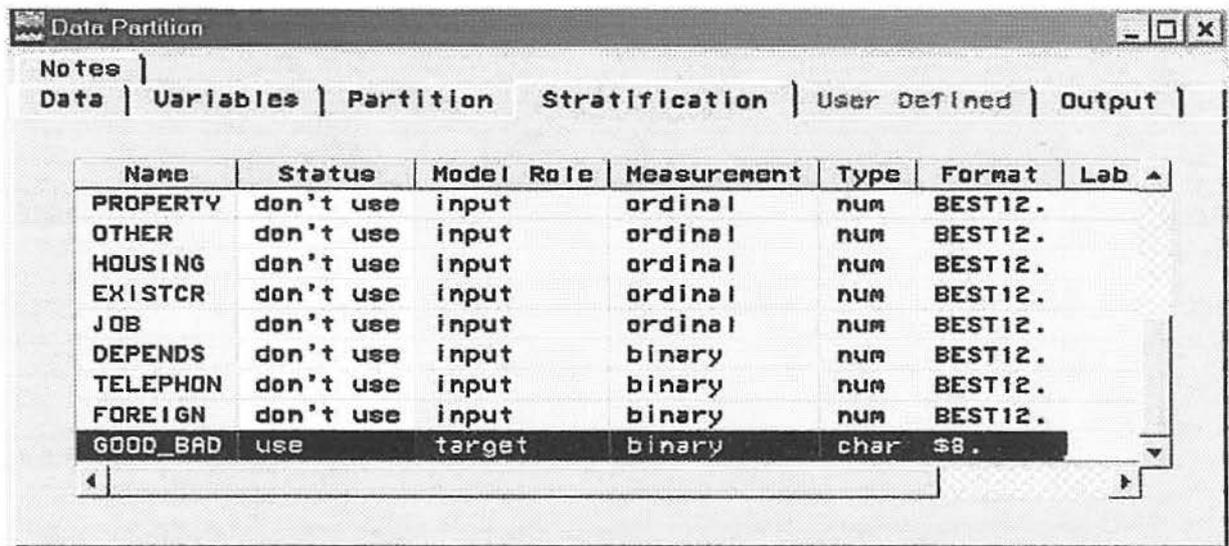


FIGURA 7.1.5: Escolha da variável *good_bad* como variável de estratificação.

O próximo passo é verificar a distribuição das variáveis. Variáveis com assimetria e curtose muito acentuadas devem ser transformadas para suavizar estes problemas. Esta transformação é feita através do nó de transformação de variáveis. Na Figura 7.1.6 é apresentada algumas estatísticas das variáveis intervalares. Pode-se observar a presença de uma forte assimetria na variável *amount*, pois o coeficiente de assimetria é 1,9, sendo que o valor deste coeficiente para uma distribuição simétrica é 0.

Name	Std Dev.	Missing %	Skewness	Kur
DURATION	12.059	0%	1.0942	***
AMOUNT	2822.7	0%	1.9496	***
AGE	11.375	0%	1.0207	***

FIGURA 7.1.6: Estatísticas descritivas das variáveis intervalares.

Para observar como está a distribuição dos dados da variável *amount* solicitou-se um histograma que é mostrado na Figura 7.1.7.

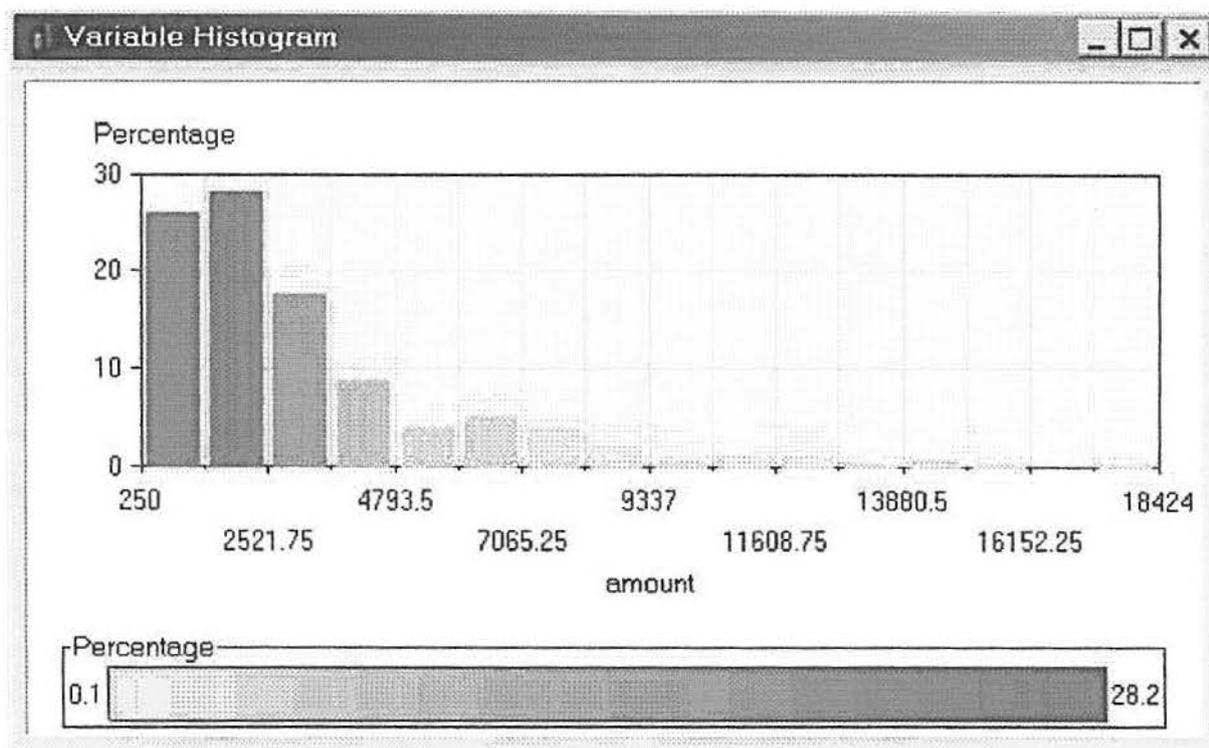


FIGURA 7.1.7: Histograma da variável *amount*.

Como pode-se observar existe mesmo uma grande assimetria dos dados da variável *amount*. Para tornar a distribuição mais simétrica aplicou-se uma transformação denominada *Maximize Normality* (maximizar a normalidade), obtendo-se a distribuição para os dados transformados apresentada na Figura 7.1.8.

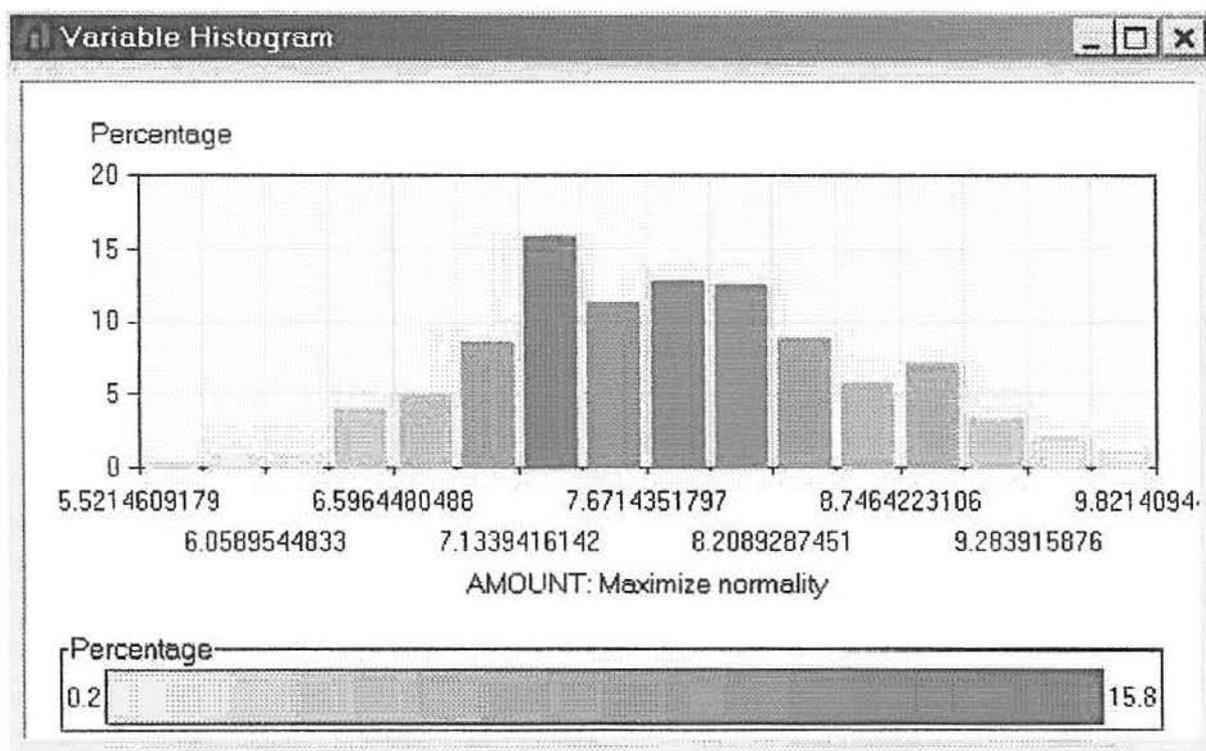


FIGURA 7.1.8: Histograma da variável transformada $\log(\text{amount})$.

Outra transformação interessante que pode ser utilizada é transformar variáveis intervalares em categórica ordinal como é o exemplo da variável *age*, onde o interesse é prever bons pagadores em grupos de idades. Criou-se assim 4 grupos da variável *age* através da transformação bucket (grupos). O histograma da variável *age* dividida em 4 categorias é apresentado na Figura 7.1.9.

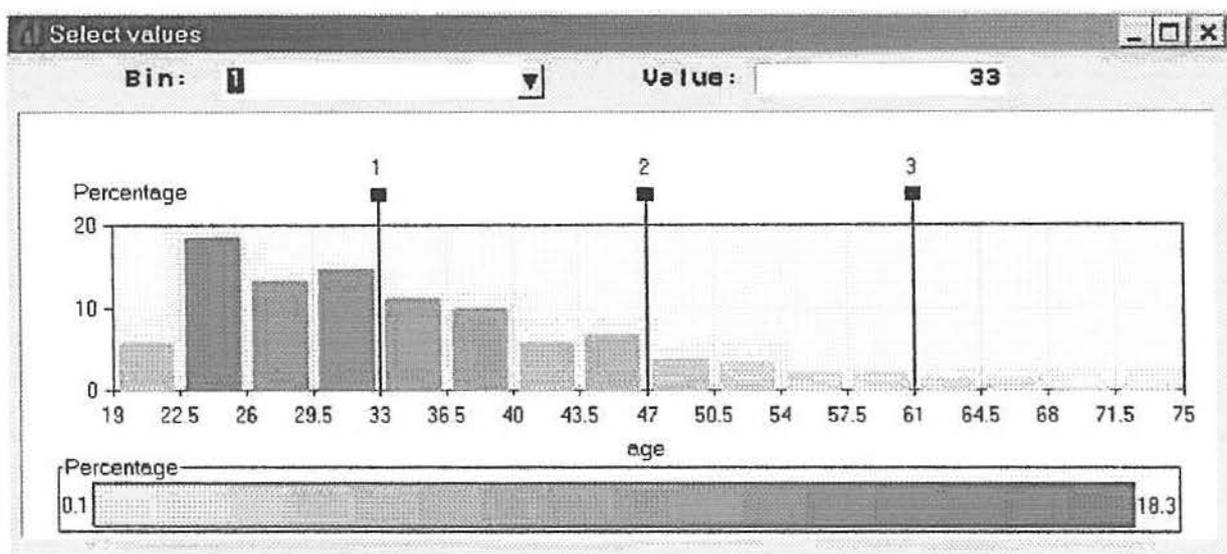


FIGURA 7.1.9: Histograma da variável *age* dividida em 4 categorias.

7.2. ANÁLISE E RESULTADOS

Para proceder a análise dos dados, primeiramente cria-se o nó de Redes Neurais, onde será ajustado o modelo de um Perceptron multi-camada. Escolheu-se a opção a interface avançada do módulo de redes neurais para que fosse possível a criação da rede perceptron com uma camada escondida. (Figura 7.2.1)

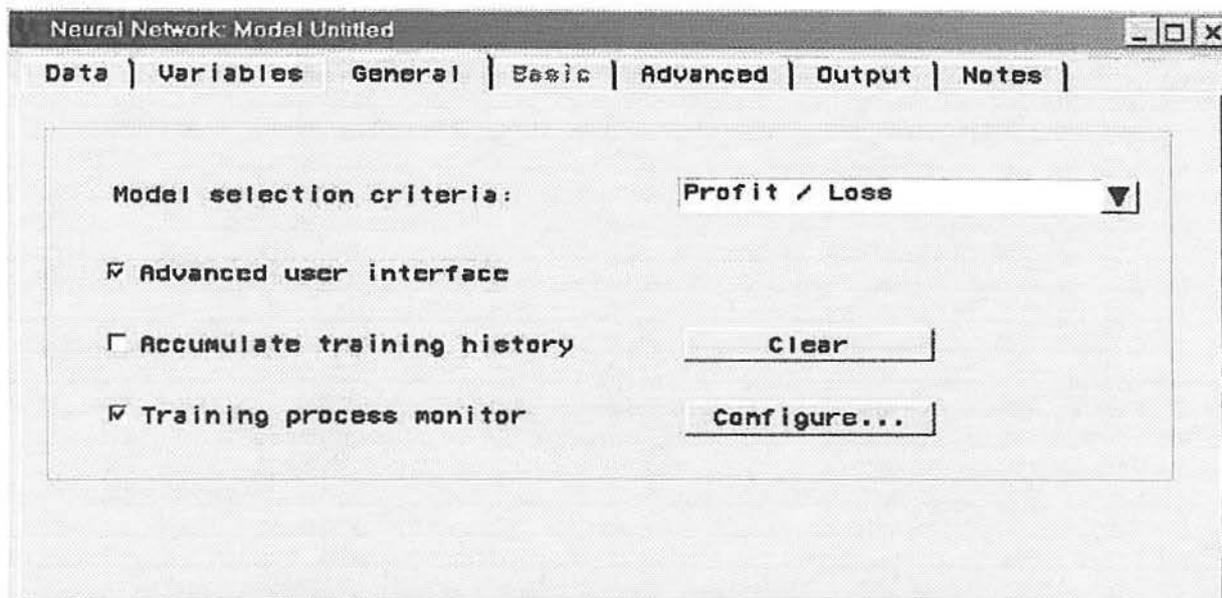


FIGURA 7.2.1: Seleção do modelo de Redes Neurais.

Logo após, começa-se a escolher o perfil da rede como o número de camada ocultas bem como o número de nós escondidos. Na Figura 7.2.2 pode-se observar a estrutura da rede que será criada. Os nós de entrada estão divididos pelo tipo de variável classificatória. O nó de saída é o resultado previsto para a variável alvo *good_bad*.

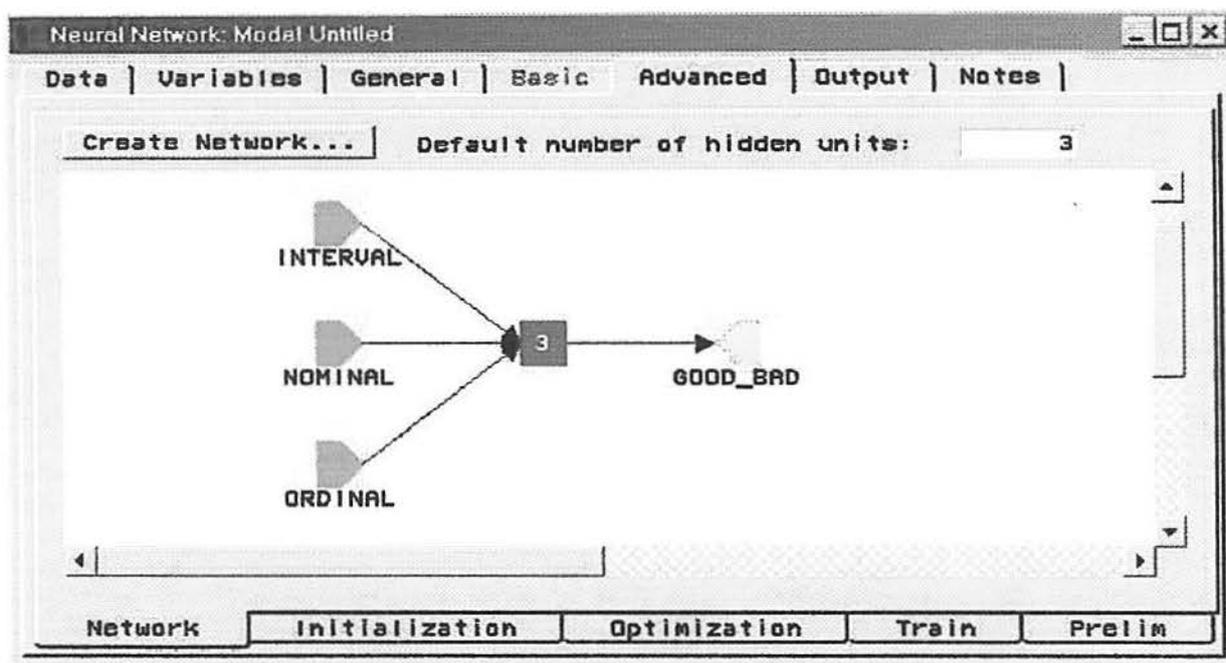


FIGURA 7.2.2: Arquitetura da rede com uma camada escondida composta de 3 nós.

Na opção *Create Network...* tem-se a arquitetura da rede, onde escolhe-se qual o modelo de rede neural a ser utilizado. Neste exemplo, utilizou-se uma rede perceptron multicamada. (Figura 7.2.3)

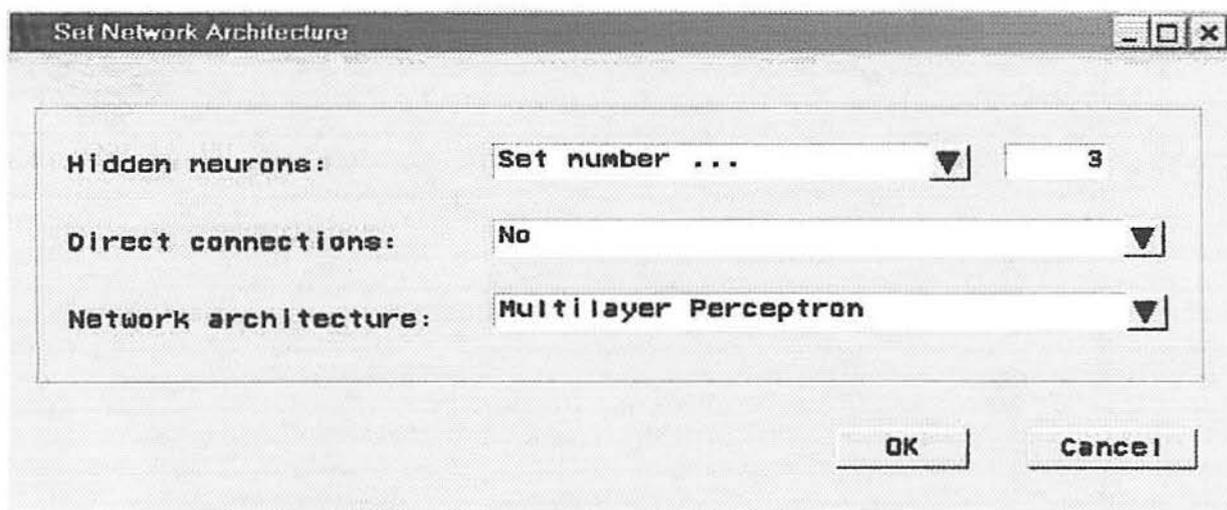


FIGURA 7.2.3: Escolha da arquitetura da rede neural.

No item *Hidden* do menu pode-se escolher, dentre várias opções, qual será a função de ativação e a função somatória, bem como se a rede terá o fator de bias ou não. (Figura 7.2.4)

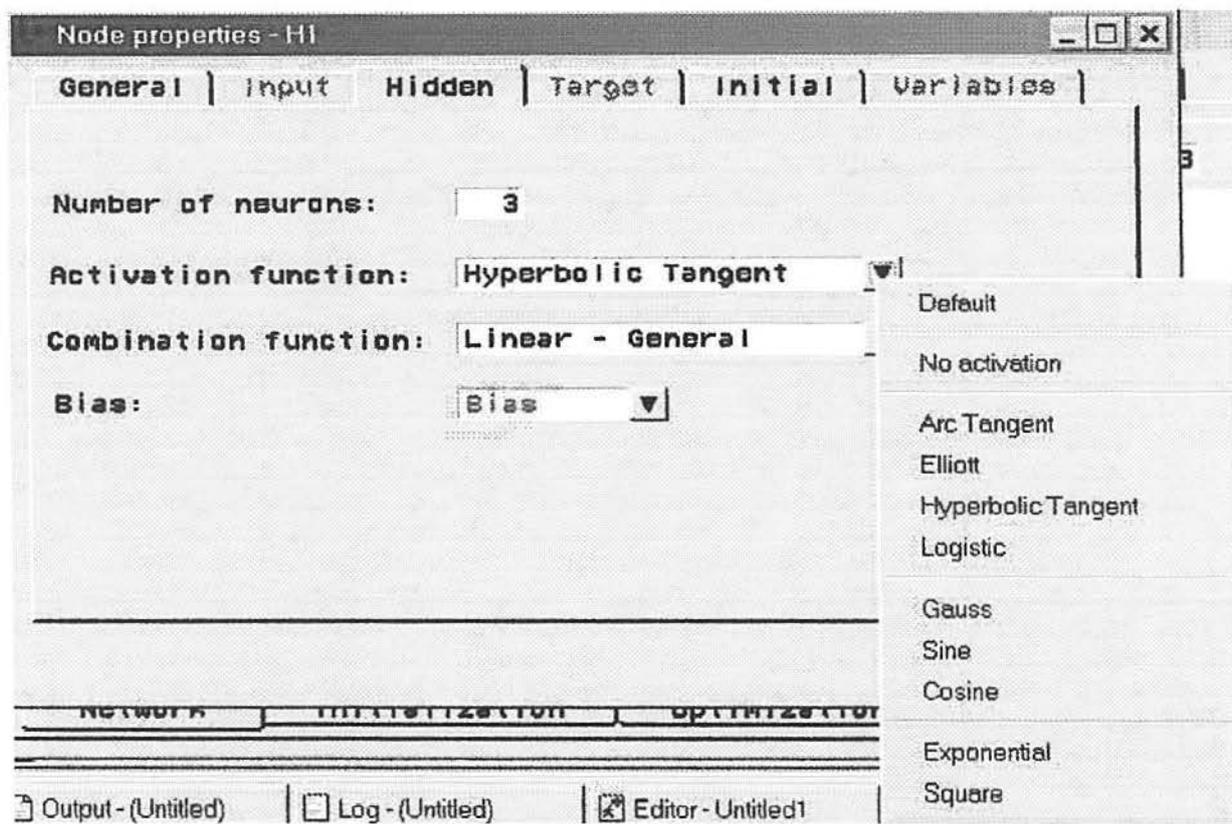


FIGURA 7.2.4: Determinação dos formatos dos componentes dos nós escondidos.

Por fim, salva-se o modelo de rede neural construído para logo após rodar e obter os resultados da análise. Dá-se o nome do modelo e a descrição como é mostrado na Figura 7.2.5.

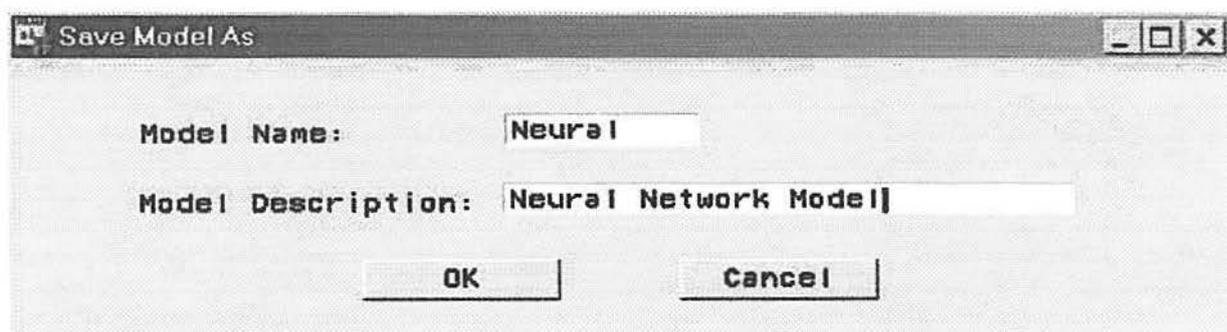


FIGURA 7.2.5: Nome e descrição do modelo de rede neural construído.

Na Figura 7.2.6 é apresentado a estrutura final do procedimento de Redes Neurais no Enterprise Miner com os nós utilizados na análise, restando apenas rodar o procedimento e avaliar os resultados obtidos.

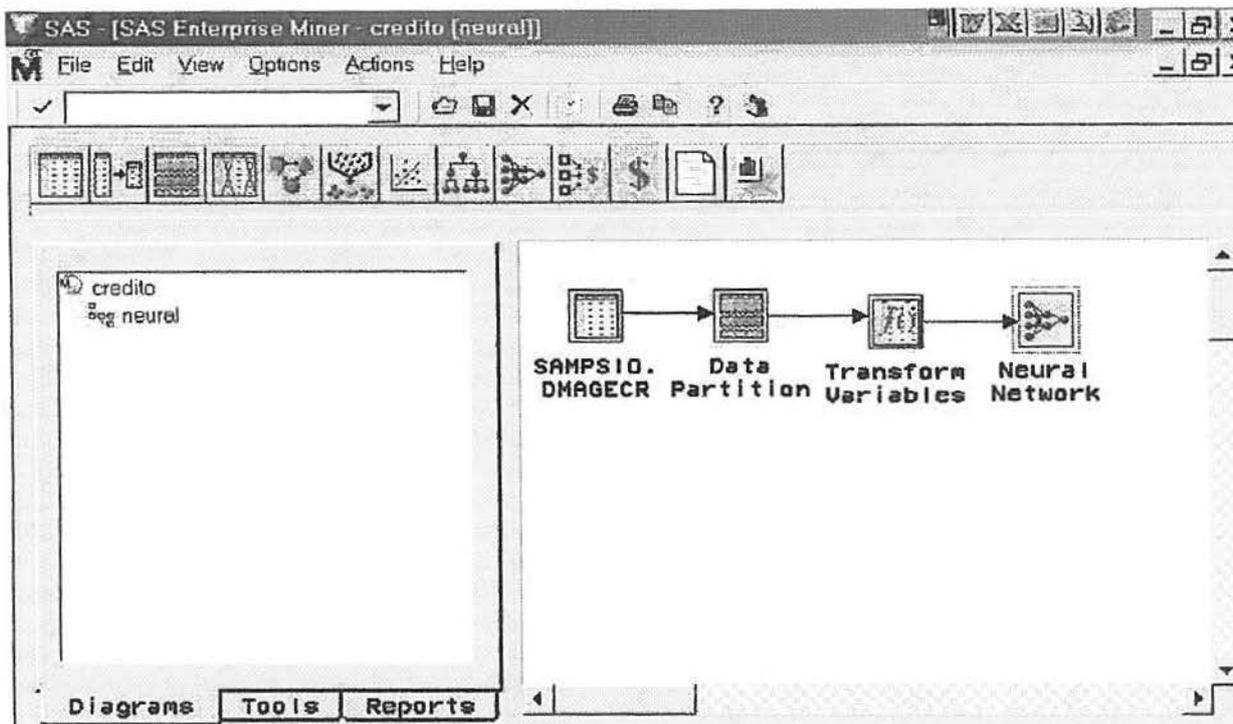


FIGURA 7.2.6: Nós do procedimento de Redes Neurais.

Enquanto a rede está rodando, aparece uma janela da evolução do erro até a otimização da rede para os dados de treinamento e validação que é mostrada na Figura 7.2.7.

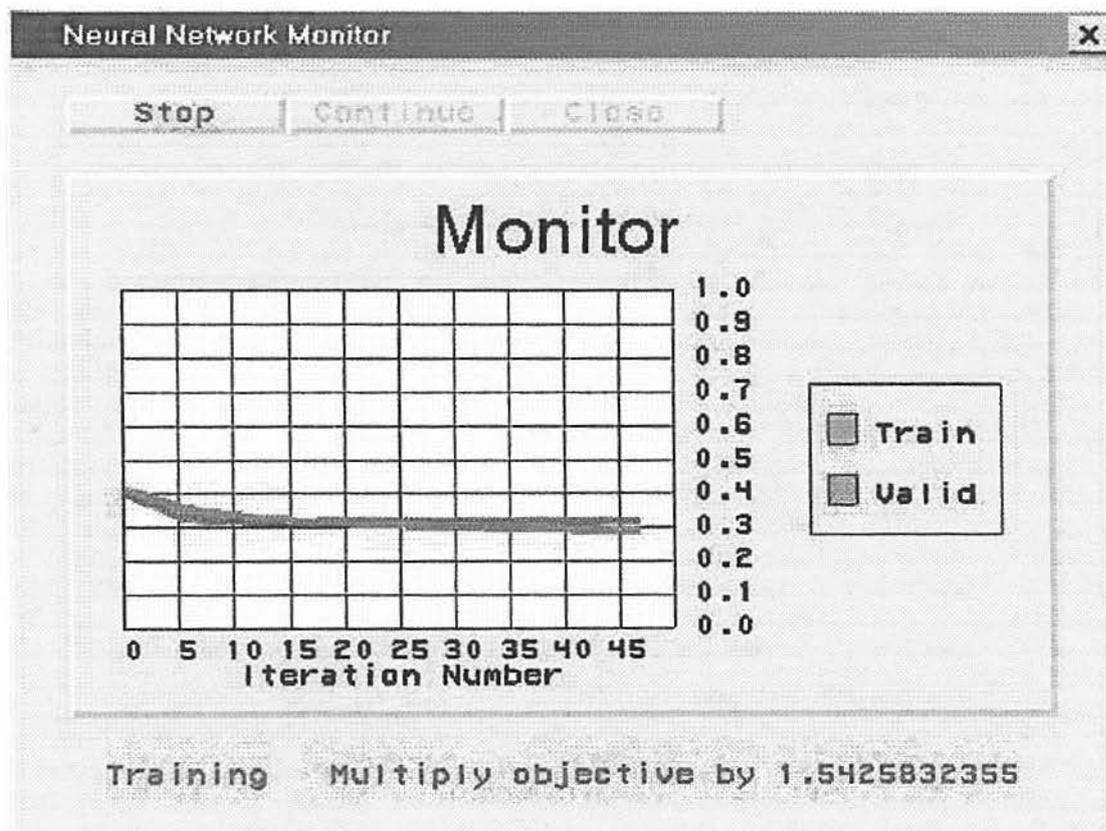


FIGURA 7.2.7: Monitoramento da evolução do erro das amostras de treinamento e validação.

A rede para quando tiver completado 100 iterações, porém basta pedir para que o processo continue para que o treinamento seja continuado. O primeiro resultado apresentado é gráfico com o erro médio para cada iteração dos dados de treinamento e validação que é mostrado na Figura 7.2.8. Para este exemplo, o ótimo para o erro médio foi alcançado na iteração 42. A rede é treinada até que um padrão seja reconhecido pela rede. É possível notar que as linhas do erro médio de treinamento e validação divergem após a 42ª iteração.

Cada vez que se roda a rede novamente, uma nova aleatorização dos dados é criada e utilizada para o treinamento da rede. Por este motivo, os resultados obtidos podem diferir a cada vez que a rede é treinada.

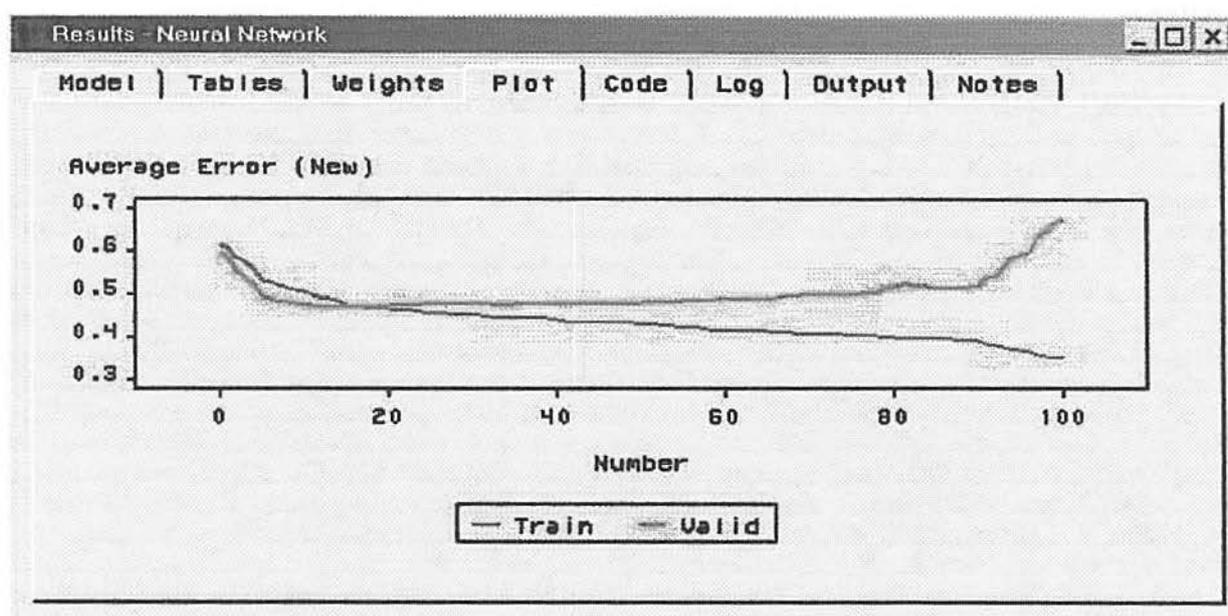


FIGURA 7.2.8: Erro médio de classificação para os dados de treinamento e validação.

A perda média para cada iteração para os dados de treinamento e validação é apresentada na Figura 7.2.9. Nota-se que a perda média é minimizada na 42ª iteração.

Na Figura 7.2.10 é apresentado valor obtido para a perda média para os dados de treinamento e validação na 42ª iteração. A perda média foi de 55 centavos para os dados de treinamento e de 59 centavos para os dados de validação. Esta perda média foi ajustada pela probabilidade a priori especificada no perfil da variável alvo *good_bad*.

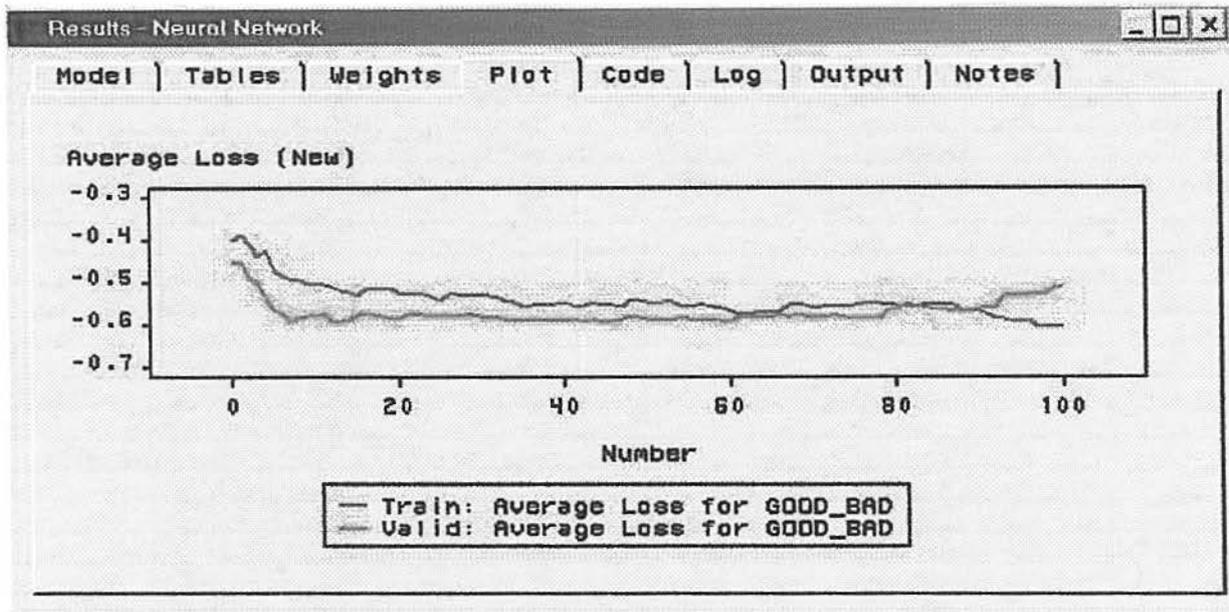


FIGURA 7.2.9: Perda média na predição de bom ou mau pagador para os dados de treinamento e validação.

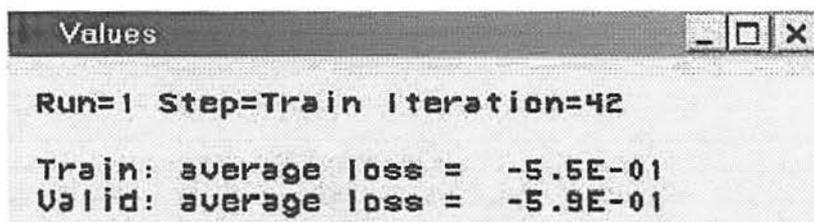


FIGURA 7.2.10: Valor obtido para a perda média na 42ª iteração.

Um último resultado obtido é apresentado na Tabela 7.2.1. Trata-se da tabela de cruzamento dos valores observados com os valores preditos para a variável alvo bom ou mau pagador (*good_bad*).

TABELA 7.2.1: Tabela cruzada de classificação: observado x predito

Observados	Preditos		
	Bom Pagador	Mau Pagador	Total
Bom Pagador	551	149	700
Mau Pagador	86	214	300
Total	637	363	1000

Com base nestes resultados pode-se concluir que a rede obteve um erro de classificação de 23,5% (235 casos classificados incorretamente em 1000 casos totais).

Analisando os resultados separadamente tem-se na Tabela 7.2.2 a classificação da rede para a amostra de treinamento e na Tabela 7.2.3 a classificação da rede para a amostra de validação.

TABELA 7.2.2: Tabela cruzada de classificação para amostra de treinamento

Observados	Decisão		
	Aceita	Rejeita	Total
Bom Pagador	309	92	401
Mau Pagador	47	133	180
Total	356	225	581

TABELA 7.2.3: Tabela cruzada de classificação para amostra de validação

Observados	Decisão		
	Aceita	Rejeita	Total
Bom Pagador	242	57	299
Mau Pagador	39	81	120
Total	281	138	419

É importante também lembrar que para este mesmo exemplo foi procedido a análise de regressão logística e a análise de árvore de decisão que juntamente com a metodologia de redes neurais, são técnicas utilizadas no procedimento de Data Mining. Para decidir qual das três técnicas conseguiu resolver o problema de classificação de maneira mais eficiente construiu-se um gráfico de comparação das três técnicas que é apresentado na Figura 7.3.4.

Os percentis do eixo dos X representam os casos sorteados da esquerda para direita pelos indivíduos que tem mais chance de ser predito como bom pagador no modelo. O eixo vertical representa a porcentagem de resposta corretamente classificada em cada percentil. Assim, comparando as linhas de porcentagem de resposta para cada técnica pode-se concluir que a metodologia de rede neural foi sempre superior a técnica de árvore de resposta e foi superior na maioria dos percentis à análise de regressão logística. Para conferir este resultado

construiu-se também o gráfico comparativo da perda esperada para cada uma das três técnicas que é apresentado na Figura 7.3.5.

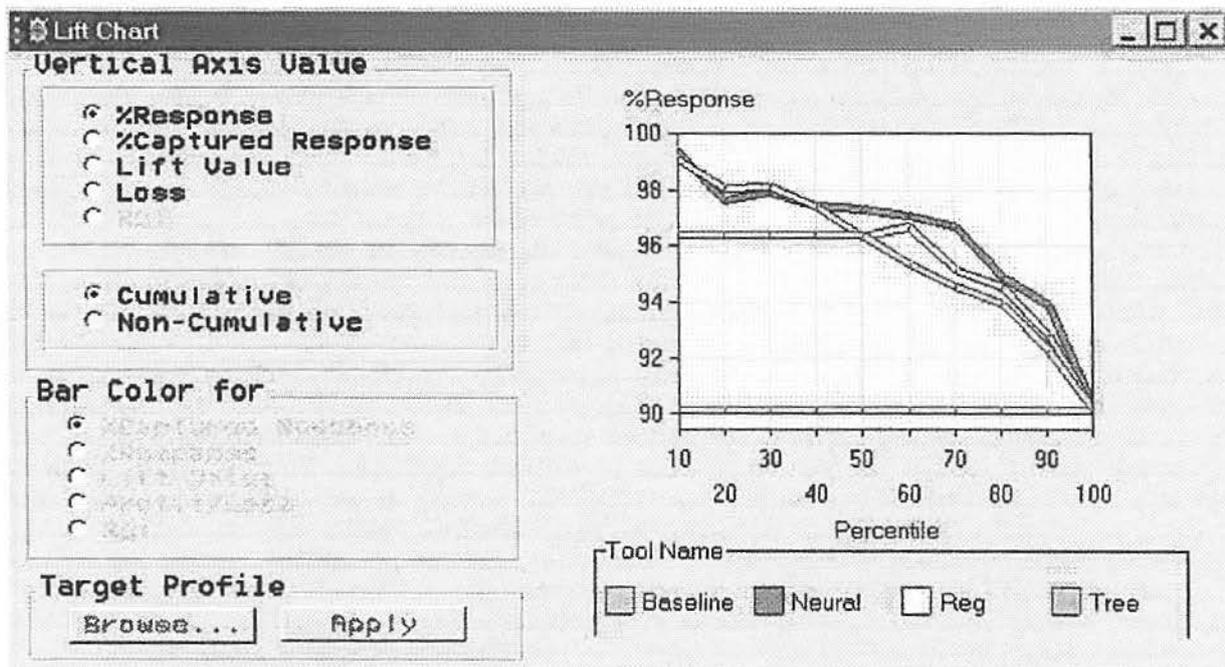


FIGURA 7.3.4: Comparação das três técnicas através da porcentagem de classificação correta.

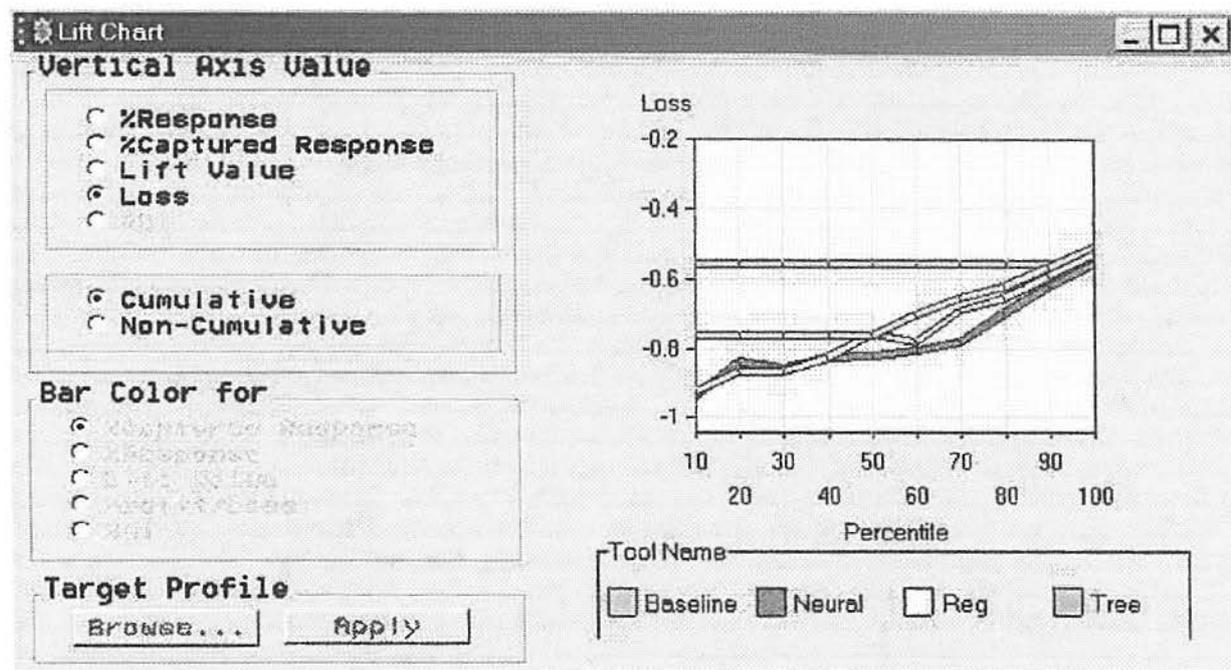


FIGURA 7.3.5: Comparação das três técnicas através da perda esperada com a classificação.

Com base no gráfico da Figura 7.3.5 pode-se observar que a perda esperada foi sempre menor para o modelo de redes neurais em comparação com a técnica de árvore de decisão e,

as redes neurais obtiveram resultados melhores na maioria dos percentis quando compara com a análise de regressão logística

Então, para este exemplo o modelo que melhor explicou os dados foi o de rede neural. É importante salientar que cabe ao estatístico responsável pela execução da análise do banco de dados saber utilizar as ferramentas disponíveis no Data Mining e escolher qual a que melhor soluciona o problema a ser resolvido.

7.3. INTEGRAÇÃO DA REDE

Este é o tópico mais interessante da análise por ser o mais prático do ponto de vista da aplicação em tomada de decisão no que diz respeito a aceitar ou não um pedido de crédito. A integração da rede será feita através de um novo banco de dados contendo 75 observações de novos clientes. Neste banco de dados existem apenas as 20 variáveis explicativas que serão utilizadas pela rede para prever a variável resposta bom ou mau pagador e, assim, tomar a decisão de aceitar ou rejeitar o fornecimento do crédito.

A Figura 7.3.1 apresenta a perda esperada na tomada de decisão de aceitar ou rejeitar uma concessão de crédito. Clientes que tem valores de perda esperada negativa representam um risco de crédito bom para o banco, ou seja, o banco vai lucrar com estes clientes. A barra vermelha representa os clientes com risco de mau crédito, ou seja, seus pedidos de concessão de crédito devem ser rejeitados ou o banco poderá perder dinheiro apostando neles.

Após a rede rodar, obteve-se 64% de clientes com pedido de crédito aceito e 36% de clientes com pedido de crédito rejeitado por terem grande risco de serem maus pagadores. (Figura 7.3.2)

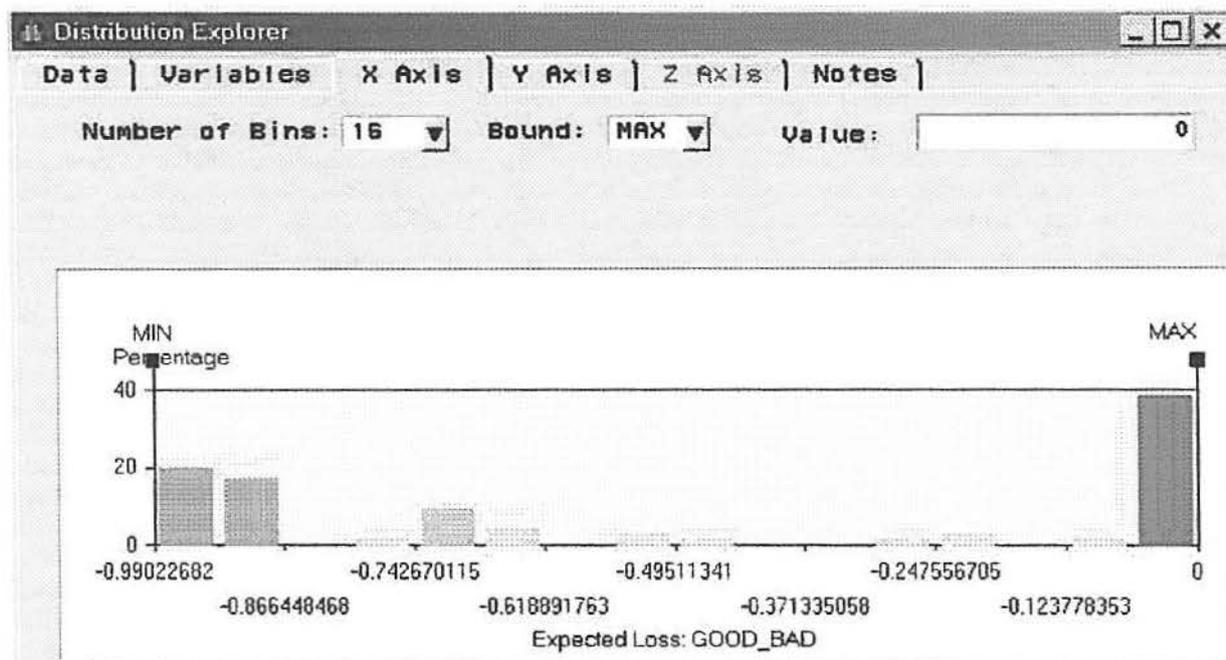


FIGURA 7.3.1: Gráfico da perda esperada pela concessão do crédito.

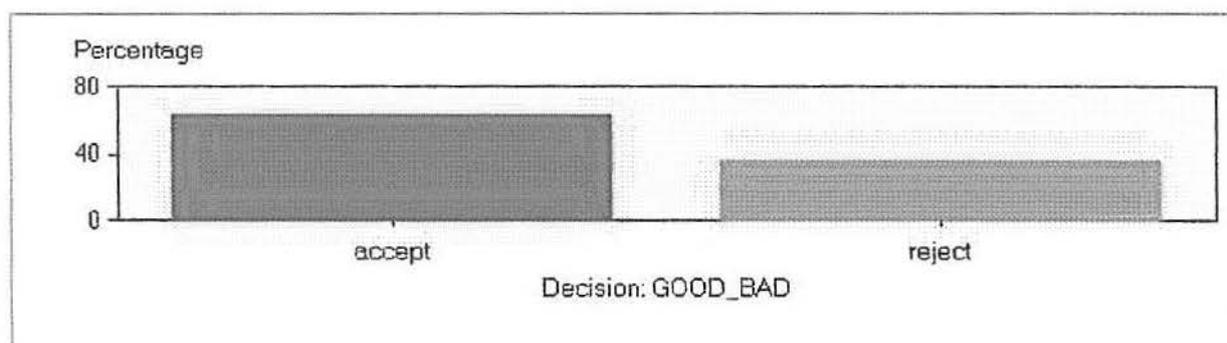


FIGURA 7.3.2: Gráfico da proporção de casos de crédito aceitos ou rejeitados.

Pode-se juntar as informações dos dois gráficos apresentados em um único gráfico tridimensional apresentado na Figura 7.3.3 onde pode-se ver claramente ao rejeitar um crédito o banco nem perde nem ganha e que quanto maior o número de créditos aceitos maior é o ganho esperada para o banco.

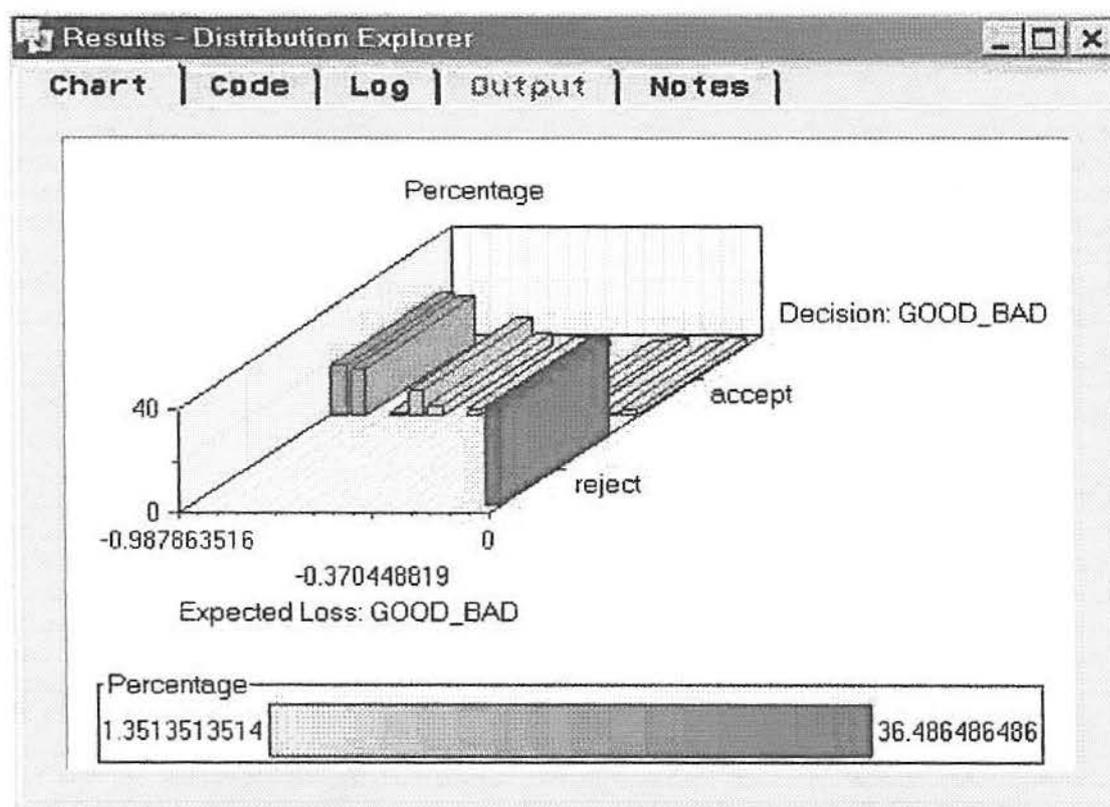


FIGURA 7.3.3: Gráfico da perda esperada pela decisão de aceitação ou rejeição do crédito.

Finalmente, após a análise dos resultados obtidos pela rede neural devidamente treinada e validada, pode-se selecionar então quais serão os clientes que terão seus pedidos de crédito aceitos. A Tabela 7.3.1 fornece a informação da decisão de concessão de crédito para cada um dos 75 clientes, bem como, a perda esperada que o banco terá com cada um deles.

Com esta tabela de resultados preditos pela rede, pode-se observar que a perda é zero quando a decisão é rejeitar o crédito e que quando toma-se a decisão de aceitar e conceder o crédito a perda esperada é negativa, mostrando que o banco só espera lucrar ao conceder o financiamento àqueles clientes que a rede classificou como bom pagador.

TABELA 7.3.1: Decisão de aceitação do crédito e perda esperada.

Código do Cliente	Decisão Esperada	Perda Esperada	Código do Cliente	Decisão Esperada	Perda Esperada
325	accept	-0.69061	921	accept	-0.73290
858	accept	-0.91849	65	accept	-0.87344
151	accept	-0.97277	694	accept	-0.01491
348	accept	-0.21071	844	accept	-0.29493
212	accept	-0.98786	211	accept	-0.98641
990	accept	-0.70823	535	accept	-0.89320
166	accept	-0.92598	7	accept	-0.92564
952	accept	-0.06478	518	accept	-0.64279
495	accept	-0.65280	759	accept	-0.95903
563	accept	-0.00845	446	accept	-0.95072
894	accept	-0.70598	649	reject	0
454	accept	-0.94973	288	reject	0
516	accept	-0.90892	303	reject	0
703	accept	-0.78107	783	reject	0
120	accept	-0.51265	938	reject	0
312	accept	-0.21902	897	reject	0
998	accept	-0.95345	814	reject	0
246	accept	-0.88966	32	reject	0
160	accept	-0.99023	375	reject	0
378	accept	-0.95294	51	reject	0
934	accept	-0.96494	640	reject	0
588	accept	-0.71063	970	reject	0
768	accept	-0.96530	384	reject	0
633	accept	-0.64683	705	reject	0
207	accept	-0.96726	667	reject	0
95	accept	-0.68433	497	reject	0
621	accept	-0.47460	935	reject	0
148	accept	-0.91970	877	reject	0
178	accept	-0.91246	482	reject	0
40	accept	-0.52271	107	reject	0
135	accept	-0.89882	826	reject	0
28	accept	-0.73816	305	reject	0
137	accept	-0.96617	721	reject	0
359	accept	-0.94285	467	reject	0
682	accept	-0.96335	690	reject	0
84	accept	-0.87590	388	reject	0
774	accept	-0.92554	594	reject	0
244	accept	-0.90319			

8. CONCLUSÃO

Este trabalho descreveu os principais tópicos referentes à redes neurais, desde seu início até propostas de implementação em diversas áreas de aplicação. No referencial histórico detalhou-se desde o surgimento, propostas iniciais de redes neurais, conflitos e progressos neste tema. Foi descrito o funcionamento de células humanas, na qual se baseou todo o funcionamento das unidades componentes das redes neurais. Seguiu-se com um quadro comparativo entre o cérebro humano e o computador, e entre máquinas tradicionais e máquinas baseadas no conceito de redes neurais. Devido a grande aplicabilidade pode-se entender os diversos motivos que levam inúmeros estudiosos a trabalhar e desenvolver redes neurais.

Descreveu-se ainda, na Arquitetura da Rede, aspectos referentes à descrição destas a partir de suas características de desenvolvimento e aplicação. Mostrou-se alguns conceitos e teorias que descrevem o aprendizado de uma rede neural juntamente com uma coletânea de aplicações das mesmas.

Todas as informações aqui expostas levam a crer que o campo de redes neurais artificiais é acima de tudo extremamente vasto e promissor, necessitando do conhecimento estatístico para auxiliar na modelagem da arquitetura e na escolha do algoritmo de aprendizagem a ser utilizado. Trata-se de um assunto recente, mas que adquiriu muita credibilidade, e devido a constantes descobertas relacionadas a Redes Neurais, tornou-se um campo bastante atrativo para profissionais de domínios distintos, reforçando sua característica multidisciplinar.

Acredita-se que o mais coerente é entender Redes Neurais Artificiais como ferramenta estatística, sendo uma opção de solução de problemas e não metodologia competidora.

Mostrou-se também que pensar uma rede neural como uma caixa preta é um mito que deve ser desmistificado, pois viu-se que existe uma metodologia para obtenção dos pesos da rede, exemplificado no trabalho através do algoritmo de retropropagação.

Por fim, acredita-se que este trabalho tem consistência e pode ser um bom referencial, tanto teórico quanto aplicado, para aqueles que quiserem uma introdução no estudo da metodologia de Redes Neurais Artificiais, metodologia esta extremamente importante e útil para análise de dados das diferentes áreas do conhecimento.

9. REFERÊNCIAS BIBLIOGRÁFICAS

- ALENCAR, R.J.N. (1997). *Avaliação da Segurança Estática de Sistemas de Energia Elétrica usando Mapeamento Auto-organizável*. (Mestrado – Centro Tecnológico - Engenharia Elétrica / UFPA), 123p.
- ALMEIDA, F.C. (1993). *L'Evaluation des risques de défaillance des entreprises à partir des réseaux de neurones insérés dans les systèmes d'aide à la décision*. Grenoble: Universidade de Grenoble, Ecole Supérieure des Affaires. (Doutorado - Ciências de Administração).
- ARBIB, M.A. (1987). *Brains, Machines, and Mathematics*. 2ªed. New York: Springer-Verlag.
- BAXT, W.G. (1990). Use of an Artificial Neural Network for Data Analysis in Clinical Decision-Making: The Diagnosis of Acute Coronary Occlusion. *Neural Computation*, vol.2, p.480-489.
- BAXT, W.G. (1991). Use of an Artificial Neural Network for the Diagnosis of Myocardial Infarction. Occlusion. *Annals of Internal Medicine*, vol.115, p.843-848.
- BISHOP, C.M. (1995). *Neural Networks for Pattern Recognition*. New York: Oxford University Press.
- BUNTINE, W.L. & WEIGEND, A.S. (1991). Bayesian Back-Propagation. *Complex Systems*, vol.5, p.603-643.
- CAIRO, L.N.Jr & YONEYAMA, T. (1997). *Inteligência Artificial em Automação e Controle*. Edgar Blucher. São Paulo, S.P.
- CHURCH, K.B. & CURRAM, S.P. (1996). Forecasting consumer's expenditure: a comparison between econometric and neural network models. *International Journal of Forecasting*, vol.12, p.255-267.

- DINIZ, C.A. & LOUZADA-NETO, F. (2000). Data Mining: Uma Introdução. *XIV Simpósio Nacional de Probabilidade e Estatística*. Caxambú, M.G.
- DUDA, R.O. & HART, P.E. (1973). *Patterns Classification and Scene Analysis*. New York: John Wiley.
- EDELMAN, G.M. (1987). *Neural Darwinism*. New York: Basil Books.
- FELDMAN, J.A. (1992). Natural computation and artificial intelligence. *Plenary Lecture presented at the International Joint Conference on Neural Networks*, Baltimore.
- FERNANDES, L.G.L., PORTUGAL, M.S. & NAVAUX, P.O.A. (1994). Um Estudo Experimental do Poder Preditivo das Redes Neurais Artificiais Comparado a Métodos Econométricos Tradicionais. *I Simpósio Brasileiro de Redes Neurais*. Caxambu – MG, p.139-144.
- FERNANDES, L.G.L., PORTUGAL, M.S. & NAVAUX, P.O.A. (1996). O Problema da Escolha da Topologia da Rede Neural na Previsão de Séries Temporais. *III Simpósio Brasileiro de Redes Neurais*. Recife – PE, p.227-234.
- FISCHLER, M.A. & FIRSCHEIN, O. (1987). *Intelligence: The Eye, The Brain, and The Computer*. Reading, MA: Addison-Wesley.
- FISHER, R.A. (1938). The statistical utilization of multiple measurements. *Annals of Eugenics*, 8, p.376-388.
- FLURY, B. & RIEDWYL, H. (1990). *Multivariate Statistics: A Practical Approach*, London: Chapman & Hall.
- FREEMAN, J.A. (1994). *Simulating Neural Networks with Mathematica*. Reading, MA: Addison-Wesley.
- FUJITA, H., KATAFUCHI, T., UEHARA, T. & NISHIMURA, T. (1992). Application of Artificial Neural Network to Computer-Aided Diagnosis of Coronary Artery Disease in Myocardial Spect Bull's-Eye Images. *Journal of Nuclear Medicine*, vol.33, nº2, p.272-276.

- GISMONDI, R.C., INFANTOSI, A.F.C. & ALMEIDA, R.M.V.R. (1999). Modelagem da Mortalidade Infantil em Municípios Brasileiros por Redes Neurais e Regressão Linear Múltipla.
- GORMAN, R.P. & SEJNOWSKI, T.J. (1998). Analysis of Hidden Units in a Layered Network to Classify Sonar Targets. *Neural Networks*, vol.1, p.75-89.
- HAIR, J.F., ANDERSON, R.E., TATHAM, R.L. & BLACK, W.C. (1998). *Multivariate Data Analysis*. 5ª ed. New Jersey. Prentice Hall, 730p.
- HARTMANN, F.R. (1996). *Redes Neurais Artificiais: Conceitos Básicos e Análise*. São Leopoldo, R.S. (Monografia apresentada para obtenção do grau de Bacharel em Análise de Sistemas/UNISINOS), 90p.
- HARVEY, A.C. (1989). *Forecasting Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge, 458p.
- HAYKIN, S. (2000). *Redes neurais: princípios e prática*. Trad. Paulo Martins Engel. 2ªed. Porto Alegre. Bookman, 900p.
- HAWLEY, D., JOHNSON, J.D. & RAINA, D. (1990). Artificial Neural Systems: A New Tool for financial Decision-Making. *Financial Analyst Journal*, p.63-72.
- HECHT-NIELSEN, R. (1990). *Application of Counterpropagation Networks*. Neural Network, vol.1.
- HEBB, P. (1944) *Organization of Behavior*. New York. John Willey & Sons.
- HERTZ, J., KROGH, A., & PALMER, R.G. (1991). *Introduction to the Theory of Neural Computation*. Santa Fe Institute Studies in the Sciences of Complexity (vol.1), Redwood City, CA: Addison-Wesley.
- HOSMER, D.W. & LEMESHOW, S. (1989). *Applied Logistic Regression*. Nova York: Wiley.

- HRUSCHKA, H. (1993). Determining Market Response Functions by Neural Network modeling: A Comparison to Econometric Techniques. *European Journal of Operational Research*, vol.66, p.27-35.
- HUTCHINSON, J.M. (1994). *A Radial Basis Function Approach to Financial Time Series Analysis*, Ph.D. dissertation, Massachusetts Institute of Technology.
- KERLIRZIN, P. & VALLET, F. (1993). Robustness in multilayer perceptrons. *Neural Computation*, vol.5, p.473-482.
- KOVÁCS, Z.L. (1996). *Redes Neurais Artificiais: Fundamentos e Aplicações*. Edição Acadêmica. São Paulo, S.P, 163p.
- KROSE, B. J. & SMAGT, P.V. (1993). *An Introduction to Neural Networks*. University of Amsterdam.
- LINSKER, R. (1988). Self-organization in a perceptual network. *Computer*, vol.21, p.105-117.
- LIPPMANN, R.P. (1987). An Introduction to Computing with Neural Nets. *IEEE ASSP Magazine*, p.4-22.
- LIPPMANN, R.P. (1989). Review of neural networks for speech recognition. *Neural Computation*, vol.1, p.1-38.
- McCUULLOCH, W.S. & PITTS, W. (1943). A logical Calculus of the Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics*, vol.5, p.115-133.
- MEAD, C.A. (1989). *Analog VLSI and Neural Systems*. Reading, MA: Addison-Wesley.
- MENDEL, J.M. & McCLAREN, R.W. (1970). Reinforcement-learning control and pattern recognition systems. in *Adaptive, Learning, and Pattern Recognition Systems: Theory and Applications*, vol.66, p.287-318, New York: Academic Press.
- MINSKY, M.L. & PAPERT, S.A. (1969). *Perceptrons*. Cambridge, MA: MIT Press.
- NETER, J., WASSERMAN, W. & KUTNER, M.H. (1990). *Applied Linear Statistical Models*. Homewood, IL: Richard D. Irwin.

- OSÓRIO, F.S. (1991). *Um Estudo sobre Reconhecimento Visual de Caracteres Através de Redes Neurais*. Porto Alegre: CPGCC. (Mestrado – Instituto de Informática/UNISINOS).
- PACHECO, M.L.C., NADAL, J. & PANERAI, R.B. (1996). *Redes Neurais Artificiais Na Avaliação de Múltiplas Tecnologias de Saúde*.
- PEREIRA, B.B. & RODRIGUES, C.V.S. (1998). *Redes Neurais em Estatística. XIII Simpósio Nacional de Probabilidade e Estatística*. Caxambú, M.G.
- POLI, R., CAGNONI, S., LIVI, R., COPPINI, G. & VALLI, G. (1991). A Neural Network Expert System for Diagnosing and Treating Hypertension. *Computer*, p.64-71.
- QIAN, N. & SEJNOWSKI, T.J. (1988). Predicting the Secondary Structure of Globular Proteins Using Neural Networks Models. *Journal of Molecular Biology*, vol.202, p.865-884.
- REEKE, G.N.Jr., FINKEL, L.H. & EDELMAN, G.M. (1990). Selective recognition automata. in *An Introduction to Neural and Electronic Network*, S.F. Zornetzer, J.L. Davis, and C. Lau, p.203-226, New York: Academic Press.
- RIPLEY, B.D. (1993). Statistical Aspects of Neural Networks. In *Networks and Chaos – Statistical and Probabilistic Aspects*, eds. O. Barndorff-Nielsen, J.Jensen, and W.Kendall, London: Chapman & Hall, p.40-123.
- RIPLEY, B.D. (1994). Neural Networks and Related Methods for Classification. *Journal of the Royal Statistical Society B*, vol.56, nº3, p.409-456.
- ROSENBLATT, F. (1958). The Perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, vol.65, p.386-408.
- ROSENBLATT, F. (1962). *Principles of Neurodynamics*, Washington, DC: Spartan Books.
- RUMELHART, D.E., HINTON, G.E. & WILLIAMS, R.J. (1986). Learning representations of back-propagation errors. *Nature (London)*, vol.323, p.533-536.
- RUMELHART, D.E., & WEIGEND, S.A. (1990). *Predicting the Future: A Connectionist Approach*. Stanford PDP-90-01, PARC-SSL-90-20.

- SÁ, C.B.P. & NOBRE, F.F. (1996). Previsão de Séries Temporais Epidemiológicas Utilizando Redes Neurais.
- SAGE, A.P., ed (1990). *Concise Encyclopedia of Information Processing in Systems and Organizations*. New York: Pergamon.
- SARLE, W.S. (1994). Neural Networks and Statistical Methods. In *Proceedings of the 19th Annual SAS Users Group International Conference*.
- SMITH, M. (1993). *Neural Networks for Statistical Modeling*. New York: Van Nostrand Reinhold.
- STUDENMUND, A.H. (1992). *Using Econometrics: A Practical Guide*. New York: Harper Collins.
- TESAURO, G. (1990). Neurogammon Wins Computer Olympiad. *Neural Computation*, vol.1, p.321-323.
- TSOUKALAS, L.H., UHRING, R.E., WILEY, J. & SONS. (1996). *Fuzzy and Neural Approaches in Engeneering*.
- VALIANT, L.G. (1984). A theory of the learnable. *Communications of the Association for Computing Machinery*, vol.27, p.1134-1142.
- VAPNIK, V.N. (1992). Principles of risk minimization for learning theory. *Advances in Neural Information Processing Systems*, vol.4, p.831-838, San Mateo, CA: Morgan Kaufmann.
- VAPNIK, V.N. (1998). *Statistical Learning Theory*. New York: Wiley.
- WALTZ, D. (1997). Neural nets and AI: Time for a synthesis. plenary talk. *International Conference on Neural Networks*, vol.1, p.xiii, Houston.
- WARNER, B. & MISRA, M. (1996). Understanding Neural Network as Statistical Tools. *The Americam Statistician*, vol.50, nº4, p.284-293.

- WERBOS, P.J. (1991). Links Between Artificial Neural Networks (ANN) and Statistical Pattern Recognition. In *Artificial Neural Networks and Statistical Pattern Recognition: Old and New Connections*, eds.I. Sethi and ^a Jain, Elsevier Science, p.11-31.
- WHITE, H. (1989). Learning in artificial neural networks: A statistical perspective. *Neural Computation*, vol.1, p.425-464.
- WIDROW, B. & HOFF, M.E. (1960). Adaptive switching circuits. *IRE WESCON Convention Record*, p.96-104.
- WU, F.Y. & YEN, K.K. (1992). Applications of Neural Network in Regression Analysis. In *Proceedings of the 14th Annual Conference on Computers and Industrial Engineering*.