

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE INFORMÁTICA  
CURSO DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**Utilização de Redes  
Neurais na Análise e Previsão  
de Séries Temporais**

por

LUIZ GUSTAVO LEÃO FERNANDES

Dissertação submetida à avaliação, como  
requisito parcial, para a obtenção do grau  
de Mestre em Ciência da Computação

Prof. Philippe O. A. Navaux  
Orientador

Marcelo Savino Portugal  
Co-orientador

Porto Alegre, maio de 1995.



UFRGS

SABI



05231379

**UFRGS**  
**INSTITUTO DE INFORMÁTICA**  
**BIBLIOTECA**

## CIP - CATALOGAÇÃO NA PUBLICAÇÃO

Fernandes, Luiz Gustavo Leão.

Utilização de redes neurais na análise e previsão de séries temporais /  
Luiz Gustavo Leão Fernandes. - Porto Alegre: CPGCC da UFRGS, 1995.  
76p.: il.

Dissertação (mestrado) - Universidade Federal do Rio Grande do Sul.  
Curso de Pós-Graduação em Ciência da Computação, Porto Alegre, 1995.  
Orientador: Navaux, Philippe O. A.

1. Redes Neurais. 2. *Backpropagation* 3. Prognósticos econômicos.  
4. Séries temporais. I. Navaux, Philippe O. A. II. Título.

UFRGS INSTITUTO DE INFORMÁTICA BIBLIOTECA			
N.º CHAMADA 681.3.011(043) F363U		N.º REG.: 32421	
ORIGEM: D		DATA: 24/07/96	PREÇO: R\$ 25,00
FUNDO: II	ORIGEM: D	DATA: 19/07/96	PREÇO: R\$ 25,00
FUNDO: II	FORN.: II		

Inteligência artificial. SBU/II  
Redes neurais  
Series temporais  
Backpropagation

ENPq 1.03.01 00-3

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
Reitor: Prof. Hélgio Trindade  
Pró-Reitor de Pesquisa e Pós-Graduação: Prof. Cláudio Scherer  
Diretor do Instituto de Informática: Prof. Roberto Tom Price  
Coordenador do CPGCC: Prof. José Palazzo Moreira de Oliveira  
Bibliotecária-Chefe do Instituto de Informática: Zita Prates de Oliveira

## AGRADECIMENTOS

Uma dissertação de mestrado é certamente um grande desafio a qualquer pessoa que pretende realizá-la. Comigo não foi diferente. Não tanto pelas dificuldades técnicas, mas pelas dificuldades relacionadas ao aspecto humano. Escrever uma dissertação de mestrado requer antes de tudo tenacidade, espírito de luta e fé no ideal inicial. Muitos obstáculos aparecem pelo caminho e é preciso acreditar no sonho para seguir em frente e torná-lo real. Mas seria muita pretensão dizer que esse caminho é percorrido sozinho. Arrisco dizer que ninguém conseguiria concluir uma dissertação de mestrado sem o apoio e a dedicação de pessoas especiais que colaboram das mais variadas maneiras com a nossa luta. Dessa forma, eu não poderia deixar de citar as pessoas cujo apoio considero fundamental e sem as quais eu não conseguiria tornar realidade este sonho.

Em primeiro lugar quero agradecer aos meus pais Valcir e Irene por terem sempre acreditado em mim, mesmo nos momentos mais difíceis. O carinho, a dedicação e a confiança que sempre demonstraram me deram forças para atingir meus sonhos e crescer como pessoa.

Quero agradecer também aos meus orientadores Philippe Navaux e Marcelo Portugal, pelo apoio e conhecimento que me repassaram. Ambos sempre estiveram dispostos a auxiliar nas dificuldades que existiram ao longo do percurso.

Existem ainda aqueles que, senão ajudaram diretamente na elaboração deste trabalho, em muito contribuíram com seu apoio e amizade. Primeiro gostaria de agradecer ao Paulo Fernandes, que apesar de meu primo, na verdade considero como um irmão mais velho que nunca tive. Com certeza é uma das pessoas mais incríveis que já conheci. Quero agradecer também ao Leandro Gazola (que há muito tempo é um grande companheiro de todas as horas), ao Giovanni Rocco e ao Rodrigo V. Nunes pela amizade sincera e por alguns momentos de convivência que acabaram se tornando fundamentais na minha vida.

Por fim, um agradecimento todo especial a minha namorada Kelly Christine, pelo amor, carinho, incentivo e por tudo que ela representou e representa na minha vida desde o dia em que a conheci. Espero, um dia, poder retribuir tudo isto.

Muito obrigado a todos !!

# SUMÁRIO

LISTA DE ABREVIATURAS .....	6
LISTA DE FIGURAS .....	7
LISTA DE FÓRMULAS .....	8
LISTA DE TABELAS .....	9
RESUMO .....	10
ABSTRACT .....	11
1 INTRODUÇÃO .....	12
1.1 Objetivos .....	12
1.2 Descrição dos Capítulos .....	13
2 REDES NEURAIIS ARTIFICIAIS .....	15
2.1 Componentes das Redes Neurais Artificiais .....	16
2.1.1 Unidades de Processamento .....	16
2.1.2 Estado de Ativação .....	17
2.1.3 Função de Saída .....	17
2.1.4 Padrão de Interconexão .....	17
2.1.5 Regra de Propagação .....	18
2.1.6 Regra de Ativação .....	19
2.1.7 Regra de Aprendizado .....	20
2.1.8 Ambiente .....	21
2.2 Aprendizado em Redes Neurais .....	21
2.2.1 Dinâmica: Convergência e Estabilidade .....	24
2.2.1.1 Convergência .....	25
2.2.1.2 Estabilidade Global .....	25
2.2.2 Principais Modelos de Redes Neurais .....	26
2.3 O Modelo Perceptron Multi-camadas .....	26
2.3.1 Características Básicas .....	26
2.3.2 Aprendizado .....	27
2.3.2.1 Algoritmo de Aprendizado - <i>Backpropagation</i> .....	28
2.3.2.2 Problemas no Aprendizado .....	29
3 SÉRIES TEMPORAIS .....	31
3.1 Conceitos Básicos .....	31
3.2 Panorama dos Principais Métodos de Previsão de Séries de Tempo .....	32
3.2.1 Modelos Univariados .....	32
3.2.1.1 Decomposição Clássica .....	33
3.2.1.2 Modelos Automáticos ou "Caixa-preta" .....	33
3.2.1.3 Modelos Box & Jenkins (ARIMA) .....	34
3.2.1.4 Modelos Bayesianos de Previsão .....	35
3.2.2 Modelos de Função de Transferência .....	36
3.2.2.1 Modelos Estruturais .....	36
3.2.3 Modelos Multivariados .....	36
3.3 Modelo Estrutural Clássico .....	37
3.3.1 Modelo Estrutural Básico .....	38

3.3.2 O Filtro de Kalman .....	40
<b>4 REDES NEURAIS E PREVISÃO: PROPOSTA DE UMA METODOLOGIA PARA MODELAGEM DA ARQUITETURA .....</b>	<b>41</b>
4.1 Determinação da Camada de Entrada da Rede .....	42
4.2 Determinação da Camada Escondida de Rede.....	43
4.3 Metodologia de Modelagem Adotada.....	44
<b>5 ANÁLISE E PREVISÃO ECONÔMETRICA DAS SÉRIES TEMPORAIS .....</b>	<b>45</b>
5.1 Série A: Índice Mensal de Passageiros de Linhas Aéreas .....	46
5.1.1 Características da Série .....	47
5.1.2 Resultados .....	48
5.2 Série B: Índice Pluviométrico Anual em Fortaleza.....	50
5.2.1 Características da Série .....	51
5.2.2 Resultados .....	52
5.3 Série C: Índice Mensal da Produção Industrial do Rio Grande do Sul.....	53
5.3.1 Características da Série .....	53
5.3.2 Resultados .....	55
<b>6 MODELAGEM E PROGNÓSTICOS DE SÉRIES TEMPORAIS UTILIZANDO REDES NEURAIS .....</b>	<b>57</b>
6.1 Série A: Índice Mensal de Passageiros de Linhas Aéreas .....	57
6.1.1 Modelagem.....	57
6.1.2 Resultados .....	59
6.1.3 Validação .....	60
6.2 Série B: Índice Pluviométrico Anual em Fortaleza.....	62
6.2.1 Modelagem.....	62
6.2.2 Resultados .....	64
6.2.3 Validação .....	65
6.3 Série C: Índice Mensal da Produção Industrial do Rio Grande do Sul.....	66
6.3.1 Modelagem.....	66
6.3.2 Resultados .....	67
6.3.3 Validação .....	68
<b>7 CONCLUSÃO .....</b>	<b>71</b>
7.1 Análise dos Resultados Obtidos.....	72
7.2 Trabalhos Futuros.....	72
<b>BIBLIOGRAFIA.....</b>	<b>74</b>

## LISTA DE ABREVIATURAS

RNAs	- Redes Neurais Artificiais
UCM	- <i>Unobservable Components Models</i>
ARIMA	- <i>Auto Regressive Integrate Moving Average</i>
IA	- Inteligência Artificial
FLE	- Filtro Linear Estável
FLI	- Filtro Linear Instável
PDP	- Processamento Distribuído e Paralelo
RMSE	- <i>Root Mean Square Error</i>
MAE	- <i>Mean Absolute Error</i>

## LISTA DE FIGURAS

FIGURA 2.1 - Neurônio como unidade limiar .....	17
FIGURA 2.2 - Exemplo de RNA.....	18
FIGURA 2.3 - Exemplos de funções de ativação .....	19
FIGURA 2.4 - Representação de um mínimo local .....	25
FIGURA 2.5 - Topologia genérica do perceptron multi-camadas .....	27
FIGURA 3.1 - Esquema ilustrativo dos modelos univariados.....	33
FIGURA 3.2 - Diagrama teórico do modelo Box & Jenkins .....	34
FIGURA 3.3 - Esquema ilustrativo dos modelos de Função de Transferência .....	36
FIGURA 4.1 - Topologia de Varfis e Versino .....	43
FIGURA 5.1 - Gráfico do índice mensal (em milhares) de passageiros de avião.....	46
FIGURA 5.2 - Gráfico do logaritmo da série de passageiros de avião .....	47
FIGURA 5.3 - Gráfico da componente sazonal - Série A.....	47
FIGURA 5.4 - Gráfico da componente tendência - Série A .....	48
FIGURA 5.5 - Gráfico da componente irregular - Série A.....	48
FIGURA 5.6 - Gráfico comparativo dos resultados do método UCM - Série A .....	50
FIGURA 5.7 - Gráfico da série do índice pluviométrico anual em Fortaleza.....	50
FIGURA 5.8 - Gráfico da componente ciclo estimada - Série B .....	51
FIGURA 5.9 - Gráfico da componente irregular - Série B .....	51
FIGURA 5.10 - Gráfico comparativo dos resultados do método UCM - Série B .....	53
FIGURA 5.11 - Gráfico da série de produção industrial do Rio Grande do Sul.....	53
FIGURA 5.12 - Gráfico da componente sazonal - Série C.....	54
FIGURA 5.13 - Gráfico da componente irregular - Série C .....	54
FIGURA 5.14 - Gráfico comparativo dos resultados do método UCM - Série C .....	56
FIGURA 6.1 - Arquitetura para previsão do valor futuro $\ln(Y_t)$ - Série A.....	59
FIGURA 6.2 - Gráfico comparativo dos resultados obtidos pela RNA - Série A .....	60
FIGURA 6.3 - UCM x RNA: um passo à frente - Série A .....	61
FIGURA 6.4 - UCM x RNA: um passo à frente - Série A .....	61
FIGURA 6.5 - Comparação das medidas de erro - Série A .....	62
FIGURA 6.6 - Gráfico do ciclo utilizado como entrada da RNA.....	63
FIGURA 6.7 - Arquitetura do modelo de Rede Neural - Série B .....	63
FIGURA 6.8 - Gráfico comparativo dos resultados obtidos pela RNA - Série B.....	64
FIGURA 6.9 - UCM x RNA: um passo à frente - Série B.....	65
FIGURA 6.10 - UCM x RNA: cinco passos à frente - Série B.....	65
FIGURA 6.11 - Comparação das medidas de erro - Série B .....	66
FIGURA 6.12 - Arquitetura do modelo de Rede Neural - Série C .....	67
FIGURA 6.13 - Gráfico comparativo dos resultados obtidos pela RNA - Série C.....	68
FIGURA 6.14 - UCM x RNA: 1 passo à frente - Série C.....	69
FIGURA 6.15 - UCM x RNA: sete passos à frente - Série C.....	69
FIGURA 6.16 - Comparação das medidas de erro - Série C .....	70

## LISTA DE FÓRMULAS

FÓRMULA 2.1 - Regra de propagação .....	18
FÓRMULA 2.2 - Função limiar rampa.....	20
FÓRMULA 2.3 - Função sigmóide.....	20
FÓRMULA 2.4 - Regra Delta.....	21
FÓRMULA 2.5 - Aprendizado por correção de erros.....	23
FÓRMULA 2.6 - Aprendizado por reforço .....	23
FÓRMULA 2.7 - Princípio da convergência com probabilidade 1 .....	25
FÓRMULA 2.8 - Princípio da convergência pela diferença dos quadrados.....	25
FÓRMULA 2.9 - Modificação no valor dos pesos .....	28
FÓRMULA 2.10 - Regra delta com fator de momento .....	29
FÓRMULA 3.1 - Decomposição clássica de séries de tempo .....	33
FÓRMULA 3.2 - Formulação geral dos MAE .....	34
FÓRMULA 3.3 - Filtro Linear Estável .....	34
FÓRMULA 3.4 - Filtro Linear Instável .....	34
FÓRMULA 3.5 - Estrutura ARIMA .....	35
FÓRMULA 3.6 - Componentes não-observáveis .....	38
FÓRMULA 3.7 - Modelo estrutural Básico de séries de tempo .....	38
FÓRMULA 3.8 - (1) a (4) - Descrição das componentes do modelo .....	38
FÓRMULA 3.9 - Equações reescritas para a componente sazonalidade.....	39
FÓRMULA 3.10 - Modelo na forma de estado de espaços particular.....	39
FÓRMULA 3.11 - Modelo na forma de estado de espaços .....	39
FÓRMULA 3.12 - Equações do Filtro de Kalman .....	40
FÓRMULA 5.1 - RMSE.....	46
FÓRMULA 5.2 - MAE.....	46

## LISTA DE TABELAS

TABELA 2.1 - Diferenças entre o computador e o cérebro humano.....	15
TABELA 5.1 - Resultados UCM - Série A.....	48
TABELA 5.2 - Medidas de erro UCM -Série A.....	49
TABELA 5.3 - Resultados UCM - Série B.....	52
TABELA 5.4 - Medidas de erro UCM - Série B.....	52
TABELA 5.5 - Resultados UCM - Série C.....	55
TABELA 5.6 - Medidas de erro UCM - Série C.....	55
TABELA 6.1 - Resultados RNA - Série A.....	59
TABELA 6.2 - Medidas de erro RNA - Série A.....	60
TABELA 6.3 - Resultados RNA - Série B.....	64
TABELA 6.4 - Medidas de erro RNA - Série B.....	64
TABELA 6.5 - Resultados RNA - Série C.....	67
TABELA 6.6 - Medidas de erro RNA - Série C.....	68
TABELA 7.1 - Desempenho de cada método por série estudada.....	72

## RESUMO

Este trabalho é um estudo a respeito da aplicação de Redes Neurais Artificiais (RNAs), mais especificamente do modelo perceptron multi-camadas com aprendizado por retro-propagação de erros, à previsão de valores futuros de Séries Temporais.

O estudo foi realizado através da realização de previsões a partir de uma determinada arquitetura de rede neural, a qual é construída com base na análise estatística da série, para três séries reais. A primeira representa o índice mensal de passageiros das linhas aéreas americanas entre janeiro de 1960 e dezembro de 1971, a segunda corresponde ao índice pluviométrico anual da cidade de Fortaleza no estado do Ceará entre 1849 e 1984, e a terceira trata do índice mensal de produção industrial do estado do Rio Grande do Sul entre janeiro de 1981 e julho de 1993. As duas primeiras séries são exemplos clássicos utilizados no estudo dos modelos estatísticos aplicados à previsão de Séries Temporais.

Os resultados obtidos com as RNAs foram comparados aos prognósticos realizados pelo método econométrico que apresenta os melhores resultados para o problema da previsão de Séries Temporais: o método da decomposição da série em suas componentes básicas não-observáveis (tendência, sazonalidade, ciclo e irregular). Tais resultados mostraram que as RNAs podem apresentar excelentes níveis de precisão em seus prognósticos, indicando sua adaptação ao problema da previsão de valores futuros de Séries Temporais.

**PALAVRAS-CHAVES:** Redes Neurais, *backpropagation*, prognósticos econômicos, séries temporais.

**TITLE: "TIME SERIES PREDICTION USING ARTIFICIAL NEURAL NETWORKS"**

## **ABSTRACT**

This work presents a study of the prediction power of Artificial Neural Networks (ANN) in comparison with prediction capability of traditional Time Series models, more specifically the Unobservable Components Models (UCM). The data used to perform the study was the monthly american airlines passengers, the annual rainfall in Fortaleza, Brazil and the monthly gross industrial output for the state of Rio Grande do Sul, Brazil. The results show that Artificial Neural Networks can outperform the forecasts of Unobservable Components Models.

**KEYWORDS:** Neural networks, backpropagation, economic forecast, time series models.

# 1 INTRODUÇÃO

As Redes Neurais Artificiais (RNAs) funcionam conceitualmente de forma similar ao cérebro humano, tentando reconhecer regularidades e padrões de dados. São capazes de aprender com a experiência e partir para a generalização baseadas no seu conhecimento prévio. Embora biologicamente inspiradas, as RNAs encontraram aplicações em muitas áreas diferentes. Para este trabalho a área de especial interesse é a realização de prognósticos de valores futuros de Séries Temporais.

As RNAs possuem a habilidade de aprender a partir de exemplos, o que as torna ferramentas poderosas para as mais variadas aplicações. Uma aplicação que vem se tornando bastante popular é a utilização de redes neurais para a realização de prognósticos a respeito de valores futuros de Séries Temporais. Porém, quando se fala na utilização de Redes Neurais para a previsão de Séries Temporais, é necessário estabelecer-se um referencial que possibilite a avaliação da qualidade dos resultados obtidos comparando-os com resultados de outros métodos existentes, em geral métodos estatísticos. Existem diversos estudos, dentre os quais podem ser citados [ZAN93] [REF91], atualmente que comprovam, através da comparação de resultados obtidos, que as redes neurais podem realizar previsões mais precisas que o método estatístico mais tradicional de previsão de Séries Temporais, o modelo ARIMA [BOX76].

Neste trabalho, o modelo ARIMA não foi considerado como um *benchmark* apropriado para comparação com as previsões realizadas pelas RNAs. A principal razão para isto é a ausência no modelo ARIMA de um mecanismo que possibilite aprendizado através de erros. Analisando o panorama atual da área de previsão de Séries Temporais impõe-se que um *benchmark* melhor para as previsões das RNAs seriam os modelos estruturais clássicos, também chamados de *Unobservable Components Models* (UCM) [HAR89]. O motivo desta escolha é que as estimações destes modelos são baseadas no filtro de Kalman [KAL60], que apresenta um mecanismo de aprendizado baseado na análise de erros cometidos em previsões anteriores. Desta forma os UCM apresentam resultados invariavelmente melhores quando comparados aos do modelo ARIMA [HAR89], tornando-se portanto um teste mais forte para a capacidade de previsão das RNAs.

Uma vez estabelecida a motivação para a realização deste trabalho, serão apresentados a seguir os objetivos que se pretende alcançar ao final do mesmo.

## 1.1 Objetivos

O objetivo fundamental deste trabalho é comprovar que as Redes Neurais Artificiais são capazes de realizar prognósticos confiáveis de valores futuros de Séries Temporais.

Para alcançar este objetivo pretende-se elaborar soluções utilizando Redes Neurais para a previsão de Séries Temporais clássicas. Tal etapa pretende verificar o comportamento das previsões dos modelos de Redes Neurais em séries já analisadas com relativo sucesso por outros métodos de previsão estatísticos. Tais séries serão modeladas também pelo método UCM, para que os resultados obtidos possam ser comparados com os das RNAs. A seguir será feito um modelo de RNA para a previsão de uma série não-clássica, ou seja uma série que não aparece na literatura de previsão de

Séries Temporais e seus resultados serão comparados com os do método UCM para a mesma série.

Um segundo objetivo deste trabalho é realizar um estudo que possibilite definir uma forma padronizada de modelar Redes Neurais Artificiais para realização de prognósticos de Séries Temporais. Esta etapa é importante, pois a maior dificuldade para a utilização de Redes Neurais para a previsão de Séries Temporais é a ausência de informações sobre a série a ser modelada, ou seja quais os aspectos da série são relevantes para que a previsão de seus valores futuros possa ser realizada com maior precisão. A proposta que será apresentada mais adiante baseia-se na utilização das informações geradas pela etapa de análise da série do modelo UCM no auxílio à definição dos valores que devem ser informados à rede para que ela consiga captar melhor as regularidades da série que está tentando prever.

Finalmente, pretende-se, ao final dos experimentos, oferecer uma análise comparativa das previsões realizadas pelas RNAs e pelo método UCM. Tal análise visa avaliar qual dos dois métodos atinge resultados mais próximos da realidade para as séries estudadas.

## 1.2 Descrição dos Capítulos

Esta seção pretende fornecer ao leitor uma visão geral da estruturação adotada para a elaboração deste trabalho.

Além desta introdução, este trabalho está dividido em mais três partes, quais sejam:

- embasamento teórico;
- trabalho desenvolvido;
- conclusão.

A primeira parte apresenta capítulos de embasamento teórico necessários para a melhor compreensão do problema que se está tentando resolver, bem como da metodologia adotada como solução. Nesta primeira parte estão incluídos os capítulos 2 e 3.

No capítulo 2 são apresentados conceitos fundamentais os conceitos fundamentais sobre Redes Neurais Artificiais. Iniciando por uma rápida análise de seus componentes básicos, passando pela discussão do conceito de aprendizado e um breve apresentação dos principais modelos de Redes Neurais existentes. Por fim, é abordado o modelo que será adotado por este trabalho para a realização de prognósticos de Séries Temporais e será apresentado o algoritmo de aprendizado utilizado pelo modelo.

O capítulo 3 pretende fornecer ao leitor um panorama sobre Séries Temporais. Em primeiro lugar são apresentados os conceitos básicos. A seguir um apanhado sobre os principais métodos de previsão de Séries Temporais é realizado. Finalmente, é apresentado o modelo estatístico que será utilizado por este trabalho, chamado de Modelo Estrutural Clássico, ou *Unobservable Components Models - UCM*.

A segunda parte do trabalho, formada pelos capítulos 4, 5 e 6, apresenta a forma como foram atingidos os resultados finais deste trabalho.

O quarto capítulo apresenta um rápido estudo sobre as metodologias mais interessantes existentes na literatura a respeito de previsão de Séries Temporais utilizando-se Redes Neurais. Neste capítulo encontra-se a proposta deste trabalho para a determinação de uma metodologia para a realização da modelagem de Séries Temporais através da definição de três etapas: estudo das características da série, determinação da camada de entrada da rede e determinação da camada de saída.

No quinto capítulo, são apresentadas as três séries que terão seus valores futuros previstos. A etapa inicial de para utilização do método UCM é realizada. Nela, uma análise sobre a características de cada série é apresentada, revelando informações sobre as séries que serão fundamentais posteriormente na hora da modelagem das mesmas através de RNAs. Por fim, para cada série, são mostrados os resultados obtidos pelo método UCM.

O sexto capítulo traz a parte final e principal do trabalho, as previsões realizadas pelas redes neurais para as três séries escolhidas. Este capítulo apresenta, para cada uma das três séries, a fase de modelagem da arquitetura e dos parâmetros das RNAs, os resultados obtidos e uma comparação com os resultados do método UCM.

Finalmente, a conclusão apresenta alguns comentários sobre o trabalho realizado, sobre os objetivos alcançados e sobre trabalhos futuros que poderão ser executados.

## 2 REDES NEURAIS ARTIFICIAIS

Neste capítulo é feita uma revisão sobre Redes Neurais Artificiais. Inicialmente serão descritos os componentes básicos destes modelos e como se interrelacionam. A seguir é introduzido o conceito de aprendizado em Redes Neurais. Um rápido apanhado sobre os principais modelos de Redes Neurais é apresentado a seguir, e por fim, é apresentado o modelo perceptron multi-camadas e seu algoritmo de aprendizado que serão utilizados neste trabalho para a previsão de Séries Temporais.

As Redes Neurais Artificiais se fundamentam nos estudos sobre a estrutura do cérebro humano para tentar emular sua forma inteligente de processar informação. Alguns estudos da neurofisiologia consideram que a riqueza computacional do cérebro humano vem do grande número de neurônios que estão interconectados por uma rede complexa de sinapses [CAV88].

Estima-se que a quantidade de neurônios existentes no mesmo está na casa dos bilhões. A velocidade de processamento destes componentes é baixa se comparada com a velocidade dos computadores tradicionais. Esta deficiência é superada pela imensa quantidade de neurônios existentes que operam em paralelo [SIM90]. Estima-se que existam cerca de  $10^{11}$  a  $10^{14}$  neurônios operando em paralelo no cérebro humano. Cada um destes está conectado através de  $10^3$  a  $10^4$  sinapses em média [COT85].

Tais características permitem ao cérebro humano executar rapidamente certas funções (por exemplo, reconhecer fisionomias) que os computadores convencionais não conseguem realizar com o mesmo desempenho. Na TABELA 2.1, baseada em Cottrell [COT85] e Simpson [SIM90], é apresentada uma comparação das principais diferenças existentes entre os computadores tradicionais e o cérebro humano. Esta comparação permite dar uma idéia mais clara sobre a capacidade adaptativa do cérebro humano, em contraste com a rigidez e precisão dos computadores convencionais.

Uma possível definição, de Redes Neurais Artificiais (RNAs), também chamadas de modelos conexionistas de computação, redes neuronais ou sistemas de processamento paralelo distribuído (PDP), é apresentada a seguir:

*“Um modelo conexionista é uma estrutura de processamento de informação distribuída e paralela. Ela é formada por unidades de processamento, comumente chamadas de nós, neurônios ou células, interconectadas por arcos unidirecionais, também chamados de ligações, conexões ou sinapses. Os nós possuem memória local e podem realizar operações de processamento de informação localizada. Cada célula possui uma única saída (axônio), a qual pode se ramificar em muitas ligações colaterais (cada ramificação possuindo o mesmo sinal de saída do neurônio). Todo o processamento que se realiza em cada unidade deve ser completamente local, isto é, deve depender apenas dos valores correntes dos sinais de entrada que chegam dos neurônios através das conexões. Estes valores atuam sobre os valores armazenados na memória local da célula”.* [HEC89]

TABELA 2.1 - Diferenças entre o computador e o cérebro humano

Características	Computador	Cérebro humano
Elementos computacionais	processadores	neurônio simples
Velocidade de processamento	$10^{-9}$ segundos	$10^{-3}$ segundos
Tipo de processamento	serial	paralelo
Confiabilidade dos elementos	confiável	não-confiável
Tolerância a falhas	quase nenhuma	grande
Tipo de sinal	precisos, simbólicos	imprecisos
Tipo de controle	centralizado	distribuído
Armazenamento de informação	substituível	adaptável

Os principais elementos utilizados na descrição de RNAs são a representação distribuída, as operações locais e o processamento não-linear [SIM90]. Estes atributos especificam duas aplicações básicas dos modelos conexionistas: situações onde poucas decisões tem que ser tomadas a partir de uma grande quantidade de dados e situações onde um complexo mapeamento não-linear deve ser aprendido. Geralmente, o ferramental matemático usado nesta tecnologia inclui: equações diferenciais, sistemas dinâmicos, álgebra linear, probabilidade e estatística.

## 2.1 Componentes das Redes Neurais Artificiais

De acordo com Rumelhart [RUM86a], uma Rede Neural Artificial pode ser descrito por oito elementos principais:

- um conjunto de unidades de processamento;
- um estado de ativação;
- uma função saída;
- um padrão de interconexão;
- uma regra de propagação;
- uma regra de ativação;
- uma regra de aprendizado;
- um ambiente onde o sistema deve funcionar.

### 2.1.1 Unidades de Processamento

As células, nós ou neurônios são o meio de representação do conhecimento, por exemplo conceitos de um domínio, existente na rede neural. Os nós podem representar pontos (pixels), caracteres (letras, números), palavras ou outros conceitos, dependendo da aplicação. Na FIGURA 2.1 temos uma ilustração de um neurônio como unidade limiar. As entradas que chegam a ele representam os dendritos. Cada dendrito possui um sinal que é adicionado ( $\Sigma$ ). Depois da adição, o sinal é processado através da função limiar  $f(\ )$ , a qual produz um sinal de saída. Nesta figura, o neurônio limiar pode ser considerado como uma representação simplificada dos neurônios biológicos.

Em geral, segundo o nível de conhecimento representado pelas RNAs, existem dois tipos de representação do conhecimento possíveis: a localizada e a distribuída. A representação localizada corresponde a um único neurônio representando um conceito. Na representação distribuída, o padrão de ativação de um conjunto de unidades é que possui significado [RUM86a]. Neste trabalho, as unidades de processamento da rede são designadas pela letra  $u$ , seguida de um índice  $i$  que indica a posição que o neurônio ocupa na rede.

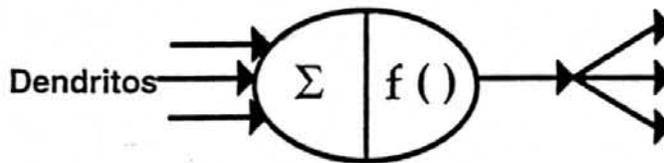


FIGURA 2.1 - Neurônio como unidade limiar

### 2.1.2 Estado de Ativação

Cada célula  $u_i$  da rede computa um estado de ativação, que é um valor numérico líquido de saída. O cálculo desta ativação é computado a partir das ativações das células conectadas diretamente a este nó, e dos correspondentes pesos destas conexões.

O estado de ativação de todas as unidades da rede, ou seja, o estado de ativação do sistema, especifica o que está sendo representado nas rede em um determinado instante  $t$  qualquer. Este estado de ativação do sistema pode ser representado por um vetor  $a(t)$ . Os valores das ativações existentes na rede podem ser discretos, por exemplo assumindo os valores  $\{0,1\}$  ou  $\{-1,0,1\}$ , como também podem ser contínuos, assumindo valores no intervalo  $[0,1]$  ou  $[-1,1]$ , que são computados pela regra de ativação a ser vista posteriormente [LIP87] [GAL88].

### 2.1.3 Função de Saída

As unidades interagem entre si através de um valor que é transmitido pelas sinapses. Este valor é determinado pela ativação da unidade estimuladora. Formalmente, o valor de saída é dado por uma função do tipo  $o_i(t) = g(a_i(t))$ .

### 2.1.4 Padrão de Interconexão

Pode-se representar o padrão de interconexão da rede por uma matriz de pesos  $w$ , onde um elemento  $w_{ij}$  corresponde à influência da célula  $u_i$  sobre a célula  $u_j$ . Conexões com pesos positivos, chamadas de excitatórias, indicam o reforço na ativação do neurônio  $u_j$ . Sinapses com pesos negativos, chamadas de inibitórias, indicam inibição na ativação da célula  $u_j$ . O conjunto das ligações excitatórias e inibitórias existentes na rede determina o comportamento da mesma.

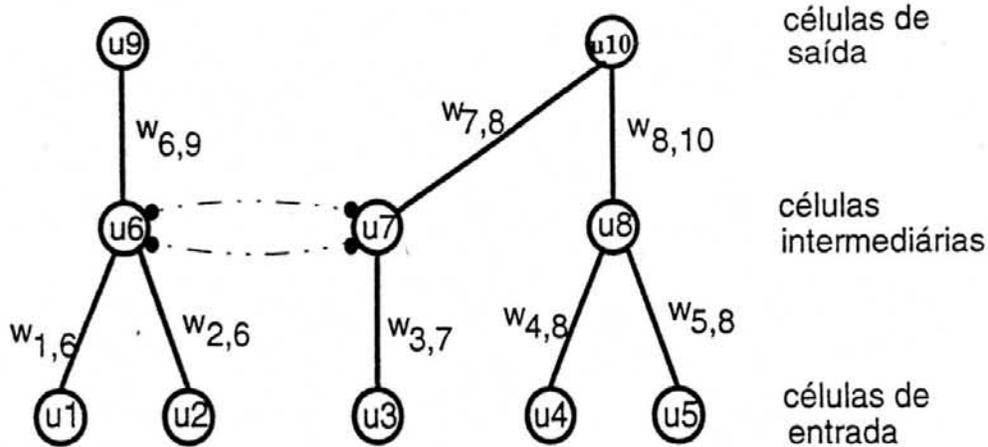


FIGURA 2.2 - Exemplo de RNA

Topologicamente, as RNAs podem ser organizadas em camadas. A camada de entrada da rede é instanciada externamente e não recomputa suas saídas. Portanto, não existem arcos de entrada em suas células. Os valores resultantes das células pertencentes a camada de saída são considerados os resultados finais da rede como um todo. Na FIGURA 2.2 as unidades  $u_9$  e  $u_{10}$  são consideradas as saídas da rede. As células que não pertencem nem à camada de entrada e nem à de saída são chamadas de intermediárias ou ocultas.

Observando a FIGURA 2.2, quando  $u_{10}$  é ativado o valor de sua ativação é determinado pelas ativações de  $u_3$ ,  $u_4$ ,  $u_5$ ,  $u_7$  e  $u_8$  e os pesos  $w_{4,8}$ ,  $w_{5,8}$ ,  $w_{3,7}$ ,  $w_{7,8}$  e  $w_{8,10}$ . Podem existir arcos que conectam nós da mesma camada, sendo chamados de sinapses intra-camadas. As ligações inter-camadas conectam células de camadas diferentes. No caso da FIGURA 2.2, os arcos que unem as unidades  $u_6$  e  $u_7$  são sinapses intra-camadas inibitórias. A ligação que conecta o nó  $u_7$  ao nó  $u_{10}$  é uma inter-camada excitatória.

Com relação aos arcos que conectam as unidades  $u_6$  e  $u_7$  eles são chamados de ligações **recorrentes**. Estas conexões formam um **ciclo**, voltando para o neurônio de origem (aquele que foi ativado primeiro). Podem existir sinapses que liguem as unidades de diferentes camadas de modo a formar ciclos entre as mesmas. Por exemplo, uma ligação que conectasse a célula  $u_{10}$  a  $u_5$  formaria um outro ciclo na rede da FIGURA 2.2. Portanto, as RNAs podem ser classificadas em redes cíclicas, que possuem ciclos, e acíclicas que não possuem ciclos [LIP87].

### 2.1.5 Regra de Propagação

Cada célula  $u_i$  computa sua nova ativação através de uma regra de propagação. Em geral, ela é definida como sendo uma função soma da entrada líquida dos pesos (**net**) das células  $u_j$  que estão diretamente conectadas a  $u_i$  conforme a FÓRMULA 2.1:

$$net_i = F \left( \sum_{j=1}^n w_{i,j} \times u_j - \theta_i \right)$$

FÓRMULA 2.1 - Regra de propagação

Onde  $u_j$  é o estado da  $j$ -ésima unidade,  $w_{ij}$  é o peso da conexão da  $i$ -ésima para a  $j$ -ésima unidade e  $\theta_i$  é o limiar da  $i$ -ésima unidade. Este limiar, que pode ser nulo inclusive, deve ser superado para que ocorra a ativação da célula.

Existem variações da regra de propagação que utilizam os conceitos de lógica nebulosa proposta por Zadeh [ZAD65]. Geralmente, nestas definições empregam-se os operadores de máximo e mínimo. Estes operadores atuam sobre o chamado **produto fuzzy**, entre as entradas e os pesos dos nós, executando as operações E e OU próprias dos conjuntos nebulosos. Neste caso, os pesos podem ser interpretados como graus de pertinência do conceito representado na entrada para o conceito representado na saída da célula.

### 2.1.6 Regra de Ativação

É necessário uma regra que calcule o valor de ativação de uma unidade no instante  $t$ . É preciso uma função  $f$  que calcule a nova ativação  $a(t)$  utilizando as entradas líquidas (**net**). Geralmente, esta função possui a forma  $a_i(t+1) = f(a_i(t), \text{net}_i(t))$ , onde  $f$  é a função de ativação, também chamada de função limiar. Esta função mapeia os neurônios de entrada para um intervalo pré-especificado de saída. As quatro funções de ativação mais utilizadas são linear, rampa, salto e sigmóide [SIM90]. Uma ilustração das quatro é apresentada na FIGURA 2.3.

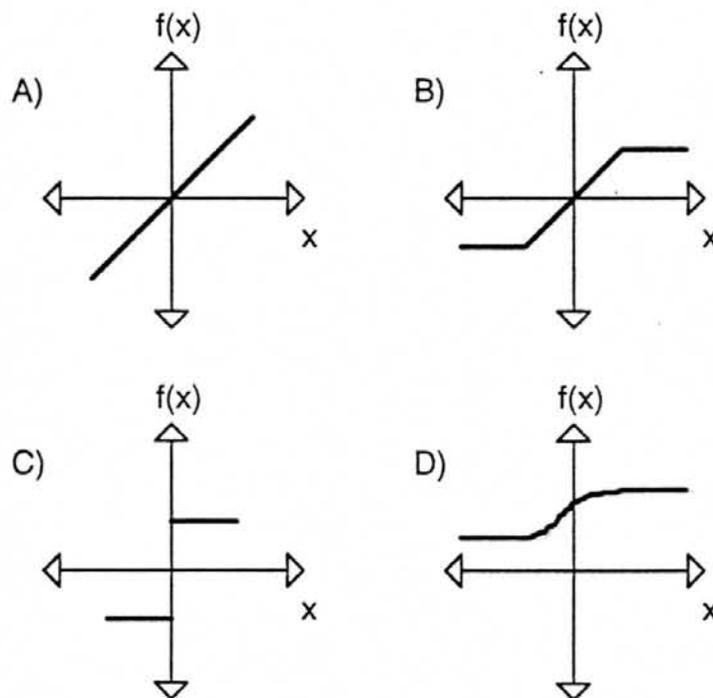


FIGURA 2.3 - Exemplos de funções de ativação

A **função linear** (FIGURA 2.3a) possui a seguinte equação:  $f(x) = \alpha x$ , onde  $\alpha$  é uma constante real que regula a intensidade da atividade de  $x$ .

A FIGURA 2.3b representa a **função limiar rampa**. Ela é uma função linear limitada pelo intervalo  $[-y, +y]$ , definida pelas seguintes equações:

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

FÓRMULA 2.2 - Função limiar rampa

Onde  $y$  indica os valores de saída máximo e mínimo, sendo chamados de pontos de saturação.

A **função salto** (FIGURA 2.3c) responde ao sinal de entrada emitindo o valor  $+1$  se o somatório for positivo e  $-1$  nos demais casos.

A FIGURA 2.3d representa a **função sigmóide**. Esta função limiar é monótona, não-decrescente e sua resposta é gradual e não-linear. A função sigmóide mais conhecida é a função logística, cujos pontos de saturação são 0 e 1. É definida pela equação da FÓRMULA 2.3.

$$S(x) = (1 + e^{-x})^{-1}$$

FÓRMULA 2.3 - Função sigmóide

## 2.1.7 Regra de Aprendizado

A modificação do processamento ou da estrutura de conhecimento de uma rede neuronal envolve modificar seu padrão de interconexão [RUM86a]. Em princípio, isto pode ser feito de três maneiras:

- desenvolvimento de novas conexões;
- perda de conexões existentes na rede;
- modificação dos pesos das conexões já existentes.

Quando o padrão de interconexão for uma matriz de pesos  $w$ , os dois primeiros itens podem ser simulados através do último. Tomando-se uma ligação com peso zero, e modificando-o para um valor positivo ou negativo, equivale a desenvolver esta sinapse. Da mesma forma, alterar o peso de uma conexão para zero significa desconectá-la.

Portanto, as regras de aprendizado alteram os pesos das sinapses das redes através da experiência.

Em geral, as regras de aprendizado podem ser consideradas como uma variante da Regra de Hebb [HEB49]. Ele estabeleceu o princípio de que a alteração da eficiência sináptica é a base do aprendizado, segundo o postulado apresentado a seguir:

*Quando o axônio A se encontra próximo do axônio B de forma a poder excitá-lo, e o faz repetidas vezes, algum processo desconhecido provoca o crescimento de sinapses entre as células A e B, facilitando assim a excitação de B por A.*

Especificamente, se uma unidade  $u_j$  recebe uma entrada de outra  $u_i$ , e ambas estão fortemente ativas, o peso  $w_{i,j}$  (de  $u_i$  para  $u_j$ ) deve ser fortalecido. Uma extensão desta idéia é apresentada na equação a seguir:

$$\delta w_{i,j} = g(a_i(t), t_i(t)) \cdot h(o_j(t), w_{i,j})$$

Onde  $t_i$  é uma espécie de unidade de entrada **instrutora** (professor) da unidade  $u_j$ . Esta equação estabelece que a mudança no peso da conexão de  $u_j$  para  $u_i$  é resultado do produto da função  $g()$  (ativação de  $u_i$  e sua entrada instrutora  $t_i$  pela função  $h()$  (valor de saída de  $u_j$  e peso  $w_{i,j}$ ).

Uma variação desta regra apresenta :

$$h(o_j(t), w_{i,j}) = o_j(t) \quad \text{e} \quad g(a_i(t)) = \eta \cdot (t_i(t) - a_i(t))$$

#### FÓRMULA 2.4 - Regra Delta

Onde  $\eta$  é uma constante de proporcionalidade que representa o rateio de aprendizado. Esta regra é chamada de **regra delta**, pois o aprendizado é proporcional a **diferença** (delta) entre a ativação realmente encontrada e a ativação fornecida pela instrutora. Esta regra é uma generalização da regra de convergência do *perceptron*, para o qual **teorema de convergência do perceptron** foi provado.

Devido a sua importância, a próxima seção é dedicada ao aprendizado no contexto dos modelos conexionistas. Considerações acerca da dinâmica dos modelos conexionistas (convergência e estabilidade) são apresentadas na seção seguinte.

### 2.1.8 Ambiente

O último componente de RNAs é o ambiente onde a rede deve funcionar. É necessário especificar a natureza do ambiente, estabelecendo os possíveis padrões de entrada e saída. Em alguns modelos, por exemplo o PDP [RUM86a], o ambiente é representado como uma função estocástica que varia ao longo do tempo sobre um espaço de padrões de entrada.

Geralmente, o ambiente é caracterizado como uma distribuição de probabilidade estável sobre um conjunto de padrões de entrada. Esta distribuição pode ser independente, ou não, de entradas ou de respostas passadas do ambiente.

## 2.2 Aprendizado em Redes Neurais

Nesta seção uma análise sobre as técnicas de aprendizado utilizadas em redes neurais é feita. A seguir serão apresentados os conceitos de convergência e estabilidade. Finalmente, serão citados alguns dos modelos clássicos encontrados na literatura.

Na sua essência, o conceito de aprendizado envolve mudança associada a aperfeiçoamento. Carbonell [CAR89] define o conceito de aprendizado, dentro do conceito da IA, como a habilidade de realizar tarefas novas que não podiam ser realizadas anteriormente, ou melhorar a realização de tarefas antigas, como resultado de mudanças produzidas pelo processo de aprendizado.

Uma classificação, apresentada por Rumelhart e Zipser [RUM85], relacionou a **função do aprendizado** em modelos conexionistas. Eles distinguiram quatro paradigmas que são apresentados abaixo acompanhados de uma breve descrição:

- **Auto-associador:** um conjunto de padrões é repetidamente apresentado e o sistema o armazena. Posteriormente, um padrão, ou parte de um, semelhante aos originais é apresentado. O sistema retorna o padrão original através de um procedimento de término padrão. Este é um processo de auto-associação onde um padrão é associado consigo mesmo, de tal forma que uma versão modificada do original pode servir de deixa para o procedimento de recuperação;
- **Associador de Padrões:** inicialmente, um conjunto de pares de padrões é apresentado ao sistema. Posteriormente, quando um membro do par é apresentado, o sistema faz a associação, produzindo o elemento correspondente do par;
- **Classificador:** neste caso há um número fixo de categorias nas quais são classificados os padrões de entrada. O objetivo é ensinar a rede a classificar corretamente os padrões de entrada, de tal forma que, quando apresentado um padrão, mesmo parcialmente modificado, a rede saiba classificá-lo corretamente;
- **Detector de Regularidades:** existe uma população de padrões estímulo, sendo que cada membro possui uma possibilidade associada. O sistema desenvolve uma representação das características dos estímulos desta população, a qual captura as propriedades mais salientes dos padrões de entrada. Não existe, a priori, um conjunto de categorias no qual os padrões desta população possam ser classificados.

Posteriormente, Rumelhart [RUM86a] apresentou uma outra classificação semelhante a esta apresentada acima. Neste caso existem dois tipos de aprendizado: o associativo e o detector de regularidades.

Nesta classificação, os três primeiros da tabela acima foram agrupados sobre o rótulo de **aprendizado associativo**. Neste aprendizado, dois padrões a serem associados são apresentados, e a rede deve aprender a mapeá-los (no caso do auto-associador o objetivo é mapeá-lo em si mesmo). No caso do detector de regularidades, não existe, a priori, um conjunto de classes determinado para separar os padrões de entrada. Não é fornecido um padrão de saída, ele deve ser descoberto.

A classificação de Rumelhart [RUM86a] se assemelha à de [LIP87] que classifica as redes quanto ao tipo de controle realizado durante o aprendizado (**supervisionado** ou **não supervisionado**). No supervisionado existem, basicamente, rótulos que especificam a classe correta para os padrões de entrada (instâncias), enquanto que no aprendizado não supervisionado não existem tais rótulos. A classificação de Rumelhart e Zipser [RUM85] é mais específica do ponto de vista de como são processados (armazenados e recuperados) os padrões durante o treinamento das redes.

Uma descrição geral das diversas **técnicas de aprendizado** utilizadas nas RNAs, inspirada em Simpson [SIM90], é apresentada a seguir:

- **Correção de Erros:** aprendizado supervisionado que ajusta os pesos das conexões entre nós na proporção da diferença entre os valores desejados e computados de cada nó da camada de saída;

- **Reforço:** aprendizado supervisionado onde os pesos são recompensados quando o sistema executa ações apropriadas e punidos caso ele não execute;
- **Estocástico:** aprendizado supervisionado que usa processos aleatórios, probabilidade e relações de energia para ajustar os pesos dos arcos;
- **Sistemas Hardwired:** as conexões e respectivos pesos são pré-determinados, semelhante a um autômato de estado finito;
- **Regra de Hebb:** aprendizado onde o ajuste dos pesos das conexões é realizado em função da relação de valores dos dois nós que ela conecta. Pode ser aplicado tanto ao aprendizado supervisionado quanto ao aprendizado não-supervisionado;
- **Competitivo e Cooperativo:** aprendizado não-supervisionado onde os processos competitivo e cooperativo são descritos em termos de redes com conexões recorrentes auto-excitáveis. Estes arcos podem ser inibidores dos nós vizinhos (competitivo) e ou excitadores dos vizinhos (cooperativo);
- **Sistemas Conectados Aleatoriamente (SCA):** aprendizado não-supervisionado utilizado para suportar a teoria de que a mente é uma rede conectada aleatoriamente quando vista do nível macroscópico.

No aprendizado por correção de erros existia um problema devido a sua incapacidade de estender o aprendizado para uma rede com mais de duas camadas. Especificamente, o valor do erro acumulado em cada unidade intermediária, que deveria ser creditado para as unidades de saída, não era definido. Com o advento dos modelos *backpropagation* [RUM86a] este problema foi solucionado. Este algoritmo consiste numa generalização da regra delta (apresentada no capítulo anterior). Através da derivação da regra delta obtém-se [RUM86b]:  $\delta w_{i,j} = \eta \cdot \delta_j \cdot o_i$ , onde:

$$\begin{aligned} \delta_j &= (t_j - o_j) \cdot f'_j(\text{net}_j) && \text{se a célula } j \text{ for de saída, ou} \\ \delta_j &= f'_j(\text{net}_j) \cdot \sum_k \delta_k \cdot w_{j,k} && \text{se a célula for intermediária} \end{aligned}$$

FÓRMULA 2.5 - Aprendizado por correção de erros

Onde  $f$  é a função de ativação e  $f'$  é sua derivada. As unidades de saída usam a diferença entre os valores obtido e desejado, como na regra delta. As unidades intermediárias precisam utilizar os valores  $\delta$  previamente calculados. Para estas células, este cálculo é realizado de maneira recorrente. Desta forma, na primeira fase, a entrada é ativada e espera-se a ativação da saída, calculando-se  $\delta_j$  para os nós de saída. Na segunda fase, a partir do cálculo de  $\delta_j$ , é feita a propagação dos sinais de erro para as camadas anteriores, efetuando-se as mudanças nos pesos. Apesar da pouca plausibilidade biológica, os resultados obtidos pelo *backpropagation* aumentaram consideravelmente o interesse em RNAs.

O aprendizado por reforço pode ser modelado pela equação:

$$\delta w_{i,j} = \alpha \cdot [r - \theta] \cdot e_{i,j}$$

FÓRMULA 2.6 - Aprendizado por reforço

Onde  $r$  é um valor escalar de sucesso/falha informado pelo ambiente,  $\theta_j$  é o valor do limiar de reforço para o  $j$ -ésimo neurônio de saída,  $e_{ij}$  é a elegibilidade canônica do peso da  $i$ -ésima para a  $j$ -ésima unidade de processamento, e  $\alpha$  é uma constante ( $0 < \alpha < 1$ ) que regula o rateio de aprendizado. A elegibilidade canônica é função da distribuição de probabilidade selecionada que é usada para determinar se o valor de saída computado é igual ao valor desejado.

O aprendizado competitivo cooperativo pode ser descrito como um sistema de equações diferenciais. A idéia fundamental deste aprendizado é possibilitar que os neurônios da rede disputem, de alguma maneira, para obtenção do direito de responder a determinados estímulos. Este tipo de aprendizado permite que as unidades se especializem, e passem a atuar como detectores de características ou classificadores de padrões sem auxílio externo [RUM85].

O aprendizado estocástico permite criar RNAs que escapem do problema de **mínimos locais** (descrito na próxima seção), utilizando um procedimento de diminuição de energia chamado *simulated annealing* [KIR83].

Os sistemas conectados aleatoriamente (SCA) fundamentam-se em um princípio de organização de caminhos de sensoriamento na mente. Este princípio postula que o posicionamento dos neurônios é ordenado e, geralmente, reflete alguma característica física do estímulo que está sendo percebido (por exemplo áreas visuais, áreas auditivas) [LIP87]. Desta forma o cérebro humano pode ser considerado como uma coleção estruturada de neurônios [KOH88].

## 2.2.1 Dinâmica: Convergência e Estabilidade

Dois importantes conceitos regem a dinâmica das redes neurais: a convergência e a estabilidade global. Biologicamente, **convergência** é a formação de similaridades sucessivas entre organismos ou associações distintas. Dentro do contexto dos modelos conexionistas, a convergência está relacionada com a minimização do erro (eventual) entre as saídas computadas e desejadas dos neurônios. Em geral, o conceito de convergência está associado ao aprendizado supervisionado.

O conceito de **estabilidade** está fortemente relacionado com a manutenção, ou retorno, a um estado de equilíbrio, sólido e estável. No contexto das RNAs, a estabilidade global é a estabilização de todas as ativações das unidades que compõem a rede, independente de qualquer entrada. Geralmente, o conceito de estabilidade está relacionado com a **chamada recursiva** existente nas redes cíclicas. Durante a execução de uma determinada tarefa, estas redes executam, eventualmente, vários ciclos de processamento até atingir um estado estável.

Quando se estuda a questão da convergência e da estabilidade em sistemas dinâmicos é comum se utilizar o conceito de **atrator**. Associando a cada RNA uma **superfície de energia**, de modo que, quando excitada com um padrão de atividade inicial, representado por um ponto sobre a referida superfície, a rede evoluirá dinamicamente, reduzindo a cada instante o valor da sua energia computacional, tendendo para o ponto de menor energia mais próximo que possa atingir. O atrator seria justamente este ponto de mínimo da superfície. O ponto de mínimo mais baixo de toda a superfície é chamado de **mínimo global**. Os demais pontos de mínimo existentes na rede são chamados de **mínimos locais**. Uma representação de um mínimo local é apresentado na FIGURA 2.4.

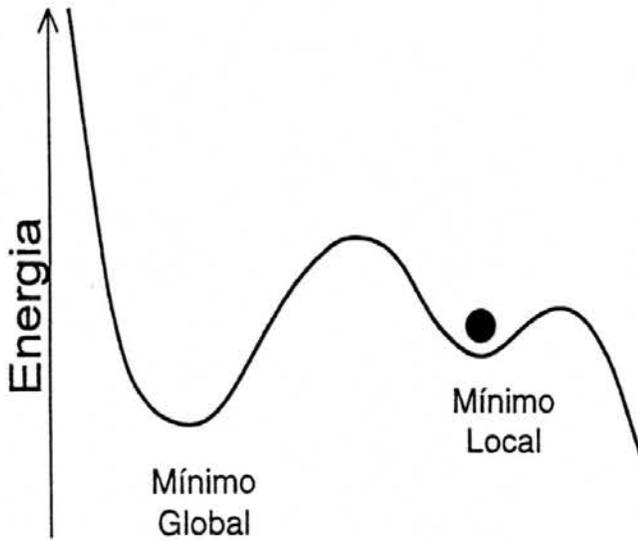


FIGURA 2.4 - Representação de um mínimo local

### 2.2.1.1 Convergência

A convergência é definida como uma seqüência de números  $x_n$  que tendem a um limite  $x$ , se dado um  $\varepsilon > 0$  pode obter-se um inteiro  $n_0$  tal que  $|x_n - x| < \varepsilon$  para todo  $n > n_0$ . Existem dois métodos de convergência que se fundamentam neste princípio: **convergência com probabilidade 1** e **convergência pela diferença dos quadrados**. A convergência com probabilidade 1 é definida como sendo

$$\lim_{n \rightarrow \infty} P\{x_n = x\} = 1$$

FÓRMULA 2.7 - Princípio da convergência com probabilidade 1

Onde  $P\{x\}$  representa a probabilidade de  $x$ . A convergência pela diferença dos quadrados é definida como sendo

$$\lim_{n \rightarrow \infty} E\{|x_n - x|^2\} = 0$$

FÓRMULA 2.8 - Princípio da convergência pela diferença dos quadrados

Onde  $E\{x\}$  representa o valor esperado de  $x$ . A convergência é uma forma de analisar se um procedimento de aprendizado qualquer está capturando corretamente o mapeamento entre os dados apresentados. Tal afirmativa é verdadeira se o mapeamento convergir para um valor, ou para um conjunto, fixo. Um grau de erro é utilizado para medir o desempenho do procedimento de aprendizado durante a realização desta tarefa.

### 2.2.1.2 Estabilidade Global

As RNAs globalmente estáveis são definidas como sistemas dinâmicos que mapeiam rapidamente todas as entradas em atratores. Estes atratores (pontos de convergência, equilíbrio, limites, etc) são locais onde a informação pode ser deliberadamente armazenada. Isto não garante que os sistemas fazem o mapeamento da maneira desejada.

Existem três teoremas principais que são usados para provar a estabilidade em RNAs. Cada um destes teoremas utiliza o **método direto de Lyapunov** [SIM90]. Este método estuda a estabilidade de encontrar certas funções das variáveis de sistemas

dinâmicos cujas derivadas, em relação ao tempo, possuem propriedades específicas. Os três teoremas são os seguintes:

- **Teorema de Cohen-Grossberg:** usado para mostrar a estabilidade em auto-associadores não-adaptativos. Este tipo de RNA opera em tempo contínuo, utiliza chamada recursiva e modifica os pesos das conexões que são desprezíveis, em relação aos valores de ativação dos neurônios, podendo ser consideradas constantes [COH83];
- **Teorema de Cohen-Grossberg-Kosko:** utilizado para descrever a estabilidade em auto-associadores adaptativos. É uma extensão do primeiro teorema. Neste caso, os auto-associadores adaptativos são redes que não possuem a restrição de ser um sistema que não pode aprender quando estiver recebendo informação [KOS88];
- **Teorema ABAM:** usado para descrever as ativações de redes neurais de duas camadas globalmente estáveis, que podem aprender e se adaptar ao mesmo tempo (heteroassociadores). Este teorema representa um caso específico do teorema anterior [KOS88].

## 2.2.2 Principais Modelos de Redes Neurais

Existem muitos modelos de redes neurais, e uma quantidade razoável de publicações que se dedicam a classificá-los [LIP87], [HIN89], [SIM90]). Dentre estes modelos, foram selecionados cinco que são apresentados a seguir em conjunto com os pesquisadores que os desenvolveram e o ano de início da publicação:

- **Perceptron** - proposto por F.Rosenblatt em 1957;
- **Perceptron de Multi-Camadas (Backpropagation)** - proposto por P.Werbos, D.Parker, D.Rumelhart em 1974 [RUM86c];
- **Classificador de Carpenter-Grossberg (Sistema ART)** - proposto por G.Carpenter, S.Grossberg em 1988 [CAR88];
- **Rede de Kohonen** - proposto por T.Kohonen em 1980 [KOH82];
- **Rede de Hopfield** - proposto por J.Hopfield em 1982 [HOP82].

A seguir será apresentado com mais detalhes o modelo do perceptron multi-camadas com algoritmo de aprendizado por retropropagação de erros, pois será este o modelo adotado neste trabalho para a previsão de séries temporais.

## 2.3 O Modelo Perceptron Multi-camadas

Como já vimos na seção anterior, o algoritmo de retropropagação é utilizado para o treinamento do modelo dos perceptrons multi-camadas.

### 2.3.1 Características Básicas

O modelo dos perceptrons multi-camadas com aprendizado baseado no algoritmo de retropropagação é o paradigma mais utilizado em áreas como reconhecimento de padrões, processamento de fala e na previsão de séries temporais [SIM 90].

A base teórica deste modelo foi apresentada originalmente em 1974 por Paul Werbos, David Parker e David Rumelhart. Em 1986 Rumelhart e James McClelland publicaram o livro "*Parallel Distributed Processing*" [RUM86a], cujo oitavo capítulo é inteiramente dedicado a análises e aprendizado deste modelo.

O modelo do perceptron multi-camadas utiliza uma topologia de três ou mais camadas. As conexões entre as unidades são do tipo intercamadas e são direcionadas da camada de entrada para a camada de saída (FIGURA 2.5).

No modelo dos perceptrons multi-camadas, entre uma camada e outra, existe uma matriz de pesos. A notação que será utilizada é  $w_{ij}$  para o peso da conexão que parte da unidade  $i$  para a unidade  $j$ . A **regra de propagação** é a combinação entre as saídas de cada unidade e a matriz de pesos e é realizada através da soma ponderada de cada sinal que chega via conexões pelo respectivo peso. O **estado de ativação** assume valores contínuos e devido a isto a **regra de ativação** das unidades utiliza como função de ativação uma função do tipo sigmóide.

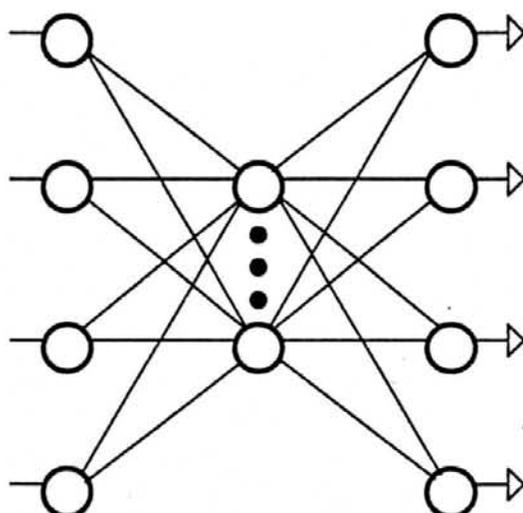


FIGURA 2.5 - Topologia genérica do perceptron multi-camadas

A função sigmóide também se faz necessária pela **regra de aprendizado** que utiliza uma função de ativação contínua, não-decrescente e diferenciável.

### 2.3.2 Aprendizado

Antes de enunciar o procedimento para o aprendizado do modelo perceptron multicamada faz-se necessário estabelecer que:

- **Regra Delta** : é o método de treinamento de Widrow e Hoff para redes sem camada escondida utilizado no modelo do perceptron;
- **Regra Delta Generalizada** : procedimento de aprendizado por minimização de erro. Consiste da aplicação do método do gradiente decrescente ao problema de múltiplas camadas;
- **Retropropagação** : propagação do erro gerado na camada de saída para trás na rede. Utiliza a regra da cadeia no cálculo da derivada do erro em relação aos pesos.

O aprendizado representado pelo algoritmo de retropropagação é do tipo **supervisionado**. Ou seja, apresenta pares de entrada e saída. Utiliza o vetor de entrada e

produz sua própria saída, comparando-a com a saída alvo (real). Se não houver diferença, não há mudanças; caso contrário, os pesos serão modificados com o objetivo de modificar esta diferença. A regra de modificação dos pesos para um determinado padrão  $p$  é dada pela seguinte fórmula:

$$\Delta_p w_{i,j} = \eta \cdot \delta_{p,j} \cdot \theta_{p,i}$$

FÓRMULA 2.9 - Modificação no valor dos pesos

Onde o padrão  $p$  corresponde ao conjunto de entradas da rede. Por exemplo, no caso de séries temporais,  $p$  corresponderá aos valores que influenciarão na realização do prognóstico do próximo valor da série. O termo  $\eta$  representa uma constante de proporcionalidade (também conhecida como taxa de aprendizado). O termo  $\delta_{p,j}$  representa a diferença entre a saída alvo e a saída da rede - erro local da unidade  $j$  para o padrão  $p$ . É esse o termo que será retropropagado para as camadas anteriores e será calculado utilizando-se o algoritmo do gradiente descendente. E finalmente o termo  $\theta_{p,i}$  é a saída produzida pela rede para o padrão  $p$  na conexão  $i$ .

A derivação desta regra de aprendizado utiliza o algoritmo *Steepest Descent*, ou seja o gradiente decrescente na superfície do espaço dos pesos [RUM86a] [BEA90]. Tal método apresenta um comportamento oscilatório, precisando de várias iterações para convergir em direção à solução, embora ele convirja rapidamente no início do processo, quando longe da solução.

### 2.3.2.1 Algoritmo de Aprendizado - *Backpropagation*

Este algoritmo apresenta duas fases bem identificadas. A primeira fase é a responsável pelo processo de treinamento da rede. É onde ocorre o ajuste dos pesos das conexões entre os neurônios. A segunda fase representa a validação da fase anterior, ou seja é onde ocorre a verificação se o treinamento foi satisfatório através da comparação de saídas geradas pela rede com valores reais.

#### i) Fase de treinamento da rede

- inicializar pesos e tendências com valores aleatórios dentro de certo intervalo;
- apresentar vetor de entrada com a respectiva saída desejada:

$$x_p = (x_0, x_1, \dots, x_n) \text{ e } t_p.$$

- calcular a saída:

$$Y_{p,j} = \sum_{i=0}^{n-1} w_{i,j} \theta_{p,i} \text{ em cada camada e o valor } \theta_{p,j} \text{ na última camada,}$$

sendo  $\theta_{p,i}$  na camada inicial igual a  $x_i$ .

- adaptar pesos, começando pela camada de saída:

$$w_{i,j}(t+1) = w_{i,j}(t) + \alpha \cdot \delta_{p,j} \cdot \theta_{p,i}, \text{ onde } w_{i,j}(t) \text{ é o peso da unidade } i$$

para a unidade  $j$  no tempo  $t$ . O termo  $\alpha$  é a taxa de aprendizado. O algoritmo do gradiente descendente requer apenas que a modificação dos pesos seja proporcional ao negativo da derivada do erro relativo ao padrão  $p$  com relação a cada peso. Esta constante de proporcionalidade é a taxa de aprendizado. Valores altos para esta constante implicam grandes modificações nos pesos. Na prática, a taxa de aprendizado dever

ser tomada tão grande quanto possível para acelerar a convergência do algoritmo, mas deve-se evitar as oscilações. O valor ótimo de  $\alpha$  depende do problema. Para superfícies de erro do tipo planície, um valor maior de  $\alpha$  resultará numa convergência mais acelerada, enquanto que para superfícies íngremes poderá gerar oscilações. Finalmente, o termo  $\delta_{p,j}$  representa o erro para o padrão  $p$  na unidade  $j$ . Na unidade de saída o erro é calculado por:

$\delta_{p,j} = (t_{p,j} - \theta_{p,j}) \cdot \theta_{p,j} \cdot (1 - \theta_{p,j})$ , enquanto que para as unidades das camadas intermediárias (escondidas), o erro é calculado por:

$$\delta_{p,j} = \theta_{p,j} \cdot (1 - \theta_{p,j}) \cdot \sum_k \delta_{p,k} \cdot w_{k,j}$$

- apresentar nova entrada até que todos os padrões sejam apresentados e o erro  $E$  esteja tão pequeno quanto se deseja.

## ii) Fase de validação do treinamento

Nesta fase deve-se conservar na rede o último conjunto de pesos resultante da fase de treinamento e apresentar vetores de entrada sem valor de saída associado. O valor de saída deverá ser calculado pela rede.

### 2.3.2.2 Problemas no Aprendizado

O maior problema da utilização do algoritmo de retropropagação ocorre quando a rede neural obtém uma solução que corresponde a um mínimo local da função do erro. Dentro do estudo do modelo de retropropagação e seu aprendizado, muitos se dedicam à pesquisa de quão freqüente e dentro de quais condições um mínimo local ocorre. Existem diversas alternativas que visam diminuir a ocorrência de mínimo local. As mais utilizadas são:

- **diminuição da taxa de aprendizado:** se a taxa de alteração dos pesos é diminuída progressivamente, então o algoritmo do gradiente decrescente está apto a alcançar uma solução melhor [BEA90].
- **adição de unidades internas:** é possível considerar que quando um mínimo local é apresentado como solução duas classes diferentes são classificadas como a mesma. Para evitar isso, ao serem colocadas mais unidades na camada escondida, a rede fará um reconhecimento melhor, diminuindo a ocorrência de mínimo local.
- **introdução de um termo de momento:** termo introduzido para que se possa aumentar a taxa de aprendizado sem que ocorra oscilação:

$$\Delta w_{i,j}(t+1) = \alpha \cdot \delta_{p,j} \cdot \theta_{p,j} + \beta \cdot (w_{i,j}(t) - w_{i,j}(t-1))$$

FÓRMULA 2.10 - Regra delta com fator de momento

Onde  $\beta$  é o fator de momento ( $0 < \beta < 1$ ) e determina o efeito da modificação do peso passado na direção do espaço de pesos. O termo de momento atua também no aumento da velocidade de convergência. Em muitas simulações realizadas por Rumelhart [RUM86a] o termo de momento assume o valor **0.9**.

Outro problema enfrentado pela retropropagação é a quebra de simetria, ou seja se os pesos começam todos com o mesmo valor e se a solução só pode ser alcançada com pesos diferentes, o sistema não poderá aprender [RUM90]. Isto ocorre porque o erro é retropropagado em proporção ao valor dos pesos, o que significa que todas as unidades escondidas conectadas diretamente às unidades de entrada e saída cometem erros iguais. Como a modificação dos pesos depende deste erro, os pesos destas unidades em relação às unidades de saída serão sempre os mesmos.

### 3 SÉRIES TEMPORAIS

Neste capítulo é feita uma apresentação dos conceitos básicos introdutórios de séries temporais. Bem como um apanhado geral de algumas das diversas metodologias estatísticas existentes na literatura de previsão de séries temporais, com o objetivo de situar o método estatístico que este trabalho adotou dentro da área de previsão de séries temporais. Finalmente será feita uma análise mais cuidadosa sobre o Modelo Estrutural Clássico de previsão de Séries Temporais, uma vez que este será o modelo utilizado como *benchmark* para comparação dos resultados das previsões feitas pelas Redes Neurais Artificiais.

A área de análise e previsão de séries temporais, atualmente um campo de pesquisa bem claro e definido dentro da Estatística, desenvolveu-se rapidamente a partir da década de 70. Sem dúvida alguma foi a partir dos trabalhos de Box & Jenkins em 1970 [BOX70] que a área despertou o interesse dos estatísticos. Quase que simultaneamente, surge em 1971 o lado bayesiano dos modelos de previsão proposto por Harrison & Stevens [HAR71]. Assim, tanto do lado "clássico" quanto do "bayesiano" toda a década de 70 foi dedicada à implementação prática dos modelos propostos e as conseqüentes melhorias metodológicas sugeridas pela aplicação real destes métodos.

#### 3.1 Conceitos Básicos

Existe uma grande classe de fenômenos (físicos, econômicos, etc.) cujo processo observacional e a conseqüente quantificação numérica, produz uma seqüência de dados distribuídos no tempo. A esta seqüência de dados ordenados segundo o parâmetro tempo, denominamos **série temporal**.

A forma mais simples de conceituar uma série temporal, denotada por  $Z_t$ , onde  $t=1,2,\dots,n$  é de interpretar  $Z_t$  como sendo um conjunto de observções discretas, observadas em tempos equidistantes que apresentam uma **depêndencia serial** entre as mesmas.

Seja  $Z^T = (Z_1, Z_2, \dots, Z_T)$  o vetor representando  $T$  observações passadas da série  $Z_t$ , conhecido também como **série histórica** ou **dados**. Estamos interessados na determinação das relações de dependência temporal da série através de uma análise estatística detalhada de sua série histórica  $Z^T$ . A este estudo denominamos **análise de séries temporais**.

A análise de uma série temporal é composta por quatro etapas:

- investigar o **mecanismo gerador** da série temporal;
- descrever o **comportamento** da série;
- procurar **periodicidades relevantes** nos dados;
- fazer **previsões sobre valores futuros** da série.

Os três primeiros objetivos acima levam a determinação das relações de dependência de  $Z^T$  (**Modelo Matemático**) que é uma descrição probabilística de uma série temporal. Após isto, pode-se partir para a previsão da série temporal, por exemplo,

a determinação dos prováveis valores que assumirão as variáveis aleatórias futuras  $Z_{T+1}$ ,  $Z_{T+2}$ , ...,  $Z_{T+\ell}$  sendo  $\ell$  o horizonte de previsão máximo. Utilza-se a seguinte notação:

- $Z_T(j)$ , sendo  $T$  a origem da previsão e
- $j = 1, \dots, \ell$  o horizonte de previsão.

Com respeito ao horizonte de previsão, dependendo de seu valor, as previsões obtidas são classificadas como de **curto**, **médio** ou **longo prazo**. Os métodos apresentados neste trabalho se referem basicamente a previsões de **curto prazo**, ou seja são métodos que permitem uma razoável previsão do futuro próximo através da extrapolação direta do ocorrido no passado.

Os métodos de previsão na sua totalidade são diferentes procedimentos computacionais para calcular a previsão de um valor futuro a partir de combinações de valores passados. Os procedimentos de previsão utilizados na prática variam muito, podendo ser simples e intuitivos ou mais quantitativos e complexos.

## 3.2 Panorama dos Principais Métodos de Previsão de Séries de Tempo

Na classificação proposta em [SOU89] os métodos de previsão de séries de tempo são agrupados em três categorias baseadas no número de séries temporais envolvidas na modelagem. As categorias são:

- **modelos univariados** : modelos que se baseiam numa única série histórica;
- **modelos função de transferência** : modelos nos quais a série de interesse é explicada pelo seu passado histórico e por outras séries temporais não correlacionadas entre si;
- **modelos multivariados** : modelos que envolvem mais de uma série temporal sem qualquer imposição com relação à direção de causalidade entre elas.

### 3.2.1 Modelos Univariados

Os métodos aqui considerados são aqueles que baseiam-se somente na informação referente à série temporal em estudo. Os métodos apresentados a seguir são classificados de duas categorias:

- métodos baseados na estatística clássica - levam em consideração somente a informação contida na série histórica  $Z^T$ .
- métodos baseados na estatística bayesiana - além da informação contida em  $Z^T$ , permitem a inclusão de outras informações relevantes não contidas na série histórica.

A FIGURA 3.1 abaixo mostra um esquema ilustrativo que caracteriza a classe dos modelos univariados.

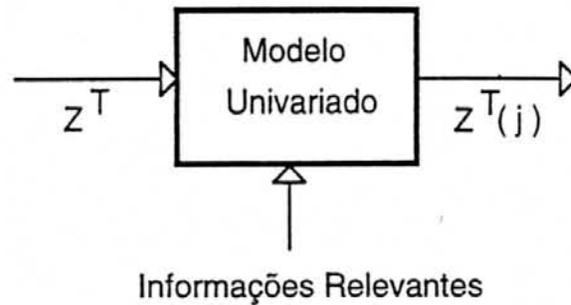


FIGURA 3.1 - Esquema ilustrativo dos modelos univariados

Serão apresentados a seguir alguns dos principais métodos de previsão dentro da categoria dos modelos univariados.

### 3.2.1.1 Decomposição Clássica

Esta é considerada a pioneira das formulações univariadas e consiste naquela classe de modelos que considera  $Z_t$  como composta de **componentes não-observáveis** distintas tais como:  $T_t$  (**tendência**),  $S_t$  (**sazonalidade**),  $C_t$  (**ciclo**) e  $e_t$  (**ruído ou erro aleatório**). Assumindo-se para  $Z_t$  o seguinte modelo geral:

$$Z_t = f(T_t, S_t, C_t, e_t)$$

FÓRMULA 3.1 - Decomposição clássica de séries de tempo

Onde a forma da FÓRMULA 3.1 especifica a maneira pela qual estes componentes se relacionam para originar o processo observável  $Z_t$ . Dentre as várias formas possíveis, destacam-se os seguintes modelos de decomposição:

- **modelo aditivo** -  $Z_t = T_t + S_t + C_t + e_t$ ;
- **modelo multiplicativo** -  $Z_t = T_t \cdot S_t \cdot C_t \cdot e_t$ ;
- **modelo misto** -  $Z_t = T_t \cdot S_t \cdot C_t + e_t$ ;

Algumas observações se fazem necessárias quanto ao modelo acima. A primeira é que normalmente, conforme [SOU89], a componente cíclica  $C$  é omitida nas equações acima pelo fato desta ser relativamente insignificante quando comparada com  $T$  e  $S$ . E a segunda é que evidentemente o modelo multiplicativo recai no aditivo via aplicação de uma transformação logarítmica.

Estes modelos utilizam meios aritméticos para extrair sucessivamente estimadores destes componentes. É interessante observar que, mesmo depois do advento das facilidades computacionais, a idéia central da "**decomposição da série temporal nas suas componentes não-observáveis**" persiste formando a base das abordagens mais atuais conforme será visto ao longo deste trabalho.

### 3.2.1.2 Modelos Automáticos ou "Caixa-preta"

São considerados automáticos todas metodologias que podem ser diretamente programadas no computador e que requerem mínima intervenção humana. Estes métodos foram desenvolvidos no início dos anos 60 e podem ser vistos como métodos de ajustamento de curvas com parâmetros seqüencialmente atualizados no tempo. Os principais modelos desta classe são:

- modelo de médias móveis;

- modelo de ajustamento sazonal;
- método de amortecimento (ou alisamento) sazonal (MAE).

Os MAE se constituem na formulação mais conhecida desta classe e ainda são bastante utilizados nos dias de hoje. A formulação geral dos MAE é a seguinte:

$$Z_t = \mu(t) \cdot C_{m_t+ks} + Ca_{t+ks} + e_t$$

#### FÓRMULA 3.2 - Formulação geral dos MAE

Onde  $\mu(t)$  é a tendência (usualmente uma função polinomial no tempo),  $s$  é o comprimento do período sazonal,  $C_{m_t+ks}$  são os fatores sazonais multiplicativos e  $Ca_{t+ks}$  são os fatores sazonais aditivos.

Dentre os principais MAE existentes destacam-se os seguintes:

- Brown : para séries não sazonais;
- Holt 2 parâmetros : para séries não-sazonais e com tendência linear;
- Winters : para séries sazonais;
- Souza e Eprecht : generalização dos MAE acima.

#### 3.2.1.3 Modelos Box & Jenkins (ARIMA)

Este é o mais importante trabalho na área de análise e previsão de séries temporais, responsável pelo grande desenvolvimento e a correspondente formalização a área. O trabalho de Box & Jenkins [BOX76] propôs uma classe geral de modelos lineares conhecida como ARIMA (Auto Regressive Integrate Moving Average) para as séries temporais.

Teoricamente Box & Jenkins assumem que a série temporal  $Z_t$  é uma realização particular de um processo estocástico gerado pela passagem sucessiva de um processo ruído branco  $a_t$  a uma seqüência de dois filtros lineares: um estável e o outro instável, como mostra a FIGURA 3.2 abaixo:

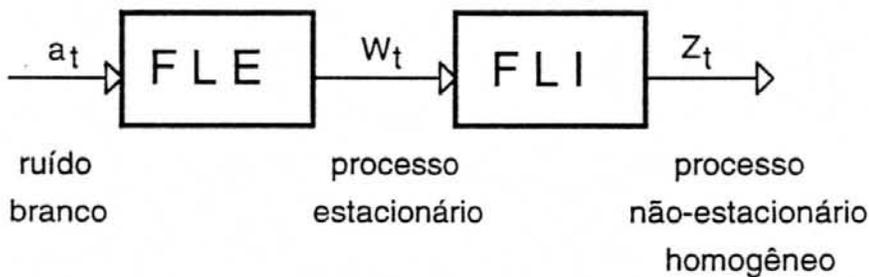


FIGURA 3.2 - Diagrama teórico do modelo Box & Jenkins

Onde:

$$FLE: \frac{\theta(B)}{\phi(B)} \equiv \frac{1 - \theta_1 B - \dots - \theta_q B^q}{1 - \phi_1 B - \dots - \phi_p B^p}$$

#### FÓRMULA 3.3 - Filtro Linear Estável

$$FLI: \nabla^{-d} = (1 - B)^{-d}$$

#### FÓRMULA 3.4 - Filtro Linear Instável

Sendo  $B$  o operador de retardo (tal que,  $B^k X_t = X_{t-k}$ ),  $\theta(B)$  o polinômio das médias móveis (MA),  $\phi(B)$  o polinômio auto-regressivo (AR) e  $\nabla$  o operador de diferença simples (tal que,  $\nabla Z_t = Z_t - Z_{t-1}$ ).

Box & Jenkins definem a estrutura ARIMA (p, d, q) pela equação:

$$\phi(B) \cdot \nabla^d Z_t = \theta(B) \cdot a_t$$

#### FÓRMULA 3.5 - Estrutura ARIMA

Onde  $d$  é o grau do operador diferença simples.

A construção do modelo é baseada num ciclo iterativo, no qual a escolha da estrutura do modelo é baseada nos próprios dados. Os estágios da metodologia são os seguintes:

- do modelo, onde uma classe geral de modelos é considerada;
- identificação do modelo, a qual é realizada com base na análise das funções de autocorrelação e de autocorrelação parcial;
- estimação dos parâmetros do modelo, objetivando minimizar o erro quadrático médio;
- verificação do modelo ajustado através de uma análise de resíduos estimados  $a_t$ , para que se possa saber se o modelo é adequado para as previsões;
- previsão dos valores futuros da série através do modelo devidamente testado.

#### 3.2.1.4 Modelos Bayesianos de Previsão

A metodologia bayesiana foi introduzida por Harrison & Stevens (1971) [HAR71]. Nela é permitido incorporar à série histórica informações transmitidas pelo analista de séries temporais, intervenções subjetivas, etc .... As características básicas são:

- representação paramétrica ou em espaço de estados - os modelos são analisados em sua forma estrutural, atribuindo-se a cada parâmetro uma interpretação específica;
- abordagem Markoviana - propicia um tipo de estimação essencialmente recursivo, que a cada instante atualiza a estimativa dos parâmetros combinando a estimativa anterior com a informação daquele instante;
- informação probabilística - os parâmetros do modelo são interpretados como variáveis aleatórias e não simplesmente quantidades desconhecidas;
- incerteza do modelo - além da incerteza dos parâmetros do modelo, a natureza bayesiana do método permite que seja levado em consideração que o modelo em funcionamento num dado instante de tempo pode não ser adequado no instante seguinte.

A formulação matemática do modelo bayesiano de Harrison & Stevens é feita através de um sistema de duas equações estocásticas que juntas formam o Modelo Linear Dinâmico (MLD), descrevendo:

- o processo gerador das observações, em termos de um vetor de parâmetros estocásticos (estados) acrescidos de um ruído branco. Isto é feito pela **Equação das Observações**;
- a evolução dos parâmetros no tempo, que se efetua através de um processo Markoviano, conhecida como **Equação do Sistema**.

### 3.2.2 Modelos de Função de Transferência

Os métodos apresentados na seção anterior consideram a descrição do comportamento futuro da série explicado somente pelo seu passado. Nos modelos de função de transferência a série é descrita não só pelo seu passado, como também por outras séries explicativas, conforme ilustrado na FIGURA 3.3 abaixo.

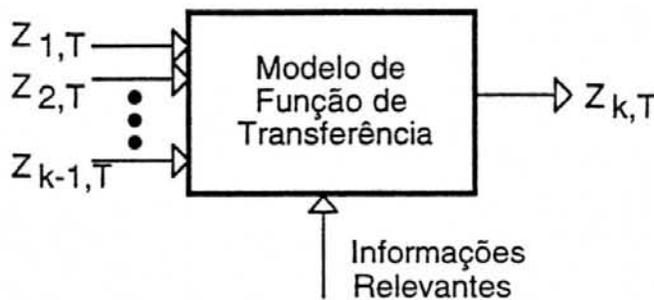


FIGURA 3.3 - Esquema ilustrativo dos modelos de Função de Transferência

É importante observar que nos modelos desta classe levam em consideração o sentido de causalidade unidirecional, ou seja as séries  $Z_{1,t}$ ,  $Z_{2,t}$ , ...,  $Z_{k-1,t}$  **causam**  $Z_{k,t}$ . Existem basicamente três metodologias de funções de transferência em uso corrente: duas baseadas na estatística clássica (Box & Jenkins e Estrutural de Harvey) e a bayesiana de Harrison & Stevens.

#### 3.2.2.1 Modelos Estruturais

Formalmente introduzidos na literatura por Harvey [HAR84], os Modelos Estruturais podem ser considerados de maneira geral, como a versão clássica da formulação em Modelos Lineares Dinâmicos utilizada por Harrison & Stevens. Com efeito, a formulação de Harvey considera o mesmo sistema de equações na descrição do modelo e considera a possibilidade de inclusão de variáveis "causais" (ou exógenas) que passam a fazer parte do vetor das variáveis independentes.

A diferença fundamental entre as duas abordagens está na forma pela qual os hiperparâmetros do modelo são considerados. Enquanto nos bayesianos estes são designados utilizando-se fatores de desconto, na formulação de Harvey estes são estimados através dos dados históricos. O modelo estrutural clássico será formalmente abordado na seção subsequente. Finalmente, deve-se ressaltar que alguns autores classificam os modelos estruturais como modelos univariados, bastando para tal não considerar o efeito das variáveis exógenas.

### 3.2.3 Modelos Multivariados

Situação na qual um modelo de uma única equação não é adequado para descrever o comportamento de duas ou mais séries, devido a presença de mais de uma variável endógena na formulação do modelo.

A literatura estatística relativa aos modelos multivariados é bastante escassa. Destacam-se:

- o modelo MARMA de Box e Tiao;
- formulação em espaço de estado;
- formulação bayesiana multivariada.

### 3.3 Modelo Estrutural Clássico

Nesta seção será apresentado o modelo estrutural clássico (também conhecido como modelo das componentes não-observáveis ou *Unobservable Components Models - UCM*). Este modelo de previsão de séries temporais servirá como *benchmark* para os modelos de redes neurais apresentados à frente neste trabalho, uma vez que seus resultados vêm sendo comprovadamente superiores aos resultados de previsões realizadas com os modelos ARIMA [HAR89].

A análise univariada de séries temporais através da decomposição nas suas respectivas componentes básicas não-observáveis (**tendência, sazonalidade, ciclo e ruído aleatório**) é, sem dúvida, uma metodologia bastante conhecida no meio estatístico. Tal formulação tem como atração principal a clara interpretação dos seus diversos componentes, o que permite uma interação mais íntima entre o analista e o modelo. Porém, a formulação revolucionária dos modelos ARIMA de Box & Jenkins, na década de 70, provocou um certo abandono da idéia inicial da decomposição da série. No entanto, o modelo ARIMA final identificado para uma série apresenta, em geral, difícil interpretação de sua estrutura especialmente quando são aplicados à série operadores diferenças, como ocorre na grande maioria das séries econômicas.

Neste panorama, os modelos estruturais propostos por [HAR84] representam uma reação aos modelos ARIMA, voltando a centrar a análise na decomposição da série de tempo em seus componentes básicos não-observáveis.

A literatura a respeito dos modelos estruturais cresceu significativamente nos últimos anos, devido principalmente à introdução do Filtro de Kalman [KAL60] à econometria. O Filtro de Kalman ajudou a tornar estes modelos operacionais, fornecendo uma maneira simplificada de estimar os componentes não-observáveis que variam com o tempo em uma série. Os modelos estruturais foram desenvolvidos tanto na forma estatística clássica quanto bayesiana. A diferença fundamental entre essas duas abordagens está na estimação das variâncias ou hiperparâmetros. Nos modelos bayesianos as variâncias observacionais e de evolução são obtidas através do aprendizado da variância e de fatores de desconto [WES89]. Por outro lado, nos modelos clássicos, os hiperparâmetros são estimados por máxima probabilidade. Neste trabalho será seguida a abordagem clássica, uma abordagem mais detalhada deste modelo é apresentada em [HAR89].

O ponto de partida deste modelo é, então, a decomposição de uma série de tempo em seus componentes básicos que são:

- a tendência ( $\mu$ ) - que pode ser encarada como a parte da série de tempo que acusa um movimento regular através de um período muito longo de tempo. É identificada por um movimento persistente da série em alguma direção;

- o ciclo ( $\psi_t$ ) - que pode ser entendido como o componente da série de tempo que apresenta um movimento em torno da tendência ao longo do tempo. Identificado quando o intervalo de observações da série for maior ou igual a um ano;
- a sazonalidade ( $\gamma_t$ ) - os fenômenos sazonais são aqueles que ocorrem regularmente em certos períodos, que em geral são anuais. As séries temporais com componente sazonal presente são aquelas cujas observações são intra-anuais, ou seja, mensais, trimestrais, semanais, etc;
- o ruído aleatório ou componente irregular ( $\varepsilon_t$ ) - componente que aparece para absorver tudo aquilo que não pode ser explicado pelo modelo, isto é, reflete movimentos não sistemáticos na série.

Uma série de tempo  $Y_t$  pode ser então representada da seguinte forma:

$$Y_t = \mu_t + \gamma_t + \psi_t + \varepsilon_t.$$

FÓRMULA 3.6 - Componentes não-observáveis

O modelo completo deve especificar também o comportamento de cada um dos componentes individuais. Um tratamento completo do modelo estrutural clássico está além do escopo deste trabalho. Será discutido a seguir o "modelo estrutural básico de séries temporais" (o modelo tendência mais ciclo é discutido em [POR93]).

### 3.3.1 Modelo Estrutural Básico

No modelo estrutural básico de séries temporais apenas três componentes dos citados acima são utilizados, uma vez que o componente ciclo é omitido. A tendência é modelada como um passeio aleatório com um desvio de tempo variante enquanto o próprio desvio é um passeio aleatório. O componente de sazonalidade é modelado pela combinação de ondas de seno e cosseno. Segundo [POR94], o modelo pode ser expresso como:

$$y_t = \mu_t + \gamma_t + \varepsilon_t$$

FÓRMULA 3.7 - Modelo estrutural Básico de séries de tempo

Onde:

$$1) \mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t$$

$$2) \beta_t = \beta_{t-1} + \zeta_t$$

$$3) \gamma_t = \sum_{j=1}^{s/2} \gamma'_{jt}$$

onde  $\gamma'_{jt}$  é formado por:

$$4) \begin{bmatrix} \gamma_{jt} \\ \gamma^*_{jt} \end{bmatrix} = \begin{bmatrix} \cos \lambda_j & \text{sen} \lambda_j \\ -\text{sen} \lambda_j & \cos \lambda_j \end{bmatrix} \begin{bmatrix} \gamma_{j,t-1} \\ \gamma^*_{j,t-1} \end{bmatrix} + \begin{bmatrix} \omega_{jt} \\ \omega^*_{jt} \end{bmatrix}$$

FÓRMULA 3.8 - (1) a (4) - Descrição das componentes do modelo

Sendo  $\varepsilon_t$ ,  $\eta_t$  e  $\zeta_t$  são erros com distribuição normal com variâncias  $\sigma_\varepsilon^2$ ,  $\sigma_\eta^2$  e  $\sigma_\zeta^2$ ,  $\omega_t$  e  $\omega_t^*$  são erros distribuídos normalmente com variâncias comuns  $\sigma_\omega^2$ ,  $\lambda_j$  é a frequência e  $\gamma_{jt}^*$  aparece por construção.

Com objetivo de simplificação vamos supor que estamos lidando com dados trimestrais, de forma que  $s = 4$ . Logo, a frequência é  $\lambda_j = 2\pi j/4$  para  $j = 1$  e  $2$ , o que resulta em  $\lambda_j = \pi, \pi/2$ . Utilizando alguns recursos trigonométricos básicos, podemos reescrever 3.8.4 como:

$$\begin{aligned}\gamma_{1t} &= \gamma_{1t-1}^* + \omega_{1t} \\ \gamma_{1t}^* &= -\gamma_{1t-1} + \omega_{1t}^* \\ \gamma_{2t} &= -\gamma_{2t-1} + \omega_{2t} \\ \gamma_{2t}^* &= -\gamma_{2t-1}^* + \omega_{2t}^*\end{aligned}$$

FÓRMULA 3.9 - Equações reescritas para a componente sazonalidade

Para que se possa aplicar o Filtro de Kalman às equações 3.8.1, 3.8.2, 3.8.3 e 3.9 é necessário colocar este modelo na forma de espaço de estados. Isto pode ser feito como mostrado nas equações 3.10.1 e 3.10.2 abaixo:

$$\begin{aligned}1) y_t &= [1 \ 0 \ 1 \ 0 \ 1] \alpha_t + \varepsilon_t \\ 2) \alpha_t &= \begin{bmatrix} \mu_t \\ \beta_t \\ \gamma_{1t} \\ \gamma_{1t}^* \\ \gamma_{2t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \\ \gamma_{1t-1} \\ \gamma_{1t-1}^* \\ \gamma_{2t-1} \end{bmatrix} + \begin{bmatrix} \eta_t \\ \zeta_t \\ \omega_{1t} \\ \omega_{1t}^* \\ \omega_{2t} \end{bmatrix}\end{aligned}$$

FÓRMULA 3.10 - Modelo na forma de estado de espaços particular

Esta especificação possui um componente para representar o nível da série ( $\mu_t$ ) e outro para representar a tendência ou coeficiente angular ( $\beta_t$ ). Portanto, séries sem tendência terão  $\beta_t = 0$ , enquanto no caso de uma tendência constante teremos  $\beta_t = \beta$ . As equações 3.10.1 e 3.10.2 são apenas uma representação particular de estado de espaços. De forma geral um estado de espaço pode ser escrito como:

$$\begin{aligned}1) Y_t &= X_t \alpha_t + \varepsilon_t \\ 2) \alpha_t &= M_t \alpha_{t-1} + S_t c_t + R_t u_t\end{aligned}$$

FÓRMULA 3.11 - Modelo na forma de estado de espaços

Onde  $\alpha_t$  é o vetor de estado,  $Y_t$  é o vetor de medidas,  $c_t$  é o vetor de entrada e  $X_t$ ,  $M_t$ ,  $S_t$ ,  $R_t$  são as matrizes conhecidas com as dimensões apropriadas.

Também assumimos que:

$$E(\varepsilon_i, \varepsilon_j) = \delta_{ij} \sigma^2 H_i \quad i, j = 1, \dots, T$$

$$E(u_i, u_j) = \delta_{ij} \sigma^2 Q_i \quad i, j = 1, \dots, T$$

$$E(\varepsilon_t, u_t) = 0 \quad t = 1, \dots, T$$

$$E(\alpha_0, \varepsilon_t) = 0 \quad t = 1, \dots, T$$

$$E(\alpha_0, u_t) = 0 \quad t = 1, \dots, T$$

Onde  $\delta_{ij}$  é o delta de Kronecker. Estas condições apenas estabelecem que  $\varepsilon_t$  e  $u_t$  têm significado zero e as matrizes de covariância  $\sigma^2 H_i$  e  $\sigma^2 Q_i$ , respectivamente, são não-autocorrelacionadas, não-correlacionadas entre elas e não-correlacionadas com o vetor de estados no tempo zero.

### 3.3.2 O Filtro de Kalman

O filtro de Kalman é um algoritmo de aprendizado que inclui equações de previsão 3.12.1 e 3.12.2 e equações de atualização 3.12.3 e 3.12.4. A idéia básica atrás do filtro de Kalman é que a estimação é realizada recursivamente em dois passos. No primeiro passo a melhor estimação no tempo  $t$ , utilizando toda a informação disponível até o tempo  $t-1$ , é obtida. Então esta estimação é atualizada utilizando a nova informação que se torna disponível no tempo  $t$ . Note que o ganho do filtro é utilizado para melhorar os prognósticos do vetor de estados. A qualquer momento o vetor de estados é tal que o  $Y_t$  previsto e o atual são diferentes, o erro é incorporado ao novo vetor de estados estimado para torná-lo mais preciso. Em outras palavras a equação 3.12.4 diz que o valor esperado do vetor de estado, consideradas todas as informações até o tempo  $t$ , pode ser dividido em dois componentes:

- o valor esperado deste vetor, consideradas todas informações até o tempo  $t-1$ ;
- mais um coeficiente vezes o erro de prognóstico de  $Y_t$ .

$$1) a_{t|t-1} = M_t a_{t-1} + S_t c_t$$

$$2) \Sigma_{t|t-1} = M_t \Sigma_{t-1} M_t' + R_t Q_t R_t'$$

$$3) a_{t|t} = a_{t|t-1} + K_t (Y_t - X_t a_{t|t-1})$$

$$4) \Sigma_{t|t} = \Sigma_{t|t-1} - \Sigma_{t|t-1} X_t' F_t^{-1} X_t \Sigma_{t|t-1}$$

$$5) K_t = \Sigma_{t|t-1} X_t' F_t^{-1}$$

FÓRMULA 3.12 - Equações do Filtro de Kalman

Onde  $K_t$  é o ganho do filtro e  $F_t = X \Sigma_{t|t-1} X_t' + H_t \dots$

Finalmente, resta dizer que os modelos estruturais possuem várias vantagens em relação aos modelos ARIMA. Uma vez que cada um dos componentes varia com o tempo, modificações estruturais na tendência ou nos componentes sazonais podem ser mais facilmente tratadas por este tipo de modelo. Além disso, os modelos estruturais são mais informativos, permitindo uma interpretação direta de seus componentes.

## 4 REDES NEURAI E PREVISÃO: PROPOSTA DE UMA METODOLOGIA PARA MODELAGEM DA ARQUITETURA

Este capítulo faz uma análise a respeito da aplicação do modelo dos perceptrons multi-camadas com aprendizado por retropropagação de erros aplicado ao problema de previsão de séries temporais. Inicialmente é apresentada a forma como uma rede neural se adapta ao problema da previsão de Série Temporais. Posteriormente, através de um estudo de trabalhos realizados nesta área, é mostrada a metodologia que este trabalho propõe adotar para a modelagem da arquitetura (topologia) da rede neural (Perceptron Multi-Camadas) que seja capaz de realizar prognósticos confiáveis sobre valores futuros 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 bibliografia é o modelo dos perceptrons multi-camadas com aprendizado baseado na retropropagação de erros (*backpropagation*). Este modelo tem se mostrado apropriado para previsões, embora não exista ainda uma definição da arquitetura ótima da rede neural para cada problema.

Como será visto a seguir, este trabalho pretende, a partir da análise da série temporal em questão, encontrar a arquitetura mais apropriada para a rede a ponto de conseguir prognósticos tão bons ou melhores que os obtidos pelos métodos estatísticos mais utilizados obtendo, ainda, considerações a respeito da modelagem.

Inicialmente, é necessário estabelecer de que forma uma rede neural se ajusta como um modelo de previsão de séries temporais:

- as **unidades de entrada** são compostas por informações relevantes para a previsão. Como, por exemplo, valores passados da própria série, dados de outra série que influenciam na previsão, características bem definidas do comportamento da série, etc;
- os **pesos** são parâmetros do modelo e serão estimados através do aprendizado da rede que leva em conta pares de entradas com as respectivas saídas alvos (observações reais da série temporal);
- as **unidades da camada escondida** servem de ligação entre as camadas de entrada e saída. Sua função é fundamental para que o melhor conjunto de pesos seja atingido ao final da fase de treinamento, porque a não-linearidade do modelo se localiza na função de ativação (função de transferência) das unidades escondidas [RUM90];
- a **camada de saída** é composta por apenas uma unidade que fornecerá a informação desejada: a previsão;
- o **treinamento dos pesos** representa a estimação dos parâmetros do modelo. Uma vez que o aprendizado é supervisionado, podemos, a partir do conjunto de pesos obtidos na fase de treinamento, gerar na unidade de saída previsões de um, dois, ..., n passos à frente do último ponto da série temporal que fez parte do conjunto de informações de treinamento.

Existem ainda os parâmetros constantes que são fornecidos a rede neural. São eles: **taxa de aprendizado**, o **termo de momento** e o **tamanho do intervalo de variação dos pesos**.

Diversos pesquisadores têm trabalhado com a utilização do modelo dos perceptrons multi-camadas com aprendizado por retropropagação de erros na previsão de séries temporais. Em geral a maior preocupação destes pesquisadores é quanto a determinação da **melhor topologia da rede**. Existem ainda os que estudam **métodos de aceleração e convergência do conjunto de pesos**.

Neste trabalho a preocupação maior foi com a determinação da melhor arquitetura da rede. A questão que se procurou responder foi qual o tamanho ótimo para a rede de treinamento? Neste sentido são apresentados resumidamente a seguir dois estudos empíricos que se dedicaram a estudar soluções para a questão da topologia da rede.

## 4.1 Determinação da Camada de Entrada da Rede

No trabalho de Varfis e Versino [VAR90], o qual tratava de séries econômicas mensais, é encontrada uma proposta a qual procurara modelar a entrada da rede de maneira a captar as componentes da série temporal (veja capítulo 3 para maiores detalhes). Desta forma as unidades foram divididas em três grupos:

- 1º) valores passados da série;
- 2º) valores passados referentes ao mesmo mês em anos anteriores visando captar tendências ou ciclos da série;
- 3º) grupo de doze unidades binárias para captar a sazonalidade da série. Exemplo: para o mês de janeiro o grupo de doze neurônios assumiria 100000000000, para fevereiro 010000000000 e assim sucessivamente até o mês de dezembro.

A FIGURA 4.1 apresenta esquema de como ficaria a topologia da rede utilizando-se a abordagem de Varfis.

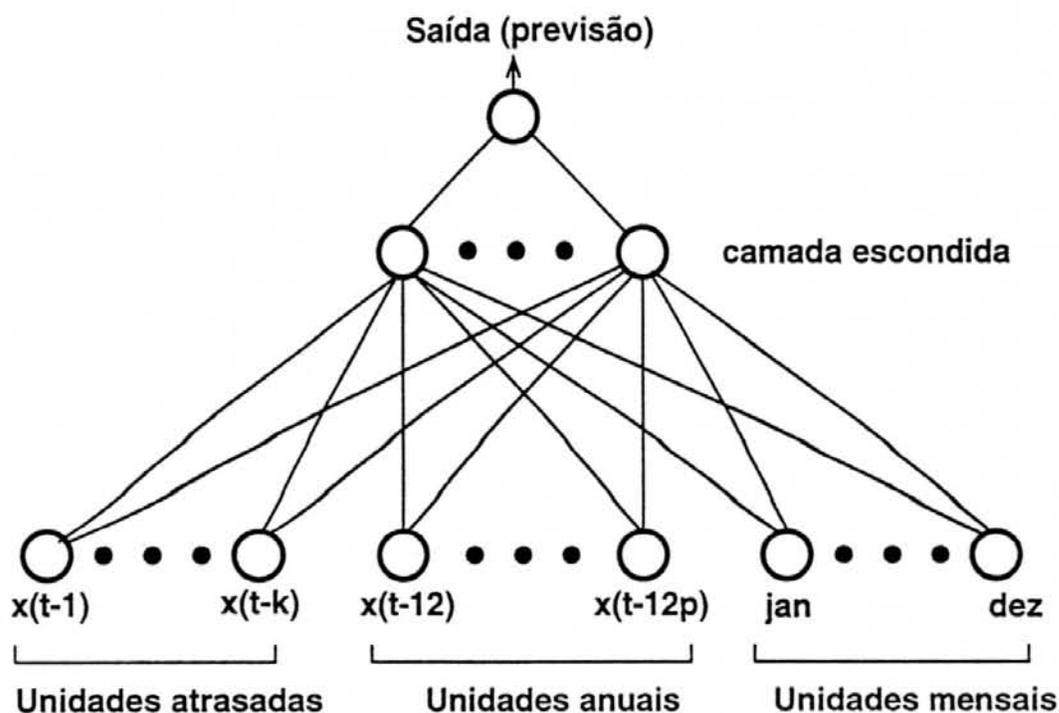


FIGURA 4.1 - Topologia de Varfis e Versino

## 4.2 Determinação da Camada Escondida de Rede

Muitos autores como [CAU90] apresentam o número de unidades das camadas escondidas como incógnita e simplesmente afirmam que "o tamanho da camada escondida é escolhido usando o bom senso: se a camada for muito grande, a rede estará memorizando os padrões e com isso perdendo a capacidade de generalização, por outro lado, se a camada for muito pequena, a rede levará muito mais iterações para atingir a precisão desejada."

Já em [RUM90], Rumelhart e Weigend afirmam que "a rede que melhor generaliza os dados é a menor rede apta a realizar o treinamento dos dados."

Os autores propõem um método que é uma extensão do método do gradiente descendente. A idéia é começar com uma rede que é muito grande para o problema, mas associar um custo a cada conexão da rede. Este custo seria decisivo para a eliminação ou não de um determinado peso.

A função custo proposta é a soma de dois termos:

- o primeiro, é a soma dos quadrados do erro;
- o segundo, descreve um custo para cada peso na rede.

Se um certo desempenho de um conjunto de dados para treinamento pode ser obtido com poucos pesos, a função custo irá encorajar a redução e eventual eliminação de tantos pesos quantos forem possíveis.

### 4.3 Metodologia de Modelagem Adotada

Segundo [LI 80], existem evidências que métodos estatísticos tradicionais podem ser utilizados para a obtenção de uma arquitetura de rede neural que apresente resultados melhores que os próprios métodos estatísticos, porque a rede não depende de hipóteses clássicas, sendo assim mais robusta.

Baseado nesta afirmação e nos estudos apresentados anteriormente é possível propor uma metodologia para a modelagem de arquiteturas de redes neurais que visam realizar previsões de Séries Temporais. Tal metodologia, adotada por este trabalho, é composta de três fases:

- 1ª) realizar a análise das componentes da série a ser modelada utilizando-se do método UCM.

Nesta fase são identificadas as componentes não-observáveis presentes na série em questão, quais sejam: **tendência**, **sazonalidade** e **ciclo**. Também é realizada uma análise da componente **irregular** ou **aleatória** da série a qual permite uma antecipação do nível de precisão das previsões realizadas. Pois quanto maior for a presença da componente irregular numa série, mais difícil se torna identificar um padrão que possibilite a realização de prognósticos com erros reduzidos.

- 2ª) definição do número de unidades da camada de entrada obtida a partir de uma análise dos dados.

Uma vez feita a análise dos dados da série, tem-se os subsídios necessários para que se possa estabelecer o número de neurônios que farão parte da camada de entrada e que tipo de informação estará contida neles. O objetivo é conseguir representar nestes neurônios da camada de entrada as informações a respeito do comportamento da série. Esta representação visa permitir que a fase de aprendizado da rede neural capte as características mais relevantes da série.

- 3ª) definição de unidades da camada escondida obtida a partir da afirmação de Rumelhart e Weigend.

Finalmente, a terceira fase tem o objetivo de estabelecer o número de neurônios da camada escondida da arquitetura. Nesta fase, a busca é feita de forma empírica, ou seja busca-se a melhor arquitetura a partir dos resultados obtidos pelos experimentos realizados, tendo como linha a seguir a afirmação de Rumelhart e Weigend.

## 5 ANÁLISE E PREVISÃO ECONOMÉTRICA DAS SÉRIES TEMPORAIS

Neste capítulo serão apresentadas as três séries escolhidas para a realização dos experimentos com Redes Neurais. Será feita uma decomposição de cada uma das séries em seus componentes básicos não-observáveis (**tendência, sazonalidade, ciclo e ruído aleatório**) e a seguir, utilizando a método estrutural clássico de previsão de séries temporais, serão apresentados os resultados obtidos. Resultados estes que servirão como parâmetro de comparação com os resultados obtidos nos experimentos com Redes Neurais.

A escolha das séries para a realização deste trabalho foi baseada na intenção de se realizar um estudo bastante abrangente. Desta forma, pretende-se trabalhar com séries que apresentem, se tomadas em conjunto, todas as componentes não-observáveis possíveis. Com isso, tem-se a possibilidade de modelar através de Redes Neurais cada uma destas componentes não-observáveis. Isto é do interesse deste trabalho, uma vez que pretende-se verificar se as Redes Neurais podem ser usadas como um método válido para a realização de prognósticos de Séries Temporais em geral.

As séries escolhidas são:

- **Série A:** índice mensal de passageiros de linhas aéreas nos EUA;
- **Série B:** índice pluviométrico anual da cidade de Fortaleza;
- **Série C:** índice mensal da produção industrial do estado do Rio Grande do Sul.

Cada uma delas terá suas características analisadas separadamente a seguir, bem como serão mostradas as previsões realizadas para seus valores futuros. Para isto será utilizado o Modelo Estrutural Clássico para previsão de Séries Temporais, também conhecido como *Unobservable Components Model - UCM*. Daqui para frente neste trabalho foi feita a opção pela denominação abreviada: simplesmente **UCM**. As previsões feitas através do método UCM foram realizadas utilizando-se o software **STAMP** [HAR89].

Antes da apresentação das séries, é importante fazer uma referência as formas para a realização da previsão de séries de tempo que serão utilizadas neste trabalho: previsão um passo à frente e previsão múltiplos passos à frente. Estes dois tipos de previsão serão utilizados tanto para o método econométrico quanto para os prognósticos realizados com Redes Neurais.

- **Previsão um passo à frente:** este termo se refere a previsões que são realizadas tendo como dados de entrada somente valores da série temporal que está sendo analisada.
- **Previsão múltiplos passos à frente:** para que se consiga realizar previsões vários passos no futuro, a última previsão realizada é introduzida como a última observação da série real servindo, desta forma, como dado de entrada para próxima previsão a ser executada.

Com o objetivo de estabelecer parâmetros palpáveis de comparação, serão adotadas neste trabalho duas formas de medidas de erros dos prognósticos realizados para os valores futuros das séries:

- **Root Mean Square Error (RMSE)**: representado pela raiz quadrada da divisão do somatório dos quadrados dos erros ( $\epsilon$ ) pelo número de prognósticos realizados ( $n$ ).

$$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n \epsilon_i^2}$$

FÓRMULA 5.1 - RMSE

- **Mean Absolute Error (MAE)**: representado pela divisão do somatório dos módulos dos erros ( $\epsilon$ ) pelo número de prognósticos realizados ( $n$ ).

$$MAE = \frac{1}{n} \cdot \sum_{i=1}^n |\epsilon_i|$$

FÓRMULA 5.2 - MAE

## 5.1 Série A: Índice Mensal de Passageiros de Linhas Aéreas

Este conjunto de dados é uma das séries clássicas utilizadas em testes de modelos de previsão de séries temporais [BRO63] [BOX76]. A série corresponde ao índice mensal de passageiros nas linhas aéreas dos Estados Unidos entre o mês de janeiro de 1960 e dezembro de 1971. A FIGURA 5.1 mostra o gráfico dos valores da série inteira. O número de passageiros está em milhares.

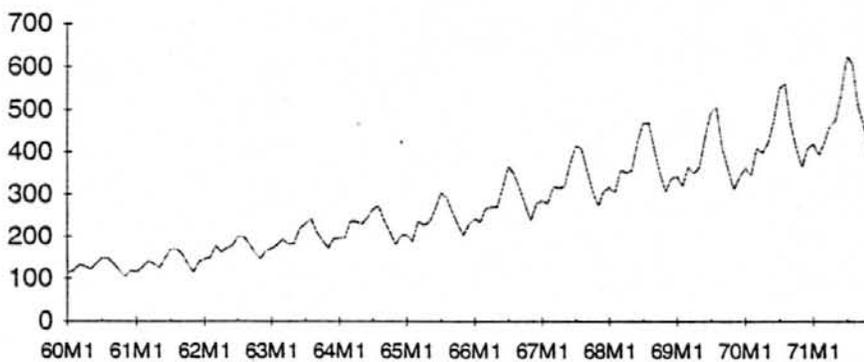


FIGURA 5.1 - Gráfico do índice mensal (em milhares) de passageiros de avião

Esta série apresenta características que a tornam um excelente teste para modelos de previsão, pois praticamente não apresenta ruídos aleatórios na sua composição. Em [HAR89] o autor sugere que se trabalhe com a série que corresponde ao logaritmo natural da série real, para evitar a tendência que a variância residual da série original tem de aumentar ao longo do tempo. Sendo assim, o novo formato da série (com o qual será feita a previsão utilizando o método UCM) é apresentado na FIGURA 5.2.

Para esta série, o conjunto de observações de janeiro de 1960 a dezembro de 1970 será utilizado como o conjunto de ajuste do modelo (fase de aprendizado), perfazendo um total de 132 observações. Enquanto que o período correspondente ao ano de 1971 (do mês de janeiro ao mês de dezembro) será utilizado para as previsões (validação do modelo).

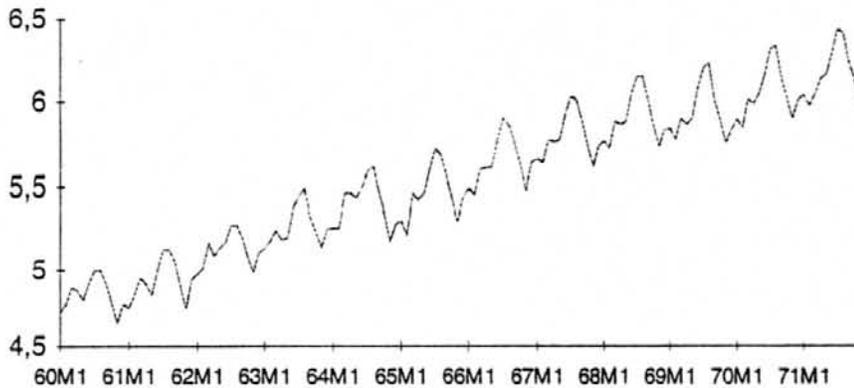


FIGURA 5.2 - Gráfico do logaritmo da série de passageiros de avião

### 5.1.1 Características da Série

Através das análises estatísticas necessárias para a modelagem da série pelo método UCM foi possível obter a característica dos componentes básicos não-observáveis da série cujos valores representam o logaritmo natural da Série A.

Após a análise dos dados foi possível a obtenção das componentes da série. Inicialmente, é importante ressaltar que a série não apresenta o componente ciclo. Já a componente sazonal aparece com a mesma constância em todo o período das observações, FIGURA 5.3.

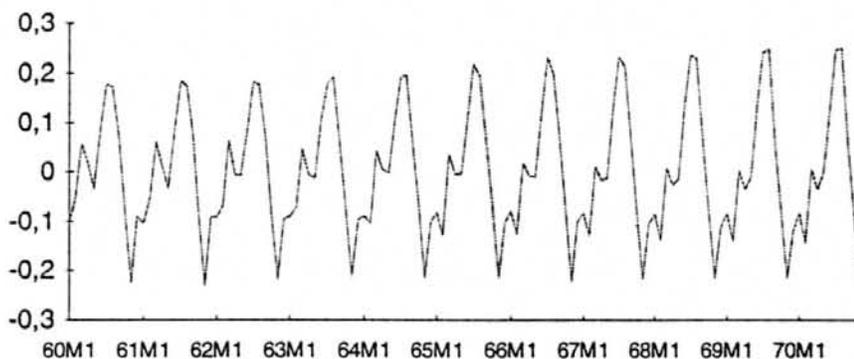


FIGURA 5.3 - Gráfico da componente sazonal - Série A

O fator sazonal mostra um período de pico anual no número de passageiros que viajaram de avião nos Estados Unidos durante os meses de julho/agosto (período de férias), assim como mostra um período de redução no número de passageiros de avião entre os meses de setembro e novembro. Com as festas de final de ano o número de viajantes aumenta novamente e segue sua escalada rumo ao pico anual com a chegada do verão norte-americano no meio do período.

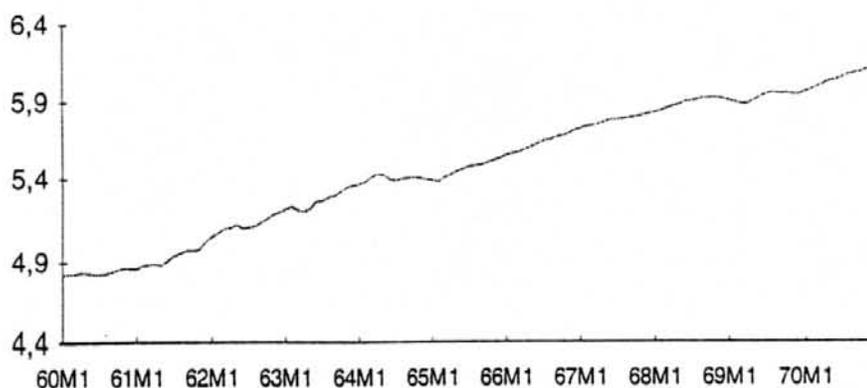


FIGURA 5.4 - Gráfico da componente tendência - Série A

A componente **tendência** também aparece com uma regularidade bastante grande. Ela apresenta um crescimento da média dos valores da série, praticamente constante ao longo do período das observções. Podendo ser considerada como uma tendência linear de crescimento, FIGURA 5.4.

De fato, ao analisarmos o gráfico da série, podemos facilmente perceber o incremento no número médio de passageiros que viajam de avião por ano. Fato que é bastante natural considerando o grande crescimento deste meio de transporte durante a década de setenta.

Por último é apresentado o gráfico (FIGURA 5.5) da componente **irregular** da série, que representa tudo aquilo que não pode ser explicado estatisticamente durante a análise da série, ou seja trata-se do componente aleatório da série.

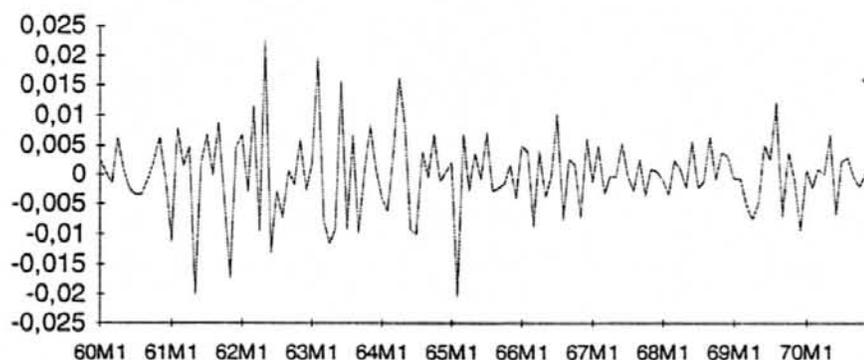


FIGURA 5.5 - Gráfico da componente irregular - Série A

É interessante ressaltar que a componente irregular apresenta valores sempre bastante próximos de zero. Tal fato já garante que as previsões dos valores futuros desta série deverão apresentar erros bastante baixos.

### 5.1.2 Resultados

Os resultados mostrados na TABELA 5.1 confirmam o que a análise do gráfico da componente irregular da série já indicava, ou seja erros muito pequenos.

TABELA 5.1 - Resultados UCM - Série A

Período	Real	Um Passo à Frente	Erro Percentual	Doze Passos à Frente	Erro Percentual
jan.71	6.0331	6.0494	-0.27%	6.0494	-0.27%
fev.71	5.9687	5.9870	-0.31%	5.9959	-0.45%
mar.71	6.0379	6.1361	-1.60%	6.1554	-1.91%
abr.71	6.1334	6.0507	1.35%	6.1234	0.16%
mai.71	6.1570	6.1381	0.31%	6.1677	-0.17%
jun.71	6.2823	6.2899	-0.12%	6.3029	-0.33%
jul.71	6.4329	6.4119	0.33%	6.4345	-0.02%
ago.71	6.4069	6.4432	-0.56%	6.4494	-0.66%
set.71	6.2305	6.2356	-0.08%	6.2652	-0.55%
out.71	6.1334	6.1091	0.40%	6.1384	-0.08%
nov.71	5.9661	5.9985	-0.54%	6.0152	-0.82%
dez.71	6.0684	6.0919	-0.39%	6.1211	-0.86%

O maior erro absoluto dentre todos os prognósticos realizados não chega sequer 2% (previsão doze passos à frente, mês de março).

A TABELA 5.2 mostra uma comparação das medidas de erro para as duas previsões realizadas com o método UCM para a Série A. Ambas apresentam valores bastante baixos, porém há uma discordância do tipo de previsão que apresenta a menor medida de erro. Utilizando RMSE a previsão um passo à frente se apresenta melhor, enquanto que se for utilizada MAE, a previsão doze passos à frente. Veremos ao decorrer do trabalho que este é o único caso que apresenta este tipo de discordância.

TABELA 5.2 - Medidas de erro UCM - Série A

Medida	Um Passo à Frente	Doze Passos à Frente
RMSE	0.6889	0.7213
MAE	0.5217	0.5008

Finalmente, como ilustração é apresentado na FIGURA 5.6 o gráfico comparativo entre as duas formas de previsão e os valores reais da série.

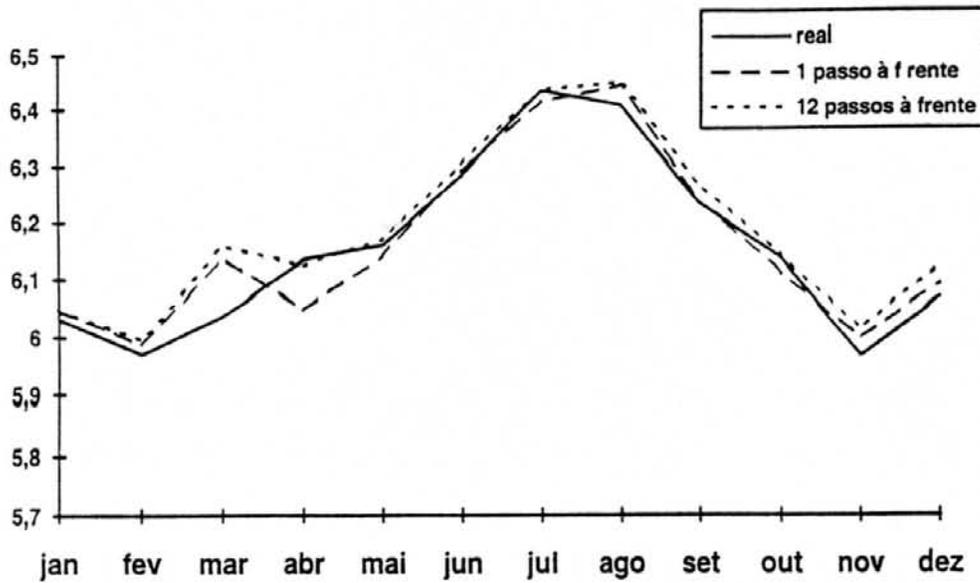


FIGURA 5.6 - Gráfico comparativo dos resultados do método UCM - Série A

## 5.2 Série B: Índice Pluviométrico Anual em Fortaleza

Esta série apresenta um conjunto de 134 observações do índice pluviométrico anual da cidade de Fortaleza, Ceará, Brasil, no período entre 1849 e 1984. A unidade de medida é centímetros. Como esta série é provavelmente a maior disponível do índice pluviométrico de uma região que sofre frequentemente com secas severas, ela tem sido objeto de um grande número de estudos na área de prognósticos de séries temporais [MOR85] [KAN86]. A FIGURA 5.7 mostra o gráfico da Série B.

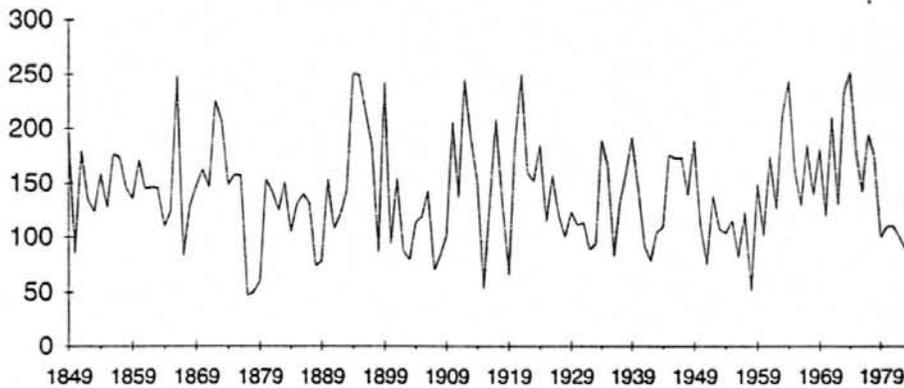


FIGURA 5.7 - Gráfico da série do índice pluviométrico anual em Fortaleza

Para esta série, o conjunto de observações entre os anos de 1849 e 1979 será utilizado como conjunto de ajuste do modelo (fase de aprendizado). Os cinco anos restantes, 1980 a 1984, serão utilizados para a realização das previsões (validação do modelo).

### 5.2.1 Características da Série

Novamente utilizando as considerações estatísticas necessárias para a modelagem da série pelo método UCM, foi possível obter o comportamento das componentes não-observáveis da série.

A primeira conclusão que se tira após a análise é a de que a Série B não apresenta as componentes de **tendência e sazonalidade**, fato que já antecipa previsões não tão acuradas quanto as que foram conseguidas na Série A. Já a componente **ciclo** pode ser estimada pelo método UCM como podemos ver no gráfico da FIGURA 5.8.

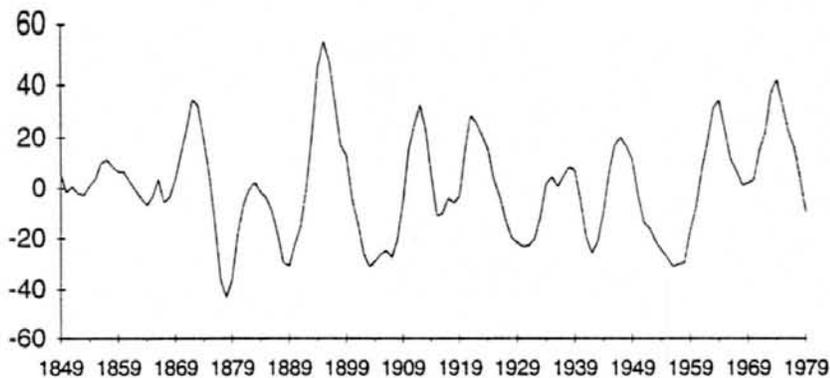


FIGURA 5.8 - Gráfico da componente ciclo estimada - Série B

Muitos dos pesquisadores que estudam esta série sustentam que ela apresenta uma composição de dois ciclos distintos, um correspondendo a um período de 13 anos e outro a um período de 26 anos.

Finalmente, resta a componente irregular da série que informa a quantidade de ruído aleatório que a série possui.

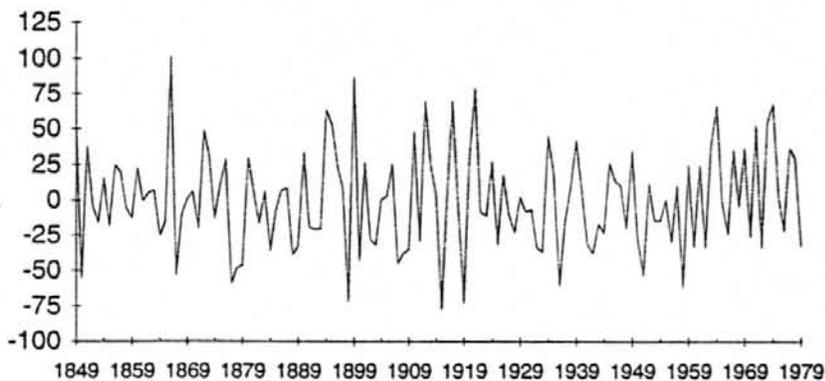


FIGURA 5.9 - Gráfico da componente irregular - Série B

Ao ser feita a análise do gráfico mostrado na FIGURA 5.9 pode-se ter uma noção da dificuldade que os métodos enfrentarão para a realização de prognósticos que apresentem erros pequenos. Isto é detectável porque pode-se verificar que em algumas observções da série o ruído aleatório pode atingir até 50% do índice.

## 5.2.2 Resultados

A análise dos resultados mostrados na TABELA 5.3 mostra a dificuldade que o método UCM teve para captar a componente ciclo da série. Os erros aparecem sempre acima de 10% chegando ao caso extremo da previsão um passo à frente para o ano de 1984, onde o erro atinge praticamente 40%. Porém, deve-se considerar que erros desta ordem em algumas observações são bastante razoáveis devido ao forte presença da componente irregular na série. Entretanto, tal fato não atenua a presença de erros acima de 10% em todas os prognósticos realizados pelo método UCM nas duas formas de previsão apresentadas.

TABELA 5.3 - Resultados UCM - Série B

Período	Real	Um Passo à Frente	Erro Percentual	Doze Passos à Frente	Erro Percentual
1980	109.5	127.47	-14.10%	127.47	-14.10%
1981	110.0	122.46	-10.17%	126.36	-12.95%
1982	100.4	122.74	-18.20%	128.16	-21.66%
1983	88.4	123.57	-28.46%	131.74	-32.90%
1984	198.1	123.67	37.57%	136.00	31.35%

A TABELA 5.4 apresenta as medidas de erro. Ambas medidas confirmam que o método UCM se sai melhor, nesta série, na previsão um passo à frente.

TABELA 5.4 - Medidas de erro UCM - Série B

Medida	Um Passo à Frente	Cinco Passos à Frente
RMSE	23.90	24.09
MAE	21.70	22.59

De qualquer forma, as medidas de erro apresentadas, tanto para a previsão um passo à frente quanto para a previsão cinco passos à frente, apresnetam-se bastante elevadas. Tal fato leva a conclusão de que os prognósticos realizados pelo método UCM para a Série B devem ser tomados com cautela, pois podem apresentar grandes distorções da realidade.

A dificuldade do método UCM para a realização de prognósticos para a Série B pode ser melhor visualizada no gráfico comparativo apresentado na FIGURA 5.10, onde estão registrados os valores reais da série e os prognósticos realizados.

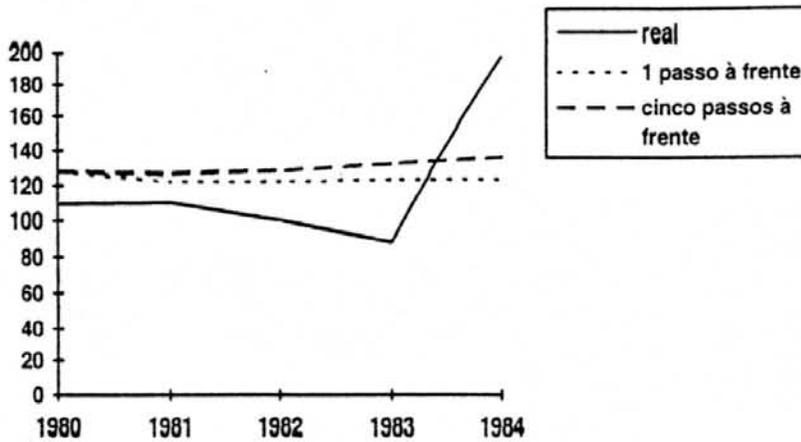


FIGURA 5.10 - Gráfico comparativo dos resultados do método UCM - Série B

### 5.3 Série C: Índice Mensal da Produção Industrial do Rio Grande do Sul

Ao contrário das duas séries anteriores, que há muitos anos vêm sendo utilizadas como objeto de estudo no meio de previsões de Séries Temporais, a série que representa o índice mensal da produção industrial do Rio Grande do Sul começou a ser estudada recentemente [FER94][POR94] na tentativa de se realizar previsões sobre seus valores futuros.

Esta série, apresentada no gráfico da FIGURA 5.11, apresenta um total de 151 observações que iniciam no mês de janeiro de 1981 e terminam no mês de julho de 1993. Uma análise rápida do gráfico já nos permite perceber que a série é estacionária (não apresentando tendência) e sazonal.

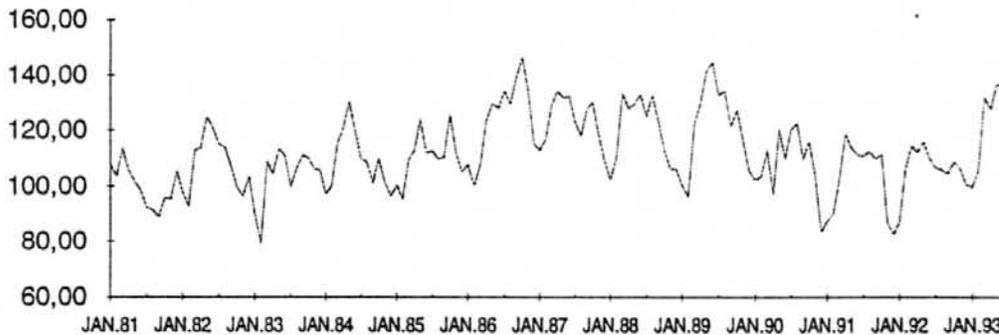


FIGURA 5.11 - Gráfico da série de produção industrial do Rio Grande do Sul

Para esta série o conjunto de observações correspondente ao período de janeiro de 1981 a dezembro de 1992 será utilizado como conjunto de ajuste do modelo (fase de aprendizado). Enquanto que o período compreendido entre janeiro e julho de 1993 será utilizado para as previsões (validação do modelo).

#### 5.3.1 Características da Série

O método UCM mais uma vez fornece as considerações estatísticas necessárias para que se possa obter o comportamento das componentes não-observáveis da série.

Com a análise da série percebe-se que a série apresenta uma componente **tendência** estacionária ao longo do tempo. Ou seja, não uma tendência importante de crescimento ou queda na produção industrial no estado do Rio Grande do Sul no período representado na série. A componente **ciclo**, por sua vez, não está presente nesta série. A componente **sazonal** aparece bem marcada ao longo do período (FIGURA 5.12). O gráfico mostra que a produção industrial do Rio Grande do Sul apresenta anualmente um período de pico nos meses entre março e agosto atingindo seu ponto máximo geralmente no mês de maio. Por outro lado a série mantém sua regularidade sazonal apresentando seu período de produção mais baixa no período entre os meses de novembro e dezembro.

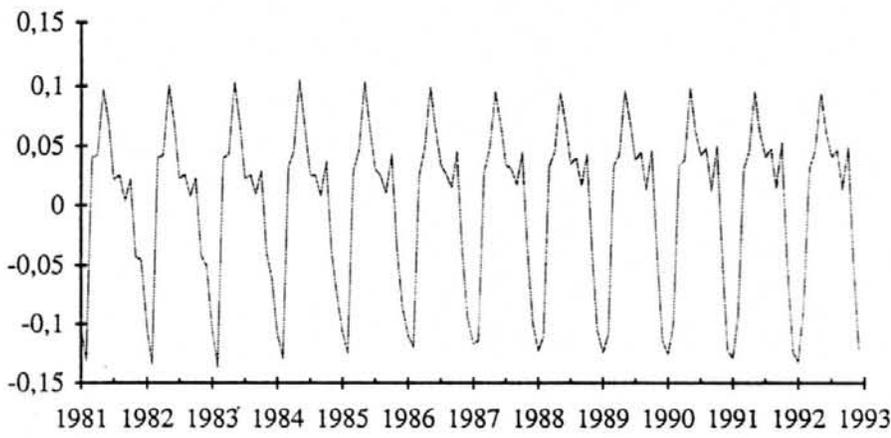


FIGURA 5.12 - Gráfico da componente sazonal - Série C

A FIGURA 5.13 apresenta o gráfico da componente **irregular**. Através da análise do gráfico desta componente pode-se esperar que o método apresente prognósticos como erros bastante baixos, devido a pouca presença de ruído aleatório na série.

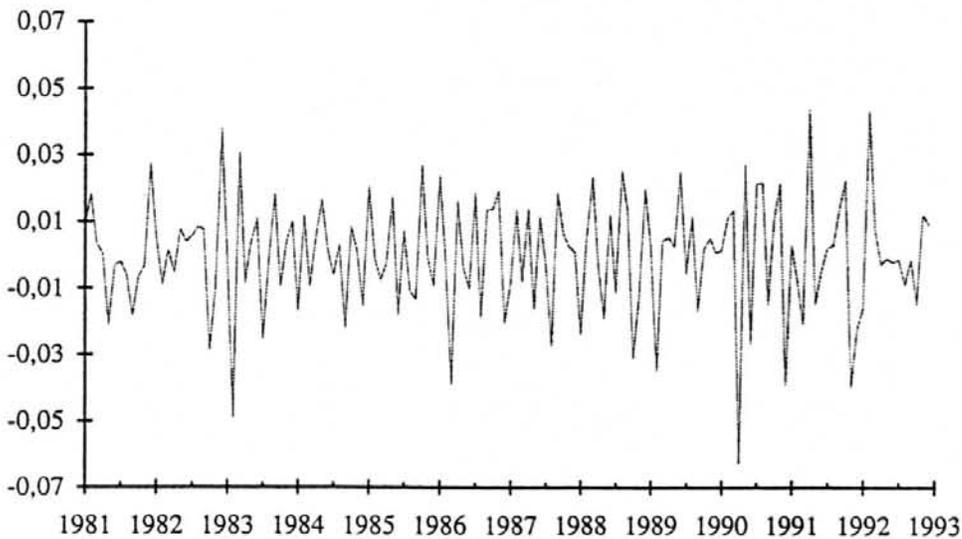


FIGURA 5.13 - Gráfico da componente irregular - Série C

### 5.3.2 Resultados

Os prognósticos apresentados pelo método UCM para a série C se apresentam de forma distinta nas duas formas de previsões adotadas. Ambas as previsões apresentam o mês de março como o mais problemático devido a este ser um mês cujo índice apresenta uma modificação elevada. Porém, ao analisar-se a TABELA 5.5, fica claro que o método UCM não consegue se recuperar do erro ocorrido no prognóstico do mês de março na previsão sete passos à frente, fato este que não ocorre na previsão um passo à frente.

Na previsão um passo à frente, o método UCM foi capaz de ajustar seus prognósticos ao substancial aumento da produção industrial após o mês de março, retornando a obter erros baixos nos seu prognósticos para os meses subseqüentes. Esta capacidade apresentada é devida a característica adicionada ao método, pela presença do Filtro de Kalman, de aprendizado com a reincorporação do erro cometido no último prognóstico.

TABELA 5.5 - Resultados UCM - Série C

Período	Real	Um Passo à Frente	Erro Percentual	Sete Passos à Frente	Erro Percentual
jan.93	99.70	98.69	1.01%	98.69	1.01%
fev.93	105.33	104.10	1.15%	103.39	1.84%
mar.93	131.95	118.04	10.54%	116.27	11.88%
abr.93	127.39	129.24	-1.46%	117.91	7.44%
mai.93	136.98	134.46	1.84%	123.77	9.65%
jun.93	135.56	131.39	3.07%	119.63	11.75%
jul.93	131.84	131.68	0.12%	117.13	11.16%

A TABELA 5.6 apresenta claramente a grande diferença na realização das previsões pelo método UCM para a Série C. Pode-se afirmar tranquilamente que o método não conseguiu realizar previsões confiáveis utilizando a forma sete passos à frente. Enquanto que atingiu níveis satisfatórios na realização de prognósticos uma passo à frente.

TABELA 5.6 - Medidas de erro UCM - Série C

Medida	Um Passo à Frente	Sete Passos à Frente
RMSE	5.65	11.86
MAE	3.55	10.28

Através do gráfico comparativo da FIGURA 5.14 pode-se acompanhar a evolução dos prognósticos. Note a grande precisão que a previsão um passo à frente consegue atingir após um grande erro no mês de março. O mesmo não ocorre com a previsão sete passos à frente que não se recupera mais após março.

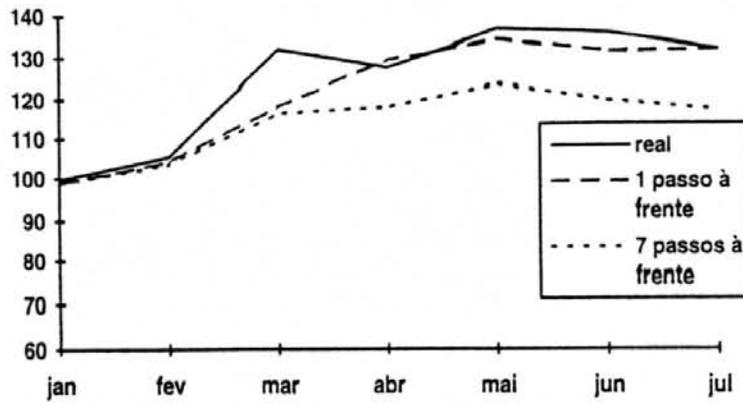


FIGURA 5.14 - Gráfico comparativo dos resultados do método UCM - Série C

## 6 MODELAGEM E PROGNÓSTICOS DE SÉRIES TEMPORAIS UTILIZANDO REDES NEURAIAS

Neste capítulo serão mostradas as fases de modelagem, obtenção de resultados e validação para cada uma das três séries apresentadas no capítulo anterior.

A apresentação da fase de modelagem de cada uma das séries incluirá uma breve discussão sobre as opções consideradas, em termos de arquitetura de rede e determinação dos valores de parâmetros, e o modelo de Rede Neural efetivamente adotado para a série em questão.

Os resultados obtidos serão apresentados nas suas duas formas de previsão: um passo à frente e múltiplos passos à frente para cada uma das séries. Estes resultados serão comparados aos valores das observações originais, para que se possa obter os erros cometidos em cada prognóstico realizado. A seguir, serão apresentadas as medidas de erro - RMSE e MAE - para as duas formas de previsão. E, por fim, um gráfico comparativo dos resultados será mostrado. Os testes foram realizados com a utilização do software **Neural Works** [NEU 91].

A validação será feita com a comparação entre os resultados obtidos pelo método econométrico UCM e os resultados obtidos pelas RNAs para as mesmas previsões. A comparação final será apresentada em gráficos contendo os prognósticos e as medidas de erros RMSE e MAE.

### 6.1 Série A: Índice Mensal de Passageiros de Linhas Aéreas

Já foram vistas no capítulo anterior (seção 5.1) todas as informações importantes a respeito da composição desta série. É importante, no entanto, ressaltar que tais informações obtidas no momento da modelagem da série pelo método UCM serão de fundamental importância para a modelagem da série utilizando Redes Neurais. Pois, será através destas informações que serão formadas as principais opções de composição da camada de entrada da Rede Neural.

Os conjuntos de treinamento e teste da rede são idênticos aos utilizados pelo método UCM.

#### 6.1.1 Modelagem

Conforme já foi visto no capítulo 4 existem dois aspectos básicos que devem ser considerados na modelagem de uma Rede Neural que será utilizada na realização de prognósticos de Séries Temporais: a determinação da arquitetura da rede e a determinação dos parâmetros utilizados no treinamento.

A arquitetura da rede é determinada em etapas, primeiro a camada de entrada e depois a camada escondida, uma vez que já se sabe que a camada de saída terá sempre um neurônio que representa o prognóstico final da rede.

No conjunto de testes realizados ao longo deste trabalho para a Série A, existiram três opções para a camada de entrada da rede, sempre com o objetivo de fornecer à rede o maior número de informações possíveis para que a ela pudesse captar a

presença das componentes não-observáveis ativas da série com a qual irá trabalhar. As opções de configuração para a camada de entrada da Série A foram:

i) **caso 1: 14 neurônios de entrada;**

- um neurônio contendo o valor da observação imediatamente anterior à atual que se quer prever (neurônio atual - 1);
- doze neurônios binários (ver seção 4.3.1), utilizados com o objetivo de auxiliar a rede a identificar a componente sazonal da Série A;
- um neurônio contendo uma seqüência linearmente crescente, utilizado com o objetivo de auxiliar a rede a identificar a componente tendência da série (neurônio de tendência).

ii) **caso 2: 3 neurônios de entrada**

- um neurônio atual - 1;
- um neurônio atual - 12, para a rede identificar a sazonalidade da série;
- um neurônio de tendência;

iii) **caso 3: idêntico ao caso 1, porém utilizando o logaritmo natural das observações da série. Como já foi visto quando da realização da previsão dos valores futuros da Série A utilizando o método UCM, a transformação da Série A na série cujos valores são correspondentes ao logaritmo natural das observações da série real é feita com a intenção de evitar a tendência que a variância residual da série original tem de aumentar ao longo do tempo.**

Já camada intermediária da rede foi determinada a partir da orientação de Caudill [CAU90] e Rumelhart e Weigend [RUM90] (seção 4.3.2). Ou seja, a opção foi por uma rede com uma camada intermediária com o menor número de unidades que não permitisse que a rede perdesse sua capacidade de generalização. Os experimentos mostraram que neste caso, para as redes com 14 neurônios na camada de entrada, o número de 6 neurônios na camada intermediária foi o mais adequado.

Os testes mostraram que a arquitetura com a camada de entrada baseada na proposta do **caso 3** foi a que obteve o melhor desempenho. Os resultados obtidos pela arquitetura baseada no **caso 1** também foram satisfatórios, porém a utilização do logaritmo natural da série no **caso 3** provou ser um acerto, pois permitiu um desempenho superior nas previsões realizadas.

Já a proposta da arquitetura baseada no **caso 2** não demonstrou a mesma eficiência em captar a componente sazonal da série vista nos **casos 1** e **3**. Levando a conclusão de que a componente sazonal de uma série é melhor captada pela rede quando da utilização na sua camada de entrada do conjunto de doze neurônios sazonais.

Seguindo a linha exposta acima a arquitetura que obteve os melhores resultados na realização dos prognósticos da Série A é mostrada na FIGURA 6.1. Sua configuração apresenta 14 neurônios na camada de entrada, 6 na intermediária e 1 na camada de saída. Segundo [ZAN93], podemos denominar esta arquitetura pela forma 14-6-1.

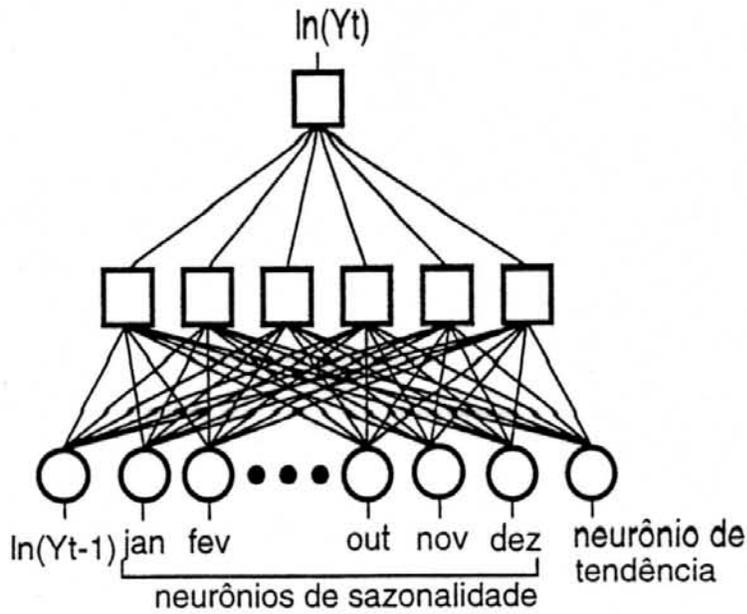


FIGURA 6.1 - Arquitetura para previsão do valor futuro  $\ln(Y_t)$  - Série A

Esta arquitetura foi treinada com os seguintes parâmetros :

- o conjunto de observações foi apresentado 15000 vezes para a rede;
- constante de aprendizado ( $\alpha$ ) = 0.9;
- momento ( $\beta$ ) = 0.6 nas 10000 épocas iniciais e 0.9 nas 5000 restantes;
- função de transferência do tipo sigmóide.

### 6.1.2 Resultados

Os resultados obtidos para a série correspondente ao logaritmo natural da Série A, podem ser visualizados na TABELA 6.1. Como se pode perceber, os erros apresentados nos prognósticos realizados apresentam seu ponto máximo no valor 1.64%, ficando na sua grande maioria abaixo de 1% tanto para a previsão um passo à frente como doze passos à frente.

TABELA 6.1 - Resultados RNA - Série A

Período	Real	Um Passo à Frente	Erro Percentual	Doze Passos à Frente	Erro Percentual
jan.71	6.0331	6.0095	0.39%	6.0095	0.39%
fev.71	5.9687	5.9084	1.01%	5.9346	0.57%
mar.71	6.0379	6.0599	-0.36%	6.0858	-0.78%
abr.71	6.1334	6.0328	1.64%	6.0394	1.53%
mai.71	6.1570	6.1209	0.59%	6.1319	0.41%
jun.71	6.2823	6.2583	0.38%	6.2855	-0.05%
jul.71	6.4329	6.3865	0.72%	6.4095	0.36%
ago.71	6.4069	6.3799	0.42%	6.4399	-0.51%
set.71	6.2305	6.2136	0.27%	6.2139	0.27%
out.71	6.1334	6.0732	0.98%	6.0870	0.76%
nov.71	5.9661	5.9839	-0.29%	5.9699	-0.06%
dez.71	6.0684	6.0459	0.37%	6.0516	0.28%

O fato do modelo de Redes Neurais ter conseguido resultados tão expressivos confirma sua capacidade de realizar prognósticos acurados, uma vez que esta série vem sendo utilizada ao longo do tempo como teste para os mais diversos métodos econométricos justamente por apresentar um comportamento bastante regular. Esta característica da série permite que se verifique a capacidade do modelo testado de identificar suas componentes não-observáveis, principalmente pela baixíssima presença de ruído aleatório na sua formação.

TABELA 6.2 - Medidas de erro RNA - Série A

Medida	Um Passo à Frente	Doze Passos à Frente
RMSE	0.7321	0.6271
MAE	0.6183	0.4975

A TABELA 6.2 faz a comparação entre as medidas de erro das duas formas utilizadas para a execução da previsão.

Apesar dos resultados obtidos pelas duas formas de previsão terem sido satisfatórios, convém ressaltar a pequena superioridade da precisão apresentada pela previsão doze passos à frente.

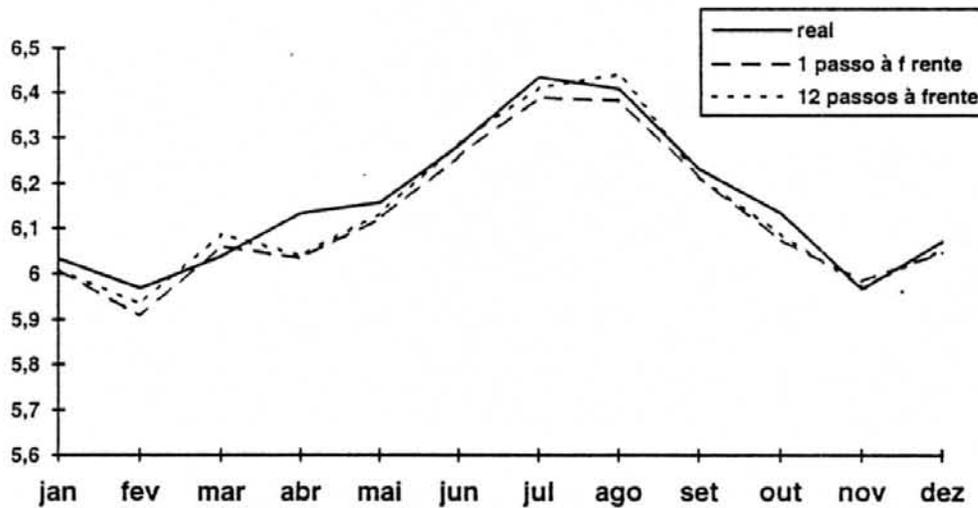


FIGURA 6.2 - Gráfico comparativo dos resultados obtidos pela RNA - Série A

Finalmente é apresentado o gráfico da FIGURA 6.2 onde se pode acompanhar visualmente a evolução dos resultados obtidos para ambas as formas de previsão comparados com os valores reais da série. Note que praticamente em todo o período de previsão a rede consegue acompanhar o movimento da série real.

### 6.1.3 Validação

Nesta fase serão mostrados gráficos comparativos dos resultados obtidos pelas previsões obtidas pelo método UCM e pela Rede Neural para as duas formas de previsão, bem como um gráfico comparando as medidas de erro apresentadas por ambos os métodos.

O gráfico da FIGURA 6.3 mostra a comparação para a previsão um passo à frente. Para esta série, as previsões um passo à frente realizadas pelo método UCM mostraram-se um pouco melhores que as da RNA.

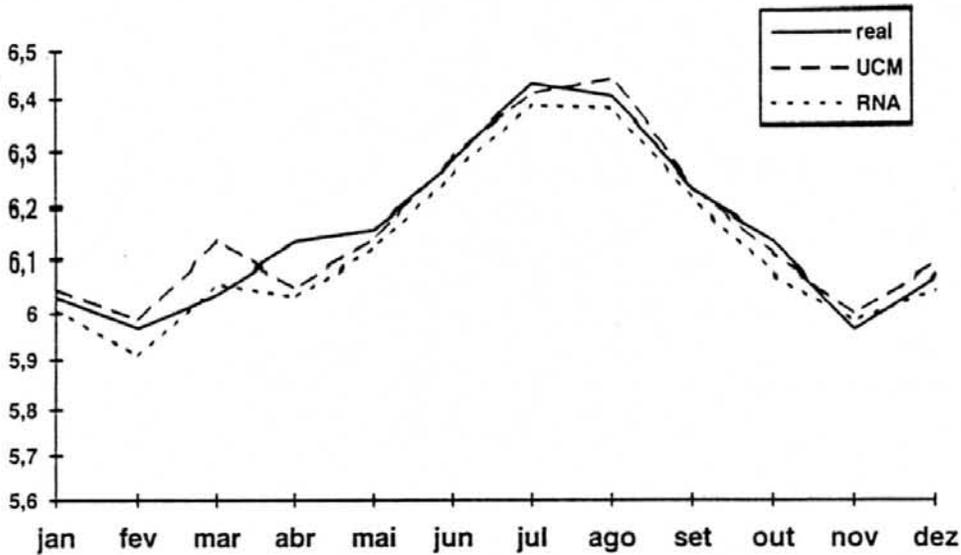


FIGURA 6.3 - UCM x RNA: um passo à frente - Série A

Já o gráfico mostrado na FIGURA 6.4 mostra os resultados dos dois métodos para a previsão doze passos à frente. Neste caso a vantagem é do modelo de Redes Neurais.

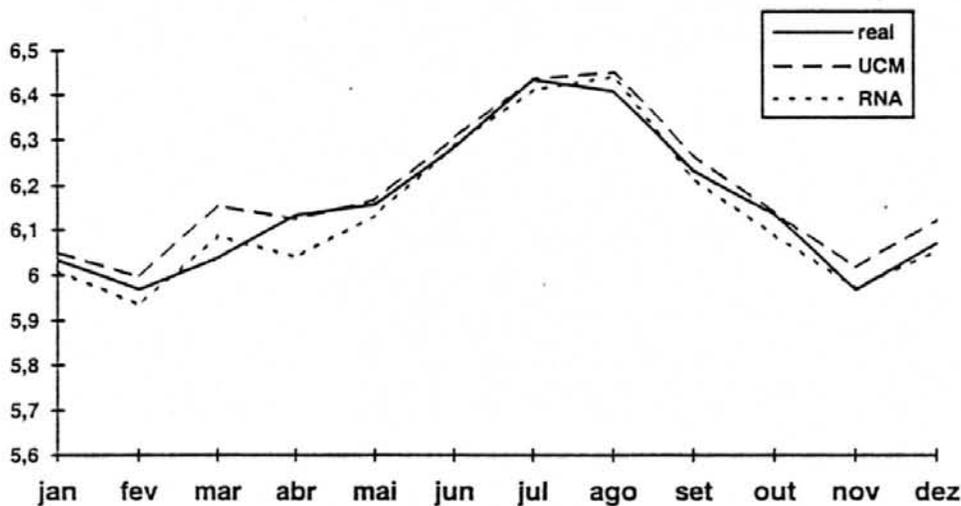


FIGURA 6.4 - UCM x RNA: um passo à frente - Série A

Finalmente, o gráfico da FIGURA 6.5 permite que se faça a avaliação de qual foi o método que obteve o melhor conjunto de previsões.

Como pode ser percebido nesta série não houve diferenças que possam ser consideradas relevantes para que se possa considerar um método superior ao outro, uma vez que ambos os métodos apresentaram prognósticos muito bons.

Ainda assim, devemos ressaltar que para previsões um passo à frente o método UCM bateu o modelo de Redes Neurais por pequena margem. Porém, para previsões doze passos à frente o modelo de Redes Neurais foi o melhor. E dentre as quatro possibilidades de previsão apresentadas foi a Rede Neural com previsões doze passos à frente que obteve, por pequena margem, os melhores resultados.

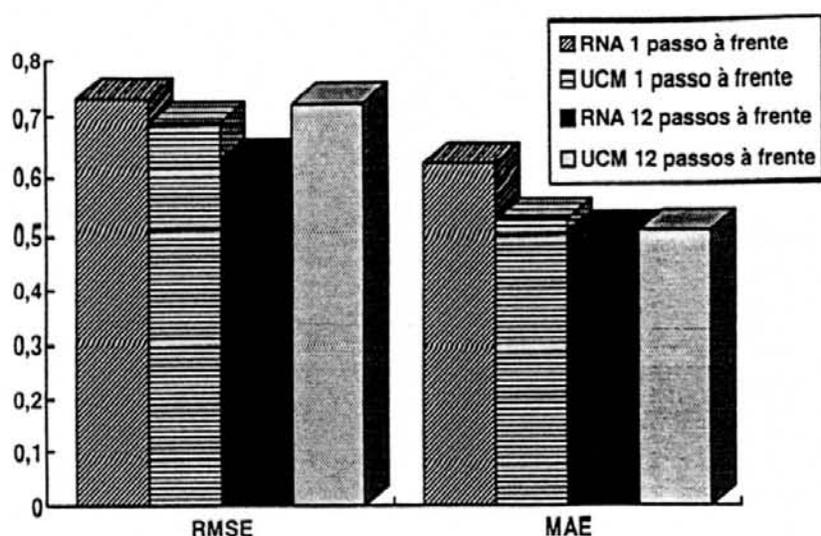


FIGURA 6.5 - Comparação das medidas de erro - Série A

## 6.2 Série B: Índice Pluviométrico Anual em Fortaleza

A única componente não-observável da Série B que deve ser considerada por uma rede que pretenda realizar prognósticos sobre seus valores futuros é a componente **ciclo**, conforme visto na seção 5.2. As componentes **tendência** e **sazonal** não estão presentes na composição da Série B.

Os conjuntos de dados de treinamento e teste utilizados para a previsão desta série são os mesmos utilizados pelo método UCM. Ou seja, o período entre 1849 e 1979 foi utilizado para a fase de aprendizado e o período entre 1980 e 1984 para teste.

### 6.2.1 Modelagem

Como já foi visto anteriormente muitos pesquisadores sustentam que a série correspondente ao índice pluviométrico anual da cidade de Fortaleza contém dois ciclos distintos. Um deles correspondendo a um período de 13 anos e outro a um período de 26 anos. Através da utilização do Teorema de Fourier, que prova que qualquer período pode ser modelado através da combinação de curvas de seno e cosseno, Morettin *et alii* [MOR85] sugerem a seguinte equação para estimar a componente **ciclo** da Série B:

$$Y_t = 142 + 20.3 \cdot \cos(0.24t + 1.99) + 25.6 \cdot \cos(0.48t + 2.2)$$

onde  $Y_t$  assume os valores do ciclo conforme a variação de  $t$  ao longo do tempo.

Tal equação gera o ciclo apresentado no gráfico da FIGURA 6.6, que segundo [MOR85] é uma aproximação suficiente para estimar a componente ciclo da Série B.

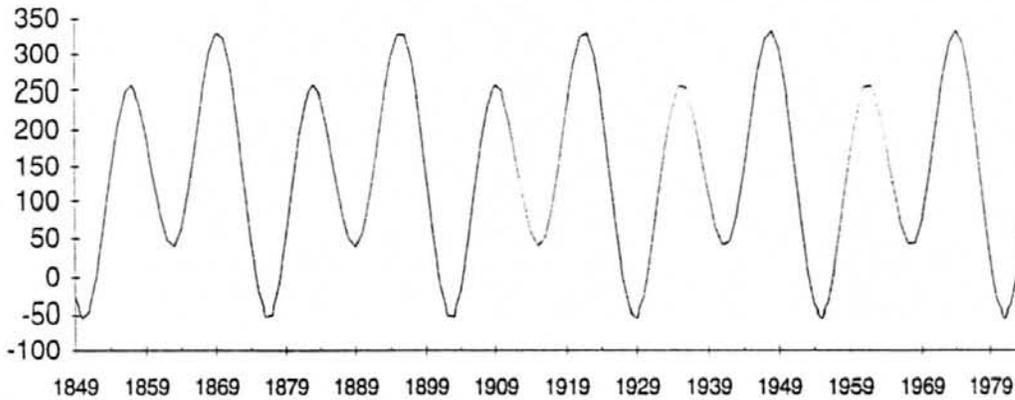


FIGURA 6.6 - Gráfico do ciclo utilizado como entrada da RNA

Esta introdução se fez necessária para demonstrar a origem e os valores que serão informados ao neurônio de ciclo que fará parte da camada de entrada da arquitetura da rede proposta para prever os valores futuros desta série. Neurônio este que será o único da referida camada. Porque uma série que apresenta o ciclo como única componente não-observável presente além da componente irregular não justifica a colocação de um neurônio contendo a observação imediatamente anterior a que se quer prever na camada de entrada da rede, já que esta informação não acrescenta nada para a que a rede aprenda o ciclo.

A FIGURA 6.7 mostra a arquitetura escolhida da rede. Note que foram necessários quatro neurônios na camada intermediária para que a rede não perdesse sua capacidade de generalização. A configuração final da arquitetura é 1-4-1.

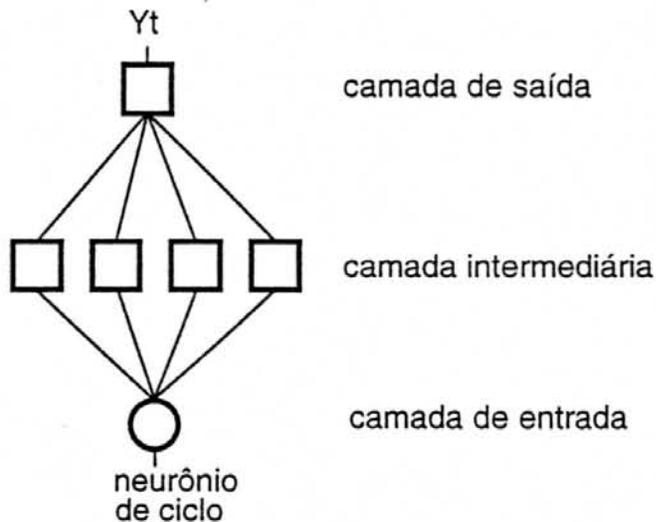


FIGURA 6.7 - Arquitetura do modelo de Rede Neural - Série B

Esta arquitetura foi treinada com os seguintes parâmetros :

- o conjunto de observações foi apresentado 20000 vezes para a rede;
- constante de aprendizado ( $\alpha$ ) = 0.9;
- momento ( $\beta$ ) = 0.6 nas 15000 épocas iniciais e 0.9 nas 5000 restantes;
- tamanho do intervalo de variação dos pesos igual a 2;
- função de transferência do tipo sigmóide.

## 6.2.2 Resultados

Dentro do esperado, devido a análise realizada da componente irregular da série quando da previsão utilizando o método UCM, os resultados das previsões utilizando Redes Neurais para a Série B (mostrados na TABELA 6.3) apresentam alguns erros bastante significativos.

TABELA 6.3 - Resultados RNA - Série B

Período	Real	Um Passo à Frente	Erro Percentual	Doze Passos à Frente	Erro Percentual
1980	109.5	109.67	-0.16%	109.67	-0.16%
1981	110.0	108.92	0.98%	103.94	5.51%
1982	100.4	104.35	-3.78%	107.63	-6.72%
1983	88.4	122.34	-27.74%	118.97	-25.69%
1984	198.1	137.40	30.64%	132.65	33.04%

O grande problema enfrentado pela rede foi a oscilação extremamente acentuada da série real nos anos de 1983 e 1984. Tais anos apresentaram índices pluviométricos completamente fora do ciclo tradicional. O ano de 1983 foi um ano de seca na região de Fortaleza, extraordinariamente seguido por um período de grande quantidade de chuva para os padrões da região. Porém, a análise das três primeiras previsões, tanto um passo à frente quanto cinco passos à frente, permite que se conclua que a rede conseguiu aprender o ciclo proposto. Mantendo seus erros em patamares abaixo de 5% neste três primeiros prognósticos.

TABELA 6.4 - Medidas de erro RNA - Série B

Medida	Um Passo à Frente	Cinco Passos à Frente
RMSE	18.57	19.12
MAE	12.66	14.22

A TABELA 6.4 mostra a comparação das medidas de erro das previsões um passo à frente e cinco passos à frente realizadas pela rede. Como se pode perceber, para esta série a previsão um passo à frente superou a previsão cinco passos à frente.

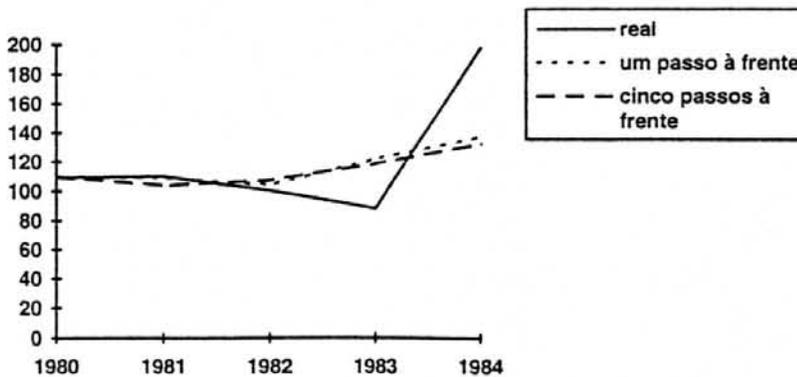


FIGURA 6.8 - Gráfico comparativo dos resultados obtidos pela RNA - Série B

O gráfico da FIGURA 6.8 mostra claramente a dificuldade do modelo em perceber as modificações acentuadas ocorridas nos anos de 1983 e 1984. Entretanto, através do gráfico podemos visualizar a confirmação de que a rede conseguiu realizar prognósticos muito bons para os três primeiros anos do conjunto de teste.

### 6.2.3 Validação

A seguir será feita a comparação dos resultados obtidos pelos métodos UCM e Redes Neurais para a Série B, bem como, uma análise das medidas de erro para encontrar o método que apresentou o menor erro nos prognósticos realizados no computo geral.

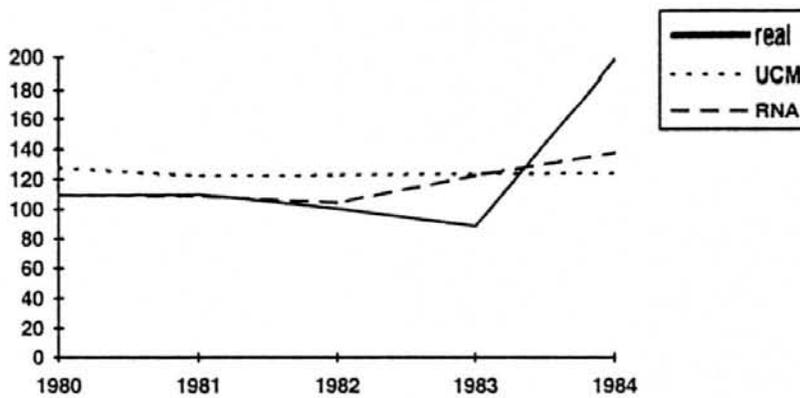


FIGURA 6.9 - UCM x RNA: um passo à frente - Série B

A FIGURA 6.9 apresenta o gráfico comparativo dos resultados um passo à frente. A análise do gráfico já é suficiente para se perceber que o modelo de Rede Neural superou largamente as previsões realizadas pelo método UCM. Mesmo que ambos tenham apresentado erros bastante consideráveis no final do teste, a rede apresentou três prognósticos com um precisão razoável, enquanto que o UCM em momento algum apresentou erro abaixo de 10%. Já no gráfico da FIGURA 6.10 é possível verificar o mesmo comportamento para a previsão cinco passos à frente, ou seja mais uma vez os prognósticos da rede neural foram bem melhores.

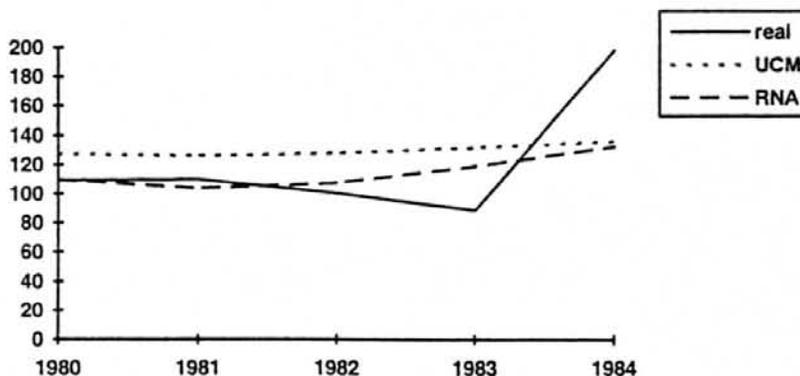


FIGURA 6.10 - UCM x RNA: cinco passos à frente - Série B

No gráfico comparativo da FIGURA 6.11 pode-se perceber claramente que para esta série o modelo de redes neurais utilizado superou amplamente o método UCM. Seus prognósticos foram bem melhores nas duas formas de previsão. Cabendo, desta vez, à previsão um passo à frente feita pela Rede Neural a menor medida de erro.

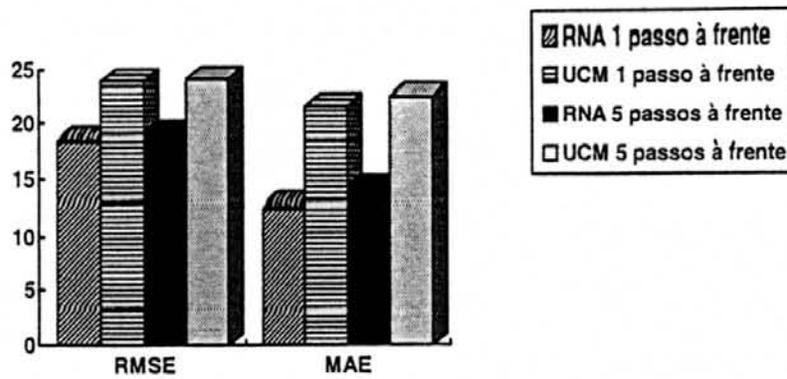


FIGURA 6.11 - Comparação das medidas de erro - Série B

### 6.3 Série C: Índice Mensal da Produção Industrial do Rio Grande do Sul

Como já foi mencionado anteriormente, esta a Série C não é uma série clássica já estudada por muitos especialistas da área de previsão de Séries Temporais anteriormente. Ou seja, não se trata de uma série-exemplo. A intenção, ao se escolher esta série, foi comprovar que os métodos apresentados neste trabalho podem ser utilizados com eficácia para qualquer série, desde que esta apresente as condições mínimas de análise.

A Série C corresponde ao índice de produção industrial do estado do Rio Grande do Sul entre janeiro de 1981 e julho de 1993 e já foi apresentada na seção 5.3. Porém, lembrando, esta série é caracterizada pela presença marcante da componente sazonal e apresenta uma componente **tendência com comportamento estacionário**.

Os conjuntos de dados de treinamento e teste utilizados para a previsão desta série são os mesmos utilizados pelo método UCM. Ou seja, o período entre janeiro de 1981 e dezembro de 1992 foram utilizados para a fase de aprendizado e o período entre janeiro de 1993 e julho de 1993 para a fase de teste.

#### 6.3.1 Modelagem

A modelagem da arquitetura desta rede seguiu a mesma linha da modelagem da Série A. Ou seja, primeiro foram determinados quais seriam os neurônios da camada de entrada da rede e, posteriormente, através da realização dos experimentos, estabeleceu-se o número de neurônios da camada intermediária que melhor generalizou a rede.

Como a Série C apresenta praticamente apenas a componente sazonal, a determinação dos neurônios da sua camada de entrada preocupou-se basicamente com a informação à rede neural dos fatores sazonais. Como na Série A, isto feito através de doze neurônios binários que informam à rede a qual mês do ano pertence a observação que está sendo aprendida. Com o objetivo de captar pequenas variações ocorridas em curtos períodos foi introduzido o neurônio atual -1 que representa o valor da observação do mês anterior ao atual.

A camada intermediária da rede foi determinada após a realização de alguns experimentos. A composição que mostrou melhor generalização da rede apresenta

apenas dois neurônios na camada intermediária. A FIGURA 6.12 apresenta a arquitetura final da rede que é da forma 13-2-1.

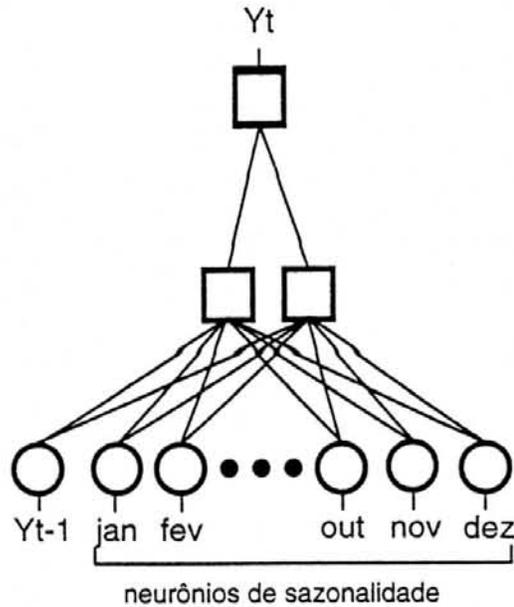


FIGURA 6.12 - Arquitetura do modelo de Rede Neural - Série C

Esta arquitetura foi treinada com o seguinte conjunto de parâmetros:

- o conjunto de observações foi apresentado 15000 vezes para a rede;
- constante de aprendizado ( $\alpha$ ) = 0.9;
- momento ( $\beta$ ) = 0.6 nas 10000 épocas iniciais e 0.9 nas 5000 restantes;
- tamanho do intervalo de variação dos pesos igual a 2;
- função de transferência do tipo sigmóide.

### 6.3.2 Resultados

De acordo com a análise da componente **irregular** da Série C, realizada durante a etapa preparação da série para a realização dos prognósticos pelo método UCM, a qual apresenta participação pequena na composição da série, os resultados obtidos apresentam erros em geral bastante baixos.

TABELA 6.5 - Resultados RNA - Série C

Período	Real	Um Passo à Frente	Erro Percentual	Sete Passos à Frente	Erro Percentual
jan.93	99.70	98.76	0.95%	98.76	0.95%
fev.93	105.33	100.05	5.27%	103.12	2.14%
mar.93	131.95	113.73	16.02%	124.95	5.60%
abr.93	127.39	126.49	0.71%	134.56	-5.33%
mai.93	136.98	134.03	2.20%	137.40	-0.30%
jun.93	135.56	129.74	4.48%	139.50	-2.82%
jul.93	131.84	127.79	3.16%	131.67	0.13%

Através da análise da TABELA 6.5 é possível perceber que, tal qual o método UCM, a rede apresentou dificuldade em prever a grande modificação no índice

apresentada no mês de março. Porém, em ambos casos (um passo à frente e sete passos à frente) a rede apresentou excelente capacidade de recuperação, uma vez que tornou a realizar prognósticos confiáveis.

É importante ressaltar que na previsão sete passos à frente o erro máximo apresentado nos prognósticos da rede esteve sempre em torno de 5%, fato este que não aconteceu com os prognósticos realizados pelo método UCM. A precisão apresentada na previsão sete passos à frente superou as expectativas.

TABELA 6.6 - Medidas de erro RNA - Série C

Medida	Um Passo à Frente	Sete Passos à Frente
RMSE	7.75	4.17
MAE	5.45	3.12

Na TABELA 6.6 acima, são mostradas as medidas de erro para as duas formas de previsão. Note que desta vez a previsão da rede neural múltiplos passos à frente (neste caso, sete) superou a previsão um passo à frente largamente. Isto claramente pode ser observado na análise da TABELA 6.5 onde se pode perceber que os erros da previsão sete passos à frente são menores para quase todas as previsões, exceção feita ao mês de abril.

O gráfico da FIGURA 6.13 ilustra melhor a comparação entre as duas formas de previsão.

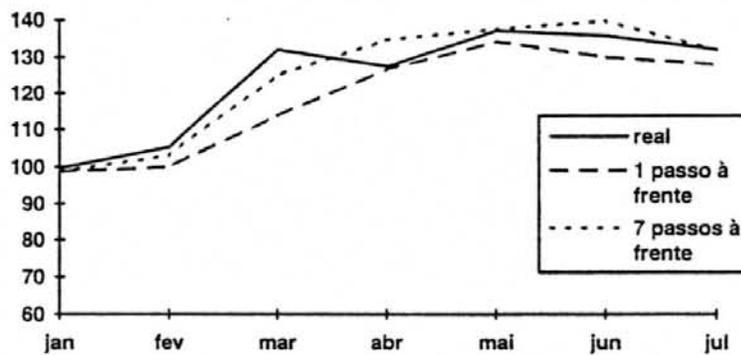


FIGURA 6.13 - Gráfico comparativo dos resultados obtidos pela RNA - Série C

As curvas mostram claramente que as duas formas de previsão conseguem acompanhar a curva das observações reais. Apesar do erro da previsão do mês de março em ambas as formas de previsão, tanto em uma quanto na outra ocorre a recuperação praticamente nos mesmos níveis. A grande diferença entra as duas formas de previsão ocorre justamente no erro apresentado no mês de março. Enquanto para a previsão um passo à frente o erro em março foi superior a 16%, na previsão sete passos à frente o erro para o mesmo mês o erro é de 5.60%.

### 6.3.3 Validação

A seguir será feita a comparação dos resultados obtidos pelos métodos UCM e Redes Neurais para as previsões realizadas para a Série C, bem como, uma análise das

medidas de erro para encontrar o método que apresentou o menor erro nos prognósticos realizados no computo geral.

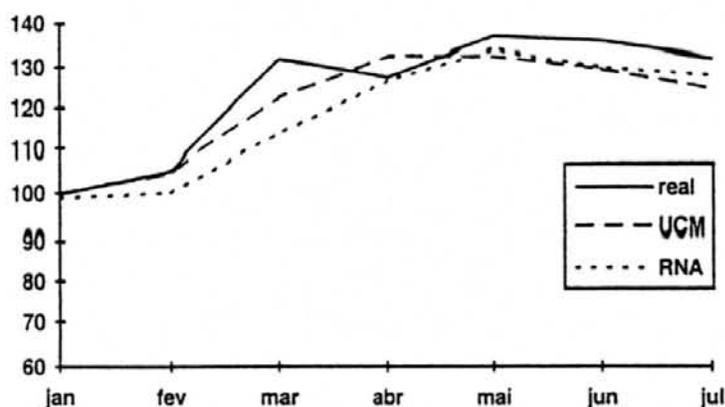


FIGURA 6.14 - UCM x RNA: 1 passo à frente - Série C

Acompanhando o gráfico da FIGURA 6.14, nota-se que o método UCM realizou prognósticos quase sempre mais próximos da curva das observações reais. Demonstrando uma melhor performance no computo geral para a previsão um passo à frente. Já na previsão sete passos à frente o fenômeno se inverte. Para este caso, embora de forma bastante apertada, a rede neural consegue realizar prognósticos mais próximo da curva das observações reais em praticamente todos os meses (exceção feita ao mês de abril), conforme podemos observar no gráfico da FIGURA 6.15 abaixo.

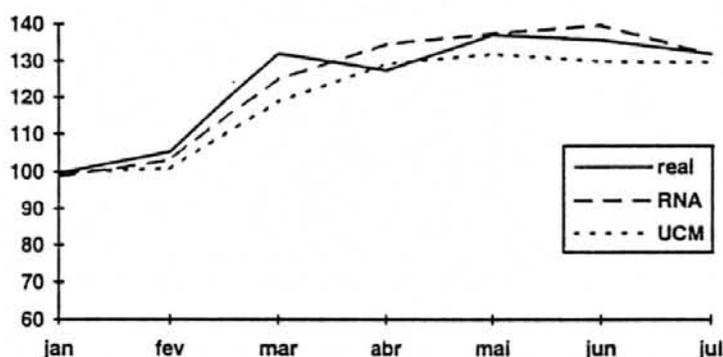


FIGURA 6.15 - UCM x RNA: sete passos à frente - Série C

Com a finalidade de estabelecer qual dos métodos apresenta os melhores prognósticos considerando as duas formas de previsão utilizadas por cada um dos métodos, é apresentado o gráfico da FIGURA 6.16 que mostra a comparação das duas medidas de erro, RMSE e MAE, para cada forma de previsão.

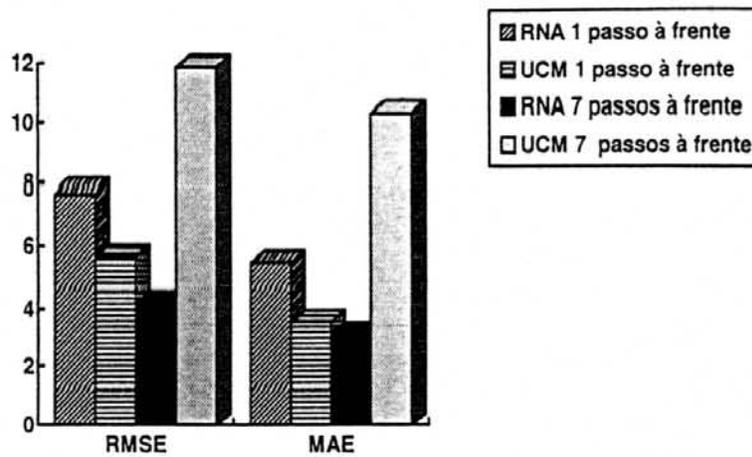


FIGURA 6.16 - Comparação das medidas de erro - Série C

Neste gráfico pode-se perceber que, mais uma vez, no cômputo geral as previsões realizadas pelas Redes Neurais são as melhores. A rede neural com previsão sete passos à frente superou o teste oferecido pelo método UCM na forma de previsão um passo à frente. Também vale ressaltar que mais uma vez a pior das quatro previsões foi a realizada pelo método UCM múltiplos passos à frente (neste caso, sete). O que indica uma certa dificuldade deste método para esta forma de previsão. Justamente a forma onde as redes neurais conseguem sempre resultados excelentes.

## 7 CONCLUSÃO

O objetivo inicial deste trabalho era comprovar que as Redes Neurais Artificiais seriam capazes de realizar prognósticos confiáveis de valores futuros de Séries Temporais.

O passo inicial deste trabalho foi a definição do modelo de Redes Neurais Artificiais a ser utilizado. Consultando a literatura da área, foi possível identificar o modelo dos Perceptrons Multi-Camadas com aprendizado por Retropropagação de Erros como o mais utilizado na realização de prognósticos de valores futuros de Séries Temporais.

Uma vez definido qual modelo de Redes Neurais Artificiais seria utilizado, a preocupação passou a ser a definição de qual método estatístico tradicional seria utilizado como parâmetro de comparação. Ou seja, buscava-se um método capaz de oferecer resultados tão bons que exigissem das RNAs resultados bastante expressivos. Muitos trabalhos, como já foi visto anteriormente, já haviam comparado o poder preditivo das RNAs com um dos métodos estatísticos de previsão mais tradicionais, o modelo ARIMA de Box & Jenkins [BOX76]. Tais trabalhos invariavelmente conseguiram comprovar que as Redes Neurais conseguiam resultados melhores que os do método ARIMA. Para este trabalho, porém, os resultados obtidos pelo método representado pelo modelo ARIMA não foram considerados como um parâmetro de comparação satisfatório. Isto porque, durante o estudo realizado sobre os métodos estatísticos existentes, foi possível constatar que o Modelo Estrutural Clássico (*Unobservable Components Models - UCM*) invariavelmente oferecia previsões mais acuradas que as do modelo ARIMA [HAR89]. Em vista disso, este trabalho optou por adotar o método UCM como parâmetro a ser igualado, ou quem sabe superado pelas RNAs.

Para que fosse executada tal comparação optou-se pela realização de experimentos envolvendo três séries reais de índices numéricos. Duas séries consideradas clássicas dentro da literatura de métodos estatísticos de previsão de Séries Temporais, isto é, séries que já foram estudadas e trabalhadas por muitos pesquisadores. E uma série não-clássica, ou seja uma série que ainda não aparecia na literatura. As séries escolhidas foram as seguintes:

- **Série A:** índice mensal de passageiros de avião nas linhas aéreas americanas (tendência crescente e sazonalidade);
- **Série B:** índice pluviométrico anual na cidade de Fortaleza, estado do Ceará, Brasil (ciclo);
- **Série C:** índice mensal de produção industrial do estado do Rio Grande do Sul, Brasil (sazonalidade e tendência estacionária).

Por fim, era preciso definir uma metodologia empírica para a realização da modelagem da arquitetura da Rede Neural que seria utilizada na realização dos prognósticos de cada série acima relacionada.

Após a execução de cada um dos passos acima mencionados, este trabalho passou para a fase seguinte que foi a da realização dos experimentos.

## 7.1 Análise dos Resultados Obtidos

A tabela abaixo mostra o resultado final alcançado pelos experimentos realizados por este trabalho.

TABELA 7.1 - Desempenho de cada método por série estudada

Modelo Utilizado	Série A	Série B	Série C
RNA um passo à frente	3 <sup>a</sup>	1 <sup>a</sup>	3 <sup>a</sup>
RNA vários passos à frente	1 <sup>a</sup>	2 <sup>a</sup>	1 <sup>a</sup>
UCM um passo à frente	2 <sup>a</sup>	3 <sup>a</sup>	2 <sup>a</sup>
UCM vários passos à frente	4 <sup>a</sup>	4 <sup>a</sup>	4 <sup>a</sup>

Como se pode perceber após a análise da TABELA 7.1 o objetivo principal deste trabalho foi plenamente alcançado. Para as três séries estudadas, a previsão realizada pelas Redes Neurais Artificiais alcançou melhores resultados na comparação com os prognósticos do método estatístico UCM.

Se a análise for feita tomando-se as quatro diferentes formas de previsão utilizadas neste trabalho, a que se sai melhor sem dúvida alguma é a representada pelas previsões realizadas pelas RNAs múltiplos passos à frente. Tanto nos prognósticos realizados para a Série A como para a Série C esta forma de previsão obteve os melhores resultados. Entretanto, nestas duas séries o método UCM com previsão um passo à frente foi o que obteve o segundo melhor resultado exigindo uma performance muito boa da RNA e superando as previsões realizadas pelas RNAs na forma um passo à frente. E aparecendo com os piores resultados em ambas, encontra-se o método UCM na forma de previsão múltiplos passos à frente.

Diferentemente das duas séries anteriores, a Série B apresentou uma superioridade muito grande nos prognósticos realizados pelas RNAs, tanto na forma um passo à frente quanto múltiplos passos à frente. Isto talvez se deva ao fato desta ser a única série que apresentava a característica marcante do componente ciclo. E, pelos resultados, as RNAs demonstraram uma capacidade bem superior de aprender o ciclo na comparação com o método UCM.

Este trabalho não possui a pretensão de estabelecer que as Redes Neurais Artificiais, mais especificamente o modelo Perceptron Multi-Camadas com aprendizado por Retropropagação de Erros, apresentam-se como o melhor método de realização prognósticos de valores futuros de Série Temporais. Pois o trabalho realizado não pode ser considerado abrangente o suficiente para uma afirmação de tamanha importância. Porém, em face do apresentado, pode-se dizer com segurança que as RNAs são ferramentas poderosas para a realização de previsão de Séries Temporais, capazes de realizar prognósticos com o mesmo nível de precisão ou até maior que um dos métodos estatísticos mais respeitados, embora não possam oferecer considerações estatísticas sobre a série devido a sua natureza.

## 7.2 Trabalhos Futuros

A principal dificuldade na utilização de Redes Neurais Artificiais na previsão de Séries Temporais ainda é a determinação da arquitetura ótima da rede.

Ainda não se possui uma metodologia consistente que apresente a melhor configuração da rede para cada série proposta. A metodologia proposta por este trabalho no capítulo 4, apresenta seu fator limitante na determinação do número de neurônios da camada intermediária da rede. Uma sugestão que pode vir a ser estudada em trabalhos futuros é a utilização de modelos de RNAs com poda automática de neurônios cuja importância se mostra menor ao longo do aprendizado da rede. Um modelo que desperta particular interesse é o **Modelo Neural Combinatório** [MAC92].

## BIBLIOGRAFIA

- [BEA91] BEALE, R. ; JACKSON, T. **Neural Computing: An Introduction**. New York: Adam Hilger Publ, 1991. 279 p.
- [BOX76] BOX, G. E. ; JENKINS, G. M. **Time Series Analysis, Forecasting and Control**. San Francisco: Holden-Day, 1976. 321 p.
- [BRO63] BROWN, R.G. **Smoothing, Forecasting and Prediction**. Englewood Cliffs: Prentice Hall, 1963. 214 p.
- [CAR89] CARBONELL, J. G. **Introduction: Paradigms for Machine Learning, Artificial Intelligence**, Amsterdam, v. 40, p. 1-9, Apr.1989.
- [CAR88] CARPENTER, G. ; GROSSBERG, S. The ART of Adaptive Pattern Recognition by a Self-organizing Neural Network. **IEEE Computer**, New York, v.21, n.3, p.77-88, Mar. 1988.
- [CAV88] CARVALHO, L.A.V. **Redes Neurais e a Tradição Conexionista da Inteligência Artificial**. Rio de Janeiro: COPPE/UFRJ, 1988 (Relatório Interno).
- [CAU90] CAUDILL, M. Neural Network Primer. **AI Expert**, San Francisco, p.1-64, 1990. Special Issue.
- [COH83] COHEN, M. ; GROSSBERG, S. Absolute Stability of Global Pattern Formation and Parallel Memory Storage by Competitive Neural Networks. **IEEE Transactions on Systems, Man and Cybernetics**, New York, v.13, p. 815-825, Jan.1983.
- [COT85] COTTRELL, G.W. **A Connectionist Approach to Word Sense Disambiguation**. Rochester: University of Rochester, 1985. Tese de Doutorado.
- [FER94] FERNANDES, L. G. ; PORTUGAL, M. S. ; NAVAU, P. Um Estudo Experimental do Poder Preditivo das Redes Neurais Artificiais Comparado a Métodos Econométricos Tradicionais. In: SIMPÓSIO BRASILEIRO DE REDES NEURAS, 1., 1994, Caxambu, Minas Gerais. **Anais ... [S.l.:s.n.]**, 1994. p. 56-61.
- [GAL88] GALLANT, S. I. Connectionist Expert Systems, **Communications of the ACM**, New York, v. 31, n. 2, p. 152-168, Feb.1988;
- [HAR71] HARRISON, P. J. ; STEVENS, C. F. A Bayesian Approach to Short Term Forecasting , **Operational Research Quarterly** , Cambridge, n. 22, p.341-362, 1971.
- [HAR84] HARVEY, A. C. A Unified View of Statistical Forecasting Procedures, **Journal of Forecasting**, Cambridge, n.3, p. 245-283, 1984.
- [HAR89] HARVEY, A. C. **Forecasting Structural Time Series Models and the Kalman Filter**. Cambridge: Cambridge University Press, 1989. p. 458.
- [HEB49] HEBB, D. O. **The Organization of Behavior**. New York: John Wiley & Sons, 1949.

- [HEC89] HECHT-NIELSEN, R. Applications of Counterpropagation Networks. **Neural Networks**, Elmsford, NY, v.1, n.1, p.131-140, May 1989.
- [HIN89] HINTON, G. E. Connectionist Learning Procedures. **Artificial Intelligence**, Amsterdam, v. 40, p. 185-234, Apr. 1989.
- [HOP82] HOPFIELD, J. J. Neural Networks and Physical Systems with Emergent Collective Computational Abilities. **Proceedings of National Academy of Sciences**, Washington, v. 79, p. 2554-2558, Apr.1982.
- [KAH60] KAHLMAN, R. E. A New Approach to Linear Filtering and Prediction Problems. **Transactions of the ASME**, [S.l.], n. 57, p.357-361, 1960.
- [KAN86] KANE, R. P. ; TRIVEDI, N.B. Are Droughts Predictable? **Climate Change**, [S.l.], n. 8, p.208-223, 1986.
- [KIR83] KIRKPATRICK, S. ; GELLAT, C. D. ; VECCHI, M. P. Optimization by Simulated Annealing. **Science**, [S.l.], n. 220, p. 671-680, 1983.
- [KOH82] KOHONEN, T. Clustering Taxonomy and Topological Maps of Patterns. In: INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION, 6., 1982. **Proceedings ...** [S.l.:s.n.], 1982. p.114-128.
- [KOS88] KOSKO, B. Feedback Stability and Unsupervised Learning, In: INTERNATIONAL CONFERENCE ON NEURAL NETWORKS, 1988, San Diego. **Proceedings ...** [S.l.:s.n.], 1988. p.141-152.
- [LIP87] LIPPMAN, R. P. An Introduction to Computing with Neural Nets. **IEEE Acoustics, Speech and Signal Processing Magazine**, New York, v.3, n.4, p. 4-22, Apr. 1987.
- [MAC92] MACHADO, R. J. ; ROCHA, A. F. Evolutive Fuzzy Neural Networks. In: INTERNATIONAL CONFERENCE ON FUZZY SYSTEMS, 1992, San Diego. **Proceedings ...** [S.l.:s.n.], 1992.
- [MEL91] MELO, M. P. **Redes Neurais Artificiais: uma Aplicação à Previsão de Preços de Derivados de Petróleo**. Rio de Janeiro: Departamento de Informática - PUC, 1991. 143 p. Dissertação de Mestrado.
- [MIN69] MINSKY, M.; PAPPERT, S. **Perceptrons: an Introduction to Computational Geometry**. Cambridge: MIT Press, 1969. 666 p.
- [MOR85] MORETTIN, P.A. ; MESQUITA, A.R. ; ROCHA, J.G.C. Rainfall at Fortaleza in Brazil Revisited. In: ANDERSON, O. D. et al. (Eds.). **Time Series Analysis: Theory and Practice** . Amsterdam: North Holland, 1985. v.6.
- [NEU91] NEURALWARE INCORPORATION. **Neural Works Professional: User Guide**. USA: Neuralaware Inc., 1991. 320 p.
- [POR93] PORTUGAL, M. S. Measures of Capacity Utilization: Brazil 1920-1988. **Análise Econômica**, Porto Alegre, v.11, n.9, p. 89-102, Sept.1993.
- [POR94] PORTUGAL, M. S. Neural Networks Versus Time Series: a Forecasting Exercise. In: INTERNATIONAL SYMPOSIUM ON FORECASTING, 14., 1994, Estocolmo, Suécia. **Proceedings ...** Estocolmo:[s.n.], 1994.
- [REF91] REFENES, A. N. Constructive Learning and its Applications to Currency Exchange Rate Forecasting. In TURBAN, E. ; TRIPPI, R. (Eds.). **Neural**

**Networks Applications in Investment and Finance Services.** New York: Probus Publishing, 1991.

- [RUM85] RUMELHART, D.E. ; ZIPSER, D. **Feature Discovery by Competitive Learning.** La Jolla: University of California at San Diego, Institute of Cognitive Science, 1985. 151 p.
- [RUM86] RUMELHART, D. E. ; McCLELLAND, J. **Parallel Distributed Processing.** Cambridge: MIT Press, 1986. v. 1, 547 p.
- [RUM86a] RUMELHART, D. E. ; HINTON, G. E. ; WILLIAMS, R. J. **Learning Internal Representations by Error Propagation.** La Jolla: University of California at San Diego, Institute of Cognitive Science, 1986. (ICS Report, 8506).
- [RUM90] RUMELHART, D. E. ; WEIGEND, A. **Predicting the Future: a Connectionist Approach.** Stanford: [s.n.] , 1990. (PDP-90-01, PARC-SSL-90-20).
- [SIM90] SIMPSON, P. **Artificial Neural Systems: Foundations, Paradigms, Applications and Implementations.** New York: Pergamon Press Inc., 1990.
- [SOU89] SOUZA, R. C. Modelos Estruturais para a Previsão de Séries Temporais, Abordagens Clássica e Bayesiana. In: COLÓQUIO BRASILEIRO DE MATEMÁTICA, 17., 1993. **Anais ...** [S.l.:s.n.], 1993.
- [SOU93] SOUZA, R. C. ; ZANDONADE, E. **Forecasting Via Neural Networks: A Comparative Study.** Rio de Janeiro: Department of Electrical Engineering, Catholic University of Rio de Janeiro (PUC-RJ), 1993.
- [VAR90] VARFIS, A.; VERSINO, C. **Univariate Economic Time Series Forecasting.** Cambridge: Cambridge University Press, 1990. 225 p.
- [WAS89] WASSERMAN, P. D. **Neural Computing: Theory and Practice.** New York: Van Nostrand Reinhold, 1989. 230 p.
- [WES89] WEST, M. ; HARRISON, J. **Bayesian Forecasting and Dynamic Models.** New York: Springer Verlag, 1989. 237 p.
- [ZAD65] ZADEH, L. A. Fuzzy Sets. **Information and Control.** [S.l.], v.8, p.338-353, 1965.
- [ZAN93] ZANDONADE, E. **Aplicação da Metodologia de Redes Neurais em Previsão de Séries Temporais.** Rio de Janeiro: Departamento de Engenharia Elétrica, PUC-RJ, 1993. 139 p. Dissertação de Mestrado.

**Informática**



**UFRGS**

*Utilização de Redes Neurais na Análise e Previsão de Séries Temporais.*

por

Luiz Gustavo Leão Fernandes

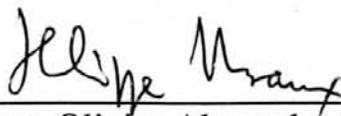
Dissertação apresentada aos Senhores:

  
\_\_\_\_\_  
Prof. Dr. Atos Freitas Grawunder (PPGE/UFRGS)

  
\_\_\_\_\_  
Prof. Dra. Beatriz de Faria Leão

  
\_\_\_\_\_  
Prof. Dr. Dante Augusto Couto Barone

Vista e permitida a impressão.  
Porto Alegre, 12/07/96.

  
\_\_\_\_\_  
Prof. Dr. Philippe Olivier Alexandre Navaux,  
Orientador.

  
\_\_\_\_\_

Prof. Flávio Rech Wagner  
Coordenador do Curso de Pós-Graduação  
em Ciência da Computação - CPGCC  
Instituto de Informática - UFRGS