

UNIVERSIDADE FEDERAL DE PERNAMBUCO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

**O USO DE REDES NEURAS ARTIFICIAIS NA PREVISÃO DE
TENDÊNCIAS NO MERCADO DE AÇÕES**

DISSERTAÇÃO SUBMETIDA À UFPE
PARA OBTENÇÃO DO GRAU DE MESTRE
POR

FELIPPE AQUINO DE MOURA

Orientador: Prof. Francisco de Souza Ramos, Ph.D.

RECIFE, DEZEMBRO/2006

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

M929u

Moura, Felipe Aquino de

O uso de redes neurais artificiais na previsão de tendências no mercado de ações / Felipe Aquino de Moura. – Recife: O Autor, 2006.

xiv, 105 f.; il., gráfs., tabs.

Dissertação (Mestrado) – Universidade Federal de Pernambuco. CTG. Programa de Pós-Graduação em Engenharia de Produção, 2006.

Inclui referências bibliográficas, glossário e apêndice.

1. Engenharia de Produção. 2. Redes Neurais – Mercado de Ações. I. Título.

658.5 CDD (22.ed.)

UFPE/BCTG/2007-

37



UNIVERSIDADE FEDERAL DE PERNAMBUCO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

PARECER DA COMISSÃO EXAMINADORA
DE DEFESA DE DISSERTAÇÃO DE MESTRADO ACADÊMICO DE

FELIPPE AQUINO DE MOURA

***“O Uso de Redes Neurais Artificiais na Previsão de Tendências
no Mercado de Ações”***

ÁREA DE CONCENTRAÇÃO: PESQUISA OPERACIONAL

A comissão examinadora, composta pelos professores abaixo, sob a presidência do primeiro, considera o candidato FELIPPE AQUINO DE MOURA **APROVADO**.

Recife, 26 de dezembro de 2006.

Prof. FRANCISCO DE SOUSA RAMOS, Docteur (UFPE)

Prof. ENRIQUE ANDRÉS LÓPEZ DROGUETT, PhD (UFPE)

Prof. ALEXANDRE STAMFORD DA SILVA, Doutor (UFPE)

Para Agamenon e Verônica

AGRADECIMENTOS

- Aos meus pais e irmãs, que sempre me ajudaram nos momentos difíceis. O que enfrentei e venci na vida foi por vocês. Espero um dia retribuir tudo que fizeram.
- A todos meus familiares, por sempre estarem presentes.
- Ao meu tio Toninho, que me mostrou o caminho da engenharia quando eu ainda era uma criança, me incentivando a estudar na Escola Técnica. Sem seu apoio não teria chegado até aqui.
- À minha querida namorada, Camila, que teve a paciência e compreensão de sacrificar inúmeros fins de semana, sabendo que se tratava de um bem maior para o meu futuro. Além de ter me ajudado auxiliando na correção gramatical deste texto.
- À Joaquim e Rose que tão bem me receberam em sua casa, fornecendo uma estada para a elaboração desta dissertação.
- Aos meus grandes amigos, Eduardo e Felipe, que ajudaram a “diminuir” a distância entre Recife e Natal. Sem a colaboração deles, este trabalho não teria sido concluído.
- Ao meu orientador, Prof. Francisco Ramos, por seu apoio e orientação acadêmica, e principalmente por sempre acreditar em minha capacidade, apesar de todas os percalços neste caminho.
- Aos colegas de mestrado, que se tornaram amigos nesta jornada.
- Ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) e à CAPES (Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pelo financiamento do meu curso de pós-graduação.
- À UFPE – Universidade Federal de Pernambuco, instituição que me deu a oportunidade de realizar esta pós-graduação.

RESUMO

O mercado de ações é considerado uma opção de investimento de alto retorno, dominado pela incerteza e volatilidade. A realização da previsão do movimento deste mercado não é uma tarefa simples, pois está sujeito a diversos fatores econômicos, políticos e até mesmo psicológico. Os tradicionais métodos estatísticos e as análises existentes (técnica e fundamentalista) não se mostram capazes de identificar as relações não-lineares entre as diversas variáveis que compõem o preço de uma ação e os seus movimentos de alta e baixa, sendo necessárias o uso de técnicas mais avançadas como Redes Neurais Artificiais.

Redes Neurais Artificiais (RNAs) são uma ferramenta que simulam a habilidade de aprendizado do cérebro humano. Redes neurais possuem, entre outras, a capacidade de modelar funções não-lineares em ambientes complexos e com informações com ruídos ou parciais. Conseqüentemente têm sido cada vez mais utilizadas para realizar previsões, inclusive no mercado de ações.

Neste trabalho serão desenvolvidos modelos de redes neurais, com o intuito de realizar previsões de valores presentes e futuros de ações e suas tendências futuras de alta e baixa. Foram avaliadas diferentes formas de arquitetura, utilizando sempre como base uma rede direta *perceptron* multi-camadas (MLP). O estudo foi realizado primeiramente para a previsão diária e futura da ação preferencial da Petrobras e posteriormente estendido para a previsão de tendência de um e dois dias futuros deste ativo e do índice da Bolsa de Valores de São Paulo (Ibovespa).

Os modelos estudados apresentaram um elevado grau de acerto na previsão de tendências de alta e baixa dos ativos em questão, sendo possível concluir que redes neurais podem ser utilizadas pelo investidor para auxiliá-lo no gerenciamento de sua carteira de investimentos.

Palavras-chave: previsão, mercado de ações, redes neurais.

ABSTRACT

Stock markets are considered a high return investment option, dominated by uncertainty and volatility. The forecast of the movement of that market is a difficult task, because is influenced by many economical, political and even psychological factors. The traditional statistical methods and the known analysis (technical and fundamental) are not capable to identify the non-linear relations between the variables that compose the price of a stock and the trends of the market, became necessary the use of more advanced techniques like Artificial Neural Networks (ANNs).

Artificial Neural Networks (ANNs) are a tool that simulate the ability of learning of the human brain. Neural Networks have, among others, the capacity of model non-linear functions in complex environments and with noisy or partial information. Thus, its use in forecast is growing, including stock markets.

In this work will be developed models of neural networks to predict futures values of stocks and its trends. Different forms of architecture were considered, always using a multi-layer perceptron (MLP). The initial simulation had the objective of forecasting the daily and future value of the preferential stock of the Petrobras, and lately had been extended for the forecasting of the trends of one and two futures days of the former stock and of the index of the Bolsa de Valores de São Paulo (Ibovespa).

The studied models revealed a high level of accuracy in trends prediction of the assets analyzed, thus neural networks can be used by investors to assist them in the management of theirs investments portfolio.

Keywords: forecasting, stock market, neural networks.

SUMÁRIO

AGRADECIMENTOS	iv
RESUMO	v
ABSTRACT	vi
SUMÁRIO.....	vii
LISTA DE FIGURAS	x
LISTA DE GRÁFICOS.....	xi
LISTA DE TABELAS	xiii
1. INTRODUÇÃO.....	1
1.1. Motivação	1
1.2. Objetivos da Dissertação	2
1.3. Organização da Dissertação.....	3
2. METODOLOGIA.....	5
2.1. Introdução.....	5
2.2. Redes Biológicas: a motivação.....	5
2.3. Neurônios artificiais: um breve histórico	6
2.4. Redes Neurais Artificiais: o neurônio artificial.....	8
2.4.1. Tipos de Função de Ativação	9
2.5. Arquitetura de Redes Neurais.....	12
2.6. Processo de Aprendizado de RNAs.....	14
2.6.1. Aprendizado supervisionado	15
2.6.2. Aprendizado não-supervisionado	16
2.6.3. Aprendizado Híbrido	16
2.7. Projeto de uma rede neural artificial.....	16
2.7.1. Coleta e separação dos dados	17
2.7.2. Seleção dos parâmetros da RNA	17
2.7.3. Treinamento.....	17
2.7.4. Teste e Validação.....	18
2.8. Algoritmo <i>Backpropagation</i>	18
2.8.1. Perceptron.....	18
2.8.2. Perceptron multi-camadas	18
2.8.3. O algoritmo <i>backpropagation</i>	19
2.9. Algoritmos Genéticos	21
2.9.1. Introdução aos algoritmos genéticos	21
2.9.2. Funcionamento de um algoritmo genético	22
2.9.3. Algoritmo genético simples.....	24
2.10. Aplicações de RNA	25

2.11.	Software.....	26
3.	REVISÃO BIBLIOGRÁFICA.....	27
3.1.	Introdução.....	27
3.2.	Motivações e limitações para o uso de RNAs em previsões.....	27
3.3.	Metodologia de RNAs para previsão.....	30
3.3.1.	Arquitetura da rede neural.....	30
3.3.1.1.	Decisão do número de neurônios na camada de entrada.....	31
3.3.1.2.	Decisão do número de camada ocultas e de seu número de neurônios.....	31
3.3.1.3.	Decisão do número de neurônios na camada de saída.....	32
3.3.2.	Função de transferência.....	33
3.3.3.	Algoritmo de treinamento.....	33
3.3.4.	Tratamento dos dados.....	34
3.3.5.	Separação dos dados em conjuntos de treinamento e teste.....	35
3.3.6.	Medidas de performance.....	35
3.4.	Aplicações de RNAs como ferramentas de previsão.....	36
3.4.1.	Aplicações em mercados financeiros.....	39
4.	JUSTIFICATIVA.....	41
4.1.	O mercado de ações.....	41
4.2.	Análises do mercado de ações.....	41
4.2.1.	Hipótese do Mercado Eficiente.....	41
4.2.2.	Análises Técnica e Fundamentalista.....	42
4.2.2.1.	Análise Técnica.....	42
4.2.2.2.	Análise Fundamentalista.....	43
4.3.	Volatilidade do mercado de ações.....	43
4.4.	O mercado de ações brasileiro.....	44
5.	ESTIMAÇÕES E RESULTADOS.....	47
5.1	Introdução.....	47
5.2	Descrição do modelo de RNA utilizado.....	47
5.2.2.	Seleção e processamento dos dados.....	47
5.2.2.	Arquitetura e topologia.....	50
5.2.3.	Algoritmo de treinamento.....	50
5.2.4.	Medição de performance de uma RNA.....	50
5.3.	Estimações.....	51
5.3.1.	Previsão do valor atual da ação Petrobras PN.....	51
5.3.1.1.	Resultados.....	52
5.3.2.	Previsão do valor futuro da ação Petrobras PN.....	56
5.3.2.1.	Resultados.....	58
5.3.3.	Previsão da variação do valor futuro da ação Petrobras PN.....	62
5.3.3.1.	Resultados.....	63
5.4.	Resultados e discussões.....	64
5.4.1.	Previsão da tendência diária futura de ações.....	64
5.4.1.1.	Resultados.....	67
5.4.2.	Previsão da tendência de ações em dois dias futuros.....	73
5.4.2.1	Resultados.....	75
6.	CONCLUSÕES.....	87
6.1.	Comentários iniciais.....	87

6.2.	Conclusões.....	88
6.3.	Sugestões para trabalhos futuros	89
	REFERÊNCIAS BIBLIOGRÁFICAS	91
	GLOSSÁRIO.....	97
	A GRÁFICOS DAS VARIÁVEIS MACROECONÔMICAS UTILIZADAS – APÓS TRANSFORMAÇÃO LINEAR.....	99

LISTA DE FIGURAS

Figura 2.1: Esquema de um neurônio.....	6
Figura 2.2: o modelo de McCulloch e Pitts.....	7
Figura 2.3: modelo não-linear de um neurônio.....	8
Figura 2.4: Rede direta de uma única camada (adaptada de Hayken, 1994).....	13
Figura 2.5: Rede direta de múltiplas camadas (adaptada de Hayken, 1994).....	13
Figura 2.6: Modelo simplificado de uma rede recorrente (adaptada de Hayken, 1994).	14
Figura 2.7: Aprendizado Supervisionado (adaptada de Hayken, 1994).	15
Figura 2.8: ajuste dos pesos pela regra Delta (adaptada de Mello, 2004).	20
Figura 2.9: Fluxograma simplificado de um AG (adaptada do manual do Matlab).....	22
Figura 2.10: Exemplo de uma matriz para a população inicial (adaptada do manual do Matlab).	23
Figura 2.11: Criação de uma nova geração em um AG (adaptado do manual do Matlab).	24

LISTA DE GRÁFICOS

Gráfico 2.1: função degrau.	9
Gráfico 2.2: função linear.	10
Gráfico 2.3: função sigmoidal.	10
Gráfico 2.4: função tangente hiperbólica.	11
Gráfico 2.5: função sigmoidal linear.	11
Gráfico 2.5: função tangente hiperbólica linear.	11
Gráfico 4.1: Volume negociado na Bovespa (Fonte: Bovespa)	45
Gráfico 5.1: gráfico das entradas utilizadas.....	49
Gráfico 5.2: resultado do Modelo 1.....	53
Gráfico 5.3: resultado do Modelo 2.....	54
Gráfico 5.4: resultado do Modelo 3.....	58
Gráfico 5.5: resultado do Modelo 4.....	59
Gráfico 5.6: resultado do Modelo 5.....	60
Gráfico 5.7: resultado do Modelo 6.....	63
Gráfico 5.8: resultado do Modelo 7.....	67
Gráfico 5.9: resultado do Modelo 8.....	68
Gráfico 5.10: resultado do Modelo 9.....	69
Gráfico 5.11: resultado do Modelo 10.....	70
Gráfico 5.12: resultado do Modelo 11.....	71
Gráfico 5.13: resultado do Modelo 12 para 1 dia.	75
Gráfico 5.14: resultado do Modelo 12 para 2 dias.	76
Gráfico 5.15: resultado do Modelo 13 para 1 dia.	77
Gráfico 5.16: resultado do Modelo 13 para 2 dias.	78
Gráfico 5.17: resultado do Modelo 14 para 1 dia.....	79
Gráfico 5.18: resultado do Modelo 14 para 2 dias	80
Gráfico 5.19: resultado do Modelo 15 para 1 dia.....	81
Gráfico 5.20: resultado do Modelo 15 para 2 dias	82

Gráfico A.1: Cotação diária da Petrobras PN.....	99
Gráfico A.2: Preço de venda do petróleo Brent.....	100
Gráfico A.3: valor da taxa Selic.	101
Gráfico A.4: valor do câmbio comercial para venda.....	102
Gráfico A.5: valor do Ibovespa em seu fechamento.	103
Gráfico A.6: valor do fechamento do Dowjones.....	104
Gráfico A.7: valor do fechamento da Nasdaq.	105

LISTA DE TABELAS

Tabela 3.1: resumo de publicações envolvendo previsões com RNAs	38
Tabela 5.1: variáveis de entrada utilizadas.....	48
Tabela 5.2: variáveis de entrada dos modelos 1 e 2.	52
Tabela 5.3: arquitetura dos modelos 1 e 2.....	52
Tabela 5.4: pesos das entradas utilizadas	55
Tabela 5.5: performance dos modelos 1 e 2.....	55
Tabela 5.6: variáveis de entrada dos modelos 3, 4 e 5.	57
Tabela 5.7: arquitetura dos modelos 3, 4 e 5.....	57
Tabela 5.8: pesos das entradas utilizadas.	61
Tabela 5.9: performance dos modelos 3,4 e 5.	61
Tabela 5.10: variáveis de entrada do modelo 6.	62
Tabela 5.11: arquitetura do modelo 6.....	62
Tabela 5.12: pesos das entradas utilizadas.	64
Tabela 5.13: performance do modelo 6.....	64
Tabela 5.14: variáveis de entrada dos modelos 7, 8, 9, 10 e 11	65
Tabela 5.15: arquitetura dos modelos 7, 8, 9, 10 e 11.....	66
Tabela 5.16: pesos das entradas utilizadas nos modelos 7, 8, 9, 10 e 11.	72
Tabela 5.17: performance dos modelos 7, 8, 9, 10 e 11.....	72
Tabela 5.18: índice de acertos com 95% de confiança.....	72
Tabela 5.19: variáveis de entrada dos modelos 12, 13, 14 e 15.	74
Tabela 5.20: arquitetura dos modelos 12, 13, 14 e 15.....	74
Tabela 5.21: pesos das entradas utilizadas nos modelos 12 e 13.	83
Tabela 5.22: pesos das entradas utilizadas nos modelos 14 e 15.	83
Tabela 5.23: performance dos modelos 12 e 13.....	84
Tabela 5.24: performance dos modelos 14 e 15.....	84
Tabela 5.25: acertos com 95% de confiança.	84

1. INTRODUÇÃO

1.1. Motivação

O desenvolvimento de modelos para a previsão de valores futuros de ações é tão antigo quanto a própria bolsa de valores (Bueno, 2000). Os investidores sempre buscaram maneiras de aumentar seus ganhos financeiros através da identificação dos ativos que apresentavam a melhor oportunidade de compra ou venda.

Durante a década de 1960 foi desenvolvida a Hipótese do Mercado Eficiente que propõe que os preços do mercado de ações são definidos pela informação disponível, não sendo possível adquirir ganhos extraordinários. Porém, uma corrente importante de pensamento defende que é possível extrair altos retornos, apenas analisando as informações públicas do mercado e de suas empresas.

A seleção das ações que compõem uma carteira, e o momento correto de vender ou comprar estes ativos é decisivo para o investidor. Com o advento da globalização e o crescimento da capacidade computacional, se tornou fundamental para os investidores a utilização de ferramentas estatísticas e de opções que possam incrementar a rentabilidade do investidor.

Porém esta não é uma tarefa simples. Mesmo os métodos de análise existentes e os modelos econométricos tradicionais não garantem uma elevada taxa de retorno por um período de tempo consistente, devido ao comportamento não-linear do mercado de ações. Existe um elevado grau de incerteza nas relações entre as variáveis de interesse e a flutuação de uma delas influencia o restante (Bressan et al., 2005 e Kim, 2003), e os métodos lineares tradicionais não conseguem acompanhar esta volatilidade do mercado (Bódis, 2004). Logo, a utilização de modelos que consigam captar relações não-lineares surge como uma interessante alternativa de análise.

A utilização de métodos como algoritmos genéticos e redes neurais artificiais (RNAs) vêm crescendo pois possuem a característica de identificar e mapear as relações não-lineares entre variáveis, de trabalhar com dados incompletos ou com ruídos e de não requerer nenhuma suposição *a priori* sobre os dados (Vellido et al., 1999).

Estudos recentes comprovam a eficiência de RNAs, em comparação aos métodos lineares, em problemas de previsão onde não existe padrões determinados. Fora do âmbito econômico, mas em ainda em problemas de previsão relacionados, Comrie (1997) mostrou que redes neurais apresentam resultados um pouco superiores quando comparadas a métodos

de regressão tradicionais, na previsão do nível de ozônio em diferentes cidades. Em 2003, Cho realizou uma comparação entre três métodos para realizar a previsão de turistas em Hong Kong: o suavizamento exponencial (*exponential smoothing - ExSm*), ARIMA e RNAs. A pesquisa concluiu que RNAs são um método eficiente para realizar previsões, principalmente em casos onde não existe um padrão definido, como no caso da oscilação do número de turistas oriundos do Japão devido à crise asiática.

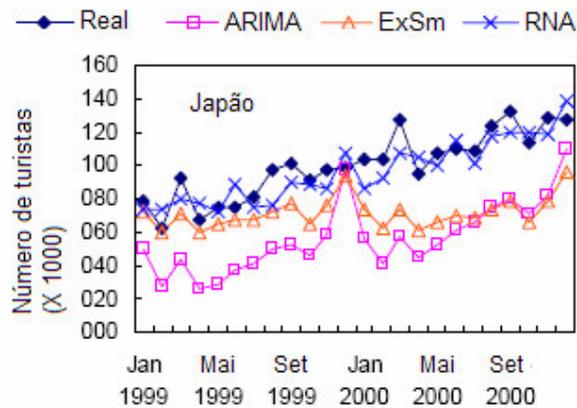


Figura 1.1: Previsão da chegada de turistas do Japão para Hong Kong (adaptada de Cho, 2003).

No campo da economia, é grande o interesse de RNAs para realizar previsões. Bódis (2004) realiza uma comparação entre ARIMA e RNAs na realização do retorno obtido da previsão da taxa de câmbio entre o dólar norte-americano e o dólar australiano. O retorno obtido por redes neurais foi mais de quatro vezes superior. Leung et al. (2000) concluíram que RNAs apresentam melhor desempenho na previsão de índices de ações ao realizar uma comparação com diversos métodos estatísticos.

Neste cenário cresce o interesse pelo uso de inteligência artificial para auxiliar na tomada de decisões em estratégias financeiras.

1.2. Objetivos da Dissertação

Existem hoje diversos métodos de realizar previsões, e um dos mais recentemente explorados é o uso de redes neurais.

O objetivo geral deste trabalho é utilizar redes neurais para desenvolver modelos capazes de realizar a previsão de tendências de alta ou baixa no mercado de ações.

São objetivos específicos:

- Utilizar redes neurais para realizar a previsão das tendências da ação Petrobras Preferencial Nominativa e do Índice da Bolsa de Valores de São Paulo.
- Identificar as variáveis que foram mais importantes para a realização das previsões citadas anteriormente.
- Analisar modelos de arquitetura de redes neurais que melhor se apliquem à realização das previsões de tendências de ações no geral.

1.3. Organização da Dissertação

A dissertação está organizada em seis capítulos.

O primeiro capítulo traz as motivações para o trabalho e traça seus objetivos.

O segundo capítulo trata da metodologia utilizada no trabalho, as redes neurais artificiais, primeiro com um apanhado histórico e em seguida realizando uma descrição de suas principais características.

O terceiro capítulo é realizada uma revisão da literatura, procurando relacionar trabalhos anteriores que utilizaram RNAs para previsão.

O objetivo do quarto capítulo é apresentar o conceito do mercado de ações, suas técnicas de análises, sua volatilidade e o panorama brasileiro.

O quinto capítulo trata das estimações e resultados do trabalho.

Por fim, o sexto capítulo traz alguns comentários e conclusões do trabalho, além de sugestões para trabalhos futuros.

Ao final deste trabalho é possível encontrar um glossário com os termos utilizados.

2. METODOLOGIA

2.1. Introdução

A metodologia utilizada para a realização da previsão de preços de ações foi o emprego de redes neurais artificiais (RNAs) e algoritmos genéticos (AGs).

As redes neurais artificiais procuram utilizar um modelo simplificado das redes biológicas, visando reproduzir algumas funções cerebrais. As principais semelhanças são o uso de processamento paralelo e o meio como os neurônios artificiais se comunicam. Apesar de possuírem uma complexidade bem menor do que as redes biológicas, as RNAs conseguem reproduzir com fidelidade alguns comportamentos dos seres humanos, como reconhecimento de padrões e aprendizado de tarefas simples, adaptando-se bem a ambientes e situações em que são sujeitas.

Algoritmos genéticos são rotinas computacionais que procuram emular o processo de evolução biológica para resolver problemas de otimização. Neste trabalho, AGs foram utilizados como uma ferramenta complementar ao treinamento das RNAs, visando otimizar parâmetros da estrutura das redes neurais.

2.2. Redes Biológicas: a motivação

O cérebro humano contém cerca de 10 bilhões de neurônios, sua célula básica. Cada neurônio está conectado com milhares de outros, realizando um processamento contínuo e paralelo. Por este motivo o cérebro é considerado o mais poderoso e fascinante processador existente.

Os neurônios são compostos por três componentes básicos: o corpo da célula (ou soma), os dendritos e o axônio. Os dendritos medem poucos milímetros, o corpo da célula alguns milésimos de milímetros, enquanto o axônio constitui a maior parte do neurônio, podendo alcançar alguns centímetros:

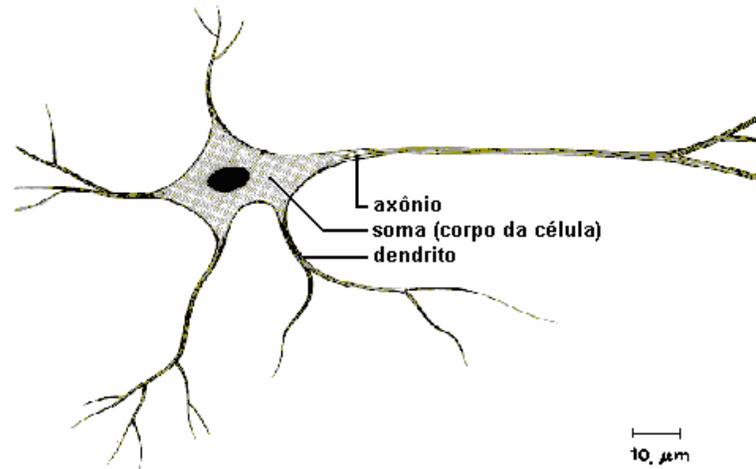


Figura 2.1: Esquema de um neurônio.

Os dendritos recebem os impulsos nervosos de outros neurônios (informação) e os transportam até o corpo da célula, que processam estas informações recebidas, gerando novos impulsos em diferentes frequências, que serão transmitidos através do axônio para os dendritos dos neurônios seguintes. Esta conexão entre um axônio e um dendrito é chamada de sinapse. As sinapses controlam as transmissões dos impulsos, ou seja, o fluxo de informação no cérebro. Cada neurônio está conectado a cerca de 10^3 a 10^4 neurônios, totalizando aproximadamente 10^{14} a 10^{15} interconexões. São através das sinapses que são estabelecidas as redes neurais biológicas. Todas as funções pelas quais é responsável o cérebro humano, como pensamento, emoção e aprendizado, são realizadas graças às conexões sinápticas.

Contudo, toda a complexidade das funções cerebrais ainda não foi desvendada pelo ser humano. O que se conhece é apenas como funcionam as sinapses e que por meio delas, com a variação da frequência e intensidade dos impulsos, e da criação e destruição de sinapses, o cérebro humano é capaz de aprender e de se adaptar, desempenhando todas as suas tarefas.

2.3. Neurônios artificiais: um breve histórico

O primeiro modelo de um neurônio artificial foi proposto por McCulloch e Pitts em 1943. Nele, o neurônio possuía como entradas x_1, x_2, \dots, x_n (dendritos), que para emular as sinapses tinham pesos associados a elas (w_1, w_2, \dots, w_n), cujos valores podiam ser positivos ou negativos, amplificando ou reduzindo o sinal, diferentemente das sinapses do cérebro, apenas positivas. O neurônio possuía apenas uma saída que era excitada quando a soma dos impulsos

$(\sum x_i \cdot w_i)$ ultrapassa um limiar de excitação θ (*threshold*). O corpo do neurônio era um mecanismo simples que apenas realizava o somatório e decidia se o neurônio deveria ou não ser excitado.

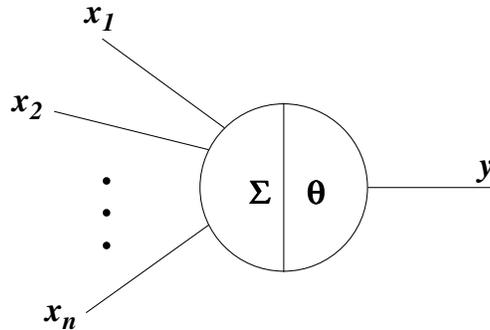


Figura 2.2: o modelo de McCulloch e Pitts.

As principais limitações do modelo de McCulloch e Pitts é que a rede possuía apenas uma camada e que os pesos eram fixos. Porém, a partir deste modelo, vários outros foram propostos, sendo a principal modificação, a implementação de diferentes funções de ativação, produzindo qualquer saída, e não apenas a saída binária (excitada ou não-excitada) proposta por McCulloch e Pitts.

Em 1949, Donald Hebb mostrou que o aprendizado das redes neurais é conseguido através da mudança dos pesos de entrada. A regra de Hebb, como é conhecida, é hoje utilizada em algoritmos de aprendizado. Em 1958, Frank Rosenblatt propôs um novo modelo, o *perceptron*. Nele as RNAs possuíam sinapses ajustáveis e podiam ser treinadas para classificar certos tipos de padrões. A topologia de RNA e o algoritmo de aprendizado proposto por Rosenblatt ainda hoje são utilizados.

Depois, as RNAs passaram por um período de ostracismo e pessimismo em relação a sua implementação, já que esta tarefa requeria uma complexidade computacional bem elevada para a época. Apenas em 1982, Hopfield publicou um artigo ressaltando as propriedades associativas das RNAs, retomando as pesquisas na área. Em 1986, foi publicado o algoritmo de treinamento *backpropagation* – BP (retro-propagação) (Rumelhart et al., 1986), que juntamente com o avanço da microeletrônica, possibilitaram a afirmação das RNAs nas comunidades de pesquisas.

2.4. Redes Neurais Artificiais: o neurônio artificial

As redes neurais artificiais são sistemas de processamento paralelo, que simulam a estrutura do cérebro humano, para calcular certas funções matemáticas lineares ou não. Cada RNA é composta de unidades de processamento de informação (neurônios ou nodos), dispostas em uma ou mais camadas e interligadas por um elevado número de conexões. Cada conexão possui um peso sináptico, responsável pelo armazenamento do conhecimento (Braga et al., 2000).

Um modelo não-linear de um neurônio pode ser representado pela figura 3, e é formado pelos seguintes componentes:

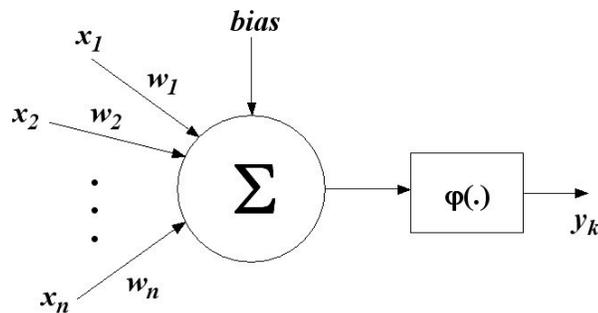


Figura 2.3: modelo não-linear de um neurônio.

- Um conjunto de entradas x_i 's.
- Um peso sináptico w_i correspondente a cada entrada x_i .
- Um somador, que realiza a adição das entradas ponderadas pelos pesos.
- Uma função de ativação ou transferência, que restringe a amplitude da saída do neurônio a um valor finito.
- Um "bias", que tem o efeito de aumentar ou diminuir a ação da função de ativação.

Em termos matemáticos, o neurônio pode ser descrito pelas seguintes equações:

$$u_k = \sum_{j=1}^n w_{kj} \cdot x_j \quad (2.1)$$

$$y_k = \varphi(u_k + b_k) \quad (2.2)$$

Onde:

- x_1, x_2, \dots, x_n são os sinais de entrada do neurônio k.
- $w_{k1}, w_{k2}, \dots, w_{kn}$ são os pesos do neurônio k.
- u_k é o resultado do somador.
- b_k é o *bias*.
- $\varphi(\cdot)$ é a função de ativação.
- y_k é o sinal de saída do neurônio k.

2.4.1. Tipos de Função de Ativação

Existem diversos tipos de função de ativação para uma RNA. Os principais são os seguintes:

- **Função degrau.** A mais simples função de ativação. Pode ser definida como:

$$f(x) = \begin{cases} 1 & \text{se } x \geq 0 \\ 0 & \text{se } x < 0 \end{cases} \quad (2.3)$$

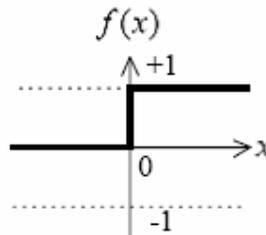


Gráfico 2.1: função degrau.

- **Função linear.**

$$f(x) = x, \forall x \quad (2.4)$$

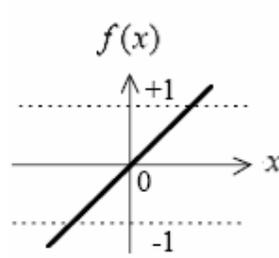


Gráfico 2.2: função linear.

- **Função sigmoial.** O tipo de função de ativação mais utilizado. A função assume valores entre 0 e 1 , e pode ser definida como:

$$f(x) = \frac{1}{1 + \exp(-x)} \quad (2.5)$$

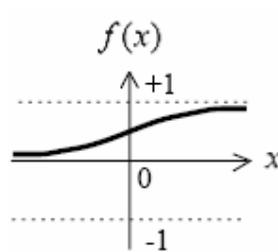


Gráfico 2.3: função sigmoial.

- **Função tangente hiperbólica.** Outro tipo de função de ativação bastante utilizado. A função assume valores entre -1 e +1, e pode ser definida como:

$$f(x) = \frac{2}{1 + \exp(-2x)} - 1 \quad (2.6)$$

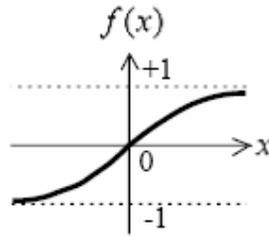


Gráfico 2.4: função tangente hiperbólica.

- **Função sigmoideal linear.** Uma variação da função sigmoideal.

$$f(x) = \begin{cases} x, & \text{se } 0 \leq x \leq 1 \\ 0, & \text{c.c.} \end{cases} \quad (2.7)$$

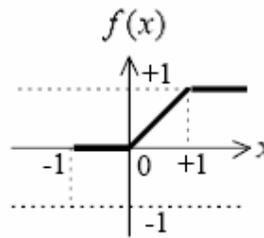


Gráfico 2.5: função sigmoideal linear.

- **Função tangente hiperbólica linear.** Uma variação da função tangente hiperbólica.

$$f(x) = \begin{cases} x, & \text{se } -1 \leq x \leq 1 \\ 0, & \text{c.c.} \end{cases} \quad (2.8)$$

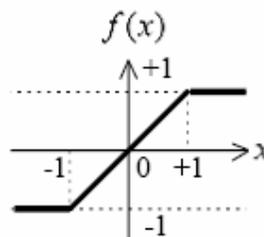


Gráfico 2.5: função tangente hiperbólica linear.

2.5. Arquitetura de Redes Neurais

A escolha da arquitetura da RNA é de fundamental importância para a resolução do problema, pois cada topologia se ajusta melhor a um determinado tipo de problema. Os parâmetros que fazem parte da escolha da arquitetura são os seguintes (Braga et al., 2000): número de camadas da rede, número de neurônios em cada camada e o tipo de conexão entre os neurônios da rede.

Quanto ao número de camadas podemos classificar as RNAs como:

- Redes de uma única camada: apenas um neurônio faz o processamento da informação entre uma entrada e a saída.
- Redes de camadas múltiplas: existem neurônios adicionais realizando o processamento entre uma entrada e a saída. Estas camadas adicionais são chamadas de camadas escondidas ou camadas ocultas (*hidden layers*). Com a adição de uma ou mais camadas ocultas, a RNA pode extrair resultados mais complexos do problema a que é submetida. Esta habilidade se faz necessária o quanto maior for o tamanho da camada de entrada (Haykin, 1994).

Quanto aos tipos de conexões dos neurônios podemos classificar as RNAs como:

- **Rede direta ou *feedforward*.** A saída de um neurônio não pode ser utilizada como entrada de uma camada anterior à camada em que está localizado este neurônio, ou seja, a rede não possui ciclos. A rede direta é bastante utilizada com o algoritmo *back-propagation*.

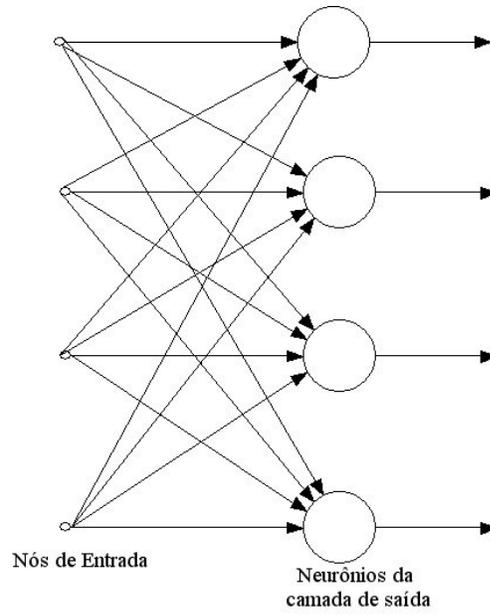


Figura 2.4: Rede direta de uma única camada (adaptada de Hayken, 1994).

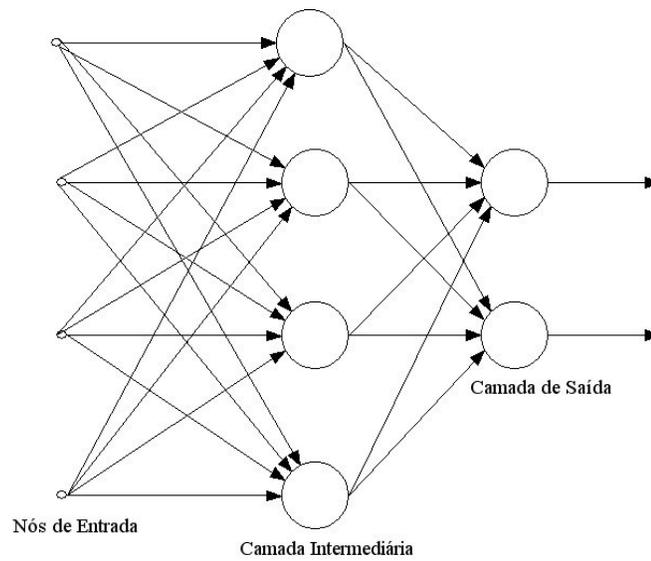


Figura 2.5: Rede direta de múltiplas camadas (adaptada de Hayken, 1994).

- **Rede recorrente ou cíclica ou *feedback*.** A saída de algum neurônio é utilizada como entrada de outro neurônio de uma camada anterior à camada do primeiro, ou seja, a rede possui ciclos. A rede recorrente é mais indicada para a solução de problemas que envolvem processamento temporal.

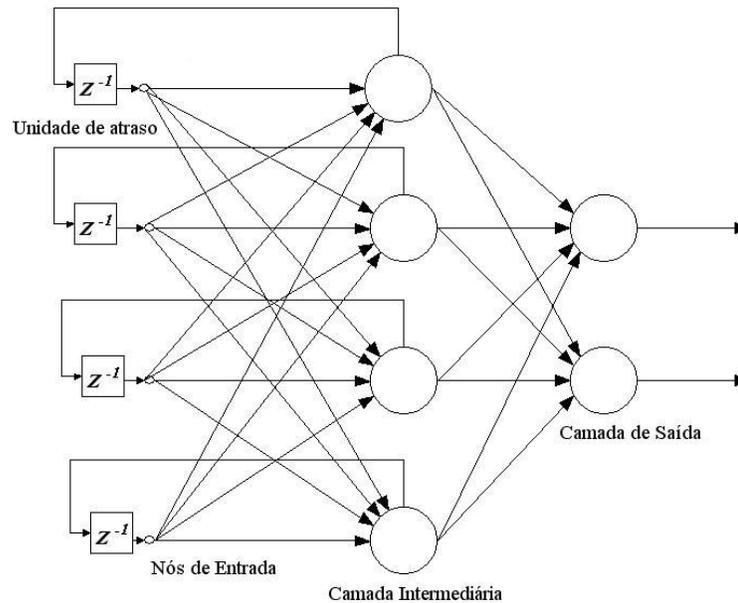


Figura 2.6: Modelo simplificado de uma rede recorrente (adaptada de Hayken, 1994).

2.6. Processo de Aprendizado de RNAs

A etapa de aprendizagem é a primeira a ser realizada por uma RNA na solução de um problema, sendo uma das mais importantes. Haykin (1994) define aprendizado no contexto de RNAs como:

“Learning is a process by which the free parameters of a neural network are adapted through a continuing process of stimulation by the environment in which the network is embedded. The type of learning is determined by the manner in which the parameter changes take place”

“Aprendizado é um processo pelo qual, os parâmetros de uma rede neural são adaptados através de um processo contínuo de estimulação do ambiente no qual a rede está

inserida. O tipo de aprendizado é determinado pela maneira na qual ocorre as mudanças dos parâmetros”

É nessa etapa que a rede cria uma representação do ambiente em que está inserida. Isto ocorre através de um processo iterativo de ajuste dos pesos das conexões entre os neurônios, que acontece devido à estimulação que a rede recebe do ambiente ao seu redor. Graças a estas mudanças em sua estrutura, uma RNA é capaz de responder a novas situações do ambiente.

Uma das escolhas mais importantes para a elaboração de uma RNA é a de seu “algoritmo de aprendizado”. É ele que vai decidir como os pesos das conexões serão alterados, que conexões serão criadas e que conexões serão destruídas, alterando a topologia da rede. Esta escolha deve ser feita baseada no problema ou tarefa a ser resolvida. Por exemplo, o algoritmo BP geralmente é utilizado com redes diretas, aprendizado supervisionado e para o reconhecimento de padrões.

Os métodos de aprendizado de RNAs se dividem em duas categorias: o aprendizado supervisionado e o aprendizado não-supervisionado.

2.6.1. Aprendizado supervisionado

O aprendizado supervisionado utiliza em sua estrutura uma espécie de instrutor que confere o quanto a rede está próxima de uma solução aceitável, adaptando na concepção do treinamento os pesos entre os neurônios, de modo a prover uma menor diferença entre as saídas desejadas e a obtida. A figura 9 representa este mecanismo de aprendizado:

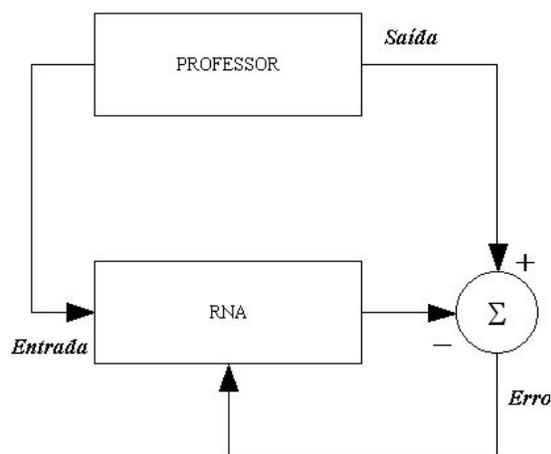


Figura 2.7: Aprendizado Supervisionado (adaptada de Hayken, 1994).

A rede recebe como informações iniciais de treinamento diversas situações para as entradas do problema, e o comportamento desejado na saída para cada situação. Desta forma, o Instrutor direciona o processo de treinamento. No treinamento, a rede tem sua saída atual comparada com a saída desejada, sendo calculado o erro da resposta corrente. Então, os pesos são ajustando, visando minimizar o erro. Esta operação é repetida em cada etapa de treinamento, tentando assim obter a solução, caso haja uma possível. Após o treinamento, os pesos estão ajustados para que a rede seja submetida a situações de teste.

A soma dos erros quadráticos é utilizada com frequência como medida de desempenho da rede e também como a função de custo a ser minimizada pelo algoritmo de treinamento (Braga et al., 2000).

A principal desvantagem do aprendizado supervisionado é a limitação de aprender novas estratégias que estejam fora do escopo do treinamento inicial. O principal exemplo de algoritmo para aprendizado supervisionado é o *backpropagation* (Braga et al., 2000).

Ainda é possível utilizar algoritmos de otimização para melhorar o desempenho de algoritmos supervisionados. Por exemplo, pode-se utilizar algoritmos genéticos para otimizar parâmetros de um algoritmo BP.

2.6.2. Aprendizado não-supervisionado

No aprendizado não-supervisionado, diferentemente do aprendizado supervisionado, a rede não possui conhecimento *a priori* das saídas desejadas, recebendo no seu treinamento apenas informações das entradas. Através de algoritmos de aprendizado baseados em conceitos de vizinhança e agrupamento, a rede procura distinguir classes de padrões diferentes dos dados apresentados à rede. Isto é possível apenas quando existe redundância nos dados de entrada, para que a rede possa identificar padrões ou características no problema analisado.

2.6.3. Aprendizado Híbrido

É uma combinação dos dois métodos citados anteriormente. Uma parte dos pesos é determinada por aprendizado supervisionado, enquanto a outra é determinada pelo não-supervisionado.

2.7. Projeto de uma rede neural artificial

Uma vez que se decide resolver um problema utilizando RNAs, são necessárias basicamente quatro etapas para se desenvolver a aplicação. São elas:

2.7.1. Coleta e separação dos dados

Esta etapa consiste em coletar os dados relativos ao problema e sua separação em dois conjuntos de dados. O primeiro conjunto será utilizado no treinamento da rede, enquanto que o uso do segundo é verificar o comportamento da rede na resolução de problemas semelhantes ao desejado, observando se o desempenho da rede está de acordo com o esperado.

É necessária uma cuidadosa análise sobre o problema, pois os dados coletados devem ser significativos e cobrir da maneira mais ampla possível o domínio do problema, fazendo com que a rede seja capaz de desenvolver soluções para as exceções e limites do problema.

Após a separação dos conjuntos de dados, geralmente eles são ordenados aleatoriamente e podem ser sujeitos a um pré-processamento, como normalização, escalonamento, etc.

2.7.2. Seleção dos parâmetros da RNA

Nesta etapa é definido o tipo de topologia que será empregado para a resolução do problema em questão. Aqui são determinados o número de camadas da rede, o número de neurônios em cada camada, o tipo de conexões entre estes neurônios, além das funções de ativação e as variáveis de treinamento da rede. Na maioria das vezes são necessárias diversas etapas de testes, modificando alguns destes parâmetros, para que a rede obtenha resultados satisfatórios.

A determinação destes parâmetros é vital para o bom desempenho da rede. Porém é uma tarefa que requer experiência. Existem metodologias que indicam que tipo de RNA será mais adequado para a resolução de um determinado problema (Haykin, 1994).

2.7.3. Treinamento

Esta é a principal etapa no desenvolvimento de uma RNA. É aqui que a rede aprende sobre o ambiente e se adapta, modificando seus pesos, para resolver o problema.

O primeiro passo é a escolha dos valores iniciais dos pesos da rede. Uma escolha correta diminui o tempo necessário para o treinamento e evita que a rede sature prematuramente. Uma boa opção para inicialização da rede é a escolha de números aleatórios uniformemente distribuídos (Zhang et al., 1998).

Então, a rede é submetida ao conjunto de dados escolhidos para esta etapa. De acordo com os padrões apresentados à rede pelos dados, ela modifica seus pesos e conexões através de uma regra pré-estabelecida de aprendizado (algoritmo de aprendizado). Após o treinamento, as RNAs são capazes de identificar padrões e de prever o valor de determinadas

variáveis para situações que não foram apresentadas no treinamento, extraindo informações sobre os dados que não estão explícitas e realizando mais do que apenas uma simples relação entrada-saída (Braga et al., 2000).

Outro ponto importante do treinamento da rede, diz respeito ao tempo de treinamento. Caso não seja escolhido um bom critério de parada, pode ocorrer um *overtraining* (Haykin, 1994), o que significa que a rede se especializou no conjunto de dados de treinamento, perdendo sua capacidade de generalização.

2.7.4. Teste e Validação

Nesta última etapa, o conjunto separado previamente para teste é submetido a RNA e, então, é validado. Caso o resultado não seja satisfatório, deve-se retornar às etapas iniciais, analisando novamente o projeto da RNA.

2.8. Algoritmo *Backpropagation*

2.8.1. Perceptron

Conforme citado anteriormente, em 1958, Rosenblatt introduziu uma forma simples de rede neural desenvolvida para resolver problemas de classificação, o *perceptron*. Basicamente, o *perceptron* é constituído por uma única camada, similar ao modelo apresentado na figura 3.3. O *perceptron* tornou-se a forma pioneira de rede neural (Haykin, 1994). Rosenblatt demonstrou que o *perceptron* era capaz de resolver problemas linearmente separáveis, ou seja, ele pode ser treinado com padrões retirados de duas classes linearmente separáveis e ser capaz de convergir para uma superfície de decisão formada por um hiperplano entre estas duas classes.

2.8.2. Perceptron multi-camadas

O *perceptron* proposto por Rosenblatt, apresentava a oportunidade de resolver apenas problemas linearmente separáveis. Então, em 1986, Rumelhart et al. propuseram o algoritmo de treinamento *backpropagation*. Neste trabalho foi demonstrado que seria possível treinar redes multi-camadas direta, introduzindo o conceito de camadas ocultas e as redes neurais que ficaram conhecidas como redes *perceptrons* multi-camadas diretas (MLP – multi-layer perceptrons). Os MLPs são os modelos de RNAs que mais são utilizados em aplicações práticas (Zhang et al., 1998), sendo os mais citados em toda a literatura.

2.8.3. O algoritmo backpropagation

Como o próprio nome diz, o algoritmo de *backpropagation* consiste em realizar a retro-propagação (realimentação) do erro gerado na diferença entre a saída da rede e o valor desejado, visando minimizar o erro total da saída gerada pela rede.

O treinamento envolve três etapas: a propagação dos dados da camada de entrada para a camada de saída da rede, a retro-propagação do erro obtido e o ajuste dos pesos associados à conexão das redes (Fausset, 1994). O erro no instante n (e) pode ser obtido de acordo com a seguinte fórmula:

$$e_j(n) = d_j(n) - y_j(n) \quad (2.3)$$

Onde, d_j é a saída desejada para o neurônio j , e y_j é a saída da rede para este neurônio.

O valor a ser minimizado é a energia média do erro, definida como a média das energias instantâneas (Ambrósio, 2002):

$$\mathcal{E}(n) = \frac{1}{2} \sum_{j \in C} e_j^2(n) \quad (2.4)$$

$$\mathcal{E}_{med} = \frac{1}{N} \sum_{n=1}^N \mathcal{E}(n) \quad (2.5)$$

Onde, C é o conjunto dos neurônios e N é o número de padrões apresentados à rede em seu treinamento. A minimização do erro é realizada através dos ajustes do peso, aplicando a regra Delta, baseada no método do gradiente descendente (Widrow e Hoff, 1960 apud. Haykin, 1994). A regra Delta procura identificar a direção de mudança do peso de um neurônio da camada j a um neurônio na camada k (w_{jk}), com o intuito de reduzir a energia do erro.

$$\Delta w_{jk}(n) = \lambda \cdot \delta_k(n) \cdot x_j(n) \quad (2.6)$$

$$w_{jk}(n+1) = w_{jk}(n) + \Delta w_{jk}(n) \quad (2.7)$$

Onde, x_j representa a saída do neurônio j , λ é uma constante positiva que determina o

tamanho do passo ou a taxa de aprendizagem em cada iteração e δ_k é o gradiente local, que pode ser definido como:

$$\delta_k(n) = e_k(n) \cdot y_k(n) \cdot (1 - y_k(n)) \quad (2.8)$$

Ou seja, o gradiente local de cada neurônio é calculado através do produto do erro pela derivada da saída obtida neste neurônio. O ajuste dos pesos é realizado no sentido contrário ao gradiente, conforme mostrado na figura 2.8:

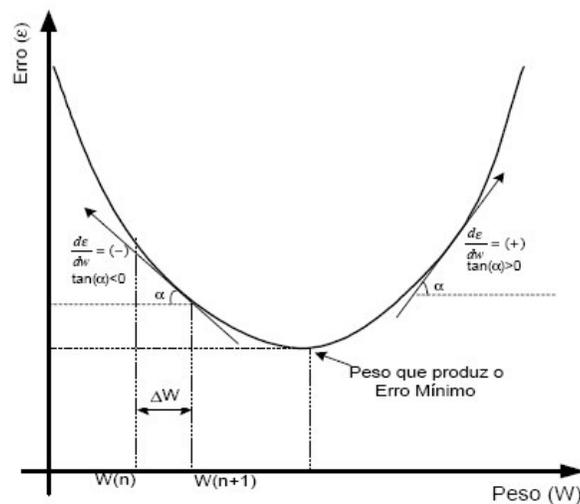


Figura 2.8: ajuste dos pesos pela regra Delta (adaptada de Mello, 2004).

O algoritmo BP ainda pode ser otimizado com a introdução da taxa de *momentum*. Esta taxa é uma constante que determina o efeito das mudanças anteriores dos pesos. Seu objetivo é evitar que a rede não se atenha a mínimos locais durante seu treinamento, reduzindo assim o tempo gasto nesta atividade.

A escolha dos parâmetros de um BP é decisiva para um melhor desenvolvimento de sua performance e agilidade. Estratégias de como incrementar o BP vêm sendo desenvolvidas, criando novas variações do BP e algoritmos complementares que possam otimizar a escolha de valores como a taxa de aprendizagem (Schiffmann et al., 1994). Uma das técnicas utilizadas atualmente é o uso de algoritmos genéticos para realizar a otimização

destes parâmetros.

2.9. Algoritmos Genéticos

Algoritmos genéticos são utilizados para a resolução de problemas de otimização - que exigem soluções adaptativas - procurando acumular conhecimento sobre o ambiente para gerar as soluções (Goldberg, 1953). Podem ser utilizados como algoritmo de aprendizado de redes neurais, gerando pesos e conexões através de processos evolutivos e buscando uma solução ótima.

Em 1960, John Holland descreveu pela primeira vez o que seria um AG e, nas décadas seguintes, continuou o desenvolvimento de um método de resolução de problemas de otimização que utilizasse o fenômeno de “adaptação” existente na natureza, ou seja, como os indivíduos de uma determinada população evoluem a cada nova geração (Holland, 1975).

Algoritmos genéticos podem ser utilizados como o próprio algoritmo de treinamento de uma RNA ou para otimizar algum parâmetro específico da rede, como o número de elementos em cada camada e como ferramenta complementar do algoritmo de treinamento. Neste trabalho, AGs foram utilizados otimizar dois parâmetros do algoritmo *backpropagation*: taxa de aprendizagem e taxa de *momentum*.

2.9.1. Introdução aos algoritmos genéticos

O princípio básico do funcionamento dos algoritmos genéticos é o uso de um critério de seleção voltado para a busca de pontos com "alta aptidão" em uma função objetiva, ou seja, pontos nos quais, a função a ser minimizada (ou maximizada) tem valores relativamente baixos (ou altos).

Os AGs empregam uma estratégia de busca paralela e estruturada. Apesar de aleatórios, eles não são caminhos desordenados, pois exploram informações históricas para encontrar novos pontos de busca onde são esperados melhores desempenhos. Isto é feito através de processos iterativos, onde cada iteração é chamada de geração.

Durante cada iteração, os princípios de seleção e reprodução são aplicados a uma população de candidatos que pode variar, dependendo da complexidade do problema e dos recursos computacionais disponíveis.

Através da seleção, determinam-se quais indivíduos conseguirão se reproduzir, gerando um número determinado de descendentes para a próxima geração, com uma probabilidade determinada pelo seu índice de aptidão. Em outras palavras, os indivíduos com maior adaptação relativa têm maiores chances de se reproduzir, porém não exclusivamente, a

fim de manter a diversidade da população.

2.9.2. Funcionamento de um algoritmo genético

O funcionamento de um AG pode ser resumido pelo fluxograma simplificado abaixo (Goldberg, 1953):

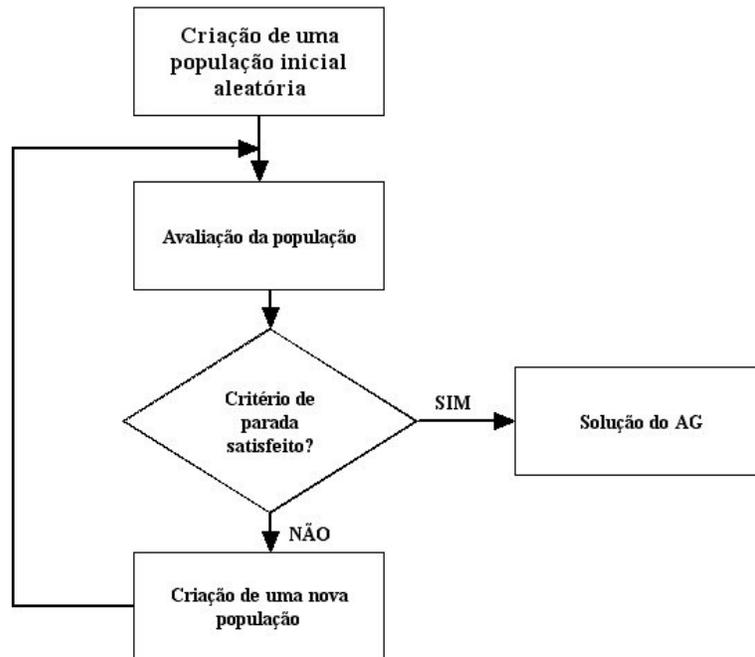


Figura 2.9: Fluxograma simplificado de um AG (adaptada do manual do Matlab).

- **População inicial:** o primeiro passo do algoritmo é criar uma população inicial aleatória. Na figura 9, podemos ver um exemplo de uma matriz desta população:

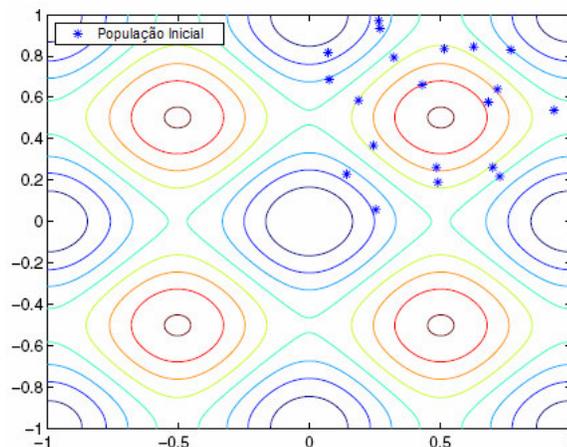


Figura 2.10: Exemplo de uma matriz para a população inicial (adaptada do manual do Matlab).

- **Avaliação da população:** em seguida, cada membro da população é avaliado, sendo calculado o resultado da função objetiva a ser otimizada. Caso o resultado da população seja satisfatório como solução, o algoritmo chega ao fim. Caso contrário, segue para a próxima etapa.

- **Criação da nova geração:** uma nova geração é criada através de três processos: seleção, *crossover* (cruzamento) e mutação. Eles são utilizados para assegurar que a nova geração seja totalmente nova, mas possua, de alguma forma, características de seus pais, ou seja, a população se diversifica e mantém características de adaptação adquiridas pelas gerações anteriores:
 - **Seleção:** os indivíduos com melhor performance na geração atual sobrevivem automaticamente para a próxima geração.

 - **Cruzamento ou Crossover:** o algoritmo realiza cruzamento entre indivíduos da geração atual para produzir filhos para a próxima geração. O cruzamento é realizado combinando aleatoriamente “genes” de cada pai, isto é, designando coordenadas e características para os filhos. Assim como na natureza, este processo está ligado a uma probabilidade do cruzamento acontecer, chamada de probabilidade de cruzamento ou *crossover*.

 - **Mutação:** o algoritmo modifica aleatoriamente “genes” de um indivíduo da geração atual, criando um novo indivíduo para a próxima geração. Este processo também está ligado a uma probabilidade de acontecer uma mutação, chamada de probabilidade de mutação.

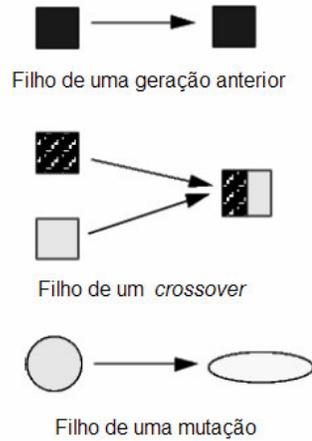


Figura 2.11: Criação de uma nova geração em um AG (adaptado do manual do Matlab).

- **Crítérios de parada:** existem condições que um algoritmo genético pode utilizar como um ou mais critérios de parada, sendo usualmente pré-definidos:
 - Número de gerações.
 - Limite de tempo.
 - Valor da função objetiva.
 - Número de gerações em que o valor da função objetiva não muda dentro de um limite pré-estabelecido.
 - Período de tempo em que o valor da função objetiva não muda dentro de um limite pré-estabelecido.

2.9.3. Algoritmo genético simples

A estrutura proposta por Goldberg é conhecida como algoritmo genético simples, ou Simple Genetic Algorithm (SGA). O SGA trabalha com um número fixo de indivíduos na população ao longo das gerações, sendo gerados inicialmente de maneira aleatória. Posteriormente, os melhores indivíduos, ou seja, aqueles com melhor desempenho, são escolhidos para a próxima geração, seja por seleção, cruzamento ou mutação. A escolha é realizada por um processo chamado de Roleta. Neste método cada indivíduo i tem uma probabilidade de seleção p_{sel} , dada pelo seu desempenho f_i :

$$p_{sel} = \frac{f_i}{\sum f_i} \quad (2.9)$$

Durante a seleção, os indivíduos com desempenho acima da média possuem maior chance de serem selecionados para a próxima geração, enquanto que os com pior desempenho podem desaparecer. Após a seleção, os indivíduos são sorteados para cruzamento e mutação, de acordo com as probabilidades atribuídas a cada processo.

Neste trabalho, será utilizado o algoritmo genético simples com o processo de seleção Roleta.

2.10. Aplicações de RNA

As RNAs podem ser treinadas para resolver problemas em diversos campos, como reconhecimento de padrões, classificação, aproximação de funções e previsão de eventos futuros.

Atualmente, as RNAs são utilizadas para resolver problemas com uma complexidade elevada para sistemas computacionais convencionais ou para os seres humanos. Entre as aplicações estão:

- **Economia:** previsão no mercado de ações, seleção de carteiras de investimentos, análise de riscos, etc.
- **Aeronáutica:** simuladores de vôo, sistemas de controle de aeronaves, sistemas de piloto automático, etc.
- **Automotivo:** sistemas de pilotagem automática, análise de seguros.
- **Entretenimento:** animação, efeitos especiais, etc.
- **Eletrônica:** controle de processos, *layout* de circuitos impressos, modelagem de sistemas não-lineares, processamento de sinais e de imagens, etc.
- **Telecomunicações:** compressão de dados e imagens, previsão de erros em códigos, supressão de ruídos, etc.
- **Medicina:** análise de células cancerígenas, análise de ECG e EEG, design de próteses, redução de custos em hospitais, etc.
- **Militar:** reconhecimento de objetos e alvos, correção de rotas de mísseis, processamento de sinais de radares e sensores, identificação de imagens, etc.
- **Inteligência artificial:** reconhecimento de voz, compressão de voz, sistemas de visão, controle de trajetórias, etc.

2.11. Software

Existe uma gama grande de *softwares* para a simulação de RNAs. O *software* escolhido foi o *NeuroSolutions*. Trata-se de uma opção onde é possível escolher o número de neurônios em cada camada, o número de camadas da rede, que algoritmo de aprendizado utilizar, que tipo de problema será resolvido, entre alternativas. Suas ferramentas de visualização permitem que o usuário verifique o comportamento da rede sem que seja necessário esperar até o fim de seu treinamento, além de permitir que parâmetros sejam alterados durante a execução da mesma.

3. REVISÃO BIBLIOGRÁFICA

3.1. Introdução

Recentemente, Redes Neurais Artificiais (RNAs) têm demonstrado uma poderosa capacidade de classificação de padrões e de previsões (Vellido et al., 1999 e Zhang et al., 1998). RNAs são utilizadas com sucesso em uma gama de campos de pesquisa como economia, indústria e física (Widrow et al., 1994). O interesse pelo tema é evidenciado pelo número crescente de artigos publicados e análises realizadas na literatura (Vellido et al., 1999).

Uma das áreas de maior destaque é o uso de RNAs para realizar previsões (Sharda, 1994). Nos últimos anos tem crescido a pesquisa do uso de RNAs neste tipo de aplicação, que durante muito tempo foi dominada por métodos lineares (Zhang et al., 1998). Os métodos lineares possuem uma implementação e interpretação simples, porém com diversas limitações, pois não são capazes de identificar relações não-lineares entre variáveis (Zhang et al., 1998).

A estrutura não-linear de uma RNA identifica relações complexas em diversos problemas reais, modelando funções lineares ou não (Vellido et al., 1999 e Zhang et al., 1998). Por isso, redes neurais se apresentam como uma interessante e promissora ferramenta de previsão – em alternativa aos tradicionais métodos estatísticos – apesar de não ser possível determinar, com um nível elevado de confiabilidade, os fatores que são decisivos em sua performance (Hill et al., 1994 e Zhang et al., 1998).

3.2. Motivações e limitações para o uso de RNAs em previsões

RNAs apresentam diversas características que as tornam atrativa para o uso em previsões. São elas:

- Aproximam qualquer função contínua (Hornik, 1991). RNAs possuem uma capacidade mais flexível e geral para identificar funções do que os métodos estatísticos tradicionais, que frequentemente possuem limitações para estimar as relações entre a entrada (dados passados e presentes) e a saída (dados futuros), devido a complexidade que envolve tal o processo. RNAs apresentam-se como bons métodos alternativos a estas limitações (Zhang et al., 1998).

- São não-lineares. Por isso, possuem a capacidade de aprender relações não-lineares entre variáveis, mesmo que estas sejam difíceis de serem descritas (Vellido et al., 1999 e Zhang et al., 1998), diferentemente das abordagens lineares tradicionais de previsão de séries temporais (Zhang et al., 1998).
- Podem ser tratadas como um método não-paramétrico (Vellido et al., 1999 e Zhang et al., 1998). Portanto não é necessária nenhuma suposição *a priori* sobre a distribuição e/ou a relação das variáveis de entrada e saída. A RNA aprende com a experiência dos exemplos que lhe são fornecidos, sendo útil em diversos problemas práticos em que apenas os dados são fornecidos, não sendo possível obter informações sobre as leis subjacentes que regem a geração destes dados (Zhang et al., 1998).
- A habilidade de lidar com ruídos e rupturas no conjunto de dados (Vellido et al., 1999 e Zhang et al., 1998). Uma rede neural é capaz de inferir corretamente em uma amostra de dados, mesmo que esta possua ruídos ou que existam períodos em que não há dados disponíveis.
- Podem ser altamente automatizadas, diminuindo o envolvimento humano (Vellido et al., 1999).

Ao mesmo tempo em que conseguem extrair relações implícitas entre entradas e saídas, estas não são explicadas trivialmente, o que dificulta conclusões sobre os resultados obtidos por uma RNA. Além disso, é impossível identificar se o projeto escolhido é o que melhor se ajustou ao problema e quais correções são necessárias para se obter melhores resultados.

As desvantagens das RNAs são:

- RNAs são “caixas pretas”. É praticamente impossível explicar a relação - identificada pela rede - entre as variáveis, pois estas são geralmente independentes. Portanto, a extração de regras com os resultados obtidos é uma tarefa extremamente complexa (Vellido et al., 1999 e Zhang et al., 1998).
- A seleção da topologia e dos parâmetros de uma RNA não possui um método específico, sendo um problema de “tentativa e erro” (Vellido et al., 1999). Não existe

uma metodologia que melhor defina a escolha dos parâmetros que fazem parte do projeto de uma rede neural (Vellido et al., 1999, Zhang et al., 1998): escolha das variáveis de entrada para o problema; separação dos dados entre treinamento e teste; o tipo de tratamento que será utilizado nos dados; a arquitetura da RNA; o número de camadas ocultas e o número de neurônios em cada camada; o algoritmo de treinamento, etc. O que podem ser encontradas são melhores práticas relacionadas a um tipo de problema (Zhang et al., 1998). Portanto, esta se torna uma etapa extenuante, onde o tempo investido não pode ser definido.

- O processo de aprendizagem pode consumir muito tempo (Vellido et al., 1999). O tempo que uma RNA leva em seu treinamento depende de diversos fatores como, o número de variáveis de entrada, o número de períodos de treinamento, a topologia e o algoritmo escolhido, além da complexidade do problema. Porém não existe nenhuma relação linear entre os fatores citados e o tempo gasto no treinamento da RNA, podendo variar indefinidamente. Todavia, pode-se observar que um fator decisivo é a escolha do algoritmo de treinamento. Algoritmos mais complexos como Algoritmos Genéticos, apesar de otimizarem as escolhas dos parâmetros (Santos et al., 1999), consomem mais tempo para o treinamento.
- RNAs não possuem a habilidade de identificar mínimos locais (Vellido et al., 1999). Conseqüentemente, durante o treinamento a rede pode se ater a um vale, não conseguindo mais retornar ao mesmo patamar de valores do restante do período.
- RNAs não possuem propriedades estatísticas clássicas, portanto não estão sujeitas à teste de hipóteses e intervalos de confiança (Vellido et al., 1999).
- RNAs são dependentes da qualidade/quantidade dos dados disponíveis (Vellido et al., 1999). Apesar de não ser possível tirar conclusões sobre o melhor tamanho de uma amostra de dados (Zhang et al., 1998), nem todos os conjuntos de dados se ajustam à uma solução via redes neurais.

3.3. Metodologia de RNAs para previsão

Apesar de não possuir nenhuma metodologia reconhecida como padrão para a definição dos parâmetros de um projeto de uma RNA - conforme apresentado acima, em suas desvantagens - é possível encontrar na literatura referências a métodos que obtiveram melhor desempenho com estilos específicos de problemas (Zhang et al., 1998).

Na literatura existem vários exemplos de como construir uma RNA para previsão e de como tratar os dados utilizados para treinamento e teste. Connor (1994) utilizou uma rede recorrente em comparação a um modelo de ARMA. Gent e Sheppard (1992) e Kuan e Liu (1995) também empregaram redes recorrentes em suas pesquisas (Zhang et al., 1998). Lapedes e Farber (apud. Vellido et al., 1999) foram os primeiros a utilizar uma rede direta multi-camadas para aplicações em previsão e, desde então, o foco da grande maioria dos trabalhos está no uso de redes diretas, com uso de *perceptrons* de múltiplas camadas (MLP) (Vellido et al., 1999 e Zhang et al., 1998).

3.3.1. Arquitetura da rede neural

Conforme descrito no capítulo anterior, são necessários os seguintes parâmetros para a definição da arquitetura de uma rede neural: número de camadas da rede, número de neurônios em cada camada e o tipo de conexão entre os neurônios da rede.

Como citado anteriormente, a rede direta é o tipo de conexão mais utilizado em aplicações de previsão. Conseqüentemente sobram três escolhas a serem realizadas quanto à arquitetura:

- O número de neurônios na camada de entrada.
- O número de camadas ocultas e o número de neurônios em cada camada.
- O número de neurônios na camada de saída.

A seleção destes parâmetros é dependente do problema em questão. Existem diversos métodos de cálculo destes parâmetros, porém são de complexa implementação e não garantem um resultado ótimo (Zhang et al., 1998). Recentemente algoritmos genéticos têm sido utilizados no intuito de otimizar o projeto da topologia de uma RNA com bons resultados (Santos et al., 1999).

3.3.1.1. Decisão do número de neurônios na camada de entrada

O número de neurônio na camada de entrada corresponde ao número de variáveis no vetor de entrada do problema. Em problemas de previsão simples, este número pode ser óbvio ou fácil de ser determinado. Porém, em situações que envolvem cenários mais complexos, não existe uma maneira sistemática de determinar este valor. Tanto um número pequeno quanto um número grandes de entradas pode afetar o desempenho da rede (Zhang et al., 1998). A grande maioria dos autores determina o número de entradas de maneira empírica ou por tentativa, e existem tanto opiniões favoráveis ao uso de um número grande de entradas quanto ao uso de um pequeno número destas.

Esta é uma escolha crítica no projeto de uma RNA. O importante é utilizar entradas que possuam alguma relação implícita com as variáveis de saídas (como por exemplo, utilizar o preço de venda do petróleo Brent para a previsão de ativos da Petrobras). Uma maneira de decidir sobre que entradas utilizar é observar os pesos de cada entrada, após o treinamento da RNA. As entradas com os maiores pesos possuem maior influência no processamento da rede.

3.3.1.2. Decisão do número de camada ocultas e de seu número de neurônios

O número de camadas ocultas e seus componentes são decisivos para a performance de uma RNA. São nos neurônios da camada oculta, que são processados os mapeamentos não-lineares entre entradas e saídas e que são identificados os padrões nos dados. Perceptrons sem camadas ocultas são equivalentes aos modelos lineares tradicionais (Zhang et al., 1998).

A maioria dos autores utiliza apenas uma camada oculta para fins de previsão (Zhang et al., 1998). Porém, tem-se observado que RNAs com duas camadas ocultas apresentam melhores resultados para determinados tipos de problemas (Zhang et al., 1998 e Barron, 1994). Srinivisan et al. (1994) concluiu que o uso de duas camadas ocultas possibilita uma arquitetura mais compacta e com uma maior eficiência no treinamento do que ao utilizar apenas uma camada oculta. Zhang (1994) observou que redes com duas camadas ocultas possuem maior precisão realizando previsões. Lippmann (1987), Cybenko (1988) e Lapedes e Farber (1988) concluíram que duas camadas ocultas são suficientes para a resolução da grande maioria dos problemas envolvendo RNAs, inclusive previsão. De acordo com Zhang(1998), uma camada oculta é suficiente para a arquitetura da maioria dos problemas de previsão, porém, em determinados problemas duas camadas ocultas levam a melhores resultados.

A escolha do número de neurônios em cada camada oculta também é uma tarefa complexa, sem um padrão a ser seguido na literatura. A maneira mais comum é por experimentação e tentativa e erro. Em geral, redes com um pequeno número de neurônios possuem uma melhor capacidade de generalização, porém um número pequeno de neurônios pode ser insuficiente para modelar o problema (Zhang et al., 1998). Youngohc (1993) observou que o aumento do número de componentes na camada de entrada resulta em um aumento de performance, porém apenas até certo ponto. Diversas regras empíricas foram desenvolvidas para determinar o número de componentes em cada camada oculta, geralmente relacionadas ao número de entrada, porém sem um padrão que possa ser seguido para todos os tipos de problema (Zhang et al., 1998).

3.3.1.3. Decisão do número de neurônios na camada de saída

Esta é uma decisão simples, diretamente relacionada com o problema em análise. Geralmente o número de componentes na camada de saída é igual ao número de saídas desejadas. Porém, na literatura existem duas maneiras de realizar previsões (Zhang et al., 1998): um período a frente, no qual se utiliza um único neurônio na camada de saída e se realiza a previsão de um único período, e a previsão de múltiplos períodos, mais comumente encontrada na literatura. Dentro da previsão de múltiplos períodos, existem dois métodos (Zhang et al., 1998). O primeiro é o método iterativo - no qual para cada saída existe apenas um componente na camada de saída e esta é utilizada como entrada no período seguinte. Podemos observar as equações utilizadas neste método iterativo, similares a métodos iterativos tradicionais como Box-Jenkins (Zhang et al., 1998):

$$\bar{x}_{t+1} = f(x_t, x_{t-1}, \dots, x_{t-n}), \quad (3.1)$$

$$\bar{x}_{t+2} = f(x_{t+1}, x_t, x_{t-1}, \dots, x_{t-n+1}), \quad (3.2)$$

.

.

.

$$\bar{x}_{t+k} = f(x_{t+k-1}, x_{t+k-2}, \dots, x_{t+1}, x_t, x_{t-1}, \dots, x_{t-n+k-1}), \quad (3.3)$$

Onde x_t é a observação no período t , \bar{x}_t é a previsão para o período t , f é a função estimada pela RNA.

O segundo é chamado de método direto, no qual o número de períodos é igual ao número de componentes na camada de saída e cada componente é responsável por um período:

$$\bar{x}_{t+1} = f_1(x_t, x_{t-1}, \dots, x_{t-n}), \quad (3.4)$$

$$\bar{x}_{t+2} = f_2(x_t, x_{t-1}, \dots, x_{t-n}), \quad (3.5)$$

.

.

.

$$\bar{x}_{t+k} = f_k(x_t, x_{t-1}, \dots, x_{t-n}), \quad (3.6)$$

Onde as f_1, \dots, f_k são funções determinadas por cada k -ésimo neurônio.

O uso mais freqüente na literatura é do método iterativo. Weigend et al. (1992) e Hill et al. (1994) concluíram que os resultados do método iterativo são superiores ao método direto. Porém, Zhang (1994) obtém melhores resultados com o método direto. No presente trabalho será utilizado método direto.

3.3.2. Função de transferência

Conforme citado no capítulo 2, a função de transferência ou ativação determina a relação entre os neurônios de entrada e saída de uma rede. Em geral, a função de ativação introduz um grau de não-linearidade importante para o desempenho de uma RNA (Zhang et al., 1998).

É possível a utilização de uma função de transferência para as diferentes camadas de uma rede neural. Porém, a grande maioria dos autores utiliza a mesma função de transferência em todas as camadas para obter melhores resultados (Sharda, 1994 e Zhang et al., 1998). A função mais escolhida é a sigmoideal, porém também se observam bons resultados com a tangente hiperbólica como função de ativação (Zhang et al., 1998).

3.3.3. Algoritmo de treinamento

O algoritmo de treinamento é uma importante decisão no projeto de uma RNA. A escolha mais utilizada para soluções de problema com RNAs é o algoritmo de *backpropagation*, inclusive em questões de previsão (Vellido et al., 1999, Zhang et al., 1998 e

Lawrence, 1997). Redes supervisionadas que utilizam alternativas ao BP são raras na literatura (Vellido et al., 1999). Mesmo assim é comum o uso de outros algoritmos para otimizar alguns elementos do BP e de utilizações de variações (Vellido et al., 1999, Zhang et al., 1998, Lawrence, 1997 e Kim e Han, 2000), visando principalmente reduzir o tempo de treinamento, além de incrementar a performance da rede.

3.3.4. Tratamento dos dados

Não existem conclusões na literatura sobre a necessidade de tratamento dos dados e que tipo de tratamento utilizar. Shanker et al. (1996) conclui que um pré-processamento dos dados é benéfico em termos de classificação e erro médio quadrático, porém este benefício diminui com o aumento do tamanho da amostra. Porém, um tratamento dos dados pode levar a um retardo no processo de treinamento.

Em problemas de previsão, a melhor maneira de normalizar os dados é definindo todas as variáveis dentro do mesmo intervalo (Zhang et al., 1998 e Kim e Han, 2000). O intervalo a ser definido depende em geral da função de transferência da camada de saída. Usualmente é utilizado [0,1] para funções sigmóides e [-1,1] para a função tangente hiperbólica (Kim e Han, 2000 e Zhang et al., 1998). Existem diversas maneiras de realizar transformações em dados para RNAs na literatura. As utilizadas com maior frequência são (Zhang et al., 1998),:

- Transformação linear para [0,1]: $x_n = \frac{(x_0 - x_{\min})}{(x_{\max} - x_{\min})}$ (3.7)

- Transformação linear para [a,b]: $x_n = \frac{(b-a)(x_0 - x_{\min})}{(x_{\max} - x_{\min})} + a$ (3.8)

- Normalização estatística: $x_n = \frac{(x_0 - \bar{x})}{s}$ (3.9)

- Normalização simples: $x_n = \frac{x_0}{x_{\max}}$ (3.10)

Onde x_n é o dado normalizado, x_0 é o dado original; x_{\min} , x_{\max} , \bar{x} e s são respectivamente, o mínimo, o máximo, a média e o desvio padrão de uma linha ou coluna selecionada.

Em geral, a normalização das variáveis de entrada e saída pode ser realizada de maneira independente (Zhang et al., 1998). É importante observar que, após a transformação, os resultados obtidos por uma RNA estarão na escala transformada e, dependendo do problema analisado, é necessário transformá-los de volta à escala original para poder interpretá-los.

3.3.5. Separação dos dados em conjuntos de treinamento e teste

Conforme mencionado anteriormente, o primeiro passo no projeto de uma RNA é coletar os dados que serão utilizados pela RNA. Após a coleta é necessário dividir os dados em dois conjuntos distintos: o primeiro será utilizado para o treinamento da rede e o segundo para testar sua performance. É comum o uso de um terceiro conjunto, chamado de validação (Weigend et al., 1992), utilizado para determinar um ponto de parada no processo de treinamento. A seleção dos conjuntos de treinamento e teste é fundamental para o desempenho de uma RNA (Zhang et al., 1998).

A primeira decisão a ser realizada é o tamanho da amostra a ser coletada. Novamente não existe nenhuma conclusão definitiva na literatura sobre este tema. A quantidade de dados depende da arquitetura da rede, do algoritmo de treinamento utilizado, da complexidade do problema e da quantidade de ruído nos dados coletados. Nas tradicionais análises estatísticas, quanto maior o tamanho da amostra coletada, melhor é o seu desempenho (Zhang et al., 1998). Nam e Schaefer (Zhang et al., 1998) realizaram teste com diferentes tamanhos de amostra e concluíram que o desempenho da rede melhora com o aumento do tamanho da amostra. Porém, em RNAs, uma amostra de tamanho elevado para sua estrutura pode levar a um *overtraining*, perdendo sua capacidade de generalização (Haykin, 1994).

Outra importante decisão é a divisão dos dados entre treinamento e teste. Também não existe uma solução geral para este problema, sendo encontradas regras na literatura como 0,9:0,1, 0,8:0,2, 0,7:0,3, etc (Zhang et al., 1998). Porém, o consenso é que a divisão entre os conjuntos não é o mais importante, e sim o tamanho de cada um deles. É necessário que o conjunto de treinamento seja suficiente para que a RNA atinja a performance desejada e que a mesma possa ser avaliada de maneira satisfatória.

3.3.6. Medidas de performance

Existem diversas medidas de performance em RNAs, como o tempo de treinamento, o erro médio quadrático em cada etapa de treinamento, erro nas etapas de validação, etc. Nas aplicações de previsão, entretanto, a medida de performance mais importante para uma RNA

é a acurácia da previsão após a etapa de treinamento (Zhang et al., 1998). Porém, não existe nenhuma medida de precisão aceita universalmente na literatura (Zhang et al., 1998). O grau de acurácia em questões de previsão é geralmente medido em termos do erro entre o valor desejado e o valor obtido na previsão. Existem diversas medidas de acurácia em previsões na literatura, com vantagens e limitações (Zhang et al., 1998). As mais utilizadas são:

- Desvio absoluto médio (MAD): $\frac{\sum |e_t|}{N}$ (3.11)

- Soma dos erros quadráticos (SSE): $\sum (e_t)^2$ (3.12)

- Erro médio quadrático (MSE): $\frac{\sum (e_t)^2}{N}$ (3.13)

- A raiz do erro médio quadrático (RMSE): \sqrt{MSE} (3.14)

- Erro médio percentual absoluto (MAPE): $\frac{1}{N} \sum \left| \frac{e_t}{y_t} \right| (100)$ (3.15)

Onde e_t é o erro individual de cada previsão, y_t é o valor atual e N é o número de períodos avaliados. A medida mais usada para acurácia na literatura é o erro médio quadrático (MSE).

3.4. Aplicações de RNAs como ferramentas de previsão

RNAs têm sido utilizadas como uma ferramenta de previsão em diversas áreas de aplicação, tornando difícil cobrir toda a literatura acerca desta tema (Vellido et al., 1999).

Algumas destas aplicações incluem (Vellido et al., 1999 e Zhang et al., 1998) contabilidade (previsão de ganhos financeiros, previsão de quebras e falência), mercado financeiro (previsão do movimento, de ganhos futuros e do risco do mercado de ações, taxas de juros, preços futuros de ações, etc), marketing (previsão de escolha do consumidor, *market share*, tendências de mercado, demanda e vendas de um novo produto), economia (previsão de ciclos econômicos, recessões, inflação, produção industrial, crescimento do PIB, títulos do tesouro americano), engenharia (previsão de demanda de eletricidade, da demanda de uso do

telefone, controle de inventário, sucesso de um novo produto), ramos imobiliários (demanda por construção de novas residências, valores futuros de casas), turismo (previsão do número de turistas, demanda de viagens, número de passageiros), transporte (tráfego terrestre e aéreo) e assuntos envolvendo o meio-ambiente (previsão do nível de ozônio e sua concentração, qualidade do ar). A Tabela 3.1 apresenta um resumo de alguns artigos publicados (Vellido et al., 1999, Zhang et al., 1998 e Araújo e Vasconcelos, 1999).

Tabela 3.1: resumo de publicações envolvendo previsões com RNAs

Problemas estudados	Publicações
Ganhos contábeis	Callen et al. (1996), Dhar eChou (2001), Spears e Leis (1997)
Ciclos econômicos e recessão	Qi (2001)
Falência, quebra ou saúde financeira	Yang (1999), Zhang et al. (1999), Mckee Greenstein (2000), Anandarajan et al. (2001), Atiya (2001), Mozer e Smolensky (1997)
Despesas de consumo	Church e Curram (1996)
Preços de <i>commodities</i> , preços ótimos	Kohzadi et al. (1996), Yao et al. (2000)
Marketing	Agrawal e Schorling (1996), West et al. (1997), Aiken e Bsat (1999), Wang (1999), Jiang et al. (2000), Vellido et al. (1999), Setiono et al. (1998)
Demanda de Eletricidade	Elkateb et al. (1998), Darbellay e Slama (2000), Hippert et al. (2001), Araújo Jr. e Vasconcelos (1999)
Taxa de câmbio	Zhang e Hu (1998), Leung et al. (2000a), Nag e Mitra (2002)
Inflação, produção industrial, crescimento do PIB	Tkacz (2001), Chen et al. (2001a), Tseng et al. (2001)
Turismo	Nam e Schaefer (1995), de Carvalho et al. (1998), Law(2000)
Controle de inventário	Bansal e Vadhavkar (1998), Partovi e Anandarajan (2002)
Concentração e nível de ozônio, previsão ambiental, qualidade do ar	Prybutok et al. (2000), Ruiz Surez et al. (1995), Murtagh et al. (2000), Kolehmainen et al. (2001)
Demanda e venda de produtos, vendas no varejo	Ansuj et al. (1996), Luxhoj et al. (1996), Charytoniuk et al. (2000), Alon et al. (2001), Kuo (2001), Zhang e Qi (2002), Dutta, Shekhar e Wong (1994), Hruschka (1993)
Sucesso de projetos	Thieme et al. (2000), Zhang et al. (2003)
Setor imobiliário	Hua (1996), Goh (1998), Nguyen e Cripps (2001)
Mercado financeiro	Wang e Leu (1996), Wittkemper e Steiner (1996), Desai e Bharati (1998), Saad et al. (1998), Qi (1999), Leung et al. (2000b), Chen et al. (2003)
Tráfego	Doughetry e Cobbett (1997), Kirby et al. (1997), Chen et al. (2001b), Dia (2001), Qiao et al. (2001)
Títulos do tesouro americano	Cheng et al. (1996)

O primeiro uso de RNAs para previsão foi realizado por Hu em 1964 (Zhang et al., 1998), que utilizou uma rede linear para realizar previsões climáticas. Devido à falta de um algoritmo de treinamento para redes de múltiplas camadas na época, a pesquisa foi bastante

limitada. Werbos em 1974 (Zhang et al., 1998) foi o primeiro a realizar um estudo comparativo entre RNAs e métodos estatísticos de regressão para previsões. Porém, foi apenas em 1986, com a introdução do algoritmo *backpropagation* (Rumelhart et al., 1986), que RNAs passaram a ser utilizadas como uma ferramenta poderosa para previsões.

Em 1987, Lapedes e Farber através do uso de simulações, concluíram que redes neurais diretas podem ser utilizadas para modelar e para prever séries temporais caóticas e não-lineares. Desde então, diversos problemas de previsão foram resolvidos por redes neurais, sendo realizadas diversas revisões da literatura: Dougherty (1995); Zhang (1998); Wong and Selvi (1998) e Fadlalla e Lin (2001) trataram especificamente de aplicações financeiras; Vellido (1999) tratou de aplicações em economia e Maier e Dandy (2000) de previsão de reservas de água.

3.4.1. Aplicações em mercados financeiros

Conforme citado no primeiro capítulo, O mercado financeiro é afetado por um grande número de fatores econômicos, políticos e psicológicos, altamente relacionados e que interagem entre si de uma maneira bastante complexa. A maior parte destas relações dificilmente podem ser expressas através de regras determinadas (Thawornwong et al., 2001, Bogullu et al., 2002, Yoon e Swales, 1993 e Kim e Han, 2000).

Weigend (1990 e 1991) estudou a topologia de uma RNA para resolver problemas de previsão de séries temporais reais. A partir deste ponto, o uso de RNAs para aplicações de previsões em mercados financeiros aumentou consideravelmente, principalmente por tratarem com séries temporais não-lineares que podem ser mapeadas por RNAs (Lawrence, 1997). Além disto, RNAs oferecem uma resposta dinâmica que possibilita uma rápida tomada de decisão e torna-se um fator decisivo para a obtenção competitividade no mercado (Freitas e Silva, 1999).

Kuvayev (1996) realizou um resumo das aplicações de redes neurais em previsões de mercados financeiros. Kulkarni (1996) utilizou redes neurais para prever os valores futuros do Índice Standard & Poor's 500. Zekić (1998) analisou a metodologia utilizada em RNAs para previsões. Thawornwong et al.(2002) e Bogullu et al. (2002) realizaram estudos para prever a tendência de índices de ações com RNAs.

É comprovado que nos últimos anos, o uso de redes neurais para previsão em mercados financeiros vêm crescendo, principalmente no uso de valores de ações, retornos de investimentos e falência (Vellido et al., 1999), (Zhang et al., 1998). A capacidade de

identificar relações não-lineares faz com que as RNAs destaquem-se como uma atrativa técnica para estas aplicações, sendo uma alternativa aos métodos tradicionais.

4. JUSTIFICATIVA

4.1. O mercado de ações

O mercado de ações ou bolsa de valores como também é conhecido, são sociedades civis sem fins lucrativos onde são negociados títulos públicos conhecidos como ações.

As ações são títulos de renda variável, emitidos por sociedades anônimas, representando a menor fração de seu capital. O investidor que realiza a compra de uma ação se torna co-proprietário da empresa eminente (acionista).

Segundo Bueno (2000), as ações de companhias abertas, de acordo com os direitos de seus titulares podem ser:

- **Ordinárias.** São aquelas que conferem ao acionista direito de voto em assembleias gerais da empresa e que proporcionam uma participação nos resultados (dividendos).
- **Preferenciais.** Proporcionam ao acionista uma prioridade no reembolso do capital, no caso da dissolução da sociedade, e no recebimento dos dividendos, e podem garantir o direito a voto.

As ações são conhecidas e negociadas como:

- **ON.** Ordinária nominativa.
- **PN.** Preferencial nominativa.

A rentabilidade das ações é variável, composta pelos dividendos, benefícios concedidos pela empresa e da eventual venda da ação. As ações são conversíveis a dinheiro a qualquer momento, bastando realizar a transação na bolsa de valores.

4.2. Análises do mercado de ações

4.2.1. Hipótese do Mercado Eficiente

A Hipótese do Mercado Eficiente (Efficient Market Hypothesis - EMH) atesta que em qualquer período, o preço de uma ação captura toda a informação disponível sobre esta ação. Como toda a informação é utilizada pelos participantes do mercado, as variações ocorrem de

maneira aleatória de acordo com a incidência de novas informações, e a aplicação destas informações pelos analistas leva ao equilíbrio do mercado (Lawrence, 1997). Conseqüentemente, se torna impossível desenvolver métodos de “ganhar” do mercado, sem que o equilíbrio seja restabelecido.

Existem três níveis descritos de eficiência no mercado (Fama, 1970):

- **Forma fraca de eficiência.** Os preços históricos dos ativos não conteriam nenhuma informação útil na obtenção de retornos extraordinários para o investidor.
- **Forma semiforte de eficiência.** Nenhum investidor conseguiria retornos extraordinários com base em informações públicas. O mercado entraria em equilíbrio rapidamente.
- **Forma forte de eficiência.** Nenhum investidor conseguiria retornos extraordinários, mesmo com informações privadas, pois os preços já refletem toda a informação disponível, mesmo a que não é pública.

A EMH é importante, pois contradiz todas as formas de análise, ou seja, seria impossível desenvolver modelos para obter retornos extraordinários no mercado, e que qualquer modelo desenvolvido possui o mesmo desempenho de tentativas aleatórias (Lawrence, 1997).

Existe uma contradição na literatura sobre a validade da EMH (Lawrence, 1997 e Bruni, 1998). Porém, a grande parte dos trabalhos a rejeita, principalmente devido a três aspectos. O primeiro, a EMH supõe que toda a informação estará disponível a qualquer momento sem custo para todo participante do mercado. O segundo é que os preços das ações não variam apenas de acordo com as informações disponíveis, mas também com aspectos psicológicos dos participantes ou através de grandes investidores manipulando o mercado. E finalizando, a existência de participantes que conseguem obter grandes retornos constantemente, sejam por informações privilegiadas, ou por grandes recursos computacionais ou ainda de uma maneira ilícita, etc.

4.2.2. Análises Técnica e Fundamentalista

A análise do mercado de ações segue duas abordagens distintas, que são denominadas análise técnica e análise fundamentalista (Lawrence, 1997).

4.2.2.1. Análise Técnica

No início do século XX, Charles Henry Dow foi o primeiro a formular uma teoria sobre o estudo do movimento dos preços, utilizando gráficos. Dow, juntamente com Edward

Davis Jones e Charles Milford Bergstresser, fundaram o *Wall Street Journal* e a *Dow Jones & Company*, ainda hoje referências em análises financeiras. Os estudos de Dow deram origem à análise técnica do mercado de ações.

A análise técnica utiliza a idéia de que os movimentos de tendência de alta ou baixa de uma ação são determinados por padrões recorrentes, sendo possível prever estes movimentos. Esta análise utiliza-se de gráficos de transações como preços (de abertura, fechamento, máximo e mínimo) e volume negociado para identificar padrões que auxiliem na decisão de compra ou venda de um determinado título.

Apesar de ser criticada por ser subjetiva (diferentes pessoas podem interpretar os gráficos de maneiras diferentes) e por requerer bastante conhecimento para sua realização, a análise técnica é utilizada por 90% dos principais participantes do mercado.

4.2.2.2. Análise Fundamentalista

A análise fundamentalista procura medir o valor intrínseco de um ativo, através do estudo das informações do mercado e da empresa emitente, como ganhos e rentabilidade desta empresa, sua credibilidade, expectativa de taxas de juros, crescimento do setor em que a empresa está inserida, etc. Desta maneira, a análise fundamentalista traça as perspectivas futuras do mercado de ações e que ativos proporcionarão um maior retorno para os investidores.

A maior vantagem versus à análise técnica é que a fundamentalista apresenta melhor desempenho no longo prazo para previsão de estabilidade e crescimento, pois possui capacidade de prever certas tendências antes de estas aparecem nos gráficos. Porém também é uma análise subjetiva, além de ser necessário um grande custo para identificar os fatores que influenciarão em um determinado ativo e mensurar o tamanho desta influência. Fazendo com que a análise fundamentalista seja restrita a um grupo de investidores.

4.3. Volatilidade do mercado de ações

De acordo com a EMH, em um mercado com agentes racionais, os valores das ações mudariam apenas com a chegada de novas informações. Entretanto, o que é observado na prática é que os movimentos de determinadas ações apresentam grandes flutuações que não podem ser explicadas apenas pela EMH (Shiller, 1981). Estudos mostram pouca ou nenhuma correlação entre a chegada de novas informações no mercado e a variabilidade de preços (Milanez, 2003). Ou seja, existe uma volatilidade que não pode ser expressa apenas pelas informações disponíveis no mercado. E certas relações entre variáveis inseridas no mercado,

que determinam a tendência de alta ou baixa de uma ação não podem ser facilmente determinadas.

A volatilidade do mercado ajuda a explicar a imprevisibilidade no mercado de ações. Mesmo as análises supracitadas não garantem que o investidor obterá altos retornos de maneira consistente.

4.4. O mercado de ações brasileiro

O Brasil é considerado hoje um dos mercados emergentes mais promissores para investimentos, e um dos principais motivos é a série de mudanças em sua economia.

Até o início da década de 90, a estrutura econômica brasileira possuía um elevado nível de intervenção do governo, sendo extremamente fechada. O fluxo de capital internacional era controlado e havia o problema da inflação crescente (Bueno, 2000). Então, o país começou a adotar uma série de medidas de desregulamentação, modernização e a reduzir o papel estatal, visando abrir sua economia e atrair maiores investimentos internacionais. Logo o mercado de capitais passou a ter uma maior importância em substituição ao mercado de crédito.

Com o advento do Plano Real em 1994, o Brasil controlou a inflação e atingiu a estabilidade de preços, criando um ambiente favorável para o amadurecimento da bolsa de valores. Hoje, o único retardador deste processo é a elevada taxa de juros (Bueno, 2000). Caso os juros reais se apresentem como uma boa alternativa de investimentos, os investidores optarão por investir em títulos de renda fixa e não variável.

Contudo, mesmo com os juros elevados, o Brasil vem passando por um momento favorável para aplicação em bolsa de valores, popularizando os investimentos em ações, mesmo que através de fundos de investimento.

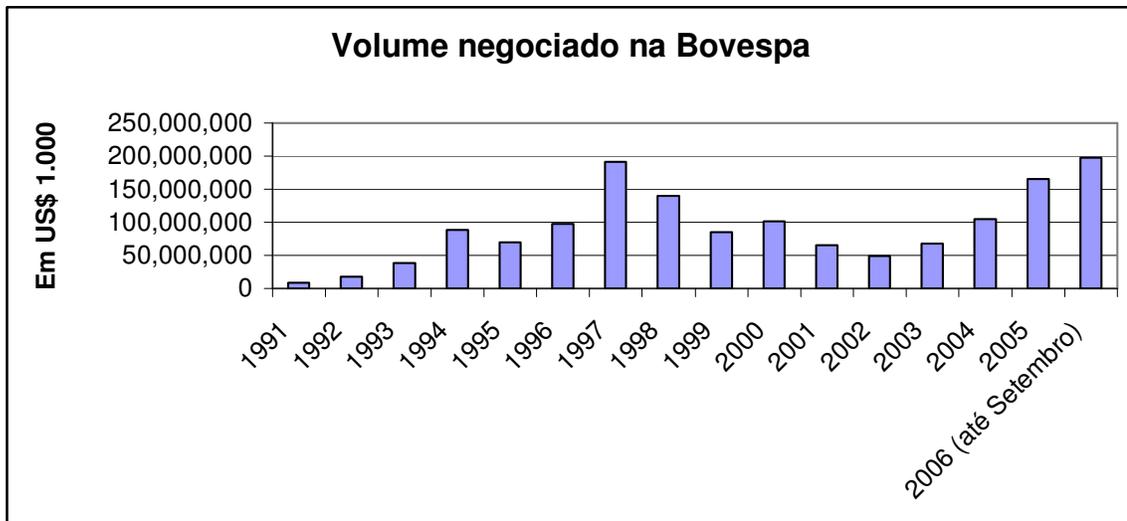


Gráfico 4.1: Volume negociado na Bovespa (Fonte: Bovespa)

Portanto, cresce o interesse por estratégias que possam ocasionar em uma maior rentabilidade para o investidor. Assim, é possível justificar a importância de se desenvolver modelos que utilizem redes neurais como ferramentas para a previsão de valores e tendências futuras no mercado de ações.

5. ESTIMAÇÕES E RESULTADOS

5.1 Introdução

Neste capítulo, serão apresentados o desenvolvimento de redes neurais e seus resultados experimentais, de acordo com o descrito nos capítulos anteriores. O objetivo é a realização da previsão dos valores e das tendências de alta ou baixa de ações, especificamente a Petrobras PN e o índice Ibovespa, referente à Bolsa de Valores de São Paulo (BOVESPA).

Foram realizadas cerca de duzentas simulações com diferentes arquiteturas de RNAs, baseados inicialmente na “tentativa e erro”, porém posteriormente identificando a melhor topologia para o problema. Os que apresentaram melhores resultados possuíam redes diretas utilizando o algoritmo de *backpropagation*. Ao final foram selecionados quinze modelos para discussão.

A escolha dos modelos visou mostrar a evolução do trabalho, focando naqueles que apresentaram os melhores resultados e maior relevância na construção dos resultados finais. Foram divididos em cinco seções, com diferentes abordagens para o problema analisado.

5.2 Descrição do modelo de RNA utilizado

5.2.2. Seleção e processamento dos dados

Os dados utilizados como entrada para as RNAs são as séries diárias de variáveis macroeconômicas, obtidas no *site* do Instituto de Pesquisa Econômica Aplicada (IPEA).

O período escolhido foi do dia 02 de Maio de 1997 ao dia 20 de Julho de 2006, pois as cotações de todas as variáveis de interesse estavam disponíveis a partir desta data. Foram excluídos os dias em que uma das variáveis não estava cotada, sendo considerado que a exclusão de um dia é menos prejudicial à importância da continuidade das séries temporais do que o fato de uma variável ser considerada nula. No apêndice A é possível encontrar a tabela com os valores e datas de todas as variáveis utilizadas.

Desta maneira o conjunto total de dados ficou com 2.195 períodos de tempo. Durante as simulações, este conjunto foi separado de diferentes formas nos dados que foram utilizados no treinamento e no teste, porém procurando mantê-lo próximo à proporção de 90% para treinamento e 10% para teste.

Visando reduzir a variância e uniformizar a escala de análise, todos os dados de entrada foram submetidos a uma transformação linear, de modo a ficarem limitados ao intervalo de +1 a -1, utilizando a seguinte fórmula:

$$x^* = \frac{2x - [\max(x) + \min(x)]}{\max(x) - \min(x)} \quad (5.1)$$

Além da transformação, verificou-se que os resultados eram incrementados com a inclusão de variáveis dependentes que aumentaram a capacidade de generalização da rede, melhorando sua performance. Optou-se pelo escalonamento de algumas variáveis de entrada e além do uso de variáveis percentuais e de diferença. Na tabela 5.1 estão descritas todas as variáveis utilizadas:

Tabela 5.1: variáveis de entrada utilizadas.

Variáveis de Entrada	Descrição
SelicN	Valor da taxa Selic no período t
CambioVendaN	Taxa de câmbio R\$/US\$ comercial no período t
DowjonesN	Fechamento da Dowjones no período t
DowjonesN-1	Fechamento da Dowjones no período (t-1)
DowjonesN-2	Fechamento da Dowjones no período (t-2)
DowjonesN-3	Fechamento da Dowjones no período (t-3)
BovespaN	Fechamento da Bovespa no período t
BovespaN-1	Fechamento da Bovespa no período (t-1)
BovespaN-2	Fechamento da Bovespa no período (t-2)
BovespaN-3	Fechamento da Bovespa no período (t-3)
dBovespaN1	Diferença entre o período t e (t-1) da Bovespa
dBovespaN2	Diferença entre o período t e (t-2) da Bovespa
NasdaqN	Fechamento da Nasdaq no período t
NasdaqN-1	Fechamento da Nasdaq no período (t-1)
NasdaqN-2	Fechamento da Nasdaq no período (t-2)
NasdaqN-3	Fechamento da Nasdaq no período (t-3)
BrentN	Fechamento do preço do petróleo Brent no período t
BrentN-1	Fechamento do preço do petróleo Brent no período (t-1)
BrentN-2	Fechamento do preço do petróleo Brent no período (t-2)
BrentN-3	Fechamento do preço do petróleo Brent no período (t-3)
Petr4N	Fechamento da Petrobrás PN no período t
Petr4N-1	Fechamento da Petrobrás PN no período (t-1)
Petr4N-2	Fechamento da Petrobrás PN no período (t-2)
Petr4N-3	Fechamento da da Petrobrás PN no período (t-3)
dPetr4N1	Diferença entre o período t e (t-1) da Petrobrás PN
dPetr4N2	Diferença entre o período t e (t-2) da Petrobrás PN
VarPetr	Variação percentual da Petr4 entre o período t e período (t-1)

No gráfico 5.1, podemos observar todas as variáveis macroeconômicas após a transformação linear. No apêndice A, constam os gráficos de cada variável utilizada, onde é possível analisá-las individualmente.

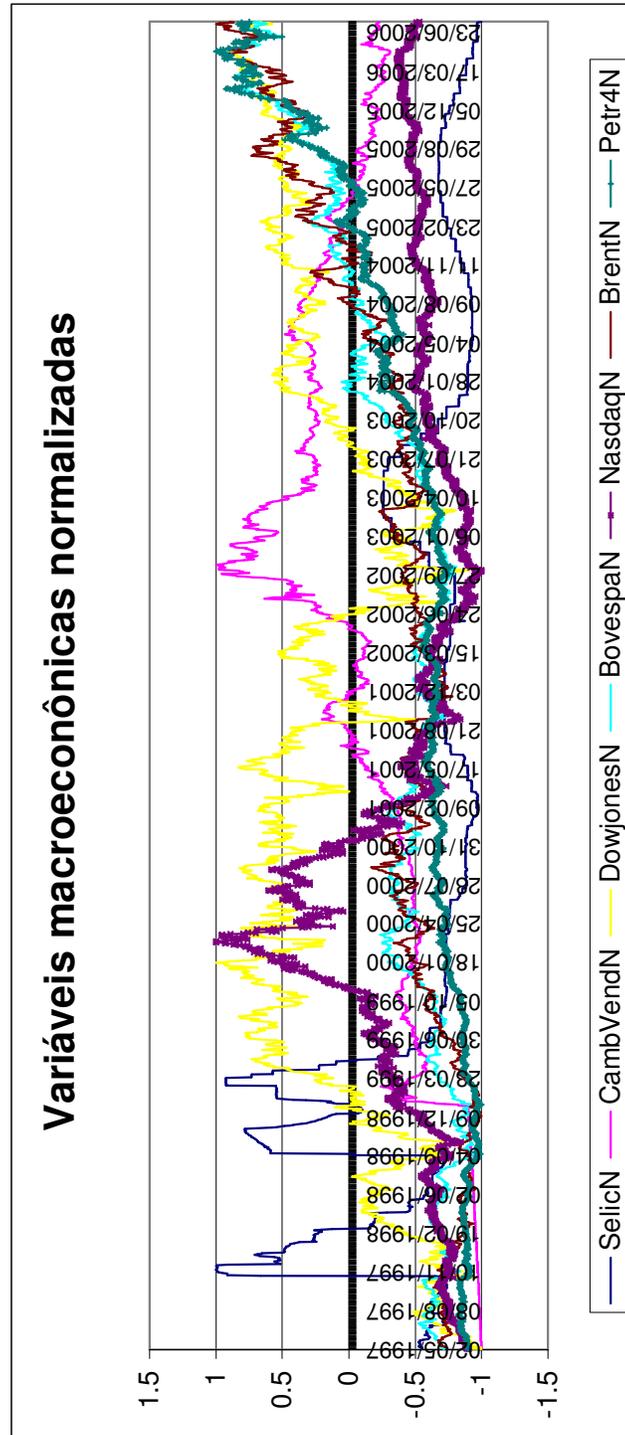


Gráfico 5.1: gráfico das entradas utilizadas

5.2.2. Arquitetura e topologia

A arquitetura escolhida para as redes neurais foi uma rede *feedforward* baseada no *perceptron* multi-camadas (MLP) com aprendizado supervisionado, pois, conforme foi descrito, apresenta os melhores resultados com previsão.

Após uma série de testes iniciais, optou-se pelo uso de duas camadas ocultas em todas as RNAs desenvolvidas, por apresentar uma melhor relação de desempenho *versus* o tempo de treinamento. Devido à falta de metodologias na literatura, foi adotada a sugestão inicial do *software* para o número de elementos em cada camada. Após testes observou-se que esta sugestão inicial apresentava resultados satisfatórios e que o desempenho da rede melhora quando é aumentado o número de componentes em cada camada oculta para simulações com mais períodos de treinamento e menor número de variáveis de entrada, porém até um certo limite que depende da natureza do problema.

Foram utilizadas como função de ativação a sigmoideal, a sigmoideal linear (*linearsig*) e a tangente hiperbólica linear (*lineartanh*).

5.2.3. Algoritmo de treinamento

O algoritmo de treinamento utilizado foi o *backpropagation* (BP). A taxa de aprendizagem utilizada foi de 0,1 e foi incluído um *momentum* de 0,7. Além disto, foi utilizado na otimização destes parâmetros um algoritmo genético simples conforme descrito anteriormente, com uma probabilidade de cruzamento de 0,9 e de uma probabilidade de mutação de 0,01. Foram utilizadas 1.000 iterações em cada treinamento e, como critério de parada, o caso em que o erro médio quadrático (MSE) fosse menor que 0,01. No início de cada treinamento, todos os pesos eram escolhidos aleatoriamente.

Todos os parâmetros são padrões do *Neurosolutions* e após testes foram validados como boas opções de desempenho.

5.2.4. Medição de performance de uma RNA

Como medição da performance das redes neurais, foram utilizadas as seguintes variáveis:

- Erro médio quadrático (MSE):
$$MSE = \frac{\sum_{i=0}^N (d_i - x_i)^2}{N} \quad (5.2)$$

- Erro absoluto médio (MAE): $MAE = \frac{1}{N} \sum_{i=1}^N |x_i - d_i|$ (5.3)

- Erro absoluto mínimo.
- Erro absoluto máximo.

- Coeficiente de correlação linear (r): $r = \frac{\sum (x_i - \bar{x}) (d_i - \bar{d})}{\sqrt{\frac{\sum (d_i - \bar{d})^2}{N}} \sqrt{\frac{\sum (x_i - \bar{x})^2}{N}}}$ (5.4)

Onde, x é a saída da rede, d é o valor desejado e N é o número de amostras.

Além disso, foram analisados os pesos da rede, visando determinar quais variáveis foram mais importantes para a tomada de decisão da RNA.

5.3. Estimações

5.3.1. Previsão do valor atual da ação Petrobras PN

Em princípio, foi decidido realizar a previsão do valor atual da ação Petrobras PN (PETR4). Foram desenvolvidos dois modelos diferentes:

Modelo 1. O objetivo deste modelo foi a previsão do valor diário da PETR4. Foi levado em consideração todo o conjunto de dados, sendo utilizado os primeiros 2.000 períodos (de 02 de Maio de 1997 até 21 de Setembro de 2005) para treinamento, enquanto os 195 períodos restantes (22 de Setembro de 2005 até 20 de Julho de 2006) foram utilizados para teste.

Modelo 2. Este modelo tem o mesmo objetivo e estrutura do anterior, porém não foi utilizado todo o conjunto de dados, foram considerados apenas os primeiros 1.900 períodos (até 27 de Abril de 2003), com o objetivo de diminuir a variação entre os conjuntos de treinamento e teste, pois próximo a este período o valor da PETR4 apresentou um alto índice de crescimento. O conjunto de dados foi repartido em 1.375 períodos de treinamento e 175 períodos de teste.

A tabela 5.2 apresenta as variáveis de entrada de cada modelo, enquanto que a tabela 5.3 apresenta as características da arquitetura de cada modelo:

Tabela 5.2: variáveis de entrada dos modelos 1 e 2.

Entrada	Modelo 1	Modelo 2
DowjonesN	x	x
BovespaN	x	x
NasdaqN	x	x
BrentN	x	x

Tabela 5.3: arquitetura dos modelos 1 e 2.

Característica	Nº de neurônios na primeira camada	Nº de neurônios na segunda camada	Função de ativação
Modelo 1	40	20	sigmoidal
Modelo 2	35	17	sigmoidal

5.3.1.1. Resultados

Após os treinamentos e simulações foram encontrados os seguintes resultados:

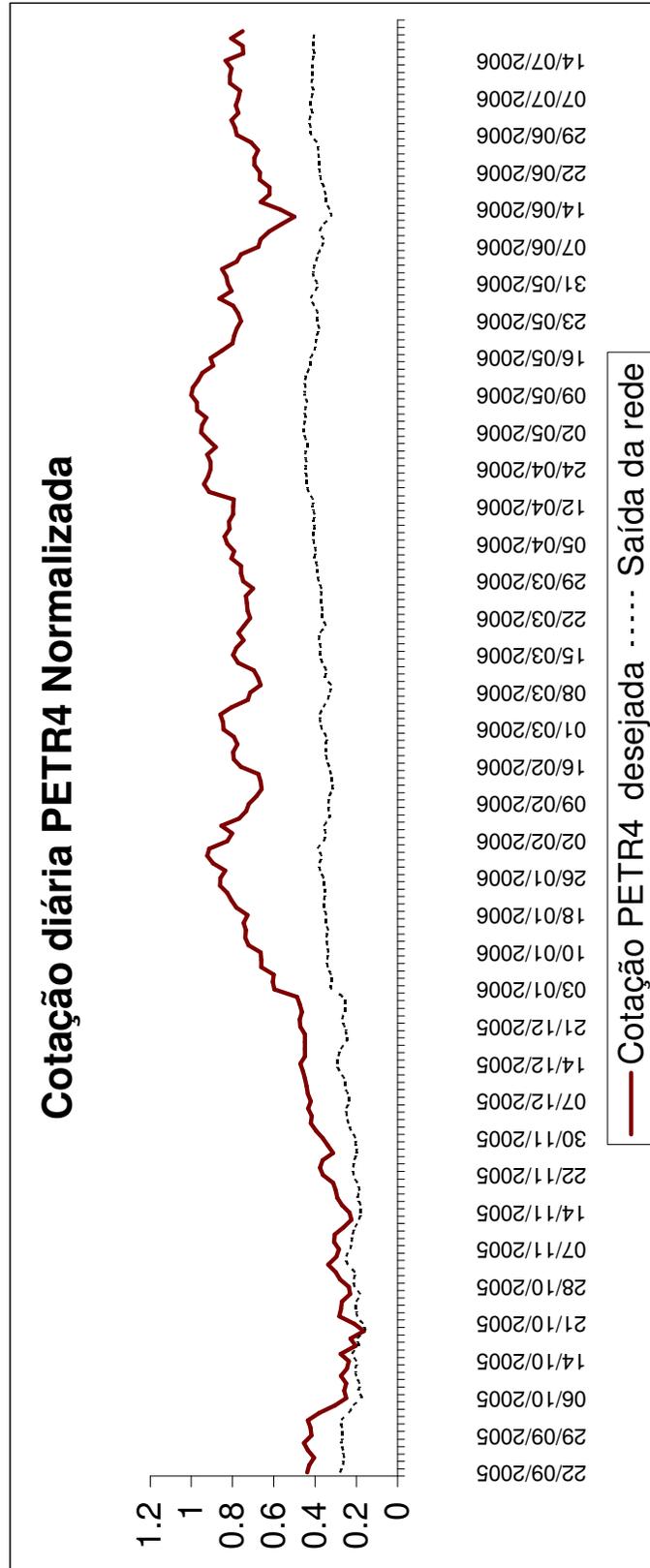


Gráfico 5.2: resultado do Modelo 1.

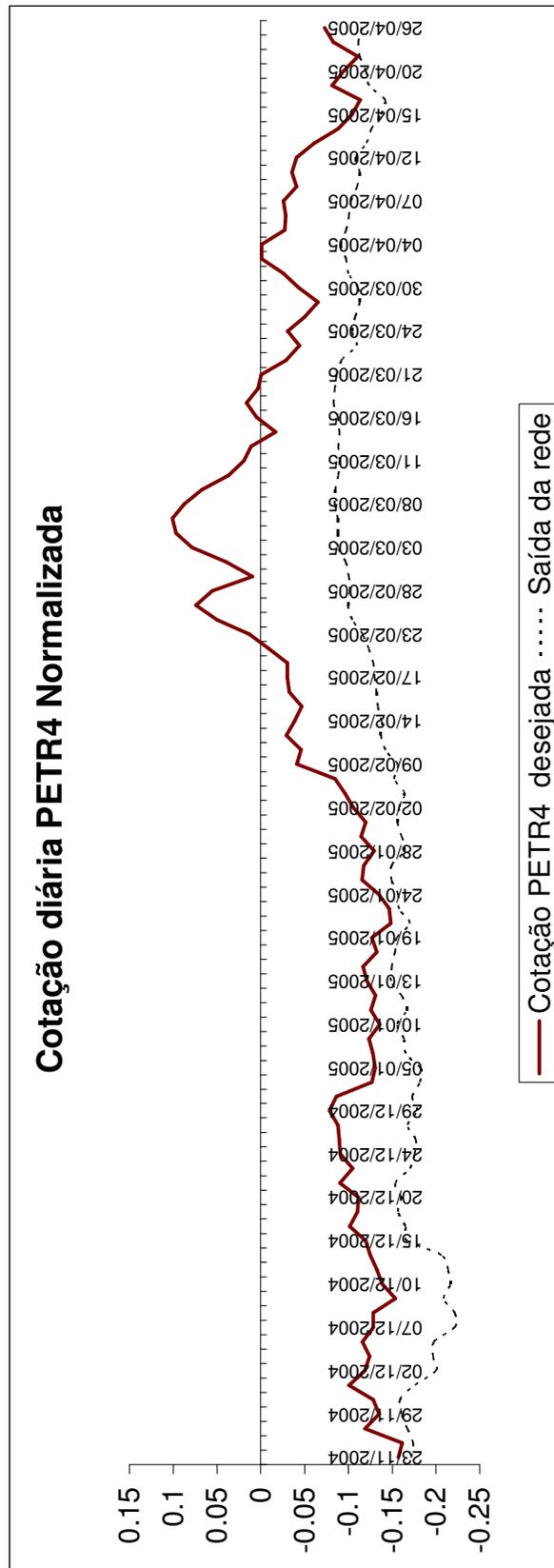


Gráfico 5.3: resultado do Modelo 2.

Tabela 5.4: pesos das entradas utilizadas

Entrada	Modelo 1	Modelo 2
DowjonesN	1.187170	0.979841
BovespaN	0.948595	1.341925
NasdaqN	0.207402	0.082491
BrentN	2.547143	2.125729

Tabela 5.5: performance dos modelos 1 e 2.

	Modelo 1	Modelo 2
MSE	0.115727	0.006989
MAE	0.306507	0.073005
Erro min abs	0.005865	0.002049
Erro max abs	0.555133	0.188648
r	0.960831	0.825687

Como é possível observar no apêndice A, o valor de PETR4 apresenta uma elevada taxa de crescimento médio no período coletado. Como o conjunto de treinamento é formado pelos primeiros períodos, a RNA só consegue apresentar bons resultados quando o valor da saída desejada é próximo aos valores de treinamento.

No modelo 2 foram desconsiderados os últimos 295 períodos, visando corrigir esta oportunidade. O problema foi reduzido e este apresentou menores erros em relação ao modelo 1, conforme a tabela 5.3. Porém a rede ainda proporcionou uma defasagem em relação ao valor real.

As variáveis que apresentaram o maior peso e conseqüentemente a maior importância na tomada de decisão da RNA, encontram-se destacadas em negrito, sendo este o método utilizado no restante do trabalho. Em ambos os modelos, a BrentN apresentou-se como a variável de maior importância, seguida pela BovespaN. Este é um resultado esperado, visto que ambas as variáveis possuem um grande peso na formação do preço da ação da Petrobras e também apresentaram crescimento médio similar ao da Petr4. Porém, quando se tratou de realizar previsões de períodos futuros, os modelos testados – nos quais a variável BrentN foi inserida - não apresentaram bons resultados, sendo esta desconsiderada daqui por diante.

Apesar das oportunidades de majoração das RNAs, também é possível identificar que as redes conseguiram modelar boa parte dos movimentos de subida e descida da PETR4 e ambos os modelos apresentaram um bom índice de correlação linear, com maior destaque para o primeiro modelo.

5.3.2. Previsão do valor futuro da ação Petrobras PN

Em seguida, foram desenvolvidas RNAs com objetivos mais voltados à uma aplicação prática: a previsão do preço da PETR4 no próximo dia. Conforme citado no capítulo anterior, o interesse do investidor está em obter os valores de ações em planos futuros. Foram desenvolvidos três modelos:

Modelo 3. O objetivo deste modelo foi realizar a previsão do valor da PETR4 no período $(t+1)$. Foram utilizados escalonamentos nas variáveis de entrada, para fornecer uma memória de curto prazo a rede (os últimos três períodos). Portanto, o conjunto de dados ficou reduzido em 4 períodos (já que no último período não existe a informação do período futuro), sendo iniciado em 07 de Maio de 1997. O conjunto de treinamento foi formado pelos 2.000 primeiros períodos e o conjunto de teste pelos 191 períodos restantes.

Modelo 4. Apresenta o mesmo objetivo do modelo 3, porém com a redução do conjunto de dados para 1.500 períodos, sendo os primeiros 1.375 o conjunto de treinamento e os 125 restantes o conjunto de dados.

Modelo 5. Similar ao modelo anterior, porém com o uso do mesmo conjunto de treinamento do modelo 3. Entretanto, o conjunto de testes utilizado foram apenas os 51 períodos subsequentes (22 de Setembro de 2005 à 12 de Dezembro de 2005). Foram utilizadas menos variáveis de entrada, mas com a introdução da variável $dPetr4N1$, que não foi utilizada nos modelos anteriores.

Tabela 5.6: variáveis de entrada dos modelos 3, 4 e 5.

Entrada	Modelo 3	Modelo 4	Modelo 5
DowjonesN	x	x	x
DowjonesN-1	x	x	
DowjonesN-2	x	x	
DowjonesN-3	x	x	
BovespaN	x	x	x
BovespaN-1	x	x	
BovespaN-2	x	x	
BovespaN-3	x	x	
NasdaqN	x	x	x
NasdaqN-1	x	x	
NasdaqN-2	x	x	
NasdaqN-3	x	x	
Petr4N	x	x	x
Petr4N-1	x	x	x
Petr4N-2	x	x	x
Petr4N-3	x	x	x
dPetr4N1			x
VarPetr	x		

Tabela 5.7: arquitetura dos modelos 3, 4 e 5.

Característica	Nº de neurônios na primeira camada	Nº de neurônios na segunda camada	Função de ativação
Modelo 3	11	5	sigmoidal
Modelo 4	8	4	sigmoidal
Modelo 5	22	1	sigmoidal

5.3.2.1. Resultados

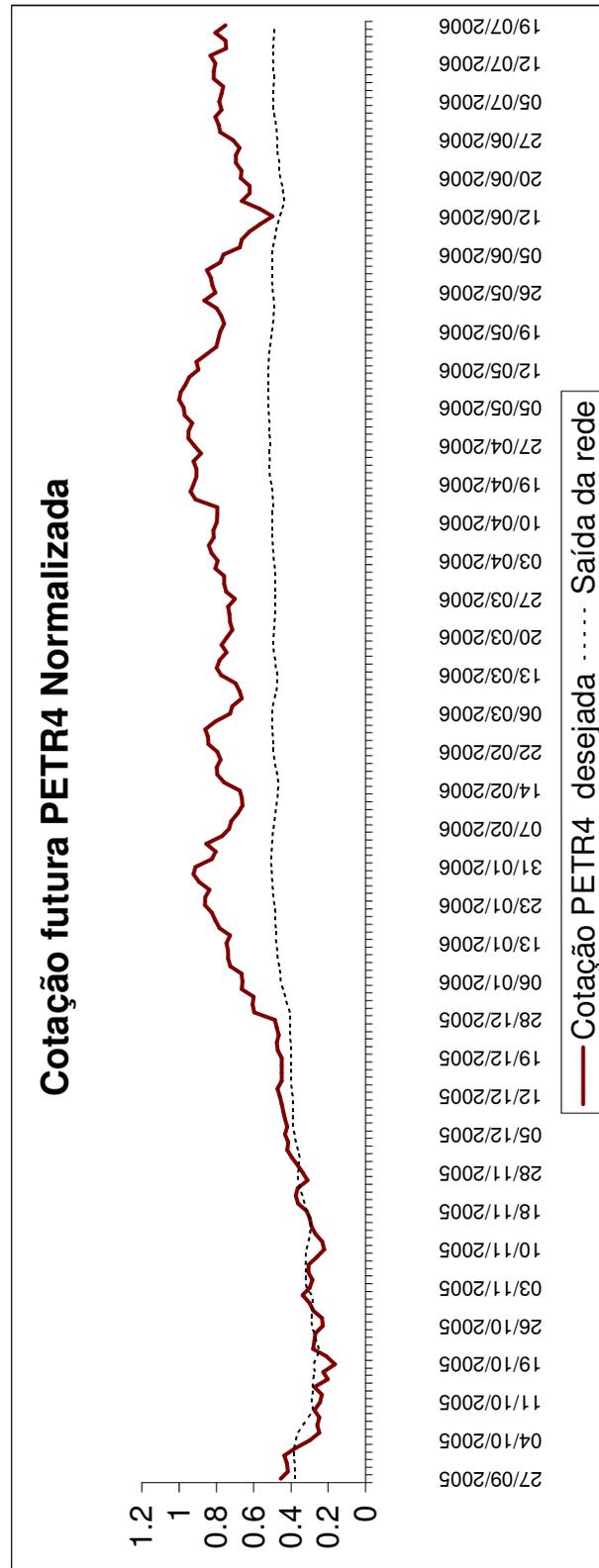


Gráfico 5.4: resultado do Modelo 3.

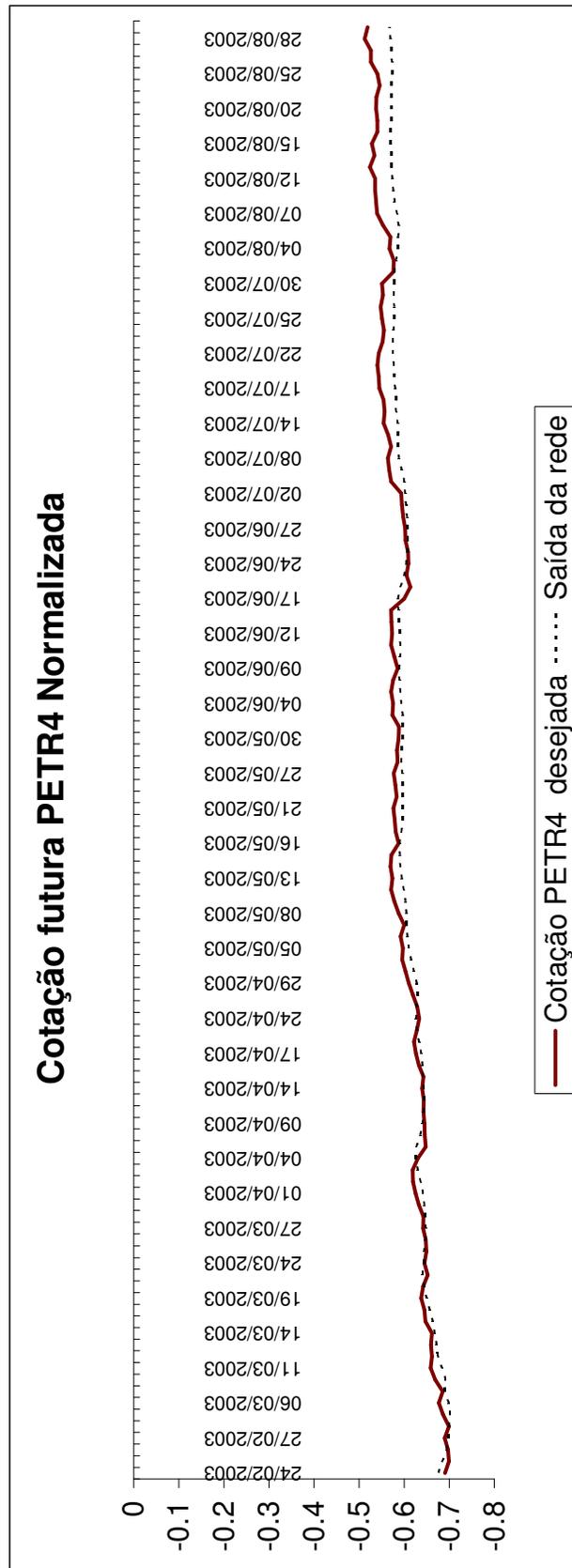


Gráfico 5.5: resultado do Modelo 4.

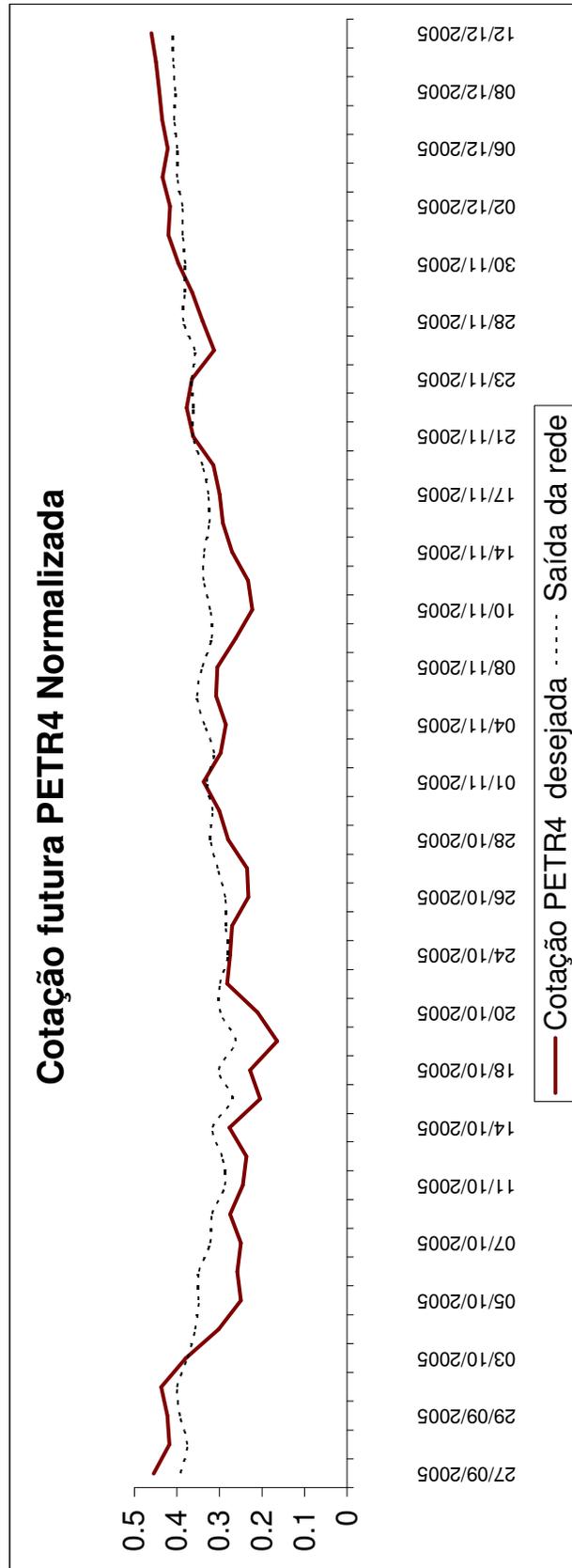


Gráfico 5.6: resultado do Modelo 5.

Tabela 5.8: pesos das entradas utilizadas.

Entrada	Modelo 3	Modelo 4	Modelo 5
DowjonesN	0.305010	0.026431	0.557014
DowjonesN-1	-0.047538	0.209493	
DowjonesN-2	0.167239	0.513727	
DowjonesN-3	0.639922	-0.060965	
BovespaN	-0.082980	0.256794	0.574073
BovespaN-1	0.196275	-0.268033	
BovespaN-2	0.358277	0.208426	
BovespaN-3	-0.379986	0.202333	
NasdaqN	0.277181	0.052560	0.603703
NasdaqN-1	-0.101924	0.348581	
NasdaqN-2	-0.233857	-0.371445	
NasdaqN-3	-0.375189	-0.103349	
Petr4N	0.969312	1.632402	0.633895
Petr4N-1	1.098339	1.527214	0.958284
Petr4N-2	1.078422	0.697100	1.099667
Petr4N-3	0.745590	0.601660	0.914065
dPetr4N1			0.223200
VarPetr	-2.030162		

Tabela 5.9: performance dos modelos 3,4 e 5.

	Modelo 3	Modelo 4	Modelo 5
MSE	0.063437	0.000476	0.002571
MAE	0.213203	0.017863	0.042994
Erro min abs	0.001532	0.000036	0.000045
Erro max abs	0.479726	0.057101	0.104689
r	0.959352	0.959062	0.910285

Podemos observar a mesma oportunidade da seção anterior no modelo 3, no qual a rede não conseguiu majorar os valores das saídas, pois não foi treinada nestes valores. Quando foram cortados tais períodos de maior cotação da PETR4, o desempenho das RNAs melhorou de maneira considerável, apresentando menores taxas de erro.

As variáveis com maior importância para as RNAs foi a Petr4N e seus respectivos escalonamentos, sendo que para as redes. No caso do modelo 3, a variável VarPetr obteve o maior peso na rede, porém em simulações percebeu-se que melhores resultados eram obtidos sem esta variável, sendo ela desconsiderada nas simulações seguintes. Ficou decidido que nos modelos futuros apenas a variável a ser prevista será escalonada, pois foi percebido em simulações que desta maneira a rede necessita de uma estrutura menor na camada oculta, reduzindo o tempo de treinamento.

É possível realizar uma comparação entre os modelos 3 e 5, que foram treinados com o mesmo conjunto de dados, porém no modelo 5 foi introduzida a variável de entrada dPetr4N1 e foi observado apenas os 51 primeiros períodos de teste. Observa-se que o modelo 5 conseguiu identificar melhor as tendências de alta e baixa da PETR4, mesmo com a variável dPetr4N1 obtendo uma importância relativa pequena na análise dos pesos no modelo 5.

5.3.3. Previsão da variação do valor futuro da ação Petrobras PN

Devido às oportunidades de majoração da RNA para a previsão do preço de uma ação, e como as redes apresentaram um bom desempenho na identificação das tendências de alta ou baixa, foi decidido utilizar um modelo de RNA para realizar a previsão do valor da variação futura da PETR4.

Modelo 6. O objetivo deste modelo foi realizar a previsão da variação percentual entre o período $(t+1)$ e o período t da ação PETR4. Foram utilizados os mesmos conjuntos de treinamento e teste do modelo 3, porém foi desconsiderado o último período de tempo do conjunto de teste, por não se obter a informação da variação futura, ficando este com 190 períodos. Foram utilizadas as mesmas variáveis de entrada do modelo anterior, com a introdução de dPetr4N2.

Tabela 5.10: variáveis de entrada do modelo 6.

Entrada	Modelo 6
BovespaN	x
DowjonesN	x
NasdaqN	x
Petr4N	x
Petr4N-1	x
Petr4N-2	x
Petr4N-3	x
dPetr4N1	x
dPetr4N2	x

Tabela 5.11: arquitetura do modelo 6.

Característica	Nº de neurônios na primeira camada	Nº de neurônios na segunda camada	Função de ativação
Modelo 6	20	10	sigmoidal

5.3.3.1. Resultados

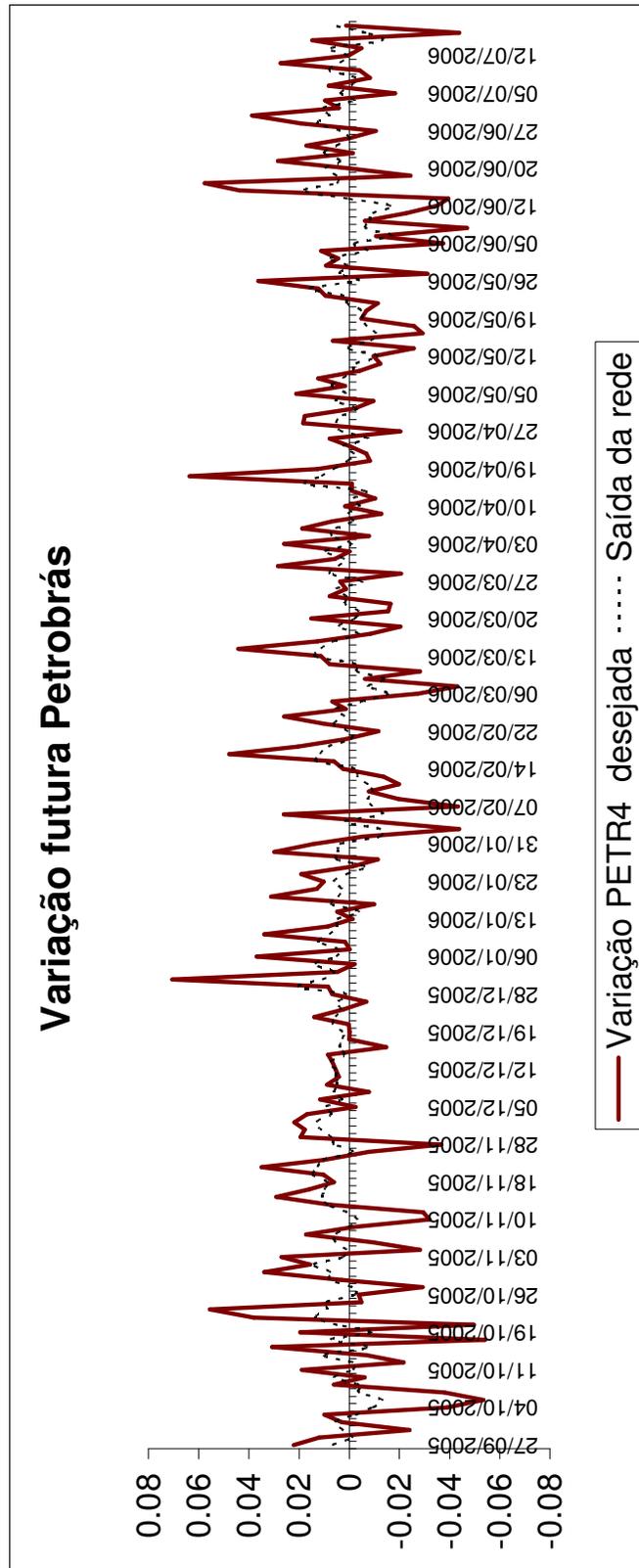


Gráfico 5.7: resultado do Modelo 6.

Tabela 5.12: pesos das entradas utilizadas.

Entrada	Modelo 6
BovespaN	-0.243540
DowjonesN	0.349654
NasdaqN	0.213304
Petr4N	0.871121
Petr4N-1	0.079262
Petr4N-2	-0.691873
Petr4N-3	-0.289726
dPetr4N1	0.262281
dPetr4N2	-0.037527

Tabela 5.13: performance do modelo 6.

	Modelo 6
MSE	0.000400
MAE	0.015407
Erro min abs	0.000068
Erro max abs	0.061710
r	0.435366

Novamente é possível observar que a RNA não conseguiu atingir o mesmo patamar das saídas desejadas. Porém fica claro que - com a introdução das variáveis de diferenças, que passaram a ser de importância no processo de aprendizagem e decisão - a rede consegue prever de maneira satisfatória se o valor da ação da Petrobras PN vai subir ou cair no dia seguinte.

Uma maneira de medir o nível de acerto na tendência de alta ou baixa é utilizar intervalos de confiança. A seguinte metodologia foi adotada: as saídas da rede foram separadas em valores positivos e negativos e foram realizados dois intervalos de confiança para cada conjunto. Logo, no caso do modelo 6, foi observado que com 95% de confiança a rede acertou 61.25% das tendências de alta ou baixa da Petrobras PN. Esta metodologia de intervalo de confiança será utilizada daqui por diante nos modelos de previsão de tendência de ações.

5.4. Resultados e discussões

5.4.1. Previsão da tendência diária futura de ações

Os bons resultados na previsão de tendências de ações utilizando RNAs, proporcionaram uma motivação para desenvolver modelos específicos apenas para a previsão

de se no dia seguinte a ação irá subir ou cair, funcionando como um teste lógico. Foram desenvolvidos cinco modelos para este fim:

Modelo 7. O objetivo deste modelo foi prever se a ação PETR4 irá subir ou cair no período de tempo $(t+1)$, utilizando como valor desejado o teste lógico: 0=cair e 1=subir. Foram utilizados os mesmos conjuntos de treinamento e teste do modelo 6.

Modelo 8. Possui a mesma proposta do modelo anterior, apenas com a mudança da função de ativação para linearsig.

Modelo 9. Muda o teste lógico e a função de ativação em relação ao modelo 8. O teste lógico passa a ser: -1=cair e 1=subir. A função de transferência utilizada foi a lineartanh.

Modelo 10. Idêntico ao modelo 9, mas com a introdução de novas variáveis de entrada: CambioVendaN e SelicN.

Modelo 11. Este modelo teve como objetivo prever a alta ou baixa do Ibovespa no período $(t+1)$. O teste lógico utilizado foi -1=cair e 1=subir, sendo o conjunto de dados o mesmo do modelo 7.

Tabela 5.14: variáveis de entrada dos modelos 7, 8, 9, 10 e 11

Entrada	Modelo 7	Modelo 8	Modelo 9	Modelo 10	Modelo 11
DowjonesN	x	x	x	x	x
BovespaN	x	x	x	x	x
BovespaN-1					x
BovespaN-2					x
BovespaN-3					x
dBovespaN1					x
dBovespaN2					x
NasdaqN	x	x	x	x	x
Petr4N	x	x	x	x	
Petr4N-1	x	x	x	x	
Petr4N-2	x	x	x	x	
Petr4N-3	x	x	x	x	
dPetr4N1	x	x	x	x	
dPetr4N2	x	x	x	x	
CambioVendaN				x	x
SelicN				x	x

Tabela 5.15: arquitetura dos modelos 7, 8, 9, 10 e 11.

Característica	Nº de neurônios na primeira camada	Nº de neurônios na segunda camada	Função de ativação
Modelo 7	20	10	sigmoidal
Modelo 8	20	10	linearsig
Modelo 9	20	10	lineartanh
Modelo 10	16	8	lineartanh
Modelo 11	18	9	lineartanh

5.4.1.1. Resultados

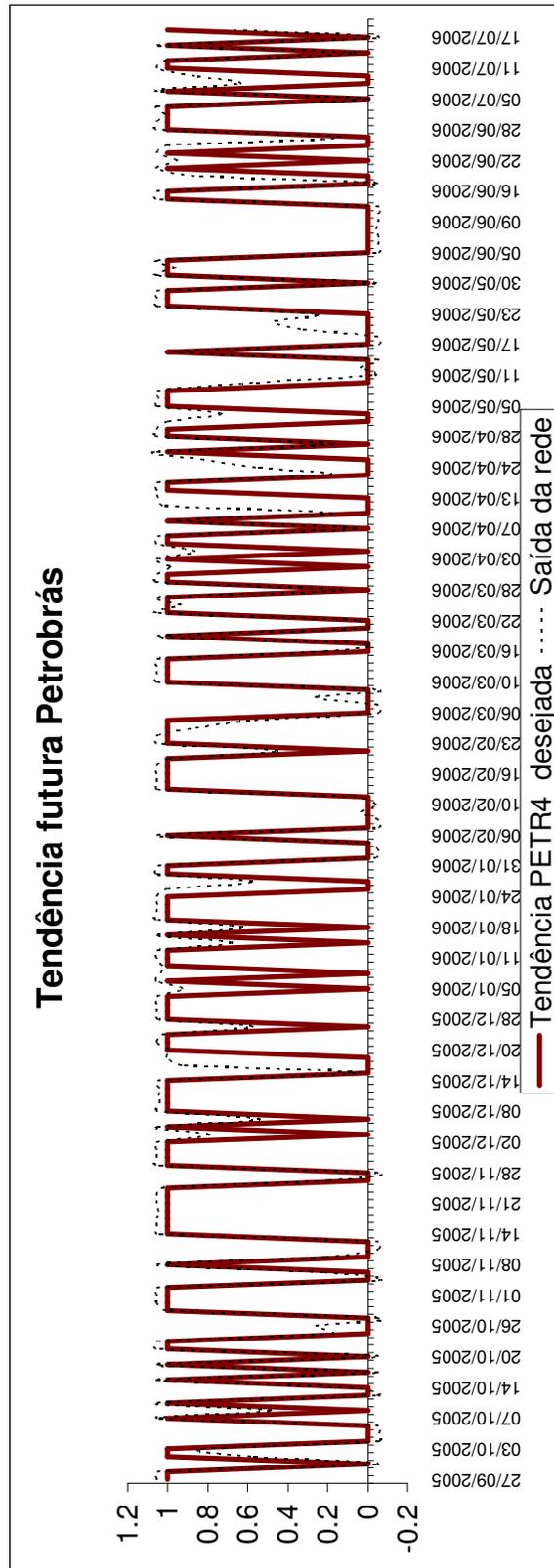


Gráfico 5.8: resultado do Modelo 7.

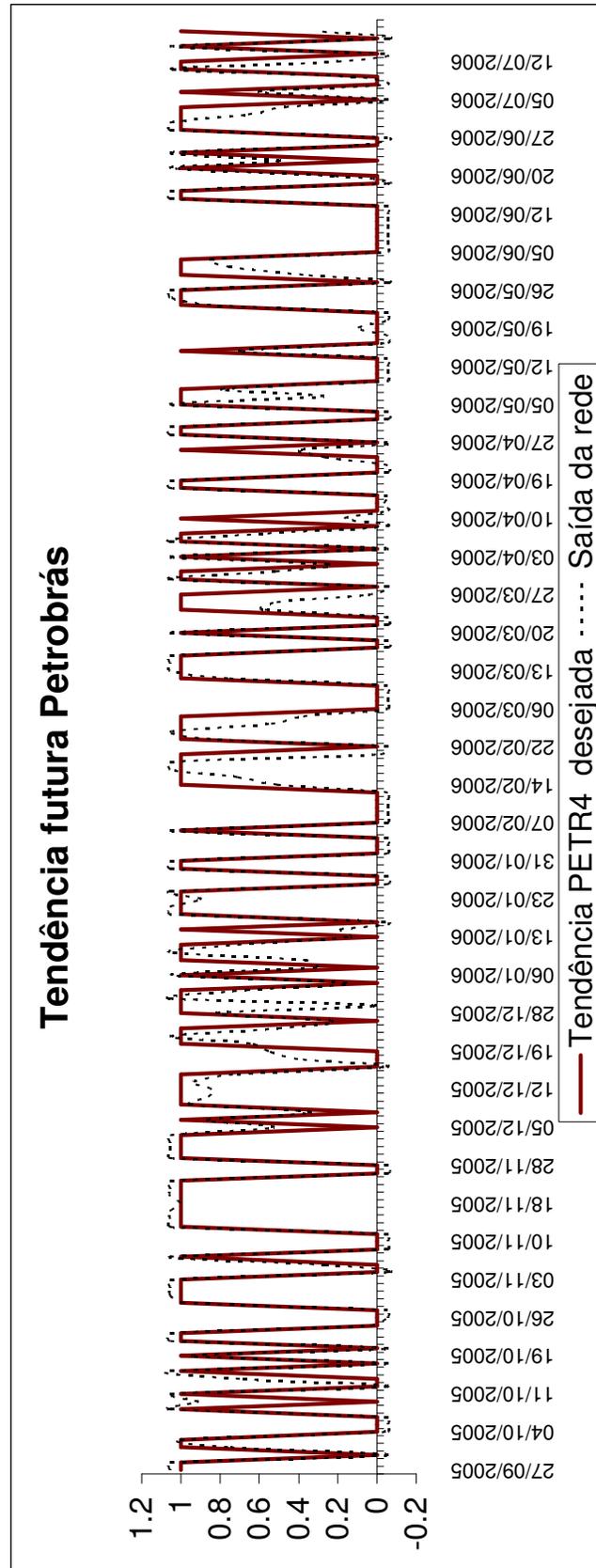


Gráfico 5.9: resultado do Modelo 8.

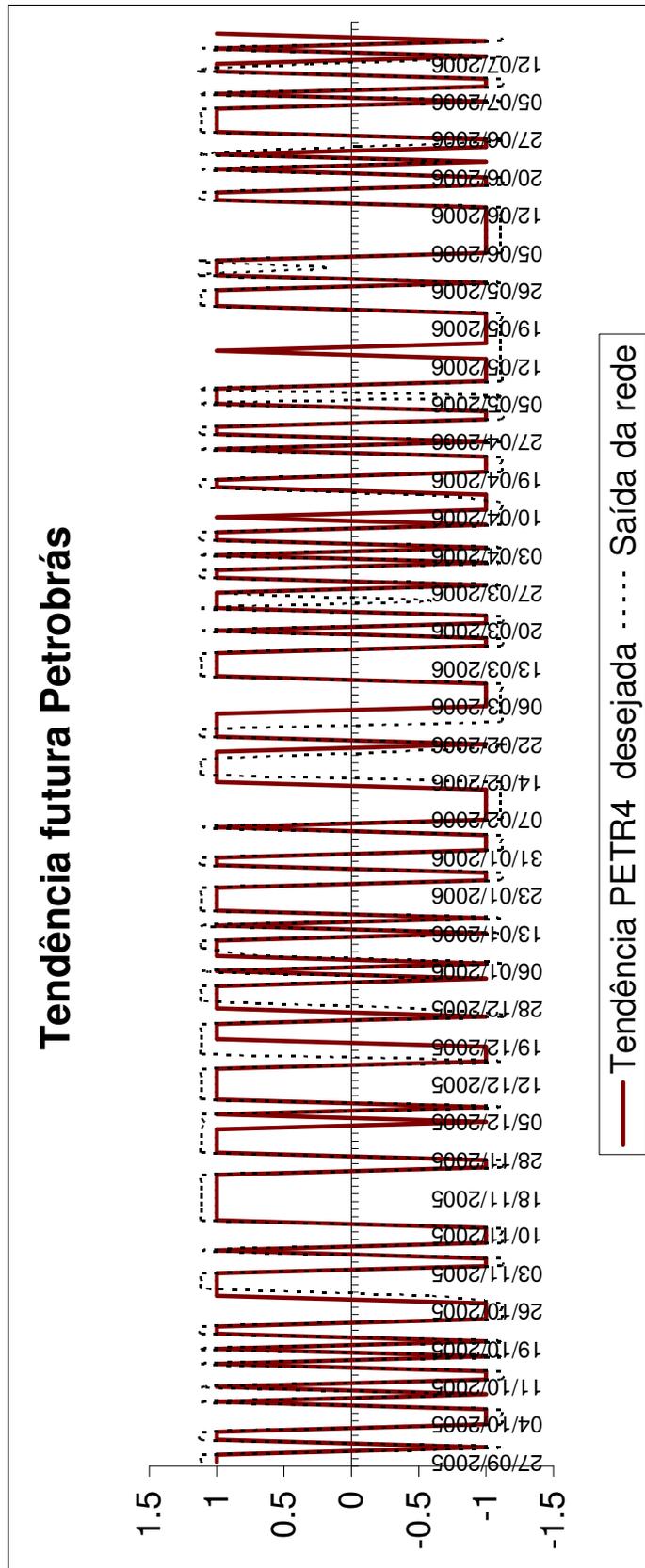


Gráfico 5.10: resultado do Modelo 9.

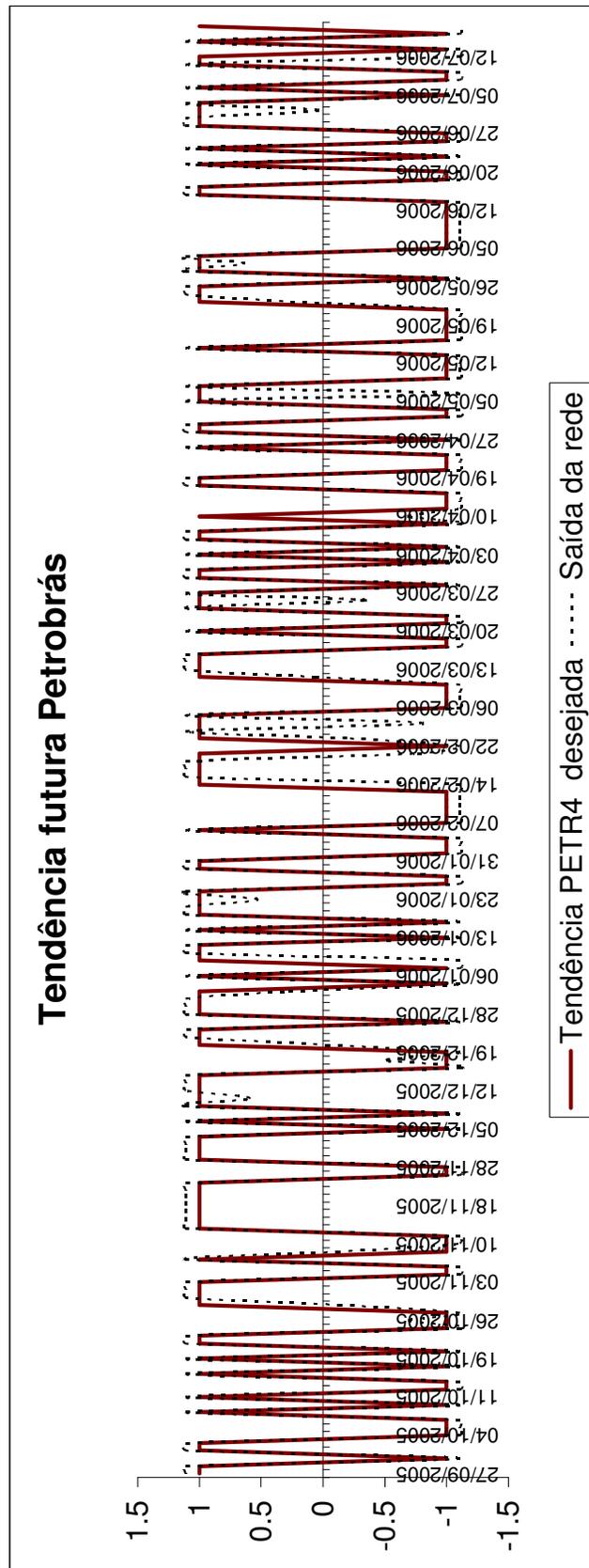


Gráfico 5.11: resultado do Modelo 10.

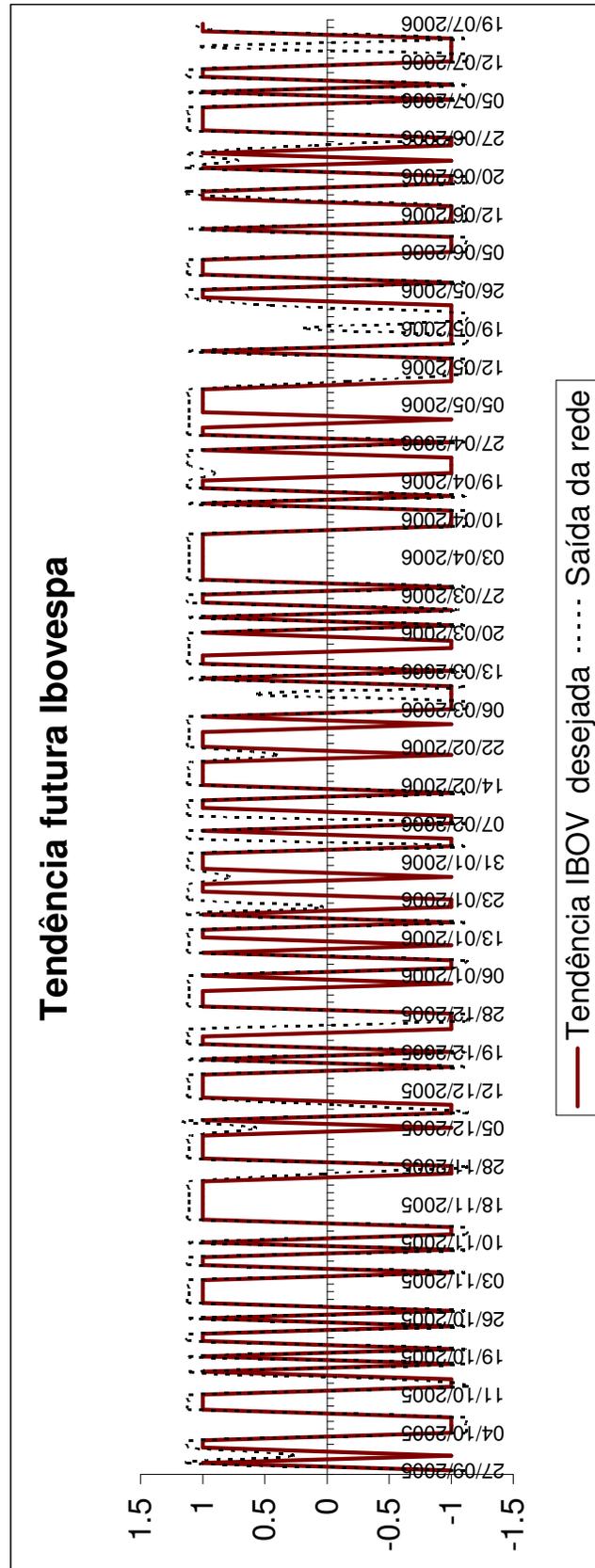


Gráfico 5.12: resultado do Modelo 11.

Tabela 5.16: pesos das entradas utilizadas nos modelos 7, 8, 9, 10 e 11.

Entrada	Modelo 7	Modelo 8	Modelo 9	Modelo 10	Modelo 11
DowjonesN	6.297415	0.513191	1.041090	-0.290596	-2.738052
BovespaN	2.726254	-1.386837	0.351053	3.156574	164.512466
BovespaN-1					-161.749969
BovespaN-2					-1.765962
BovespaN-3					1.054506
dBovespaN1					-1.380684
dBovespaN2					2.026145
NasdaqN	-3.131164	-1.196158	-4.165770	-0.928643	1.226541
Petr4N	61.402111	42.238346	213.649902	217.685654	
Petr4N-1	-45.960632	-40.043224	-212.778107	-218.958359	
Petr4N-2	-15.571846	-6.360564	-7.196335	-6.887551	
Petr4N-3	-2.470294	2.862858	0.691939	2.337707	
dPetr4N1	1.874781	-0.170147	1.845131	-1.164528	
dPetr4N2	0.261655	-0.149753	1.205409	-1.483412	
CambioVendaN				2.832320	-1.756478
SelicN				0.766941	-1.425588

Tabela 5.17: performance dos modelos 7, 8, 9, 10 e 11.

	Modelo 7	Modelo 8	Modelo 9	Modelo 10	Modelo 11
MSE	0.105051	0.072824	0.304722	0.231054	0.453883
MAE	0.176458	0.163133	0.261325	0.239952	0.340589
Min Abs Error	0.001490	0.000791	0.066832	0.105738	0.017898
Max Abs Error	1.047731	1.018421	2.111111	2.111111	2.111111
r	0.818047	0.858002	0.864554	0.901288	0.809441

Tabela 5.18: índice de acertos com 95% de confiança.

	Índice de acertos com 95% de confiança
Modelo 7	72.10%
Modelo 8	67.89%
Modelo 9	91.57%
Modelo 10	91.05%
Modelo 11	90.52%

Os modelos apresentados mostraram um bom desempenho na previsão das tendências de alta ou baixa da PETR4 e do Ibovespa. Pode-se observar claramente o salto de qualidade que as RNAs apresentaram ao utilizar a função de ativação lineartanh e o uso de -1 para indicar queda (modelos 9, 10 e 11), conforme indicado na tabela 5.13.

Os modelos 7, 8 e 9 utilizaram a mesma base de dados, porém, com o uso da função lineartanh no modelo 9, a rede foi capaz de acertar com 95% de confiança 91.57% das tendências, um excelente resultado quando aplicado de maneira prática no mercado financeiro.

No caso do modelo 10, é possível analisar que a adição das variáveis CambioVendaN e SelicN não acarretou em um aumento de desempenho e sim em um pequeno decréscimo. Também não se mostrou importante para a rede ao considerarmos seus pesos. Realizando uma análise dos pesos dos modelos discutidos, as variáveis a serem previstas e seus escalonamentos foram as que apresentaram maior importância.

5.4.2. Previsão da tendência de ações em dois dias futuros

Visando obter resultados mais complexos e testar o desempenho de RNAs em previsão de mais de uma variável, foram desenvolvidos modelos para a previsão da tendência de alta ou baixa de um e dois futuros. Quatro modelos foram projetados:

Modelo 12. O objetivo deste modelo foi prever se a ação PETR4 irá subir ou cair nos períodos de tempo $(t+1)$ e $(t+2)$, utilizando como valor desejado o teste lógico: -1=cair e 1=subir. Foram utilizados os mesmos conjuntos de dados do modelo 6, porém com a exclusão do último período de teste, por não se obter a informação do período $(t+2)$, ficando este com 189 períodos.

Modelo 13. Idêntico ao modelo anterior, apenas com a adição das variáveis dBovespaN1 e dBovespaN2.

Modelo 14. Similar ao modelo 12, porém com objetivo de prever a tendência do Ibovespa.

Modelo 15. Idêntico ao modelo anterior, com exclusão de algumas variáveis de entrada.

Tabela 5.19: variáveis de entrada dos modelos 12, 13, 14 e 15.

Entrada	Modelo 12	Modelo 13	Modelo 14	Modelo 15
DowjonesN	x	x	x	
BovespaN	x	x	x	x
BovespaN-1			x	x
BovespaN-2			x	x
BovespaN-3			x	x
dBovespaN1		x	x	x
dBovespaN2		x	x	x
NasdaqN	x	x	x	
Petr4N	x	x		
Petr4N-1	x	x		
Petr4N-2	x	x		
Petr4N-3	x	x		
dPetr4N1	x	x		
dPetr4N2	x	x		
CambioVendaN			x	
SelicN			x	

Tabela 5.20: arquitetura dos modelos 12, 13, 14 e 15.

Característica	Nº de neurônios na primeira camada	Nº de neurônios na segunda camada	Função de ativação
Modelo 12	18	9	lineartanh
Modelo 13	15	7	lineartanh
Modelo 14	16	8	lineartanh
Modelo 15	25	12	lineartanh

5.4.2.1 Resultados

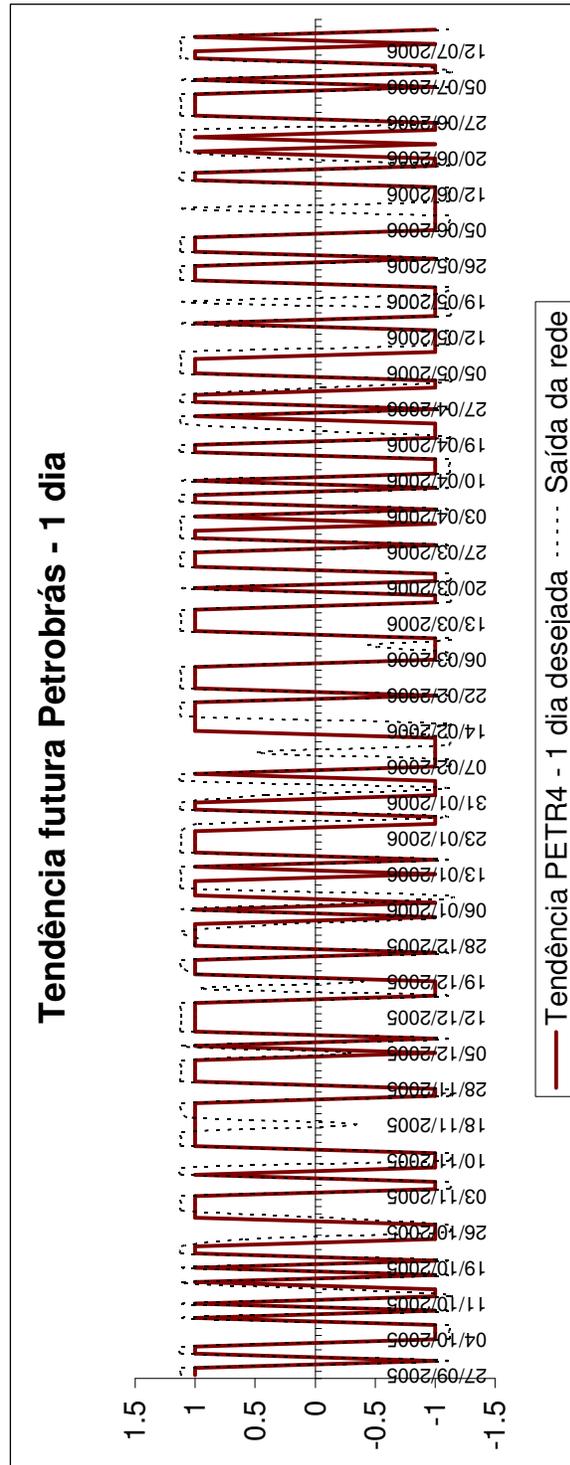


Gráfico 5.13: resultado do Modelo 12 para 1 dia.

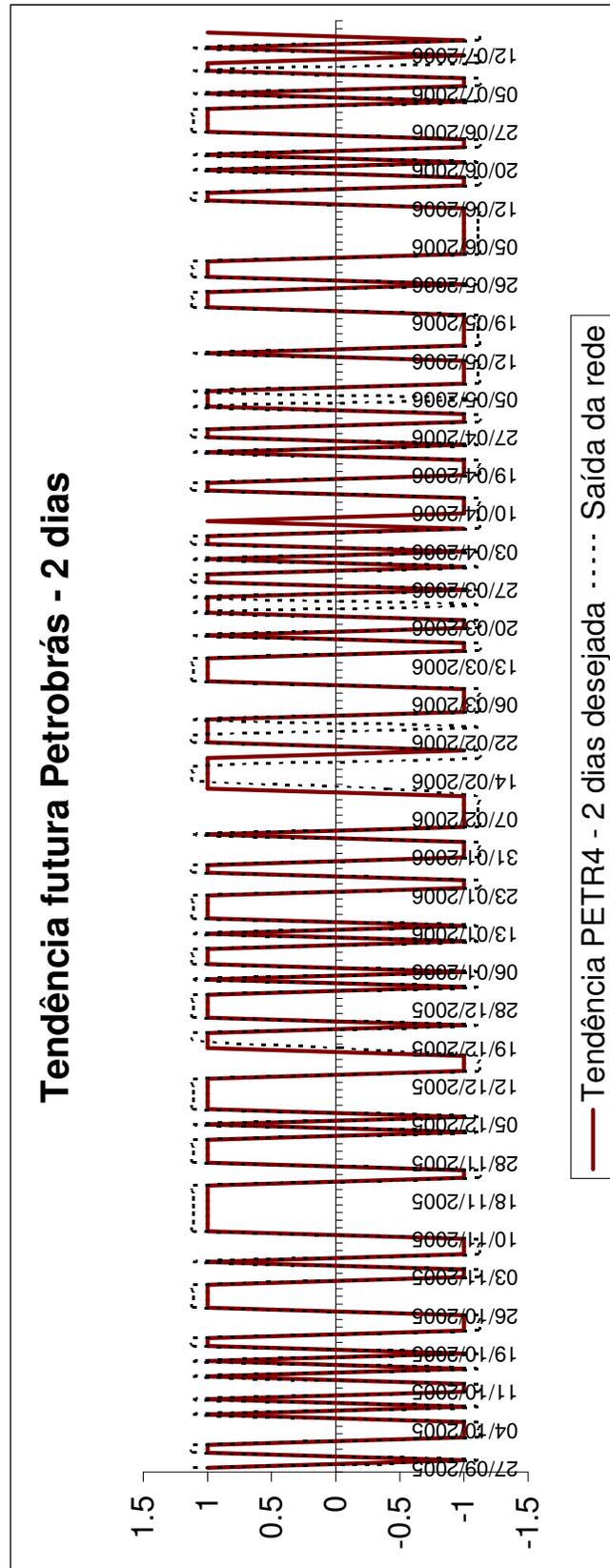


Gráfico 5.14: resultado do Modelo 12 para 2 dias.

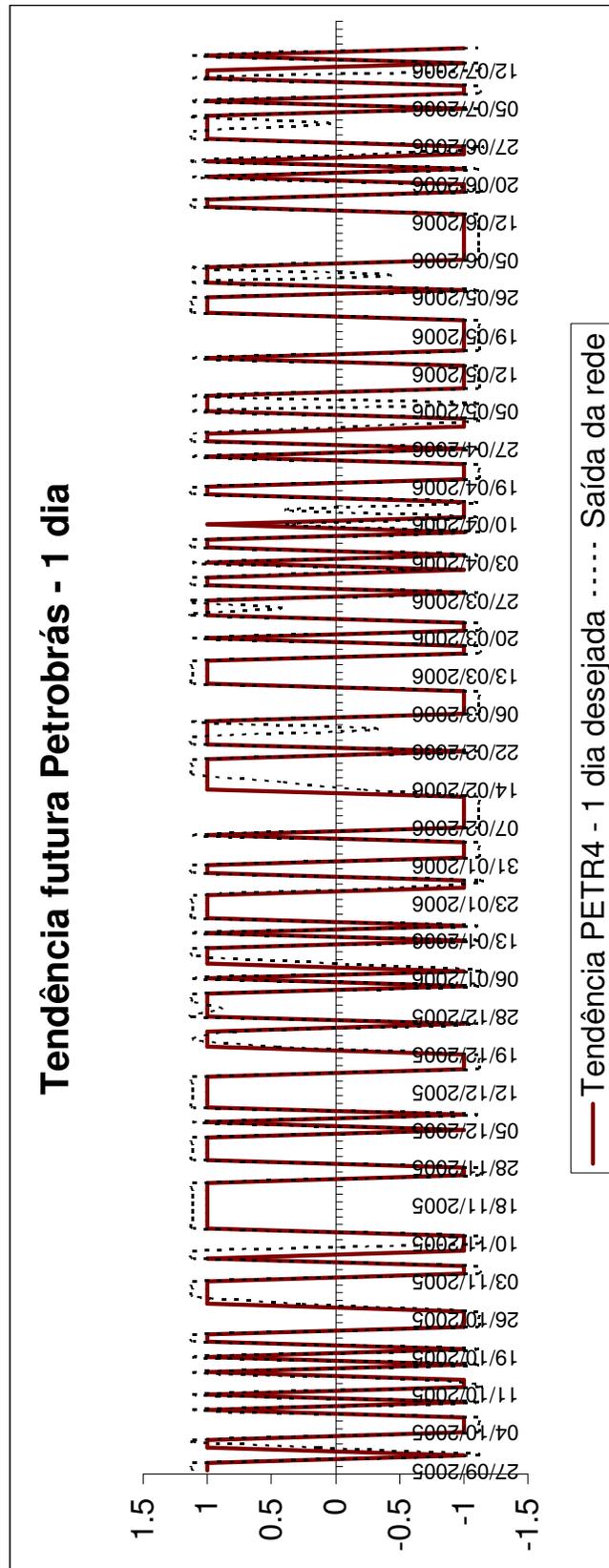


Gráfico 5.15: resultado do Modelo 13 para 1 dia.

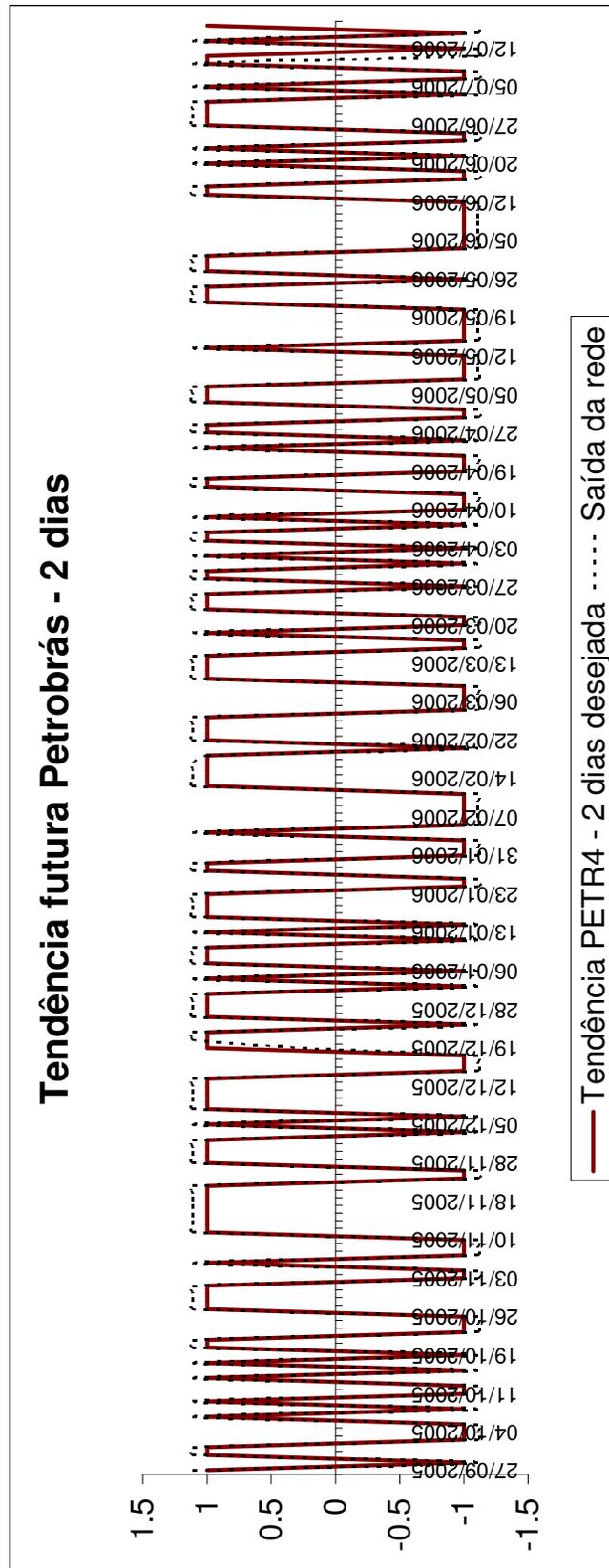


Gráfico 5.16: resultado do Modelo 13 para 2 dias.

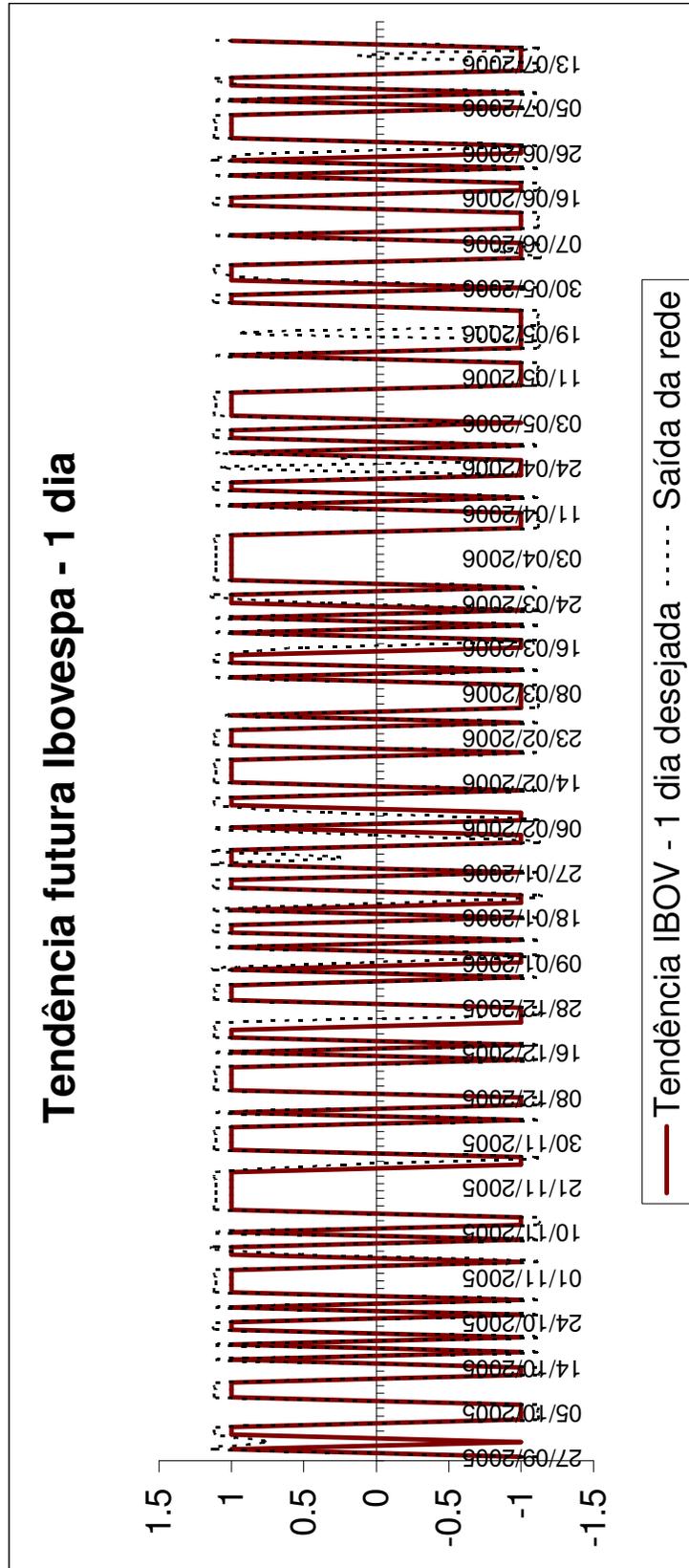


Gráfico 5.17: resultado do Modelo 14 para 1 dia

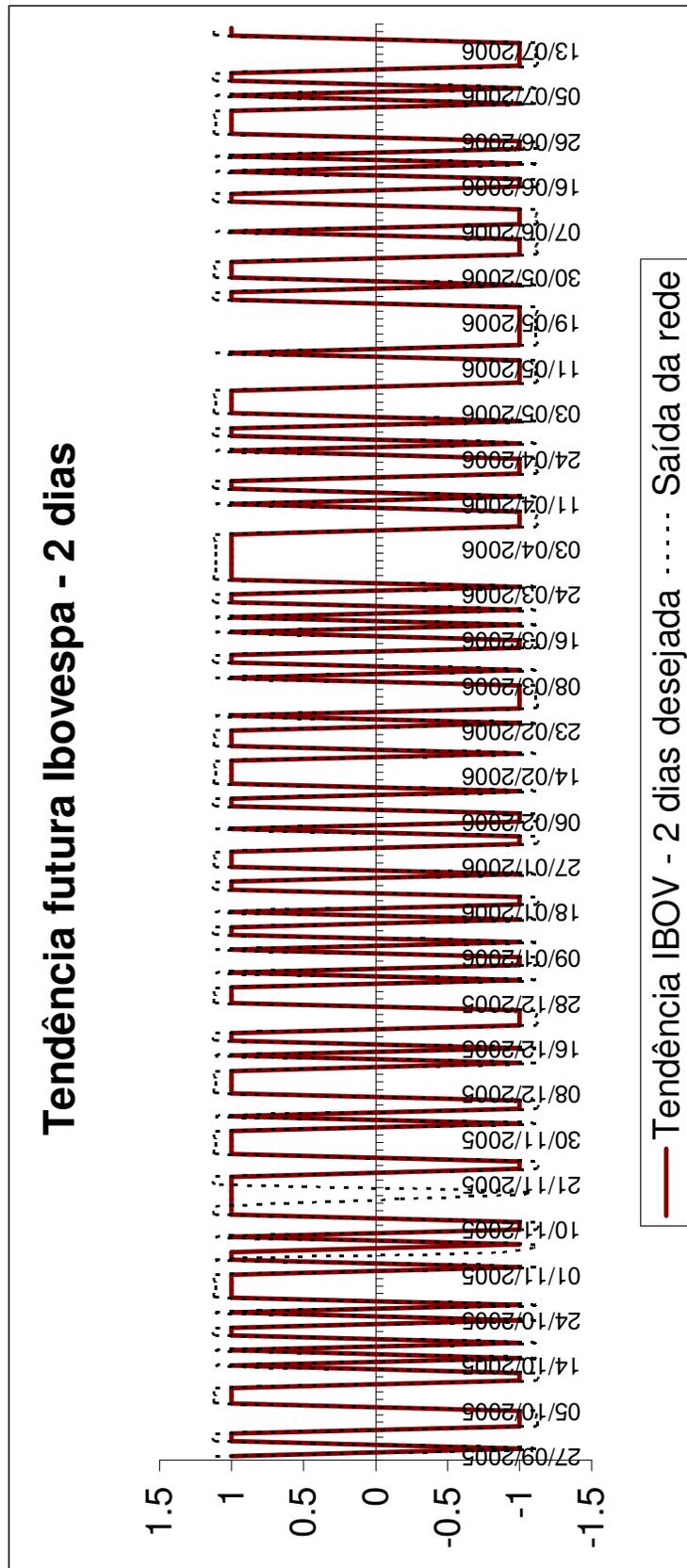


Gráfico 5.18: resultado do Modelo 14 para 2 dias

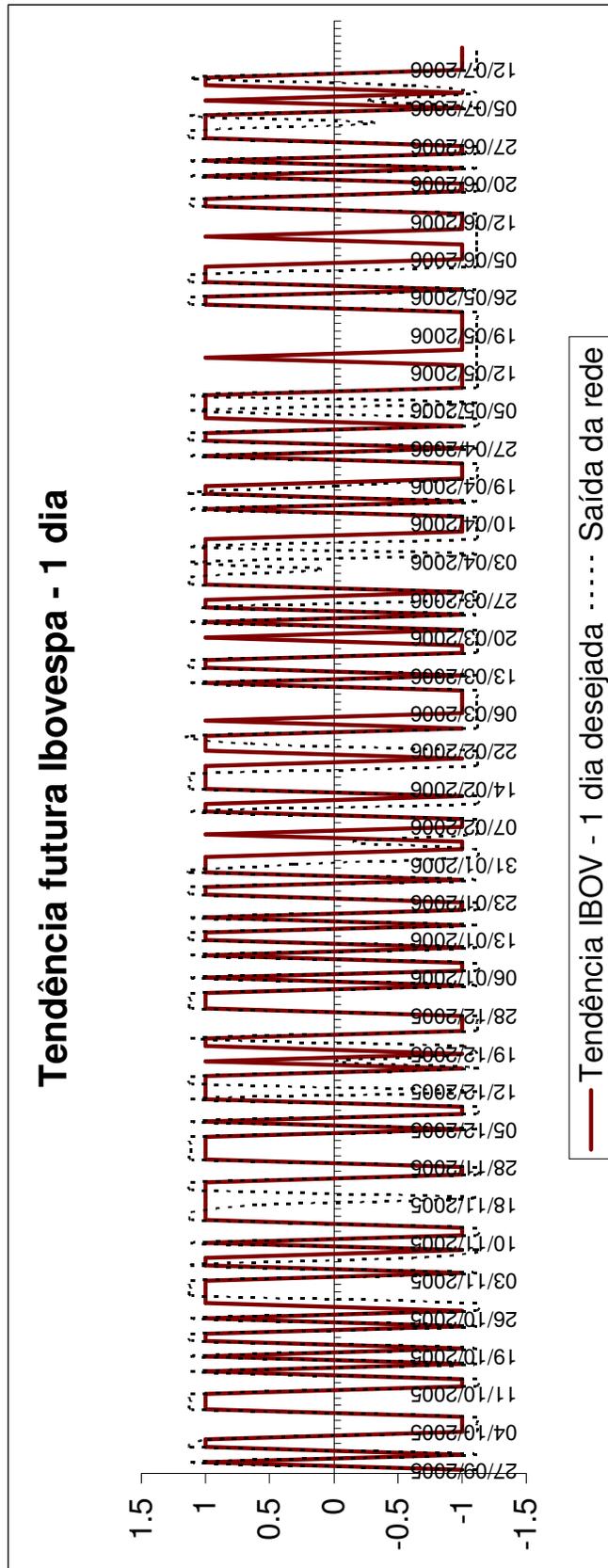


Gráfico 5.19: resultado do Modelo 15 para 1 dia

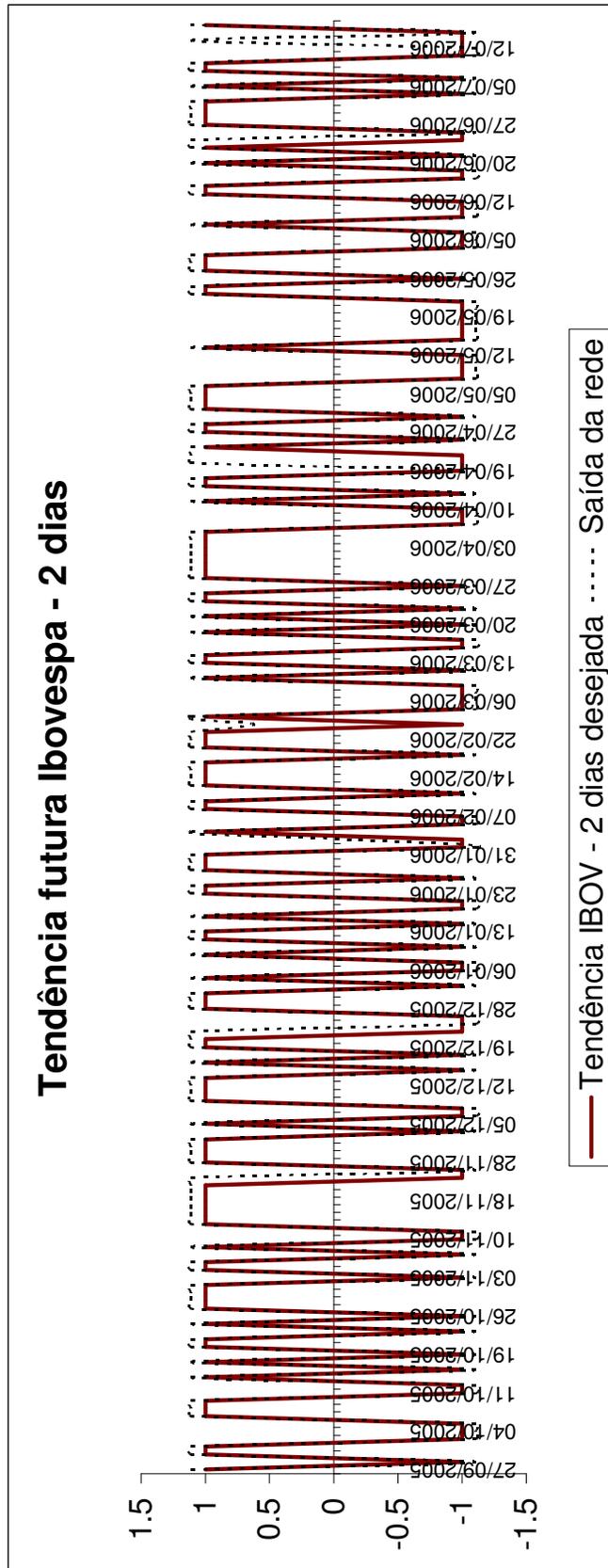


Gráfico 5.20: resultado do Modelo 15 para 2 dias

Tabela 5.21: pesos das entradas utilizadas nos modelos 12 e 13.

	Modelo 12		Modelo 13	
	Petr4 - 1 dia	Petr4 - 2 dias	Petr4 - 1 dia	Petr4 - 2 dias
DowjonesN	-0.182517	0.229334	-2.484539	1.588775
BovespaN	-0.390266	0.153714	2.747420	0.917480
BovespaN-1				
BovespaN-2				
BovespaN-3				
dBovespaN1			0.853354	3.085698
dBovespaN2			-1.589178	-1.354521
NasdaqN	0.030509	0.993381	0.747921	-0.780914
Petr4N	-1.007616	4.124803	217.136246	-2.655094
Petr4N-1	0.457615	2.524864	-215.034439	1.443604
Petr4N-2	0.764492	0.159299	-3.462045	-6.086920
Petr4N-3	-0.257531	0.176337	-2.363471	6.622989
dPetr4N1	1.221988	-3.711599	-1.386933	128.626480
dPetr4N2	0.446777	-0.387761	0.073140	0.114460
CambioVendaN				
SelicN				

Tabela 5.22: pesos das entradas utilizadas nos modelos 14 e 15.

	Modelo 14		Modelo 15	
	Ibovespa - 1 dia	Ibovespa - 2 dias	Ibovespa - 1 dia	Ibovespa - 2 dias
DowjonesN	0.522344	-0.306408		
BovespaN	167.199478	-11.308463	105.340248	-1.573831
BovespaN-1	-167.033829	8.143661	-105.899330	0.033405
BovespaN-2	-2.276578	-1.896884	-4.681610	3.961773
BovespaN-3	1.305504	4.199389	-0.032573	-1.335096
dBovespaN1	2.110259	105.478172	4.131535	87.229462
dBovespaN2	-0.488181	2.285720	1.314967	3.023567
NasdaqN	1.293449	3.486175		
Petr4N				
Petr4N-1				
Petr4N-2				
Petr4N-3				
dPetr4N1				
dPetr4N2				
CambioVendaN	-1.057814	-1.244990		
SelicN	-0.217391	-0.738685		

Tabela 5.23: performance dos modelos 12 e 13.

	Modelo 12		Modelo 13	
	Petr4 - 1 dia	Petr4 - 2 dias	Petr4 - 1 dia	Petr4 - 2 dias
MSE	0.457601	0.195290	0.153517	0.063251
MAE	0.341199	0.198688	0.207935	0.136334
Erro min abs	0.000629	0.111111	0.081354	0.004630
Erro max abs	2.111111	2.111111	2.111111	2.111111
r	0.794043	0.921393	0.929967	0.977316

Tabela 5.24: performance dos modelos 14 e 15.

	Modelo 14		Modelo 15	
	Ibovespa - 1 dia	Ibovespa - 2 dias	Ibovespa - 1 dia	Ibovespa - 2 dias
MSE	0.167856	0.061129	0.521018	0.172927
MAE	0.208518	0.137334	0.361029	0.188025
Erro min abs	0.042824	0.111111	0.013490	0.111111
Erro max abs	2.111111	2.057636	2.111111	2.111111
r	0.926436	0.978124	0.790624	0.929818

Tabela 5.25: acertos com 95% de confiança.

	Índice de acertos com 95% de confiança	
	1 dia	2 dias
Modelo 12	88.42%	98.42%
Modelo 13	88.89%	100.00%
Modelo 14	89.94%	98.41%
Modelo 15	92.59%	95.76%

Novamente a rede conseguiu obter um excelente índice de acerto das tendências de alta ou baixa da PETR4 e Ibovespa, chegando a 100% no caso da tendência de dois dias para o modelo 13. É interessante observar que as RNAs apresentaram um desempenho superior para a previsão de dois dias, sendo o desempenho para a previsão de um dia inferior aos modelos apresentados na seção anterior. Outro resultado importante é que a performance na previsão para dois dias aumentou ao serem inseridas mais variáveis nos modelos, mesmo estas apresentando pouca importância relativa, em oposição aos resultados encontrados na seção anterior.

As variáveis a serem previstas foram as que apresentaram maior importância para a análise de um dia, enquanto que as variáveis de diferença se mostraram mais significativas para a análise de dois dias.

6. CONCLUSÕES

6.1. Comentários iniciais

Com o aumento dos recursos computacionais e da facilidade com que novas informações são disponibilizadas, cresceu o interesse sobre métodos de previsão de ativos futuros em bolsa de valores que apresentem vantagens competitivas para os investidores.

Esta dissertação obteve resultados que demonstraram ser possível atingir um elevado nível de acerto das tendências no mercado de ações – acima de 90% para a previsão de um dia e acima de 98% para a previsão de dois dias, ambas com 95% de confiança – utilizando-se para este fim redes neurais artificiais. Entre os benefícios encontrados em relação às análises tradicionais, podemos citar:

- Uso de dados disponíveis facilmente e a necessidade de processamentos simples destes dados.
- A velocidade com que são extraídos os resultados, principalmente após o treinamento das redes.
- Após um estudo inicial, o uso de redes neurais se torna mais simples e prático do que realizar as análises técnicas e fundamentalistas, além de outros métodos estatísticos.

Contudo, não é uma tarefa simples analisar como cada variável envolvida no processo influenciaram de maneira positiva ou negativa os resultados. Acaba-se ficando “refém” da tediosa experiência de tentativa e erro.

Apesar de continuarem cercadas de incertezas, diversos autores defendem o uso de redes neurais e suas vantagens. E mesmo sendo “caixas-pretas”, existem problemas particulares onde os métodos tradicionais não podem ser aplicados e RNAs podem se apresentar como uma interessante alternativa. Para Refenes (1994), redes neurais são muito mais que apenas uma “moda passageira”, afirmando que “existe muito a ganhar tanto para as ciências computacionais como para as estatísticas ao deixarem de tratar problemas de previsão, classificação e reconhecimento de padrões como áreas exclusivas de conhecimento destes campos”.

6.2. Conclusões

Observando os modelos estudados, é possível concluir que redes neurais são ferramentas que podem ser utilizadas pelo investidor na decisão de compra e venda de ações.

As RNAs apresentaram uma oportunidade de não conseguir resultados satisfatórios para padrões em que não foram apresentadas no treinamento. Isto ocorreu para o caso de previsões de valores, onde as RNAs apresentaram oportunidades de majoração de valores que não lhes foram apresentados em seu treinamento, não conseguindo atingir o mesmo patamar das cotações reais. Porém para o caso de previsão das tendências como um teste lógico (-1 significando baixa e +1 significando alta), onde não se encontrava a mesma oportunidade, os resultados encontrados foram excelentes. Ao encontrar a correta arquitetura – duas camadas ocultas, função de ativação tangente hiperbólica linear – obteve-se índices de acertos superiores a 90% com uma confiança de 95% para a previsão da tendência de um dia futuro. No caso da previsão da tendência de dois dias futuros, apesar do rendimento da previsão no primeiro dia ter sido reduzida para cerca de 88%, a taxa de acerto foi superior a 98%, chegando a 100% no caso do modelo 13.

O histórico passado das variáveis analisadas se mostraram como de fundamental importância para as redes na previsão das tendências. Outras variáveis importantes observadas foram aquelas que representam as diferenças entre períodos passados da entrada a ser prevista. A única referência encontrada na literatura sobre a utilização de variáveis de diferença como entradas em RNAs é o uso bem sucedido para auxiliar no processo de tomada de decisão de agentes artificiais em ambientes oligopolísticos (Baldassare, 1996). Porém para a previsão de tendências, a inclusão destas entradas proporcionou uma melhora de desempenho das RNAs, conforme é possível observar realizando uma comparação entre os modelos 3 e 5.

Outro importante aspecto notado é a de que certas entradas apresentam um maior peso relativo, porém sua exclusão do modelo pode levar a melhores resultados, como nos casos da variável BrentN nos modelos 1 e 2 e da variável VarPetr no modelo 3. Logo, é possível criticar o uso de pesos de uma rede como referência de importância de uma dada variável.

Observou-se que o uso de duas camadas ocultas foi suficiente para resolver todos os modelos apresentados, sendo que um aumento nesta estrutura não melhorou o desempenho da rede e apenas levou a um maior desperdício de tempo de treinamento. Para o uso de um maior número de variáveis de entrada, recomenda-se um menor número de componentes nas

camadas ocultas, e para o uso de maiores conjuntos de dados um número maior de neurônios. Porém, esta análise pode não ser válida para problemas de diferentes naturezas.

Para o caso da previsão da tendência de um dia futuro, o rendimento foi inferior quando o número de entradas aumentou, enquanto que para a previsão de dois dias futuros, as redes apresentaram rendimento superior com um número maior de entradas. Isso pode ser facilmente observado realizando uma comparação entre os resultados dos modelos 14 e 15, os quais tiveram como distinção apenas o número de entradas. No modelo 15 foram retiradas as entradas SelicN e CambioVendaN, e foi verificado que a taxa de acertos para o primeiro dia aumentou, enquanto que a taxa de acertos para dois dias foi reduzida.

6.3. Sugestões para trabalhos futuros

Esta seção tem como objetivo propor sugestões para trabalhos futuros que possam aprimorar e dar continuidade ao estudo realizado neste trabalho. São elas:

- Dar continuidade à pesquisa reaplicando o modelo, para a previsão de ações de outras empresas e outros índices de mercados de ações, de preferência testando sua validade em ativos que apresentem maior variabilidade e risco do que a Petrobras.
- Realizar previsões de tendências no longo prazo, procurando medir o movimento de um determinado ativo em uma semana e/ou mês.
- Realizar análises estatísticas mais profundas sobre os resultados apresentados, procurando outras maneiras de validar o acerto ou erro da rede.
- Realizar um estudo sobre a metodologia utilizada e procurar encontrar estratégias de obter melhores resultados, variando o número de camadas ocultas e o número de elementos, os parâmetros do algoritmo de treinamento, etc.
- Realizar estudos com aprendizados não-supervisionados e redes recorrentes para a previsão de valores de tendências de ações.
- Realizar uma comparação entre o modelo e as análises técnicas e fundamentalistas, além de outros métodos econométricos.

- Incrementar o modelo de redes neurais, com o uso de ferramentas suplementares como *wavelets* e lógica *fuzzy*, verificando se ocorre uma melhora de desempenho.
- Realizar simulações de ganhos no mercado de ações.

REFERÊNCIAS BIBLIOGRÁFICAS

AMBROSIO, P. E. (2002) – Redes neurais artificiais no apoio ao diagnóstico diferencial de lesões intersticiais pulmonares. (Mestrado – Universidade de São Paulo).

ARAÚJO JR, A.L. & VASCONCELOS G.C (1999). Previsão de demanda máxima mensal baseada em redes neurais. *XV Seminário Nacional de Produção e Transmissão de Energia Elétrica (SNPTEE'99)*. Foz do Iguaçu - PR, Brasil.

BOGULLU, V.K.; ENKE, D. & DAGLI, C. (2002). Using Neural Networks and Technical Indicators for Generating Stock Trading Signals. *Proceedings of the 2002 Artificial Neural Networks in Engineering Conference (ANNIE '02)*, ASME, pp. 721-726, St. Louis, Missouri.

BOVESPA – Bolsa de valores de São Paulo. <http://www.bovespa.com.br>. Último acesso em 23/10/2006.

BRAGA, A. P.; CARVALHO, A.C. de & LUDERMIR, T. B. (2000). Redes Neurais Artificiais: Teoria e aplicações. *Livros Técnicos e Científicos..*

BRESSAN, A. A.; BRESSAN, V. G. F. & LAMOUNIER, W. M. (2005). Estratégias de Investimento em Futuros de Ibovespa utilizando Redes Neurais Artificiais. *In: II Encontro Norte-Nordeste de Finanças, 2005, Recife-PE*. Anais do II Encontro Norte-Nordeste de Finanças, v. Único. p. 1-16.

BRUNI, A. L. & FAMÁ, R. (1998). Eficiência, Previsibilidade dos Preços e Anomalias em Mercados de Capitais: Teoria e Evidências. *Caderno de Pesquisas em Administração, São Paulo*, v.1, n.7.

BÓDIS, L. (2004). Financial time series forecasting using artificial neural networks. (Mestrado – Babeş-Bolyai University).

BUENO, A. E. (2000). Análise empírica do “Dividend Yield” das ações brasileiras.

(Mestrado – Universidade de São Paulo).

CHO, V. (2003). A comparison of three different approaches to tourist arrival forecasting. *Tourism Management*, 24, 323–330.

COMRIE, A. C. (1997). Comparing neural network and regression models for ozone forecasting. *Journal of the Air and Waste Management Association*, 47:653 – 663.

CONNOR, J. T.; MARTIN, R. D. & ATLAS, L.E. (1994). Recurrent neural networks and robust time series prediction. *IEEE Transaction on Neural Networks* 51 (2), 240–254.

CYBENKO, G. (1988). Continuous Valued Neural Networks with Two Hidden Layers are Sufficient. *Technical Report*, Tuft University.

FADLALLA, A. & LIN, C.-H. (2001). An analysis of the applications of neural networks in finance. *Interfaces*, 31(4), 112-122.

FAMA, E. F. (1970). Efficient capital markets: a review of theory and empirical work. *Journal of Finance*, v. 25, n.2, p. 338-417.

FAUSETT, L. (1994). Fundamentals of Neural Networks – Architecture, Algorithms, and Applications. *Prentice Hall International, Inc.*

FREITAS, A. A. C. de & SILVA, I. N. da. (1999). Análise Técnica de Títulos Financeiros Através de Redes Neurais Artificiais. *Proceedings of the IV Brazilian Conference on Neural Networks - IV Congresso Brasileiro de Redes Neurais* pp. 067-071, ITA, São José dos Campos - SP - Brazil

GOLDBERG, D. E. (1953). Genetic algorithms in search optimization and machine learning. *Addison-Wesley, c1989. 412p.*

GENETIC ALGORITHM AND DIRECT SEARCH TOOLBOX. <http://www.mathworks.com>.
Último acesso em 24/05/2006.

HAYKIN, Simon (1994). *Neural Networks: A Comprehensive Foundation*. Macmillan Publishing Company.

HEBB, D. O. (1949). *The Organization of Behavior*. Wiley.

HILL, T.; MARQUEZ, L.; O'CONNOR, M. & REMUS, W. (1994). Artificial neural networks for forecasting and decision making. *International Journal of Forecasting* 10, 5–15.

HOLLAND, J. H. (1975). *Adaptation in Natural and Artificial Systems*. MIT Press, Second edition.

HOPFIELD, J. J. (1982). Neural Networks and physical systems with emergent collective properties. *Proc. Nat. Acad. Sci.*, 79:2554-8.

HORNIK, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks* 4, 251–257.

IPEA – Instituto de Pesquisa Econômica Aplicada. <http://www.ipea.gov.br>. Último acesso em 23/10/2006.

KIM, K. (2003). Financial time series forecasting using support vector machines. *Neurocomputing* 55, 307-319.

KIM, K. & HAN, I. (2000). Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. *Expert Syst. Appl.* 19 (2) 125–132.

KULKARNI, A. S. (1996). *Application of Neural Networks to Stock Market Prediction, Technical Report*.

KUVAYEV, L. (1996). Predicting Financial Markets with Neural Networks, *Review Paper, Seminar in Capital Markets*.

LAPEDES, A. & FARBER, R. (1987). Nonlinear signal processing using neural networks: prediction and system modeling. *Technical Report LA-UR-87-2662*, Los Alamos National Laboratory, Los Alamos, NM.

LAPEDES, A. & FARBER, R. (1988). How neural nets work. In: *Anderson, D.Z., (Ed.), Neural Information Processing Systems*, American Institute of Physics, New York, pp. 442–456.

LAWRENCE, R. (1997). Using Neural Networks to Forecast Stock Market Prices. <http://www.cs.uiowa.edu/~rlawrenc/research/Papers/nn.pdf>. Último acesso em 24 de Outubro de 2006.

LEUNG MT, DAOUK H, CHEN AS. (2000). Forecasting stock indices: a comparison of classification and level estimation models. *International Journal of Forecasting* 16:173–90.

LIPPMANN, R.P. (1987). An introduction to computing with neural nets, *IEEE ASSP Magazine*, April, 4–22.

MAIER, H. R. & DANDY, G. C. (2000). Neural networks for the prediction and forecasting of water resource variables: A review of modeling issues and applications. *Environmental Modeling and Software*, 15, 101-124.

MELLO, M. T. DE (2004). Aplicação de Redes Neurais Artificiais no processo de precificação de ações. (Monografia – Universidade Federal de Pelotas).

MILANEZ, D. Y. (2003). Finanças Comportamentais no Brasil. (Mestrado – Universidade de São Paulo).

MCCULLOCH, W. S. & PITTS, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115-113.

NEURAL NETWORK TOOLBOX FOR USE WITH MATLAB. <http://www.mathworks.com>. Último acesso em 24/05/2006.

NEUROSOLUTIONS – GETTING START MANUAL. *Http://www.nd.com*. Último acesso em 19/05/2006.

ROSENBLATT, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev.*, 65:386-408.

RUMELHART, D. E.; HINTON, G. E. & WILLIAMS, R.J. (1986). Learning internal representation by back- propagating errors. In: Rumelhart, D.E., McClelland, J.L., *the PDP Research Group (Eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. MIT Press, MA.

RUMELHART, D. E.; HINTON, G. E. & WILLIAMS, R. J. (1986). Learning representations by back-propagation errors. *Nature*, 323: 533-536.

SANTOS, R. T.; NIEVOLA, J. C.; FREITAS, A. A. & LOPES, H.S. (1999). Extração de Regras de Redes Neurais via Algoritmos Genéticos. *Anais do IV Congresso Brasileiro de Redes Neurais*, pp. 158-163, ITA, São José dos Campos.

SCHIFFMANN, W. H.; JOOST, M. & WERNER, R. (1994). Optimization of the Backpropagation Algorithm for Training Multilayer Perceptrons. *Technical report*. Institute of Physics, University of Koblenz.

SHANKER, M.; HU, M. Y. & HUNG, M.S. (1996). Effect of data standardization on neural network training. *Omega* 24 (4), 385–397.

SHARDA, R. (1994). Neural networks for the MS/OR analyst: An application bibliography. *Interfaces* 24 (2), 116–130.

SHILLER, R. (1981). Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review*, v.71, n.3, p.421-436.

SRINIVASAN, D.; LIEW, A. C. & CHANG, C.S. (1994). A neural network short-term load forecaster. *Electric Power Systems Research* 28, 227–234.

VELLIDO, A.; LISBOA, P. J. G. & VAUGHAN, J. (1999). Neural networks in business: a survey of applications (1992–1998). *Expert Systems with Applications*, 17 (1), 51–70.

WEIGEND, A.S.; HUBERMAN, B.A. & RUMELHART, D.E. (1990). Predicting the future: A connectionist approach. *International Journal of Neural Systems 1*, 193–209.

WEIGEND, A.S.; HUBERMAN, B.A. & RUMELHART, D.E. (1992). Predicting sunspots and exchange rates with connectionist networks. In: *Casdagli, M., Eubank, S. (Eds.), Nonlinear Modeling and Forecasting*. Addison-Wesley, Redwood City, CA, pp. 395–432.

WIDROW, B.; RUMELHART, D.E. & LEHR, M.A. (1994). Neural networks: Applications in industry, business and science. *Communications of the ACM 37* (3), 93–105.

WONG, B. K. & SELVI, Y. (1998). Neural network applications in finance: A review and analysis of literature (1990-1996). *Information and Management*, 34, 129-139.

Y. YOON & G. SWALES. (1993). Predicting stock price performance: A neural network approach. In *Neural Networks in Finance and Investing*, chapter 19, pages 329–342. Probus Publishing Company, 1993.

ZEKIĆ, M. (1998). Neural Network Applications in Stock Market Predictions . A Methodology Analysis. In *B. Aurer, R. Logo.ar, Vara.din (Eds.), Proceedings of the 9th International Conference on Information and Intelligent Systems*, pp. 255-263.

ZHANG, X. (1994). Time series analysis and prediction by neural networks. *Optimization Methods and Software 4*, 151–170.

ZHANG, G.; PATUWO, B. E. & HU, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14 (1), 35–62.

GLOSSÁRIO

- **Algoritmos genéticos:** são rotinas computacionais que procuram emular o processo de evolução biológica para resolver problemas de otimização.
- **Algoritmo de treinamento:** também pode ser chamado de algoritmo de aprendizado. Ele decide como os pesos das conexões serão alterados, que conexões serão criadas e que conexões serão destruídas, alterando a topologia da rede e fazendo com que a rede seja capaz de responder a diferentes situações do ambiente em que está inserida.
- **Arquitetura de uma RNA:** é a definição de como será a estrutura da RNA. Aqui são definidos: o número de camadas da rede, número de neurônios em cada camada e o tipo de conexão entre os neurônios da rede.
- **Backpropagation:** algoritmo de treinamento mais utilizado em RNAs. Consiste em realizar a retro-propagação (realimentação) do erro gerado na diferença entre a saída da rede e o valor desejado, visando minimizar o erro total da saída gerada pela rede.
- **Camadas de uma RNA:** blocos onde se encontram os neurônios da rede neural. Podem ser de entrada e saída e ainda intermediárias (ou ocultas).
- **Camada oculta:** também pode ser chamada de camada intermediária ou escondida. Realizam o processamento entre a entrada e a saída.
- **Função de ativação:** também pode ser chamada de função de transferência. Restringe a amplitude da saída do neurônio a um valor finito, realizando uma transformação no sinal.
- **Neurônios artificial:** modelo matemático simplificado de um neurônio biológico. Formam a rede neural.
- **Neurosolutions:** *software* para a simulação de RNAs.

- **Momentum:** parâmetro do algoritmo *backpropagation*. Uma taxa que é uma constante que determina o efeito das mudanças anteriores dos pesos.
- **Perceptron:** forma pioneira de rede neural que posteriormente foi incrementada e passou a ser utilizada com o algoritmo *backpropagation*.
- **Pesos:** têm a função de definir a força das conexões entre as variáveis de entrada e as de saída. É através da mudança dos pesos que acontece o aprendizado das redes neurais.
- **Probabilidade de cruzamento:** parâmetro de um algoritmo genético. Define a probabilidade de acontecer um cruzamento entre dois genes.
- **Probabilidade de mutação:** parâmetro de um algoritmo genético. Define a probabilidade de acontecer uma mutação em um gene.
- **Redes Neurais Artificiais (RNAs):** são sistemas de processamento paralelo, que simulam a estrutura do cérebro humano, para calcular certas funções matemáticas lineares ou não.
- **Rede direta ou *feedforward*:** modelo de conexão entre os neurônios de uma RNA. A saída de um neurônio não pode ser utilizada como entrada de uma camada anterior à camada em que está localizado este neurônio.
- **Taxa de aprendizagem:** parâmetro do algoritmo *backpropagation*.
- **Topologia de uma RNA:** o mesmo que arquitetura.
- **Treinamento de uma RNA:** a principal etapa no desenvolvimento de uma RNA. É aqui que a rede aprende sobre o ambiente e se adapta, modificando seus pesos, para resolver o problema. O aprendizado acontece através de um algoritmo de treinamento.

**A GRÁFICOS DAS VARIÁVEIS MACROECONÔMICAS UTILIZADAS –
APÓS TRANSFORMAÇÃO LINEAR**

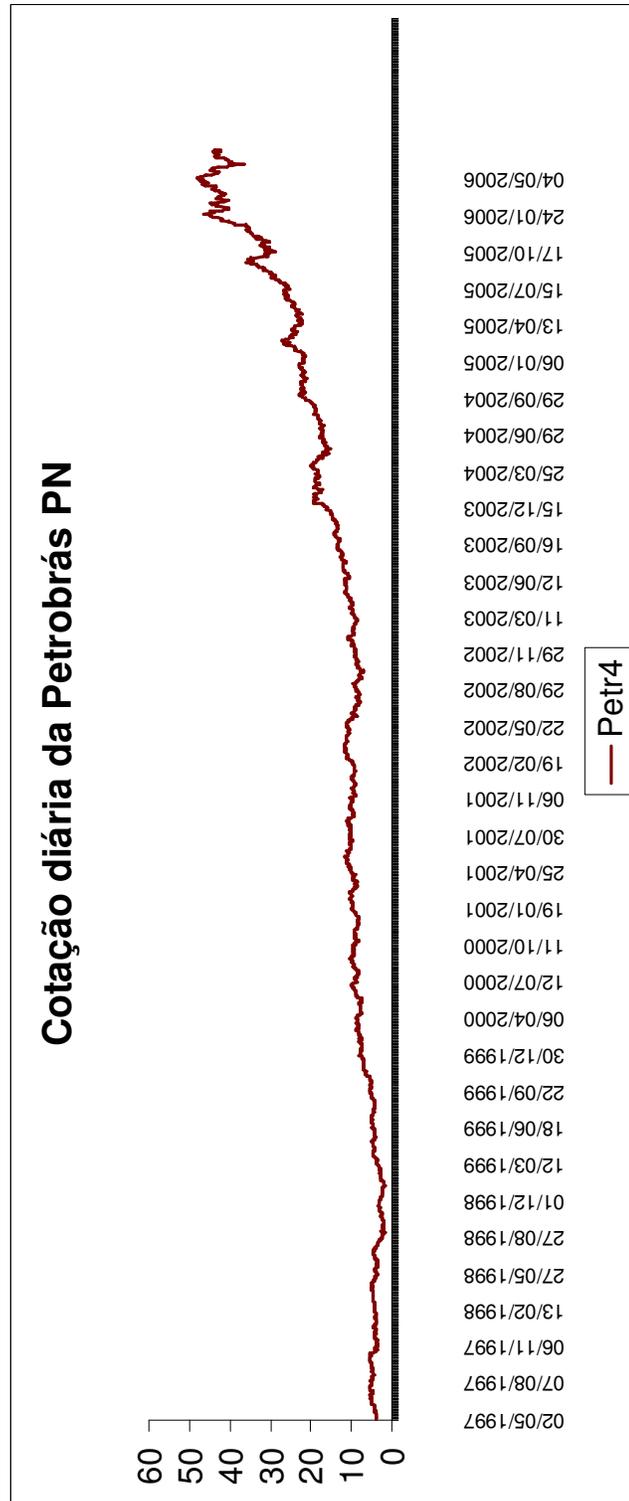


Gráfico A.1: Cotação diária da Petrobras PN.

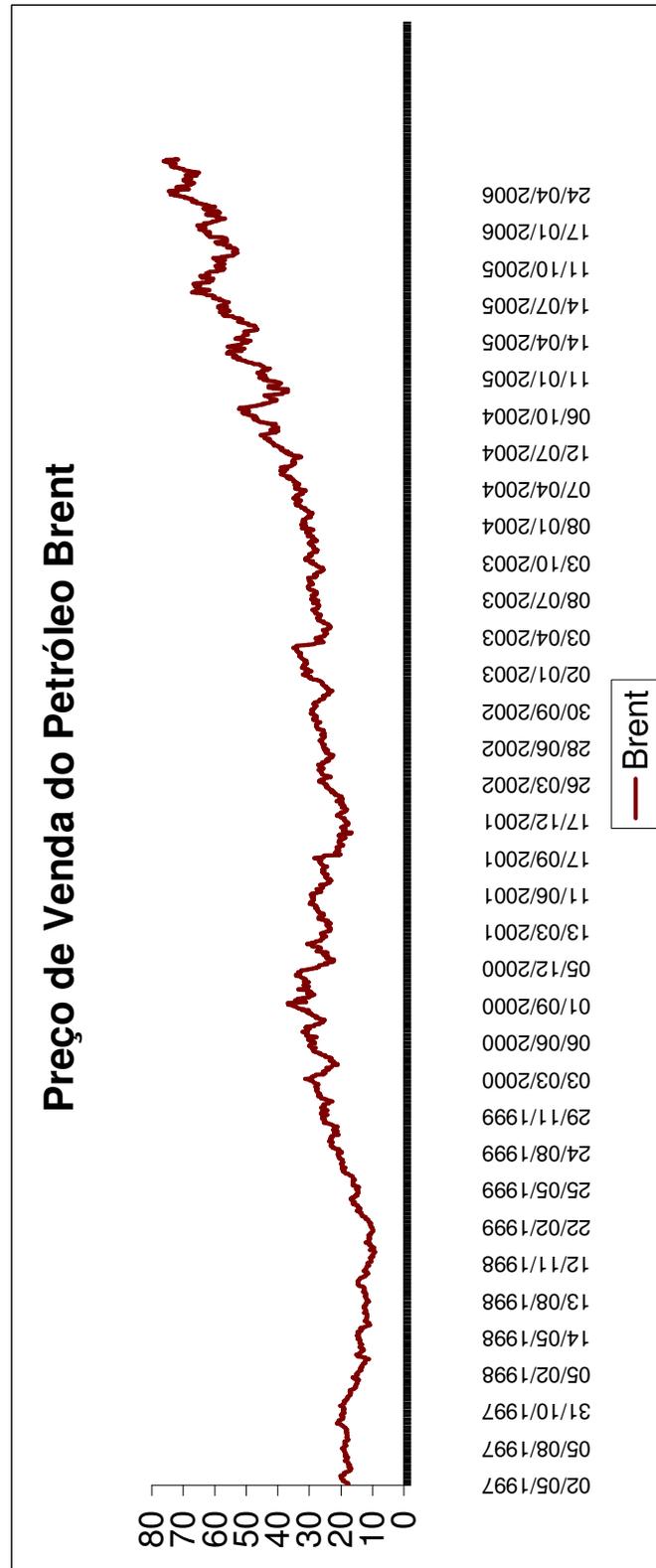


Gráfico A.2: Preço de venda do petróleo Brent.

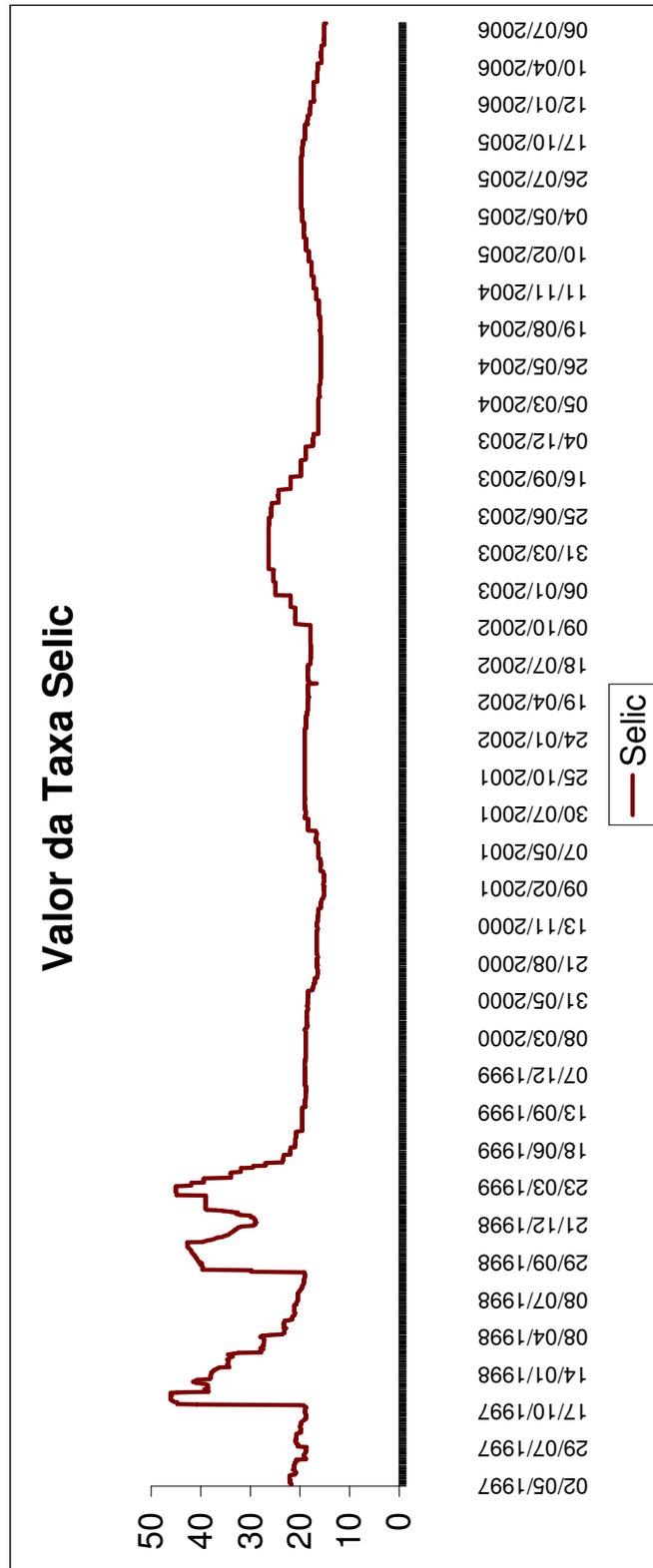


Gráfico A.3: valor da taxa Selic.

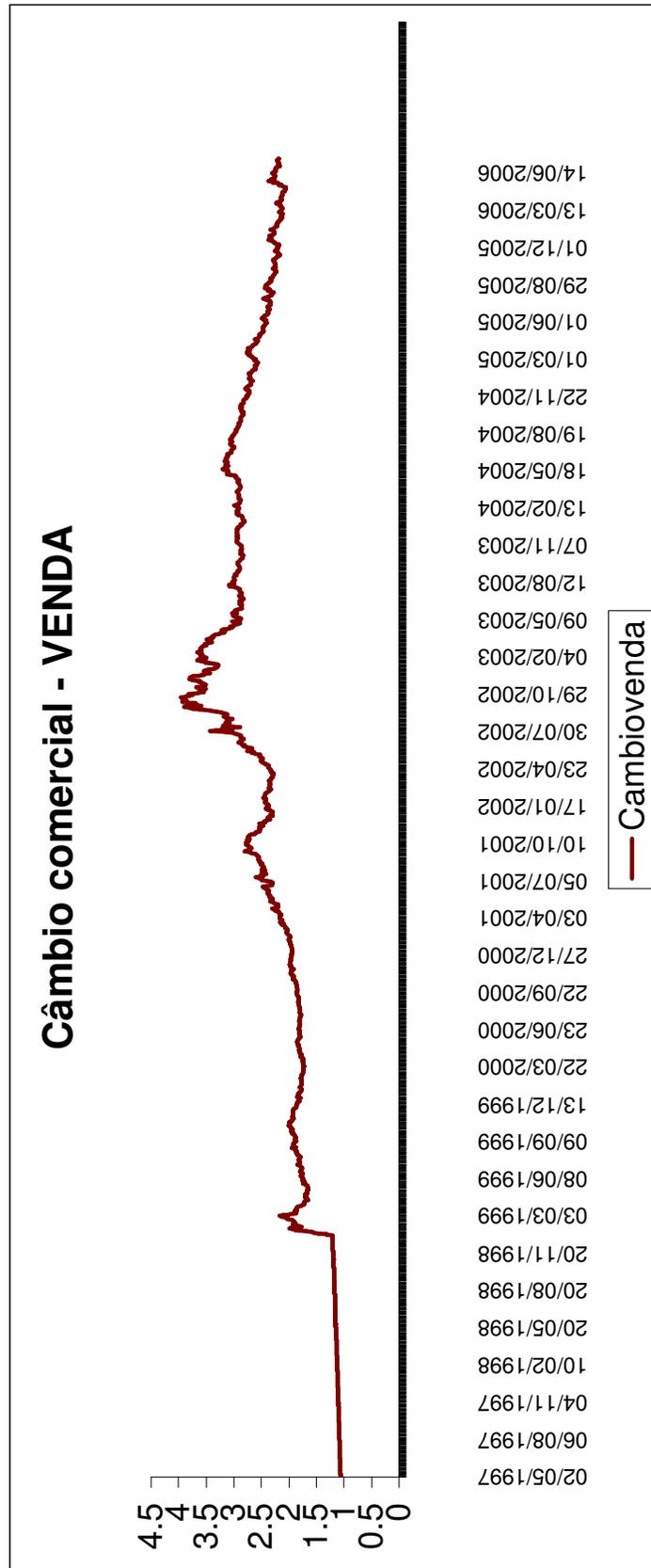


Gráfico A.4: valor do câmbio comercial para venda.

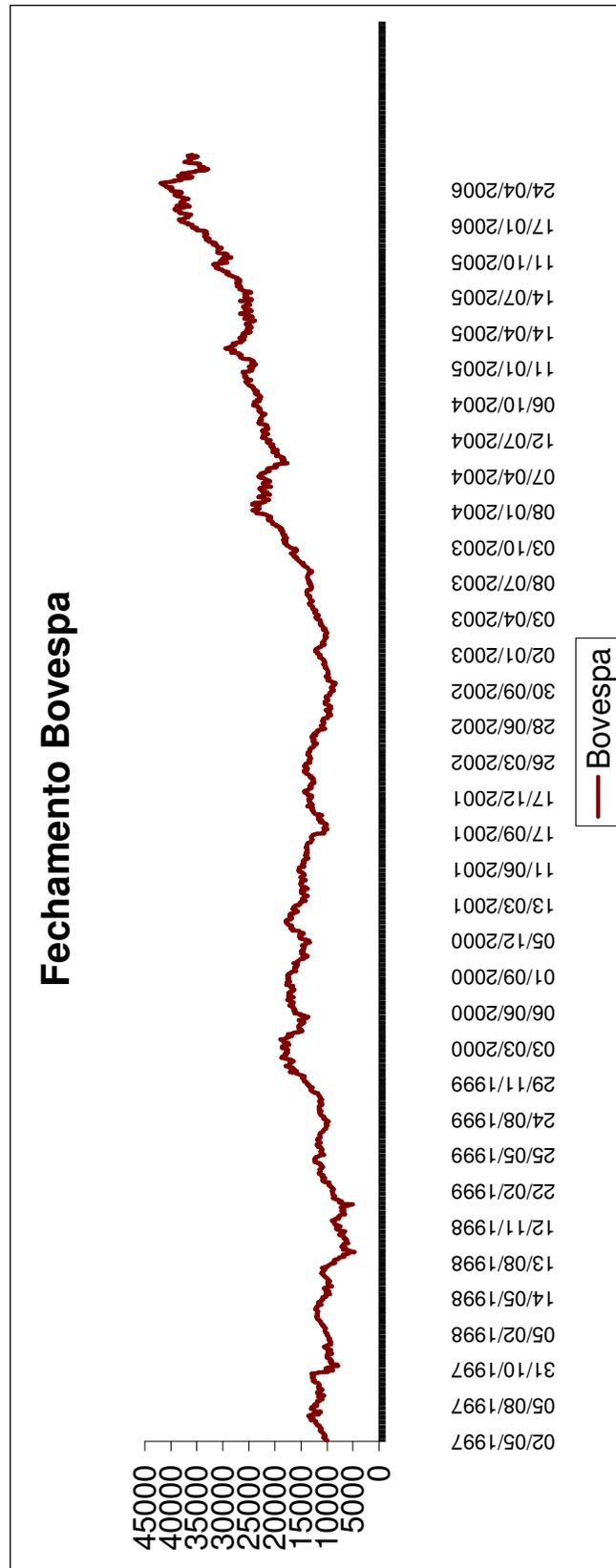


Gráfico A.5: valor do Ibovespa em seu fechamento.

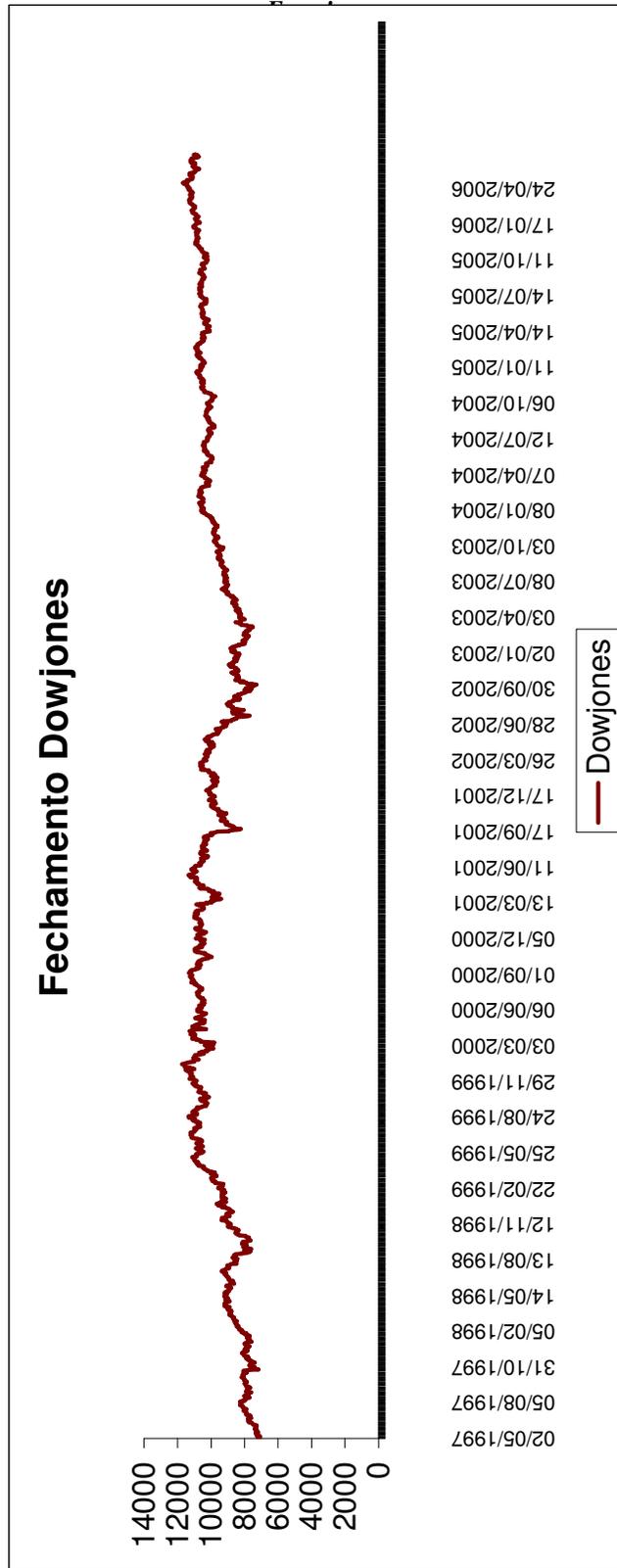


Gráfico A.6: valor do fechamento do Dowjones.

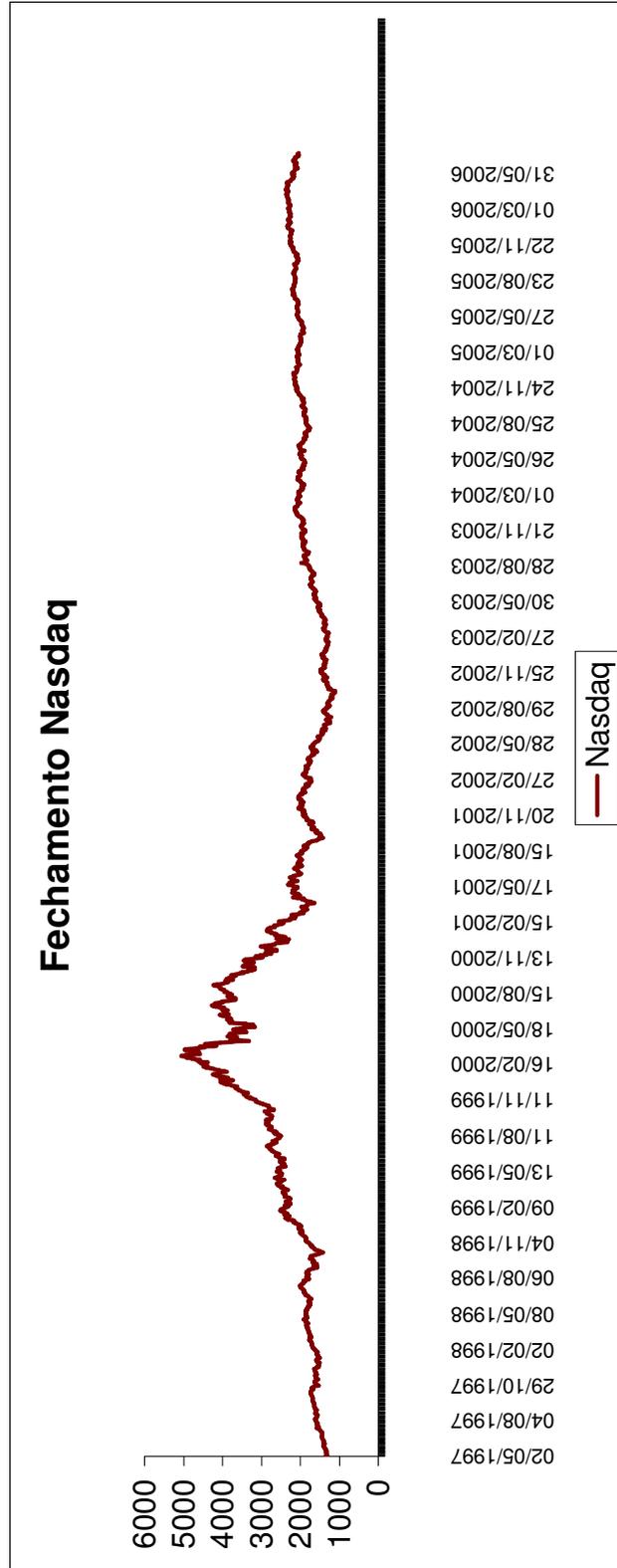


Gráfico A.7: valor do fechamento da Nasdaq.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)