

GABRIEL ROTOLO NASCIMENTO

**PREVISÃO DE PREÇOS DO MERCADO
SUCRO-ALCOOLEIRO UTILIZANDO
REDES NEURAIS**

Trabalho de Formatura

Engenharia de Produção da Escola

Politécnica da USP

São Paulo

2004

GABRIEL ROTOLO NASCIMENTO

**PREVISÃO DE PREÇOS DO MERCADO
SUCRO-ALCOOLEIRO UTILIZANDO
REDES NEURAIIS**

Trabalho de Formatura

Engenharia de Produção da Escola

Politécnica da USP

Orientadora:

Prof^a: Celma de Oliveira Ribeiro

São Paulo

2004

FICHA CATALOGRÁFICA

Nascimento, Gabriel Rotolo

**Previsão de preços do mercado sucro-alcooleiro
utilizando redes neurais. São Paulo, 2004.
100p.**

**Trabalho de Formatura – Escola Politécnica da
Universidade de São Paulo. Departamento de Engenharia
de Produção.**

**1. Redes Neurais 2. Mercado Futuro 3. Commodities.
I. Universidade de São Paulo. Escola Politécnica.
Departamento de Engenharia de Produção II.t**

AGRADECIMENTOS

Gostaria de agradecer primeiramente aos meus pais pelo apoio, confiança e por sempre me incentivarem a superar meus limites, não somente na elaboração deste trabalho, mas também durante todos os anos do curso de Engenharia de Produção e na minha vida.

Agradeço ainda à Prof^ª. Celma Ribeiro, por ter exigido o máximo de mim, por ter me mostrado o fascinante mundo da pesquisa e por ter me estimulado à querer sempre aprender mais, conseguindo fazer desse trabalho um abrir de portas para o mundo do conhecimento.

À Carlos Widonsck, por ter apresentado esse tema e ter ajudado, com muita paciência, a esclarecer todas dúvidas relacionadas com o mercado sucro-alcooleiro.

À Graziela Maragni por ter me ajudado com muito carinho à superar momentos difíceis desses anos e por ter tido tanta paciência ao corrigir meus textos.

Por fim à todos da Zoom Asset que me ensinaram muito mais que conhecimentos práticos e sim uma nova forma de enxergar a vida.

RESUMO

Esse trabalho de formatura visa estudar diferentes tipos de ferramentas de previsão, desde métodos tradicionais até modelos mais complexos como redes neurais, com o objetivo de construir um modelo de previsão de preços de contratos futuros de açúcar. A finalidade desse trabalho é auxiliar produtores de cana-de-açúcar e usineiros no processo de decidir qual produto vai ser mais vantajoso produzir (álcool ou açúcar) antes do fim da colheita.

Primeiramente será feita uma análise detalhada do mercado sucro-alcooleiro, para tentar entender as peculiaridades que fazem desse setor tão atraentes para pesquisas. Depois, será apresentada as teorias necessárias para elaboração de um modelo de previsão, se utilizando principalmente de conceitos de redes neurais. Por fim esse modelo será aplicado nas condições reais de funcionamento desse mercado com o objetivo de se achar instrumentos que ajudem a prever o comportamento do preço do açúcar.

ABSTRACT

This work seeks to study different types of forecasting tools, from traditional methods until more complex models as neural networks, to build a forecasting model to predict future prices of sugar. The purpose of that work is to help producing of sugar-cane in the deciding process which product will be more advantageous to produce (alcohol or sugar) before the end of the crop.

Firstly it will be made a detailed analysis of the sugar and alcohol markets, to try to understand the peculiarities that make this sector such attractive for researches. Then, it will be presented the necessary theories to elaborate a forecast model, using mainly concepts of neural networks. That model will finally be applied in the real conditions of this market, with the purpose to foresee the behavior of the sugar price.

Índice

1. INTRODUÇÃO	1
1.1.Introdução.....	2
1.2. Objetivo Do Trabalho	3
 2. ANÁLISE DO MERCADO SUCRO-ALCOOLEIRO.....	4
2.1. Introdução.....	5
2.2. Mercado da Cana de Açúcar	6
2.2.1. História	6
2.2.2.Produção	8
2.2.3.A Industria Sucro Alcooleira.....	9
2.2.4.A Pesquisa e Desenvolvimento.....	10
2.2.5.Posicionamento Estratégico do mercado da cana.....	11
2.3. Mercado do Álcool.....	16
2.3.1. História do Pro-álcool	15
2.3.2. Perspectiva desse Mercado.....	18
2.3.3. O legado do Pro-álcool	19
2.4. Mercado de Açúcar	20
2.4.1. Oferta e Demanda Mundial.....	20
2.4.2. Oferta e Demanda Brasileira.....	22
2.4.3. Mercado Livre	23
2.5. Formação do preço do Açúcar e do Álcool	25
 3. MODELOS DE PREVISÃO.....	28
3.1. Introdução.....	29
3.2a. Métodos De Previsão	30
3.2b. Séries Temporais.....	32
3.2.1. Sazonalidade	32
3.2.2. Tendência	33

3.2.3. Componente aleatória ou residual.....	34
3.3. Medidas de Desempenho.....	34
3.4. Métodos Quantitativos de Previsão.....	35
3.4.1. Regressão Linear Simples.....	36
3.4.2. Media Móvel Simples (MMS).....	37
3.4.3. Alisamento Exponencial Simples (AES).....	39
3.4.4. Alisamento Exponencial Sazonal.....	40
3.5. Métodos Causais de Previsão.....	41
3.5.1. Regressão Linear Simples.....	41
3.5.2. Regressão Múltipla.....	43
3.6. Análise dos Resíduos	44
3.6.1. Teste Kolmogorov-Smirnov.....	45

4. REDES NEURAIS APLICADOS NA PREVISÃO..... 46

4.1. Introdução a Redes Neurais Artificiais.....	48
4.2. Historia das RNAs.....	49
4.3. Componentes de uma RNAs.....	51
4.4. Arquitetura das Redes Neurais.....	55
4.5. Aprendizado de uma RNAs.....	57
4.5.1. Aprendizado Supervisionado.....	58
4.5.2. Aprendizado Não-Supervisionado.....	59
4.6. Classificação dos Modelos.....	60
4.7. Modelos de Redes MLP (Backpropagation).....	60
4.7.1. Algoritmo de Treinamento.....	63
4.7.2. Métodos de Aceleração e de Convergência	65
4.7.3. Generalizações.....	68
4.7.4. Limitações do Backpropagation em uma RNA.....	70
4.7.5. Tratamento dos Dados.....	71

4.8. Medidas de Desempenho de um RNA.....	74
4.9. Exemplo de uma RNA.....	76
5. EXPERIMENTOS DOS MODELOS	82
5.1. Introdução.....	82
5.2. Escolha das variáveis	83
5.3a. Tratamento dos Dados.....	86
5.3b. Medidas de Desempenho.....	87
5.4. Utilização da Média Móvel	91
5.4.1. <i>Teste de Normalidade</i>	92
5.4.1. <i>Resultados da Média Móvel</i>	94
5.5. Utilização das Redes Neurais.....	95
5.5.1. <i>Entrada das Redes Neurais</i>	96
5.5.2. <i>Treinamento</i>	97
5.5.3. <i>Saída da Rede</i>	98
5.5.4. <i>Resultados da Redes Neurais</i>	99
5.6. Aplicação Prático dos Resultados.....	101
6. CONCLUSÕES	102
6. REFERÊNCIAS BIBLIOGRÁFICAS	105

1.INTRODUÇÃO

Nos últimos anos, o Brasil tem sido destaque no cenário mundial devido à crescente produção de insumos agrícolas, dentre as quais se encontram a cana-de-açúcar e seus derivados. O consumo mundial de açúcar vem crescendo com taxas superiores às taxas da economia global, liderados principalmente pelos países em desenvolvimento como a Índia, China e Rússia, mas também pelos tradicionais EUA e países europeus.

Para atender a essa forte demanda, o Brasil tornou-se, nos últimos dois anos, o maior produtor mundial de açúcar, superando tradicionais produtores como a Índia e a União Européia. Porém, diferentemente de outras mercadorias, o comércio mundial de açúcar é marcado por forte intervencionismo dos países participantes, que se utilizam práticas protecionistas, como a imposição de sobretaxas e de cotas sobre as importações.

Por estas razões e pelo fato do mercado de cana-de-açúcar ser marcado por práticas não muito transparentes ao longo da história, os preços do açúcar e do álcool sofrem uma alta oscilação no decorrer dos anos. Essas características fazem desse mercado uma boa opção para realização de estudos que possam descrever a dinâmica do negócio tornando-o menos incerto mais atraente para novos investidores.

Uma alternativa viável e já utilizada pelos produtores e usineiros, para se proteger da oscilação dos preços, é procurar instrumentos do mercado financeiro. Assim eles têm a possibilidade de saber antes do final da colheita se vão ter lucro ou prejuízo no seu fluxo de caixa final. Mas mesmo em posse dessas ferramentas, os produtores e usineiros carecem de informações que os auxiliem nesse processo de tomada de decisão. Logo, o trabalho tentará apresentar algumas alternativas para que se consiga diminuir essas incertezas.

1.2. Objetivo do Trabalho

O objetivo desse trabalho é construir ferramentas para auxiliar produtores de cana-de-açúcar e usineiros a tomarem decisões mais precisas quanto ao percentual da produção de cana será alocado para fabricação de açúcar e álcool.

Para fazer essa escolha, produtores e usineiros dispõem de poucas informações que realmente podem auxiliá-los nesse processo. Uma de suas ferramentas é avaliar os preços dessas mercadorias e verificar qual é mais vantajosa produzir. Como produtor e usineiros estão negociando um ativo que ainda não possuem (essa escolha é feita antes da colheita de cana-de-açúcar ser realizada), eles correm o risco de que a estrutura dos preços seja diferente, quando a cana tiver sido colhida, de tal forma que a decisão a ser tomada deveria ser outra.

Inicialmente, vamos construir um modelo de previsão que possa indicar de forma clara qual a tendência do preço do açúcar num futuro próximo, verificando se a produção de açúcar será financeiramente vantajosa. Para a construção desse modelo serão utilizadas ferramentas estatísticas e serão analisadas variáveis que podem influenciar na formação do preço do açúcar.

2. ANÁLISE DO MERCADO SUCRO- ALCOOLEIRO

2.1. Introdução

O setor sucro-alcooleiro viveu nas últimas décadas momentos bem distintos economicamente. Quem vê, hoje em dia, o Brasil se destacando mundialmente na produção e exportação de açúcar, quem acompanha o aumento das exportações de álcool e a grande produtividade de cana por hectare talvez não imagine as dificuldades que esse segmento passou para chegar a essa situação.

Depois das crises do petróleo na década de 70 (1973 e 1979), a cana-de-açúcar apareceu como uma solução energética imediata para esse problema. Desde então, a produção de álcool combustível, a partir da cana-de-açúcar, passou a contar com apoio do governo e esse setor viveu seus tempos áureos com o programa Pro-álcool, que vinha tanto como uma solução para a crise do petróleo como para a alta flutuação dos preços internacionais do açúcar (Goldenberg e Moreira, 1997).

Porém, em meados dos anos 90, com o desabastecimento do mercado interno de álcool e fim dos incentivos governamentais, o declínio do Pro-álcool foi imediato e levou a uma crise sem precedentes para toda a cadeia produtiva de açúcar e álcool. Foram anos enfrentando preços baixos, queda nas vendas de veículos movidos a álcool e de concorrência desleal por conta da crescente sonegação fiscal no mercado de combustíveis.

Isso tudo, não lembrava a proposta inicial do álcool como sendo uma alternativa energética, economicamente viável, ecologicamente correta e genuinamente brasileira. O Proálcool, que nasceu como o mais arrojado projeto mundial de desenvolvimento, pesquisa e uso em larga escala de um combustível de fonte renovável, caiu em descrédito.

O programa Proálcool deixou bons resultados dentre os quais vale ressaltar o salto em produtividade que teve a lavoura de cana-de-açúcar nessa época, o que permitiu, que

com a melhora na economia mundial trouxesse uma retomada sustentável do mercado de álcool combustível, e na conquista de novos mercados para o nosso açúcar.

Hoje em dia, o setor sucro-alcooleiro vive uma nova realidade. Com o mundo todo preocupado com problemas ambientais, como aquecimento global, e o petróleo atingindo preços recordes no mercado internacional causados por incertezas na continuidade de atender a crescente demanda mundial, o álcool combustível volta a ter destaque como fonte energética. Assim, a experiência de mais de 30 anos com o álcool combustível colocou o Brasil numa posição de destaque no mercado internacional de biocombustíveis.

As exportações de álcool devem ultrapassar a barreira dos 2 bilhões de litros este ano. A adoção da mistura do combustível renovável na gasolina pelos Estados Unidos, países europeus e asiáticos, fez do Brasil como o único país capaz de suprir a crescente demanda por este biocombustível.

Já no mercado de açúcar, o Brasil tem crescido e cada vez mais tem se consolidado como uma potência nesse setor, e ainda com as recentes vitórias no âmbito da Organização Mundial do Comércio - OMC fizeram com que o maior e mais competitivo produtor mundial possa vislumbrar a abertura de novos mercados para o produto brasileiro.

2.2. Mercado Cana-de-açúcar

2.2.1. História

A história da cana-de-açúcar no Brasil caminha junto com a colonização e desenvolvimento do nosso país. As primeiras mudas vieram em 1532 e por causa do solo fértil e do clima tropical quente e úmido e a cultura dessa planta rendeu bons frutos à

muita gente. Um dos grandes beneficiários foi Portugal, que espalhou o açúcar brasileiro por toda a Europa ganhando muito dinheiro com nova colônia.

Em meados de 1600, Portugal e suas colônias foram anexados ao reino Espanhol, formando a conhecida União Ibérica, e as riquezas promovidas pela cultura canavieira acabaram por levar à invasão holandesa no Nordeste brasileiro. Depois de algumas décadas e algumas guerras, o governo português retomou o poder no nordeste, porém a economia açucareira brasileira não conseguiu se recuperar e dar continuidade às altas receitas dessa economia, muito por causa dos já conhecidos protecionismos e subsídios (empecilhos de longa data para a expansão dos mercados ao produto brasileiro) e o surgimento do açúcar de beterraba na Europa representaram o fim do “ciclo do açúcar” no Brasil, que conheceu uma fase de decadência iniciada no século XVIII e encerrada ao final do século XIX.

Com a crise de 1929, iniciada pelo crash da bolsa de valores de Nova Iorque, houve uma queda dos preços internacionais prejudicando o desempenho das exportações do açúcar e abrindo espaço para a intervenção do Estado na economia açucareira. Depois desses acontecimentos o governo brasileiro incentivou o consumo de álcool combustível e tornou obrigatória a mistura de 5% de etanol na gasolina utilizada no País, em 1931 (Informações do Ministério da Ciência e Tecnologia).

É da mesma época a criação do IAA – Instituto do Açúcar e do Alcool, com o objetivo de assegurar o equilíbrio do mercado, inclusive com a formação de estoques, o estabelecimento de cotas de produção de açúcar e o controle amplo sobre a comercialização.

O crescimento da produção de etanol expandiu a cultura da cana-de-açúcar no Sudeste, especialmente em São Paulo, com o produto ganhando mais espaço na mistura carburante diante das dificuldades de importação de petróleo, na Segunda Guerra Mundial.

O apoio do governo à produção de álcool se intensificou com as duas crises internacionais do petróleo, em 1973 e 1979. O etanol chegou a superar o açúcar no mix de fabricação, de meados dos 1980 até a década dos 1990. Mas a interferência estatal estava com os dias contados. Ao final dos anos 1990, o mercado estava livre e, desde então, desenvolve sua auto-regulamentação.

2.2.2. Produção

O Brasil é o maior produtor mundial de cana-de-açúcar, seguido por Índia, Austrália e Tailândia, sendo responsável por quase 30% da produção de cana do mundo. A tabela abaixo dá uma idéia da grandiosidade dos números envolvidos nesse mercado.

Movimenta	R\$36 bilhões
Representa	3% do PIB
Produção de Cana	340 milhões de tonelada
Investe	R\$ 3,5 bilhões / ano
Compoem-se	302 Usinas e Destilarias

Tabela 1.1. Dados do mercado de cana-de-açúcar (Fonte: ÚNICA)

Em média, nas últimas cinco safras, 52% dessa produção destinou-se às fábricas de etanol (anidro e hidratado) e 48% às de açúcar (refinado, cristal e demerara). A produção é realizada tanto no Centro-Sul quanto no Norte-Nordeste do País, em dois períodos de safra distintos, ocupando 2,4% da área agricultável do solo brasileiro, perto de 5,5 milhões de hectares.

Na Região Norte-Nordeste a safra ocorre entre os meses de setembro e março; na Região Centro-Sul, desenvolve-se entre maio e novembro. A cultura canavieira dessa área representa cerca de 85% da produção brasileira e está compreendida entre os estados de São Paulo, Paraná, Goiás, Mato Grosso, Mato Grosso do Sul, Minas Gerais, Rio de Janeiro e Espírito Santo. Responsáveis pelos 15% restantes da produção de cana-de-açúcar, na Região Nordeste alinham-se principalmente os estados de Alagoas, Pernambuco, Paraíba, Sergipe, Rio Grande do Norte e Bahia.

As 10 maiores usinas do país chegam a processar até quase 7 milhões de toneladas de cana por ano, mais de 2% da produção total do Brasil, dentre elas quase todas operam com equipamentos fabricados por empresas nacionais de bens de capital, cuja tecnologia permitiu ao País alcançar rendimento industrial invejável. Se destinada apenas à fabricação de álcool, cada tonelada de cana-de-açúcar moída resultaria hoje em 89 litros de etanol hidratado ou 85 litros de etanol anidro; direcionada exclusivamente à produção açucareira, renderia 118 kg de açúcar e 10 litros de álcool do mel residual.

Entretanto, em regime normal de operação de mercado, o rendimento médio nacional para cada tonelada de cana-de-açúcar moída fornece 71 kg de açúcar, 42 litros de álcool ou 11,5 toneladas de açúcares totais recuperáveis por hectare de cana-de-açúcar cultivada. (Dados ÚNICA)

2.2.3. A indústria Sucro-Alcooleira

A agroindústria sucro-alcooleira do país gera 1 milhão de empregos diretos e abriga 60 mil produtores rurais que fornecem cana-de-açúcar. Em mais de 960 municípios, cerca de 17% do total do Brasil, a economia tem forte presença dessa atividade, com processo permanente de geração descentralizada de empregos e renda. O setor mantém diretamente mais de 600 escolas, mais de 200 creches e mais de 300 ambulatórios médicos em todo o Brasil.

Além dos efeitos sociais diretos, a ocupação intensiva de mão-de-obra rural promove um óbvio efeito positivo na redução dos fluxos migratórios para as cidades e na redução do inchaço populacional urbano. O Estado de São Paulo, que produz 60% da cana-de-açúcar é referência de modernidade quanto às relações com os trabalhadores. Quase todos têm carteira de trabalho e oferecem todos os direitos da legislação trabalhista. Em média, recebem remuneração e benefícios 3,5 vezes maiores do que o salário mínimo nacional (equivalente hoje a US\$ 83,62), na lavoura – onde os trabalhadores têm baixo

nível de qualificação e escolaridade – e 5,3 vezes maiores nas ocupações industriais. Entre esses benefícios estão assistências médica, odontológica, oftalmológica e farmacêutica; seguro de vida, refeição, cestas básicas, vales para refeições e transporte.

Os produtores rurais, pequenos e médios proprietários, em sua maioria, são remunerados de acordo com uma fórmula paramétrica que considera o teor de açúcar total da matéria-prima e os preços do açúcar e do álcool nos mercados interno e externo. O valor pago pela cana-de-açúcar no Brasil representa 60% do faturamento da indústria. No Estado de São Paulo, na safra 2003/2004, o produtor recebeu em média US\$ 10.35 por tonelada de cana-de-açúcar fornecida às unidades industriais.

2.2.4. Pesquisa e Desenvolvimento

A partir de 1979, o setor sucroalcooleiro tem investido cerca de US\$ 40 milhões por ano em pesquisa e desenvolvimento, desde que a maior cooperativa de produção do País criou a sua entidade de investigação e estudos, o CTC – Centro de Tecnologia Copersucar – que está prestes a se transformar em Centro de Tecnologia Canavieira, consolidado em uma associação entre os atuais proprietários e grande parte dos produtores de cana-de-açúcar, açúcar e álcool do País.

Em São Paulo, a cultura de cana-de-açúcar é a atividade agrícola com menores índices de contaminação dos solos e das águas em todo o mundo por utilizar as menores quantidades de agroquímicos – fertilizantes e defensivos – das Américas. A parte maior dos resíduos industriais é processada e utilizada nas operações de irrigação e fertilização dos cultivos.

O setor sucro-alcooleiro está integrado ao Programa Genoma da Cana-de-Açúcar, que envolve mais de 150 pesquisadores de várias instituições na tarefa dedicada de mapear e

caracterizar os genes da planta, criando as condições para o surgimento de espécies mais produtivas, adaptáveis a diversos tipos de solo e clima, resistentes a pragas e doenças.

O País utiliza avançado programa de processamento de imagens de satélite para planejar ações estratégicas; monitorar a cultura da cana-de-açúcar e o uso da terra; identificar áreas potenciais de expansão da lavoura; construir novas usinas e identificar variedades vegetais. Produz rotineiramente análises detalhadas de composição do etanol hidratado ou anidro para cumprir as especificações brasileiras e internacionais, procedimento também estendido ao açúcar, de forma a certificar o produto quanto à segurança de produtores e consumidores.

Nas pesquisas brasileiras incluem-se o desenvolvimento do plástico biodegradável ou PHB, obtido a partir do açúcar, já fabricado em escala industrial, produto bem acolhido na Europa e na Ásia. Ensaia-se também a produção de álcool a partir dos resíduos da cana, programa desenvolvido com indústrias de base locais e instituições públicas. Conhecido como DHR – Dedini Hidrólise Rápida, o empreendimento pretende elevar a capacidade de produção de etanol no País, com o aproveitamento total da matéria-prima.

2.2.5. Posicionamento Estratégico do mercado da cana

Nesse item será apresentada de forma introdutória uma abordagem estratégica do mercado de cana-de-açúcar através de um modelo proposto por Michael E. Porter (1980) que se baseia na interação sistêmica de algumas classes de fatores determinantes na competitividade nacional e internacional do setor estudado. O objetivo dessa análise é permitir um melhor entendimento da capacidade de competir das empresas que compõe esse setor.

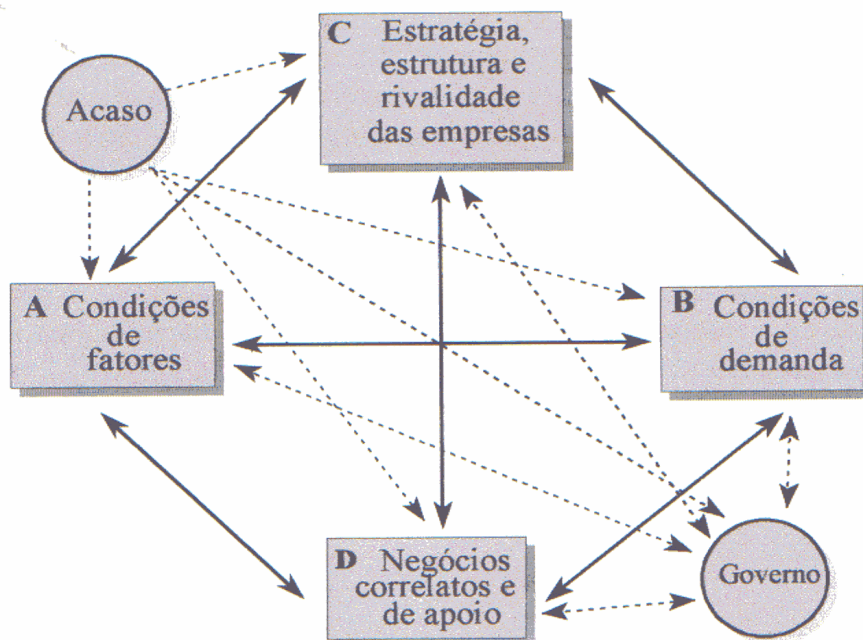
O modelo proposto é o chamado de *diamante de competitividade* que aborda os seguintes fatores:

- A. Condições de fatores de produção e de infra-estrutura*
- B. Condições de demanda*
- C. Estratégia estrutura e rivalidade das empresas participantes*
- D. Negócios Correlatos e de apoio*

Além desses fatores, ainda existem duas classes de fatores considerados menos importantes, mas que são assuntos pertinentes quando se faz uma análise do mercado de cana.

- E. Ação ou inação do governo*
- F. Acaso ou eventos aleatórios*

A figura 1.1. ilustra bem o relacionamento desses fatores, e mais adiante será explicitado de forma mais detalhada a abertura desses fatores no mercado estudado.



O diamante da competitividade nacional $\Rightarrow$ internacional.

Figura 1.1. Diamante da competitividade (Porter-1980).

A. Condições de Fatores de produção e de infra-estrutura

O mercado de cana-de-açúcar brasileiro apresenta algumas vantagens que o deixa em ótima situação para competir com o mercado internacional, entre elas:

- *Grande disponibilidade de solos*: Conforme foi citado no item 1.1.2 o Brasil destina cerca de 5 milhões de hectares (2,4% da área agricultável do solo brasileiro) para a produção de cana-de-açúcar;
- *Chuva e insolação*: O clima brasileiro proporciona ótimas condições para o plantio da cana;
- *Abundância de mão de obra*: Por abrigar uma população extensa, com uma boa parte no campo, a mão-de-obra no Brasil pode ser considerada abundante e barata o que ajuda a reduzir significativamente os custos de produção.
- *Infra – estrutura*: Como a cultura da cana foi iniciada logo nos primórdios do desenvolvimento do país e por causa dos grandes avanços da produtividade na década de 80, o Brasil desfruta de uma ótima infraestrutura para o plantio da cana;
- *Pesquisa e Desenvolvimento*: Item 1.2.4 mostra bem as vantagens dessa área.

B. Condições de Demanda

A demanda brasileira por produtos como açúcar e álcool, já seria por si só suficientemente grande para ser atendido pelo setor produtivo brasileiro, porém, quando se leva em conta a internacionalização dos mercados, a demanda internacional por esses produtos é um universo muito mais amplo, que mesmo sendo marcado por forte protecionismo, não deixa de ser uma boa oportunidade para a produção nacional.

- *Mercado Interno*:
 - Mercado Pouco exigente;
 - Tamanho significativo.
- *Mercado Externo*:
 - Aumenta ainda mais o mercado;
 - Obtenção de economia de escala.

C. Estratégia e Estrutura/Rivalidade

De forma simplificada podemos separar a estrutura de custos desse mercado da seguinte forma:

- 60% do custo proveniente da parte agrícola;
- Mercado com muitas cooperativas regionais.

Se tratando da Estratégia / Rivalidade do mercado da cana no Brasil, ficam bem claros alguns pontos:

- *Setor tradicionalmente controlado*: O setor de energia no Brasil em geral, é marcado por forte intervenção do governo sendo tratado como questão estratégica interna porém com o fim do Pro-álcool o setor ficou desregulamentado sem muita preocupação das autoridades;
- *Pouca rivalidade*: Produtores e usineiros ficaram marcados nas décadas de 80 e início de 90, pela falta de competição no que se refere a preço. Por esse motivo, esse setor perdeu muita credibilidade com a população em geral, por apresentarem aumentos de preços, muitas vezes considerados abusivos, sem motivos claros.
- *Competir por ganhos de escala e não diferenciação*. Como o açúcar e álcool são produtos que não apresentam muita diferenciação entre os concorrentes, é comum usineiros competirem por ganhos em escala e produtividade.

D. Negócios correlatos e de apoio

Pode-se identificar nesse vértice do diamante da competitividade o aumento significativo na importância da tecnologia na produção da cana-de-açúcar, está comprovado que é cada vez mais importante investir em novas técnicas de plantio e colheita, novos processos de produção do açúcar e do álcool para continuar a crescer e se destacar mundialmente nesse setor.

E. Ação e inação do governo

É importante frisar as melhorias significativas nesse mercado desde de otimização nos processo produtivos até na profissionalização das empresas participantes depois do programa Pro-álcool, no qual o governo subsidiou e investiu bastante produtores e usineiros.

F. Acaso ou eventos aleatórios

Por se tratar de uma atividade rural, onde a terra, o plantio e o clima são fatores essenciais para o sucesso de uma colheita, os produtores acabam ficando muito dependentes de acontecimento aleatórios da natureza.

2.3. Mercado do Álcool***2.3.1. História Pro-Álcool***

O Programa Brasileiro de Álcool Combustível (Pro-Álcool) nasceu em meados dos anos 70 como uma solução imediata para a primeira “crise do petróleo” ocorrida em 1973. As conseqüências desse choque foram muito graves para o país. Naquela época o Brasil importava praticamente 80% do petróleo consumido e num único ano, de 1973 para 1974, as despesas com a importação de combustível saltaram de US\$ 600 milhões para mais de US\$ 2 bilhões.

Uma vez disponível, o "combustível verde" foi aproveitado de duas formas. No início, o álcool anidro era misturado à gasolina. Liderando esse processo de mistura de dosagens crescentes, a gasolina distribuída na cidade de São Paulo passou de uma percentagem de mistura de 20%, em 1977, para 22%, em 1980 e foi adotada em todo o país.

No início dos anos 80, surgiram os motores que dispensavam totalmente a gasolina, utilizando unicamente álcool hidratado. Nos primeiros anos, eram simples adaptações

cujo desempenho muitas vezes deixava a desejar. Foi preciso esperar 1979 e o segundo choque do petróleo para que as montadoras colocassem no mercado os modelos especialmente fabricados para funcionar a álcool e nesse momento o álcool combustível superou todas as previsões.

Em 1984, os carros a álcool constituíam 94,4% da produção das montadoras, ou seja, 19 de cada 20 carros que saíam das linhas de montagem. Havia uma política definida que remunerava adequadamente o produtor de álcool e mantinha uma relação diferenciada entre os preços do álcool e da gasolina (informação da ANFAVEA).

Vários foram os fatores que incentivavam essa mudança no perfil da frota de automóveis brasileira, dentre tantos se pode citar:

- *Economia interna:* O programa incentivaria a produção em larga escala do álcool combustível (através de financiamento, subsídios, benefícios fiscais, entre outros), o que estimularia a economia do Brasil. O investimento na produção, aumentaria o número de empregos no campo melhorando a situação econômica do país.
- *Balança Comercial:* O objetivo do programa era reduzir a importação de petróleo para transformação em gasolina, pois o preço dessa commodity estava muito alto acarretando em um déficit comercial muito grande. Com a fabricação de álcool combustível, além de diminuir significativamente o déficit comercial, o programa serviria para trocar dividas em dólar por investimento em moeda local (Moreira e Goldemberg – 1997)
- *Meio ambiente:* A utilização do álcool como combustível de automóvel, reduz significativamente a emissão de CO₂ se comparado com a emissão de combustíveis fósseis.

De fato, além dessas vantagens, o programa foi responsável por grandes melhorias no campo: criou-se mais de 700 mil postos de trabalho no campo, com um baixo custo de

investimento (US\$20 mil cada); aumentou a produtividade e a tecnologia na colheita da cana; redução de custo no processo produtivo, o que tornou o Brasil como um grande “*player*” mundial no mercado de açúcar; mostrou para o mundo a possibilidade de implementar um programa de fonte de energia renovável em menos de 10 anos.

Porém, no final da década de 80, a participação dos carros a álcool na produção anual de veículos despencou dos níveis de 80% para menos de 1% no ano de 1996 (vide Figura 1.2), vários foram os fatores dessa brusca mudança no perfil automobilístico brasileiro e que prejudicaram o programa:

- Em 1986, o preço do petróleo caiu bruscamente (caiu pela metade), deixando assim a gasolina à preços mais acessíveis. Logo, o álcool que era subsidiado e custava até 40% menos que a gasolina, passou a ficar mais caro, relativamente, deixando de ser tão atrativa para os consumidores e deixou de ser tão vantajosa a ponto de o governo subsidiar e garantir a compra da produção.
- Entre o final dos anos 80 e início da década de 90, houve uma série de desabastecimentos de álcool combustível em algumas regiões importantes do país. Esses acontecimentos minaram a confiança do consumidor que temiam não poder utilizar seu veículo.
- Como a Petrobrás garantia a compra do etanol dos produtores, em épocas de desabastecimentos a empresa tinha que importar álcool etanol, à preços maiores que no Brasil, para suprir a demanda externa, isso causou um enorme déficit para empresa. E para manter a paridade do preço da gasolina/álcool, o governo sobretaxava a gasolina para cobrir os subsídios do álcool, o que ficou conhecido como “Subsidio Cruzado”.
- Com a popularização dos carros populares, causada pela a redução dos impostos desses carros, houve um enorme avanço nas vendas desse segmento, porém por restrições técnicas não se produzia “carros

populares” (motor 1.0) movidos à álcool o que fez piorar ainda mais a situação desse mercado.

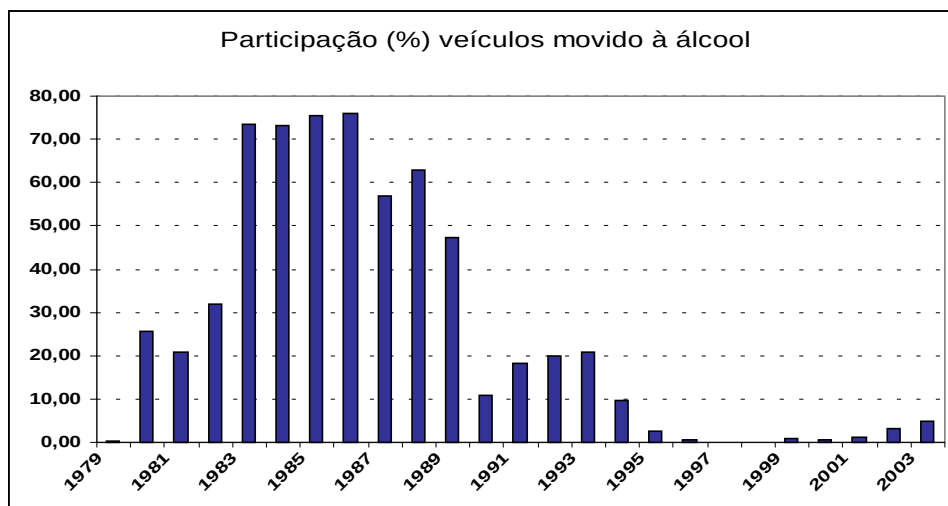


Figura 1.2: Participação dos veículos movidos à álcool no Brasil (Fonte:ANFAVEA)

2.3.2. Perspectiva desse mercado

Com os preços do petróleo ultrapassando a barreira de US\$50, com tantas incertezas envolvendo o futuro do abastecimento do petróleo por causa dos conflitos geoeconômicos do Oriente médio e por causa da preocupação crescente da preservação ambiental, o álcool combustível volta ao cenário como uma boa opção de energia limpa. Uma dessas manifestações é o aparecimento dos veículos com bi-combustível, movido tanto à gasolina quanto à álcool, que começaram a ser vendidos no início de 2003 e vêm conquistando mercado rapidamente nos últimos meses graças ao baixo preço do álcool combustível quando comparada ao atual preço da gasolina.

Os novos veículos resgataram a afinidade dos brasileiros pelo etanol, um combustível que polui menos, melhora o desempenho e aumenta a vida útil dos automóveis. Além disso, com os carros bi-combustíveis o consumidor decide conforme os preços vigentes e suas necessidades o combustível que vai escolher utilizar.

Uma outra oportunidade existente, é a possível adesão de países desenvolvidos em programa de mistura do álcool anidro na gasolina por motivos ambientais. Tal acontecimento seria muito benéfico para o país pois o Brasil já possui toda uma capacidade instalada e “know how” para atender essa demanda. Se países como Estados Unidos, Japão adicionarem cerca de 2 a 5% de álcool em seus combustíveis, isso já representaria um grande avanço para indústria brasileira.

2.3.3. O legado do Pro-álcool

O principal objetivo do Pro-álcool era aumentar a produção de álcool combustível para atender às novas necessidades do mercado, isso tudo sem causar prejuízo na produção de açúcar, produto essencial à economia brasileira. Nesse ponto seu objetivo foi realizado, pois os esforços tecnológicos desenvolvidos para o álcool beneficiaram também a produção do açúcar brasileiro, que se tornou altamente competitivo no mercado mundial. O avanço mostrou-se especialmente impressionantes em São Paulo, onde os custos de produção são atualmente os menores do mundo.

Segundo a Fundação Getúlio Vargas, os ganhos de produtividade têm sido constantes, a queda dos custos de produção do álcool na região Centro-Sul desde o início do programa foram em média de 3,2% ao ano. No Nordeste, ficaram em torno de 1,9% por ano.

Além da melhor tecnologia de cultivo, as pesquisas levaram ao desenvolvimento das variedades da cana-de-açúcar utilizadas na região Centro-Sul. Elas têm sido responsáveis por um crescimento da produtividade dos canaviais entre 15% e 20% nos últimos anos. Atualmente, os pesquisadores paulistas concentram esforços no campo da biologia molecular, com o objetivo de acelerar o processo de seleção de variedades de cana-de-açúcar mais eficientes.

As usinas estão extraíndo cada vez mais açúcar da cana-de-açúcar processada. O índice de extração de açúcar, que era de 89%, alcança hoje 97%. Paralelamente, disseminou-se a técnica de queima de bagaço para produção de energia, tornando 95% das usinas do

estado de São Paulo auto-suficientes em eletricidade. Parte dessa energia já é inclusive direcionada para a rede elétrica estadual.

Outra grande melhoria refere-se à melhoria do desempenho dos motores a álcool, que sendo centro de estudos nas últimas décadas, solucionou grande parte dos problemas do passado, como problemas de acionamento do motor em dias de frio e o funcionamento dos motores para trabalharem em regime de compressão mais elevada.

Esses pontos ilustram um pouco dos avanços tecnológicos e sociais nos últimos anos desse setor, sendo que o Pro-álcool teve papel fundamental nesse processo.

2.4. Mercado do Açúcar

2.4.1. Oferta e Demanda Mundial

Brasil é o maior produtor de açúcar do mundo e, sem demonstrar acomodação, sua produção vem crescente consistentemente ano após ano. Seus maiores competidores são a Índia, União Européia, China e Estados Unidos. Já em termos de consumo, o Brasil está na segunda posição do ranking dos consumidores mundiais desse produto, ficando atrás da Índia e sendo seguida pelos Estados Unidos e pela China.

O intervalo entre as safras 1994/95 e 2003/04 mostra algumas evidências interessantes sobre a produção e o consumo de açúcar no âmbito mundial. Já no começo dos anos 90 o Brasil figurava entre os 10 principais produtores; mas desde a safra 1996/97 disputa a primeira posição no ranking com a Índia. Na safra 2002/03, a exemplo da temporada anterior, ocupou a liderança. De modo geral, o crescimento global no consumo do produto está solidificado e com a expansão populacional aumenta ao redor de 1,5% ao ano. O aumento da produção praticamente ficou concentrado na América do Sul e na Ásia, e permanece estável nos demais continentes.

1000 ton

Países	94/95	95/96	96/97	97/98	98/99	99/00	00/01	01/02	02/03	03/04*
Brasil	12.714	15.190	14.636	18.042	19.376	20.100	17.100	20.400	23.760	22.400
Índia	15.851	17.943	14.093	13.923	16.893	20.219	20.490	20.475	20.100	18.800
União										
Européia	16.560	16.973	18.133	18.935	17.591	19.187	18.078	16.032	18.284	17.432
China	5.901	6.770	7.339	8.745	9.708	7.525	6.849	8.305	9.488	9.532
USA	7.254	6.698	6.587	7.178	7.903	8.203	7.956	7.172	7.620	7.797
Tailândia	5.513	6.323	6.098	4.325	5.478	5.721	5.107	6.397	6.813	6.800
México	4.650	4.667	4.840	5.492	4.993	4.979	5.220	5.168	5.038	5.200
Austrália	4.791	5.632	6.009	5.818	5.150	5.448	4.162	4.662	5.350	4.900
Cuba	3.419	4.504	4.316	3.285	3.851	4.060	3.500	3.700	2.200	2.200
África do Sul	1.774	2.106	2.409	2.533	2.646	2.685	2.895	2.542	2.930	2.655
Polônia	1.492	1.717	2.435	2.286	2.236	1.962	2.188	1.674	2.250	1.900
Ucrânia	3.598	3.804	2.935	2.280	2.038	1.720	1.687	1.800	1.550	1.900
Filipinas	1.705	1.853	1.883	1.866	1.682	1.620	1.805	1.900	2.000	2.100
Indonésia	2.221	2.102	2.162	1.720	1.622	1.690	1.800	1.725	1.755	1.770
Outros	36.684	34.200	36.640	37.916	37.825	55.631	40.322	40.672	42.968	33.242
Total	115.920	122.229	122.546	124.939	130.425	136.603	130.432	134.662	143.275	138.634

Tabela 1.1: Produção Mundial de Açúcar (FONTE: USDA)

Data	95/96	96/97	97/98	98/99	99/00	00/01	01/02	02/03	03/04**
Índia	14.820	15.697	16.700	16.977	17.296	17.845	19.760	20.750	21.500
Brasil	8.100	8.500	8.800	9.100	9.100	9.250	9.450	9.640	9.980
China	8.040	8.268	9.012	8.907	8.476	8.650	9.050	9.122	9.194
USA	8.667	8.866	8.903	9.079	9.049	9.306	9.249	8.699	9.177
Rússia	5.000	5.100	4.960	4.995	6.130	6.840	7.040	6.600	6.200
México	4.140	4.240	4.240	4.400	4.576	4.623	5.082	5.266	5.283
Paquistão	3.090	3.050	3.200	3.210	3.300	3.450	3.450	3.500	3.500
Indonésia	2.900	3.100	3.150	2.800	3.200	3.300	3.350	3.400	3.450
Egito	1.647	1.890	1.920	1.950	1.940	2.030	2.290	2.250	2.255
Japão	2.520	2.448	2.418	2.358	2.142	2.293	2.277	2.314	2.250
Ucrânia	2.250	2.150	1.900	2.050	1.937	2.100	2.020	2.185	2.200
Irã	1.850	1.870	1.900	2.000	2.100	1.956	1.981	2.100	2.056
Filipinas	1.970	1.900	1.918	1.900	1.930	1.940	1.950	1.980	2.010
Turquia	1.900	1.960	2.020	2.080	2.050	1.850	1.850	1.900	1.950
Tailândia	1.560	1.651	1.756	1.800	1.650	1.750	1.832	1.900	1.900
União									
Européia	16.972	18.001	19.047	17.561	19.175	18.194	15.973	18.504	14.529*
Outros	40.265	41.108	42.270	42.688	43.890	43.741	43.885	36.529	55.691
Total	116.275	119.476	122.778	124.067	127.422	130.285	132.560	133.304	138.596

Tabela 1.2: Consumo Mundial de Açúcar (FONTE: USDA)

De acordo com a Tabela 1.1, em 2003 a produção mundial de açúcar foi cerca de 140 milhões de toneladas, deste total 17% vem do mercado brasileiro, porém o mercado

açucareiro é tradicionalmente marcado por fortes intervenções protecionistas dos principais países consumidores. Para tentar proteger produtores locais, países como EUA, Rússia e a UE adotam políticas de sobretaxas às importações, em alguns casos impõem cotas de importação. Essas práticas impedem que o mercado mundial de açúcar funcione livremente.

Portanto, somente 30% do consumo mundial do açúcar (42mi ton/ano) é comercializado livremente entre as nações, desse total o Brasil exporta 13mil ton/ano, deixando-o em posição privilegiada no mercado internacional.

2.4.2. Oferta e Demanda Brasileira

Devido a sazonalidade da cana de açúcar, a produção brasileira de açúcar, apresenta grande variação de preços entre safra e entressafra. Por necessidade de caixa das usinas, podemos dizer que no início da safra (abril/maio para o centro-sul e setembro/outubro para o nordeste), os preços são deprimidos, pois há uma oferta maior do produto. Com o andamento da colheita ao longo do ano, os preços tendem a subir. O detalhe é que, neste setor, produz-se em seis meses o que se comercializa em doze.

A produção de açúcar vem aumentando ano a ano, devido a maior participação do açúcar brasileiro no mercado internacional. Não fossem os subsídios de alguns países, como os europeus, esta participação seria ainda maior, em vista dos custos de produção no Brasil ser um dos menores do mundo.

(mil toneladas métricas)

Safras	Estoque inicial	Produção	Importação	Consumo	Exportação	Estoque Final
89/90	2.045	8.000	234	7.281	1.421	1.576
90/91	1.577	8.952	-	6.924	1.409	2.196
91/92	2.196	9.510	115	7.400	1.771	2.650
92/93	2.650	11.104	62	7.459	2.837	3.519
93/94	3.519	10.534	20	7.760	3.727	2.585
94/95	2.585	12.714	24	8.033	4.986	2.304
95/96	2.304	15.190	-	8.385	5.502	3.607
96/97	3.607	14.636	-	8.939	5.921	3.383
97/98	3.383	18.042	-	9.172	8.051	4.202
98/99	4.203	18.300	-	9.100	8.750	1.010
99/00	1.010	20.100	-	9.100	11.300	710
00/01	710	17.100	-	9.250	7.700	860
1/2	860	20.400	-	9.450	11.600	210
02/03	210	23.760	-	9.640	14.230	100
3/4	100	22.400	-	9.980	12.420	100

Tabela 1.3: Oferta e Demanda Brasileira de Açúcar (FONTE: USDA)

2.4.3. Mercado Livre

Como foi mencionado anteriormente, o mercado sucro-alcooleiro é marcado por fortes intervenções de países ricos com o propósito de protegerem sua economia agrícola interna. Dentre outros países, os da União Européia, particularmente, impõem severas sobretaxas alfandegárias e cotas de importação para evitarem a entrada de açúcar dos grandes países produtores, isso porque a produção de açúcar européia é extremamente ineficiente e não tem condições de competir em igualdade com outros países e só conseguiu sobreviver economicamente durante tanto tempo através de subsídios agrícolas, sobretaxando o açúcar importado e muitas vezes impondo cotas de importação por países que deseja importar açúcar para UE.

Um dos motivos da ineficiência produtiva européia está no fato deles produzirem açúcar a partir da beterraba, enquanto no Brasil produz açúcar a partir da cana. O custo de

produção do Brasil para uma tonelada de açúcar à base de cana é US\$ 98, a mesma tonelada produzida pela Europa, à base de beterraba, custa até sete vezes mais¹.

Além disso, a cana-de-açúcar também garante maior produtividade de açúcar por hectare, enquanto no Brasil, o rendimento é de 9,8 sacas por tonelada de cana, na Europa é de 8,8 sacas por tonelada de beterraba. Na Rússia, a proporção cai para 3 sacas por tonelada de beterraba.¹

Três são os pilares de sustentação desse regime que garantem uma renda elevada aos agricultores europeus:

- i) Cotas de produção repartidas por país e por usina;
- ii) Preço garantido da beterraba e do açúcar três vezes superior ao preço internacional (€600 contra €200);
- iii) Proteção tarifária que bloqueia a entrada de importações.

Por causa dessas peculiaridades, apenas 30% do consumo mundial do produto é comercializado livremente, o Brasil ao exportar 13 milhões de toneladas em 2003, ocupa pouco mais de 30% do mercado livre mundial. No entanto, a dimensão desse mercado e o número de parceiros comerciais são certamente menores do que poderiam ser na ausência da proteção excessiva.

Os principais resultados dessas políticas protecionistas de países como Rússia, EU, EUA e o Japão são:

- Redução do consumo global de açúcar;
- Estímulo à produção do produto em países que não tem condições de competir sem ajuda governamental;
- Alta restrição do comércio mundial;

¹ Esses números pertencem a um estudo da Agroindústria Canavieira de São Paulo (UNICA)

- Contribuir para depreciação dos preços de açúcar no mercado internacional.

Nesse cenário, obviamente, o Brasil defende fortemente comércio livre e justo para o açúcar, se posicionando contra as barreiras às importações, favorecimento de preços internos, subsídios às exportações e outros mecanismos de proteção. A principal frente de batalha do Brasil para combater essas irregularidades é a OMC (Organização Mundial do Comércio) onde, juntamente com outros países produtores que também são prejudicados pelo protecionismo, tem promovido uma campanha mundial para eliminação gradual dessas barreiras.

Além dessa perspectiva de melhoras no comércio mundial através da OMC, os produtores brasileiros têm uma boa perspectiva pela frente, a China, país com a maior população mundial e com as maiores taxas de crescimento econômico das últimas duas décadas, ainda apresenta um consumo per capita muito baixo. Enquanto nos Estados Unidos e CEE o consumo de açúcar é de 30 kg por habitante/ano, os chineses não passam de 6 kg por habitante/ano. Por isso, com os avanços do capitalismo na china é bem possível que o consumo de açúcar cresce melhorando ainda mais as vendas do Brasil.

2.5. Formação do Preço de Açúcar e do Alcool

Um dos problemas na formação do preço do açúcar reside no fato de não existir um mercado formal de compra e venda desse produto à vista. Isso faz com que quase toda negociação seja balizada nos preços do mercado futuro.

No Brasil, contratos futuros são negociados na Bolsa de Mercadorias e Futuros (BM&F) e representam um compromisso de compra ou de venda para uma data futura previamente estabelecida. No caso dos produtos agrícolas, os contratos futuros não

têm liquidez ¹ suficiente para que os usineiros possam utilizar esse recurso de forma rotineira. Assim os preços das Bolsas de Valores americanas acabam indicando as tendências das commodities no comércio mundial.

De acordo com conceitos básicos de finanças o preço de um contrato futuro (F) nada mais é que o preço desse ativo à vista (S) corrigido pela taxa de juros (i) até o fim do contrato (em dias úteis - du), representado pela fórmula abaixo.

$$F = S \times (1 + i)^{(du/252)} \quad (2.1)$$

Por exemplo, para calcular o preço futuro (6 mês adiante) de uma ação XYZ, que está cotada na bolsa por um preço R\$30,00 e a taxa de juros nesse mercado é de 25%a.a.,

temos: $P_{\text{futuroXYZ}} = 30 \times (1 + 0,25)^{\frac{126}{252}} = R\$ 34,29$

O problema desse conceito para o mercado de açúcar (commodities em geral) é a inexistência de um banco de dados de preço à vista, como a negociação de ações em bolsa de valores do exemplo acima, no qual acarreta em uma distorção da fórmula (3.1), de forma que o preço futuro seja também, ou muitas vezes em demasia, influenciado por outras variáveis como: expectativa em relação à safra, previsão climática oferta e demanda futura.

Outro ponto importante é definir quem são os agentes que operam nesse mercado:

- **Vendedores:** São aqueles que geralmente possuem o produto físico, como produtores rurais e cooperativas e usineiros. Na maioria das vezes temem uma queda nos preços futuros e por isso assumem posições vendidas. Caso o preço caia, eles ganham a diferença e caso subam eles possuem o produto físico para

¹ Um ativo apresenta uma boa liquidez quando os agentes que operam nesse mercado conseguem entrar ou sair de posições com relativa facilidade, ou seja, conseguem comprar e vender esse ativo sem terem que pagar um ágio ou deságio (*spread*) muito alto.

entregar não assumindo grandes riscos no meio tempo. Essa operação de fixar o preço através de venda de contratos futuro é conhecida como *hedge* de venda.

- **Compradores:** São aqueles que se preocupam com uma alta nos preços e, geralmente, utilizam o produto como matéria-prima, por exemplo, as agroindústrias. Essa operação é a “ponta” contrária da operação acima, é conhecida como *hedge* de compra.
- **Especuladores:** São aqueles que não têm o produto físico, não participam da atividade agropecuária, mas atuam nos Mercados Futuros com o objetivo de aferir lucro com a operação. A participação dos especuladores dá liquidez ao mercado futuro.

3. MODELOS DE PREVISÃO

3.1. Introdução

Esse capítulo tem como propósito apresentar algumas técnicas estatísticas de previsão, com o objetivo de elaborar um modelo para previsão de preços do contrato futuro do açúcar negociados na NYBOT (New York Board of Trade) ¹.

A utilização de ferramentas estatísticas para previsão de preços de ativos é algo rotineiro e adotado há muito tempo por acadêmicos e profissionais do mercado financeiro. Por ser relativamente fácil conseguir séries temporais de preços, dos mais diversos ativos existentes, a previsão do comportamento futuro com base nos preços históricos do passado tem se tornado cada vez mais difundida e valorizada no mercado.

Para Montgomery; Johnson (1976), o propósito da previsão é reduzir o risco na tomada de decisão. Previsões na maioria das vezes são erradas, porém as experiências com esses erros nos levam a melhorar a acurácia da previsão e assim eliminar algumas perdas resultantes da incerteza no processo de decisão. Logo, é importante que os métodos estatísticos de previsão possam não só prever o comportamento futuro de uma série, mas também tentem encontrar algum padrão nos erros obtidos no processo. Dessa forma pode-se aprender com erros do passado e diminuir a imprecisão nas novas previsões no futuro.

Como a série temporal estudada neste trabalho consiste de preços do contrato futuro de açúcar da *NYBOT*, e considerando que o açúcar é uma commodity agrícola, deve-se levar em conta uma sazonalidade natural desse produto que são os momentos de safra e entressafra das colheitas do açúcar.

Sendo o Brasil o maior produtor de açúcar do mundo, épocas de maior e de menor oferta desse produto no país afetam diretamente o comércio no mundo. Por isso, é importante

¹Os preços utilizados como base para construção das séries temporais foram os preços do contrato futuro de açúcar (Sugar 11) do 1º vencimento. Isso porque os contratos mais longos têm uma liquidez menor resultando em uma maior volatilidade do preço, o que pode levar a uma distorção nos dados da série estudada.

verificar se esses fatos afetam diretamente na formação do preço mundial do açúcar. E também verificar se há sazonalidade nos períodos de safra e entressafra da cana-de-açúcar.

Do mesmo modo que vamos analisar se há presença ou não de sazonalidade nos dados estudados, temos que, definido um intervalo de tempo, analisar também se os dados apresentam algum tipo de tendência ao longo desse período. Conseguindo identificar padrões como tendência e sazonalidade no passado, pode-se escolher o método de previsão mais adequado para a construção do modelo, aumentando a precisão dos resultados obtidos.

Para verificar se essas hipóteses são válidas e para poder melhor analisar os dados estudados, serão apresentados conceitos necessários para construção do modelo proposto de previsão dos preços do açúcar de NYBOT. Essas definições são importantes para que possamos usar as ferramentas matemáticas da melhor forma possível.

3.2. Métodos de Previsão

As diversas técnicas de previsão desenvolvidas ao longo dos anos, podem ser divididas basicamente em duas grandes categorias: quantitativas e qualitativas. O diagrama a seguir mostra as variáveis que serão levadas em conta no processo de elaboração do modelo de previsão.

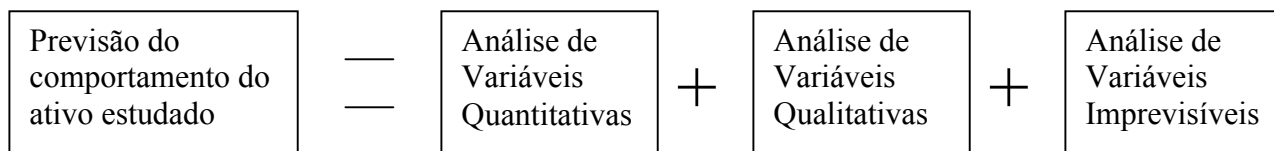


Figura 3.1. – Diagrama do Modelo de Previsão

Sendo que:

Variáveis Quantitativas: Nesse caso, existe informação quantitativa suficiente, ou seja, existe registro de séries históricas dos dados a serem estudados:

- *Séries Temporais*: prever a continuidade de uma série histórica de preços de açúcar do passado no futuro.
- *Causais*: entender como outras variáveis, que não são relacionadas diretamente com a commodity estudada, podem afetar o preço do contrato futuro de açúcar. Por exemplo: preço do álcool, cotação do dólar em reais ou ainda a taxa de juros de curto e longo prazo.

Variáveis Qualitativas: Existe pouca ou nenhuma informação quantitativa disponível, porém há alguma informação qualitativa. Exemplos:

- Prever a safra de cana de açúcar.
- Prever o clima (muita chuva, seca, etc) na época de safra.

Variáveis Imprevisíveis: Existe pouca ou nenhuma informação disponível, mas podem afetar diretamente as séries estudadas. Exemplos:

- Prever a manipulação do preço do açúcar por parte dos produtores, através da formação de cartéis.
- Prever aumento na taxa de importação, por parte do governo.

O Modelo descrito acima foi baseado no “Categorias de Métodos de Previsão e Exemplos de Aplicações” de Makridakis; Wheelwright; Hyndman (1998).

3.2. Séries Temporais

Série temporal é definida como uma seqüência de observações na qual é necessário definir uma variável específica e um período de tempo determinado. Por exemplo: consumo de energia mensal do Estado de São Paulo ou série de preços diários de uma ação negociada na Bolsa de Valores. Analisar uma série temporal é tentar descrever o processo gerado pela seqüência de dados e para prevê-la é necessário representar, através de modelos matemáticos, o seu comportamento no futuro.

Um modelo clássico para séries temporais supõe que a série - Z_t , pode ser escrita como uma soma de três componentes, Tendência, Sazonalidade e Componente aleatória ou residual.

$$z_t = T_t + S_t + \varepsilon_t \quad t=1,2,\dots,N \quad (3.0)$$

Para previsões de curto prazo, como será feito neste trabalho, normalmente utiliza-se a hipótese de que o futuro seja uma continuação do passado, principalmente do mais recente, ou seja, tendências e sazonalidades observadas devem permanecer no futuro.

3.2.1. Sazonalidade - S_t

Sazonalidade são mudanças no comportamento dos dados ao longo do tempo que se repetem em intervalos regulares, geralmente em períodos até um ano. Por exemplo: aumento vendas de passagens aéreas no verão ou aumento nas vendas do comércio em épocas festivas.

Podemos testar a existência de sazonalidade ou não em uma série histórica testando as hipóteses:

$$\begin{aligned} H_0: & \text{Não existe sazonalidade} \\ H_1: & \text{Existe sazonalidade} \end{aligned} \quad (3.1)$$

Os teste utilizados são:

- a) Teste de Kruskal – Wallis (método não paramétrico)
- b) Teste de Friedman (método não paramétrico)
- c) Análise de Variância (método paramétrico)

É importante ressaltar que antes de realizar esses testes é necessário tirar a componente de tendência, quando ela existir, da série temporal.

3.2.2. *Tendência* - T_t

A componente de Tendência é caracterizada por movimentos suaves de declínio ou de crescimento dos valores de uma série temporal e são causadas geralmente por fatores que são medidos durante períodos longos de tempo.

Por exemplo, se construirmos um gráfico com os preços de apartamentos de cidades como Nova Iorque e Londres dos últimos dez anos, podemos verificar que eles mostram uma tendência de alta, ou seja, os preços vêm subindo de forma constante ao longo dos anos.

Para verificar se esta tendência é significativa na série, além da inspeção gráfica, podemos utilizar testes de hipóteses estatísticas.

$$\begin{aligned} H_0: & \text{Não existe tendência} \\ H_1: & \text{Existe tendência} \end{aligned} \quad (3.2)$$

Os teste utilizados são:

- a) Teste de Seqüências (Wald-Wolfowitz)
- b) Teste do Sinal (Cox-Stuart)
- c) Teste baseado no coeficiente de correlação de Spearman

É importante ressaltar que antes de realizar esses testes é necessário tirar a componente de sazonalidade, quando ela existir, da série temporal.

Para um maior aprofundamento dos testes veja Morettin, Tolloi (1987)

3.2.3. Componente aleatória ou residual - ε_t

Supõe-se que ε_t seja componente aleatória ou ruído branco com média $\mu^2_\varepsilon = 0$ e variância σ^2_ε .

3.3. Medidas de Desempenho

Várias medidas podem ser usadas para avaliar um modelo de previsão. Entre as mais importantes estão aquelas que conseguem analisar os erros da previsão.

Sendo z_t o valor da observação no período “t” e $\hat{z}_t$ a previsão da observação para o período “t”, determinado por algum modelo estatístico, o erro de previsão é definido como:

$$\varepsilon_t = z_t - \hat{z}_t \quad (3.3)$$

Em métodos de previsão o erro “e” é uma variável aleatória com média $E(\varepsilon) \approx 0$ ¹ e variância σ^2_ε .

¹ A média dos erros - $E(\varepsilon)$ só é igual a zero para seqüências nos quais os erros não apresentam uma tendência ou viés. Em caso dos erros apresentarem tendências ou viés existem métodos estatísticos para estimá-los, que não serão citados nesse trabalho.

Para se medir a precisão de um método de previsão, são utilizadas algumas medidas de erro, entre elas:

- *Erro absoluto médio* (EAM)

$$EAM = \frac{1}{N} \times \sum_{t=1}^N |\varepsilon_t| \quad (3.4)$$

- *Erro quadrático* (EQ)

$$EQ = \sum_{t=1}^N \varepsilon_t^2 = \sum_{t=1}^N (z_t - \hat{z}_t)^2 \quad (3.5)$$

- *Erro quadrático médio* (EQM)

$$EQM = \frac{1}{N} \times \sum_{t=1}^N \varepsilon_t^2 = \frac{1}{N} \times \sum_{t=1}^N (z_t - \hat{z}_t)^2 \quad (3.6)$$

3.4. Métodos Quantitativos de Previsão

Os métodos quantitativos de previsão visam mostrar como fazer para encontrar um padrão no comportamento dos dados e representá-los através de equações matemáticas, estimando os parâmetros que melhor representem a série histórica.

Segundo Makridakis; Wheelwright; Hyndman (1998), existem três condições que devem ser respeitadas para se fazer uma previsão quantitativa:

1. Informações sobre o passado devem estar disponíveis;
2. Estas informações podem ser quantificadas;

3. Pode-se assumir que alguns aspectos do comportamento passado vão se repetir no futuro (Hipótese da Continuidade - Assumption of Continuity)

A seguir, serão mostrados os métodos de previsão mais usuais do meio acadêmico levados em conta para elaboração do modelo de previsão utilizado. Os métodos estarão dispostos de forma genérica, sem aprofundamento teórico. Para maior detalhamento matemático consultar Morettin, Tolloí (1987); Montgomery; Johnson (1976) e Makridakis; Wheelwright; Hyndman (1998).

3.4.1. Regressão Linear Simples

Muitas vezes uma série temporal pode ser representada como uma função linear (de 1º grau) do tempo, descrita da seguinte forma:

$$y_t = \beta_1 + \beta_2 t + \varepsilon_t \quad (3.7)$$

Como β_1 e β_2 são constantes desconhecidas e “ ε ” é o resíduo com média 0 e variância σ_e^2 então podemos estimar a reta através de

$$\hat{y}_t = b_1 + b_2 t \quad (3.8)$$

onde b_1 é o interceptor da reta; b_2 o coeficiente angular da reta e a diferença entre o valor previsto e o observado é chamado de resíduo.

$$\varepsilon_t = y_t - \hat{y}_t \quad (3.9)$$

Para estimar os valores de β_1 e β_2 pode-se utilizar o *Método dos Mínimos Quadrados* (MMQ), ou seja, acha-se b_1 e b_2 de tal forma que a soma dos quadrados

$$MMQ = \sum_{t=1}^N \varepsilon^2_t = \sum_{t=1}^N (y_t - b_1 - b_2 t)^2 \quad (3.10)$$

seja minimizada.

Na figura 3.1 fica claro como funciona o processo de previsão através da reta de regressão. Com os pontos $t = 1, 2, \dots, N$ constrói-se a reta de regressão $y_t = b_1 + b_2 t$ e para prever um valor $\hat{y}(t = N + h)$ sendo $h > 0$ tem-se $\hat{y}_{t=(N+h)} = b_1 + b_2 \times (N + h)$.

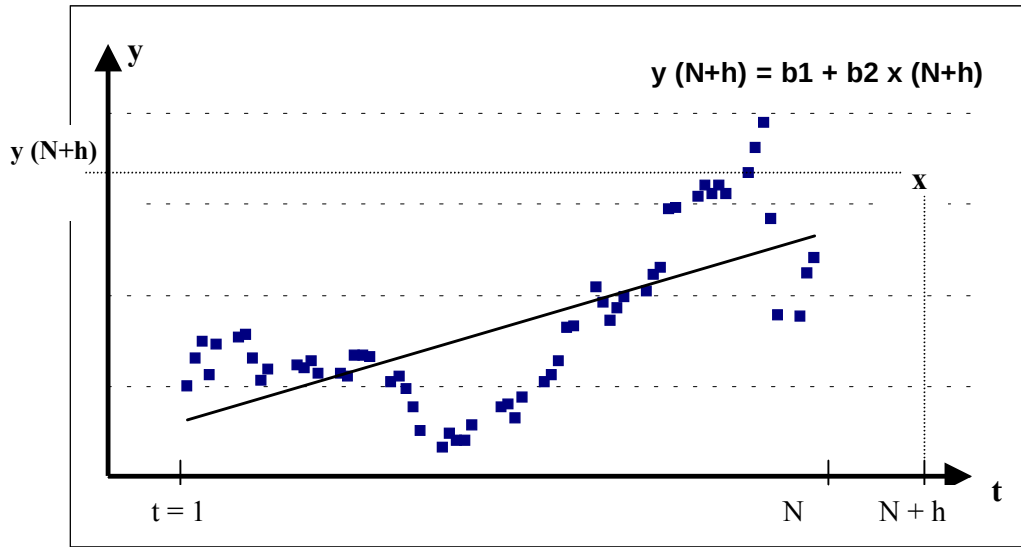


Figura 3.1. – Reta de Regressão

3.4.2. Média Móvel Simples (MMS)

O Método *MMS* consiste em utilizar a média aritmética das r observações mais recentes

$$M_t = \frac{Z_t + Z_{(t-1)} + \dots + Z_{(t-r+1)}}{r} \quad (3.11)$$

Esse método é chamado de média móvel porque, a cada período, a nova observação substitui a r -ésima para o cálculo da média.

A previsão do valor futuro da série é a última média calculada, isto é,

$$\hat{z}_{(t+h)} = M_t \quad \text{para } h = 1, 2, 3, \dots \quad (3.12)$$

As propriedades do método MMS dependem do valor de r utilizado para se determinar a média. Se r for um número grande, previsão reagirá mais devagar às mudanças mais recentes. Se r for um número pequeno, a previsão reagirá mais rapidamente às mudanças mais recentes da série. No caso extremo de $r = 1$, a previsão será o último valor da série.

Assim o valor de r deve ser de tal forma que minimize o erro de previsão (3.3), para isso utiliza-se o Método dos Mínimos Quadrados (MMQ).

$$MMQ = \sum_{t=1}^N \varepsilon^2_t = \sum_{t=1}^N (z_t - \hat{z}_t)^2 \quad (3.13)$$

Segundo Morettin, Tolloi (1987), as principais vantagens desse método são:

- i) Fácil aplicação;
- ii) É aplicável para um pequeno número de observações;
- iii) Permite uma grande flexibilidade devido à variação de r de acordo com o padrão da série.

E suas principais desvantagens são:

- i) Necessidade de armazenar pelo menos $(r-1)$ observações;
- ii) Dificuldade de achar o valor de r .

3.4.3. Alisamento Exponencial Simples (AES)

O Alisamento Exponencial Simples (*AES*) é provavelmente o método mais utilizado para previsão de valores no futuro imediato, pela sua simplicidade teórica e prática e também pela sua razoável precisão.

O *AES* é similar ao modelo da média móvel (3.4.2), porém em vez de considerar todos os r valores da série igualmente, ele dá um peso maior às observações mais recentes. Outra diferença é que são utilizados todos os valores da série histórica para se achar a previsão, com coeficientes de ponderação que decrescem exponencialmente, mostrados pela expressão seguinte.

$$\hat{z}_t = \alpha \times z_{(t-1)} + (1-\alpha) \times \hat{z}_{(t-1)} \quad (3.14)$$

Ou

$$\hat{z}_t = \alpha \times z_{(t-1)} + (1-\alpha)^2 \times z_{(t-2)} + (1-\alpha)^3 \times z_{(t-3)} + \dots \quad (3.15)$$

Onde :

α = Constante de Alisamento $0 \leq \alpha \leq 1$;

z_t = Valor “real” do dado no instante “t”;

$\hat{z}_t$ = Valor “previsto” do dado no instante “t”.

Pode-se perceber que nesse método, a previsão é uma média ponderada entre o valor “real” do instante “t-1” e o valor da “previsão” também do instante “t-1”. Sendo que quanto maior for a constante de alisamento α , maior será o peso dado aos valores mais recentes.

Uma grande vantagem desse modelo em relação ao *MMS* está na necessidade de armazenar somente z_t , $\hat{z}_t$ e α . A principal desvantagem está na dificuldade para encontrar o valor da constante de alisamento que melhor descreva o modelo.

Para encontrar a “melhor” constante de alisamento α , utiliza-se o Método dos Mínimos Quadrados para minimizar os erros quadráticos

$$MMQ = \sum_{t=1}^N \varepsilon_t^2 = \sum_{t=1}^N (z_t - \hat{z}_t)^2 = \sum_{t=1}^N [z_t - (\alpha \times z_{(t-1)}) - (1 - \alpha) \times \hat{z}_{(t-1)}]^2 \quad (3.16)$$

Usualmente varia-se o valor da variável α até encontrar o ótimo da função, sem empregar métodos de otimização mais sofisticados.

3.4.4. Alisamento Exponencial Sazonal

Este método é também chamado de Alisamento Exponencial Sazonal de Holt-Winters, sendo adequado para seqüência de dados que apresentem sazonalidade.

Suponha que uma série temporal com sazonalidade possa ser escrita da seguinte forma:

$$z_t = \mu_t + T_t + S_t + \varepsilon_t \quad (3.17)$$

Onde:

μ_t = componente permanente

T_t = componente de tendência linear

S_t = fator sazonal

ε_t = componente de erro aleatório

Considerando que a série temporal tenha uma sazonalidade de s períodos, ou seja, no caso de uma sazonalidade anual tem-se $s = 12$, e que se a componente de tendência não for

necessária pode ser tirada da equação (3.17) podemos estimar os parâmetros anteriores da seguinte forma:

$$\hat{S}_t = D \times (z_t - \hat{z}_t) + (1 - D) \times \hat{S}_{t-s} \quad 0 < D < 1 \quad (3.18)$$

$$\hat{z}_t = A \times (z_t - \hat{S}_{t-s}) + (1 - A) \times (\hat{z}_{t-1} + \hat{T}_{t-1}) \quad 0 < A < 1 \quad (3.19)$$

$$\hat{T}_t = C \times (\hat{z}_t - \hat{z}_{t-1}) + (1 - C) \times \hat{T}_{t-1} \quad 0 < C < 1 \quad (3.20)$$

No qual D, A e C são constantes de alisamento que obtidas pelo *MMQ*. Deste modo às previsões para os valores futuros da série são:

$$\hat{z}_{t+h} = \hat{z}_{t+h} + \hat{T}_t + \hat{S}_{t+h-s} \quad (3.21)$$

A desvantagem desse método está em achar as constantes de alisamento que melhor descrevem a função.

3.5. Métodos Causais de Previsão

3.5.1. Regressão Linear Simples

A construção de uma reta de regressão linear é similar ao item 3.4.1, porém é feita com uma finalidade diferente. Em vez de elaborar uma reta de regressão com objetivo de prever o comportamento futuro, vamos achar uma relação entre duas variáveis através de uma reta de regressão linear simples.

A construção da equação que deve ser encontrada é feita da mesma forma que em (3.8); (3.9) e (3.10), porém em vez do tempo, temos a outra variável que queremos relacionar. Depois que a equação é traçada pelo *MMQ* devemos analisar a adequação do modelo através de uma análise de variância aplicada à regressão e uma análise de resíduos.

A análise de variância é feita hoje em dia por qualquer aplicativo estatístico. Para isso é preciso conhecer algumas variáveis de saída desse método que ajudam a interpretar os resultados da forma mais correta.

Dentre elas está o erro-padrão (ou variância residual), que mede a dispersão dos valores em volta da reta de regressão, e é estimada da seguinte forma:

$$S^2_{Res} = \frac{\sum (y_i - \hat{y}_i)^2}{n - 2} \quad (3.23)$$

Na mesma linha encontra-se a variância devido ao modelo de regressão:

$$S^2_{Reg} = \sum (\hat{y}_i - \mu_{yi})^2 \quad (3.24)$$

Outra importante variável de saída que é utilizada para se medir o quanto duas variáveis estão relacionadas entre si é o coeficiente de correlação ρ , que é obtido da seguinte forma:

$$\rho_{x,y} = \frac{Cov(x,y)}{\sigma_x \times \sigma_y} \quad (3.25)$$

Onde:

$$Cov(x,y) = \frac{1}{N} \times \sum_{i=1}^n (x_i - \mu_x) \times (y_i - \mu_y)$$

μ é a média dos valores

σ é o desvio padrão dos valores

O valor de ρ pode variar entre -1 até +1. Se o valor da correlação for muito alto, próximo de +1, significa que quando uma variável aumenta a outra também aumenta. Se o valor da correlação for muito baixa, próximo de -1, significa que quando uma variável diminui a outra aumenta e vice-versa. Se a correlação for igual a zero, significa que as duas variáveis não têm nenhuma relação linear entre si.

Uma informação que decorre da correlação é o coeficiente R^2 , que nada mais é do que a correlação ao quadrado.

$$R^2 = (\rho_{x,y})^2 \quad (3.26)$$

Logo, pode-se verificar se a equação descreve de forma adequada as variáveis x e y , utilizando-se do teste de F-Snedecor ou “teste-F”, que testa a significância do modelo.

$$F = \frac{S^2_{Reg}}{S^2_{Res}} \quad (3.27)$$

Quanto maior for o resultado **(3.27)** mais adequada será a equação ao modelo de regressão.

Na análise de resíduos, uma forma rápida de verificar a aderência da equação é construir um gráfico com os resíduos padronizados (divisão dos resíduos pelo seu desvio padrão) e verificar se esses estão distribuídos na forma de uma curva normal.

3.5.2. Regressão Múltipla

Neste caso, a resposta é uma função de mais de uma variável e admite-se que esta função seja do tipo:

$$y = \beta_0 + \sum_{i=1}^k (\beta_i \times x_i) + \varepsilon_i \quad ; \quad \text{sendo } k \text{ o número de variáveis} \quad (3.28)$$

que será estimada pela equação

$$\hat{y} = b_0 + \sum_{i=1}^k b_i \times x_i \quad ; \quad \text{sendo } k \text{ o número de variáveis} \quad (3.29)$$

Como foi feito na regressão linear simples, para se achar os coeficientes b_k utiliza-se o método de minimizar os erros quadráticos através do *MMQ*.

Uma maneira de incluir dados qualitativos na análise de regressão múltipla é através das variáveis binárias (dummy variables). Sua criação é simples: quando ocorre o evento em questão seu valor é 1, caso contrário assume valor 0. Esse recurso pode ser usado no modelo em casos como se a colheita de cana for boa a variável é igual a 1, caso contrário igual a 0. Para verificar a adequação da equação emprega-se a análise de variância e de resíduos de forma similar à descrita anteriormente.

A dificuldade em encontrar a melhor equação consiste nas escolhas das variáveis que entrarão no modelo de regressão, pois nem todas as variáveis que estão envolvidas no processo terão uma influência significativa na saída final do modelo.

3.6. Análise dos Resíduos

No processo de construção de um modelo de previsão, é importante estudar o comportamento dos erros de previsão e analisar se existe alguma tendência ou sazonalidade dos resíduos ao longo do tempo ou se esses têm um comportamento aleatório.

Para isso será mostrado a teoria de um dos métodos mais utilizados para avaliar se duas amostras tem distribuição semelhantes.

3.6.1. *Teste Kolmogorov – Smirnov*

O principio do teste baseia-se na comparação da curva de frequência acumulada dos dados com a curva da função de distribuição normal padrão. Com isso testa-se as hipóteses

H_0 = Os dados apresentam distribuição normal

H_1 = Os dados não apresentam distribuição normal

O teste rejeita H_0 quando a máxima diferença entre um valor de x e a distribuição normal é significativa a um determinado nível de confiança. Matematicamente pode ser escrito da seguinte forma

$$D_{\max} = \max(|F_x - G_x|) \quad (3.30)$$

onde F_x é função acumulada dos dados e G_x é a função acumulada normal.

Se $D_{\max} > D_{\text{critico}}$ rejeitamos H_0 .

Por ser relativamente fácil de se aplicar e por apresentar resultados claros e confiáveis, o teste Kolmogorov - Smirnov é utilizado em diversas áreas de estudo em que se precisa comparar duas amostras de dados.

4. Redes Neurais Aplicados na Previsão de Séries Temporais

Nos últimos anos, uma série de fatores tem permitido um avanço considerável na utilização de técnicas matemáticas e computacionais para a criação de modelos a partir de bases de dados. Algumas técnicas, apesar de conceitualmente apresentadas há quatro ou cinco décadas, como as redes neurais (MacCulloch and Pitts, 1943), não podiam ser experimentadas de maneira efetiva, pois não havia capacidade computacional para tanto. Com os avanços tecnológicos crescentes na área de informática, modelos matemáticos mais complexos como redes neurais e algoritmos genéticos já podem ser utilizados facilmente com o auxílio de aplicativos modernos tais como o *MatLab* e *Cran-R*.

As Redes Neurais Artificiais (RNA), para previsão de séries temporais, diferenciam-se dos modelos tradicionais de previsão por serem modelos não paramétricos envolvendo algoritmos de aprendizado. Esses algoritmos são baseados em uma modelagem da estrutura de interconexões similar ao cérebro humano, com o objetivo de reconhecer padrões de comportamento de uma série, e assim conseguir prever, da maneira mais eficiente possível, valores futuros da mesma. Características particulares de uma série de tempo, tais como sazonalidade, tendência e ciclo, podem ser aprendidas por uma RNA, possibilitando, que em previsões futuras, essas características já estejam incorporadas nos resultados finais.

Os estudos de redes neurais aplicados na área das finanças não objetivam substituir métodos tradicionais de previsão de dados, visam sim se tornar mais uma ferramenta complementar que ajuda a fomentar os estudos matemáticos nesse amplo campo que é mercado financeiro.

Este capítulo apresenta as ferramentas necessárias para construção de um modelo de previsão utilizando redes neurais, que posteriormente serão aplicadas na previsão de uma commodity agrícola. É importante realçar que o uso de RNA, na previsão de preços de ativos do mercado financeiro, já é objeto de estudo desde a década de 80, tendo evoluído significativamente e se tornado cada vez mais difundido entre profissionais e acadêmicos do meio (Refenes, 1995).

4.1. Introdução a Redes Neurais Artificiais (RNA)

Desde o início do século passado tenta-se construir uma máquina que consiga funcionar similarmente ao cérebro, com isso os estudos no campo da Inteligência Artificial vêm progredindo junto com a modernização de processadores dos computadores existentes. Desde então, diversos cientistas se aprofundaram nesse assunto resultando em teorias sólidas, que serão parcialmente expostas a seguir, que não ficam limitadas ao universo acadêmico, mas também são usualmente utilizadas no mundo real dos negócios.

As Redes Neurais Artificiais fundamentam-se nos estudos sobre estrutura do cérebro humano para tentar imitar sua forma inteligente de processar informação. Como o sistema nervoso é formado por um número muito grande de neurônios (casa dos bilhões), o cérebro humano consegue executar rapidamente certas funções que os computadores não conseguem, isso porque os neurônios cerebrais trabalham de forma paralela e não linear, sendo conectados por uma complexa e eficiente rede de sinapses. Deste modo o sistema nervoso consegue suprir deficiências de sua velocidade de processamento, já que é relativamente baixa se comparada a computadores tradicionais.

Segundo Refenes (1995), existem pelo menos dois principais motivos para a continuada pesquisa em redes neurais artificial:

- É sabido que as redes neurais são um instrumento poderoso para modelar e entender o comportamento humano cognitivo. Esse motivo tem estimulado pesquisas em neurociência, anatomia, psicologia e ciência biológicas para que se consiga uma melhor visão do comportamento do cérebro humano.
- Os modelos de RNA são reconhecidos por terem um alto poder de reconhecer padrões (“*patterns*”), e com isso conseguem superar outras técnicas tradicionais de modelagens. Esse motivo tem atraído

pesquisadores de diversas áreas, incluindo matemática, economia e finanças.

4.2. Histórico das RNAs

As primeiras informações mencionando redes neurais na computação são de 1943, em artigos publicado pelo neurofisiologista McCulloch e o matemático Pitts, em que sugeriam a construção de uma máquina inspirada no cérebro humano que simulava o comportamento de um neurônio real. Esse modelo serviu como base para a maioria dos modelos de redes neurais existentes.

Muitos outros artigos e livros surgiram desde então, porém, por um longo período de tempo, pouco resultado foi obtido. Em 1949 Donald Hebb escreveu um livro com intitulado de "The Organization of Behavior" (A Organização do Comportamento) que perseguia a idéia de que o condicionamento psicológico clássico está presente em qualquer parte dos animais pelo fato de que esta é uma propriedade de neurônios individuais. Suas idéias não eram completamente novas, mas Hebb foi o primeiro a propor uma lei de aprendizagem específica para as sinapses dos neurônios. Este primeiro e corajoso passo serviu de inspiração para que muitos outros pesquisadores perseguissem a mesma idéia. Embora muito tenha sido estudado e publicado nos anos que seguiram (1940-1950), estes serviram mais como base para desenvolvimento posterior que para o próprio desenvolvimento.

Também proveniente deste período de tempo foi a construção do primeiro neuro computador, denominado Snark, por Marvin Minsky, em 1951. O Snark operava com sucesso a partir de um ponto de partida técnico, entretanto, ele nunca executou qualquer função de processamento de informação interessante, mas serviu de inspiração para as idéias de estruturas que o sucederam.

Em 1956 no "Dartmouth College" nasceram os dois paradigmas da Inteligência

Artificial, a simbólica e o conexionista. A Inteligência Artificial Simbólica tenta simular o comportamento inteligente humano desconsiderando os mecanismos responsáveis por tal.

Já a Inteligência Artificial Conexionista acredita que construído um sistema que simule a estrutura do cérebro, este sistema apresentará inteligência, ou seja, será capaz de aprender, assimilar, errar e aprender com seus erros.

O primeiro neuro computador a obter sucesso (Mark I Perceptron) surgiu entre 1957 e 1958, criado por Frank Rosenblatt, Charles Wightman e outros. Devido à profundidade de seus estudos, suas contribuições técnicas e de sua maneira moderna de pensar, muitos o vêem como o fundador da neuro computação na forma em que a temos hoje. Seu interesse inicial para a criação do *perceptron* era o reconhecimento de padrões.

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

Em meados dos anos 80, o livro “*Parallel Distributed Processing*” escrito por Rumelhart e McClelland (1986) causou grande impacto na computação e ciências biológicas aumentando significativamente o número de pesquisas sobre o assunto. Tanto é que um ano depois (1987) ocorreu em São Francisco a primeira conferência de redes neurais em tempos modernos, a IEEE “*International Conference on Neural Networks*”, e também foi formada a *International Neural Networks Society* (INNS).

Com o passar dos anos, o assunto ganhou bastante credibilidade não só de acadêmicos, mas também por profissionais das mais diversas áreas de conhecimento que passaram a utilizar modelos de redes neurais para melhorar suas capacidades analíticas. Um caso

clássico de utilização de redes neurais com relativo sucesso na área financeira foi com Bergenson e Wunsch, que através de um sistema híbrido baseado em RNA conseguiram gerar um retorno de 660% em 25 meses a partir de uma aplicação de US\$10.000,00 (veja Bergenson e Wunsch II, 1996).

4.3. Componentes de uma RNA

Como foi levantado anteriormente, as redes neurais tendem a imitar o processo de funcionamento do sistema nervoso, logo o neurônio tem um papel chave nesse processo.

Abaixo temos uma comparação de um neurônio do ponto de vista biológico e do ponto de vista computacional (as funções de uma componente de uma RNA equivalente serão descritas posteriormente).

Componentes de uma célula Neural	Funções Biológicas	Componentes de um RNA equivalente
Dendrito	Recebe os estímulos de outros neurônios	Sinapses
Corpo da célula	Coleta e combina as informações vindas de outros neurônios	Somador
Axônio	Transmite estímulos para outra célula	Função de Ativação

Tabela 4.1. Comparação Neurônio Biológico e Artificial.

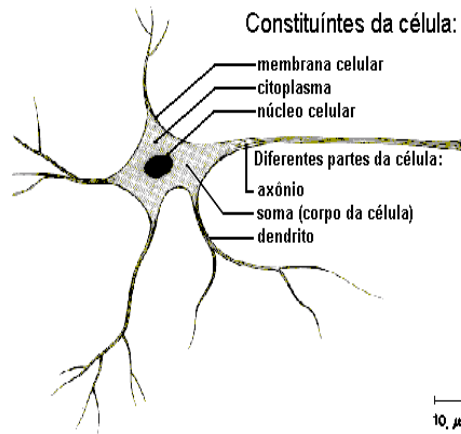


Figura 4.1: Constituintes de uma célula neural

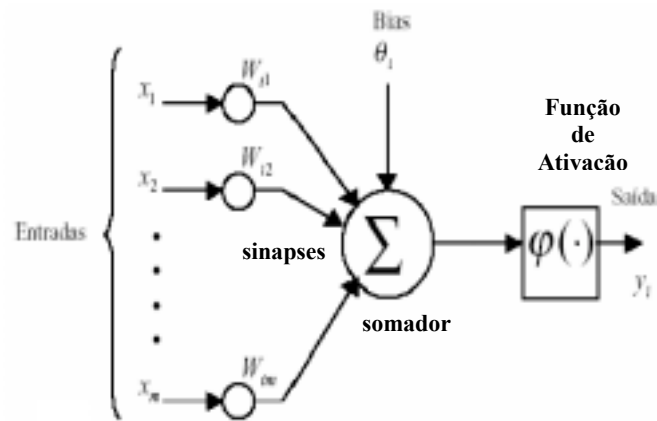


Figura 4.2: Modelo de um Neurônio Artificial (Haykin, 2000)

Funções dos componentes de uma RNA:

- a) As *sinapses* de uma RNA, como os dendritos, recebem os estímulos ou dados de entradas de um neurônio. Cada valor de entrada é multiplicado pelo seu respectivo peso conforme a equação

$$u_j = \sum_{i=1}^n w_{ij} \times x_i \quad (4.1)$$

onde:

u_j é o sinal produzido por uma sinapse j

w_{ij} é o peso correspondente da entrada i na sinapse j

x_i é o valor de entrada i

n é o número de sinapses no neurônio

- b) O *somador* é responsável por somar os sinais de entrada do neurônio, já ponderados pelos respectivos pesos sinápticos. No modelo neuronal da figura 4.2 ao somador é aplicado um viés (*bias*) externo nos sinais de

entrada, representado por θ_j . Esse viés tem o objetivo de aumentar ou diminuir os valores da entrada na função de ativação.

$$v_j = u_j + \theta_j \quad (4.2)$$

- c) A *função de ativação* é a função que determina o valor de saída após a iteração dos neurônios entre si. Essa função é utilizada para restringir a amplitude dos valores de saída de um neurônio ou em muitos casos, a função de saída é uma função de limiar que só emite sinal quando o seu estado ultrapassa um certo limite. A função de ativação é representada pela função

$$y_t = \varphi(v_j) \quad (4.3)$$

onde

y_t é a saída do neurônio;

$\varphi(.)$ é a função de ativação que é determinada pelos sinais produzidos pelas sinapses u_j e pela viés externo θ_j

Os valores de ativação existentes na rede podem ser discretos, por exemplo, assumindo os valores $\{0,1\}$ ou $\{-1,0,1\}$, como também podem ser contínuos, assumindo valores no intervalo $[0,1]$ ou $[-1,1]$.

Essa transformação que função de ativação $\varphi(.)$ realiza pode ser feita matematicamente por funções de ativação não lineares, como as seguintes. (Refenes, 1995):

- i. **Função de limiar** (“*Hard Limiter*”): Produz valores em $\{0,1\}$, dependendo se o total de valores de entrada for maior que um valor limite (*threshold value*). Essa função é utilizada por ser relativamente simples.

$$\varphi(a) = \begin{cases} 1, & \text{se } a \geq 0 \\ 0, & \text{se } a < 0 \end{cases} \quad (4.4)$$

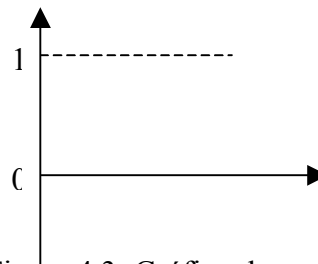


Figura 4.3: Gráfico de uma função limiar

- ii. **Função Sigmóide:** Essa função é a mais utilizada na construção de uma função de ativação de uma rede neural principalmente por ser mais complexa, diferenciável e por se comportar bem em várias aplicações.

$$\varphi(a) = \frac{1}{1 + e^{-a}} \quad (4.5)$$

onde a é o parâmetro de inclinação da função Sigmóide.

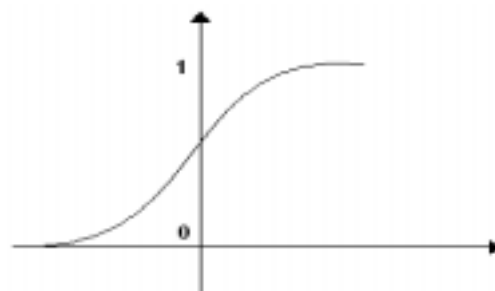


Figura 4.4: Gráfico de uma função Sigmóide

- iii. **Função tangente hiperbólica.** A função tangente hiperbólica é dada pela seguinte fórmula

$$\varphi(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}} \quad (4.6)$$

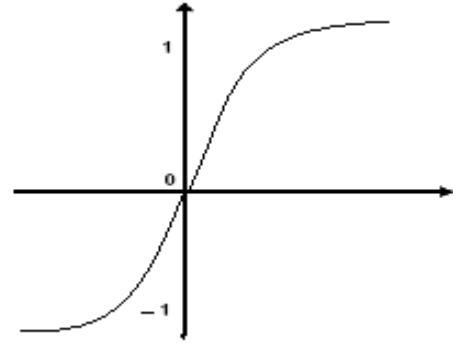


Figura 4.5: Gráfico de uma função tangente hiperbólica

Nas últimas páginas, mostramos como se cria um neurônio artificial se baseando em um neurônio biológico natural, e ao mesmo tempo foi-se mapeando matematicamente as funções de um cérebro humano.

Nos próximos itens mostraremos como vários neurônios artificiais estão conectados e distribuídos em uma rede neural, ou seja, apresentaremos características e propriedades de rede. Após isso, é possível verificar como funciona a dinâmica de funcionamento da rede, como podemos treinar essa rede (processo de aprendizado) para tentar se comportar como uma mente humana.

4.4. Arquitetura das Redes Neurais Artificiais

Segundo Braga, Ludermir e Carvalho (2000), definir a arquitetura de uma rede neural consiste em determinar sua organização estrutural: número de camadas da rede e de neurônios em cada camada, e tipo de conexão entre os neurônios e a topologia da rede.

Normalmente, as redes neurais são estruturadas em camadas, com um ou mais neurônios. Todas elas possuem camada de entrada, que recebe os sinais de entrada, e camada de saída, que armazena o resultado final da rede. As camadas intermediárias (*hidden layers*) localizam-se entre estas duas conforme Fig.4.6.

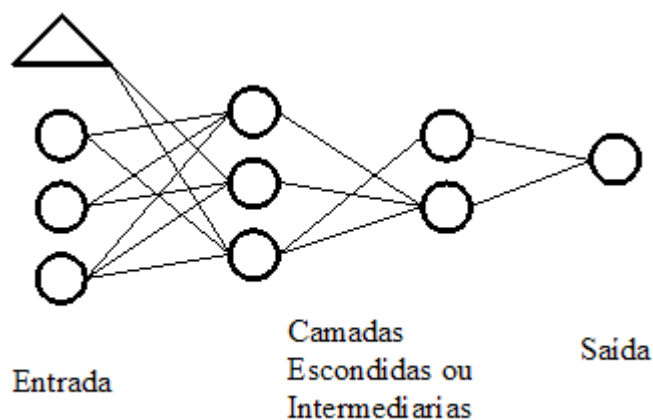


Figura 4.6. Modelo de rede neural acíclica com 2 camadas escondidas.

Os neurônios artificiais também são conhecidos como “nós”. Os nós de entrada contêm as informações para a rede neural processar, os nós intermediários não são diretamente observáveis, pois estão na camada escondida e são responsáveis pelas características não lineares da rede. Já os nós de saída entregam os valores de acordo com as características da rede.

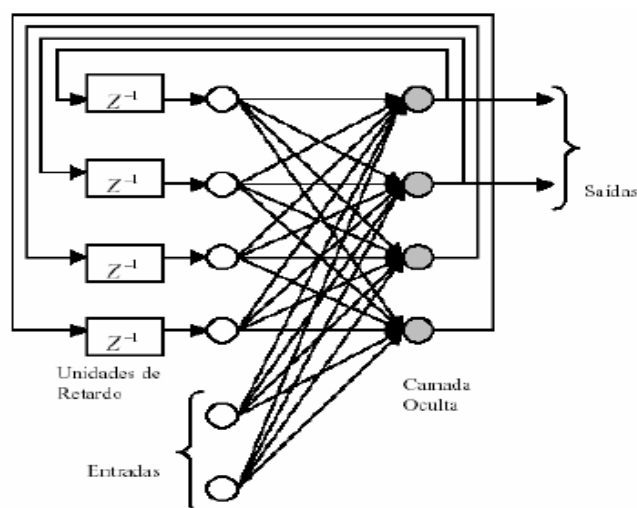


Figura 4.7. Modelo de rede neural cíclica (Haykin, 2000)

Quanto ao tipo de conexão, as redes neurais podem ser do tipo *feedforward* (acíclica) ou recorrentes (cíclica). A cíclica diferencia-se da acíclica por possuir pelo menos um link de realimentação (feedback), pelo qual a saída de algum neurônio de uma camada k é utilizada como entrada para um neurônio de camada de ordem menor ou igual a k , vide Fig. 4.7.

As redes neurais podem ser completamente conectadas, quando todos os neurônios estão conectados entre si, ou simplesmente não serem completamente conectados se pelo menos dois neurônios não apresentam ligações entre si. A Figura 4.6 ilustra uma rede neural acíclica completamente conectada.

4.5. Aprendizado em uma RNA

O principal objetivo do processo de aprendizagem de uma rede neural artificial é formar associações através de padrões observados nas séries de dados de entrada e com isso melhorar o desempenho da rede. Esse aprendizado é realizado na prática através de um processo iterativo que visa ajustar os valores dos pesos sinápticos da entrada, com a finalidade de fazer com que os valores da saída seja o mais próximo possível do desejado.

Com isso, pode-se perceber que o processo de aprendizagem é um elemento crucial para que uma rede consiga obter resultados significativos, pois é necessário que a rede “aprenda” regras, identificando padrões nas séries de dados, e não “decore” exemplos que já foram vistos anteriormente.

Três conceitos são importantes para melhor compreender o processo de aprendizagem:

- **Algoritmo de treinamento ou algoritmo de aprendizado:** conjunto de procedimentos (regras) utilizados para ajustar os parâmetros (pesos

sinápticos) das redes neurais, de maneira que possam realizar determinada função;

- **Número de épocas:** número de vezes que os padrões de treinamento serão apresentados às redes neurais, a fim de que se faça a atualização dos pesos;
- **Taxa de aprendizado:** controla a intensidade das alterações dos pesos. Uma alta taxa de aprendizado acelera o processo, mas pode reduzir a capacidade de generalização da rede neural.

Existem muitos tipos de algoritmos de aprendizado específicos para determinados modelos de redes neurais, estes algoritmos diferem entre si principalmente pelo modo como os pesos são modificados. Dentre esses é importante detalhar dois processos de aprendizado:

4.5.1. Aprendizado Supervisionado

Nesse tipo de aprendizado para cada valor inicial de entrada existe um valor de saída conhecido, ou seja, o modelo é sempre alimentado por um par de valores: valor de entrada e valor de saída. O objetivo é minimizar a diferença entre o valor atual e o valor desejado (chamado de erro ou resíduo) através de uma retro-alimentação do erro alterando os pesos sinápticos até que esse o resíduo seja menor do que um valor pré-estabelecido. Esse processo de aprendizado é feito através de um algoritmo de otimização que minimize o erro quadrático médio (EQM) da rede. O EQM é definido como o quadrado da diferença do valor desejado ($target - t_k$) pela a atual saída da rede ($output - y_k$) para 1 específico dado de entrada.

$$E = \frac{1}{p} \sum_{k=1}^p (t_k - y_k)^2 \quad (4.7)$$

onde p é a quantidade de dados que serão treinados na série.

A minimização do erro (E) é baseada no método dos gradientes, esses métodos consistem em mudar os pesos W_{ij} depois de cada passo de treinamento de toda a amostra de dados. Essas mudanças dos pesos ocorrem na direção contrária do passo da função.

$$W_{(t+1)} = W_t - \varepsilon \frac{\partial E}{\partial W} \quad (4.8)$$

onde ε é um parâmetro definido como tamanho do passo do gradiente ou taxa de aprendizado (*learning rate*)

Um segundo método é chamado “gradiente estocástico”. Essa abordagem consiste em mudar os pesos depois de cada apresentação de treinamento dos dados usando uma função parcial relativa a esse passo

$$W_{(t+1)} = W_t - \varepsilon \frac{\partial E_k}{\partial W} \quad (4.9)$$

4.5.2. Aprendizado Não – Supervisionado

Nesse tipo de aprendizado a rede não recebe nenhum tipo de valor desejado de saída para calibrar o modelo. Ele se orienta somente pelos dados de entrada e tem como objetivo capturar regularidades nos padrões dos dados de entrada sem interferência externa.

4.6. Classificação dos Modelos

Existem muitos modelos de redes neurais, e uma quantidade razoável de publicações que se dedicam a classificá-los. Dentre esses modelos, mostraremos alguns e depois detalharemos o modelo de rede mais utilizado na neuro computação.

- i. Perceptron - proposto por F. Rosenblatt em 1957;
- ii. Perceptron de Multi-Camadas (“*Multi Layer Perceptron - MLP*”) popularmente conhecido por Backpropagation - proposto por P. Werbos, D.Parker, D.Rumelhart em 1974;
- iii. Classificador de Carpenter-Grossberg (Sistema ART) - proposto por G.Carpenter, S.Grossberg em 1978;
- iv. Rede de Kohonen - proposto por T. Kohonen em 1980;
- v. Rede de Hopfield - proposto por J. Hopfield em 1982.

4.7. Modelos de Redes MLP (Backpropagation)

As redes MLP são redes acíclicas (*feedforward*) com pelo menos uma camada de neurônios intermediários (*hidden layer*). O algoritmo de treinamento mais popular aplicado às redes perceptron de Multi-Camadas é o backpropagation, que segue o paradigma de aprendizado supervisionado (4.5.1.), no qual são fornecidos valores de entrada às redes neurais com as respectivas saídas desejadas.

O algoritmo Backpropagation consiste, basicamente, em determinar as variações nos pesos sinápticos da rede neural, tendo como objetivo minimizar o erro obtido na saída, através do aprendizado dos valores de entrada e valores pretendidos. Seu principal diferencial em relação a outros algoritmos está na capacidade de assimilação dos erros pela camada intermediária. Esse método é um algoritmo de gradiente descendente que minimiza o erro quadrático total

$$E = \frac{1}{p} \sum_{k=1}^p \sum_i (t_k^i - y_k^i)^2 \quad (4.10)$$

onde i varia por toda saída da rede MLP.

Por meio desse algoritmo, a aprendizagem é realizada em duas etapas. Na primeira, a *forward-propagation* (passo para frente), os valores de saída da rede são calculados a partir dos valores de entrada fornecidos. Na segunda, a *backward-propagation* (passo para trás), os pesos associados a cada conexão são atualizados conforme as diferenças entre os valores de saída obtidos e os valores desejados, da última camada à camada de entrada. A figura 4.8 mostra de maneira esquemática o funcionamento da rede MLP.

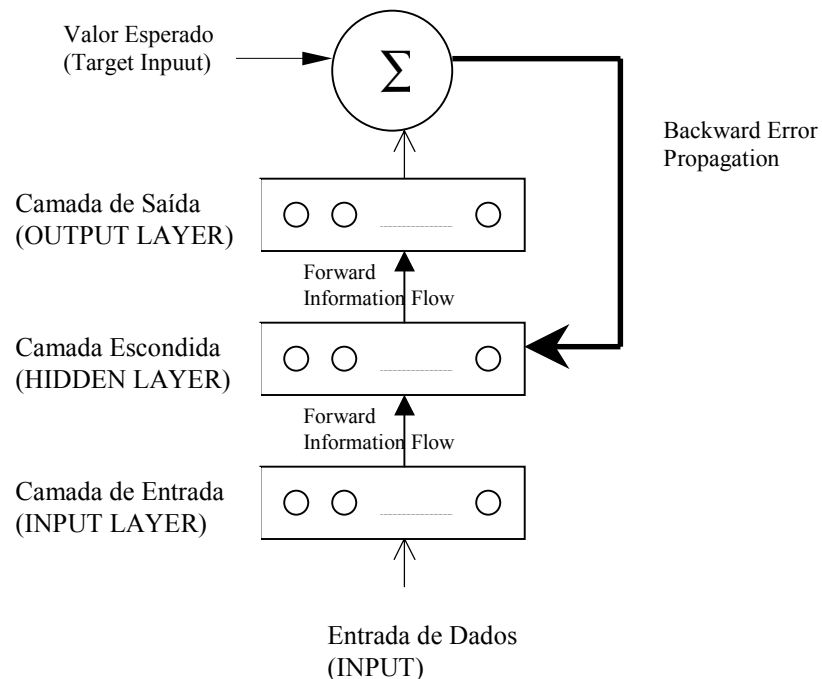


Figura 4.8. Rede MLP (Fu 1994).

Matematicamente essa mudança dos pesos através da comparação das saídas com os valores desejados é possível através da derivada parcial do erro relacionada ao peso. O

k – *ésimo* exemplo presente na rede é definido como $\{x^i_k\}$. As derivadas parciais são feitas utilizando a regra da cadeia e o peso do neurônio i vindo do neurônio j é atualizado pela seguinte fórmula:

$$\Delta W_{ij} = -\varepsilon \sum_{k=1}^p \frac{\partial E_k}{\partial W_{ij}} = -\varepsilon \sum_{k=1}^p \frac{\partial E_k}{\partial V^i_k} \frac{\partial V^i_k}{\partial W_{ij}} = -\varepsilon \sum_{k=1}^p \delta^i_k x^i_k \quad (4.11)$$

onde ε é o tamanho do passo na direção oposta do gradiente.

Para o k – *ésimo* exemplo, a mudança do peso seria:

$$\Delta W_{ij} = -\varepsilon \frac{\partial E_k}{\partial W_{ij}} = -\varepsilon \delta^i_k x^i_k \quad (4.12)$$

O fator δ^i_k é computado usando os fatores δ^i_k da próxima camada, propagando assim o erro passo para trás (*backward*).

Nas saídas teremos:

$$\delta^i_k = -\varepsilon \frac{\partial E_k}{\partial v^i_k} = -\frac{\partial E_k}{\partial x^i_k} \frac{\partial x^i_k}{\partial v^i_k} = (d^i_k - x^i_k) f'(v^i_k) \quad (4.13)$$

Na camada escondida o fator δ^i_k é:

$$\delta^i_k = \frac{\partial E_k}{\partial v^i_k} = -\frac{\partial E_k}{\partial v^h_k} \frac{\partial v^h_k}{\partial v^i_k} \quad (4.14)$$

E como

$$\frac{\partial E_k}{\partial v_k^h} = \delta_k^h \quad e \quad \frac{\partial v_k^h}{\partial v_k^i} = w_{hi} f'(v_k^i) \quad (4.15)$$

Obtém-se:

$$\delta_k^i = \frac{\partial E_k}{\partial v_k^i} = -f'(v_k^i) \sum_h w_{hi} \delta_k^h \quad (4.16)$$

Normalmente se usam duas versões do algoritmo, a versão estocástica atualiza os pesos depois de cada apresentação de um exemplo da rede e a versão por época (*epoch ou off-line*) que atualiza os pesos depois de apresentar toda a série de treinamento.

4.7.1. Algoritmo de treinamento

O algoritmo de treinamento baseado em retropropagação trabalha normalmente com o método de descida por gradiente. Para a maioria dos problemas práticos, esse método é muito lento. Assim, ao longo do tempo foram sendo criados diversos algoritmos de alto desempenho para o treinamento de redes neurais com retropropagação, que convergem entre dez e cem vezes mais rapidamente que o algoritmo de descida por gradiente tradicional.

Estes algoritmos de alta performance podem ser divididos basicamente em duas categorias segundo Baestaens, Bergh e Wood (1994):

1. Os que usam técnicas heurísticas, desenvolvidas a partir de uma análise cuidadosa do algoritmo de descida por gradiente básico;
2. Os que utilizam métodos de otimização numérica, para acelerar a convergência dos pesos da rede.

Dentro do primeiro grupo, vale citar três diferentes métodos: (i) o uso de momentum, ou seja, a inclusão de um parâmetro adicional na equação de mudança de pesos da rede, de modo a acelerar a convergência em regiões onde o gradiente é pronunciado, e vice-versa em regiões onde a rede aprende pouco; (ii) o algoritmo com Taxa de Aprendizagem Variável, onde essa pode variar ao longo do processo de treinamento, evitando os problemas de instabilidade advindos de uma escolha a priori errônea deste parâmetro; e (iii) o algoritmo de Retropropagação Resiliente, que tenta evitar o problema de magnitude das derivadas parciais usadas para realizar a atualização dos pesos da rede (note-se nas equações do “passo para trás” do algoritmo backpropagation, que se usa constantemente as derivadas da função de ativação), e faz uso apenas do sinal destas derivadas, que indicam apenas a direção que se deve tomar na superfície de erro, deixando o problema da magnitude de variação dos pesos da rede para um outro parâmetro. De um modo geral, estas três modificações no algoritmo básico tendem a melhorar consideravelmente a performance do processo de treinamento, sem incorrer em substancial aumento dos recursos computacionais utilizados.

Já no segundo grupo, há também três técnicas diferentes que permitem obter ganhos no processo de aprendizado: Gradiente Conjugado, quasi-Newton e Levenberg-Marquardt.

O algoritmo do Gradiente Conjugado parte da premissa que, embora uma função decresça mais rapidamente na direção do negativo de seu gradiente (esta é a premissa do algoritmo básico de descida pelo gradiente), não necessariamente esta direção é que produz a melhor convergência global. Assim, o algoritmo faz uma busca numérica nas direções conjugadas à negativa do gradiente, tentando otimizar o resultado. Há diversas maneiras de realizar tal busca, e daí surgem diversas variantes do algoritmo: Fletcher-Reeves, Polak-Ribière, Powell-Beale, Golden Search, entre outros.

Os métodos quasi-Newton partem da mesma premissa básica das técnicas de gradiente conjugado, apenas utilizam métodos de otimização numérica diferentes, baseados em variantes do algoritmo de Newton. Da mesma maneira, o algoritmo de Levenberg-

Marquardt busca otimização numérica com uma aproximação do método de Newton (em linhas gerais, o método de Newton envolve o cálculo do Hessiano de uma função – os métodos quasi-Newton e Levenberg-Marquardt buscam aproximar o Hessiano). Segundo Demuth e Beale (2001), este último tem se mostrado ser o algoritmo mais rápido para o treinamento de redes alimentadas adiante (*feedforward*) de tamanho moderado (algumas centenas de pesos).

4.7.2. Métodos de Aceleração e de Convergência

- Inicialização dos Pesos

Antes de começar qualquer processo de aprendizado, os pesos de uma rede neural devem ter um valor inicial. O objetivo dessa inicialização é começar o processo de minimização do erro com uma solução inicial que possa, dentro do possível, economizar tempo de aprendizado e de convergência. Uma alternativa inicial seria colocar todos os pesos da rede com valor zero, porém isso resultaria em erros na hora de derivar as equações dos nós das camadas escondidas, não modificando os pesos e, por conseguinte, fazendo com que a rede neural não seja capaz de resolver qualquer problema.

Uma abordagem clássica para inicialização de pesos é escolher pequenos valores de forma aleatória para assegurar que a rede consiga realizar suas derivadas. Uma alternativa proposta por Fu (1994) é que em uma rede no qual as derivadas da função sigmodal fique no intervalo de $[-3, 3]$, uma escolha apropriada seria colocar valores iniciais do peso fiquem no intervalo de $[-\frac{3}{\sqrt{N}}, \frac{3}{\sqrt{N}}]$, onde “N” é o número de entradas da rede neural. Porém, essa escolha inicial resultará em mudanças relativamente pequenas nos pesos. Logo, nesses casos, deve-se fazer uma escolha que leve em conta os valores iniciais e também os critérios de mudanças dos pesos.

Essas regras são feitas para assegurar que a rede comece a funcionar na parte “linear” de sua função e não na parte plana horizontal (onde a derivada é igual a zero) da função. No

entanto, não há garantias que esses valores iniciais levarão ao mínimo global da função, mas é importante, principalmente em método de aprendizado como backpropagation, que a solução inicial fique presa em uma região de mínimo local.

A figura 4.9 mostra uma superfície de erro de uma rede neural. Observe que nela existem diversos de mínimos locais, o que pode levar o método a achar um mínimo local da função custo (EQM) e desistir de procurar a melhor solução do problema (mínimo global).

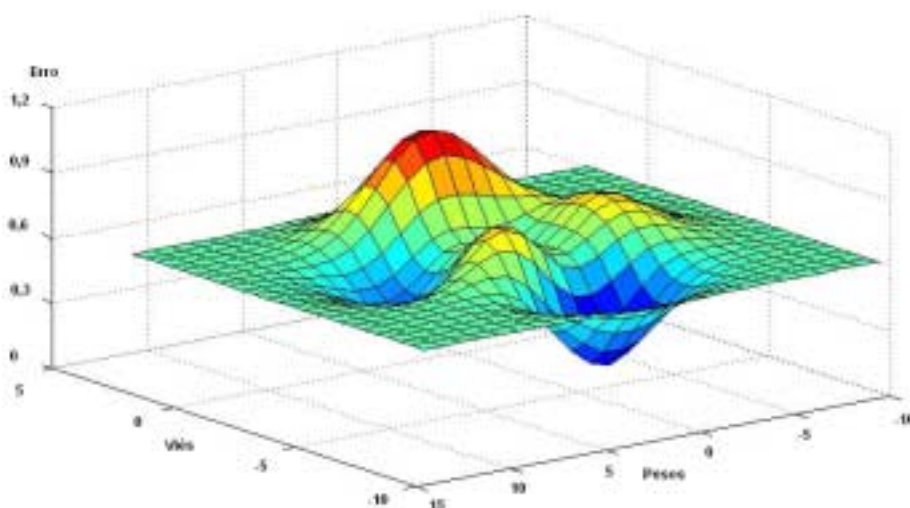


Figura 4.9: Superfície de erro de uma rede neural fictícia

▪ Evitando Mínimo Local

Como foi mencionado anteriormente, na superfície do erro de uma RNA geralmente contêm mínimos locais e isso cria um grande obstáculo para o aprendizado de redes neurais, especialmente no algoritmo backpropagation. Mesmo tendo autores que questionam que mínimos locais podem achar soluções satisfatórias (Fahlman, 1992) geralmente a melhor estratégia seria evitá-los e garantir a convergência do algoritmo de aprendizado para soluções globais.

- Treinamento seqüencial e por Lote

Em uma aplicação prática do algoritmo de retropropagação, o aprendizado resulta das muitas apresentações de um determinado conjunto de exemplos de treinamento para o perceptron de múltiplas camadas. Como mencionado anteriormente, uma apresentação completa do conjunto de treinamento inteiro é denominada uma época. O processo de aprendizagem é mantido em uma base de época em época até os pesos sinápticos e os níveis de viés se estabilizarem e o erro médio quadrado sobre todo o conjunto de treinamento convergir para um valor mínimo. É uma boa prática tornar aleatória a ordem de apresentação dos exemplos de treinamento (Haykin, 2001), de uma época para a seguinte.

Quando se utiliza o modo estocástico do backpropagation, como é conhecido o treinamento seqüencial, os pesos são atualizados após cada exemplo do treinamento da rede. Com isso, a rede pode perder muito tempo indo para trás e pra frente na busca da melhor solução para o problema

Uma forma simples de evitar esse fenômeno descrito acima é somar as mudanças de pesos nos exemplos de treinamento. Esse modo de treinamento é chamado de treinamento por lote (*batching*), que tem como objetivo computar a média das mudanças dos pesos. A principal característica do treinamento por lote é que as mudanças dos pesos só são realizadas após a apresentação de todos exemplos de treinamento da época.

Para uma época particular, define-se a função de custo como:

$$E_{\text{médio}} = \frac{1}{2N} \sum_{n=1}^N \sum e_j^2 \quad (4.17)$$

onde o sinal de erro e_j é relativo ao neurônio de saída j do exemplo de treinamento n .

- *Momentum*

Outro método utilizado para se alterar a procura pela direção é somar algumas das antigas direções do gradiente atual. Esse método é chamado *momentum* e é implementado pela seguinte expressão:

$$\Delta W_{ij}(t+1) = \mu \Delta w_{ij}(t) - (1-\mu) \varepsilon \frac{\partial E}{\partial W_{ij}} \quad (4.18)$$

onde μ é a taxa de treinamento

De acordo com essa equação, no modo por lote, o ajuste do peso w_{ij} é feito somente após o conjunto de treinamento inteiro ter sido apresentado à rede.

Do ponto de vista operacional, o modo seqüencial de treinamento é preferível em relação ao modo por lote, porque requer menos armazenamento local para cada conexão sináptica (Haykin, 2001). Além disso, dado que os parâmetros são apresentados à rede de uma forma aleatória, o uso de ajuste de pesos de padrão torna a busca no espaço de pesos estocástica, o que torna menos provável que o algoritmo de treinamento fique preso em um mínimo local.

4.7.3. Generalizações

Um dos problemas que podem ocorrer durante o treinamento de uma rede neural é o chamado excesso de treinamento (*overfitting*). Isso ocorre quando o erro de treinamento fica muito pequeno, porém quando um novo dado é apresentado à rede o erro fica muito grande. Nesse caso a rede “decorou” os exemplos de treinamento, mas não aprendeu a generalizar para novas situações.

Existem algumas técnicas que podem ser utilizadas para evitar que tal processo ocorra. Essas técnicas podem ter como objetivo detectar quando uma rede está produzindo resultados muito precisos, identificar quando há excesso de treinamento, detectar casos

de más generalizações, ou ainda, achar problemas decorrentes de segmentação dos dados. Abaixo são mostradas essas técnicas com mais detalhes.

- Tamanho da Amostra de Dados

Um método simples para melhorar a generalização de uma rede é aumentar o número de exemplos na base de dados. Como em qualquer observação física, o aumento do número de observações leva a uma diminuição proporcional dos efeitos de ruídos (Fu 1994), e ainda, várias observações do mesmo exemplo vão levar a uma rede a aprender esses acontecimentos ao invés de “decorar” o último acontecimento.

Porém na prática, é difícil se saber de antemão o que é um tamanho grande de dados para uma aplicação específica.

- Regularização

Um segundo método para melhorar a generalização de uma rede é chamado de “Regularização”. Sua implementação consiste em somar um termo de penalidade na função custo para evitar movimentos bruscos na curva de treinamento.

Na prática isso consiste em somar a função do Erro Quadrático Médio um termo que seria a média da soma dos quadrados dos pesos e vieses (SQPV) da rede.

$$E = EQM = \frac{1}{N} \sum_{i=1}^N (e_i)^2 \quad (4.19)$$

$$E' = \gamma \times EQM + (1 - \gamma) SQPV \quad (4.20)$$

sendo γ a taxa de performance e

$$SQPV = \frac{1}{N} \sum_{j=1}^N w_j^2 \quad (4.21)$$

Uma consequência em utilizar esse método é que esta função forçará os pesos e vieses da rede a serem pequenos e isso fará com que a resposta da rede seja suave bem

ajustada. O problema da Regularização está na dificuldade de se determinar o valor ótimo do parâmetro taxa de performance.

- Parada Antecipada (*Early Stopping*)

Nessa técnica os dados disponíveis são subdivididos em três subconjuntos: i) Subconjunto de treinamento, utilizado para calcular o gradiente e atualizar os pesos e vieses da rede; ii) Subconjunto de validação, o erro de validação é monitorado durante o processo de treinamento. Normalmente o erro de validação decrescerá na fase inicial de treinamento, porém quando ocorrer excesso de treinamento o erro de validação começará a aumentar. Quando esse aumento ocorrer durante um número específico de iterações, o treinamento é interrompido e o resultado serão os pesos e vieses do menor erro de validação encontrado no treinamento; iii) Subconjunto de testes, nessa etapa os pesos já foram calculados e servem como prévia dos resultados.

Em todos os casos mostrados acima, quando aplicado de maneira correta, podem fazer com que a rede consiga generalizar suas propriedades de aprendizado. Na técnica de regularização, é importante que se treine a rede até que atinja convergência, ou seja, os parâmetros EQM e SQPV devem chegar bem perto de valores constantes. Na técnica de parada antecipada, é importante utilizar um algoritmo que não convirja muito rapidamente e também se devem escolher os pontos do subconjunto de validação de forma que abranja representativamente todos pontos do subconjunto de treinamento.

4.7.4. Limitações do *Backpropagation* em uma RNA

O algoritmo de descida por gradiente é geralmente muito lento porque ele necessita de uma baixa taxa de aprendizado para estabilizar a rede e achar os melhores resultados. Já utilizando o conceito de *Momentum* a rede converge mais rapidamente do que na descida do gradiente, isso porque se pode utilizar taxas de aprendizado maiores enquanto a rede se mantém estável, mais ainda é um algoritmo muito lento para aplicações práticas. Segundo Demuth e Beale (2001) em redes de pequeno e médio porte é mais usual

utilizar-se do treinamento Levenberg-Marquardt por ser um algoritmo que converge mais rapidamente que os outros.

Uma rede MLP é capaz de desempenhar tão bem quanto outras técnicas computacionais, porém, teoricamente, enquanto uma rede neural treinada é sempre capaz de atingir resultados, os algoritmos de retropropagação e suas variações, apresentadas nesse capítulo, podem não conseguir convergir para uma solução ótima ficando preso em mínimos locais.

As redes MLP são muito sensíveis à quantidade de neurônios existentes na camada escondida, enquanto poucos neurônios podem levar a uma rede ficar pouco treinada (*underfit*), muito neurônios contribuem para um excesso de treinamento. Então é importante encontrar o ponto em que a rede consiga atingir bons resultados sendo capaz de generalizar soluções.

4.7.5. Tratamento dos dados

Segundo Refenes (1995), um dos resultados mais importantes no trabalho com redes neurais é fazer com que uma rede consiga aproximar o maior número de funções possíveis, ou seja, dado um número suficientemente grande de dados de entrada, o processo de treinamento conseguirá achar um padrão entre o conjunto de variáveis existentes (sejam elas dependentes ou independentes).

Este um resultado é muito importante, pois garante que uma rede neural consiga resolver uma grande variedade de problemas. Porém essa possibilidade de generalização tem um ponto fraco, uma vez que pode acontecer que durante o processo de treinamento a rede tenda a achar relações entre os dados onde não existam. Portanto, o processo de selecionar e tratar as variáveis de um problema é uma parte essencial do processo de construção de uma rede neural.

Independentemente da eficiência do algoritmo de aprendizagem em termos de convergência, generalização e estabilidade, o indicador de desempenho de uma RNA depende da relevância das variáveis independentes escolhidas e na qualidade dos dados utilizados. Quando se utiliza poucas variáveis independentes na entrada de dados faz com que diminua excessivamente o espaço de busca, podendo introduzir vieses no processo de modelagem, que em geral levam a generalizações pobres. Já quando se utiliza variáveis em excesso, existe a possibilidade de aumentar a dimensionalidade do espaço de busca o que vai introduzir erros de generalizações do algoritmo de busca.

Logo, a seleção das variáveis de entrada é uma etapa importante do processo de modelagem antes de se iniciar a computação propriamente dita. Outra etapa essencial é o tratamento dos dados, que nada mais é do que estabelecer algumas rotinas de pré-processamento e de pós-processamentos, de modo que a rede neural trabalhe de forma mais eficiente, dentre as mais importantes pode-se citar:

- Normalização dos dados

Normalização é um processo de padronização do intervalo que ao qual os dados podem pertencer, com isso transforma-se todos os dados de forma que pertençam a um intervalo pré-estabelecido, como por exemplo [0,1], isso sem perder as características individuais de cada variável. A normalização dos dados é necessária para remover a possibilidade dos parâmetros ficarem presos em um determinado intervalo e também para fazer com que os dados de entrada fiquem dentro dos limites das funções de ativações.

Uma normalização muito utilizada consiste em remover as médias dos valores de entrada e dos valores desejados (média igual a zero), e transformar seus desvios padrões em unitários através da seguinte equação:

$$x'_t = \frac{x_t - \bar{x}}{s_x} \quad (4.22)$$

$$s'_x = \sqrt{\frac{\sum_{t=1}^N (x_t - \bar{x})^2}{N-1}} \quad (4.23)$$

onde $\bar{x}$ e s'_x são as médias e os desvios padronizados.

- Retirar a correlação dos dados

Na método, os conjuntos de séries de entrada são analisados e as séries que são correlacionadas são retiradas da amostra de dados. Um exemplo simples dessa aplicação seria o caso de existir uma série de conjunto de entradas que é múltiplo do outro conjunto, nesse caso retira-se uma das entradas.

- Retirada dos “*Outliers*”

Uma prática muito comum é encontrar os *ouliers* (valores extremos que destoam do resto da série de dados) de uma série e retirá-los para que não haja interferência na média e na variância dos dados estudados.

Um método para encontrar os *outliers* é calcular a distância de “Mahalanobis” D^2 na série de dados¹.

$$D^2 = \sum_{i=1}^N \sum_{j=1}^P (z^i_t - \bar{z}^i) v^{ij} (z^j_t - \bar{z}^j) \quad (4.24)$$

onde P é o número de variáveis, N é o número total de padrões de treinamento, $\bar{z}^i_t$ é a média da variável estudada z^i_t e v^{ij} é o elemento do inverso da matriz de covariância das P variáveis.

Além do pré-processamento, é obviamente fundamental realizar um pós-processamento dos resultados, de maneira a ter resultados interpretáveis dada uma saída qualquer da

¹Para maiores informações desse método, veja Refenes, 1995

rede neural. O pós-processamento segue as mesmas etapas que o pré-processamento, mantendo as mesmas características para que no final se possa melhor analisar os resultados.

4.8. Medidas de desempenho de uma RNA

Enquanto o Erro Quadrático Médio (EQM) é uma medida de desempenho perfeitamente aceitável, na prática o objetivo final de quem está testando o modelo é saber se os resultados são consistentes o suficiente para garantir a confiabilidade do modelo, e para isso é importante conferir mais de uma medida de desempenho.

- Estimadores de Desempenho

Uma medida de desempenho muito popular também fora do universo das redes neurais é o coeficiente de correlação (R). Esse coeficiente mede a correlação linear entre os valores desejados (*target- t_k*) e os valores de saída (*output- y_k*).

$$R = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}} \quad (4.25)$$

onde $\bar{x} = 1/N \sum_{i=1}^N x_i$, $\bar{y} = 1/N \sum_{i=1}^N y_i$, $R^2 \in [0,1]$ e $R \in [-1,1]$

O coeficiente R^2 é utilizado de forma de $R^2 = 1$ representa uma correlação perfeita e $R^2 = 0$ significa que os dados não tem correlação.

Uma outra medida de desempenho que é bastante similar ao EQM, com a diferença de se incorporar a complexidade do modelo é o critério de informação Bayesiano dado pela seguinte equação:

$$B = \ln \left[\frac{\sum_{i=1}^N (x_i - y_i)^2}{n} \right] + \frac{\ln(n)}{n} \times k \quad (4.26)$$

onde k é o número de pesos da rede neural e n é o número de dados da série estudada. Modelos que tem o menor B apresentam as melhores previsões e as melhores capacidades de generalizações.

▪ *Lucro Estimado*

Quando se trata de previsões de preços de ativos financeiros uma boa medida de desempenho da rede é simplesmente verificar se com as previsões que estão se realizando é possível, teoricamente, obter lucro ou não com a operação. Se o modelo prevê para amanhã um preço maior do que o preço de hoje, é vantajoso comprar esse ativo e vice-versa. Em termos matemáticos temos o retorno r dessa estratégia representada abaixo:

$$r = \sum p_t \times (x_{t+1} - x_t) \quad (4.27)$$

$$\text{onde } p_t = \begin{cases} 1 ; \text{ se } (y_{t+1} - x_t) > 0 \\ -1 ; \text{ se } (y_{t+1} - x_t) < 0 \\ 0 ; \text{ se } (y_{t+1} - x_t) = 0 \end{cases}$$

Quanto mais positivo o retorno r for melhor é a capacidade de previsão do modelo.

4.9. Exemplo de uma RNA

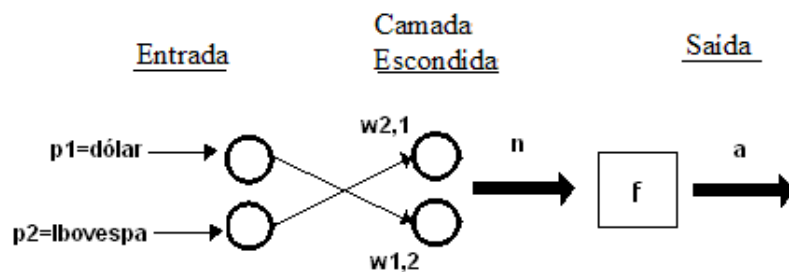
Para ficar mais claro o funcionamento de uma rede neural artificial, adiante será apresentado um exemplo¹ de uma rede feedforward acíclica com uma camada escondida formada com neurônios. Para treinar a rede se utilizará o algoritmo backpropagation.

Considere que se deva prever o preço do dia seguinte de uma ação XYZ negociada na BOVESPA (Bolsa de Valores do estado de São Paulo) e para obter essa previsão dispõem-se dos preços da última semana do Índice Ibovespa e da cotação do dólar. A empresa XYZ é uma empresa que exporta boa parte dos seus produtos por isso o preço de sua ação é muito relacionado com o preço do dólar.

A tabela abaixo apresenta os dados de entrada da rede neural artificial (série da cotação do dólar e do Índice Ibovespa) e dos valores desejados (preços da empresa XYZ). Logo depois, foi esquematizado a topologia da rede utilizada nesse exemplo.

	Inputs		Target
	Dólar	Ibovespa	XYZ
Segunda	2,82	21.020	7,65
Terça	2,84	20.850	7,58
Quarta	2,85	21.200	7,72
Quinta	2,81	21.500	7,55
Sexta	2,83	21.300	7,61

¹ Este exemplo foi feito pelo próprio autor do trabalho, sendo meramente ilustrativo para um melhor entendimento do funcionamento de uma RNA. O exemplo foi realizado no *Matlab*.



$$n = w_{1,2} \times p_1 + w_{2,1} \times p_2$$

$$a = f(Wp)$$

sendo que W é a matriz dos pesos $w_{1,2}$ e $w_{2,1}$; p_1 e p_2 são as entradas do problema; a é a saída (*output*) da rede e f é a função de ativação.

- **Normalização dos dados:** Primeiramente será normalizado os dados para deixarmos com média igual a zero e variância unitária.

Inputs		Target
Dólar	lbovespa	XYX
-0,63	-0,61	-1,79
0,63	-1,29	0,45
1,26	0,10	0,46
-1,26	1,30	0,45
0,00	0,50	0,45

- **Inicializar o pesos:** Depois de montada a rede, é preciso inicializar os pesos $w_{1,2}$ e $w_{2,1}$ com valores pequenos e aleatórios.
- **Cálculo da função de ativação:** No exemplo será utilizado a função de ativação

sigmóide ($f = \frac{1}{1 + e^{-a}}$).

- **Treinamento dos pesos:** Depois de inicializado os pesos e calculada a função de ativação, inicia-se o processo de treinamento pelo algoritmo backpropagation mostrado no item 4.7. O processo de treinamento é dividido em três etapas: treinamento, validação e teste conforme o gráfico abaixo.

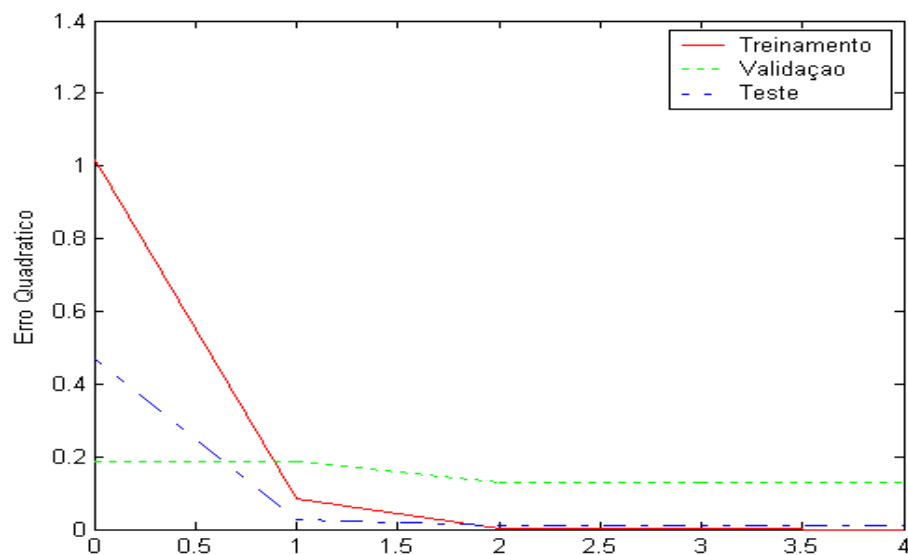


Figura 4.10: Gráfico do resultado do treinamento , validação e teste. O Eixo “x” são as etapas do processo, nesse caso os dias da semana.

Geralmente as curvas do gráfico 4.10 caminham mais juntas, mas como no exemplo são utilizadas um conjunto de dados pequeno, os resultados da rede ficam um pouco distorcidos.

- **Saída da Rede Neural:** Os resultados obtidos na rede são apresentadas no gráfico 4.11, nele é possível verificar os preços previstos e comparar com os preços esperados.

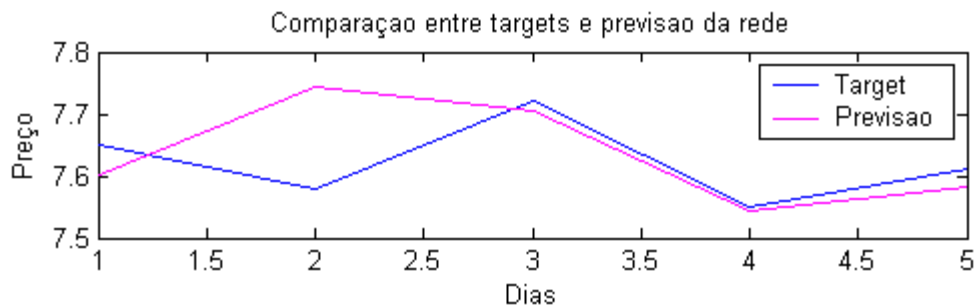


Figura 4.11: Gráfico comparativo dos preços previstos com os preços reais da rede neural

- **Análise dos resultados:** Para analisar se o treinamento foi bem realizado considera-se duas medidas: o erro quadrático médio (4.6) e o coeficiente de correlação (4.25) entre a saída da rede (*output*) com os resultados desejados (*target*).

No gráfico abaixo observa-se no eixo “x” e os valores desejados e no eixo “y” os valores encontrados, através de uma regressão linear encontra-se a reta que melhor ajusta os dados e consequentemente o coeficiente R .

Erro Quadrático Médio = 0,006

$R = 38,5\%$

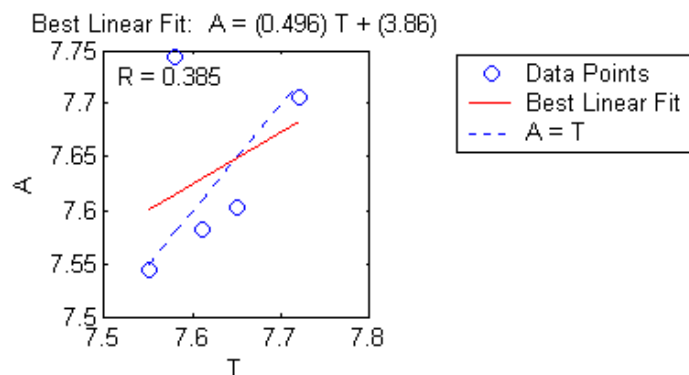


Figura 4.12: Gráfico de Correlação dos preços previstos com os preços reais da rede neural

Conforme o gráfico 4.12, os resultados obtidos não foram muito bons ($R^2 = 15\%$), isso ocorreu principalmente por que no exemplo feito foi-se utilizado uma pequena série de dados (somente de 1 semana). Numa simulação real é preciso que se use uma sequência maior de dados para que a rede consiga realizar um bom treinamento sendo capaz de melhores generalizações.

Com todos os resultados obtidos nesse exemplo, se quisermos prever o preço das ações XYZ para a semana seguinte (segunda-feira) utilizaríamos os pesos $w_{1,2} = 3,432$ e $w_{2,1} = 2,3016$ obtidos no treinamento.

5. EXPERIMENTOS DOS MODELOS

5.1. Introdução

Depois de expor as teorias necessárias para a construção de um modelo, utilizando desde métodos de previsão tradicionais até modelos mais complexos como redes neurais, é possível mostrar, passo a passo, como esses conceitos podem ser aplicados na prática para auxiliar no processo de tomada de decisão dos usineiros e produtores.

A proposta desse trabalho é auxiliar os usineiros e produtores de cana-de-açúcar a decidir o mix de produção alguns meses antes do final da colheita, ou seja, qual porcentagem de cana-de-açúcar vai ser utilizada para produzir álcool e qual vai virar açúcar. Como foi dito no capítulo 2, hoje em dia é utilizado, dentre outros fatores, o preço açúcar e do álcool negociados no mercado futuro das bolsas de valores do Brasil e de Nova Iorque, para determinar qual será mais vantajoso financeiramente na época da colheita.

Nesse capítulo, será exposto como os preços futuros de açúcar podem ser modelados matematicamente, para auxiliarem tanto no processo de tomada de decisão da escolha do mix de produção quanto no ajustes que podem ser realizados ao longo do tempo até a entrega da safra.

O capítulo está estruturado da seguinte forma: primeiramente será discutido quais variáveis serão utilizadas no modelo, mostrando como cada uma delas pode influenciar nos resultados finais; o tratamento preliminar realizado nesses dados para que haja uma homogeneidade entre eles; por fim, será apresentado os resultados da saída de cada modelo utilizado, comparando-os através medidas de desempenho para identificar aquele que melhor se ajusta à realidade.

Utilizaremos os seguintes modelos de previsão:

- i. *Média Móvel* (apresentada no item 3.4.2). Serão testados modelos com diferentes intervalos (*leg*) de tempo, para que se possa testar desde os modelos que respondem mais lentamente às mudanças na série de dados (*leg* maior) até os que respondem rapidamente à essas mudanças (*leg* menor).
- ii. *Redes Neurais* (apresentada no capítulo 4). Serão testadas algumas estruturas de rede, através da variação do número de camadas escondidas e do número de neurônios nessas camadas. Com isso, podem-se verificar as diferenças nos resultados obtidos.

5.2. Escolha das Variáveis do Modelo

Para escolher quais variáveis entrarão no modelo de redes neurais, é necessário identificar quais são os fatores, ou quais circunstâncias, podem afetar seu objeto de estudo, que no caso é o preço do contrato futuro de açúcar negociado na NYBOT¹. Depois de identificar as variáveis, as redes neurais irão quantificar quais delas mais afetam o comportamento da série estudada, através do treinamento dos pesos de cada neurônio.

- I. *Preço do açúcar do Brasil*: Como existem inúmeros empecilhos protecionistas que impedem que o açúcar brasileiro seja comercializado livremente nos países desenvolvidos, nem sempre um acontecimento nos EUA afeta diretamente os preço do açúcar brasileiro naquele momento. Logo a série de preços do contrato futuro de açúcar negociados na BMF² pode ser um bom indicador do preço do açúcar internacional.
- II. *Preço do álcool*: Como os dois produtos são derivados da cana-de-açúcar, existe uma relação natural entre ambos, ou seja, se a procura pelo álcool

¹ New York Board of Trade, é ao lugar são negociados a maioria das commodities agrícolas dos EUA.

² Bolsa de Mercadorias e Futuros é onde se negociam os contratos futuros das commodities brasileiras.

aumenta elevando seu preço existe uma probabilidade razoável de que o preço de açúcar também suba ao longo do tempo. Pelo fato do contrato futuro de álcool praticamente não ter liquidez nas bolsas americanas, será utilizado o contrato futuro de álcool da BMF.

- III. *Cotação do Dólar*: O Brasil é o maior produtor e maior exportador de açúcar do mundo, então o fato da moeda brasileira valorizar ou desvalorizar-se pode ser um grande evento. No caso de desvalorização da moeda local, o açúcar do Brasil fica mais barato no mercado internacional afetando a cotação dos contratos americanos. No caso de valorização ocorre o contrário da situação descrita acima.
- IV. *Índice CRB (Commodity Research Bureau)*: CRB é muito utilizado nos Estados Unidos, ele é um índice que acompanha o comportamento das commodities em geral e é composto por uma cesta de 25 commodities onde seus preços são ponderados e acompanhados diariamente pelo mercado. Sua importância reside no fato de que nem sempre o movimento das commodities andam juntas, em alguns casos o movimento de alta ou queda de um produto hoje pode influenciar outro produto no futuro.
- V. *Índice Dow Jones*: Esse índice é composto pelas ações das principais empresas industriais americanas, mudanças bruscas em seu comportamento influenciam direta ou indiretamente as bolsas e preços de ativos do mundo inteiro. Com ele, é possível identificar se o mercado está otimista ou pessimista em relação ao futuro.
- VI. *Número de Contratos em aberto dos contratos futuros de açúcar da NYBOT*: Número de contratos em aberto significa quantos contratos estão em mãos dos investidores. Por exemplo: se eu compro um contrato de ação hoje,

alguém que está na ponta contrária vendeu essa ação, ou seja, existe um contrato em aberto. Se eu vender essa ação para a mesma pessoa que comprei o contrato que foi aberto anteriormente vai ser fechado, se eu vender para um outro investidor esse contrato continua aberto só que com um dono diferente.

No mercado financeiro esse número representa o grau de exposição que os investidores tem em relação a um determinado ativo, se existem muitos contratos em aberto significa que existem muitas pessoas que acreditam que o preço da ação vai subir (os que compraram) e muitos que acreditam que o preço da ação vai cair (os que venderam). Na prática, quanto maior for o número de contratos em aberto, em relação à média histórica, maior é a chance de uma mudança brusca no preço, pois os “comprados” podem querer liquidar suas posições rapidamente por qualquer motivo, e vice-versa, afetando o comportamento geral dos ativos em questão.

VII. *Volume de Negociação do contrato futuro de açúcar da NYBOT:* Esse número indica o quanto se está negociando, em termos financeiros, do ativo por dia. A identificação de dias atípicos, como nos contratos aberto, pode identificar mudanças repentinas no mercado em geral.

VIII. *Dias até o vencimento do contrato futuro.* Esses dias nada mais são do que os dias que faltam para o contrato futuro vencer. No caso em estudo, serão utilizados os dias úteis, ou seja, não contando os feriados, até o prazo de vencimento do contrato.

Todas as variáveis acima serão testadas na aplicação das redes neurais para previsão dos preços dos preços do contrato futuro de açúcar da NYBOT. Além dessas, outras duas serão adicionadas nas séries de variáveis de entrada, são elas a média dos preços dos últimos 200 dias e a média dos preços dos últimos 30 dias de negociação do contrato

futuro de açúcar NYBOT. Esses dois valores são importantes para suavizar a série de preços desejada eliminando eventuais oscilações bruscas ocorridas ao longo do tempo.

Na figura 5.1, ilustra a entrada dessas variáveis na rede neural.

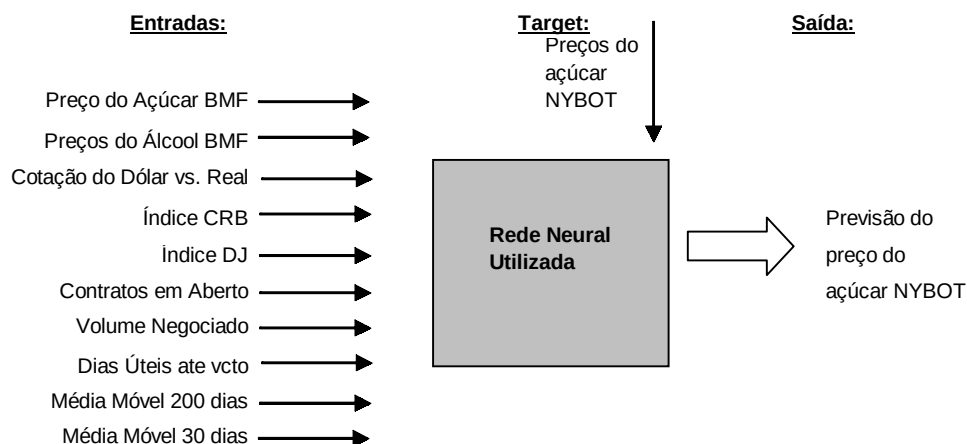


Figura 5.1: Variáveis do Modelo Neuronal utilizado.

5.3. Tratamento dos Dados

Depois da escolha das variáveis que serão utilizadas para a aplicação do modelo de previsão, foi necessário conseguir séries históricas desses dados. Todos os dados foram obtidos através de uma ferramenta muito utilizada no mercado financeiro, Bloomberg, que contém um banco de dados completo e confiável de diversos ativos e índices negociados no mundo inteiro.

Porém, nem todas as séries estavam perfeitamente completas, ou havia determinadas inconsistências entre elas, o que determinou a necessidade de realização de um tratamento nos dados, antes que esses fossem usados na aplicação dos modelos apresentados a seguir.

Os dois principais problemas eram: ausência de cotações em determinadas datas, e datas em que apenas um dos mercados opera. No primeiro caso, a causa é a ausência de negócios em determinados dias. A solução é adotar o preço médio do dia anterior como o preço daquele dia em que não houve negócios – ou seja, a melhor aproximação do valor daquela ação é o último preço médio dela. Afinal, não é possível saber a razão pela qual não houve negócios, e deve-se trabalhar com a informação que está disponível. Para o segundo problema, a solução adotada foi ignorar as cotações dos mercados nessas datas – por exemplo, no dia 31 de dezembro de 2002, o Dow Jones, CRB e NYBOT estavam abertos, assim como o dólar, enquanto a Bovespa estava fechada, logo não há cotações nessa data para as ações. Assim, ignorou-se os dados desta data, e supôs-se que os mercados não estivessem abertos.

Finalizado este trabalho preliminar de seleção, coleta e tratamento dos dados, nas seções a seguir serão apresentados a aplicação dos modelos e os principais resultados obtidos.

5.3. Medidas de desempenho do modelo de previsão

Essa é com certeza a etapa mais importante do trabalho, nela serão aplicados na prática os conceitos teóricos e análises apresentadas nos primeiros capítulos. Os testes serão realizados da seguinte forma: utilizaremos modelos já apresentados para prever o preço de açúcar do dia seguinte, e a acurácia da previsão será verificada através de algumas medidas de desempenho mostradas a seguir.

- *Erro Quadrático Médio (EQM)*: O EQM, como foi apresentado inúmeras vezes ao longo do trabalho, é uma medida que verifica se os resultados da previsão estão próximos do valor desejado, dando maior peso àquelas previsões que distam muito do valor real. (veja expressão matemática 3.6)

- *Coefficiente de Correlação (R)*: Esse índice mede a correlação linear entre os valores previstos e entre os valores desejados. Este conceito foi apresentado no item 4.8.
- *Lucro Estimado (LE)*: Uma boa medida de desempenho em previsões de ativos financeiros é simplesmente verificar se com elas é possível, teoricamente, obter lucro ou não com a operação. Se o modelo prevê para amanhã um preço maior do que o preço de hoje, é vantajoso comprar esse ativo e vice-versa. Por exemplo, com equação (4.9), podemos fazer uma pequena simulação para melhor entender o processo de estimar esse lucro teórico:

Data	Preço Real	Preço Previsto	LE	LE Acum.
05/07/04	7,96			
06/07/04	8,13	7,98	0,17	0,17
07/07/04	8,22	8,08	-0,09	0,08
08/07/04	8,16	8,07	0,06	0,14
09/07/04	8,27	8,12	-0,11	0,03
12/07/04	8,33	8,23	-0,06	-0,03

Tabela 5.1: simulação do LE

O 1º valor de LE é calculado pela variação dos preços reais $8,13 - 7,96 = 0,17$, se a previsão acertar que o preço vai subir, multiplica-se esse valor por 1, se errar multiplica-se por -1. O primeiro valor previsto era 7,98 (maior que 7,96, logo previsão de subida), como o valor real foi 8,13 (realmente o preço subiu) o LE fica positivo, no caso de queda de preço acontece o inverso. Ou seja, esse indicador penaliza se o modelo errar a direção do preço e gratifica se acertar, a magnitude dessa penalidade ou gratificação depende da variação real ocorrida no preço do dia. O Lucro Estimado Acumulado ao longo das previsões indica se o modelo tem acertado mais do que errado em suas previsões.

- *Retorno Financeiro*: Essa medida tem como objetivo mostrar que, se o investidor seguir os resultados do modelo de previsão, vai ganhar dinheiro ou não com a

operação. Esse indicador se diferencia da medida anterior por quantificar esse ganho ou perda realizando a operação de acordo com a previsão.

Funciona da seguinte forma: no início da operação há disponibilidade de 100 u.m. (unidades monetárias), que podem ser utilizados para comprar ou vender um determinado ativo financeiro, neste caso específico, o contrato de açúcar futuro. No começo do dia aloca-se uma percentagem desse dinheiro para comprar ou vender contrato futuro de açúcar e no fim do dia essa operação é liquidada, ou nos jargões do mercado financeiro, é zerada. A decisão de comprar/vender e a percentagem do capital que deve ser alocada nessa operação é determinada pelo resultado de saída do modelo de previsão.

Por exemplo: se o preço de fechamento do dia anterior for \$10, e após rodar o modelo de previsão, o resultado para o dia de amanhã é \$11, eu aloco uma parte do capital disponível na compra desse ativo; se for \$9, aloco outra parte na venda.

E para determinar o percentual do capital que será alocado em cada operação, estabeleceu-se a seguinte regra: quanto maior for a variação do preço previsto em relação o fechamento do dia anterior, maior será a alocação do capital disponível para a operação. Em outras palavras se o modelo está prevendo que a alta do dia seguinte é de 8%, maior é a probabilidade do preço do ativo realmente subir, por esse motivo deve ser alocado uma maior parte do dinheiro do que se a previsão de alta fosse de 1%.

Na prática, foi estabelecido que a cada 1% de previsão de alta deverá ser alocado 20% do capital disponível na compra e a cada 1% de previsão de queda deverá vender 20% do dinheiro disponível em contrato de açúcar futuro. Para previsão de 2% de será alocado 40% e assim sucessivamente até que uma previsão de 5%

aloque 100% do patrimônio. Para evitar alavancagem¹ financeira, mesmo com previsões de retornos acima de 5%, somente era aplicado os 100% disponíveis.

A simulação abaixo tem como objetivo exemplificar a metodologia dessa medida de desempenho:

Data	Previsão do Preço	Preço Real	Previsão de Retorno	Retorno Real	% de Alocação	Retorno Financeiro do dia	Retorno Financeiro Acum.
02/07/04	7,89	7,96					100,00
05/07/04	7,92	7,96	-0,50%	0,00%	-20%	0,00	100,00
06/07/04	7,98	8,13	0,25%	2,14%	0%	0,00	100,00
07/07/04	8,08	8,22	-0,62%	1,11%	-20%	-0,22	99,78
08/07/04	8,07	8,16	-1,82%	-0,73%	-40%	0,29	100,07
09/07/04	8,12	8,27	-0,49%	1,35%	0%	0,00	100,07
12/07/04	8,23	8,33	-0,48%	0,73%	0%	0,00	100,07
13/07/04	8,21	8,13	-1,44%	-2,40%	-20%	0,48	100,55
14/07/04	8,38	8,23	3,08%	1,23%	60%	0,74	101,29
15/07/04	8,39	8,24	1,94%	0,12%	40%	0,05	101,34

Tabela 5.2 – Funcionamento da medida de Retorno Financeiro

Observe na tabela 5.2, que o resultado do modelo de previsão do preço do açúcar para o dia 07/07 foi \$8,08, uma previsão de queda de 0,62% em relação ao dia anterior ($8,08/8,13 = -0,62\%$), isso significa que devo investir 20% dos \$100. disponíveis na venda do açúcar futuro. Porém, como o preço do açúcar fechou o dia cotado a \$8,22, uma alta de 1,11%, o investidor teve uma perda de \$0,22 ($\$20 \times -1,1\%$) com a operação. Já no dia 14/7, o modelo resultava em uma previsão de alta de 3,08%, devo investir 60% do meu capital, como o retorno real ao fim do dia foi de 1,23%, o investidor ganhou \$0,74.

Como pode ser observado esse indicador tem como pressuposto a utilização de operações de “*day-trade*”, ou seja, o investidor compra e vende o ativo no mesmo dia, iniciando o dia seguinte sempre sem nenhum ativo na carteira. Para tornar essa medida mais realista, é interessante considerar a variação do preço do

¹ alavancar significa investir mais do que se tem disponível, incorrendo em maiores riscos.

ativo do dia e não a variação do preço de fechamento, ou seja, em vez de calcular

a variação real pela fórmula $\frac{P_{D0}}{P_{D-1}} - 1$, passe-se a calcular pela variação do preço

de fechamento em relação ao preço de abertura $\frac{P_{fec}}{P_{abe}} - 1$.

5.4.Utilização da Média Móvel

Para realizar previsão através do método da Média Móvel, foram utilizados sete intervalos distintos de tempo, que tem com objetivos captar diferentes suavizações dos preços do açúcar futuro. Primeiro testou-se um *leg* de 6 meses, depois de 3 meses, 2 meses, 1 mês, 15 dias, 1 semana e 2 dias.

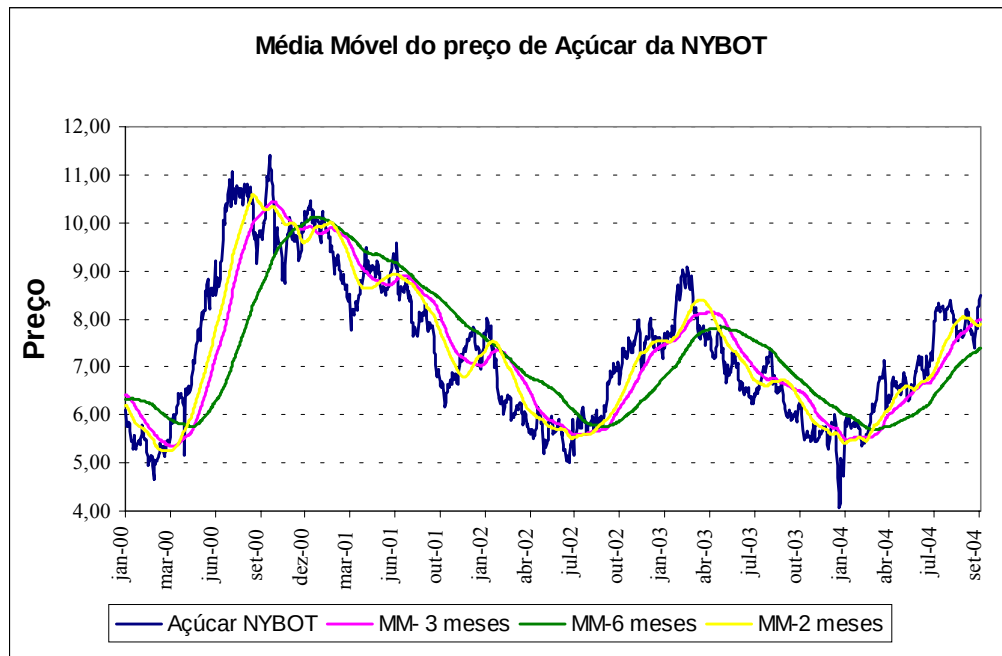


Fig 5.2- Média Móvel em diferentes intervalos de tempo (6meses, 3 meses, 2meses)

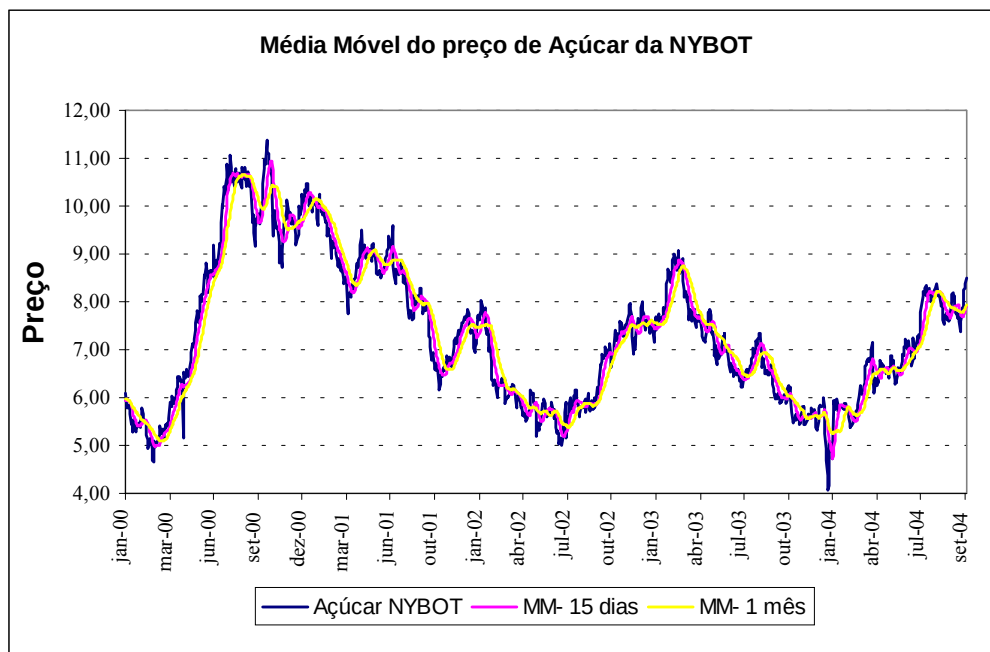


Fig 5.3- Média Móvel em diferentes intervalos de tempo (1 mês, 15 dias)

Observe no gráfico da figura 5.2 quando se usa média móvel com *legs* de tempo mais longos, como o de 6 meses, responde as mudanças no preço do açúcar NYBOT de forma muito mais lenta que nos *legs* menores. Essa suavização das series é importante porque se em uma semana houve algum evento atípico que derrubou a cotação do açúcar do NYBOT, nas curvas suavizadas esses eventos passam despercebidos.

Já no gráfico da figura 5.3, as curvas da média móvel, por terem um *leg* de tempo menor, respondem mais rápido às mudanças nas séries ficando bem parecidas ao longo do tempo.

5.3.1 Teste de Normalidade

Antes de apresentar os resultados de previsão da média móvel, é importante fazer uma breve análise de resíduos, para verificar qual o comportamento dos erros de previsão. Para isso, foi aplicado o Teste Kolmogorov – Smirnov (apresentado no capítulo 3) nos resíduos de todas a média móveis.

Esse teste foi feito com o auxílio do software *Matlab*, e todos resultados mostraram normalidades dos erros de previsão com 95% de confiança. Abaixo são apresentadas as saídas desse teste da Média Móvel de 3 meses.

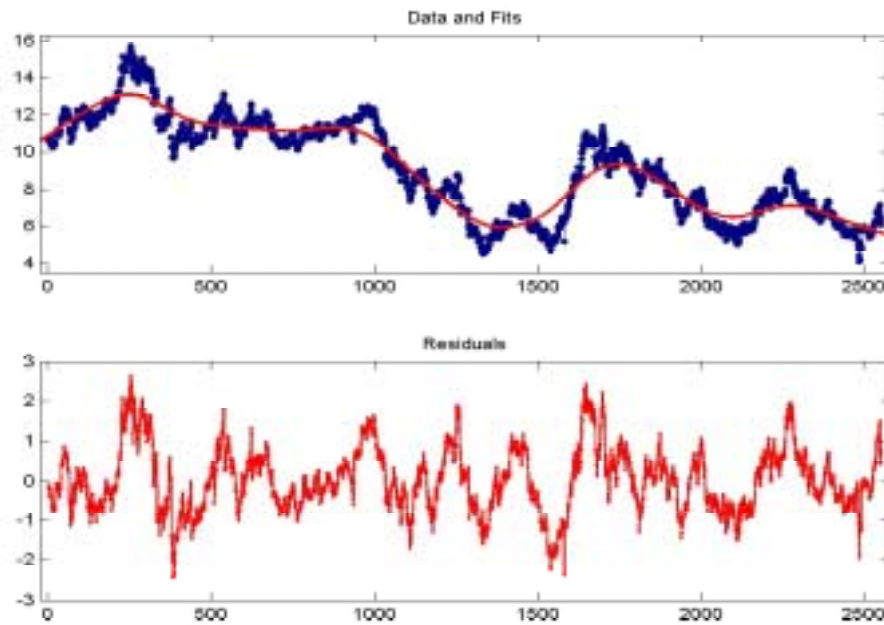


Figura 5.4. Gráfico dos Resíduos da Média Móvel

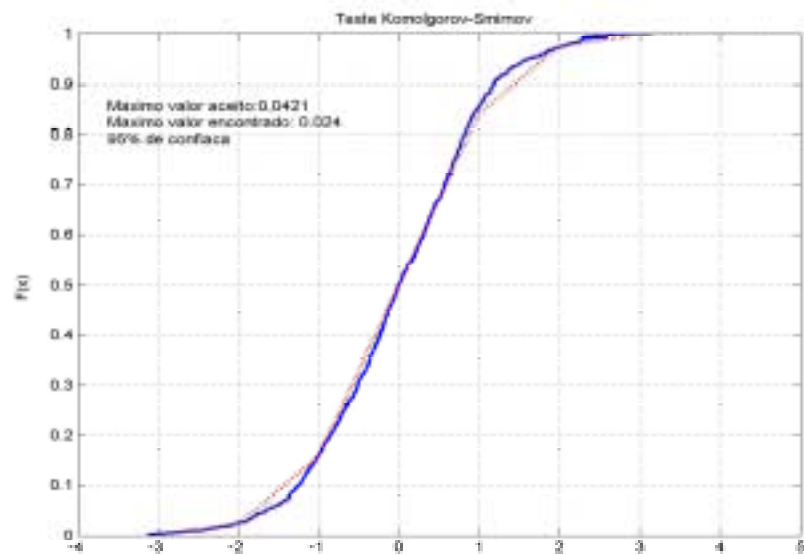


Figura 5.5. Saída do Teste Komolgorov - Smirnovv.

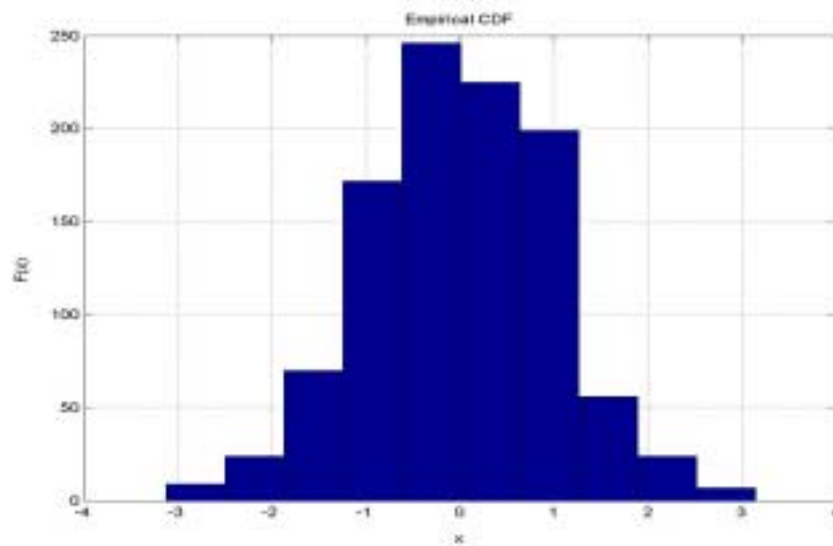


Figura 5.6 – Histograma dos Resíduos da Média Móvel.

5.3.2. Resultados da Média Móvel

Para aplicar o modelo de previsão utilizando Média-Móvel para poder avaliar a acurácia da previsão, simulou-se durante três meses (2/jul/04 até 30/set/04) o modelo para obter os resultados apresentados na tabela abaixo.

	EQM	R	Retorno Financeiro	Lucro Estimado
MM - 6 meses	1,326	-48,4%	R\$ 87,67	-0,43
MM - 3 meses	0,559	-48,4%	R\$ 100,94	1,40
MM - 2 meses	0,372	-55,6%	R\$ 100,79	1,33
MM - 1 mês	0,193	-22,8%	R\$ 99,41	-0,03
MM - 15 dias	0,169	-13,7%	R\$ 101,65	0,43
MM - 7 dias	0,060	52,5%	R\$ 98,99	-1,27
MM - 2 dias	0,022	83,7%	R\$ 99,47	-0,36

Tabela 5.3 – Resultados da Média Móvel

Pode-se observar que os resultados não vieram muito bons, como já era de se esperar quanto menor for o *leg* de tempo do modelo, mais perto a previsão vai ficar da realidade. Essa teoria pode ser comprovada, pela tabela 5.3, o Erro Quadrático médio e o coeficiente de correlação foram diminuindo cada vez que se diminuía o intervalo da

média móvel. Porém, isso não indica que o modelo está prevendo melhor, tanto é que o menor EQM e o maior coeficiente de correlação (Média Móvel de 2 dias) não apresentaram os melhores lucros estimados e retornos financeiros. No quesito retorno financeiro, o melhor resultado veio na média móvel de 15 dias e no lucro estimado mo de 3 meses.

5.5.Utilização das Redes Neurais

Nessa etapa será testada a rede neural para previsão de preços do contrato futuro de açúcar da NYBOT, vamos mostrar todos os passos desde como foi feito para construir a rede, o tratamento dos dados até os resultados obtidos. Todos testes foram feitos com o auxílio do software *Matlab*.

Como foi apresentado no capítulo 4, redes neurais são ferramentas ideais para a construção de modelos de previsão de ativos financeiros. Suas propriedades se adequam bem às características das séries temporais de preços: flexibilidade (as redes lidam bem com situações de mudança bruscas no comportamento dos preços), robustez (há uma enormidade de situações possíveis no mercado, e o reconhecimento de padrões é tarefa absolutamente não-trivial) e capacidade de capturar padrões não-lineares complexos.

Logo, para que a rede funcione da forma desejada, teve que se levar em conta detalhes na construção de sua topologia e no tipo de treinamento utilizado. Para realizar os testes foram considerados dois tipos de rede: i) com uma camada escondida; ii) com duas camadas escondidas. Em cada uma dessas opções testou-se de 3 até 6 neurônios na camada escondida. E para garantir uma maior confiabilidade do processo de treinamento cada uma dessas opções foi testada 10 vezes (vide tabela 5.4).

Tipos de Redes (1 ou 2 camadas escondidas)	Número de Neuronios (3 a 6)	Número de teste em cada rede	Total de teste realizados
2	4	10	80

Tabela 5.4 – Testes realizados na RNA

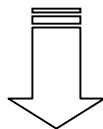
O treinamento da rede neural construída foi realizado utilizando o algoritmo Levenberg-Marquardt. A superioridade desse algoritmo de treinamento se dá tanto pela alta performance apresentada, quanto pela boa velocidade para treinar a rede.

5.5.1 Entrada das Redes Neurais

Os dados de entrada da rede são todas as variáveis apresentadas no item 5.2 (10 no total), no intervalo de tempo de Abril de 200 até outubro de 2004. Não se conseguiu um intervalo de tempo maior, principalmente por algumas variáveis não eram negociados antes dessa data, como o preço de álcool da BMF.

Data	Alcool BMF	Açucar BMF	Indice CRB	Real vs Dolar	Indice DJ	Cont. aberto	Volume Neg	MM- 200 dias	MM - 30 dias	Dias até Vcto
3/4/2000	381,00	7,35	212,48	1,74	11.221,93	62.506,00	20.337,00	6,00	5,23	19
4/4/2000	374,00	7,35	211,36	1,75	11.164,84	59.823,00	11.371,00	6,00	5,26	18
5/4/2000	370,00	7,35	212,22	1,75	11.033,92	57.123,00	18.129,00	6,00	5,29	17
6/4/2000	370,00	7,30	210,72	1,75	11.114,27	53.701,00	13.718,00	6,00	5,31	16
7/4/2000	374,00	7,30	209,91	1,74	11.111,48	51.425,00	10.888,00	6,00	5,35	15
10/4/2000	379,00	7,35	208,60	1,74	11.186,56	49.987,00	7.806,00	6,00	5,38	14
11/4/2000	380,00	7,50	208,29	1,74	11.287,08	49.336,00	14.486,00	6,00	5,43	13
12/4/2000	382,00	7,75	210,75	1,76	11.125,13	42.883,00	13.203,00	6,00	5,46	12
13/4/2000	382,00	7,80	212,07	1,78	10.923,55	38.355,00	17.342,00	6,00	5,50	11
14/4/2000	385,00	7,81	211,11	1,79	10.305,77	34.674,00	13.418,00	6,00	5,54	10

(Continuação...)



21/9/2004	830,00	8,53	7,59	277,93	2,86	10.244,93	42.203,00	51.090,00	6,83	7,77	7,93	9
22/9/2004	850,00	8,61	7,76	279,49	2,88	10.109,18	34.222,00	64.365,00	6,84	7,77	7,94	8
23/9/2004	852,00	8,70	7,88	280,06	2,88	10.038,90	31.146,00	43.258,00	6,85	7,77	7,95	7
24/9/2004	860,00	8,75	7,79	277,77	2,87	10.047,24	27.463,00	98.230,00	6,85	7,76	7,96	6
27/9/2004	880,00	9,00	8,25	281,27	2,87	9.988,54	22.461,00	56.687,00	6,86	7,77	7,97	5
28/9/2004	885,00	8,95	8,28	281,73	2,87	10.077,40	17.324,00	48.992,00	6,87	7,80	7,97	4
29/9/2004	900,00	8,95	8,39	284,40	2,85	10.136,24	13.845,00	47.766,00	6,88	7,83	7,98	3
30/9/2004	900,00	8,92	8,50	284,98	2,86	10.080,27	13.845,00	47.766,00	6,88	7,85	7,98	2

Tabela 5.5 – Dados de entrada da RNA

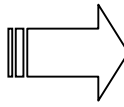
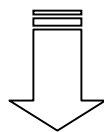
Data	Açúcar NYBOT		Data	Açúcar NYBOT
3/4/2000	5,91		27/9/2004	8,25
4/4/2000	5,81		28/9/2004	8,28
5/4/2000	6,02		29/9/2004	8,39
6/4/2000	5,96	(Continuação...)	30/9/2004	8,50

Tabela 5.6 – Dados dos valores desejados (*targets*) da RNA

Depois de entrar com os dados na rede, o primeiro passo para tratar esses dados é normalizá-los para ficarem com média zero e desvio padrão unitária.

Data	Alcool BMF	Açúcar BMF	Indice CRB	Real vs Dolar	Indice DJ	Cont. aberto	Volume Neg	MM- 200 dias	MM - 30 dias	Dias até Vcto
3/4/2000	-1,82	-0,92	-0,58	-1,70	1,22	-0,47	2,44	-1,19	-1,60	-0,52
4/4/2000	-1,88	-0,92	-0,65	-1,69	1,55	-0,60	0,32	-1,19	-1,59	-0,57
5/4/2000	-1,92	-0,89	-0,69	-1,68	1,49	-0,67	-0,40	-1,19	-1,57	-0,62
6/4/2000	-1,95	-0,87	-0,66	-1,67	1,34	-0,73	0,14	-1,19	-1,55	-0,66

(Continuação...)



27/9/2004	1,33	-0,09	2,05	0,39	0,19	-1,58	3,22	-0,43	0,23	-1,21
28/9/2004	1,37	-0,11	2,06	0,38	0,29	-1,71	2,61	-0,42	0,25	-1,26
29/9/2004	1,46	-0,11	2,17	0,35	0,35	-1,80	2,51	-0,42	0,27	-1,31
30/9/2004	1,46	-0,13	2,19	0,36	0,29	-1,80	2,51	-0,42	0,29	-1,35

Tabela 5.6 – Dados de entrada normalizados.

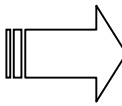
Data	Açúcar NYBOT		Data	Açúcar NYBOT
3/4/2000	-1,11		27/9/2004	0,54
4/4/2000	-1,11		28/9/2004	0,56
5/4/2000	-1,18		29/9/2004	0,63
6/4/2000	-1,03	(Continuação...)	30/9/2004	0,71

Tabela 5.7 – *Targets* normalizados.

5.5.2 Treinamento

Depois de normalizar todos os dados, se inicia o processo de treinamento para encontrar o conjunto de pesos que melhor se ajuste a série de dados apresentada. Ao total são utilizados 1131 dias de dados (3 de abril/2000 até 30 de setembro/2004), desse total metade é utilizado para treinamento, um quarto é utilizado para validação dos dados e os 25% restante dos dados são utilizados para as previsões propriamente ditas.

O gráfico da figura 5.7 exemplifica bem essas três fases (treinamento, validação e previsão), mostrando como o erro quadrático médio vai diminuindo ao longo do processo de ajustes de peso.

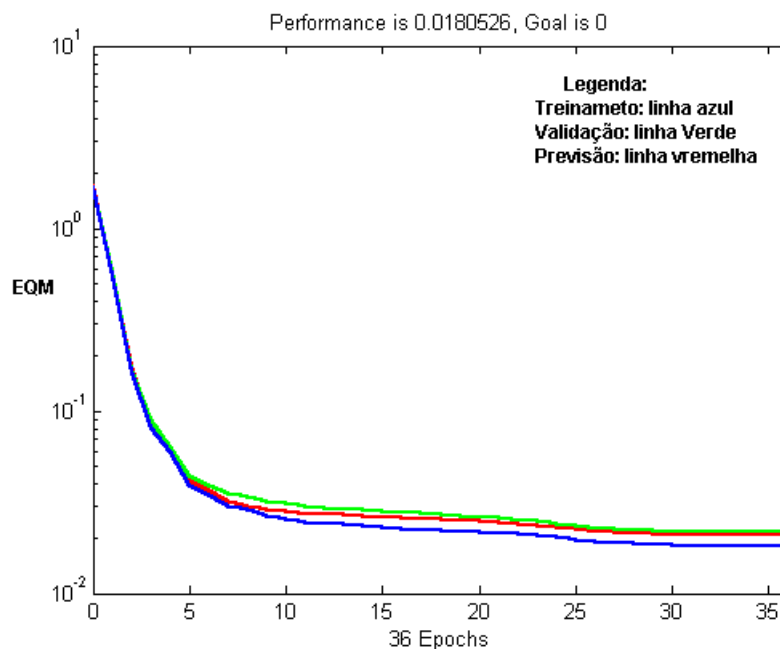


Figura 5.7 – Acompanhamento do EQM durante as três fases (treinamento, validação e previsão)

5.5.3 Saída da Rede

Por fim, os resultados da previsão ficam ajustados e são liberados pela rede neural. Do mesmo modo que o início se teve que normalizar os dados, na saída os dados também tem que passar por um tratamento para voltarem a seus formatos originais.

Data	Previsão de preços do açúcar NYBOT
3/4/2000	7,89
4/4/2000	7,92
5/4/2000	7,98
6/4/2000	8,08

(Continuação...)

Data	Previsão de preços do açúcar NYBOT
27/9/2004	8,16
28/9/2004	8,25
29/9/2004	8,41
30/9/2004	8,42

Tabela 5.8 – Saída da RNA.

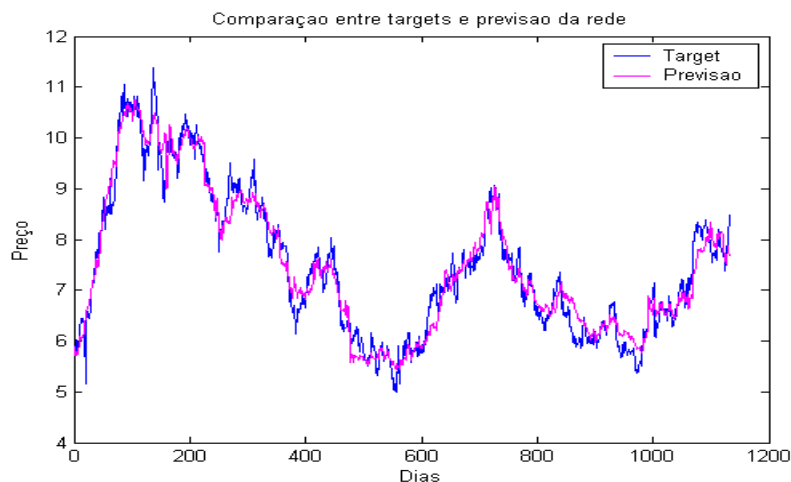


Figura 5.8 – Gráfico de Previsão vs Real

Com isso é possível construir um gráfico (veja figura 5.8), no qual se pode comparar os preços previstos com os preços reais.

5.5.4. Resultados das Redes Neurais

Do mesmo modo que foi mostrado nos itens acima, esse processo é realizado em todos os testes (variando o número de camadas escondidas, variando os números de neurônios variando, etc) com o objetivo de obter saídas diferentes e tentar encontrar àquela que melhor consiga prever o comportamento do preço do contrato futuro de açúcar NYBOT. Depois de ter em posse todas as saídas da rede, iniciar-se-á o processo de análise dos resultados através das medidas de desempenho.

A tabela 5.9 mostra todos um exemplo de teste que foi realizado. Nessa tabela pode-se observar, de forma clara, todos os resultados atingidos por essa configuração de rede (2 camadas escondidas com 3 neurônios nela). A realização de 10 testes de cada configuração foi para assegurar que os resultados obtidos não foram fruto do acaso, tendo todos os 10 testes dessa configuração tira-se a médias desses resultados para se ter números mais consistentes.

Preços diários - 10 entradas - 2 hidden c/ 3neuronios				
Teste	EQM	R	Retorno Financeiro	Lucro Estimado
1	0,1339	98,20%	R\$ 106,77	9,03
2	0,0936	97,60%	R\$ 105,71	1,95
3	0,0936	97,80%	R\$ 105,71	1,95
4	0,1059	97,50%	R\$ 105,56	0,62
5	0,0805	98,40%	R\$ 116,37	1,06
6	0,0532	97,90%	R\$ 116,72	2,68
7	0,0510	98,30%	R\$ 110,47	0,64
8	0,0790	97,60%	R\$ 105,22	2,98
9	0,0690	98,60%	R\$ 109,44	2,01
10	0,1025	96,70%	R\$ 106,95	1,26
Média	0,0862	97,86%	R\$ 108,89	2,42

Tabela 5.9 – Resultados da Rede Neural.

A tabela 5.10 mostra os resultados (média dos 10 testes) obtidos em todas configurações de rede.

Nº camadas escondidas	Nº neuronios nessa camada	EQM	R	Retorno Financeiro	Lucro Estimado
1	3	0,1072	98,10%	R\$ 103,68	1,02
1	4	0,0580	98,33%	R\$ 116,09	2,01
1	5	0,0429	98,70%	R\$ 110,30	1,29
1	6	0,0280	98,82%	R\$ 122,76	3,14
2	3	0,0862	97,86%	R\$ 108,89	2,42
2	4	0,0768	98,24%	R\$ 111,87	1,59
2	5	0,0399	98,50%	R\$ 114,76	2,19
2	6	0,0665	98,65%	R\$ 112,14	1,40

Tabela 5.10 – Todos os resultados da Previsão da RNA

Pelos resultados fica claro que a rede com 1 camada escondida, composta por 6 neurônios, obteve os melhores resultados de previsão do preço futuro do açúcar. Essa configuração conseguiu o melhor resultado em todas as mediadas de desempenho

testadas, conseguindo obter um lucro de 22,76% em 3 meses de teste, o que é um número bem expressivo em termos financeiros.

5.6. Aplicação Prática dos Resultados (Exemplo de Hedge dinâmico)

Imagine que um produtor ou usineiro, que daqui a 3 meses vai ter sua safra pronta para colheita, e decide se proteger de futuras oscilações do preço do açúcar vendendo contratos futuro de açúcar da NYBOT. Utilizando-se desses instrumentos de previsão, ele poderá fazer ajustes na sua posição para se proteger das oscilações dos preços ao longo do tempo, e ainda obter ganhos financeiros durante o processo. Isso possibilitará que ele tenha mais liberdade e tranquilidade na hora de ter que escolher quanto de açúcar e de álcool irá produzir, podendo se utilizar dessas técnicas apresentadas no trabalho para aumentar seu ganho futuro.

6. CONCLUSÕES

Os resultados obtidos no capítulo anterior mostraram o alto poder de predição das redes neurais, principalmente em identificar padrões no comportamento de séries temporais de preços de ativos financeiros. Mostraram também que diferentes configurações das redes neurais produzem resultados diversos, necessitando assim de inúmeros testes para identificar àquela configuração que melhor se adapte ao seu modelo. Mesmo, apresentado diferenças, os modelos de redes neurais se mostraram ser muito superiores aos modelos feitos com conceitos de média móvel.

Pelos resultados do modelo de previsão de redes neurais conseguiu-se prever de forma razoável o comportamento dos preços de contratos futuros de açúcar, podendo ser utilizadas em inúmeras ocasiões e por diferentes públicos, entre eles, por produtores e usineiros para se protegerem da oscilação futura do preço de açúcar, que era proposta inicial do trabalho.

Para complementar as conclusões desse trabalho, seria interessante tentar aplicar esse modelo aos preços futuros de álcool, podendo assim obter informações mais completas dos dois produtos provenientes da cana-de-açúcar.

Por melhor que tenham sido os resultados, é importante frisar que o mercado financeiro continua sendo um mercado imprevisível, logo se tentar replicar o mesmo modelo em diferentes situações não existe garantia de se obter os mesmos resultados. Porém, o que se pode com certeza afirmar é a boa ferramenta que é as redes neurais para reconhecer padrões de séries temporais.

Esse trabalho com certeza abre caminhos para um maior aprofundamento nos estudos de ferramentas matemáticas para serem aplicadas em finanças. Uma série de outras teorias poderiam ser exploradas complementando os resultados desse trabalho, dentre elas um maior aprofundamento na teoria de redes neurais ou até utilizar-se de outras ferramentas avançadas de previsão como lógica *fuzzy*, teoria do caos, dinâmica não-linear, etc.

Aprofundar os estudos nesse sentido, vão com certeza contribuir para um maior reconhecimento de que a modelagem matemática pode ser muito bem utilizado no mercado financeiro.

7. Referências bibliográfica

ALVIN, A. C. **Previsão de demanda no varejo.** 2003. 95p. Trabalho de Formatura- Escola Politécnica, Universidade de São Paulo. São Paulo, 2003.

BAESTAENS, D.E., BERGH, W. M., WOOD, D. **Neural Network Solution for Trading in Financial Markets.** United London, Pitman Publishing, 1994.

BERGERSON, K e WUNSH II, D.C. **A Commodity Trading Model Base on a Neural Network - Expert System Híbrido.** In: NEURAL NETWORKS IN FINANCE AND INVESTMENTS, Robert Trippi and Efrain Turban, Editors. Chicago: Irwin, pp.589-611, 1996.

BRAGA, A., LUDEMIR, T., CARVALHO, A. **Redes Neurais Artificiais: Teoria e Aplicações.** Rio de Janeiro; Livros Técnicos e Científicos, 2000.

BUCHANANAN, W.K.; HODGES, P.; THEIS, J. **Witch way the natural gas price; an attempt to predict the direction of natural gas spot price movements using trader position.** In: ENERGY ECONOMICS 23, pg.279-293, 2001.

CAMPBELL, J. Y. et al. **The Econometrics of Financial Markets.** Princeton: Princeton University Press, 1997.

DEMUTH, H. e BEALE, M. **Neural Network Toolbox – for use with MATLAB – User’s Guide Version 4.** Natick. The Mathworks, Inc., 2001.

FALHMAN, S.E. **Comments on comp.ai.neural-nets.** item 2198 – 1992.

FERNANDES, L. G. e PORTUGAL, M. S. **Redes Neurais Artificiais e Previsão de Séries Econômicas: Uma Introdução.** Textos para discussão n.95/1, Curso de Pós-Graduação em Economia, UFRGS 1995

FU, L. **Neural Networks in Computer Finance.** Florida. McGraw-Hill, Inc. 1994.

HARVEY, A. C. **Forecasting, structural time series models and the Kalman filter.** 1st paperback ed. Cambridge. New York, Cambridge University Press, 1990, c1989.

HAYKIN, S. **Redes Neurais Princípios e Fundamentos.** São Paulo. Bookman, 2001.

MAKRIDAKIS, S.; WHEELWRIGHT, S.; HYNDMAN, R. **Forecasting: Methods and Applications**. 3rd edition. New York, John Wiley & Sons, 1998.

MONTGOMERY, D. C. e JOHNSON, L. A. **Forecasting and time series analysis**. New York: McGraw-Hill, c1976.

MOREIRA, J. R. e GOLDEMBERG, J. **The alcohol program**. In: ENERGY POLICY 27, pg. 229-425, 1999.

MORETTIN, P. A. e TOLOI, C. M. **Series temporais**. 2^a ed. São Paulo. Atual, 1987.

PARREIRAS, L.R.F. **Modelo Genético- Neural de Gestão de uma carteira de ações**. 2003. 138p. Trabalho de Formatura- Escola Politécnica, Universidade de São Paulo. São Paulo, 2003.

PORTER, M. E. **Competitive Strategy: techniques for analysing industries and competitors**. New York. Free Press c1980

REFENES, A. P. **Neural Network in the Capital Market**. England: John Wiley & Sons Ltd. 1995

RIBEIRO, C. O. **Um modelo de previsão hierárquico com indicadores técnicos**. 1997. 130p. Tese (Doutorado) - Escola Politécnica, Universidade de São Paulo. São Paulo, 1997.

SMITH, C.E. **Exchange rate variation, commodity price variation and implications for international trade**. In: JOURNAL OF INTERNATIONAL MONEY AND FINANCE 18, pg. 471-491, 1997.

SORENSEN, C. **Modeling seasonality in agricultural commodity futures**. In: THE JOURNAL OF FUTURES MARKETS 22(5), pg. 393-426; ABI/INFORM Global; may 2002.

ZHANG, P.G.; QI, Min. **Neural Network forecasting for seasonal and trend time series**. In: EUROPEAN JOURNAL OF OPERATION RESEARCH 160, pg. 501-514, nov 2003.