

UNIVERSIDADE DE SÃO PAULO  
ESCOLA POLITÉCNICA  
DEPARTAMENTO DE ENGENHARIA DE TELECOMUNICAÇÕES E  
CONTROLE

JARBAS AQUILES GAMBOGI

**APLICAÇÃO DE REDES NEURAIS NA TOMADA  
DE DECISÃO NO MERCADO DE AÇÕES**

São Paulo  
2013

JARBAS AQUILES GAMBOGI

# APLICAÇÃO DE REDES NEURAIS NA TOMADA DE DECISÃO NO MERCADO DE AÇÕES

Dissertação apresentada ao Departamento de Engenharia de Telecomunicações e Controle da Universidade de São Paulo para obtenção do título de Mestre em Engenharia Elétrica.

Área de Concentração: Engenharia de Sistemas  
Orientador: Prof. Dr. Oswaldo Luiz do Valle Costa

São Paulo  
2013



UNIVERSIDADE DE SÃO PAULO

## FOLHA DE APROVAÇÃO

# APLICAÇÃO DE REDES NEURAIIS NA TOMADA DE DECISÃO NO MERCADO DE AÇÕES

JARBAS AQUILES GAMBOGI

Dissertação apresentada ao Departamento de Engenharia de Telecomunicações e Controle da Universidade de São Paulo para obtenção do título de Mestre em Engenharia Elétrica.

Área de Concentração: Engenharia de Sistemas

Banca Examinadora constituída por:

Doutor OSWALDO LUIZ DO VALLE COSTA – Orientador  
Universidade de São Paulo

Doutor EMILIO DEL MORAL HERNANDEZ  
Universidade de São Paulo

Doutor WANDERLEI LIMA DE PAULO  
Universidade de São Paulo

São Paulo, 2013

# Dedicatória

À minha esposa Adriana pela sua compreensão, incentivo e apoio fundamentais para a conclusão deste trabalho.

# Agradecimentos

Ao Professor Dr. Oswaldo Luiz do Valle Costa pela paciência e disponibilidade na orientação desta Dissertação.

À Escola Politécnica da USP pela oportunidade de realização do curso de mestrado e a todos os professores do curso pelos ensinamentos transmitidos.

À Élia Yathie Matsumoto pelas profícuas discussões sobre o tema.

*“The market can stay irrational longer than you can stay solvent”.*

John Maynard Keynes

# Resumo

Este trabalho apresenta um sistema de *trading* que toma decisões de compra e de venda do índice Standard & Poor's 500, na modalidade seguidor de tendência, mediante o emprego de redes neurais artificiais multicamadas com propagação para frente, no período de 5 anos, encerrado na última semana do primeiro semestre de 2012. Geralmente o critério usual de escolha de redes neurais nas estimativas de preços de ativos financeiros é o do menor erro quadrático médio entre as estimativas e os valores observados. Na seleção das redes neurais foi empregado o critério do menor erro quadrático médio na amostra de teste, entre as redes neurais que apresentaram taxas de acertos nas previsões das oscilações semanais do índice Standard & Poor's 500 acima de 60% nessas amostras de teste. Esse critério possibilitou ao sistema de *trading* superar a taxa anual de retorno das redes neurais selecionadas pelo critério usual e, por larga margem, a estratégia de compre e segure no período. A escolha das variáveis de entrada das redes neurais recaiu entre as que capturaram o efeito da anomalia do momento dos preços do mercado de ações no curto prazo, fenômeno amplamente reconhecido na literatura financeira.

Palavras-chave: Redes Neurais, Mercado de Ações, Sistema de *Trading*

# Abstract

This work presents a trend follower system that makes decisions to buy and sell short the Standard & Poor's 500 Index, by using multilayer feedforward neural networks. It was considered a period of 5 years, ending in the last week of the first half of 2012. Usually a neural networks choice criterion to forecast financial asset prices is based on the least mean square error between the estimated and observed prices in the test samples. In this work we also adopted another criterion based on the least mean square error for those neural networks that had a hit rate above 60% of the Standard & Poor's 500 Index weekly change in the test sample. This criterion was shown to be the most appropriate one. The neural networks input variables were chosen among those technical indicators that better captured the anomaly of the short term momentum of prices. The annual rate of return of the trading system based on those criteria surpassed those selected by the usual criteria, and by a wide margin the buy-and-hold strategy. The neural networks inputs were chosen to capture the momentum anomaly of the prices on the short term that is fully recognized in the financial literature.

Keywords: Artificial Neural Network, Stock Market, Trading System



# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Revisão Bibliográfica</b>	<b>5</b>
2.1	Revisão Histórica . . . . .	5
2.2	Aplicações em Finanças . . . . .	8
<b>3</b>	<b>Redes Neurais Artificiais</b>	<b>16</b>
3.1	Introdução às Redes Neurais . . . . .	16
3.2	Arquitetura e Funções de Ativação . . . . .	17
3.3	Normalização dos Dados . . . . .	19
3.4	Treinamento e Critérios de Desempenho . . . . .	20
3.5	Algoritmo . . . . .	22
<b>4</b>	<b>Metodologia</b>	<b>24</b>
4.1	O Índice Standard & Poor's 500 . . . . .	24
4.2	Períodos de Treinamento, Validação, Teste e de Verificação . . . . .	26
4.3	Funcionamento do Sistema de <i>Trading</i> . . . . .	27
4.4	Seleção de Atributos . . . . .	30
4.5	Variáveis de Entrada das Redes Neurais do Sistema de <i>Trading</i> . . . . .	34
4.6	A Técnica do Janelamento Móvel . . . . .	36
4.7	Critério de Escolha da Rede Neural . . . . .	37
4.7.1	Primeira Etapa da Seleção das Configurações . . . . .	38
4.7.2	Segunda Etapa da Seleção da Configuração e Redes Neurais . . . . .	40

<b>5</b>	<b>Resultados dos Experimentos</b>	<b>41</b>
5.1	Pré-Seleção das Configurações com Janela Única . . . . .	41
5.2	Seleção da Configuração, do Tamanho da Janela de Verificação e das Amostras de Treinamento, Validação e Teste e das Redes Neurais . . .	44
5.3	Métricas de Desempenho do Sistema de <i>Trading</i> . . . . .	46
<b>6</b>	<b>Conclusões e Recomendações</b>	<b>51</b>
	<b>Referências Bibliográficas</b>	<b>54</b>
	<b>APÊNDICE A</b>	<b>59</b>
	<b>APÊNDICE B</b>	<b>62</b>
	<b>APÊNDICE C</b>	<b>65</b>

# Lista de Figuras

3.1	Arquitetura com uma camada escondida . . . . .	18
4.1	Técnica do janelamento móvel . . . . .	37
5.1	RMSE_R e retornos anuais com janelamento único . . . . .	42
5.2	Evolução da carteira do sistema de <i>trading</i> . . . . .	43
5.3	Evolução da carteira do sistema de <i>trading</i> . . . . .	47
5.4	Distribuição dos retornos mensais . . . . .	49

# Lista de Tabelas

5.1	Resultados com janelamento único . . . . .	42
5.2	Resultados com janelamento Móvel . . . . .	44
5.3	Resultados por critério e configuração . . . . .	46
5.4	Resultados do sistema de <i>trading</i> . . . . .	46
1	Tabulação parcial dos resultados . . . . .	62

# Capítulo 1

## Introdução

A previsão do comportamento dos mercado financeiro sempre foi alvo de grande interesse pelos investidores, pois avaliações acuradas se traduzem em elevados retornos. Os mercados financeiros são movidos por inúmeras variáveis macroeconômicas, inclusive por expectativas irracionais dos investidores Shiller (2005). Embora o acesso às informações tenha aumentado significativamente com o advento da internet, a tarefa de interpretá-las, de avaliar os seus impactos e de apurar o que já foi apreçado pelos mercados é muito complexa e árdua, inclusive para os investidores institucionais que contam com recursos materiais e humanos altamente especializados. Por essa razão redes neurais artificiais, muitas vezes integradas com outras ferramentas da inteligência artificial, têm sido amplamente estudadas para servirem de instrumentos de apoio nas decisões sobre investimentos devido à capacidade de mapear funções complexas.

O presente trabalho tem o objetivo de propor um sistema de *trading* seguidor de tendência (*trend follower*) de curto prazo que apresente boas métricas de desempenho financeiro, mediante o emprego de redes neurais artificiais multicamadas com propagação para frente supervisionadas, aplicadas no mercado de ações norte-americano. Um sistema seguidor de tendência é inerentemente volátil, todavia o que se espera é que a

curva do capital empregado (*equity line*) ao longo do tempo seja consistentemente ascendente e que não dependa de operações isoladas, o que é possível com uma boa taxa de acerto nas previsões das oscilações do ativo financeiro objeto das transações. Neste trabalho o sistema de *trading* foi testado no índice Standard & Poor's 500 durante o período de cerca de 5 anos, compreendido entre as semanas findas em 13/07/2007 e 29/06/2012, com operações realizadas na frequência semanal, baseadas em previsões feitas por redes neurais sobre o comportamento do índice com uma semana de antecedência. O sistema está sempre exposto ao mercado quer na ponta de compra do índice quanto na de venda. Com a finalidade de avaliar o desempenho da rede neural não foi utilizado mecanismo de perda máxima (*stop loss*) que interrompe uma operação quando o prejuízo ultrapassa um limite previamente fixado, presente nos sistemas operacionais utilizados no cotidiano e, tampouco, se fez uso de alavancagem, artifício corriqueiramente utilizado nos fundos multimercados (*hedge funds*). Para a configuração vencedora (entradas e número de neurônios na camada escondida), entre as testadas, foram exploradas as inclusões do custo de corretagem (praticadas no mercado brasileiro) e alavancagem nas operações (*leverage*).

Geralmente a escolha das redes neurais utilizadas como aproximadoras de funções é baseada no critério do menor erro quadrático médio (MSE) entre os valores observados, ou em métricas similares, e os previstos pelas redes neurais nas amostras de teste (*out of sample*). Embora seja desejável acurácia nas previsões dos preços de um ativo financeiro, um sistema de *trading* somente é ganhador se acertar as oscilações dos preços do ativo, a despeito de diferenças materiais entre as previsões e as cotações observadas. Este trabalho propõe que a seleção de redes neurais seja feita pelo critério da menor raiz quadrada do erro quadrático médio relativo (RMSE\_R), métrica similar à do menor erro quadrático médio, pertencente ao subconjunto das simulações que registraram taxas de acerto das oscilações semanais do índice Standard & Poor's 500 superiores a 60% nas amostras de teste. No período estudado esse critério sugerido apresentou resultado financeiro superior ao usualmente empregado.

Na literatura não há um consenso sobre as variáveis de entrada mais apropriadas de uma rede neural com a finalidade deste trabalho. Na maioria dos experimentos, que empregam redes neurais na predição de preços no curto prazo, as entradas utilizadas têm sido tomadas emprestadas da Escola Técnica (indicadores técnicos), tendo sido raros os autores que utilizaram indicadores macroeconômicos obtidos através de pesquisas (*survey*). Os indicadores técnicos são as informações extraídas do próprio mercado, tais como preço, volume de transações, contratos em aberto de derivativos do ativo objeto, entre outros. Os indicadores macroeconômicos frequentemente sofrem revisões, pois normalmente os dados divulgados são preliminares, a metodologia costuma ser alterada e, conseqüentemente, toda a base de dados, e a frequência mínima de divulgação é semanal. Por outro lado, os dados macroeconômicos apreçados pelo mercado diariamente, como taxa de juros, não apresentam as inconveniências mencionadas, embora também tenham sido pouco utilizados. Esses últimos também apresentam problemas que serão discutidos no Capítulo 4 que trata sobre a metodologia empregada neste trabalho. As variáveis de entrada das redes neurais foram escolhidas com base na tese de que a sequência dos preços apresenta o viés de realimentar a tendência no curto prazo, fenômeno conhecido como anomalia do momento dos preços, embora esse efeito tende a se extinguir no longo prazo. Jegadeesh e Titman (1993) se referem à anomalia do efeito momento, que colide com a Hipótese do Mercado Eficiente (HME) em sua forma fraca proposta por Fama (1970), ao constatar uma diferença de 1,5% ao mês entre os retornos médios das ações com elevados e baixos momentos medidos nos 12 meses precedentes. No longo prazo, contudo, esse efeito desaparece e é revertido. Por outro lado, Fama e Litterman (2012) reconheceram que o efeito momento nos preços das ações é o que mais tem embaraçado a HME. Pesquisadores do recente ramo conhecido como Finanças Comportamentais têm observado que sistematicamente as decisões dos investidores se desviam dos postulados da HME. Rau (2010) afirma que os vieses dos investidores são sistemáticos e previsíveis. Um dos postulados da Escola Técnica, em linha com a tese da Finanças Comportamentais, reza que os preços dos

ativos financeiros se movem em tendências (Murphy (1999)) que, uma vez instaladas, o mais provável é que elas se mantenham ao invés de serem revertidas.

Este trabalho está dividido em 6 capítulos, incluída esta introdução. O Capítulo 2 apresenta a revisão bibliográfica sobre redes neurais e discorre sobre a evolução das técnicas empregadas, focada nas aplicações na área financeira. O Capítulo 3 descreve de forma sumária uma rede neural artificial. O Capítulo 4 aborda a metodologia usada nos experimentos do sistema de *trading*. O Capítulo 5 expõe os resultados dos experimentos e a análise e o Capítulo 6 as conclusões e recomendações finais.



# Capítulo 2

## Revisão Bibliográfica

Neste capítulo a Seção 2.1 é dedicada à revisão da literatura sobre a evolução das redes neurais artificiais e a Seção 2.2 é voltada exclusivamente às aplicações no ramo financeiro.

### 2.1 Revisão Histórica

Segundo Haykin (2009) em 1911 Ramón y Cajál introduziu a ideia de que os neurônios compõem a estrutura de um cérebro. McCulloch e Pitts (1943) descreveram o mecanismo de funcionamento dos neurônios e propuseram que os eventos neurológicos e as relações entre eles poderiam ser tratados por meio de uma lógica proposicional. Hebb (1949) desenvolveu o conceito fundamental de uma lei de aprendizagem específica que passou a ser conhecida como o Postulado de Aprendizagem de Hebb. Por essa lei quando um axônio de uma célula A está suficientemente próximo de uma célula B para excitá-la e repetidamente leva cabo a excitação, algum processo de crescimento ou mudança metabólica ocorre em uma ou em ambas as células de tal modo que a

eficiência da célula A em excitar a célula B aumenta. Em 1959 Bernard Widrow e Marcian Hoff desenvolveram o ADALINE (Adptive Linear Combiner), cujos pesos e vieses eram atualizados durante o treinamento na proporção do erro da estimação, técnica conhecida como regra do Delta. Entretanto, esse dispositivo somente era capaz de resolver problemas linearmente separáveis. Rosenblatt (1958) desenvolveu uma rede neural usada para classificar padrões linearmente separáveis que consistia de apenas um neurônio com pesos e vieses sinápticos ajustáveis, conhecido como Perceptron. Minsky e Papert (1969) demonstraram através de uma abordagem matemática as limitações do Perceptron. As críticas de Minsky e Papert tiveram um impacto negativo sobre os pesquisadores a ponto de por fim a um ciclo de entusiasmo que somente foi restabelecido na década de 80, quando em 1988 esses autores reconheceram avanços promissores na área que reacenderam o interesse sobre redes neurais numa era em que grandes empresas como a IBM desenvolviam computadores pessoais mais poderosos que foram introduzidos no meio empresarial. Werbos (1974) propôs em sua tese de doutorado a técnica do *backpropagation* que consiste em um método de cálculo, implementada no modo reverso, da derivada do erro entre a saída desejada e a da rede neural em relação aos seus pesos e vieses, com a finalidade de ajustá-los. O termo *backpropagation* é frequentemente usado para denotar o algoritmo do gradiente descendente implementado pela técnica do *backpropagation* (Haykin (2009)). Essa terminologia pode gerar confusões uma vez que uma vasta gama de algoritmos usam essa técnica. Essa técnica somente foi popularizada em meados da década de 80 por Rumelhart et al. (1986), que sugeriu incluir na regra do Delta um termo, denominado *momentum*, para aumentar a taxa de aprendizagem e a convergência do algoritmo, além de evitar o risco de instabilidade. O algoritmo *backpropagation* superou as falhas do Perceptron anteriormente apontadas por Minsky e Papert ao viabilizar o aprendizado de redes neurais multicamadas. Esse algoritmo se baseia na minimização do erro entre a saída desejada e a observada mediante a aplicação da regra do Delta, em que o vetor erro caminha na direção oposta ao do gradiente, cujo cálculo é feito pelo método *backpropagation*

anteriormente comentado. Hopfield (1982) propôs uma rede neural, para a memorização de imagens, composta por um conjunto de neurônios cujas saídas realimentam os demais. O treinamento consiste no armazenamento de padrões por meio de memória associativa para posterior reconhecimento de imagens distorcidas. Kohonen (1982) sugeriu um novo método de auto-organização que consiste na transformação de vetores de entrada com dimensões arbitrárias em uma mapa topológico, tipicamente de 2 dimensões. Segundo Haykin (2009) o modelo de Kohonen foi precedido pelo de Willshaw-von Malsburg, porém ele é mais geral no desempenho da redução da dimensionalidade dos vetores de entrada. Esses 3 modelos talvez tenham sido os mais influentes nas décadas de 80 e 90. Entretanto, outros tipos de modelos surgiram, tais como a rede de função de base radial (Park e Sandberg (1991) e Chng et al. (1996)), redes com transformada *wavelet* (Delyon et al. (1996)), a rede polinomial (Shin e Ghosh (1995)) e a recorrente (Connor et al. (1994)). As novas concepções sobre redes neurais estimularam o Instituto de Engenheiros Elétricos e Eletrônicos (IEEE-Institute of Electrical and Electronic Engineers) a patrocinar a partir de 1987 eventos anuais sobre redes neurais. Ao mesmo tempo muitos periódicos surgiram nos anos subsequentes.

Maass (1997) propôs a divisão dos estágios de desenvolvimento das redes neurais artificiais em 3 gerações. Segundo o autor a primeira geração consistiu de neurônios (*McCulloch-Pitts threshold neurons*) que enviam um sinal binário se a soma de suas entradas ponderadas superar os limiares (*threshold*) de tensão. Neurônios da segunda geração não incorporam esse mecanismo para computar os sinais de saída, mas funções de ativação contínuas, mais comumente as funções sigmoidais. Essa geração é computacionalmente mais poderosa do que a primeira devido à superioridade de sua capacidade de aprendizagem. Exemplos típicos dessa geração são as redes neurais multicamadas com propagação para a frente e as recorrentes. Novas descobertas pela neurociência encorajaram a exploração de alternativas sobre os esquemas de codificação das redes neurais. A terceira geração, as redes neurais pulsadas, usa códigos de pulso (*spike*), mecanismos pelos quais os neurônios recebem e enviam pulsos (informações),

empacotando-os em um meio de transmissão único. Ainda segundo Maass (1997), essas redes neurais podem aumentar substancialmente a quantidade de informações transmitida com maior velocidade e eficiência, computar qualquer função da segunda geração com menor número de neurônios e, por serem mais plausíveis biologicamente, poderiam usufruir dos avanços na neurociência. Muitas linhas de pesquisas ainda estão abertas; um dos desafios é descobrir regras de aprendizagem eficientes que poderiam extrair as vantagens das redes neurais pulsadas. Essa nova concepção ainda não foi aplicada no ramo das finanças, até onde se sabe. Maiores informações sobre essa nova geração podem ser encontradas em Maass e Bishop (2001), trabalho que reuniu dezenas de colaboradores.

## 2.2 Aplicações em Finanças

A popularização do algoritmo *backpropagation* na década de 80 gerou um fértil período nas pesquisas sobre aplicações de redes neurais em finanças nos anos 90. Contribuições em **previsões de falências** incluem Odom e Sharda (1990), pioneiros nessa aplicação, que compararam os resultados com a análise de discriminante, com vantagens para a rede neural MLP (*Multilayer Perceptron*); Salchenberger et al. (1992) que avaliaram os desempenhos da rede neural MLP com a regressão logística e constataram vantagens da primeira opção; Tam e Kiang (1992) que confrontaram os resultados com a análise de discriminante, regressão logística e com o algoritmo k-vizinhos mais próximos, todos com desempenhos inferiores ao da rede MLP; e Back et al. (1996) que empregaram algoritmo genético para selecionar os atributos dos vetores de entrada da rede MLP. Em **classificações do risco de títulos de crédito** corporativos (setor de telefonia) Singleton e Surkan (1990) empregaram uma rede neural MLP (*backpropagation*) e a análise múltipla de discriminante com vantagens para a primeira. Em **previsões de cotações de ativos financeiros** (ações da IBM) um dos primeiros

trabalhos, com ambições mais prospectivas, foi realizado por White (1988), que reconheceu as limitações do escopo de seu trabalho. Esse autor sugeriu indicadores técnicos e macroeconômicos antecedentes como atributos dos vetores de entrada da rede e o uso de redes recorrentes. Zhang e Hu (1998) utilizou uma rede MLP para previsão da cotação da libra esterlina frente ao dólar e concluiu que a seleção dos atributos dos vetores de entrada e a arquitetura são os elementos críticos para melhorar o poder das predições no mercado de moedas. Roman e Jameel (1996) empregou uma rede neural recorrente (Elman) na previsão dos retornos de índices representativos de diversos mercados internacionais de ações e concluiu que, embora seja mais apta a capturar informações temporais do que a rede MLP (*backpropagation*), os resultados não foram substancialmente superiores. Uma possível razão sugerida pelo autor teria sido a deficiência da base de dados que poderia não conter informações suficientes para melhorar a qualidade da predição. Malliaris e Salchenberger (1996), mediante uso do *backpropagation* modificado (*fast backpropagation*), utilizou uma rede neural MLP na **previsão da volatilidade implícita das opções** lançadas sobre o índice Standard & Poor's 100 e apontou vantagens sobre o tradicional modelo *Black-Sholes*. Kryzanowski et al. (1993) usou o algoritmo de classificação de padrões (ações com desempenho positivo versus negativo) da rede neural (binária) Boltzmann, que emprega a técnica de otimização estocástica denominada *simulated annealing*, para **alocação de carteira** (*portfolio*) tendo como entradas dados contábeis e indicadores macroeconômicos nos 4 anos precedentes. O autor relatou taxa de acerto médio de 72%. Trabalho que ganhou notoriedade no início dos anos 90 foi o de Kimoto et al. (1990) que usou redes MLP e algoritmo *backpropagation* com taxa de aprendizagem modificada. Foi utilizada uma arquitetura modular com variáveis de entrada compostas por indicadores técnicos e macroeconômicos. As saídas de cada rede neural eram binárias (compra e venda) e a média desses valores foi utilizada como regra - abaixo de 0,5 era compra e acima era venda - para as operações de um **sistema de trading** para o índice Topix (Japão), sendo um dos primeiros a avaliar a curva do capital empre-

gado (*equity line*) e utilizar a técnica do janelamento móvel. Além de bater o retorno da estratégia de compre e segure a metodologia empregada superou à da regressão linear múltipla. Kaastra e Boyd (1996) propôs um **guia de procedimentos**, que tem servido como referência até os dias atuais, constituído de 8 etapas que deveriam ser observadas em modelos de previsão, com sugestões para a seleção das variáveis, coleta de dados, pré-processamento dos dados, treinamento e validação, topologia, critérios de avaliação de desempenho, parâmetros do algoritmo *backpropagation* e implementação. Zhang et al. (1998) realizou **comparações de resultados** de predições obtidos por vários autores que empregaram redes MLP em diversas áreas, inclusive financeira, e modelos lineares e relatou discrepâncias e inconsistências. Em alguns casos em que as redes MLP obtiveram resultados inferiores deveu-se, segundo o autor, à linearidade da função mapeada. Em outros casos atestou que a razão recaiu em más escolhas das estruturas em relação à natureza dos problemas. De maneira geral, na década de 90 prevaleceu o uso de redes MLP com o algoritmo *backpropagation*, com diferentes variações, e no caso particular de previsão de preços de ativos financeiros as seleções das entradas eram feitas por tentativa e erro, a partir de uma lista padrão de indicadores técnicos, sem o respaldo na literatura financeira. Os critérios de desempenho populares para a escolha das redes eram, frequentemente, o MSE (*mean square error*), RMSE (*root mean square error*), MAE (*mean absolute error*), o MAPE (*mean absolute percentage error*) e o SSE (*sum of square error*).

No parágrafo anterior foram relacionadas algumas das aplicações mais utilizadas na área financeira. Com o foco direcionado apenas para os novos modelos de predição de preços de ativos financeiros e sistemas de *trading*, que é o alvo deste trabalho, na década de 2000 em diante a tendência foi criar redes neurais híbridas com contribuições de outras técnicas de inteligência artificial e estatísticas. Embora não tenha mantido o mesmo *status* o algoritmo *backpropagation* continuou sendo muito usado, com variações entre as quais a que inclui funções de custo que penalizam erros de predição da direção do preço do ativo financeiro. Os algoritmos *Levenberg-Marquardt*, *Bayesian*

*Regularization* e genéticos ganharam mais espaço, nesse último caso tanto como algoritmo para os ajustes dos pesos e vieses, como na extração de atributos dos vetores de entrada e na otimização da topologia da rede. Os critérios de seleção de redes em previsões de preços de ativos financeiros continuaram sendo os tradicionais, embora os de avaliação do desempenho incorporaram a taxa de acerto em muitos estudos, especialmente os mais recentes, porém sem ser o indicador decisivo na seleção da rede neural. A preferência por redes supervisionadas com propagação para a frente (MLP) se manteve, embora as recorrentes ganharam mais espaço. A combinação de redes neurais com a técnica *fuzzy* também mereceu atenção na literatura.

Raposo e Cruz (2002) empregaram uma rede neural *fuzzy* para a **alocação de carteira** de ações. Um conjunto de atributos de entradas foi constituído de indicadores técnicos e financeiros recomendados por especialistas da área e a dimensão desse conjunto foi reduzida pela técnica da Análise de Componentes Principais (PCA). Foi usada uma rede MLP de 5 camadas, treinada trimestralmente (periodicidade da divulgação dos balanços). A segunda camada mapeou os graus de pertinência (fuzzificação) das entradas (primeira camada), a terceira as regras de fuzzificação e a quarta a defuzzificação. A saída da rede (quinta camada) gerou 3 sinais, quais sejam, de compra, venda e manutenção, conforme a oscilação dos preços das ações superavam um intervalo de preços, ficavam abaixo ou se mantinham dentro dele, respectivamente. Na fase de teste a taxa de acerto atingiu 70% para o modelo composto pelos indicadores fornecidos pelos especialistas e cerca de 78% após a extração dos atributos pelo método PCA. Kim (2004) sugeriu uma rede neural MLP *backpropagation* para a **predição** do índice KOSPI 200 da Coreia do Sul na frequência diária. Os atributos, compostos por indicadores técnicos, foram selecionados inicialmente por um algoritmo genético entre um rol de indicadores. Em seguida foi efetuada uma transformação que criou um novo conjunto de atributos com base em conhecimento *a priori*, cujo desempenho foi superior ao da transformação linear dos atributos. Na amostra de teste obteve-se uma taxa de acerto na previsão pelo modelo proposto de 73% e de 57% pelo da transformação

linear. O autor reconheceu como limitações do método proposto eventual subjetivismo na transformação com conhecimento *a priori* e o fato de não ter sido levado em conta a correlação entre as variáveis explicativas e a dependente. Embora não tenha sido explicitado o critério de seleção das redes neurais testadas a taxa de acerto na direção dos preços foi o único critério de avaliação divulgado. Klassen (2005) investigou a eficácia do uso de alguns indicadores técnicos na predição dos preços semanais de fechamento dos índices norte-americanos Dow Jones Industrial e Nasdaq Composite, combinados com os últimos 4 preços de fechamento semanais (ao todo 5 entradas), sem que tenha sido justificada a escolha desse rol de entradas. O critério usado na avaliação foi a diferença percentual das previsões dos preços em relação aos observados e concluiu que o MACD (*Moving Average Convergence and Divergence*) era o mais recomendado (detalhes sobre o MACD no Capítulo sobre a metodologia). Nos experimentos a autora empregou uma rede MLP e o algoritmo *Levenberg-Marquardt*. Nesta dissertação as entradas escolhidas para a rede neural foram inspiradas nas conclusões dessa autora, pois os resultados foram de encontro com as constatações de Fama (1970) sobre a existência da anomalia do efeito momento dos preços, que as entradas selecionadas capturam. Matsumoto (2009) empregou uma rede neural MLP, com camada de saída de 3 neurônios e algoritmo *Bayesian Regularization*, para a **classificação da rentabilidade futura** de empresas brasileiras entre ruim, média e ótima. As entradas foram constituídas de indicadores dos balanços patrimoniais que apresentaram os melhores testes estatísticos para os modelos da regressão linear múltipla e logística ordenada (*ordered logistic regression*). Na amostra de verificação a rede neural apresentou taxa de acerto de 70,7% contra 69,8% pela regressão logística e 35,1% pela regressão linear múltipla. O método VAR (*value at risk*) é muito usado pelos gestores de carteiras de investimento (*portfolio*) para estimar, em um determinado horizonte de tempo, a perda máxima dentro de um intervalo de confiança devido à exposição aos riscos do mercado. Na **estimativa do VAR**, Chang et al. (2008) usou o modelo ARMA (*Auto regressive Moving Average*) para extrair as variáveis estatisticamente significantes,



compostas de indicadores técnicos, e servirem como entradas de uma rede neural MLP *backpropagation* na predição de índices de ações. Tilakaratne et al. (2009) levantou um dos problemas que aflige os usuários de redes neurais MLP que consiste em diferentes soluções possíveis de serem encontradas, devido à randomização dos pesos e vieses no início do treinamento e da possibilidade de que esse seja encerrado em diferentes mínimos locais. A solução sugerida foi usar o algoritmo AGOP, proposto por Mammadov (2004), que tem o objetivo de buscar uma **solução global na minimização da função de custo MSE** (*mean square error*). Além disso, essa função de custo foi modificada para corrigir os parâmetros livres (pesos e vieses) pelo acerto das predições da direção do preço do ativo financeiro, ao invés de uma penalização proposta por outros autores. Essa versão foi aplicada ao índice de ações AORD, representativo do mercado de ações australiano, e o critério de avaliação de desempenho foi a taxa média de acerto nas predições da direção do índice nas 6 janelas móveis usadas no período de predição. Segundo os autores os resultados obtidos foram ligeiramente superiores aos da abordagem convencional. Martinez et al. (2009) afirmaram que avaliar a contribuição de uma rede neural para um investidor pela acurácia na predição de preços de ativos financeiros não é o caminho adequado, sendo que o mais recomendado seria através de um **sistema de *trading***, pelo qual a rede neural poderia ser avaliada por métricas financeiras. Constatou ainda que a literatura apresenta uma lacuna sobre uma abordagem sistemática na seleção das variáveis de entrada da rede. Após uma avaliação preliminar foram escolhidas como entradas os preços de abertura, máximo, mínimo e fechamento nos 5 períodos precedentes à predição e alguns indicadores técnicos, tais como médias móveis exponenciais com diversos parâmetros, totalizando um máximo de 33 entradas, para um sistema de *trading* intradiário na frequência de 15 minutos. A rede neural MLP apresentou 2 saídas: estimativa do preço máximo e mínimo. A regra básica de *trading*, em que foi incluída o mecanismo de perda máxima (*stop loss*), foi a seguinte (houve outras regras adicionais): a indicação é de compra se o último fechamento conhecido é inferior à previsão da cotação mínima pela rede

neural e de venda se for superior à previsão do preço máximo. Foram consideradas taxas de corretagem de *homebrokers* e o custo do *slippage* (diferença entre o preço sugerido pelo sistema *trading* e o implementado na prática, se desfavorável). As métricas de avaliação foram a taxa nominal anualizada de retorno, perda máxima, e outras geralmente utilizadas na avaliação de um sistema de *trading*. Os resultados foram considerados bons, todavia, os autores reconheceram o tradicional desconforto provocado pelo mecanismo de caixa preta das redes neurais. Maciel e Ballini (2010) empregaram redes neurais MLP *backpropagation* na **predição** de diversos índices de ações internacionais (inclusive o *Standard & Poor's*) e constaram que os resultados foram superiores ao modelo autoregressivo GARCH (*General Autoregressive Conditional Heterocedasticity*). Os log-retornos (propensos a apresentarem distribuições normais) dos índices defasados (*lag*) por uma 1 a 3 unidades de tempo serviram como entradas das redes neurais. As amostras foram divididas em treinamento, validação e teste na proporção de 80%, 5% e 15%. As taxas de aprendizagem do algoritmo, do número de camadas e de neurônios variaram conforme o índice objeto da predição. Os critérios de avaliação de desempenho da rede neural foram  $R^2$  (coeficiente de determinação), RMSE (*root mean square error*) e taxa de acerto da direção do índice. A taxa de acerto em todas as amostras para o índice Standard & Poor's foi de 76,23% pela rede neural e 57,76% pelo modelo GARCH. Os resultados nas amostras de validação e teste confirmaram a superioridade das redes neurais. Wei e Cheng (2012) propôs um modelo híbrido, implementado por uma rede neural recorrente (Elman), com o propósito de prever o índice TAIEX, representativo do mercado de ações de Taiwan, durante 6 anos. Com a finalidade de reduzir a dimensão do vetor de entrada, composto de indicadores técnicos, uma matriz de correlação entre esses indicadores e o índice TAIEX serviu para efetuar a primeira filtragem. Em seguida foi usada a regressão passo a passo (*stepwise regression*), tendo como critério a estatística t de *Student*, e uma árvore decisória a partir de um critério de ganho de informação. Pelo critério de avaliação do RMSE o método proposto superou o modelo autoregressivo convencional AR(1) e os modelos

---

de séries temporais *fuzzy* sugeridos por Chen (1996) e Yu (2005).

## Capítulo 3

# Redes Neurais Artificiais

Neste Capítulo 3 são descritos de forma sumária os princípios que regem o funcionamento de uma rede neural artificial multicamadas com propagação para a frente (*feed forward*) sem retroalimentação (*perceptron* multicamadas), a arquitetura básica, os procedimentos consolidados na literatura sobre o tratamento dos dados de entrada e saída, treinamento e critérios de desempenho da rede neural e o algoritmo empregado neste trabalho.

### 3.1 Introdução às Redes Neurais

Uma rede neural artificial é um processador paralelo constituído de unidades de processamento denominados neurônios, interligados entre si através de conexões sinápticas, com capacidade de armazenar conhecimento experimental mediante a interação com o meio ambiente (Haykin (2009)). Conforme são estruturados os neurônios uma rede pode ser classificada como: (a) camada única ou multicamadas com propagação para a frente sem retroalimentação dos sinais (*perceptron* multicamadas), sendo que,

no último caso, todos os neurônios de uma camada são conectados com os da camada subsequente; (b) recorrentes, quando constituídas com ou sem camadas escondidas em que há retroalimentação dos sinais para os próprios neurônios geradores dos sinais ou somente para os demais. De acordo com o aprendizado as redes se subdividem em supervisionadas, aprendizado com reforço e não supervisionadas. As primeiras requerem que os vetores de entradas sejam associados às saídas desejadas no treinamento. A segunda categoria desenvolve habilidades de aprender mediante a interação contínua com o ambiente até que uma função de custo ou de desempenho seja minimizada; a terceira categoria aprende através de regras competitivas entre os neurônios na codificação das entradas. Segundo Haykin (2009), Cybenko (1988,1989) foi quem demonstrou rigorosamente que uma rede neural com uma camada escondida, que empregue funções de ativação  $\varphi(\cdot)$  contínuas, monotônicas, crescentes e limitadas, além de um viés, é um regressor universal não linear. O seu trabalho foi precedido por pesquisas feitas por Weierstrass (1885), Kolmogorov (1957), Sprecher (1965) e Gallant e White (1988).

## 3.2 Arquitetura e Funções de Ativação

A Figura 3.1 abaixo ilustra a arquitetura de uma rede multicamadas com propagação para frente supervisionada, com uma camada de entrada com 3 sinais e 1 viés com sinal de intensidade unitária, uma camada escondida com 3 neurônios e 1 viés, com funções de ativação tangentes hiperbólicas, e uma camada de saída com 1 neurônio com função de ativação linear, que foi a arquitetura básica utilizada neste trabalho.

O vetor na camada de entrada de dimensão M correspondente a um exemplo (amostra) é expresso por:

$$\mathbf{X} = [1 \ x_1 \ x_2 \ \cdots \ x_k \ \cdots \ x_M]^T \quad (3.1)$$

em que 1 (um) é o sinal unitário do viés,  $x_k$  são os sinais de entrada com  $k = 1, \dots, M$ ,

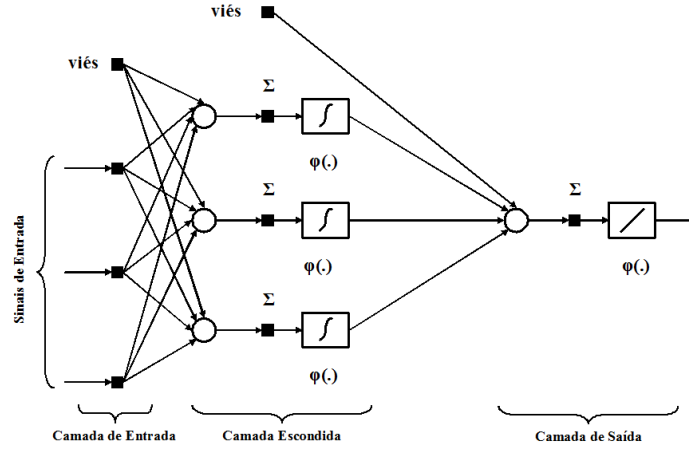


Figura 3.1: Arquitetura com uma camada escondida

sendo que cada  $k$  é denominado **atributo** (*feature*), e  $T$  a notação de vetor ou matriz transposta. Na Figura 3.1 o valor de  $M$  é igual a 3. O critério de escolha dos elementos  $x_k$  do vetor de entrada  $\mathbf{X}$  é descrito na Seção 4.7, que trata sobre a metodologia empregada nos experimentos. O vetor de entrada  $\mathbf{X}_i$ , correspondente à  $i$ -ésima amostra, é associado à saída desejada  $d_i$  e o objetivo da rede neural supervisionada é mapear e aprender essa associação na fase de treinamento. Quanto ao número de neurônios na camada escondida não há consenso na literatura. Neste trabalho, após experimentos preliminares, foi adotada a regra  $\frac{M}{2}$  e  $\frac{M-1}{2}$  como o número de neurônios nessa camada, sendo  $M$  a dimensão dos vetores de entrada (atributos), conforme definição anterior.

Uma rede (treinada) multicamadas com propagação para frente com 1 neurônio na camada de saída, com diversas camadas escondidas, pode ser vista como funções de ativação  $\varphi(\cdot)$  aninhadas, com saída  $y$  expressa por (Haykin (2009)):

$$y = \varphi\left(\sum_r w_{sr} y_r\right) \quad \dots \quad y = \varphi\left(\sum_r w_{sr} \varphi\left(\sum_j w_{rj} y_j\right)\right)$$

$$y = \varphi\left(\sum_r w_{sr} \varphi\left(\sum_j w_{rj} \varphi\left(\dots \varphi\left(\sum_k w_{lk} x_k\right)\right)\right)\right) \quad (3.2)$$

em que  $w_{sr}$  é o peso sináptico entre o  $r$ -ésimo neurônio da última camada escondida e o único neurônio  $s$  da camada de saída,  $y_r$  é o sinal de saída do  $r$ -ésimo neurônio da

camada escondida, e assim por diante.

### 3.3 Normalização dos Dados

Como as entradas da rede neural apresentam naturezas diferentes é necessário que os seus valores sejam normalizados para evitar que os neurônios da camada escondida com funções de ativação tangente hiperbólica, utilizadas neste trabalho, não operem na região de saturação e que os valores de grandezas diferentes não influenciem de maneira diversa os ajustes dos pesos e vieses da rede neural durante o treinamento. As saídas desejadas  $d_i$  também devem ser normalizadas para que seja possível o correto mapeamento entre as entradas normalizadas e as saídas correspondentes. Para essas finalidades os dados de entrada da amostra de treinamento da rede e as saídas desejadas correspondentes são ajustados para se situarem dentro intervalo  $[-1,+1]$ . Neste trabalho as normalizações foram feitas por meio da função *mapminmax* da plataforma MATLAB® versão R2009b utilizada nos experimentos, conforme a equação abaixo:

$$x_N = \frac{(x_{Nmax} - x_{Nmin})(x - x_{min})}{x_{max} - x_{min}} + x_{Nmin} \quad (3.3)$$

em que  $x_{Nmax}$  e  $x_{Nmin}$  são iguais aos extremos do intervalo da normalização,  $+1$  e  $-1$  no caso deste trabalho, respectivamente,  $x_{max}$  e  $x_{min}$  são os valores máximo e mínimo de cada atributo relativos à amostra de **treinamento** e  $x_N$  é o valor normalizado. Tratamento semelhante é dado à saída  $d$  correspondente.

Conforme será comentado com mais detalhes no Capítulo 4 sobre a metodologia, as entradas da rede neural foram constituídas de uma série temporal de preços defasados (*lag*) do índice Standard & Poor's 500 e de indicadores técnicos. Os intervalos de variação das  $i$ -ésimas entradas  $\mathbf{X}_i$  e das saídas desejadas correspondentes  $d_i$  são desconhecidos no **período de verificação**, que é o espaço temporal em que a rede

neural treinada irá estimar os valores de  $d_i$  que servirão de subsídios para o sistema de *trading*. Por essa razão, os mesmos valores  $x_{max}$  e  $x_{min}$  da equação (3.3), utilizados na normalização das entradas das amostras de treinamento e das saídas desejadas correspondentes, são aplicáveis às normalizações dos atributos e saídas desejadas da amostra do período de verificação. Como neste trabalho, durante o período de verificação, sabia-se que o índice Standard & Poor's 500 permaneceu dentro do intervalo da amostra de treinamento não foi necessário alterar o intervalo de normalização  $[-1, +1]$ , que são os parâmetros padrões (*default*) da função *mapminmax* do MATLAB. Todavia, sem o conhecimento *a priori* dos preços do ativo financeiro objeto de transação pelo sistema de *trading* no período de verificação, é recomendável que o intervalo de normalização  $[-1, +1]$  seja alterado para  $[-0,75, +0,75]$ , ou seja, que  $x_{Nmax}$  e  $x_{Nmin}$  sejam iguais a  $+0,75$  e  $-0,75$ , sob pena das funções de ativação operarem nas regiões de saturação. Não tem sido incomum fortes períodos de tendência de alta e de baixa, como em 2009 quando o índice IBOVESPA subiu  $+86\%$ , que poderia colocar em risco a normalização no intervalo  $[-1, +1]$  caso coincidissem com o período de verificação.

### 3.4 Treinamento e Critérios de Desempenho

O treinamento de uma rede neural consiste em minimizar uma função de custo, mediante a alteração dos pesos e vieses por um algoritmo, cujos valores iniciais são escolhidos aleatoriamente com o objetivo de facilitar o trabalho do algoritmo na busca do valor mínimo da função de custo na amostra de treinamento, denominado mínimo global. A função de custo típica de uma rede multicamadas com propagação para frente, com um neurônio na camada de saída, com apresentações das amostras de treinamento ao algoritmo feitas no modo sequencial, pelo qual os pesos e vieses são



alterados após cada apresentação, é:

$$\xi(\mathbf{W}, i) = \frac{(d_i - y(\mathbf{X}_i, \mathbf{W}))^2}{2} \quad (3.4)$$

em que  $\mathbf{W}$  é o vetor de pesos e vieses,  $d_i$  a saída desejada correspondente à  $i$ -ésima amostra de treinamento,  $y(\mathbf{X}(i), \mathbf{w})$  é a função descrita na equação (3.2), ambos correspondentes ao  $i$ -ésimo exemplo. Cada ciclo de apresentação à rede de todas as amostras do treinamento é denominado época (*epoch*). No modo de apresentação das amostras de treinamento por lote (*batch*), em que os pesos são alterados somente após completada uma época, a função de custo típica é:

$$\xi(\mathbf{W}) = \frac{1}{N} \sum_{i=1}^N \xi(\mathbf{W}, i) \quad (3.5)$$

em que  $N$  é a quantidade de exemplos da amostra de treinamento. O treinamento é interrompido - critério de parada - quando a função de custo ou a sua taxa de variação atinge no treinamento um valor previamente estipulado pelo usuário. É comum também estipular um número máximo de iterações do algoritmo no treinamento e, nesse caso, deve-se tomar o cuidado de escolher um número elevado para que o treinamento não seja parado precocemente. Uma vez atingido um dos critérios de parada as amostras de teste são apresentadas à rede treinada.

Com a finalidade de capacitar a rede de generalização, ou seja, de não se especializar na amostra de treinamento, essa é subdividida em **treinamento propriamente dita** e **teste**, sendo que essa última não participa do treinamento e serve para avaliar o desempenho da rede por um critério de desempenho. Para que o algoritmo cubra melhor o domínio da função a ser mapeada, ditada pela equação (3.2), as amostras de treinamento e teste são randomizadas mesmo em se tratando de uma série temporal de preços. Os critérios típicos de avaliação do desempenho da rede na amostra de

teste são o MSE (*mean squared error*), RMSE (*root mean squared error*), o RMSE relativo (RMSE\_R) e o MAE (*mean absolute error*). O treinamento é repetido até que a métrica de desempenho na amostra de teste estipulada pelo usuário seja atingida. Neste trabalho foi escolhida a métrica de desempenho RMSE\_R:

$$RMSE\_R = \sqrt{\frac{1}{N} \sum_{i=1}^N \left( \frac{d_i - y_i}{d_i} \right)^2} \quad (3.6)$$

cujos termos já foram definidos neste Capítulo. O critério de seleção da rede neural geralmente é o de desempenho. Porém, neste trabalho, além do RMSE\_R, foi introduzido o critério de taxa de acerto da rede neural, na amostra de teste, da direção do preço do ativo financeiro objeto da transação, que em última instância é o que tornará lucrativo um sistema de *trading*.

Uma técnica usada com a finalidade de tornar a rede com melhor capacidade de generalização, adotada nos experimentos deste trabalho, é a **validação cruzada**. Ela consiste em subdividir a amostra de treinamento em **treinamento propriamente dita**, **validação** e **teste**, também randomizadas com a mesma finalidade citada anteriormente. Neste trabalho as proporções dessas amostras foram 60%, 20% e 20%, respectivamente. Durante o treinamento é apresentada periodicamente à rede a amostra de validação e esse é interrompido assim que a função de custo da amostra de validação começa a subir, instante em que os pesos e vieses são fixados. Passo subsequente, o desempenho da rede é avaliado na amostra de teste, conforme procedimento anteriormente descrito. A desvantagem da validação cruzada é a quantidade de amostra exigida que poderia ser usada no treinamento e teste.

### 3.5 Algoritmo

Neste trabalho foi empregado o algoritmo *Bayesian Regularization* (BR) proposto

por Mackay (1992), que é uma extensão do algoritmo *Levenberg-Marquardt*, mas se diferencia pela função de custo  $F_c(\mathbf{W})$  a ser minimizada dada por:

$$F_c(\mathbf{W}) = \beta \xi(\mathbf{W}) + \alpha E(\mathbf{W}) \quad (3.7)$$

em que  $\beta$  e  $\alpha$  são fatores de regularização obtidos por inferência bayesiana (ver **Apêndice A**),  $E(\mathbf{W})$  é a soma quadrática dos pesos e vieses e  $\xi(\mathbf{W})$  (equação (3.5)). Como  $\xi(\mathbf{W})$  incorpora toda a amostra de treinamento o algoritmo BR funciona pelo modo por lote (*batch*). A ideia do algoritmo BR consiste em minimizar  $F_c(\mathbf{W})$  de modo que a resposta da rede seja suave, apresente melhores métricas de desempenho na presença de amostras com dados singulares (*outliers*) no treinamento, particularmente quando há baixas agudas em um curto período de tempo, como foi o caso do índice Standard & Poor's 500 em setembro de 2008, e dispense o emprego da validação cruzada, embora ela tenha sido usada neste trabalho devido à imperiosa necessidade de generalização da rede. Esse algoritmo utiliza a técnica da retropropagação (*backpropagation*) que consiste no cálculo da derivada parcial da função  $y(\mathbf{X}, \mathbf{W})$  da equação (3.2) em relação a  $\mathbf{W}$  para cada exemplo (amostra)  $\mathbf{X}$ , mediante o recurso da regra da cadeia, a começar pela camada de saída.

# Capítulo 4

## Metodologia

Um sistema de *trading* é um conjunto de regras que presumidamente busca identificar um quadro favorável do mercado para a obtenção de lucros através de negociações de um ativo financeiro. No cotidiano a maioria dos sistemas inclui regras que limitam eventuais perdas (*stop loss*), previamente definidas, quando elas ultrapassam um percentual calculado sobre o capital empregado acumulado. Além disso, é comum estipular regras para a realização de lucros (*profit taking*). Neste trabalho o sistema de *trading* é do tipo seguidor de tendência (*trend follower system*), não foi desenvolvido nenhum mecanismo de limitação dos prejuízos, bem como de realização de lucros, e as regras são definidas a partir da saída de uma rede neural. O ativo financeiro escolhido para os experimentos foi o índice Standard & Poor's 500.

### 4.1 O Índice Standard & Poor's 500

O índice mais representativo do mercado de ações norte-americano, o de maior capitalização no mundo, é o Standard & Poor's 500 constituído por 500 empresas

domiciliadas nos EUA. No sítio [www.standardandpoors.com/home/en/us/](http://www.standardandpoors.com/home/en/us/) é descrita a metodologia de apuração do índice desenvolvida pela empresa Standard & Poor's. Ele é calculado pela soma dos valores de capitalização das ações que o compõem e que não pertencem à categoria *closed held shares*, que são as ações detidas pelo grupo de controle acionário. O resultado é dividido por um fator (divisor) com 2 objetivos: (a) tornar a cotação do índice tratável, caso contrário seria apreçado em trilhões de dólares; (b) manter a continuidade do índice nos casos de mudança em sua composição ou de eventos societários das empresas que o compõem, tais como bonificações em ações, dividendos, desdobramentos, fusões, etc. O ajuste do divisor, na hipótese (b), é melhor compreendido por meio de um exemplo. Suponhamos que o índice tenha fechado em 1200 pontos no  $i$ -ésimo pregão e que no pregão seguinte algumas ações que o compõem distribuam dividendos. Os valores teóricos dessas ações no pregão seguinte ao  $i$ -ésimo pregão são iguais às cotações desse pregão deduzidas dos valores dos respectivos dividendos (preços ex-dividendos). O divisor é ajustado de tal forma que o valor teórico da carteira do índice Standard & Poor's 500 na abertura do pregão seguinte ao  $i$ -ésimo pregão seja exatamente igual a 1200, apurado pelas cotações do  $i$ -ésimo pregão, ajustadas pelas distribuições dos dividendos na forma mencionada anteriormente.

A maneira mais simples de negociar o índice Standard & Poor's 500 é através da compra ou da venda de seu índice futuro. O índice futuro é o índice à vista, no caso do Standard & Poor's 500, acrescido até a data de vencimento do contrato da taxa anual de juros livre de risco e descontado da taxa anual dos dividendos das empresas que o compõem Hull (1991), dado por:

$$F = A \left( \frac{1+r}{1+q} \right)^t \quad (4.1)$$

em que  $F$  é o índice futuro,  $A$  o índice à vista,  $r$  é a taxa anual livre de risco,  $q$  a taxa anual de dividendos das empresas componentes do índice e  $t$  o número de anos a decorrer até o vencimento do contrato. A título de observação, no caso do IBOVESPA  $q = 0$  na fórmula acima, pois os dividendos são reinvestidos na carteira do índice.

No mercado norte-americano os vencimentos dos contratos ocorrem 4 vezes por ano. No dia do vencimento a operação em curso é expirada, mas pode ser renovada mediante a implementação da mesma operação no contrato relativo ao próximo vencimento. A operação exige um depósito de garantia tanto na compra quanto na venda, equivalente a um percentual do valor do contrato, que é mantido na corretora até o final da operação na forma de dinheiro em espécie ou de títulos especificados pela Bolsa de Futuros. Diariamente a posição é avaliada a preços de mercado e o investidor deverá depositar na corretora a diferença (ajuste) entre os fechamentos diários sucessivos no caso de perda ou receber essa diferença no caso de ganho. Uma posição comprada ou vendida é ganhadora quando a cotação do índice futuro sobe ou cai, respectivamente. Caso contrário, a operação é perdedora.

Neste trabalho foi empregado o índice à vista no lugar do índice futuro devido à facilidade de obtenção de sua base de dados no sítio [www.yahoo.com](http://www.yahoo.com) (código GSPC). Se fosse utilizado o índice futuro ao invés do à vista, os valores dos contratos futuros expirados deveriam ser ajustados pela equação (4.1) desde os respectivos vencimentos até a data da última amostra do período de verificação. Como essa base de dados ajustada não é facilmente obtida, o ajuste de toda a base do índice futuro resultaria em um trabalho bastante árduo, sem que fossem colhidos benefícios uma vez que a metodologia não seria alterada com o emprego do índice à vista.

## 4.2 Períodos de Treinamento, Validação, Teste e de Verificação

Foram usados os dados dos fechamentos semanais do índice Standard & Poor's 500 entre as semanas findas em 06/04/1990 e 29/06/2012, totalizando 1160 semanas. As amostras de **treinamento**, **validação** e **teste** ficaram compreendidas entre as sema-

nas findas entre 06/04/1990 e 06/07/2007, totalizando 900 semanas. O período de **verificação**, durante o qual foram feitas as estimativas dos fechamentos semanais do índice Standard & Poor's 500 pelas redes neurais e o sistema de *trading* foi experimentado, ficou compreendido entre as semanas findas entre 13/07/2007 e 29/06/2012, totalizando 260 semanas ou cerca de 5 anos. O período de verificação foi escolhido de modo a cobrir fases distintas do mercado de ações norte-americano. Nesse período houve uma fase de baixa aguda nos preços devido à crise mundial entre 2007 e 2008, seguida de um período de alta até a semana finda em 29/06/2012, permeado por fortes recuos intermediários. Como no período de verificação o índice Standard & Poor's 500 permaneceu dentro do intervalo da amostra de treinamento, validação e teste, a normalização dos vetores de entrada foi feita no modo padrão (*default*) do MATLAB. O procedimento adotado somente foi possível porque nos experimentos o comportamento do índice no período de verificação era conhecido *a priori*. Quando não for esse o caso, ou seja, quando os valores do índice Standard & Poor's 500 no período de verificação somente são conhecidos *a posteriori*, é recomendável que o intervalo da normalização das entradas e saídas da rede neural se situe entre  $[-0,75, +0,75]$ .

### 4.3 Funcionamento do Sistema de *Trading*

Neste trabalho o sistema de *trading* tem o objetivo de operar uma carteira que emula o índice Standard & Poor's 500 para auferir lucros de forma consistente e contínua. As operações de compra e venda da carteira são realizadas somente no fechamento dos últimos pregões de cada semana pelos preços de fechamentos do índice, quando for o caso, ou seja, quando o sistema de *trading* ativar uma operação. Como regra o sistema de *trading* se limita a negociar 1 contrato do índice à vista no final do último pregão de cada semana, conforme será esclarecido nesta Subseção, e durante a semana o sistema permanece inativo, ou seja, permanece com a posição estabelecida no final

da semana anterior. Conhecido o dado  $d_i$  de fechamento do índice no último pregão da  $i$ -ésima semana, a operação implementada pelo sistema de *trading*, no fechamento dessa  $i$ -ésima semana, será de compra de 1 contrato se a estimativa  $y_{i+1}$  da rede neural do fechamento na semana seguinte for superior ao último fechamento conhecido  $d_i$ . Caso contrário a operação será de venda. Uma operação de compra ou de venda é mantida de uma semana para outra, qualquer que seja o seu resultado financeiro, se a indicação da rede neural não for alterada. Em caso contrário é assumida a posição inversa à anteriormente em curso. O sistema de *trading* permanece sempre ativo, ou seja, ou ele está comprado ou vendido em 1 contrato e uma operação de compra será sempre sucedida por uma de venda e vice-versa. A variação  $V_{i+1}$  do valor de mercado da posição (*equity*) entre a  $i$ -ésima semana e a seguinte é dada por:

$$V_{i+1} = s_i |d_{i+1} - d_i| \quad (4.2)$$

em que  $d_i$  é o valor de 1 contrato no fechamento da  $i$ -ésima semana, quer a posição seja comprada ou vendida,  $d_{i+1}$  é a cotação de fechamento na semana seguinte, e  $s_i$  é dado por:

$$s_i = \begin{cases} +1 & \text{se } \text{sign}(d_{i+1} - d_i) = \text{sign}(y_{i+1} - d_i) \\ -1 & \text{se } \text{sign}(d_{i+1} - d_i) \neq \text{sign}(y_{i+1} - d_i) \end{cases} \quad (4.3)$$

em que *sign* é a função que assume valor igual a +1 se seu argumento for positivo, valor igual a -1 se for negativo, e igual a 0 se o argumento for nulo. A estimativa  $y_{i+1}$  da rede neural do fechamento do índice Standard & Poor's 500 na semana seguinte à  $i$ -ésima semana é igual a  $y$  da equação (3.2). Em outras palavras, o sistema de *trading* é ganhador se a rede neural acertar a oscilação, positiva ou negativa, do índice Standard & Poor's 500 de uma semana para a outra, mesmo na situação em que a estimativa da rede neural se desvie materialmente do preço efetivo. Em caso contrário é perdedor. Neste trabalho assumiu-se que a estimativa da rede neural  $y_{i+1}$  da cotação



de fechamento semanal  $d_{i+1}$  e a implementação de uma operação, no caso da indicação da rede neural ser alterada em relação à semana anterior, são feitas no mesmo instante do fechamento do pregão na  $i$ -ésima semana (**ver Apêndice B para maiores detalhes**).

Não foram considerados emolumentos da bolsa, taxas de corretagem e tributações sobre os lucros obtidos. A tributação depende da estrutura jurídica do fundo de investimento e do regime tributário do país. No Brasil a tributação das operações de um fundo multimercado exclusivo fechado é postergada até o resgate das cotas e, nesse caso, o imposto de renda a pagar pode ser reinvestido. Em paraísos fiscais como nas Ilhas Virgens Britânicas a tributação é nula. Atualmente, devido à maior concorrência entre as corretoras, as taxas de corretagem caíram substancialmente. Os investidores individuais pagam cerca de apenas R\$ 20 por operação e um grande investidor geralmente negocia uma taxa mensal fixa que se dilui entre todas as demais operações nos diversos mercados. Em 05/04/2013 a corretagem e emolumentos de uma operação real (realizada pelo autor deste trabalho) no valor de R\$ 253918,00 com índice futuro do IBOVESPA na bolsa BM&FBovespa custou 0,076% desse valor (taxa pode ser reduzida

mais adiante e, portanto, não foi necessário investir além do valor da primeira operação. Além das restrições mencionadas, ausências de alavancagem e reinvestimentos dos lucros no aumento das operações ou na renda fixa e das margens que permaneceram ociosas, que limitaram os ganhos do sistema de *trading*, não foi usado o mecanismo de perda máxima (*stop loss*), que interrompe uma operação no caso das perdas superarem um percentual previamente estipulado, critério que tem sido usado em trabalhos acadêmicos, conforme mencionado na Seção 2.2. Na Seção 5.3 foram feitas **considerações sobre a inclusão de custos operacionais** e sobre o emprego de alavancagem moderada.

## 4.4 Seleção de Atributos

Kaastra e Boyd (1996) comenta que há 3 possibilidades de classes de atributos na previsão de preços de ativos financeiros mediante o emprego de rede neural. De um lado há os indicadores técnicos, extraídos do próprio mercado, tais como preços de ativos, volumes das transações, etc, que são usados pela Escola Técnica (Murphy (1999)), os contábeis (dados dos balanços patrimoniais das empresas) e os macroeconômicos (taxa de juros no mercado secundário, por exemplo). Os 2 primeiros são prontamente disponíveis e o cuidado consiste em checar a reputação do fornecedor dos dados (*vendor*). Quantos às entradas de natureza macroeconômica a preocupação consiste em checar se a série temporal é consistente ao longo do tempo, se são sujeitas a revisões periódicas e ter ciência de que sempre são divulgadas com atraso. Uma vez escolhidos os atributos é necessário avaliar a necessidade de transformação das entradas e respectivas saídas para minimizar eventuais ruídos que mascaram as tendências e adequar a distribuição de ocorrências das variáveis às assimetrias e às caudas gordas (*fat tails*). Nesse último caso, as transformações mais comuns são a primeira diferença, como forma de eliminar eventual tendência, e o logaritmo da variável para eliminar ou

reduzir as caudas gordas da distribuição. Outros procedimentos sugeridos consistem na redução do número de variáveis redundantes e no retreinamento periódico da rede neural (janelamento móvel) de modo que ela se adapte às novas condições de mercado. O emprego de médias móveis simples ou exponenciais para suavizar a tendência dos preços também está no rol das recomendações. Outra recomendação dos autores consiste em eliminar do treinamento as amostras correspondentes às variações muito pequenas dos preços, de modo que a rede neural se especialize nas oscilações maiores que apresentam impactos mais significativos sobre o sistema de *trading* e sobre os pesos e vieses da rede neural. Recomendam também testar várias combinações de indicadores técnicos com o argumento de que alguns deles são úteis somente quando combinados com outros.

Poderiam ser acrescentados aos comentários dos autores acima mencionados as dificuldades no emprego de dados macroeconômicos como entradas da rede neural, que geralmente são obtidas através de pesquisas ou coletas, com revisões rotineiras (não causadas por mudança na metodologia), que é a regra e não a exceção, e a periodicidade mínima semanal com atraso, as constantes mudanças no padrão sazonal das séries dessazonalizadas, como ocorreu a partir de 2008 na economia norte-americana devido aos dados singulares (*outliers*) gerados pela crise econômica, e as revisões sobre a metodologia de cálculo.

Diz-se que um indicador é antecedente se ele antecipa a ocorrência da variável dependente. Como exemplos de alterações recentes na metodologia de cálculo de importantes indicadores antecedentes relativos à economia norte-americana, *The Conference Board*, que divulga o indicador macroeconômico antecedente composto (*macroeconomic leading indicator*) mensal da economia norte-americana com atraso de pelo menos 1 mês, em janeiro de 2012 alterou substancialmente a metodologia de cálculo ao substituir alguns de seus componentes. Como consequência a base de dados da série foi substancialmente alterada. Outro importante indicador antecedente, o componente Novas Ordens (*New Orders*), divulgado pelo *Institute for Supply Management*, também revi-

sou essa série no mês de fevereiro de 2013. Como consequência a série mensal nos 12 meses precedentes foi afetada.

A seleção das variáveis de entradas de uma rede neural empregadas na previsão de preços de ativos financeiros depende do horizonte da previsão, se anual, mensal, semanal ou um horizonte mais curto. Nos horizontes mais longos poderiam ser usados indicadores macroeconômicos antecedentes obtidos por meio de pesquisas, desde que tomados os devidos cuidados acima mencionados, e os extraídos de mercados correlatos, obviamente não sujeitos à revisões, com a propriedade de antecipação, tais como os diferenciais (*spreads*) entre a taxa de juros dos títulos federais de longo prazo e os corporativos de alto risco (*junk bonds*) com a mesma duração (*duration*), recentemente incorporado no indicador antecedente composto da *The Conference Board*. Taxa de juros dos títulos federais também é uma opção, como foi o caso da escolha feita por Kimoto et al. (1990) na implementação de um sistema de *trading* para o índice TOPIX do Japão. Nesse experimento o período de verificação foi encerrado às vésperas do início da crise no país e do colapso do mercado de ações (começo de 1990), cuja economia continua estagnada desde então. Como forma de reanimá-la o governo tem praticado uma política monetária relaxada (*quantitative easing*) que derrubou as taxas de juros dos títulos federais para próximo de 0% ao ano, patamar que permanece por mais de uma década sem que tenha havido nesse período oscilações materiais. Como consequência essa variável perdeu a eficácia como regressor e, portanto, como entrada de uma rede neural na previsão do índice de ações do país.

Até meados da década de 2000 a escolha de variáveis de entrada das redes neurais com a finalidade de predição de preços de ativos financeiros seguia o critério da tentativa e erro, sem uma formalização consistente conforme se pode constatar pela recomendação de Kaastra e Boyd (1996) de testar pura e simplesmente combinações de indicadores. Em parte isso era devido ao reinado da Hipótese do Mercado Eficiente (HME) proposta por Fama (1970) que somente foi abalada após as sucessivas bolhas nos preços das ações, a começar pela conhecida *dotcom* na bolsa eletrônica Nasdaq

nos EUA que estourou em 2000. A partir de então a literatura sobre o uso de redes neurais na predição de preços passou a registrar o emprego de variáveis que capturam a anomalia do efeito momento dos preços dos ativos financeiros. Contudo, na maioria dos estudos a escolha dessas variáveis foram feitas sem o amparo de uma fundamentação na literatura financeira.

Entende-se por **momento** a taxa de variação do preço de um ativo por unidade de tempo:

$$Momento = \left( \frac{P_{i+n} - P_i}{n} \right) \quad (4.4)$$

em que  $P_i$  é o preço de um ativo financeiro no  $i$ -ésimo instante e  $n$  o número de unidades de tempo escolhido pelo usuário. Jegadeesh e Titman (1993), pesquisadores do ramo das Finanças Comportamentais, constataram uma diferença entre os retornos médios das ações com elevados e baixos momentos, medidos nos 12 meses precedentes, de 1,5% ao mês, a favor das primeiras. No longo prazo, contudo, esse efeito desaparece e é revertido. Essa tese encontra abrigo na Escola Técnica (Murphy (1999)) que defende que, uma vez instalada uma tendência nos preços das ações, o mais provável é que ela continue no curto prazo ao invés de ser revertida. Esse fenômeno é conhecido como

## 4.5 Variáveis de Entrada das Redes Neurais do Sistema de *Trading*

Conforme mencionado anteriormente os atributos do vetor de entrada das redes neurais foram escolhidos de modo a capturar a anomalia do efeito momento nos preços das ações. A escolha dos atributos foi inspirada no trabalho de Klassen (2005) que empregou os fechamentos semanais nas 4 semanas anteriores à semana da estimativa dos índices Dow Jones Industrial e Nasdaq Composite e o índice técnico MACD (*moving average convergence and divergence*) na semana imediatamente anterior. O trabalho citado teve o propósito de testar um leque de indicadores técnicos na predição dos índices mencionados pelo critério de desempenho do menor erro percentual da estimativa da rede na amostra de teste, sendo que a preferência recaiu sobre o MACD. A autora não apresentou uma justificativa da escolha do rol de indicadores que foram testados e não foi proposto nenhum sistema de *trading*. O indicador técnico ROC (*rate of change*) não esteve presente no rol dos testados, razão pela qual ele foi incluído neste trabalho por ser também um indicador de momento muito utilizado (Murphy (1999)).

Neste trabalho a função mapeada pela rede neural foi:

$$d_{i+1} = f(d_i, \Psi d_{i-1}, \Omega d_{i-2}, \Gamma d_{i-3}, \Theta MACD_i, \Phi ROC_i) \quad (4.5)$$

em que  $d_{i+1}$  é o preço de fechamento do índice Standard & Poor's 500 na semana seguinte à  $i$ -ésima semana, a ser previsto pela rede neural. Os atributos do vetor de entrada  $\mathbf{X}$  na expressão (3.1), são  $d_{i-k}$ ,  $k = 0, \dots, 3$ , ou seja, os preços de fechamento do índice nos últimos pregões das 4 semanas imediatamente anteriores à semana cujo preço de fechamento é estimado, o  $MACD_i$  e o  $ROC_i$  que são indicadores técnicos referentes à  $i$ -ésima semana, ambos definidos mais adiante. Os parâmetros  $\Psi$ ,  $\Omega$ ,  $\Gamma$ ,  $\Theta$  e  $\Phi$  podem assumir valores iguais a 0 ou 1, conforme foram ou não incluídos na lista dos atributos os fechamentos do índice Standard & Poor's 500 nas 3 semanas

imediatamente anteriores à última semana e os indicadores MACD e ROC. Para efeito deste trabalho denomina-se **configuração** o conjunto dos atributos das entradas da rede neural, a arquitetura ilustrada na Figura 3.1 e as funções de ativação. Na única camada escondida foi usada a função tangente hiperbólica (*tansig*) e na de saída ( $y_i$ ), composta de 1 neurônio, a função linear (*pureline*). Portanto, a saída da rede neural não é binária, tais como +1 para compra e -1 para venda. Uma vez estimado  $y_i$  aplicam-se as regras do sistema de *trading* discutidas na Seção 4.3. Em todos os experimentos foi empregado o algoritmo *Bayesian Regularization* e uma quantidade de neurônios variável na camada escondida igual a  $\frac{M}{2}$  e  $\frac{M-1}{2}$ , que apresentou melhores resultados na amostra de teste, abordagem sugerida por Kaastra e Boyd (1996). Foi observado nos inúmeros experimentos que o emprego de 3 neurônios na única camada escondida ou o acréscimo de uma segunda camada escondida prejudicaram o desempenho do sistema de *trading*. Como foi detectada forte influência do último fechamento semanal  $d_i$ , conhecido no momento da decisão sobre a implementação de eventual operação pelo sistema de *trading*, essa variável esteve presente em todas as configurações testadas e, portanto, o seu parâmetro é sempre igual a 1. Os indicadores técnicos  $MACD_i$  e  $ROC_i$ , referentes à  $i$ -ésima semana, são dados por (Murphy (1999)):

$$MACD_i = MME_i(p_1) - MME_i(p_2) \text{ e } ROC_i = d_i - d_{i-p_3-1} \quad (4.6)$$

em que  $MME_i(p_1)$  e  $MME_i(p_2)$  designam as médias móveis exponenciais do índice Standard & Poor's 500 na  $i$ -ésima semana com parâmetros  $p_1$  e  $p_2$ . A  $MME_i(p_1)$  é dada por:

$$MME_i(p_1) = \left( d_i - MME_{i-1}(p_1) \right) \frac{2}{1 + p_1} + MME_{i-1}(p_1) \quad (4.7)$$

em que, para  $i = 1$ ,  $MME_1(p_1)$  é igual a  $d_1$ . Neste trabalho os parâmetros  $p_1$  e  $p_2$

são, respectivamente, 8 e 17 sugeridos por Klassen (2005) e  $p_3$  é igual a 10, neste caso porque a rede neural foi informada na  $i$ -ésima semana sobre a sequência dos últimos 4 fechamentos semanais. Se os valores do  $MACD_i$  ou do  $ROC_i$  aumentam (diminuem) em uma tendência de alta (baixa) significa que o movimento do índice se fortalece (*convergence*) e, em caso contrário, se enfraquece (*divergence*).

## 4.6 A Técnica do Janelamento Móvel

A técnica do janelamento móvel consiste no treinamento periódico de redes neurais empregadas nas previsões correspondentes a períodos (janelas) de verificação que se movem ao longo do tempo. Significa que a janela única no período de verificação de cerca de 5 anos, mencionada anteriormente, é fatiada em diversas janelas e para cada uma é escolhida uma rede neural. Kaastra e Boyd (1996) justificam o emprego dessa técnica como a maneira de a rede neural se adaptar às condições mais recentes do mercado. Outra razão consiste em reduzir a possibilidade de a função de ativação tangente hiperbólica da rede neural treinada operar em sua região de saturação no período total de verificação, pois no cotidiano, sem o conhecimento *a priori* da amostra desse período, nada se sabe sobre o intervalo dos atributos dessa amostra. Conforme foi comentado na Seção 3.3 os parâmetros da normalização são definidos pela amostra de treinamento propriamente dito e aplicados na normalização da amostra do período de verificação. Portanto, com o janelamento móvel a chance de o índice pontuar, no período de verificação, muito acima ou abaixo dos valores máximos ou mínimos registrados no treinamento é menor e desejável. Neste trabalho essa possibilidade inexistiu, pois o índice Standard & Poor's 500 se manteve no período de verificação dentro dos intervalos das amostras de treinamento, tanto no caso do janelamento único, quanto no caso do janelamento móvel.

A Figura 4.1 ilustra como as amostras de verificação, de treinamento, validação e teste se movem ao longo do tempo. Nos experimentos deste trabalho as janelas



móveis dos sucessivos períodos de verificação cobriram as 260 semanas no intervalo entre 13/07/2007 e 29/06/2012. O tamanho máximo das amostras de treinamento, validação e teste foi de 900 semanas, sendo a primeira semana finda em 06/04/1990.

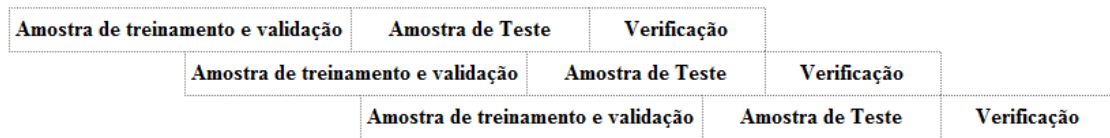


Figura 4.1: Técnica do janelamento móvel

As amostras de treinamento e validação têm como objetivo o treinamento da rede neural, procedimento por meio do qual os pesos e vieses são fixados; a rede neural é escolhida pelo critério de seleção na amostra de teste (desconhecida pela rede) e, finalmente, é avaliada no período de verificação, que pode ser visto como o período de vida útil da rede neural. Como as amostras de treinamento, validação e teste foram randomizadas para melhor cobrir o domínio da função mapeada, essas amostras deixaram de ser constituídas por uma série temporal das variáveis de entrada.

## 4.7 Critério de Escolha da Rede Neural

Neste trabalho foram usados os dados dos fechamentos semanais do índice Standard & Poor's 500 entre as semanas findas em 06/04/1990 e 29/06/2012, totalizando 1160 semanas. Na pré-seleção das configurações as amostras de treinamento, validação e teste ficaram compreendidas entre as semanas findas entre 06/04/1990 e 06/07/2007, totalizando 900 semanas. O período de verificação, no qual foram feitas as estimativas dos fechamentos semanais do índice Standard & Poor's 500 pelas redes neurais testadas, ficou compreendido entre as semanas findas entre 13/07/2007 e 29/06/2012, totalizando 260 semanas ou cerca de 5 anos.

Foi empregada uma notação para facilitar a identificação das configurações testadas. Como exemplo de notação uma rede neural com 3 neurônios na camada escondida e

com todos os atributos na equação (4.5), com os parâmetros  $\Psi$ ,  $\Omega$ ,  $\Gamma$ ,  $\Theta$  e  $\Phi$  iguais a 1 (o parâmetro de  $d_i$  é sempre igual a 1), a notação da configuração é 1-1-1-1-1-N3. Para atributos  $d_i$ ,  $d_{i-1}$ ,  $MACD_i$  e com 1 neurônio na camada escondida a notação da configuração é 1-1-0-0-1-0-N1.

#### 4.7.1 Primeira Etapa da Seleção das Configurações

Devido à enorme carga computacional exigida pelo emprego da técnica do janelamento móvel, a escolha da configuração vencedora foi feita em 2 etapas. Na primeira etapa as configurações candidatas a integrar o sistema de *trading* foram pré-selecionadas por meio de critérios abordados mais adiante, testadas no período de verificação entre 13/07/2007 e 29/06/2012, de 260 semanas ou cerca de 5 anos, sem o emprego do janelamento móvel. Foram experimentadas nessa fase 10 configurações: 1-1-1-1-1-0-N3, 1-1-1-1-1-0-N2, ambas sem o  $ROC_i$ , 1-1-1-1-0-1-N3, 1-1-1-1-0-1-N2, ambas sem o  $MACD_i$ , 1-1-1-0-1-0-N2, sem o  $ROC_i$ , 1-1-1-0-0-1-N2, sem o  $MACD_i$ , 1-1-0-0-1-0-N1, sem o  $ROC_i$ , 1-1-0-0-0-1-N1, sem o  $MACD_i$ , 1-0-0-0-1-1-N2 e 1-0-0-0-1-1-N1, ambas com  $MACD_i$  e  $ROC_i$ .

Geralmente o critério de escolha da rede neural na previsão de preços das ações é o do menor RMSE\_R na amostra de teste ou métricas congêneres. Todavia, embora seja uma condição desejável ela poderá não ser suficiente e ensejar desempenhos insatisfatórios do sistema de *trading*, pois um desvio material na estimativa é inconsequente desde que a oscilação semanal do índice seja corretamente prevista. O sistema de *trading* somente é lucrativo se a rede neural acerta a direção do preço do ativo financeiro, a despeito de um desvio material na estimativa, ou seja, de um RMSE\_R elevado. Foi observado nos experimentos que uma taxa de acerto nas previsões das oscilações semanais do índice acima de 60% na amostra de teste era geralmente associada a um maior retorno anual nominal do sistema de *trading* no período de verificação. Na escolha das configurações capazes de apresentar bons retornos no período de verificação foram geradas 1000 redes neurais para cada uma das 10 configurações elencadas ante-

riormente nesta Subseção. Cada uma dessas redes neurais foi escolhida entre 5 obtidas por um laço na sintaxe do MATLAB em cada simulação (ver **Apêndice C**). Entre essas 5 redes neurais geradas em cada simulação foi escolhida para integrar o rol das 1000 redes neurais (por configuração) a que apresentou o menor MSE (*mean squared error*), com parâmetros *default* do MATLAB, na amostra de treinamento. Com os pesos fixados nos treinamentos foram apurados para cada uma das 1000 redes neurais o RMSE\_R na amostra de teste, a taxa de acerto nas previsões das oscilações semanais do índice pelas redes neurais nessa amostra de teste e a taxa de retorno anual nominal do sistema de *trading* no período de verificação.

Para cada configuração, entre as 10 inicialmente selecionadas, foram escolhidas as redes neurais que atenderam os 2 critérios relacionados abaixo, sendo 10 redes neurais para cada critério, totalizando 20 redes neurais, 2 para cada configuração:

(1) menor RMSE\_R na amostra de teste pertencente ao subconjunto que apresentou taxa de acerto superior a 60% nessa mesma amostra de teste, critério proposto neste trabalho;

(2) menor RMSE\_R na amostra de teste, sem a restrição da taxa de acerto, critério usualmente empregado.

As configurações correspondentes às redes neurais com o menor RMSE\_R na amostra de teste e com a maior taxa anual de retorno no período de verificação pelos critérios 1 e 2 foram pré-selecionadas. Portanto, foram escolhidas até 4 configurações (uma configuração poderá atender os 2 critérios) para serem testadas pelo janelamento móvel, que será abordado na subseção seguinte.

O percentual de 60% estabelecido no critério 1 acima foi escolhido de modo que fosse superior a 50%, condição que aumentaria a chance do sistema de *trading* ser ganhador, que fosse plausível ser atingido por uma quantidade razoável de redes neurais

nas 1000 simulações e ser realista. Entre as 1000 simulações da configuração escolhida (Seção 5.2) 73 redes neurais, ou 7,3%, atingiram taxa de acerto superior a 60% na amostra de teste, sendo que a maior taxa de acerto foi de 66,1%.

### 4.7.2 Segunda Etapa da Seleção da Configuração e Redes Neurais

Seja  $J$  o **tamanho da janela móvel** de tempo do período de verificação em número de semanas e  $T$  um **múltiplo** de  $J$  que define o tamanho das amostras de treinamento, validação e teste, ou seja, o produto entre  $J$  e  $T$ . Nos experimentos, para cada uma das até 4 configurações selecionadas pelos critérios 1 e 2 na Subseção 4.7.1, foram testados  $J$  igual a 52 semanas, equivalente a 1 ano, e  $T$  iguais a 6 e 17. Como o período total de verificação é de 260 semanas, nesse caso foram escolhidas 5 redes neurais ( $260/52$ ) para as 5 janelas móveis. Para  $J$  igual a 26 semanas, equivalente a 1 semestre, foram testados  $T$  iguais a 34, 17 e 6. Nesse caso foram escolhidas 10 redes neurais ( $260/26$ ) para as 10 janelas móveis. Para  $J$  igual a 13 semanas, equivalente a 1 trimestre, foram testados  $T$  iguais a 69 e 34. Nesse caso foram escolhidas 20 redes neurais ( $260/13$ ) para as 20 janelas móveis. Nos experimentos, uma vez estabelecido  $J$ , esse termo foi mantido constante durante o período de verificação total compreendido entre 13/07/2007 e 29/06/2012.

Entre as até 4 configurações, tamanhos de período de verificação  $J$  e tamanhos das amostras de treinamento, validação e teste  $JT$ , foram escolhidas a configuração e respectivas redes neurais que apresentaram a maior taxa anual de retorno no período de verificação nessa segunda etapa.

# Capítulo 5

## Resultados dos Experimentos

A escolha da configuração e respectivas redes neurais para compor o sistema de *trading* foi efetuada em 2 etapas devido à enorme exigência computacional da técnica do janelamento móvel. Na primeira etapa, conforme foi comentado no Capítulo 4, Seção 4.7, foram selecionadas as configurações que atenderam os critérios estipulados sem o emprego da técnica do janelamento móvel. Entre essas configurações foi escolhida, na segunda etapa, a configuração que apresentou o melhor rendimento no período de verificação com o emprego do janelamento móvel, bem como foi identificada o tamanho mais apropriado das janelas. Os resultados são apresentados a seguir.

### 5.1 Pré-Seleção das Configurações com Janela Única

O quadro à esquerda na Figura 5.1 mostra as taxas anuais de retornos nominais **no período de verificação** de cerca de 5 anos das redes neurais de cada uma das 10 configurações inicialmente selecionadas na Subseção 4.7.1 que atenderam o critério 1 e os respectivos RMSE\_R nas amostras de teste. A configuração 1-1-0-0-1-0-N1 apresentou o menor RMSE\_R e a configuração 1-1-1-0-1-0-N2 a maior taxa anual de

retorno e são indicadas por setas na Figura 5.1.

O quadro à direita mostra os retornos anuais no período de verificação das redes neurais de cada uma das mesmas 10 configurações e os respectivos RMSE\_R nas amostras de teste que atenderam o critério 2 da Subseção 4.3. A configuração 1-0-0-0-1-1-N2 apresentou o menor RMSE\_R e a configuração 1-1-1-0-1-0-N2 a maior taxa anual de retorno no período de verificação e são indicadas por setas na mesma Figura.

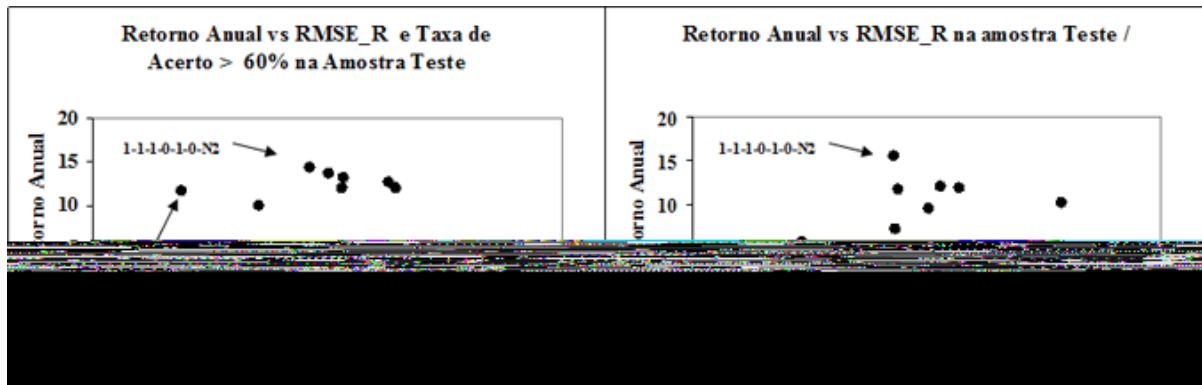


Figura 5.1: RMSE\_R e retornos anuais com janelamento único

A Tabela 5.1 abaixo dispõe os resultados ilustrados na Figura 5.1, para cada configuração e critérios, e em negrito as configurações escolhidas para serem testadas pela técnica do janelamento móvel, indicadas por setas na Figura 5.1.

Configuração	Critério 1		Critério 2	
	RMSE_R	Retorno	RMSE_R	Retorno
1-1-1-1-1-0-N3	0,017067	13,03%	0,017057	5,23%
1-1-1-1-1-0-N2	0,017261	12,58%	0,017076	10,20%
1-1-1-1-0-1-N3	0,017059	11,96%	0,016511	9,51%
1-1-1-1-0-1-N2	0,016705	9,98%	0,016367	7,24%
<b>1-1-1-0-1-0-N2</b>	0,016927	<b>14,34%</b>	0,016357	<b>15,44%</b>
1-1-1-0-0-1-N2	0,017005	13,55%	0,016640	11,75%
<b>1-1-0-0-1-0-N1</b>	<b>0,016377</b>	11,59%	0,016377	11,59%
1-1-0-0-0-1-N1	0,017293	11,96%	0,016561	11,99%
<b>1-0-0-0-1-1-N2</b>	0,017673	1,32%	<b>0,015969</b>	5,71%
1-0-0-0-1-1-N1	0,017531	4,09%	0,016698	3,34%

Tabela 5.1: Resultados com janelamento único

As configurações pré-selecionadas, entre as que atenderam os critérios 1 e 2, foram as que apresentaram o menor RMSE\_R e a maior taxa anual de retorno. A configu-

ração 1-0-0-0-1-1-N2, que apresentou o menor RMSE\_R pelo critério 2, obteve uma taxa anual de retorno nominal de apenas 5,71% e a curva dos rendimentos acumulados (*equity line*) do sistema de *trading* não apresentou uma tendência claramente ascendente ao longo do período de verificação, conforme se pode constatar na Figura 5.2, em que é confrontada com a evolução do índice Standard & Poor's 500. Além disso, pode-se notar que os ganhos significativos foram excessivamente concentrados entre o início das operações e o mês de julho de 2008 (falta de consistência). Por essas 2 razões ela foi descartada como opção para ser testada pela técnica do janelamento móvel.

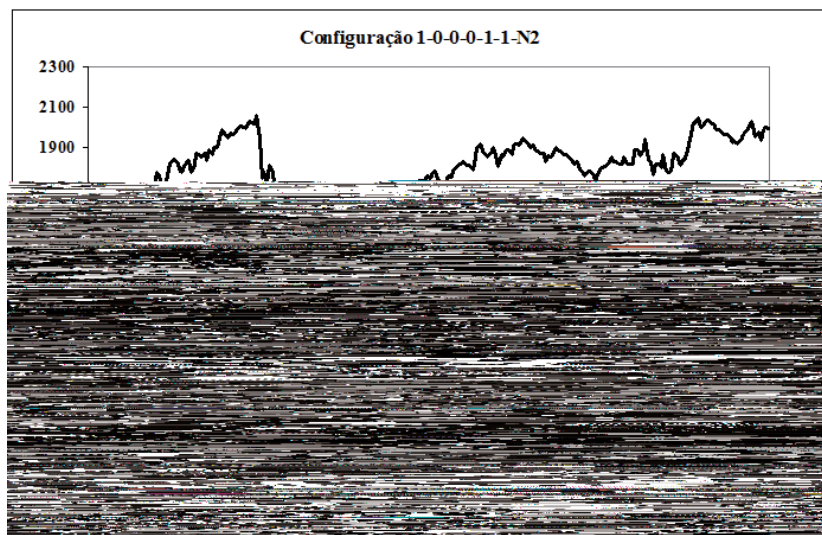


Figura 5.2: Evolução da carteira do sistema de *trading*

A configuração 1-1-1-0-1-0-N2 atendeu os critérios 1 e 2 com a maior taxa anual de retorno e a 1-1-0-0-1-0-N1 apresentou o menor RMSE\_R pelo critério 1. Por outro lado, as curvas do capital empregado no período de verificação de ambas as configurações atenderam o quesito desejado, qual seja, uma evolução continuamente ascendente, embora volátil como é natural em um sistema seguidor de tendência (*trend follower system*). Sobraram, então, essas 2 configurações para serem testadas pela técnica do janelamento móvel.

## 5.2 Seleção da Configuração, do Tamanho da Janela de Verificação e das Amostras de Treinamento, Validação e Teste e das Redes Neurais

A Tabela 5.2 abaixo mostra as taxas anuais nominais de retorno do sistema de *trading*, das redes neurais correspondentes a cada uma das 2 configurações selecionadas na Seção 5.1, quais sejam 1-1-0-0-1-0-N1 e 1-1-1-0-1-0-N2, com combinações de diferentes tamanhos de janelas de verificação  $J$  e de treinamento, validação e teste  $JT$ .

Janelamento		Configurações			
J	Número de Semanas	1-1-0-0-1-0-N1		1-1-1-0-1-0-N2	
		Critério		Critério	
T	Múltiplo	1	2	1	2
J-T	13-69	9,31%	8,97%	8,32%	5,79%
J-T	13-34	8,05%	7,15%	-1,67%	-5,07%
J-T	26-34	10,08%	13,42%	10,78%	8,93%
J-T	26-17	7,49%	-0,54%	0,51%	9,74%
J-T	26-06	-1,03%	-7,03%	-12,94%	-9,52%
J-T	<b>52-17</b>	<b>15,42%</b>	9,70%	9,70%	8,84%
J-T	52-06	-2,43%	-2,86%	-0,25%	-4,84%

Tabela 5.2: Resultados com janelamento Móvel

O **critério final** de escolha da configuração e das respectivas redes neurais foi o da **maior taxa anual de retorno nominal no período de verificação** entre 13/07/2007 e 29/06/2012, sendo que a escolhida foi 1-1-0-0-1-0-N1 que apresentou taxa anual de 15,42%, com tamanho de janela  $J$  de verificação igual a 52 semanas, portanto com 5 redes neurais (260/52), e de  $T$  igual a 17, totalizando 884 (52x17) semanas de amostras de treinamento, validação e teste. A melhor opção, com atributos do vetor de entrada  $d_i$ ,  $d_{i-1}$ ,  $MACD_i$  e 1 neurônio na camada escondida, sendo  $i$  a semana do último fechamento conhecido, atendeu o critério do menor RMSE\_R na amostra de teste entre as que apresentaram taxas de acerto nessa amostra acima de 60%, critério 1 sugerido neste trabalho.

Entre as 7 opções de janelamento  $J - T$  para as 2 configurações pré-selecionadas,



totalizando 14 opções, mostradas na Tabela 5.2, somente em 2 casos a taxa anual de retorno nominal no período de verificação pelo critério 1 foi inferior à obtida pelo critério 2, usualmente empregado. Esses 2 casos foram registrados com janelamento  $J - T$  igual a 26-34 para a configuração 1-1-0-0-1-0-N1, cuja taxa anual de retorno foi de 13,42%, e  $J - T$  igual a 26-17 para a configuração 1-1-1-0-1-0-N2, cuja taxa anual de retorno foi de 9,74%. Constata-se que a redução de  $T$ , múltiplo de  $J$  que define o tamanho das amostras de treinamento, validação e teste, piorou o desempenho do sistema de *trading* pelos 2 critérios em 15 entre 16 casos. A exceção ocorreu com a configuração 1-1-1-0-1-0-N2 em que a redução de  $T$  de 34 para 17 pelo critério 2 elevou a taxa anual de retorno de 8,93% para 9,74%. O fato de terem sido realizadas 1000 simulações para cada tamanho de janela e em cada simulação ter sido criado um laço de 5 iterações para encontrar a rede neural com o menor MSE (*mean squared error*) na amostra de treinamento, embora tenha exigido uma expressiva carga computacional, foi um procedimento adequado. Os experimentos foram realizados 2 vezes para a configuração escolhida 1-1-0-0-1-0-N1, com  $J - T$  igual a 52-17, sendo que as métricas de desempenho mostradas a seguir na Seção 5.3 foram iguais nas 2 rodadas até a segunda casa decimal.

A Tabela 5.3 abaixo resume os retornos, para as 2 configurações selecionadas, correspondentes aos tamanhos de janelas móveis de verificação  $J$  iguais a 13, 26 e 52 semanas, associados aos maiores tamanhos das amostras de treinamento, validação e teste, e os respectivos critérios 1 e 2 de escolha. Constata-se que entre as 6 possibilidades de configuração e janelamento, 5 obtidas pelo critério 1, proposto como alternativa neste trabalho, apresentaram as maiores taxas anuais de retorno no período de verificação. Esses resultados mostram que o critério 1 sugerido neste trabalho, que consistiu na escolha das redes neurais que apresentaram o menor RMSE\_R nas amostras de testes pertencentes ao subconjunto com taxas de acerto das oscilações do índice Standard & Poor's 500 acima de 60% nessas amostras, é uma alternativa viável ao critério usualmente empregado, baseado no menor RMSE\_R nas amostras de teste.

J-T	Configurações			
	1-1-0-0-1-0-N1		1-1-1-0-1-0-N2	
	Retorno	Critério	Retorno	Critério
52-17	15,42%	1	9,70%	1
26-34	13,42%	2	10,78%	1
13-69	9,31%	1	8,32%	1

Tabela 5.3: Resultados por critério e configuração

### 5.3 Métricas de Desempenho do Sistema de *Trading*

A Tabela 5.4 mostra as métricas de desempenho do sistema de *trading* da configuração 1-1-0-0-1-0-N1, tamanho de janela  $J$  igual a 52 semanas e  $T$  igual a 17 (884 semanas), referentes ao período de verificação entre 13/07/2007 e 29/06/2012, que apresentou nesse período uma taxa anual (geométrica) de retorno de 15,42%, diante de um retorno anual da estratégia compre e segure de -2,30% (negativo).

Métricas de Desempenho									
Taxa de Retorno	Taxa de Acerto Operações	Taxa de Acerto Semanal	Período Médio Operações	<i>Payout</i>	Índice Sharpe	Volatilidade Anual	Percentual Operações Compras	Melhor Retorno	Pior Retorno
15,42%	75,28%	59,62%	2,92	2,75	0,436	9,78%	49,44%	4,68%	-18,51%

Tabela 5.4: Resultados do sistema de *trading*

O intervalo entre 2 operações sucessivas foi de 2,92 semanas em média e nesse intervalo o sistema de *trading* permaneceu inativo. Uma operação é mantida de uma semana para a outra, mesmo com prejuízo, se o sistema não alterar a recomendação de compra ou de venda do índice. As operações vitoriosas, no total de 67, perduraram por 172 semanas e as perdedoras, no total de 22, por 88 semanas. Para cada 100 operações realizadas 75,28 foram bem-sucedidas. Convém mencionar que a taxa de acerto das operações de 75,28% durante o período de verificação não se compara com a taxa de acerto das previsões semanais das redes neurais no período de

**verificação** que foi de 59,62%, pois a base de cálculo é diferente, além do que 1 operação bem-sucedida (malsucedida) pode conter durante a sua duração erros (acertos) em uma ou mais semanas, conforme se pode constatar no **Apêndice B**. Nesse Apêndice foram detalhados os cálculos dos rendimentos semanais e das operações nas primeiras e últimas operações no período de verificação. A razão entre a soma dos lucros obtidos nas operações lucrativas e a soma dos prejuízos nas operações perdedoras foi de 2,75 vezes (*payout*). Ou seja, para cada U\$ 1 de prejuízo foram auferidos ganhos de U\$ 2.75. As operações de compra de 1 contrato do índice representaram 44,76% dos resultados financeiros, enquanto que as operações de venda contribuíram com 55,24%. As quantidades de operações de compra e de venda foram balanceadas, divididas em 49,44% e 50,56%, respectivamente. O melhor retorno de uma operação foi de 4,68% e o pior foi um prejuízo de -18,51%, ambos calculados sobre os valores das posições financeiras acumuladas pelo sistema de *trading*, apurados até às operações imediatamente anteriores, sendo que essas taxas não são anualizadas.

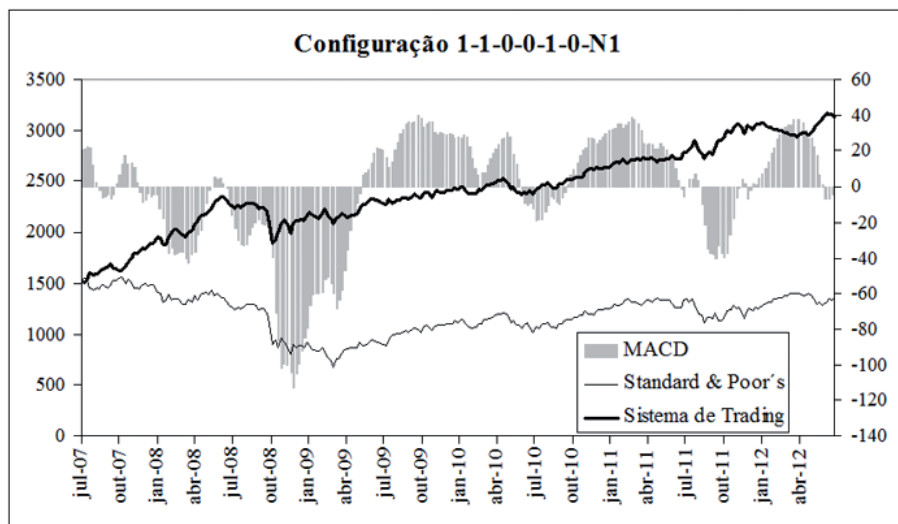


Figura 5.3: Evolução da carteira do sistema de *trading*

A Figura 5.3 mostra, na escala à esquerda, a evolução do capital nominal empregado no sistema de *trading* em 1 contrato do índice Standard & Poor's 500 no período de verificação, antes da apuração do imposto sobre a renda e custos operacionais, con-

frontadas com a do índice em si, relativas à configuração 1-1-0-0-1-0-N1, e o histograma do MACD na escala à direita. A curva do capital empregado no sistema de *trading*, mostrada na Figura 5.3, contrasta visivelmente com a exibida na Figura 5.2. Enquanto que a curva do sistema de *trading* apresenta um comportamento razoavelmente suave e com viés de alta persistente, exceto no mês de setembro de 2008 quando foi registrada a perda de -18,51%, a curva da configuração rejeitada, mostrada na Figura 5.2, apresenta o formato de um dente de serra e sem viés de alta definido.

Se fossem deduzidos os custos com emolumentos e corretagens à alíquota total de 0,076% (Subseção 4.3) incidentes em cada operação a taxa anual de retorno cairia para 14,17% ao ano, sem o reinvestimento dos lucros e aplicação na renda fixa das sobras de caixa ociosas. Por outro lado, se fosse incorporada uma alavancagem de 50%, considerada normal para os padrões dos fundos multimercados, o resultado financeiro e os custos aumentariam na mesma proporção e, conseqüentemente, a taxa anual de retorno subiria para 19,23% ao ano, sem o reinvestimento dos lucros e aplicação na renda fixa das sobras de caixa ociosas. O efeito colateral negativo da alavancagem seria o aumento do risco medido pela volatilidade anual dos retornos mensais que subiria de 9,78% para 14,21%.

Quando o MACD é crescente (decrecente) em uma tendência de alta (baixa) do índice Standard & Poor's 500 diz-se que a tendência ganha momento, caso contrário perde momento ou força. Por outro lado, se o MACD permanece por longo tempo orbitando em torno da linha do zero diz-se que a tendência está sem momento, equivalente a uma indefinição sobre a direção dos preços. Geralmente um sistema seguidor de tendência não é lucrativo em um mercado com baixo momento, como no segundo semestre de 2007 (início do período de verificação). Entretanto, surpreendentemente, o sistema de *trading* apresentou o seu melhor desempenho nessa fase. Outra situação que é desfavorável a esse tipo de sistema de *trading* ocorre quando o MACD muda de direção subitamente e de maneira aguda, não dando tempo suficiente para a rede neural detectar a mudança dos ventos e responder adequadamente. Isso ocorreu na crise

econômica mundial no segundo semestre de 2008 e foi responsável pela pior fase do sistema proposto. Uma outra explicação plausível para a perda significativa registrada nesse período pode ser devido ao treinamento inadequado, pois a rede neural não foi capaz de atribuir maior importância aos eventos de cauda (*tail events*) em relação aos movimentos de preços com baixa oscilação, mais comuns e que geram resultados financeiros triviais. Por outro lado, no segundo semestre de 2011 o MACD se encontrava em visível tendência de alta, porém o sistema não foi capaz de usufruir desse instante que geralmente lhe é favorável. Com as exceções acima mencionadas, de forma geral o sistema de *trading* apresentou um desempenho bastante satisfatório e uma evolução do capital empregado menos volátil do que normalmente se espera de sistemas seguidores de tendência.

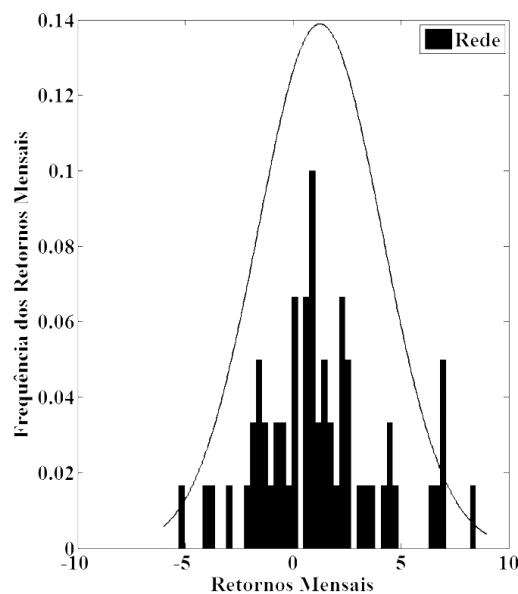


Figura 5.4: Distribuição dos retornos mensais

A Figura 5.4 mostra o histograma dos retornos mensais medidos pelas variações percentuais dos rendimentos entre as últimas sextas-feiras (ou o dia útil imediatamente anterior) de cada mês, cuja média aritmética foi de 1,24% ao mês, sem inclusão dos custos. A distribuição apresentou uma curtose de 3,2793, assimetria positiva e desejável de 0,40179, índice *Sharpe* (razão entre o rendimento esperado e o desvio padrão, com o

rendimento mínimo - *benchmark* - igual a 0% ao ano) igual a 0,436 (medida de gestão de risco). A distribuição gaussiana na figura apresenta a mesma média e desvio padrão da distribuição dos rendimentos mensais.

## Capítulo 6

# Conclusões e Recomendações

Os resultados dos experimentos mostraram que redes neurais podem servir de instrumentos de apoio nas decisões de compra e venda no mercado de ações, desde que as escolhas dos atributos, da arquitetura, do critério de seleção das redes neurais, do tamanho das janelas móveis e das amostras de treinamento, validação e teste sejam apropriadas. O período de verificação nos experimentos foi longo, cerca de 5 anos, marcado por fortes oscilações, e o sistema de *trading* apresentou uma taxa anual de retorno de 15,42% diante de um revés de -2,30% da estratégia compre e segure, que tem como referência (*benchmark*) o índice Standard & Poor's 500. Por outro lado, a curva do capital empregado apresentou viés ascendente, propriedade fundamental de um bom sistema, com baixa volatilidade em se tratando de um sistema seguidor de tendência. Apesar do desempenho bastante satisfatório, uma avaliação mais conclusiva do sistema de *trading* somente poderia ser feita após a metodologia ser testada em períodos de verificação com duração ainda mais longa, que abrangesse prolongada fase com baixo momento nos preços e que fosse empregada em outros mercados financeiros, como de moedas e *commodities*.

As variáveis de entrada da rede neural, usadas na implementação das operações de

compra e de venda a descoberto do índice Standard & Poor's 500, foram escolhidas com o propósito de capturar a anomalia do efeito momento dos preços no curto prazo, reconhecida por Fama e Litterman (2012). Um dos indicadores de momento utilizado foi o MACD (*moving average convergence divergence*), cujos parâmetros foram os mesmos empregados por Klassen (2005). Uma ideia a ser explorada seria usar um algoritmo genético no período de treinamento para otimizar esses parâmetros pelo critério da taxa de acerto da rede neural sobre a direção do índice Standard & Poor's 500. A otimização de parâmetros de indicadores técnicos com base em um curto período de treinamento quase sempre é prejudicial aos sistemas de *trading* devido à perda de generalização do sistema. Porém, como a duração do treinamento nos experimentos foi muito longa, de 884 semanas (52x17), equivalente a 17 anos, esse risco é mitigado. Um ponto negativo no uso de redes neurais com as entradas usadas neste trabalho foi o tamanho excessivo exigido para as amostras de treinamento, validação e teste que poderiam impedir o emprego da metodologia em índices com reduzida base de dados. Por essa razão convém explorar a metodologia na frequência diária e intradiária que apresentam bases de dados maiores. Como parte da amostra foi usada para a validação cruzada, poderia ser testada a ideia de usar o algoritmo *Bayesian Regularization* sem o emprego dessa técnica. Convém também explorar a viabilidade de usar, no lugar da série temporal de preços empregada neste trabalho, os logaritmos dos retornos, pois esses tendem a apresentar uma distribuição gaussiana e, eventualmente, requisitar uma base de dados de treinamento menor.

Os resultados mostraram que o critério de seleção das redes neurais empregado nos experimentos, que consistiu na escolha das redes neurais que apresentaram o menor RMSE\_R nas amostras de testes pertencentes ao subconjunto com taxas de acerto das oscilações do índice Standard & Poor's 500 acima de 60% nessas amostras, é uma alternativa ao critério usualmente empregado, baseado no menor RMSE\_R (ou métricas congêneres) nas amostras de teste. Foram testados tamanhos de janelas móveis de 52, 26 e 13 semanas, e diversos tamanhos de amostras para o treinamento, sendo que a



escolha recaiu sobre a janela de maior duração, ou seja, de 52 semanas e base de dados para o treinamento 17 vezes o tamanho da janela móvel. O critério sugerido neste trabalho para a seleção das redes neurais apresentou melhor desempenho em relação ao critério usual para quase todos os tamanhos de janelas móveis, quando foi usada a base de dados de treinamento mais longa possível. Variáveis de entrada das redes neurais diferentes das que foram usadas neste trabalho poderão apresentar um tamanho ótimo da janela móvel diferente da que foi obtida no período de verificação.

No cotidiano é recomendável que se faça a normalização dos atributos de entrada e respectivas saídas desejadas no intervalo  $[-0,75, +0,75]$  com a finalidade de reduzir a possibilidade de as redes neurais operarem na região de saturação das funções de ativação nos períodos de verificação, cujos intervalos de oscilações do índice não são conhecidos *a priori*, como ocorre no cotidiano. Neste trabalho optou-se pela normalização no intervalo  $[-1, 1]$  porque no período de verificação, previamente conhecido, o índice Standard & Poor's 500 permaneceu no intervalo das amostras de treinamento, validação e teste. Como em apenas uma operação foi registrada uma perda de -18,51%, é recomendável explorar a sugestão de Kaastra e Boyd (1996) de tornar as redes neurais mais especializadas em detectar as grandes oscilações na direção correta, mediante a eliminação de amostras de treinamento correspondentes às oscilações muito pequenas do índice Standard & Poor's 500.

# Referências Bibliográficas

- Back, B.; Litinen, T.; Sere, K. e van Wezel, M. (1996). Choosing bankruptcy predictors using discriminant analysis, logit and genetic algorithms. *Technical Report-Centre for Computing Science*, 40.
- Chang, H. C.; Chou, J.-H.; Chen, C.-T. e Hsieh, C.-S. (2008). Hybrid method of using neural networks and arma model to forecast value at risk (var) in the chinese stock market. *J. Stat. Manag. Syst.*, 11:1093–1108.
- Chen, S. M. (1996). Forecasting enrollments based on fuzzy time-series. *Fuzzy Sets Systems*, 81:311–319.
- Chng, E.; S., S. C. e Mulgrew, B. (1996). Gradient radial basis function networks for nonlinear and nonstationary time series prediction. *IEEE Transactions on Neural Networks*, 7:190–194.
- Connor, J.; Martin, R. D. e E., A. L. (1994). Recurrent neural networks and robust time series prediction. *IEEE Transactions on Neural Networks*, 5:240–254.
- Delyon, B.; Juditsky, A. e Benveniste, A. (1996). Accuracy analysis for wavelet approximations. *IEEE Transactions on Neural Networks*, 6:332–348.
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 5:427–465.
- Fama, E. F. e Litterman, R. (2012). An experienced view on markets and investing. *Financial Analysts Journal*, 68:15–19.

- Haykin, S. (2009). *Neural Networks and Neural Machines*. Pearson Education, Inc.
- Hebb, D. (1949). *The Organization of Behavior*. John Wiley & Sons, Inc.
- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *National Academy of Science*, 79:2554–2558.
- Hull, J. (1991). *Introduction to Futures and Options Markets*. Prentice Hall, New Jersey.
- Jegadeesh, N. e Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48:65–91.
- Kaastra, I. e Boyd, M. (1996). Designing a neural network for forecasting financial and economic time series. *Neurocomputing*, 10:215–236.
- Kim, K. J. (2004). Artificial neural networks with feature transformation based on domain knowledge for the prediction os stock index futures. *International Journal of Intelligent Systems in Accounting and Finance Management*, 12:167–176.
- Kimoto, T.; Asakawa, K.; Yoda, M. e Takeoka, M. (1990). Stock market prediction system with modular neural networks. *Proceedings of the IEEE International Joint Conference on Neural Networks*, pp. 11–16.
- Klassen, M. (2005). Effect of some technical indicators in stock market time-series forecasting using neural networks. *Informatika*, 5:75–79.
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43:59–69.
- Kryzanowski, L.; Galler, M. e Wright, D. W. (1993). Using artificial neural networks to picking stocks. *Financial Analyst Journal*, pp. 21–27.
- Maass, W. (1997). The third generation of neural network models. *Neural Networks*, 10:1659–1671.

- Maciel, L. S. e Ballini, R. (2010). Neural networks applied to stock market forecasting: An empirical analysis. *Journal of the Brazilian Neural Network Society*, 8:3–22.
- Mackay, J. C. D. (1992). Bayesian interpolation. *Neural Computational*, 4:415–447.
- Malliaris, M. e Salchenberger, L. (1996). Using neural networks to forecast the s&p 100 implied volatility. *Neurocomputing*, 10:183–195.
- Mammadov, M. A. (2004). A new global optimization algorithm based on dynamical systems approach. in *Proceedings of the 6th International Conference on Optimization: Techniques and Applications (ICOTA 04)*.
- Martinez, L. C.; da Hora, D. N.; de M. Palotti, J. R.; Jr., W. M. e Pappa, G. L. (2009). From an artificial neural network to a stock market day-trading system: A case study on the bm&f bovespa. *The 2009 International Joint Conference on Neural Networks*.
- Matsumoto, É. Y. (2009). Aplicação de redes neurais na classificação de rentabilidade futura de empresas. *Anais do IX Congresso Brasileiro de Redes Neurais-SBRN*.
- McCulloch, W. e Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115–133.
- Minsky, M. e Papert, S. (1969). *Perceptrons*. MIT Press, Cambridge.
- Murphy, J. (1999). *Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications*. New York Institute of Finance, New York.
- Odom, M. D. e Sharda, R. (1990). A neural network model for bankruptcy prediction. *Proceedings of the IEEE International Conference on Neural Networks, San Diego*, pp. 163–168.
- Park, J. e Sandberg, I. (1991). Universal approximation using radial basis function networks. *Neural Computation*, 3:246–257.

- Raposo, R. e Cruz, A. J. O. (2002). Stock market prediction based on fundamentalist analysis with fuzzy-neural networks. *Proceedings of 3rd WSES International Conference on Fuzzy Sets & Fuzzy Systems (FSFS'02), Neural Networks and Applications (NNA'02), Evolutionary Computation (EC'02)*.
- Rau, R. (2010). *Behavioral Finance: Investors, Corporations and Markets*, pp. 333-349. John Willy & Sons, New Jersey.
- Roman, J. e Jameel, A. (1996). Backpropagation and recurrent neural networks in financial analysis of multiple stock market returns. *Proceedings of the 29th Annual Hawaii International Conference on System Sciences*.
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65:386–408.
- Rumelhart, D. E.; Hinton, G. E. e R.J., W. (1986). Learning representations of back-propagation errors. *Nature*, 323:533–536.
- Salchenberger, L. M.; Cinar, E. M. e Lash, N. A. (1992). Neural networks: A new tool for predicting thrift failures. *IEEE Transactions on Neural Networks*, 23:899–916.
- Shiller, R. J. (2005). *Irrational Exuberance*. Princeton Univerty Press, 2nd ed., New Jersey.
- Shin, Y. e Ghosh, J. (1995). Ridge polynomial networks. *IEEE Transactions on Neural Networks*, 6:610–622.
- Singleton, J. e Surkan, A. (1990). Neural networks for bond rating improved by multiple hidden layers. *Proceedings of the IEEE International Conference on Neural Networks*, pp. 163–168.
- Tam, K. e Kiang, M. Y. (1992). Managerial applications of neural networks: The case of bank failure predictions. *IEEE Transactions on Neural Networks*, 38:926–947.

- Tilakaratne, C. D.; Mammadov, M. A. e Morris, S. A. (2009). Modified neural network algorithms for predicting trading signals of stock market indices. *Journal of Applied Mathematics and Decision Sciences*.
- Wei, L. Y. e Cheng, C. H. (2012). A hybrid recurrent neural networks model based on synthesis features to forecast the taiwan stock market. *International Journal of Innovative Computing, Information and Control*, 8.
- Werbos, P. J. (1974). *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. Phd dissertation, Harvard University, Cambridge.
- White, H. (1988). Economic prediction using neural networks: The case of ibm daily stock returns. *Proceedings of the IEEE International Conference on Neural Networks*, 2:609–624.
- Yu, H. K. (2005). Weighted fuzzy time-series models for taiex forecasting. *Physica A*, 349:451–458.
- Zhang, G. e Hu, M. Y. (1998). Neural network forecasting of the british pound/us dollar exchange rate. *Elsevier*, 26:495–506.
- Zhang, G.; Patuwo, B. E. e Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14:35–62.

# APÊNDICE A

A apresentação do algoritmo *Bayesian Regularization* (BR) a seguir segue a abordagem feita por Foresee e Hagan (1997). O objetivo do algoritmo BR é minimizar a função de custo  $F_c(\mathbf{W})$  definida pela equação (3.7), reproduzida abaixo:

$$F_c(\mathbf{W}) = \beta \xi(\mathbf{W}) + \alpha E(\mathbf{W})$$

em que  $\beta$  e  $\alpha$  são fatores de regularização obtidos por inferência bayesiana,  $E(\mathbf{W})$  é a soma quadrática dos pesos e vieses e  $\xi(\mathbf{W})$  da equação 3.5.

Sejam  $\mathbf{H}$  a matriz Hessiana  $P \times P$ , sendo  $P$  o número de pesos e vieses, igual à derivada parcial de segunda ordem de  $\xi(\mathbf{W})$  em relação a  $\mathbf{W}$ ,  $\mathbf{I}$  a matriz identidade,  $\lambda$  um parâmetro de pequeno valor (*damping factor*) para evitar que  $\mathbf{H}$  se aproxime da condição de singularidade,  $\mathbf{g}$  o gradiente  $P \times 1$ , a derivada parcial de primeira ordem de  $\xi(\mathbf{W})$  em relação a  $\mathbf{W}$ . De acordo com o método *Levenberg-Marquardt* (Haykin (2009)) tem-se:

$$\Delta \mathbf{W} = (\mathbf{H} + \lambda \mathbf{I}) \mathbf{g}$$

em que  $\mathbf{H}$  é aproximada na forma abaixo para reduzir as exigências computacionais, procedimento aceitável nas proximidades dos mínimos locais ou globais:

$$\mathbf{H} = \frac{1}{N} \sum_{i=1}^N \left( \frac{\partial y(\mathbf{X}(i), \mathbf{W})}{\partial \mathbf{W}} \right) \left( \frac{\partial y(\mathbf{X}(i), \mathbf{W})}{\partial \mathbf{W}} \right)^T$$

em que  $N$  é o quantidade de amostras e  $y$  é definido pela equação 3.2, reproduzida abaixo:

$$y = \varphi\left(\sum_r w_{sr} \varphi\left(\sum_j w_{rj} \varphi\left(\cdots \varphi\left(\sum_k w_{lk} x_k\right)\right)\right)\right)$$

Por outro lado, considerando os pesos e vieses  $\mathbf{W}$  como variáveis aleatórias, a função de distribuição de  $\mathbf{W}$  é fornecida pela regra de *Bayes*. De acordo com Foresee e Hagan (1997), exposto de forma sumária, tem-se o seguinte desenvolvimento:

$$P(\mathbf{W}|D, \alpha, \beta, R) = \frac{P(D|\mathbf{W}, \beta, R)P(\mathbf{W}|\alpha, R)}{P(D|\alpha, \beta, R)}$$

em que  $D$  representa a amostra de treinamento,  $\beta$  e  $\alpha$  os parâmetros de regularização e  $R$  uma rede neural em particular.  $P(\mathbf{W}|\alpha, R)$  é a densidade de probabilidade *a priori* que representa o conhecimento sobre os pesos e vieses antes da amostra de treinamento ser coletada,  $P(D|\mathbf{W}, \beta, R)$  é a verossimilhança que representa a densidade de probabilidade da amostra  $D$  dado  $\mathbf{W}$ .  $P(D|\alpha, \beta, R)$  é o fator de normalização que garante que a probabilidade total seja igual a 1. Ao assumir distribuições gaussianas para as densidades  $P(D|\mathbf{W}, \beta, R)$  e  $P(\mathbf{W}|\alpha, R)$  obtém-se  $P(\mathbf{W}|D, \alpha, \beta, R)$ , probabilidade *a posteriori*, cujos parâmetros dependem de  $\beta$ ,  $\alpha$  e do fator de normalização. Sob a ótica bayesiana  $P(\mathbf{W}|D, \alpha, \beta, R)$  deveria ser maximizada, que equivale a minimizar a função de custo  $\xi(\mathbf{W})$ . Para otimizar a escolha dos parâmetros de regularização  $\beta$  e  $\alpha$  considera-se que:

$$P(\alpha, \beta|D, R) = \frac{P(D|\alpha, \beta, R)P(\alpha, \beta|R)}{P(D|R)}$$

e assume-se que  $P(\alpha, \beta|R)$  apresenta densidade uniforme e que, portanto,  $P(\alpha, \beta|D, R)$  é maximizada pela maximização de  $P(D|\alpha, \beta, R)$ , cuja densidade é por hipótese também guassiana, sendo o fator de normalização de  $P(\mathbf{W}|D, \alpha, \beta, R)$ . Após o desenvolvi-



mento da equação acima, sendo  $n$  igual ao número de pesos e vieses da rede (parâmetros livres) e  $Tr(\mathbf{H}^{-1})$  igual ao traço de  $\mathbf{H}^{-1}$ , chega-se a:

$$\alpha = \frac{\gamma}{2 E(\mathbf{W})}, \beta = \frac{(N - \gamma)}{2 \xi(\mathbf{W})} \text{ e } \gamma = n - 2 \alpha Tr(\mathbf{H}^{-1})$$

Após a inicialização dos pesos e vieses é estabelecido valores iniciais para  $\alpha = 0$  e  $\beta = 1$ . O algoritmo BR ajusta os pesos e vieses iterativamente para minimizar  $F_c(\mathbf{W})$ . Assim que todas as amostras de treinamento são apresentadas resolve-se as equações referentes a  $\Delta \mathbf{W}$  e  $\mathbf{H}$ , essa última pelo método da retropropagação (*backpropagation*). O fator  $\lambda$  é iniciado com um valor baixo, geralmente em um décimo, e é alterado a cada iteração por um fator, dependendo se a redução de  $\xi(\mathbf{W})$  for rápida, situação em que se pode usar  $\lambda$  menor, ou lenta e, nesse caso, um  $\lambda$  maior. A derivada de primeira ordem de  $y(\mathbf{X}, \mathbf{W})$  é calculada pela técnica da retropropagação. Em seguida computa-se o número de parâmetros livres  $n$  fornecido por  $\gamma$ , recalcula-se  $\beta$  e  $\alpha$ . Recalcula-se a função de custo  $F_c(\mathbf{W})$  e repete-se o procedimento descrito até a convergência da função de custo.

# APÊNDICE B

A tabela abaixo mostra o detalhamento parcial dos cálculos dos rendimentos semanais e das operações para a configuração escolhida.

Semana	Cotação Fech. (Pontos)	Previsão Rede (Pontos)	Direção		Lucro Semanal \$	Operação		Número Operação	Custos \$
			SPX	Rede		Lucrativa \$	Perdedora \$		
0	1530,44			-1					1,16
1	1552,50	1520,55	1	-1	-22,06				
2	1534,10	1542,00	-1	-1	18,40				
3	1458,95	1530,18	-1	1	75,15	71,49		1	2,22
4	1433,06	1467,01	-1	1	-25,89				
5	1453,64	1435,99	1	-1	20,58		-5,31	2	2,21
6	1445,94	1449,68	-1	1	7,70	7,70		3	2,20
7	1479,37	1446,48	1	-1	33,43	33,43		4	2,25
8	1473,99	1472,80	-1	-1	5,38				
9	1453,55	1472,99	-1	1	20,44	25,82		5	2,21
10	1484,25	1455,60	1	-1	30,70	30,70		6	2,26
-	-	-	-	-	-	-	-	-	-
249	1370,26	1394,64	-1	-1	27,82				
250	1378,53	1369,49	1	-1	-8,27				
251	1403,36	1374,98	1	-1	-24,83				
252	1369,10	1397,38	-1	-1	34,26				
253	1353,39	1368,89	-1	-1	15,71				
254	1295,22	1352,73	-1	1	58,17		-17,41	85	1,97
255	1317,82	1299,73	1	-1	22,60	22,60		86	2,00
256	1278,04	1316,17	-1	1	39,78	39,78		87	1,94
257	1325,66	1281,94	1	-1	47,62	47,62		88	2,02
258	1342,84	1322,12	1	-1	-17,18				
259	1335,02	1340,65	-1	-1	7,82				
260	1362,16	1334,81	1		-27,14		-36,50	89	1,04

Tabela 1: Tabulação parcial dos resultados

A **primeira coluna** enumera as semanas do período de verificação, que compreende as de número 1 a 260, e inclui a semana anterior ao início desse período (número igual a 0). A **segunda** e a **terceira colunas** as cotações de fechamento do índice Standard &

Poor's 500 e as projeções correspondentes feitas pelas redes neurais **com 1 semana de antecedência**. Os conteúdos das segunda e terceira colunas geraram os resultados das demais colunas. O número 1 na **quarta coluna** (Direção SPX) indica se o índice subiu (segunda coluna) em relação à semana anterior e o -1 se caiu. Se a estimativa da rede, correta ou não, para a semana seguinte é de alta é atribuído o número 1 na **quinta coluna** (Direção Rede) que resultará, no instante da estimativa, em uma operação de compra (*long*) e, em caso contrário, o número -1 que resultará em uma operação de venda (*short*). Obviamente quando o sistema altera a recomendação em relação à semana anterior é necessário imediatamente fechar a operação em curso, no instante da estimativa, e abrir simultaneamente uma outra na direção contrária. Por outro lado, se a recomendação de compra ou de venda é mantida, a despeito de eventual prejuízo na semana, nenhuma operação é realizada. A **sexta coluna** (Lucro Semanal) mostra os resultados semanais do sistema de *trading* sendo que eventual prejuízo não obriga o cancelamento da operação em curso. A **sétima coluna** (Operação Lucrativa) mostra os resultados das operações bem-sucedidas iguais às somas dos resultados semanais durante os períodos em que as operações foram mantidas. A **oitava coluna** os prejuízos das operações malsucedidas. A **nona coluna** enumera as operações no total de 89 e, finalmente, a **décima coluna** relaciona os custos com corretagens e emolumentos, incidentes sobre os valores da segunda coluna.

Um exemplo ajuda a entender o critério de apuração dos lucros. Na semana que antecedeu o início do período de verificação (semana 0) a rede neural estimou o fechamento do índice na semana 1 em 1520,55. Como esse valor foi inferior ao fechamento conhecido até então, de 1530,44, o sistema iniciou uma operação de venda (aposta em queda da cotação do índice). Como na semana 1 o índice subiu 22,06 pontos o sistema auferiu um prejuízo no mesmo montante. Entretanto, na semana 2 a operação de venda foi mantida porque a estimativa de 1542,00 foi inferior ao fechamento na semana anterior de 1552,50. Como o índice caiu o sistema de *trading* obteve um lucro na semana 2 de \$ 18,40, que compensou parcialmente a perda na semana anterior. Na semana 3

a rede alterou a sua recomendação tendo projetado alta do índice Standard & Poor's 500 na semana seguinte e, portanto, a operação de venda foi zerada e os resultados semanais acumulados enquanto a operação vigeu foi de \$ 71,49. Na semana 260 a rede não efetuou nenhuma estimativa, pois o período de verificação havia sido encerrado e a operação em curso foi compulsoriamente zerada (com prejuízo de \$ 36,50), sem a abertura de uma nova operação.

O número de vezes em que as 5 redes neurais acertaram as estimativas semanais é obtido pela quantidade de valores positivos na sexta coluna, Lucro Semanal. Em 260 projeções o sistema acertou a tendência semanal do índice Standard & Poor's 500 em 155 tentativas, que representa uma taxa de acerto de 59,62%. Em um total de 89 operações 67 foram vitoriosas com ganhos de \$ 2521,24 e 22 operações foram malsucedidas com prejuízos de \$ 916,98, representando uma taxa de acerto de operações de 75,28% e um *payout* de 2,75 vezes (valor absoluto da razão entre ganhos e perdas). O resultado líquido das operações foi de \$ 1604,26 - sem o cômputo dos custos de corretagens e emolumentos - que propiciou no período de verificação uma taxa anual (composta) de retorno sobre o capital inicial de \$ 1530,44 de 15,42% ao ano (ano com 52 semanas).

Os custos (corretagens e emolumentos) foram calculados à alíquota de 0,076% (ver Seção 4.3) incidentes sobre os valores das operações, constantes da segunda coluna (Cotação Fechamento em Pontos). Com as exceções das primeira e última operações, as demais incorreram em custos de fechar a operação anterior e abrir uma nova na direção oposta e, portanto, os custos foram dobrados. Se os custos fossem computados, correspondentes a \$ 166,56 no período de verificação, a taxa anual de retorno cairia para 14,17% ao ano, sem o reinvestimento dos lucros e aplicação na renda fixa das sobras de caixa ociosas. Por outro lado, se fosse usada uma alavancagem de 50% e incluídos os custos, percentual baixo para os padrões dos fundos multimercados, o resultado aumentaria na mesma proporção e a taxa anual de retorno subiria para 19,23% ao ano, sem o reinvestimento dos lucros e aplicação na renda fixa das sobras de caixa ociosas. No caso de alavancagem a volatilidade anual subiria para 14,21%.

# APÊNDICE C

Abaixo são apresentadas as sintaxes: (a) criação de uma rede neural empregada neste trabalho; (b) escolha da rede com o menor MSE (*mean squared error*) no treinamento para cada uma das 1000 simulações; (c) cálculo do RMSE\_R da amostra de teste; (d) saídas da rede no período de verificação. A plataforma utilizada foi o MATLAB® versão R2009b.

(a) criação da rede neural com randomização das amostras e dos pesos da rede neural

```
net = newff(inputs_Train,targets_Train,nhidden,{},'trainbr'); % inputs_Train são as
% amostras de treinamento, validação e teste previamente fornecidas; targets_Train
% as saídas desejadas correspondentes, nhidden o número de neurônios na camada
% escondida, { } as funções de ativação default (tansig na camada escondida e purelin
% na de saída), 'trainbr' o algoritmo Bayesian Regularization;
```

(b) escolha da rede neural com o menor MSE (*mean squared error*) no treinamento, entre 5 tentativas, para compor 1 das 1000 simulações

```
NTrain = 5; % laço com 5 iterações
```

```
Best_tperf = Inf;
```

```
for i = 1:NTrain
```

```
    [net,tr] = train(net,inputs_Train,targets_Train); % executa treinamento da
```

```
% rede criada; tr grava dados do treinamento com amostras randomizadas
    if tr.tperf(end) < Best_tperf % compara MSE (end) com o da rede com melhor
% performance anterior à iteração
        Best_Net = net; % escolhe a rede com menor MSE após a iteração
        Best_tr = tr; % armazena dados do treinamento e teste, cujas amostras
% foram randomizadas
    end
end
```

(c) cálculo do RMSE\_R da amostra de teste

```
ind = Best_tr.testInd; % índice que identifica as amostras randomizadas de teste
inputs_Test = inputs_Train(ind,:); % amostra randomizada de teste
targets_Test = targets_Train(ind); % saídas desejadas correspondentes à amostra ran-
% domizada de teste
outputs_Test = sim(Best_Net,inputs_Test); % saídas da rede referentes à amostra
% de teste
RMSE_test = rmse(targets_Test,outputs_Test); % RMSE é uma rotina para o cál-
% culo do RMSE_R, descrita a seguir:
```

```
function val = rmse(YObs,Yhat)
Error = abs((YObs-Yhat)./YObs);
val = sqrt(mean(Error.*Error));
```

(d) saídas da rede no período de verificação

```
outputs_verif = sim(Best_Net,inputs_Verif); % saídas da rede Best_Net, forneci-
% das as entradas do período de verificação (inputs_Verif)
```