



Universidade Federal Rural do Rio de Janeiro
Instituto Três Rios
Departamento de Ciências Econômicas e Exatas - DCEEX

RAFAEL VICENTE RIVETTI

**MERCADO DE CAPITAIS: CAPACIDADE DE PREVISÃO COM O USO DE
REDES NEURAIIS**

TRÊS RIOS

2016



Universidade Federal Rural do Rio de Janeiro
Instituto Três Rios
Departamento de Ciências Econômicas e Exatas - DCEEX

RAFAEL VICENTE RIVETTI

**MERCADO DE CAPITAIS: CAPACIDADE DE PREVISÃO COM O USO DE
REDES NEURAIIS**

Monografia de conclusão de curso apresentada ao curso de Ciências Econômicas do Instituto Três Rios da Universidade Federal Rural do Rio de Janeiro, como requisito parcial à obtenção do título de Bacharel em Ciências Econômicas.

Orientador: Prof. Dr. José Leonardo Ribeiro Macrini

TRÊS RIOS

2016

RAFAEL VICENTE RIVETTI

**MERCADO DE CAPITAIS: CAPACIDADE DE PREVISÃO COM O USO DE
REDES NEURAIAS**

Monografia de conclusão de curso submetida ao curso de Ciências Econômicas do Instituto Três Rios da Universidade Federal Rural do Rio de Janeiro, UFRRJ, como requisito parcial à obtenção do título de Bacharel em Ciências Econômicas e aprovada pela seguinte banca examinadora:

Prof. José Leonardo Ribeiro Macrini
Universidade Federal Rural do Rio de Janeiro

Prof. Cid de Oliva Botelho Junior
Universidade Federal Rural do Rio de Janeiro

Prof. Roberto Salvador Santolin
Universidade Federal Rural do Rio de Janeiro

TRÊS RIOS
2016

Dedico este trabalho à minha família, sem a qual não teria condições de chegar até aqui e me formar.

AGRADECIMENTOS

Agradeço a Deus, à minha família e a todos que de alguma forma contribuíram para o meu crescimento neste período de construção de conhecimento e descobertas.

Agradeço ao meu orientador por ter aceitado me orientar, por sempre se mostrar solícito quando precisei e por sua generosidade em compartilhar seu conhecimento.

Agradeço também à minha banca examinadora, por ter aceitado o convite e dispor de seu tempo para avaliar meu trabalho de monografia.

Embora ninguém possa voltar atrás e fazer um novo começo, qualquer um pode começar agora e fazer um novo fim.

Chico Xavier

RESUMO

Este trabalho tem como objetivo analisar a Teoria dos Mercados Eficientes, que defende não ser possível fazer qualquer tipo de previsão dentro do mercado de capitais, como por exemplo, no mercado de ações. Buscando questionar esta defesa, estimamos os valores futuros do IBOVESPA (Índice da Bolsa de Valores de São Paulo) através das técnicas de previsão com um modelo de Redes Neurais, visto se tratar de uma série temporal de movimentação não-linear, de modo que o método das redes neurais se faz adequado. Ou seja, avaliar-se-á a capacidade das Redes Neurais na previsão de valores futuros em um mercado de movimentação não-linear, contrapondo a teoria que defende não ser possível prever os movimentos futuros de índices ou preços de ações.

Palavras-Chave: Redes Neurais. Previsão. Ibovespa. Mercados Eficientes.

ABSTRACT

This work aims to analyze the theory of efficient market, which advocates not be possible to make any kind of prediction in the capital market, such as in the stock market. Seeking question this defense, we estimate future values of the IBOVESPA (Index of the São Paulo Stock Exchange) through forecasting techniques with a model of neural networks, as this is a time series of non-linear movement, so that the method of neural networks becomes appropriate. That is, to evaluate will the ability of neural networks in forecasting future values in a non-linear motion market, opposing the theory that argues not be possible to predict future movements in rates or stock prices.

Keywords: Neural Networks. Forecast. Ibovespa. Efficient markets.

SUMÁRIO

INTRODUÇÃO.....	11
CAPÍTULO 1 – O MERCADO DE CAPITAIS E A TEORIA DA IMPREVISÃO.....	12
1.1. O Mercado de Capitais	12
1.2. O Mercado de Ações	12
1.3. O Índice Bovespa	13
1.4. A Teoria do Caos	14
1.5. A Teoria dos Mercados Eficientes	15
1.5.1. Eficiência Fraca.....	16
1.5.2. Eficiência Semiforte.....	17
1.5.3. Eficiência Forte.....	17
1.6. Imprevisibilidade no Mercado de Ações.....	17
CAPÍTULO 2 – REDES NEURAIS.....	19
2.1. Rede Neural Biológica	19
2.2. Redes Neurais Artificiais - RNAs.....	20
2.2.1. O modelo de um neurônio.....	21
2.2.2. Funções de Ativação.....	23
2.2.2.1. Função Limiar	23
2.2.2.2. Função Linear	24
2.2.2.3. Função Linear por Partes.....	25
2.2.2.4. Função Sigmóide.....	25
2.2.2.5. Função Tangente Hiperbólica	27
2.2.2.6. Função Softmax	27
2.2.2.7. Função Gaussiana.....	28
2.2.3. Arquitetura das Redes Neurais	29
2.2.3.1. Redes Alimentadas Adiante com Camada Única.....	30
2.2.3.2. Redes Alimentadas Diretamente com Múltiplas Camadas.....	31
2.2.3.3. Redes Recorrentes	31
2.2.4. Aprendizado das Redes Neurais	32
2.2.4.1. Aprendizado Supervisionado	33
2.2.4.2. Aprendizado Não Supervisionado.....	34
2.2.4.3. Aprendizado Híbrido	35
2.2.4.4. O Algoritmo de Backpropagation	35
2.2.5. Exemplo Prático da Aritmética nas Redes Neurais	41

	10
2.2.6. Generalização	42
CAPÍTULO 3 – A PREVISÃO DO IBOVESPA	44
3.1. A Escolha do Ativo	44
3.2. Por que usar apenas a série temporal do Ibovespa e não outras variáveis?	44
3.3. Características do Experimento	45
3.4. Resultados.....	46
3.5. Validação do Modelo	49
CONCLUSÃO.....	51
REFERÊNCIAS BIBLIOGRÁFICAS	53

INTRODUÇÃO

A Teoria dos Mercados Eficientes sempre foi alvo de muitos debates, visto se tratar de um assunto polêmico ao defender a impossibilidade de previsão de, por exemplo, uma série de preços, tendo como base a afirmativa de que todas as informações são conhecidas pelos participantes do mercado, ou seja, a informação é perfeita, e que este conhecimento anularia a capacidade de previsão, onde não seria possível obter retornos extraordinários ao longo do tempo.

Os preços no mercado de capitais, como os das ações ou índices de ações, por não terem um movimento linear, produzem dificuldades no que se refere a previsões, visto que a maioria dos modelos existentes se prestam a realizar previsões de modo linear, o que não é adequado para esta finalidade.

Diante disto, o modelo de Rede Neural será utilizado de modo a buscar prever o comportamento futuro dos valores do Índice da Bolsa de Valores de São Paulo (IBOVESPA), pois é um modelo não linear sendo aplicado em uma série histórica não linear. Logo, a intenção é modelar através de Redes Neurais a predição de valores futuros do Ibovespa em um mercado de movimentação não linear, no sentido de que o mercado de ações é sabidamente não linear e, como a maioria das técnicas se prestam a modelos lineares, o uso das Redes Neurais se torna indispensável, visto se tratar de um conjunto de técnicas não lineares.

Você deve ter ouvido a proposição de que os preços das ações seguem um passeio aleatório. É um termo técnico, mas com uma interpretação simples. Algo – pode ser uma molécula ou o preço de um ativo – segue um passeio aleatório se cada passo que dá pode ser tanto para cima quanto para baixo. Portanto, seus movimentos são imprevisíveis. (BLANCHARD, 2007, p. 303)

Temos então dois lados, uma teoria que defende não ser possível realizar qualquer tipo de previsão de séries de preços, e uma teoria que defende que é possível realizar esta previsão, ainda que não de modo perfeitamente preciso.

Portanto, temos que um ponto importante é que o uso das Redes Neurais na previsão de séries temporais pode ser uma solução para reduzir a grande incerteza existente em relação ao movimento futuro dos preços e questionar a Teoria dos Mercados Eficientes, onde chegar-se-á a uma conclusão em face de um embate de duas teorias contraditórias.

CAPÍTULO 1 – O MERCADO DE CAPITAIS E A TEORIA DA IMPREVISÃO

1.1. O Mercado de Capitais

Primeiramente, é importante termos uma definição do que é o mercado de capitais, de modo que possamos entender o meio onde atuaremos. Diante disto, podemos defini-lo da seguinte forma:

Um conjunto de instituições e de instrumentos que negociam com títulos e valores mobiliários¹, objetivando a canalização dos recursos dos agentes compradores para os agentes vendedores. Ou seja, o mercado de capitais representa um sistema de distribuição de valores mobiliários que tem o propósito de viabilizar a capitalização das empresas e dar liquidez aos títulos emitidos por elas. (PINHEIRO, 2009, p. 174)

Este mercado é uma das subdivisões do Mercado Financeiro, em que está inserido o mercado de ações. Com isto, temos que engloba bolsas de valores, corretoras e diversas instituições financeiras, tendo a função de captar e canalizar poupança da sociedade para atividades econômicas das empresas.

1.2. O Mercado de Ações

A BM&FBOVESPA (2011, p. 2) define o mercado de ações como “um sistema criado para facilitar a capitalização das empresas, contribuindo para a geração de riqueza à sociedade”.

Mueller (1996) analisa que uma empresa em processo de expansão precisa se financiar para realizar seus investimentos, mas muitas vezes não é possível a partir de recursos próprios e, nesta situação, uma saída é recorrer ao mercado de ações, onde as empresas abrem seu capital e emitem ações no mercado, vendendo parte da empresa, e os investidores compram estas ações na espera de lucros futuros, tornando-se sócios.

Segundo a CVM (2014), ações são a menor fração do capital social de uma empresa, ou seja, é um título negociável que representa a menor parte de uma empresa. No Brasil, as ações podem ser Ordinárias Nominativas, que são aquelas que dão direito à voto nas assembleias das empresas, e Preferências Nominativas, onde seus detentores têm preferência na distribuição dos lucros.

¹ Valores Mobiliários são títulos de propriedade ou de crédito, com características e direitos padronizados, emitidos por agente público ou privado, como por exemplo, ações, debentures, bônus de subscrição, cupom cambial, cotas de fundos de investimentos, contratos derivativos, entre outros.

Estes títulos são negociados nos ambientes das bolsas de valores, e a BM&FBOVESPA (2016) classifica as bolsas de valores como sendo os locais que oferecem as condições para a realização de negócios de compra e venda de títulos e valores mobiliários, além de outros ativos. No caso do Brasil, temos a BM&FBOVESPA – A Nova Bolsa, fruto da fusão entre a Bovespa – Bolsa de Valores de São Paulo e a BM&F – Bolsa de Mercadorias e Futuros.

As ações são títulos de renda variável, isto é, a partir do momento em que as ações de uma empresa entram em negociação no mercado, seu preço pode tomar qualquer direção, e a cada dia de negócios oscilam mudando de preço, para cima ou para baixo, formando uma série histórica de valores, ou seja, uma série temporal, como podemos ver a seguir.

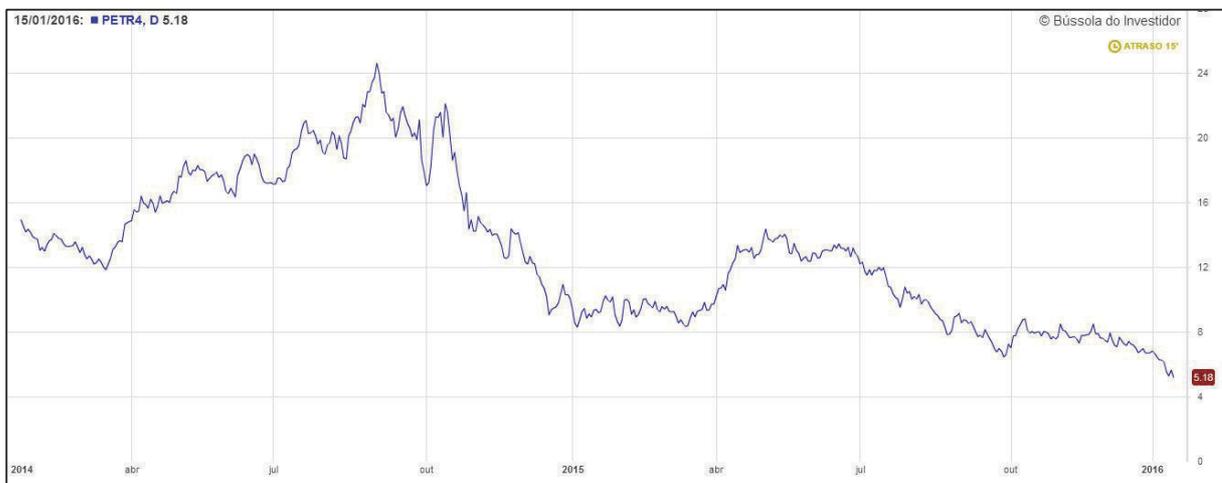


FIGURA 1.1 - gráfico diário das ações Petrobrás PN de 22/01/2014 a 15/01/2016.

Fonte: Bússola do Investidor (2015).

1.3. O Índice Bovespa

Um índice de ações da bolsa é indicador capaz de mostrar o desempenho de um conjunto de ações, fornecendo informações de performance ao longo do tempo de um determinado grupo de títulos conforme os preços destes varia, sendo um resultado comum do comportamento diferente de uma cesta de ações em determinado período.

De acordo com a BM&FBOVESPA (2015, p. 2), o Índice da Bolsa de Valores de São Paulo (IBOVESPA) tem como objetivo “ser o indicador do desempenho médio das cotações dos ativos de maior negociabilidade e representatividade do mercado de ações brasileiro”. Funciona como um termômetro do mercado de ações brasileiro, sendo considerado o mais importante indicador de performance média das cotações das ações de sua composição.

É um índice importante por utilizar critérios de inclusão como, por exemplo, ter presença em 95% dos pregões, elevada liquidez, ter participação no total financeiro negociado maior ou igual a 0,1%, entre outros. De modo a ser um índice representativo das principais ações do mercado, conta em sua composição com cerca das 60 maiores empresas do mercado brasileiro, cada uma em sua devida proporção de acordo com determinados critérios, ou seja, é uma média ponderada. Sua cotação é dada em pontos, onde em sua criação iniciou a partir de uma base e, conforme o tempo essa base foi sendo alterada de acordo com o desempenho médio das ações em sua composição. Veja um exemplo da movimentação do Ibovespa.

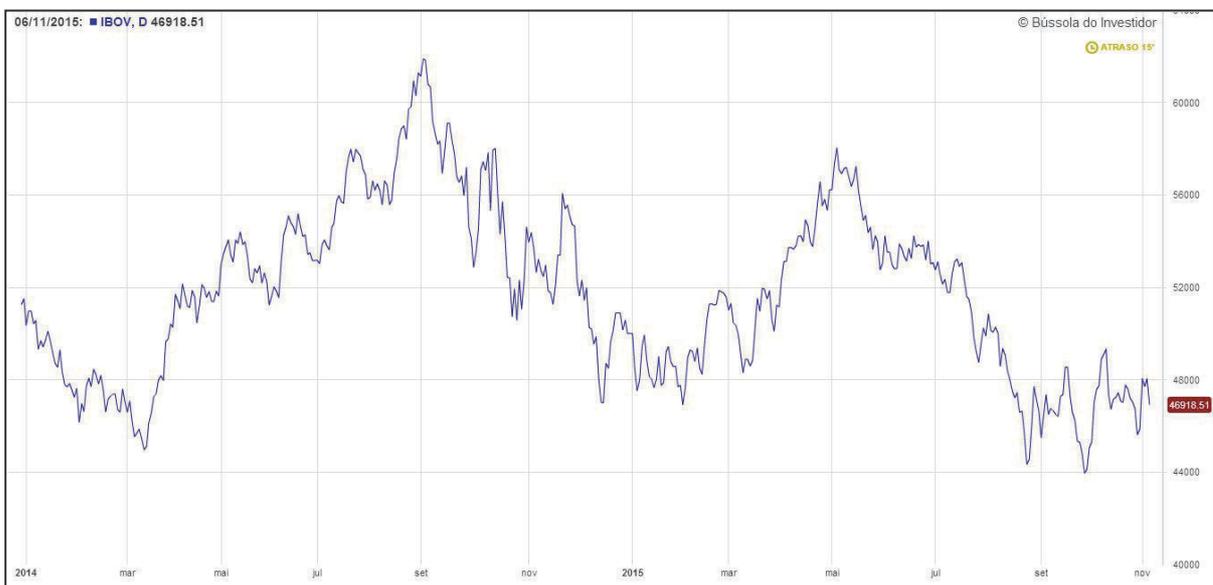


FIGURA 1.2 - gráfico diário do Ibovespa de 02/01/2014 a 07/11/2015.

Fonte: Bússola do Investidor (2015).

1.4. A Teoria do Caos

Ceretta (2003) refere-se a tal teoria como uma disciplina ainda em desenvolvimento, voltada para sistemas não-lineares complexos, podendo ser explicada a partir da análise destes três termos: sistemas, não-linearidade e complexidade.

Afirma que “um sistema é uma relação de interdependência e inter-relacionamento entre partes”, e dá como exemplo uma pilha de pedras, onde essas características podem ser vistas quando se remove uma pedra da base e há um desmoronamento, que teria como objetivo estabelecer um novo equilíbrio, logo, é um sistema dinâmico onde uma alteração leva a uma reorganização.

A não-linearidade é um arcabouço matemático que representa o comportamento do sistema na realidade. Um modelo linear representa uma relação de proporcionalidade que é constante entre variáveis, isto é, quando houver uma alteração em uma variável, isto causará uma alteração proporcional em outra, onde esta relação pode ser representada por uma linha reta. Diferentemente, a não-linearidade refere-se à falta de proporcionalidade constante, portanto, a mudança em uma variável produziria alterações de modo não proporcional em outra variável e, diante disto, a melhor forma de se representar esta relação não seria através de uma reta, mas sim de uma curva.

A complexidade está ligada ao problema de criar um modelo para prever o comportamento futuro de um sistema real. Temos como exemplo o fato de que é relativamente fácil prever o tempo gasto para se dirigir de uma cidade a outra, porém, não é nada trivial prever o comportamento de uma montanha de pedras desmoronando.

Temos então que o caos é um fenômeno que se tornaria imune a testes de previsibilidade, tendo um comportamento estocástico dentro de um sistema determinístico, pois na prática, sempre existirá o acaso, que no longo prazo se demonstra como uma distorção nos resultados.

Os preços das ações trabalhariam dentro desta perspectiva, uma vez que a partir do momento em que observamos determinadas cotações não seria possível determinar qual movimento tomariam nos tempos posteriores.

1.5. A Teoria dos Mercados Eficientes

A Hipótese dos Mercados Eficientes – HME, nas palavras de Fama (apud COSTA JR., 1990, p. 80): “é o mercado onde os preços refletem totalmente as informações disponíveis”, que coloca como postulado que todas as informações estão disponíveis e são incorporadas a todo momento, logo, não seria possível realizar uma previsão, e os preços das ações seguiriam um caminho aleatório. Samuelson (apud CERETTA, 2003, p. 1) diz que a HME é uma teoria de equilíbrio para o mercado, em que toda e qualquer informação é disponível e pública e, por este motivo, ninguém seria capaz de obter lucros extraordinários, visto que os preços dos ativos negociados sempre refletem plena e instantaneamente as informações disponíveis sobre os mesmos, portanto, seu preço de mercado seria sua melhor estimativa.

A HME tem como base a crença de que os investidores são racionais, e os que não são, tendo uma participação aleatória, teriam suas participações no mercado canceladas entre si, não produzindo qualquer efeito nos preços dos ativos negociados no mercado.

As hipóteses de sustentação da Teoria dos Mercados Eficientes seriam as seguintes:

a) a concorrência deve ser perfeita, ou seja, deve haver participantes em número suficiente nos mercados e nos ativos financeiros para impedir que uma decisão isolada afete os preços; b) as preferências dos investidores são estáveis, de forma que as expectativas racionais sejam formadas e, assim, se maximize as utilidades esperadas; c) as expectativas dos investidores são homogêneas, uma vez que se supõem os investidores racionais e com igual acesso às informações do mercado; d) como novas informações sobre os ativos financeiros surgem aleatoriamente, os ajustes nos portfólios dos investidores devem ser instantâneos; e) não há fricções, ou seja, os ativos são homogêneos, divisíveis e não envolvem custos de transação; e f) as informações disponíveis são processadas de maneira ótima pelos agentes. Porém, o autor pondera que sob esses pressupostos, prevalecem os resultados do modelo de passeio aleatório (*random walk*), no qual as mudanças de preços nos ativos financeiros são imprevisíveis e independem dos preços passados. (FAMA apud CERETTA, 2003)

Penteado (2003) coloca que, diante disto, não haveria possibilidade de previsão dos preços de um ativo a partir de preços passados e essa incapacidade de prever levou alguns autores a defenderem que os preços das ações poderiam subir ou cair em igual probabilidade, isto é, 50%, logo, os preços se distribuiriam de acordo com um *random walk*, caminho aleatório, e o preço futuro independeria do preço passado. Malkiel (apud PENTEADO, 2003, p. 8) afirma que um caminho aleatório é ‘aquele em que passos ou direções futuras não podem ser previstos com base em atos anteriores’. Aplicando o termo ao mercado de ações, temos que mudanças no curto prazo em seus preços não podem ser previstas, muito menos mudanças no longo prazo.

De acordo com Ikeda e Rabelo Junior (2004), podemos ter três tipos de eficiência de mercado, isto é, três formas: Eficiência Fraca, Eficiência Semiforte e Eficiência Forte. Isto ocorre porque as informações são separadas em subgrupos de acordo com a velocidade com que afetam os preços de ativos, tendo informações sobre preços históricos, informação pública e toda informação, e a incorporação de cada forma implica distintas formas de eficiência.

1.5.1. Eficiência Fraca

Ocorre quando os ativos absorvem completamente a informação passada dos preços e, diante disto, investidores não conseguiriam auferir lucros extraordinários com estratégias baseadas em preços históricos, não sendo possível estabelecer padrões e, com a informação de preços passados sendo de fácil obtenção, qualquer participante do mercado teria a capacidade de obter lucro e esta competição, no limite, faria a possibilidade de lucro desaparecer.

1.5.2. Eficiência Semiforte

Ocorre quando os preços dos ativos representam toda informação pública disponível, como demonstrações contábeis, demonstrativos, informações históricas, etc. Nessa forma de eficiência, nenhum indivíduo poderia obter lucros expressivos com base em informações públicas.

Em comparação com a forma fraca, a semiforte necessita que aqueles que investem, além de terem acesso a preços históricos, precisariam ter maior embasamento teórico para analisar com maior consistência as empresas e a economia ou qualquer objeto de investimento, o que implica maior dispêndio de tempo e custos, o que poderia limitar atuações.

1.5.3. Eficiência Forte

Ocorre quando engloba as informações públicas, a de preços passados e as informações não disponíveis, portanto, nenhum investidor teria possibilidade de obter lucros anormais usando qualquer informação, seja ela disponível ou não, como informações confidenciais por exemplo, pois qualquer informação relacionada a uma ação estará incorporada, onde mesmo os maiores investidores não teriam condições de transformar as informações em ferramenta de obtenção de lucros acima da média. Esta é a mais restritiva das hipóteses.

1.6. Imprevisibilidade no Mercado de Ações

O modelo básico de imprevisão no mercado de ações é o caminho aleatório citado anteriormente. Ceretta (2003) coloca que o chamado movimento browniano foi descrito pelo botânico Robert Brown ao constatar a forma como o pólen das plantas se dispersava sobre a água na forma de grande número de partículas sem tendenciosidade, podendo seguir qualquer caminho, sendo então um caminho aleatório.

Explica que o estatístico Maurice Kendall estudou o comportamento dos preços das ações e de índices da bolsa de valores de Londres e concluiu que seus movimentos seguiam este movimento browniano, isto é, um caminho aleatório, e que nenhuma informação a respeito de preços, sazonalidades, etc. ajudaria a prever o movimento futuro.

Diante disto, temos que a teoria do caos, a hipótese dos mercados eficientes e as observações empíricas de que o mercado de ações se movimenta de acordo com um movimento browniano, totalmente aleatório e impossível de prever, formam o arcabouço teórico que

defende a imprevisibilidade dentro deste mercado, afirmando que nenhum investidor seria capaz de realizar uma previsão consistente a respeito de preços futuros de modo a se beneficiar com lucros, independentemente das informações que possuir.

CAPÍTULO 2 – REDES NEURAIS

2.1. Rede Neural Biológica

Haykin (2001) explica que o cérebro humano é como um computador, pois tem a capacidade de processar informações de modo complexo e não-linear. Sua estrutura é formada por células denominadas neurônios, que são responsáveis por realizar processamentos, mas de modo muito mais rápido que um computador.

Wasserman (apud MUELLER, 1996) afirma que o sistema nervoso humano, que é composto por uma rede de neurônios, constitui a unidade básica do cérebro, responsável pelos fenômenos conhecidos como pensamento, emoção e cognição, além da execução das funções sensorio-motoras e autônomas. Cada neurônio tem a capacidade de receber, processar e enviar sinais eletroquímicos ao longo das fibras nervosas no sistema de comunicação cerebral.

Cada neurônio é conectado a vários outros por meio de estruturas chamadas dendritos e axônios, formando uma espécie de teia, ou rede neural. Beale (apud MUELLER, 1996) descreve os dendritos como sendo “uma complexa rede de prolongamentos – recebem impulsos nervosos de outros neurônios e os conduzem ao corpo celular ou núcleo”, onde há o somatório das informações, criando novos impulsos. Os axônios são responsáveis pela transmissão dos sinais para os próximos neurônios, e esta conexão é feita através das sinapses, que são zonas ativas de contato entre as terminações nervosas. Maren (apud RODRIGUES, 2009) argumenta que essas sinapses que formam as conexões entre os neurônios constituem a saída do mesmo por meio do axônio e os dendritos, e este espaço sináptico gera uma resposta através de impulsos elétricos, logo, há uma propagação e uma retro propagação de um neurônio para outro através de substâncias chamadas neurotransmissores.

É a força sináptica da conexão neural que capacita o cérebro a armazenar conhecimento e torná-lo disponível para uso, ou seja, aprendizado. Um exemplo desta capacidade é o resultado que um indivíduo tem ao estudar, pois ao ler sobre determinado assunto absorve aquele conhecimento e o armazena, tornando-o disponível para uso quando necessário, como por exemplo, uma prova.

A seguir, temos uma representação visual de um neurônio biológico contendo as informações descritas a respeito do mesmo.

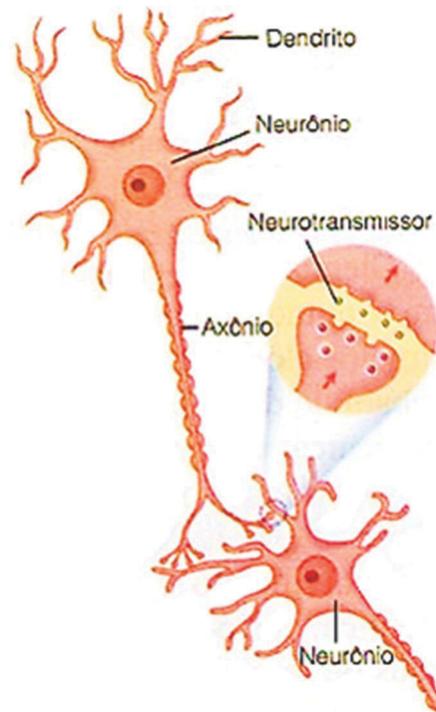


FIGURA 2.1 – Modelo de neurônio biológico.

Fonte: Google imagens.

2.2. Redes Neurais Artificiais - RNAs

As Redes Neurais Artificiais são um modelo de inteligência artificial, e Cavalheiro (2011) explica que esta é produzida pelo homem para prover às máquinas habilidades que simulam a inteligência humana. As RNAs buscam simular, por meio de máquinas como um computador, o funcionamento do cérebro biológico humano, e o fato de ter a competência de tratar dados ruidosos, incompletos ou sem precisão, faz com que sua capacidade de predição a dados não-lineares seja adequada, inclusive no mercado de capitais.

Haykin (2001, p. 28), define que “uma rede neural é um processador [...] que tem a propensão natural para armazenar conhecimento [...] e torná-lo disponível para o uso”. Desta forma, através do algoritmo de aprendizagem, a rede obtém o conhecimento através da série temporal escolhida e realiza a previsão com base neste conhecimento, seja linear ou não-linear. Desta forma, a rede neural se parece com cérebro humano em dois aspectos: (1) o conhecimento é adquirido pela rede a partir de seu ambiente de um processo de aprendizagem; (2) forças de conexão entre neurônios (pesos sinápticos) são utilizadas para armazenar o conhecimento adquirido e torná-lo disponível para uso.

A capacidade de aprender através de exemplos e de generalizar a informação aprendida é a capacidade que a rede neural tem em relação a qualquer outro método. Este é um modelo distribuído composto por unidades chamadas de “neurônios”, mas que ao invés de biológicos, são constituídos de funções matemáticas não-lineares (tipicamente sigmóides e tangentes hiperbólicas). A combinação destas unidades, através de parâmetros estimados a partir dos dados, é o que confere a capacidade deste modelo de inferir relações não-lineares de complexidade arbitrária. Ou seja, a RNA busca reproduzir o comportamento funcional das redes neurais biológicas, em unidades computacionais simples que são uma representação matemática de um neurônio.

2.2.1. O modelo de um neurônio

Haykin (2001) refere-se a um neurônio como sendo “uma unidade de processamento de informação que é fundamental para a operação neural”. É composto por, além dos sinais de entrada, um conjunto de pesos sinápticos, função de ativação e bias (viés).

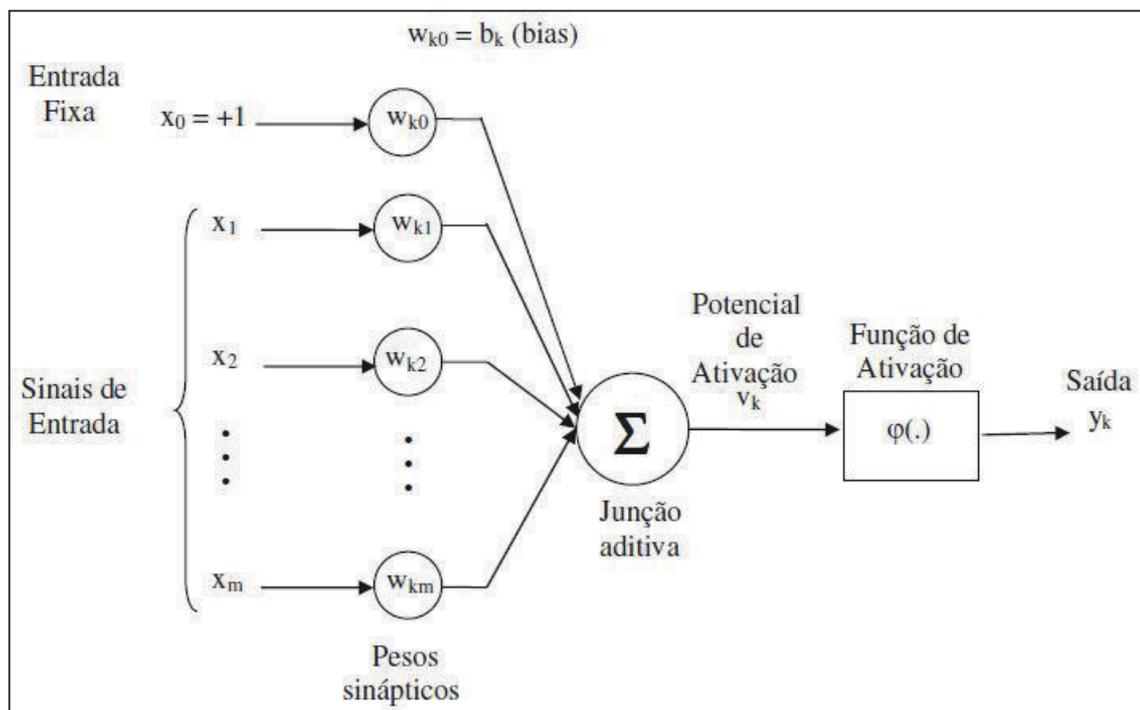


FIGURA 2.2 – Modelo de um neurônio artificial.

Fonte: Haykin (2001).

Diante disto, temos algumas características do neurônio que precisam ser descritas. Observamos os valores de entrada do neurônio, dados pela variável X . Também temos um

conjunto de sinapses ou elos de conexão, cada uma tendo um peso ou força própria, que são os pesos descritos pela variável W . Haykin (2001) descreve que “um sinal x_j na entrada da sinapse j conectada ao neurônio k é multiplicada pelo peso sináptico w_{kj} ”, em que o índice k refere-se ao neurônio em questão e o índice j ao ponto de entrada da sinapse a qual o peso se refere. Diferente da sinapse cerebral, o peso sináptico do neurônio artificial pode assumir valores positivos como negativos.

Temos uma função de ativação, que é uma fórmula matemática que recebe o somatório ponderado do produto entre as entradas com os pesos sinápticos, que tem efeito de inibição ou de excitação sobre os valores de entrada, gerando uma saída. A função de ativação $\varphi(\cdot)$ define a saída do neurônio em função do nível de atividade na sua entrada. É ela quem restringe a amplitude de saída de um neurônio.

Observamos também um *bias* (viés), demonstrado por b_k , tendo o efeito de aumentar ou reduzir a entrada líquida da função de ativação, visto que pode ser positivo ou negativo. O *bias* é um neurônio especial que tem a função de expandir os graus de liberdade, gerando melhor adaptação da rede em relação ao conhecimento fornecido a ela, representa uma sinapse de peso fixo $w_{k0} = b_k$, conectado a uma entrada fixa de valor +1. É um parâmetro externo.

Portanto, temos a seguinte estrutura de funcionamento: Um neurônio artificial recebe uma série de sinais de entrada, X . Cada sinal é ponderado pelo correspondente peso sináptico, isto é, o sinal é multiplicado pelo peso sináptico, onde o índice k, j se refere a ligação do sinal j para o neurônio k . O neurônio artificial processa a soma ponderada pelos pesos sinápticos e as entradas, adiciona o termo de *bias*, e estes são levados à função de ativação, resultando no sinal de saída do neurônio k , y_k .

A representação matemática de um neurônio é dada por:

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

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

Onde u_k é o combinador linear obtido pelo somatório do produto entre as entradas e os pesos sinápticos, e, de acordo com a FIGURA anterior, temos que

$$v_k = u_k + b_k \quad (2.3)$$

$$y_k = \varphi(v_k) \quad (2.4)$$

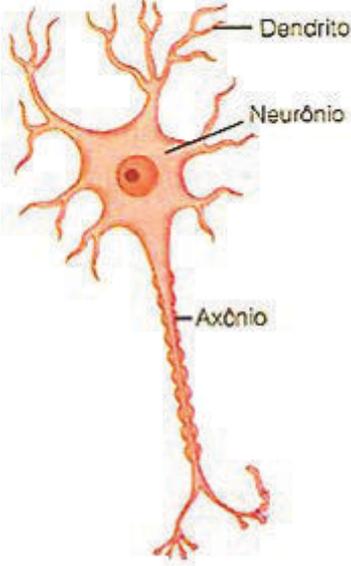
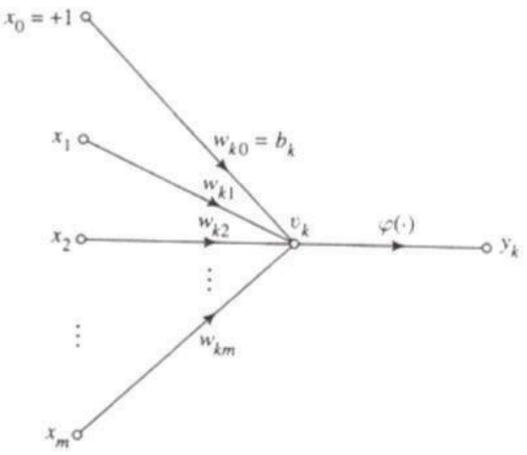
Parâmetros	Neurônio Biológico	Neurônio Artificial
Modelo		
Entrada	Dentritos	$x_0, x_1, x_2, x_3, \dots, x_m$
Saída	Axônio	y_k
Pesos	Comportamento das sinapses	$w_{k0}, w_{k1}, w_{k2}, w_{k3}, \dots, w_{km}$
Efeito de uma sinapse	Neurônio pós-sináptico	$u_k = \sum_{j=1}^m w_{kj} x_j$

TABELA 2.1 – Comparação entre o neurônio biológico e o artificial.

Fonte: O autor.

2.2.2. Funções de Ativação

As funções de ativação representam a saída de um neurônio em termos do campo local induzido v_k , sendo utilizadas para saídas lineares, não lineares, problemas de classificação, etc. Algumas das mais utilizadas, são:

2.2.2.1. Função Limiar

Também chamada de *função de Heaviside* ou *função degrau*, tem como característica ser positiva e direta, assumindo a propriedade de ser *tudo-ou-nada*, podendo ser descrita da seguinte forma:

$$\varphi(v_k) = \begin{cases} 1, & \text{se } v_k \geq 0 \\ 0, & \text{se } v_k < 0 \end{cases} \quad (2.5)$$

Veja a representação gráfica:

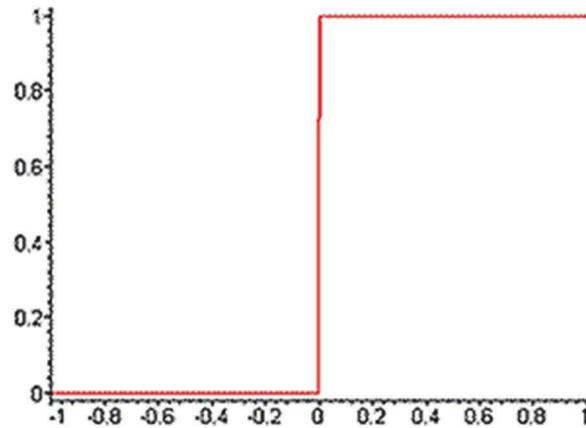


FIGURA 2.3 – Gráfico da função limiar.

Como corresponde à saída do neurônio k , que utiliza a função limiar, essa saída por ser expressa da seguinte forma:

$$y_k = \begin{cases} 1, & \text{se } v_k \geq 0 \\ 0, & \text{se } v_k < 0 \end{cases} \quad (2.6)$$

2.2.2.2. Função Linear

A função linear se caracteriza por uma função que segue a seguinte lei:

$$y_k = \alpha \cdot v_k \quad (2.7)$$

Em que y_k é a saída da função, com $\alpha \cdot v_k$ sendo o coeficiente angular, onde v_k é a variável independente e α o valor do coeficiente. Graficamente, fica da seguinte forma:

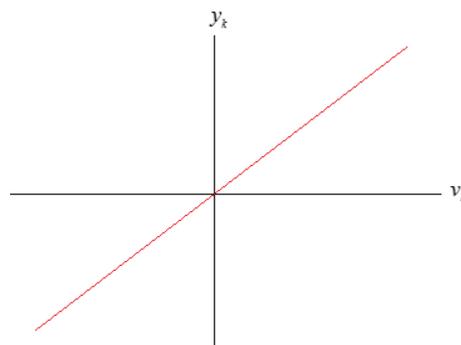


FIGURA 2.4 – Gráfico da função linear.

2.2.2.3. Função Linear por Partes

Decorre do fato de podermos restringi-la para produzir valores constantes em uma determinada faixa. Haykin (2001) coloca que caso a região linear não entre em saturação, surge um combinador linear; e se o fator de amplificação é infinitamente grande, a função linear se reduz à *função de limiar*. Matematicamente, temos:

$$\varphi(v) = \begin{cases} 1, & \text{se } v \geq +\frac{1}{2} \\ v, & \text{se } -\frac{1}{2} < v < +\frac{1}{2} \\ 0, & \text{se } v \leq -\frac{1}{2} \end{cases} \quad (2.8)$$

Graficamente:

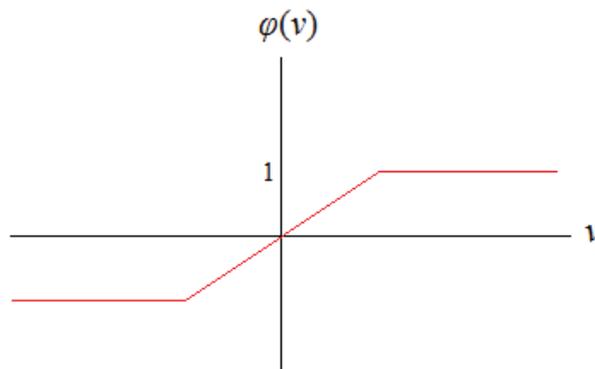


FIGURA 2.5 – Gráfico da função linear por partes.

2.2.2.4. Função Sigmóide

É uma das funções de ativação mais usadas em redes neurais. Possui a forma de S, variando numa escala entre 0 e 1, sendo definida como uma função estritamente crescente, exibindo balanceamento adequado entre comportamento linear e não-linear. O principal exemplo de função sigmóide é a *função logística*:

$$\varphi(v_k) = \frac{1}{1 + \exp(-\alpha \cdot v_k)} \quad (2.9)$$

Onde α é a inclinação da função, sendo $\alpha > 0$.

Graficamente, temos:

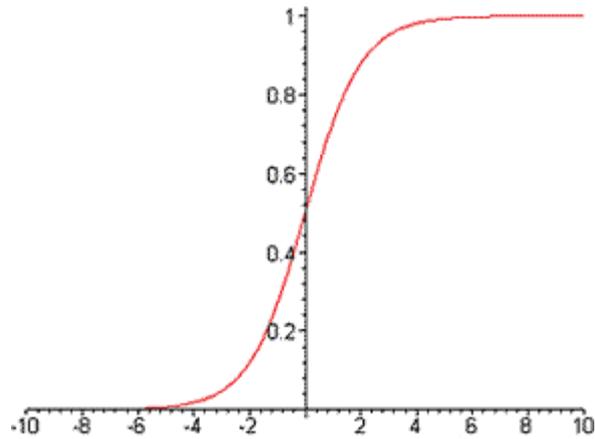


FIGURA 2.6 – Gráfico da função sigmóide.

Se α variar, teremos sigmóides com diferentes inclinações, sendo que sua inclinação na origem é dada por $\alpha/4$. No limite, se a inclinação tende ao infinito, a sigmóide se torna a função limiar, mas ao invés de assumir o valor 0 ou 1, assume um intervalo contínuo entre 0 e 1. Outra característica é que esta é uma função diferenciável, enquanto a limiar não é.

$$\varphi(v_k) = \frac{e^{\alpha \cdot v_k}}{1 + e^{\alpha \cdot v_k}} = \frac{1}{1 + e^{-\alpha \cdot v_k}} \quad (2.10)$$

$$\frac{\partial \varphi}{\partial v_k} = \alpha \cdot v_k(1 - v_k) > 0 \quad (2.11)$$

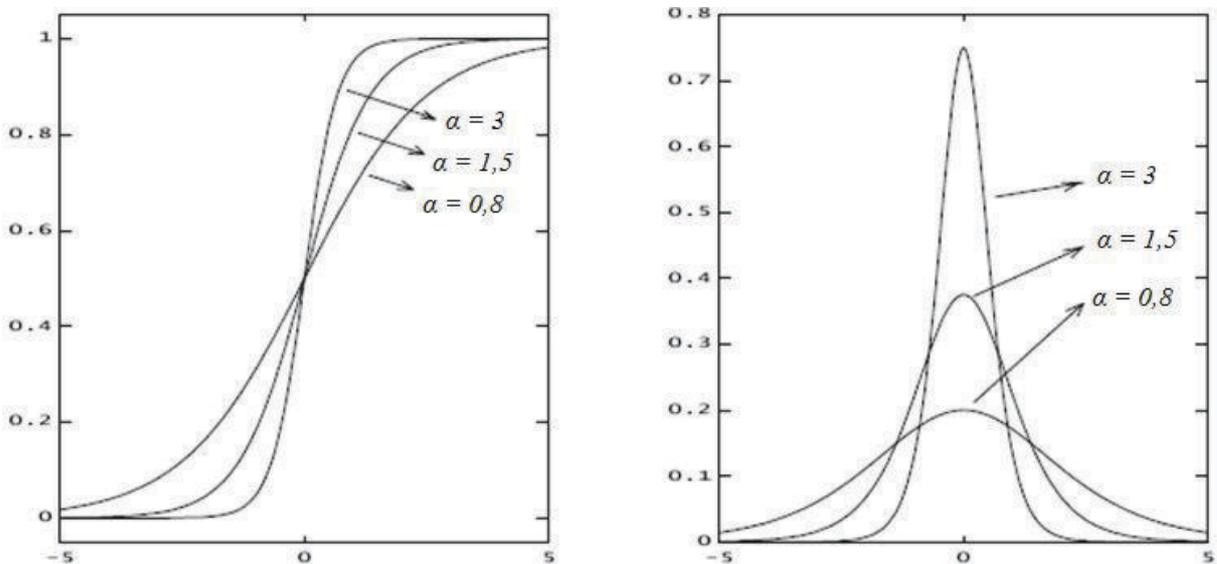


FIGURA 2.7 – Gráficos da função sigmóide com diferentes inclinações e sua diferenciação.

2.2.2.5. Função Tangente Hiperbólica

As funções de ativação limiar, linear por partes e sigmóide nos dão um intervalo entre 0 e 1, porém, em certos momentos é necessário que a função de ativação se estenda para um intervalo entre -1 e 1, assumindo uma forma antissimétrica em relação à origem, sendo uma função ímpar do campo local induzido. Com isso, a função tangente hiperbólica se faz adequada, onde a capacidade de trazer valores negativos traz benefícios analíticos. Matematicamente, se apresenta da seguinte forma:

$$\varphi(v_k) = \tanh(\alpha \cdot v_k) = \frac{e^{\alpha \cdot v_k} - e^{-\alpha \cdot v_k}}{e^{\alpha \cdot v_k} + e^{-\alpha \cdot v_k}} = \frac{1 - e^{-2\alpha \cdot v_k}}{1 + e^{-2\alpha \cdot v_k}} \quad (2.12)$$

$$\frac{\partial \varphi}{\partial v_k} = \alpha \cdot v_k (1 - v_k^2) > 0 \quad (2.13)$$

Graficamente, temos:

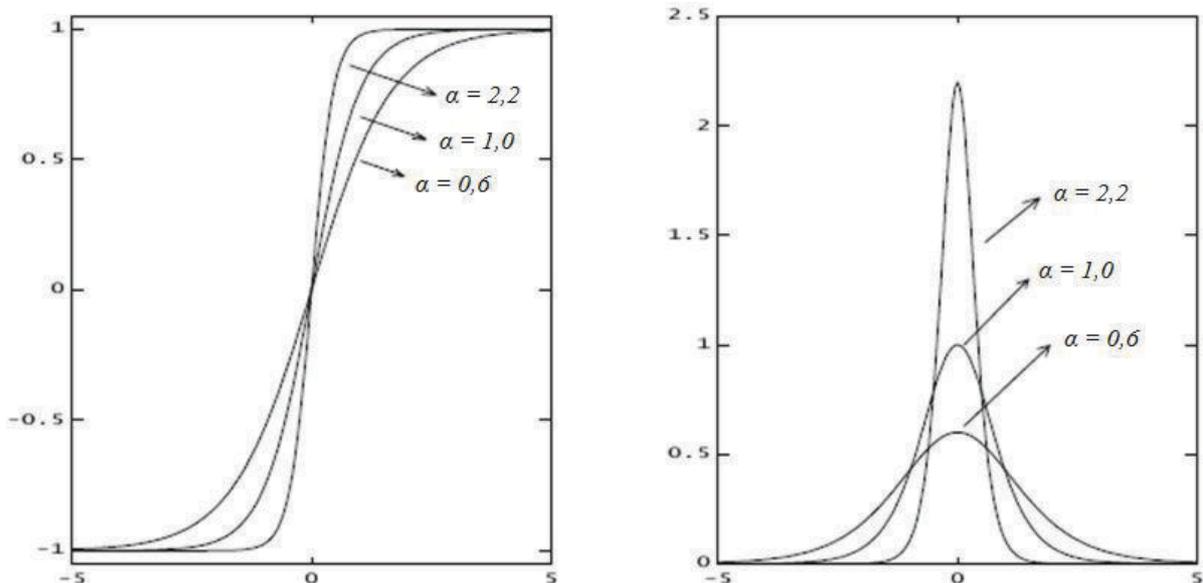


FIGURA 2.8 – Gráficos da função tangente hiperbólica com diferentes inclinações e sua diferenciação.

2.2.2.6. Função Softmax

A função Softmax é utilizada em classificações que envolvem várias classes diferentes, por exemplo, rating de crédito. Matematicamente e graficamente se apresenta como:

$$y_k = \varphi(v_k) = \frac{e^{v_k}}{\sum_k e^{v_k}} \quad (2.14)$$

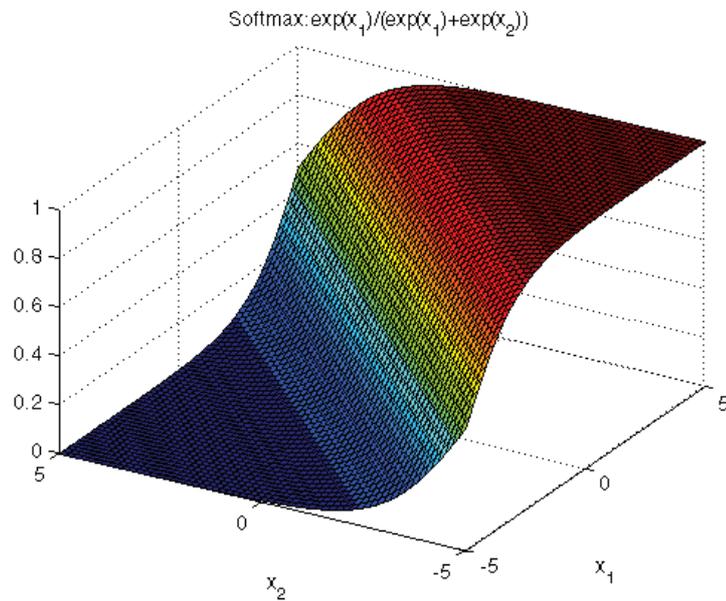


FIGURA 2.9 – Gráfico da função softmax.

2.2.2.7. Função Gaussiana

Esta é a famosa função da distribuição Normal, onde os dados se distribuem em torno de uma média, possuindo algumas características, como: as variáveis podem variar em um intervalo de $-\infty$ a $+\infty$, tendo como parâmetros uma média μ e um desvio padrão σ . Quanto mais próximos da média, mais prováveis são os valores, quanto mais distantes, mais improváveis.

Sua função é escrita da seguinte forma:

$$\varphi(v_k) = \frac{e^{-\frac{1}{2}\left(\frac{v_k - \mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}} \quad (2.15)$$

Essa função distribui seus valores de forma simétrica a partir da média, e seu gráfico forma uma curva chamada “curva de sino”, como pode ser visto no gráfico a seguir.

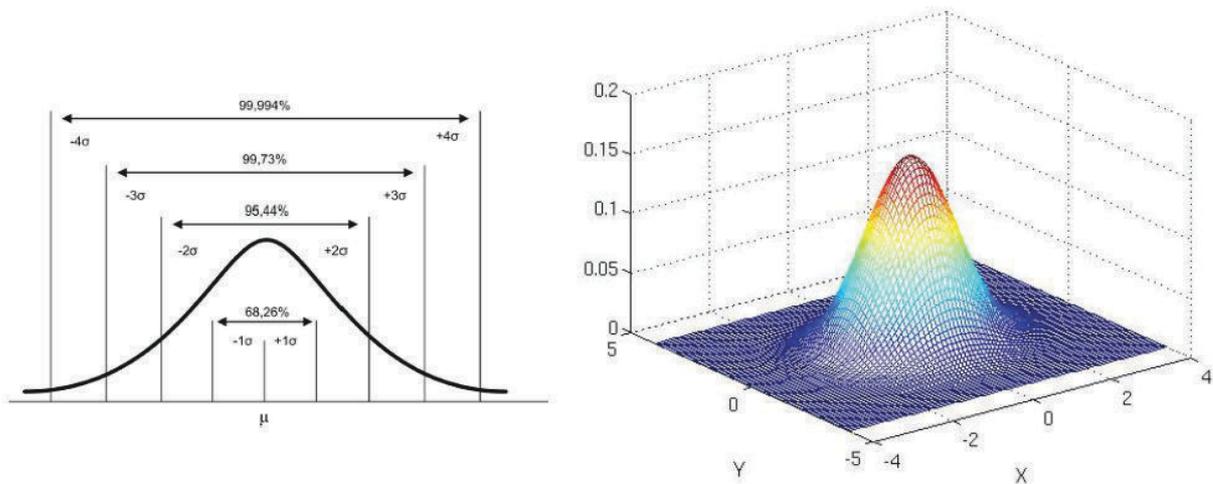


FIGURA 2.10 – Gráficos da função gaussiana unidimensional e bidimensional.

2.2.3. Arquitetura das Redes Neurais

Diz respeito ao modo como os neurônios estão organizados na estrutura da rede, o que está ligado ao tipo de objetivo que se deseja com a rede e ao algoritmo de aprendizado. Rodrigues (2009) coloca que as redes neurais artificiais podem se configurar em diversas arquiteturas, dividindo-se basicamente em 2 categorias:

- *Feedforward*, que são redes alimentadas adiante, com propagação direta, que pode ser de uma camada ou múltiplas camadas, sendo que o resultado da saída de um neurônio não pode ser usado como entrada de um neurônio anterior. Nesse tipo de rede, os sinais se propagam em apenas uma direção a partir da entrada, passando pelas camadas de neurônios até a saída, os pesos vão sendo ajustados em um processo itinerante de ida (propagação) e volta (retro propagação) com o objetivo de reduzir ao máximo o erro.
- *Feedbackward*, que são as Redes Recorrentes (ou Redes Realimentadas), em que o laço ocorre entre as conexões de realimentação, possuindo pelo menos uma interconexão realimentando a saída de neurônios para outros neurônios da rede (conexão cíclica). Haykin (2001) afirma que este tipo de rede se distingue da anterior por ter, pelo menos, um laço de realimentação, podendo ter apenas uma camada de neurônios, onde estes alimentam seu sinal de saída de volta para as entradas dos outros neurônios.

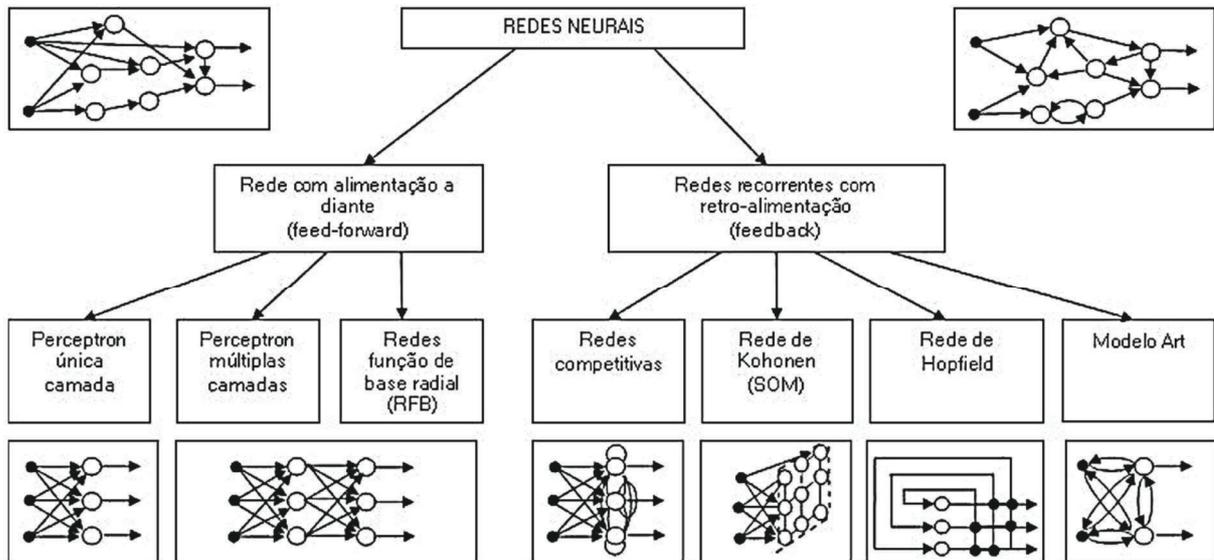


FIGURA 2.11 – Uma taxonomia das RNA (JAIN et al., 1995 apud RODRIGUES, 2009)

Uma rede neural possui uma camada escondida (também chamada de camada oculta ou intermediária) quando a saída de um neurônio não conduz a uma resposta final, mas sim para outros neurônios antes de gerar a saída final. A rede é organizada em camadas.

2.2.3.1. Redes Alimentadas Adiante com Camada Única

Haykin (2001) coloca que esta é uma forma simples, onde temos uma camada de entrada de valores que são projetados para a camada de saída, gerando uma resposta. Veja uma representação de rede neural de camada única a seguir:

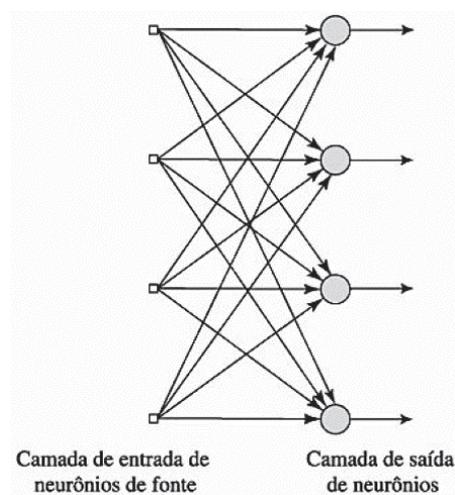


FIGURA 2.12 – Rede alimentada adiante com camada única (HAYKIN, 2001).

2.2.3.2. Redes Alimentadas Diretamente com Múltiplas Camadas

Caracteriza-se por uma rede que possua uma ou mais camadas ocultas, possuindo neurônios ocultos, cuja função é realizar um processamento entre a entrada e a saída, sendo que a saída de um neurônio oculto é a entrada para os neurônios da camada seguinte, isto é, os sinais da segunda camada (primeira camada oculta) são usados como entrada para a terceira camada, e assim sucessivamente até a saída da rede com a resposta global final.

Quanto mais camadas ocultas a rede tiver, maior sua capacidade de extrair estatísticas de ordem elevada, mas uma rede com muitos neurônios não significa que seja melhor do que uma rede com poucos neurônios, pois tudo depende do objetivo e da tarefa que a rede precisa executar. Geralmente quando se tem uma camada de entrada muito grande, mais neurônios são adicionados à rede.

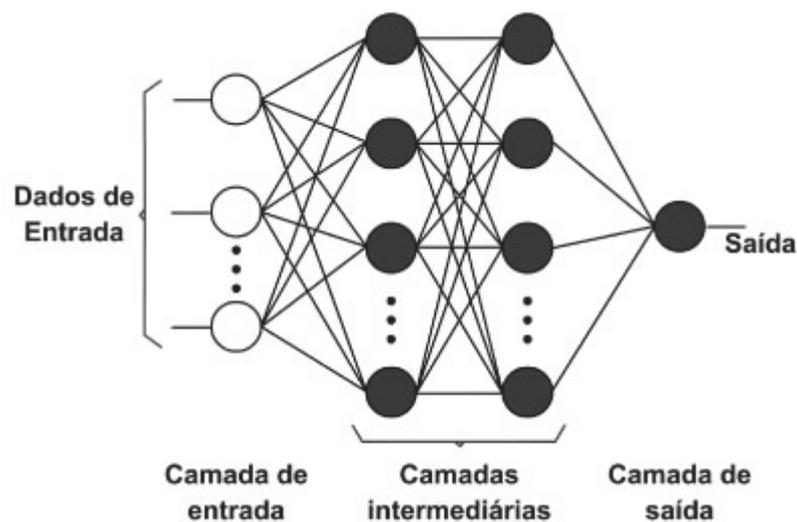


FIGURA 2.13 – Rede neural alimentada adiante com 2 camadas ocultas.

Geralmente as redes neurais do tipo *feedforward* são utilizadas para problemas de classificação e para previsão.

2.2.3.3. Redes Recorrentes

Como mencionado anteriormente, a rede recorrente (*feedback*) difere da rede *feedforward* por realizar ao menos um laço de realimentação, isto é, a saída de um neurônio tem a capacidade de se tornar a entrada de um neurônio anterior.

Observe a representação de um tipo de rede *feedback*:

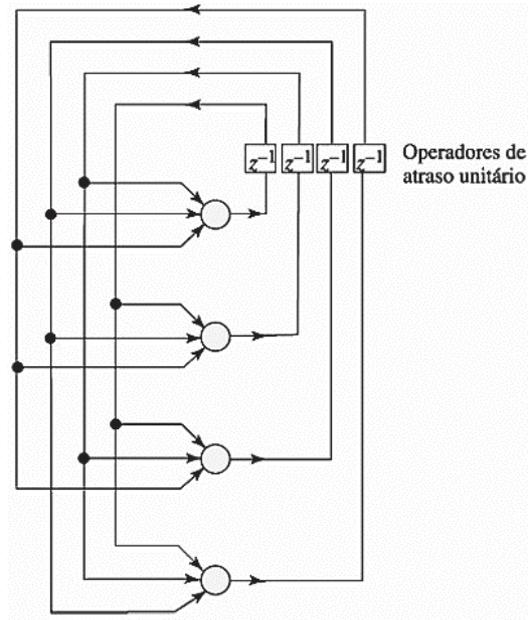


FIGURA 2.14 – Rede recorrente sem laços de auto-realimentação e sem neurônios ocultos (HAYKIN, 2001).

A FIGURA 2.14 mostra uma rede com apenas uma camada neuronal, e cada neurônio alimenta seu sinal de saída de volta à entrada de cada neurônio, não ocorrendo auto-realimentação, pois esse seria o caso se a saída do neurônio realimentasse sua própria entrada.

2.2.4. Aprendizado das Redes Neurais

Na fase de aprendizado (ou treinamento) da rede, a mesma recebe os dados, isto é, uma série de exemplos, podendo a rede fazer interpolações e extrapolações com aquilo que aprendeu. Este processo de aprendizagem ocorre pelo treinamento da rede por meio de iterações e apresentações de exemplos ou padrões que servem como parâmetros para a atualização e ajuste dos pesos na rede, que ao final do processo guardam o conhecimento adquirido e o torna disponível para uso. O conhecimento é adquirido através desse processo de aprendizado e a informação é armazenada em “densidades de conexão” conhecidas como pesos sinápticos ou simplesmente pesos.

Para isto, a rede utiliza algoritmos de aprendizagem. Macrini (2000) define um algoritmo de aprendizado como “um conjunto de procedimentos bem definidos para adaptar os parâmetros de uma rede neural para que a mesma possa aprender uma determinada função”, sendo que diferem apenas na maneira como o ajuste dos pesos é feito.

Haykin (2001) diz que para entender este processo de aprendizagem da rede é necessário conhecer o paradigma de aprendizado, de modo a se saber qual informação da rede está

disponível; e deve-se conhecer o algoritmo de treinamento. Há três paradigmas de aprendizagem: supervisionado, não supervisionado e híbrido.

O processo de aprendizagem implica na seguinte sequência de eventos: 1) a rede neural é estimulada pelo ambiente através de exemplos extraídos do mesmo. 2) a rede neural sofre modificações nos seus parâmetros livres como resultado desta estimulação pelo processo iterativo. 3) a rede neural responde de uma maneira nova ao ambiente, devido as modificações ocorridas na sua estrutura interna no passo 2.

2.2.4.1. Aprendizado Supervisionado

Também conhecido como “aprendizado com professor”, é o método em que a entrada e a saída desejadas são fornecidas para a rede. Haykin (2001) coloca que o objetivo é ajustar os parâmetros da rede com o objetivo de encontrar uma conexão entre os pares de entrada e saída fornecidos. Cada entrada é comparada à saída calculada e os pesos são ajustados de modo a que o erro (diferença entre a entrada e a saída) seja o menor possível. O algoritmo mais utilizado é conhecido como Retropropagação do Erro (*Backpropagation*).

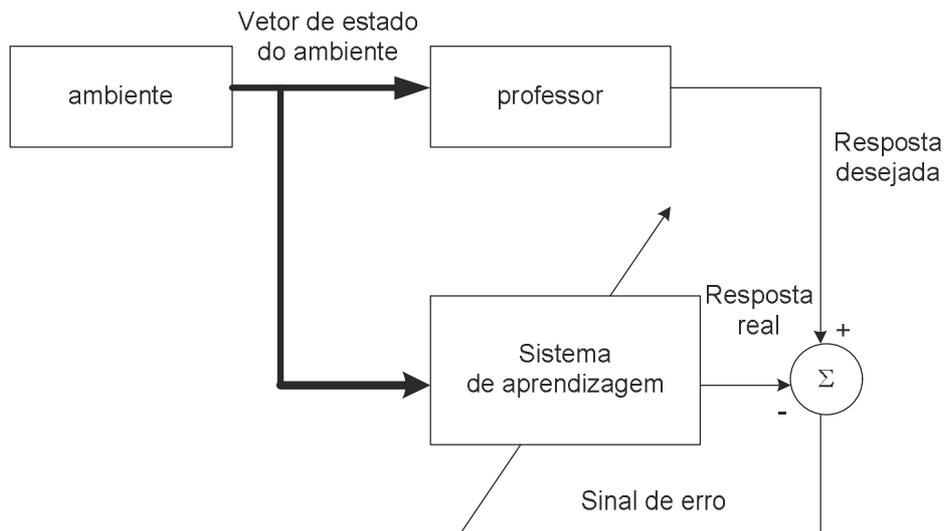


FIGURA 2.17 – Diagrama em blocos da aprendizagem com um professor (HAYKIN, 2001).

Consiste em que o professor tenha o conhecimento do ambiente, e fornece o conjunto de exemplos de entrada-resposta desejada. Com esse conjunto, o treinamento é feito usando a regra de aprendizagem por correção de erro.

Escolhe-se um conjunto de pares (entrada-saída desejada) conhecidos que representem o ambiente (conjunto de treinamento). Inicializa-se os parâmetros independentes da rede com valores arbitrários. Alimenta-se a rede com as entradas escolhidas. Calcula-se o erro na saída da rede com relação a saída desejada. Modifica-se os parâmetros independentes de acordo com critério baseado no erro de saída. À medida em que o erro de saída é analisado, os pesos são ajustados de forma a diminuir a diferença entre a saída encontrada em relação a saída desejada, até encontrar um erro de saída bastante aceitável.

2.2.4.2. Aprendizado Não Supervisionado

Neste tipo de aprendizagem, também chamada de *auto-organizada*, não se utilizam exemplos previamente conhecidos, ou seja, não há um supervisor para acompanhar o processo de aprendizagem, onde somente as entradas estão disponíveis. Uma medida da qualidade da representação do ambiente pela rede é estabelecida, e os parâmetros independentes são modificados de modo a otimizar esta medida.

Para implementar um esquema de aprendizado não supervisionado pode-se usar uma regra de competição. Uma primeira camada recebe os dados enquanto as unidades de uma segunda camada disputam o privilégio de responder aos estímulos. Em outras palavras, uma rede neural típica com aprendizado competitivo é uma rede de uma única camada (uni ou bidimensional) em que todos os neurônios recebem a mesma entrada. Cada neurônio computa o seu nível de ativação multiplicando o seu vetor de pesos pelo vetor de entrada, da maneira usual. O neurônio que tiver o maior nível de ativação é chamado de vencedor e apenas ele terá atividade diferente de zero na saída da rede, ou seja, o padrão de entrada que estiver sendo apresentado à rede provocará a ativação de apenas um neurônio da rede neural.

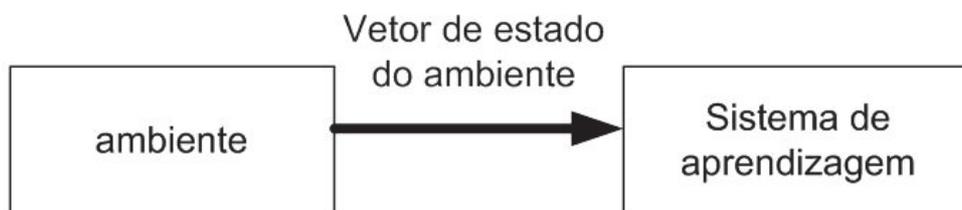


FIGURA 2.18 – Diagrama em blocos da aprendizagem não supervisionada (HAYKIN, 2001).

2.2.4.3. Aprendizado Híbrido

O aprendizado híbrido tem como característica a combinação do aprendizado supervisionado e não supervisionado, onde uma parte dos pesos sinápticos é determinada pelo aprendizado supervisionado, enquanto a outra parte é alcançada pelo aprendizado não supervisionado.

2.2.4.4. O Algoritmo de Backpropagation

De acordo com Macrini (2000), é um modelo de aprendizado supervisionado que usa um sistema de correção de erros de modo a ajustar os pesos da rede através das observações de entrada e saída desejada, ou seja, realiza o treinamento da rede ajustando os pesos. Esse treinamento é realizado em duas fases que percorrem o sentido da rede, chamadas fase *forward* e fase *backward*, onde a primeira define a saída para uma dada entrada na rede, e a outra utiliza a saída desejada para atualizar os pesos da rede.

O *backpropagation* consiste em definir a saída da rede na primeira fase; comparar as saídas geradas com as saídas desejadas; atualizar os pesos na segunda fase. O seu objetivo principal é minimizar o erro médio, por isso altera os pesos da rede de modo a ter uma saída mais próxima possível da saída desejada, onde os ajustes são realizados pelo método do *gradiente descendente* de uma função, que percorre na direção e sentido em que a função tem taxa de variação máxima, o que garante que a rede segue na direção de maior redução do erro.

Haykin (2001) descreve que ao considerarmos que o neurônio j é uma unidade de saída, temos que o erro na iteração “ n ” é dado por:

$$e_j(n) = d_j(n) - y_j(n) \quad (2.16)$$

em que $d_j(n)$ é a saída desejada e $y_j(n)$ é a saída estimada. A soma dos erros quadráticos na camada de saída pode ser escrita como

$$\xi(n) = \frac{1}{2} \sum_{j \in C} e_j^2(n) \quad (2.17)$$

onde C é o conjunto de todos os neurônios da camada de saída. Temos agora também que o erro médio com relação aos N padrões (exemplos) existentes no conjunto de treinamento é:

$$\xi_{med} = \frac{1}{N} \sum_{n=1}^N \xi(n) \quad (2.18)$$

Como o objetivo é minimizar o erro médio, haverá modificações nos pesos padrão a padrão. A média destas modificações (sobre todos os exemplos) é uma estimativa da modificação que se obteria caso se minimizasse sobre todo o conjunto. Tome como exemplo a seguinte arquitetura de rede:

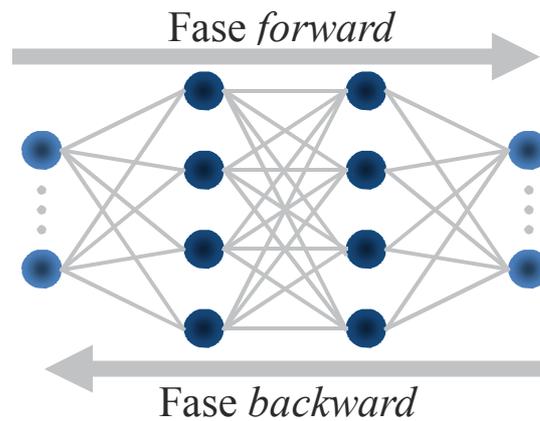


FIGURA 2.19 – Rede com 2 entradas, 2 camadas escondidas e uma camada de saída com 2 neurônios.

Fonte: O autor.

Na fase *forward*, a entrada é apresentada à primeira camada de neurônios da rede e propagada em direção às saídas. Os neurônios desta primeira camada calculam seus sinais de saída e propagam à camada seguinte. A última camada oculta calcula seus sinais de saída e os envia à camada de saída final, que calcula os valores de saída da rede.

Inicia-se a fase *backward*, onde a camada de saída calcula o erro da rede, calculando o termo de correção dos pesos. Este erro é enviado para a última camada oculta, que por sua vez calcula seu erro, calculando o termo de correção dos pesos. Este erro é então enviado para a primeira camada oculta que realiza o mesmo processo, e assim cada unidade atualiza seus pesos, e este processo é repetido até a rede neural aprender o padrão de entrada.

Se considerarmos o índice j como camada de saída e i como a camada anterior. A entrada do neurônio j é

$$v_j(n) = \sum_{i=0}^p w_{ji}(n) y_i(n) \quad (2.18)$$

onde p é o número total de ligações (excluindo o bias) que chegam ao neurônio j . O peso w_{j0} corresponde a uma entrada fixa $y_0 = +1$, sendo o termo de polarização. A saída do neurônio j na iteração n , que é a função de ativação da camada de saída é dada por

$$y_j(n) = \phi_j(v_j(n)) \quad (2.19)$$

São feitas modificações nos pesos $\Delta w_{ji}(n)$ proporcionais ao gradiente instantâneo, onde o algoritmo corrige o peso sinápticos. Com a regra da cadeia, o gradiente pode ser descrito da seguinte forma:

$$\frac{\partial \xi(n)}{\partial w_{ji}(n)} = \frac{\partial \xi(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial w_{ji}(n)} \quad (2.20)$$

Haykin (2001, p. 189) diz que “a derivada parcial $\partial \xi(n)/\partial w_{ji}(n)$ representa um *fator de sensibilidade*, determinando a direção de busca no espaço de pesos, para o uso sináptico w_{ji} .” Temos então as seguintes derivadas

$$\begin{array}{cccc} \frac{\partial \xi(n)}{\partial e_j(n)} = e_j(n) & \frac{\partial e_j(n)}{\partial y_j(n)} = -1 & \frac{\partial y_j(n)}{\partial v_j(n)} = \phi'_j(v_j(n)) & \frac{\partial v_j(n)}{\partial w_{ji}(n)} = y_i(n) \\ (a) & (b) & (c) & (d) \end{array} \quad (2.21)$$

o que resulta em

$$\frac{\partial \xi(n)}{\partial w_{ji}(n)} = -e_j(n) \phi'_j(v_j(n)) y_i(n) \quad (2.22)$$

sendo o apóstrofe representante de uma diferenciação em relação ao argumento. Estes cálculos nos levam ao fato de que a correção dos pesos é feita pela chamada Regra Delta, dada por

$$\Delta w_{ji}(n) = -\eta \frac{\partial \xi(n)}{\partial w_{ji}(n)} \quad (2.23)$$

onde o sinal de menos foi introduzido para que se caminhe no sentido decrescente no espaço de pesos. O parâmetro η é a taxa de aprendizado do algoritmo de retropropagação. O uso da equação (2.22) em (2.23) nos leva a

$$\Delta w_{ji}(n) = \eta \delta_j(n) y_i(n) \quad (2.24)$$

onde

$$\delta_j(n) = - \frac{\partial \xi(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} = - e_j(n) \phi'_j(v_j(n)) \quad (2.25)$$

Então, se j é uma unidade de saída, sabemos como atualizar os pesos que ligam esta unidade a camada anterior. Quando se tratar de uma unidade pertencente a alguma camada oculta, redefinimos $\delta_j(n)$:

$$\delta_j(n) = - \frac{\partial \xi(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} = - \frac{\partial \xi(n)}{\partial y_j(n)} \phi'_j(v_j(n)) \quad (2.26)$$

Agora, no caso da camada oculta, reescrevemos a definição de erro, trocando o índice j pelo índice k :

$$\xi(n) = \frac{1}{2} \sum_{k \in C} e_k^2(n) \quad (2.27)$$

onde k é um neurônio de saída. O índice j se refere agora a neurônios da camada oculta. Diferenciando (2.27) em relação ao sinal $y_j(n)$, temos:

$$\frac{\partial \xi(n)}{\partial y_j(n)} = \sum_{k \in C} e_k \frac{\partial e_k(n)}{\partial y_j(n)} \quad (2.28)$$

Aplicando a regra da cadeia e reescrevendo (2.28), obtemos:

$$\frac{\partial \xi(n)}{\partial y_j(n)} = \sum_{k \in C} e_k(n) \frac{\partial e_k(n)}{\partial v_k(n)} \frac{\partial v_k(n)}{\partial y_j(n)} \quad (2.29)$$

mas temos que

$$e_k(n) = d_k(n) - y_k(n) = d_k(n) - \phi_k(v_k(n)) \quad (2.30)$$

o que nos leva a

$$\frac{\partial e_k(n)}{\partial v_k(n)} = - \phi'_k(v_k(n)) \quad (2.31)$$

Por outro lado, observamos que para o neurônio k , o campo local induzido é

$$v_k(n) = \sum_{j=0}^q w_{kj}(n) y_j(n) \quad (2.32)$$

onde q representa o número de ligações (excluindo o bias) que chegam ao neurônio k . Derivando (2.32) em relação a $y_j(n)$, temos:

$$\frac{\partial v_k(n)}{\partial y_j(n)} = w_{kj}(n) \quad (2.33)$$

Fazendo as substituições de (2.29) e (2.23) em (2.27), chegamos à derivada parcial desejada:

$$\frac{\partial \xi(n)}{\partial y_j(n)} = - \sum_k e_k(n) \phi'_k(v_k(n)) w_{kj}(n) = - \sum_k \delta_k(n) w_{kj}(n) \quad (2.34)$$

Finalmente, aplicando a equação (2.34) em (2.26), obtemos a fórmula de retropropagação para o gradiente local $\delta_j(n)$, dada por

$$\delta_j(n) = \phi'_j(v_j(n)) \sum_k \delta_k(n) w_{kj}(n) \quad (2.35)$$

onde o neurônio j é oculto.

Em resumo, Haykin (2001) observa primeiro que a correção $\Delta w_{ji}(n)$ aplicada à sinapse que conecta o neurônio i ao j é dada pela regra delta

$$\begin{pmatrix} \text{Correção} \\ \text{de peso} \\ \Delta w_{ji}(n) \end{pmatrix} = \begin{pmatrix} \text{Parâmetro da} \\ \text{taxa de aprendizagem} \\ \eta \end{pmatrix} \cdot \begin{pmatrix} \text{Gradiente} \\ \text{local} \\ \delta_j(n) \end{pmatrix} \cdot \begin{pmatrix} \text{Sinal de entrada} \\ \text{do neurônio } j \\ y_i(n) \end{pmatrix} \quad (2.36)$$

além do fato de que o gradiente local $\delta_j(n)$ depende se o neurônio j é de saída ou oculto.

Segue um fluxograma descrevendo o funcionamento de uma rede neural em cada uma de suas etapas:

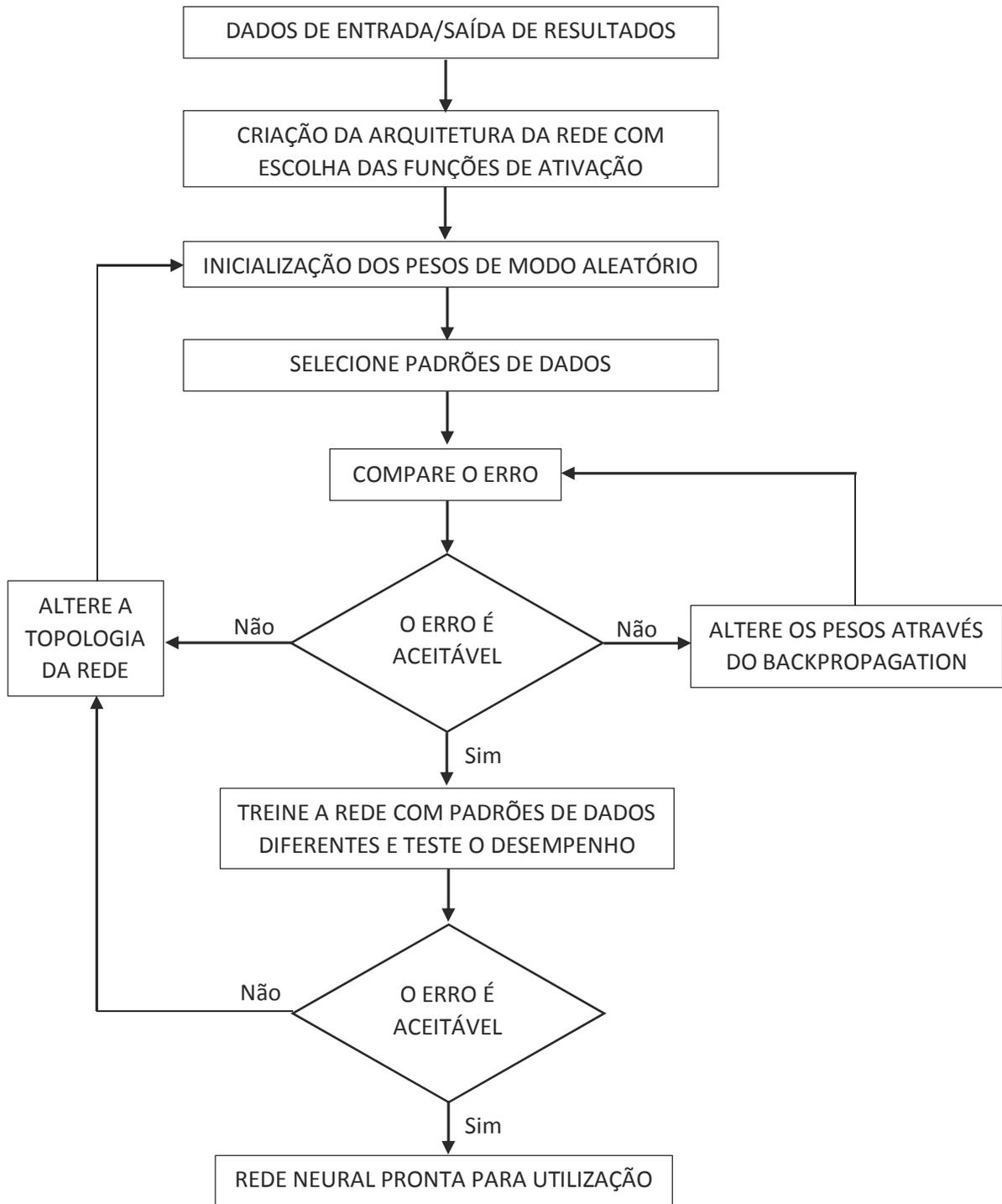


FIGURA 2.20 – Fluxograma de uma rede neural.

Fonte: Elaboração própria.

2.2.5. Exemplo Prático da Aritmética nas Redes Neurais

Suponha uma rede neural *feedforward* contendo 2 entradas, 1 camada escondida com 3 neurônios e 1 neurônio na camada de saída. Sua representação fica da seguinte forma:

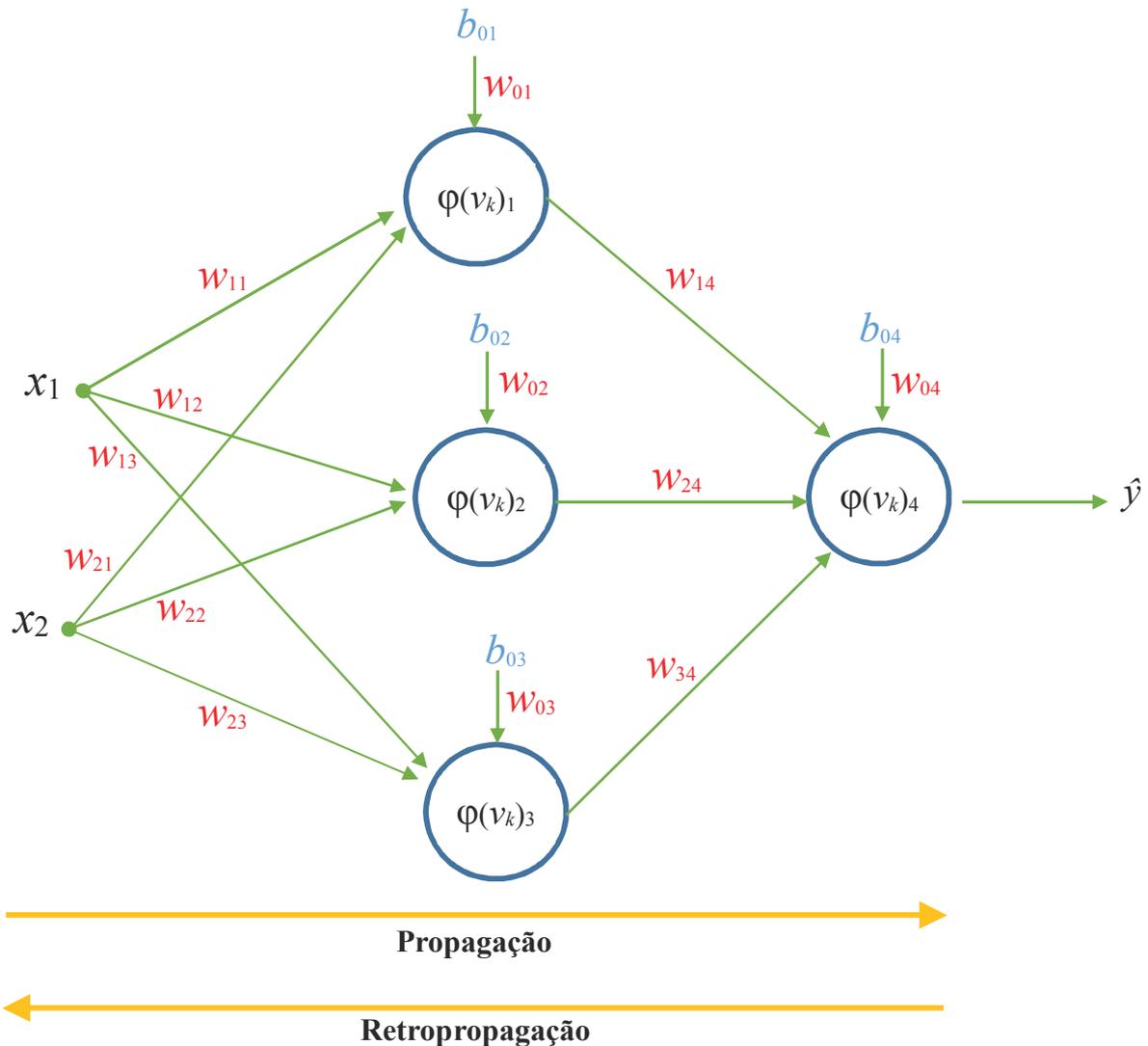


FIGURA 2.21 – Representação prática da arquitetura neural em cálculo.

Fonte: Elaboração própria.

Onde,

- $\varphi(v_k)$ é a função de ativação, também chamada de função de transferência. Vamos considerar que ela seja a função sigmóide: $\varphi(v_k) = 1/[1 + \exp(-v_k)]$;
- W são os pesos sinápticos. A rede modela os pesos do modo a ter um valor estimado mais próximo possível do valor real;
- b são os *bias* (viés).

$v_{k1} = \sum(w_{i1} \cdot x_i) + b_{01} \cdot w_{01}$, implicando

$$\varphi(v_k)_1 = \frac{1}{1 + \exp - [\sum(w_{i1} \cdot x_i) + b_{01} \cdot w_{01}]} \quad (2.37)$$

$v_{k2} = \sum(w_{i2} \cdot x_i) + b_{02} \cdot w_{02}$, implicando

$$\varphi(v_k)_2 = \frac{1}{1 + \exp - [\sum(w_{i2} \cdot x_i) + b_{02} \cdot w_{02}]} \quad (2.38)$$

$v_{k3} = \sum(w_{i3} \cdot x_i) + b_{03} \cdot w_{03}$, implicando

$$\varphi(v_k)_3 = \frac{1}{1 + \exp - [\sum(w_{i3} \cdot x_i) + b_{03} \cdot w_{03}]} \quad (2.39)$$

$v_{k4} = \sum(w_{i4} \cdot x_i) + b_{04} \cdot w_{04} = w_{14} \cdot \varphi(v_k)_1 + w_{24} \cdot \varphi(v_k)_2 + w_{34} \cdot \varphi(v_k)_3 + b_{04} \cdot w_{04}$, implicando

$$\varphi(v_k)_4 = \frac{1}{1 + \exp - (\sum w_{i4} \cdot \varphi(v_k)_i + b_{04} \cdot w_{04})} \quad (2.40)$$

O resultado de $\varphi(v_k)_4$ é a estimação da rede, \hat{y} . Os valores são estimados através de um processo iterativo dado pelo algoritmo de retropropagação que ajusta todos os pesos da rede para que o erro seja mínimo, isto é, para que a estimação seja um valor mais próximo possível do valor real observado. Isto se dá pelo movimento de propagação (da entrada até a saída) e pelo movimento de retropropagação (da saída até a entrada), de modo que a saída final \hat{y} seja a melhor possível.

2.2.6. Generalização

Haykin (2001) descreve que temos duas fases numa rede neural, uma de treinamento (*in sample*) e uma de testes (*out of sample*), quando se pretende obter a generalização. A primeira fase obtém o conhecimento a partir dos dados fornecidos à rede, ou seja, aprende com o comportamento do ambiente. A segunda fase testa a qualidade do que a rede assimilou na fase anterior, tendo como objetivo produzir inferências como projeções, classificações, previsões, etc., e este conhecimento aprendido pela rede está contido em suas sinapses, ou seja, nos seus

pesos. Deste modo, temos que uma rede neural trabalha com uma parte dos dados para treinamento e outra para validação dos resultados.

A rede tem êxito em sua generalização quando gera um resultado representativo ao resultado obtido na fase de treinamento com o aprendizado da relação entrada-saída. Outro ponto importante é utilizar uma arquitetura simples, apropriada, com base nas entradas e saídas, que podemos descrever pela expressão $N > W/\xi$, tendo N como o volume de exemplos da primeira fase, W a quantidade de pesos e ξ é o erro admitido na segunda fase.

De modo a analisar o comportamento da rede, é recomendado considerar outros testes, como entradas especiais e análise dos pesos atuais, como valores muito pequenos que podem ser considerados insignificantes e excluídos. Da mesma forma, valores muito grandes podem significar que houve um processo de *over-training* na rede.

Enfim, com o fim da fase de treino e generalização (validação), a rede pode ser aplicada ao ambiente que se deseja, sendo importante que se monitore o desempenho da rede periodicamente para que se possa realizar uma manutenção quando for necessário, como por exemplo, um retreinamento.

CAPÍTULO 3 – A PREVISÃO DO IBOVESPA

3.1. A Escolha do Ativo

Nesta fase, teremos o experimento propriamente dito. O ativo escolhido para a previsão foi o Ibovespa (Índice da Bolsa de Valores de São Paulo), que como visto anteriormente, possui cotações em pontos, o que dispensa o uso de métodos matemáticos para deflacionar a série, visto não se tratar de valores em moeda corrente.

Outro ponto é que o índice representa uma carteira teórica com as principais ações do mercado brasileiro, as de maior liquidez e negociabilidade, o que em si reflete o mercado como um todo, além de ser um ativo menos propenso a contaminações como manipulações de preços, já que movimenta bilhões de reais por dia no somatório financeiro nos negócios com as ações das empresas em sua carteira teórica. Portanto, a previsão será realizada com a série histórica do Ibovespa.

3.2. Por que usar apenas a série temporal do Ibovespa e não outras variáveis?

Elder (2011), ao estudar o comportamento da psicologia no mercado financeiro, observa que o principal líder do mercado é o preço, e coloca:

Cada preço é o consenso momentâneo entre os participantes do mercado sobre o valor de determinado ativo. Revela a votação mais recente a respeito do valor de determinado objeto de negociação. Qualquer investidor pode dar o seu voto, emitindo uma ordem de compra ou venda ou recusando-se a negociar ao preço vigente. Cada barra no gráfico reflete a batalha entre touros e ursos. Quando os touros se sentem mais vigorosos, eles compram com mais ansiedade e empurram o mercado para cima. Quando os ursos se sentem mais vigorosos, eles vendem com mais disposição e puxam o mercado para baixo. Cada preço reflete a ação ou inação de todos os operadores de mercado. (ELDER, 2011, p. 66)

Deste modo, temos que os preços, ou os pontos de índice, refletem o consenso do mercado a respeito de determinado ativo, funcionando como um visor da psicologia de massa, visto que o preço das ações não é baseado na forma como uma empresa trabalha, mas sim em como a psicologia do mercado a enxerga, ou seja, o preço reflete o comportamento do mercado, o seu julgamento.

Corroborando este entendimento, Matsura (2006) aborda a Teoria de Dow, em que um dos pressupostos é o de que os índices (preços) descontam tudo. Isto quer dizer que toda e qualquer informação relevante está contida nos preços, seja a informação daqueles muitos bem informados ou daqueles pouco ou nada informados, onde todos os milhares de participantes se

encontram no mercado e realizam compras e vendas por milhões de motivos distintos, e a união destes forma um consenso, que é o preço de mercado de um ativo ou índice, logo, não é necessário saber o porquê os preços se movem, mas sim como eles se movem, visto ser impossível ter acesso a todas as milhões de informações que os participantes do mercado tem acesso e equacioná-las de modo a chegar em uma previsão, o que por si só já vai contra a Teoria dos Mercados Eficientes.

Portanto, a informação mais importante que precisamos ter acesso é o próprio preço do ativo ou índice em questão, visto que este representa toda a massa de mercado, informando sobre tendência, volatilidade, entre outros fatores, tudo em uma única variável.

3.3. Características do Experimento

A previsão do IBOVESPA será estimada utilizando uma Rede Neural *feedforward*. Este tipo de modelo não-linear tem sido utilizado com sucesso em uma gama extensiva de aplicações desde o final da década de 1980.

Rede Neural Artificial, como visto anteriormente, é um modelo distribuído composto por unidades chamadas na literatura de “neurônios”, constituídas de funções não-lineares. A combinação destas unidades, através de parâmetros estimados a partir dos dados, é o que confere a capacidade deste modelo de inferir relações não-lineares de complexidade arbitrária.

Na forma utilizada neste estudo, estas unidades foram organizadas em camadas, incluindo uma camada oculta, que não está diretamente conectada à saída do modelo. Estas conexões entre as unidades, ou neurônios, são chamadas de pesos. Estes pesos são os parâmetros do modelo que são ajustados por um algoritmo iterativo através dos dados. Uma vez ajustados os pesos, a rede tem a capacidade de representar a relação dos dados de entrada com a variável de saída, neste caso a previsão do IBOVESPA. A capacidade de aprender através dos dados da amostra e de generalizar (fora-da-amostra) informação gerada em ambientes não-lineares complexos, é sem dúvida a grande vantagem das Redes Neurais.

Na Rede implementada, utilizou-se o algoritmo de Regularização Bayesiana, que de acordo com MacKay (1992), consiste em adicionar um termo de penalização (regularização) à função objetivo, de forma que o algoritmo de estimação faça com que os parâmetros irrelevantes convirjam para zero, reduzindo assim o número de parâmetros efetivos utilizados no processo.

A rede obtém conhecimento através de um processo de aprendizado, onde a informação é armazenada em “densidades de conexão” conhecidas como pesos sinápticos, e isto ocorre

através do algoritmo de *backpropagation*, capaz de ajustar iterativamente os pesos de modo que se atinja o objetivo proposto.

Em termos práticos, temos um modelo de rede neural que utilizará os dados de uma série histórica com 120 observações diárias, de 18/05/2015 a 06/11/2015, diferenciada em dois períodos. O treinamento será realizado com as primeiras 100 (cem) observações desta série, tendo um treinamento supervisionado, isto é, forneceremos as entradas e as saídas que queremos para estas 100 observações, com isso, a rede fará o cálculo de modo a chegar na menor diferença, no menor erro, entre a entrada e a saída desejada e irá aprender com estes dados, se moldando às características e padrões de movimentação do Ibovespa.

Após o treinamento, o próximo passo é a generalização, que consiste na capacidade da rede de reconhecer o ambiente que origina os dados não utilizados no treinamento. Isto significa que uma rede neural consegue produzir saídas adequadas para entradas não observadas no seu processo de treinamento.

Posteriormente, para comprovar o aprendizado e a capacidade de previsão da rede, as últimas 20 (vinte) observações de entrada serão fornecidas à rede, e a mesma irá gerar as saídas de previsão.

A arquitetura da rede conta com duas entradas defasadas do tipo $Z_{t-1} + Z_{t-2}$, cinco neurônios na camada oculta, utilizando a função de ativação não linear sigmóide, tendo ao fim um neurônio na camada de saída com função linear, visto que o resultado de saída tem natureza linear.

3.4. Resultados

A seguir, a FIGURA 3.1 mostra o resultado que a rede gerou no processo de treinamento. O gráfico em azul representa os valores reais do Ibovespa, já o gráfico vermelho representa os valores estimados com o treinamento, isto é, a partir do conhecimento adquirido, gerou a estimacão. Temos que esta é uma rede com treinamento supervisionado, ou seja, ao ter as entradas e as saídas desejadas, a rede gera o melhor valor possível de saída. Este gráfico contém as primeiras 100 observações.

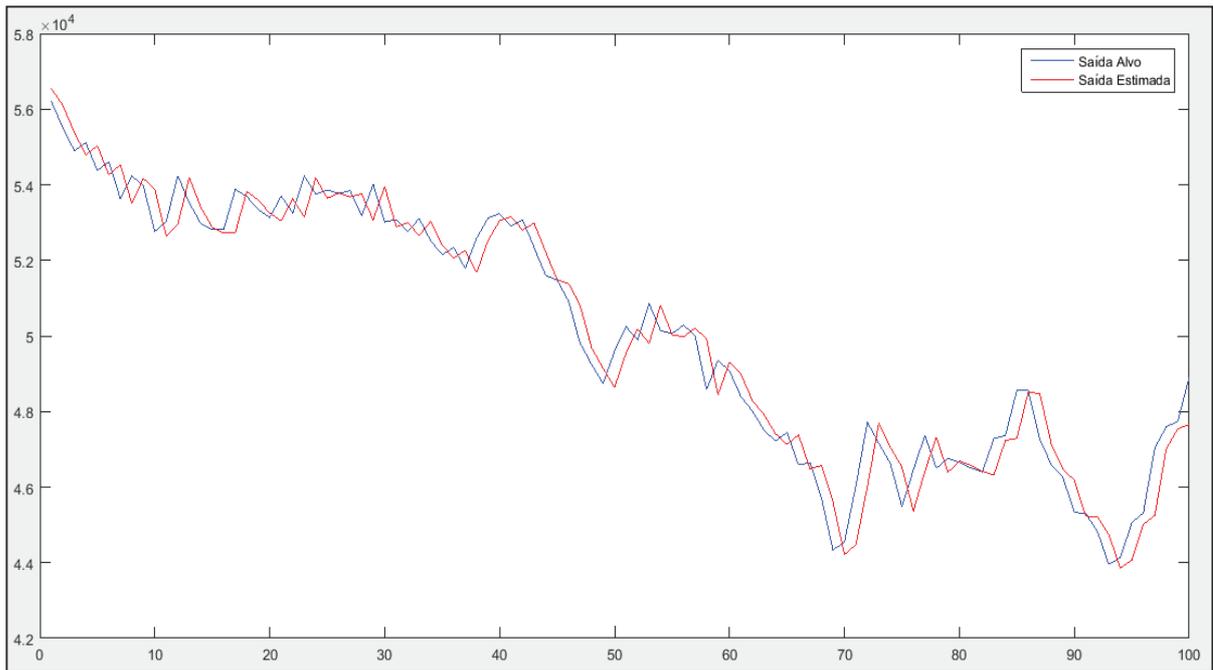


FIGURA 3.1 – Gráfico do Ibovespa e sua estimação a partir do treinamento.

Fonte: MatLab.

Em seguida podemos observar a FIGURA 3.2, mostrando a estimação efetiva do Ibovespa para as últimas 20 observações, ou seja, o resultado da generalização, a previsão efetiva.

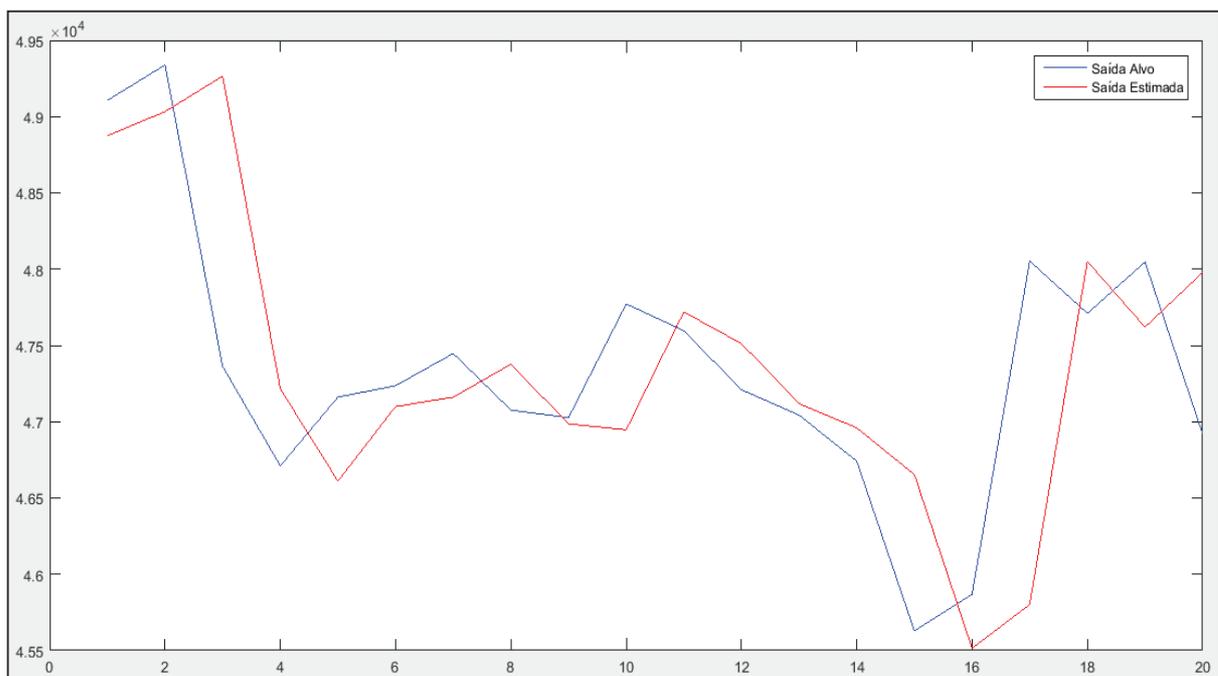


FIGURA 3.2 – Gráfico estimado do Ibovespa e as observações reais.

Fonte: MatLab.

Isto ocorre a partir do fato de que foi fornecido para a rede as últimas 20 (vinte) observações de entrada, de modo a estimar 20 (vinte) valores de saída, que são as previsões efetivas exibidas pelo gráfico em vermelho de acordo com os dados de entrada. Portanto, temos que este gráfico estimado foi o que a rede previu que iria acontecer e, o gráfico azul foi o que realmente aconteceu.

Concatenando as FIGURAS 3.1 e 3.2, temos um gráfico completo com os valores reais do Ibovespa em azul e as estimações em vermelho, como pode ser visto na FIGURA 3.3 a seguir.

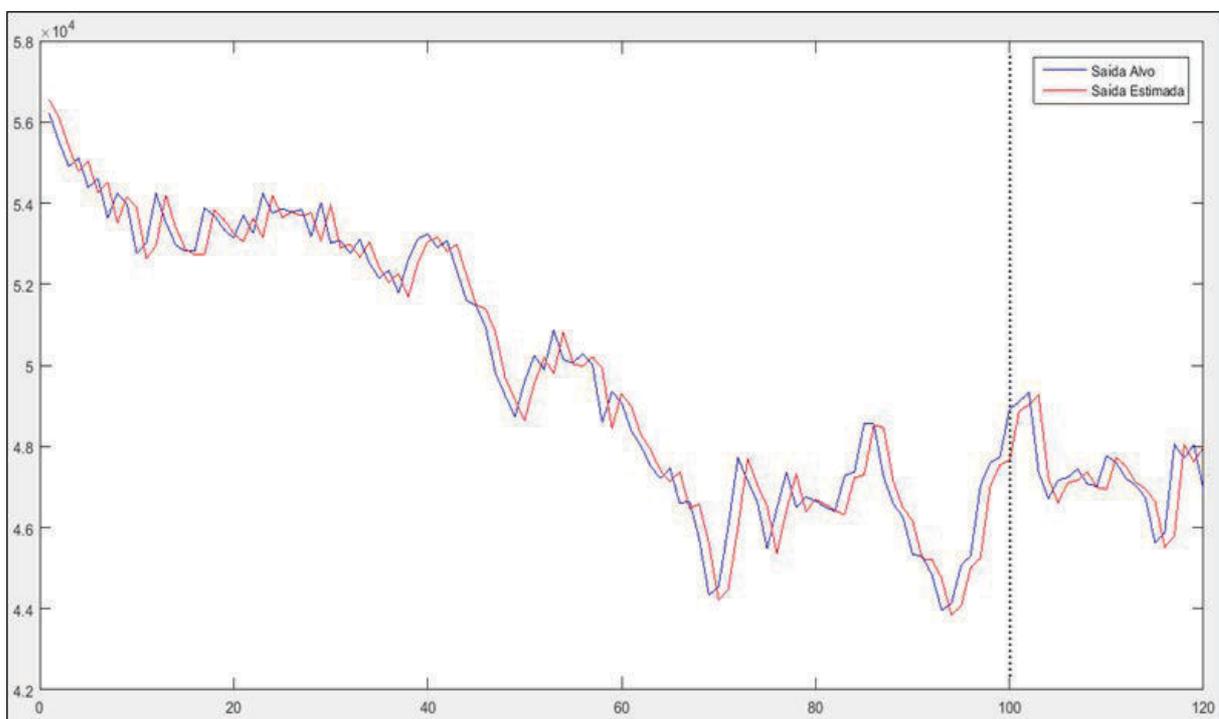


FIGURA 3.3 – Previsão do Ibovespa e seus valores reais.

Fonte: O autor via MatLab.

A linha pontilhada vertical mostra a divisão entre as observações e a estimativa de treinamento e a efetiva previsão, comparando com os valores reais. O gráfico azul representa a série de valores reais observados do Ibovespa. Em vermelho, na parte esquerda, temos a estimativa gerada no treinamento e, na direita, os valores previstos gerados com o conhecimento adquirido pela rede neural.

Podemos perceber que a capacidade de gerar valores de previsão com qualidade se manteve em relação às estimativas obtidas na fase de treinamento, o que demonstra que a rede foi capaz de aprender adequadamente com a série temporal e, posteriormente, prever movimentos futuros.

3.5. Validação do Modelo

Observando os resultados a partir dos gráficos, é possível perceber que a previsão não é perfeita, assim como nenhuma o é, existindo então um termo de erro, que é a diferença entre o valor real e o valor estimado, visto pela diferença entre os gráficos.

A tabela a seguir tem a representação dos valores reais e estimados do Ibovespa na generalização, incluindo o termo de erro para cada valor de previsão:

Valor Real	Valor Estimado	Erro	Erro %
49107	48752	355	0,72%
49338	48901	437	0,89%
47363	49147	-1784	3,77%
46710	46946	-236	0,51%
47161	46530	631	1,34%
47236	47175	61	0,13%
47447	47157	290	0,61%
47077	47368	-291	0,62%
47026	46929	97	0,21%
47772	46956	816	1,71%
47597	47748	-151	0,32%
47209	47438	-229	0,49%
47043	47054	-11	0,02%
46741	46952	-211	0,45%
45628	46666	-1038	2,27%
45869	45505	364	0,79%
48054	46117	1937	4,03%
47710	48239	-529	1,11%
48047	47500	547	1,14%
46919	47895	-976	2,08%

TABELA 3.1: valores reais, estimados e o termo de erro.

Fonte: Elaboração própria com dados Bovespa e MatLab.

Diante das observações de erros, dos valores observados e estimados, foi possível calcular medidas estatísticas para a avaliação do modelo, verificando os valores do Erro Percentual Absoluto Médio (MAPE) e o U-THEIL, dados respectivamente pelas seguintes equações:

$$MAPE = \frac{\sum_{i=1}^N |e(i)|}{N} \times 100 \quad (3.1)$$

$$U-THEIL = \sqrt{\frac{\frac{1}{N-1} \sum_{i=1}^{N-1} e_i^2}{\frac{1}{N-1} \sum_{i=1}^{N-1} (Z(i) - Z(i-1))}} \quad (3.2)$$

A medida de erro U-THEIL tem a característica de poder comparar o erro de um modelo testado em relação ao método ingênuo de previsão (ou método Naive), que consiste em considerar o valor observado no tempo t com outro no tempo t-1. Ou seja, o método ingênuo é aquele que diz que a previsão do período seguinte é o último valor observado. Se U-THEIL < 1, a previsão do modelo é melhor do que o método ingênuo, se U-THEIL > 1, a sua previsão do modelo é pior do que o método ingênuo.

O valor obtido para o MAPE foi de 1,16%, que representa um erro bastante baixo, caracterizando uma boa capacidade de previsão do modelo, e tendo um U-THEIL de 0,92, confirmando que a previsão do modelo é consistente, tendo um baixo nível de erro, o que valida o modelo.

CONCLUSÃO

É sabido que séries econômicas como séries de índices de preços, inflação entre outras, muitas das vezes apresentam um comportamento que dificulta uma modelagem através de modelos puramente lineares quando abordada a questão da eficiência preditiva dos modelos.

A metodologia desenvolvida nesta monografia contemplou esta realidade através de uma estratégia de construção de um modelo não linear para o índice Ibovespa na tentativa de encontrar um modelo que tratasse os aspectos não lineares intrínsecos ao mundo real e aos fenômenos econômicos de tal forma que pudesse otimizar os resultados das previsões.

Acredita-se que este objetivo foi alcançado na medida em que Erro Percentual Absoluto Médio (MAPE) encontrado foi de 1,16%, e que podemos considerar um nível de erro bastante baixo, o que também se justifica por termos encontrado a medida U-THEIL < 1 , confirmando que o modelo não linear adotado (redes neurais) apresenta um resultado altamente satisfatório, prevendo de modo bastante significativo a tendência dos valores futuros.

Da mesma forma, observamos que o gráfico dos dados estimado é similar ao gráfico dos dados reais, havendo apenas pequena defasagem em alguns pontos, sem nenhum tipo de discrepância, mostrando que o modelo de Redes Neurais é robusto na tarefa de previsão do Índice Bovespa. Sendo assim, observa-se que o modelo de Redes Neurais constitui uma ferramenta poderosa e adequada para tal objetivo.

Estes resultados demonstram que os argumentos sustentados pela Teoria dos Mercados Eficientes e apresentados neste trabalho são questionáveis, pois o modelo foi capaz de realizar uma previsão satisfatória que a teoria defende não ser possível. Assim, é possível argumentar que a teoria não possui alicerce robusto, pois o modelo de previsão não linear das Redes Neurais foi capaz de invalidar argumentos como o de não ser possível auferir lucros com base no estudo de séries de preços, visto que estas seguiriam um caminho aleatório impossível de prever.

Tendo em vista estas considerações, pode-se sugerir ainda, para trabalhos futuros, um modelo híbrido, com o intuito de melhorar ainda mais a capacidade de previsão do modelo. Poderia ser utilizado um modelo linear para estimar a componente de tendência e sazonalidade da série, caso existam, e um modelo não linear para estimar o erro de previsão resultante do modelo anterior. Este modelo híbrido, além de reduzir de forma significativa o erro de previsão *out of sample*, também apresenta uma outra importante vantagem: mantém a possibilidade do tomador de decisões se antecipar diante de uma realidade desconhecida.

Além disso, comprovou-se que, em termos de eficiência preditiva, o uso das Redes Neurais para previsão em séries econômicas pode constituir uma poderosa ferramenta auxiliar na tomada de decisão.

Por fim, ressalta-se que a metodologia de construção de modelos utilizada para prever o índice Ibovespa, pode ser aplicada, em princípio, a qualquer série temporal de interesse, na área econômica ou não.

REFERÊNCIAS BIBLIOGRÁFICAS

BM&FBOVESPA. **Introdução ao Mercado de Capitais**. Disponível em: <https://corretora.miraeasset.com.br/global/bz/po/downloads/cursosOnline/Introdu%C3%A7%C3%A3o_ao_Mercado_de_Capitais.pdf>. Acesso em: 16 jan. 2016.

BM&FBOVESPA. **MERCADO A VISTA DE AÇÕES: Conceitos e Definições**. São Paulo, Instituto Educacional, 2011.

BM&FBOVESPA. **Metodologia do Índice Bovespa**. Disponível em: <<http://www.bmfbovespa.com.br/indices/ResumoIndice.aspx?Indice=ibovespa&=Opca&idoma=pt-br>>. Acesso em: 21 dez. 2015.

BLANCHARD, Olivier. **Macroeconomia**. São Paulo: Pearson Prentice Hall, 2007.

BÚSSOLA DO INVESTIDOR. Disponível em: <<https://www.bussoladoinvestidor.com.br/graficos/>>. Acesso em: 11 nov. 2015.

CAVALHEIRO, Everton Anger. et al. Aplicação de Redes Neurais Polinomiais na Previsão do Ibovespa e Merval. **Revista Desenvolvimento em Questão**, vol. 9, nº. 18, p. 196-224. Rio Grande do Sul, jul/set. 2011.

CERETTA, Paulo Sergio. **Eficiência no Mercado de Ações Latino-Americano: contribuição da teoria do caos**. 2003. Disponível em: <http://www.anpad.org.br/~anpad/eventos.php?cod_evento=1&cod_edicao_subsecao=48&cod_evento_edicao=7&cod_edicao_trabalho=1891>. Acesso em: 01 nov. 2015.

COSTA JR., N. C. A.. Sazonalidades do IBOVESPA. **Revista de Administração de Empresas**, vol. 30, no. 3, p. 80. São Paulo, jul/set. 1990.

CVM. **Mercado de Valores Mobiliários Brasileiro**. 3ª ed. Rio de Janeiro, Comissão de Valores Mobiliários, 2014.

ELDER, Alexander. **Como se Transformar em um Operador e Investidor de Sucesso:** entenda a psicologia do mercado financeiro. 15ª ed. Rio de Janeiro, Elsevier, 2011.

GOOGLE IMAGENS. Disponível em: <<http://www.google.com>>. Acesso em: 15 nov. 2015.

HAYKIN, Simon. **Redes Neurais:** Princípios e Prática. 2ª ed. Porto Alegre, Bookman, 2001.

IKEDA, Ricardo Hirata; RABELO JUNIOR, Tarcísio Saraiva. Mercados Eficientes e Arbitragem: Um estudo sob o enfoque das finanças comportamentais. **Revista Contabilidade & Finanças - USP**, vol. 15, n. 34, p. 97 - 107, São Paulo, janeiro/abril 2004.

MACRINI, José Leonardo Ribeiro. **MODELO DE PREVISÃO DE CARGA UTILIZANDO REDES NEURAIAS:** OTIMIZAÇÃO CAMADA A CAMADA. 2000. 73 f. Dissertação (Mestrado em Engenharia Elétrica) – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2000.

MacKay, D.J.C. **Bayesian interpolation.** Neural Computation, p. 415-447. 1992.

MATLAB, **MathWorks.** Disponível em: <<http://www.mathworks.com/products/matlab/>>. Acesso em: 11 nov. 2015.

MATSURA, Eduardo. **Comprar ou Vender?** Como Investir na Bolsa Utilizando Análise Gráfica. 4ª ed. Rio de Janeiro, Saraiva, 2006.

MUELLER, Alessandro. **Uma Aplicação de Redes Neurais Artificiais na Previsão do Mercado Acionário.** 1996. 103 f. Dissertação (Mestrado em Engenharia) – Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 1996.

PENTEADO, Marco Antônio de Barros. **Uma avaliação estatística da análise gráfica no mercado de ações brasileiro à luz da teoria dos mercados eficientes e das finanças comportamentais.** 2003. 124 f. Dissertação (Mestrado em Administração) – Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo, 2003.

PINHEIRO, Juliano Lima. **Mercado de Capitais: Fundamentos e Técnicas**. São Paulo, Atlas, 2009.

RODRIGUES, Alcantaro Lemes. **Redes Neurais Artificiais Aplicadas na Previsão de Preços do Mercado Spot de Energia Elétrica**. 2009. 179 f. Dissertação (Mestrado em Engenharia) – Programa de Pós-graduação em Energia, Universidade de São Paulo, São Paulo, 2009.