

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
CAMPUS GOVERNADOR VALADARES
BACHARELADO EM CIÊNCIAS ECONÔMICAS

**O MERCADO FINANCEIRO E
INTELIGÊNCIA ARTIFICIAL: UMA
COMPARAÇÃO DE MODELOS DE
APRENDIZAGEM AUTOMÁTICA COM
MODELOS TRADICIONAIS DE
PRECIFICAÇÃO**

GUSTAVO LIRA UGARTE

GOVERNADOR VALADARES
SETEMBRO, 2021

**O MERCADO FINANCEIRO E
INTELIGÊNCIA ARTIFICIAL: UMA
COMPARAÇÃO DE MODELOS DE
APRENDIZAGEM AUTOMÁTICA COM
MODELOS TRADICIONAIS DE
PRECIFICAÇÃO**

GUSTAVO LIRA UGARTE

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
CAMPUS GOVERNADOR VALADARES
INSTITUTO DE CIÊNCIAS SOCIAIS APLICADAS
Bacharelado em ciências econômicas

Orientador: Alex Sander de Moura

GOVERNADOR VALADARES

SETEMBRO, 2021

O MERCADO FINANCEIRO E INTELIGÊNCIA ARTIFICIAL: UMA
COMPARAÇÃO DE MODELOS DE APRENDIZAGEM
AUTOMÁTICA COM MODELOS TRADICIONAIS DE
PRECIFICAÇÃO

GUSTAVO LIRA UGARTE

MONOGRAFIA SUBMETIDA AO CORPO DOCENTE DO CAMPUS GOVERNADOR
VALADARES DA UNIVERSIDADE FEDERAL DE JUIZ DE FORA, COMO PARTE
INTEGRANTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU
DE BACHAREL EM CIÊNCIAS ECONÔMICAS.

Aprovada por:

Alex Sander de Moura
Doutor

Werley Gomes Facco
Doutor

Willian Vieira de Paula
Mestre

GOVERNADOR VALADARES

08 DE *SETEMBRO*, 2021

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Lira Ugarte, Gustavo.

O mercado financeiro e inteligência artificial: uma comparação de modelos de aprendizagem automática com modelos tradicionais de precificação / Gustavo Lira Ugarte. -- 2021.

75 f.

Orientador: Alex Sander de Moura

Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Juiz de Fora, Campus Avançado de Governador Valadares, Instituto de Ciências Sociais Aplicadas - ICSA, 2021.

1. Redes neurais. 2. Precificação de opções. 3. Black-Scholes. I. Sander de Moura, Alex, orient. II. Título.



UNIVERSIDADE FEDERAL DE JUIZ DE FORA

ATA DE DEFESA DE TRABALHO DE CONCLUSÃO DE CURSO

Às 10 horas do dia 10 de setembro de 2021, por webconferência, conforme Resolução N° 24/2020 do Conselho Superior (CONSU), foi instalada a banca do exame de Trabalho de Conclusão de Curso para julgamento do trabalho desenvolvido pelo(a) discente Gustavo Lira Ugarte, matriculado(a) no curso de bacharelado em Ciências Econômicas. O(a) Prof.(a) Dr. Alex Sander de Moura, orientador(a) e presidente da banca julgadora, abriu a sessão apresentando os demais examinadores, os professores: Dr. Werley Gomes Facco e Me. Willian Vieira de Paula.

Após a arguição e avaliação do material apresentado, relativo ao trabalho intitulado: **O MERCADO FINANCEIRO E A INTELIGÊNCIA ARTIFICIAL: UMA COMPARAÇÃO DE MODELOS DE APRENDIZAGEM AUTOMÁTICA COM MODELOS TRADICIONAIS DE PRECIFICAÇÃO**, a banca examinadora se reuniu em sessão fechada considerando o(a) discente Gustavo Lira Ugarte:

Aprovado(a)

Reprovado(a)

Nada mais havendo a tratar, foi encerrada a sessão e lavrada a presente ata que vai assinada eletronicamente pelos presentes.

Governador Valadares, 08 de setembro de 2021.

Prof.(a) Dr. Alex Sander de Moura

Prof.(a) Dr. Werley Gomes Facco

Prof.(a) Me. Willian Vieira de Paula



Documento assinado eletronicamente por **Alex Sander de Moura, Professor(a)**, em 08/09/2021, às 10:51, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **WERLEY GOMES FACCO, Usuário Externo**, em 08/09/2021, às 10:56, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Willian Vieira de Paula, Professor(a)**, em 10/09/2021, às 13:58, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf (www2.ufjf.br/SEI) através do ícone Conferência de Documentos, informando o código verificador **0488034** e o código CRC **58A8BF5A**.

Resumo

O objetivo deste trabalho é investigar a aplicação de redes neurais na precificação de opções. Diferentes arquiteturas e parâmetros foram testados a fim de obter melhor desempenho e maior precisão na predição. Foram utilizados dois mecanismos de previsão de volatilidade para comparar o desempenho da rede neural com o modelo paramétrico Black-Scholes. Também foi comparado o desempenho da rede neural com redes neurais modulares mais avançadas e uma técnica completamente nova foi testada para adicionar hipóteses de predição racional às redes neurais, bem como, foi mostrada a importância de adicionar opções virtuais que satisfaçam essas hipóteses para melhor treinar as redes neurais. Além disso, o tamanho e o número de camadas ocultas foram testados para fornecer uma referência e uma forma estruturada de como escolher os parâmetros da rede neural. Foi comprovado também que a previsão de redes neurais supera os métodos paramétricos tradicionais.

Palavras-chave: Redes neurais, Redes neurais modulares, Black-scholes, Precificação de opções. .

Abstract

The aim of this work is to investigate the application of neural networks in option pricing. Different architectures and parameters were tested in order to obtain better performance and greater prediction accuracy. Two volatility prediction mechanisms were used to compare the performance of the neural network with the parametric Black-Scholes model. Neural network performance was also compared with more advanced modular neural networks and a completely new technique was tested to add rational prediction hypotheses to neural networks, as well as the importance of adding virtual options that satisfy these hypotheses to better train the neural networks. In addition, the size and number of hidden layers were tested to provide a reference and a structured way of choosing neural network parameters. It was also proven that the prediction of neural networks surpasses traditional parametric methods.

Keywords: Neural Networks, Modular Neural Networks, Black-scholes, Option Pricing.

Conteúdo

Lista de Figuras	5
Lista de Tabelas	6
Lista de Abreviações	7
1 Introdução	8
1.1 Hipóteses	9
2 Referencial Teórico	11
2.1 Precificação de opções	11
2.1.1 Modelos tradicionais de precificação	11
2.1.2 A fórmula Black-Scholes	11
2.1.3 Gregas e Delta Hedging	13
2.2 Mecanismos de previsão de volatilidade	14
2.2.1 Volatilidade Histórica	14
2.2.2 Volatilidade Implícita	15
2.2.3 Volatility smile	16
2.2.4 Previsão de volatilidade com os modelos ARCH	16
2.3 Inteligência artificial	20
2.3.1 Neurônios Biológicos	20
2.3.2 Redes neurais como ferramenta de aproximação não linear	21
2.3.3 Perceptron e redes neurais feedforward	21
2.3.4 Funções de Ativação	23
2.4 Redes neurais modulares	24
2.4.1 Módulos de rede neural modular	25
2.4.2 Outras arquiteturas modulares de rede neural	26
2.5 Suposições de Predição Racional	29
2.5.1 Motivação	29
2.5.2 Requisitos para previsões racionais	29
2.5.3 Verificando a racionalidade	32
3 Parte Empírica	35
3.0.1 Descrição dos dados	35
3.0.2 Simulações	46
3.0.3 Testes de resolução	46
3.0.4 Opções virtuais	52
4 Resultados	54
4.1 Previsão de volatilidade	54
4.2 Previsão com redes neurais	56
4.3 Previsão com redes neurais modulares	60
4.4 Padrão proposto para escolha da arquitetura	64
4.5 Discussão dos resultados	66

5 Conclusão

68

Bibliografia

71

Lista de Figuras

2.1	Rede modular escolhendo módulos de acordo com moneyness m e tempo até o vencimento T	27
3.1	Preços de fechamento diário S&P500 entre 2004-2007	37
3.2	Retorno do índice S&P500 entre 2004-2007	37
3.3	Volatilidade do índice S&P500 entre 2004-2007	37
3.4	Preço de strike $X = 1100$	39
3.5	Preço de strike $X = 1200$	40
3.6	Preço de strike $X = 1300$	40
3.7	Preço de strike $X = 1400$	41
3.8	Preço de strike $X = 1100$	41
3.9	Preço de strike $X = 1200$	42
3.10	Preço de strike $X = 1300$	42
3.11	Preço de strike $X = 1400$	43
3.12	Opções de compra	44
3.13	Opções de venda	44
3.14	Previsão fora da amostra ruim	48
3.15	Previsão fora da amostra boa	49
4.1	Função de autocorrelação	54
4.2	Função de autocorrelação parcial	55
4.3	Comparação do modelo de volatilidade GARCH (1,1) com a volatilidade histórica de curto prazo	55
4.4	Erro absoluto 1	58
4.5	Erro percentual absoluto 1	58
4.6	Precisão de previsão 1	58
4.7	Histograma de diferença 1	58
4.8	Erro absoluto 2	59
4.9	Erro percentual absoluto 2	59
4.10	Precisão de previsão 2	59
4.11	Histograma de diferença 2	59

Lista de Tabelas

3.1	Resumo do conjunto de dados S&P500	45
3.2	Parâmetros usados para simular preços de opções de compra	47
3.3	Estimativa MQO de MAE dos parâmetros da rede neural	50
3.4	Estimativa MQO de RMSE dos parâmetros da rede neural	50
4.1	Estimativa do modelo GARCH (1,1) com base no preço de fechamento do ativo subjacente (nível de significância 0.05 *, nível de significância 0.01 **, nível de significância 0.001 ***)	55
4.2	Comparação de precisão de previsão para diferentes modelos com vários números de opções virtuais para opções de compra.	60
4.3	Comparação de precisão de previsão para diferentes modelos com vários números de opções virtuais para opções de venda.	61
4.4	Comparação da precisão de previsão da rede neural modular com a rede feedforward simples cumprindo as suposições de previsões racionais	63
4.5	Precisão de previsão para módulos específicos na rede neural modular . . .	63

Lista de Abreviações

ATM - At the Money

DM - Diebold-Mariano

EDP - Equações Diferenciais Parciais

EQM - Erro Quadrático Médio

EMA - Erro médio Absoluto

ITM - In the Money

MAE - Erro Médio Absoluto

MAPE - Erro Médio Absoluto Percentual

MQO - Mínimos Quadrados Ordinários

OTM - Out of the Money

RNA - Rede Neural Artificial

TRA - Teoria Ressonante Adaptativa

UFJF - Universidade Federal de Juiz de Fora

ULE - Unidade Linear Exponencial

ULR - Unidade Linear Retificada

1 Introdução

Uma grande quantidade de literatura foi escrita desde que (BLACK; SCHOLES, 1973) desenvolveram sua fórmula de precificação de opções, transformando o problema de precificação de opções em uma tarefa de resolver uma equação diferencial parcial parabólica com uma condição inicial. Esta tarefa pode ser realizada sob vários pressupostos e a solução pode ser encontrada de forma fechada. No entanto, algumas premissas são violadas na prática ou são pelo menos questionáveis. Portanto, vários autores como (SCOTT, 1987) buscaram otimizar o modelo, o que levou a modelos mais precisos. Apesar de alguns sucessos específicos, premissas irrealistas, porém cruciais, como distribuição lognormal de preços e continuidade das negociações, ainda eram necessárias.

Outra abordagem foi feita por (HUTCHINSON; LO; POGGIO, 1994) que usou um modelo de rede neural artificial (RNA). Este modelo foi capaz de explicar a forma como os dados de entrada afetam o preço da opção de saída, sem qualquer forma funcional predeterminada de relacionamento. Descobriu-se que os modelos de RNA aplicados em opções de estilo americano tiveram um desempenho melhor do que a fórmula comumente usada de Black-Scholes, que serviu de referência. Este artigo inovador foi seguido por (GEIGLE; ARONSON, 1999), (GHAZIRI; ELFAKHANI; ASSI, 2000), (SAITO; JUN, 2000), e muitos outros que compararam o desempenho da abordagem não paramétrica da RNA com a fórmula de Black-Scholes com resultados variados nas opções de estilo europeu e americano.

Embora muitos trabalhos de pesquisa tenham sido escritos sobre este tópico, poucos deles analisam se as duas abordagens podem ser combinadas ou aplicadas em derivados não padronizados. Além disso, não se sabe muito sobre as frequências de amostragem adequadas de dados, bem como, existem várias maneiras de estimar o parâmetro de volatilidade que é essencial para o desempenho do modelo. Conforme afirma (HAHN, 2013), também não há consenso sobre as estratégias utilizadas na decisão de como selecionar tamanho, arquitetura, parâmetros, funções de aprendizagem e avaliação para RNAs, apesar da extensa pesquisa em aprendizado de máquina. Todos os fatores listados acima

contribuem para a ambiguidade das conclusões dos artigos publicados e não são tratados sistematicamente.

Como não existe uma fórmula fechada nas condições gerais de mercado e as redes neurais são uma forma alternativa aos modelos tradicionais, o aprimoramento da estimativa de precificação de opções é essencial. Este trabalho tem como objetivo não apenas comparar o desempenho de redes neurais padrão com a fórmula de Black-Scholes, mas também usar um conceito de redes neurais modulares e compará-lo com a previsão de rede neural feedforward simples. Além disso, tem-se como objetivo examinar e propor uma forma estruturada de como escolher o tamanho e a arquitetura da rede neural especificamente para a precificação de opções. Uma nova técnica usada por (YANG; ZHENG; HOSPEDALES, 2017) de adicionar suposições de predição racionais também é investigada e o objetivo do trabalho é também mostrar se a técnica de adicionar opções virtuais melhora a predição. Assim, o trabalho visa verificar as seguintes hipóteses de acordo com (YANG; ZHENG; HOSPEDALES, 2017).

1.1 Hipóteses

Se a rede neural artificial tem melhor desempenho na previsão de dados do que a fórmula paramétrica de Black-Scholes.

As redes neurais modulares superam as redes neurais de feedforward na precificação de opções.

Ao incluir opções virtuais, supera as redes neurais feedforward mais simples na tarefa de precificação de opções, cumprindo assim, as suposições de predição racional.

O poder de previsão das redes neurais fora da amostra depende do mecanismo de previsão de volatilidade usado e pode produzir resultados variados.

Além disso, um padrão estruturado quanto à escolha do tamanho e da arquitetura deve ser desenvolvido refletindo os resultados deste trabalho. A estrutura deste trabalho está dividida da seguinte forma: O capítulo **2** discute pesquisas anteriores feitas na área de precificação de opções usando redes neurais e apresenta a estrutura teórica tanto para a precificação de opções quanto para as redes neurais. O capítulo **3** apresenta a análise de dados, simulações numéricas e testes feitos para encontrar a escolha certa de parâmetros

e resultados finais. O trabalho termina com uma discussão dos resultados no capítulo 4 e sugere possíveis melhorias no capítulo 5.

2 Referencial Teórico

2.1 Precificação de opções

2.1.1 Modelos tradicionais de precificação

Os modelos de precificação de opções são minuciosamente investigados pelos pesquisadores de uma ampla gama de áreas. A popularidade do tema se dá não apenas pela relação com as bolsas de valores, mas também pelo fato de ocorrer na fronteira entre a pesquisa acadêmica e o uso prático. Para os acadêmicos, é uma ferramenta perfeita que avalia o que exatamente está acontecendo nos mercados financeiros, enquanto para os profissionais este tópico é considerado uma ferramenta para obter lucro.

Um dos artigos mais influentes escrito nesta área é o artigo de (BLACK; SCHOLES, 1973), que desenvolveu sua famosa fórmula que fornece uma estimativa do preço das opções europeias. Muitos outros pesquisadores tentaram melhorar este modelo. Entre eles, (SCOTT, 1987), que tentou otimizar as principais suposições feitas por Black & Scholes. Apesar de vários sucessos ao tentar otimizar o modelo de Black & Scholes, os modelos ainda exigiam suposições muito estritas que os tornavam geralmente malsucedidos. Uma boa revisão sobre os modelos paramétricos já existentes foi escrita por (BATES, 2003).

2.1.2 A fórmula Black-Scholes

O modelo de Black-Scholes foi desenvolvido para a precificação de qualquer derivativo dependente de uma ação sem dividendos. Esse modelo pressupõe que os preços das ações sigam um movimento browniano (EINSTEIN, 1956), em que as variações proporcionais dos preços de uma ação, em um curto período de tempo, possuem uma distribuição normal. Com isso, os preços de uma ação apresentariam uma distribuição lognormal.

A fórmula de Black-Scholes considera as seguintes hipóteses: o comportamento do preço da ação segue o modelo lognormal, com taxa de retorno esperada da ação e

volatilidade do preço da ação constantes; não existem custos operacionais nem impostos e todos os títulos são perfeitamente divisíveis; a ação não paga dividendos até o vencimento da opção; não há oportunidades de arbitragem sem risco; a negociação com títulos é contínua; é possível captar ou emprestar à mesma taxa de juro livre de risco; a taxa de juro livre de risco de curto prazo é constante.

A precificação de opções é realizada pela aplicação da fórmula de Black-Scholes, que é o resultado da seguinte equação "equação diferencial parcial parabólica"

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS \frac{\partial V}{\partial S} - rV = 0 \quad (2.1)$$

onde V é o preço do derivativo, S é o preço da ação, σ é o desvio padrão dos retornos das ações, r é a taxa de juros livre de risco e t é o tempo. A motivação para a solução desta equação está na minimização do risco nos mercados de opções, comprando ou vendendo o ativo-objeto exatamente na quantidade necessária para eliminar o risco, o que se denomina delta hedging continuamente revisado. Como a equação parabólica pode ser resolvida pela teoria das Equações Diferenciais Parciais (EDP), dadas as condições iniciais e de contorno específicas, as seguintes fórmulas de Black-Scholes são obtidas para a precificação de opções europeias de compra e venda

$$C(S, t) = N(d_1)S - N(d_2)Ke^{-rt} \quad (2.2)$$

$$P(S, t) = -N(-d_2)Xe^{rt} - N(-d_1)S \quad (2.3)$$

onde C é o preço de compra da opção, P é o preço de venda da opção, $N(\cdot)$ é a função de distribuição cumulativa da distribuição normal padrão, t é o tempo de expiração, X é o preço de strike (valor pago no ato do exercício da opção para aquisição do ativo-objeto previsto em contrato), S é o preço inicial do ativo e σ é a volatilidade do preço da ação.

$$d_1 = \frac{1}{\sigma\sqrt{t}} \left[\ln\left(\frac{S}{X}\right) + \left(r + \frac{\sigma^2}{2}\right)t \right] \quad (2.4)$$

$$d_2 = d_1 - \sigma\sqrt{t} \quad (2.5)$$

Este modelo é amplamente utilizado devido à sua simplicidade, porém na prática existem alguns pressupostos aos quais o modelo se limita. Primeiro, o modelo subestima a possibilidade de movimentos extremos do mercado. Em segundo lugar, supõe a execução instantânea e livre de negociações. Terceiro, o modelo assume também que o processo é estacionário, o que gera um risco de volatilidade. Por último, existe uma suposição de tempo contínuo e negociação contínua que geralmente não é satisfeita.

2.1.3 Gregas e Delta Hedging

O processo de precificação de opções pode ser analisado também pela análise das chamadas gregas. Em matemática financeira, é descrita como uma sensibilidade do preço das opções à mudança de algum parâmetro do qual o preço depende. Essas sensibilidades são usadas em particular na gestão de risco. Essas sensibilidades estão tanto no modelo de Black-Scholes quanto nos modelos descritos pelas redes neurais artificiais.

A grega mais importante é Delta definida como $\Delta = \frac{\partial V}{\partial S}$ onde as variáveis são as mesmas da fórmula de Black-Scholes. O Delta mede a taxa de variação do preço da opção com base em suas ações subjacentes. O coeficiente Δ_c para opções de compra está no intervalo $[0, 1]$, enquanto o coeficiente Δ_p para opções de venda está no intervalo $[-1, 0]$. Esses dois valores estão mutuamente interconectados e para o mesmo ativo subjacente com o mesmo preço de strike e prazo de validade, em que

$$\Delta_c - \Delta_p = 1 \quad (2.6)$$

que é chamada de paridade put-call. A paridade put-call requer menos suposições do que o modelo Black-Scholes, pois o mais importante é a existência de um contrato a termo. É uma ferramenta útil para fazer o hedge (mecanismo de proteção da carteira contra as oscilações do mercado) ter neutralidade delta. O portfólio delta neutro é aquele em que o valor do portfólio permanece inalterado no caso de uma pequena mudança no preço das ações subjacentes. O portfólio delta neutro é obtido se os componentes delta positivo e

negativo se compensam e se “neutralizam” (HAUG, 2000).

Gamma está relacionada a Delta e é definida por $\Gamma = \frac{\partial^2 V}{\partial S^2} = \frac{\partial \Delta}{\partial S}$. Gamma quer dizer qual a velocidade de mudança do Delta, ou seja, a sua aceleração. Quanto maior o Gamma, mais rápido o delta está evoluindo.

Vega definida como $\nu = \frac{\partial V}{\partial \sigma}$ mede a quantidade de ganho ou perda de dinheiro, em caso de um aumento ou uma diminuição da volatilidade. Outra grega é a Theta, que é definida como $\Theta = -\frac{\partial V}{\partial r}$ que é importante pois mede a sensibilidade do preço da opção conforme o tempo passa. Ao longo do tempo, caso o preço comece a se aproximar do preço de exercício, a opção irá valer mais e vice-versa. Se o preço não se aproximar do preço de exercício conforme vai se aproximando o vencimento, a opção terá de valer menos.

A última grega é a Rho, que é definida como $\rho = \frac{\partial V}{\partial r}$ se refere a mudanças na taxa de juros livre de risco (Selic no Brasil), já que mudanças nessa variável afetam o preço da opção. Tende a ter uma relevância menor já que as taxas de juros não tendem a sofrer alterações bruscas.

2.2 Mecanismos de previsão de volatilidade

Um dos parâmetros mais importantes no processo de precificação de opções é a volatilidade do ativo subjacente, devido à sua não observabilidade nos dados de preço, enquanto outros parâmetros, como preço das ações, preço de exercício, tempo até o vencimento e a taxa de juros, normalmente são observáveis diretamente. Por ser uma medida de como o preço de um ativo oscila em torno de sua média, é importante usar o valor correto na previsão do preço. Isso influencia significativamente o valor da opção no tempo, uma vez que as expectativas de alta volatilidade no futuro geralmente tendem a aumentar os preços das opções. Existem várias maneiras de prever a volatilidade, que veremos à seguir.

2.2.1 Volatilidade Histórica

Para estimar a volatilidade este método depende do uso de dados históricos de preços de ações em uma determinada frequência de tempo. É obtido como o desvio padrão

do retorno da ação subjacente ao longo de algum período. O retorno das ações é calculado na escala logarítmica dos preços de fechamento

$$r_t = \ln \left(\frac{S_t}{S_{t-1}} \right), \quad (2.7)$$

onde S_t é o preço da ação subjacente no período t e r_t é o retorno entre os períodos $t-1$ e t . A volatilidade histórica é então obtida pela fórmula estatística padrão para o segundo momento estatístico, a variância, e tomando sua raiz quadrada

$$\sigma_{HV} = \sqrt{\frac{1}{T-1} \sum_{t=1}^T T r_t - r}, \quad (2.8)$$

onde σ_{HV} denota o estimador de volatilidade histórica, T é o número de períodos durante os quais a volatilidade é medida, r é o valor médio dos retornos das ações r_t . A escolha do intervalo de tempo durante o qual a volatilidade é medida é questionável e há uma regra de ouro para usar T como um número de dias ao qual a estimativa de volatilidade será aplicada. Também é possível calcular a volatilidade histórica a partir do desvio padrão da média móvel dos retornos, que trata melhor o valor alterado do valor médio dos retornos.

Nos casos em que a volatilidade histórica σ_{HV} é maior, ela decorre do período instável anterior com maior amplitude nos movimentos dos retornos dos preços. Por outro lado, os períodos calmos de negociação estável ou baixa atividade no mercado podem ser caracterizados pelo baixo valor do estimador de volatilidade histórica σ_{HV} .

A estimativa de volatilidade histórica é freqüentemente usada para a comparação transversal entre várias ações ou uma ação específica para todo o mercado geral (ou possivelmente alguma parte dele). Também é possível comparar duas volatilidades históricas em diferentes períodos de tempo, o que fornece informações se a volatilidade mudou recentemente de forma positiva ou negativa.

2.2.2 Volatilidade Implícita

Na precificação de opções, o mecanismo de previsão de volatilidade comumente usado é a medida de volatilidade implícita. A sua aplicação surge a partir do modelo Black-Scholes, uma vez que utiliza muitas informações de mercado, incluindo a previsão

do preço da opção, e tenta explicar as expectativas do mercado incluindo a volatilidade futura. A medida de volatilidade implícita, então, calcula a última variável desconhecida restante, que é a volatilidade da opção. Da mesma forma que a volatilidade histórica, para encontrar o tamanho relativo da volatilidade, a medida deve ser aplicada para vários períodos no passado, a fim de avaliar as expectativas do mercado.

2.2.3 Volatility smile

Na maioria dos casos, a volatilidade implícita calculada a partir dos preços das opções no mercado constitui um padrão persistente que varia de acordo com o tempo restante até ao vencimento e o seu preço de exercício, enquanto o modelo Black-Scholes assume uma volatilidade constante. Esse padrão também é conhecido como “volatility smile”, em que o nome deriva do formato da função de volatilidade implícita, dados os preços de exercício, uma vez que tem formato convexo com um mínimo global.

O “volatility smile” é um fenômeno empírico que aumenta, por exemplo, nos mercados de opções de moeda e implica que as opções deep out-of-the-money e deep in-the-money exibem preços mais elevados do que os valores teóricos dados pelo modelo Black-Scholes. Nos mercados de ações e índices de ações, a volatilidade implícita geralmente tende a ter uma relação negativa com o preço de exercício, o que significa que, com o aumento do preço de exercício, a volatilidade implícita diminui, este fenômeno é chamado de inclinação da volatilidade e decorre do uso de alavancagem. Durante a queda do preço do ativo subjacente, a alavancagem aumenta e, por sua vez, a volatilidade também aumenta, razão pela qual a probabilidade de que o preço diminua é maior. Isso se aplica da mesma forma também vice-versa.

2.2.4 Previsão de volatilidade com os modelos ARCH

Conforme mencionado em (FIGLEWSKI, 1997), existem muitas razões pelas quais a previsão clássica a partir da volatilidade histórica não é correta. Os problemas decorrem da correlação serial dos retornos, da volatilidade variável no tempo, da estimativa ruidosa da média e também do fato de os retornos não serem normalmente distribuídos. A volatilidade variável com o tempo pode ser tratada pelos modelos que

envolvem a previsão da volatilidade ao longo do tempo.

O modelo básico que assume a volatilidade invariante no tempo pode ser especificado da seguinte forma

$$r_t = E(r_t) + \epsilon_t \quad (2.9)$$

$$E(\epsilon) = 0 \quad (2.10)$$

$$Var(\epsilon) = \sigma_t^2 \quad (2.11)$$

onde o modelo mais simples assumiria variância constante

$$\sigma_t^2 = C. \quad (2.12)$$

Em que r_t é logaritmo do retorno do ativo financeiro em t , E é a esperança, ϵ_t representa os termos de erro e σ_t^2 é a variância condicional de r_t

Em vez deste modelo simples, a volatilidade poderia ser calculada pelo modelo Autoregressive Conditional Heteroskedasticity (ARCH), que lida com a volatilidade variável no tempo de forma bastante eficaz. A variação no período t é dada por

$$\sigma_t^2 = \beta_0 + \beta_1 \epsilon_{t-1}^2 \quad (2.13)$$

para o modelo ARCH mais simples que está refletindo um atraso dos erros na estimativa da variância. Se adicionarmos mais defasagens, obteríamos o modelo ARCH (q) com $q \in \mathbb{N}$. O modelo é dado por

$$\sigma_t^2 = \beta_0 + \beta_1 \epsilon_{t-1}^2 + \dots + \beta_q \epsilon_{t-q}^2 \quad (2.14)$$

A fim de atingir a estabilidade numérica do processo estocástico, ele não deve ser raiz unitária e, portanto, a soma dos coeficientes deve ser menor que um

$$\sum_{i=1}^q \beta_i \leq 1 \quad (2.15)$$

Em aplicações práticas, o parâmetro q é geralmente muito baixo, pois um número maior de parâmetros diminui a precisão da previsão fora da amostra. Efeitos de maior duração também são melhor descritos pelos modelos GARCH (Generalized ARCH) que refletem nas variáveis dependentes também a variância dos períodos anteriores. O modelo GARCH mais simples e amplamente utilizado é o GARCH (1,1) e é fornecido pela seguinte especificação

$$\sigma_t^2 = w + \alpha_1 \sigma_{t-1}^2 + \beta_1 \epsilon_{t-1}^2 \quad (2.16)$$

onde w , α_1 e β_1 são os parâmetros de estimação e ϵ_{t-1} é o termo residual da regressão (2.31). Este modelo tem novamente possível generalização para GARCH (p, q) dado por

$$\sigma_t^2 = w + \alpha_1 \sigma_{t-1}^2 + \dots + \alpha_p \sigma_{t-p}^2 + \beta_1 \epsilon_{t-1}^2 + \dots + \beta_q \epsilon_{t-p}^2 \quad (2.17)$$

onde todo α_i e β_j para todo i e j também são coeficientes de regressão. O estado estacionário de longo prazo no caso de GARCH (1,1) é dado por

$$\sigma_{LR}^2 = \frac{w}{1 - \alpha_1 - \beta_1} \quad (2.18)$$

já que o valor esperado de ϵ^2 é σ^2 e, portanto, o modelo geralmente converge para o valor de estado estacionário (se os coeficientes somam menos de 1). Os modelos GARCH, entretanto, apresentam uma forte deficiência dada pelo fato de serem difíceis de ajustar e também restringem o impacto do choque para ser independente de seu sinal (FIGLEWSKI, 1997).

Entende-se que em alguns mercados os choques não são simétricos, visto que os traders geralmente reagem de forma diferente em mercados altistas e baixistas. Este problema foi resolvido por (NELSON, 1991) que modelou um logaritmo de variâncias precedentes no chamado EGARCH (Exponential GARCH). Embora neste trabalho este modelo não seja utilizado, vale ressaltar que ele estende a aplicação dos modelos GARCH,

pois permite o tratamento assimétrico para movimentos positivos e negativos, bem como pode diminuir a negatividade da variável dependente, o que poderia ser uma melhoria da abordagem utilizada neste trabalho. A especificação do modelo é dada por

$$\log\sigma_t^2 = w + \alpha_1 \log\sigma_{t-1}^2 + \cdots + \alpha_p \log\sigma_{t-p}^2 + \beta_1 g(\epsilon_{t-1}) + \cdots + \beta_q g(\epsilon_{t-q}) \quad (2.19)$$

onde $g(\epsilon_t) = \theta_{\epsilon_t} + \lambda(\epsilon_t - E(\epsilon_t))$ e θ e λ são os coeficientes de estimação. A forma funcional da função $g(\epsilon_t)$ permite que a magnitude e o sinal tenham diferentes efeitos sobre a volatilidade, o que é útil na estimativa de preços (FIGLEWSKI, 1997). Outra possibilidade é usar TARCH (Threshold ARCH) modelo desenvolvido por (ZAKOIAN, 1994) que usa diferentes coeficientes para diferentes sinais de movimentos de preços com o modelo especificado por

$$\sigma_t^2 = w + \alpha\sigma_{t-1}^2 + \beta_1\epsilon_{t-1}^2 + \gamma_1\epsilon_{t-1}^2 I_{t-1} \quad (2.20)$$

em que $I_{t-1} = 0$ se $\epsilon_{t-1} \geq 0$ e $I_{t-1} = 1$ se $\epsilon_{t-1} < 0$

Os modelos da família ARCH são amplamente usados em economia e finanças e o foco principal é a explicação da variância dentro da amostra ao invés da previsão fora da amostra (FIGLEWSKI, 1997). O modelo é ajustado realizando estimativa de máxima verossimilhança para os termos ϵ_t com uso frequente da distribuição normal para os preços, mas às vezes a distribuição t é usada para permitir o efeito de caudas grandes. No entanto, a suposição de lognormalidade do modelo não se ajusta à situação real o suficiente por causa das grandes mudanças de preços muito frequentes, que muitas vezes são explicadas pelas grandes caudas na distribuição dos retornos (FIGLEWSKI, 1997).

O problema com os modelos ARCH geralmente é fazer a estimativa, pois os modelos requerem grandes conjuntos de dados para treinar, caso contrário, eles não se comportam bem o suficiente. Outro problema com os modelos ARCH pode ser que pode haver um choque de longa duração, que não pode estar envolvido no modelo, pois diminuiria o poder de explicação do modelo. Além disso, os coeficientes estimados econometricamente podem se tornar negativos, razão pela qual a simplicidade do modelo é, neste caso, favorecida. Por último, os modelos são projetados para prever a variação para um ou alguns

passos à frente, em vez de uma previsão de longo prazo, o que os torna inadequados para previsões de longo prazo (FIGLEWSKI, 1997).

2.3 Inteligência artificial

2.3.1 Neurônios Biológicos

Redes neurais artificiais são entidades de processamento de dados que possuem processamento de informações e funções semelhantes às redes neurais biológicas, cérebros humanos ou animais. O uso de estruturas semelhantes ao cérebro no cálculo de vários problemas sempre foi muito procurado devido à sua adaptabilidade e capacidade de resolução de problemas. Os sistemas neurais biológicos são emaranhados naturalmente envolvidos de neurônios, células cerebrais, capazes de transferir o impulso através da rede de neurônios interconectados mutuamente. Um neurônio biológico consiste em uma célula com muitas pequenas partes chamadas dendritos que recebem os sinais de outros neurônios. O sinal coletado passa por um axônio que é outra parte do corpo do neurônio e no final do axônio existem conexões chamadas sinapses que conectam os axônios a dendritos de outros neurônios (AZAM, 2007).

O cérebro é geralmente considerado uma caixa preta com muitas funcionalidades não totalmente compreendidas. Ele obtém sinais de entrada e retorna uma resposta de saída que, como um subproduto, fortalece as ligações entre neurônios específicos que eram usados com frequência. O cérebro geralmente decide cumprir objetivos específicos pela maneira mais provável de como realizar a tarefa dada, o que erroneamente dá uma intuição de que o cérebro funciona como um todo (AZAM, 2007).

Há evidências da neuropsicologia e da neurobiologia de que algumas partes do cérebro são independentes de outra e que algumas funções cerebrais específicas podem ser realizadas mesmo sem estar conectadas a outras partes. Os neurobiologistas há muito pensam que as partes específicas do cérebro são diferenciadas em módulos distintos, cooperando mutuamente entre si e os avanços na neurobiologia confirmam esses pensamentos e apontam para o fato de que os cérebros humano e animal são constituídos por diferentes módulos com diferentes composições fisiológicas que são responsáveis por várias tarefas.

2.3.2 Redes neurais como ferramenta de aproximação não linear

Para prever os preços das opções, (HUTCHINSON; LO; POGGIO, 1994) usaram pioneiramente uma abordagem de Rede Neural Artificial (RNA). Modelos de RNA são naturalmente capazes de identificar a relação empírica entre dados de entrada e saída sem conhecimento prévio sobre a relação funcional de forma fechada, o que a torna uma ferramenta útil para a tarefa de precificação de opções.

O uso de modelos de redes neurais também é chamado de abordagem não paramétrica e apresenta várias vantagens em comparação aos modelos paramétricos (Black-Scholes etc.). Na abordagem não paramétrica, não há suposições como lognormalidade dos retornos ou continuidade do tempo de negociação. Além disso, os modelos são adaptativos e podem reagir imediatamente a choques inesperados, alterando a forma funcional. Contudo, existe um problema com esta abordagem que é a sua ineficácia nos casos em que não há dados suficientes. A RNA deve ser treinada em conjuntos de dados muito grandes e, portanto, requer um grande banco de dados de preços históricos. A abordagem, portanto, não é muito boa no caso de opções pouco negociadas ou outros derivativos.

As RNAs são modelos estatísticos distribuídos paralelamente que processam as informações do conjunto de dados de entrada e generalizam as relações entre os dados para prever as novas situações. O uso de RNAs para os dados de série temporal não é vinculativo, pois pode ser usado também para aplicações transversais. O uso de RNAs como um método não paramétrico é uma forma bastante recente de prever relações funcionais não lineares desconhecidas. (HORNIK, 1991) mostrou que as redes feedforward podem aproximar qualquer função arbitrária com uma precisão arbitrária, que abriu a porta para todas as diversas aplicações para modelagem não linear.

2.3.3 Perceptron e redes neurais feedforward

Perceptron é a forma mais básica de rede neural. Consiste em apenas um elemento que classifica se a entrada pertence ou não à classe de destino. Ele usa a seguinte função

$$\phi(x) = \begin{cases} 1, & \text{se } w \cdot x + b > 0 \\ 0, & \text{caso contrário} \end{cases} \quad (2.21)$$

onde w é o vetor de pesos, x são os valores de entrada e b é a tendência que desloca o limite de decisão, que permite ao perceptron formar várias saídas, $w \cdot x$ representa o produto escalar de dois vetores. À medida que a função assume os valores 0 ou 1, o perceptron é um neurônio artificial que usa a função de Heaviside como sua função de ativação.

Uma rede feedforward é uma rede neural, onde as conexões entre os elementos da rede são organizadas em camadas separadas e cada camada de neurônios pode influenciar apenas sua saída na próxima camada, enquanto a camada da qual extrai os dados de entrada permanece fixa. Este procedimento garante a transferência unilateral de dados, razão pela qual a rede neural é chamada de feedforward.

As camadas entre as camadas de entrada e saída são chamadas de camadas ocultas e cada camada oculta pode ter qualquer número de neurônios (nós). O número de camadas ocultas depende da natureza do problema de aproximação e existem várias técnicas empíricas para escolhê-lo. A rede feedforward é caracterizada pela inexistência de quaisquer ciclos ou loops fechados na rede. Em comparação com o perceptron, a rede feedforward pode ter uma saída contínua onde a função matemática que processa as entradas é chamada de função de ativação e é comumente considerada como a função logística (ou sigmóide) devido à existência de sua derivada na forma fechada

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2.22)$$

Como um dos objetivos do aprendizado de máquina é encontrar a melhor forma funcional possível de entradas, fornecendo previsões para outras situações, o processo de aprendizado durante a avaliação dos dados é importante. Em redes neurais multicamadas, o algoritmo comumente usado é a retropropagação, que calcula a contribuição de erro depois que os dados são processados. O algoritmo de retropropagação então calcula as diferenças entre o destino do padrão de treinamento e o valor real. Com base nas diferenças, os pesos são atualizados para encontrar o gradiente do peso que é conseqüentemente dividido e sua fração é subtraída. Este problema de otimização resulta após várias etapas em um estado estacionário, onde os pesos em camadas específicas são fixados. A abordagem mais comum para otimizar esses pesos é usar um esquema de otimização de gradiente descendente.

2.3.4 Funções de Ativação

Ao construir uma rede feedforward, muitas vezes é questionável não apenas como escolher o tamanho e a arquitetura da rede, mas também quais devem ser as funções de ativação entre as camadas individuais. A suposição que deve ser feita na escolha da função de ativação é a continuidade da função de ativação e a existência de derivadas em todos os pontos. Mesmo que o processo feito pelo neurônio biológico esteja muito próximo da definição matemática da função de ativação sigmóide, também existem outras possibilidades de como escolher a função de ativação no caso de camadas mais ocultas que podem em alguma aplicação funcionar melhor. As funções a seguir representam todas as funções de ativação usadas e comparadas neste trabalho e serão denotadas por ϕ_i , onde i é um coeficiente que determina a forma funcional (GLOROT; BORGES; BENGIO, 2011)

$$\phi_1(x) = \frac{\exp(x)}{1 + \exp(x)} = \frac{1}{1 + e^{-x}}, \quad (2.23)$$

$$\phi_2(x) = \ln(1 + \exp(x)), \quad (2.24)$$

$$\phi_3(x) = \begin{cases} 1, & \text{se } x > 0 \\ 0, & \text{caso contrário,} \end{cases} \quad (2.25)$$

$$\phi_4(x) = \begin{cases} 0,01, & \text{se } x > 0 \\ 0, & \text{caso contrário,} \end{cases} \quad (2.26)$$

$$\phi_5(x) = \begin{cases} 1, & \text{se } x > 0 \\ \exp(x) - 1, & \text{caso contrário.} \end{cases} \quad (2.27)$$

A função (2.23) é a função sigmoideal (ou logística) $\phi_1(x)$ e ela converge para 1 (transferência de sinal completo) conforme x se aproxima de $+\infty$ e 0 (sem transferência de sinal) conforme x se aproxima de $-\infty$. Devido a essas propriedades de convergência, sua derivada simples e sua simplicidade, é muito utilizado por muitos pesquisadores.

A segunda função (2.24) é a função softplus que oferece uma propriedade inte-

ressante para a parte positiva da entrada, pois ela cresce assintoticamente linearmente à medida que x se aproxima de $+\infty$ e se comporta como uma função exponencial para a entrada negativa. É também uma aproximação suave da próxima função chamada Unidade Linear Retificada (ULR). Esta função é denotada (2.25) e ignora completamente a entrada com sinal negativo, enquanto a entrada positiva é tratada linearmente. A vantagem da função de ativação ULR é que ela separa rapidamente a entrada e funciona bem como um filtro.

A próxima função de ativação (2.26) é chamada de função de ativação Leaky ULR, que é quase idêntica à função ULR anterior, mas também uma pequena fração de entrada negativa é transmissível através do neurônio. A última função de ativação utilizada neste trabalho (2.27) é a função de ativação da Unidade Linear Exponencial (ULE) que torna as ativações médias mais próximas de zero o que acelera o aprendizado. (GLOROT; BORGES; BENGIO, 2011)

2.4 Redes neurais modulares

A estrutura da rede neural modular é de extrema importância. Ele decompõe problemas complexos em módulos menores que interagem entre si. A fim de trazê-lo mais para o contexto de precificação de opções, de acordo com (JAND; SUN; MIZUTANI, 1997) a função de ativação foi especificada para a arquitetura de rede neural modular como uma média ponderada de saídas de módulos individuais

$$\varphi(x_1, x_2) = w_1 c_{1t}(x_1, x_2) + \dots + w_M c_M(x_1, x_2) = \sum_{k=1}^M w_k c_k(x_1, x_2), \quad (2.28)$$

onde M denota o número de módulos que são aproximados pelas funções c_1, \dots, c_M e seus pesos w_1, \dots, w_M . Os pesos w_k podem ser estimados por vários métodos. (MCCULLAGH; NELDER, 1989) e (BRIDLE, 1990) propuseram uma função de ativação softmax que eles definiram com pesos dados como

$$w_k = \frac{\exp c_k}{\sum_{j=1}^M \exp(c_j)} \quad (2.29)$$

$$\sum_{k=1}^M w_k = 1 \quad (2.30)$$

Outra possibilidade é utilizar modelos de mistura gaussiana, que é uma forma de caracterizar cada módulo por distribuição gaussiana paramétrica. O modelo é então considerado uma mistura dessas distribuições. Os modelos de mistura gaussiana também podem ser considerados como redes neurais com portas que passam sua saída para portas de multiplicação, onde todos os módulos usados são ponderados.

2.4.1 Módulos de rede neural modular

Este trabalho, no entanto, segue uma abordagem ligeiramente diferente da que é mostrada inicialmente por (GRADOJEVIC; GENCAJ; KUKOLJ, 2009) que escolhe os módulos específicos usando parâmetros de opção como *moneyness*(liquidez) ($\frac{S}{X}$) e tempo até o vencimento t . Em tal configuração, também é uma boa abordagem dividir as opções em grupos diferentes de acordo com os parâmetros nomeados e alternar os módulos da rede neural de acordo com o subconjunto de onde os dados de entrada vêm.

Essa técnica garante um tratamento melhor dos dados de entrada do que a ponderação padrão de módulos específicos, conforme mostrado na equação (2.16). Os módulos particulares são considerados como redes neurais feedforward e a saída geral da rede modular é estimada pela seguinte equação no caso de uma camada oculta

$$c_k = \phi_k(\beta_{k0} + \sum_{j=1}^{q_k} \beta_{kj} \psi_k(\alpha_{kj0} + \sum_{i=1}^s \alpha_{kij} x_i)), k = 1, \dots, M \quad (2.31)$$

em que, q_k denota o número de nós ocultos no k -ésimo módulo, α_{kij} e β_{kj} denotam os pesos de conexão entre as camadas adjacentes para os módulos e os subscritos iguais a zero denotam os vieses da rede neural. ϕ_k e ψ_k são funções de ativação que dão flexibilidade ao modelo visto que os tipos desta função são considerados funções sigmóides (logísticas) para esta arquitetura modular. Esta opção de função de ativação pode lidar com quaisquer funções de entrada, conforme declarado na configuração de feedforward simples.

A fórmula de precificação de opções é a mesma que em (GRADOJEVIC; GENCAJ; KUKOLJ, 2009)

$$C = \phi(S, X, \tau) \quad (2.32)$$

onde C é o preço da opção de compra, S é o preço do ativo subjacente, X é o preço de exercício e τ é o prazo de vencimento. Da mesma forma como em (GRADOJEVIC; GENCAJ; KUKOLJ, 2009), pode-se supor homogeneidade de grau um em termos de preços, o que significa que podemos reorganizar a função normalizando a equação por X

$$\frac{C}{X} = \phi\left(\frac{S}{X}, 1, \tau\right) \implies c = \phi(s, \tau) \quad (2.33)$$

onde $s := \frac{S}{X}$ e $c := \frac{C}{X}$.

Isso é particularmente útil neste trabalho, pois essa normalização é muito útil em termos de aumento do conjunto de dados e, portanto, uma simplificação da inferência. É questionável se esta suposição é completamente válida, já que pode haver razões psicológicas no mercado que tornam duas situações com as mesmas variáveis normalizadas completamente diferentes.

A arquitetura de rede neural modular usada divide as opções de acordo com o moneyness (3 grupos) e o tempo para o vencimento (3 grupos) que constituem no total 9 categorias, portanto, o número de módulos treinados é exatamente o número dessas categorias. Esses módulos são então preenchidos com o moneyness $m = \frac{S}{X}$, tempo até o vencimento τ e volatilidade σ e, com base nas duas primeiras variáveis, o módulo ótimo correspondente ao moneyness dado e o tempo até o vencimento são escolhidos. A visão esquemática deste processo está representada na figura 2.1, onde uma das 9 redes neurais calcula a saída de acordo as entradas.

2.4.2 Outras arquiteturas modulares de rede neural

As arquiteturas de redes neurais modulares foram bastante investigadas por muitos pesquisadores no campo da ciência da computação, no entanto, no campo de precificação de opções não foi explicado completamente qual arquitetura escolher e por quê. A fim de cumprir o requisito de ter módulos distintos que operem para toda a computação de forma distinta, apenas algumas arquiteturas correspondem à modularidade necessária.

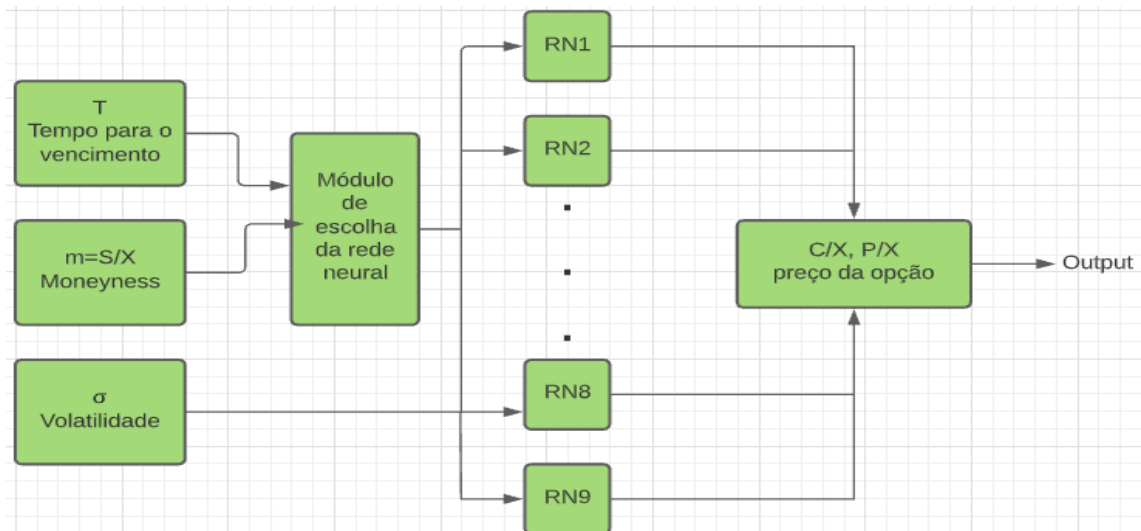


Figura 2.1: Rede modular escolhendo módulos de acordo com moneyness m e tempo até o vencimento T

(AZAM, 2007)

Uma das primeiras arquiteturas sugeridas, chamada de módulos desacoplados, usa a aprendizagem não supervisionada e supervisionada em estágios processados sequencialmente. O primeiro estágio compreende uma rede de Teoria Ressonante Adaptativa (TRA) que categoriza os dados de entrada em vários clusters sem supervisão. Na segunda parte, os modelos são treinados separadamente com o supervisor de acordo com os clusters criados na primeira etapa. A saída final é dada pela saída da função de ativação máxima dos módulos separados.

Outra opção é usar o chamado modelo de “other-output”, ela difere da arquitetura anterior apenas pelo fato de a saída ser apenas binária, o que controla se a entrada é classificada em uma determinada região do espaço de entrada.

A próxima arquitetura de rede neural modular é chamada de rede hierárquica, que é bastante semelhante aos modelos anteriores, pois tem dois níveis de hierarquia, ambos com supervisor, que cria clusters no primeiro nível de hierarquia, e o segundo nível hierárquico, faz processos de aprendizagem e classificação pelos módulos especialistas.

Entre várias outras extensões ou alternativas aos modelos mencionados, a arquitetura de rede modular mais amplamente aceita é chamada de “adaptive mixture of local experts” (JACOBS; CHISTOFFERSEN, 2004). Esta arquitetura é bastante simples e possui uma estrutura intuitiva por se aproximar da analogia biológica. Ele se baseia

na teoria das misturas gaussianas e nos algoritmos de aprendizagem local, que tentam ajustar localmente a capacidade às propriedades do sistema de treinamento. Este método também é usado por (YANG; ZHENG; HOSPEDALES, 2017) que é de uma forma ligeiramente diferente seguida neste trabalho por não utilizar-se a rede modular com a mesma arquitetura.

A última arquitetura de rede neural modular apresentada é chamada de “hierarchical mixture of experts”, introduzida por (JORDAN; JACOBS, 1994). Ela tem potencial para ser amplamente utilizada para fins de classificação e regressão, como nos casos de redes neurais artificiais multicamadas feedforward. A arquitetura é funcionalmente derivada dos modelos de “dividir” e “conquistar” e é capaz de realizar tarefas como árvores de classificação e regressão junto com splines de regressão adaptativa multivariada. Esses algoritmos de ajuste dividem o espaço de entrada em tais sequências aninhadas de regiões que mais tarde ajudam a classificar e regredir os dados (AZAM, 2007).

A “hierarchical mixture of experts” é composta por dois tipos de módulos, redes neurais especializadas e redes de portas, onde as redes de especialistas estão no nível inferior da hierarquia e dividem o problema em partes menores para as quais as soluções podem ser encontradas. As sub-soluções específicas são então unidas pelas chamadas redes de portas e formam a solução final. A diferença entre a rede neural artificial de portas e especialista é que as redes de portas transformam os valores de entrada em níveis de confiança para as saídas, enquanto as redes especializadas usam os dados para encontrar a estimativa do valor de saída. (AZAM, 2007).

Como mostrado acima, existe uma ampla gama de diferentes redes neurais modulares que podem ser utilizadas em várias aplicações, incluindo precificação de opções e, como tal, este trabalho é capaz de cobrir apenas a rede neural modular simples como feito em (GRADOJEVIC; GENCAJ; KUKOLJ, 2009).

2.5 Suposições de Predição Racional

2.5.1 Motivação

Para obter o melhor modelo de rede neural possível, levanta-se naturalmente a questão de saber se pode haver alguma informação extra, que melhoraria o desempenho do modelo. Um modelo de rede neural geral, independentemente de sua arquitetura, depende puramente dos dados de entrada fornecidos e, portanto, emprega a vantagem de um teorema de aproximação universal. Este teorema diz que a rede neural com uma camada oculta pode se aproximar de uma função contínua arbitrária em subconjuntos compactos de números reais, dado que a função de ativação da rede neural é contínua e diferenciável. (AZAM, 2007)

Embora haja um forte suporte para o uso de redes neurais simples devido a esse teorema, há uma necessidade de um conjunto de dados de treinamento grande, para garantir que o procedimento de treinamento converta rapidamente e a função possa ser ensinada. Além disso, na precificação de opções, não há garantia de que a função, que generalizaria as condições de mercado arbitrárias, ainda exista, uma vez que o preço da opção depende do comportamento de compradores e vendedores nos mercados que é limitado por muitos fatores, como taxas de negociação, não continuidade da negociação, oferta limitada de fundos dos traders, fenômenos psicológicos, como mudança de comportamento dos traders, etc.

Portanto, este trabalho segue e investiga o método usado por (YANG; ZHENG; HOSPEDALES, 2017) que aplicou seis condições que qualquer modelo de precificação de opções significativo deve satisfazer. No entanto, este método é usado para uma nova escolha de arquitetura e, portanto, a verificação da racionalidade teve que ser derivada de uma forma completamente nova na subseção 2.5.3.

2.5.2 Requisitos para previsões racionais

Algumas condições são definidas para atender aos pressupostos básicos sobre o comportamento dos negociantes nos mercados de opções. Se essas condições não fossem atendidas e o instrumento de mercado tivesse um volume de negociação suficiente, a regra

de não arbitragem levaria imediatamente a uma correção de mercado que atenderia às premissas.

O preço da opção de compra é determinado matematicamente com precisão para qualquer situação de mercado, seguindo a definição de (YANG; ZHENG; HOSPEDALES, 2017) como

$$C(S, X, \tau) = e^{-r\tau} \int_0^{\infty} \max(0, S_T - X) f(S_T | S_t, \tau) dS_T, \quad (2.34)$$

em que r é a taxa livre de risco, o preço da opção no momento do vencimento T é estimado no tempo t com a função de densidade de probabilidade neutra ao risco condicional $f(x | S_t, \tau)$ e $\tau = T - t$ é a diferença de tempo. Isso pode ser entendido muito facilmente como uma soma descontada de todos os fluxos de lucro possíveis de uma opção ponderada por sua probabilidade. A integral se combina ao preço da ação subjacente no momento do vencimento, que pode ser arbitrariamente grande, porém, com probabilidade insignificante de preços de ações irrealistas. Denotamos o preço da opção C , mas a taxa livre de risco e o prazo de desconto são obtidos independentemente, portanto, estimamos pela rede neural apenas o seguinte preço C sem o fator de desconto

$$C(S, X, \tau) = \int_0^{\infty} \max(0, S_T - X) f(S_T | S_t, \tau) dS_T \quad (2.35)$$

Portanto, para opções de compra, as seguintes seis condições de Yang et al. (2017) tem que se manter para ter suposições de previsão racionais:

A primeira condição garante um preço crescente em relação ao preço de exercício, uma vez que todos os outros parâmetros são fixos. Isso parece razoável, pois com o aumento do preço de exercício X , a diferença do preço do ativo subjacente S só pode diminuir no caso da opção de compra “in the money”, onde o preço da opção é a diferença $S - X$. Matematicamente, se tomarmos uma derivada de (2.23), obtemos

$$\frac{\partial C}{\partial X} = \int_0^X f(S_T | S_t, \tau) dS_T - 1 \quad (2.36)$$

que na verdade é uma função de distribuição cumulativa $P(S_T \leq X)$ que não pode exceder 1 devido às suas propriedades estatísticas.

$$\frac{\partial C}{\partial X} \leq 0. \quad (2.37)$$

A segunda condição garante a convexidade do preço da opção em relação ao preço de exercício

$$\frac{\partial^2 C}{\partial X^2} \geq 0, \quad (2.38)$$

A terceira condição garante que quanto maior o tempo de vencimento, maior a probabilidade de ultrapassar o preço de exercício e, portanto, o preço da opção não deve diminuir

$$\frac{\partial C}{\partial \tau} \geq 0, \quad (2.39)$$

A quarta condição diz que, quando o preço de exercício é alto o suficiente, há uma probabilidade zero de obter qualquer lucro com a opção e, portanto, o preço da opção deve ser zero

$$\lim_{X \rightarrow +\infty} C(S_t, X, \tau) = C(S_t, +\infty, \tau) = 0. \quad (2.40)$$

A quinta condição trata do preço da opção no momento do exercício, neste momento o preço deve corresponder exatamente ao seu valor teórico, que é para opções típicas de função de rampa

$$C(S, X, \tau = 0) = \max(0, S_t - X). \quad (2.41)$$

Por último, a sexta condição fornece um limite inferior e superior para o preço da opção, uma vez que não se pode exceder o preço da ação subjacente sem a violação da regra de não arbitragem, caso contrário, um investidor faria uma arbitragem comprando o ativo subjacente e vendendo a opção de compra simultaneamente e fechando ambos no momento do vencimento da opção

$$\max(0, S_t - X) \leq C(S_t, X, \tau) \leq S_t. \quad (2.42)$$

Para que essas condições sejam válidas, (YANG; ZHENG; HOSPEDALES, 2017) também declara algumas suposições. É necessário assumir a função de preço da opção C seja duas vezes diferenciável em relação a X e uma vez diferenciável em relação a τ .

2.5.3 Verificando a racionalidade

Opondo-se a (YANG; ZHENG; HOSPEDALES, 2017) este trabalho usa um modelo diferente para a rede neural, pois eles usam uma rede neural fechada complicada que trata as entradas individualmente com diferentes funções de ativação. Para ser compatível com os pressupostos de racionalidade, os autores usam a função sigmóide $\phi(x) = \frac{e^x}{e^x + 1}$ para o moneyness inverso $\frac{1}{m}$ e função softplus $\phi_2(x) = \log(1 + e^x)$ para tempo até vencimento τ . O modelo deles combina ambas as funções de ativação em cada etapa e como este trabalho usa arquiteturas completamente diferentes, esta verificação das condições de racionalidade não pode ser usada e deve ser modificada para corresponder aos modelos usados neste trabalho.

Na literatura de precificação de opções, a função sigmóide é frequentemente usada para a função de ativação da rede neural, visto que é a escolha mais previsível no aprendizado de máquina. No entanto, poucos artigos de pesquisa sobre precificação de opções usam a função de ativação softplus, que tem uma vantagem clara: seus valores devem ser positivos. Esta é uma propriedade muito desejável para a camada de saída da rede neural, pois o preço da opção deve ter um valor positivo naturalmente. Portanto, neste trabalho é proposto uma arquitetura totalmente nova e original para ter uma função arbitrária em camadas ocultas (sigmóide) e função softplus como saída.

Pode-se argumentar que esta é a mesma escolha que (YANG; ZHENG; HOSPEDALES, 2017) fizeram, no entanto, há uma diferença, pois neste trabalho a função sigmóide foi usada em camadas ocultas e a função softplus na camada de saída, ao passo que eles usaram em sua rede de camada única e ambas as funções combinadas.

A rede neural com uma camada oculta, respeitando a escolha de funções de

ativação, teria o preço da opção de compra da seguinte forma

$$c = \frac{C}{X} = \phi_2(a + \sum_j w_j \phi_1(b_j + \sum_i w_{(ij)x_i})) \quad (2.43)$$

onde x_i representa as variáveis de entrada, como moneyness m , tempo até vencimento τ e volatilidade σ , a e b_j são os termos de erro da rede neural, w_j e $w_{(ij)}$ são os pesos para cada neurônio i e j . No caso de mais de uma camada oculta, as funções seriam aninhadas, o que não mudaria a inferência, uma vez que todas as suposições se manterão trivialmente usando a regra da cadeia para obter as derivadas e devido a uma boa escolha da função de ativação, a derivada de um número arbitrário de funções não muda o sinal do resultado.

Para verificar a primeira suposição, considera-se

$$\frac{\partial C}{\partial X} = \frac{\partial C}{\partial c} \frac{\partial c}{\partial X} = X \frac{\partial c}{\partial m} \frac{\partial m}{\partial X} = X \sum_j w_j \phi_1(\cdot) w_{(ij)} \phi_2(\cdot) \left(-\frac{S}{X^2}\right) \leq 0 \quad (2.44)$$

em que é usada a regra da cadeia e (\cdot) representa a função interna para tornar a equação mais simples de ler e i representa o índice de neurônio usado para a entrada moneyness. As derivadas das funções de ativação são sempre positivas como o seguinte:

$$\phi_1'(x) = \phi(x)(1 - \phi(x)) \geq 0 \text{ e } \phi_2'(x) = \phi_1(x) \geq 0 \text{ para } x \rightarrow \infty.$$

A segunda suposição é um pouco mais complicada, pois requer o cálculo da segunda derivada da equação 2.30.

$$\begin{aligned} \frac{\partial^2 C}{\partial X^2} &= \frac{\partial}{\partial X} \left(\frac{\partial C}{\partial c} \frac{\partial c}{\partial m} \frac{\partial m}{\partial X} \right) = \frac{\partial}{\partial X} \left(-\frac{\partial c}{\partial m} \frac{S}{X} \right) \\ &= \frac{\partial}{\partial m} \left(-\frac{\partial c}{\partial m} m \right) \frac{\partial m}{\partial X} = \left(\frac{\partial c}{\partial m} + \frac{\partial^2 c}{\partial m^2} m \right) \frac{m^2}{S}. \end{aligned} \quad (2.45)$$

Como a última etapa da derivação compara um valor conhecido da primeira derivada, para o qual foi mostrado que é positivo, é suficiente assumir implicitamente os erros b_j na forma funcional de c de modo que a segunda derivada $\frac{\partial^2 c}{\partial m^2}$ também seja positiva, o que pode ser feito assumindo erros altos o suficiente. Então, todo o termo será positivo e, portanto, $\frac{\partial^2 C}{\partial X^2}$.

As demais suposições, no entanto, não são tão complicadas de provar. Da mesma forma que a verificação da primeira suposição, obtém-se a validade da terceira suposição

da seguinte forma,

$$\frac{\partial C}{\partial \tau} = \frac{\partial C}{\partial c} \frac{\partial c}{\partial \tau} = X \frac{\partial c}{\partial \tau} = X \sum_j w_j \phi_1'(\cdot) w_{ij} \phi_2'(\cdot) \geq 0. \quad (2.46)$$

A única exceção é que i representa a entrada do tempo até o vencimento.

A quarta suposição pode ser derivada usando a seguinte ideia. Se $X \rightarrow +\infty$, então, em um cenário onde $C > 0$, levaria automaticamente a $c = \frac{C}{X}$, pois C é limitado por S e como c é diferente de zero para vários tempo até o vencimento τ , portanto, é válido que $C = 0$. O quinto e o sexto pressupostos não são facilmente alcançados pela escolha da arquitetura de rede. Portanto, foram geradas opções virtuais a fim de treinar a rede neural para as duas últimas suposições. Essas opções geradas não existem nos dados reais, no entanto, sua presença no conjunto de treinamento pode treinar a rede neural melhor no limite superior e inferior do preço da opção.

A quinta condição é garantida sintetizando todos os S_t do conjunto de dados para muitos preços de strike X possíveis no momento em que o tempo até o vencimento é zero $\tau = 0$. A última condição também é aprendida a partir dos dados gerando opções virtuais com $X \rightarrow 0$ para muitos τ diferentes, isso é numericamente problemático, pois se $X = 0$ for escolhido, o moneynesse m será divergente, o que pode ser resolvido com um X diferente de zero muito pequeno.

3 Parte Empírica

3.0.1 Descrição dos dados

O conjunto de dados de opções usado neste trabalho consiste em opções de compra e venda no estilo europeu sobre o índice SPX (representado por S&P500). A razão pela qual este índice de opções foi escolhido é que ele pertence aos ativos de opções mais líquidos do mercado norte-americano e é uma escolha adequada para a avaliação da previsão do preço da opção. A principal vantagem é que este índice é negociado como uma opção europeia, que é mais fácil de realizar a tarefa de avaliação, uma vez que não pode ser exercido mais cedo do que no momento do vencimento. Além disso, o índice reflete a situação das 500 principais ações listadas nos EUA, que é amplo o suficiente para capturar, pelo menos parcialmente, a situação de todo o mercado mundial. Essa escolha de índice também é melhor do que as opções de ações negociadas em grandes empresas, pois não é tão suscetível aos choques negativos que acontecem em vários setores e, em vez disso, cobre a situação geral do mercado.

Os dados consistem em dados diários no período de junho de 2004 a junho de 2007. O conjunto de dados originalmente tinha 491819 preços de opções compostos por opções de compra e venda ao longo de 761 dias de negociação. O conjunto de dados utilizado neste trabalho corresponde aos dados de cotações de opções no CBOE, que são dados de escolha comum na literatura de precificação de opções. O conjunto de dados original foi extraído do conjunto de dados usado em (BARUNÍKOVA; BARUNÍK, 2011), que investigou o desempenho da rede neural feedforward básica. Isso dá a este trabalho a possibilidade de se relacionar com o referido artigo, embora as medidas sejam ligeiramente diferentes e em diferentes subconjuntos de dados e, portanto, não são diretamente comparáveis.

A possibilidade de expandir o conjunto de dados por dados mais novos não foi escolhida devido à indisponibilidade desses dados na mesma escala na fonte original, bem como seu custo relativamente alto em diferentes fontes de dados. Além disso, se mais fontes de dados forem combinadas, a confiabilidade dos dados pode não ser garantida

como consistente o suficiente em todo o conjunto de dados. Além disso, se dados mais novos fossem examinados, haveria um possível risco de vieses existentes nos preços, uma vez que a crise financeira global de 2008 ocorreu no período subsequente.

O conjunto de dados consiste em vários dados relativos aos preços de opções diários, como a melhor oferta diária e o preço de venda para cada preço de exercício listado em determinada data. O conjunto de dados fornece também informações do tempo até o vencimento para cada opção, que se estende de 1 dia a 3 anos.

Para obter a taxa de juros livre de risco, que é uma entrada necessária para o modelo de preços de Black-Scholes, obtém-se os dados do site oficial do Federal Reserve sobre a taxa de rendimento mensal dos títulos do Tesouro de cupom zero, que servem como proxy da taxa de juros livre de risco.

Como é necessário usar apenas um preço de entrada na fórmula de Black-Scholes, o preço da opção é avaliado como o valor médio do preço de compra e venda fornecido no conjunto de dados.

Da mesma forma que em (BARUNÍKOVA; BARUNÍK, 2011), foram aplicados filtros de exclusão nas opções com menos de seis dias de vencimento, pois pode ocorrer o viés relacionado à liquidez mencionado por (BAKSHI; CAO; CHEN, 1997). Além disso, as cotações do preço da opção inferiores a \$0,4 não foram usadas, pois os efeitos do preço discreto poderiam afetar a inferência geral. Por último, as premissas da seção 2.5 foram verificadas para cada opção e, como havia a possibilidade no mercado de fazer a arbitragem, essas opções foram limpas do conjunto de dados para não enviesar a inferência.

O conjunto de dados resultante consiste em 203671 opções de compra e 150184 opções de venda, que é um conjunto de dados razoavelmente grande para treinar os vários cenários e parâmetros da rede neural examinados neste trabalho.

A fim de compreender melhor as características do conjunto de dados, as seguintes figuras resumindo algumas propriedades chave do conjunto de dados foram criadas. A sequência de figuras a seguir mostra três características diferentes dos dados usados.

A figura 3.1 mostra os preços de fechamento do índice S&P500 no período de junho de 2004 a junho de 2007, que corresponde exatamente aos 761 dias de negociação que correspondem ao conjunto de dados da opção. A faixa de preço começa em \$1063

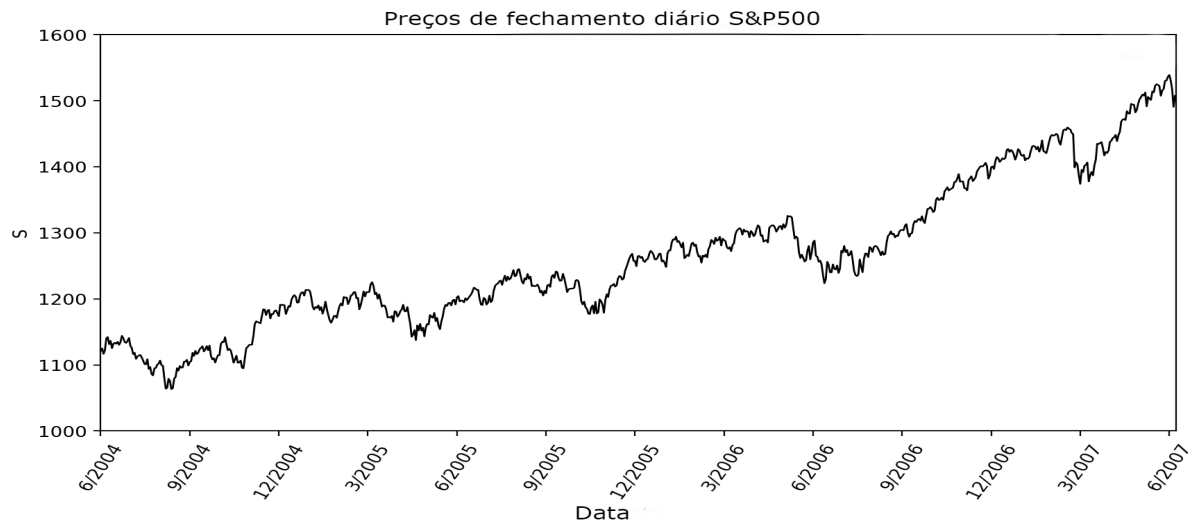


Figura 3.1: Preços de fechamento diário S&P500 entre 2004-2007

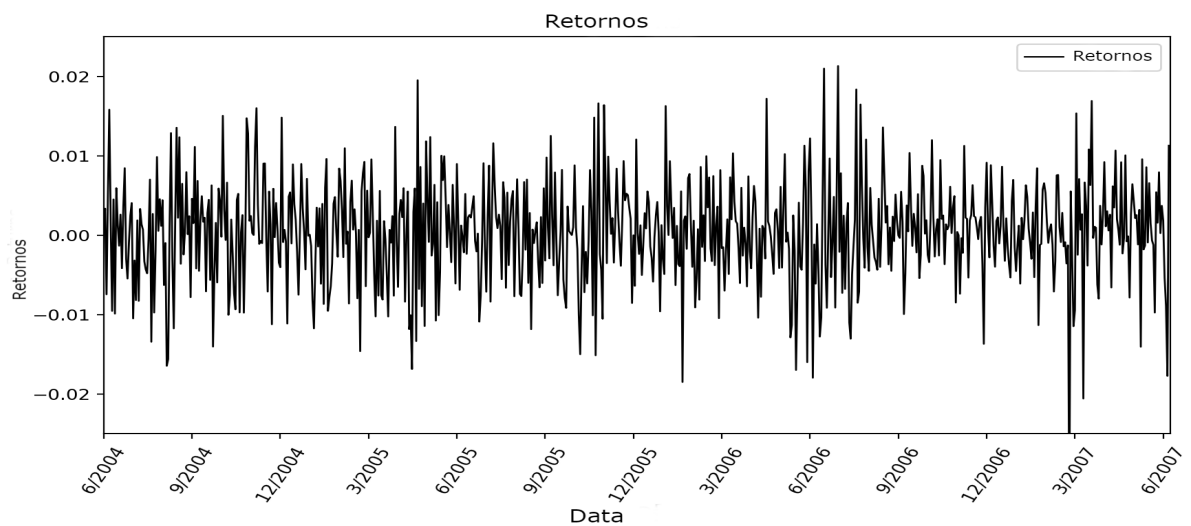


Figura 3.2: Retorno do índice S&P500 entre 2004-2007

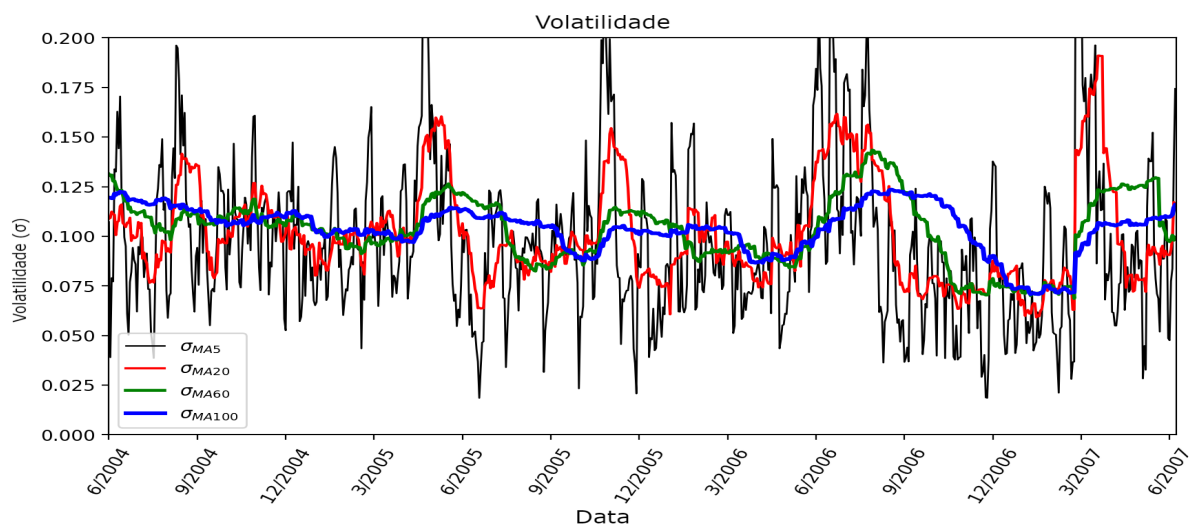


Figura 3.3: Volatilidade do índice S&P500 entre 2004-2007

por ação do índice S&P500 e termina em \$1539. O preço de fechamento cresce ao longo do período examinado, no entanto, como o preço não é tratado como um valor absoluto diretamente e todos os preços das opções são normalizados, esta tendência não produz um efeito que afetaria significativamente a inferência para nossa estimativa, uma vez que as opções são comparadas umas com as outras com base em suas outras características, como moneyness m e tempo até o vencimento τ , que foram distribuídas de forma semelhante em todo o conjunto de dados, pois existem opções com diferentes tempos de vencimento e variações na liquidez em cada dia de negociação.

Com base nesses preços de fechamento diário, os retornos foram calculados de acordo com a fórmula (2.7). A partir da figura 3.2 podemos ver que os retornos estão geralmente oscilando em torno da média próxima a 0. Como medida de volatilidade, os estimadores da volatilidade histórica foram escolhidos com a variância calculada a partir dos valores das médias móveis ao longo de quatro períodos distintos.

A volatilidade de curto prazo é estimada de acordo com a fórmula (2.8) pela variação das médias móveis no intervalo de 5 e 20 dias sobre os retornos e são denotadas por σ_{MA5} e σ_{MA20} .

Por outro lado, a volatilidade de longo prazo é estimada por intervalos de tempo mais longos. Conforme sugerido por (HAHN, 2013), os 60 dias e 100 dias são escolhidos e, portanto, as volatilidades históricas de longo prazo são denotadas como σ_{MA60} e σ_{MA100} . Todas essas quatro medidas de volatilidades históricas são representadas na figura 3.3, onde as medidas de longo prazo são naturalmente defasadas após os períodos de maior volatilidade de curto prazo.

Nas figuras 3.4 e 3.5 a seguir, observa-se uma forma funcional de rampa, típica para opções de compra e venda conforme é mostrado na estrutura de dados. Além disso, as figuras também representam o valor de tempo das opções, pois cada cor individual distingue os diferentes tempos de vencimento (nas figuras a seguir denotadas como T). Quanto mais próxima do vencimento a opção está, mais semelhante ela é à forma teórica.

As figuras de 3.4 a 3.7 mostram os dados para os preços das opções de compra que têm preços de strike \$1100, \$1200, \$1300 e \$1400, respectivamente. As opções de compra no conjunto de dados também têm preços de strike diferentes, com a menor diferença

entre dois preços de strike de \$5. A maioria das observações são, no entanto, as opções que correspondem à diferença de \$25 nos preços de strike (ou seja, 1000, 1025, 1050, etc.). O menor e o maior preço de strike no conjunto de dados são \$400 e \$2000 para as opções de compra.

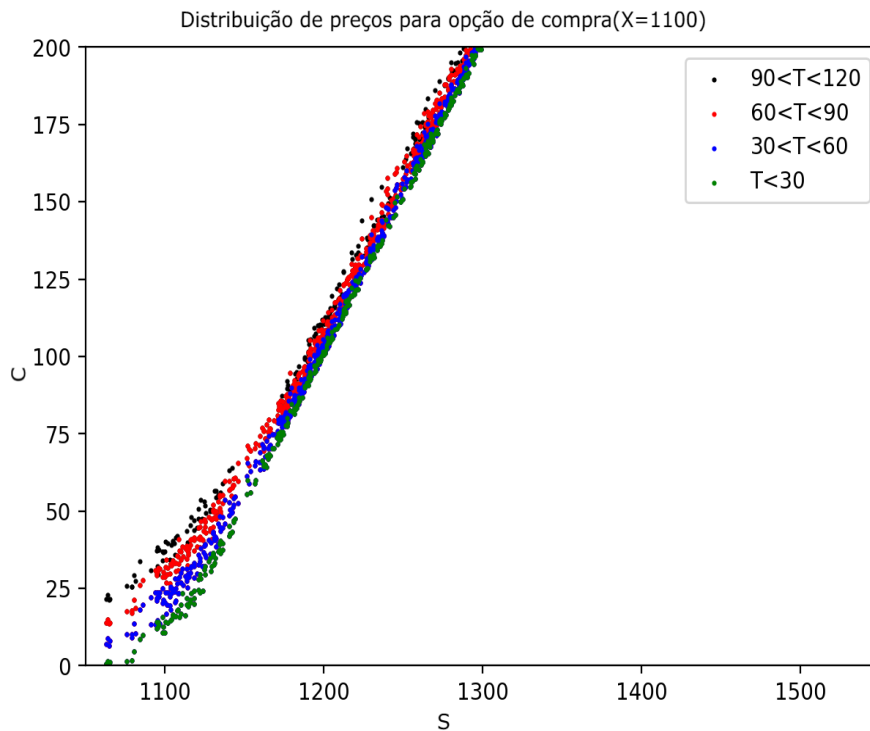
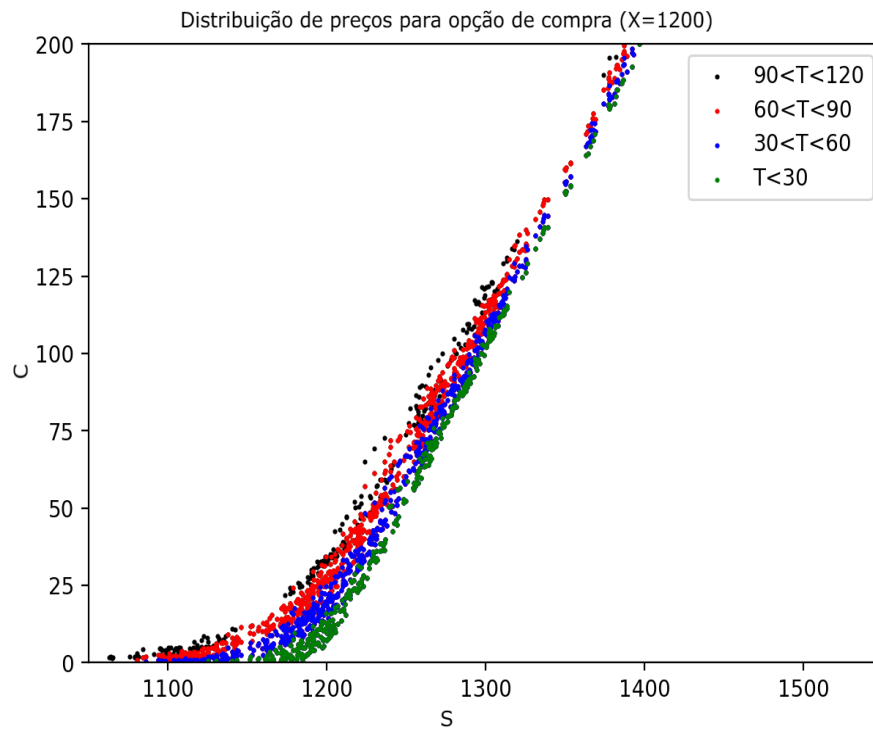
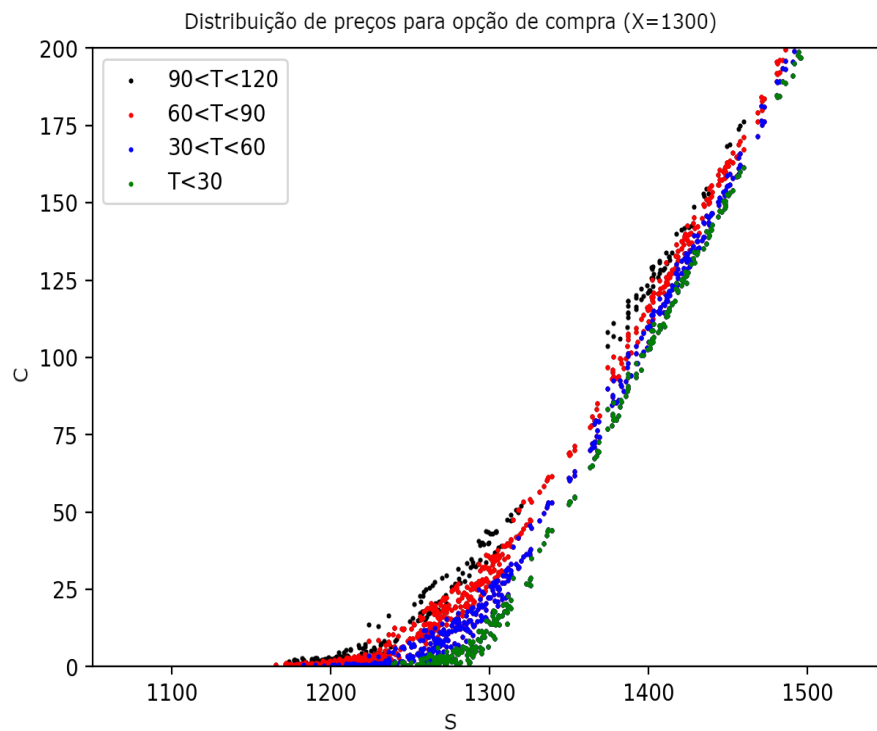


Figura 3.4: Preço de strike $X = 1100$

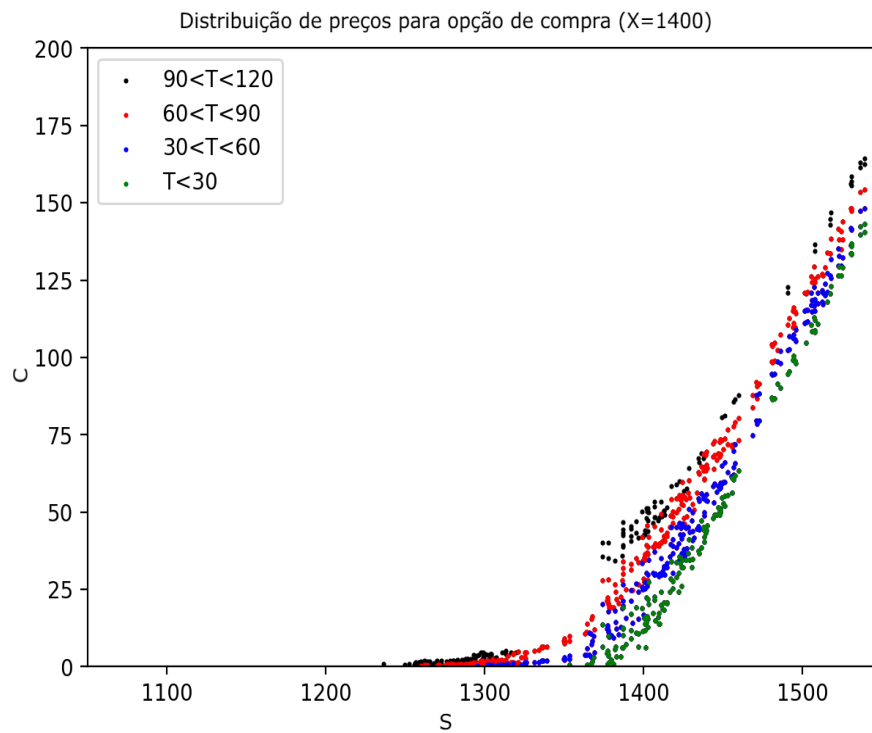
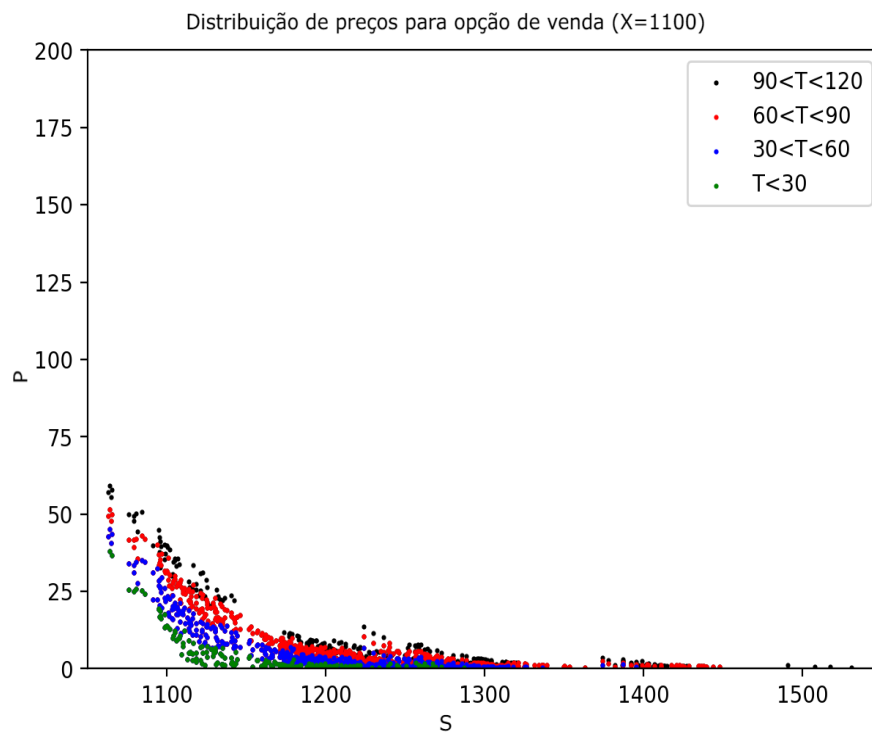
Da mesma forma, as opções de venda são menos negociadas neste período e os dados estão disponíveis para as negociações que oferecem preços de strike entre \$500 e \$1700. Para os preços de strike mais negociados, as opções de venda foram representadas nas figuras de 3.8 a 3.11 em que são mostrados os dados para os preços das opções de venda, novamente com preços de exercício \$1100, \$1200, \$1300 e \$1400, respectivamente. O motivo de representar ambas as opções de compra e as opções de venda para esses preços de strike, é para mostrar as opções de compra e venda que estavam no at the money (ATM) (situação em que o preço de strike de uma opção é idêntico ao preço de mercado atual do título) em algum momento durante o período que estudamos.

Em ambas as figuras anteriores havia subconjuntos de dados perceptíveis que eram mais ricos em termos de densidade de dados e preços, para os quais havia menor número de observações. Isso ocorre devido à limpeza dos dados que retirou alguns dados

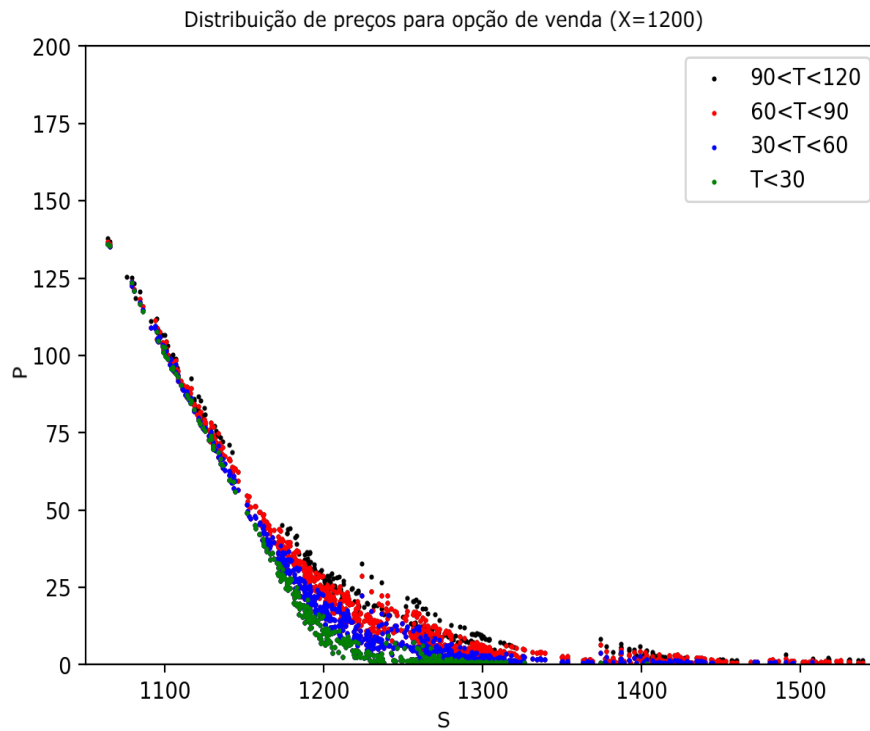
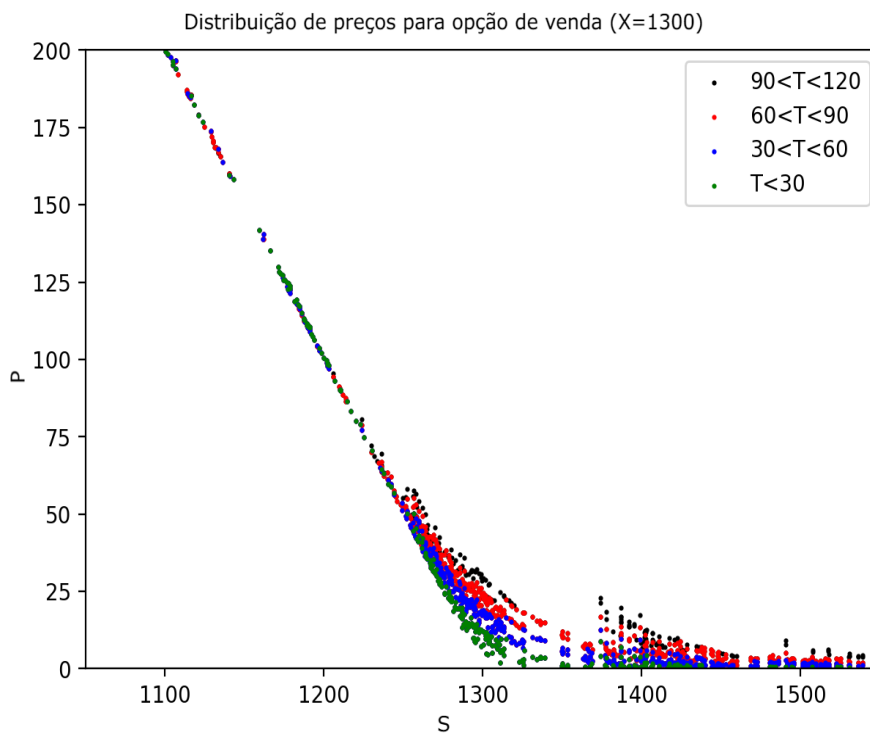
Figura 3.5: Preço de strike $X = 1200$ Figura 3.6: Preço de strike $X = 1300$

e, portanto, é necessário normalizar de alguma forma as opções restantes.

Como já foi mencionado acima, o preço do ativo é dividido pelo preço de strike a fim de obter o moneyness $m = \frac{S}{X}$ da opção, dividimos as opções em três subconjuntos de

Figura 3.7: Preço de strike $X = 1400$ Figura 3.8: Preço de strike $X = 1100$

acordo com a liquidez e em três subconjuntos de acordo com o tempo para o vencimento τ (em figuras denotado por T). Todas as combinações nos fornecem 9 subconjuntos para opções de compra e 9 subconjuntos para opções de venda. Cada um desses subconjuntos

Figura 3.9: Preço de strike $X = 1200$ Figura 3.10: Preço de strike $X = 1300$

é estimado por um módulo de rede neural diferente e, portanto, é necessário conhecer as características desses subconjuntos.

A maioria das opções com um volume relativamente alto tem seu moneyness m

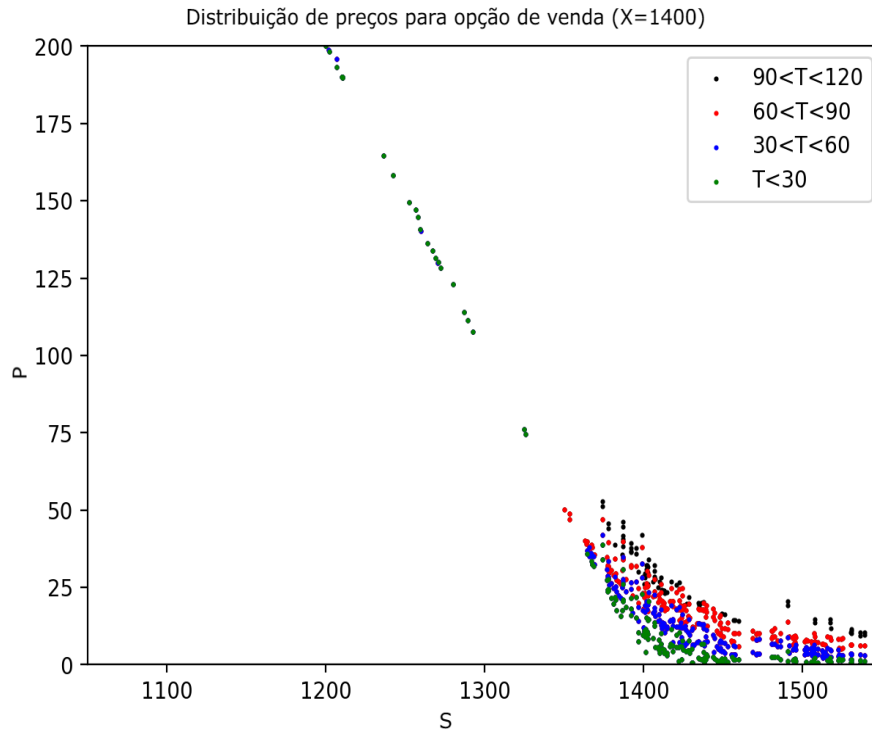


Figura 3.11: Preço de strike $X = 1400$

na faixa de 0.8-1.3, que inclui mais de 84% de todas as opções em nosso conjunto de dados. Os limites dos subconjuntos são escolhidos para que cada um dos módulos seja grande o suficiente e para que o conjunto de treinamento da rede não seja muito pequeno para cada módulo distinto. Os valores $m = 0.97$ e $m = 1.05$ foram escolhidos como valores limites para a divisão do subconjunto, pois é feito da mesma forma por (GRADOJEVIC; GENCAJ; KUKOLJ, 2009).

Da mesma forma, isso também foi feito com a divisão das opções, de acordo com o tempo de vencimento τ . Os limites da divisão foram estabelecidos em 60 dias e 180 dias até o vencimento, o que separa as opções de curto e longo prazo por um subconjunto razoavelmente amplo de opções intermediárias.

A fim de proporcionar uma melhor visualização de como as opções estão distribuídas nos módulos, as figuras 3.12 e 3.13 mostram a distribuição das opções de compra e de venda. A figura 3.12 mostra uma divisão de opções de compra em subconjuntos separados por uma linha vertical (divisão por moneyness m) e pelo tempo até o vencimento τ (na figura denotado por T) representada na figura por cores diferentes. A figura 3.13 mostra exatamente a mesma divisão, exceto que as opções in the money (ITM) estão no

lado esquerdo do gráfico a partir da definição de opções de venda.

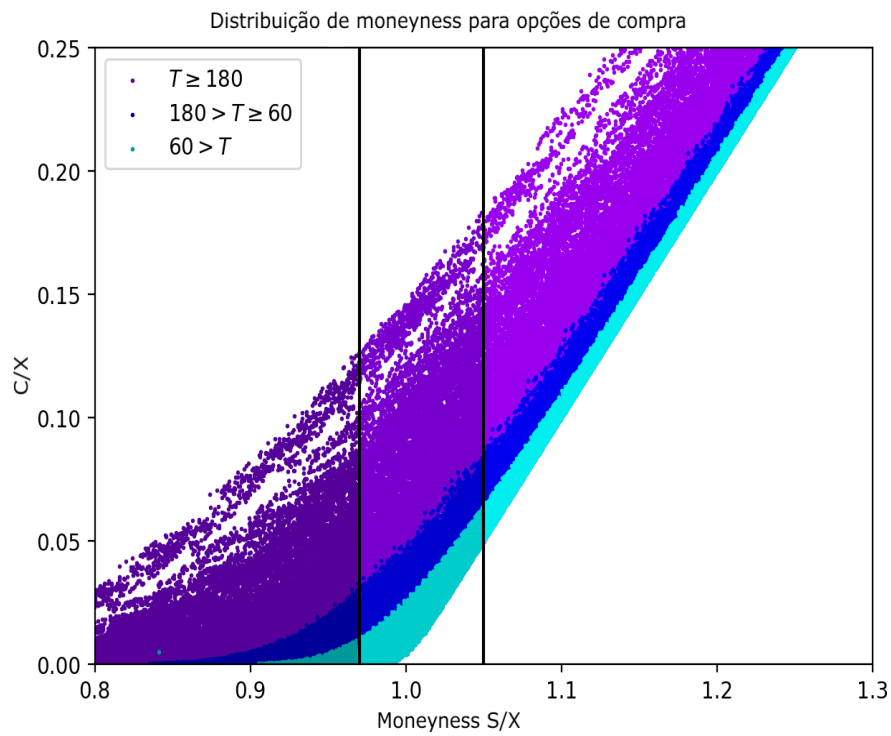


Figura 3.12: Opções de compra

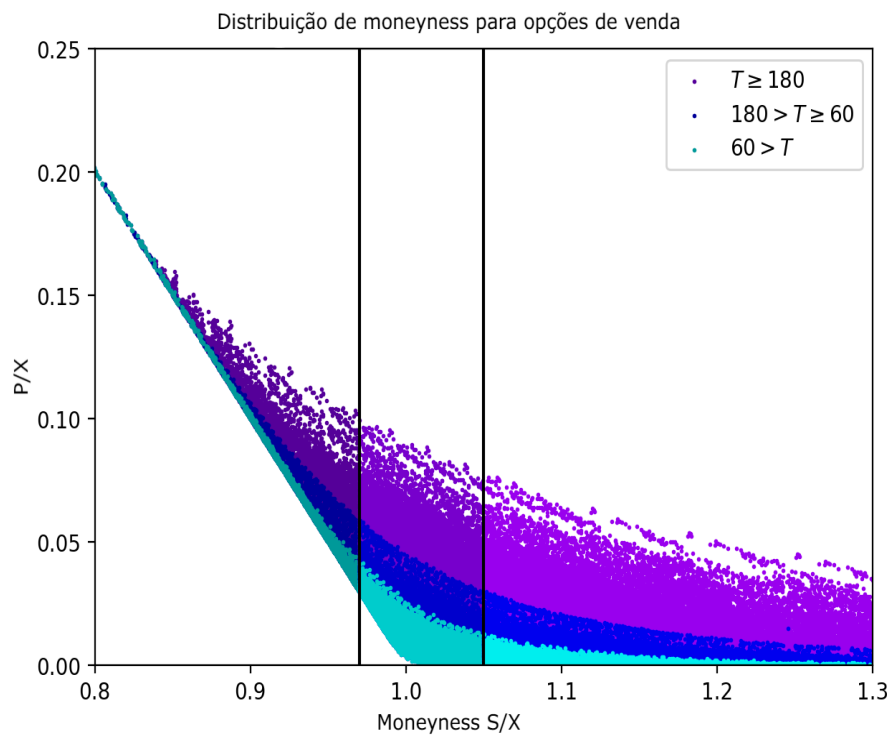


Figura 3.13: Opções de venda

Para mostrar as estatísticas detalhadas dos subconjuntos individuais que são usados na de rede neural modular, usamos a tabela 3.1. Ele descreve o tamanho dos

		Compra		
Dias para vencimento	<60	60-180	≥ 180	Todas as opções
Moneyness (S/X)		Prêmio da opção médio		
OTM (<0.97)	\$2.39	\$7.52	\$30.67	\$21.05
ATM(0.97 - 1.05)	\$28.10	\$44.79	\$105.91	\$53.58
ITM(≥ 1.05)	\$211.08	\$264.71	\$326.38	\$275.03
Moneyness (S/X)		Observações		
OTM(<0.97)	6,769	9,378	26,311	42,458
ATM (0.97 - 1.05)	25,053	12,810	14,347	52,210
ITM (≥ 1.05)	34,627	26,028	48,348	109,003
		Venda		
Dias para vencimento	<60	60-180	≥ 180	Todas as opções
Moneyness (S/X)		Prêmio da opção médio		
ITM (<0.97)	\$84.43	\$81.66	\$101.50	\$92.42
ATM (0.97 - 1.05)	\$14.70	\$27.49	\$56.81	\$29.69
OTM(≤ 1.05)	\$2.12	\$5.45	\$15.96	\$10.77
Moneyness (S/X)		Observações		
ITM(<0.97)	4,230	4,120	8,600	16,950
ATM (0.97 - 1.05)	24,080	12,787	14,336	51,203
OTM(≤ 1.05)	16,377	18,932	46,722	82,031

Tabela 3.1: Resumo do conjunto de dados S&P500

módulos individuais em termos de dados de opções e, para cada módulo, fornece um prêmio médio de opção.

Dois fenômenos teóricos podem ser observados na tabela. O primeiro mostra que o valor temporal da opção cresce com o tempo até o vencimento tanto para as opções de compra quanto para as opções de venda e, assim, o prêmio cresce. O segundo fenômeno observável é o preço crescente da opção à medida que aumenta sua liquidez. Isso é intuitivo, pois quanto maior o moneyness significa que a opção tem maior probabilidade de expirar in the money (ITM) e, portanto, o preço é mais alto.

O menor dos subconjuntos são os subconjuntos com 4230 e 4120 observações de opções de venda, que são as opções com curto prazo de vencimento e estão in the money (ITM). O subconjunto com o menor tempo até o vencimento é o menor também para a opção de compra com 6769 observações, no entanto, essas opções estão out of the money (OTM).

Por outro lado, o maior subconjunto é, em ambos os casos de compra e venda, o subconjunto com opções que tem moneyness $m < 0.97$ e tempo até o vencimento $\tau \geq 180$. Isso ocorre devido à possibilidade de negociar várias cotações de opções com vencimento

futuro superior a 180 dias.

Esta tabela também mostra o fenômeno empírico do volatility smile descrito na seção 2.2.3 que é perceptível principalmente no caso de opções de venda, pois as opções at the money (ATM) têm uma volatilidade implícita muito menor do que as opções que são OTM e ITM. No caso de opções de compra, isso não é observável na tabela devido à escolha de limite dos módulos, mas as opções que estão deep in the money ou deep out of the money mostram o mesmo fenômeno empírico nos dados.

3.0.2 Simulações

Embora existam várias regras empíricas sobre como escolher os parâmetros de redes neurais na literatura de aprendizado de máquina, a literatura sobre a seleção de modelo de rede neural, especialmente em precificação de opções, é bastante rara. Existe uma variedade de algoritmos complicados para geração automática de rede neural, por exemplo, usando redes neurais evolutivas que tira proveito de algoritmos genéticos. No entanto, neste trabalho, o objetivo é ajustar os parâmetros explicitamente para o problema de precificação de opções e propor uma abordagem padronizada facilmente reproduzível. Com base nisso, foram rodadas uma série de simulações com diversos tamanhos de rede para descobrir os parâmetros ótimos da rede. O poder preditivo da rede neural é testado durante o processo medindo o erro quadrático médio (EQM) e o erro médio absoluto (EMA) especificamente para o problema de precificação de opções. Depois que os erros são calculados, os parâmetros ideais são escolhidos para minimizar os erros. A fim de descobrir que tamanho de rede é grande o suficiente para não subequipar a rede neural e não superequipá-la ao mesmo tempo, de forma independente no conjunto de dados de opções, seguimos um caminho semelhante ao apresentado em (CULKIN; DAS, 2017) que é explicado na próxima subseção.

3.0.3 Testes de resolução

Nesta subseção, o objetivo principal é descobrir o número ideal de nós e o número de camadas ocultas na rede neural. Embora possa ser provado que uma rede de camada oculta pode aproximar qualquer função contínua arbitrária pelo teorema de aproximação

universal, é empiricamente preferido usar redes neurais mais profundas em casos em que a tarefa de aprendizagem é realizada para objetos difíceis, por exemplo caracteres manuscritos ou reconhecimento facial, o que ajuda a rede a aprender um padrão difícil mais rapidamente.

Nos testes a seguir tentaremos descobrir qual é o número ideal de camadas ocultas e o número ideal de nós para precificação de opções. Da mesma forma que (CULKIN; DAS, 2017), geramos novos dados com parâmetros que correspondem ao nosso verdadeiro conjunto de dados, com a exceção de que seu preço corresponde precisamente à fórmula original de Black-Scholes. Os pontos são gerados de maneira uniforme e aleatória na faixa de parâmetros do conjunto de dados de opção original. Os parâmetros são escolhidos de acordo com a tabela 3.2.

Parâmetro	Faixa
Preço da ação (S)	\$1100 – \$1500
Preço de Strike (X)	\$1000 – \$1600
Vencimento (T)	1 dia - 1000 dias
Data	1/6/2004 - 31/5/2007
Taxa de juros (r)	Baseada no ativo
Volatilidade (σ)	Baseada no ativo

Tabela 3.2: Parâmetros usados para simular preços de opções de compra

Para testar o efeito dos parâmetros, tivemos que fixar a função de ativação como função de ativação sigmóide e fixamos a função softplus na camada de saída, conforme proposto anteriormente como uma nova arquitetura de bom desempenho. A razão para corrigir isso é que comparar 360 modelos de redes neurais é computacionalmente difícil e, na prática, não é possível testar todos os outros cenários possíveis de escolhas de parâmetros.

Com base nesses parâmetros, estimamos os preços das opções e os comparamos na amostra e no poder de previsão fora da amostra. O conjunto de treinamento, conjunto de validação e conjunto de teste é dividido na proporção 60:20:20. A previsão dentro da amostra se mostrou muito boa para previsões da rede neural, portanto, este trabalho se concentra mais na capacidade de generalizar o relacionamento e o desempenho fora da amostra.

A variável dropout representa o número de nós que são eliminados da arquitetura

tura da rede neural quando a etapa de retropropagação é concluída, o que é usado na literatura de aprendizado de máquina para aumentar a precisão. Para algumas opções de parâmetros, o poder de predição na predição fora da amostra foi, no entanto, muito fraco, ao passo que também houve predições de correspondência muito boas que demonstramos nas figuras 3.14 e 3.15 a seguir.

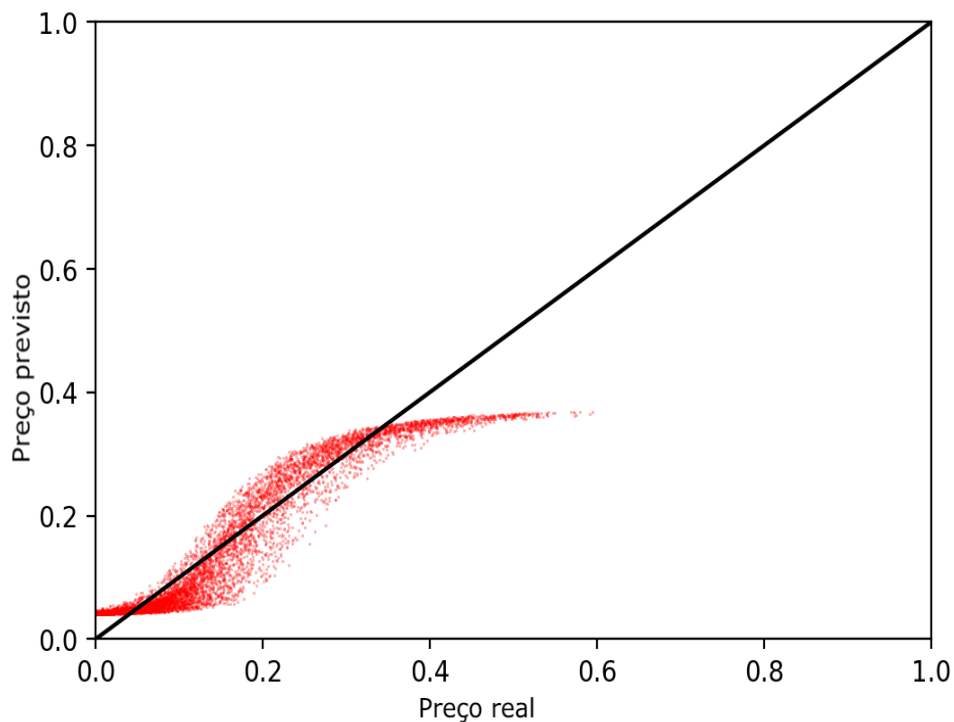


Figura 3.14: Previsão fora da amostra ruim

A previsão na figura 3.14 é gerada usando os seguintes parâmetros: 4 camadas ocultas, 40 nós, 10% de dropout, tamanho da amostra de treinamento: 10000. A previsão na figura 3.15 é gerada usando os seguintes parâmetros: 4 camadas ocultas, 80 nós, 0% de dropout, tamanho da amostra de treinamento: 40.000

Para avaliar o quão boa é a previsão fora da amostra da rede neural, medimos o Erro Médio Absoluto (MAE) e o Erro Quadrático Médio (RMSE) para cada modelo e, em seguida, comparamos os modelos entre si. O Erro Absoluto Médio Percentual (MAPE) não foi usado para essas simulações, pois as previsões não precisas para preços pequenos tornam esta medida muito enviesada. A fim de encontrar o efeito dos parâmetros do modelo, uma regressão de Mínimos Quadrados Ordinários (MQO) foi estimada com base

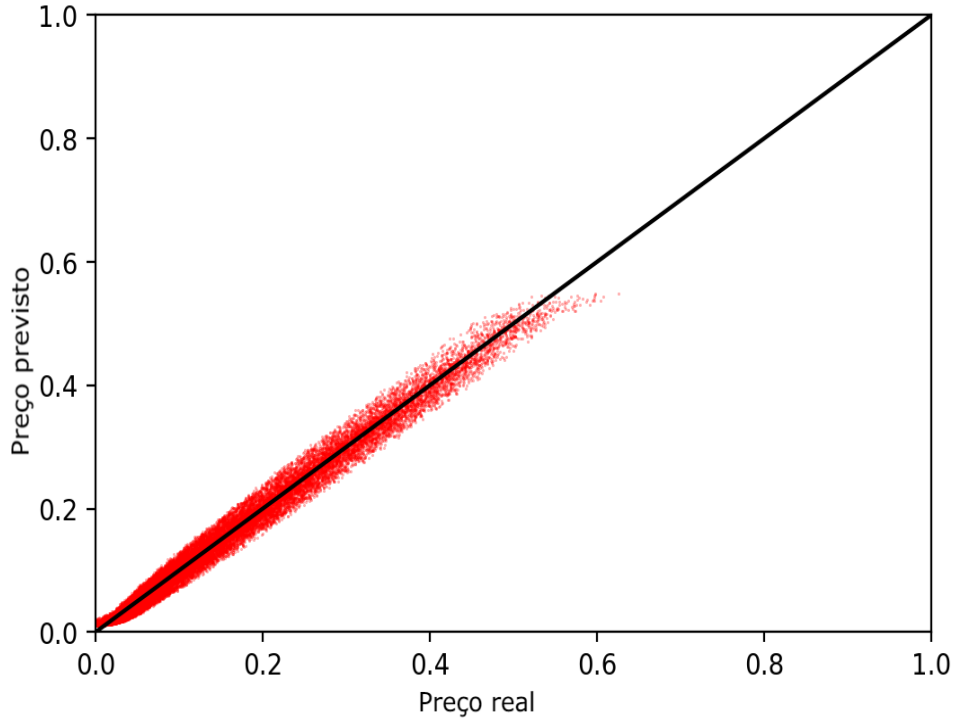


Figura 3.15: Previsão fora da amostra boa

nas seguintes fórmulas

$$MAE_i = \alpha + \beta_1 \text{hidden}_i + \beta_2 \text{nodes}_i + \beta_3 \text{size}_i + \beta_4 \text{dropout}_i + \epsilon_i \quad (3.1)$$

$$RMSE_i = \alpha + \beta_1 \text{hidden}_i + \beta_2 \text{nodes}_i + \beta_3 \text{size}_i + \beta_4 \text{dropout}_i + \epsilon_i \quad (3.2)$$

Usando o Teste de Erro de Especificação de Equação de Regressão de Ramsey (RESET), descobriu-se que esses modelos foram especificados incorretamente, o que também é bastante intuitivo, pois o efeito do número de camadas ocultas está relacionado ao número de nós e também ao tamanho da amostra de treinamento. Portanto, melhoramos o modelo adicionando algumas das variáveis quadradas e termos mistos para tornar os modelos especificados corretamente.

$$\begin{aligned} MAE_i = & \alpha + \beta_1 \text{hidden}_i + \beta_2 \text{hidden}_i^2 + \beta_3 \text{nodes}_i \\ & + \beta_4 \text{nodes}_i^2 + \beta_5 \text{size}_i + \beta_6 \text{size}_i^2 + \beta_7 \text{dropout}_i + \\ & \beta_8 \text{hidden}_i \text{nodes}_i + \beta_9 \text{hidden}_i \text{size}_i + \beta_{10} \text{nodes}_i \text{size}_i + \epsilon_i \end{aligned} \quad (3.3)$$

$$\begin{aligned}
EQM_i = & \alpha + \beta_1 \text{hidden}_i + \beta_2 \text{hidden}_i^2 + \beta_3 \text{nodes}_i \\
& + \beta_4 \text{nodes}_i^2 + \beta_5 \text{size}_i + \beta_6 \text{size}_i^2 + \beta_7 \text{dropout}_i + \\
& \beta_8 \text{hidden}_i \text{nodes}_i + \beta_9 \text{hidden}_i \text{size}_i + \beta_{10} \text{nodes}_i \text{size}_i + \epsilon_i
\end{aligned} \tag{3.4}$$

Como passo seguinte, executamos o teste de Breusch-Pagan que é usado para testar a heteroscedasticidade em modelos MQO. A homocedasticidade foi rejeitada em ambos os modelos, portanto estimamos ambos os modelos de regressão com erros padrão de White robustos. Os resultados do modelo estão nas tabelas 3.3 e 3.4.

Número de observações		R^2	
360		0.5808	
MAE	β	Erro padrão robusto	t
hidden	$-1.89 \cdot 10^2$	$6.51 \cdot 10^3$	-2.90**
<i>hidden</i> ²	$3.37 \cdot 10^3$	$1.19 \cdot 10^3$	2.74**
nodes	$-2.15 \cdot 10^3$	$2.01 \cdot 10^4$	10.71***
<i>nodes</i> ²	$1.26 \cdot 10^5$	$1.61 \cdot 10^6$	7.83***
size	$-5.23 \cdot 10^6$	$7.41 \cdot 10^7$	7.05***
<i>size</i> ²	$5.61 \cdot 10^{11}$	$1.32 \cdot 10^{11}$	4.25***
dropout	$1.57 \cdot 10^2$	$1.51 \cdot 10^2$	1.04
<i>hidden</i> · <i>nodes</i>	$-7.63 \cdot 10^5$	$3.75 \cdot 10^5$	-2.04*
<i>hidden</i> · <i>size</i>	$2.23 \cdot 10^7$	$8.62 \cdot 10^8$	2.59**
<i>nodes</i> · <i>size</i>	$1.30 \cdot 10^8$	$3.40 \cdot 10^9$	3.82***
intercept	0.190	0.012	16.42***

Tabela 3.3: Estimativa MQO de MAE dos parâmetros da rede neural

Número de observações		R^2	
360		0.5851	
RMSE	β	Erro padrão robusto	t
hidden	$-2.19 \cdot 10^2$	$8.17 \cdot 10^3$	-2.68**
<i>hidden</i> ²	$3.91 \cdot 10^3$	$1.50 \cdot 10^3$	2.62**
nodes	$-2.79 \cdot 10^3$	$2.53 \cdot 10^4$	-11.04***
<i>nodes</i> ²	$1.64 \cdot 10^5$	$2.03 \cdot 10^6$	8.10***
size	$-6.55 \cdot 10^6$	$9.29 \cdot 10^7$	-7.05***
<i>size</i> ²	$7.10 \cdot 10^{11}$	$1.66 \cdot 10^{11}$	4.27***
dropout	$3.31 \cdot 10^2$	$1.89 \cdot 10^2$	1.75
<i>hidden</i> · <i>nodes</i>	$-9.38 \cdot 10^5$	$4.73 \cdot 10^5$	-1.98*
<i>hidden</i> · <i>size</i>	$2.59 \cdot 10^7$	$1.08 \cdot 10^8$	2.39**
<i>nodes</i> · <i>size</i>	$1.70 \cdot 10^8$	$4.26 \cdot 10^9$	3.99***
intercept	0.237	0.015	16.24***

Tabela 3.4: Estimativa MQO de RMSE dos parâmetros da rede neural

Ambas as regressões estimam variáveis dependentes de forma muito semelhante portanto, não é surpreendente que os resultados sejam muito semelhantes em termos de

coeficientes e também em termos de erros padrão e significância. Portanto, analisaremos apenas a primeira tabela estimando o MAE. A partir dos coeficientes estimados, vemos que todas as variáveis, exceto dropout, são significativamente diferentes de zero, pelo menos a um nível de significância de 0,05 e algumas delas são significativas mesmo a um nível de significância de 0,001. O coeficiente R^2 é 0,5808, o que sugere que o modelo explica razoavelmente bem a variância das variáveis independentes.

O sinal nos termos quadráticos mostra convexidade nas variáveis hidden, nodes e size. A escolha ideal dos parâmetros Hidden e Nodes pode ser feita encontrando um número de nós e camadas ocultas correspondentes ao MAE mínimo. Os resultados da estimativa sugerem, encontrando o efeito mínimo da variável hidden, que o número de camadas ocultas que minimizam o erro é $hidden = 2.8$. Da mesma forma, ao encontrar o efeito mínimo da variável nodes, o número ideal de nós é $nodes = 85$.

Portanto, com base nos resultados dessa estimativa, definimos o número de camadas ocultas nos modelos futuros como 2-3 e o número de nós em cada camada em cerca de 50-80 dependendo do tamanho da rede de treinamento nos módulos de rede usados. Na escolha do tamanho da rede também é importante pensar nas regras empíricas reconhecidas na literatura de aprendizado de máquina que, geralmente penalizam um número muito grande de nós e um número grande de camadas. No entanto, a escolha de fixar os parâmetros deve ser feita a fim de tornar as comparações entre os diferentes modelos numericamente viáveis, pois há muitos parâmetros que poderiam influenciar a estimativa.

Com base nesta estimativa, também sugere-se não usar a técnica de dropout (que foi usada, por exemplo, em (CULKIN; DAS, 2017) para a aplicação de precificação de opções) nos modelos futuros, uma vez que não teve um efeito significativo fora da amostra para previsão do preço da opção. Além disso, o sinal do coeficiente foi positivo, sugerindo que quanto maior a razão de dropout, maior o erro de predição fora da amostra. Portanto, essa técnica não será usada nos modelos a seguir. Para a escolha de opções de compra, os resultados da aplicação desta estratégia no conjunto de dados real para diferentes números de nós e camadas ocultas estão resumidos na tabela A.2 que mostra que os erros estão aumentando com um número muito alto de nós. Isso é mostrado no conjunto de dados real que legitima parcialmente o método de simulação que usamos neste capítulo para

estimar o tamanho da rede neural.

3.0.4 Opções virtuais

De acordo com o artigo escrito por (YANG; ZHENG; HOSPEDALES, 2017), é útil restringir as redes neurais na tarefa de precificação de opções adicionando suposições sobre a função de precificação de opções. Os autores mencionados sugeriram seis melhorias, que são necessárias para fazer previsões racionais a fim de obter melhor precisão na precificação de opções usando redes neurais. Neste trabalho, uma discussão a respeito dessas premissas foi feita na parte teórica (seção 2.5). A quinta e a sexta suposições dificilmente podem ser cumpridas na rede neural apenas pela escolha da arquitetura ou funções de ativação. O uso prático dessas duas suposições é feito primeiramente por (YANG; ZHENG; HOSPEDALES, 2017). Neste trabalho, foi identificado que adicionar as chamadas opções virtuais pode, em alguns casos, melhorar significativamente o desempenho da rede.

O cumprimento da quinta e sexta premissas é feito pela geração de novas opções (virtuais) originalmente não existentes. No caso da quinta premissa, essas opções atendem à condição de não arbitragem quando seu prazo de vencimento é zero. Isso significa que o formato das opções de preço geradas correspondem à função de rampa de opções. Isso garante que a rede neural aprenda onde está o limite inferior do preço e não superestime ou subestime as opções com pouco tempo até o vencimento.

Por outro lado, cumprir a sexta premissa é mais complicado. Isso leva a uma geração de opções virtuais com preço de exercício zero, onde os limites inferior e superior do preço da opção coincidem. Isso significa que essa opção deve custar exatamente o mesmo que o ativo, pois não há risco de cair abaixo do preço do ativo. Em outras palavras, essas opções virtuais são as mais caras. Muitas dessas opções podem ser geradas para qualquer tempo arbitrário até o vencimento τ e correspondendo a qualquer preço do ativo.

Há um problema prático com essas opções, em que as opções com preço de strike X zero têm moneyness que tende ao infinito $\lim_{x \rightarrow 0} m = +\infty$. Este problema pode ser, no entanto, mitigado pela suposição de um preço de exercício muito baixo, mas positivo (no

caso \$ 1). Esta é uma aproximação, mas como todas as outras opções em nosso conjunto de dados têm um preço de ativo na faixa entre \$1000-\$1600, isso gera instantaneamente uma opção com moneyness muito alto, mas é exatamente isso que a rede neural precisa aprender. Este problema também pode ser mitigado usando moneyness inverso $\frac{1}{m} = \frac{X}{S}$ como uma entrada nas redes neurais, o que foi feito por (YANG; ZHENG; HOSPEDALES, 2017). No entanto, neste trabalho, o moneyness padrão m é usado da maneira que foi descrito anteriormente na parte teórica, visto que a maioria dos outros artigos de pesquisa também o usa.

Portanto, as opções virtuais estão estendendo o conjunto de dados fornecido, adicionando novas opções respeitando a regra teórica de não arbitragem no momento da expiração da opção ou simulando a opção mais cara possível. É possível identificar a diferença em relação às simulações anteriores, onde geramos opções completamente novas que atendiam exatamente à fórmula de Black e Scholes. As opções geradas na subseção anterior não podem ser adicionadas ao conjunto de dados real, pois a fórmula Black-Scholes exata não é válida para os dados de opções reais. O conjunto de dados foi gerado apenas para descobrir qual é aproximadamente o tamanho certo da rede para aprender a relação entre a entrada e a saída, que é muito semelhante (mas não idêntica) à fórmula de Black-Scholes.

Por outro lado, as opções virtuais geradas de acordo com os pressupostos mencionados, são totalmente compatíveis com os dados reais e podem ser adicionadas aos dados reais no conjunto de treino da rede neural para melhorar o processo de aprendizagem da rede neural. Os resultados desse treinamento aprimorado são investigados no próximo capítulo.

4 Resultados

4.1 Previsão de volatilidade

Para estimar a volatilidade, usamos dois métodos neste trabalho. Primeiro, a volatilidade histórica é estimada medindo o desvio padrão da média móvel em uma janela histórica que é descrita tanto na parte teórica quanto na parte de descrição de dados. A duração escolhida dessas janelas foi de 5 dias de negociação de curto prazo e 20 e 60 dias de negociação de longo prazo. A rede neural que aprende esse parâmetro de volatilidade usa todas as quatro medidas de volatilidade.

Por outro lado, no caso de um mecanismo de previsão de volatilidade diferente usando GARCH (1,1), a rede neural aprende apenas esta medida que, no entanto, se mostrou para alguns modelos uma entrada melhor do que a volatilidade histórica. Nas figuras 4.1 e 4.2, há um teste de autocorrelação que confirma que GARCH (1,1) está parametrizado corretamente.

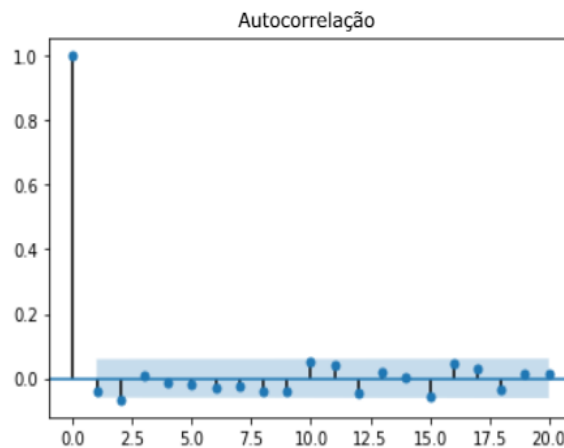


Figura 4.1: Função de autocorrelação

Os parâmetros resultantes do modelo GARCH (1,1) foram estimados em uma amostra de 761 dias de negociação com os seguintes resultados na tabela 4.1.

A tabela 4.1 mostra que os retornos médios em torno dos quais a volatilidade é medida são quase desprezíveis em $4.60 \cdot 10^{-4}$. Os resultados importantes do modelo são suas magnitudes α_1 e β_1 que mostram que o modelo é mais determinado pelo resíduo da

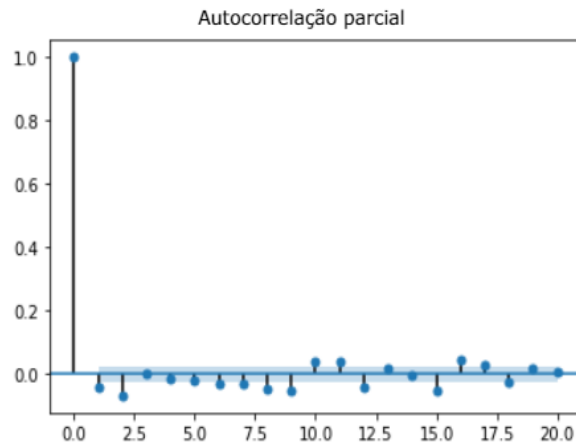


Figura 4.2: Função de autocorrelação parcial

Número de observações		761	
σ GARCH(1, 1)	β	Erro Padrão Robusto	t
Retornos médios	$4.60 \cdot 10^{-4}$	$4.64 \cdot 10^{-5}$	9.93***
ω	$4.25 \cdot 10^{-6}$	$1.78 \cdot 10^{-11}$	$2.38 \cdot 10^5$ ** *
α_1	$5.00 \cdot 10^{-2}$	$2.02 \cdot 10^{-2}$	2.48
β_1	0.85	$2.94 \cdot 10^{-2}$	28.90

Tabela 4.1: Estimativa do modelo GARCH (1,1) com base no preço de fechamento do ativo subjacente (nível de significância 0.05 *, nível de significância 0.01 **, nível de significância 0.001 ***)

equação média do que pela volatilidade passada, pois β_1 é 17 vezes maior do que α_1 .

A previsão de volatilidade por GARCH (1,1) está representada na figura 4.3, onde pode ser visto que corresponde relativamente bem à volatilidade histórica de curto prazo da média móvel de cinco dias σ_{MA5} .

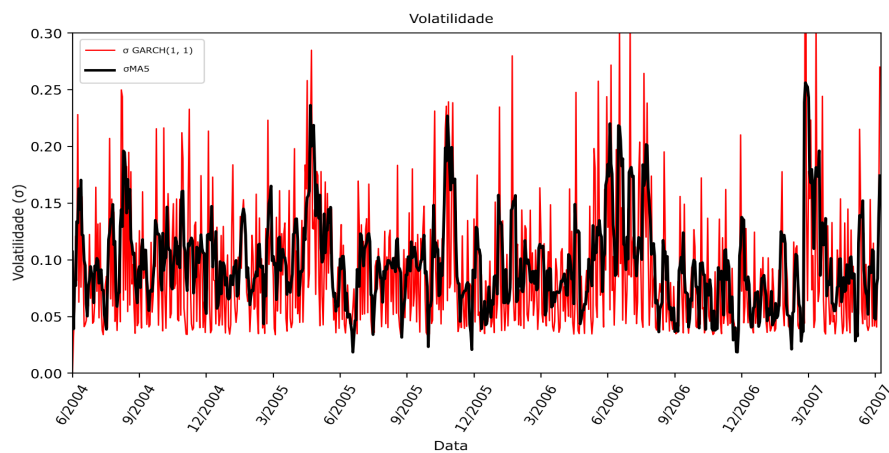


Figura 4.3: Comparação do modelo de volatilidade GARCH (1,1) com a volatilidade histórica de curto prazo

4.2 Previsão com redes neurais

Comparamos o poder de previsão de várias arquiteturas de redes neurais com e sem o uso de opções virtuais para treinamento do conjunto de dados. O tamanho do conjunto de dados de treinamento, validação e teste, foi ao longo deste trabalho, escolhido na proporção 60:20:20. Estimamos três arquiteturas diferentes com 2 camadas ocultas fixas e 50 nós em cada uma delas. Essa escolha do tamanho foi feita após a obtenção dos resultados da seção de simulações.

A tabela 4.2 mostra a previsão das opções de compra, o efeito do procedimento de treinamento aprimorado pode ser visto em todas as três medidas MAE, RMSE e MAPE. O poder de predição é avaliado pelo teste Diebold Mariano, que é usado para comparação fora da amostra de dois modelos (em nosso caso, predição de rede neural e fórmula de Black-Scholes). A estatística Diebold-Mariano (DM) (DIEBOLD; MARIANO, 1995) testa se as duas amostras comparadas têm a mesma precisão preditiva ou não. A hipótese nula desse teste é que ambas as amostras têm o mesmo poder preditivo e, portanto, se a hipótese puder ser rejeitada, o primeiro modelo supera o segundo modelo. Mais detalhes sobre este modelo podem ser encontrados em (DIEBOLD; MARIANO, 1995).

Em todos os casos estudados, as redes neurais foram muito melhores na previsão de preços do que a fórmula de Black-Scholes, o que é mostrado pelo fato da estatística Diebold-Mariano ser grande e positiva, o que significa que tem um desempenho de previsão fora da amostra significativamente melhor do que no caso de fórmula clássica de Black-Scholes. O nível de significância no qual as previsões da rede neural são melhores do que Black-Scholes é em todos os casos 0,001.

Isso significa que todos os modelos usados estão na previsão fora da amostra superando o modelo Black-Scholes e, portanto, confirmando nossa primeira hipótese de melhor previsão fora da amostra de redes feedforward do que modelos paramétricos. No caso da previsão de volatilidade GARCH (1,1), o teste Diebold Mariano tem valor menor, mas ainda significativo, pois essa volatilidade é usada também para a previsão de Black-Scholes, enquanto no caso de volatilidade histórica o σ_{MA100} de longo prazo foi escolhido para a avaliação do preço Black-Scholes.

Nas seguintes tabelas 4.2 para opções de compra e 4.3 para opções de venda, vários

efeitos podem ser observados. Estas duas tabelas contêm os resultados mais importantes deste trabalho, pois confirmam duas das hipóteses apresentadas. Primeiro, todos os três erros MAE, RMSE e MAPE diminuem drasticamente à medida que aumentamos as opções virtuais para cenários com função de saída softplus e função de ativação especial. Isso confirma a terceira hipótese, pois os erros são muito menores usando opções virtuais do que se não as usarmos.

Ambos os cenários tiveram melhores previsões fora da amostra para a volatilidade histórica se não houvesse opções virtuais, no entanto, conforme o número de opções virtuais aumentava, o preditor de volatilidade GARCH (1,1) começou a prever melhor o preço da opção do que a volatilidade histórica. Isso mostra que a quarta hipótese também é válida, em que o desempenho da previsão difere conforme o mecanismo de volatilidade muda. No caso das opções de venda, a volatilidade histórica superou sistematicamente o GARCH (1,1) em quase todos os casos, exceto para o cenário com saída linear, que no entanto se mostrou o pior dos três modelos examinados com erro entre 2 e 4 múltiplos de as outras duas arquiteturas em quase todas as configurações.

No entanto, para a escolha da função de saída linear e função de ativação sigmóide nas camadas ocultas, o efeito de adicionar opções virtuais foi o mais alto, embora geralmente não tivesse um bom desempenho. A melhor escolha geral foi um cenário em que a função de saída softplus foi usada com a função de ativação sigmóide que tinha o menor erro de todas as opções de compra quando 76100 opções virtuais foram usadas para treinar o conjunto de dados.

A fim de demonstrar a precisão da predição por nossas escolhas de arquitetura, representamos as medidas de predição nas figuras 4.4 a 4.7 a seguir para uma saída softplus e função de ativação sigmóide nas camadas ocultas. Descrevemos também as mesmas medidas para a escolha das funções de ativação do artigo de (CULKIN; DAS, 2017). Para as figuras 4.8 a 4.11 temos a previsão de preços de opções de compra com funções especiais de ativação sem o uso de opções virtuais. A forma gráfica das figuras é mostrada da mesma forma que no artigo mencionado.

As figuras demonstram que a rede neural é muito bem-sucedida no aprendizado da

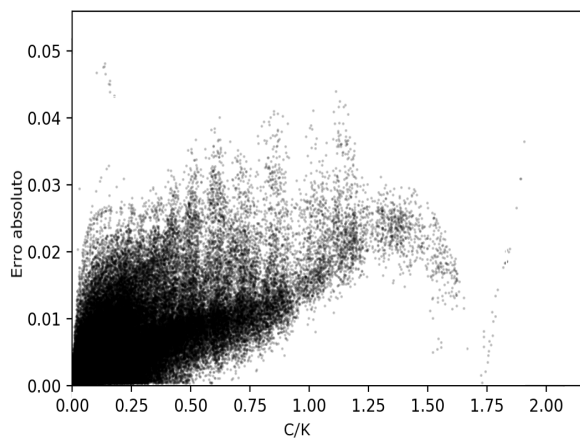


Figura 4.4: Erro absoluto 1

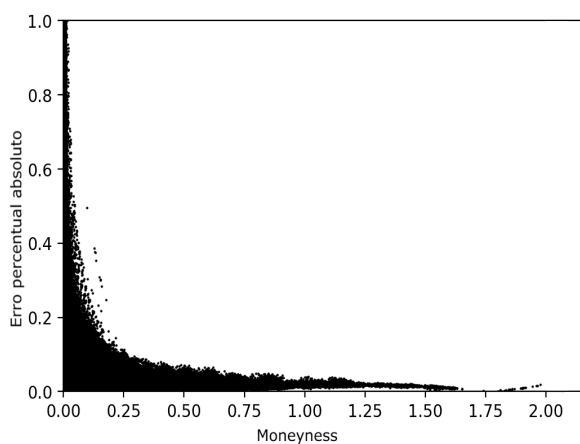


Figura 4.5: Erro percentual absoluto 1

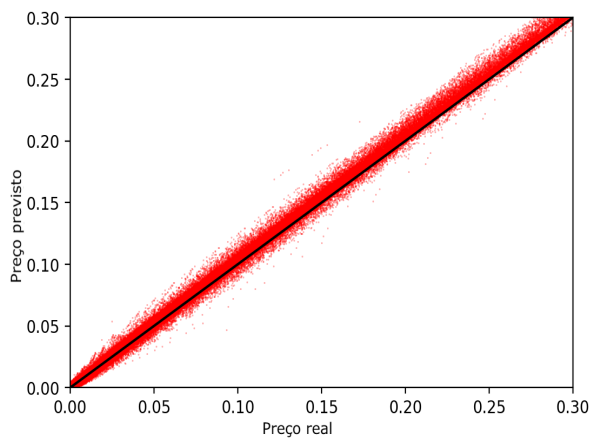


Figura 4.6: Precisão de previsão 1

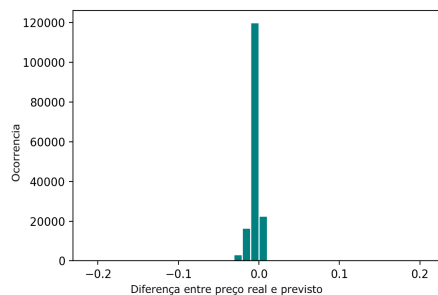


Figura 4.7: Histograma de diferença 1

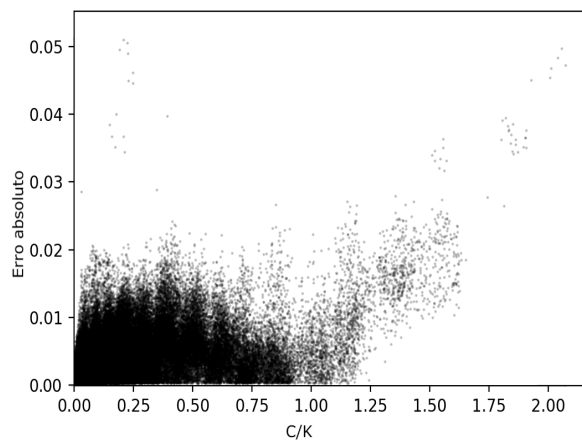


Figura 4.8: Erro absoluto 2

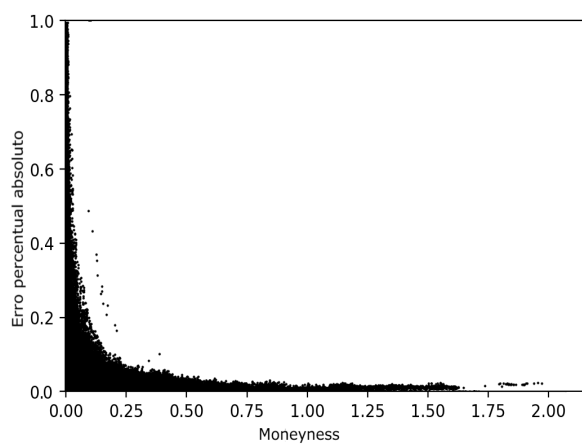


Figura 4.9: Erro percentual absoluto 2

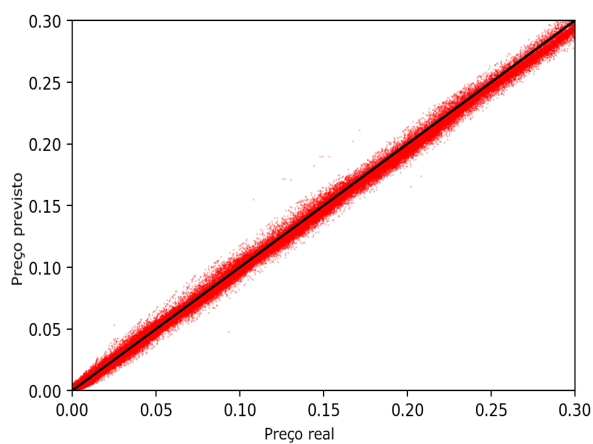


Figura 4.10: Precisão de previsão 2

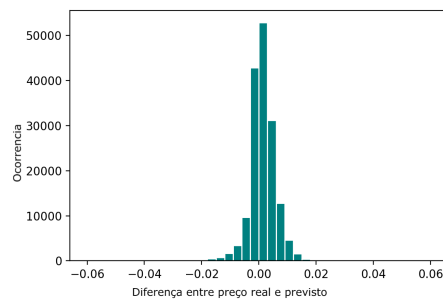


Figura 4.11: Histograma de diferença 2

Ativação	Output	Volatilidade	Virtual	MAE	RMSE	MAPE	Teste DM
sigmoid	linear	σ HV	0	0.00798	0.00902	0.693	23.77***
sigmoid	softplus	σ HV	0	0.00331	0.00461	0.129	39.51***
especial	exp	σ HV	0	0.00407	0.00583	0.119	37.25***
sigmoid	linear	σ HV	7610	0.00359	0.00466	0.446	38.96***
sigmoid	softplus	σ HV	7610	0.00315	0.00453	0.109	39.71***
especial	exp	σ HV	7610	0.00448	0.00615	0.117	34.95***
sigmoid	linear	σ HV	30440	0.00513	0.00639	0.675	34.83***
sigmoid	softplus	σ HV	30440	0.00307	0.00429	0.106	39.94***
especial	exp	σ HV	30440	0.00416	0.00639	0.171	35.82***
sigmoid	linear	σ HV	76100	0.00614	0.00762	0.514	31.09***
sigmoid	softplus	σ HV	76100	0.003	0.00431	0.103	40.14***
especial	exp	σ HV	76100	0.00322	0.00486	0.109	38.84***
sigmoid	linear	σ GARCH(1, 1)	0	0.00399	0.00524	0.48	8.48***
sigmoid	softplus	σ GARCH(1, 1)	0	0.00522	0.00734	0.147	8.15***
especial	exp	σ GARCH(1, 1)	0	0.00429	0.00622	0.144	8.48***
sigmoid	linear	σ GARCH(1, 1)	7610	0.00456	0.0058	0.525	8.41***
sigmoid	softplus	σ GARCH(1, 1)	7610	0.00325	0.00459	0.129	8.56***
especial	exp	σ GARCH(1, 1)	7610	0.00306	0.00464	0.128	8.7***
sigmoid	linear	σ GARCH(1, 1)	30440	0.00578	0.00833	0.366	7.95***
sigmoid	softplus	σ GARCH(1, 1)	30440	0.00342	0.00454	0.123	8.57***
especial	exp	σ GARCH(1, 1)	30440	0.00376	0.0051	0.14	8.63***
sigmoid	linear	σ GARCH(1, 1)	76100	0.00302	0.00428	0.198	8.6***
sigmoid	softplus	σ GARCH(1, 1)	76100	0.00287	0.00432	0.089	8.74***
especial	exp	σ GARCH(1, 1)	76100	0.00293	0.00436	0.098	8.59***

Tabela 4.2: Comparação de precisão de previsão para diferentes modelos com vários números de opções virtuais para opções de compra.

forma funcional correta e que seu preço previsto quase corresponde ao preço real observado nos dados de mercado. Ambos os modelos nas figuras são apresentados para opção de compra sem o uso de opções virtuais. As opções geralmente têm um erro percentual absoluto alto para um moneyness pequeno, pois as pequenas diferenças do preço correto resultam em um erro percentual alto. Nas figuras 4.7 e 4.11 é mostrado o histograma da diferença entre a previsão e o preço real.

4.3 Previsão com redes neurais modulares

Na seção anterior, comparamos três opções diferentes de rede neural feed forward com o uso de duas medidas de volatilidade. Nesta seção, comparamos a arquitetura feedforward com a arquitetura da rede neural modular, conforme descrito na seção 2.4. A rede neural modular divide os dados de entrada em subconjuntos separados de acordo

Ativação	Output	Volatilidade	Virtual	MAE	RMSE	MAPE	Teste DM
sigmoid	linear	σ_{HV}	0	0.0038	0.00523	0.592	78.09***
sigmoid	softplus	σ_{HV}	0	0.00262	0.00426	0.25	83.77***
especial	exp	σ_{HV}	0	0.00281	0.00426	0.302	83.63***
sigmoid	linear	σ_{HV}	7610	0.00402	0.00557	0.944	75.76***
sigmoid	softplus	σ_{HV}	7610	0.00266	0.00413	0.257	83.81***
especial	exp	σ_{HV}	7610	0.00277	0.00421	0.277	82.54***
sigmoid	linear	σ_{HV}	30440	0.00368	0.00519	0.792	79.76***
sigmoid	softplus	σ_{HV}	30440	0.0027	0.0041	0.286	83.72***
especial	exp	σ_{HV}	30440	0.00287	0.00423	0.284	82.37***
sigmoid	linear	σ_{HV}	76100	0.00323	0.00455	0.731	82.47***
sigmoid	softplus	σ_{HV}	76100	0.00223	0.0035	0.217	84.76***
especial	exp	σ_{HV}	76100	0.00273	0.00411	0.275	83.73***
sigmoid	linear	$\sigma_{GARCH}(1, 1)$	0	0.00602	0.00748	1.2	8.12***
sigmoid	softplus	$\sigma_{GARCH}(1, 1)$	0	0.00297	0.00501	0.296	8.41***
especial	exp	$\sigma_{GARCH}(1, 1)$	0	0.00346	0.00503	0.345	8.15***
sigmoid	linear	$\sigma_{GARCH}(1, 1)$	7610	0.00556	0.0077	1.251	8.09***
sigmoid	softplus	$\sigma_{GARCH}(1, 1)$	7610	0.00304	0.00466	0.323	8.45***
especial	exp	$\sigma_{GARCH}(1, 1)$	7610	0.003	0.00458	0.345	8.18***
sigmoid	linear	$\sigma_{GARCH}(1, 1)$	30440	0.00399	0.00545	0.73	8.37***
sigmoid	softplus	$\sigma_{GARCH}(1, 1)$	30440	0.00282	0.00449	0.275	8.43***
especial	exp	$\sigma_{GARCH}(1, 1)$	30440	0.00298	0.00479	0.309	8.19***
sigmoid	linear	$\sigma_{GARCH}(1, 1)$	76100	0.00346	0.00519	0.566	8.4***
sigmoid	softplus	$\sigma_{GARCH}(1, 1)$	76100	0.00273	0.00437	0.251	8.46***
especial	exp	$\sigma_{GARCH}(1, 1)$	76100	0.00281	0.00449	0.257	8.2***

Tabela 4.3: Comparação de precisão de previsão para diferentes modelos com vários números de opções virtuais para opções de venda.

com o moneyness m e tempo até o vencimento τ . Esta separação é feita afim de separar os dados de entrada em 9 subconjuntos definidos pelas três escolhas de moneyness $m < 0.97, 0.97 \leq m < 1.05, 1.05 \leq m$, e por três escolhas de tempo até o vencimento $\tau < 60, 60 \leq \tau < 180, 180 \leq \tau$.

A rede neural avaliou o padrão nos dados para cada subconjunto separadamente e aprendeu para todos os 9 subconjuntos de previsão diferentes. Cada um desses subconjuntos foi processado por um módulo diferente, que é basicamente uma rede neural autônoma com a escolha de 30 nós em duas camadas ocultas (ligeiramente menor para não super ajustar a rede neural, pois os subconjuntos têm tamanho menor). Depois que todos os módulos da rede são treinados, os dados de teste chegam a toda a rede que alterna o módulo para que o módulo de avaliação correto seja escolhido, gerando o preço da opção. Toda a arquitetura da rede neural modular é representada na parte teórica.

Os resultados são comparados de acordo com as mesmas medidas das seções an-

teriores, ou seja, MAE, RMSE, MAPE e poder de predição em comparação com a fórmula clássica de Black-Scholes pela estatística Diebold-Mariano. Na tabela 4.4 a seguir temos uma avaliação da arquitetura da rede neural modular em comparação com a arquitetura feedforward simples favorecida anteriormente com função de ativação sigmóide nas camadas ocultas e função softplus na saída. Pode-se observar que a previsão fora da amostra em comparação com a fórmula de Black-Scholes é um pouco melhor, pois o teste Diebold-Mariano mostra um valor mais alto. Além disso, ele também supera a rede neural feedforward mencionada em todas as três medidas de erro. Isso confirma a segunda hipótese, que diz que o desempenho da rede neural modular será melhor do que o desempenho da rede feedforward simples.

No entanto, a rede neural modular não é melhor do que a rede neural feedforward em termos de MAE e RMSE se compararmos no caso em que 76100 opções virtuais foram adicionadas. A tabela mencionada leva à conclusão de que, se adicionarmos uma pequena quantidade de opções virtuais, a rede neural modular ainda supera a rede neural feedforward, no entanto, se a quantidade de opções virtuais adicionadas for grande o suficiente como 76100 (que representa mais de 46% do conjunto de dados de opção de compra original e mais de 63% do conjunto de dados de opção de venda), a previsão tem medidas de erro mais baixas para a arquitetura mais simples com opções virtuais adicionadas em vez da rede neural modular.

No entanto, vale a pena mencionar que ambos os modelos têm muito bom poder de previsão e também há a possibilidade de mesclar essas duas abordagens em uma que geraria opções virtuais em alguns modelos específicos, sendo sugerida como uma futura extensão do tema.

Na seguinte tabela 4.5, temos o desempenho dos módulos individuais na estimativa da rede neural modular. Os módulos são classificados de acordo com o moneyness (primeiros três módulos $m < 0.97$, os três módulos do meio $0.97 \leq m < 1.05$ e últimos três módulos $1.05 \leq m$) e também em cada trio de módulos, eles são classificados por tempo até o vencimento em ordem crescente (ou seja, o módulo 1 tem $\tau < 60$, o módulo 2 tem $60 \leq \tau < 180$, etc.).

Como pode ser visto na tabela, os módulos que são treinados em amostras de

Opção	Arquitetura	MAE	RMSE	MAPE	Teste DM	Virtual	Tamanho
Call	modular	0.00308	0.00446	0.085	41.03***	0	162939
Call	softplus	0.00331	0.00461	0.129	39.51***	0	162939
Call	softplus	0.00315	0.00453	0.109	39.71***	7610	170549
Call	softplus	0.00307	0.00429	0.106	39.94***	30440	193379
Call	softplus	0.00300	0.00431	0.103	40.14***	76100	239039
Put	modular	0.00249	0.00400	0.229	84.50***	0	120149
Put	softplus	0.00262	0.00426	0.25	83.77***	0	120149
Put	softplus	0.00266	0.00413	0.257	83.81***	7610	127759
Put	softplus	0.00270	0.00410	0.286	83.72***	30440	150589
Put	softplus	0.00223	0.00350	0.217	84.76***	76100	196249

Tabela 4.4: Comparação da precisão de previsão da rede neural modular com a rede feedforward simples cumprindo as suposições de previsões racionais

Opção	Módulo	MAE	RMSE	MAPE	Teste DM	Tamanho do Treinamento
Compra	módulo 1	0.009	0.00915	0.523	7.48***	5415
Compra	módulo 2	0.00438	0.00536	0.182	9.04***	7502
Compra	módulo 3	0.0026	0.00347	0.175	11.03***	21048
Compra	módulo 4	0.01377	0.01609	0.235	8.02***	20042
Compra	módulo 5	0.00142	0.00168	0.009	42.8***	10248
Compra	módulo 6	0.006	0.00645	0.204	32.8***	11477
Compra	módulo 7	0.00183	0.00242	0.013	49.4***	27701
Compra	módulo 8	0.00254	0.00313	0.013	46.59***	20822
Compra	módulo 9	0.00312	0.00337	0.013	43.75***	38678
Venda	módulo 1	0.02775	0.03723	0.441	4.64***	3384
Venda	módulo 2	0.02186	0.02706	0.374	5.24***	3296
Venda	módulo 3	0.01663	0.01989	0.213	6.96***	6880
Venda	módulo 4	0.00602	0.00646	0.193	33.25***	19264
Venda	módulo 5	0.00625	0.00685	0.185	34.41***	10229
Venda	módulo 6	0.00617	0.00705	0.172	46.28***	11468
Venda	módulo 7	0.00133	0.0016	0.126	89.82***	13101
Venda	módulo 8	0.00352	0.00447	0.152	78.9***	15145
Venda	módulo 9	0.00281	0.00411	0.185	85.68***	37377

Tabela 4.5: Precisão de previsão para módulos específicos na rede neural modular

tamanho pequeno têm um desempenho relativamente ruim em comparação com aqueles com amostras maiores. Os três primeiros módulos 1-3 são os principais para os casos de opção de compra e venda. Quanto maior for o moneyness, no entanto, maior também é a amostra de treinamento no módulo e, no final, resulta em uma previsão fora da amostra muito melhor para os módulos 7 a 9, que superam até mesmo os melhores cenários da rede feedforward simples.

4.4 Padrão proposto para escolha da arquitetura

Um dos objetivos deste trabalho é descobrir como estimar o tamanho certo e a forma funcional correta da rede neural. Isso é muito desejado pela literatura existente, já que a maioria dos artigos de pesquisa sobre precificação de opções se concentram mais na comparação dos métodos paramétricos e não paramétricos do que na escolha dos parâmetros das redes neurais. Este trabalho segue os métodos feitos em vários artigos recentes sobre o assunto e se propõe a estabelecer um padrão a partir dos resultados obtidos.

Como o número de parâmetros que podem ser ajustados no caso de predição de redes neurais é relativamente grande, o método proposto na escolha dos parâmetros ótimos é dividir a escolha da arquitetura em duas etapas principais e depois fazer os ajustes necessários.

O primeiro grande passo é avaliar o tamanho certo da rede, o que pode ser feito usando algumas das regras empíricas usadas no aprendizado de máquina ou uma nova regra empírica pode ser estimada pelo método usado na seção 3.0.3. Este método simula opções completamente novas que correspondem ao conjunto de dados real apenas pela escolha dos parâmetros do ativo, mas nada do conjunto de dados real da opção em relação ao preço da opção foi usado. O preço das opções geradas é calculado pela fórmula de Black-Scholes.

A próxima etapa é estimar uma série de modelos de rede neural com uma arquitetura de escolha do próprio usuário, mas com um número variável de camadas e nós ocultos em cada camada. Avaliamos também a técnica de dropout que descarta uma parte dos nós que têm pouco ou nenhum peso após a retropropagação. A etapa seguinte é

executar uma regressão MQO do erro de predição fora da amostra nos parâmetros da rede neural, que avalia o efeito da adição de camadas ocultas ou número de neurônios em cada camada. É importante executar vários modelos com tamanhos diferentes de amostras de treinamento, pois com o aumento da amostra de treinamento o erro tende a diminuir.

O segundo grande passo é escolher a função de ativação correta e a função de saída. Isso depende parcialmente da resposta que o usuário obtém na primeira etapa e oferece uma grande quantidade de possibilidades de como escolher a função de ativação ideal. Os resultados deste trabalho analisaram três escolhas diferentes e mostraram que é muito melhor usar a função softplus como a função de saída em vez da saída linear. O desempenho da função softplus é mostrado como o melhor neste trabalho, no entanto, apenas um pouco melhor e de sucesso semelhante ao da arquitetura escolhida por (CULKIN; DAS, 2017), que usou a função de saída exponencial e uma escolha especial de funções de ativação para cada camada.

Um passo importante na escolha da função de ativação e de saída é encontrar a forma funcional da função de ativação que respeita as hipóteses de predição racional mencionadas por (YANG; ZHENG; HOSPEDALES, 2017). Essas hipóteses foram verificadas individualmente na seção 2.5 deste trabalho para nossa escolha de arquitetura. Este procedimento inclui a geração de opções virtuais para atender a duas dessas premissas. Essa técnica provou ser muito eficaz e melhorou de maneira geral o desempenho de todos os modelos com diferentes funções de ativação e saídas. Outra possibilidade é impor uma estrutura modular, o que também melhorou o desempenho do modelo que usamos comparativamente com a adição das opções virtuais.

Um dos possíveis aprimoramentos desse modelo é encontrar uma maneira de combinar a regra empírica com as regras normalmente usadas no aprendizado de máquina. Outro pode ser combinar o uso de redes neurais modulares com o cumprimento de hipóteses racionais que resultariam em melhor desempenho em alguns módulos pela presença de opções virtuais nos subconjuntos de dados. Isso não foi examinado por este trabalho, mas certamente produziria resultados interessantes.

4.5 Discussão dos resultados

A primeira hipótese se mostrou verdadeira em todos os testes que foram feitos com qualquer arquitetura de rede neural testada. O poder de previsão fora da amostra foi medido pelo teste Diebold-Mariano e esta medida provou que a previsão de preço da rede neural era significativamente melhor do que a fórmula de Black-Scholes.

A segunda hipótese foi confirmada na seção 4.3, onde foi mostrado, que a rede neural modular com os mesmos parâmetros da rede neural feed forward superou a última caso não houvesse opções virtuais adicionadas. A adição de opções virtuais melhorou ligeiramente o desempenho da rede neural feedforward e, em seguida, tornou-se mais precisa até mesmo do que a rede neural modular.

A terceira hipótese examinou se o cumprimento das premissas de predição racional melhora a rede. Foi mostrado na parte teórica na seção 2.5, que a escolha da arquitetura com a função softplus na saída e função de ativação sigmóide nas camadas ocultas satisfaz pelo menos 5 das 6 hipóteses de predição racional.

Isso é comparado na seção de resultados 4.2, onde as opções virtuais são adicionadas e duas outras configurações, foram comparadas à arquitetura mencionada com a função softplus na saída. A terceira hipótese mostrou-se correta, pois as opções virtuais melhoraram o desempenho em pelo menos 10% em termos de MAE e RMSE. A última quarta hipótese também se mostra válida pela comparação de dois métodos de previsão de volatilidade. A volatilidade histórica que foi baseada na janela móvel de preços históricos de 5, 20, 60, ou 100 dias de negociação, mostrou que pode ser em vários casos superada pelo método GARCH (1,1). No entanto, nenhum dos métodos de previsão de volatilidade se mostrou predominantemente melhor para todos os tipos de opções. No caso de opções de venda, a previsão de preço usando a volatilidade histórica apresentou um erro geral inferior em termos de todas as três medidas MAE, RMSE e MAPE.

Ao final, foi formulado o modelo da seção anterior, o que mostra ser razoável executar várias simulações de testes com os dados gerados pela fórmula de Black-Scholes, que avalia o tamanho ideal da arquitetura da rede neural na precificação de opções. Este teste de resolução não foi realizado no caso de rede neural modular, o que também pode ser considerado uma expansão deste trabalho para tornar o modelo mais confiável. Após

a estimativa do tamanho ideal, as várias opções de função de ativação e função de saída foram comparadas nos dados reais com número fixo de camadas e nós ocultos. Essa abordagem tornou toda a estimativa da rede neural significativamente mais fácil em termos de poder computacional, pois há um número significativamente menor de escolhas possíveis. Embora essa abordagem tenha simplificado significativamente a escolha da arquitetura, ainda há muitas opções possíveis, pois apenas um pequeno subconjunto de todas as opções de arquitetura foi investigado neste trabalho.

5 Conclusão

Já se passaram mais de 40 anos desde que a famosa fórmula de precificação de opções foi desenvolvida por Black e Scholes. Muitas melhorias em sua fórmula foram feitas por pesquisadores acadêmicos e profissionais desde então. Um dos pontos críticos para a teoria de precificação de opções veio com o lançamento do artigo com uma nova abordagem adotada por (HUTCHINSON; LO; POGGIO, 1994) que usou um método de rede neural para prever o preço da opção.

Este método foi seguido por muitos outros pesquisadores, pois as redes neurais são capazes de reconstruir com muita precisão qualquer função contínua arbitrária apenas a partir da estrutura dos dados que são fornecidos. Muitos pesquisadores interessados em precificação de opções compararam os resultados da fórmula clássica de Black-Scholes e a nova abordagem não paramétrica usando redes neurais para vários tipos de opções e diferentes conjuntos de dados.

No entanto, a literatura atual sobre o tema ainda não apresenta respostas claras sobre as questões relacionadas aos problemas de aprendizado de máquina, como a escolha ótima da arquitetura e o tamanho da rede neural. Existem também vários tipos de volatilidade que podem ser usados como um mecanismo de previsão que desempenha um papel fundamental na precificação de opções.

Como foi mostrado por uma vasta literatura que na maioria dos casos as redes neurais são melhores mecanismos de previsão do que a fórmula de Black-Scholes de forma fechada, é muito importante ter essa abordagem o mais precisa possível para descobrir o verdadeiro preço das opções.

Isso leva este trabalho a examinar métodos que são testados com o objetivo específico de aumentar o poder de predição das redes neurais. Isso é feito por meio da investigação de várias melhorias na configuração de rede feedforward mais simples.

A contribuição deste trabalho inclui os seguintes resultados. Primeiramente, um novo modelo na escolha do tamanho e arquitetura da rede neural foi proposto com base nos resultados deste trabalho. Esse modelo permite avaliar o quão grande deve ser a

arquitetura para minimizar o erro de previsão de preço.

Em segundo lugar, foi confirmado que todas as escolhas da arquitetura de rede neural feitas neste trabalho levaram a um maior poder de previsão e, portanto, a uma melhor previsão de preços. Além disso, a principal contribuição está na avaliação do efeito de adicionar a hipótese de previsão racional ao modelo que força a rede neural a aprender alguns conceitos teóricos-chave da teoria de precificação de opções e, portanto, tem melhor desempenho na precificação.

Em Terceiro, um conceito de redes neurais modulares foi revisado e mostrado que essa abordagem leva a uma melhor previsão do que as redes simples de feed forward.

E por último foi mostrado que o desempenho da rede neural pode variar em relação ao mecanismo de previsão de volatilidade utilizado. Neste trabalho, a volatilidade histórica e o método de estimativa GARCH foram usados e foi mostrado que embora ambos os mecanismos produzam resultados geralmente melhores para redes feedforward do que para a fórmula de Black-Scholes, nenhuma das volatilidades usadas é significativamente dominante sobre a outra e, portanto, em alguns dos modelos investigados é melhor usar a volatilidade histórica, enquanto em outros modelos a estimativa de volatilidade GARCH é melhor.

Embora muitas melhorias possíveis para redes neurais simples tenham sido examinadas, ainda há uma grande quantidade de trabalho pela frente na precificação de opções em termos de uso de redes neurais mais avançadas e arquiteturas de redes neurais modulares melhores. Uma das melhorias sugeridas diretamente aos resultados deste trabalho é combinar a abordagem de geração de opções virtuais com a modularidade, o que implicaria que alguns dos módulos da rede neural modular teriam um poder de predição muito melhor.

Além disso, como cada módulo na rede neural modular estimou seu próprio mecanismo de previsão, em que cada dois módulos vizinhos devem satisfazer a previsão em seus limites e estar alinhados com a previsão dos módulos opostos. A resposta para essa condição também poderia ser encontrada no método da teoria dos elementos finitos amplamente utilizado na modelagem de equações diferenciais parciais. Além disso, alguns autores sugerem o uso de uma classe diferente de algoritmos genéticos para realizar a

precificação de opções ao invés das escolhas que fizemos neste trabalho.

Também existem vários mecanismos de previsão de volatilidade, alguns dos quais foram resumidos na parte teórica, no entanto, este trabalho investigou este tópico apenas superficialmente e, portanto, seria interessante usar modelos mais sofisticados em combinação com os outros conceitos usados neste trabalho em trabalhos futuros.

Em suma, vários métodos foram examinados por este trabalho, entretanto, o campo de precificação de opções ainda pode ser investigado de forma muito mais profunda com o uso de redes neurais mais avançadas e diferentes intervalos de tempo dos dados. Este trabalho foi um avanço do artigo (UGARTE; MOURA; FACCO, 2021) no estudo de precificação de opções utilizando a linguagem Python. Conforme já mencionado, ainda há muito a ser explorado nesta área e pesquisas mais avançadas serão desenvolvidas futuramente.

Bibliografia

- AZAM, F. Biologically inspired modular neural networks. *McGraw-Hill New York*, 2007.
- BAKSHI, G.; CAO, C.; CHEN, Z. Empirical performance of alternative option pricing models. *The Journal of finance*, v. 52, n. 5, p. 2003–2049, 1997.
- BARUNÍKOVA, M.; BARUNÍK, J. Neural networks as a semiparametric option pricing tool. 2011.
- BATES, D. S. Empirical option pricing: A retrospection. *Journal of Econometrics*, v. 116, n. 1-2, p. 387–404, 2003.
- BLACK, F.; SCHOLES, M. The pricing of options and corporate liabilities. *Journal of political economy*, v. 81, n. 3, p. 637–654, 1973.
- BRIDLE, J. S. Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In: *In Neurocomputing*. [S.l.: s.n.], 1990. p. 227–236.
- CULKIN, R.; DAS, S. R. Machine learning in finance: The case of deep learning for option pricing. *Journal of Investment Management*, v. 15, n. 4, p. 92–100, 2017.
- DIEBOLD, F.; MARIANO, R. S. Comparing predictive accuracy. p. 253–263, 1995.
- EINSTEIN, A. Threshold heteroskedastic models. *Investigations on the Theory of the Brownian Movement*, 1956.
- FIGLEWSKI, S. Forecasting volatility. *Financial markets, institutions & instruments*, v. 6, n. 1, p. 1–88, 1997.
- GEIGLE, D. S.; ARONSON, J. E. An artificial neural network approach to the valuation of options and forecasting of volatility. *Journal of Computational Intelligence in finance*, v. 7, n. 6, p. 19–25, 1999.
- GHAZIRI, H.; ELFAKHANI, S.; ASSI, J. Neural networks approach to pricing options. *Neural Network World*, v. 10, n. 1, p. 271–277, 2000.
- GLOROT, X.; BORGES, A.; BENGIO, Y. Deep sparse rectifier neural networks. In: *Proceedings of the fourteenth international conference on artificial intelligence and statistics*. [S.l.: s.n.], 2011. p. 315–323.
- GRADOJEVIC, J.; GENCAJ, R.; KUKOLJ, D. Option pricing with modular neural networks. *IEEE Transactions on Neural Networks*, v. 20, n. 6, p. 626–637, 2009.
- HAHN, T. Option pricing using artificial neural networks : an australian perspective. *Bond University*, 2013.
- HAUG, E. G. The complete guide to option pricing formulas, volume 2. *Ph.D thesis, Virginia Tech*, 2000.

- HORNIK, K. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, v. 4, n. 2, p. 251–257, 1991.
- HUTCHINSON, J. M.; LO, A. W.; POGGIO, T. A nonparametric approach to pricing and hedging derivative securities via learning networks. *Journal of Finance*, v. 49, n. 3, p. 851–889, 1994.
- JACOBS, K.; CHISTOFFERSEN, P. Which garch model for option valuation? *Management science*, v. 50, n. 9, p. 1204–1221, 2004.
- JAND, J. S. R.; SUN, T.; MIZUTANI, E. Neuro-fuzzy and soft computing; a computational approach to learning and machine intelligence. 1997.
- JORDAN, M. I.; JACOBS, R. A. Hierarchical mixtures of experts and the em algorithm. *Neural computation*, v. 6, n. 2, p. 181–214, 1994.
- MCCULLAGH, P.; NELDER, J. A. *Generalized linear models*,. [S.l.]: CRC press, 1989. v. 37.
- NELSON, D. B. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, p. 347–370, 1991.
- SAITO, N.; JUN, L. Neural network option pricing in connection with the black and scholes model. In: *Proceedings of the Fifth Conference of the Asian Pacific Operations Research Society*. Singapore, 2000: [s.n.], 2000.
- SCOTT, L. O. Option pricing when the variance changes randomly: Theory, estimation, and an application. *Journal of Financial and Quantitative analysis*, v. 22, n. 4, p. 419–438, 1987.
- UGARTE, G. L.; MOURA, A. S. de; FACCO, W. G. Simulando movimento browniano geom etrico em python. *Proceeding Series of the Brazilian Society of Computational and Applied Mathematics*, 2021.
- YANG, Y.; ZHENG, Y.; HOSPEDALES, T. M. Gated neural networks for option pricing: Rationality by design. *AAAI*, p. 52–58, 2017.
- ZAKOIAN, J. M. Threshold heteroskedastic models. *Journal of Economic Dynamics and control*, v. 18, n. 5, p. 931–955, 1994.