

**INTELIGÊNCIA ARTIFICIAL APLICADA POR MEIO DE DEEP LEARNING
(REDES NEURAIS) E PELO MÉTODO TRADICIONAL ECONOMETRICO DE
REGRESSÃO POR MÍNIMOS QUADRADOS ORDINÁRIOS**

***INTELIGENCIA ARTIFICIAL APLICADA MEDIANTE DEEP LEARNING (REDES
NEURONALES) Y EL MÉTODO TRADICIONAL ECONOMETRICO DE REGRESIÓN
POR MÍNIMOS CUADRADOS ORDINARIOS***

***ARTIFICIAL INTELLIGENCE APPLIED BY MEANS OF DEEP LEARNING
(NEURAL NETWORKS) AND BY THE TRADITIONAL ECONOMETRIC METHOD OF
ORDINARY LEAST SQUARES REGRESSION***



Leandro Pereira da SILVA¹
e-mail: leandro.p.silva@unesp.br

Como referenciar este artigo:

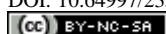
SILVA, Leandro Pereira da. Inteligência artificial aplicada por meio de Deep Learning (redes neurais) e pelo método tradicional econométrico de regressão por mínimos quadrados ordinários. **Rev. Iniciativa Econômica**, Araraquara, v. 11, n. 00, e025003, 2025. e-ISSN: 2358-5951. DOI: 10.64997/2358-5951-19880



| Submetido em: 06/12/2024
| Revisões requeridas em: 09/12/2024
| Aprovado em: 16/07/2025
| Publicado em: 23/12/2025

Editor: Prof. Dr. Gustavo Pereira Serra

¹ Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), Araraquara – SP – Brasil. Doutorando em Economia na Universidade Paulista (UNESP).



RESUMO: O campo de aplicação para instrumentos que envolvem algoritmos de inteligência artificial (IA) está cada vez mais abrangente no mundo atual. Sendo que em praticamente todas as ciências, esse ramo de aprendizado avança em várias áreas e especificidades. Dessa forma, esta pesquisa teve como objetivo verificar, em termos aplicados e didáticos, se a metodologia tradicional econométrica de regressões por mínimos quadrados ordinários (MQO) e a metodologia de redes neurais (RN) têm o mesmo desempenho em situações diversas de testes empíricos para série de dados sobrepostos em algoritmos de inteligência artificial. O problema da pesquisa foi saber se o método MQO e o método RN, em geral, apresentam resultados satisfatórios em estimativas estatísticas e para certos tipos de algoritmos de IA. A hipótese da pesquisa é que os métodos tradicionais, como o MQO, funcionam melhor para dados estruturados, enquanto redes neurais destacam-se com dados complexos/aleatórios. Em prévia, os testes econométricos quantitativos realizados demonstraram que a utilização das metodologias RN e MQO são abordagens que podem ser utilizadas em conjunto quando aplicadas em algoritmos de IA. Os resultados quantitativos ainda demonstraram que a técnica RN, em geral, apresenta melhores resultados para dados com características de variáveis aleatórias e o método tradicional econométrico MQO apresenta, em geral, melhores resultados para quando os dados observados têm características de variáveis “mais comportadas” e menos aleatórias em termos matemáticos. Sendo ainda, que o critério principal utilizado para mensuração de desempenho foi o indicador de Mínimos Quadrados Médio (MSE - *Mean Squared Error*), além também dos indicadores de Raiz Quadrada da Média dos Erros ao Quadrado (RMSE - *Root Mean Squared Error*), Erro Absoluto Médio (MAE - *Mean Absolute Error*) e o Coeficiente de Determinação R^2 .

PALAVRAS-CHAVE: Algoritmos de IA. Redes Neurais. Econometria.

CLASSIFICAÇÃO JEL: B23, C5, C01.

RESUMEN: El campo de aplicación de los instrumentos que involucran algoritmos de inteligencia artificial (IA) es cada vez más amplio en el mundo actual, ya que en prácticamente todas las ciencias esta rama del aprendizaje avanza en múltiples áreas y especificidades. De este modo, esta investigación tuvo como objetivo verificar, en términos aplicados y didáticos, si la metodología econométrica tradicional de regresiones por mínimos cuadrados ordinarios (MCO) y la metodología de redes neuronales (RN) presentan el mismo desempeño en distintas situaciones de pruebas empíricas con series de datos superpuestas en algoritmos de inteligencia artificial. El problema de investigación consistió en determinar si el método MCO y el método RN, en general, producen resultados satisfactorios en estimaciones estadísticas y para ciertos tipos de algoritmos de IA. La hipótesis de la investigación es que los métodos tradicionales, como el MCO, funcionan mejor con datos estructurados, mientras que las redes neuronales se destacan en el tratamiento de datos complejos o aleatorios. De manera preliminar, las pruebas econométricas cuantitativas realizadas demostraron que el uso de las metodologías RN y MCO constituye enfoques que pueden emplearse de forma conjunta cuando se aplican en algoritmos de IA. Los resultados cuantitativos también evidenciaron que la técnica de RN, en general, presenta mejores resultados para datos con características de variables aleatorias, mientras que el método econométrico tradicional MCO, por lo general, ofrece mejores resultados cuando los datos observados presentan características de variables “más comportadas” y menos aleatorias en términos matemáticos. Asimismo, el principal criterio utilizado para la medición del desempeño fue el indicador de Error Cuadrático Medio (MSE – *Mean Squared Error*), además de los indicadores de Raíz del Error Cuadrático Medio

(RMSE – Root Mean Squared Error), Error Absoluto Medio (MAE – Mean Absolute Error) y el Coeficiente de Determinación R^2 .

PALABRAS CLAVE: Algoritmos de IA. Redes neuronales. Econometría.

CLASIFICACIÓN JEL: B23, C5, C01.

ABSTRACT: The field of application for instruments involving artificial intelligence (AI) algorithms is increasingly broad in today's world. In practically all sciences, this branch of learning advances in several areas and specificities. Thus, this research aimed to verify, in applied and didactic terms, whether the traditional econometric methodology of ordinary least squares (OLS) regressions and the neural network (NN) methodology have the same performance in different situations of empirical tests for overlapping data series in artificial intelligence algorithms. The research problem was to know whether the OLS method and the NN method, in general, present satisfactory results in statistical estimates and for certain types of AI algorithms. The research hypothesis is that traditional methods, such as OLS, work better for structured data, while neural networks excel with complex/random data. In advance, the quantitative econometric tests carried out demonstrated that the use of the NN and OLS methodologies are approaches that can be used together when applied to AI algorithms. The quantitative results also demonstrated that the RN technique, in general, presents better results for data with characteristics of random variables. And the traditional econometric method MQO presents, in general, better results when the observed data have characteristics of “more well-behaved” and less random variables in mathematical terms. Furthermore, the main criterion used to measure performance was the Mean Squared Error (MSE) indicator, in addition to the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Determination R^2 indicators.

KEYWORDS: AI Algorithms. Neural Networks. Econometric.

JEL CLASSIFICATION: B23, C5, C01.

Introdução

Os avanços em algoritmos de Inteligência Artificial (IA) continuam prosseguindo em um ritmo cada vez maior tanto na questão teórica e acadêmica como em suas diversas aplicações práticas, sendo que hoje em dia os recursos necessários para tal prática requerem recursos humanos e teóricos como algoritmos diversos, grande volume de dados e equipamentos de hardware de alto desempenho. Já as técnicas e metodologias utilizadas que envolvem algoritmos de IA são diversas, como por exemplo, a Aprendizagem Supervisionada (*Supervised Learning* – SL), a Aprendizagem Não Supervisionada (*Unsupervised Learning* – UL), o Aprendizado de Máquina (*Machine Learning* – ML), a Aprendizagem por Reforço (*Reinforcement Learning* – RL) e o Aprendizado Profundo (*Deep Learning* – DL). Tendo ainda como principais instrumentos as redes neurais densas (DNNs), as redes neurais recorrentes (RNNs) e as redes neurais convolucionais (CNNs), além de Regressão por mínimos quadrados ordinários, Regressão Lasso, *Random Forest*, entre outros.

Esta pesquisa utilizou algoritmos de IA por meio de instrumentos de Redes Neurais (RN) e por meio do método tradicional econométrico de regressão por Mínimos Quadrados Ordinários (MQO), como etapa metodológica aplicada em exercícios estatísticos e didáticos, com a finalidade de validar o estudo. O objetivo da pesquisa foi verificar se esses dois instrumentos têm o mesmo desempenho em situações diversas de testes empíricos para diferentes séries de dados. Dessa forma, o problema de pesquisa foi: dado que o método de regressão MQO oferece, em geral, resultados satisfatórios em estimativas estatísticas e em algoritmos de IA, porque então, usar Redes Neurais, algoritmos de *Machine Learning* (ML) e *Deep Learning* (DP) para estimar o desempenho de resultados estimados? A resposta não é simples e não está “fechada” em termos acadêmicos. Porém, a hipótese é que os algoritmos de IA que usam redes neurais podem ser capazes — em algumas situações — de solucionar ineficiências onde os métodos tradicionais estatísticos — como a regressão MQO — falham, o que não necessariamente levaria a descartar o método tradicional de regressão MQO. Já a justificativa para a pesquisa deve-se em função de que os estudos que tratam de Inteligência Artificial, já há algum tempo estão sendo discutidos nos meios acadêmicos, porém, ainda estão avançando em termos de eficiência técnica. Ou seja, ainda não se tem modelos plenos de IA, como por exemplo, para uma precificação totalmente satisfatória de ativos financeiros. Dessa forma, a metodologia adotada nesta pesquisa busca realizar testes estatísticos empíricos por meio de modelos de IA que utilizam algoritmos de Aprendizagem Supervisionada e Aprendizado de Máquina.

Em prévia quanto aos resultados alcançados por meio desta pesquisa, nota-se que quando da utilização de base dados com características de variáveis de trajetórias “mais comportadas” e menos aleatórias, tem-se que os algoritmos de IA que utilizam o método econométrico tradicional MQO, em geral, apresentam melhor desempenho. Quando os dados observados têm características de variáveis com trajetórias “menos comportadas” e mais aleatórias, o algoritmo de IA que utiliza instrumentos de redes neurais têm, em geral, melhor desempenho do que se comparado com a metodologia de regressão MQO. As métricas para mensurar o desempenho e performance das técnicas de aproximações matemáticas (MQO e RN), que compara os dados e valores reais da amostra para com os seus respectivos valores estimados, foram os seguintes indicadores estatísticos: Erro Médio Quadrado (MSE), Raiz Quadrada da Média dos Erros ao Quadrado (RMSE), Erro Absoluto Médio (MAE) e o Coeficiente de Determinação (R^2). Para validar tais mensurações de desempenho, foram também realizadas nesta pesquisa as análises estatísticas formais de significância dos resultados encontrados, no qual foram utilizados o Teste *T* pareado e realizado a Análise de Robustez, entre outros testes e inferências estatísticas.

A composição estrutural deste texto — além desta introdução e do resumo — tem na sequência do texto o referencial teórico, que procura resgatar os principais aspectos dos conceitos que envolvem algoritmos de mínimos quadrados ordinários e de redes neurais, como instrumentos utilizados em modelagem de tecnologia IA. O referencial teórico ainda resgata as principais técnicas e métodos abordados para mensurar o desempenho de aproximações lineares e não lineares de funções matemáticas e regressões diversas, tendo destaque os indicadores MSE, RMSE, MAE e o Coeficiente de determinação R^2 . Também será abordado nesse momento da pesquisa críticas e limitações da modelagem utilizada.

Na sequência estrutural do texto é apresentada a seção de metodologia, tendo destaque os tipos diferentes de modelagens de IA utilizadas nesta pesquisa, em que nesta seção é demonstrada a construção dos modelos MQO e RN, com a finalidade de realizar as devidas comparações a serem consideradas neste estudo. Esta pesquisa ainda traz a seção de resultados apontando as principais observações e inferências realizadas pelos métodos quantitativos utilizados, e serão apresentados os testes estatísticos formais de significância, Teste *t* Pareado, Teste de normalidade dos resíduos (Shapiro-Wilk), Teste de Homoscedasticidade (Breusch-Pagan) e a análise de robustez. Ao final do texto, tem-se as considerações finais e o referencial bibliográfico.

Referencial Teórico

Esta pesquisa tem a característica de observar metodologias diferentes para se avaliar o desempenho de resultados notados por certos métodos quantitativos utilizados em algoritmos de IA — no qual foi delimitado os possíveis tipos de algoritmos e metodologias utilizadas em Aprendizado Profundo (*Deep Learning* - *DL*) via Redes Neurais (RN) e o método tradicional de regressão por meio de Mínimos Quadrados Ordinários (MQO) —, sendo mensurados pelo parâmetro estatístico de Erro Médio Quadrado (MSE - *Mean Squared Error*), adotado nesta pesquisa como o principal indicador de desempenho dos algoritmos de IA. Para efeitos de comparação e aprofundamento da análise, são também utilizados os indicadores de Raiz Quadrada da Média dos Erros ao Quadrado (RMSE - *Root Mean Squared Error*), Erro Absoluto Médio (MAE - *Mean Absolute Error*) e o Coeficiente de Determinação R^2 .

Para uma visão geral dos principais algoritmos aplicados em IA, de acordo com Hipisch (2020) e Galit (2023), os tipos mais comuns de aprendizagem utilizados nessas metodologias são: i. Aprendizagem Supervisionada (*Supervised Learning* – *SL*); ii. Aprendizagem Não Supervisionada (*Unsupervised Learning* - *UL*); iii. Aprendizado de Máquina (*Machine Learning* - *ML*); iv. Aprendizagem por Reforço (*Reinforcement Learning* – *RL*); e v. Aprendizado Profundo (*Deep Learning* - *DL*). Os principais instrumentos utilizados são: a. Redes neurais densas (DNNs); b. Redes neurais recorrentes (RNNs); c. Redes neurais convolucionais (CNNs); d. Regressão por mínimos quadrados (MQO); e. Regressão Lasso, f. Random Forest, e outros.

Regressão por mínimos quadrados ordinários (MQO)

Muitos problemas estatísticos começam com a relação de dados entre variáveis de interesse. Por exemplo, pode-se querer entender o quanto a taxa de juros básica de um determinado país influencia no índice de inflação. Neste caso, pode-se assumir que a taxa de juros é uma variável explicativa (x) que ajuda a entender o impacto que esta causa na variável independente (y) — que nesse caso é a inflação. Assim, de acordo com Lykkegaard *et al.* (2025), Neto (2013), Gujarati (2011), e, especificamente, Wooldridge (2023, p. 20-23), pode-se resolver essa ambiguidade por meio de uma equação que relaciona y a x , como:

$$y = \beta_0 + \beta_1 x + \mu \quad (1),$$

que é um modelo de regressão linear simples, em que μ é o termo de erro ou perturbação da relação e representa outros fatores — além de x —, que afetam y , no qual assumimos que μ tem média ou esperança igual a zero ($E(\mu) = 0$).

Assim, de acordo com Wooldridge (2023, p. 20-23), a derivação dos mínimos quadrados depende de uma amostra da população de dados, que é utilizada para estimar os parâmetros β_0 e β_1 da equação (1) acima. Ao considerar $\{(x_i, y_i): i = 1, \dots, n\}$ como uma amostra aleatória, tem-se:

$$y_i = \beta_0 + \beta_1 x_i + \mu_i \quad (2),$$

para cada valor de i , em que μ_i é o termo de erro para a observação i (Lykkegaard *et al.* 2025; Gujarati, 2011; Neto, 2013; Wooldridge, 2023, p. 24). Assim, ao assumir:

$$E(\mu) = 0 \quad (3)$$

Adicionado as devidas propriedades de variância:

$$\text{Cov}(x, \mu) = E(x\mu) = 0, \quad (4)$$

Para as variáveis observáveis x e y e dos parâmetros desconhecidos β_0 e β_1 , as equações (3) e (4) tem-se:

$$(y - \beta_0 - \beta_1 x) = 0 \quad (5)$$

e

$$E[x(y - \beta_0 - \beta_1 x)] = 0 \quad (6)$$

Como há dois parâmetros desconhecidos para estimar, pode-se esperar que as equações (5) e (6) pudessem ser usadas para obter bons estimadores de β_0 e β_1 (Lykkegaard *et al.* 2025; Wooldridge, 2023, p. 25). Dada uma amostra de dados, pode-se escolher as estimativas $\hat{\beta}_0$ e $\hat{\beta}_1$ para resolver as equivalentes amostrais de (5) e (6):

$$n^{-1} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad (7)$$

$$n^{-1} \sum_{i=1}^n x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad (8)$$

Resolvendo para β_0 e β_1 , e usando as propriedades básicas do operador somatório, a equação (7) pode ser reescrita como:

$$\underline{y} = \hat{\beta}_0 + \hat{\beta}_1 \underline{x} \quad (9),$$

em que $\underline{y} = n^{-1} \sum_{i=1}^n y_i$ é a média amostral de y_i e igualmente para \underline{x} . Essa equação permite escrever $\hat{\beta}_0$ em termos de $\hat{\beta}_1$, \underline{y} e \underline{x} :

$$\hat{\beta}_0 = \hat{\beta} - \hat{\beta}_1 \underline{x}. \quad (10)$$

Resolvendo e simplificando o processo matemático por meio de Wooldridge (2023, p. 26), tem-se como inclinação estimada a covariância amostral entre x_i e y_i dividida pela variância amostral de x_i , que pode ser observada usando álgebra simples ou reescrever $\hat{\beta}_1$ como:

$$\hat{\beta}_1 = \hat{p}_{xy} \cdot (\hat{\sigma}_y / \hat{\sigma}_x). \quad (11),$$

em que \hat{p}_{xy} é a correlação amostral entre x_i e y_i , e $\hat{\sigma}_x$ e $\hat{\sigma}_y$ correspondem aos desvios padrão da amostra. Dessa forma, as estimativas observadas em (9) e (11) são chamadas de estimativas de mínimos quadrados ordinários (MQO) de β_0 e β_1 , no qual Wooldridge (2023, p. 27) e Lykkegaard *et al.* (2025) observam que o resíduo para a observação i é a diferença entre o valor verdadeiro de y_i e seu valor estimado:

$$\hat{u}_i = y_i - \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i. \quad (12)$$

Ao apontar os principais aspectos construtivos do modelo MQO, necessita-se agora de parâmetros indicativos para verificar a eficiência de tal modelo, sendo que para essa pesquisa foi utilizado o MSE como principal indicador de desempenho. Além desses, também foram utilizados os indicadores RMSE, MAE e R2 — esses indicadores também são utilizados para testar a eficiência do próximo processo que envolve Redes Neurais.

Redes Neurais

Os algoritmos e os instrumentos utilizados em tecnologias de IA são diversos, como é o caso, por exemplo, das regressões MQO, o método Monte Carlo, *Random Forest*, entre outros. Porém, uma abordagem que vem se destacando muito no meio acadêmico é a utilização de Aprendizado Profundo (*Deep Learning*) por meio de redes neurais (RN), em que, apesar da complexidade que envolve essa técnica, as redes neurais foram inspiradas no funcionamento do cérebro humano — e é por essa razão que se tem o nome de inteligência artificial (IA). No qual, assim como nos humanos, a experiência relativa sobre um determinado assunto se estabelece em um processo de reconhecimento desse contexto e se efetiva como uma reação quanto ao aspecto que foi relacionado. De acordo com Galit (2023, p. 276), as redes neurais são modelos de classificação e previsão que têm sido bem-sucedidas em termos de precisão preditiva. Segundo Scheuch, Voigt e Weiss (2023, p. 181):

As redes neurais propagam informações de uma camada de entrada, através de uma ou múltiplas camadas ocultas, até uma camada de saída. E embora o número de unidades (neurônios) na camada de entrada seja igual à dimensão dos preditores, a camada de saída geralmente consiste em um neurônio (para regressão) ou vários neurônios para classificação. A camada de saída prevê os dados futuros, semelhante ao valor ajustado em uma análise de regressão. As redes neurais têm bases teóricas como aproximadores universais para qualquer associação preditiva suave. Sua complexidade, no entanto, classifica as redes neurais entre as ferramentas de *Machine Learning* menos transparentes, menos interpretáveis e mais altamente parametrizadas. A especificação de uma rede neural requer, no mínimo, uma postura sobre a profundidade (número de camadas ocultas), a função de ativação, o número de neurônios, a estrutura de conexão das unidades (densa ou esparsa) e a aplicação de técnicas de regularização. para evitar sobre ajuste. Finalmente, aprender significa escolher parâmetros ótimos com base na otimização numérica, o que muitas vezes requer a especificação de uma taxa de aprendizagem apropriada.

Scheuch, Voigt e Weiss (2023, p. 181) ainda citam que “cada neurônio aplica uma função de ativação não linear ao seu sinal agregado antes de enviar sua saída para a próxima camada” e isso pode ser observado através da equação (13) abaixo.

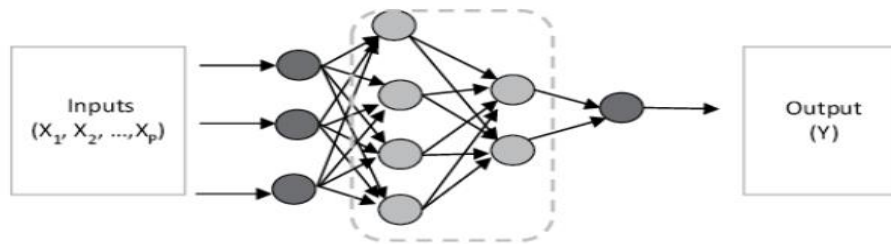
$$x_k^l = f \left(\theta_0^k + \sum_{j=1}^{N^l} z_j \theta_{l,j}^k \right) \quad (13),$$

em que θ são os parâmetros a serem ajustados, N^l é o número de unidades e z_j são as variáveis de entrada que podem ser os dados brutos ou, no caso de múltiplas camadas

encadeadas, o resultado de uma camada anterior $z_j = x_k - I$ (Scheuch; Voigt; Weiss, 2023, p. 181).

As redes neurais, em geral, têm a propriedade de conectar dados de saída (*output* – y), com dados de entrada (*input* – x_1, x_2, \dots, x_n). Na figura 1 (abaixo), um arranjo de redes neurais é esquematizado, em que cada círculo apontado na figura indica possíveis ligações entre os neurônios (Sicsú; Samartini; Barth, 2023, p. 288).

Figura 1 – Esquema de uma rede neural artificial



Fonte: Sicsú, Samartini e Barth (2023, p. 288).

Nota: Figura 8.1 extraída do livro Técnicas de Machine Learning.

O neurônio da rede neural artificial é a unidade básica de processamento, em que a rede é estabelecida quando “um neurônio recebe como *inputs* as saídas dos neurônios de camadas anteriores (s_1, s_2, \dots, s_k)”. Já “treinar uma rede neural significa determinar um conjunto de pesos que permite prever o *output* Y como menor erro possível” (Sicsú; Samartini; Barth, 2023, p. 291), em que a função de ligação que permite conectar as entradas com as saídas é tida por meio de funções. As funções mais comuns utilizadas nesses processos são: i) lineares; ii) logística (sigmóide); iii) tangente hiperbólica; iv) ReLU (Rectified linear unit); v) Softmax (logística generalizada); e vi) SWISH² (Sicsú; Samartini; Barth, 2023, p. 291-292; Haykin, 2007, p. 36; Galit, 2023). Haykin (2007, p. 37) ainda cita que se pode descrever um neurônio k a partir das equações (14) e (15) abaixo.

$$u_k = j = 1w_{kj}x_j \quad (14)$$

e

$$y_k = \varphi(u_k + b_k) \quad (15),$$

em que x_1, x_2, \dots, x_m são os valores de entrada; $w_{k1}, w_{k2}, \dots, w_{km}$ são os pesos sinápticos do neurônio k ; u_k é a saída do combinador linear; b_k são as bias; $\varphi(.)$ é a função de ativação; e

² Função desenvolvida pela empresa Google.

y_k é o valor de saída do neurônio, sendo que o uso das bias b_k tem o efeito de aplicar uma transformação a fim de uma saída u_k do combinador linear (Haykin, 2007, p. 37).

Dessa forma, ao observar os principais aspectos teóricos que envolvem os processos de redes neurais (RN) e a modelagem de regressão MQO, a fundamentação teórica desta pesquisa agora é conduzida para os parâmetros que irão procurar avaliar o desempenho desses processos citados anteriormente.

Indicador de desempenho MSE (Mean Squared Error – Erro Quadrado Médio)

Esse estudo procura comparar o desempenho da utilização das metodologias utilizadas em algoritmos de IA, que envolve redes neurais e a regressão MQO. O principal indicador utilizado nesta pesquisa para mensurar tal desempenho foi o MSE. Wooldridge (2023, p. 49) utiliza um sinônimo deste indicador, em que ele utiliza a terminologia de Erro Padrão da Regressão (EPR) ou simplesmente $\hat{\sigma}$. O mesmo autor ainda diz que esse indicador pode ser chamado de outros nomes, como por exemplo, erro padrão da estimativa. Esse tipo de mensuração fornece uma medida geral de quão “bom” o modelo se ajusta aos dados. Gujarati (2011, p. 817) e Neto (2013, p. 165) utilizam a terminologia MSE como estimador de Erro Quadrado Médio Mínimo, onde o MSE de um estimador $\hat{\theta}$ pode ser definido como:

$$SE(\hat{\theta}) = (\hat{\theta} - \theta)^2 \quad (16),$$

em que esse estimador contrasta com a variância de $\hat{\theta}$, que se define como:

$$ar(\hat{\theta}) = E[\hat{\theta} - E(\hat{\theta})]^2 \quad (17)$$

Simplificando por meio de Gujarati (2011, p. 824), tem-se que:

$$MSE(\hat{\theta}) = \text{variância de } \hat{\theta} \text{ mais o quadrado do viés.} \quad (18a)$$

Ainda segundo Gujarati (2011, p. 824), a diferença entre MSE e variância é que:

à $var(\hat{\theta})$ mensura a dispersão da distribuição de $\hat{\theta}$ em torno da sua média, ou valor esperado, enquanto o $MSE(\hat{\theta})$ mensura a dispersão em torno do valor verdadeiro do parâmetro. Naturalmente, se o viés é zero, $MSE(\hat{\theta}) = var(\hat{\theta})$. O critério MSE mínimo consiste em escolher um estimador cujo MSE seja o

menor em um conjunto de estimadores concorrentes. Observe que, mesmo se tal estimador for encontrado, há um trade-off envolvido — para obter uma variância mínima, podemos ter de aceitar algum viés. Na prática, o critério MSE mínimo é utilizado quando o critério do melhor não viesado é incapaz de produzir estimadores com variâncias menores.

Hilpisch (2020, p. 10) sintetiza o significado do termo MSE e cita que “para determinados valores de entrada x_1, x_2, \dots, x_N e valores de saída y_1, y_2, \dots, y_N , o MSE minimiza os valores reais de saída e os respectivos valores de saída aproximados”, que podem ser observados pela equação (18b) a seguir.

$$\frac{1}{N} \sum_{n=1}^N (y_n - \hat{f}(x_n))^2 \quad (18b)$$

Neto (2013, p. 165) cita que entre dois estimadores não enviesados, o mais eficiente é o que apresenta menor variância. No entanto, se compararem dois estimadores quaisquer, usa-se o erro quadrático médio, no qual este é a média da diferença entre o valor do estimador e do parâmetro ao quadrado. De acordo com Doane e Seward (2014, p. 440), o MSE é também utilizado na chamada estatística F, que é a razão da variância devido ao tratamento pela variância devido ao erro.

Outro indicador de desempenho dos modelos MQO e RN utilizados nessa pesquisa é o Erro Absoluto Médio (MAE), que de acordo com Ramasubramanian e Moolayil (2019, p. 154), é uma métrica que é independente da direção da variável. Ou seja, “o MAE fornecerá o erro absoluto médio (diferença entre os valores previstos e reais da variável dependente) na previsão, indiferente à direção do erro (positivo ou negativo)”. Isso pode ser observado na equação (18c) a seguir.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (18c),$$

em que y_i é o valor da i -ésima observação da variável dependente e \hat{y}_i é o valor previsto.

Um outro indicador que também pode ser utilizado para mensurar o desempenho dos modelos, segundo Ramasubramanian e Moolayil (2019, p. 155), é a Raiz do Erro Quadrático Médio (RMSE), que apesar de ser semelhante ao MAE, tem um diferencial, pois, “o RMSE calcula o valor ao quadrado, o que adiciona mais peso aos valores de erro mais altos antes de calcular a raiz quadrada”. A equação (18d) a seguir demonstra a fórmula da RMSE.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (18d)$$

em que $(y_i - \hat{y}_i)$ representa a diferença entre os valores reais e estimados da variável dependente para a i -ésima observação.

O tradicional coeficiente R^2 também é um recurso utilizado nesse estudo. O R^2 mede a porcentagem (valor entre 0 e 1) da variância na variável de resposta explicada pelo modelo linear (Gujarati, 2011; Ramasubramanian; Moolayil, 2019, p. 156; Wooldridge, 2023). Além desses instrumentos de mensuração, também são realizados os testes estatísticos formais de validação e significância, Teste T pareado, Teste de Normalidade dos Resíduos (Shapiro-Wilk), Teste de Homocedasticidade (Breusch-Pagan) e análises de robustez dos modelos utilizados, que serão tratados ao final da próxima seção de metodologia. Mas antes de detalhar a construção metodológica dos modelos utilizados nesta pesquisa, faremos a seguir uma breve crítica, assim como as abordaremos as devidas limitações deste estudo.

Críticas e limitações

Um dos grandes desafios acadêmicos em pesquisa aplicada é desenvolver estudos e modelos que tenham a capacidade preditiva dos dados utilizados, sendo que a academia trata desse assunto em diversos aspectos e dentre ramos distintos das ciências. Atualmente, a intensificação do uso de IA para tentar resolver esses problemas de previsão de dados se apoiou em diversos aspectos estruturais e tecnológicos, sendo que ainda existem limitações diversas, e entre estas tem-se as limitações de *hardware* como ferramenta computacional para a realização das pesquisas. Modelos RN de Aprendizado por Reforço é um exemplo do quanto o poder computacional é importante, limitando muito os modelos avançados de RN em função da capacidade computacional dos pesquisadores (Chollet, 2017; Galit, 2023; Goodfellow, 2016; Hilpisch, 2020; Scheuch; Voigt; Weiss, 2023).

Um outro ponto de possível crítica quanto ao modelo de Redes Neurais (RN) é a questão da quantidade de *rounds* (rodadas) que a modelagem utiliza e sua relação direta com o chamado problema de *overfitting*. Ao estudar algoritmos de IA, o *overfitting* significa que o modelo treinou em excesso a partir dos dados existentes para treinamento e isso pode levar a uma

espécie de “vício” ou viés quando o modelo é exposto a novos dados ou mesmo dados reais inseridos posteriormente à modelagem, além disso à uma piora de desempenho em situações de previsão — por exemplo, em alguns modelos IA de Aprendizado por Reforço, além dos dados reais inseridos para treinamento, são também adicionados dados sintéticos para evitar viés no modelo. Assim, a técnica de Validação Cruzada pode ser um meio para evitar o *overfitting*, sendo que essa metodologia utiliza múltiplas populações de dados de treinamento para validação e testes do modelo. Outras técnicas que também podem ser utilizadas para evitar o *overfitting* são técnicas de Abandono, Regularização e Ensacamento (*Dropout*, *Regularization* e *Bagging*) (Hilpisch, 2020; Scheugh; Voigt; Weiss, 2023).

Já os modelos MQO podem ser considerados também limitados por seu desempenho não tão eficiente em modelos mais complexos. Os esforços em trabalhar com aproximações lineares e não lineares de graus elevados muitas vezes não são suficientes para se obter uma acurácia precisa dos modelos preditivos, sendo essa falta de assertividade um obstáculo para pesquisas com esses modelos. Porém, apesar dos modelos não terem uma precisão máxima na maior parte dos casos, os modelos podem ajudar ao mínimo de se ter uma ideia da trajetória das variáveis observadas (Krishnamoorthy; Alok, 2022). Para isso, em seguida, no texto tem-se a seção de metodologia que procura demonstrar a construção de toda modelagem envolvida no referencial teórico, que serve de base para entender os resultados alcançados nesta pesquisa.

Metodologia

A metodologia desta pesquisa consiste em demonstrar as etapas de construção dos modelos, assim como demonstrar como testar a utilização dos instrumentos RN e MQO. A partir disso, o objetivo geral do trabalho consiste em verificar se esses dois instrumentos, aplicados em metodologias (*Machine Learning* e *Deep Learning*) de IA, têm o mesmo desempenho em situações diversas para uma série de dados distintos. Sintetizando, a metodologia tratada aqui procura resolver o problema de pesquisa por meio de aplicações desses métodos quantitativos para com as bases de dados didáticos aqui apontados.

Regressão por meio de Mínimos Quadrados Ordinários (MQO)

Uma situação típica encontrada em estudos que envolvem métodos quantitativos, e em especial, metodologias estatísticas e econométricas é apontar para uma possível diferenciação entre a matemática como ciência, e o aprendizado de estatística. Hipisch (2020, p. 101),

corroborar com esse tema ao citar que não existe uma resposta geral para esta questão, mesmo considerando a estatística um subcampo da matemática.

Dessa forma, uma maneira didática e apoiada em Hipisch (2020), para demonstrar a diferenciação entre as metodologias MQO e RN aplicadas em algoritmos de IA, é preciso aplicar essas técnicas em uma função matemática inicialmente, e observar os resultados alcançados pelos métodos envolvidos. Porém, de forma geral, quando se estuda estatística e econometria, não é comum partir a análise por meio de funções matemáticas para resolver os problemas envolvidos. O comum é ter uma base de dados e então rodar, por exemplo, uma regressão e através dessa regressão procurar estimar uma função ou equação matemática que explique a variável independente de interesse.

Assim, suponha a seguinte função matemática³ (19) abaixo:

$$f: R \rightarrow R, y = \alpha x^2 - \theta x^3 \quad (19)$$

Nota-se que a função (19) tem um monômio quadrático e outro cúbico, no qual essa função transforma um valor de entrada x em um valor de saída y ou transforma uma série de valores de entrada x_1, x_2, \dots, x_N em uma série de valores de saída y_1, y_2, \dots, y_N (Hipisch, 2020).

Novamente, nota-se a diferença do processo puramente matemático, onde tem-se uma função e quando adicionado valores de entrada x_i ao longo dessa função por meio de um critério ou restrição estabelecido, tem-se valores resultantes y_i de saída. Já o processo de aprendizagem estatística e econométrica, em geral, funciona ao contrário — no qual, tem-se os valores de x_i de entrada e dado estes como variáveis explicativas, procura-se estimar a variável independente y_i a ser explicada. Por exemplo, para saber estimar o quanto a variável explicativa preço (x_i) impactou na variável a ser explicada e estimada vendas (y_i) de veículos leves no Brasil nos últimos dez anos, pode-se rodar os dados da série temporal dos preços dos veículos ao longo dos anos, para ter como resposta ou saída dos valores para a variável vendas.

É claro que seria muito mais simples resolver questões aplicadas, por exemplo, em finanças, se já existisse uma função exata da trajetória da variável observada, ou ainda, o quanto seria fácil precificar o preço de um ativo financeiro se houvesse, com certeza, a matemática — a função da trajetória do preço desse ativo. Na verdade, de acordo com Hipisch (2020) e Kissinger, Schimidt e Huttenlocher (2021), esse é o objetivo maior da chamada Singularidade

³ Função baseada na expressão matemática dada por Hipisch (2020, p. 9 e p. 101).

Financeira, situação na qual uma IA financeira⁴ teria a capacidade de precificação exata do ativo observado.

Porém, por enquanto não se tem ainda essa tecnologia tão avançada. Dessa forma, ao trabalhar com a aprendizagem estatística, presume-se que os dados (amostra) ou valores de entrada e de saída sejam fornecidos e estimados, sendo que uma das metodologias mais utilizadas para estimar a função é por meio de uma regressão. Já um problema específico de uma regressão estatística é encontrar uma função que aproxima a relação funcional entre os valores de entrada e os valores de saída de maneira otimizada. Para facilitar o entendimento e explicar a metodologia MQO, pode-se em um primeiro momento fazer uso de uma regressão linear MQO simples. Nesse formato, tem-se os valores de entrada e saída assumidos como lineares, no qual o problema é encontrar os parâmetros ideais α e β para essa equação (20) abaixo (Gujarati, 2011; Hipisch, 2020, p. 10-11; Neto, 2013; Wooldridge, 2023).

$$f: R \rightarrow R, \hat{y} = \alpha - \beta_x \quad (20)$$

Para determinar os valores de entrada x_1, x_2, \dots, x_N e valores de saída y_1, y_2, \dots, y_N , tem-se que a solução ideal, neste caso, significa que eles minimizam o Erro Quadrático Médio (MSE) entre a saída real dos valores e os valores de saída aproximados (Hipisch, 2020, p. 11):

$$\frac{1}{N} (y_n - \hat{f}(x_n))^2 \quad (21)$$

Para o caso de regressão linear simples, a solução α^*, β^* é conhecida na forma fechada, como demonstrado nas equações 22 e 23 a seguir. As barras sob as variáveis y e x indicam os valores médios da amostra (Hipisch, 2020, p. 11).

$$\beta^* = \frac{Cov(x,y)}{Var(x)} \quad (22)$$

$$\alpha^* = \underline{y} - \beta_{\underline{x}} \quad (23)$$

⁴ Artificial financial Intelligence (AFI).

A equação (21) acima é uma função de monômios superiores, quadrática e cúbica fechada, e, por consequência, não é uma equação linear simples. Nesse caso, ao aplicar o método MQO a partir da regressão linear simples (20), o resultado do MSE pode não ser tão bom. No entanto, o MSE pode ser otimizado (diminuído) — talvez até para o valor 0, isto é, para uma “estimativa perfeita”. Assim, a regressão MQO não está restrita a uma relação linear simples. Além dos termos constantes e lineares, monômios de ordem superior podem ser facilmente adicionados como funções básicas. Dessa forma, ao supor que é conhecida antecipadamente a forma matemática da função (20) e, consequentemente, ao adicionar mais funções básicas à regressão, tem-se como resultado uma “perfeita aproximação.” Ou seja, “a regressão MQO recupera os fatores (dados) exatos da equação quadrática e parte cúbica, respectivamente, da função original” (Hipisch, 2020, p. 12-13).

Redes Neurais

Nos estudos econométricos, na grande maioria das vezes não se tem o conhecimento prévio da função matemática da base de dados que se está analisando, ou seja, o que se tem são os dados de entrada da série, e a partir dessa série aplica-se regressões para realizar as devidas inferências estatísticas. Nesse sentido, “as redes neurais podem ajudar a aproximar uma ampla gama de relacionamentos funcionais” (Hipisch, 2020, p. 13). Ainda de acordo com Hipisch (2020, p. 13):

Embora a abordagem de regressão MQO para a regressão linear simples, baseia-se no cálculo de certas quantidades e parâmetros bem especificados, a abordagem da rede neural depende de aprendizado. Isto, por sua vez, significa que um conjunto de parâmetros (pesos dentro da rede neural), são primeiro inicializados aleatoriamente e depois ajustados gradualmente, dada as diferenças entre a saída da rede neural e os valores de saída da amostra. Essa abordagem permite treinar novamente (atualizar) uma rede neural de forma incremental.

Um possível bom desempenho do modelo MQO observado na equação (1) pode produzir um valor zero de MSE — ou seja, uma regressão perfeita. Já as ferramentas RN também podem ter uma boa acurácia na estimativa, como a regressão MQO, que oferece um resultado perfeito — para o caso da função matemática previamente estabelecida. Veremos mais à frente no texto os testes empíricos dessas situações na seção de resultados dessa pesquisa.

Dessa forma, pode-se pensar o porquê estudar e aplicar a metodologia de redes neurais em algoritmos de IA, se o método MQO, aparentemente, já contempla a solução do problema?

Para responder à esta questão, pode-se verificar o que pode acontecer com os desempenhos das estimativas dos métodos MQO e de rede neurais quando a função matemática não for “tão comportada e suave”, como no caso da equação (20). Para isso Hipisch (2020, p. 16), estipula e considera “um conjunto de dados de amostra aleatória, no qual tanto os recursos quanto os rótulos são escolhidos aleatoriamente”.

Nesse formato de gerar um conjunto de dados de amostra aleatória e ao criar uma regressão MQO — nos mesmos moldes da equação (20), incluindo um número variável de funções de base monominal —, tem-se que a estimativa alcançada pode não ser tão eficiente. Ou seja, ao rodar a regressão com dados de características aleatórias, o valor MSE pode ficar mais alto do que se comparado ao desempenho medido pelo MSE para o caso da equação (20). Isso se dá em função de que a aproximação de uma regressão MQO pode ser alcançada através de uma combinação apropriada de um número finito de funções básicas. Como o novo conjunto de dados de amostra foi gerado aleatoriamente, a regressão MQO pode não ter um bom desempenho nesse caso (Hipisch, 2020).

No entanto, por meio da metodologia redes neurais, os resultados de desempenho MSE podem ser satisfatórios para o caso de variáveis com características aleatórias. Segundo Hipisch (2020, p. 18), apesar do resultado não ser perfeito, os métodos de redes neurais, em geral, podem ter um melhor desempenho do que se comparado com a regressão MQO. Isso se deve ao fato de que “dada a sua arquitetura, a rede neural possui quase 200.000 parâmetros treináveis⁵ (pesos) que oferecem flexibilidade relativamente alta, especialmente quando comparado com a regressão MQO, para os quais são utilizados no máximo 15 + 1 parâmetros” (Hipisch, 2020, p. 18).

Dessa forma, as metodologias para os algoritmos de IA utilizadas na pesquisa e apresentadas nas seções anteriores podem ser mais bem observadas na seção seguinte de resultados, que demonstra os resultados apontados pelos testes quantitativos realizados. Mas antes de prosseguir para a próxima seção, são apresentados nesse momento os métodos formais de análise estatística utilizados nesta pesquisa.

⁵ O recurso computacional utilizado neste caso é o pacote Scikit-learn utilizado na base programável Python. Para um estudo mais a contexto desse processo de regressão pode-se consultar Goodfellow et al. (2016) e também por meio do endereço eletrônico a seguir. Disponível em: https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPRegressor.html. Acesso em: 17 jul. 2024.

Análise Estatística Formal (Comparação de Modelos)

O objetivo de realizar uma análise estatística formal é ir além da comparação visual ou baseada nas métricas de erro (MSE, MAE, RMSE, R^2) e realizar assim uma análise estatística para determinar se as diferenças observadas no desempenho dos modelos são estatisticamente significativas. Nesse sentido, a análise estatística busca verificar se a diferença nos erros entre os dois modelos (MQO e RN) é grande o suficiente para não ser considerada apenas uma flutuação aleatória nos dados utilizados. Para realizar tal verificação metodológica, a análise utiliza o Teste t pareado, respectivamente as interpretações do p -value Teste de Normalidade dos Resíduos (Shapiro-Wilk), Teste de Homocedasticidade (Breusch-Pagan), entre outros (Langenberg *et al.*, 2023; Kim, 2015; Talikan; Ajan, 2025).

O teste t pareado pode ser utilizado ao verificar duas amostras relacionadas ou "pareadas". Sendo que para essa pesquisa, as amostras são os resíduos (erros) dos dois modelos (MQO e RN) ao considerar os mesmos pontos de dados (x , y) do espaço amostral observado. Esse teste compara a média das diferenças entre esses pares de resíduos, ou seja, o objetivo do teste t pareado é verificar se a média das diferenças entre os resíduos dos modelos MQO e RN é significativamente diferente de zero. Assim, se a média das diferenças for diferente de zero de forma significativa, sugere-se que um modelo tem erros consistentes e maiores ou menores que o outro. Já o valor encontrado pela Estatística t representa a magnitude da diferença entre as médias das amostras em relação à sua variabilidade. Exemplificando para os exercícios estatísticos desta pesquisa: uma estatística t de valor 0.0000 sugere que não há diferença observada na média dos resíduos entre, por exemplo, um modelo Linear e um Polinomial de Grau 3 para o mesmo conjunto de dados (Langenberg *et al.*, 2023; Kim, 2015; Talikan; Ajan, 2025).

Resultados

Essa é uma pesquisa que pode ser considerada aplicada em métodos quantitativos do tipo quali-quantitativo. Assim, ao inferir os resultados observados nesta pesquisa, pode-se concluir que dependendo do grau de complexidade dos dados utilizados para realizar a estimação, os métodos MQO e RN podem ser melhor aproveitados em suas devidas situações. Para isso, são observados os testes quantitativos (1 a 4) a seguir.

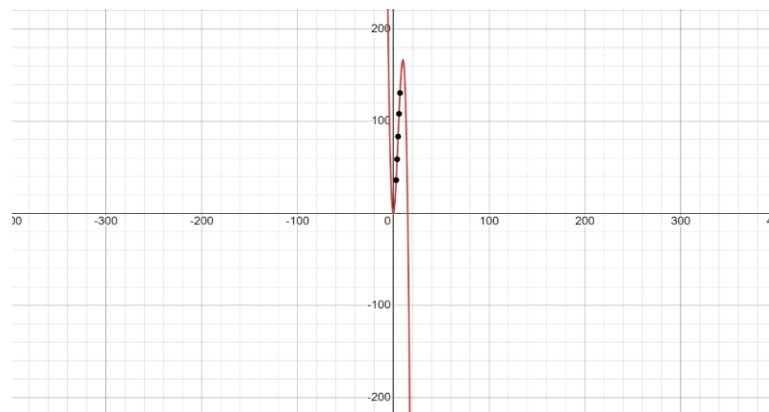
Teste Quantitativo 1 (Parte A)

Rodar as regressões pelo método MQO (parte A) e RN (parte B) para a função (24)⁶, e gerar as suas respectivas estimativas de desempenho mensurado por MSE, RMSE, MAE e R^2 , além também de realizar os testes de formais de estatística e significância.

$$f: R \rightarrow R, y = 5x^2 - \left(\frac{1}{3}\right)x^3 \quad (24)$$

A figura 2 abaixo, demonstra geometricamente a função (24).

Figura 2 – Demonstração geométrica da função matemática (24)



Fonte: Elaboração própria a partir de ferramentas geométrica do site Mathway⁷.

Assim, dado a função matemática (24) dentre um espaço amostral $-2 > x < 4$, aplica-se a regressão linear para gerar a estimativa dessa regressão, no qual o objetivo é observar o resultado de desempenho por meio dos indicadores MSE, RMSE, MAE e R^2 pelo método MQO. Posteriormente — na parte B deste teste quantitativo 1 —, aplica-se a metodologia de redes neurais na mesma função matemática para observar os mesmos indicadores de desempenho e, assim, realizar as devidas comparações de resultados.

A função matemática (24) transforma uma série de valores de entrada x_1, x_2, \dots, x_N em uma série de valores de saída y_1, y_2, \dots, y_N . Para isso, foi utilizado o software *Python*, em que o código programado implementou a função matemática e isso permitiu criar uma série de valores de entrada e saída. A figura 3 abaixo representa graficamente os valores de saída em relação aos valores de entrada para a função (24). O código *Python* calcula os valores ideais dos

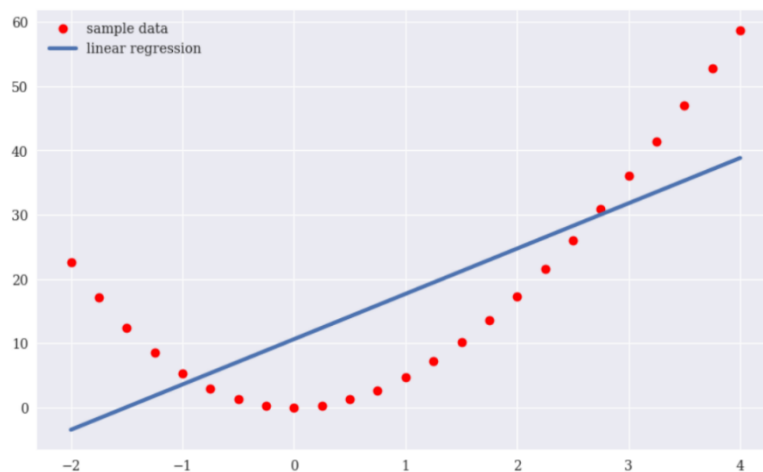
⁶ Exemplos quantitativos adaptados e gerados a partir de Hipisch (2020).

⁷ Disponível em: <https://www.mathway.com/pt/Algebra>. Acesso em: 17 jul. 2024.

parâmetros, estima linearmente (aproxima) os valores de saída e traça a linha de regressão linear ao lado dos dados da amostra (figura 3 abaixo). Vale observar que a abordagem de regressão linear simples não funciona muito bem aqui. Isso pode ser observado e é confirmado pelo relativamente alto valor do resultado MSE (136,8226) e também pelos resultados não muito satisfatórios dos indicadores RMSE, MAE e R^2 que resultam respectivamente 11.69883, 10.06950, 0.54163. Já os valores estimados de alfa e beta para essa regressão linear podem ser observados na equação (25) a seguir.

$$\hat{f}: R \rightarrow R, \hat{y} = 10,6125 - 7,0542x \quad (25)$$

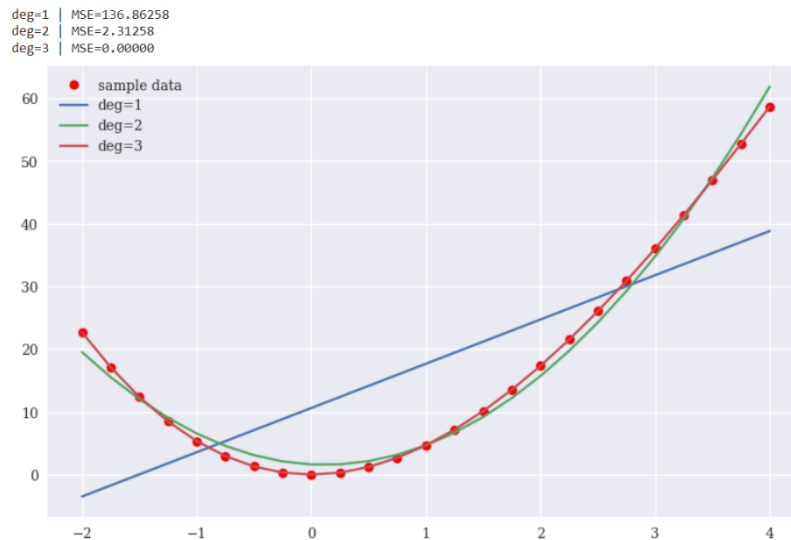
Figura 3 – Demonstração geométrica de valores plotados a partir da função (25) via software Python



Fonte: Elaboração própria e adaptado a partir de programação baseada em Hipisch (2020, p. 6).

Como os resultados de desempenho dos indicadores não foram muito satisfatórios para a regressão linear simples, estes podem ser melhorados. Além dos termos constantes e lineares, também monômios de ordem superior podem ser adicionados como funções básicas para melhorar os valores estimados. As melhorias podem ser visualizadas (figura 4 abaixo) e observadas ao realizar os ajustes dos parâmetros e monômios de ordem superior — quadrático e cúbico —, confirmadas pelos valores MSE estimados.

Figura 4 – Demonstração geométrica de valores plotados via ajuste de monômios para a regressão linear



Fonte: Elaboração própria e adaptado a partir de programação baseada em Hipisch (2020).

Nesse caso, como a função matemática utilizada (24) é “bem-comportada”, tem-se que a estimativa do valor MSE foi melhorada (diminuída), podendo ser considerada uma “estimativa perfeita”. Ou seja, ao explorar o conhecimento prévio da forma da função matemática a ser aproximada e, conseqüentemente, ao adicionar mais funções básicas — quadrática e cúbica — à regressão leva a uma “perfeita aproximação.” Assim, a regressão MQO recupera os fatores exatos da equação quadrática e cúbica da função original (24) (Hipisch, 2020, p. 12-13). Além do MSE, é notório observar a melhora também dos outros indicadores de performance. Para $\text{deg} = 1$, obteve-se: MSE: 136.86258, RMSE: 11.69883, MAE: 10.06950, e R^2 : 0.54163. Para $\text{deg} = 2$: MSE: 2.31258, RMSE: 1.52072, MAE: 1.31450, e R^2 : 0.99225. Já para $\text{deg} = 3$: MSE: 0.00000, RMSE: 0.00000, MAE: 0.00000, e R^2 : 1.00000.

Por meio de todas as métricas acima, verifica-se que o modelo MQO consegue aproximar de forma perfeita os valores da equação (24) dada neste experimento. Ou seja, o modelo MQO é “perfeito” para encontrar a trajetória de uma equação de dados bem estruturados.

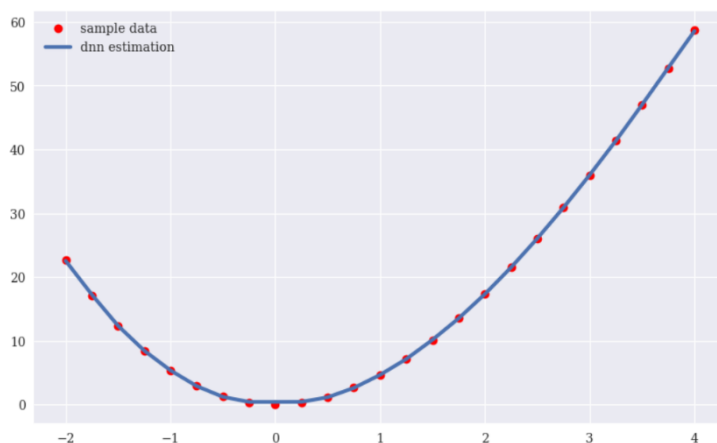
Teste Quantitativo 1 (Parte B)

Rodar as regressões pelo método RN para a mesma função (24) e gerar as suas respectivas estimativas dos indicadores de desempenho.

Uma observação válida é que esta pesquisa está sendo realizada a partir da ótica didática da ciência econômica e em específico dentre a subárea de finanças, no qual as metodologias estudadas neste trabalho têm a intenção de serem trabalhadas posteriormente dentre estas áreas de conhecimento. Dessa forma, a maior parte dos estudos dessa área envolve não se ter funções tão comportadas para serem estimadas e mesmo nem ter tais equações. São por essas razões que as metodologias que envolvem redes neurais podem ajudar a encontrar resultados positivos dentre as observações realizadas por essa ciência (Hipisch, 2020, p. 13).

Ao selecionar novamente a função matemática (24) com o objetivo de estimar o desempenho e resultado MSE, aplica-se agora a metodologia de redes neurais por meio do software Python e por meio do pacote *Scikit-learn*. Esse código utiliza um processo regressor chamado de classe *MLPRegressor* do *Scikit-learn*, que implementa uma DNN para a estimativa (Goodfellow *et al.*, 2016; Hipisch, 2020). Os resultados podem ser observados por meio da figura 5 abaixo.

Figura 5 – Demonstração geométrica de valores plotados via metodologia de redes neurais para a função (24)



Fonte: Elaboração própria e adaptado a partir de programação baseada em Hipisch (2020).

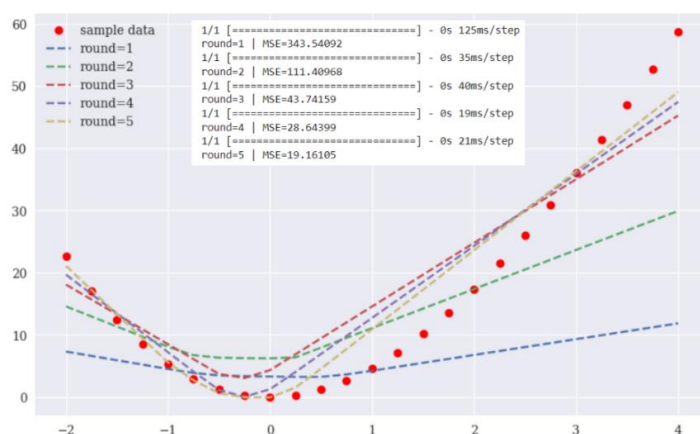
Ao observar a figura 5 acima, pode-se verificar os resultados obtidos via as métricas de desempenho: MSE: 0.25683, RMSE: 0.50679, MAE: 0.42759, R^2 : 0.13383). Fica evidente nesse exemplo que os resultados dos indicadores não foram perfeitos, se comparados com o método MQO da Parte A desse Teste Quantitativo 1. No entanto, se observarmos visualmente

o gráfico da regressão gerada sobre a base de dados pode-se notar que a solução via redes neurais foi satisfatória. Ou seja, a Análise Comparativa indica que o modelo com o menor MSE (0.00000) obteve o melhor ajuste geral — que no caso foi o modelo de Grau 3 —, para o método MQO. Já o modelo com o maior R^2 (1.0), explica a maior proporção da variância, que também foi o modelo de Grau 3.

Dessa forma, para o teste quantitativo 1, tem-se que os resultados observados por meio do processo metodológico MQO (figura 4) apresentaram um desempenho praticamente perfeito ($MSE = 0$), se comparado com o desempenho estimado por meio da metodologia de redes neurais, que apresentou um desempenho menor ($MSE = 0,25683$).

Antes de ser apresentado o teste econométrico 2, iremos rodar a mesma regressão anterior (24) por meio de uma outra metodologia de redes neurais — só que agora o pacote de aprendizagem profundo utilizado foi o *Keras*⁸ (Python). Neste, o modelo é ajustado e treinado em até 100 vezes (100 *epochs*), sendo que esse procedimento é repetido por cinco rodadas (*rounds*) e após cada rodada, a aproximação pela rede neural é atualizada e traçada (Hipisch, 2020, p. 15). A figura 6 abaixo demonstra como a aproximação melhora gradualmente a cada rodada. Este desempenho pode ser observado nos valores decrescentes do indicador MSE, em que o resultado final também não é perfeito, mas pode ser considerado satisfatório.

Figura 6 – Demonstração geométrica de valores plotados via metodologia de redes neurais para a função (24) e por meio do pacote Keras



Fonte: Elaboração própria e adaptado a partir de programação baseada em Hipisch (2020).

Em síntese, para o teste econométrico 1, tem-se que os algoritmos RN se saem quase tão bem na estimativa como a regressão MQO, que para esse caso, oferece um resultado

⁸ Para mais detalhes deste pacote de programação Python, veja Chollet (2017).

perfeito. Vale reforçar a observação de que o exercício 1 foi produzido e gerado a partir de um conjunto de dados didáticos bem definidos e por meio de uma função matemática já previamente estabelecida, fechada e “bem-comportada” (quadrática e cúbica). No entanto, no exercício quantitativo 2 — mais à frente no texto — iremos gerar um conjunto de dados a partir de uma amostra aleatória — e não a partir de uma função matemática “fechada”. Nesse caso, o objetivo é verificar se o desempenho das metodologias MQO e de redes neurais será o mesmo que a do exercício anterior.

Análise estatística formal de significância por meio dos resultados dos Testes T Pareados entre Previsões dos Modelos do Exercício Quantitativo 1

Com o objetivo de aprofundar os estudos em termos de vigor de testes estatísticos formais, é introduzido neste momento uma análise de estatística formal, na qual foram realizados os cálculos dos resíduos (erros) dos modelos utilizados nesse exercício econométrico. O Teste *t* pareado foi utilizado para comparar estatisticamente os resíduos de diferentes pares de modelos; já a Análise de Robustez será tratada mais à frente neste texto ao adicionar dados *outliers* em novos exercícios e cenários estatísticos (3 e 4) (Langenberg *et al.* 2023; Kim, 2015; Talikan; Ajan, 2025).

O teste *t* pareado compara estatisticamente as previsões de cada par de modelos com o objetivo de determinar se as diferenças observadas são estatisticamente significativas. Um *p-value* baixo — geralmente menor que 0.05 — sugere que há uma diferença estatisticamente significativa entre as previsões dos modelos comparados. A tabela 1 a seguir resume os resultados do exercício econométrico 1 (Parte A e B), em que o teste *t* pareado demonstrou que não há evidência estatisticamente significativa de diferença entre as previsões desses dois modelos para este exercício quantitativo 1 em específico.

Tabela 1 – Resultados dos Testes T Pareados entre Previsões dos Modelos (Comparação da Estatística T, e do *p-value*)

| | T-Statistic | P-Value |
|--|-------------|----------|
| Polinomial_deg_1 vs Polinomial_deg_2 | -7,32E-15 | 1,000000 |
| Polinomial_deg_1 vs Polinomial_deg_3 | -1,55E-15 | 1,000000 |
| Polinomial_deg_1 vs MLPRegressor | -9,99E-03 | 0,992108 |
| Polinomial_deg_1 vs Keras_RN | -6,54E-03 | 0,519609 |
| Polinomial_deg_2 vs Polinomial_deg_3 | 4,49E-14 | 1,000000 |
| Polinomial_deg_2 vs MLPRegressor | -7,736643 | 0,938975 |
| Polinomial_deg_2 vs Keras_RN | -1,18E+00 | 0,249445 |
| Polinomial_deg_3 vs MLPRegressor | -1,11E+00 | 0,276355 |
| Polinomial_deg_3 vs Keras_RN | -1,45E+00 | 0,159621 |
| MLPRegressor vs Keras_RN | -1,42E+00 | 0,167955 |
| Teste de Normalidade dos Resíduos (Shapiro-Wilk) | | |
| Estatística do Teste: | | 0,958891 |
| P-Value: | | 0,08003 |
| Conclusão: Não há evidência suficiente para rejeitar a hipótese nula. Os resíduos podem seguir uma distribuição normal | | |
| Teste de Homocedasticidade (Breusch-Pagan) | | |
| Estatística do Multiplicador de Lagrange: | | 0,07103 |
| P-Value (do Multiplicador de Lagrange): | | 0,78984 |
| Estatística F: | | 0,06829 |
| P-Value (do Teste F): | | 0,794496 |
| Conclusão: Não há evidência suficiente para rejeitar a hipótese nula. Assume-se HOMOCEASTICIDADE. | | |

Fonte: Elaboração própria.

Teste Quantitativo 1: Cenário expandido para observar Análise de Robustez

Com o objetivo de aprofundar a pesquisa em rigor de testes formais estatísticos por meio de realizar uma análise mais robusta e a partir da equação (24) do Teste quantitativo 1, foi aumentado o número de pontos do espaço amostral em x e y (espaço amostral original: $-2 > x > 4$). Foi adicionado o Ruído Gaussiano nessa função $f(x)$ e aumentado graus na regressão Polinomial para explorar o trade-off de variância.

Dessa forma, para a Análise de Robustez foi criado um conjunto de dados com *outliers* para testar a sensibilidade dos modelos por meio de dados anômalos. Foi treinado e avaliado os modelos (MQO e RN) com esses novos dados. Esse cenário expandido permite uma análise mais profunda do problema de regressão, comparando diferentes abordagens de modelagem e não apenas pelas métricas básicas de erro, mas também considerando a significância estatística das diferenças e a capacidade dos modelos de lidar com dados não estruturados.

Dessa forma, ao adicionar os dados de ruído (2 pontos) na equação $f: \mathbb{R} \rightarrow \mathbb{R}, y = 5x^2 - (1/3)x^3$ (24) e ao aproximar esses dados por meio de regressão MQO linear, teve-se o seguinte

desempenho mensurado: R^2 : 0.5359, MSE: 164.4848, MAE: 10.7686 e RMSE: 12.8252. Assim, ao comparar o MSE (164,48) da equação adicionado de ruído com o desempenho mensurado pelo MSE (136,86) na equação original, verifica-se a esperada piora do modelo. Ou seja, dado que a equação “bem estruturada” (24) foi adicionada de dados não estruturados — a trajetória deixou de ser “bem-comportada” para ser uma trajetória com valores foras de sua curva matemática natural. O mesmo experimento aproximado por meio de uma regressão Polinomial de Grau = 1 obteve os seguintes desempenhos de aproximação: MSE: 164.4848, MAE: 10.7686, RMSE: 12.8252, e R^2 : 0.5359. Para as aproximações de graus superiores, os resultados foram: Grau = 2: MSE: 21.2080, MAE: 3.6491, RMSE: 4.6052, e R^2 : 0.9402. Grau = 3: MSE: 19.4901, MAE: 3.6467, RMSE: 4.4148, e R^2 : 0.9450. Grau = 4: MSE: 19.4708, MAE: 3.6408, RMSE: 4.4126, e R^2 : 0.9451. Grau = 5: MSE: 19.4643, MAE: 3.6382, RMSE: 4.4118, e R^2 : 0.9451.

Para o mesmo experimento anterior com adição de ruídos na equação (24) — mas agora tendo aproximação via redes neurais o desempenho obtido pelas métricas —, foram: MSE: 16.3045, MAE: 3.2527, RMSE: 4.0379, e R^2 : 0.9540. Lembrando que o desempenho sem a adição de dados de ruído para este mesmo experimento tinham sido: MSE: 0.25683, RMSE: 0.50679, MAE: 0.42759, R^2 : 0.13383. Assim, é possível observar novas inferências estatísticas por meio desse novo cenário com dados de ruídos inseridos no modelo. Das quatro métricas observadas, o MSE e o R^2 pioraram; já o RMSE e o MAE obtiveram melhoras. Ou seja, como os dados de ruídos tem características de serem dados mais explosivos, permitiu-se que as métricas MSE e R^2 detectassem rapidamente esse problema. Se aplicarmos tal inferência em dados financeiros, por exemplo, nota-se que a características desses dados também carregam ruídos, o que pode indicar a métrica MSE como a mais indicada para esses tipos de dados.

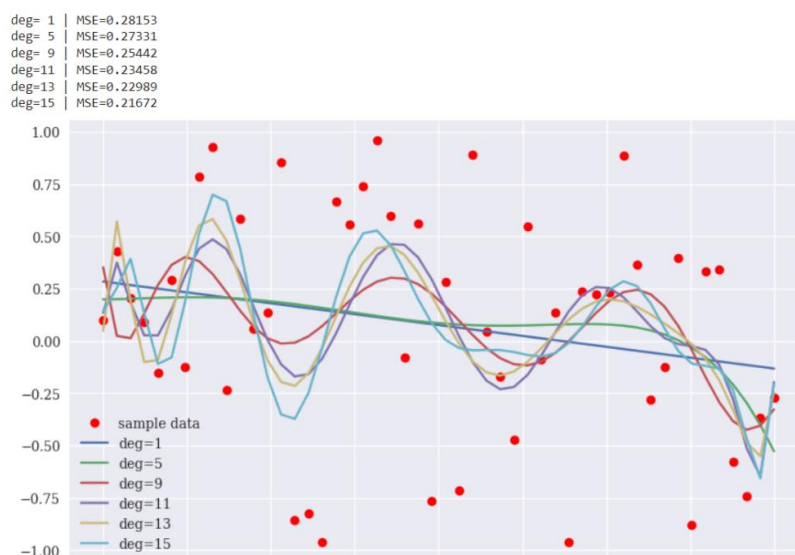
Teste Quantitativo 2 (Parte A)

Rodar as regressões pelo método MQO (parte A) e por RN (parte B) adicionado de uma semente aleatória, e gerar as suas respectivas estimativas de desempenho MSE.

A programação Python utilizada neste exercício gerou um conjunto de dados de amostra aleatória e criou a regressão MQO estimada, baseada em um número variável de funções de base monominal. Os resultados da regressão MQO deste teste quantitativo 2 (Parte A) não são tão surpreendentes, como pode ser observado na figura 7 abaixo. A regressão MQO, neste caso, assume que a aproximação pode ser alcançada através de uma combinação apropriada de um número finito de funções básicas. Como o conjunto de dados de amostra foi gerado por uma

semente aleatória, a regressão MQO não tem um bom desempenho neste caso, se comparado com o exercício 1. O novo valor MSE encontrado em $\text{deg}=15$, foi de 0,21672⁹ (Hipisch, 2020, p. 15).

Figura 7 – Demonstração geométrica de valores plotados e gerados de forma aleatória a partir do software Python



Fonte: Elaboração própria e adaptado a partir de programação baseada em Hipisch (2020).

Teste Quantitativo 2 (Parte B)

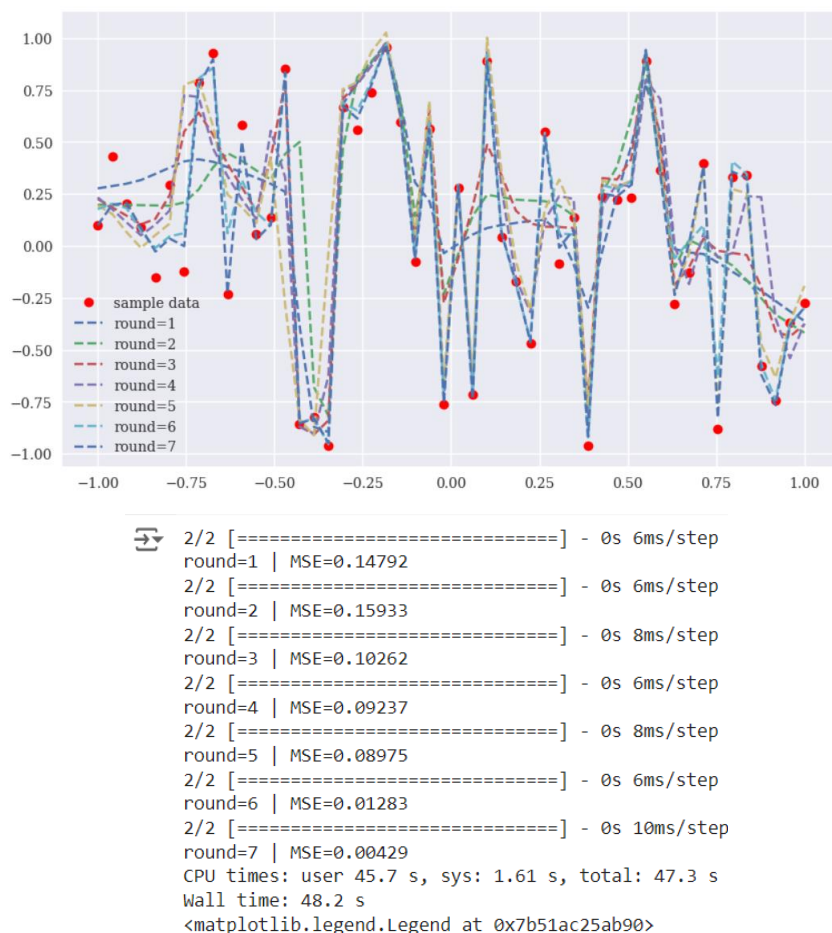
Rodar as regressões pelo método RN para a função aleatória gerada pelo software *Python* e gerar as suas respectivas estimativas de desempenho MSE.

Ao rodar o teste quantitativo 2 (parte B) com a utilização da metodologia de redes neurais para uma base de dados aleatórios, observa-se por meio da figura 8 abaixo que, embora o resultado não seja perfeito, verifica-se visualmente que nesse caso, a rede neural tem um

⁹ deg=1: MSE: 0.28153, RMSE: 0.53059, MAE: 0.44174, R^2 : 0.05054. deg=2: MSE: 0.27825, RMSE: 0.52749, MAE: 0.43112, R^2 : 0.06160. deg=3: MSE: 0.27600, RMSE: 0.52536, MAE: 0.42902, R^2 : 0.06918. deg=4: MSE: 0.27363, RMSE: 0.52310, MAE: 0.42664, R^2 : 0.07718. deg=5: MSE: 0.27331, RMSE: 0.52279, MAE: 0.42726, R^2 : 0.07826. deg=6: MSE: 0.27329, RMSE: 0.52277, MAE: 0.42756, R^2 : 0.07833. deg=7: MSE: 0.26854, RMSE: 0.51821, MAE: 0.41751, R^2 : 0.09433. deg=8: MSE: 0.25975, RMSE: 0.50965, MAE: 0.41222, R^2 : 0.12401. deg=9: MSE: 0.25442, RMSE: 0.50440, MAE: 0.40968, R^2 : 0.14196. deg=10: MSE: 0.25359, RMSE: 0.50358, MAE: 0.40733, R^2 : 0.14476. deg=11: MSE: 0.23458, RMSE: 0.48433, MAE: 0.37466, R^2 : 0.20888. deg=12: MSE: 0.23040, RMSE: 0.48000, MAE: 0.37040, R^2 : 0.22297. deg=13: MSE: 0.22989, RMSE: 0.47947, MAE: 0.37284, R^2 : 0.22470. deg=14: MSE: 0.21982, RMSE: 0.46885, MAE: 0.37593, R^2 : 0.25865. deg=15: MSE: 0.21672, RMSE: 0.46553, MAE: 0.36636, R^2 : 0.26911.

melhor desempenho do que se comparado a metodologia de regressão MQO para a estimativa dos valores de rótulo e de recursos aleatórios.

Figura 8 – Demonstração geométrica de valores plotados via metodologia de redes neurais um conjunto de dados aleatórios



Fonte: Elaboração própria e adaptado a partir de programação baseada em Hipisch (2020).

Diferentemente do exercício econométrico 1, ao observar os dados de performance do resultado das métricas de desempenho da parte B do exercício econométrico 2 (MSE: 0.00429, RMSE: 0.22561, MAE: 0.10056, R^2 : 0.82834) e ao comparar com os da parte A do mesmo exercício 2 (deg=15: MSE: 0.21672, RMSE: 0.46553, MAE: 0.36636, R^2 : 0.26911), nota-se o melhor desempenho, nesse caso, para o algoritmo que utiliza a abordagem de redes neurais. Ou seja, quando da utilização de dados gerados a partir de uma função aleatória, tem-se que a performance da metodologia de redes neurais apresenta um melhor desempenho, se comparado com a metodologia de regressão MQO.

Ao realizar o teste T pareado — comparando agora no exercício quantitativo 2 —, quanto aos métodos MQO e de RN para dados de características aleatórias, tem-se que os

resultados apresentaram significância estatística ($p\text{-value} < 5\%$) entre esses dois modelos desde o Grau 1 ao 15 do modelo linear e polinomial MQO em comparação ao modelo RN (Tabela 2).

Tabela 2 – Resultados dos Testes T Pareados entre Previsões dos Modelos (Comparação da Estatística T, e do $p\text{-value}$)

| | T-Statistic | P-Value |
|---|-------------|----------|
| Polinomial_deg_1 vs Keras_RN | 4,31E+00 | 0,000780 |
| Polinomial_deg_2 vs Keras_RN | 4,10E+00 | 0,000206 |
| Polinomial_deg_3 vs Keras_RN | 3,81E+00 | 0,000386 |
| . | . | . |
| . | . | . |
| . | . | . |
| Polinomial_deg_13 vs Keras_RN | 2,21E+00 | 0,032163 |
| Polinomial_deg_14 vs Keras_RN | 2,06E+00 | 0,044677 |
| Polinomial_deg_15 vs Keras_RN | 2,02E+00 | 0,048714 |
| Teste de Normalidade dos Resíduos (Shapiro-Wilk) | | |
| Estatística do Teste: | | 0,95891 |
| P-Value: | | 0,08003 |
| Conclusão: Não há evidência suficiente para rejeitar a hipótese nula. Os resíduos podem seguir uma distribuição normal | | |
| Teste de Homocedasticidade (Breusch-Pagan) | | |
| Estatística do Multiplicador de Lagrange: | | 0,07103 |
| P-Value (do Multiplicador de Lagrange): | | 0,78984 |
| Estatística F: | | 0,06829 |
| P-Value (do Teste F): | | 0,794496 |
| Conclusão: Não há evidência suficiente para rejeitar a hipótese nula. Assume-se HOMOCEASTICIDADE. | | |

Fonte: Elaboração própria.

Por meio da parte inferior da figura 8, pode-se conferir que foram utilizados 7 *rounds* nesse experimento. Também é possível observar que o valor MSE melhora (diminui) com a adição de subsequentes *rounds*. O que poderíamos erroneamente sugerir que quanto maior for o número de *rounds*, melhor serão os resultados da modelagem. No entanto, infelizmente existe o problema de *overfitting*. Assim, para demonstrar a questão do problema de *overfitting* foi utilizado o mesmo experimento anterior (teste quantitativo 2), no qual agora foi adicionado um total de 9 *rounds*, obtendo como resultado um valor de MSE de 0,00014, ou seja, houve melhora de desempenho com o aumento do número de *rounds*. Vale observar que o fato de aumentar o número de *rounds*, por exemplo, até uma quantidade muito grande, não é uma mera questão de melhoria contínua no modelo. Tem-se além do problema de *overfitting* a capacidade computacional de realizar tal processo computacional, ou seja, o tempo de processamento vai aumentando de forma relevante. Por exemplo, continuando com o mesmo experimento anterior e ao aumentar o número de *rounds* para 11, teve-se como resultado de desempenho MSE um valor de 0,00040; já o tempo de processamento foi de 4min43seg — para 7 *rounds*, o tempo foi de 43,3 segundos. Ou seja, o modelo piorou devido ao problema de *overfitting*.

Uma observação importante quanto ao uso de redes neurais é trabalhar e adaptar os parâmetros às características dos dados utilizados. Ou seja, não existe uma parametrização ideal para todo tipo de modelagem; o ideal é testar qual parametrização melhor se adapta aos dados estudados. Outro ponto importante quanto à adição em excesso de quantidade de *rounds*, pode ser melhor verificada quando os dados de análise apresentam valores muito fora do normal (*outliers*), por exemplo, como veremos em outro exercício e cenário a seguir.

Teste Quantitativo 3. Análise de Robustez. Nova função/cenário: $f: \mathbb{R} \rightarrow \mathbb{R}, y = 2x^2 + 1$ (Com e sem a adição de dados outliers)

Em uma situação de um novo cenário e exercício, foi considerado uma nova função matemática (26) abaixo. Essa equação também é composta por dados em seu espaço amostral ($-2 > x < 4$), sem a adição de dados *outliers* em um primeiro momento. Em seguida, também foi considerado neste exercício a adição de dados do tipo *outliers* com o objetivo de aprofundar nas inferências estatísticas, nos testes estatísticos formais de significância e na robustez da modelagem, em que foram adicionados dois dados *outliers* $(-3, 5)$ no espaço amostral da mesma função (26).

$$f: \mathbb{R} \rightarrow \mathbb{R}, y = 2x^2 + 1 \quad (26)$$

Ao observar os resultados do teste quantitativo 3 (tabela 3 abaixo), verifica-se de forma clara que o método MQO de grau 1 (linear) produziu uma aproximação não muito satisfatória ($MSE = 33,63750$) dos dados reais comparados com os dados estimados. No entanto, o mesmo método MDO de grau 2 produziu uma aproximação perfeita entre os dados reais e os dados estimados, sendo que esse resultado já poderia ser esperado dado a característica construtiva quadrático e linear da equação (26). Já a modelagem RN produziu uma aproximação inferior do que o modelo MQO par essa função matemática $MSE = 3,38950$ para o modelo RN Scikit-learn, e 4,54000 para o modelo RN Keras. No entanto, quando a mesma função (26) é exposta a dados *outliers*, fica evidente a vantagem que o método RN ($MSE = 6,93630$) leva sobre o método MQO (226,69500).

Tabela 3 – Resumo dos Resultados do Teste Quantitativo 3

| Métricas de Avaliação dos Modelos | MSE | RMSE | MAE | R ² |
|--|----------------|----------|---------------|----------------|
| Modelo MQO deg=1 | 33,63750 | 5,79978 | 5,00000 | 0,60721 |
| Modelo MQO deg=2 | 0,00000 | 0,00000 | 0,00000 | 1,00000 |
| Modelo MQO deg=3 | 0,00000 | 0,00000 | 0,00000 | 1,00000 |
| Modelo Polimomial deg=1 | 44,26010 | 6,65280 | 5,30810 | 0,54020 |
| Modelo Polimomial deg=2 | 8,47500 | 2,91120 | 5,50510 | 0,91200 |
| Modelo Polimomial deg=3 | 8,18950 | 2,86170 | 2,38480 | 0,91490 |
| Modelo Polimomial deg=4 | 7,48880 | 2,73660 | 2,18450 | 0,92220 |
| Modelo Polimomial deg=5 | 7,33940 | 2,70910 | 2,07520 | 0,92380 |
| Modelo Scikit-learn (Rede Neural) | 3,38950 | 1,84110 | 1,12340 | 0,96480 |
| Modelo Keras: época =1000 (Rede Neural) | 4,54000 | 2,12600 | 1,27670 | 0,95300 |
| Modelo MQO (Adicionado de outliers) deg=1 | 226,69500 | | | |
| Modelo Keras: época =1000 (Adicionado de outliers) (Rede Neural) | 6,93630 | | | |
| Testes t pareado (Comparam a diferença média dos resíduos entre pares de modelos.) | T-Statistic | P-Value | | |
| Hipótese Nula (H0): A diferença média dos resíduos é zero. | | | | |
| Hipótese Alternativa (H1): A diferença média dos resíduos não é zero. | | | | |
| Um p-valor baixo (< 0.05) sugere uma diferença estatisticamente significativa. | | | | |
| Linear vs Polimomial Grau 2 | 0,00E+00 | 1,000000 | | |
| Polimomial Grau 2 vs Scikit-learn | 1,07E+00 | 0,294200 | | |
| Scikit-learn vs Keras | -1,46E+00 | 0,157400 | | |
| Teste de Normalidade dos Resíduos (Shapiro-Wilk) | Estatística W | P-Value | | |
| Resíduos do Modelo Linear: | 0,94440 | 0,1866 | | |
| Resíduos do Modelo Polimomial (Grau 3): | 0,96350 | 0,4892 | | |
| Resíduos do Modelo Scikit-Learn: | 0,81590 | 0,0004 | | |
| Resíduos do Modelo Keras: | 0,87100 | 0,0045 | | |
| Testa se os resíduos seguem uma distribuição normal. | | | | |
| Hipótese Nula (H0): Os resíduos são normalmente distribuídos. | | | | |
| Hipótese Alternativa (H1): Os resíduos NÃO são normalmente distribuídos. | | | | |
| Um p-valor baixo (< 0.05) sugere que os resíduos NÃO são normais. | | | | |
| Teste de Homocedasticidade (Breusch-Pagan) | Estatística BP | P-Valor | Estatística F | P-Valor |
| P-valor baixo (geralmente < 0.05) para o teste de Breusch-Pagan sugere que há evidência de heterocedasticidade (a variância dos resíduos não é constante). | | | | |
| P-valor alto (geralmente >= 0.05) sugere que não há evidência suficiente para rejeitar a hipótese de homocedasticidade. | | | | |
| Resíduos do Modelo Linear: | 0,51210 | 0,47420 | 0,48100 | 0,49490 |
| Resíduos do Modelo Polimomial (Grau 3): | 4,12360 | 0,24840 | 1,38270 | 0,27560 |
| Resíduos do Modelo Scikit-Learn: | 8,19490 | 0,00280 | 12,74740 | 0,00160 |
| Resíduos do Modelo Keras: | 4,99060 | 0,02550 | 5,73650 | 0,02520 |

Fonte: Elaboração própria.

Uma observação que se pode considerar é a respeito da Significância Estatística vs. Significância Prática — em que um resultado estatisticamente não significativo (*p-valor* > 0.05) não significa necessariamente que os modelos são idênticos em termos práticos —, sendo possível haver pequenas diferenças no desempenho entre os modelos testados, que não são detectadas como estatisticamente significativas devido, por exemplo, ao tamanho da amostra ou à variabilidade dos dados. Dessa forma, é possível a utilização prática de tais modelos em alguns casos específicos.

Teste Quantitativo 4. Análise de Robustez. Nova função/cenário: $f: R \rightarrow R, y = [2/(\log x + 2)] + 3x^2 - 0,5x^3 + 10$ (Adicionados dados outliers)

Esse exercício é similar ao anterior, mas agora tem-se uma montagem de um novo cenário com uma função matemática mais complexa, sendo a nova função matemática (27) demonstrada abaixo.

$$: R \rightarrow R, y = \frac{2}{\log x + 2} + 3x^2 - 0,5x^3 + 10 \quad (27)$$

Tabela 4 – Resumo dos Resultados do Teste Quantitativo 4

| Métricas de Avaliação dos Modelos | MSE | RMSE | MAE | R ² |
|--|----------------|----------|---------------|----------------|
| Modelo MQO deg=1 | 13,16168 | 3,62790 | 3,13989 | 0,65522 |
| Modelo MQO deg=2 | 2,99772 | 1,73139 | 1,48451 | 0,92147 |
| Modelo MQO deg=3 | 0,80475 | 0,89708 | 0,60778 | 0,97892 |
| Modelo Polimomial deg=1 | 45,08290 | 4,95070 | 6,71440 | 0,38530 |
| Modelo Polimomial deg=2 | 29,09760 | 5,39420 | 4,45600 | 0,60330 |
| Modelo Polimomial deg=3 | 27,94250 | 5,28610 | 4,36060 | 0,61900 |
| Modelo Polimomial deg=4 | 27,86730 | 5,27890 | 4,35540 | 0,62000 |
| Modelo Polimomial deg=5 | 26,16270 | 5,11500 | 4,23900 | 0,64330 |
| Modelo Scikit-learn (Rede Neural) | 43,69090 | 6,60990 | 5,11340 | 0,40430 |
| Modelo Keras: época =1000 (Rede Neural) | 24,92390 | 4,99240 | 4,03200 | 0,66020 |
| Modelo MQO (Adicionado de outliers) deg=1 | 161,18550 | | | |
| Modelo Keras: época =1000 (Adicionado de outliers) (Rede Neural) | 88,97960 | | | |
| Testes t pareado (Comparam a diferença média dos resíduos entre pares de modelos.) | T-Statistic | P-Value | | |
| Hipótese Nula (H0): A diferença média dos resíduos é zero. | | | | |
| Hipótese Alternativa (H1): A diferença média dos resíduos não é zero. | | | | |
| Um p-valor baixo (< 0,05) sugere uma diferença estatisticamente significativa. | | | | |
| Linear vs Polimomial Grau 3 | 0,00E+00 | 1,000000 | | |
| Polimomial Grau 3 vs Scikit-learn | -6,72E+00 | 0,000000 | | |
| Scikit-learn vs Keras | 6,14E+00 | 0,000000 | | |
| Teste de Normalidade dos Resíduos (Shapiro-Wilk) | Estatística W | P-Value | | |
| Resíduos do Modelo Linear: | 0,94800 | 0,2264 | | |
| Resíduos do Modelo Polimomial (Grau 3): | 0,96980 | 0,6407 | | |
| Resíduos do Modelo Scikit-Learn: | 0,98070 | 0,8982 | | |
| Resíduos do Modelo Keras: | 0,97990 | 0,8841 | | |
| Testa se os resíduos seguem uma distribuição normal. | | | | |
| Hipótese Nula (H0): Os resíduos são normalmente distribuídos. | | | | |
| Hipótese Alternativa (H1): Os resíduos NÃO são normalmente distribuídos. | | | | |
| Um p-valor baixo (< 0,05) sugere que os resíduos NÃO são normais. | | | | |
| Teste de Homocedasticidade (Breusch-Pagan) | Estatística BP | P-Valor | Estatística F | P-Valor |
| P-valor baixo (geralmente < 0,05) para o teste de Breusch-Pagan sugere que há evidência de heterocedasticidade (a variância dos resíduos não é constante). | | | | |
| P-valor alto (geralmente >= 0,05) sugere que não há evidência suficiente para rejeitar a hipótese de homocedasticidade. | | | | |
| Resíduos do Modelo Linear: | 1,03000 | 0,31020 | 0,98830 | 0,33050 |
| Resíduos do Modelo Polimomial (Grau 3): | 5,12570 | 0,16280 | 1,80530 | 0,17710 |
| Resíduos do Modelo Scikit-Learn: | 1,56950 | 0,21030 | 1,54070 | 0,22700 |
| Resíduos do Modelo Keras: | 4,64970 | 0,03110 | 5,25510 | 0,03140 |

Fonte: Elaboração própria.

A equação (27) é um tipo de função que contém, obviamente, complexidade maior do que se comparada a equação (26). No entanto, de forma surpreendente, o modelo MQO de grau 3 se sobressaiu em termos de desempenho (MSE = 0,80470) em relação ao modelo RN Keras

(MSE = 24,92390) para essa equação e nesse intervalo amostral. No entanto, como no exemplo (3), quando da adição de dados não estruturados à equação, a resposta da modelagem RN Keras (MSE = 88,97960) teve melhor performance, do que o modelo MQO (MSE = 161,18550).

Conclusões

Esta pesquisa teve como objetivo verificar se a metodologia de regressões por mínimos quadrados ordinários (MQO) e a metodologia de redes neurais (RN) têm o mesmo desempenho em situações diversas de testes empíricos para série de dados aplicados em algoritmos de inteligência artificial (IA), em que o problema de pesquisa foi saber se o método de regressão MQO e o método de redes neurais, em geral, apresentam resultados satisfatórios em estimativas estatísticas e para certos tipos de algoritmos de IA. Já a hipótese da pesquisa foi que os algoritmos de IA poderiam ser capazes — em algumas situações — de encontrar ineficiências, onde os métodos tradicionais de econometria — como a regressão por meio de mínimos quadrados ordinários (MQO) — falham. A metodologia utilizada de características aplicadas nesta pesquisa foi a construção e realização de testes estatísticos empíricos por meio de modelos de IA que utilizam algoritmos de Aprendizagem Supervisionada e Aprendizado de Máquina.

Como resultados de pesquisa, tem-se que os testes estatísticos realizados demonstraram que as metodologias de redes neurais e o método tradicional econométrico de regressão por mínimos quadrados ordinários (MQO) são abordagens que podem ser utilizadas em conjunto quando aplicadas em algoritmos de IA, sendo que a técnica de redes neurais, em geral, apresenta melhores resultados para dados com características de variáveis aleatórias. O método de regressão MQO apresenta, em geral, melhores resultados para quando os dados observados têm características variáveis “mais comportadas” em suas trajetórias e menos aleatórias, onde o principal indicador utilizado foi o de Mínimos Quadrados Médio (MSE), que representou a métrica de desempenho de resultado para ambos os testes. Outros indicadores também foram utilizados na pesquisa, como foi o caso do RMSE, MAE e o Coeficiente de determinação R^2 .

A pesquisa também contemplou análise estatística formal ao avaliar a significância e robustez dos resultados, principalmente por meio do Teste T Pareado, análise de normalidade e homoscedasticidade. Em síntese, a pesquisa demonstrou que ambas as metodologias observadas são válidas em processos de criar algoritmos de inteligência artificial. Ou seja, quando a característica da base de dados ou da função matemática a ser estimada apresenta ser

mais “comportada e fechada” ou mesmo com características mais lineares, o método MQO apresenta os melhores algoritmos de IA. Já quando a característica da base de dados indica variáveis mais aleatórias ou mais complexas, o método de redes neurais apresenta melhores desempenhos em algoritmos de IA. Quanto a análise por meio de dados de *outliers* adicionados nas funções e cenários, reforçou o argumento — de forma mais robusta — de que os modelos RN tendem a ser mais densos em relação aos dados estruturados do que se comparado com o modelo MQO. Isso pode significar uma vantagem em dados reais, como por exemplo, dados econômicos e financeiros que geralmente contém anomalias.

Por fim, esta pesquisa ainda pretende se estender na aplicação de tais métodos aqui estudados, no sentido de base de dados de ativos financeiros. Essas abordagens metodológicas ainda podem ser estendidas para outros tipos de algoritmos de IA, como por exemplo, as aplicações de Aprendizado por Reforço.

REFERÊNCIAS

- CHOLLET, F. **Deep learning with Python**. Shelter Island: Manning, 2017.
- DOANE, D. P.; SEWARD, L. E. **Estatística aplicada à administração e economia**. Porto Alegre: Grupo A, 2014.
- GALIT, S. *et al.* **Machine learning for business analytics**: concepts, techniques, and applications with JMP Pro. 2. ed. Hoboken: John Wiley & Sons, 2023.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. Cambridge: MIT Press, 2016. Disponível em: <http://deeplearningbook.org>. Acesso em: 19 jan. 2026.
- GUJARATI, D. N.; PORTER, D. C. **Econometria básica**. Porto Alegre: Grupo A, 2011.
- HAYKIN, S. **Redes neurais**: princípios e prática. Porto Alegre: Grupo A, 2007.
- HILPISCH, Y. **Artificial intelligence in finance**: a Python-based guide. Sebastopol: O'Reilly Media, 2020.
- KIM, T. K. T-test as a parametric statistic. **Korean Journal of Anesthesiology**, v. 68, n. 6, p. 540, 2015.
- KISSINGER, H. A.; SCHMIDT, E.; HUTTENLOCHER, D. **The age of AI and our human future**. New York: Little Brown and Company, 2021.
- KRISHAMOORTHY, V.; ALOK, U. **Regressão linear**: suposições e limitações. QuantInsti, 2022. Disponível em: <https://blog.quantinsti.com/linear-regression-assumptions-limitations/>. Acesso em: 22 jun. 2025.
- LANGENBERG, B. *et al.* A tutorial on using the paired t test for power calculations in repeated measures ANOVA with interactions. **Behavior Research Methods**, v. 55, n. 5, p. 2467-2484, 2023.
- LYKKEGARRD, M. B. *et al.* Data compression for time series modelling: a case study of smart grid demand forecasting. **arXiv preprint**, 2025.
- NETO, A. S. **Estatística e introdução à econometria**. 2. ed. São Paulo: SRV, 2013.
- RAMASUBRAMANIAN, K.; MOOLAYIL, J. **Applied supervised learning with R**: use machine learning libraries of R to build models that solve business problems and predict future trends. Birmingham: Packt Publishing, 2019.

SCHEUCH, C.; VOIGT, S.; WEISS, P. **Tidy finance with R. Boca Raton**: Chapman and Hall/CRC, 2023.

SICSÚ, A. L.; SAMARTINI, A.; BARTH, N. L. **Técnicas de machine learning**. São Paulo: Blucher, 2023.

TALIKAN, A.; AJAN, R. **On paired samples t-test**: applications, examples and limitations. [S. l: s. n.], 2025.

WOOLDRIDGE, J. M. **Introdução à econometria**: uma abordagem moderna. São Paulo: Cengage Learning Brasil, 2023.

CRedit Author Statement

- ☐ **Reconhecimentos:** O autor agradece à Universidade Estadual Paulista (Unesp) pelo apoio institucional e infraestrutura disponibilizada para a realização deste estudo.
 - ☐ **Financiamento:** O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 .
 - ☐ **Conflitos de interesse:** O autor declara não haver conflitos de interesse.
 - ☐ **Aprovação ética:** Não aplicável. (Justificativa: O estudo baseia-se em modelagem matemática/computacional e não envolveu seres humanos ou animais).
 - ☐ **Disponibilidade de dados e material:** Os dados que suportam os resultados deste estudo estão disponíveis mediante solicitação ao autor correspondente.
-