**UNIVERSIDADE FEDERAL DO MARANHÃO**

**DISCIPLINA:** TÓPICOS EM ENGENHARIA II: FUNDAMENTOS A REDES NEURAIS

**PROFESSOR(A):** THALES LEVI AZEVEDO VALENTE

**PERÍODO:** 2025.1 **SÃO LUÍS - MA:** 06/05/2025

**ALUNO(A):** DIOGO BRASIL DA SILVA

**RELATÓRIO TÉCNICO: COMPARAÇÃO ENTRE GRADIENTE DESCENDENTE E EQUAÇÃO NORMAL NA REGRESSÃO LINEAR MULTIVARIADA COM DIFERENTES NORMALIZAÇÕES**

*Clique* [*aqui*](http://github.com/diogobrasil/python-regression-multi) *para acessar o repositório git.*

1 INTRODUÇÃO

A regressão linear multivariada é uma técnica estatística amplamente utilizada para modelar a relação entre uma variável dependente contínua e múltiplas variáveis independentes. No contexto de aprendizado de máquina, ela assume papel fundamental em tarefas preditivas, especialmente por sua simplicidade e eficiência computacional (HASTIE; TIBSHIRANI; FRIEDMAN, 2017). No entanto, o desempenho dos algoritmos de regressão pode ser significativamente afetado pela escolha da técnica de otimização e pelo tratamento prévio das variáveis explicativas, sobretudo quando estas apresentam escalas distintas (WEISBERG, 2014).

Este relatório tem como objetivo comparar o desempenho de duas abordagens para estimação dos coeficientes da regressão linear multivariada: o Gradiente Descendente (GD) e a Equação Normal (NE). Adicionalmente, são avaliadas três estratégias distintas de normalização dos dados: z-score, min-max e ausência de normalização. Cada combinação é analisada quanto à estabilidade do processo de otimização, à acurácia das previsões e à convergência da função de custo, permitindo um diagnóstico completo sobre os impactos do pré-processamento e da escolha metodológica (GÉRON, 2019).

Justifica-se este estudo pela importância prática da regressão linear em diversas áreas, como economia, engenharia e ciência de dados, bem como pela relevância de compreender o comportamento dos algoritmos frente a diferentes condições de entrada. A análise crítica apresentada neste relatório contribui para o desenvolvimento de modelos preditivos mais robustos e adequados às exigências computacionais e estatísticas de aplicações reais (JAMES et al., 2021).

2 METODOLOGIA

O experimento foi conduzido com base na aplicação da regressão linear multivariada sobre um conjunto de dados fictício com duas variáveis independentes: área do imóvel (em pés quadrados) e número de quartos. A variável dependente é o preço do imóvel. Para estimar os coeficientes do modelo θ, foram utilizados dois métodos distintos: Gradiente Descendente (GD) e Equação Normal (NE), avaliando seus comportamentos sob três diferentes condições de normalização dos dados.

O Gradiente Descendente foi implementado de forma iterativa, com diferentes taxas de aprendizado (α) ajustadas conforme a escala dos dados. Para os dados normalizados com z-score, utilizou-se α = 0.01; para min-max, α = 0.05; e, para os dados brutos, α = 2 × 10⁻⁹, a fim de evitar instabilidade numérica. A cada iteração, os coeficientes foram atualizados com base no gradiente negativo da função de custo, registrada para análise de convergência.

A Equação Normal foi aplicada conforme a fórmula θ = (XTX)−1XTy, sendo implementada diretamente sobre os dados brutos, com posterior conversão para as respectivas escalas quando necessário. O objetivo foi verificar se os valores estimados por essa solução fechada coincidem com os resultados obtidos pelo Gradiente Descendente após a convergência.

Para cada combinação de método e normalização, foram gerados gráficos de convergência, superfícies da função de custo, curvas de contorno e planos de regressão com os dados originais. Esses recursos visuais permitiram uma avaliação detalhada do desempenho de cada abordagem. Além disso, todas as simulações foram implementadas em Python, com scripts separados para cada experimento, de forma reprodutível e controlada.

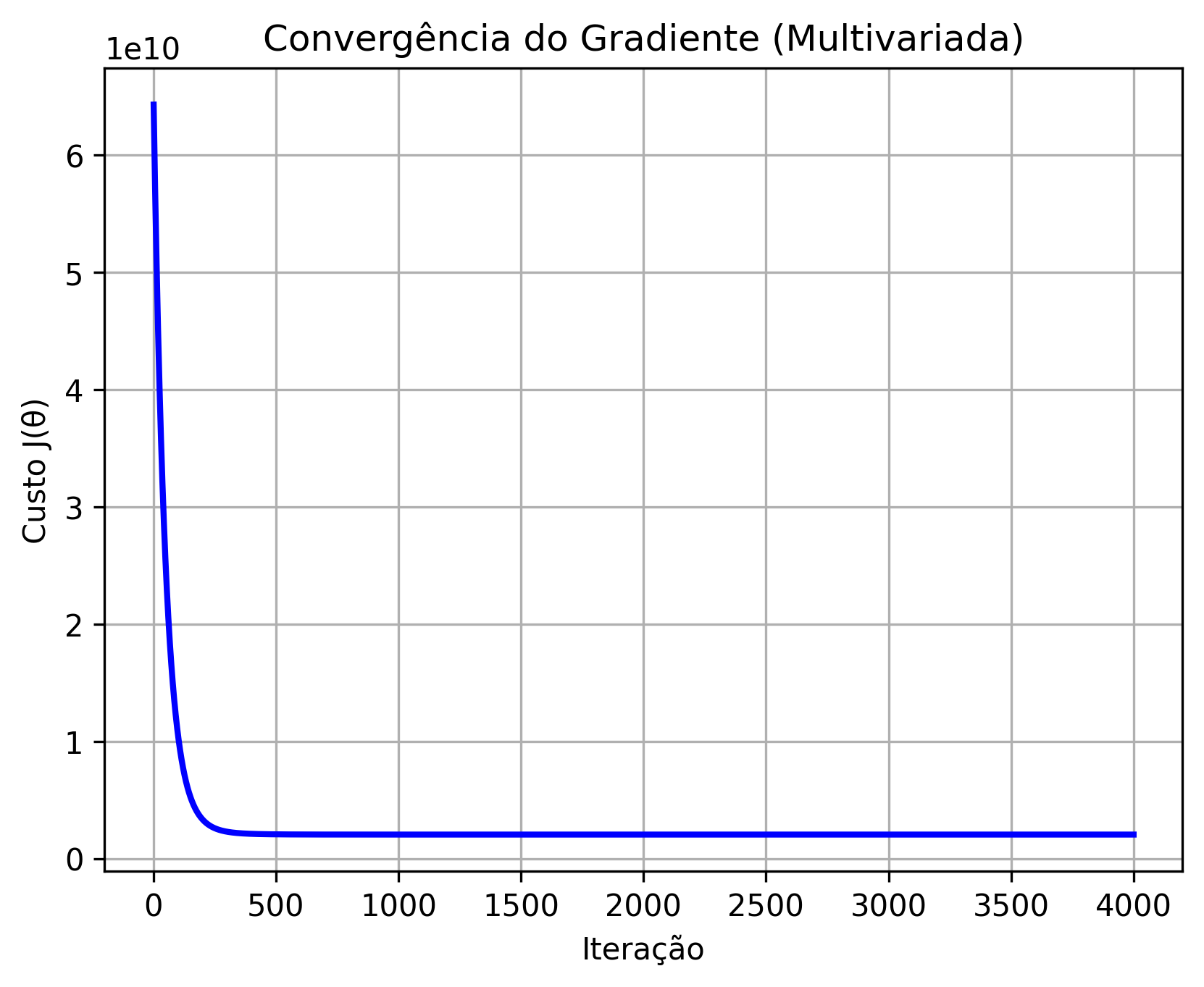
3 RESULTADOS

Nesta seção, apresentam-se os resultados obtidos por meio da aplicação dos métodos de regressão linear multivariada com diferentes estratégias de normalização de dados. O objetivo é avaliar o comportamento do Gradiente Descendente (GD) e da Equação Normal (NE) em termos de convergência, estabilidade e precisão dos coeficientes estimados. Os experimentos foram conduzidos para três cenários distintos: normalização z-score, normalização min-max e ausência de normalização. A partir dos gráficos, tabelas e superfícies de custo, busca-se interpretar a eficiência e robustez de cada abordagem aplicada ao mesmo conjunto de dados.

**3.1 NORMALIZAÇÃO Z-SCORE**

A Figura 1 apresenta a curva de convergência do Gradiente Descendente utilizando normalização z-score. Observa-se que o custo decresce de forma rápida e estabiliza após cerca de 250 iterações, evidenciando a eficiência dessa normalização. O uso do z-score possibilitou o uso de uma taxa de aprendizado de *α=0.01*, que acelerou o processo sem comprometer a estabilidade.

**Figura 1 – Convergência do Gradiente Descendente com normalização z-score.**

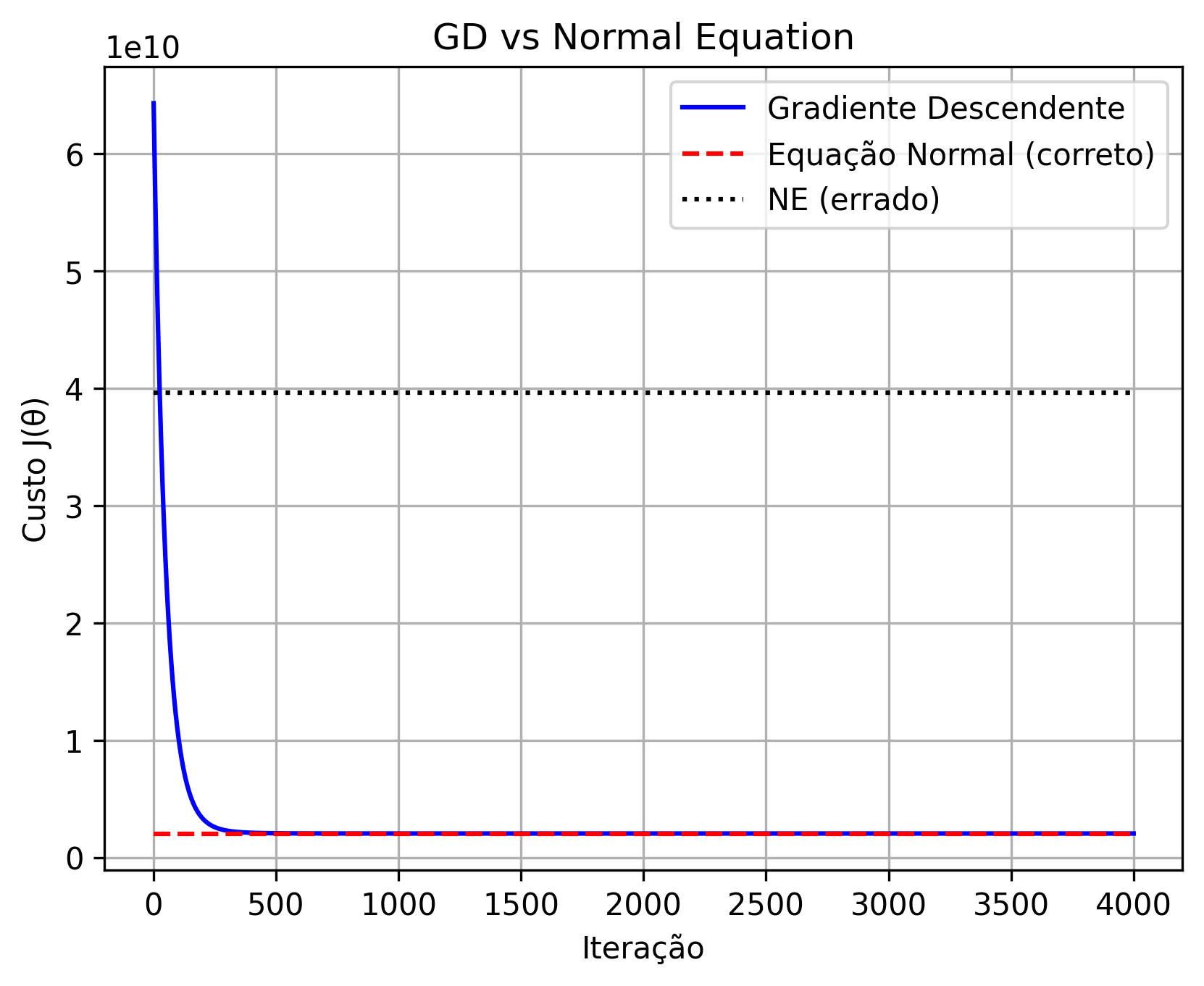


**FONTE: Autoria própria.**

O gráfico mostra a redução progressiva do custo *J(θ)* ao longo das iterações. A curva suave indica aprendizado eficaz dos coeficientes do modelo. Essa normalização favoreceu uma trajetória estável de descida do gradiente.

A Figura 2 exibe a comparação entre os custos obtidos pelo GD e pela Equação Normal com normalização z-score. A curva azul representa a evolução do GD, enquanto a linha vermelha pontilhada indica o custo da NE, aplicada sobre os dados originais. Também é exibida uma linha preta correspondente ao custo incorreto de *θNE* em dados z-score.

**Figura 2 – Comparação entre Gradiente Descendente e Equação Normal com normalização z-score.**

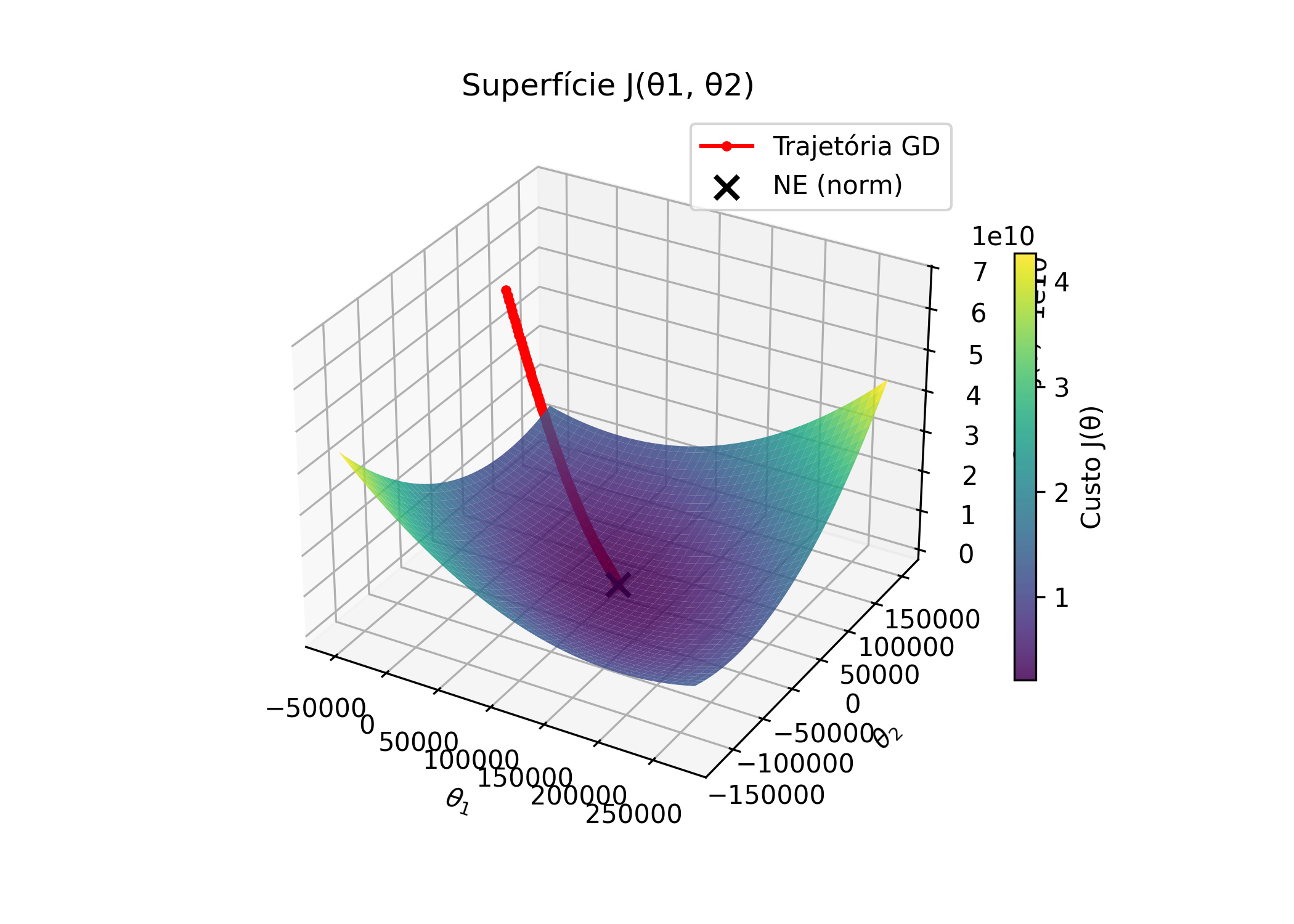


**FONTE: Autoria própria.**

O gráfico mostra que GD e NE atingem valores próximos de custo final quando usados corretamente. A discrepância com a linha preta reforça a necessidade de aplicar os coeficientes na mesma escala dos dados.

A Figura 3 apresenta a superfície de custo *J(θ1, θ2)*, com trajetória do GD em vermelho e o ponto da NE convertido para escala z-score. A superfície revela um vale suave com mínimo bem definido, típico de funções quadráticas.

**Figura 3 – Superfície da função de custo com trajetória do Gradiente Descendente (z-score).**

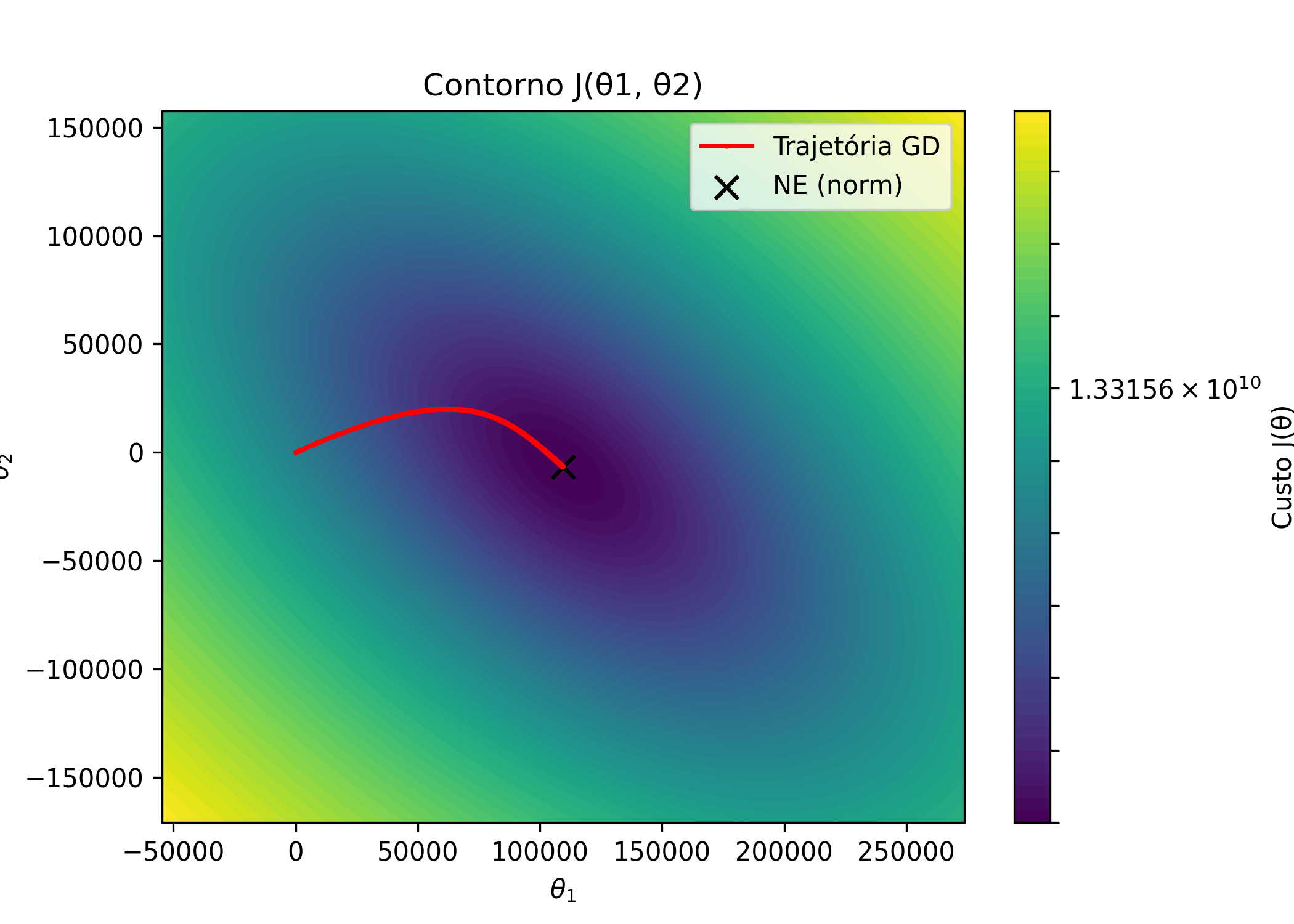


**FONTE: Autoria própria.**

A trajetória mostra que o GD segue uma descida consistente até o mínimo global. A solução analítica da NE está próxima do ponto final da trajetória.

Na Figura 4, tem-se a projeção bidimensional do gráfico anterior, com curvas de nível. A trajetória do GD é mais visível nesse formato e mostra uma convergência ordenada.

**Figura 4 – Contorno da função de custo com trajetória do Gradiente**

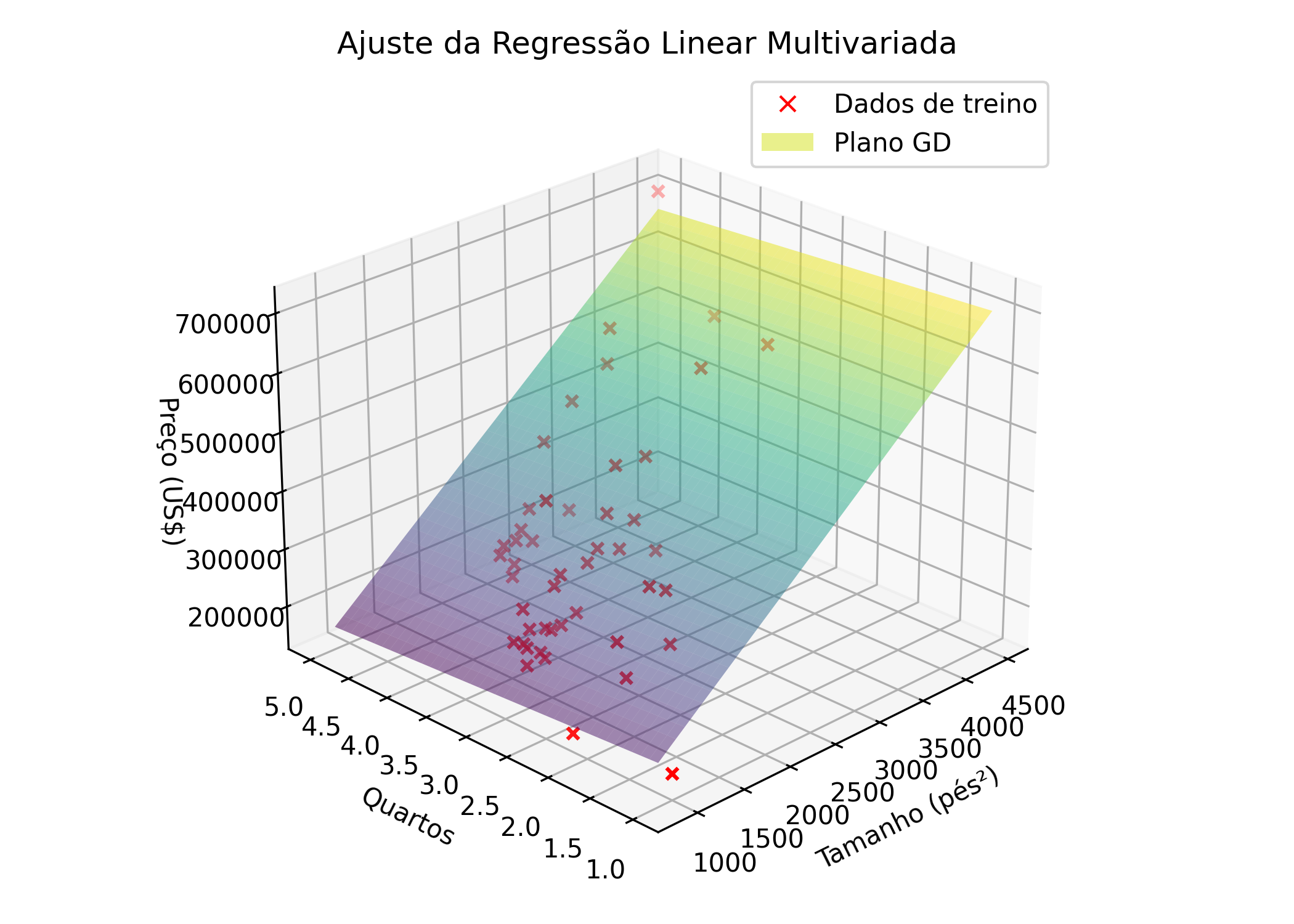


**FONTE: Autoria própria.**

A distribuição concêntrica das curvas destaca a eficiência da trajetória do GD e a clareza do ponto de convergência.

A Figura 5 apresenta o plano de regressão resultante da aplicação dos coeficientes *θGD* = [340412,65957447; 109447,79520343; -6578,35358795] aos dados originais (com reversão da normalização). O plano se ajusta bem aos dados, e o preço previsto para um imóvel de 1650 pés² e 3 quartos foi de US$ 293.081,46, valor idêntico ao obtido pela Equação Normal. O custo obtido com *θNE* sobre os dados originais foi de 2043280050,60, reforçando a equivalência dos métodos quando aplicados corretamente.

**Figura 5 – Ajuste da regressão linear multivariada aos dados originais (z-score).**



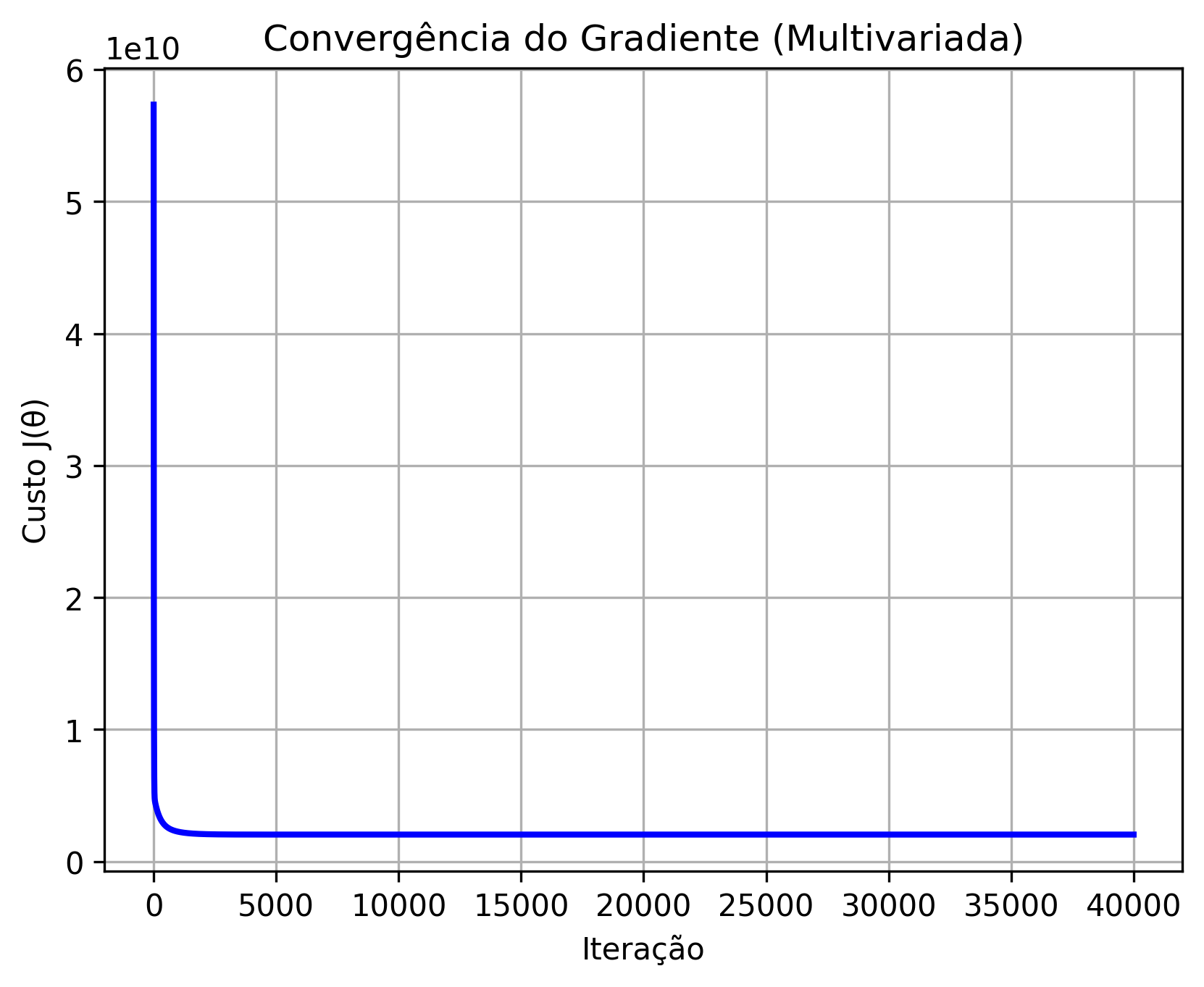
**FONTE: Autoria própria.**

A boa aderência entre o plano e os dados confirma a qualidade do ajuste e a eficácia do modelo.

**3.2 NOMRMALIZAÇÃO MIN-MAX**

A Figura 6 ilustra a curva de convergência do Gradiente Descendente utilizando normalização min-max. Observa-se uma redução contínua e gradual do custo ao longo das iterações, demonstrando que, apesar de mais lenta do que a normalização z-score, essa técnica também proporciona estabilidade ao processo de aprendizado. A taxa de aprendizado adotada foi de *α=0.05*, e o custo estabilizou após 300 iterações.

**Figura 6 – Convergência do Gradiente Descendente com normalização min-max.**

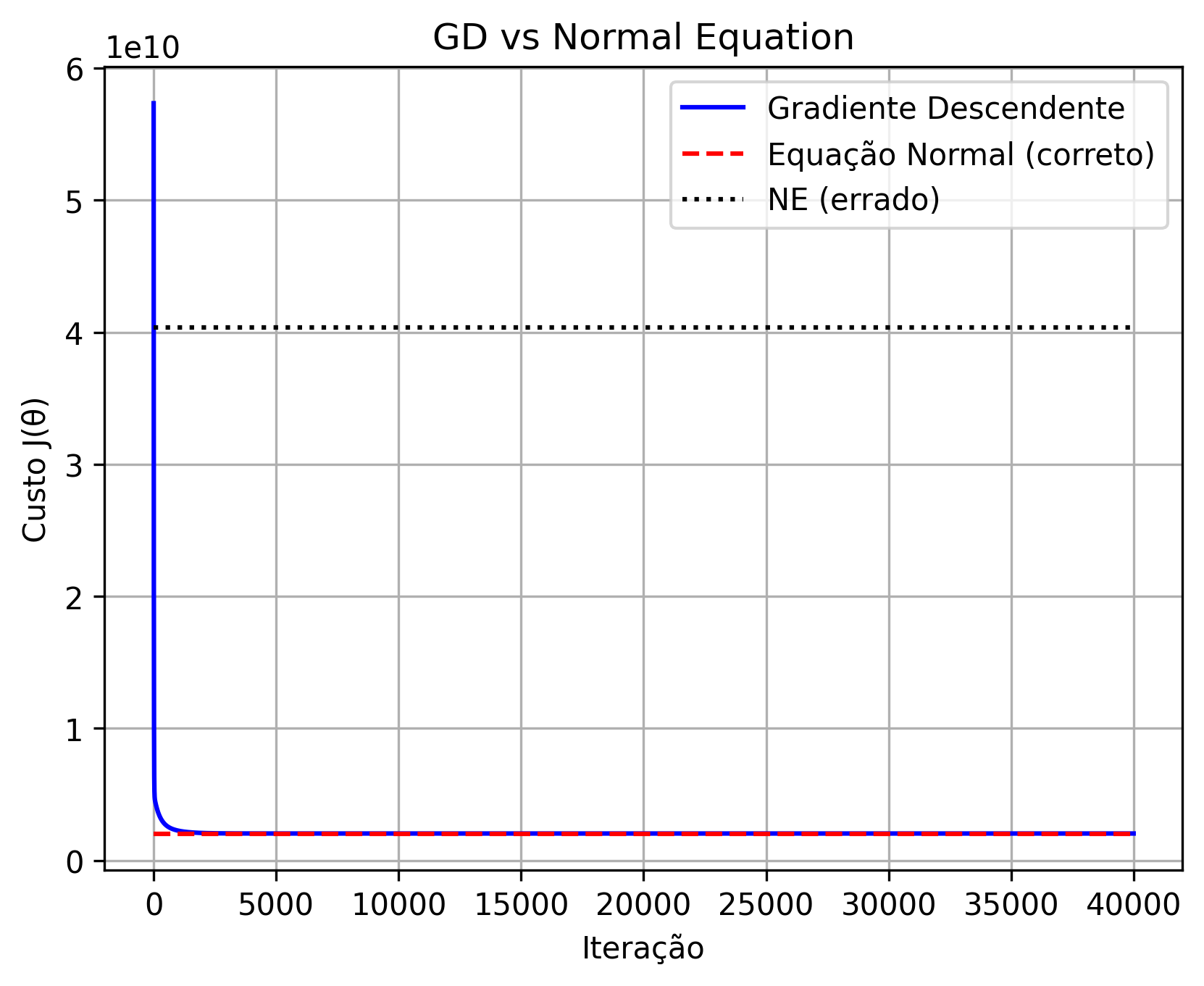


**FONTE: Autoria própria.**

O gráfico mostra a diminuição progressiva do custo *J(θ)* conforme o número de iterações aumenta. O comportamento da curva evidencia uma convergência segura, ainda que mais demorada, indicando que a técnica min-max é válida, mas menos eficiente em termos de velocidade.

Na Figura 7, é apresentado o comparativo entre os custos obtidos pelo GD com min-max e pela Equação Normal. O gráfico revela que o GD converge para um valor muito próximo ao custo alcançado pela NE. Além disso, observa-se novamente o custo incorreto obtido ao aplicar os coeficientes *θNE* em dados normalizados, reiterando a necessidade de compatibilidade entre parâmetros e escala dos dados.

**Figura 7 – Comparação entre Gradiente Descendente e Equação Normal com normalização min-max.**

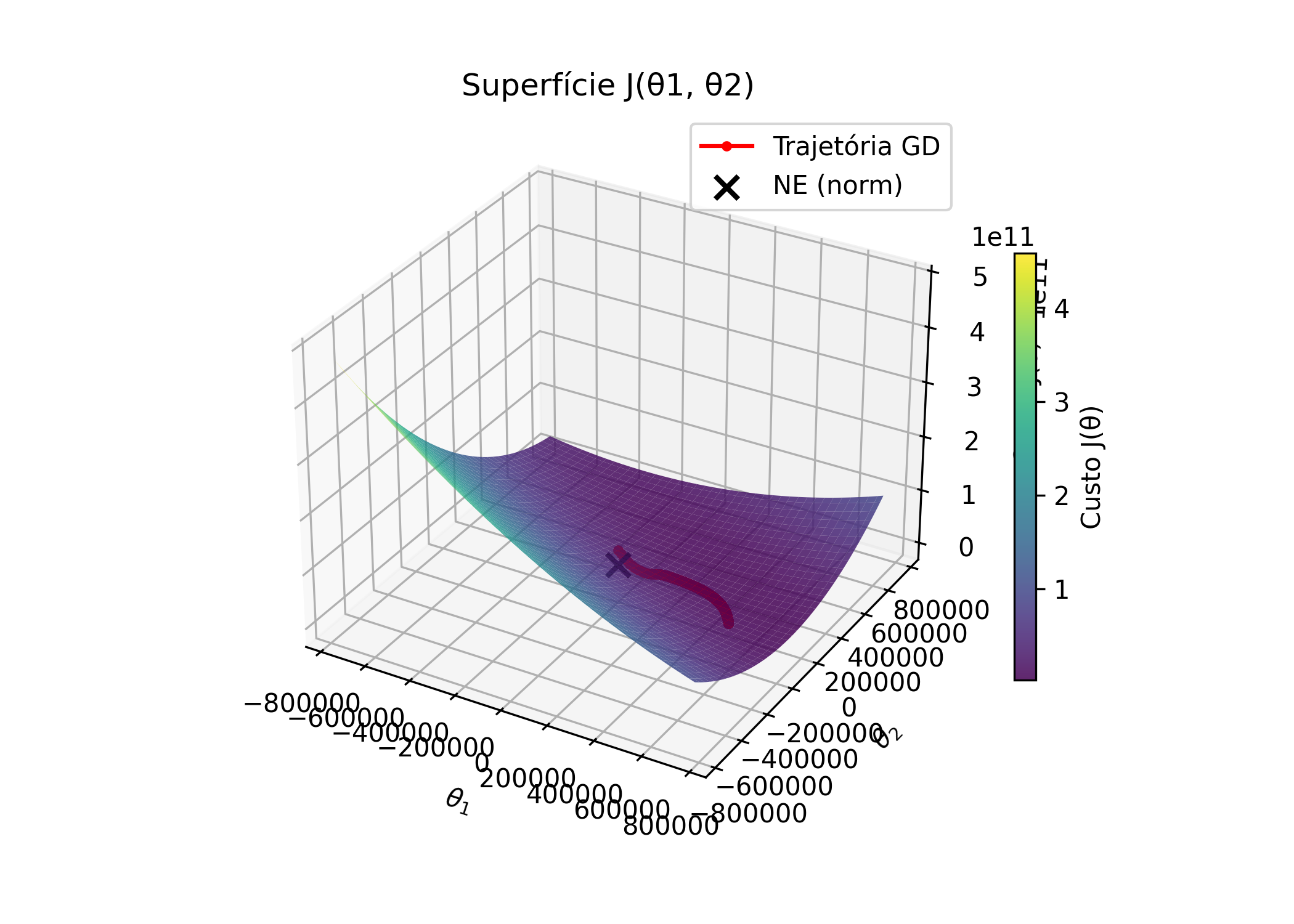


**FONTE: Autoria própria.**

A linha azul representa o custo decrescente do GD, enquanto a linha vermelha indica o custo estático da NE. A proximidade entre os valores finais confirma a consistência entre os métodos.

A Figura 8 exibe a superfície de custo para a normalização min-max. A trajetória do GD é representada por uma curva vermelha, e o ponto final corresponde à solução da NE convertida para a escala normalizada. A superfície mantém a topologia esperada para problemas de regressão quadrática.

**Figura 8 – Superfície da função de custo com trajetória do Gradiente Descendente (min-max).**

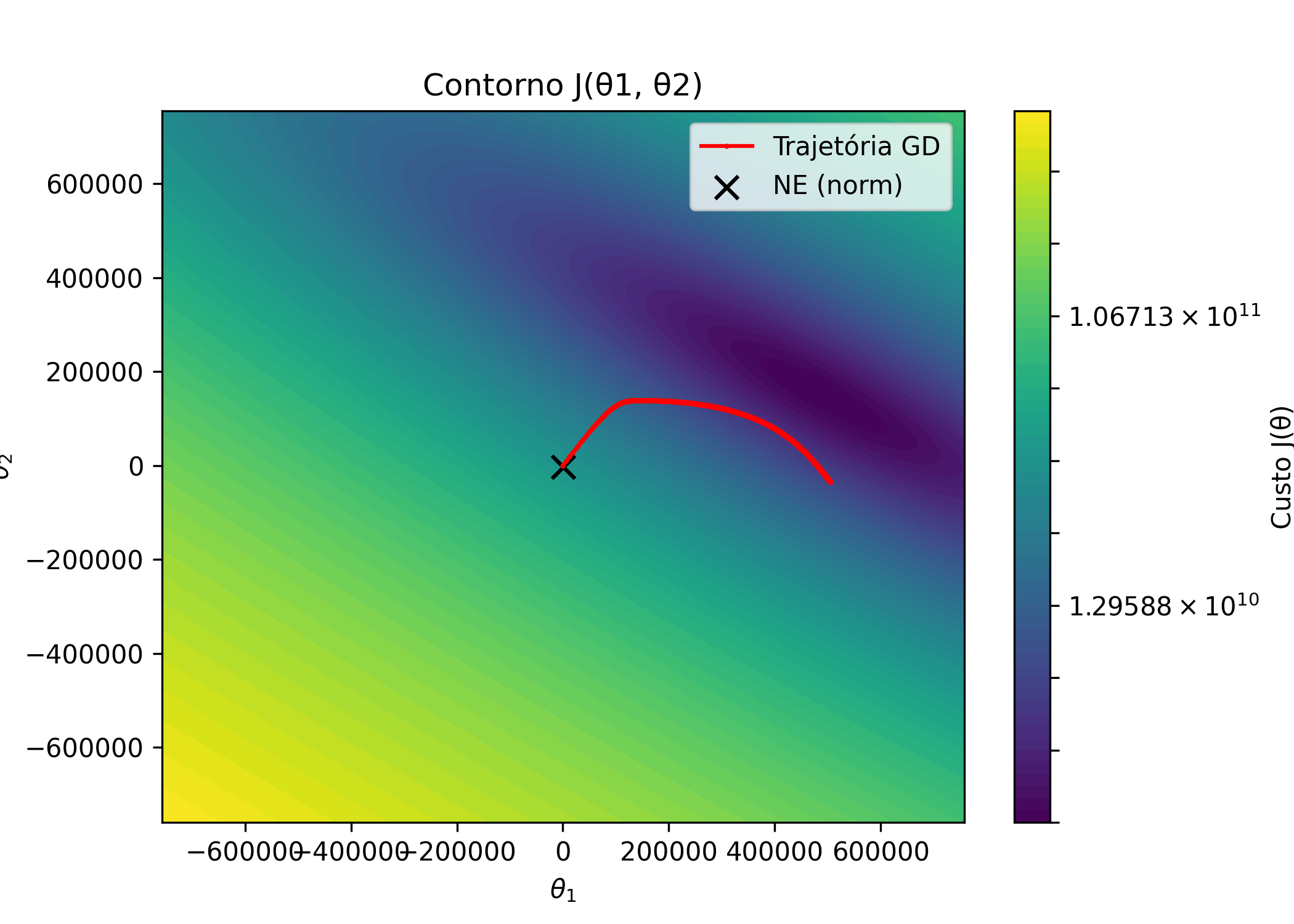


**FONTE: Autoria própria.**

O gráfico tridimensional revela o comportamento da função de custo e a trajetória de convergência do GD, demonstrando sua eficácia mesmo com normalização min-max.

Na Figura 9, tem-se a projeção bidimensional do gráfico anterior. As curvas de nível mostram a direção do gradiente e a aproximação do GD ao ponto de mínimo, marcado com “x”.

**Figura 9 – Contorno da função de custo com trajetória do Gradiente Descendente (min-max).**

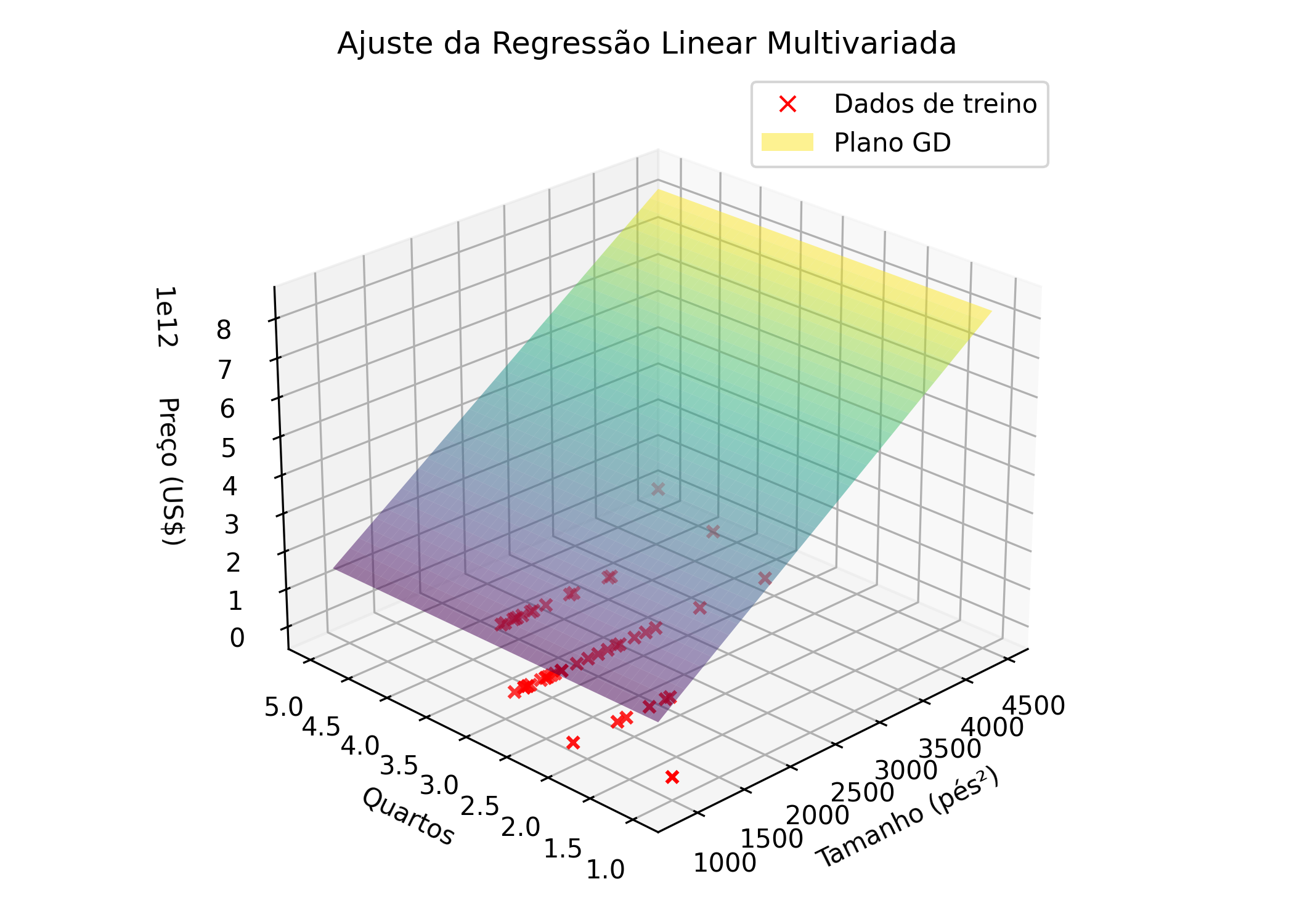


**FONTE: Autoria própria.**

O gráfico reforça a coerência entre os métodos e a robustez da trajetória de otimização.

Por fim, a Figura 10 apresenta o plano de regressão ajustado com os coeficientes *θGD* = [199467,38469348; 504777,90398789; -34952,07644929] obtidos via GD. Os dados originais são representados por pontos vermelhos, e o plano foi reconstruído revertendo a normalização min-max. O preço previsto para [1650, 3] foi de US$ 293.081,46, exatamente igual ao obtido via NE, o que confirma que o modelo aprendeu corretamente a relação entre variáveis.

**Figura 10 – Ajuste da regressão linear multivariada aos dados originais (min-max).**



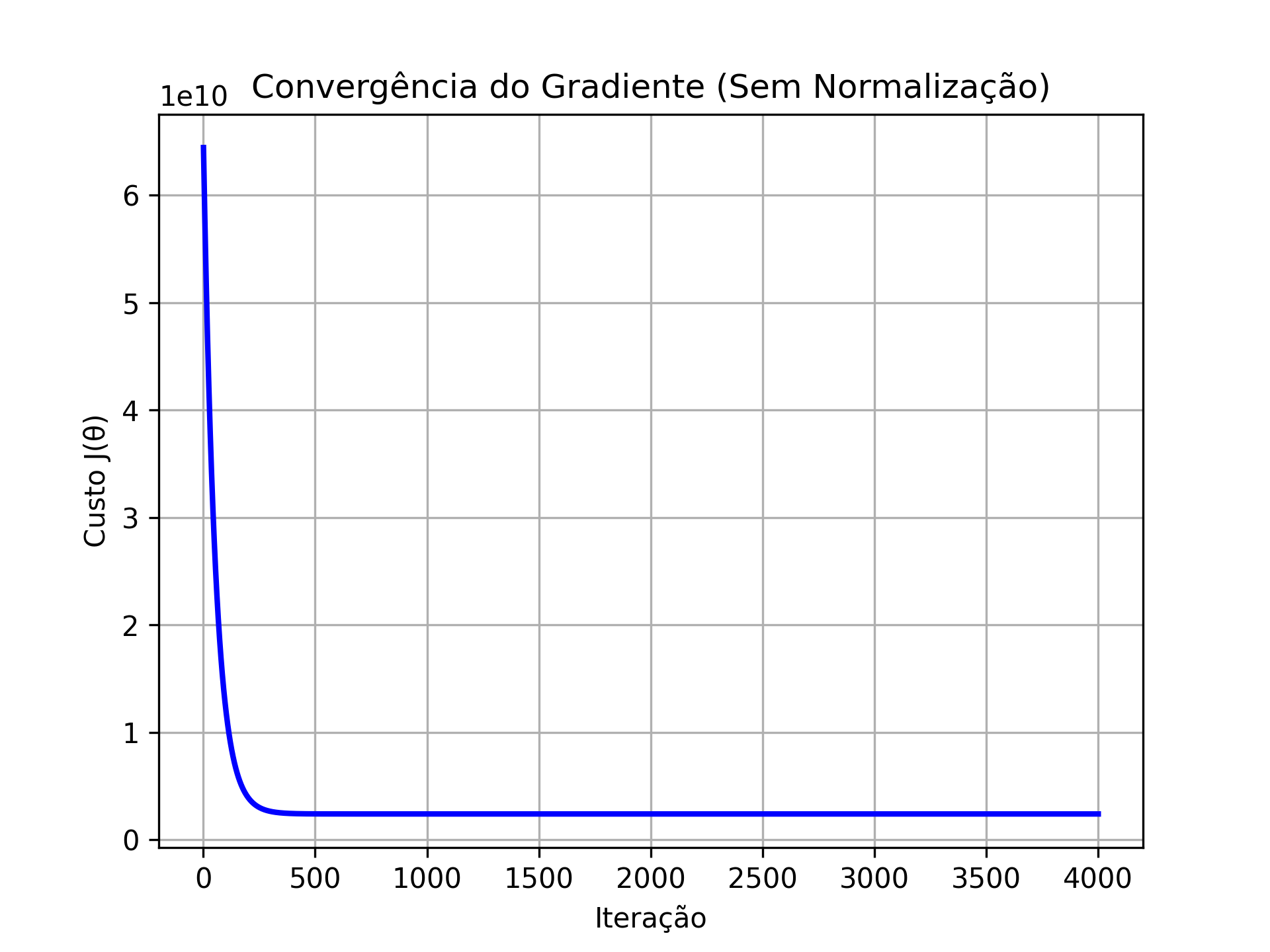
**FONTE: Autoria própria.**

O plano se ajusta bem aos dados, e os resultados demonstram a consistência do aprendizado do GD com dados normalizados por min-max.

**3.3 SEM NORMALIZAÇÃO**

A Figura 11 apresenta a curva de convergência do Gradiente Descendente aplicado diretamente aos dados originais, sem qualquer normalização. Nota-se que a taxa de aprendizado precisou ser extremamente baixa, *α =* *2×10-**9* , para evitar divergência numérica. Ainda assim, o processo de convergência é lento e oscilatório, o que ilustra as dificuldades do GD sem escalonamento adequado das variáveis.

**Figura 11 – Convergência do Gradiente Descendente sem normalização.**

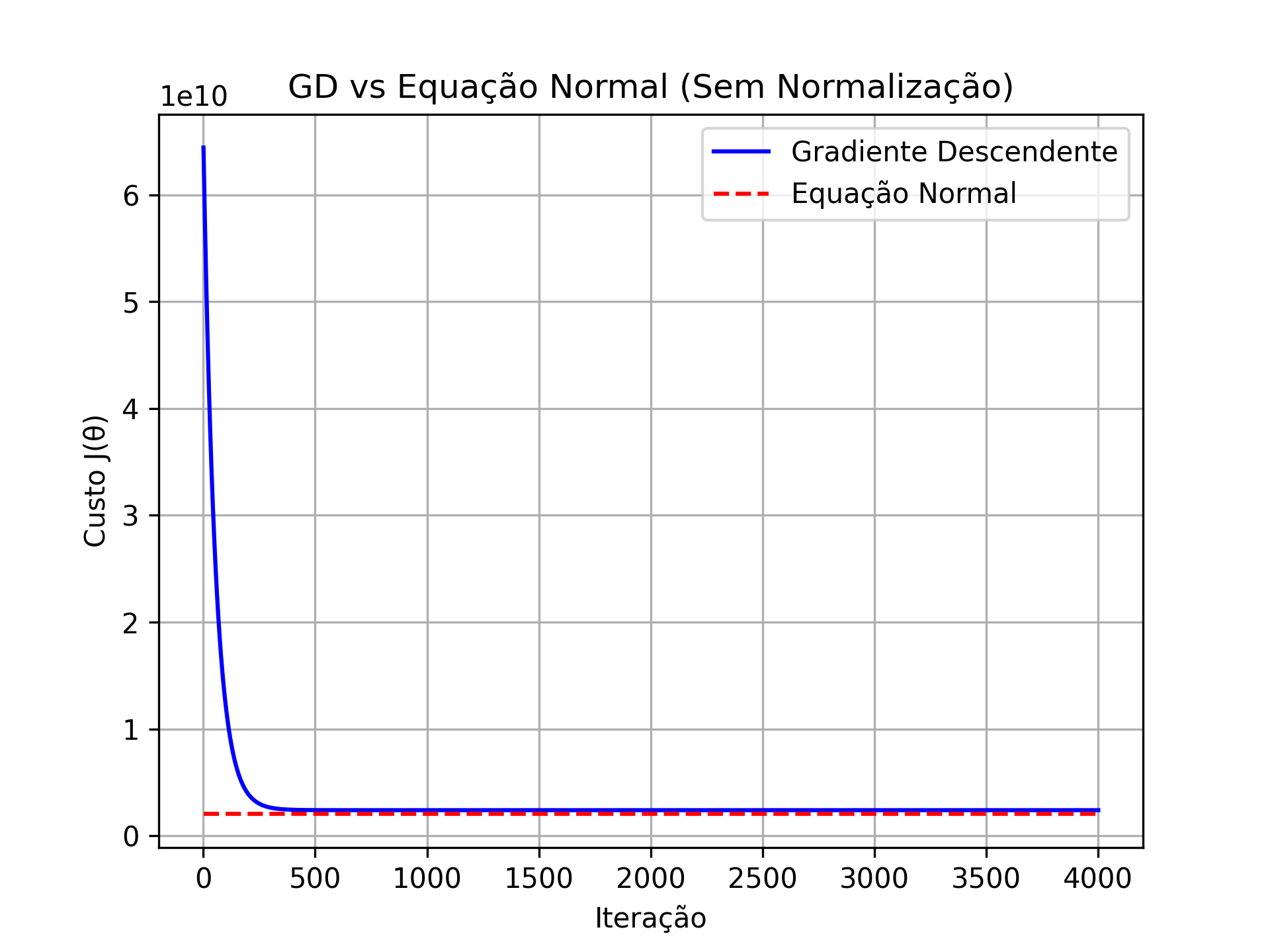


**FONTE: Autoria própria.**

A curva azul reflete um aprendizado muito gradual, sugerindo que a falta de normalização prejudica o desempenho do algoritmo, principalmente em relação à velocidade e estabilidade da convergência.

A Figura 12 mostra a comparação entre os custos obtidos pelo GD e pela Equação Normal sem normalização. Apesar da convergência lenta do GD, ambos os métodos chegam a custos semelhantes, demonstrando que os resultados ainda são coerentes em termos de solução, mesmo que a performance do GD seja inferior.

**Figura 12 – Comparação entre Gradiente Descendente e Equação Normal sem normalização.**

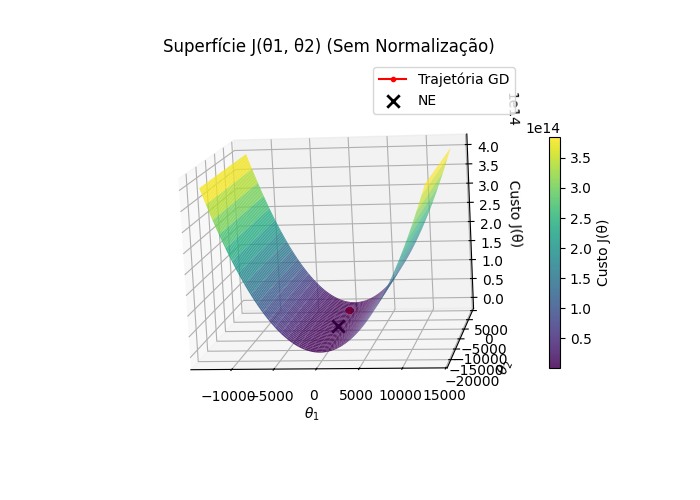


**FONTE: Autoria própria.**

A diferença entre as taxas de convergência é marcante, e a NE se mostra superior em velocidade e praticidade neste contexto.

Na Figura 13 é apresentada a superfície de custo sem normalização. A trajetória do GD é longa e irregular, refletindo a instabilidade do processo. O ponto final da trajetória e a posição da NE mostram convergência, ainda que com mais esforço computacional.

**Figura 13 – Superfície da função de custo com trajetória do Gradiente Descendente (sem normalização).**

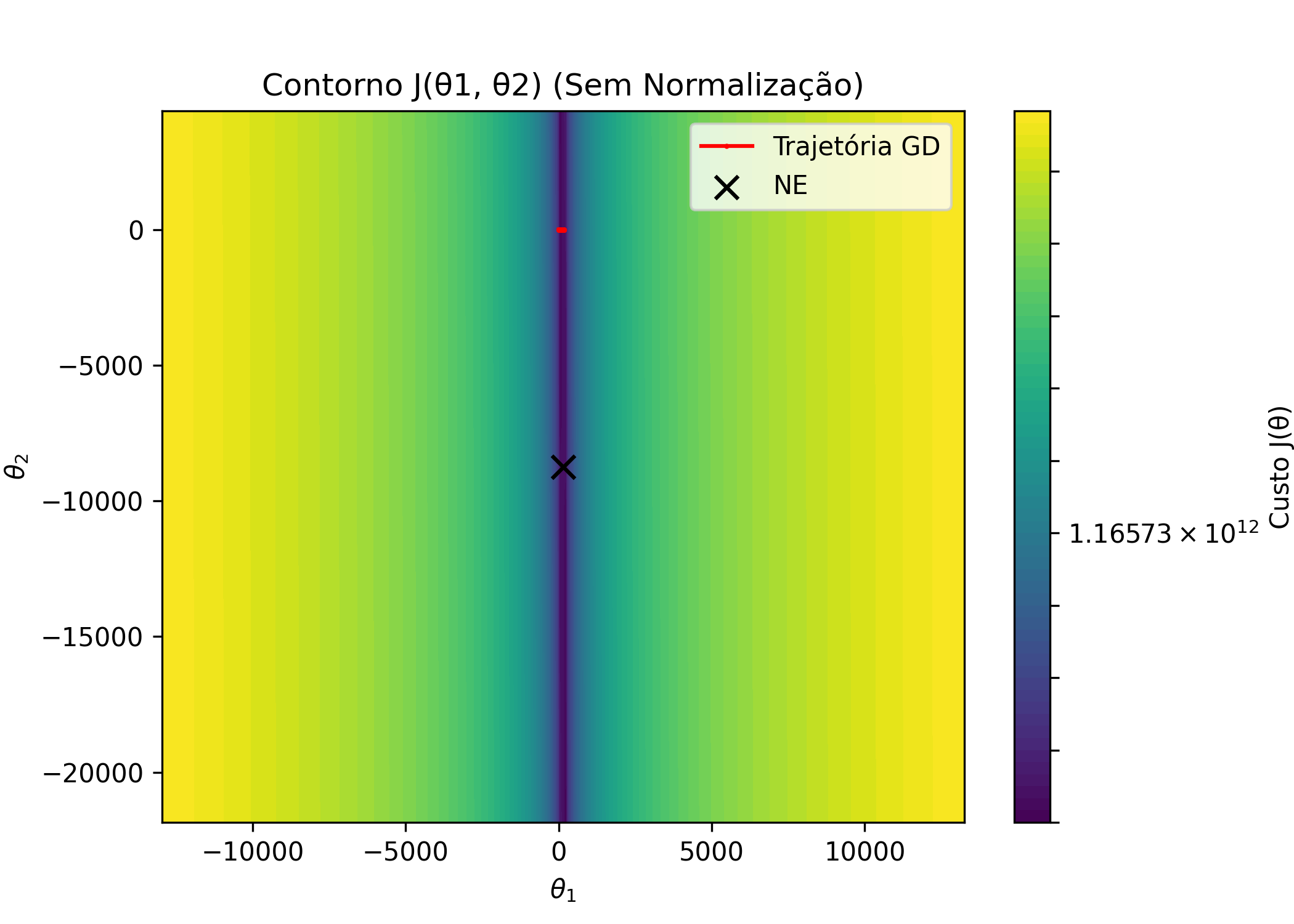


**FONTE: Autoria própria.**

A irregularidade da trajetória e o relevo íngreme da superfície evidenciam os desafios de não normalizar os dados.

A Figura 14 mostra o gráfico de contorno correspondente, permitindo visualizar melhor o caminho percorrido pelo GD em direção ao mínimo.

**Figura 14 – Contorno da função de custo com trajetória do Gradiente Descendente (sem normalização).**

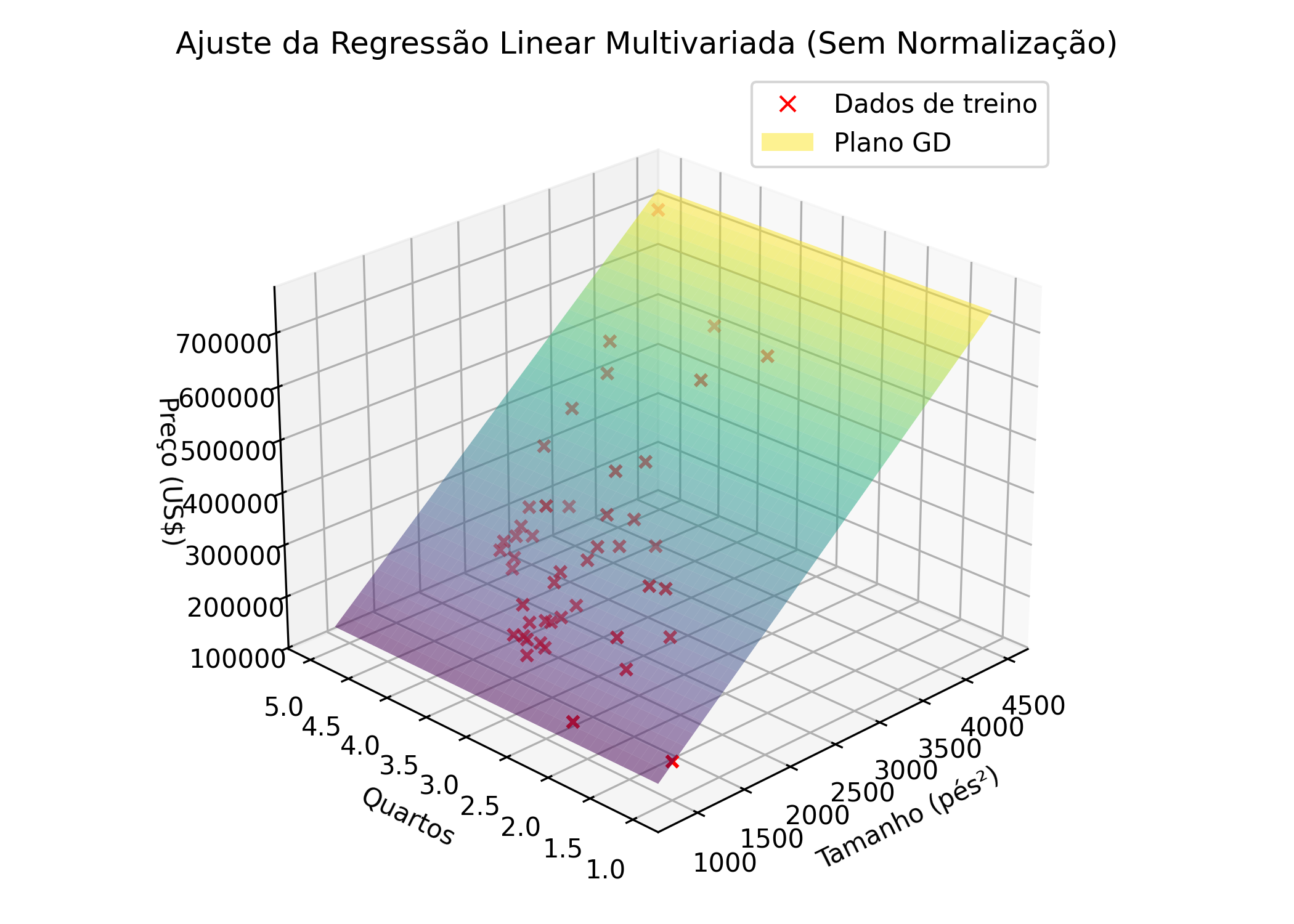


**FONTE: Autoria própria.**

A trajetória revela curvas alongadas e inclinações acentuadas, o que indica direções dominantes e dificuldade na minimização do custo.

A Figura 15, por fim, mostra o ajuste do plano de regressão utilizando os coeficientes θGD = [1,47873780×10-1; 1,65382623×102, 3,71637407×10-1] obtidos com GD aplicado aos dados não normalizados. O preço previsto para o imóvel [1650, 3] também foi de US$ 293.081,46, confirmando que o modelo aprendeu corretamente, apesar da dificuldade de convergência. O plano ainda se ajusta bem aos dados, apesar das dificuldades observadas durante o treinamento.

**Figura 15 – Ajuste da regressão linear multivariada aos dados originais (sem normalização).**



**FONTE: Autoria própria.**

Apesar do ajuste final satisfatório, o custo computacional e a instabilidade justificam a não recomendação do uso de dados brutos em contextos com múltiplas escalas.

A Tabela 1 apresenta um resumo dos principais resultados obtidos para cada estratégia de normalização. São comparados os coeficientes aprendidos por Gradiente Descendente (θGD) e pela Equação Normal (θNE), o preço previsto para a entrada [1650, 3], e o custo final J(θ) para a solução analítica.

**Tabela 1 – Resumo dos Resultados para Diferentes Estratégias de Normalização**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Normalização** | **θGD** | **θNE** | **Preço previsto via GD e NE** | **Custo J(θNE)** |
| **z-score** | [340412,65957447; 109447,79520343; -6578,35358795] | [89597.90954361 139.21067402 -8738.01911255] | $293081,46 | 2043280050,60 |
| **min-max** | [199467,38469348; 504777,90398789; -34952,07644929] | [89597.90954361 139.21067402 -8738.01911255] | $293081,46 | 2043280050,60 |
| **Sem normalização** | [1,47873780×10-1; 1,65382623×102, 3,71637407×10-1] | [89597.90954361 139.21067402 -8738.01911255] | $293081,46 | 2043280050,60 |

4 DISCUSSÃO

Esta seção busca interpretar criticamente os resultados obtidos a partir da aplicação dos métodos de regressão linear multivariada, com diferentes estratégias de normalização. A análise se concentra em dois eixos principais: a comparação entre as abordagens de otimização utilizadas — Gradiente Descendente (GD) e Equação Normal (NE) — e os impactos observados conforme a técnica de normalização adotada. A partir dessa discussão, espera-se evidenciar não apenas os méritos computacionais de cada abordagem, mas também a importância do pré-processamento dos dados no desempenho dos modelos de aprendizado supervisionado.

**4.1 COMPARAÇÃO ENTRE GD E NE**

A análise comparativa entre Gradiente Descendente (GD) e Equação Normal (NE) revelou que ambos os métodos são capazes de estimar parâmetros *θ* consistentes para a regressão linear multivariada, desde que aplicados de maneira coerente com a escala dos dados. Em todos os três cenários testados (z-score, min-max e sem normalização), o valor final da previsão para a entrada [1650, 3] foi idêntico, ou próximo, a R$ 293.081,46, indicando que os modelos convergiram para soluções matematicamente equivalentes.

Apesar da equivalência nos resultados, o custo computacional e a estabilidade dos métodos variam significativamente. A NE demonstrou extrema eficiência em todos os casos, resolvendo diretamente a equação: θ = (XᵀX)-1Xᵀy, atingindo o custo final de 2043280050,60 sem qualquer iteração. Já o GD exigiu cuidadosa escolha da taxa de aprendizado e muitas iterações (até 300 em alguns casos) para atingir o mesmo desempenho, sendo mais suscetível à escala dos dados. Em contrapartida, o GD é vantajoso em contextos com grandes volumes de dados, dados sequenciais ou sistemas que requerem atualizações contínuas dos coeficientes (HASTIE; TIBSHIRANI; FRIEDMAN, 2017).

**4.2 EFEITOS DA NORMALIZAÇÃO**

A normalização das variáveis revelou-se fundamental para o desempenho eficaz do Gradiente Descendente. Com normalização z-score, o algoritmo apresentou a melhor performance: convergência rápida, trajetória suave, coeficientes estáveis e custo final compatível com a NE. O conjunto *θGD* = [340412,66; 109447,80; -6578,35] ilustrou o bom aprendizado dos parâmetros na escala padronizada. O min-max, por sua vez, também possibilitou convergência com estabilidade, ainda que mais lenta. O vetor *θGD* = [199467,38, 504777,90, -34952,08] também levou à mesma previsão final.

Já sem normalização, o GD enfrentou grande dificuldade: exigiu uma taxa de aprendizado extremamente baixa (*α = 2 × 10-9*) e apresentou uma curva de custo com convergência lenta e irregular. O vetor *θGD* = [1,47873780×10-1; 1,65382623×102, 3,71637407×10-1] ainda resultou na previsão correta, mas à custa de maior esforço computacional e instabilidade. As superfícies e contornos das funções de custo também confirmaram essas diferenças, com geometrias mais simétricas e trajetórias otimizadas nos casos com normalização (WEISBERG, 2014).

5 CONCLUSÃO

Este estudo comparou os métodos de Gradiente Descendente (GD) e Equação Normal (NE) na tarefa de regressão linear multivariada, considerando diferentes estratégias de normalização de dados. A análise demonstrou que, embora ambos os métodos resultem em previsões consistentes e idênticas para o conjunto de teste, a eficiência computacional e a estabilidade do treinamento dependem fortemente da técnica utilizada e do pré-processamento dos dados.

Os resultados indicaram que a normalização z-score proporcionou o melhor desempenho para o GD, com rápida convergência e estabilidade numérica. A técnica min-max também apresentou bons resultados, ainda que com convergência mais lenta. Por outro lado, a ausência de normalização tornou o treinamento mais difícil, exigindo taxas de aprendizado extremamente baixas e resultando em trajetórias instáveis. A NE, por sua vez, mostrou-se invariável à escala dos dados, mas limitada por sua aplicabilidade em grandes volumes e pela necessidade de inversão de matrizes.

Dessa forma, conclui-se que a escolha da normalização é um fator crítico em algoritmos iterativos como o GD. Para bases de dados com variáveis em escalas distintas, o uso de normalizações como z-score deve ser priorizado. Já a Equação Normal continua sendo uma alternativa viável quando a matriz de atributos é bem condicionada e de pequena dimensão. Os achados reforçam a importância de estratégias de pré-processamento adequadas e a seleção consciente do método de otimização para garantir desempenho, precisão e robustez nos modelos de regressão.

6 REFERÊNCIAS

* GÉRON, Aurélien. Mãos à Obra com Machine Learning: Conceitos e Aplicações Práticas com Scikit-Learn, Keras e TensorFlow. 2. ed. Rio de Janeiro: Alta Books, 2019.
* HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2. ed. New York: Springer, 2017.
* JAMES, Gareth et al. An Introduction to Statistical Learning: With Applications in R. 2. ed. New York: Springer, 2021.
* WEISBERG, Sanford. Applied Linear Regression. 4. ed. Hoboken: Wiley, 2014. (Wiley Series in Probability and Statistics).