**Nome: João Pedro de Alcântara Lima**

**Disciplina: EECP0053 - Tópicos Em Engenharia Da Computação II - Fundamentos de Redes Neurais.**

**Assunto/Atividade: Regressão Linear**

**1. Introdução**

A regressão linear é um dos métodos mais fundamentais em aprendizado de máquina. Sua eficiência, no entanto, depende de uma otimização eficaz dos parâmetros, comumente realizada por meio da técnica de descida do gradiente. Neste trabalho, exploramos visualmente a trajetória dos parâmetros durante a execução do algoritmo e sua convergência em relação à função de custo. Através dos experimentos conduzidos, também refletimos sobre a escolha adequada da taxa de aprendizado (α) e a inicialização dos parâmetros, conectando tais conceitos ao fine-tuning em redes neurais.

## 2. Resultados e Discussão

### 2.1 Trajetória da Descida do Gradiente

A Figura 1 mostra o contorno da função de custo J(θ), com as curvas de nível representando diferentes valores dessa função. Sobre o contorno, observa-se a trajetória do algoritmo de descida do gradiente (em vermelho), iniciando de um ponto arbitrário e caminhando em direção ao mínimo.

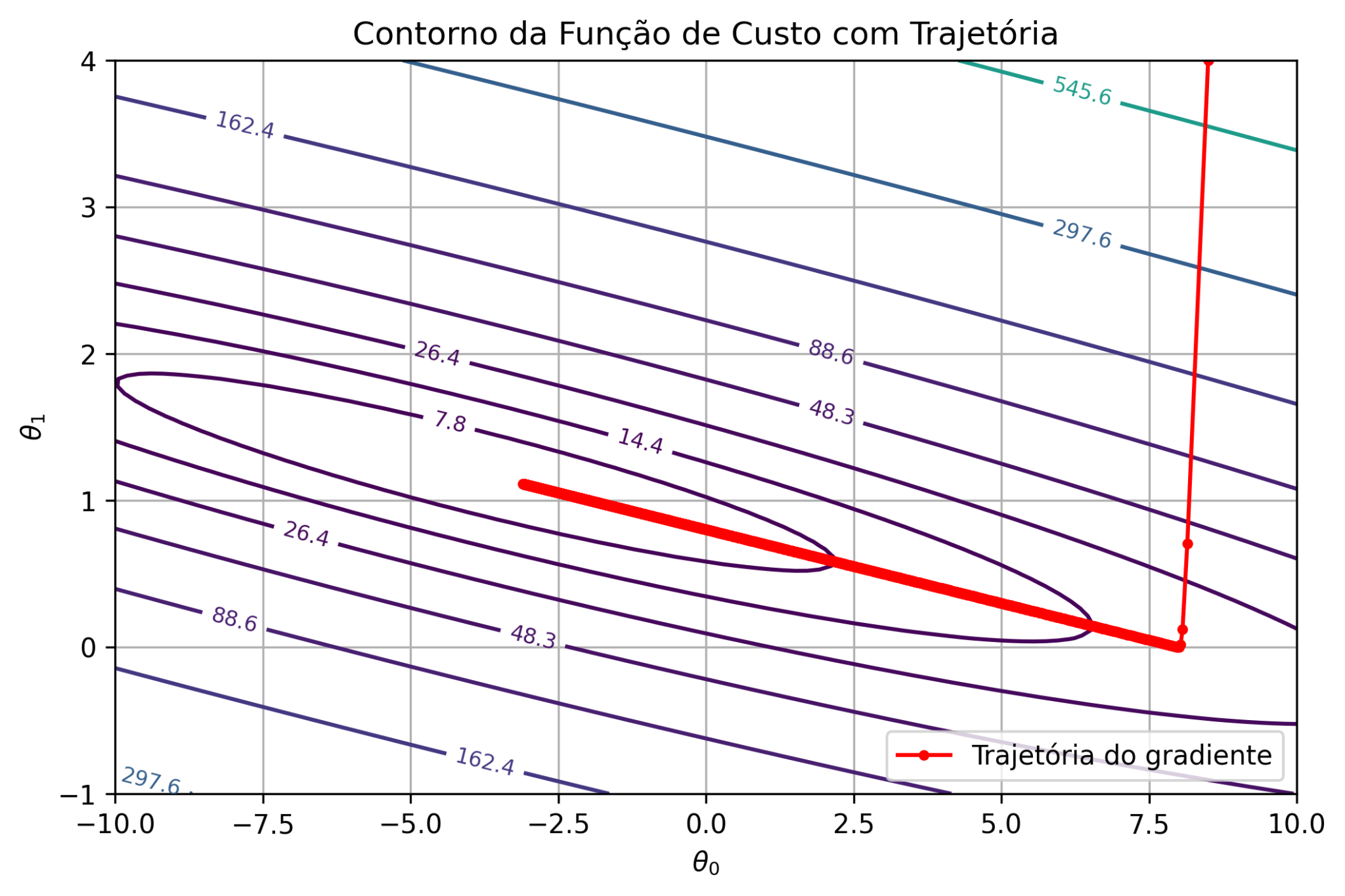


Figura 1 – Contorno da função de custo com trajetória da descida do gradiente.

### 2.2 Convergência da Função de Custo

A Figura 2 apresenta a evolução do valor da função de custo ao longo das iterações.  
Inicialmente, há uma queda brusca no custo, indicando que os primeiros passos trazem grandes ganhos. Com o passar das iterações, as melhorias tornam-se mais sutis, até que a função estabiliza próxima a um valor mínimo. Este comportamento é esperado em processos de otimização.

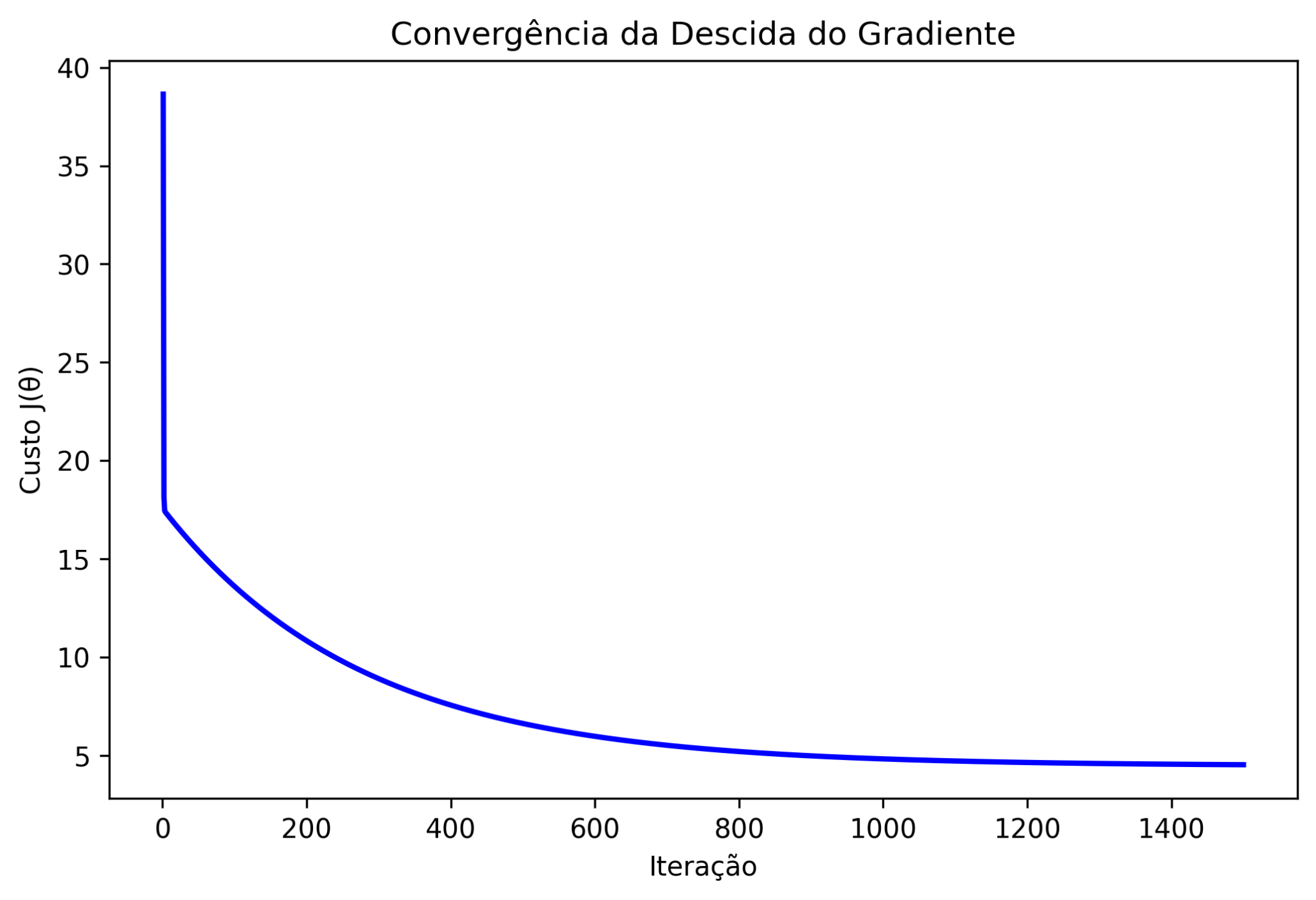


Figura 2 – Convergência da função de custo durante a descida do gradiente.

### 2.3 Visualização Tridimensional da Superfície de Custo

A Figura 3 mostra a superfície 3D da função de custo, permitindo observar como a trajetória da descida do gradiente ocorre no espaço tridimensional formado por θ₀, θ₁ e J(θ).  
A representação permite ver claramente o ponto inicial (em azul), a trajetória (em vermelho) e o ponto de convergência (em verde), ressaltando a eficiência do método na busca por mínimos locais.

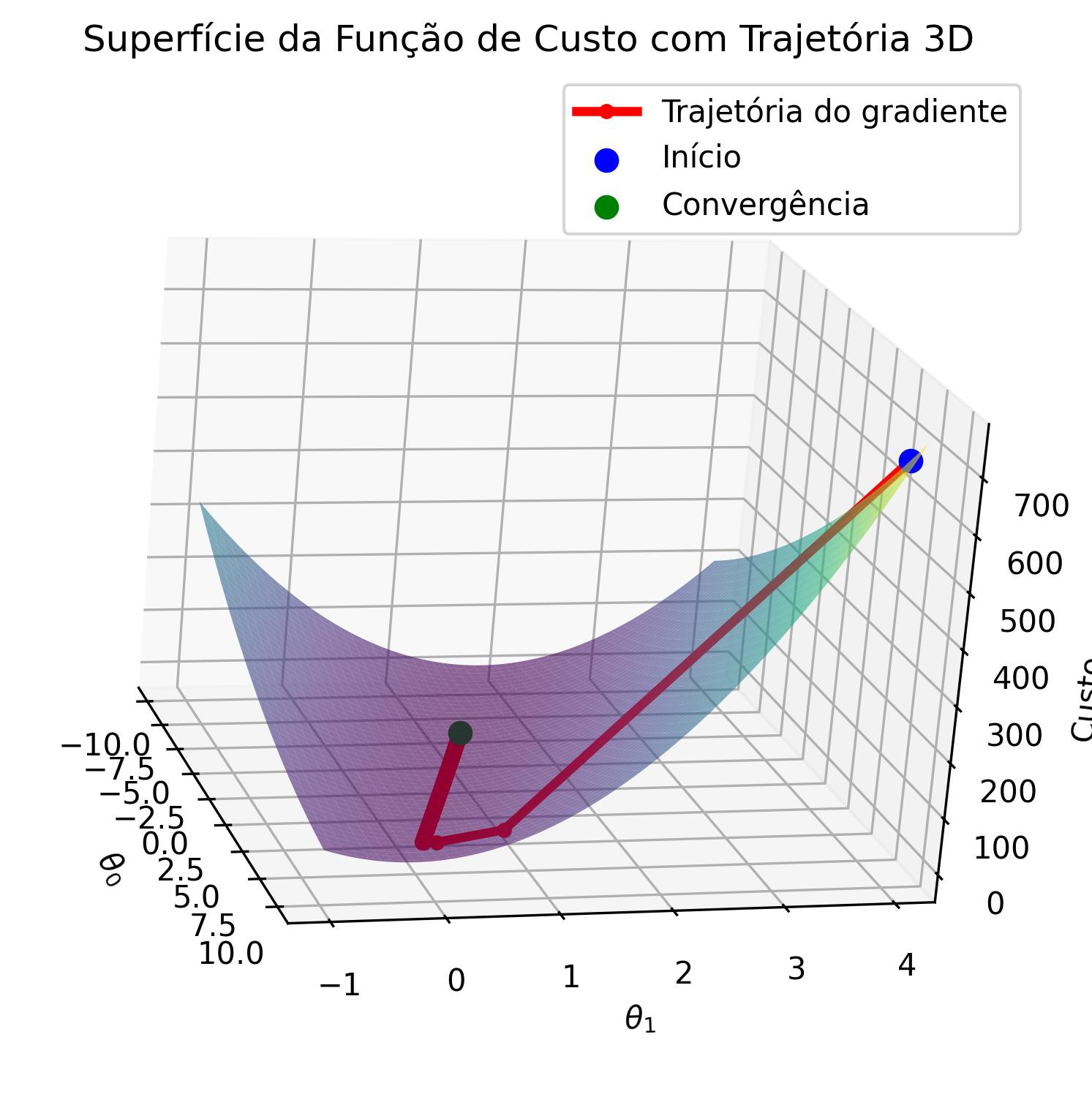


Figura 3 – Superfície da função de custo com trajetória da descida do gradiente.

### 2.4 Influência da Taxa de Aprendizado (α)

A escolha do valor de α é crítica. Quando α é muito grande, o algoritmo pode divergir, saltando sobre o mínimo e causando oscilações. Por outro lado, se α for muito pequeno, o algoritmo converge lentamente, aumentando o custo computacional. O valor ideal de α permite convergência rápida e estável, como demonstrado pelas figuras anteriores.

### 2.5 Inicialização dos Pesos e Fine-Tuning

A inicialização dos pesos (ou parâmetros θ) também influencia diretamente na eficiência da otimização. Uma inicialização inadequada pode levar a mínimos locais ou retardar a convergência. Em redes neurais profundas, é comum utilizar o chamado fine-tuning, que ajusta os pesos pré-treinados com uma pequena taxa de aprendizado, aproveitando uma boa inicialização para alcançar resultados mais precisos. Essa técnica demonstra a importância de começar a otimização a partir de um ponto já vantajoso, tal como ilustrado nas trajetórias observadas nas Figuras 1 e 3.

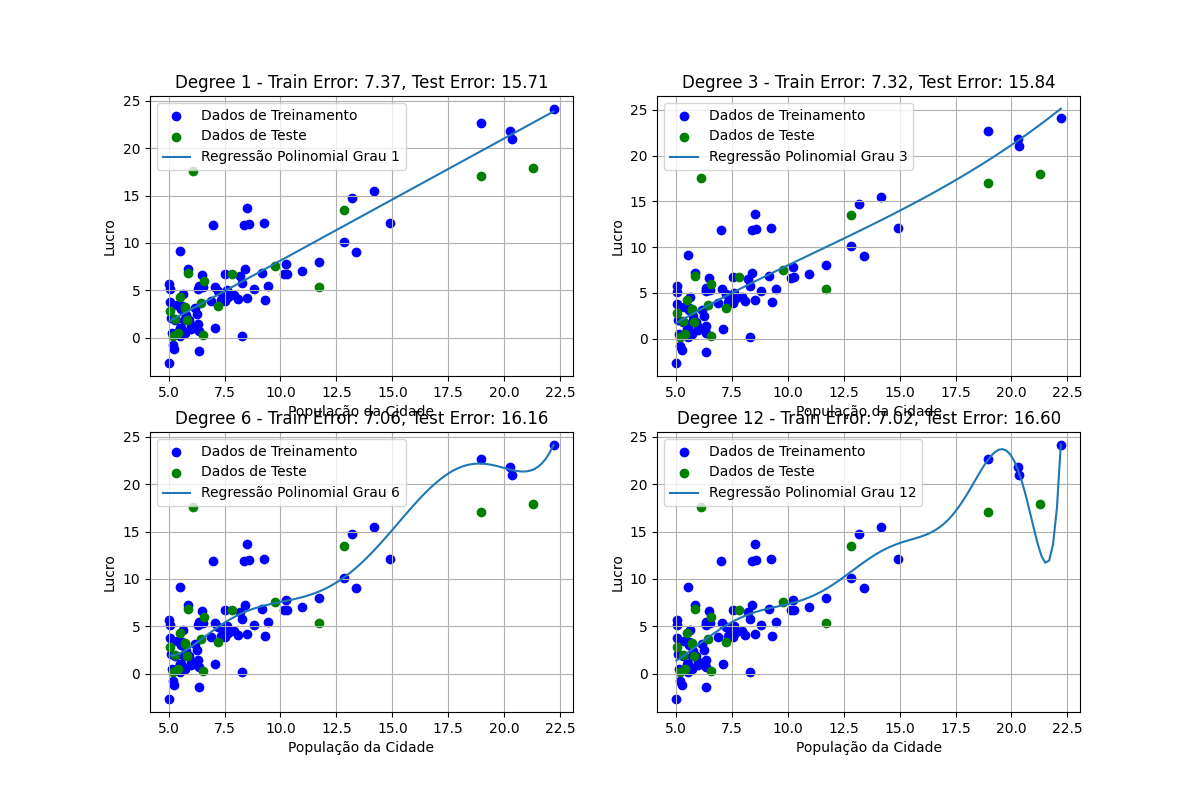
3.**Explorações adicionais**

As explorações adicionais realizadas no presente estudo tiveram como objetivo aprofundar a compreensão do comportamento dos modelos de regressão linear e polinomial, bem como dos algoritmos de otimização, especialmente o gradiente descendente. A seguir, apresentam-se os motivos que justificam cada análise realizada, juntamente com sua finalidade prática.

# 3.1 **Underfitting e Overfitting**

A exploração de modelos com diferentes graus polinomiais teve como finalidade demonstrar os fenômenos de underfitting e overfitting. O underfitting ocorre quando o modelo é simples demais para capturar a complexidade dos dados, resultando em alto erro tanto no conjunto de treinamento quanto no de validação. Já o overfitting acontece quando o modelo se ajusta excessivamente aos dados de treinamento, apresentando baixo erro nesse conjunto, mas alto erro em dados não vistos. Com isso, a análise permite escolher um grau polinomial que proporciona o melhor equilíbrio entre viés e variância, favorecendo a generalização do modelo.

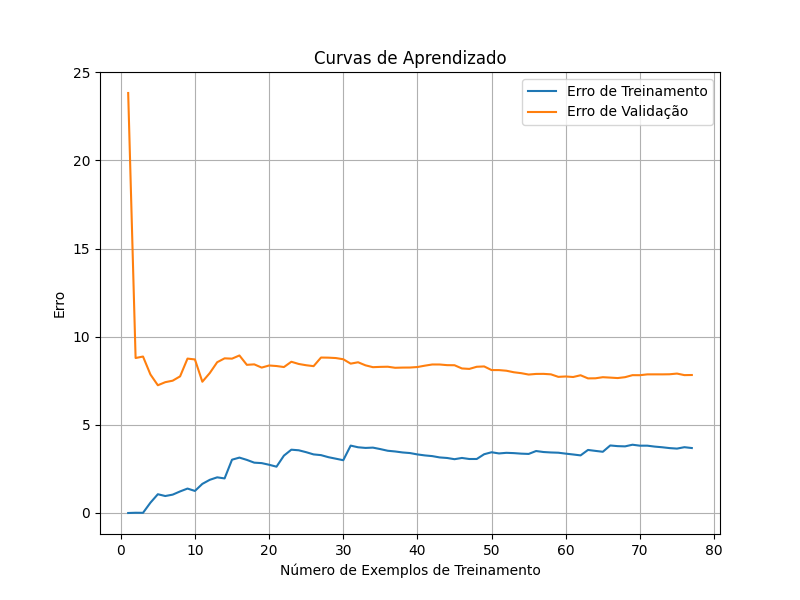
O gráfico abaixo mostra como diferentes graus polinomiais afetam o erro de treino e teste. Modelos com grau muito baixo (e.g., 1) sofrem de underfitting, enquanto modelos com grau muito alto (e.g., 12) sofrem de overfitting.



# 3.2. **Curvas de Aprendizado**

As curvas de aprendizado têm por finalidade avaliar o desempenho do modelo à medida que se aumenta o número de exemplos utilizados no treinamento. Essa análise possibilita a identificação de problemas de alta variância (overfitting) ou alto viés (underfitting), além de indicar se o aumento do conjunto de dados pode ou não melhorar a performance do modelo. Esse tipo de diagnóstico é fundamental para tomar decisões sobre a necessidade de adquirir mais dados ou de alterar a complexidade do modelo.

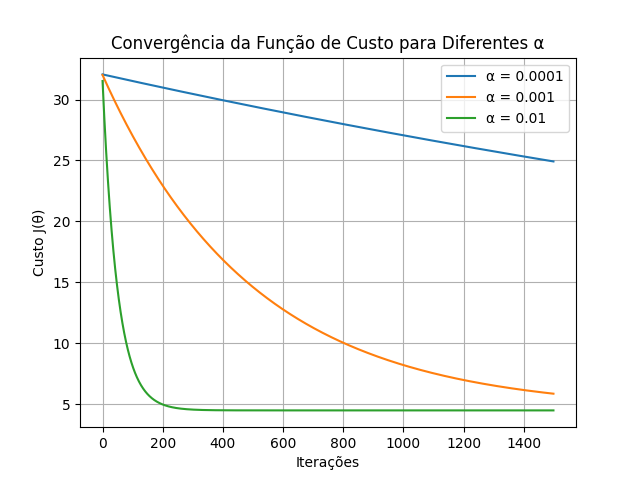
Esse gráfico mostra como o erro de treino e validação evoluem conforme aumenta o número de exemplos de treino. Se ambos os erros forem altos, temos underfitting. Se houver uma grande separação, temos overfitting. Quando ambos são baixos e próximos, o modelo está bem ajustado.



# 3.3. **Convergência com Diferentes Taxas de Aprendizado**

Esta análise foi realizada com o objetivo de investigar como a escolha da taxa de aprendizado afeta a convergência do algoritmo de otimização por gradiente descendente. Taxas de aprendizado muito pequenas podem levar a uma convergência excessivamente lenta, enquanto taxas muito altas podem tornar o processo instável, impossibilitando a convergência. Com essa exploração, é possível selecionar um valor adequado para α, garantindo um bom desempenho na minimização da função de custo.

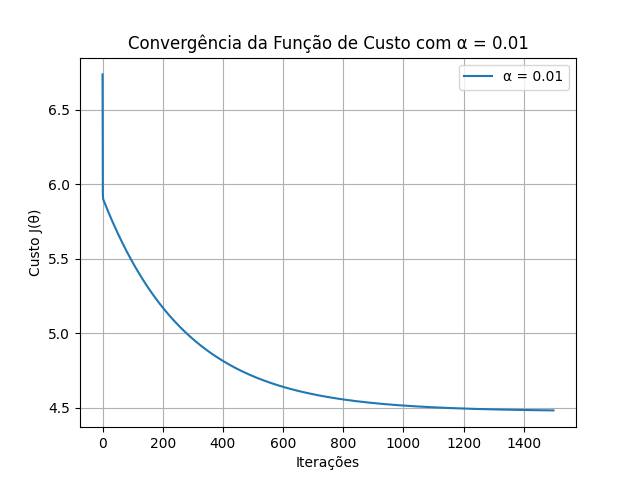
Este gráfico mostra a convergência do algoritmo de Gradiente Descendente com diferentes taxas de aprendizado (α). Taxas muito baixas resultam em convergência lenta, enquanto taxas muito altas podem impedir a convergência.



# 3.4. **Convergência com α = 0.01**

Após a escolha de uma taxa de aprendizado adequada, a evolução do valor da função de custo J(θ) ao longo das iterações foi analisada. Esse procedimento visa verificar se o algoritmo está, de fato, convergindo para um valor mínimo e qual é a velocidade de redução do erro. Isso contribui para avaliar a eficácia da otimização no treinamento do modelo.

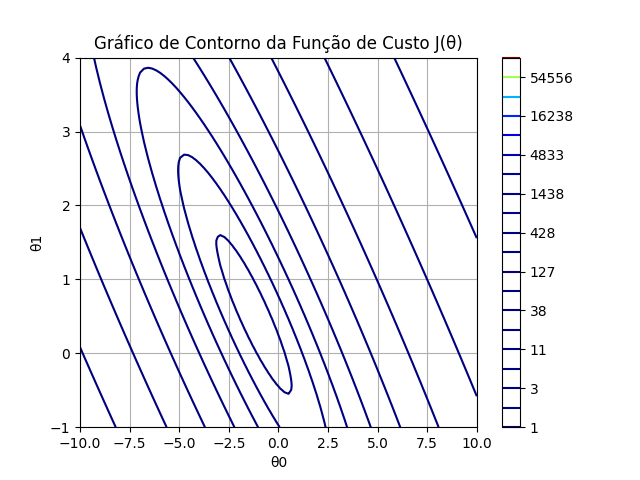
Com uma taxa de aprendizado bem escolhida (α = 0.01), o custo diminui rapidamente, mostrando uma boa convergência.



# 3.5. **Gráfico de Contorno da Função de Custo**

A construção do gráfico de contorno da função de custo em relação aos parâmetros θ₀ e θ₁ teve por objetivo ilustrar visualmente como a função de custo varia no espaço de parâmetros. Essa representação permite identificar o ponto de mínimo global da função de custo e compreender o comportamento das curvas de nível, que são fundamentais para interpretar a direção dos passos dados pelo gradiente descendente. Essa análise é particularmente útil para compreender o processo de otimização em um espaço bidimensional.

O gráfico de contorno mostra o formato da função de custo J(θ) em relação a θ0 e θ1. As linhas representam curvas de nível da função de custo, ajudando a visualizar o caminho do gradiente descendente em direção ao mínimo.



## 4. Conclusão

O presente trabalho teve como objetivo estudar e aplicar os conceitos fundamentais de regressão linear e polinomial, com ênfase na implementação do algoritmo de gradiente descendente para a otimização dos parâmetros do modelo. Por meio de uma abordagem prática, foi possível observar como diferentes configurações influenciam diretamente o desempenho dos modelos e a qualidade das predições realizadas. Inicialmente, foi implementada a regressão linear simples, utilizando o gradiente descendente como método de otimização. A análise da função de custo e sua convergência evidenciou a importância da escolha adequada da taxa de aprendizado, demonstrando que valores muito baixos acarretam em lentidão no treinamento, enquanto valores excessivamente altos podem comprometer a estabilidade do processo de otimização.

Na sequência, explorou-se a regressão polinomial com diferentes graus, o que permitiu observar de forma clara os fenômenos de underfitting e overfitting. Através da análise das curvas de aprendizado, foi possível identificar o grau polinomial que apresenta o melhor equilíbrio entre viés e variância, favorecendo a generalização do modelo para dados não vistos. Adicionalmente, foi realizada a visualização gráfica da função de custo por meio de curvas de contorno, o que proporcionou uma compreensão mais profunda do comportamento da superfície de erro e da direção dos gradientes durante a otimização. Essa análise visual reforçou a importância do gradiente descendente como ferramenta eficaz para encontrar o mínimo da função de custo.

As explorações complementares realizadas, como a evolução da função de custo e o impacto do número de exemplos de treinamento, contribuíram para consolidar o entendimento sobre o processo de aprendizagem dos modelos e suas limitações práticas. Tais análises permitiram ainda desenvolver um olhar crítico sobre o comportamento dos algoritmos de aprendizado de máquina em diferentes cenários.