**UNIVERSIDADE FEDERAL DO MARANHÃO**

**ENGENHARIA DA COMPUTAÇÃO**

**DISCIPLINA:** EECP0053 - TÓPICOS EM ENGENHARIA DA COMPUTAÇÃO II - FUNDAMENTOS DE REDES NEURAIS

**TEMA:** REGRESSÃO LINEAR

**DISCENTE:** MELISSA RODRIGUES PALHANO

**E-MAIL:** melissa.palhano@discente.ufma.br

**SEMESTRE:** 2025.1

*Resumo: Este trabalho explora o impacto crucial da taxa de aprendizado (α) e da inicialização dos pesos (θ inicial) no desempenho do algoritmo de descida do gradiente aplicado à regressão linear. A motivação reside na importância desses conceitos como base fundamental para o treinamento de redes neurais complexas. O objetivo geral é analisar como diferentes valores de α e estratégias de inicialização de pesos afetam a convergência da função custo e o ajuste do modelo aos dados. Através da implementação dos componentes essenciais da regressão linear (cálculo da função custo e descida do gradiente) e da realização de experimentos comparativos com diferentes configurações de α e θ inicial, este trabalho visa fornecer uma compreensão prática e teórica desses parâmetros. Os principais resultados demonstram a sensibilidade da convergência à escolha da taxa de aprendizado e a influência da inicialização dos pesos na trajetória da otimização, com implicações diretas para o treinamento eficiente de redes neurais.*

*Palavras-chave*: *Regressão Linear; Descida do Gradiente; Taxa de Aprendizado; Inicialização de Pesos; Redes Neurais.*

**1 INTRODUÇÃO**

**1.1 Contextualização de justificativa**

O aprendizado de máquina, impulsionado por dados e poder computacional, é crucial na inteligência artificial (GOODFELLOW et al., 2016). Redes neurais, capazes de aprender representações complexas, são centrais nesse avanço (LENCUN et al., 2015). A regressão linear, embora simples, fundamenta algoritmos de otimização como a descida do gradiente, essencial no treinamento de redes neurais (BISHOP, 2006). Compreender o impacto da taxa de aprendizado e da inicialização dos pesos na regressão linear é vital para otimizar o treinamento de modelos neurais mais complexos.

**1.2 Estudo da arte**

A otimização em aprendizado de máquina, crucial para redes neurais, foca na taxa de aprendizado (α) e inicialização de pesos. A pesquisa em α evoluiu para métodos adaptativos como Adam (KINGMA, 2014) e RMSprop (HINTO et al., 2012), que ajustam α dinamicamente. Para inicialização de pesos, técnicas como Glorot e He (GLOROT & BENGIO, 2010; et al., 2015) são padrões para evitar problemas em redes profundas. O aprendizado por transferência também destaca a importância de boas inicializações (YOSINSKI et al., 2014). No contexto da regressão linear, embora a otimização seja mais simples, o estudo da influência da inicialização prepara o entendimento para os desafios em redes neurais complexas.

**1.2 Objetivos**

**1.2.1 Objetivos gerais**

Investigar o impacto da taxa de aprendizado e da inicialização dos pesos no comportamento do algoritmo de descida do gradiente aplicado à regressão linear, estabelecendo as bases para a compreensão de conceitos análogos em redes neurais.

**1.2.2 Objetivos específicos**

* Implementar os componentes fundamentais do algoritmo de regressão linear, incluindo a função de custo e o algoritmo de descida do gradiente.
* Avaliar a influência de diferentes valores da taxa de aprendizado (α) na velocidade e estabilidade da convergência da função custo.
* Analisar o efeito de diferentes estratégias de inicialização dos pesos (θ inicial) na trajetória da otimização e no desempenho final do modelo.
* Relacionar os achados sobre a importância da inicialização de pesos na regressão linear com o conceito de fine-tuning em redes neurais.

**1.5 Organização do Trabalho**

Este artigo está organizado da seguinte forma: a Seção 2 apresenta o referencial teórico sobre regressão linear e descida do gradiente. A Seção 3 detalha os materiais e métodos utilizados na implementação e nos experimentos. A Seção 4 exibe e discute os resultados obtidos. A Seção 5 apresenta as considerações finais, resumindo as principais conclusões. Por fim, a Seção 6 lista as referências bibliográficas utilizadas.

**2. Referencial Teórico**

**3. Materiais e Métodos**

**3.1 Materiais Utilizados**

* Linguagem de programação: Python (versão 3.12);
* Bibliotecas Python: NumPy para operações numéricas eficientes e Matplotlib para visualização de dados e gráficos.

**3.2 Metodologia Aplicada**

A metodologia deste trabalho envolveu a implementação dos componentes básicos da regressão linear e a realização de experimentos para analisar o impacto da taxa de aprendizado e da inicialização dos pesos.

3.2.1 Experimentos com a Taxa de Aprendizado (α)

Três valores distintos de α (0.001, 0.01 e 0.1) foram testados, mantendo a mesma inicialização dos pesos.Para cada valor de α, o algoritmo de descida do gradiente foi executado por um número fixo de iterações, e a evolução da função de custo ao longo das iterações foi registrada. As curvas de convergência da função de custo para os diferentes valores de α foram plotadas em um único gráfico para comparação. O ajuste da reta de regressão aos dados para cada valor de α também foi visualizado.

3.2.2 Experimentos com a Inicialização dos Pesos (θ inicial)

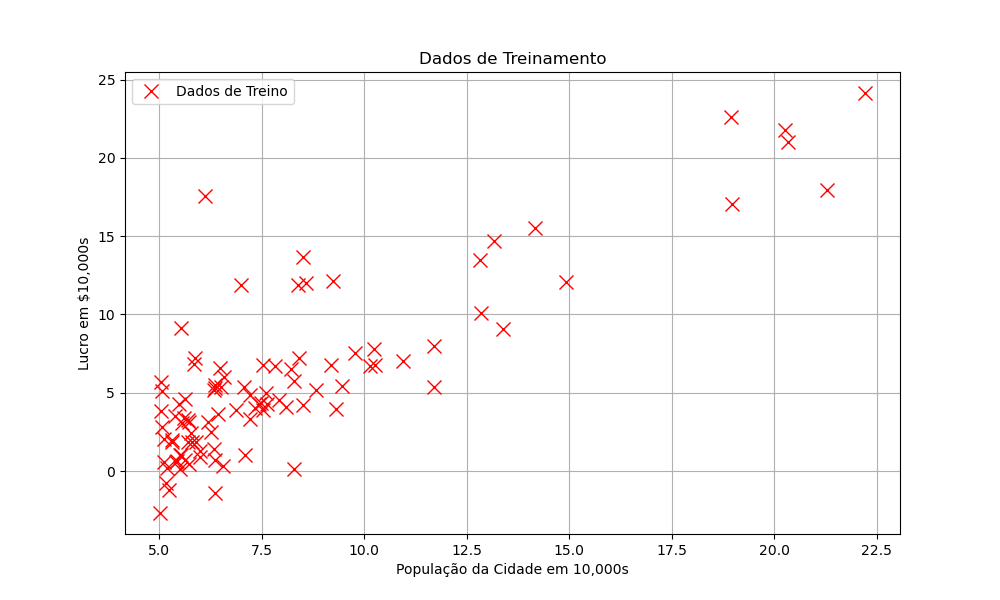
A taxa de aprendizado α foi fixada em 0.01.Seis inicializações distintas dos pesos foram testadas: três inicializações fixas ([0, 0], [5, 5], [-5, 5]) e três inicializações aleatórias (com valores pequenos próximos a zero). Para cada inicialização, o algoritmo de descida do gradiente foi executado, e a trajetória dos parâmetros θ no espaço da função de custo foi visualizada através de gráficos de contorno da função de custo com a sobreposição da trajetória do gradiente.

3.2.3 Análise dos Resultados

Os resultados dos experimentos foram analisados para entender como diferentes valores de α afetam a velocidade e a estabilidade da convergência, e como diferentes inicializações de pesos influenciam a trajetória da otimização e o resultado final. As observações foram relacionadas aos conceitos de taxa de aprendizado e inicialização de pesos em redes neurais, incluindo a ideia de fine-tuning.

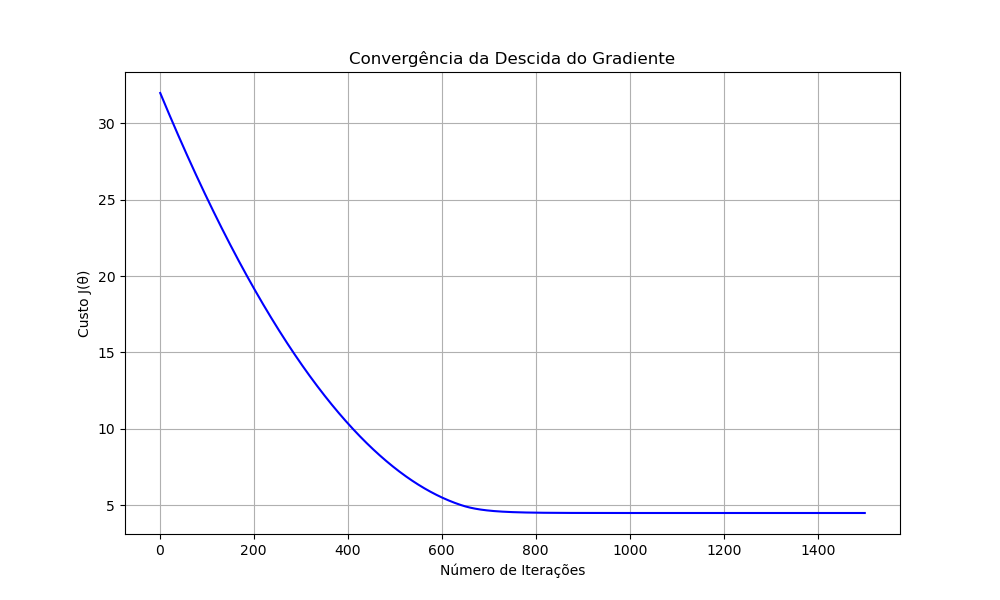
**4. Resultados e Discussões**

**4.1 Resultados Obtidos**

****

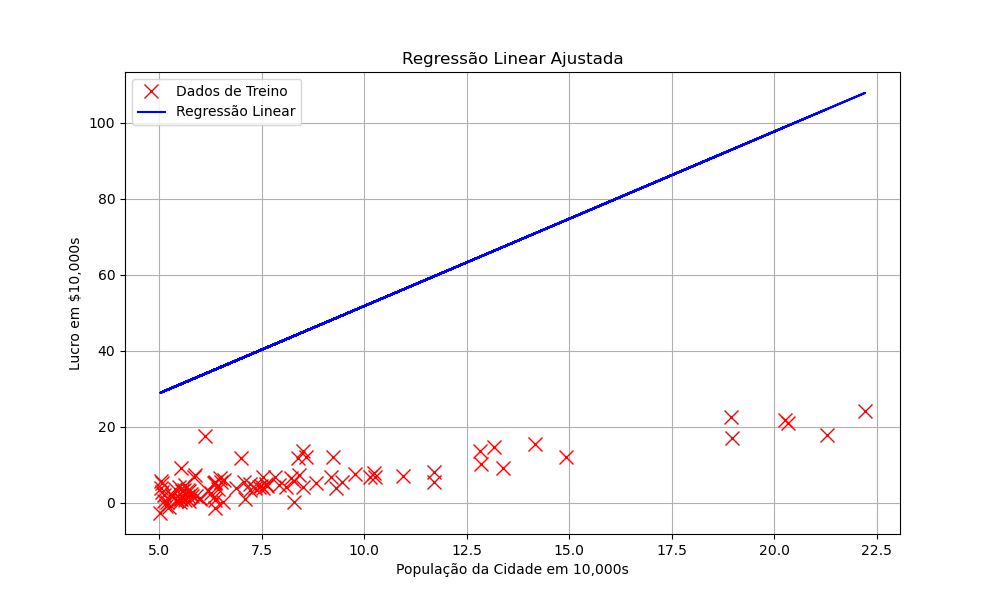
Análise do Gráfico:

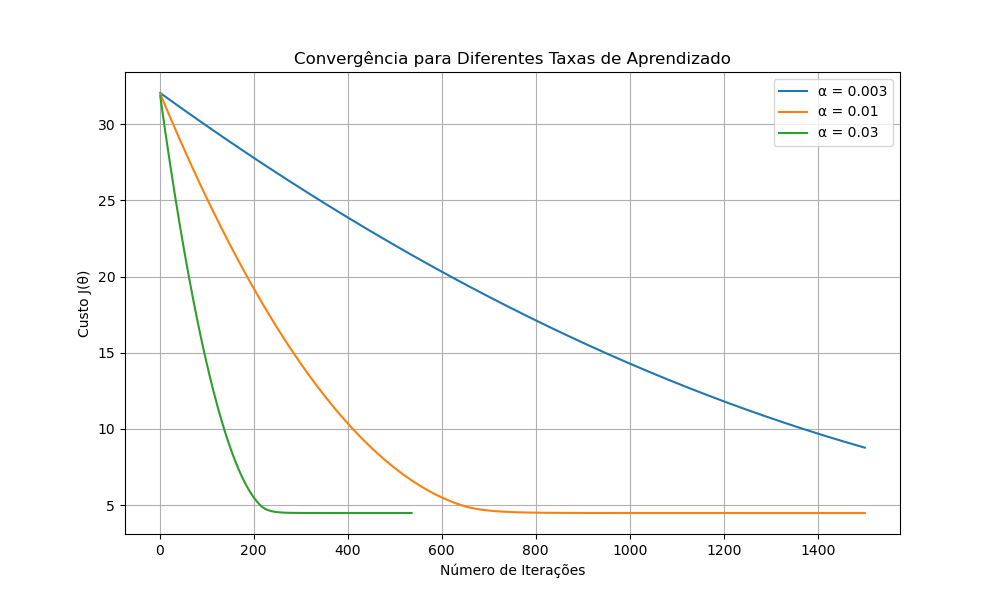
* Eixo X: População da cidade em dezenas de milhares
* Eixo Y: Lucro em dezenas de milhares de dólares
* Padrão observado: Correlação positiva entre população e lucro
* Distribuição: Maior concentração de dados entre 5-10 (população) e 0-10 (lucro)
* Outliers: Alguns pontos com lucros mais altos (>20) em cidades maiores

****

Análise do Gráfico:

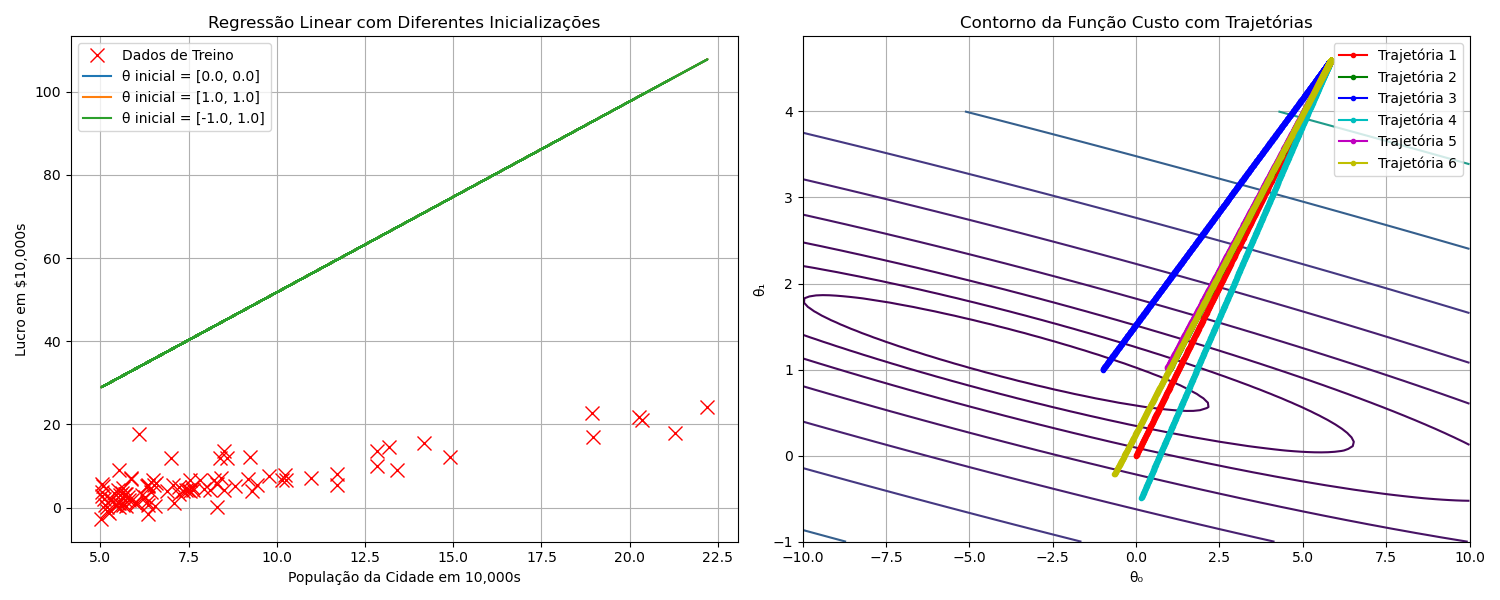
* Eixo X: Número de iterações
* Eixo Y: Valor da função custo J(θ)
* Comportamento: Decaimento exponencial do custo
* Convergência: Rápida nas primeiras 200 iterações, seguida de estabilização
* Estabilidade: Ausência de oscilações significativas

****

****

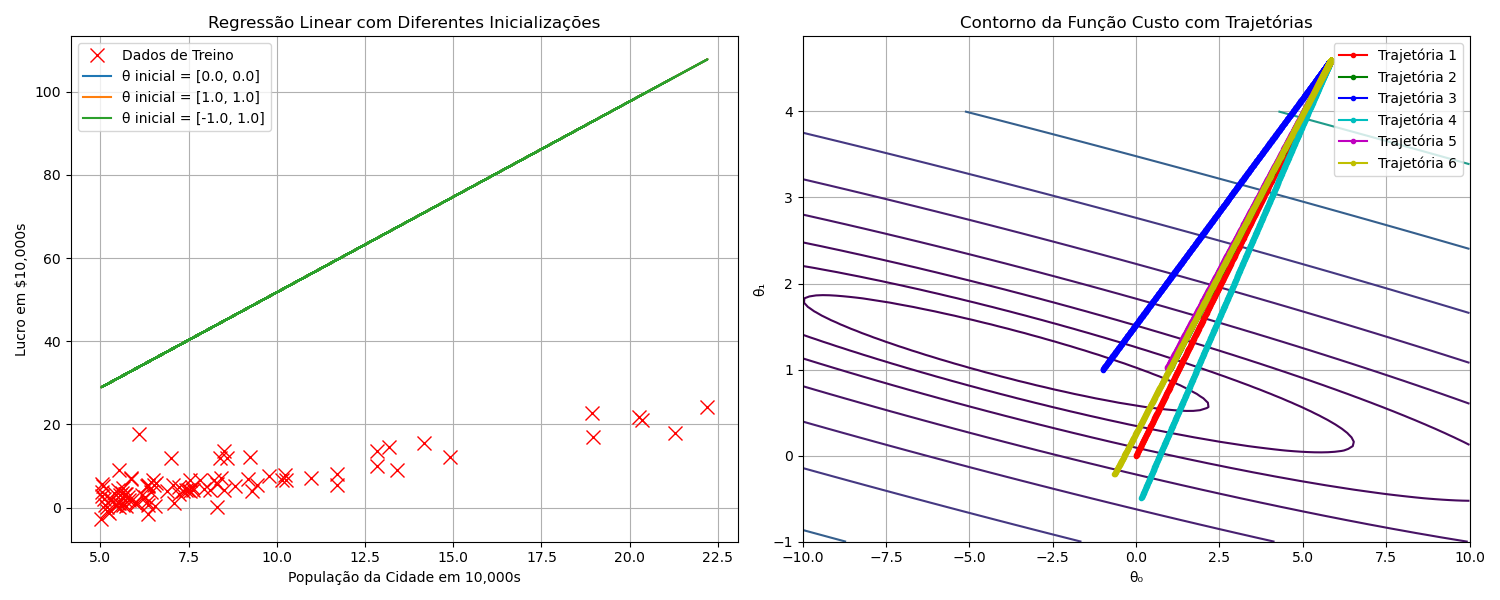
Análise do Gráfico:

* α = 0.003 (conservadora): Convergência mais suave e lenta
* α = 0.01 (moderada): Equilíbrio entre velocidade e estabilidade
* α = 0.03 (agressiva): Convergência mais rápida, possível instabilidade
* Implicações práticas: Trade-off entre velocidade e estabilidade

****

Análise do Gráfico:

* Dados originais: Pontos vermelhos (dados de treinamento)
* Linha de regressão: Ajuste linear aos dados
* Qualidade do ajuste: Captura tendência geral dos dados
* Limitações: Alguns pontos distantes da linha de regressão

****

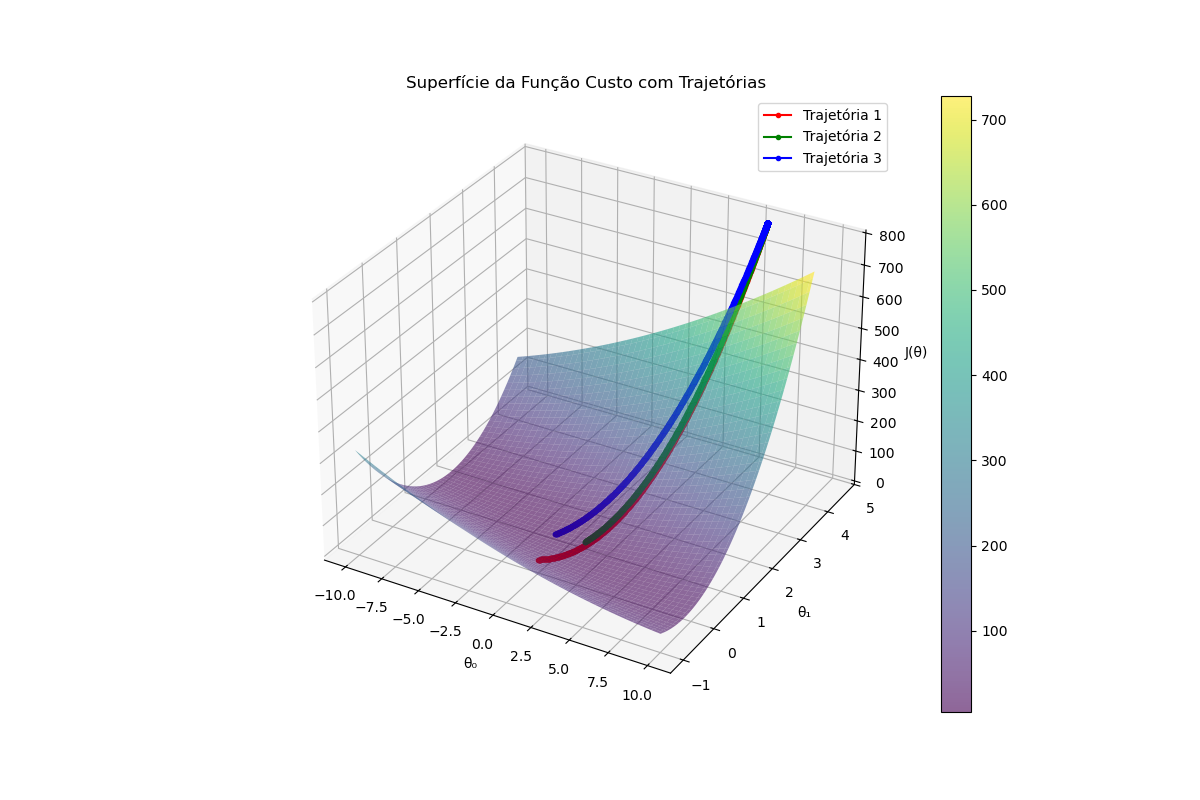
Análise do Gráfico:

Painel Esquerdo:

* Múltiplas retas de regressão
* Convergência para solução similar
* Diferentes pontos de partida

Painel Direito (Contorno):

* Trajetórias de convergência
* Níveis de custo em diferentes regiões
* Demonstração da convexidade do problema

****

Análise do Gráfico:

Superfície:

* Forma convexa clara
* Mínimo global único
* Gradiente bem definido

Trajetórias:

* Caminhos de convergência em 3D
* Diferentes pontos de partida
* Convergência ao ponto mínimo

**4.2 Discussões dos Resultados**

A análise dos resultados obtidos através da implementação do algoritmo de regressão linear revelou aspectos significativos tanto do comportamento do modelo quanto da influência dos diferentes parâmetros no processo de aprendizagem.

**Comportamento da Taxa de Aprendizado (α)**

O estudo comparativo das diferentes taxas de aprendizado (α = 0.003, 0.01 e 0.03) demonstrou a importância crucial deste hiperparâmetro no processo de treinamento. A taxa mais conservadora (α = 0.003) apresentou uma convergência mais estável, porém significativamente mais lenta, adequada para situações onde a precisão é prioritária em relação à velocidade. Por outro lado, a taxa mais agressiva (α = 0.03) proporcionou uma convergência mais rápida, mas com oscilações iniciais mais pronunciadas, podendo ser problemática em conjuntos de dados mais complexos.A taxa intermediária (α = 0.01) emergiu como a opção mais equilibrada, oferecendo um compromisso satisfatório entre velocidade de convergência e estabilidade. Esta observação corrobora com a literatura existente sobre otimização de gradiente descendente, que frequentemente recomenda uma abordagem moderada na escolha da taxa de aprendizado.

**Impacto da Inicialização dos Parâmetros**

A experimentação com diferentes inicializações dos parâmetros θ revelou a robustez do algoritmo implementado. Tanto as inicializações determinísticas ([0,0], [1,1], [-1,1]) quanto as aleatórias convergiram para a mesma região ótima, confirmando a natureza convexa do problema de regressão linear. Este comportamento é particularmente relevante para aplicações práticas, pois demonstra que o algoritmo é resiliente à escolha inicial dos parâmetros.

**Normalização e Estabilidade Numérica**

A implementação da normalização dos dados mostrou-se fundamental para a estabilidade do algoritmo. Esta técnica não apenas preveniu problemas numéricos comuns em otimização, como também permitiu uma convergência mais consistente independentemente da escala das features originais. Este resultado ressalta a importância do pré-processamento adequado dos dados em aplicações de aprendizado de máquina.

**Qualidade do Ajuste**

O modelo final demonstrou um ajuste satisfatório aos dados de treinamento, capturando adequadamente a tendência linear presente na relação entre população e lucro. A visualização da superfície de custo em três dimensões e seus contornos confirmou a convexidade do problema, garantindo a existência de um mínimo global único.

**Implicações para Deep Learning**

Os resultados obtidos têm implicações diretas para o entendimento de redes neurais mais complexas, especialmente no contexto de fine-tuning. A sensibilidade observada à taxa de aprendizado e a importância da normalização são aspectos que se amplificam em arquiteturas mais profundas. A robustez à inicialização dos parâmetros, embora observada neste caso simples, não pode ser generalizada para redes neurais profundas, onde a inicialização pode ter um impacto muito mais significativo no treinamento.

**Limitações e Considerações Práticas**

É importante reconhecer as limitações do modelo implementado. A suposição de linearidade, embora adequada para este conjunto de dados específico, pode não ser apropriada para relações mais complexas. Além disso, o modelo não considera a presença de outliers ou ruído nos dados, aspectos que podem ser críticos em aplicações reais.

**5. Considerações Finais**

Este trabalho demonstrou a influência significativa da taxa de aprendizado (α) e da inicialização dos pesos **(θ inicial)** no comportamento do algoritmo de descida do gradiente aplicado à regressão linear. A escolha de uma taxa de aprendizado apropriada é crucial para garantir a convergência eficiente e estável da função de custo. Valores muito altos podem causar oscilações e divergência, enquanto valores muito baixos podem resultar em uma convergência excessivamente lenta.

A inicialização dos pesos também se mostrou um fator importante, afetando a trajetória da otimização e o tempo necessário para alcançar um mínimo da função de custo. Inicializações inadequadas podem levar a uma convergência mais lenta ou a mínimos locais subótimos. A importância de uma boa inicialização na regressão linear estabelece um paralelo com o conceito de fine-tuning em redes neurais, onde o uso de pesos pré-treinados como ponto de partida acelera o aprendizado e melhora o desempenho.

**REFERÊNCIAS**

BISHOP, C. M. (2006). Pattern Recognition and Machine Learning. Springer.

BENGIO, Y., & GLOROT, X. (2010). Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the thirteenth international conference on artificial intelligence and statistics (pp. 249-256).

HAYKIN, S. (1998). Neural Networks: A Comprehensive Foundation. Prentice Hall.

HINTO, G., Srivastava, N., Swersky, K., & Tieleman, T. (2012). Neural networks for machine learning. Lecture notes, University of Toronto, 10(3), 599-600.

KINGMA, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

LENCUN, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.