UNIVERSIDADE FEDERAL DO MARANHÃO

CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIAS - CCET

COORDENAÇÃO DO CURSO DE ENGENHARIA DA COMPUTAÇÃO

IMPLEMENTAÇÃO E ANÁLISE ALGORÍTMO DE REGRESSÃO LINEAR MULTIVARIADA

Prof. Dr. THALES LEVI AZEVEDO VALENTE

Discente: CAIO REIS BATISTA

São Luís-MA

2025.1

1. INTRODUÇÃO

A regressão linear multivariada é uma técnica fundamental em aprendizado de máquina e estatística para modelar a relação entre múltiplas variáveis independentes (features) e uma variável dependente contínua. Este trabalho analisa a implementação e comparação de dois métodos de otimização para regressão linear multivariada: Gradiente Descendente (GD) e Equação Normal (NE), além de avaliar o impacto de diferentes técnicas de normalização de features.

O conjunto de dados utilizado contém informações sobre casas, onde as features são o tamanho em pés quadrados e o número de quartos, e a variável alvo é o preço da casa. Este problema exemplifica perfeitamente a necessidade de regressão multivariada, pois o preço de uma casa é influenciado por múltiplos fatores.

1. METODOLOGIA
   1. Métodos de Otimização

2.1.1Gradiente Descente

O Gradiente Descendente é um algoritmo iterativo que busca minimizar a função de custo J(θ) ajustando os parâmetros θ na direção oposta ao gradiente da função de custo. A implementação utiliza a seguinte atualização:

θ := θ - α \* (1/m) \* (Xᵀ \* (Xθ - y))

Onde:

- θ é o vetor de parâmetros

- α é a taxa de aprendizado

- m é o número de amostras

- X é a matriz de features (com termo de bias)

- y é o vetor de valores alvo

O algoritmo executa um número fixo de iterações, atualizando os parâmetros e calculando o custo em cada iteração.

2.1.2 Equação Normal

A Equação Normal oferece uma solução fechada para o problema de regressão linear, calculando diretamente os parâmetros ótimos sem iterações:

θ = (XᵀX)⁻¹ Xᵀ y

A implementação utiliza a pseudo-inversa (np.linalg.pinv) para lidar com casos onde a matriz XᵀX pode não ser invertível ou mal condicionada.

* 1. Técnicas de Normalização

2.2.1 Normalização Z-Score

Esta técnica normaliza as features para terem média zero e desvio padrão unitário:

X\_norm = (X - μ) / σ

Onde:

- μ é o vetor de médias de cada feature

- σ é o vetor de desvios padrão de cada feature

2.2.2 Normalização Min-Max (features\_normalizes\_by\_min\_max)

Esta técnica normaliza as features para o intervalo [0, 1]:

X\_norm = (X - min) / (max - min)

Onde:

- min é o vetor de valores mínimos de cada feature

- max é o vetor de valores máximos de cada feature

2.3 Métricas de Avaliação

Para comparar os métodos, utilizamos:

- Função de custo J(θ) = (1/2m) \* Σ(h(x) - y)²

- Convergência do custo ao longo das iterações

- Tempo de execução

- Precisão das predições

1. RESULTADOS E DISCUSSÕES
   1. Convergência do Gradiente Descendente

Gráfico, Gráfico de linhas

O conteúdo gerado por IA pode estar incorreto. Figura 1: Convergência do Custo

O gráfico de convergência do custo mostra como J(θ) diminui ao longo das iterações para o método GD. Observamos que:

- A curva apresenta uma queda acentuada nas primeiras iterações, indicando uma rápida redução inicial do custo.

- A convergência se estabiliza após aproximadamente 100 iterações, sugerindo que o algoritmo está se aproximando do mínimo global.

- A normalização z-score aplicada aos dados contribuiu significativamente para esta convergência suave e eficiente.

* 1. Comparação de Custo entre GD e NE

Gráfico, Gráfico de linhas

O conteúdo gerado por IA pode estar incorreto.

Figura 2: Comparação GD vs NE

Este gráfico compara diretamente a convergência do GD com o custo obtido pela Equação Normal:

- A linha azul representa a convergência do GD ao longo das iterações.

- A linha vermelha tracejada representa o custo obtido pela Equação Normal (correto).

- A linha preta pontilhada representa o custo incorreto que seria obtido ao aplicar os parâmetros da NE em dados normalizados.

Observamos que o GD converge para um valor muito próximo ao obtido pela Equação Normal, confirmando que ambos os métodos encontram soluções similares quando implementados corretamente. A diferença residual pode ser atribuída à natureza iterativa do GD, que se aproxima assintoticamente do mínimo global.

* 1. Plano de Regrassão 3D

Gráfico, Gráfico de superfície

O conteúdo gerado por IA pode estar incorreto.

Figura 3: Ajuste da Regressão Multivariada

Este gráfico 3D mostra o plano de regressão ajustado aos dados:

- Os pontos vermelhos representam os dados de treinamento (tamanho, quartos, preço).

- A superfície colorida representa o plano de regressão ajustado pelo modelo.

- Observamos que o plano captura a tendência geral dos dados, onde casas maiores e com mais quartos tendem a ter preços mais elevados.

- A inclinação do plano nas direções do tamanho e número de quartos corresponde aos coeficientes θ₁ e θ₂ do modelo, respectivamente.

O plano de regressão demonstra visualmente como o modelo linear multivariado aproxima a relação entre as features e o preço das casas. A qualidade do ajuste pode ser avaliada pela proximidade dos pontos de dados ao plano.

* 1. Superfício e Contorno de J(**θ**)

Gráfico, Gráfico de superfície

O conteúdo gerado por IA pode estar incorreto.

Figura 4: Superfície de Custo

Uma imagem contendo Gráfico

O conteúdo gerado por IA pode estar incorreto.

Figura 5: Contorno de Custo

Estes gráficos mostram a superfície de custo J(θ₁, θ₂) e seu contorno, mantendo θ₀ fixo:

- A superfície 3D mostra como o custo varia em função dos parâmetros θ₁ e θ₂.

- A trajetória vermelha representa o caminho percorrido pelo GD durante as iterações.

- O ponto preto marca a solução encontrada pela Equação Normal.

- O gráfico de contorno (2D) mostra as curvas de nível da superfície de custo, facilitando a visualização do caminho de convergência.

Observações importantes:

- A trajetória do GD converge em direção ao mínimo global, seguindo o gradiente negativo.

- A solução da Equação Normal coincide com o mínimo global da superfície de custo.

- A forma da superfície de custo é aproximadamente convexa, o que garante a convergência para um mínimo global único.

- A normalização das features contribui para a forma mais simétrica da superfície, facilitando a convergência do GD.

4. DISCUSSÃO

4.1 Efeito da Normalização de Features

4.1.1 Impacto no Gradiente Descendente

A normalização de features tem um impacto significativo no desempenho do GD. Sem normalização, as features com escalas diferentes causam uma superfície de custo alongada, tornando o GD ineficiente e propenso a oscilações. Isso ocorre porque:

1. Sem normalização: Features com valores maiores dominam o gradiente, causando passos desproporcionais nas diferentes direções.

2. Com normalização z-score: A superfície de custo torna-se mais simétrica, permitindo que o GD avance com passos mais equilibrados em todas as direções, acelerando a convergência.

3. Com normalização min-max: Embora também melhore a convergência, pode ser menos robusta a outliers comparada à z-score.

4.1.2 Impacto na Equação Normal

A Equação Normal é matematicamente invariante à escala das features, ou seja, teoricamente produz os mesmos resultados independentemente da normalização. No entanto, na prática:

1. Estabilidade numérica: Com features de escalas muito diferentes, a matriz XᵀX pode se tornar mal condicionada, afetando a precisão da inversão.

2. Precisão computacional: A normalização pode melhorar a estabilidade numérica da operação de pseudo-inversa, especialmente em sistemas com precisão limitada.

4.2 Comparação de Desempenho entre GD e NE

4.2.1 Tempo de Execução

- Gradiente Descendente: O tempo de execução escala com O(kn²), onde k é o número de iterações e n é o número de features. Para conjuntos de dados pequenos ou médios, pode ser mais lento que a NE.

- Equação Normal: O tempo de execução escala com O(n³) devido à inversão da matriz. Para conjuntos de dados com muitas features, pode se tornar computacionalmente proibitivo.

4.2.2 Precisão

- Gradiente Descendente: A precisão depende do número de iterações e da taxa de aprendizado. Com parâmetros adequados e normalização, pode se aproximar muito do ótimo global.

- Equação Normal: Fornece diretamente o ótimo global em uma única operação, sem necessidade de ajuste de hiperparâmetros.

4.2.3 Robustez

- Gradiente Descendente: Mais flexível e adaptável a diferentes tipos de problemas. Pode ser estendido para modelos não-lineares e conjuntos de dados muito grandes.

- Equação Normal: Limitado a problemas lineares e pode enfrentar problemas de estabilidade numérica com features colineares ou mal condicionadas.

5. CONCLUSÕES

A análise comparativa entre Gradiente Descendente e Equação Normal para regressão linear multivariada, considerando diferentes técnicas de normalização, permite concluir que:

1. Normalização de features: É crucial para o desempenho eficiente do Gradiente Descendente, mas menos impactante para a Equação Normal. A normalização z-score mostrou-se mais robusta e eficiente para o conjunto de dados analisado.

2. Escolha do método de otimização: Depende das características do problema:

- Para conjuntos de dados pequenos a médios (n < 10.000), a Equação Normal é geralmente mais rápida e precisa.

- Para conjuntos de dados grandes ou com muitas features, o Gradiente Descendente é mais escalável e eficiente.

3. Interpretabilidade do modelo: O plano de regressão 3D obtido demonstra como o modelo captura a relação entre as features e o preço das casas, permitindo predições baseadas em novos dados.

4. Convergência e estabilidade: A normalização adequada garante uma convergência mais rápida e estável do Gradiente Descendente, aproximando seu desempenho ao da Equação Normal em termos de custo final.

Este estudo reforça a importância da preparação adequada dos dados e da escolha criteriosa do método de otimização em problemas de regressão linear multivariada, considerando o tamanho do conjunto de dados, o número de features e os requisitos computacionais.

6. REFERÊNCIAS

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. \*\*The Elements of Statistical Learning\*\*. 2. ed. New York: Springer, 2009.

NG, A. \*\*Machine Learning\*\*. Stanford University, 2018. Disponível em: <https://www.coursera.org/learn/machine-learning>.

BISHOP, C. M. \*\*Pattern Recognition and Machine Learning\*\*. New York: Springer, 2006.

JAMES, G. et al. \*\*An Introduction to Statistical Learning\*\*. New York: Springer, 2013.