Regressão Linear e Gradiente Descendente

Fernando Filho
Cientista de dados

1 Introdução

A regressão linear é uma das técnicas mais fundamentais e amplamente utilizadas em estatística e aprendizado de máquina. Seu objetivo é modelar a relação entre uma variável dependente y e uma ou mais variáveis independentes x, assumindo que essa relação possa ser representada por uma equação linear. Apesar de sua simplicidade, a regressão linear é extremamente poderosa, pois permite não apenas prever valores futuros, mas também compreender o comportamento e a influência das variáveis envolvidas em um processo.

A importância da regressão linear vai além do ajuste de dados. Ela fornece uma base conceitual para diversos algoritmos mais avançados, como regressão logística, redes neurais e modelos lineares generalizados. Além disso, em muitos contextos científicos e de engenharia, a interpretação dos coeficientes da regressão permite inferências úteis e insights práticos, como identificar relações causais ou medir o impacto de um fator sobre outro.

Para ilustrar a aplicação prática da regressão linear, utilizamos dados reais provenientes de um estudo biomecânico de pacientes ortopédicos. Esses dados fazem parte de um conjunto disponível publicamente e frequentemente usado em estudos de classificação e regressão. O dataset contém medidas angulares da coluna vertebral e da pelve de pacientes, as quais são importantes para o diagnóstico de problemas posturais, como espondilolistese e hiperlordose lombar.

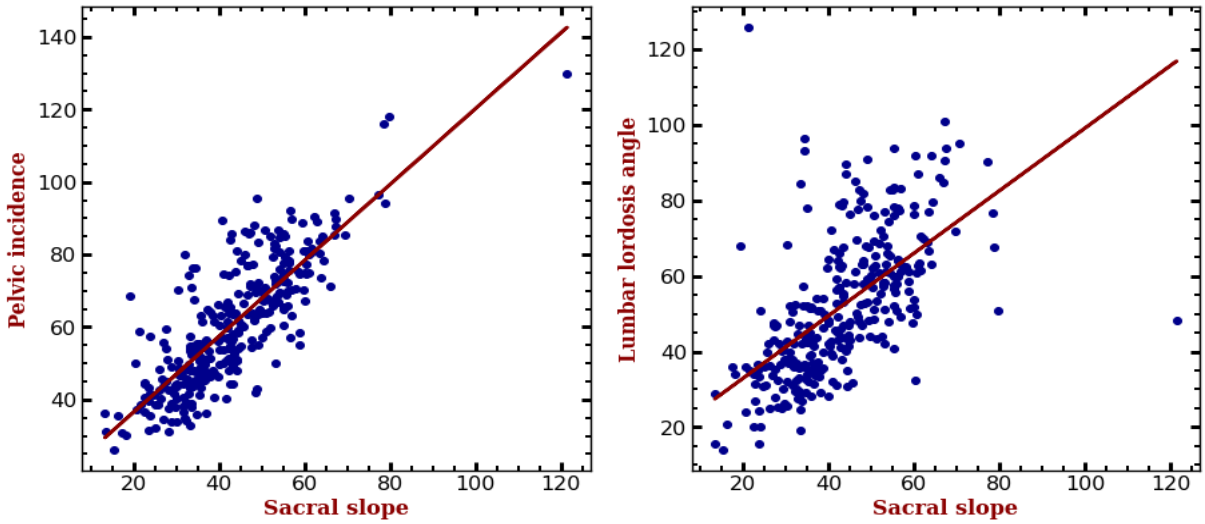

Figure 1: Incidência pélvica x Inclinação sacral e Ângulo de lordose lombar x Inclinação sacral

Na figura acima, apresentamos dois gráficos que exemplificam bem a utilidade da regressão linear na análise de dados biomédicos:

- O primeiro gráfico mostra a relação entre a **incidência pélvica** (*Pelvic incidence*) e a **inclinação sacral** (*Sacral slope*);
- O segundo gráfico mostra a relação entre o **ângulo de lordose lombar** (*Lumbar lordosis angle*) e a **inclinação sacral**.

Ambas as análises revelam padrões lineares evidentes, onde o aumento da inclinação sacral está associado a aumentos proporcionais nas outras variáveis. As linhas vermelhas nos gráficos representam os modelos ajustados via regressão linear, evidenciando o bom alinhamento com os dados.

Essas relações não são apenas matematicamente interessantes: elas têm implicações clínicas diretas. Por exemplo, compreender como a inclinação do sacro afeta o ângulo de lordose lombar pode auxiliar médicos a detectar desalinhamentos posturais precocemente, contribuindo para diagnósticos mais precisos e tratamentos personalizados.

Portanto, a regressão linear se mostra como uma ferramenta essencial não apenas em ambientes acadêmicos, mas também em aplicações práticas na medicina, engenharia, economia e ciências sociais, sendo frequentemente o primeiro passo para uma análise mais robusta e fundamentada de dados.

2 Regressão Linear Simples

Na regressão linear simples, assumimos que a relação entre as variáveis segue a equação:

$$\hat{y} = \theta_0 + \theta_1 x \tag{1}$$

onde:

- $\hat{y}$ é o valor predito,
- x é a variável independente,
- θ_0 é o intercepto,
- θ_1 é o coeficiente angular (inclinação).

O objetivo é encontrar os parâmetros θ_0 e θ_1 que minimizam o erro entre os valores previstos $\hat{y}$ e os valores reais y.

3 Função de Custo

A função de custo usada na regressão linear é o erro quadrático médio (MSE):

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)}\right)^2 \tag{2}$$

onde:

- m é o número de amostras,
- $y^{(i)}$ é o valor real,
- $\hat{y}^{(i)} = \theta_0 + \theta_1 x^{(i)}$ é o valor previsto.

4 Gradiente Descendente

Para minimizar a função de custo, utilizamos o algoritmo do **gradiente descendente**, que atualiza os parâmetros de forma iterativa:

$$\theta_j := \theta_j - \alpha \frac{\partial J(\theta)}{\partial \theta_j} \tag{3}$$

Para a regressão linear simples, as derivadas parciais são:

$$\frac{\partial J(\theta)}{\partial \theta_0} = \frac{1}{m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)} \right) \tag{4}$$

$$\frac{\partial J(\theta)}{\partial \theta_1} = \frac{1}{m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)} \right) x^{(i)} \tag{5}$$

Assim, as atualizações dos parâmetros ficam:

$$\theta_0 := \theta_0 - \alpha \cdot \frac{1}{m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)} \right) \tag{6}$$

$$\theta_1 := \theta_1 - \alpha \cdot \frac{1}{m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)} \right) x^{(i)} \tag{7}$$

onde α é a **taxa de aprendizado**, que controla o tamanho dos passos durante a otimização.

5 Conclusão

A regressão linear, apesar de ser um dos modelos mais simples da estatística e da ciência de dados, continua sendo uma ferramenta indispensável para análise exploratória, modelagem preditiva e interpretação de relações entre variáveis. Sua força reside na clareza com que expressa a dependência entre os dados e na facilidade de implementação e interpretação dos resultados.

Ao longo deste estudo, revisamos os fundamentos matemáticos da regressão linear e do algoritmo de gradiente descendente, que permite estimar os coeficientes do modelo por meio de otimização iterativa. A compreensão desses fundamentos não só reforça o domínio técnico, como também abre portas para o entendimento de técnicas mais avançadas, que têm como base os mesmos princípios.

Além disso, ilustramos sua aplicação prática por meio da análise de um conjunto real de dados biomecânicos de pacientes ortopédicos. Essa abordagem evidenciou como a regressão linear pode ser utilizada para compreender relações estruturais em dados médicos, como entre a inclinação sacral e a incidência pélvica ou o ângulo de lordose lombar. Tais análises têm grande valor clínico, pois podem apoiar diagnósticos, revelar padrões anatômicos e embasar intervenções terapêuticas.

Concluímos que, mesmo diante da crescente complexidade dos modelos modernos de aprendizado de máquina, a regressão linear permanece relevante, didática e eficaz. Seu estudo é essencial para qualquer profissional ou pesquisador que deseje extrair conhecimento de dados de maneira sólida e fundamentada.