

Fundamentos Matemáticos dos Algoritmos de Machine Learning: A Regressão Logística

Fernando Lima Filho

18 de junho de 2025

Resumo

Este documento apresenta os fundamentos matemáticos do algoritmo de Regressão Logística, um modelo fundamental para problemas de classificação binária em Machine Learning. Detalhamos a Função Sigmoidal, a formulação da hipótese, a Função de Custo baseada em Entropia Cruzada e o método de otimização via Gradiente Descendente.

1 Introdução à Regressão Logística

Apesar do nome "regressão", a Regressão Logística é um algoritmo de **classificação**. Seu objetivo principal é modelar a probabilidade de uma determinada amostra pertencer a uma classe específica. É mais comumente usada para classificação binária (onde existem apenas duas classes, como "Sim/Não", "Aprovado/Reprovado" ou "Spam/Não Spam").

A ideia central é usar uma função linear dos atributos de entrada e aplicar uma função de mapeamento não-linear para transformar a saída em uma probabilidade, que sempre estará no intervalo $[0, 1]$.

2 Da Regressão Linear à Logística

A Regressão Linear busca modelar uma saída contínua através de uma combinação linear dos atributos:

$$z = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n = \theta^T X$$

O problema é que a saída z pode assumir qualquer valor real $(-\infty, +\infty)$, o que não é adequado para representar uma probabilidade.

Para resolver isso, a Regressão Logística aplica uma função de "esmagamento" (squashing function) chamada **Função Logística** ou **Função Sigmoidal** à saída linear z .

2.1 A Função Sigmoidal

A Função Sigmoidal, denotada por $\sigma(z)$, é definida como:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

Esta função tem a propriedade de mapear qualquer entrada real para o intervalo $(0, 1)$, tornando-a perfeita para modelar probabilidades.

- Se $z \rightarrow +\infty$, então $e^{-z} \rightarrow 0$, e $\sigma(z) \rightarrow 1$.
- Se $z \rightarrow -\infty$, então $e^{-z} \rightarrow +\infty$, e $\sigma(z) \rightarrow 0$.
- Se $z = 0$, então $e^0 = 1$, e $\sigma(z) = 0.5$.

3 A Hipótese da Regressão Logística

A hipótese do modelo, $h_\theta(X)$, representa a probabilidade estimada de que a saída seja 1 ($y = 1$), dado um conjunto de atributos X e parâmetros θ . Ela é formulada aplicando a Função Sigmoidal à combinação linear dos atributos:

$$h_\theta(X) = \sigma(\theta^T X) = \frac{1}{1 + e^{-\theta^T X}}$$

Assim, $h_\theta(X)$ é interpretado como a probabilidade condicional:

$$h_\theta(X) = P(y = 1|X; \theta)$$

Como se trata de uma classificação binária, a probabilidade da outra classe ($y = 0$) é simplesmente:

$$P(y = 0|X; \theta) = 1 - P(y = 1|X; \theta) = 1 - h_\theta(X)$$

A decisão de classificação é feita com base em um limiar (geralmente 0.5):

- Se $h_\theta(X) \geq 0.5$, prevemos $y = 1$.
- Se $h_\theta(X) < 0.5$, prevemos $y = 0$.

4 A Função de Custo

Para treinar o modelo, precisamos encontrar os melhores parâmetros θ que ajustem nossos dados. Isso é feito minimizando uma **Função de Custo** $J(\theta)$.

Usar o Erro Quadrático Médio (como na Regressão Linear) não é ideal, pois a aplicação da Função Sigmoidal tornaria a função de custo não-convexa, resultando em múltiplos mínimos locais. Em vez disso, a Regressão Logística usa uma função de custo baseada no conceito de **Máxima Verossimilhança**, que resulta na chamada **Entropia Cruzada (Cross-Entropy Loss)**.

A função de custo para uma única amostra de treinamento é definida como:

$$\text{Custo}(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & \text{se } y = 1 \\ -\log(1 - h_\theta(x)) & \text{se } y = 0 \end{cases}$$

Esta função penaliza o modelo de forma inteligente:

- Se a classe real é $y = 1$, o custo aproxima-se de 0 quando a previsão $h_\theta(x)$ se aproxima de 1. Se a previsão se aproxima de 0, o custo tende ao infinito.

- Se a classe real é $y = 0$, o custo aproxima-se de 0 quando a previsão $h_\theta(x)$ se aproxima de 0. Se a previsão se aproxima de 1, o custo tende ao infinito.

Podemos escrever esta função de custo de forma mais compacta em uma única linha:

$$\text{Custo}(h_\theta(x), y) = -y \log(h_\theta(x)) - (1 - y) \log(1 - h_\theta(x))$$

A Função de Custo total, $J(\theta)$, para todo o conjunto de treinamento com m amostras é a média dos custos individuais:

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log(h_\theta(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)}))]$$

Esta função é convexa, garantindo que o Gradiente Descendente encontrará o mínimo global.

5 Otimização com Gradiente Descendente

O objetivo é encontrar os valores de θ que minimizam $J(\theta)$. Para isso, usamos o algoritmo do **Gradiente Descendente**. Ele atualiza os parâmetros θ iterativamente na direção oposta ao gradiente da função de custo.

A regra de atualização para cada parâmetro θ_j é:

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

Onde α é a **taxa de aprendizado**.

A derivada parcial da função de custo $J(\theta)$ em relação a um parâmetro θ_j é:

$$\frac{\partial}{\partial \theta_j} J(\theta) = \frac{1}{m} \sum_{i=1}^m (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

Substituindo na regra de atualização, obtemos a fórmula para o Gradiente Descendente na Regressão Logística:

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^m (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

Esta atualização é realizada simultaneamente para todos os θ_j a cada iteração, até que o algoritmo convirja para o valor mínimo de $J(\theta)$. Curiosamente, a fórmula de atualização é idêntica à da Regressão Linear, mas a diferença fundamental é que a hipótese $h_\theta(x)$ agora é a Função Sigmoide.

6 Conclusão

A Regressão Logística é uma ferramenta poderosa e interpretável para problemas de classificação. Ela serve como um bloco de construção essencial para modelos mais complexos, como as redes neurais. Sua base matemática, que conecta a álgebra linear com a teoria da probabilidade através da Função Sigmoide e da otimização via Gradiente Descendente, a torna um tópico de estudo indispensável para quem deseja aprofundar seus conhecimentos em Machine Learning.