Fundamentos Matemáticos dos Algoritmos de Machine Learning: Regressão Ridge

Fernando Lima FIlho

21 de junho de 2025

Resumo

Este documento detalha a Regressão Ridge, uma técnica de regularização aplicada à Regressão Linear para mitigar o sobreajuste (overfitting) e lidar com problemas de multicolinearidade. A análise foca na introdução da penalidade L2 na função de custo e seu impacto na estimação dos coeficientes do modelo.

1 A Necessidade de Regularização

A Regressão Linear Padrão (Mínimos Quadrados Ordinários - OLS) busca encontrar os coeficientes θ que minimizam a Soma dos Erros Quadráticos (Residual Sum of Squares - RSS):

$$J(\theta)_{OLS} = \text{RSS} = \sum_{i=1}^{m}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{m}(y_i - \theta^T X^{(i)})^2$$

Apesar de sua eficácia, o modelo OLS pode sofrer com dois grandes problemas:

1. **Sobreajuste (Overfitting):** Se o modelo tiver muitos atributos (alta dimensionalidade), ele pode se tornar excessivamente complexo, ajustando-se ao ruído dos dados de treinamento e perdendo a capacidade de generalizar para novos dados. Isso geralmente se manifesta em coeficientes θ_j com valores muito grandes.

2. **Multicolinearidade:** Quando os atributos preditores são altamente correlacionados entre si, a estimativa dos coeficientes se torna instável e muito sensível a pequenas variações nos dados.

A Regressão Ridge foi desenvolvida para solucionar esses problemas através de uma técnica chamada **regularização**.

2 A Penalidade L2: O Núcleo da Regressão Ridge

A ideia central da Regressão Ridge é adicionar um termo de penalidade à função de custo da Regressão Linear. Este termo penaliza o modelo por ter coeficientes com magnitudes elevadas. Especificamente, a Ridge utiliza a **penalidade L2**, que é a soma dos quadrados dos coeficientes.

A nova função de custo para a Regressão Ridge é:

$$J(\theta)_{Ridge} = \text{RSS} + \lambda \sum_{j=1}^{n} \theta_j^2$$

$$J(\theta)_{Ridge} = \sum_{i=1}^{m} (y_i - \theta^T X^{(i)})^2 + \lambda \sum_{j=1}^{n} \theta_j^2$$

Onde:

- $\sum_{j=1}^{n} \theta_j^2$ é a **penalidade L2**. Note que, por convenção, a penalidade não é aplicada ao coeficiente de interceptação θ_0, pois ele representa o valor base da predição.
- λ (lambda) é o **hiperparâmetro de regularização**. Ele controla a intensidade da penalidade.

O objetivo do algoritmo agora é minimizar esta nova função de custo. Isso cria um trade-off: o modelo precisa minimizar tanto o erro de previsão (RSS) quanto a magnitude de seus coeficientes (a penalidade L2).

3 O Papel do Hiperparâmetro λ

O valor de λ determina o quão forte será a regularização e é crucial para o desempenho do modelo.

- **Se $\lambda = 0$:** O termo de penalidade se anula. A Regressão Ridge se torna idêntica à Regressão Linear Padrão. O modelo terá alta variância e pode sofrer overfitting.
- **Se $\lambda \to \infty$:** A penalidade domina a função de custo. Para minimizar a função, o modelo é forçado a encolher os coeficientes θ_j para valores muito próximos de zero. Isso resulta em um modelo extremamente simples (quase uma linha horizontal), que provavelmente sofrerá de subajuste (underfitting) e alto viés.

A escolha de um λ ótimo, que equilibre o viés e a variância, é geralmente feita através de validação cruzada.

4 Encontrando os Coeficientes

Assim como na Regressão Linear, podemos encontrar os coeficientes ótimos através de um método analítico (solução de forma fechada). A solução para a Regressão Linear Padrão é dada pela Equação Normal:

$$\hat{\theta}_{OLS} = (X^T X)^{-1} X^T y$$

Para a Regressão Ridge, a adição do termo de penalidade modifica a equação, resultando em:

$$\hat{\theta}_{Ridge} = (X^T X + \lambda I)^{-1} X^T y$$

Onde I é a **matriz identidade** (uma matriz com 1s na diagonal principal e 0s no resto).

Esta modificação é matematicamente muito importante. O termo λI adiciona um valor positivo à diagonal da matriz $X^T X$, garantindo que ela seja sempre invertível. Isso resolve o problema da multicolinearidade, que na Regressão Linear Padrão pode tornar a matriz $X^T X$ singular (não-invertível).

5 A Importância da Padronização dos Dados

Como a penalidade da Regressão Ridge é baseada na magnitude dos coeficientes, o algoritmo é sensível à escala dos atributos. Se um atributo tiver uma escala muito maior que os outros (e.g., ‘salário‘ vs. ‘idade‘), seu coeficiente será penalizado de forma desproporcional.

Para garantir que a penalização seja justa para todos os atributos, é **essencial padronizar (standardize)** os dados antes de aplicar a Regressão Ridge. A técnica mais comum é a **Padronização Z-score**, que transforma os atributos para que tenham média 0 e desvio padrão 1.

6 Conclusão

A Regressão Ridge é uma melhoria robusta sobre a Regressão Linear. Ao introduzir a penalidade L2, ela efetivamente combate o sobreajuste, reduz a complexidade do modelo e fornece uma solução estável mesmo na presença de multicolinearidade. A compreensão de sua matemática revela um dos conceitos mais importantes do machine learning moderno: o trade-off entre ajuste e simplicidade, controlado pelo hiperparâmetro de regularização.