

UÉvora  
MD - MEI 2023/2024  
*Take-home exam*  
**Funções de perda de regressão e ruído**

Gonçalo Candeias Amaro  
Número 56870

19 Jan 2024

## **1 Estratégias Comuns para Funções de Perda em Modelos de Regressão**

### **1.1 Erro Quadrático Médio (MSE) e sua Sensibilidade a Outliers**

Devido à sua simplicidade e eficácia na medição da média dos quadrados dos erros ou desvios, o Erro Quadrático Médio (MSE) é uma função de perda muito usada em tarefas de regressão. Em outras palavras, o MSE mostra o desempenho do modelo em termos de erro calculando a média do quadrado da diferença entre os valores reais e previstos.

No entanto, a principal desvantagem do MSE é sua sensibilidade extrema a valores inesperados ou imprevistos em um conjunto de dados. Como o MSE aumenta ao quadrado os erros, diferenças maiores têm um efeito menor no resultado final. Isso indica que o MSE pode não ser a melhor medida de avaliação em conjuntos de dados onde os desvios não são apenas anomalias, mas também parte de variações significativas. Em situações como essa, a presença de valores extremos relativamente baixos pode prejudicar a avaliação do modelo, forçando-o a se concentrar excessivamente em ajustar esses pontos em detrimento do desempenho geral.

Essa limitação pode ser uma vantagem em contextos em que os outliers, em vez de erro ou ruído, são uma característica comum do conjunto de dados. As alternativas ao MSE, como o erro absoluto médio (MAE) ou o Huber Loss, são frequentemente consideradas para contornar esta limitação.

## 1.2 Erro Médio Absoluto (MAE): Uma Alternativa Robusta

O Erro Absoluto Médio (MAE), em contraste com o Erro Quadrático Médio (MSE), é uma opção mais confiável. Isso é particularmente verdadeiro em casos em que os conjuntos de dados contenham ausências significativas. O MAE usa uma abordagem diferente, calculando a média das diferenças absolutas entre os valores reais e previstos. Por outro lado, o MSE amplifica o impacto dos erros maiores aumentando as diferenças ao quadrado. Isso significa que, independentemente do tamanho, cada erro contribui linearmente para o MAE total.

O MAE é especialmente útil nessa característica em situações em que os outliers não devem afetar o modelo de forma desproporcional. A interpretação do MAE é mais direta do que a do MSE. O MAE mostra diretamente o erro médio no mesmo nível dos dados, enquanto o MSE fornece uma medida que está em uma escala quadrática em relação aos dados originais, dificultando a interpretação.

É importante ter em mente que o MAE tem limitações. O MAE pode ser ineficaz quando tratamos todos os erros de forma semelhante porque a punição de erros graves é mais importante do que a de erros menores. Além disso, o MAE não é diferenciável em zero, o que pode ser um problema para certos algoritmos de otimização. Finalmente, o MAE não é tão sensível quanto o MSE a mudanças menores nos dados; em certos casos, isso pode ser uma desvantagem.

## 1.3 Função de Perda de Huber: Combinando MSE e MAE

A Função de Perda de Huber, também conhecida como Perda de Huber, é um método inovador para o aprendizado de máquina que combina características do Erro Quadrático Médio (MSE) e do Erro Absoluto Médio (MAE). Esta função de perda foi desenvolvida para combinar as melhores qualidades de ambas as abordagens, equilibrando a sensibilidade aos erros menores, que é comum no MSE, com a robustez em relação aos outliers, que é comum no MAE.

A função de perda de Huber é principalmente quadrática para erros pequenos e linear para erros grandes. Um parâmetro chamado delta determina essa transição. Ele define o limiar entre o comportamento tipo MAE e tipo MSE. A função de perda penaliza mais fortemente os pequenos erros quando as diferenças entre o valor real e previsto são menores que o delta. Isso ocorre porque a função segue o modelo quadrático do MSE. A função de perda se torna linear para diferenças maiores que o delta, semelhante ao MAE, o que reduz a influência de grandes exceções.

A Função de Perda de Huber é particularmente versátil e eficaz em uma variedade de situações de aprendizado de máquina devido a esse design híbrido. Ela é particularmente útil em situações em que se deseja manter a sensibilidade a erros menores, mas sem permitir que os outliers tenham um impacto excessivamente significativo, como ocorre com o MSE.

Além disso, a Função de Perda de Huber é uma opção confiável em vários cenários de modelagem porque é menos sensível a escolhas específicas do pa-

râmetro delta do que outras funções de perda. Como a parte quadrática da função é diferenciável, o que facilita o uso de métodos de otimização baseados em gradiente, ela também ajuda a resolver alguns problemas de otimização que podem surgir com o MAE.

## 1.4 Função de Perda Quantil: Foco em Segmentos Específicos da Distribuição

A Função de Perda Quantil é um método distinto para o aprendizado de máquina, principalmente para tarefas de regressão. Sua principal característica é a capacidade de se concentrar em quantidades específicas de distribuição de dados em vez de apenas minimizar o erro médio, como fazem o MSE ou o MAE. Essa propriedade torna a Função de Perda Quantil muito útil para aplicações em que o interesse é em modelar a incerteza ou prever extremos da distribuição, como percentis superior ou inferior.

Na prática, o quantil tau, que varia de 0 a 1, é o parâmetro que configura a Função de Perda Quantil. O quantil da distribuição que o modelo deve aprender a prever é determinado por este parâmetro. Quando  $\tau = 0,5$ , a função de perda concentra na mediana dos dados. Por outro lado, quando  $\tau = 0,1$ , a função de perda concentra no décimo percentil.

A função de perda quantitativa é única devido à sua assimetria na penalização dos erros. Erros abaixo ou acima de um determinado valor são penalizados de maneira diferente. Isso significa que o modelo pode ser incentivado a prever valores que sejam consistentemente acima ou abaixo do valor real dependendo do valor de tau selecionado. Isso é particularmente útil quando subestimar ou superestimar tem consequências diferentes.

## 2 Estratégias Avançadas para Funções de Perda em Modelos de Regressão

### 2.1 Regressão Ridge e Lasso: Regularização e Controle de Ruído

As técnicas de Regressão Ridge e Lasso, que são técnicas de aprendizado de máquina e estatísticas sofisticadas, são usadas para criar modelos de regressão mais confiáveis, especialmente quando há muitas variáveis preditoras. Ambos os métodos visam diminuir a complexidade do modelo e o sobreajuste.

#### 2.1.1 Regressão Ridge: Reduzindo a Complexidade do Modelo

A Regressão Ridge, também conhecida como regularização Tikhonov, adiciona um termo de penalidade proporcional ao quadrado dos coeficientes do modelo para alterar a função de custo da regressão linear padrão. Um parâmetro de regularização conhecido como lambda é multiplicado por este termo. Esta punição visa regularizar o modelo diminuindo os coeficientes. A penalidade mais

alta e os coeficientes menores resultantes de um  $\lambda$  maior reduzem a variação do modelo e, portanto, o risco de overfitting. No entanto, o modelo pode se tornar excessivamente simplificado, resultando em underfitting, se o  $\lambda$  for muito grande. A Regressão Ridge é especialmente vantajosa em situações em que as variáveis preditoras estão correlacionadas.

### **2.1.2 Regressão Lasso: Reduzindo a Complexidade do Modelo e Realizando Seleção de Variáveis**

A Regressão Lasso (Least Absolute Shrinkage and Selection Operator) também adiciona um termo de penalidade à função de custo; no entanto, este termo é proporcional ao valor absoluto dos coeficientes, ao contrário da Ridge. Por outro lado, tem um parâmetro de regularização  $\lambda$ . O efeito único do Lasso é que ele pode reduzir os coeficientes de certas variáveis a zero. Isso significa que ele seleciona um subconjunto de variáveis explicativas mais simples e exclui as outras variáveis do modelo. Isso torna o Lasso uma ferramenta útil para regularização e seleção de características. O valor de  $\lambda$  deve ser escolhido com cuidado, assim como na Regressão Ridge, para equilibrar a complexidade do modelo e o risco de underfitting.

## **2.2 Funções de Perda Robustas: Lidando Efetivamente com Outliers**

As funções de perda robustas, como as de Tukey e Cauchy, foram criadas com o objetivo específico de reduzir o impacto dos erros imprevistos em modelos estatísticos, especialmente em análises de regressão. As funções de perda tradicionais, como o Erro Quadrático Médio (MSE), que podem ser excessivamente sensíveis a outliers, são substituídas por essas funções. Em situações em que os dados contêm valores extremos ou anômalos, as funções de perda robustas podem fornecer estimativas mais precisas e confiáveis.

### **2.2.1 Função de Perda de Tukey: Uma Abordagem Híbrida**

A característica distintiva desta função é o tratamento menos agressivo dos outliers. Ela aumenta o custo dos erros, ou penalidade, apenas até um ponto específico. A influência dos erros que excedem esse limite é efetivamente reduzida. Isso indica que os excedentes têm um efeito menor quando os pontos de dados dentro de um determinado intervalo de valores são tratados de maneira semelhante à do MSE. A Função de Perda de Tukey é especialmente útil em conjuntos de dados em que se suspeita da presença de outliers significativos que podem distorcer as estimativas.

### **2.2.2 Função de Perda de Cauchy: Uma Abordagem Mais Conservadora**

Sua abordagem ainda mais tolerante aos outliers é uma característica conhecida desta função. A penalidade para os erros aumenta de forma logarítmica, em vez

de aumentar quadraticamente. Isso resulta em uma curva de perda mais plana em comparação com a Função de Perda de Tukey, também conhecida como MSE. Como resultado, a Função de Perda de Cauchy é muito útil em situações em que os desvios são extremamente significativos ou quando o impacto que eles têm no modelo deve ser significativamente reduzido. Em problemas de visão computacional e ajuste de modelos geométricos, onde os dados podem ser particularmente propensos a grandes desvios, ela é frequentemente utilizada.