

# ML Under Modern Optimization Lens - Stable Regression -

## Exercícios

Giovanni Amorim

Abril, 2023

### 1 Regressão estável

O exercício consistiu na implementação e avaliação do algoritmo de regressão estável apresentado no capítulo "Stable Regression" do livro. Foram feitas implementações da regressão linear de mínimos quadrados com e sem regularização Lasso e seus respectivos modelos robustos.

Foram implementadas regressão linear de mínimos quadrados com e sem regularização Lasso e seus respectivos modelos robustos para avaliação comparativa. As avaliações foram feitas repetidas vezes para diminuir o impacto da separação em conjuntos de treino e teste. A base utilizada foi a *winequality - red*, para o problema de previsão de qualidade de vinho tinto ( $n=1599$ ,  $p=11$ ).

Para a regressão estável, o processo de avaliação consiste em identificar o modelo selecionado a partir de um subconjunto de validação nos dados de treino e calcular o erro quadrático relativo ao conjunto de teste. Para a versão robusta, o conjunto selecionado por otimização pode ser encontrado utilizando os valores do vetor  $u$ :

$$Val = \{i \mid u_i \geq u_{(K)}\}$$

Ou seja, os valores em  $u$  maiores ou iguais do que o  $K$ -ésimo maior valor indicam as amostras selecionadas pelo problema para cálculo do erro. Assim, o erro computado para avaliar o modelo na seleção do valor  $\lambda$  foi:

$$\|X_{Test} \beta - y_{Test}\|_2^2$$

Dessa forma, para cada rodada  $i$  e tamanho de base de validação  $K$ , esses processos foram executados para seleção de valores  $\lambda$  (para modelos Lasso e Lasso Robusto) e avaliação na base de teste (i).

A média dos erros de teste computados entre as rodadas  $i$  são apresentados na tabela a seguir:

Sem regularização:

%Treino/Validação	Aleatório	Otimizado
70/30	0.528977	0.527051
60/40	0.527016	0.526913
50/50	0.532246	0.527594

Com regularização:

%Treino/Validação	Aleatório	Otimizado
70/30	0.528977	0.527074
60/40	0.531359	0.526845
50/50	0.539702	0.527594

As tabelas a seguir trazem o desvio padrão das estimativas de Y fora da amostra entre os modelos:

Sem regularização:

%Treino/Validação	Aleatório	Otimizado
70/30	0.721743	0.722376
60/40	0.720584	0.722508
50/50	0.727098	0.723917

Com regularização:

%Treino/Validação	Aleatório	Otimizado
70/30	0.721743	0.722399
60/40	0.723634	0.722468
50/50	0.732015	0.723927

Como pode-se perceber, o modelo de regressão estável performou melhor do que a regressão Lasso com validação aleatória, mesmo com uma validação repetida, em todos os cenários.