

# ML Under Modern Optimization Lens - Boosting - Exercícios

Giovanni Amorim

Junho, 2023

## 1 Boosting - AdaBoost

O exercício consiste na implementação e análise do algoritmo adaboost em uma tarefa de classificação binária. O objetivo é avaliar a capacidade do modelo de aprender os padrões apresentados pelos dados, capacidade de generalização e analisar características internas do algoritmo.

O dataset "adult" foi utilizado, com a tarefa de previsão de classe de renda de indivíduos civis baseado em características individuais. A base original possui 32.561 amostras, porém apenas 10.000 amostras foram consideradas no exercício, para reduzir a complexidade computacional.

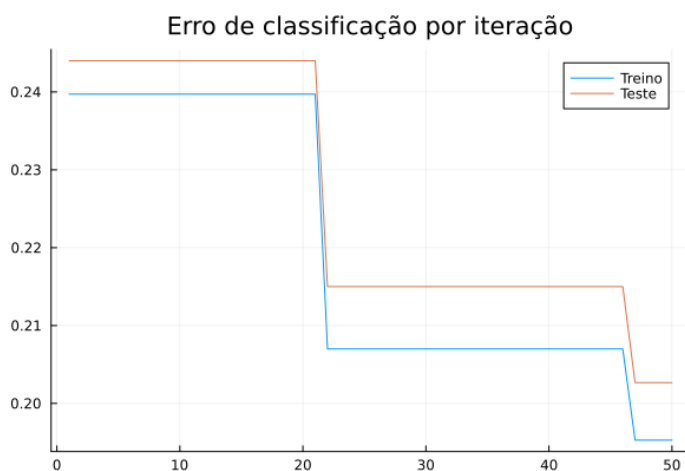
O modelo utilizado como aprendiz fraco foi a árvore de decisão de profundidade 1, pela implementação em "DecisionTree.jl".

### 1.1 Erros de treino e teste

Os erros de treino e teste por iteração são avaliados utilizando a fórmula de previsão final do modelo até a iteração corrente:

$$H_m(x) = \text{sign}\left(\sum_{i=1}^m \alpha_i f_i(x)\right)$$

Os gráficos abaixo mostram os erros avaliados por iteração:



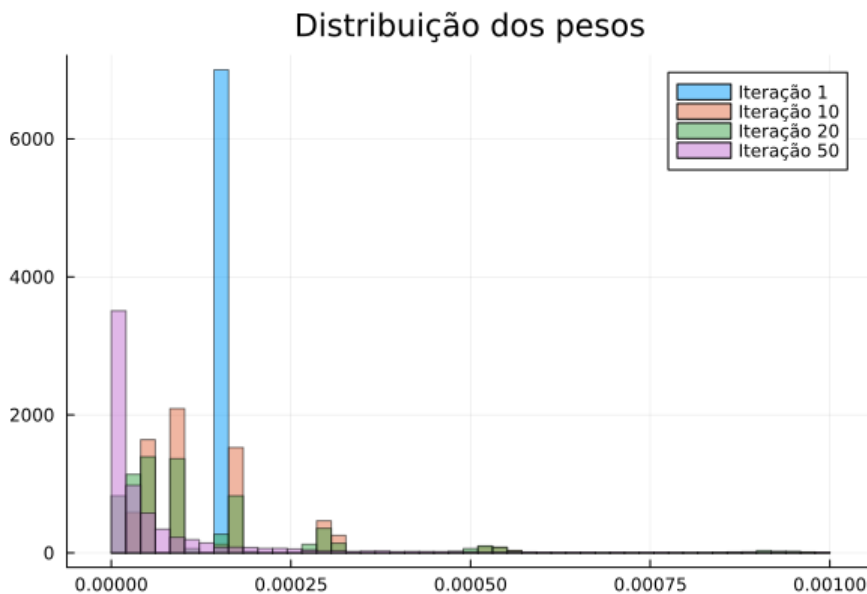
Erros medidos nas iterações 1, 10, 20 e 50:

iteração	erro em treino	erro em teste
1	23.97	24.4
10	23.97	24.4
20	23.97	24.4
50	19.52	20.26

Percebemos que a redução dos erros acontece no processo de boosting, mas em "escada", mantendo um mesmo erro durante diversas iterações até "aprender" um novo padrão útil para reduzir os erros seguintes.

## 1.2 Distribuição dos pesos das amostras

A seguir temos as distribuições dos pesos das amostras para as iterações 1, 10, 20 e 50:



Como esperado, os pesos começam todos iguais, concentrados em  $\frac{1}{N}$ . Com o passar das iterações, temos concentrações de pesos em amostras mais "difíceis" de acertar para os aprendizes treinados até o momento. Ao fim das 50 iterações, parece que os conjuntos de dados com pesos maiores diminuí e ficamos com uma distribuição com comportamento do tipo exponencial.