

Exercício 2

Uma preocupação constante na aplicação de métodos de aprendizado de máquinas é garantir que o modelo final faça boas generalizações. Para tanto, uma estratégia comum é a divisão de dados em conjuntos de treino (com fração de treino e fração de avaliação) e teste.

Suponha que tenhamos dois modelo em competição:

$$h_1(x) = b_0 + b_1x,$$

$$h_2(x) = b_0 + b_1x + b_2x^2.$$

Considerando a função custo

$$J(y_i, h(\mathbf{x})) = \frac{1}{n} \sum_{i=1}^n [y_i - \hat{h}(x_i)]^2,$$

descreva como decidir sobre o melhor modelo, detalhando onde cada conjunto de dados será utilizado e como. Utilize o método de 5-fold.

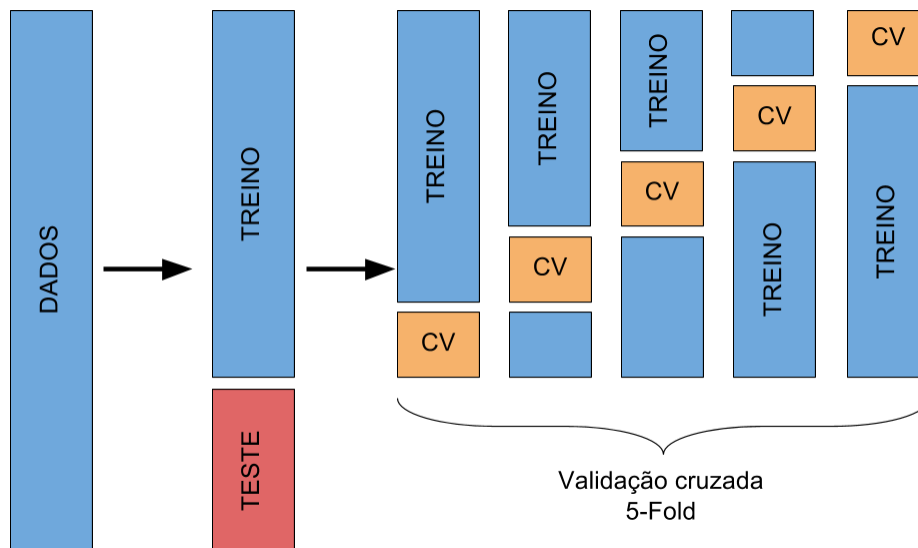


Figure 1: Validação cruzada 5-fold