

Pacotes sugeridos

dplyr
ggplot2
reshape2
mlbench
car
caret

Base “BostonHousing” do pacote mlbench

Exercício 1

Carregar os dados da base de casas de Boston em um objeto chamado “housing” e inspecionar os tipos de variáveis presentes.

Exercício 2

Explorar e visualizar a distribuição de nossas variáveis target

Exercício 3

Explorar e visualizar qualquer potencial correlação entre medv e as variáveis crim, rm, age, rad, tax e lstat

Exercício 4

Definir a semente aleatória 123 e dividir os dados em treinamento e teste, usando a proporção 75/25.

Exercício 5

Podemos ver que as variáveis crim, rm, tax e lstat podem ser bons preditores para a variável medv. Então vamos seguir em frente, e vamos ajustar um modelo linear para essas variáveis.

Exercício 6

Vamos analisar e diagnosticar o modelo ajustado.

Exercício 7

Podemos visualizar alguns problemas no nosso modelo logo de cara, com resíduos que destoam, um QQ plot bem ruim e um r-quadrado baixo.

Então vamos testar outro modelo, só que dessa vez vamos transformar a MEDV por conta da assimetria positiva que é exibida.

Exercício 8

Analisando o diagnóstico do modelo, o quê concluímos? É uma melhora em relação ao primeiro modelo?

Um dos pressupostos de um modelo linear é que a média dos resíduos é zero. Podemos testar.

Exercício 9

Crie um data frame dos nossos valores previstos e os valores originais.

Exercício 10

Vamos “plotar” esses dados e visualizar a “performance” dos nossos modelos