

# Mineração de Dados

## Lista 2

Rafael Izbicki

**Lista em trios.**

**NÃO COPIE! JUSTIFIQUE TODOS OS EXERCÍCIOS!**

**Exercício 1.** Considere o banco `Carseats` disponível no pacote `ISLR` do R. O objetivo é prever a variável `Sales` (unidades vendidas) de cadeiras de bebês para carros com base nas demais.

- (a) Divida o conjunto de dados em 60% para treinamento e 40% para teste.
- (b) Utilizando o conjunto de treinamento, ajuste uma regressão (i) via mínimos quadrados e (ii) via lasso (usando validação-cruzada no treinamento para escolher  $\lambda$ ). Qual o melhor valor de  $\lambda$  encontrado para o lasso?
- (c) Mostre os coeficientes estimados para ambos os métodos. Interprete.
- (d) Qual dos métodos acima apresentou melhores resultados? Responda essa pergunta utilizando o conjunto de teste e o melhor valor de  $\lambda$  encontrado. Mostre também intervalos de confiança para o risco estimado.
- (e) Repita os itens anteriores adicionando todas as interações duas-a-duas entre as covariáveis. Utilizar interações foi benéfico?