

## MAE 5905: Introdução à Ciência de Dados

Lista 3. Primeiro Semestre de 2020. Entregar 03/05/2020.

1. Problema 18 do Capítulo 6.
2. Problema 23 do Capítulo 6.
3. Problema 24 do Capítulo 6.
4. (a) Use a função `rnorm()` (simula valores de uma distribuição normal) do R para gerar um preditor  $X$  com  $n = 100$  observações, bem como um erro  $\varepsilon$  também de comprimento 100.  
(b) Simule um vetor de resposta  $Y$ , de comprimento  $n = 100$  de acordo com o modelo

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \varepsilon,$$

na qual os parâmetros  $\beta_i$  são constantes de sua escolha.

- (c) Considere o modelo de (b), agora com os  $\beta_i$  e  $\varepsilon$  desconhecidos,  $X$  como em (a)  $Y$  como em (b). Qual seria o melhor modelo usando  $R^2$  ajustado e BIC?
- (d) Para o modelo como em (c), obtenha os estimadores ridge e lasso. Use VC para selecionar o valor ótimo de  $\lambda$ .