

Universidade Federal do Paraná - Departamento de Estatística
Especialização em Data Science e Big Data
Prof. Cesar Augusto Taconeli
Avaliação - 19/10/2018

Vamos considerar a aplicação de uma árvore de regressão à base de dados `abalone`, do pacote `AppliedPredictiveModeling`. Os dados referem-se a 4177 espécimes de abalone, tipo de molusco encontrado ao longo das águas costeiras de todos os continentes. A variável resposta é a idade do molusco, aferida pelo número de anéis internos, que é um procedimento demorado e pouco adequado. O objetivo é ajustar um modelo que permita estimar a idade a partir de outras medidas, que são obtidas com maior facilidade. Para maiores detalhes a respeito da base, consultar a documentação e o link fornecido.

Para a análise, as primeiras 3000 linhas deverão ser usadas para ajuste, e as demais para validação.

1. Qual o tamanho da árvore (número de nós finais) selecionada por validação cruzada? Quantas são as partições?

Nota: Fixe a semente com `set.seed(1)`. Estabeleça $cp = 0.001$ para o processo de poda.

2. Quantas covariáveis aparecem no ajuste da árvore?
3. Qual a idade estimada para moluscos com:
 - a) `ShellWeight=0.18` e `ShuckedWeight=0.25`;
 - b) `ShellWeight=0.31` e `ShuckedWeight=0.45`?
4. Qual o resíduo para cada um dos dados? Considere, para o primeiro, `Rings=8` e para o segundo `Rings=10`.
5. Usando os dados de validação, calcule e apresente o valor da soma de quadrados de resíduos.