Discente: Deivison Venicio Souza

Universidade Federal do Paraná - Departamento de Estatística
Especialização em Data Sciece 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.

Usando cp = 0.001 o tamanho da árvore (nós finais) e o número de partições foram:

- Tamanho da árvore = **11 nós finais** (igual a árvore sem podar);
- Número de partições = **10 partições.**

Usando a regra de Breiman como critério, sugere-se podar usando cp = 0.013786 (observado no gráfico gerado pela função **plotcp**). Assim, a árvore após processo de poda teria:

- **Tamanho da árvore** = **8 nós finais** (igual a árvore sem podar);
- **Número de partições** = **7 partições**.

Pelo critério de poda de Breiman ter-se-ia um modelo de árvore mais parcimonioso.

2. Quantas covariáveis aparecem no ajuste da árvore?

R: Apenas duas variáveis foram usadas para construir a árvore: **ShellWeight** e **ShuckedWeight**.

3. Qual a idade estimada para moluscos com:

Para este previsão considerei o modelo com podado com cp = 0.001. Então, os resultados foram:

a) ShellWeight=0.18 e ShuckedWeight=0.25.

R: Predição de Rings = **9.071168**

b) ShellWeight=0.31 e ShuckedWeight=0.45.

R: Predição de Rings = **10.977230**

4. Qual o resíduo para cada um dos dados? Considere, para o primeiro, Rings=8 e para o segundo Rings=10.

Os resíduos são dados pela diferença entre os valores observado (reais) e estimados. Assim, para as predições da questão 3 têm-se:

- Resíduo para predição de Rings (quando ShellWeight=0.18 e ShuckedWeight=0.25) = **1.071168**

- Resíduo para predição de Rings (quando ShellWeight=0.31 e ShuckedWeight=0.45) = **1.977230**

5. Usando os dados de validação, calcule e apresente o valor da soma de quadrados de resíduos.

O valor da soma de quadrados de resíduos no conjunto de validação foi de = **10379.5**

O valor de MSE no conjunto de validação foi de = **8.81861**

O valor de RMSE no conjunto de validação foi de = **2.969614**

Obs: O código utilizado para solução deste trabalho está em anexo ao e-mail.
