

FA084 / AP522 – Atividade Pré-Aula 3

O conjunto de dados 'fa084_cana_estagio.csv' contém a produtividade de diferentes talhões (uma divisão de área, um "campo") de produção de cana-de-açúcar e o número de vezes que estes talhões foram colhidos. A produtividade corresponde a coluna TCH (Toneladas de Cana-de-açúcar por Hectare) e o número de cortes corresponde a coluna Estágio.

Atividade

Preencha o script base com a rotina abaixo:

1. Use `set.seed(RA)`. Basta substituir o RA do script pelo seu.
2. Carregue o conjunto de dados.
3. Faça a divisão em conjunto de treino (75%) e teste (25%). Use o mesmo procedimento visto em sala na aula 02.
4. Faça o sample de um número entre 1,2,3. Coloque na variável `a`.
5. Faça o sample de um número entre 4 ou 5. Coloque na variável `b`.
6. Treine dois modelos usando o conjunto de treino, um com grau `a` e outro com grau `b`.
7. Teste as predições dos dois modelos no conjunto de teste.
8. Calcule o `mean absolute error` (MAE) de cada modelo.

OBS: Não altere as posições de cada `set.seed()`. SEMPRE execute o `set.seed()` imediatamente antes de uma chamada da função `sample()`.

1. Qual modelo apresenta melhor resultado de acordo com o MAE? Justifique.
2. Construa um gráfico de TCH vs Estágio. É possível identificar o tipo de relação entre as variáveis? Você considera que os modelos escolhidos em A e B são adequados para representar a relação entre TCH e Estágio?

Entrega:

Um arquivo `.zip` com o nome: `pre_aula03_RA_primeironome.zip`, com:

- Um arquivo pdf com o nome `fa084_preaula03_RA_primeironome.pdf`
- O scrip preenchido com o nome `fa084_preaula03_RA_primeironome.R`