

Segunda prova - Introdução à ciência de dados

1 Questão: Classificação de Gênero Musical

O objetivo desta atividade é construir modelos de aprendizado de máquina capazes de **prever o gênero de uma música** com base em suas características extraídas do áudio.

Siga os passos abaixo:

1. **Importe o arquivo** contendo os dados das músicas.
2. **Filtre os dados**, mantendo apenas as observações cuja variável **Class** seja igual a **2** ou **8**. Essas duas classes representam os dois gêneros musicais que serão comparados.
3. **Realize as manipulações necessárias**.
4. **Apresente uma análise gráfica exploratória** com o objetivo de caracterizar cada um dos dois gêneros. Você pode utilizar gráficos como boxplots, histogramas ou gráficos de dispersão.
5. **Ajuste modelos de Floresta Aleatória (Random Forest)** com diferentes números de árvores: **50**, **100** e **500**.
6. **Compare o desempenho** dos modelos ajustados e comente os resultados. Qual número de árvores produziu o melhor modelo?
7. **Ajuste um modelo SVM com kernel linear**, variando o parâmetro **cost** entre os valores: **0.01**, **0.1**, **1** e **5**. Para isso, utilize a função **tune()** conforme apresentado na página 390 do livro *Introduction to Statistical Learning (ISLR)*.
8. Para o **melhor modelo SVM** encontrado, analise os resultados obtidos (acurácia, matriz de confusão, etc.) e comente.
9. **Compare os dois modelos** (Floresta Aleatória e SVM linear). Qual deles teve melhor desempenho? Justifique sua resposta com base nas métricas e nos gráficos obtidos.