



INF-0615 – APRENDIZADO DE MÁQUINA SUPERVISIONADO I

EXERCÍCIO 7

ÁRVORES DE DECISÃO E FLORESTAS ALEATÓRIAS BANKNOTE AUTHENTICATION DATASET

1 Descrição do Problema

Um dos maiores problemas que muitos bancos e empresas ainda enfrentam é detectar a utilização de cédulas falsas em transações financeiras. Neste contexto, o objetivo deste exercício é prever se uma dada cédula é verdadeira ou falsa. Para abordar este problema, temos uma base de dados com diferentes características obtidas após uma *Wavelet Transform*[1] das imagens das cédulas. Os atributos obtidos após a transformação são listados abaixo:

- **variance**: variância dos valores da imagem após a transformação;
- **skewness**: distorção da distribuição dos valores na imagem após a transformação;
- **curtosis**: curtose da distribuição dos valores da imagem após a transformação;
- **entropy**: entropia da imagem original;
- **Class**: Valor alvo que queremos prever, o qual indica se a cédula é falsa (*forgery*) ou verdadeira (*genuine*);

Esse dataset pode ser encontrado em sua configuração original em [2].

2 Tarefas

Neste exercício, pedimos que você:

1. Inspecione os dados.
2. Treine uma árvore de decisão para classificar se uma cédula é autêntica ou falsa.
3. Varie a profundidade da árvore de decisão e plote a acurácia balanceada no treino e na validação pela profundidade da árvore. Qual a melhor árvore?
4. Treine também florestas aleatórias variando o número de árvores geradas. Plote a acurácia balanceada no conjunto de treinamento e de validação pelo número de árvores na floresta. Qual a melhor floresta?
5. Calcule a matriz de confusão e a acurácia balanceada para os melhores modelos dos itens anteriores no conjunto de teste. Qual melhor generalizou?

3 Arquivos

Os arquivos disponíveis no Moodle são:

- *banknote_authentication_train.csv*: dados de treinamento;
- *banknote_authentication_validation.csv*: dados de validação;
- *banknote_authentication_test.csv*: dados de teste;
- *Ex07.R*: código que implementa a solução do exercício;

4 Referências

1. *Wavelet Transorms*. ScienceDirect.
<https://www.sciencedirect.com/topics/computer-science/wavelet-transforms>.
2. *Banknote authentication dataset* <https://archive.ics.uci.edu/ml/datasets/banknote+authentication>.