



TRABALHO 2 WINE QUALITY DATASET

1 Descrição do Dataset

A qualidade de um vinho depende de diversos fatores, como a forma que as uvas são cultivadas (viticultura) e como elas são transformadas em vinhos (vinificação). A forma como todo o processo é feito tem relação direta com diversas características químicas da bebida e, consequentemente, com o seu sabor. Uma medida da qualidade de um vinho pode ser feita baseando-se em medições químicas (pH, densidade, acidez, porcentagem de álcool) e testes sensoriais feitos por especialistas.

Neste trabalho, você irá classificar a qualidade de um vinho a partir de medições químicas. As medições disponíveis são:

- | | | |
|--------------------|------------------------|-------------|
| – Fixed acidity | – Chlorides | – pH |
| – Volatile acidity | – Free sulfur dioxide | – Sulphates |
| – Citric acid | – Total sulfur dioxide | |
| – Residual sugar | – Density | – Alcohol |

- **Quality:** a qualidade do vinho, com valor “0” (ruim) ou “1” (bom).

2 Tarefas

Pedimos que você:

1. Inspecione os dados. Quantos exemplos você tem? Qual o intervalo de valores de cada feature?
2. Normalize os dados de modo que eles fiquem todos no mesmo intervalo.
3. Treine uma regressão logística para prever a qualidade dos vinhos.
4. Classifique os dados de validação.
5. Calcule a matrix de confusão e acurácia no conjunto de validação.

3 Arquivos

Os arquivos disponíveis no Moodle são:

- *wineQuality_train.data*: conjunto de dados para treinamento;
- *wineQuality_val.data*: conjunto de dados para validação;
- *wineQuality_test.data* (**apenas durante a avaliação**): conjunto de dados retido pelo professor a ser utilizado na avaliação;

4 Avaliação

O dataset foi previamente dividido aleatoriamente em três conjuntos — treino, validação e teste — e apenas os dois primeiros serão disponibilizados para os alunos implementarem suas soluções.

No dia da avaliação, nos 40 minutos finais da aula, o professor irá disponibilizar o conjunto de teste e validar a solução de cada aluno (ou dupla de alunos). A avaliação consiste da análise dos resultados reportados sobre o conjunto de teste e possíveis perguntas a respeito das decisões tomadas pelo aluno durante a implementação de sua solução.

Observações sobre a avaliação:

- **NÃO** haverá submissão do código ou relatório do trabalho no Moodle;
- O trabalho poderá ser feito em duplas, podendo haver repetição das duplas a cada trabalho;
- Pelo menos um membro da dupla deverá estar presente no momento da avaliação;
- As notas do trabalho serão divulgadas em até uma semana após a avaliação;