

Aula 02 – Exercícios Pré-processamento de Dados

Submeter sua solução num arquivo nome_sobrenome_aula2.pdf

- 1) (0,5) Use o dataset Wine: <https://archive.ics.uci.edu/dataset/109/wine>

Instances = 178
Features = 13

Variable Name	Role	Type
class	Target	Categorical
Alcohol	Feature	Continuous
Malicacid	Feature	Continuous
Ash	Feature	Continuous
Alcalinity_of_ash	Feature	Continuous
Magnesium	Feature	Integer
Total_phenols	Feature	Continuous
Flavanoids	Feature	Continuous
Nonflavanoid_phenols	Feature	Continuous
Proanthocyanins	Feature	Continuous
Color_intensity	Feature	Continuous
Hue	Feature	Continuous
OD280_OD315_of_diluted_wines	Feature	Continuous
Proline	Feature	Integer

- Faça um boxplot analisando a média e desvio padrão dos valores de cada *feature* e se há *outliers*.
- Aplique normalização *Min-Max Scaling* nas *features*.
- Aplique PCA para reduzir a dimensionalidade e plote um gráfico 2D.
- Calcule e plote a matriz de correlação entre as *features* e indique quais atributos tem maior correlação entre si.
- Verifique se o dataset é desbalanceado.

- 2) (0,5) Na aula foram mencionados exemplos de abordagens para tratar valores ausentes, como preencher com a média, mediana ou moda da coluna, etc. Remova alguns elementos do dataset e aplique alguma dessas técnicas para prever os valores removidos. Além disso, veja a biblioteca sklearn e tente aplicar alguma das funções dessa biblioteca: <https://scikit-learn.org/stable/modules/impute.html>