

Relatório Aprendizado de Máquina Aula 1

Lucas Ribeiro da Silva - 2022055564

Universidade Federal de Minas Gerais

Belo Horizonte - Minas Gerais - Brasil

lucasrsilvak@ufmg.br

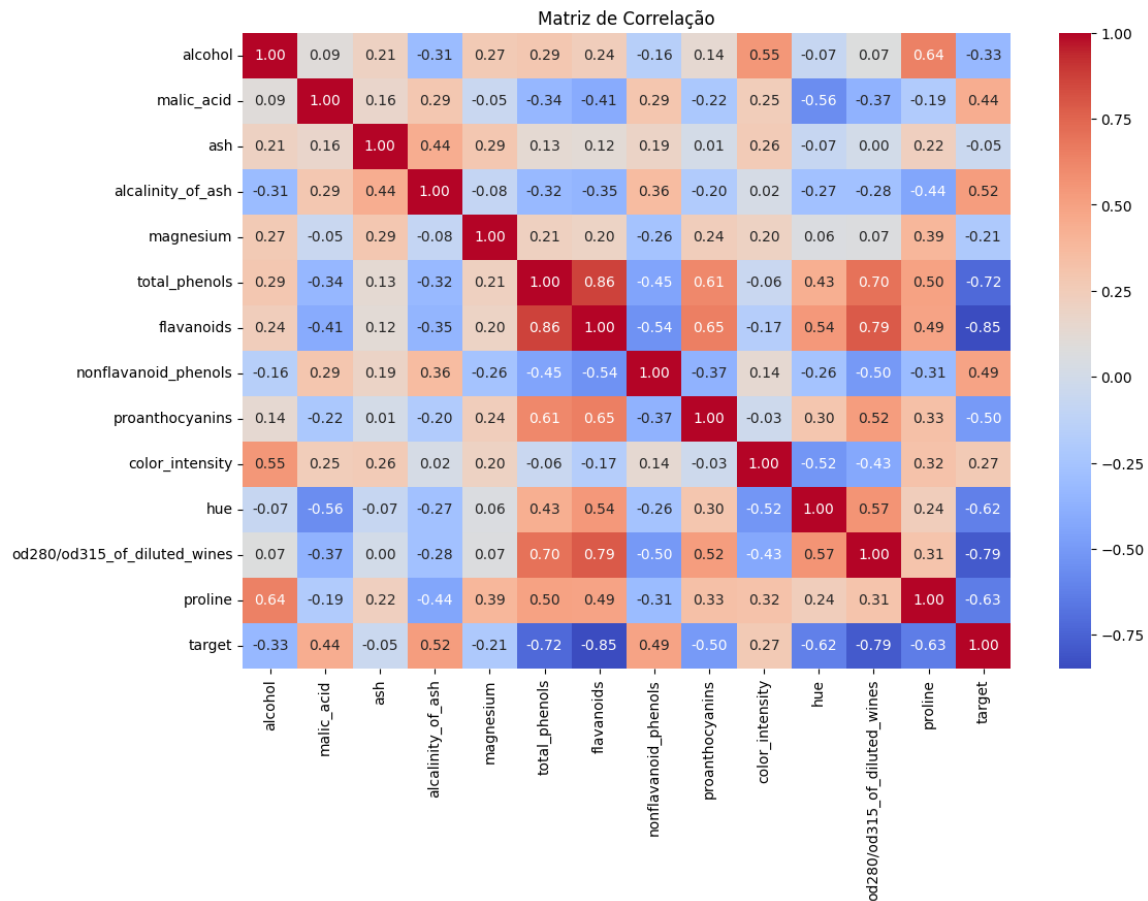
1 Carregamento de Dados

Os nomes das variáveis são: Alcohol, Malic Acid, Ash, Alcalinity of Ash, Magnesium, Total Phenols, Flavanoids, Nonflavanoid Phenols, Proanthocyanins, Color Intensity, Hue, OD280/OD315 of Diluted Wines, Proline e Target

2 Descrição Estatísticas dos Dados

Index	Count	Mean	Std	Min	25%	50%	75%	Max
Alcohol	178.0	13.00	0.81	11.03	12.36	13.05	13.68	14.83
Malic Acid	178.0	2.34	1.12	0.74	1.60	1.87	3.08	5.80
Ash	178.0	2.37	0.27	1.36	2.21	2.36	2.56	3.23
Alcalinity of Ash	178.0	19.49	3.34	10.60	17.20	19.50	21.50	30.00
Magnesium	178.0	99.74	14.28	70.00	88.00	98.00	107.00	162.00
Total Phenols	178.0	2.30	0.63	0.98	1.74	2.36	2.80	3.88
Flavanoids	178.0	2.03	1.00	0.34	1.21	2.14	2.88	5.08
Non. Phenols	178.0	0.36	0.12	0.13	0.27	0.34	0.44	0.66
Proanthocyanins	178.0	1.59	0.57	0.41	1.25	1.56	1.95	3.58
Color Intensity	178.0	5.06	2.32	1.28	3.22	4.69	6.20	13.00
Hue	178.0	0.96	0.23	0.48	0.78	0.97	1.12	1.71
OD280/OD315	178.0	2.61	0.71	1.27	1.94	2.78	3.17	4.00
Proline	178.0	746.89	314.91	278.0	500.5	673.5	985.0	1680.0
Target	178.0	0.94	0.78	0.0	0.0	1.0	2.0	2.0

3 Matriz de Correlações

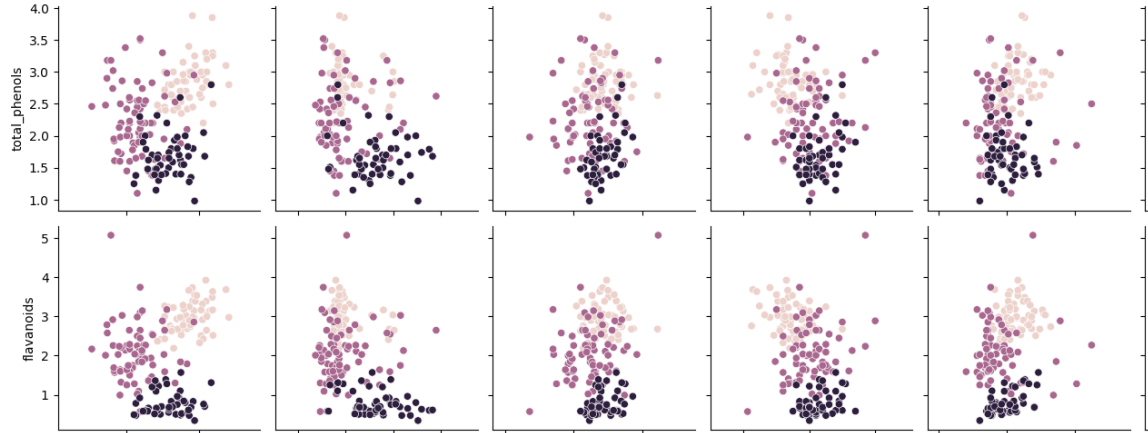


Elementos altamente correlacionados positivamente incluem Flavanoids e Phenols, Flavanoids e OD280/OD315 of Diluted Wines. Elementos altamente correlacionados negativamente incluem Flavanoids e Target, Phenols e OD280/OD315 of Diluted Wines e Target.

4 Matriz de Scatterplots

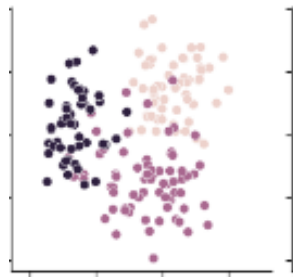
4.1 Classes Redundantes

Flavanoids e Total Phenols são classes redundantes e exibem gráficos parecidos.

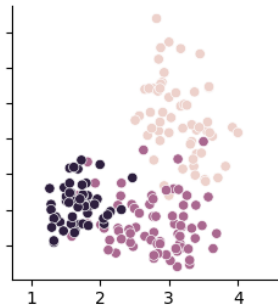


4.2 Classes Não-Redundantes

Alcohol e OD280/OD315 of Diluted Wines separa bem as classes de vinho.

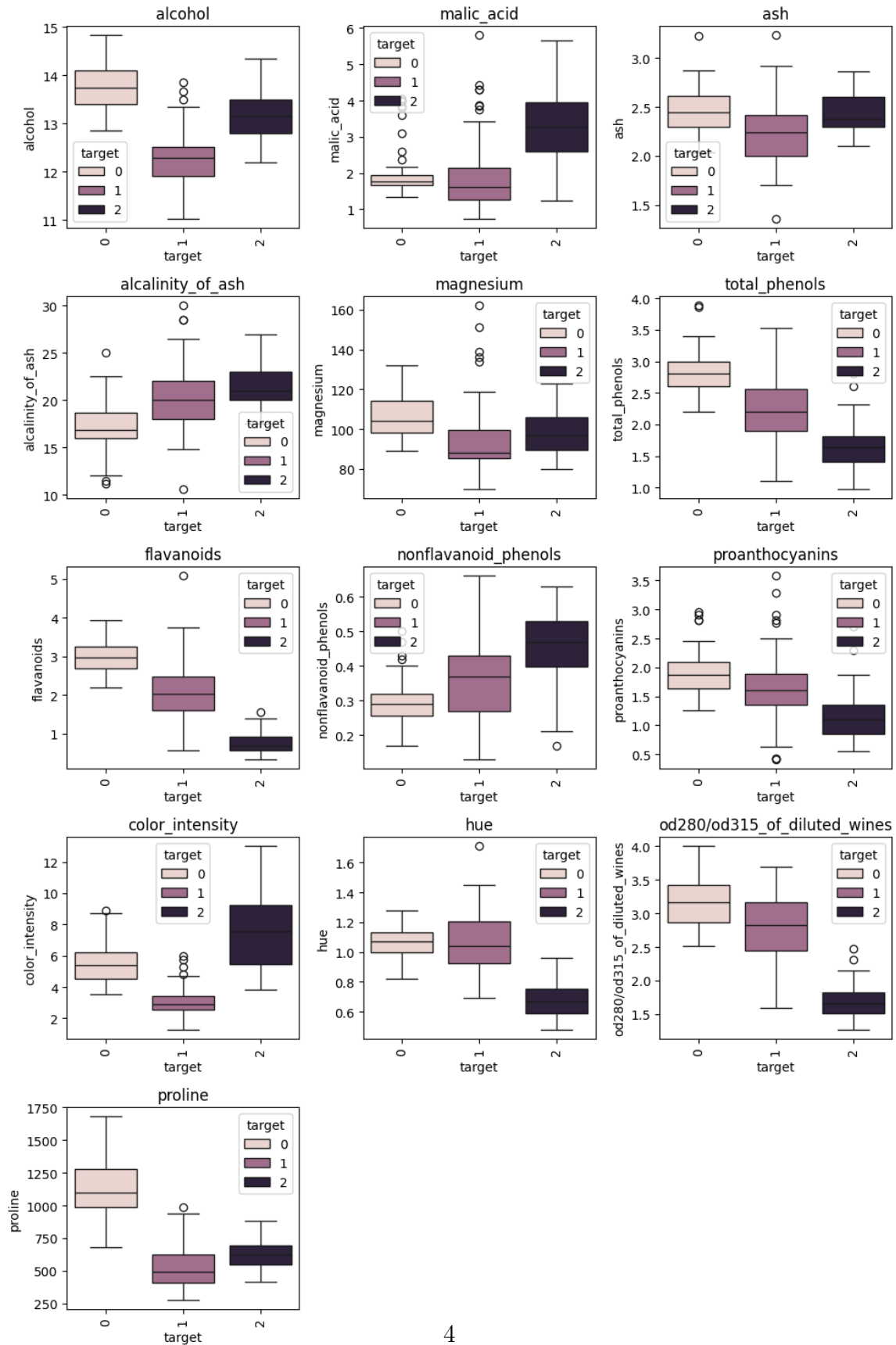


Proline e OD280/OD315 of Diluted Wines separa bem as classes de vinho.



No geral, Alcohol, Proline, OD280/OD315 of Diluted Wines separam bem as classes. Ash e Nonflavanoid Phenols separam mal as classes.

4.3 Boxplots



Podemos perceber que **Ash** e **Magnesium** se sobrepõem e por isso, são variáveis que possivelmente não ajudam muito na solução do problema, por um outro lado, **Alcohol** e **Flavanoids** possuem variáveis que se sobrepõem pouco e por isso auxiliam muito na solução do problema.

4.4 Conclusão

4.4.1 Quais variáveis apresentam alta correlação entre si?

Flavanoids e **Phenols** possuem uma alta correlação entre si e uma delas pode ser removida, nesse caso, os **Phenols** se sobrepõem mais e podem ser eliminados.

4.4.2 Há variáveis que, com base nos scatterplots e boxplots, parecem não ajudar a distinguir as classes?

Magnesium e **Ash** têm alta superposição na relação entre as variáveis e correlação quase 0 com o target, o que indica uma variável que pouco ajuda.

4.4.3 Quais variáveis você consideraria remover para otimizar o modelo de classificação, baseado nas observações feitas?

Ash, **Phenols**, **Magnesium**, **Nonflavanoid Phenols** possuem características que nos permitem inferir que uma análise dos dados sem essas variáveis ainda seria eficiente.