

Aprendizado Profundo

Professor: Julio

Nome: Alekyne Ribeiro e Ana Livia

Matricula: 541062 e 536158

Atividade 1

1. Exercício de Normalização: Considere o arquivo wine.dat. Foram feitas várias observações nos lotes de produção de vinho de uma vinícola. O teor alcólico é a segunda coluna. Normalie estes dados conforme indicado na fórmula acima.

Texto sobre a atividade

No arquivo wine.dat, que é um dataset sobre vinho que contem 12 colunas. Foi feito o processo de normalização que foi realizado em três etapas distintas. Primeiro, foi aplicada a normalização por norma constante L2, em que cada variável numérica foi dividida pela sua norma(magnitude), garantindo que os dados ficassem em uma escala comparável. Em seguida, foi realizada a normalização Min-Max, onde os dados foram escalados para um intervalo entre 0 e 1, mantendo as relações entre as variáveis. Por fim, foi aplicado o terceiro método Padronização(média=0, variância=1), transformando os dados para que tivessem média 0 e desvio padrão 1, o que é útil para modelos que assumem distribuições normais. Antes dessa atividade de aplicação de normalização, foi feita a renomeação das colunas para português. Em seguida, foi feita uma análise exploratória básica, verificando valores ausentes, mínimos, máximos, médias e desvio padrão das variáveis e por fim, a coluna ID foi removida. E para visualizar as variáveis, foram gerados um boxplot, um histograma do teor alcoólico e um heatmap da matriz de correlação, facilitando a compreensão das relações entre as variáveis.

✓ Importação das bibliotecas necessárias

```
import pandas as pd
from sklearn.metrics import confusion_matrix
import seaborn as sns
import matplotlib.pyplot as plt
```

```
import seaborn as sns
import numpy as np
```

✓ Carregamento da Base de dados

```
data = pd.read_csv('WineQT.dat')
```

✓ Análise descritiva dos dados

```
colunas_port = [['acidez fixa', 'acidez volatil', 'acido citrico', 'Açucar residual', 'Cl
```

```
data.columns = colunas_port
```

```
data.describe()
```



	acidez fixa	acidez volatil	acido citrnico	Açucar residual	Cloretos	Dioxido de enxofre livre	dio enx tot
count	1143.000000	1143.000000	1143.000000	1143.000000	1143.000000	1143.000000	1143.000000
mean	8.311111	0.531339	0.268364	2.532152	0.086933	15.615486	4.370833
std	1.747595	0.179633	0.196686	1.355917	0.047267	10.250486	3.181122
min	4.600000	0.120000	0.000000	0.900000	0.012000	1.000000	0.120000
25%	7.100000	0.392500	0.090000	1.900000	0.070000	7.000000	2.120000
50%	7.900000	0.520000	0.250000	2.200000	0.079000	13.000000	3.120000
75%	9.100000	0.640000	0.420000	2.600000	0.090000	21.000000	6.120000
max	15.900000	1.580000	1.000000	15.500000	0.611000	68.000000	28.120000

```
print(data['qualidade'].value_counts())
```



```
(qualidade,)
5      483
6      462
7      143
4       33
8       16
3         6
Name: count, dtype: int64
```

```
print('colunas do dataset')
print(data.columns)
```

```
print('Valores ausentes no dataset')
print(data.isnull().sum())
```

```
print('valor minimo')
print(data.min())
print('valor maximo')
print(data.max())
print('valor médio')
print(data.mean())
print('Desvio padrão')
print(data.std())
```



colunas do dataset

```
MultiIndex([(
            'acidez fixa',),
            (
            'acidez volatil',),
            (
            'acido citrico',),
            (
            'Açucar residual',),
            (
            'Cloretos',),
            ('Dioxido de enxofre livre',),
            ('dioxido de enxofre total',),
            (
            'Densidade',),
            (
            'pH',),
            (
            'sulfatos',),
            (
            'alcool',),
            (
            'qualidade',),
            (
            'ID',)],)
```

Valores ausentes no dataset

acidez fixa	0
acidez volatil	0
acido citrico	0
Açucar residual	0
Cloretos	0
Dioxido de enxofre livre	0
dioxido de enxofre total	0
Densidade	0
pH	0
sulfatos	0
alcool	0
qualidade	0
ID	0
dtype: int64	
valor minimo	
acidez fixa	4.60000
acidez volatil	0.12000
acido citrico	0.00000
Açucar residual	0.90000
Cloretos	0.01200
Dioxido de enxofre livre	1.00000
dioxido de enxofre total	6.00000
Densidade	0.99007
pH	2.74000
sulfatos	0.33000
alcool	8.40000
qualidade	3.00000
ID	0.00000
dtype: float64	
valor maximo	

acidez fixa	15.90000
acidez volatil	1.58000
acido citrico	1.00000
Açucar residual	15.50000
Cloretos	0.61100
Dioxido de enxofre livre	68.00000
dioxido de enxofre total	289.00000
Densidade	1.00369
pH	4.01000
sulfatos	2.00000
alcool	14.90000
qualidade	8.00000

```
# metodo 1
# Normalização por norma constante (L2)
numeric_cols = data.select_dtypes(include=[np.number]).columns
for col in numeric_cols:
    norma = np.sqrt(np.sum(data[col] ** 2))
    if norma != 0:
        data[col] = data[col] / norma

# Mostrar os dados normalizados
print("Dados normalizados:")
print(data.head())

print("\nVerificação das normas L2 após normalização:")
for col in numeric_cols:
    norma_normalizada = np.sqrt(np.sum(data[col] ** 2))
    print(f'Norma L2 de "{col}" após normalização: {norma_normalizada}')
```



Dados normalizados:

	acidez fixa	acidez volatil	acido citrico	Açucar residual	Cloretos	\
0	0.025773	0.036917	0.000000	0.019568	0.022720	
1	0.027166	0.046410	0.000000	0.026777	0.029297	
2	0.027166	0.040081	0.003556	0.023687	0.027503	
3	0.039008	0.014767	0.049791	0.019568	0.022421	
4	0.025773	0.036917	0.000000	0.019568	0.022720	

	Dioxido de enxofre livre	dioxido de enxofre total	Densidade	pH	\
0	0.017421	0.017828	0.029610	0.031321	
1	0.039593	0.035133	0.029581	0.028555	
2	0.023756	0.028316	0.029586	0.029090	
3	0.026923	0.031462	0.029616	0.028198	
4	0.017421	0.017828	0.029610	0.031321	

	sulfatos	alcool	qualidade	ID
0	0.024380	0.026485	0.025882	0.000000
1	0.029604	0.027612	0.025882	0.000032
2	0.028298	0.027612	0.025882	0.000064
3	0.025251	0.027612	0.031058	0.000096
4	0.024380	0.026485	0.025882	0.000127

```
Verificação das normas L2 após normalização:
Norma L2 de "('acidez fixa',)" após normalização: 1.0
Norma L2 de "('acidez volatil',)" após normalização: 1.0
Norma L2 de "('acido citrico',)" após normalização: 1.0
Norma L2 de "('Açucar residual',)" após normalização: 1.0
```

```

Norma L2 de "('Cloretos',)" após normalização: 1.0
Norma L2 de "('Dioxido de enxofre livre',)" após normalização: 1.0
Norma L2 de "('dioxido de enxofre total',)" após normalização: 1.0
Norma L2 de "('Densidade',)" após normalização: 1.0000000000000002
Norma L2 de "('pH',)" após normalização: 1.0
Norma L2 de "('sulfatos',)" após normalização: 1.0
Norma L2 de "('alcool',)" após normalização: 1.0
Norma L2 de "('qualidade',)" após normalização: 0.9999999999999999
Norma L2 de "('ID',)" após normalização: 1.0

```

```
# metodo 2
```

```
# min-max scaled
```

```
max_val, min_val = 1, 0
```

```
data_std = (data - data.min(axis=0)) / (data.max(axis=0) - data.min(axis=0))
```

```
data = data_std * (max_val - min_val) + min_val
```

```
print('dados com min-max:\n\n\n', data)
```



```
dados com min-max:
```

	acidez fixa	acidez volatil	acido citrico	Açucar residual	Cloretos \
0	0.247788	0.397260	0.00	0.068493	0.106845
1	0.283186	0.520548	0.00	0.116438	0.143573
2	0.283186	0.438356	0.04	0.095890	0.133556
3	0.584071	0.109589	0.56	0.068493	0.105175
4	0.247788	0.397260	0.00	0.068493	0.106845
...
1138	0.150442	0.267123	0.13	0.095890	0.106845
1139	0.194690	0.342466	0.08	0.068493	0.093489
1140	0.141593	0.328767	0.08	0.075342	0.130217
1141	0.115044	0.294521	0.10	0.089041	0.083472
1142	0.115044	0.359589	0.12	0.075342	0.105175

	Dioxido de enxofre livre	dioxido de enxofre total	Densidade	pH \
0	0.149254	0.098940	0.567548	0.606299
1	0.358209	0.215548	0.494126	0.362205
2	0.208955	0.169611	0.508811	0.409449
3	0.238806	0.190813	0.582232	0.330709
4	0.149254	0.098940	0.567548	0.606299
...
1138	0.417910	0.120141	0.416300	0.535433
1139	0.402985	0.113074	0.472834	0.535433
1140	0.462687	0.134276	0.354626	0.559055
1141	0.567164	0.159011	0.370778	0.614173
1142	0.462687	0.134276	0.396476	0.653543

	sulfatos	alcool	qualidade
0	0.137725	0.153846	0.4
1	0.209581	0.215385	0.4
2	0.191617	0.215385	0.4
3	0.149701	0.215385	0.6
4	0.137725	0.153846	0.4
...
1138	0.251497	0.400000	0.6
1139	0.293413	0.169231	0.6
1140	0.149701	0.323077	0.4
1141	0.257485	0.430769	0.6
1142	0.227545	0.276923	0.4

```
[1143 rows x 12 columns]
```

```
# metodo 3
# Padronização da variável (média=0, variância=1)
numeric_cols = data.select_dtypes(include=[np.number]).columns
for col in numeric_cols:
    media = np.mean(data[col])
    desvio_padrao = np.std(data[col])
    if desvio_padrao != 0:
        data[col] = (data[col] - media) / desvio_padrao

# Mostrar os dados padronizados
print("Dados padronizados:")
print(data.head())

print("\nVerificação da média e do desvio padrão após padronização:")
for col in numeric_cols:
    media_padronizada = np.mean(data[col])
    desvio_padrao_padronizado = np.std(data[col])
    print(f'Média de "{col}" após padronização: {media_padronizada:.4f}')
    print(f'Desvio padrão de "{col}" após padronização: {desvio_padrao_padronizado:.4f}')
```



Dados padronizados:

	acidez fixa	acidez volatil	acido citrico	Açucar residual	Cloretos	\
0	-0.521580	0.939332	-1.365027	-0.466421	-0.231395	
1	-0.292593	1.941813	-1.365027	0.050060	0.234247	
2	-0.292593	1.273492	-1.161568	-0.171289	0.107253	
3	1.653789	-1.399789	1.483400	-0.466421	-0.252560	
4	-0.521580	0.939332	-1.365027	-0.466421	-0.231395	

	Dioxido de enxofre livre	dioxido de enxofre total	Densidade	pH	\
0	-0.450467	-0.363610	0.555854	1.270695	
1	0.915920	0.643477	0.036165	-0.708928	
2	-0.060071	0.246745	0.140103	-0.325775	
3	0.135127	0.429852	0.659792	-0.964363	
4	-0.450467	-0.363610	0.555854	1.270695	

	sulfatos	alcool	qualidade	ID
0	-0.573658	-0.963382	-0.815724	-1.735618
1	0.130881	-0.593601	-0.815724	-1.733462
2	-0.045254	-0.593601	-0.815724	-1.731306
3	-0.456235	-0.593601	0.425784	-1.729150
4	-0.573658	-0.963382	-0.815724	-1.726993

```
Verificação da média e do desvio padrão após padronização:
Média de "('acidez fixa',)" após padronização: -0.0000
Desvio padrão de "('acidez fixa',)" após padronização: 1.0000
Média de "('acidez volatil',)" após padronização: 0.0000
Desvio padrão de "('acidez volatil',)" após padronização: 1.0000
Média de "('acido citrico',)" após padronização: 0.0000
Desvio padrão de "('acido citrico',)" após padronização: 1.0000
Média de "('Açucar residual',)" após padronização: 0.0000
Desvio padrão de "('Açucar residual',)" após padronização: 1.0000
Média de "('Cloretos',)" após padronização: 0.0000
Desvio padrão de "('Cloretos',)" após padronização: 1.0000
Média de "('Dioxido de enxofre livre',)" após padronização: -0.0000
```

```

Desvio padrão de "('Dioxido de enxofre livre',)" após padronização: 1.0000
Média de "('dioxido de enxofre total',)" após padronização: 0.0000
Desvio padrão de "('dioxido de enxofre total',)" após padronização: 1.0000
Média de "('Densidade',)" após padronização: 0.0000
Desvio padrão de "('Densidade',)" após padronização: 1.0000
Média de "('pH',)" após padronização: -0.0000
Desvio padrão de "('pH',)" após padronização: 1.0000
Média de "('sulfatos',)" após padronização: 0.0000
Desvio padrão de "('sulfatos',)" após padronização: 1.0000
Média de "('alcool',)" após padronização: -0.0000
Desvio padrão de "('alcool',)" após padronização: 1.0000
Média de "('qualidade',)" após padronização: 0.0000
Desvio padrão de "('qualidade',)" após padronização: 1.0000
Média de "('ID',)" após padronização: -0.0000
Desvio padrão de "('ID',)" após padronização: 1.0000

```

```
data = data.drop(columns='ID', errors='ignore')
```

```
print(data)
```

```

↩
    acidez fixa  acidez volatil  acido citrico  Açucar residual  Cloretos  \
0      0.247788      0.397260          0.00      0.068493  0.106845
1      0.283186      0.520548          0.00      0.116438  0.143573
2      0.283186      0.438356          0.04      0.095890  0.133556
3      0.584071      0.109589          0.56      0.068493  0.105175
4      0.247788      0.397260          0.00      0.068493  0.106845
...      ...      ...      ...      ...      ...
1138     0.150442      0.267123          0.13      0.095890  0.106845
1139     0.194690      0.342466          0.08      0.068493  0.093489
1140     0.141593      0.328767          0.08      0.075342  0.130217
1141     0.115044      0.294521          0.10      0.089041  0.083472
1142     0.115044      0.359589          0.12      0.075342  0.105175

    Dioxido de enxofre livre  dioxido de enxofre total  Densidade      pH  \
0              0.149254              0.098940  0.567548  0.606299
1              0.358209              0.215548  0.494126  0.362205
2              0.208955              0.169611  0.508811  0.409449
3              0.238806              0.190813  0.582232  0.330709
4              0.149254              0.098940  0.567548  0.606299
...      ...      ...      ...      ...
1138             0.417910              0.120141  0.416300  0.535433
1139             0.402985              0.113074  0.472834  0.535433
1140             0.462687              0.134276  0.354626  0.559055
1141             0.567164              0.159011  0.370778  0.614173
1142             0.462687              0.134276  0.396476  0.653543

    sulfatos  alcool  qualidade
0    0.137725  0.153846      0.4
1    0.209581  0.215385      0.4
2    0.191617  0.215385      0.4
3    0.149701  0.215385      0.6
4    0.137725  0.153846      0.4
...      ...      ...
1138  0.251497  0.400000      0.6
1139  0.293413  0.169231      0.6
1140  0.149701  0.323077      0.4
1141  0.257485  0.430769      0.6
1142  0.227545  0.276923      0.4

```

```
[1143 rows x 12 columns]
<ipython-input-21-3a1dedea2c46>:1: PerformanceWarning: dropping on a non-lexsorted mu
data = data.drop(columns='ID', errors='ignore')
```

```
from scipy.stats import zscore
```

```
# Calculando o Z-Score
```

```
z_scores = np.abs(zscore(data.select_dtypes(include=[np.number])))
outliers = (z_scores > 3).sum(axis=0)
print("Outliers por variável:\n", outliers)
```

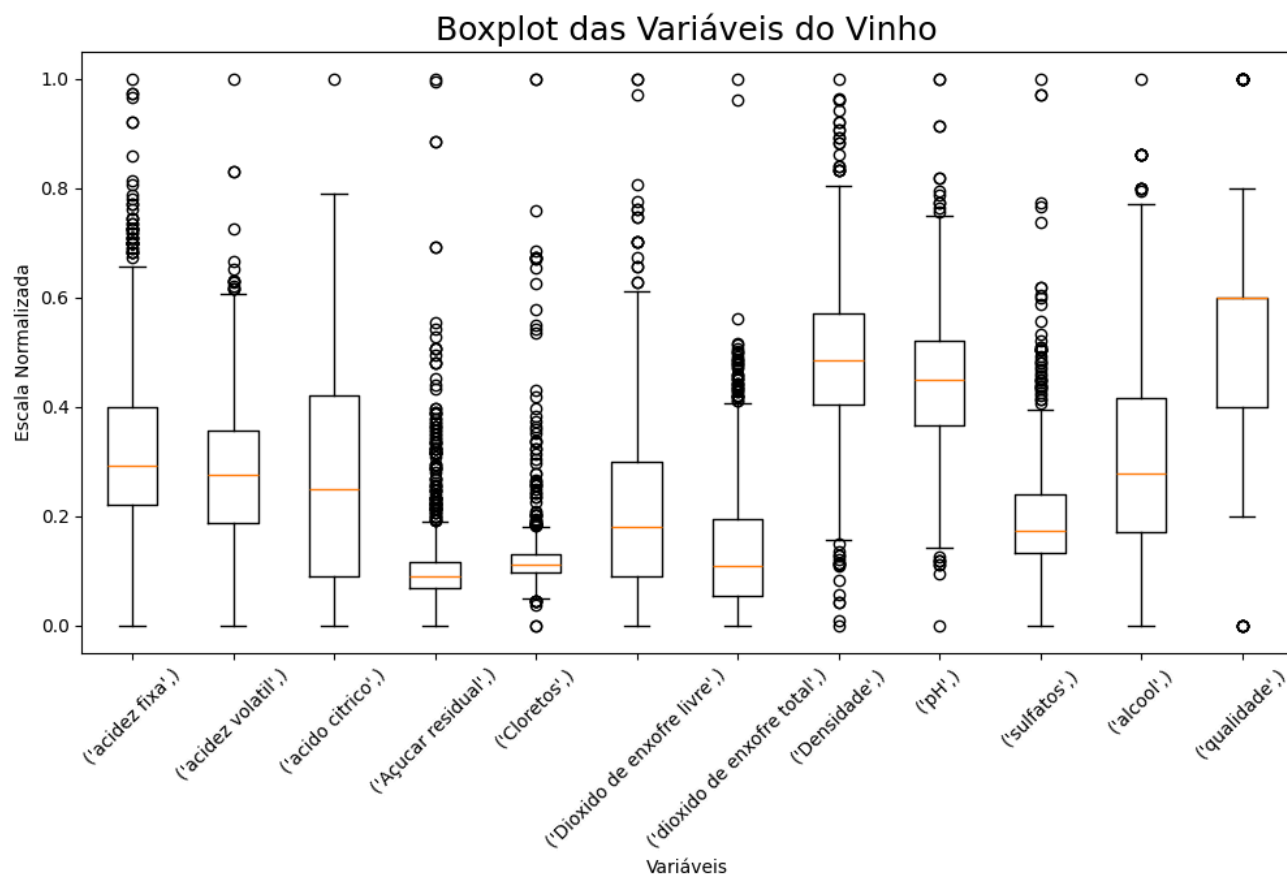
```
⇒ Outliers por variável:
  acidez fixa                9
  acidez volatil            5
  acido citrico              1
  Açúcar residual          23
  Cloretos                  21
  Dioxido de enxofre livre  13
  dioxido de enxofre total  10
  Densidade                 12
  pH                        5
  sulfatos                  21
  alcool                    7
  qualidade                  6
  ID                        0
  dtype: int64
```

```
cor_matrix = data.corr().abs()
upper_triangle = cor_matrix.where(np.triu(np.ones(cor_matrix.shape), k=1).astype(bool))
to_drop = [column for column in upper_triangle.columns if any(upper_triangle[column] > 0.
data = data.drop(columns=to_drop)
print(f'Variáveis removidas por colinearidade: {to_drop}')
```

```
⇒ Variáveis removidas por colinearidade: []
```

✓ EDA - Analise Exploratoria dos Dados

```
# Plotar o boxplot com rótulos das colunas no eixo x
plt.figure(figsize=(12, 6))
plt.boxplot(data, labels=data.columns)
plt.title("Boxplot das Variáveis do Vinho", loc="center", fontsize=18)
plt.xlabel("Variáveis")
plt.ylabel("Escala Normalizada")
plt.xticks(rotation=45)
plt.show()
```

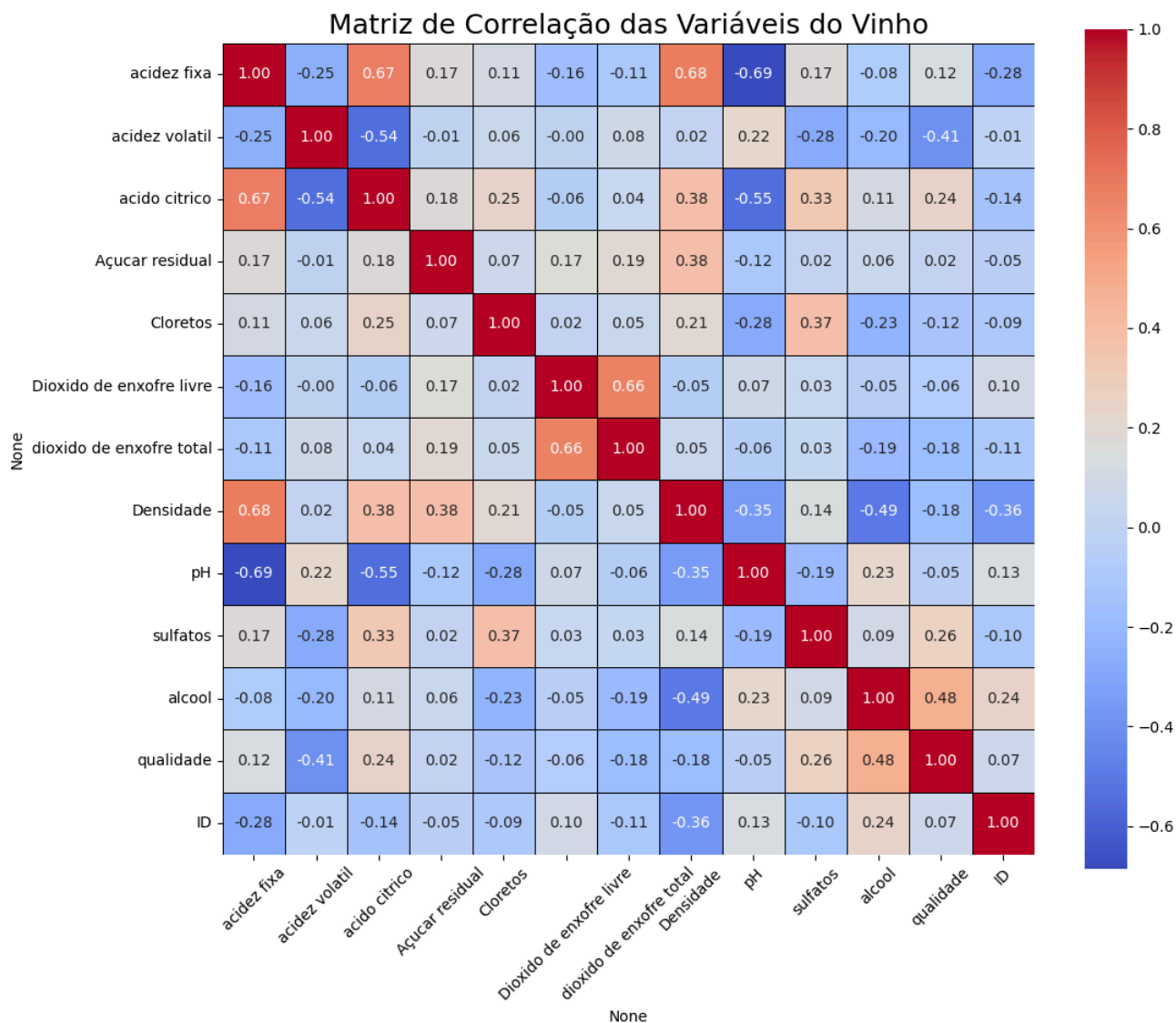



```
# Calcular a matriz de correlação
correlation_matrix = data.corr()

plt.figure(figsize=(12, 10))

sns.heatmap(correlation_matrix, annot=True, fmt=".2f", cmap='coolwarm',
            square=True, cbar=True, linewidths=.5, linecolor='black')

plt.title("Matriz de Correlação das Variáveis do Vinho", fontsize=18)
plt.xticks(rotation=45)
plt.yticks(rotation=0)
plt.show()
```



```
data['alcool'].hist(bins=20, color='lightblue', edgecolor='black')
plt.title('Distribuição do Alcool')
plt.xlabel('Teor Alcoólico')
plt.ylabel('Frequência')
plt.show()
```

