11-3-logistic-regression

April 27, 2024

Logistic Regression Analysis for the Wine Dataset

[5]:
```
pip install ucimlrepo
```

```
Requirement already satisfied: ucimlrepo in /usr/local/lib/python3.10/dist-
packages (0.0.6)
```

[6]:
```
pip install hvplot
```

```
Collecting hvplot
  Downloading hvplot-0.9.2-py2.py3-none-any.whl (1.8 MB)
                           1.8/1.8 MB
14.4 MB/s eta 0:00:00
Requirement already satisfied: bokeh>=1.0.0 in
/usr/local/lib/python3.10/dist-packages (from hvplot) (3.3.4)
Requirement already satisfied: colorcet>=2 in /usr/local/lib/python3.10/dist-
packages (from hvplot) (3.1.0)
Requirement already satisfied: holoviews>=1.11.0 in
/usr/local/lib/python3.10/dist-packages (from hvplot) (1.17.1)
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages
(from hvplot) (2.0.3)
Requirement already satisfied: numpy>=1.15 in /usr/local/lib/python3.10/dist-
packages (from hvplot) (1.25.2)
Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-
packages (from hvplot) (24.0)
Requirement already satisfied: panel>=0.11.0 in /usr/local/lib/python3.10/dist-
packages (from hvplot) (1.3.8)
Requirement already satisfied: param<3.0,>=1.12.0 in
/usr/local/lib/python3.10/dist-packages (from hvplot) (2.1.0)
Requirement already satisfied: Jinja2>=2.9 in /usr/local/lib/python3.10/dist-
packages (from bokeh>=1.0.0->hvplot) (3.1.3)
Requirement already satisfied: contourpy>=1 in /usr/local/lib/python3.10/dist-
packages (from bokeh>=1.0.0->hvplot) (1.2.1)
Requirement already satisfied: pillow>=7.1.0 in /usr/local/lib/python3.10/dist-
packages (from bokeh>=1.0.0->hvplot) (9.4.0)
Requirement already satisfied: PyYAML>=3.10 in /usr/local/lib/python3.10/dist-
packages (from bokeh>=1.0.0->hvplot) (6.0.1)
Requirement already satisfied: tornado>=5.1 in /usr/local/lib/python3.10/dist-
packages (from bokeh>=1.0.0->hvplot) (6.3.3)
```

```
Requirement already satisfied: xyzservices>=2021.09.1 in
/usr/local/lib/python3.10/dist-packages (from bokeh>=1.0.0->hvplot) (2024.4.0)
Requirement already satisfied: pyviz-comms>=0.7.4 in
/usr/local/lib/python3.10/dist-packages (from holoviews>=1.11.0->hvplot) (3.0.2)
Requirement already satisfied: python-dateutil>=2.8.2 in
/usr/local/lib/python3.10/dist-packages (from pandas->hvplot) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-
packages (from pandas->hvplot) (2023.4)
Requirement already satisfied: tzdata>=2022.1 in /usr/local/lib/python3.10/dist-
packages (from pandas->hvplot) (2024.1)
Requirement already satisfied: markdown in /usr/local/lib/python3.10/dist-
packages (from panel>=0.11.0->hvplot) (3.6)
Requirement already satisfied: markdown-it-py in /usr/local/lib/python3.10/dist-
packages (from panel>=0.11.0->hvplot) (3.0.0)
Requirement already satisfied: linkify-it-py in /usr/local/lib/python3.10/dist-
packages (from panel>=0.11.0->hvplot) (2.0.3)
Requirement already satisfied: mdit-py-plugins in
/usr/local/lib/python3.10/dist-packages (from panel>=0.11.0->hvplot) (0.4.0)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-
packages (from panel>=0.11.0->hvplot) (2.31.0)
Requirement already satisfied: tqdm>=4.48.0 in /usr/local/lib/python3.10/dist-
packages (from panel>=0.11.0->hvplot) (4.66.2)
Requirement already satisfied: bleach in /usr/local/lib/python3.10/dist-packages
(from panel>=0.11.0->hvplot) (6.1.0)
Requirement already satisfied: typing-extensions in
/usr/local/lib/python3.10/dist-packages (from panel>=0.11.0->hvplot) (4.11.0)
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.10/dist-packages (from Jinja2>=2.9->bokeh>=1.0.0->hvplot)
(2.1.5)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-
packages (from python-dateutil>=2.8.2->pandas->hvplot) (1.16.0)
Requirement already satisfied: webencodings in /usr/local/lib/python3.10/dist-
packages (from bleach->panel>=0.11.0->hvplot) (0.5.1)
Requirement already satisfied: uc-micro-py in /usr/local/lib/python3.10/dist-
packages (from linkify-it-py->panel>=0.11.0->hvplot) (1.0.3)
Requirement already satisfied: mdurl~=0.1 in /usr/local/lib/python3.10/dist-
packages (from markdown-it-py->panel>=0.11.0->hvplot) (0.1.2)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.10/dist-packages (from requests->panel>=0.11.0->hvplot)
(3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-
packages (from requests->panel>=0.11.0->hvplot) (3.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.10/dist-packages (from requests->panel>=0.11.0->hvplot)
(2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.10/dist-packages (from requests->panel>=0.11.0->hvplot)
(2024.2.2)
```

```
Installing collected packages: hvplot
Successfully installed hvplot-0.9.2
```

[7]:
```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import hvplot.pandas
from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.linear_model import LinearRegression
%matplotlib inline
```

[8]:
```python
from ucimlrepo import fetch_ucirepo

# fetch dataset
wine = fetch_ucirepo(id=109)

# data (as pandas dataframes)
X = wine.data.features
y = wine.data.targets

# metadata
print(wine.metadata)

# variable information
print(wine.variables)
```


```
{'uci_id': 109, 'name': 'Wine', 'repository_url':
'https://archive.ics.uci.edu/dataset/109/wine', 'data_url':
'https://archive.ics.uci.edu/static/public/109/data.csv', 'abstract': 'Using
chemical analysis to determine the origin of wines', 'area': 'Physics and
Chemistry', 'tasks': ['Classification'], 'characteristics': ['Tabular'],
'num_instances': 178, 'num_features': 13, 'feature_types': ['Integer', 'Real'],
'demographics': [], 'target_col': ['class'], 'index_col': None,
'has_missing_values': 'no', 'missing_values_symbol': None,
'year_of_dataset_creation': 1992, 'last_updated': 'Mon Aug 28 2023',
'dataset_doi': '10.24432/C5PC7J', 'creators': ['Stefan Aeberhard', 'M. Forina'],
'intro_paper': {'title': 'Comparative analysis of statistical pattern
recognition methods in high dimensional settings', 'authors': 'S. Aeberhard, D.
Coomans, O. Vel', 'published_in': 'Pattern Recognition', 'year': 1994, 'url': 'h
ttps://www.semanticscholar.org/paper/83dc3e4030d7b9fbdbb4bde03ce12ab70ca10528',
'doi': '10.1016/0031-3203(94)90145-7'}, 'additional_info': {'summary': 'These
data are the results of a chemical analysis of wines grown in the same region in
Italy but derived from three different cultivars. The analysis determined the
quantities of 13 constituents found in each of the three types of wines.
\r\n\r\nI think that the initial data set had around 30 variables, but for some
reason I only have the 13 dimensional version. I had a list of what the 30 or so
```

```
variables were, but a.)  I lost it, and b.), I would not know which 13 variables
are included in the set.\r\n\r\nThe attributes are (dontated by Riccardo Leardi,
riclea@anchem.unige.it )\r\n1) Alcohol\r\n2) Malic acid\r\n3) Ash\r\n4)
Alcalinity of ash  \r\n5) Magnesium\r\n6) Total phenols\r\n7) Flavanoids\r\n8)
Nonflavanoid phenols\r\n9) Proanthocyanins\r\n10)Color
intensity\r\n11)Hue\r\n12)OD280/OD315 of diluted wines\r\n13)Proline \r\n\r\nIn
a classification context, this is a well posed problem with "well behaved" class
structures. A good data set for first testing of a new classifier, but not very
challenging.           ', 'purpose': 'test', 'funded_by': None,
'instances_represent': None, 'recommended_data_splits': None, 'sensitive_data':
None, 'preprocessing_description': None, 'variable_info': 'All attributes are
continuous\r\n\t\r\nNo statistics available, but suggest to standardise
variables for certain uses (e.g. for us with classifiers which are NOT scale
invariant)\r\n\r\nNOTE: 1st attribute is class identifier (1-3)', 'citation':
None}}
                            name     role         type demographic  \
0                          class   Target  Categorical        None
1                        Alcohol  Feature   Continuous        None
2                      Malicacid  Feature   Continuous        None
3                            Ash  Feature   Continuous        None
4              Alcalinity_of_ash  Feature   Continuous        None
5                      Magnesium  Feature      Integer        None
6                  Total_phenols  Feature   Continuous        None
7                     Flavanoids  Feature   Continuous        None
8           Nonflavanoid_phenols  Feature   Continuous        None
9                Proanthocyanins  Feature   Continuous        None
10               Color_intensity  Feature   Continuous        None
11                           Hue  Feature   Continuous        None
12  0D280_0D315_of_diluted_wines  Feature   Continuous        None
13                       Proline  Feature      Integer        None

   description units missing_values
0         None  None             no
1         None  None             no
2         None  None             no
3         None  None             no
4         None  None             no
5         None  None             no
6         None  None             no
7         None  None             no
8         None  None             no
9         None  None             no
10        None  None             no
11        None  None             no
12        None  None             no
13        None  None             no
```

[10]:
```
df = pd.concat([X, y], axis = 1)
df
```

[10]:
```
     Alcohol  Malicacid   Ash  Alcalinity_of_ash  Magnesium  Total_phenols  \
0      14.23       1.71  2.43               15.6        127           2.80
1      13.20       1.78  2.14               11.2        100           2.65
2      13.16       2.36  2.67               18.6        101           2.80
3      14.37       1.95  2.50               16.8        113           3.85
4      13.24       2.59  2.87               21.0        118           2.80
..       …          …     …                 …          …             …
173    13.71       5.65  2.45               20.5         95           1.68
174    13.40       3.91  2.48               23.0        102           1.80
175    13.27       4.28  2.26               20.0        120           1.59
176    13.17       2.59  2.37               20.0        120           1.65
177    14.13       4.10  2.74               24.5         96           2.05

     Flavanoids  Nonflavanoid_phenols  Proanthocyanins  Color_intensity   Hue  \
0          3.06                  0.28             2.29             5.64  1.04
1          2.76                  0.26             1.28             4.38  1.05
2          3.24                  0.30             2.81             5.68  1.03
3          3.49                  0.24             2.18             7.80  0.86
4          2.69                  0.39             1.82             4.32  1.04
..          …                     …                …                …     …
173        0.61                  0.52             1.06             7.70  0.64
174        0.75                  0.43             1.41             7.30  0.70
175        0.69                  0.43             1.35            10.20  0.59
176        0.68                  0.53             1.46             9.30  0.60
177        0.76                  0.56             1.35             9.20  0.61

     0D280_0D315_of_diluted_wines  Proline  class
0                            3.92     1065      1
1                            3.40     1050      1
2                            3.17     1185      1
3                            3.45     1480      1
4                            2.93      735      1
..                            …        …      …
173                          1.74      740      3
174                          1.56      750      3
175                          1.56      835      3
176                          1.62      840      3
177                          1.60      560      3

[178 rows x 14 columns]
```

[11]:
```
df.columns
```

[11]:
```
Index(['Alcohol', 'Malicacid', 'Ash', 'Alcalinity_of_ash', 'Magnesium',
       'Total_phenols', 'Flavanoids', 'Nonflavanoid_phenols',
       'Proanthocyanins', 'Color_intensity', 'Hue',
       '0D280_0D315_of_diluted_wines', 'Proline', 'class'],
      dtype='object')
```

[12]:
```
df.head(20)
```

[12]:

	Alcohol	Malicacid	Ash	Alcalinity_of_ash	Magnesium	Total_phenols \
0	14.23	1.71	2.43	15.6	127	2.80
1	13.20	1.78	2.14	11.2	100	2.65
2	13.16	2.36	2.67	18.6	101	2.80
3	14.37	1.95	2.50	16.8	113	3.85
4	13.24	2.59	2.87	21.0	118	2.80
5	14.20	1.76	2.45	15.2	112	3.27
6	14.39	1.87	2.45	14.6	96	2.50
7	14.06	2.15	2.61	17.6	121	2.60
8	14.83	1.64	2.17	14.0	97	2.80
9	13.86	1.35	2.27	16.0	98	2.98
10	14.10	2.16	2.30	18.0	105	2.95
11	14.12	1.48	2.32	16.8	95	2.20
12	13.75	1.73	2.41	16.0	89	2.60
13	14.75	1.73	2.39	11.4	91	3.10
14	14.38	1.87	2.38	12.0	102	3.30
15	13.63	1.81	2.70	17.2	112	2.85
16	14.30	1.92	2.72	20.0	120	2.80
17	13.83	1.57	2.62	20.0	115	2.95
18	14.19	1.59	2.48	16.5	108	3.30
19	13.64	3.10	2.56	15.2	116	2.70

	Flavanoids	Nonflavanoid_phenols	Proanthocyanins	Color_intensity	Hue \
0	3.06	0.28	2.29	5.64	1.04
1	2.76	0.26	1.28	4.38	1.05
2	3.24	0.30	2.81	5.68	1.03
3	3.49	0.24	2.18	7.80	0.86
4	2.69	0.39	1.82	4.32	1.04
5	3.39	0.34	1.97	6.75	1.05
6	2.52	0.30	1.98	5.25	1.02
7	2.51	0.31	1.25	5.05	1.06
8	2.98	0.29	1.98	5.20	1.08
9	3.15	0.22	1.85	7.22	1.01
10	3.32	0.22	2.38	5.75	1.25
11	2.43	0.26	1.57	5.00	1.17
12	2.76	0.29	1.81	5.60	1.15
13	3.69	0.43	2.81	5.40	1.25
14	3.64	0.29	2.96	7.50	1.20
15	2.91	0.30	1.46	7.30	1.28

```
16        3.14                    0.33                1.97                6.20   1.07
17        3.40                    0.40                1.72                6.60   1.13
18        3.93                    0.32                1.86                8.70   1.23
19        3.03                    0.17                1.66                5.10   0.96

    OD280_OD315_of_diluted_wines  Proline  class
0                           3.92     1065      1
1                           3.40     1050      1
2                           3.17     1185      1
3                           3.45     1480      1
4                           2.93      735      1
5                           2.85     1450      1
6                           3.58     1290      1
7                           3.58     1295      1
8                           2.85     1045      1
9                           3.55     1045      1
10                          3.17     1510      1
11                          2.82     1280      1
12                          2.90     1320      1
13                          2.73     1150      1
14                          3.00     1547      1
15                          2.88     1310      1
16                          2.65     1280      1
17                          2.57     1130      1
18                          2.82     1680      1
19                          3.36      845      1
```

[20]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 178 entries, 0 to 177
Data columns (total 14 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   Alcohol                       178 non-null    float64
 1   Malicacid                     178 non-null    float64
 2   Ash                           178 non-null    float64
 3   Alcalinity_of_ash             178 non-null    float64
 4   Magnesium                     178 non-null    int64
 5   Total_phenols                 178 non-null    float64
 6   Flavanoids                    178 non-null    float64
 7   Nonflavanoid_phenols          178 non-null    float64
 8   Proanthocyanins               178 non-null    float64
 9   Color_intensity               178 non-null    float64
 10  Hue                           178 non-null    float64
 11  OD280_OD315_of_diluted_wines  178 non-null    float64
 12  Proline                       178 non-null    int64
```

```
 13  class                          178 non-null    int64
dtypes: float64(11), int64(3)
memory usage: 19.6 KB
```

[34]:
```python
missing_values = df.isnull().sum()
print(missing_values)
```

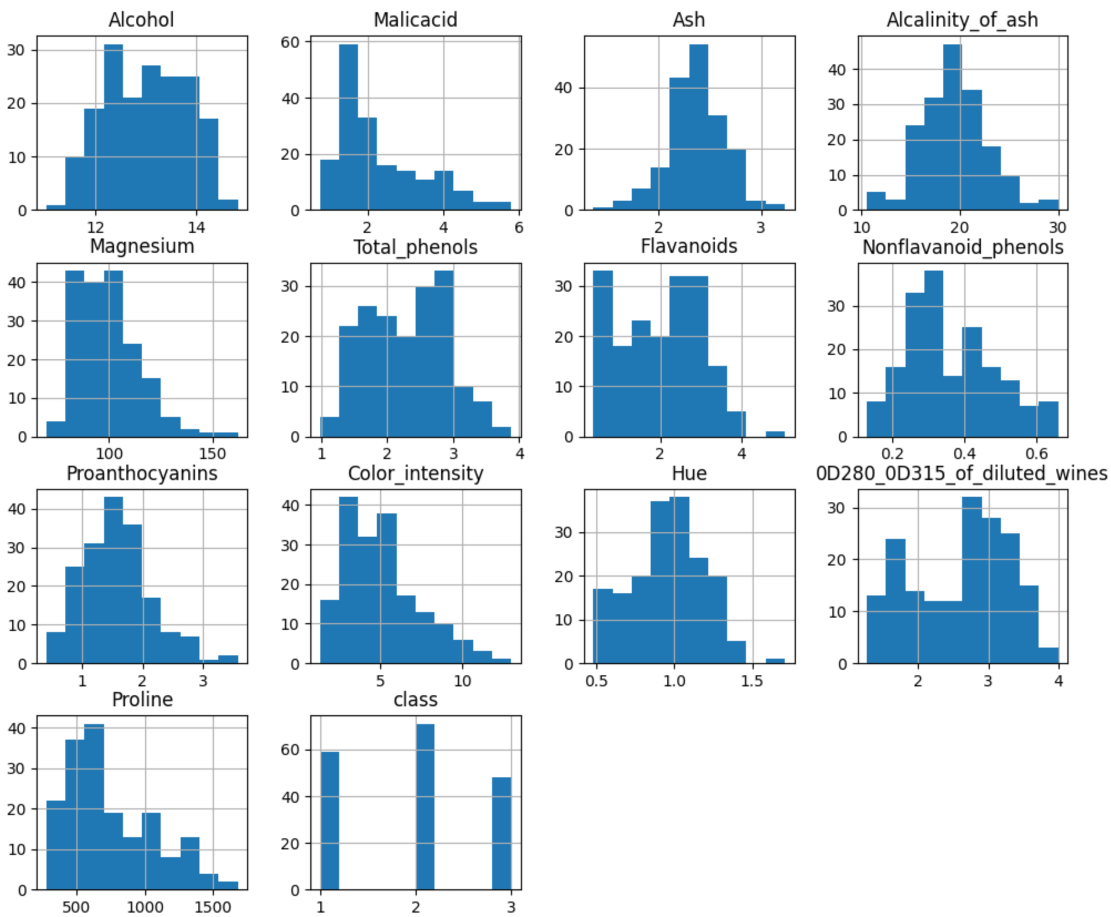
```
Alcohol                         0
Malicacid                       0
Ash                             0
Alcalinity_of_ash               0
Magnesium                       0
Total_phenols                   0
Flavanoids                      0
Nonflavanoid_phenols            0
Proanthocyanins                 0
Color_intensity                 0
Hue                             0
0D280_0D315_of_diluted_wines    0
Proline                         0
class                           0
dtype: int64
```

[35]:
```python
summary_statistics = df.describe()
print(summary_statistics)
```

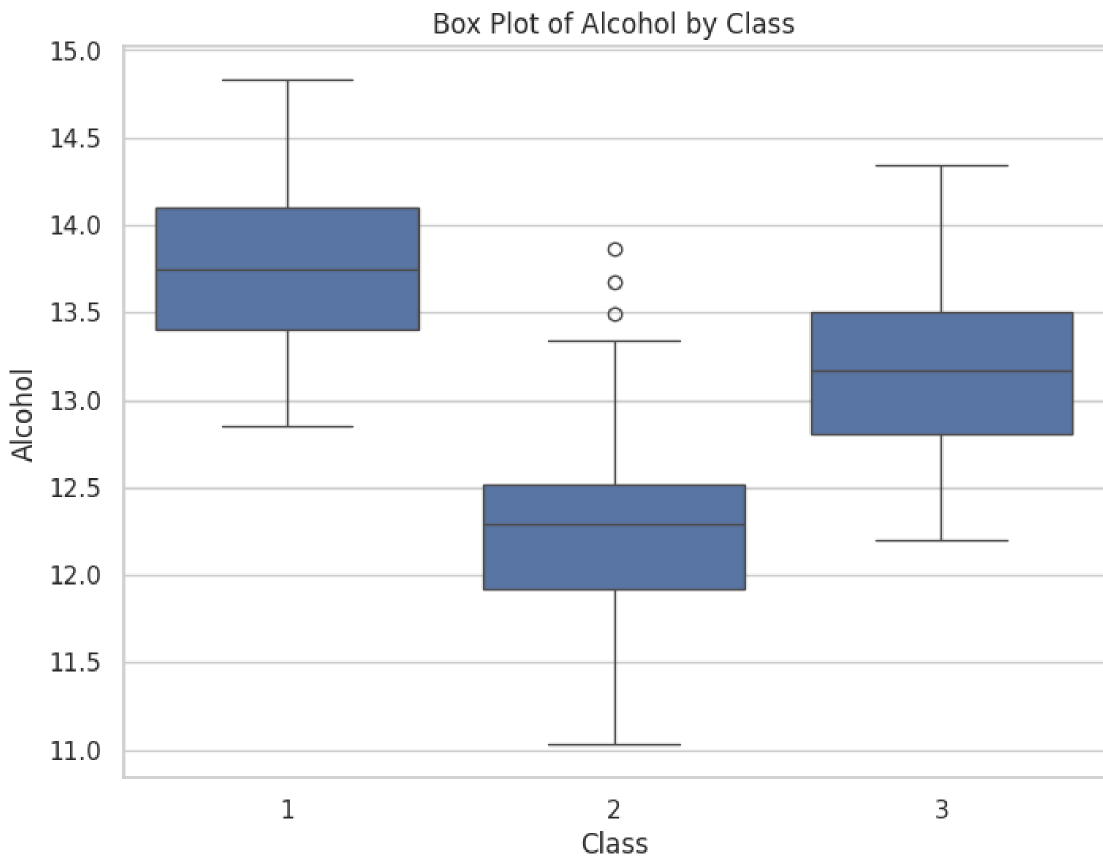
```
          Alcohol   Malicacid         Ash  Alcalinity_of_ash   Magnesium  \
count  178.000000  178.000000  178.000000         178.000000  178.000000
mean    13.000618    2.336348    2.366517          19.494944   99.741573
std      0.811827    1.117146    0.274344           3.339564   14.282484
min     11.030000    0.740000    1.360000          10.600000   70.000000
25%     12.362500    1.602500    2.210000          17.200000   88.000000
50%     13.050000    1.865000    2.360000          19.500000   98.000000
75%     13.677500    3.082500    2.557500          21.500000  107.000000
max     14.830000    5.800000    3.230000          30.000000  162.000000

       Total_phenols  Flavanoids  Nonflavanoid_phenols  Proanthocyanins  \
count     178.000000  178.000000            178.000000       178.000000
mean        2.295112    2.029270              0.361854         1.590899
std         0.625851    0.998859              0.124453         0.572359
min         0.980000    0.340000              0.130000         0.410000
25%         1.742500    1.205000              0.270000         1.250000
50%         2.355000    2.135000              0.340000         1.555000
75%         2.800000    2.875000              0.437500         1.950000
max         3.880000    5.080000              0.660000         3.580000

       Color_intensity         Hue  0D280_0D315_of_diluted_wines     Proline  \
count       178.000000  178.000000                    178.000000  178.000000
```

```
mean         5.058090    0.957449                       2.611685   746.893258
std          2.318286    0.228572                       0.709990   314.907474
min          1.280000    0.480000                       1.270000   278.000000
25%          3.220000    0.782500                       1.937500   500.500000
50%          4.690000    0.965000                       2.780000   673.500000
75%          6.200000    1.120000                       3.170000   985.000000
max         13.000000    1.710000                       4.000000  1680.000000

            class
count  178.000000
mean     1.938202
std      0.775035
min      1.000000
25%      1.000000
50%      2.000000
75%      3.000000
max      3.000000
```

[36]:
```
import matplotlib.pyplot as plt

df.hist(figsize=(12, 10))
plt.show()
```



[67]:
```
import seaborn as sns
import matplotlib.pyplot as plt

sns.set(style="whitegrid")

plt.figure(figsize=(8, 6))
sns.boxplot(x='class', y='Alcohol', data=df)
plt.title('Box Plot of Alcohol by Class')
plt.xlabel('Class')
plt.ylabel('Alcohol')
plt.show()
```



[38]:
```
correlation_matrix = df.corr()
print(correlation_matrix)
```

```
                              Alcohol  Malicacid       Ash  \
Alcohol                      1.000000   0.094397  0.211545
Malicacid                    0.094397   1.000000  0.164045
Ash                          0.211545   0.164045  1.000000
Alcalinity_of_ash           -0.310235   0.288500  0.443367
Magnesium                    0.270798  -0.054575  0.286587
Total_phenols                0.289101  -0.335167  0.128980
Flavanoids                   0.236815  -0.411007  0.115077
Nonflavanoid_phenols        -0.155929   0.292977  0.186230
Proanthocyanins              0.136698  -0.220746  0.009652
Color_intensity              0.546364   0.248985  0.258887
Hue                         -0.071747  -0.561296 -0.074667
0D280_0D315_of_diluted_wines 0.072343  -0.368710  0.003911
Proline                      0.643720  -0.192011  0.223626
class                       -0.328222   0.437776 -0.049643

                              Alcalinity_of_ash  Magnesium  Total_phenols  \
```

Alcohol	-0.310235	0.270798	0.289101
Malicacid	0.288500	-0.054575	-0.335167
Ash	0.443367	0.286587	0.128980
Alcalinity_of_ash	1.000000	-0.083333	-0.321113
Magnesium	-0.083333	1.000000	0.214401
Total_phenols	-0.321113	0.214401	1.000000
Flavanoids	-0.351370	0.195784	0.864564
Nonflavanoid_phenols	0.361922	-0.256294	-0.449935
Proanthocyanins	-0.197327	0.236441	0.612413
Color_intensity	0.018732	0.199950	-0.055136
Hue	-0.273955	0.055398	0.433681
OD280_OD315_of_diluted_wines	-0.276769	0.066004	0.699949
Proline	-0.440597	0.393351	0.498115
class	0.517859	-0.209179	-0.719163

	Flavanoids	Nonflavanoid_phenols \
Alcohol	0.236815	-0.155929
Malicacid	-0.411007	0.292977
Ash	0.115077	0.186230
Alcalinity_of_ash	-0.351370	0.361922
Magnesium	0.195784	-0.256294
Total_phenols	0.864564	-0.449935
Flavanoids	1.000000	-0.537900
Nonflavanoid_phenols	-0.537900	1.000000
Proanthocyanins	0.652692	-0.365845
Color_intensity	-0.172379	0.139057
Hue	0.543479	-0.262640
OD280_OD315_of_diluted_wines	0.787194	-0.503270
Proline	0.494193	-0.311385
class	-0.847498	0.489109

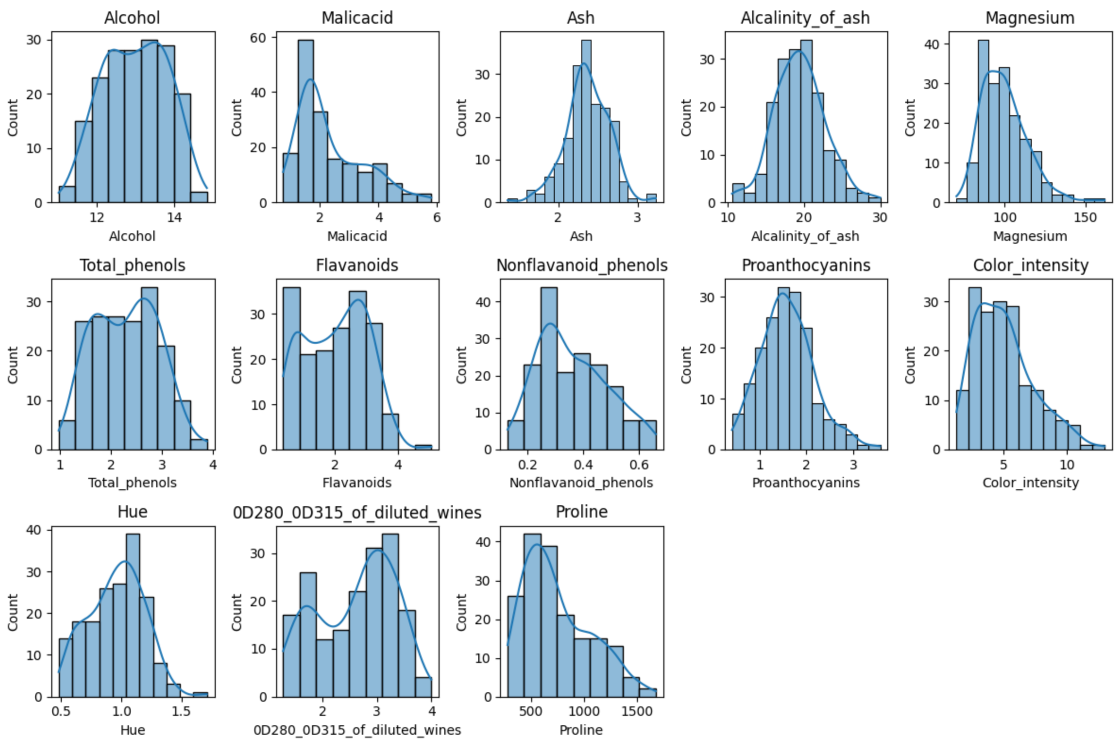
	Proanthocyanins	Color_intensity	Hue \
Alcohol	0.136698	0.546364	-0.071747
Malicacid	-0.220746	0.248985	-0.561296
Ash	0.009652	0.258887	-0.074667
Alcalinity_of_ash	-0.197327	0.018732	-0.273955
Magnesium	0.236441	0.199950	0.055398
Total_phenols	0.612413	-0.055136	0.433681
Flavanoids	0.652692	-0.172379	0.543479
Nonflavanoid_phenols	-0.365845	0.139057	-0.262640
Proanthocyanins	1.000000	-0.025250	0.295544
Color_intensity	-0.025250	1.000000	-0.521813
Hue	0.295544	-0.521813	1.000000
OD280_OD315_of_diluted_wines	0.519067	-0.428815	0.565468
Proline	0.330417	0.316100	0.236183
class	-0.499130	0.265668	-0.617369

	OD280_OD315_of_diluted_wines	Proline	class

Alcohol	0.072343	0.643720	-0.328222
Malicacid	-0.368710	-0.192011	0.437776
Ash	0.003911	0.223626	-0.049643
Alcalinity_of_ash	-0.276769	-0.440597	0.517859
Magnesium	0.066004	0.393351	-0.209179
Total_phenols	0.699949	0.498115	-0.719163
Flavanoids	0.787194	0.494193	-0.847498
Nonflavanoid_phenols	-0.503270	-0.311385	0.489109
Proanthocyanins	0.519067	0.330417	-0.499130
Color_intensity	-0.428815	0.316100	0.265668
Hue	0.565468	0.236183	-0.617369
OD280_OD315_of_diluted_wines	1.000000	0.312761	-0.788230
Proline	0.312761	1.000000	-0.633717
class	-0.788230	-0.633717	1.000000

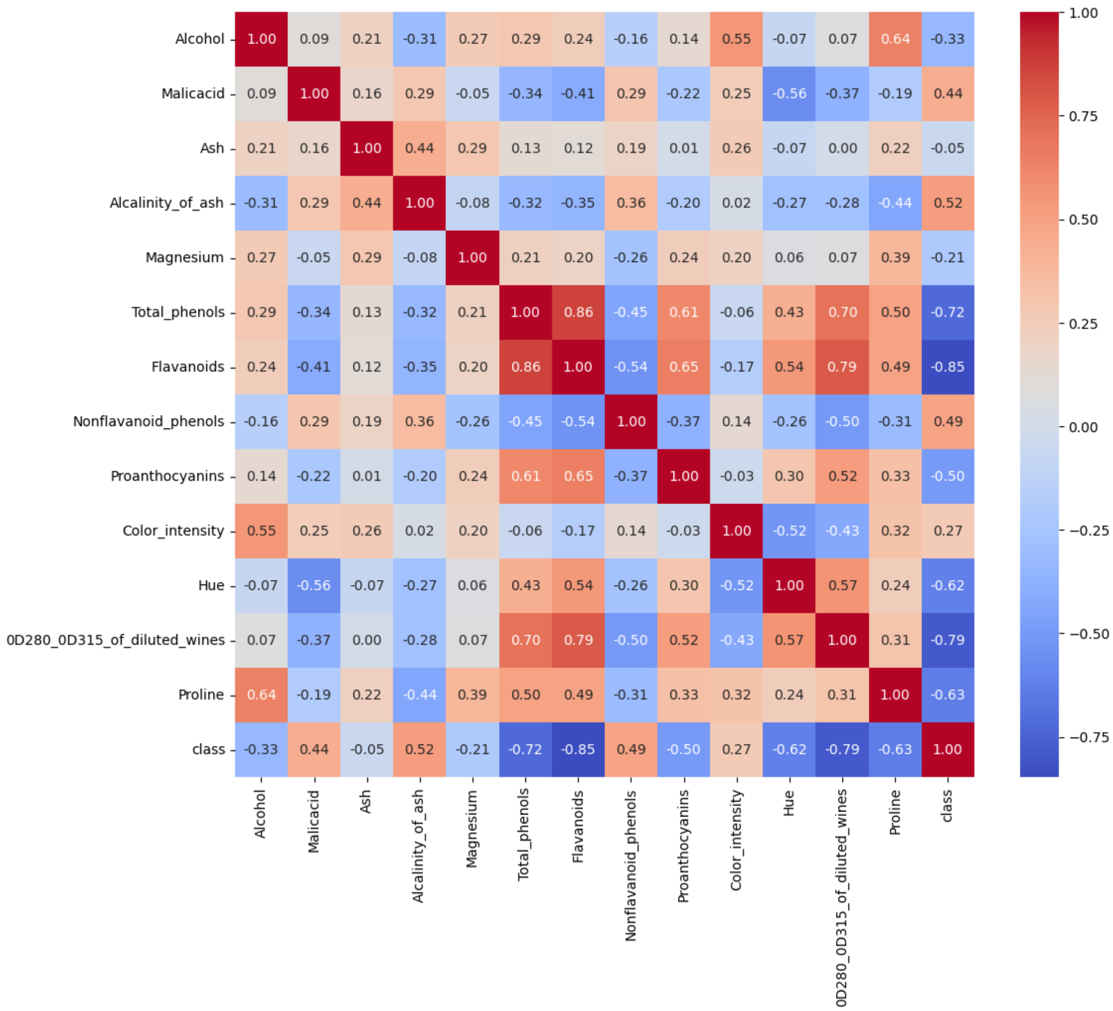
[32]:

```
plt.figure(figsize=(12, 8))
for i, col in enumerate(df.columns[:-1]):
    plt.subplot(3, 5, i+1)
    sns.histplot(df[col], kde=True)
    plt.title(col)
plt.tight_layout()
plt.show()
```



[39]:

```python
plt.figure(figsize=(12, 10))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.show()
```



[40]:

```python
class_distribution = df['class'].value_counts()
print(class_distribution)
```

```
class
2    71
1    59
3    48
Name: count, dtype: int64
```

Class 2 is the most popular, Class 1 comes next, and Class 3 is the least common among the three types

[46]:
```
feature_importance = pd.DataFrame({'Feature': X.columns, 'Coefficient':␣
↪logistic_reg.coef_[0]})
feature_importance = feature_importance.sort_values(by='Coefficient',␣
↪ascending=False)
print(feature_importance)
```

```
                         Feature  Coefficient
6                     Flavanoids     0.931024
2                            Ash     0.823180
11  OD280_OD315_of_diluted_wines     0.479091
1                      Malicacid     0.354056
9                Color_intensity     0.143029
5                  Total_phenols     0.114470
7           Nonflavanoid_phenols     0.067620
12                       Proline     0.008603
4                      Magnesium    -0.016505
8                Proanthocyanins    -0.045146
10                           Hue    -0.146752
0                        Alcohol    -0.295785
3              Alcalinity_of_ash    -0.321011
```

[45]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
↪random_state=42)

logistic_reg = LogisticRegression(max_iter=1000)
logistic_reg.fit(X_train, y_train)
y_pred = logistic_reg.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy)
report = classification_report(y_test, y_pred)
print("Classification Report:")
print(report)
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/utils/validation.py:1143:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
```

```
Accuracy: 0.9722222222222222
Classification Report:
              precision    recall  f1-score   support

           1       1.00      0.93      0.96        14
           2       0.93      1.00      0.97        14
           3       1.00      1.00      1.00         8

    accuracy                           0.97        36
```

```
   macro avg       0.98      0.98      0.98        36
weighted avg       0.97      0.97      0.97        36


/usr/local/lib/python3.10/dist-packages/sklearn/linear_model/_logistic.py:458:
ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression
  n_iter_i = _check_optimize_result(
```

[66]:
```
from sklearn.model_selection import cross_val_score
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
logistic_reg = LogisticRegression(max_iter=10000)

k = 5
cv_scores = cross_val_score(logistic_reg, X_scaled, y, cv=k)

print("Average cross-validation score:", cv_scores.mean())
```

```
Average cross-validation score: 0.9888888888888889
```

0.1 Comments

- model predicts classes with an accuracy of 98.9%. This high accuracy means model performs well
- The model resulted an accuracy of 0.97
- 0D280_0D315_of_diluted_wines have the highest positive coefficients(a high value of feature means a higher chance of belonging to those classes.)
- Alcohol and Alcalinity_of_ash have negative coefficients, indicating a negative impact on predicting the wine classes. This means that there is a lower probability of it being associated to a certain wine. (For example, wines with higher alcohol content might be less likely to belong to a particular class compared to wines with lower alcohol content.)
- Class 2: It's the most common type in your data, with 71 instances.
- Class 1: It's also common but less than Class 2, with 59 instances.
- Class 3: It's the least common, with only 48 instances.