Московский государственный технический университет им. Н.Э. Баумана

Факультет «Информатика и системы управления»

Кафедра «Автоматизированные системы обработки информациии и управления»

Отчет

Лабораторная работа № 1

По курсу Технологии машинного обучения»

ИСПОЛНИТЕЛЬ:

Кожуро Б.Е.

Группа ИУ5-65Б

ПРЕПОДАВАТЕЛЬ:

Гапанюк Ю.Е.

"__"_____2021 г.

Москва 2021

Lab1

April 20, 2021

1 Lab 1

1.1

```
[1]: import numpy as np
     import pandas as pd
     import seaborn as sns
     import matplotlib.pyplot as plt
     %matplotlib inline
     sns.set(style="ticks")
```

```
[2]: from sklearn.datasets import *
     wine = load_wine()
```

```
[3]: #
     for x in wine:
         print(x)
```

```
data
target
frame
target_names
DESCR
feature_names
```

```
[4]: #
     wine['target_names']
```

```
[4]: array(['class_0', 'class_1', 'class_2'], dtype='<U7')
```

```
[5]: #
     wine['feature_names']
```

```
[5]: ['alcohol',
      'malic_acid',
      'ash',
      'alcalinity_of_ash',
      'magnesium',
      'total_phenols',
```

```
        'flavanoids',
        'nonflavanoid_phenols',
        'proanthocyanins',
        'color_intensity',
        'hue',
        'od280/od315_of_diluted_wines',
        'proline']
```

[6]:
```
#
data = pd.DataFrame(data = np.c_[wine['data'], wine['target']],
                    columns= wine['feature_names'] + ['target'])
```

[7]: `data`

[7]:

	alcohol	malic_acid	ash	alcalinity_of_ash	magnesium	total_phenols \
0	14.23	1.71	2.43	15.6	127.0	2.80
1	13.20	1.78	2.14	11.2	100.0	2.65
2	13.16	2.36	2.67	18.6	101.0	2.80
3	14.37	1.95	2.50	16.8	113.0	3.85
4	13.24	2.59	2.87	21.0	118.0	2.80
..	…	…	…	…	…	…
173	13.71	5.65	2.45	20.5	95.0	1.68
174	13.40	3.91	2.48	23.0	102.0	1.80
175	13.27	4.28	2.26	20.0	120.0	1.59
176	13.17	2.59	2.37	20.0	120.0	1.65
177	14.13	4.10	2.74	24.5	96.0	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

	od280/od315_of_diluted_wines	proline	target
0	3.92	1065.0	0.0
1	3.40	1050.0	0.0
2	3.17	1185.0	0.0
3	3.45	1480.0	0.0
4	2.93	735.0	0.0
..	…	…	…

```
173                           1.74    740.0     2.0
174                           1.56    750.0     2.0
175                           1.56    835.0     2.0
176                           1.62    840.0     2.0
177                           1.60    560.0     2.0

[178 rows x 14 columns]
```

1.2

```
[8]: #
     print(f"    : {data.shape[1]},      {data.shape[0]}")
     data.dtypes
```

```
     : 14,       178
```

```
[8]: alcohol                       float64
     malic_acid                    float64
     ash                           float64
     alcalinity_of_ash             float64
     magnesium                     float64
     total_phenols                 float64
     flavanoids                    float64
     nonflavanoid_phenols          float64
     proanthocyanins               float64
     color_intensity               float64
     hue                           float64
     od280/od315_of_diluted_wines  float64
     proline                       float64
     target                        float64
     dtype: object
```

```
[9]: #
     data.isnull().sum()
```

```
[9]: alcohol                       0
     malic_acid                    0
     ash                           0
     alcalinity_of_ash             0
     magnesium                     0
     total_phenols                 0
     flavanoids                    0
     nonflavanoid_phenols          0
     proanthocyanins               0
     color_intensity               0
     hue                           0
     od280/od315_of_diluted_wines  0
     proline                       0
```

```
       target                   0
       dtype: int64
```

[10]: `data.describe()`

[10]:
	alcohol	malic_acid	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	od280/od315_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

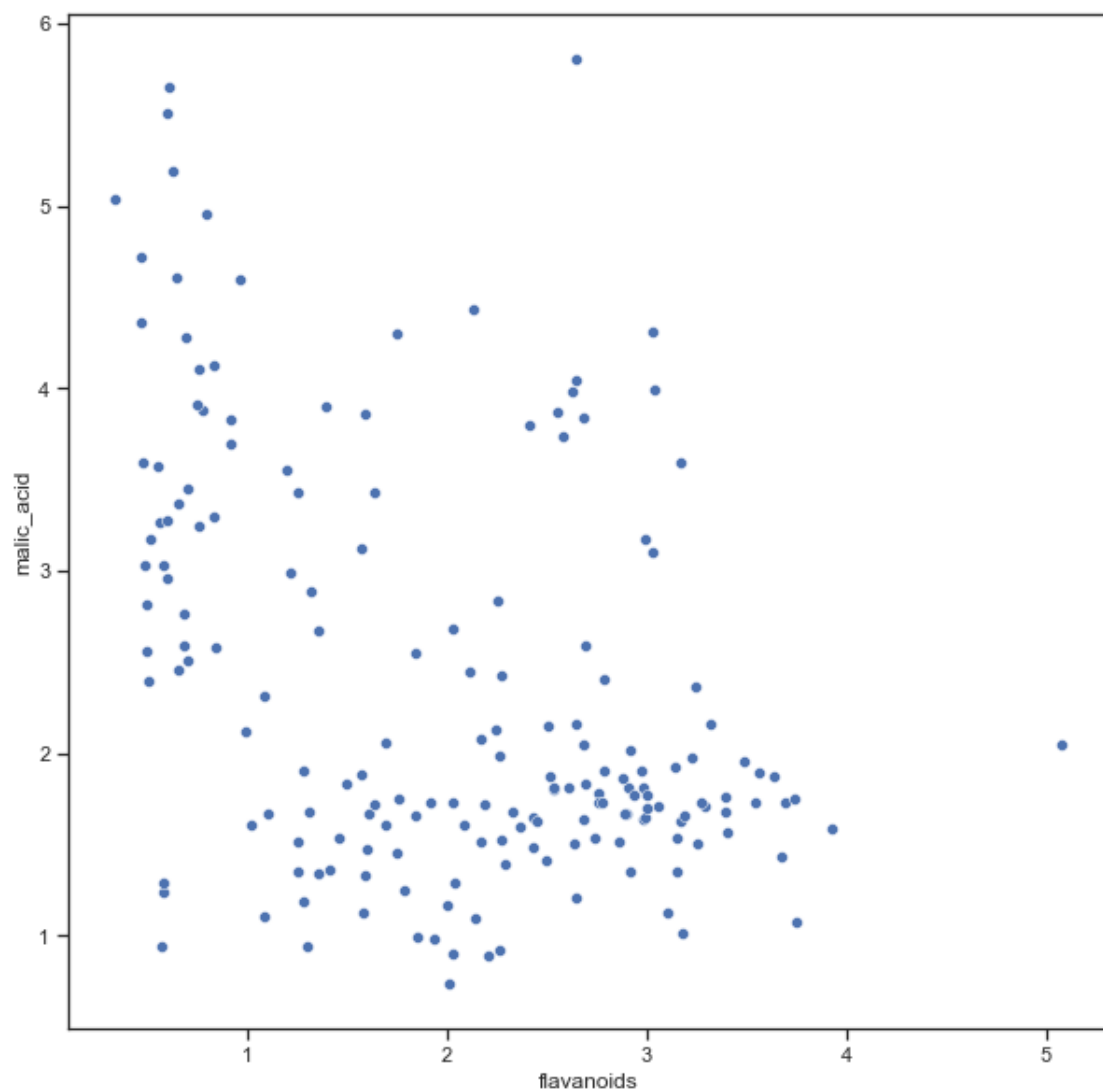|	target
count	178.000000
mean	0.938202
std	0.775035
min	0.000000
25%	0.000000
50%	1.000000
75%	2.000000
max	2.000000

[11]: `data['target'].unique()`

[11]: array([0., 1., 2.])

0, 1, 2

1.3

.
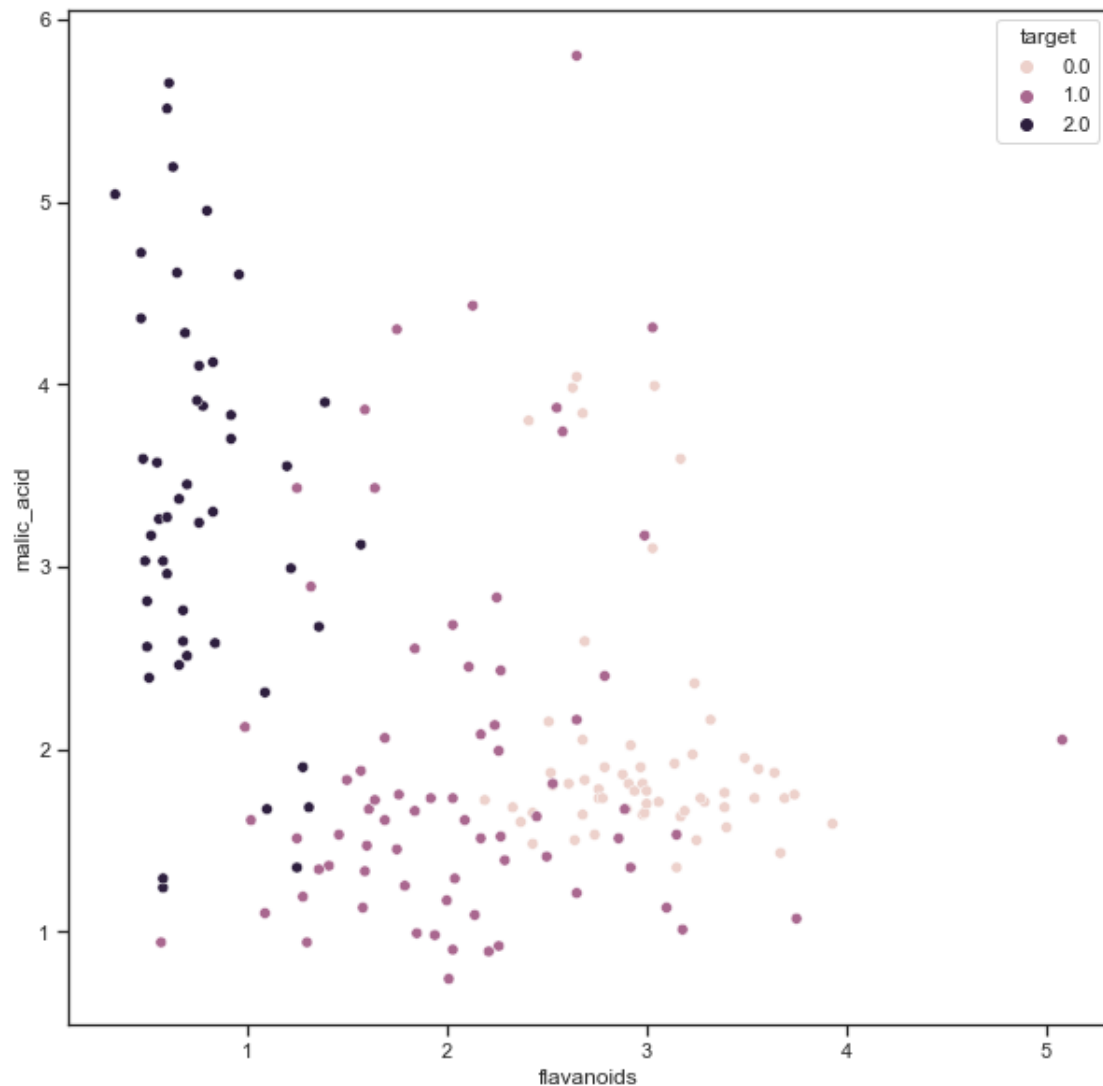
```
[12]:  ##        -
       fig, ax = plt.subplots(figsize=(10,10))
       sns.scatterplot(ax=ax, x='flavanoids', y='malic_acid', data=data)
```

[12]: <AxesSubplot:xlabel='flavanoids', ylabel='malic_acid'>

```
[13]: fig, ax = plt.subplots(figsize=(10,10))
      sns.scatterplot(ax=ax, x='flavanoids', y='malic_acid', data=data, hue =
       ↪'target')
```

```
[13]: <AxesSubplot:xlabel='flavanoids', ylabel='malic_acid'>
```
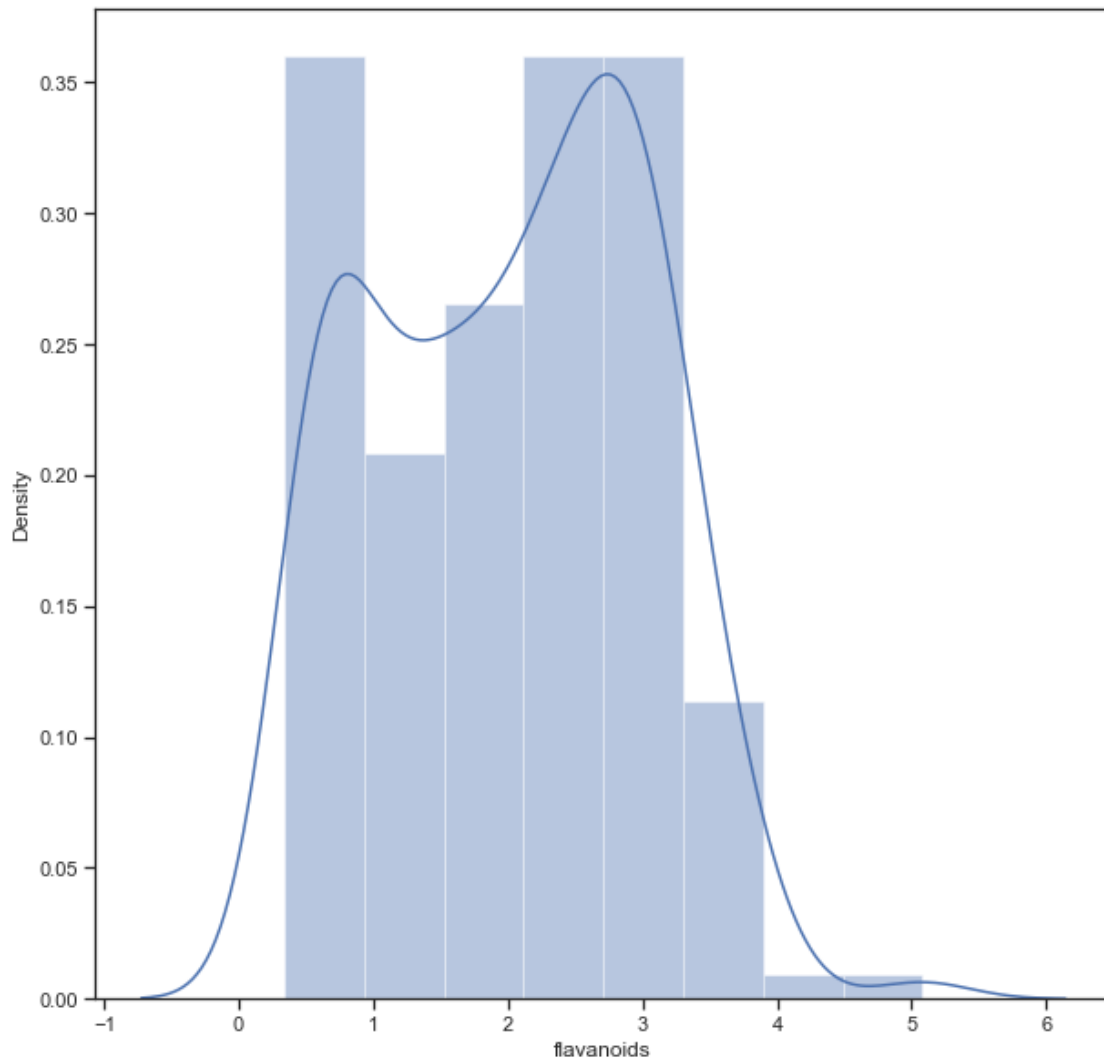


```
[14]: fig, ax = plt.subplots(figsize=(10,10))
      sns.distplot(data['flavanoids'])
```

C:\Users\ksarb\anaconda3\lib\site-packages\seaborn\distributions.py:2551:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level

```
function with similar flexibility) or `histplot` (an axes-level function for
histograms).
  warnings.warn(msg, FutureWarning)
```
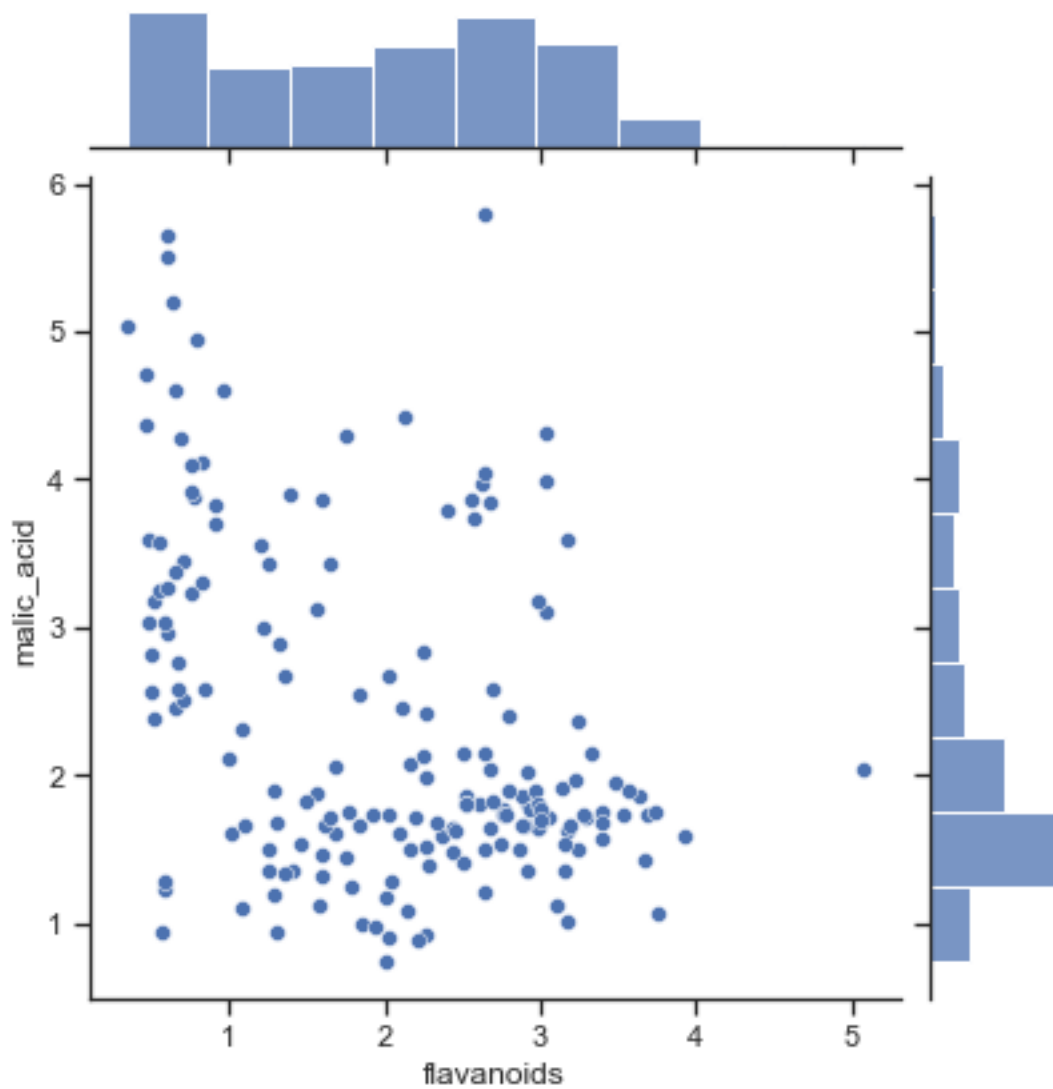
[14]: <AxesSubplot:xlabel='flavanoids', ylabel='Density'>


```
[15]: sns.jointplot(x='flavanoids', y='malic_acid', data=data)
      #    ,                                          ,                    [1,2]
```
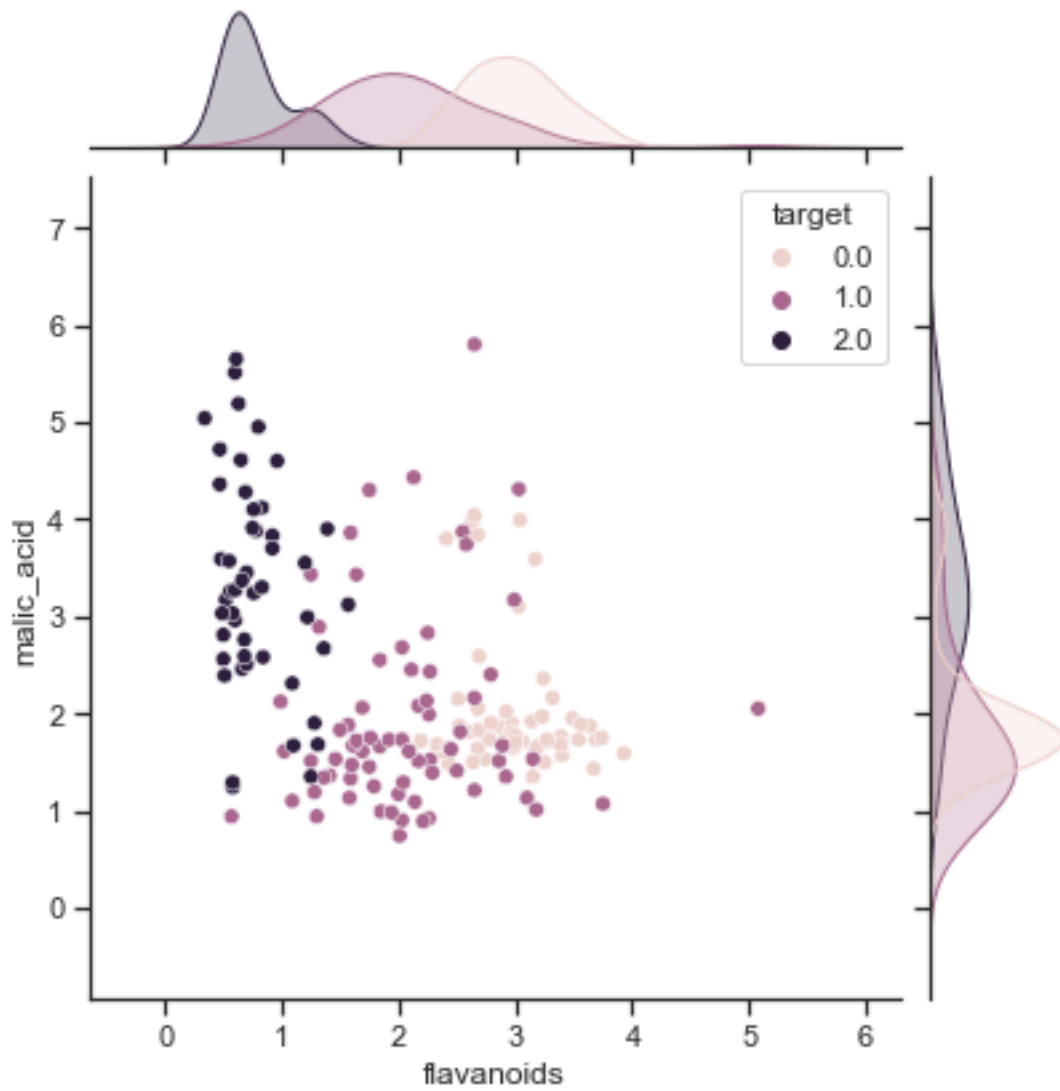
[15]: <seaborn.axisgrid.JointGrid at 0x1e421e4c9d0>

```
[16]: sns.jointplot(x='flavanoids', y='malic_acid', data=data,  hue = 'target')
```

```
[16]: <seaborn.axisgrid.JointGrid at 0x1e422461d30>
```

```
[17]: #                                    target:
      sns.violinplot(x='target', y='flavanoids', data=data)
```
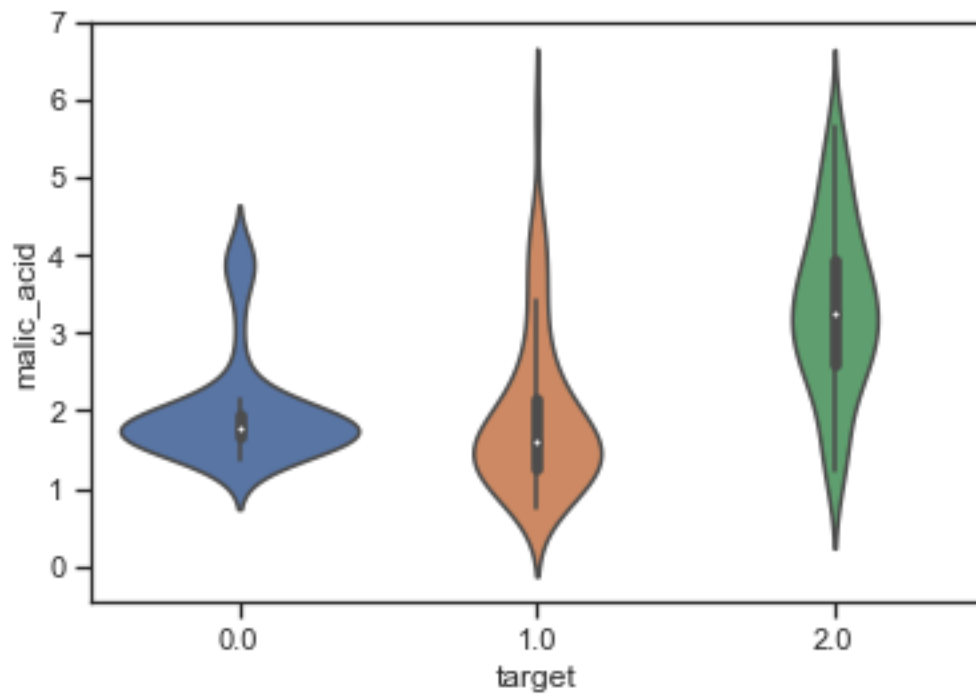
```
[17]: <AxesSubplot:xlabel='target', ylabel='flavanoids'>
```
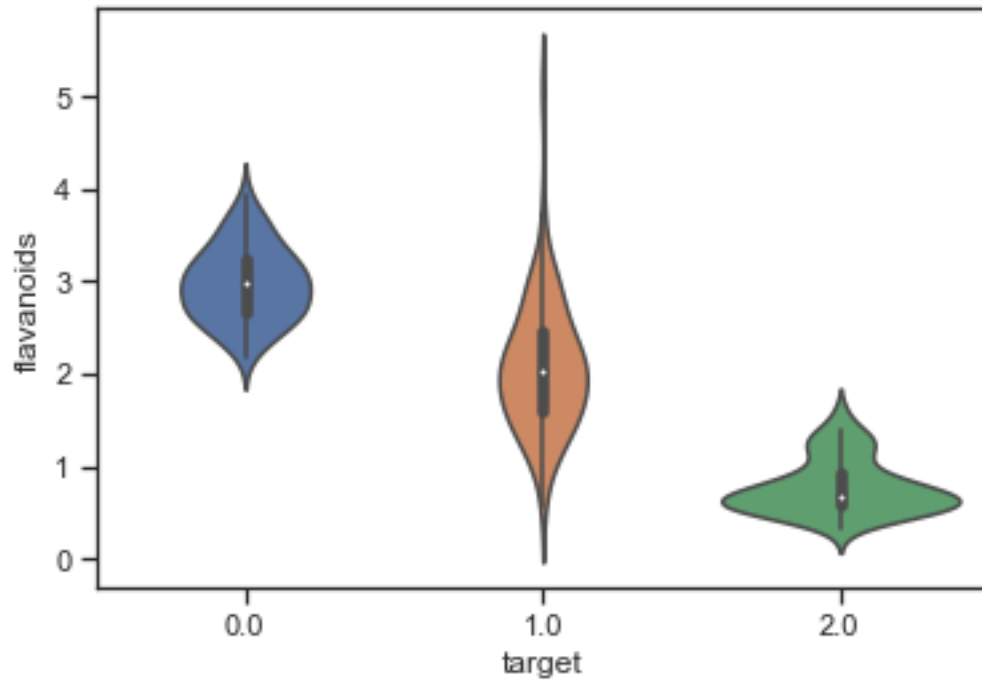
```
[18]: sns.violinplot(x='target', y='malic_acid', data=data)
```

[18]: <AxesSubplot:xlabel='target', ylabel='malic_acid'>

, sns.jointplot(x='flavanoids', y='malic_acid', data=data,
hue = 'target'). target.

```
[19]: sns.pairplot(data, hue='target')
```

[19]: <seaborn.axisgrid.PairGrid at 0x1e4224b1f10>

target , .

1.4

```
[20]: data.corr()
```

[20]:

	alcohol	malic_acid	ash
alcohol	1.000000	0.094397	0.211545
malic_acid	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
od280/od315_of_diluted_wines	0.072343	-0.368710	0.003911
proline	0.643720	-0.192011	0.223626
target	-0.328222	0.437776	-0.049643

	alcalinity_of_ash	magnesium	total_phenols
alcohol	-0.310235	0.270798	0.289101
malic_acid	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
target	0.517859	-0.209179	-0.719163

	flavanoids	nonflavanoid_phenols
alcohol	0.236815	-0.155929
malic_acid	-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

	proanthocyanins	color_intensity	hue \
od280/od315_of_diluted_wines	0.787194		-0.503270
proline	0.494193		-0.311385
target	-0.847498		0.489109

	proanthocyanins	color_intensity	hue \
alcohol	0.136698	0.546364	-0.071747
malic_acid	-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
target	-0.499130	0.265668	-0.617369

	od280/od315_of_diluted_wines	proline	target
alcohol	0.072343	0.643720	-0.328222
malic_acid	-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
target	-0.788230	-0.633717	1.000000

```python
[21]: plt.figure(figsize = (20, 20))
      sns.heatmap(data.corr(), annot=True, fmt='.3f')
```

```
[21]: <AxesSubplot:>
```
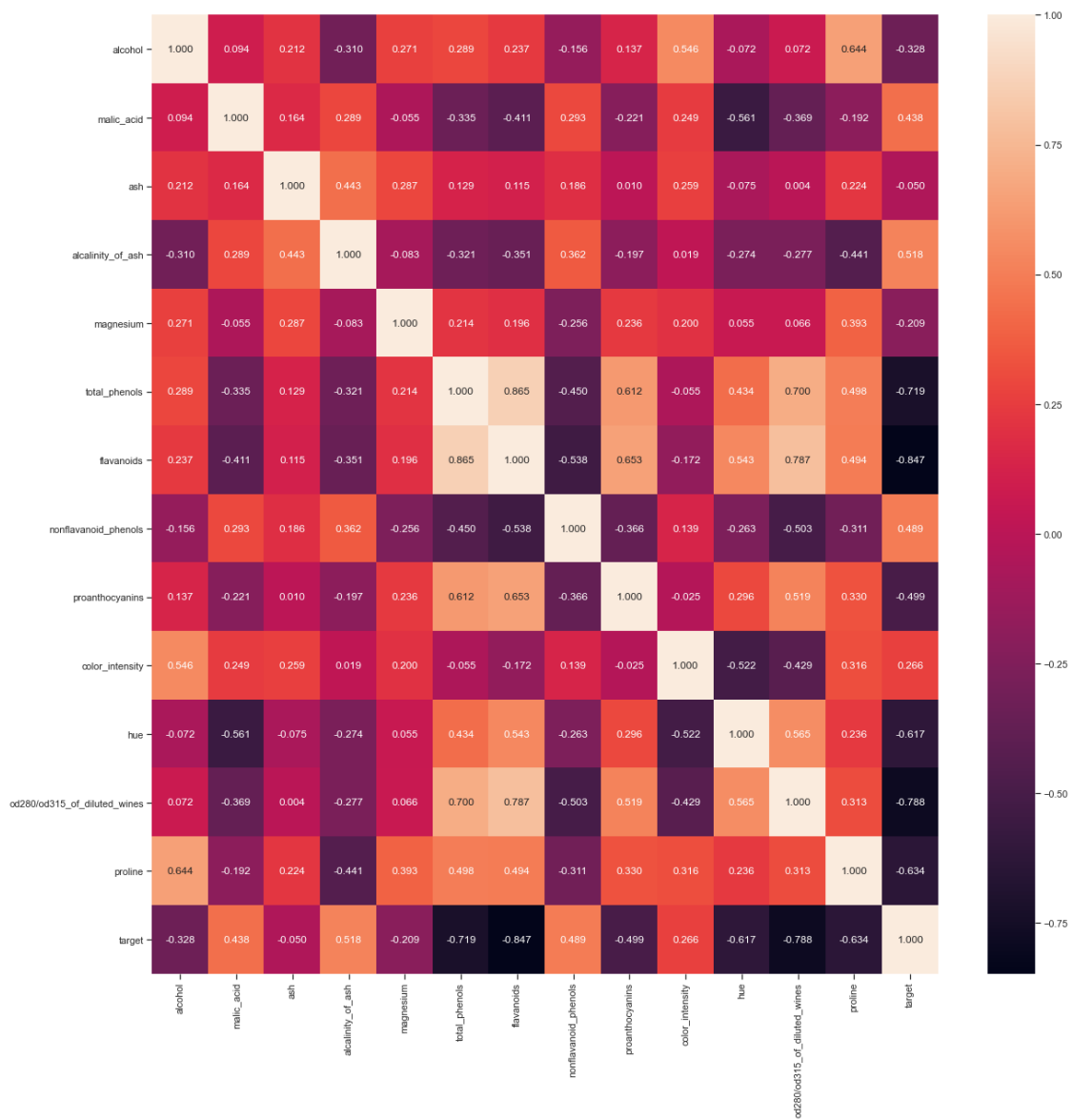
heatmap-   ,   .   target
flavanoids, total_phelons, od280.   target,
.   od280 total_phenols.   ash,
.   ,   (~0.2)
ash   0.5 target.