Московский государственный технический университет им. Н.Э. Баумана
Кафедра «Системы обработки информации и управления»

Лабораторная работа №1
по дисциплине
«Методы машинного обучения»
на тему
«Разведочный анализ данных. Исследование и визуализация данных.»

Выполнил:
Хотин П.Ю.
ИУ5-24М

Москва, 2020 год

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

from sklearn.datasets import load_wine

raw_data = load_wine()
features =
pd.DataFrame(data=raw_data['data'],columns=raw_data['feature_names'])
data = features
data['target']=raw_data['target']
data['class']=data['target'].map(lambda ind: raw_data['target_names']
[ind])
data.head()
```

	alcohol	malic_acid	ash	alcalinity_of_ash	magnesium	total_phenols	flava
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	

```
data.describe()
```

	alcohol	malic_acid	ash	alcalinity_of_ash	magnesium	total_phe
count	178.000000	178.000000	178.000000	178.000000	178.000000	178.0(
mean	13.000618	2.336348	2.366517	19.494944	99.741573	2.2(
std	0.811827	1.117146	0.274344	3.339564	14.282484	0.62
min	11.030000	0.740000	1.360000	10.600000	70.000000	0.9(
25%	12.362500	1.602500	2.210000	17.200000	88.000000	1.7(
50%	13.050000	1.865000	2.360000	19.500000	98.000000	2.3(
75%	13.677500	3.082500	2.557500	21.500000	107.000000	2.8(
max	14.830000	5.800000	3.230000	30.000000	162.000000	3.8(

```
data.shape
```

```
(178, 15)
```

Датасет включает в себя 13 атрибутов:
Alcohol

Malic acid

Ash

Alcalinity of ash

Magnesium

Total phenols

Flavanoids

Nonflavanoid phenols

Proanthocyanins

Color intensity

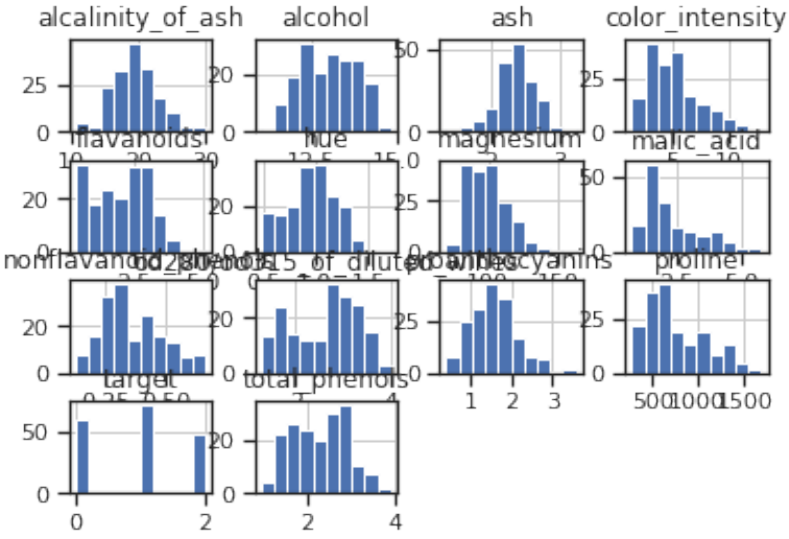
Hue

OD280/OD315 of diluted wines

Proline

```
data.hist()
```

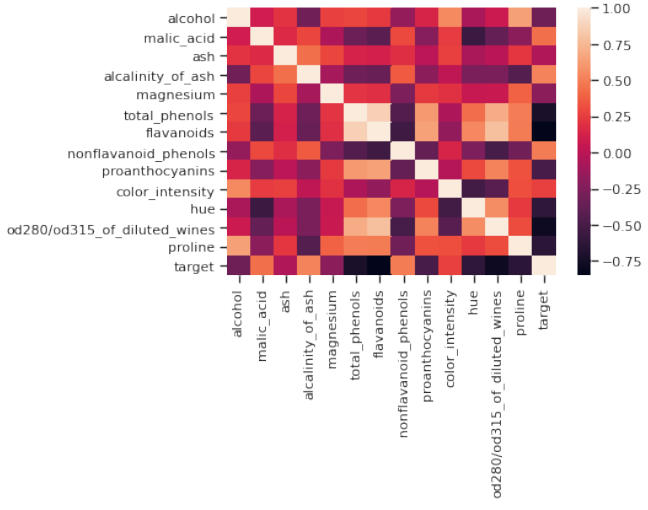
```
array([[<matplotlib.axes._subplots.AxesSubplot object at
0x7f9de315cb38>,
        <matplotlib.axes._subplots.AxesSubplot object at
0x7f9de3134358>,
        <matplotlib.axes._subplots.AxesSubplot object at
0x7f9de30df9b0>,
        <matplotlib.axes._subplots.AxesSubplot object at
0x7f9de308ffd0>],
       [<matplotlib.axes._subplots.AxesSubplot object at
0x7f9de30476a0>,
        <matplotlib.axes._subplots.AxesSubplot object at
0x7f9de3074cf8>,
        <matplotlib.axes._subplots.AxesSubplot object at
0x7f9de302c390>,
        <matplotlib.axes._subplots.AxesSubplot object at
0x7f9de2fd89b0>],
       [<matplotlib.axes._subplots.AxesSubplot object at
0x7f9de2fd8a20>,
        <matplotlib.axes._subplots.AxesSubplot object at
0x7f9de2fbe6d8>,
        <matplotlib.axes._subplots.AxesSubplot object at
0x7f9de2f6fd30>,
        <matplotlib.axes._subplots.AxesSubplot object at
0x7f9de2f2d3c8>],
       [<matplotlib.axes._subplots.AxesSubplot object at
0x7f9de2edda20>,
        <matplotlib.axes._subplots.AxesSubplot object at
0x7f9de2e9a0b8>,
        <matplotlib.axes._subplots.AxesSubplot object at
0x7f9de2e49710>,
        <matplotlib.axes._subplots.AxesSubplot object at
0x7f9de2e79d68>]],
      dtype=object)
```




```
corr = data.corr()
```

```
sns.heatmap(corr,
        xticklabels=corr.columns,
        yticklabels=corr.columns)
```

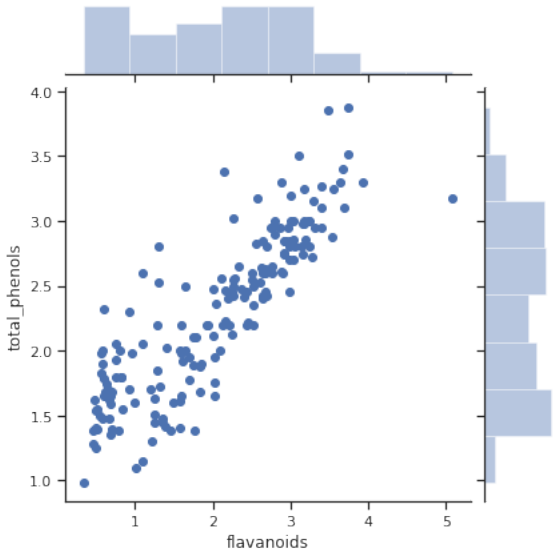
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9de26f9cf8>
```



Видим наибольшую корреляцию между *flavanoids* и total_phenols

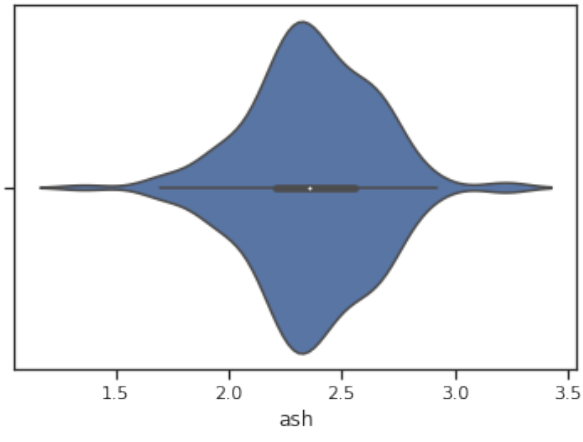
```
sns.jointplot(x='flavanoids', y='total_phenols', data=data)
```

```
<seaborn.axisgrid.JointGrid at 0x7f9de13b5cc0>
```




```
sns.violinplot(x=data['ash'])
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9ddd8fc0b8>
```



```
for i in data.target.unique():
    sns.distplot(data['alcohol'][data.target==i],
                 kde=1,label='{}'.format(i))

plt.legend()
```

```
<matplotlib.legend.Legend at 0x7f9ddaf33cf8>
```