**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:

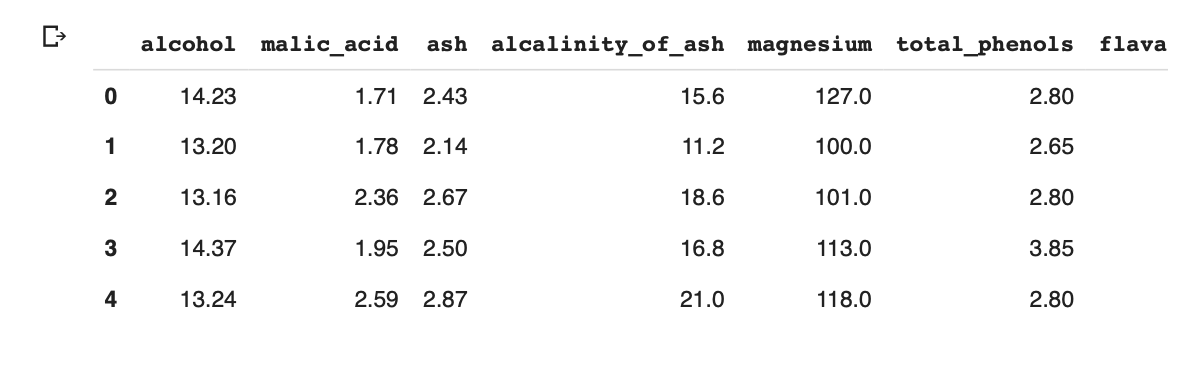
Хотин П.Ю.

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")

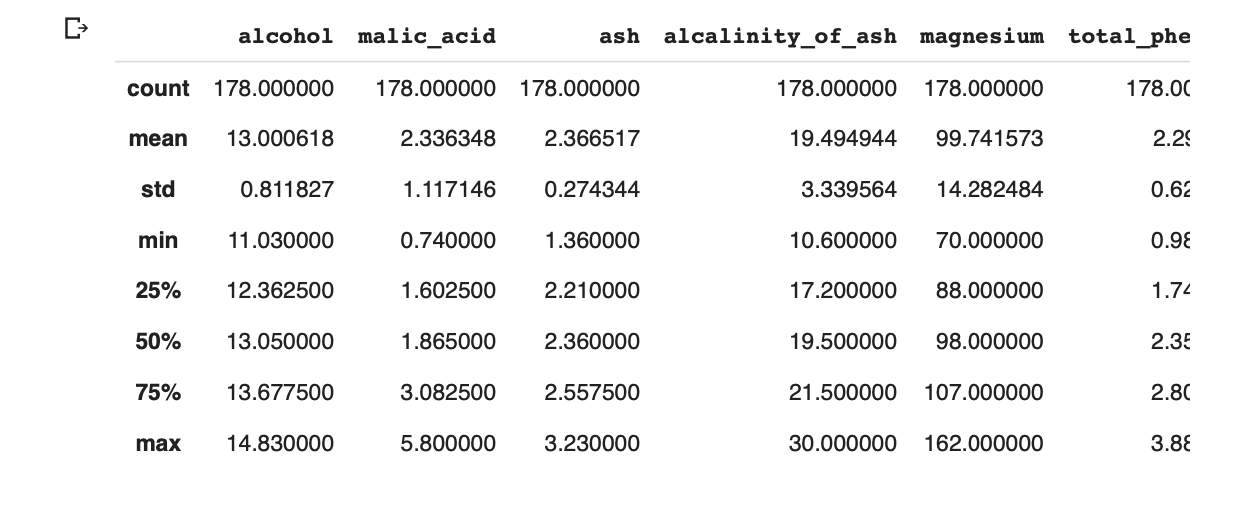
from sklearn.datasets import load\_wine

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()

data.describe()

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

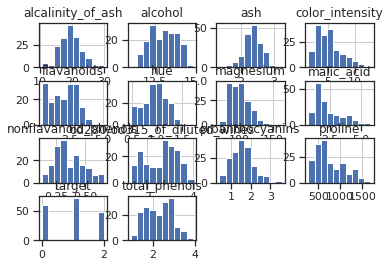
OD280/OD315 of diluted wines

Proline

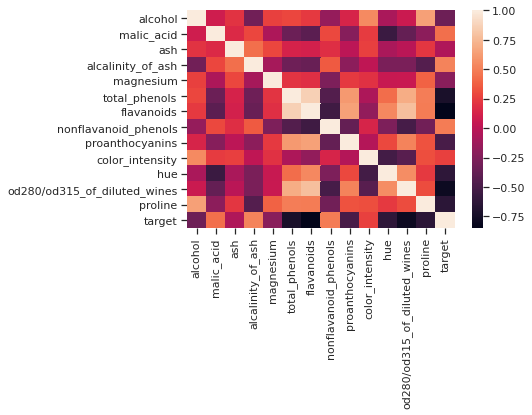
data.hist()

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()

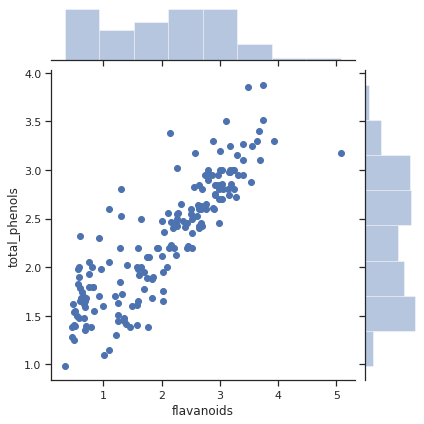
sns.heatmap(corr,   
 xticklabels=corr.columns,  
 yticklabels=corr.columns)

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

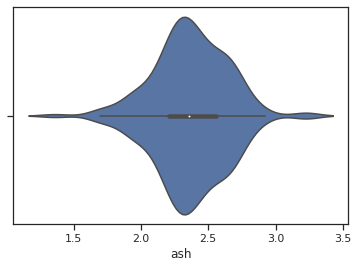


Видим наибольшую корреляцию между *flavanoids* и total\_phenols

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>