Московский Государственный Технический Университет
им. Н.Э. Баумана

Рубежный контроль №1

по курсу
Технологии Машинного Обучения

Выполнила:
Костян Алина
ИУ5-53

Проверил:
Гапанюк Ю.Е.

Москва, 2019

Задание

Для заданного набора данных проведите корреляционный анализ. В случае наличия пропусков в данных удалите строки или колонки, содержащие пропуски. Сделайте выводы о возможности построения моделей машинного обучения и о возможном вкладе признаков в модель.

Код и результаты выполнения

1. Подключим библиотеки:

```
import pandas as pd
import sklearn
import numpy as np
from sklearn.datasets import load_boston
import matplotlib.pyplot as plt
import seaborn as sns
```

2. Подготовим данные

```
boston = load_boston()
```

```
print(boston.data.shape)
```

```
(506, 13)
```

```
data=pd.DataFrame(boston.data)
```

```
data.isnull().sum()
```

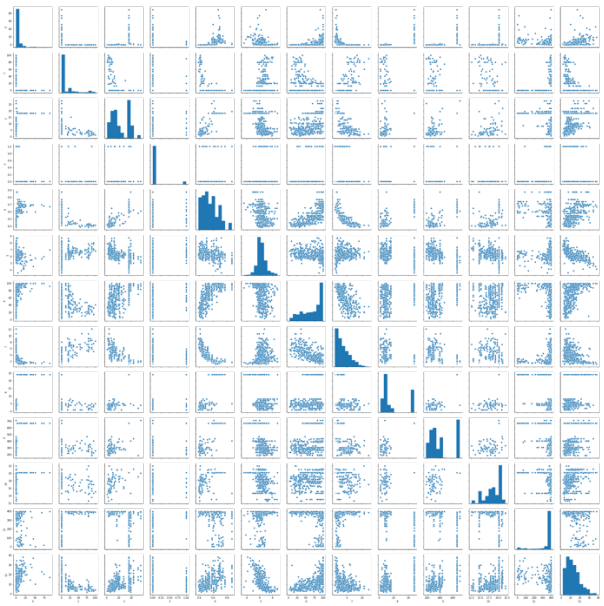
```
0     0
1     0
2     0
3     0
4     0
5     0
6     0
7     0
8     0
9     0
10    0
11    0
12    0
dtype: int64
```

3. Основные статистические характеристики набора данных

```
data.describe()
```

	0	1	2	3	4	5	6	7	8	9	10	11	
count	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.(
mean	3.613524	11.363636	11.136779	0.069170	0.554695	6.284634	68.574901	3.795043	9.549407	408.237154	18.455534	356.674032	12.6
std	8.601545	23.322453	6.860353	0.253994	0.115878	0.702617	28.148861	2.105710	8.707259	168.537116	2.164946	91.294864	7.1
min	0.006320	0.000000	0.460000	0.000000	0.385000	3.561000	2.900000	1.129600	1.000000	187.000000	12.600000	0.320000	1.7
25%	0.082045	0.000000	5.190000	0.000000	0.449000	5.885500	45.025000	2.100175	4.000000	279.000000	17.400000	375.377500	6.9
50%	0.256510	0.000000	9.690000	0.000000	0.538000	6.208500	77.500000	3.207450	5.000000	330.000000	19.050000	391.440000	11.3
75%	3.677083	12.500000	18.100000	0.000000	0.624000	6.623500	94.075000	5.188425	24.000000	666.000000	20.200000	396.225000	16.9
max	88.976200	100.000000	27.740000	1.000000	0.871000	8.780000	100.000000	12.126500	24.000000	711.000000	22.000000	396.900000	37.9

4. Построим парные диаграммы для всего датасета



5. Корреляция величин набора данных

```
data.corr()
```

	0	1	2	3	4	5	6	7	8	9	10	11	12
0	1.000000	-0.200469	0.406583	-0.055892	0.420972	-0.219247	0.352734	-0.379670	0.625505	0.582764	0.289946	-0.385064	0.455621
1	-0.200469	1.000000	-0.533828	-0.042697	-0.516604	0.311991	-0.569537	0.664408	-0.311948	-0.314563	-0.391679	0.175520	-0.412995
2	0.406583	-0.533828	1.000000	0.062938	0.763651	-0.391676	0.644779	-0.708027	0.595129	0.720760	0.383248	-0.356977	0.603800
3	-0.055892	-0.042697	0.062938	1.000000	0.091203	0.091251	0.086518	-0.099176	-0.007368	-0.035587	-0.121515	0.048788	-0.053929
4	0.420972	-0.516604	0.763651	0.091203	1.000000	-0.302188	0.731470	-0.769230	0.611441	0.668023	0.188933	-0.380051	0.590879
5	-0.219247	0.311991	-0.391676	0.091251	-0.302188	1.000000	-0.240265	0.205246	-0.209847	-0.292048	-0.355501	0.128069	-0.613808
6	0.352734	-0.569537	0.644779	0.086518	0.731470	-0.240265	1.000000	-0.747881	0.456022	0.506456	0.261515	-0.273534	0.602339
7	-0.379670	0.664408	-0.708027	-0.099176	-0.769230	0.205246	-0.747881	1.000000	-0.494588	-0.534432	-0.232471	0.291512	-0.496996
8	0.625505	-0.311948	0.595129	-0.007368	0.611441	-0.209847	0.456022	-0.494588	1.000000	0.910228	0.464741	-0.444413	0.488676
9	0.582764	-0.314563	0.720760	-0.035587	0.668023	-0.292048	0.506456	-0.534432	0.910228	1.000000	0.460853	-0.441808	0.543993
10	0.289946	-0.391679	0.383248	-0.121515	0.188933	-0.355501	0.261515	-0.232471	0.464741	0.460853	1.000000	-0.177383	0.374044
11	-0.385064	0.175520	-0.356977	0.048788	-0.380051	0.128069	-0.273534	0.291512	-0.444413	-0.441808	-0.177383	1.000000	-0.366087
12	0.455621	-0.412995	0.603800	-0.053929	0.590879	-0.613808	0.602339	-0.496996	0.488676	0.543993	0.374044	-0.366087	1.000000

Метод: pearson

	0	1	2	3	4	5	6	7	8	9	10	11	12
0	1.000000	-0.200469	0.406583	-0.055892	0.420972	-0.219247	0.352734	-0.379670	0.625505	0.582764	0.289946	-0.385064	0.455621
1	-0.200469	1.000000	-0.533828	-0.042697	-0.516604	0.311991	-0.569537	0.664408	-0.311948	-0.314563	-0.391679	0.175520	-0.412995
2	0.406583	-0.533828	1.000000	0.062938	0.763651	-0.391676	0.644779	-0.708027	0.595129	0.720760	0.383248	-0.356977	0.603800
3	-0.055892	-0.042697	0.062938	1.000000	0.091203	0.091251	0.086518	-0.099176	-0.007368	-0.035587	-0.121515	0.048788	-0.053929
4	0.420972	-0.516604	0.763651	0.091203	1.000000	-0.302188	0.731470	-0.769230	0.611441	0.668023	0.188933	-0.380051	0.590879
5	-0.219247	0.311991	-0.391676	0.091251	-0.302188	1.000000	-0.240265	0.205246	-0.209847	-0.292048	-0.355501	0.128069	-0.613808
6	0.352734	-0.569537	0.644779	0.086518	0.731470	-0.240265	1.000000	-0.747881	0.456022	0.506456	0.261515	-0.273534	0.602339
7	-0.379670	0.664408	-0.708027	-0.099176	-0.769230	0.205246	-0.747881	1.000000	-0.494588	-0.534432	-0.232471	0.291512	-0.496996
8	0.625505	-0.311948	0.595129	-0.007368	0.611441	-0.209847	0.456022	-0.494588	1.000000	0.910228	0.464741	-0.444413	0.488676
9	0.582764	-0.314563	0.720760	-0.035587	0.668023	-0.292048	0.506456	-0.534432	0.910228	1.000000	0.460853	-0.441808	0.543993
10	0.289946	-0.391679	0.383248	-0.121515	0.188933	-0.355501	0.261515	-0.232471	0.464741	0.460853	1.000000	-0.177383	0.374044
11	-0.385064	0.175520	-0.356977	0.048788	-0.380051	0.128069	-0.273534	0.291512	-0.444413	-0.441808	-0.177383	1.000000	-0.366087
12	0.455621	-0.412995	0.603800	-0.053929	0.590879	-0.613808	0.602339	-0.496996	0.488676	0.543993	0.374044	-0.366087	1.000000

Метод: kendall

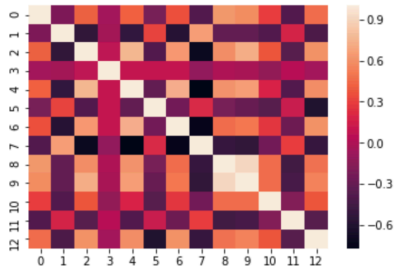
	0	1	2	3	4	5	6	7	8	9	10	11	12
0	1.000000	-0.462057	0.521014	0.033948	0.603361	-0.211718	0.497297	-0.539878	0.563969	0.544956	0.312768	-0.264378	0.454837
1	-0.462057	1.000000	-0.535468	-0.039419	-0.511464	0.278134	-0.429389	0.478524	-0.234663	-0.289911	-0.361607	0.128177	-0.386818
2	0.521014	-0.535468	1.000000	0.075889	0.612030	-0.291318	0.489070	-0.565137	0.353967	0.483228	0.336612	-0.192017	0.465980
3	0.033948	-0.039419	0.075889	1.000000	0.056387	0.048080	0.055616	-0.065619	0.021739	-0.037655	-0.115694	-0.033277	-0.041344
4	0.603361	-0.511464	0.612030	0.056387	1.000000	-0.215633	0.589608	-0.683930	0.434828	0.453258	0.278678	-0.202430	0.452005
5	-0.211718	0.278134	-0.291318	0.048080	-0.215633	1.000000	-0.187611	0.179801	-0.076569	-0.190532	-0.223194	0.032951	-0.468231
6	0.497297	-0.429389	0.489070	0.055616	0.589608	-0.187611	1.000000	-0.609836	0.306201	0.360311	0.251857	-0.154056	0.485359
7	-0.539878	0.478524	-0.565137	-0.065619	-0.683930	0.179801	-0.609836	1.000000	-0.361892	-0.381988	-0.223486	0.168631	-0.409347
8	0.563969	-0.234663	0.353967	0.021739	0.434828	-0.076569	0.306201	-0.361892	1.000000	0.558107	0.251913	-0.214364	0.287943
9	0.544956	-0.289911	0.483228	-0.037655	0.453258	-0.190532	0.360311	-0.381988	0.558107	1.000000	0.287769	-0.241606	0.384191
10	0.312768	-0.361607	0.336612	-0.115694	0.278678	-0.223194	0.251857	-0.223486	0.251913	0.287769	1.000000	-0.042152	0.330335
11	-0.264378	0.128177	-0.192017	-0.033277	-0.202430	0.032951	-0.154056	0.168631	-0.214364	-0.241606	-0.042152	1.000000	-0.145430
12	0.454837	-0.386818	0.465980	-0.041344	0.452005	-0.468231	0.485359	-0.409347	0.287943	0.384191	0.330335	-0.145430	1.000000

Метод: spearman

	0	1	2	3	4	5	6	7	8	9	10	11	12
0	1.000000	-0.571660	0.735524	0.041537	0.821465	-0.309116	0.704140	-0.744986	0.727807	0.729045	0.465283	-0.360555	0.634760
1	-0.571660	1.000000	-0.642811	-0.041937	-0.634828	0.361074	-0.544423	0.614627	-0.278767	-0.371394	-0.448475	0.163135	-0.490074
2	0.735524	-0.642811	1.000000	0.089841	0.791189	-0.415301	0.679487	-0.757080	0.455507	0.664361	0.433710	-0.285840	0.638747
3	0.041537	-0.041937	0.089841	1.000000	0.068426	0.058813	0.067792	-0.080248	0.024579	-0.044486	-0.136065	-0.039810	-0.050575
4	0.821465	-0.634828	0.791189	0.068426	1.000000	-0.310344	0.795153	-0.880015	0.586429	0.649527	0.391309	-0.296662	0.636828
5	-0.309116	0.361074	-0.415301	0.058813	-0.310344	1.000000	-0.278082	0.263168	-0.107492	-0.271898	-0.312923	0.053660	-0.640832
6	0.704140	-0.544423	0.679487	0.067792	0.795153	-0.278082	1.000000	-0.801610	0.417983	0.526366	0.355384	-0.228022	0.657071
7	-0.744986	0.614627	-0.757080	-0.080248	-0.880015	0.263168	-0.801610	1.000000	-0.495806	-0.574336	-0.322041	0.249595	-0.564262
8	0.727807	-0.278767	0.455507	0.024579	0.586429	-0.107492	0.417983	-0.495806	1.000000	0.704876	0.318330	-0.282533	0.394322
9	0.729045	-0.371394	0.664361	-0.044486	0.649527	-0.271898	0.526366	-0.574336	0.704876	1.000000	0.453345	-0.329843	0.534423
10	0.465283	-0.448475	0.433710	-0.136065	0.391309	-0.312923	0.355384	-0.322041	0.318330	0.453345	1.000000	-0.072027	0.467259
11	-0.360555	0.163135	-0.285840	-0.039810	-0.296662	0.053660	-0.228022	0.249595	-0.282533	-0.329843	-0.072027	1.000000	-0.210562
12	0.634760	-0.490074	0.638747	-0.050575	0.636828	-0.640832	0.657071	-0.564262	0.394322	0.534423	0.467259	-0.210562	1.000000

```
sns.heatmap(data.corr())
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x12ec7510>
```




```
fig, ax = plt.subplots(1, 3, sharex='col', sharey='row', figsize=(15,5))
sns.heatmap(data.corr(method='pearson'), ax=ax[0], fmt='.2f')
sns.heatmap(data.corr(method='kendall'), ax=ax[1], fmt='.2f')
sns.heatmap(data.corr(method='spearman'), ax=ax[2], fmt='.2f')
fig.suptitle('Корреляционные матрицы, построенные различными методами')
ax[0].title.set_text('Pearson')
ax[1].title.set_text('Kendall')
ax[2].title.set_text('Spearman')
```