

# Лабораторная работа № 2

Логинов Сергей

Вариант 21

1. Считаем заданный набор данных из репозитория UCI, используя Pandas

```
In [16]: import pandas as pd
import numpy as np
```

```
In [17]: url = \
"http://archive.ics.uci.edu/ml/machine-learning-databases/arrhythmia/arrhythmia.data"

my_data = pd.read_csv( url, header=None)
my_data.columns = ['V' + str(i) for i in range(0, 280)] #prefix="V" old and will be remo
print( "\n*** Начало данных:\n", my_data.head() ) # начальные данные
print( "\n*** Конец данных:\n", my_data.tail() ) # конечные данные

summary = my_data.describe() # сводка данных
print( "\n*** Сводка данных:\n", summary)
```

\*\*\* Начало данных:

	V0	V1	V2	V3	V4	V5	V6	V7	V8	V9	...	V270	V271	V272	V273	\
0	75	0	190	80	91	193	371	174	121	-16	...	0.0	9.0	-0.9	0.0	
1	56	1	165	64	81	174	401	149	39	25	...	0.0	8.5	0.0	0.0	
2	54	0	172	95	138	163	386	185	102	96	...	0.0	9.5	-2.4	0.0	
3	55	0	175	94	100	202	380	179	143	28	...	0.0	12.2	-2.2	0.0	
4	75	0	190	80	88	181	360	177	103	-16	...	0.0	13.1	-3.6	0.0	

	V274	V275	V276	V277	V278	V279
0	0.0	0.9	2.9	23.3	49.4	8
1	0.0	0.2	2.1	20.4	38.8	6
2	0.0	0.3	3.4	12.3	49.0	10
3	0.0	0.4	2.6	34.6	61.6	1
4	0.0	-0.1	3.9	25.4	62.8	7

[5 rows x 280 columns]

\*\*\* Конец данных:

	V0	V1	V2	V3	V4	V5	V6	V7	V8	V9	...	V270	V271	V272	V273	\
447	53	1	160	70	80	199	382	154	117	-37	...	0.0	4.3	-5.0	0.0	
448	37	0	190	85	100	137	361	201	73	86	...	0.0	15.6	-1.6	0.0	
449	36	0	166	68	108	176	365	194	116	-85	...	0.0	16.3	-28.6	0.0	
450	32	1	155	55	93	106	386	218	63	54	...	-0.4	12.0	-0.7	0.0	
451	78	1	160	70	79	127	364	138	78	28	...	0.0	10.4	-1.8	0.0	

	V274	V275	V276	V277	V278	V279
447	0.0	0.7	0.6	-4.4	-0.5	1
448	0.0	0.4	2.4	38.0	62.4	10
449	0.0	1.5	1.0	-44.2	-33.2	2
450	0.0	0.5	2.4	25.0	46.6	1
451	0.0	0.5	1.6	21.3	32.8	1

[5 rows x 280 columns]

\*\*\* Сводка данных:

V0	V1	V2	V3	V4	V5	\
----	----	----	----	----	----	---

count	452.000000	452.000000	452.000000	452.000000	452.000000	452.000000
mean	46.471239	0.550885	166.188053	68.170354	88.920354	155.152655
std	16.466631	0.497955	37.170340	16.590803	15.364394	44.842283
min	0.000000	0.000000	105.000000	6.000000	55.000000	0.000000
25%	36.000000	0.000000	160.000000	59.000000	80.000000	142.000000
50%	47.000000	1.000000	164.000000	68.000000	86.000000	157.000000
75%	58.000000	1.000000	170.000000	79.000000	94.000000	175.000000
max	83.000000	1.000000	780.000000	176.000000	188.000000	524.000000

	V6	V7	V8	V9	...	V270	\
count	452.000000	452.000000	452.000000	452.000000	...	452.000000	
mean	367.207965	169.949115	90.004425	33.676991	...	-0.278982	
std	33.385421	35.633072	25.826643	45.431434	...	0.548876	
min	232.000000	108.000000	0.000000	-172.000000	...	-4.100000	
25%	350.000000	148.000000	79.000000	3.750000	...	-0.425000	
50%	367.000000	162.000000	91.000000	40.000000	...	0.000000	
75%	384.000000	179.000000	102.000000	66.000000	...	0.000000	
max	509.000000	381.000000	205.000000	169.000000	...	0.000000	

	V271	V272	V273	V274	V275	V276	\
count	452.000000	452.000000	452.000000	452.0	452.000000	452.000000	
mean	9.048009	-1.457301	0.003982	0.0	0.514823	1.222345	
std	3.472862	2.002430	0.050118	0.0	0.347531	1.426052	
min	0.000000	-28.600000	0.000000	0.0	-0.800000	-6.000000	
25%	6.600000	-2.100000	0.000000	0.0	0.400000	0.500000	
50%	8.800000	-1.100000	0.000000	0.0	0.500000	1.350000	
75%	11.200000	0.000000	0.000000	0.0	0.700000	2.100000	
max	23.600000	0.000000	0.800000	0.0	2.400000	6.000000	

	V277	V278	V279
count	452.000000	452.000000	452.000000
mean	19.326106	29.473230	3.880531
std	13.503922	18.493927	4.407097
min	-44.200000	-38.600000	1.000000
25%	11.450000	17.550000	1.000000
50%	18.100000	27.900000	1.000000
75%	25.825000	41.125000	6.000000
max	88.800000	115.900000	16.000000

[8 rows x 275 columns]

1. Проведем исследование числовых признаков. Заменим пропущенные значения. Приведем к числовым типам.

```
In [18]: my_data = my_data.replace('?', np.NaN)
my_data = my_data.fillna(my_data.median())
for i in my_data:
    if my_data[i].dtype == "object":
        my_data[i] = pd.to_numeric(my_data[i])
```

```
In [19]: my_data.describe()
```

Out[19]:	V0	V1	V2	V3	V4	V5	V6	V
count	452.000000	452.000000	452.000000	452.000000	452.000000	452.000000	452.000000	452.000000
mean	46.471239	0.550885	166.188053	68.170354	88.920354	155.152655	367.207965	169.949115
std	16.466631	0.497955	37.170340	16.590803	15.364394	44.842283	33.385421	35.633072
min	0.000000	0.000000	105.000000	6.000000	55.000000	0.000000	232.000000	108.000000
25%	36.000000	0.000000	160.000000	59.000000	80.000000	142.000000	350.000000	148.000000
50%	47.000000	1.000000	164.000000	68.000000	86.000000	157.000000	367.000000	162.000000

75%	58.000000	1.000000	170.000000	79.000000	94.000000	175.000000	384.000000	179.000000
max	83.000000	1.000000	780.000000	176.000000	188.000000	524.000000	509.000000	381.000000

8 rows x 280 columns

1. В данной выборке имеется несколько категориальных признаков: V1, V2, V20-V26. Определим числовой признак, имеющий максимальную дисперсию.

```
In [20]: def max_std(data):
max_std_key = 0
max_std_value = 0
for i in data:
tmp_std = data[i].std()
if max_std_value < tmp_std:
max_std_value = tmp_std
max_std_key = i
return {max_std_key : max_std_value}
numeric_attr_with_max_std = max_std(my_data)
print(numeric_attr_with_max_std)

{'V10': 57.34637571704272}
```

1. При помощи класса SelectKBest библиотеки scikit-learn найдем два признака, имеющих наиболее выраженную взаимосвязь с признаком, имеющим максимальную дисперсию (в данном случае это V10). В данном случае я удаляю признаки с низкой дисперсией, нормализую данные и вычисляю 2 признака. Так как V10 является непрерывным признаком, то мы используем статистический критерий f\_regression.

```
In [21]: from sklearn.feature_selection import SelectKBest, f_regression
from sklearn.feature_selection import VarianceThreshold
from sklearn import preprocessing
sel = VarianceThreshold(threshold=(.8 * (1 - .8)))
X = my_data.drop(['V10'], axis=1).values
X = sel.fit_transform(X)
min_max_scaler = preprocessing.MinMaxScaler()
X = min_max_scaler.fit_transform(X)
Y = my_data["V10"].values
test = SelectKBest(score_func=f_regression, k=2)
fit = test.fit(X, Y)
```

```
In [22]: cols = test.get_support(indices=True)
df_new = my_data.iloc[:,cols]
print("\nОтобранные признаки:\n",df_new)
```

Отобранные признаки:

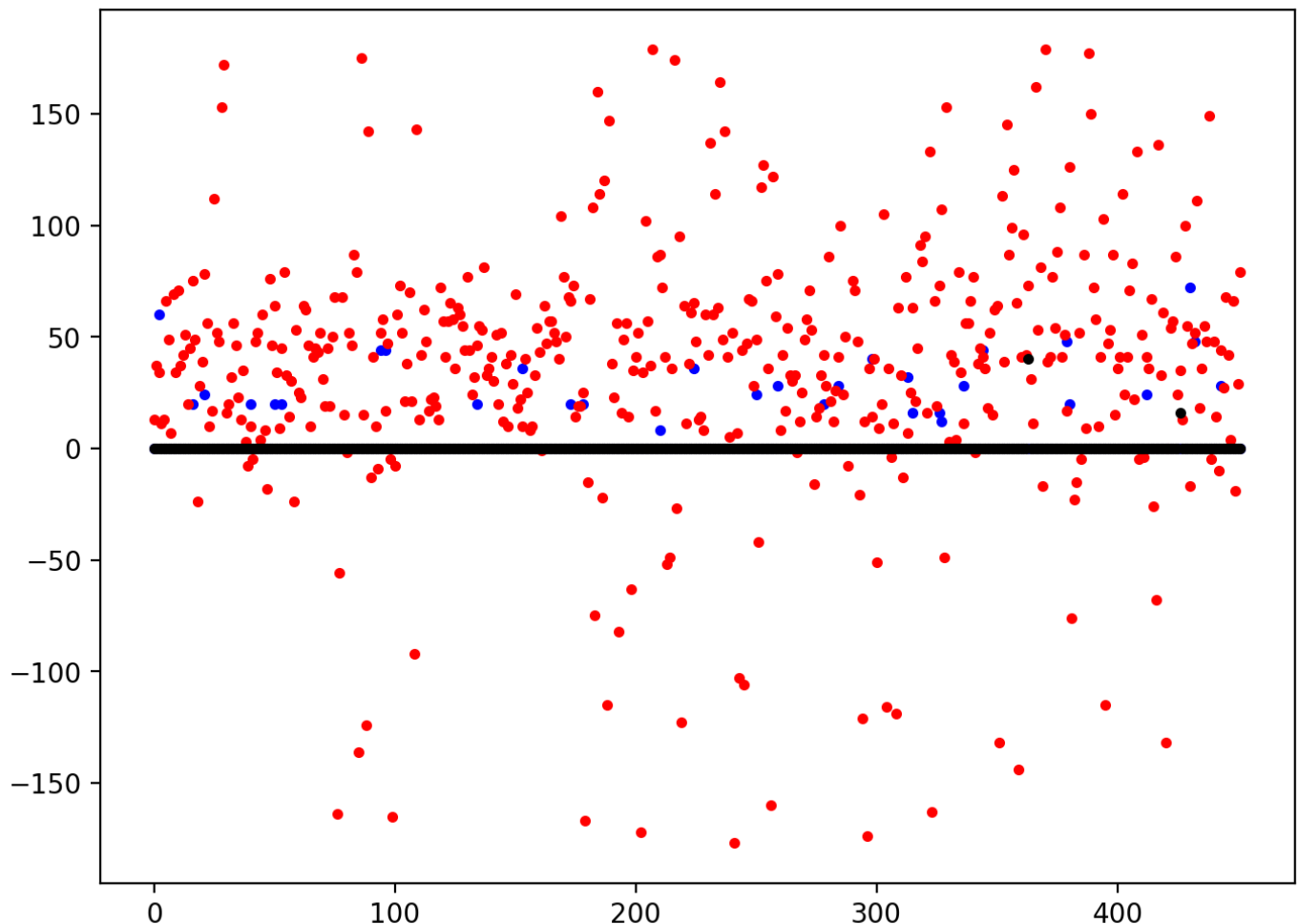
	V102	V127
0	0	0
1	0	0
2	60	0
3	0	0
4	0	0
..	...	...
447	0	0
448	0	0
449	0	0
450	0	0
451	0	0

[452 rows x 2 columns]

1. Визуализируем набор данных в виде точек плоскости с координатами, соответствующими найденным признакам.

```
In [23]: %matplotlib inline
import matplotlib.pyplot as plt

plt.figure( figsize=(8, 6), dpi=200 )
plt.plot(my_data["V102"].values, 'b. ');
plt.plot(my_data["V10"].values, 'r. ');
plt.plot(my_data["V127"].values, 'k. ');
```



1. Найдем и выведем на экран размерность метода главных компонент, для которой доля объясняемой дисперсии будет не менее 99%

```
In [24]: from sklearn.decomposition import PCA

enough_r = 1
flag = False
for r in range(1,281):
    pca = PCA( n_components = r )
    pca.fit( my_data.values )
    tmp = sum(pca.explained_variance_ratio_)
    print( "r =",r,"\tДисперсия =", tmp * 100, "%" )
    if (tmp > .99) and (not flag):
        enough_r = r
        flag = True
print("Для значения r = {} дисперсия будет не менее 99%".format(enough_r))
```

r = 1	Дисперсия = 15.661270826761758 %
r = 2	Дисперсия = 25.621834571661832 %
r = 3	Дисперсия = 34.82012218984459 %
r = 4	Дисперсия = 43.3100505312845 %
r = 5	Дисперсия = 48.68922027144323 %
r = 6	Дисперсия = 53.89765807249073 %
r = 7	Дисперсия = 58.43587886941775 %
r = 8	Дисперсия = 62.55942676916963 %
r = 9	Дисперсия = 65.88593234550922 %
r = 10	Дисперсия = 68.8291648445902 %
r = 11	Дисперсия = 71.65692750160963 %
r = 12	Дисперсия = 73.89987418544015 %
r = 13	Дисперсия = 75.99648819871088 %
r = 14	Дисперсия = 77.93009301927574 %
r = 15	Дисперсия = 79.65151827592841 %
r = 16	Дисперсия = 81.19509111067471 %
r = 17	Дисперсия = 82.53503464322446 %
r = 18	Дисперсия = 83.76036131123608 %
r = 19	Дисперсия = 84.9464448608823 %
r = 20	Дисперсия = 85.85996271381056 %
r = 21	Дисперсия = 86.7268275700529 %
r = 22	Дисперсия = 87.48950495659449 %
r = 23	Дисперсия = 88.20655780492697 %
r = 24	Дисперсия = 88.88054110387925 %
r = 25	Дисперсия = 89.51266709923107 %
r = 26	Дисперсия = 90.09712344997996 %
r = 27	Дисперсия = 90.63683627934238 %
r = 28	Дисперсия = 91.1636283842211 %
r = 29	Дисперсия = 91.66253074354998 %
r = 30	Дисперсия = 92.14066392248559 %
r = 31	Дисперсия = 92.58698803821079 %
r = 32	Дисперсия = 93.00681740887971 %
r = 33	Дисперсия = 93.40688292742202 %
r = 34	Дисперсия = 93.78174781988517 %
r = 35	Дисперсия = 94.13085710877489 %
r = 36	Дисперсия = 94.45919292561562 %
r = 37	Дисперсия = 94.76296012446637 %
r = 38	Дисперсия = 95.06019970772076 %
r = 39	Дисперсия = 95.34392702133229 %
r = 40	Дисперсия = 95.59680679353735 %
r = 41	Дисперсия = 95.83974940103413 %
r = 42	Дисперсия = 96.06335350070736 %
r = 43	Дисперсия = 96.28107209000561 %
r = 44	Дисперсия = 96.49642481187986 %
r = 45	Дисперсия = 96.70564460791225 %
r = 46	Дисперсия = 96.90121191164032 %
r = 47	Дисперсия = 97.09501681520263 %
r = 48	Дисперсия = 97.26434152078525 %
r = 49	Дисперсия = 97.42282718568379 %
r = 50	Дисперсия = 97.56718068154385 %
r = 51	Дисперсия = 97.70208321000791 %
r = 52	Дисперсия = 97.82927200939723 %
r = 53	Дисперсия = 97.95208824640429 %
r = 54	Дисперсия = 98.07017498131319 %
r = 55	Дисперсия = 98.18667037790904 %
r = 56	Дисперсия = 98.29390301387059 %
r = 57	Дисперсия = 98.39459139264939 %
r = 58	Дисперсия = 98.49482885013519 %
r = 59	Дисперсия = 98.5918170257402 %
r = 60	Дисперсия = 98.68186366580028 %
r = 61	Дисперсия = 98.76761565295502 %
r = 62	Дисперсия = 98.84632349508864 %
r = 63	Дисперсия = 98.92077295301608 %

r = 64 Дисперсия = 98.98629319742804 %  
r = 65 Дисперсия = 99.0500811850969 %  
r = 66 Дисперсия = 99.1103413413353 %  
r = 67 Дисперсия = 99.1669133448843 %  
r = 68 Дисперсия = 99.22114337472809 %  
r = 69 Дисперсия = 99.26956866353017 %  
r = 70 Дисперсия = 99.3139911607824 %  
r = 71 Дисперсия = 99.35322756048252 %  
r = 72 Дисперсия = 99.3910067961736 %  
r = 73 Дисперсия = 99.42699469020485 %  
r = 74 Дисперсия = 99.4625279545551 %  
r = 75 Дисперсия = 99.49605812078286 %  
r = 76 Дисперсия = 99.52620514474971 %  
r = 77 Дисперсия = 99.55469031438085 %  
r = 78 Дисперсия = 99.58306272425601 %  
r = 79 Дисперсия = 99.60921882632994 %  
r = 80 Дисперсия = 99.6339393422834 %  
r = 81 Дисперсия = 99.65809761836059 %  
r = 82 Дисперсия = 99.68080659172503 %  
r = 83 Дисперсия = 99.7022870163289 %  
r = 84 Дисперсия = 99.72292580491157 %  
r = 85 Дисперсия = 99.74221056279787 %  
r = 86 Дисперсия = 99.76059951246735 %  
r = 87 Дисперсия = 99.77751539618428 %  
r = 88 Дисперсия = 99.79329953466697 %  
r = 89 Дисперсия = 99.80817443632701 %  
r = 90 Дисперсия = 99.82210479829074 %  
r = 91 Дисперсия = 99.83542418389624 %  
r = 92 Дисперсия = 99.84800524130095 %  
r = 93 Дисперсия = 99.85991085343565 %  
r = 94 Дисперсия = 99.87040807521711 %  
r = 95 Дисперсия = 99.87989171324796 %  
r = 96 Дисперсия = 99.88837619299605 %  
r = 97 Дисперсия = 99.89659317885024 %  
r = 98 Дисперсия = 99.90427662361833 %  
r = 99 Дисперсия = 99.91127949000801 %  
r = 100 Дисперсия = 99.9178450065159 %  
r = 101 Дисперсия = 99.9238694805254 %  
r = 102 Дисперсия = 99.92898150122627 %  
r = 103 Дисперсия = 99.93399357903468 %  
r = 104 Дисперсия = 99.93862656324363 %  
r = 105 Дисперсия = 99.94306028038412 %  
r = 106 Дисперсия = 99.94726376619136 %  
r = 107 Дисперсия = 99.9514315071439 %  
r = 108 Дисперсия = 99.9553047046044 %  
r = 109 Дисперсия = 99.95897725404322 %  
r = 110 Дисперсия = 99.96192804511715 %  
r = 111 Дисперсия = 99.96486519940716 %  
r = 112 Дисперсия = 99.96748131768187 %  
r = 113 Дисперсия = 99.96990938992522 %  
r = 114 Дисперсия = 99.97227698814525 %  
r = 115 Дисперсия = 99.97441480155608 %  
r = 116 Дисперсия = 99.97649812075832 %  
r = 117 Дисперсия = 99.97834193551422 %  
r = 118 Дисперсия = 99.98012481244953 %  
r = 119 Дисперсия = 99.98170705536013 %  
r = 120 Дисперсия = 99.98324440381407 %  
r = 121 Дисперсия = 99.984499334931 %  
r = 122 Дисперсия = 99.98570038521636 %  
r = 123 Дисперсия = 99.98675136548638 %  
r = 124 Дисперсия = 99.98776618703353 %  
r = 125 Дисперсия = 99.98873014003041 %  
r = 126 Дисперсия = 99.98966875678504 %  
r = 127 Дисперсия = 99.99046832119852 %  
r = 128 Дисперсия = 99.99115111289207 %  
r = 129 Дисперсия = 99.99177935533969 %

r = 130	Дисперсия = 99.9923508819734 %
r = 131	Дисперсия = 99.9928901551227 %
r = 132	Дисперсия = 99.99338972186523 %
r = 133	Дисперсия = 99.99382300359959 %
r = 134	Дисперсия = 99.99421268096 %
r = 135	Дисперсия = 99.9945640460681 %
r = 136	Дисперсия = 99.99487960320823 %
r = 137	Дисперсия = 99.99518518186434 %
r = 138	Дисперсия = 99.99547881949078 %
r = 139	Дисперсия = 99.99575393890392 %
r = 140	Дисперсия = 99.99599809900461 %
r = 141	Дисперсия = 99.99623574082617 %
r = 142	Дисперсия = 99.9964511541938 %
r = 143	Дисперсия = 99.99665931869207 %
r = 144	Дисперсия = 99.9968625565659 %
r = 145	Дисперсия = 99.99704656882993 %
r = 146	Дисперсия = 99.99720971367266 %
r = 147	Дисперсия = 99.99736800634163 %
r = 148	Дисперсия = 99.9975215544825 %
r = 149	Дисперсия = 99.99766591325965 %
r = 150	Дисперсия = 99.99780570177914 %
r = 151	Дисперсия = 99.99793120757413 %
r = 152	Дисперсия = 99.99804790788801 %
r = 153	Дисперсия = 99.99815582884536 %
r = 154	Дисперсия = 99.9982610662385 %
r = 155	Дисперсия = 99.99835528294925 %
r = 156	Дисперсия = 99.99844258416651 %
r = 157	Дисперсия = 99.99852658702656 %
r = 158	Дисперсия = 99.99860038040642 %
r = 159	Дисперсия = 99.99866959409157 %
r = 160	Дисперсия = 99.99873670793556 %
r = 161	Дисперсия = 99.99880288450578 %
r = 162	Дисперсия = 99.99886742219662 %
r = 163	Дисперсия = 99.99892717943803 %
r = 164	Дисперсия = 99.99898428118618 %
r = 165	Дисперсия = 99.9990369522872 %
r = 166	Дисперсия = 99.99908603048091 %
r = 167	Дисперсия = 99.99913095180253 %
r = 168	Дисперсия = 99.99917232194863 %
r = 169	Дисперсия = 99.99921232934416 %
r = 170	Дисперсия = 99.99924978885085 %
r = 171	Дисперсия = 99.99928564628708 %
r = 172	Дисперсия = 99.99931920666643 %
r = 173	Дисперсия = 99.99935191111675 %
r = 174	Дисперсия = 99.99938275560227 %
r = 175	Дисперсия = 99.99941142866095 %
r = 176	Дисперсия = 99.99943887444417 %
r = 177	Дисперсия = 99.99946456243836 %
r = 178	Дисперсия = 99.99948959398945 %
r = 179	Дисперсия = 99.99951318029062 %
r = 180	Дисперсия = 99.9995361452634 %
r = 181	Дисперсия = 99.99955839180663 %
r = 182	Дисперсия = 99.99957960136904 %
r = 183	Дисперсия = 99.99959920263115 %
r = 184	Дисперсия = 99.99961827950254 %
r = 185	Дисперсия = 99.99963622054653 %
r = 186	Дисперсия = 99.99965297924062 %
r = 187	Дисперсия = 99.99966909033402 %
r = 188	Дисперсия = 99.99968507932263 %
r = 189	Дисперсия = 99.99970011755866 %
r = 190	Дисперсия = 99.99971457317109 %
r = 191	Дисперсия = 99.9997284395646 %
r = 192	Дисперсия = 99.99974171099082 %
r = 193	Дисперсия = 99.99975425611711 %
r = 194	Дисперсия = 99.99976636474348 %
r = 195	Дисперсия = 99.9997780791794 %

r = 196	Дисперсия = 99.99978946462197 %
r = 197	Дисперсия = 99.99980002254088 %
r = 198	Дисперсия = 99.99981017474069 %
r = 199	Дисперсия = 99.99982005016513 %
r = 200	Дисперсия = 99.99982951784239 %
r = 201	Дисперсия = 99.99983853474583 %
r = 202	Дисперсия = 99.99984725160587 %
r = 203	Дисперсия = 99.99985520475303 %
r = 204	Дисперсия = 99.99986259524533 %
r = 205	Дисперсия = 99.99986958153026 %
r = 206	Дисперсия = 99.99987628092202 %
r = 207	Дисперсия = 99.99988268776326 %
r = 208	Дисперсия = 99.99988886432345 %
r = 209	Дисперсия = 99.99989488098886 %
r = 210	Дисперсия = 99.99990068223555 %
r = 211	Дисперсия = 99.99990626677379 %
r = 212	Дисперсия = 99.99991159171725 %
r = 213	Дисперсия = 99.99991656939889 %
r = 214	Дисперсия = 99.99992149910743 %
r = 215	Дисперсия = 99.99992637102841 %
r = 216	Дисперсия = 99.99993097214168 %
r = 217	Дисперсия = 99.99993533199384 %
r = 218	Дисперсия = 99.99993956496036 %
r = 219	Дисперсия = 99.99994365824345 %
r = 220	Дисперсия = 99.99994757230229 %
r = 221	Дисперсия = 99.99995123620413 %
r = 222	Дисперсия = 99.99995472924776 %
r = 223	Дисперсия = 99.99995803405868 %
r = 224	Дисперсия = 99.99996132478094 %
r = 225	Дисперсия = 99.99996430246213 %
r = 226	Дисперсия = 99.99996713144694 %
r = 227	Дисперсия = 99.99996978298273 %
r = 228	Дисперсия = 99.99997240609402 %
r = 229	Дисперсия = 99.99997485402177 %
r = 230	Дисперсия = 99.99997713158493 %
r = 231	Дисперсия = 99.99997935500731 %
r = 232	Дисперсия = 99.99998140360093 %
r = 233	Дисперсия = 99.99998334949117 %
r = 234	Дисперсия = 99.99998493895096 %
r = 235	Дисперсия = 99.99998647995453 %
r = 236	Дисперсия = 99.99998794590365 %
r = 237	Дисперсия = 99.99998931350767 %
r = 238	Дисперсия = 99.99999053001225 %
r = 239	Дисперсия = 99.99999172629484 %
r = 240	Дисперсия = 99.99999276639106 %
r = 241	Дисперсия = 99.99999376952918 %
r = 242	Дисперсия = 99.99999464014762 %
r = 243	Дисперсия = 99.99999545081224 %
r = 244	Дисперсия = 99.9999962187938 %
r = 245	Дисперсия = 99.99999690625062 %
r = 246	Дисперсия = 99.99999744958521 %
r = 247	Дисперсия = 99.9999979732061 %
r = 248	Дисперсия = 99.99999843505694 %
r = 249	Дисперсия = 99.99999881595173 %
r = 250	Дисперсия = 99.99999912041821 %
r = 251	Дисперсия = 99.9999993877052 %
r = 252	Дисперсия = 99.9999996252144 %
r = 253	Дисперсия = 99.999999802035 %
r = 254	Дисперсия = 99.99999990854236 %
r = 255	Дисперсия = 99.99999995790529 %
r = 256	Дисперсия = 99.99999999041846 %
r = 257	Дисперсия = 99.99999999765114 %
r = 258	Дисперсия = 99.9999999929328 %
r = 259	Дисперсия = 99.9999999999997 %
r = 260	Дисперсия = 99.9999999999997 %
r = 261	Дисперсия = 99.9999999999997 %



```

r = 262          Дисперсия = 99.9999999999997 %
r = 263          Дисперсия = 99.9999999999997 %
r = 264          Дисперсия = 99.9999999999997 %
r = 265          Дисперсия = 99.9999999999997 %
r = 266          Дисперсия = 99.9999999999997 %
r = 267          Дисперсия = 99.9999999999997 %
r = 268          Дисперсия = 99.9999999999997 %
r = 269          Дисперсия = 99.9999999999997 %
r = 270          Дисперсия = 99.9999999999997 %
r = 271          Дисперсия = 99.9999999999997 %
r = 272          Дисперсия = 99.9999999999997 %
r = 273          Дисперсия = 99.9999999999997 %
r = 274          Дисперсия = 99.9999999999997 %
r = 275          Дисперсия = 99.9999999999997 %
r = 276          Дисперсия = 99.9999999999997 %
r = 277          Дисперсия = 99.9999999999997 %
r = 278          Дисперсия = 99.9999999999997 %
r = 279          Дисперсия = 99.9999999999997 %
r = 280          Дисперсия = 99.9999999999997 %
Для значения r = 65 дисперсия будет не менее 99%

```

1. Пользуясь методом главных компонент, снизим размерность набора данных до двух признаков и изобразим полученный набор данных в виде точек на плоскости, отображая точки различных классов разными цветами.

```

In [26]: X = my_data.values
pca = PCA(n_components=2)
pcad = pca.fit_transform( X )
fit = pca.fit(X)
features = fit.transform(X)
X1 = []
X2 = []
for i in features:
    X1.append(i[0])
    X2.append(i[1])
plt.figure( figsize=(8, 6), dpi=200 )
plt.plot(X1, 'b. ');
plt.plot(X2, 'r. ');

```

