

hw1.2

28 февраля 2017 г.

```
In [1]: %pylab inline
```

Populating the interactive namespace from numpy and matplotlib

1 Loading datasets

```
In [2]: from sklearn.datasets import load_breast_cancer, load_digits
import pandas as pd
from pandas import DataFrame
```

```
In [3]: breast_cancer_frame = DataFrame(load_breast_cancer().data)
pd.options.display.max_columns = 15
breast_cancer_frame.head()
```

```
Out[3]:
```

	0	1	2	3	4	5	6	...	23	\
0	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.3001	...	2019.0	
1	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.0869	...	1956.0	
2	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.1974	...	1709.0	
3	11.42	20.38	77.58	386.1	0.14250	0.28390	0.2414	...	567.7	
4	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.1980	...	1575.0	

	24	25	26	27	28	29
0	0.1622	0.6656	0.7119	0.2654	0.4601	0.11890
1	0.1238	0.1866	0.2416	0.1860	0.2750	0.08902
2	0.1444	0.4245	0.4504	0.2430	0.3613	0.08758
3	0.2098	0.8663	0.6869	0.2575	0.6638	0.17300
4	0.1374	0.2050	0.4000	0.1625	0.2364	0.07678

[5 rows x 30 columns]

```
In [4]: digits_frame = DataFrame(load_digits().data)
digits_frame.head()
```

```
Out[4]:
```

	0	1	2	3	4	5	6	...	57	58	59	60	61	62	\
0	0.0	0.0	5.0	13.0	9.0	1.0	0.0	...	0.0	6.0	13.0	10.0	0.0	0.0	
1	0.0	0.0	0.0	12.0	13.0	5.0	0.0	...	0.0	0.0	11.0	16.0	10.0	0.0	
2	0.0	0.0	0.0	4.0	15.0	12.0	0.0	...	0.0	0.0	3.0	11.0	16.0	9.0	

```

3  0.0  0.0  7.0  15.0  13.0   1.0  0.0 ...   0.0  7.0  13.0  13.0   9.0  0.0
4  0.0  0.0  0.0   1.0  11.0   0.0  0.0 ...   0.0  0.0   2.0  16.0   4.0  0.0

```

```

63
0  0.0
1  0.0
2  0.0
3  0.0
4  0.0

```

[5 rows x 64 columns]

2 Loading estimators

```
In [5]: from sklearn.model_selection import cross_val_score
```

```
In [6]: from sklearn.naive_bayes import BernoulliNB, MultinomialNB, GaussianNB
```

```
In [7]: bern = BernoulliNB()
        mult = MultinomialNB()
        gauss = GaussianNB()
```

```
In [8]: def score(X, y, estimator):
        return mean(cross_val_score(estimator, X, y))
```

3 Breast cancer dataset

```
In [9]: X, y = load_breast_cancer(return_X_y=True)
        print("BernoulliNB score is {:.3}".format(score(X, y, bern)))
        print("MultinomialNB score is {:.3}".format(score(X, y, mult)))
        print("GaussianNB score is {:.3}".format(score(X, y, gauss)))
```

```

BernoulliNB score is 0.627
MultinomialNB score is 0.895
GaussianNB score is 0.937

```

4 Digits dataset

```
In [10]: X, y = load_digits(return_X_y=True)
         print("BernoulliNB score is {:.3}".format(score(X, y, bern)))
         print("MultinomialNB score is {:.3}".format(score(X, y, mult)))
         print("GaussianNB score is {:.3}".format(score(X, y, gauss)))
```

```

BernoulliNB score is 0.826
MultinomialNB score is 0.871
GaussianNB score is 0.819

```

5 Results

1. The maximum score for breast_cancer dataset is 0.937
2. The maximum score for digits dataset is 0.871
3. Correct statements are (c) and (d)