Vyhodnocovanie

November 23, 2017

1 Obsah

1.1 Trenovacia / testovacia / validacna vzorka

1.2 Krizova validacia

1.3 Metriky vyhodnocovania

1.4 Hyperparameter tuning

1.5 Overfitting

1.6 Variance / Bias

2 Motivacia pre vyhodnocovanie klasifikatorov

Existuje sptrane vela pristupov a typov klasifikatorov. ** Ktory z nich je najlepsi? **

Rozne klasifikatory maju rozne parametre. ** Ako nastavit klasifikator tak, aby daval najlepsie vysledky? **

```
In [1]: %matplotlib inline
        import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        import seaborn

In [2]: import warnings
        warnings.filterwarnings('ignore')

In [3]: plt.rcParams['figure.figsize'] = 9, 6
```

3 Dataset

```
In [4]: from sklearn.datasets import load_breast_cancer

In [5]: data = load_breast_cancer()

        label_names = data['target_names']
        labels = data['target']
```

```
feature_names = data['feature_names']
features = data['data']
```

```
In [6]: label_names
```

```
Out[6]: array(['malignant', 'benign'],
              dtype='<U9')
```

```
In [7]: labels[-10:]
```

```
Out[7]: array([1, 1, 1, 0, 0, 0, 0, 0, 0, 1])
```

```
In [8]: feature_names
```

```
Out[8]: array(['mean radius', 'mean texture', 'mean perimeter', 'mean area',
               'mean smoothness', 'mean compactness', 'mean concavity',
               'mean concave points', 'mean symmetry', 'mean fractal dimension',
               'radius error', 'texture error', 'perimeter error', 'area error',
               'smoothness error', 'compactness error', 'concavity error',
               'concave points error', 'symmetry error', 'fractal dimension error',
               'worst radius', 'worst texture', 'worst perimeter', 'worst area',
               'worst smoothness', 'worst compactness', 'worst concavity',
               'worst concave points', 'worst symmetry', 'worst fractal dimension'],
              dtype='<U23')
```

```
In [9]: features
```

```
Out[9]: array([[  1.79900000e+01,   1.03800000e+01,   1.22800000e+02, ...,
                  2.65400000e-01,   4.60100000e-01,   1.18900000e-01],
               [  2.05700000e+01,   1.77700000e+01,   1.32900000e+02, ...,
                  1.86000000e-01,   2.75000000e-01,   8.90200000e-02],
               [  1.96900000e+01,   2.12500000e+01,   1.30000000e+02, ...,
                  2.43000000e-01,   3.61300000e-01,   8.75800000e-02],
               ...,
               [  1.66000000e+01,   2.80800000e+01,   1.08300000e+02, ...,
                  1.41800000e-01,   2.21800000e-01,   7.82000000e-02],
               [  2.06000000e+01,   2.93300000e+01,   1.40100000e+02, ...,
                  2.65000000e-01,   4.08700000e-01,   1.24000000e-01],
               [  7.76000000e+00,   2.45400000e+01,   4.79200000e+01, ...,
                  0.00000000e+00,   2.87100000e-01,   7.03900000e-02]])
```

```
In [10]: df = pd.DataFrame(features)
         df.columns = feature_names
         df['target'] = labels
         df.head()
```

```
Out[10]:    mean radius  mean texture  mean perimeter  mean area  mean smoothness  \
         0        17.99         10.38          122.80     1001.0          0.11840
         1        20.57         17.77          132.90     1326.0          0.08474
```