拆分训练集与测试集(留出法)

```
In [1]:  import numpy as np
```

```
In [2]:  X = np.loadtxt('x.txt')
         X.shape

Out[2]:  (150, 4)
```

```
In [3]:  y = np.loadtxt('y.txt')
         y.shape

Out[3]:  (150,)
```

```
In [4]:  np.random.seed(1)
         shuffle = np.random.permutation(len(X))        # 生成0到len(X)之间的
         所有不重复整数的随机排列
```

```
In [5]:  test_size = 0.25
         size = int(len(X) * test_size)
         size

Out[5]:  37
```

```
In [6]:  test_index = shuffle[:size]
         train_index = shuffle[size:]
```

```
In [7]:  test_index

Out[7]:  array([ 14,  98,  75,  16, 131,  56, 141,  44,  29, 120,  94,   5,
         102,
                 51,  78,  42,  92,  66,  31,  35,  90,  84,  77,  40, 125,
         99,
                 33,  19,  73, 146,  91, 135,  69, 128, 114,  48,  53])
```

```
In [8]:  x_train = X[train_index]
```

```
In [9]:  y_train = y[train_index]
```

```
In [10]: x_test = X[test_index]
```

```
In [11]: y_test = y[test_index]
```

```
In [12]: from ML.knn import kNN_classify
```

```
In [13]: predict_y = kNN_classify(x_train, y_train, x_test)
In [14]: y_predict = np.array(predict_y)

In [15]: y_predict

Out[15]: array([0., 1., 1., 0., 2., 1., 2., 0., 0., 2., 1., 0., 2., 1., 1.,
         0., 1.,
                1., 0., 0., 1., 1., 1., 0., 2., 1., 0., 0., 1., 2., 1., 2.,
         1., 2.,
                2., 0., 1.])

In [16]: y_test

Out[16]: array([0., 1., 1., 0., 2., 1., 2., 0., 0., 2., 1., 0., 2., 1., 1.,
         0., 1.,
                1., 0., 0., 1., 1., 1., 0., 2., 1., 0., 0., 1., 2., 1., 2.,
         1., 2.,
                2., 0., 1.])

In [17]: sum(y_predict == y_test)

Out[17]: 37

In [18]: sum(y_predict == y_test) / len(x_test)

Out[18]: 1.0

In [19]: from ML.model_selection import train_test_split

In [20]: x_train, y_train, x_test, y_test = train_test_split(X, y, test_size
         =0.25, seed=1)

In [21]: predict_y = np.array(kNN_classify(x_train, y_train, x_test))

In [22]: sum(predict_y == y_test) / len(x_test)

Out[22]: 1.0
```