k近邻算法实战鸢尾花分类

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

In [2]:
```
df = pd.read_excel('iris.xlsx')
df.head()
```

Out[2]:

	萼片长(cm)	萼片宽(cm)	花瓣长(cm)	花瓣宽(cm)	分类
0	5.1	3.5	1.4	0.2	0
1	4.9	3.0	1.4	0.2	0
2	4.7	3.2	1.3	0.2	0
3	4.6	3.1	1.5	0.2	0
4	5.0	3.6	1.4	0.2	0

In [3]:
```
X_train = df.values[:, :4]
X_train.shape
```

Out[3]:
```
(150, 4)
```

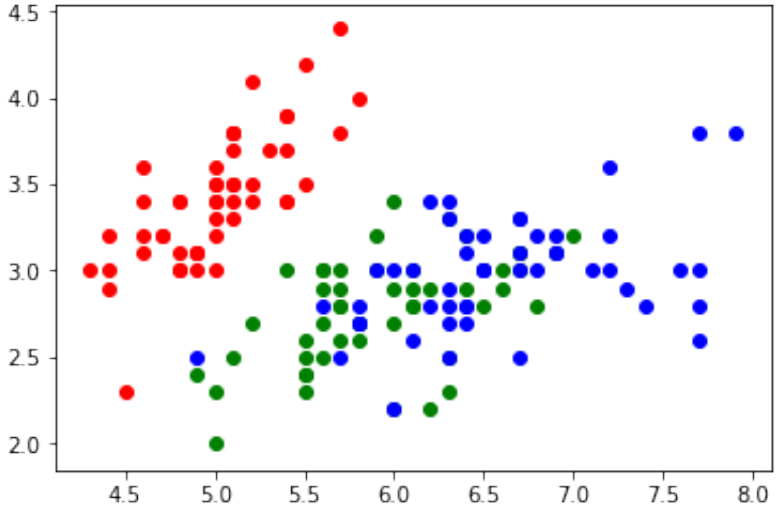
In [4]:
```
y_train = df.values[:, -1]
y_train.shape
```

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

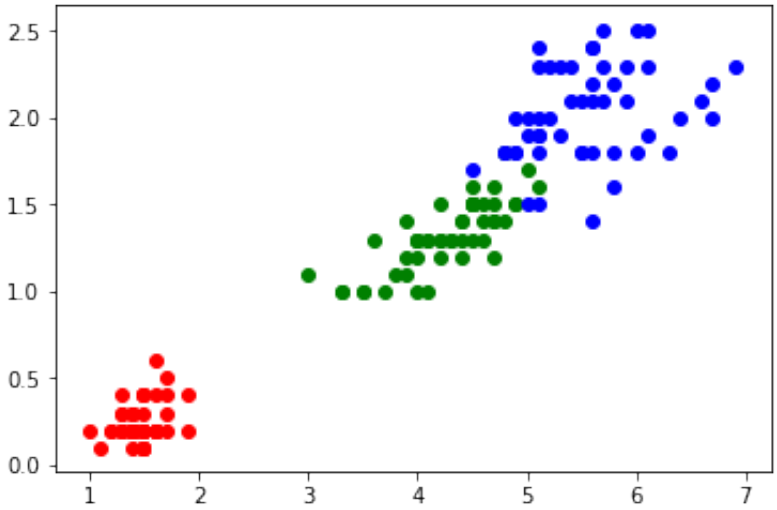
In [5]:
```
y_train
```

Out[5]:
```
array([0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 1.,
       1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
       1., 1.,
       1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
       1., 1.,
       1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.,
       2., 2.,
       2., 2., 2., 2., 2., 2., 2., 2., 2., 2., 2., 2., 2., 2., 2.,
       2., 2.,
       2., 2., 2., 2., 2., 2., 2., 2., 2., 2., 2., 2., 2., 2., 2.,
       2., 2.,
       2., 2., 2., 2., 2., 2., 2., 2., 2., 2., 2., 2., 2., 2., 2.])
```

In [6]:
```
plt.scatter(X_train[y_train==0, 0], X_train[y_train==0, 1], color='
r')
plt.scatter(X_train[y_train==1, 0], X_train[y_train==1, 1], color='
g')
plt.scatter(X_train[y_train==2, 0], X_train[y_train==2, 1], color='
b')
plt.show()
```



In [7]:
```
plt.scatter(X_train[y_train==0, 2], X_train[y_train==0, 3], color='
r')
plt.scatter(X_train[y_train==1, 2], X_train[y_train==1, 3], color='
g')
plt.scatter(X_train[y_train==2, 2], X_train[y_train==2, 3], color='
b')
plt.show()
```



In [8]:
```
from ML.knn import kNN_classify
```

In [9]:
```
predict_X = np.array([[5.1, 3.5, 1.4, 0.2], [5.9, 3. , 5.1, 1.8]])
```

In [10]:
```
kNN_classify(X_train, y_train, predict_X)
```

Out[10]:
```
[0.0, 2.0]
```