课后作业：KNN (K最近邻算法/K近邻算法)

作者：欧新宇（Xinyu OU）

【作业提交】

将分类结果保存到文本文档进行提交，同时提交源代码。

1. 测试结果命名为: ex03-结果-你的学号-你的姓名.txt
2. 源代码命名为: ex03-01-你的学号-你的姓名.py

结果文件，要求每小题标注题号，两题之间要求空一行

使用 “糖尿病预测” 数据集完成以下任务，要求如下：

1. 要求训练集和测试集的分割比例为70%:30%, 给出KNN在训练集和测试集上的分类精度
2. 给定新样本，给出该样本的类别。

样本中各个参数的值为：

- Pregnancies: 【学号//6】
- Glucose：【学号*3】
- BloodPressure：【学号*2】
- SkinThickness：【学号】
- Insulin：【学号*4】
- BMI：30+【学号/7】
- DiabetesPedigreeFunction：【学号/6】
- Age：【学号】

参考代码

```
# 加载 pandas库，并使用read_csv()函数读取糖尿病预测数据集diabetes
import pandas as pd
# data = pd.read_csv('../Datasets/diabetes.csv')  # 载入本地数据集一
# data = pd.read_csv(r'..\Datasets\diabetes.csv')  # 载入本地数据集一
data = pd.read_csv('http://ouxinyu.cn/Teaching/MachineLearning/Datasets/diabetes.csv') # 载入在线数据集

# 将数据中的特征和标签进行分离，其中第0位位索引号，第1-8位位特征，第9位为标签
X = data.iloc[:, 0:8]
y = data.iloc[:, 8]

# 以 70%:30%的比例对训练集和测试集进行拆分
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.3, random_state=10)

# 引入KNN分类模型，并配置KNN分类器，设置近邻数 = 2
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier(n_neighbors = 2)
knn.fit(X_train, y_train)
```

```

train_score = knn.score(X_train, y_train)
test_score = knn.score(X_test, y_test)

print("训练集评分:{0:.2f}；测试集评分:{1:.2f}".format(train_score, test_score))


```

```
import numpy as np
noStudent = 131
X_new = np.array([[noStudent//6, noStudent*3,
                   noStudent*2, noStudent, noStudent*4,
                   noStudent/7, noStudent/6, noStudent]])
prediction = knn.predict(X_new)
print("新样本的分类为：{}".format(prediction))
```

```
新样本的分类为：[1]
```