

# python 实验报告

## 实验名称：药材辨别

院系：人工智能学院

姓名：张运吉

学号：211300063

班级：21 级人工智能学院 AI2 班

邮箱：211300063@smail.nju.edu.cn

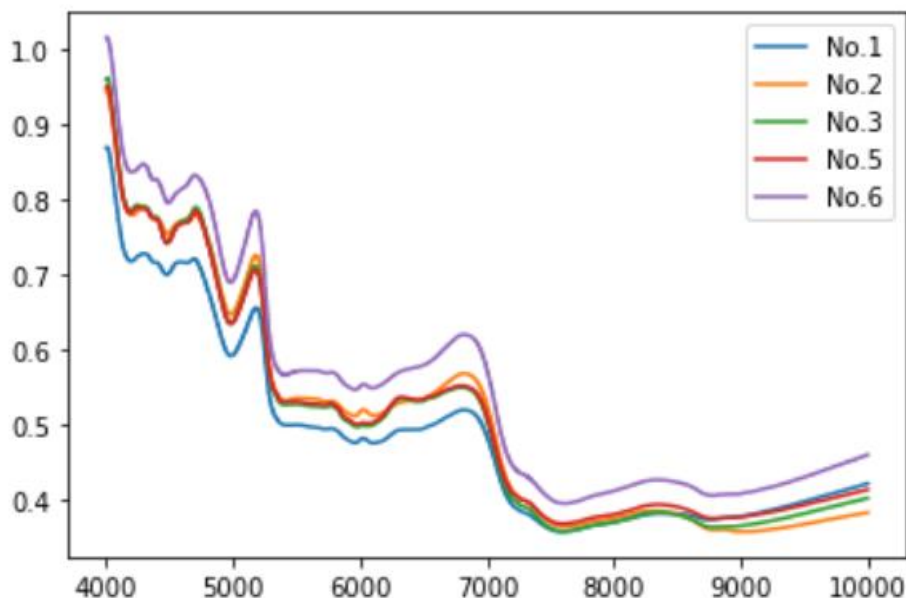
时间：2022 年 5 月 16 日

# 目录

1 可视化: .....	3
2 方案描述: .....	3
3 方案验证: .....	4
4 产地鉴别: .....	4

## 1 可视化：

将问题中编号为 1、2、3、5、6 药材的近红外光谱放在同一幅图中进行可视化。

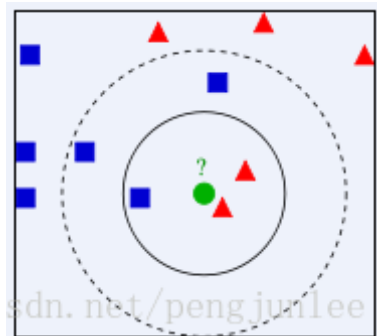


## 2 方案描述：

在网上搜索看过很多种不同的分类算法，由于知识能力受限，只能看得懂几种最简单的算法，在这里我决定使用KNN算法。

为了判断未知样本的类别，以所有已知类别的样本作为参照，计算未知样本与所有已知样本的距离，从中选取与未知样本距离最近的K个已知样本，根据少数服从多数的投票法则（majority-voting），将未知样本与K个最邻近样本中所属类别占比较多的归为一类。

如下图所示，如何判断绿色圆应该属于哪一类，是属于红色三角形还是属于蓝色四方形？如果K=3，由于红色三角形所占比例为 2/3，绿色圆将被判定为属于红色三角形那个类，如果K=5，由于蓝色四方形比例为 3/5，因此绿色圆将被判定为属于蓝色四方形类。



### 3 方案验证：

从附件中随机选出 20 个药材验证所提方案的准确性，并报告方案在该 20 个样本上的准确率。

```
In [118]: lst = []
          for i in range(10):
              arr1 = np.random.randint(228, size=(20,))
              prediction1 = knn.predict(dataset_np[arr1][:, 1:])
              acc = accuracy(prediction1, dataset_np[arr1][:,])
              lst.append(acc)
          print(sum(lst) / 10)
```

92.0

准确率约为 92%.

### 4 产地鉴别：

对问题中的药材产地进行鉴别并填写在报告中。

```
: # 预测结果
prediction = knn.predict(np.array(data)[[0, 1, 2, 3, 5, 6, 7, 9, 11, 13, 26, 31, 60, 65, 69]][:, 2:])
prediction
: array([14., 13., 17., 16.,  5.,  7.,  1., 16.,  4.,  2.,  6.,  8.,  8.,
        1., 11.]
```