

题目： 红酒分类KNN处理过程与结果

**一、实验目的**

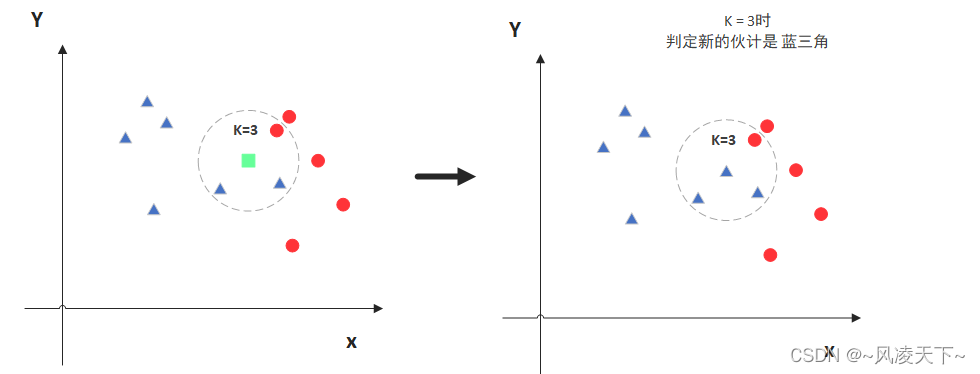
* 熟悉KNN分类算法；
* 通过本实验掌握和利用MATLAB软件，KNN算法，实现对红酒分类应用。

**二、实验环境**

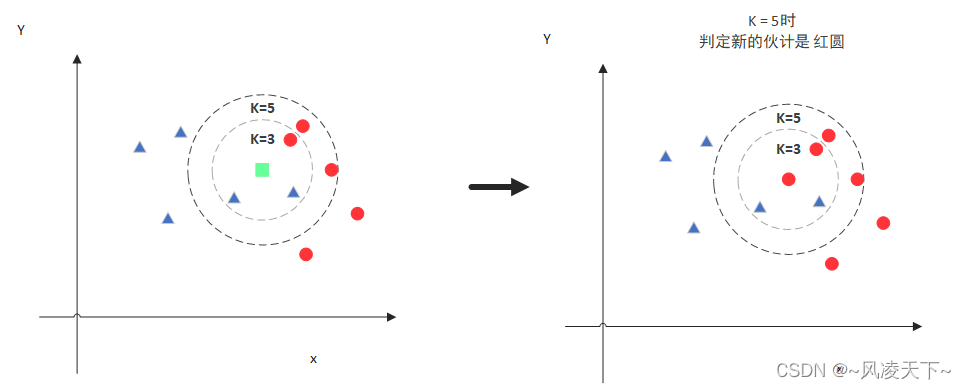
MATLAB 2023a、Win 10 系统

**三、实验原理**

[KNN](https://so.csdn.net/so/search?q=KNN&spm=1001.2101.3001.7020)是最常用的经典分类算法之一，同时，KNN算法是有监督学习中的分类算法，KNN的全称是K Nearest Neighbors，意思是K个最近的邻居，从这个名字我们就能看出一些KNN算法的蛛丝马迹了。K个最近邻居，毫无疑问，K的取值肯定是至关重要的。 =KNN的原理就是当预测一个新的值x的时候，根据它距离最近的K个点是什么类别来判断x属于哪个类别。



图中绿色的点就是我们要预测的那个点，假设K=3。那么KNN算法就会找到与它距离最近的三个点（这里用圆圈把它圈起来了），看看哪种类别多一些，比如这个例子中是蓝色三角形多一些，新来的绿色点就归类到蓝三角了。



但是，当K=5的时候，判定就变成不一样了。这次变成红圆多一些，所以新来的绿点被归类成红圆。从这个例子中，我们就能看得出K的取值是很重要的。

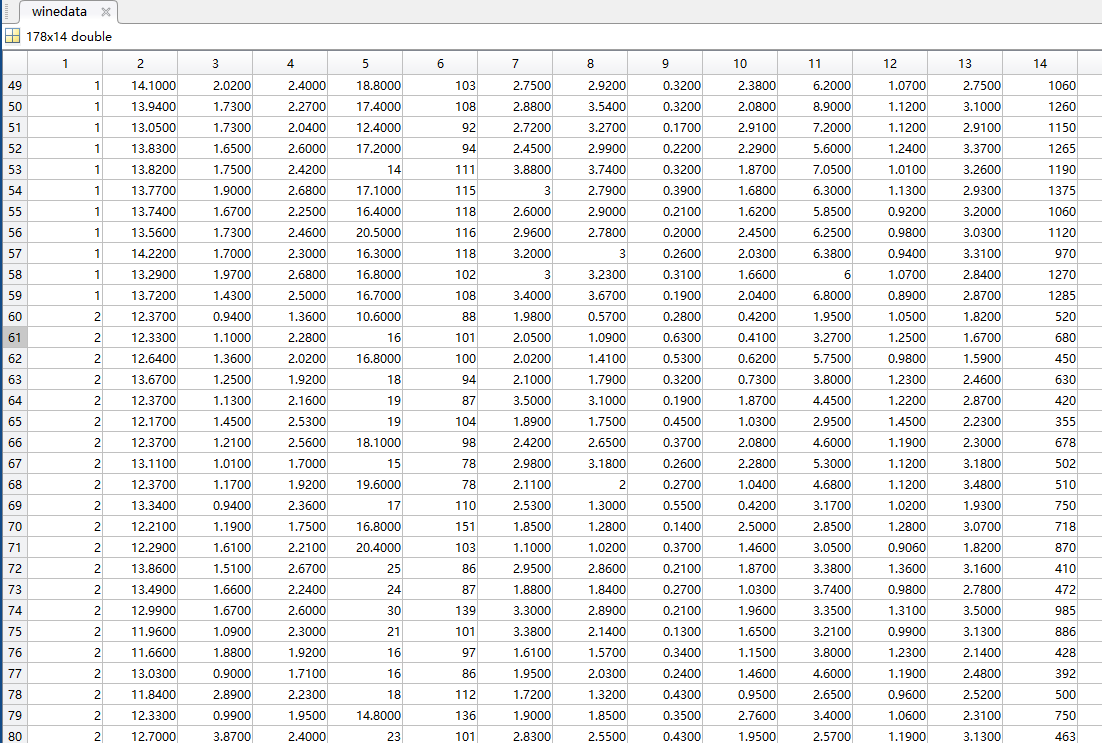
**四、实验过程**

### （1）wine.data数据加载与解析

利用matlab读取、加载与解析wine.data数据。

数据解析后，第一列为标签id,第2-14列有13个特征数据，数据大小为178\*14

输出结果截图：



### 数据分布统计分析

对178组数据进行统计，标签为1的数组有59组，标签为2的数组有71组，

标签为3的数组有48组，数据分布基本均衡

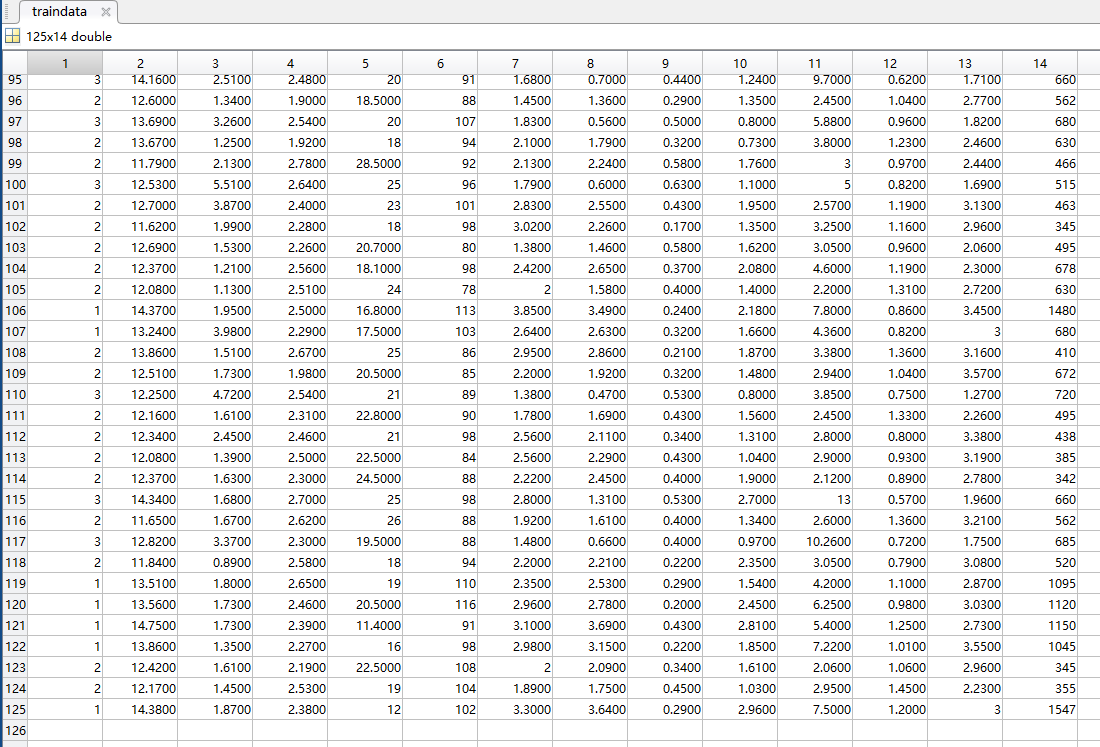
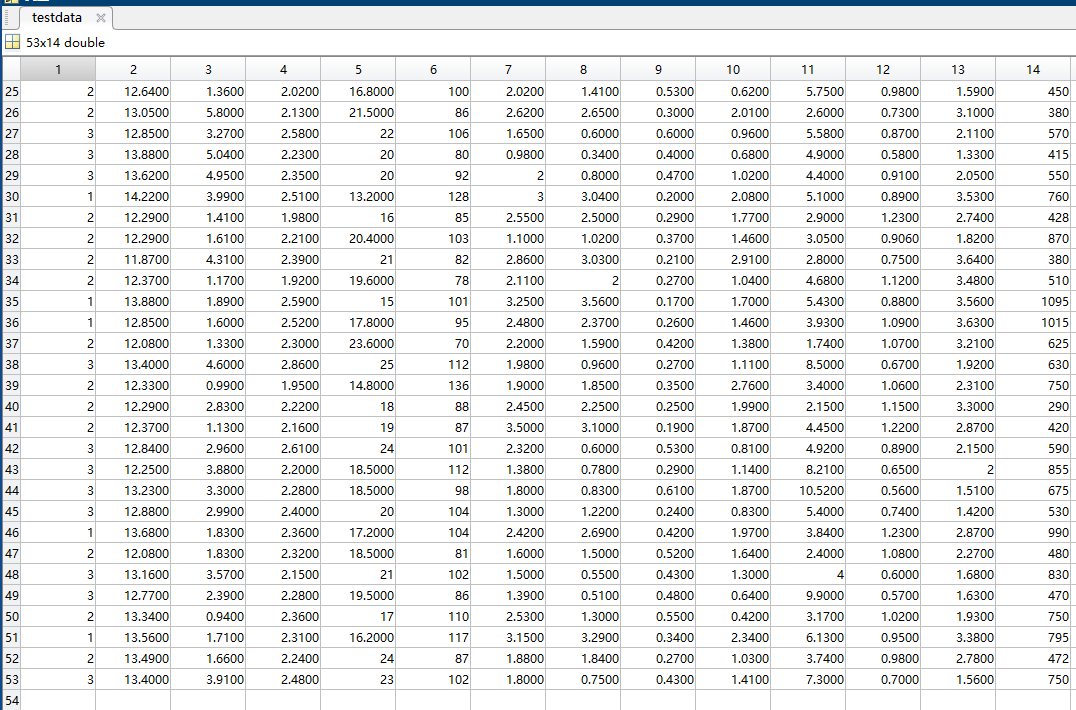
输出结果截图：



### （3）训练集与测试集划分

将178组实验数据按照7比3随机划分训练集与测试集，得到训练集大小为125\*4；测试集大小为53\*4

输出结果截图：

训练集 测试集

### （4）KNN模型训练与测试

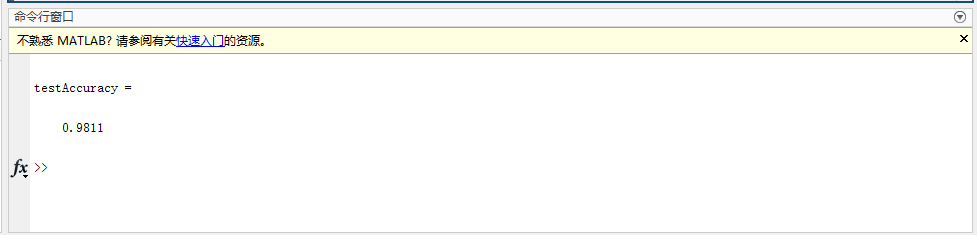
在[matlab](https://so.csdn.net/so/search?q=matlab&spm=1001.2101.3001.7020" \t "https://blog.csdn.net/weixin_44269400/article/details/_blank)中，构建了一个KNN分类的函数为mKNN.m，

[testPredictions ,testAccuracy, modelx] = mKNN(trainData,testData,k,distance);

其中输入参数分别为   
1、trainData：待训练的数据，其中第一列为标签。   
2、testData：待测试的数据，其第一列为标签。   
3、k：分类中，最近邻的个数，默认为10.   
4、distance：计算数据间距离的方法，默认为：cosine:余弦距离   
输出结果截图：

测试集准确率在98.11%；第2类标签有个别预测错误。如下图所示





**五、红酒KNN分类评估结论**

模型评估后，在53组随机测试数据中，有52组测试正确，其中第3类中有一个预测错误，整体正确率为98.11%。

