

SHANGHAI JIAO TONG UNIVERSITY



KNN实验报告

|  |  |
| --- | --- |
| 学生姓名： | 黄书航 |
| 学生学号： |  |
| 专　　业： |  |
| 课程名称： |  |
| 学院(系)： |  |

一、KNN原理

KNN算法，对一个输入，从样本空间中选取k个最相似的样本，将k个样本中大多数样本所属的那个类别作为这个输入的分类结果。

KNN算法没有训练过程，只有将训练用的正确分类样本和标签保存，在分类时将输入与所有样本进行相似度比较和邻居类别统计，在一些问题中可以取得比较好的效果。

二、实验

1．补全predict函数

代码见图1.

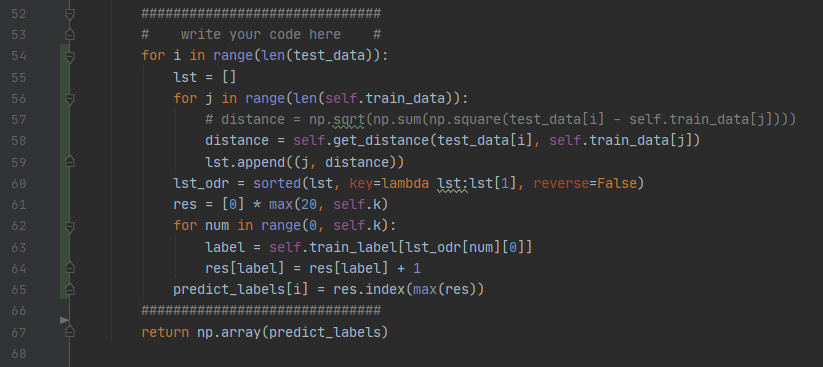
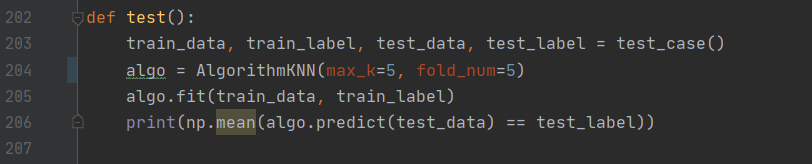


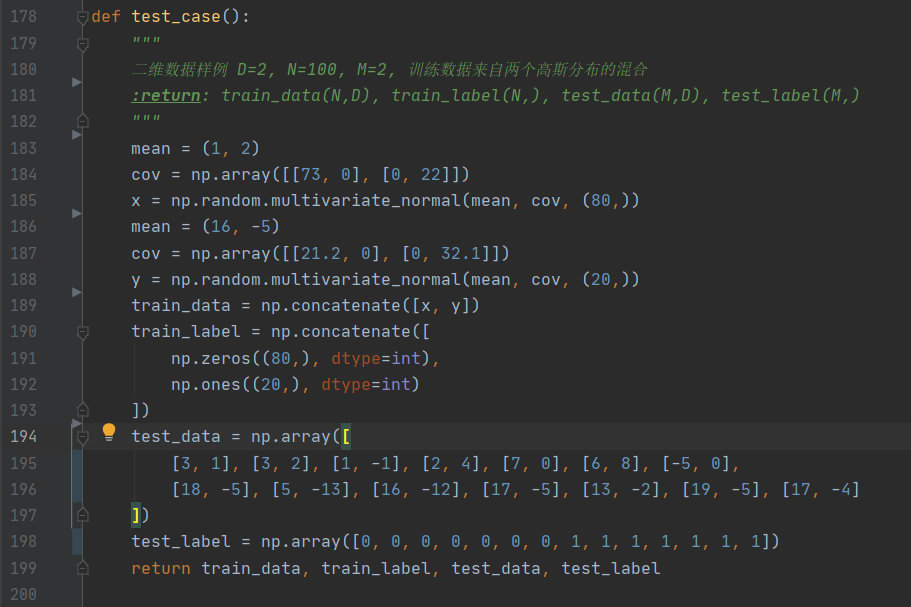
图 1

相似度的计算为输入向量和训练向量的距离或距离。

2. 实验

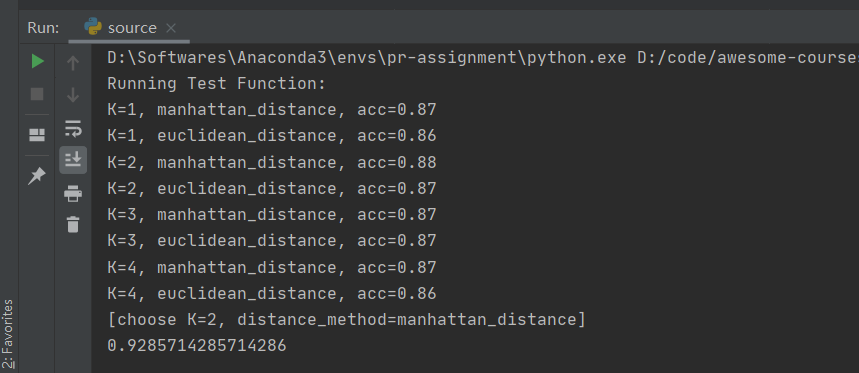
1）第一个实验为原来的代码中自带的测试函数。





训练集为自动生成的高斯分布混合而成，训练集为自行增加的11个数据。

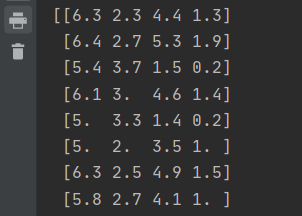
实验结果为：



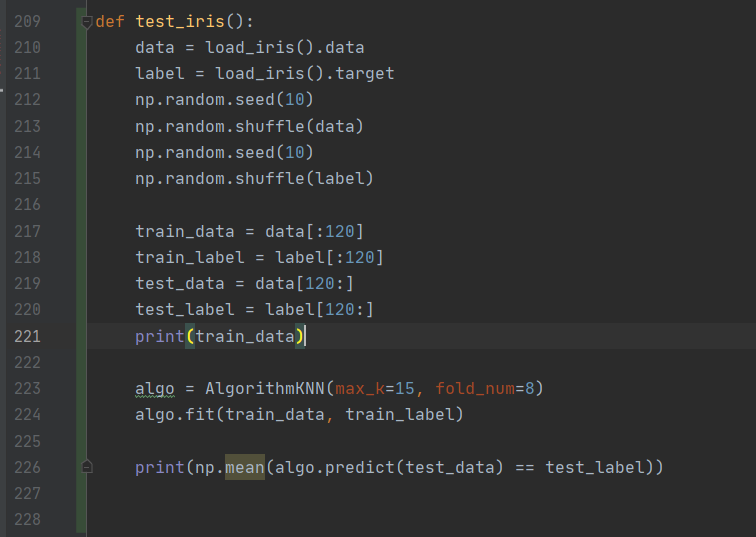
可以看到，KNN的分类准确率比较高，k从1到4变化后，在训练集准确率基本都在0.87~0.88之间，在测试集上表现也很好。

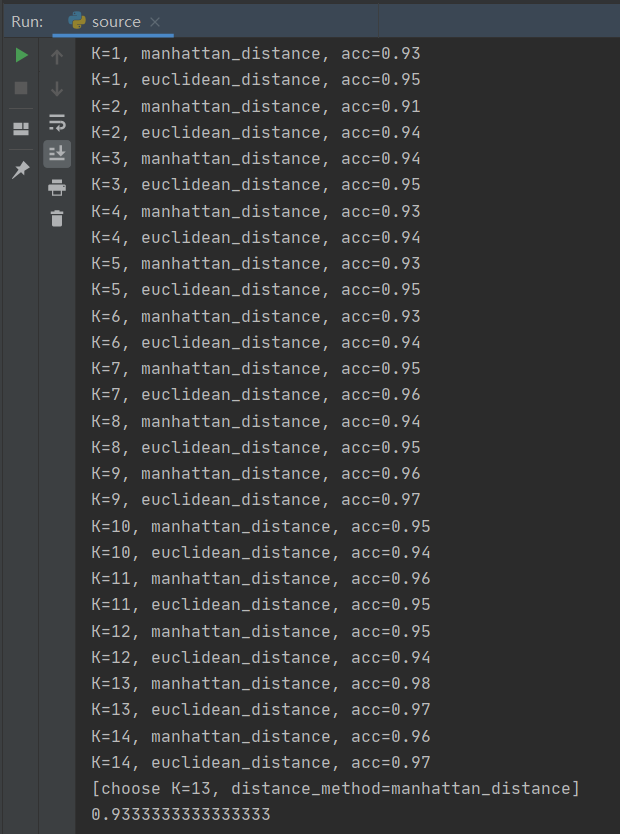
2）第二个实验采用了知名的鸢尾花数据集。

鸢尾花数据集中每条数据包括四个数据，如下图所示：



在本实验中设置训练集120条数据，测试集30条数据，K从1到14变换，采用曼哈顿和欧几里得距离来衡量相似度，实验代码和实验结果如下：





可以看到用KNN来分类，准确度基本在于0.91-0.98之间准确度很高，采取K=13和曼哈顿距离，在测试集上准确度最高，在测试集上预测结果也比较好。