**实验一（4学时—两次实验课）**

**数据集**：semeion\_train.csv, semeion\_test.csv

**实验基本要求**

编程实现kNN算法；给出在不同k值（1，3，5）情况下，kNN算法对手写数字的识别精度。

**实验中级要求**

与 Python 机器学习包中的kNN分类结果进行对比。

**实验提高要求**

画一个以k值为x轴，交叉验证集分类精度为y轴的曲线。

**实验拓展要求**

1. 数据处理：旋转、缩放、加噪音；
2. 算法：优化数据结构，减小计算量，比如KDTree；采用加权投票方法，增加邻近样本的权值，应对样本不平衡问题。

**提交地址：jipeng.huang@mail.nankai.edu.cn**