说明文档

一、实验要求：

实验目的：对验证码识别

使用数据集：train和test

二、预测使用方法：

K近邻方法

三、功能说明：

* **批量修改文件名（Batch modify file name）**：将test文件夹中的文件重新编号，方便顺序读入到程序中
* **预处理训练集（Preprocessing training set）**：预处理训练集，并将图片数据压缩成csv数据保存方便读取

read\_captcha(path)：读取图片，并且二值化，保存标签

注：因为一张图片对应5个字符，所以等距离切割，标签页分散开来

从识别验证码图片，转换至单个小图片的62分类问题（10个数字+26个大写字母+26个小写字母）

* **预处理测试集（Preprocessing test set）**：预处理测试集，原理一样与处理训练集原理一样
* **预测模型（prediction model）**：训练模型，导出预测值

in\_data():：读取保存好的训练集scv文档

in\_test\_data():读取保存好的测试集scv文档

save(answer):将预测好的值保存到out.csv

四、手动参数传入说明：

n\_components=200 #pca降维到指定属性个数

n\_neighbors = 1 #近邻数

五、函数输出结果：





