PYTHON 计算综合实验

一、目的和要求

1. 熟悉 Python 外部库扩展；

2. 熟练使用 Python 程序设计规范；

二、实验环境

1. 不限操作系统；

2. 不限开发环境；

三、实验内容（以下题目任选 1 项，验收后提交报告）

1、课内主题

A．数据分析综合应用：

（1）以文本文件格式读入文件夹\dataanalysis\label\下的 MTL_*.dat, CMTL_*.dat, CEMTL_*.dat(*表示 White 或者 Male，选择其中一种处理即可)中数据，并且分别读入 numpy 数组 MTLLabel、CMTLLabel 或者 CEMTLLabel 中，对各个数组取绝对值后按照降序排序，并且记录数据元素排序前的下标号；

（2）以文本文件格式读入文件夹\dataanalysis\train\下的 MTL_*_train.dat(*表示 White 或者 Male，选择其中一种处理即可)中的数据，并且读入 numpy 矩阵 TrainSample 中，计算矩阵的行列数（该矩阵包含了 1000 个维数为 3304 的样本的观测值，第 1-500 个样本属于第一类，第 501-1000 个样本属于第二类，每类含 500 个样本顺序保存在文件中）。根据（1）中数组的排序（3 个数组分别实验），选择最大的 k 个值（k 取 200，400，600，800，1000，…3304 维）对应的维度，把 TrainSample 中的 1000 个样本降维为 k 维，并保存到新的矩阵中 TrainSub 中；

（3）对于\dataanalysis\test\下文件作和（2）相同的处理（其中数据矩阵包含了 800 个维数为 3304 的样本，第 1-400 个样本属于第一类，第 401-800 个样本属于第二类，每类含 400 个样本顺序保存在文件中）；

（4）阅读和学习\knnexample\下面关于最近邻分类算法 Knn 的实现，用（2）中数据训练分类模型，用（3）中数据测试分类结果，统计错误率。

B．文本分析综合应用：

（1）编写模块实现中文文本中给定字或词的频率统计功能；

（2）运用（1）中功能模块分析文件“dreamofredmaison.txt”中前 80 回和后 40 回中常见文言虚实词的词频，分析结果存入文本文件；

（3）采用 Matplotlib 可视化（2）中的分析结果；

（4）运用 GUI 编制用户界面，为用户提供选择文言虚实字词的交互界面，按照用户选择采用（1）中功能实现频率统计，并且把（3）中实现的分析结果动态呈现给用户。

2、自选主题

要求：体现若干 Python 外部库的使用技能，自由选择开发任务，设计和完成算法、应用程序或系统。以下为参考选题。

参考选题

（1）Python 的编程技巧综合

要求包含下列 4 项技术的使用：Python 中动态链接库 / 多线程编程操作 / 程序打包的依赖关系 / 面向对象的编程。

（2）Python 网络通信应用设计

要求包含 Excel 数据处理、编制图形化界面设计（包含交互界面读取和更改网络参数设置）、文件输出，文件压缩；

（3）Python 舆情分析

要求：获取新冠疫情评述的相关的推特文本，选择若干主题（如疾病起源、疾病防控、数据异常）等开展分析。

（4）Python 数据分析

要求：获取新冠疫情的相关数据，选择若干主题（如变化趋势分析、预测、数据异常）等开展分析。

（5）Python 调用 Canalyzer 或者 Canape 的接口进行编程。