**本 科 实 验 报 告**

实验名称： 高斯混合聚类实现

学 员： 孙源辰 学 号： 201802001067

培养类型： 无军籍本科学员 年 级： 大三

专 业：人工智能与大数据 所属学院： 计算机学院

指导教员： 刘新旺 职 称： 副教授

实 验 室： 305-205 实验日期： 2021春季学期

国防科学技术大学训练部制

《本科实验报告》填写说明

1．学员完成人才培养方案和课程标准要所要求的每个实验后，均须提交实验报告。

2．实验报告封面必须打印，报告内容可以手写或打印。

3．实验报告内容编排及打印应符合以下要求：

（1）采用A4（21cm×29.7cm）白色复印纸，单面黑字打印。上下左右各侧的页边距均为3cm；缺省文档网格：字号为小4号，中文为宋体，英文和阿拉伯数字为Times New Roman，每页30行，每行36字；页脚距边界为2.5cm，页码置于页脚、居中，采用小5号阿拉伯数字从1开始连续编排，封面不编页码。

（2）报告正文最多可设四级标题，字体均为黑体，第一级标题字号为4号，其余各级标题为小4号；标题序号第一级用“一、”、“二、”……，第二级用“（一）”、“（二）” ……，第三级用“1.”、“2.” ……，第四级用“（1）”、“（2）” ……，分别按序连续编排。

（3）正文插图、表格中的文字字号均为5号。

一、实验目的和要求

通过python实现数据集的高斯混合聚类

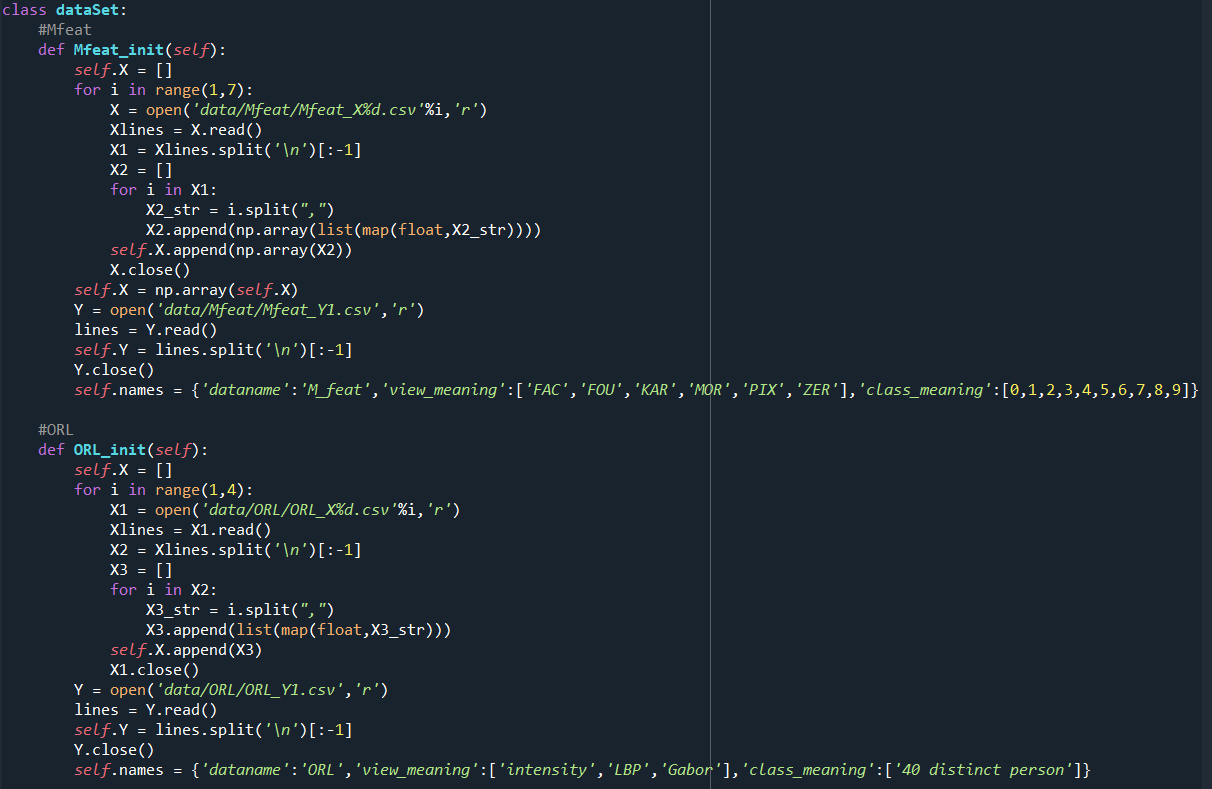
二、实验内容和原理

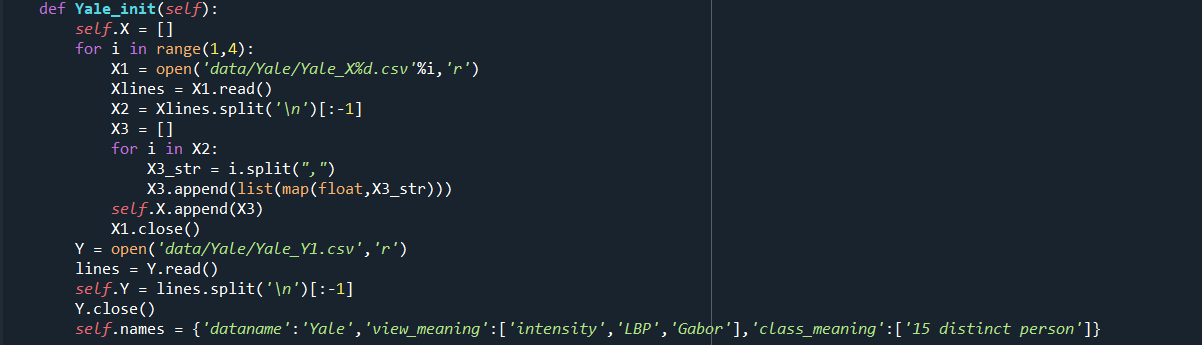
聚类是无监督学习的经典方法，通过对数据集本身的挖掘与分析，获得数据集内数据本身存在的的结构与规律。

三、操作方法与实验步骤

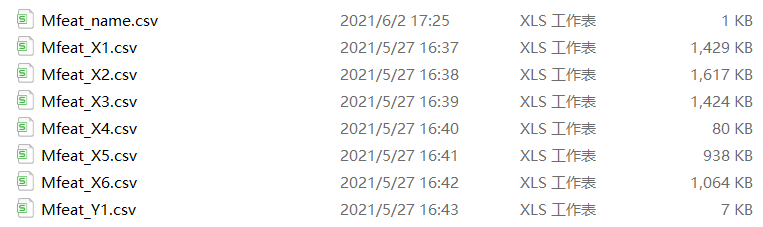
1.编程思路

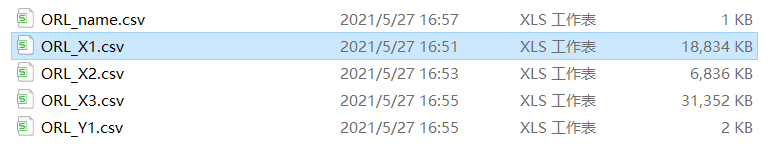
(1)数据集的初始化类

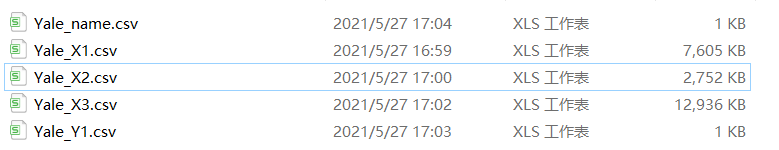




原数据集是mat格式的，在python中用h5py处理了半天见不到数据，于是我把文件放在matlab工作区里面，并且将每个数据集对应的X，Y等信息分别保存为csv文件，如下：

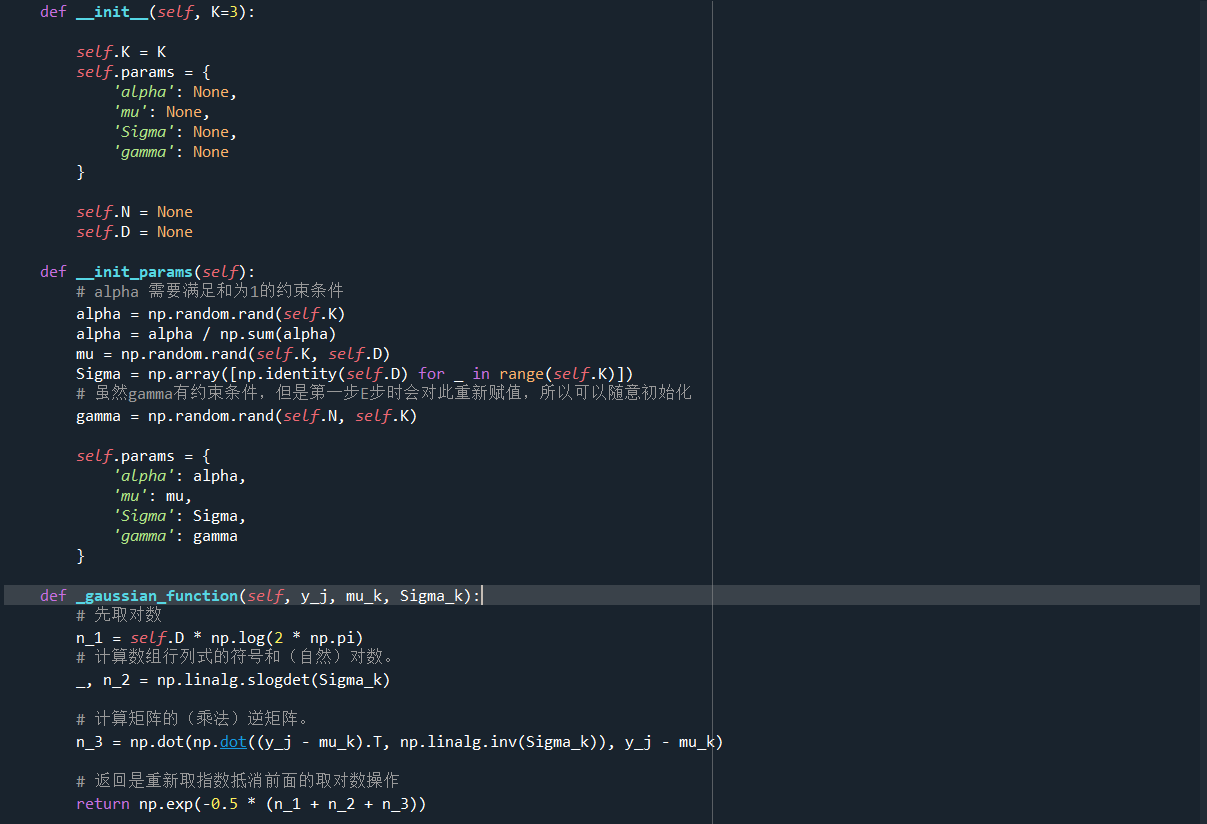






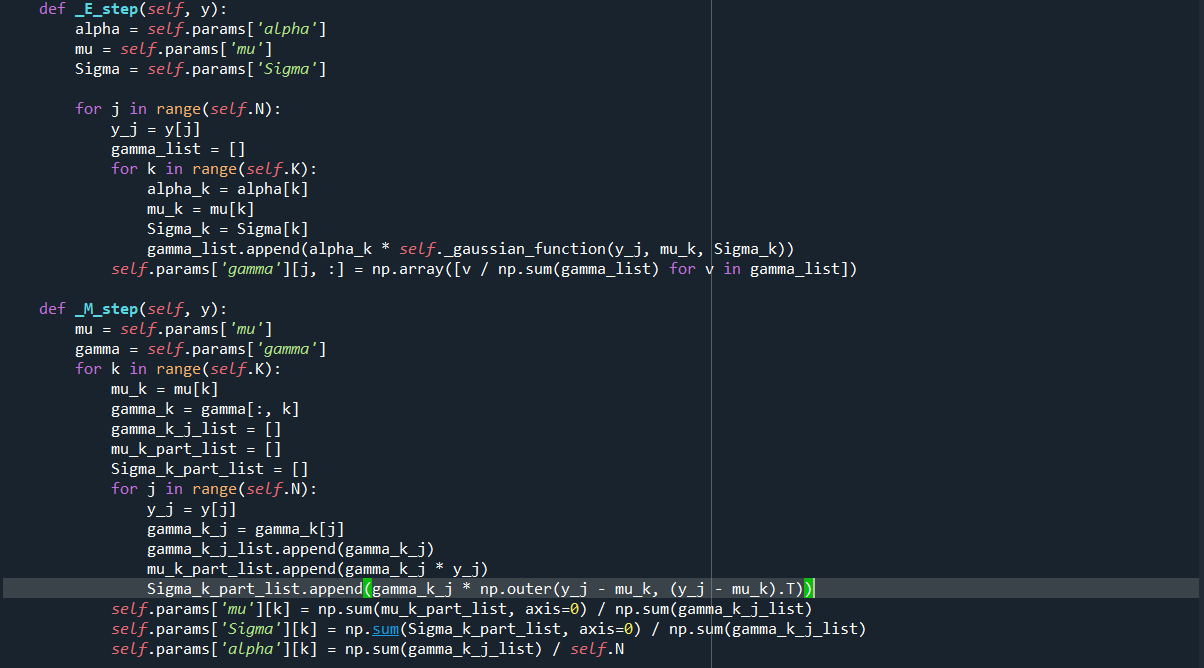
每个csv文件都是二维的。

(2)参数初始化



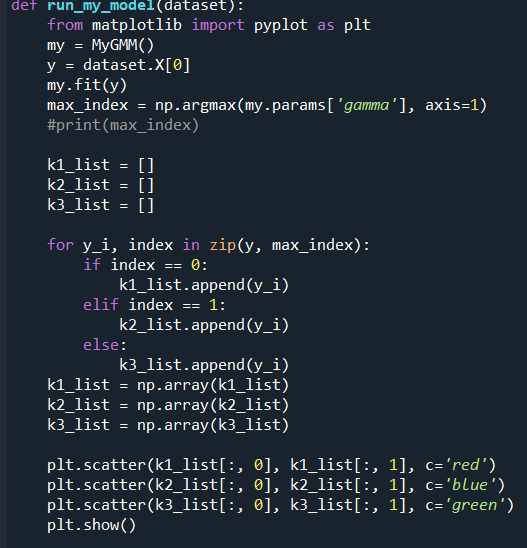
对聚类中心，参数进行初始化，这里面我们设定有三个中心。

(3)EM算法



EM算法分为E步与M步。E步是用来计算后验概率的，而M步是更新参数。

(4)可视化



四、实验结果及分析

