Assignment 1 Report

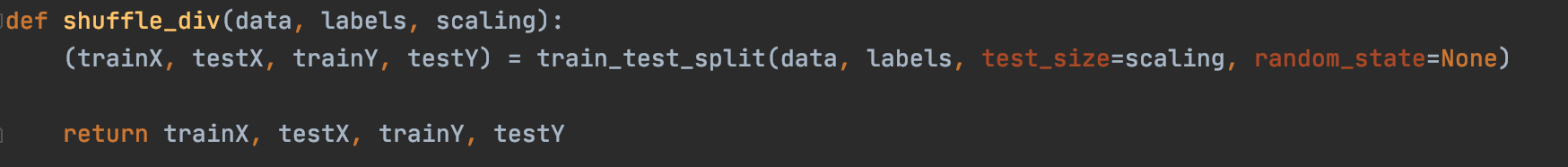
Xie Zhiyao:2009853G-II20-0195

程序设计：

1、根据Aggregation.txt中的数据有七个类别，划分带标签的数据集和不带标签的数据集 图表, 散点图

描述已自动生成

将数据集划分为有标签的数据集和没有标签等待传播的数据集，按1:1划分，由于最开始写代码的时候没有意识到在标签传播的时候不需要划分训练集和测试集合，只需要划分有标签的数据集和无标签的数据集就可以在迭代后检测标签是否被成功传播，所以Aggregation.txt中原有788个数据，按8:2划分为630个训练数据集和158个测试数据集，后续实验中未做修改，只用630个数据集划分为315个带标签的数据集和315个无标签的数据集来完成标签传播的实验。

文本

描述已自动生成

2、根据划分好的带标签数据集和不带标签的数据集，构建YL和YU矩阵，并将YL和YU拼接构建F矩阵

文本

描述已自动生成

标签集构建出来的YL矩阵：

图形用户界面

描述已自动生成

初始化值为0.5的YU矩阵：

电子设备的屏幕

描述已自动生成

3、根据划分出来的数据集（包含带标签的数据和不带标签的数据，共630条）构建概率转移矩阵P：

文本

描述已自动生成

其中根据题目要求保留每个节点的k近邻权重，其他的为0，因此概率转移矩阵P为稀疏矩阵，由于Aggregation.txt数据集中的类别数量为7，因此设置top\_k为7，取前7个最大的权重，即前7个最容易转移的点来进行迭代计算。

电子设备的屏幕

描述已自动生成

在构建好的概率转移矩阵P中，可见每个点最相似的点是自己本身，经过查表确认，在概率转移矩阵中转移到自己的点的概率是7个概率中最大的，因此可以简单判断概念转移矩阵构建正确。

4、进行F=PF迭代训练，使带标签的数据集中的标签传播到不带标签的数据集中，根据一开始划分的不带标签的数据集的真实标签YU\_label来判断一轮迭代中的标签传播结果是否准确，即是否收敛。

文本

描述已自动生成

运行结果的loss如下:

文本

描述已自动生成

迭代到一定次数后，标签不再进行传播，因此loss完全相同，此时可判断loss已经完全收敛：文本

描述已自动生成

将预测的结果中取最大值，转换为0/1的向量矩阵，返回一个预测出来的0/1向量矩阵loss\_matrix

文本

描述已自动生成

预测结果与真实的label对比：图表

低可信度描述已自动生成

观察可得，传播的标签与原标签一致，标签传播成功。