****

**实**

**验**

**报**

**告**

**课 程 大数据分析**

**题 目 项目一**

**分 数**

**学院名称 智能与计算学部**

**专 业 软件工程**

**学生姓名 黄冬梅**

**学 号 3016218170**

**年 级 2016级**

**班 级 5班**

目录

[一、实验目的 1](#_Toc527352380)

[二、实验内容 1](#_Toc527352381)

[三、实验步骤 1](#_Toc527352382)

[四、实验结果 1](#_Toc527352383)

[五、实验结论 1](#_Toc527352384)

[六、源代码 1](#_Toc527352385)

## 实验目的

分析全球不同城市不同年份的各项指标数据。数据包括城市名称、所属国家或地区、年份（从1951年到2008年）、该年份下的收入水平、该年份下的健康指数、该年份下的人口数量及该城市所处的地理位置（经度、维度）

## 实验内容

任务1

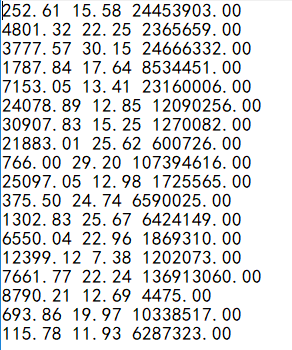
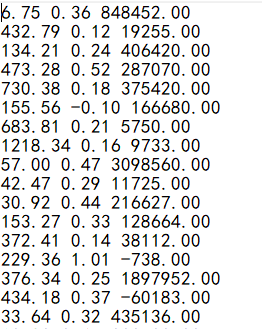
* 1. 使用2种以上聚类算法分析哪些国家具有相似的人口、人均收入、人均寿命变化情况。请考虑可以使用哪些指标，例如可以考虑原始数据、相对某一年的变化值、相对某一年的变化率等（至少使用2种指标）。
  2. 改变聚类算法的参数（每个参数至少取3个值），讨论分析不同的参数和聚类算法对结果的影响。

任务2

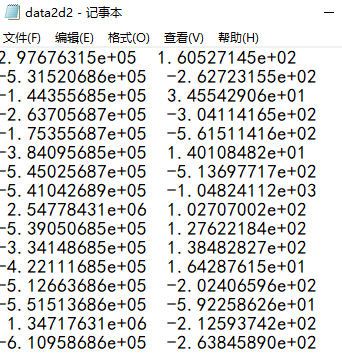
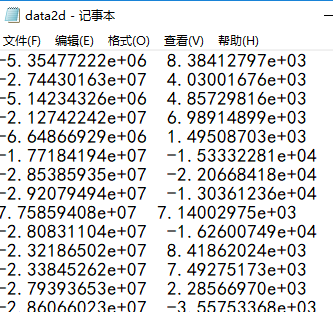
* 1. 使用2种以上降维算法分析哪些国家具有相似的人口、人均收入、人均寿命变化情况。请考虑可以使用哪些指标，例如可以考虑原始数据、相对某一年的变化值、相对某一年的变化率等（至少使用2种指标）。
  2. 改变降维算法的参数（每个参数至少取3个值），讨论分析对比不同参数和降维算法对结果的影响。

## 实验步骤

1. 使用Java将数据处理成需要的格式。将每个国家2008年的人口数、人均收入、人均寿命与原始数据做差，得到增长量，保存在out1.txt；再将2008年和2007年的数据比较得到一年的增长量,保存在out2.txt。

1. 最后得到两份三维的数据，将数据分别用PCA、LDA降维。

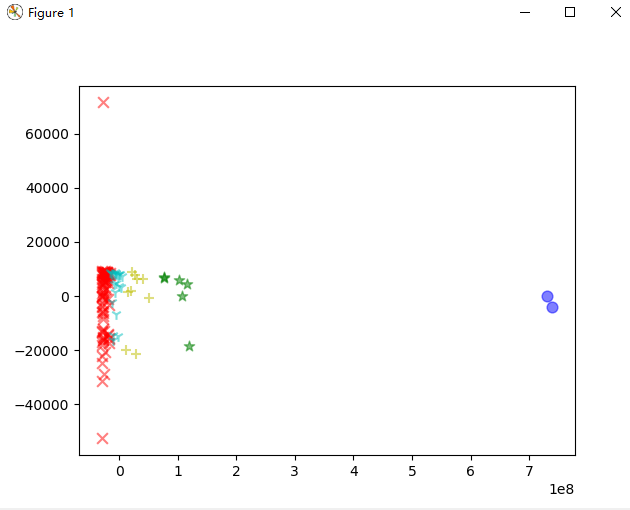


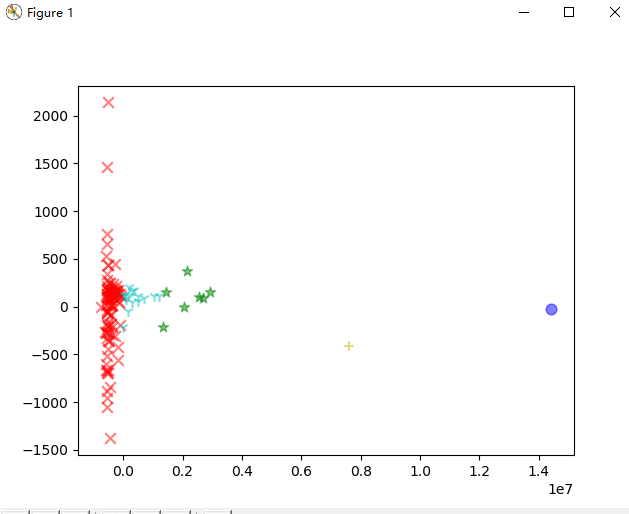
1. 使用将降维后的数据分别做K-means聚类和层次聚类。

## 实验结果

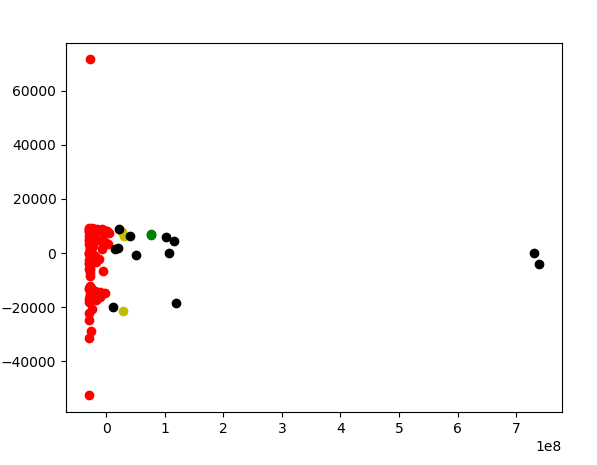
聚类结果如下:

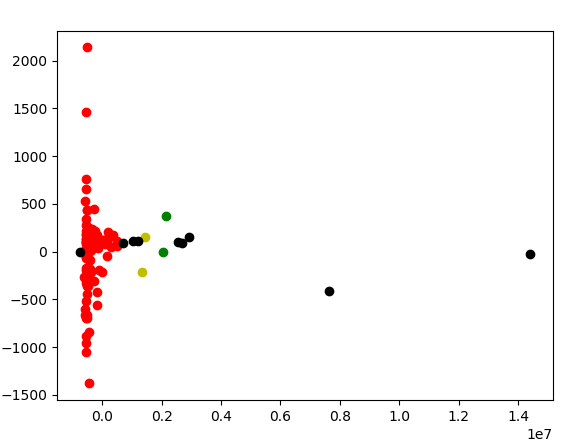
（1）out1.txt 、out2.txt进行K-means聚类





(2) 分别做层次聚类，结果如下：





## 实验结论

大部分国家的发展趋势、发展速度较一致，小部分国家的发展相比较而言较缓慢、或较快，在聚类中处于几个簇之外。

## 源代码

源代码见附件。