****

**《数据挖掘》作业一**

**题目： 主成分分析练习**

班级序号： 111151-19

学生姓名： 白思开

任课教师： 刘袁媛

**中国地质大学信息工程学院软件工程系**

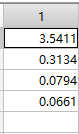
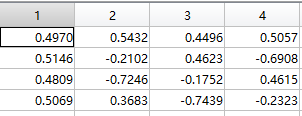
**2018年3月**

1：**数据读入**：初始数据为4\*30的矩阵,四列分别为X1,X2,X3,X4.如：身高拉成了30\*1的细条，及该矩阵的一列。将读入的举着赋给X(初始数据详见data.txt)

2.**数据处理：**包括以下步骤：(1)去除平均值（在这里我参照了网上，每一个数据减去该列数据的均值后，再除以该列数据的标准差，即为标准化过程。如果不除以标准差，最后的投影会有负值。）（2）求协方差矩阵covariance（3）计算协方差矩阵的特征值和特征向量。（4）将特征值从大到小排列（5）将数据转换到N个向特征向量构建的新空间中（即投影过程）result。（代码详见miningData.m）

**3：结果：通过计算，第一个特征值贡献率一超过90%，故选取第一个主成分进行综合评价**

特征值： 特征向量：

部分投影数据：

