**实验二 数据预处理常用方法与应用**

【实验目的】

1 掌握随机抽样的常用方法。

2 掌握变量规范化的常用方法。

3 掌握常见距离的定义及计算方法。

4 掌握二元变量相似度、连续变量余弦相似度计算方法。

5 掌握Mahalanobis距离的计算方法。

【实验类型】

设计型

【实验学时】

2学时

【实验环境】

Windows 7以上操作系统

Python3.0以上版本

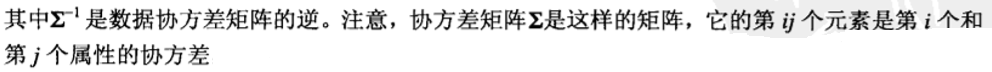
Pycharm开发环境

Spyder开发环境

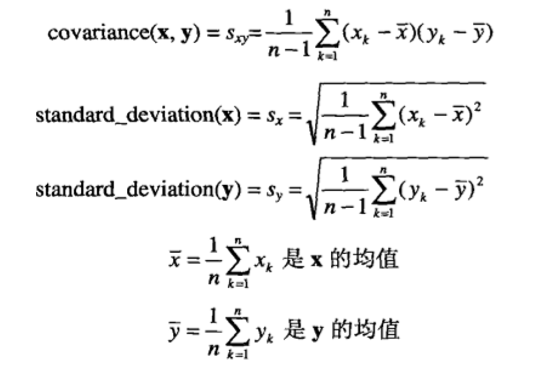
【实验要求】

八、从统计学的观点看，计算多维向量的距离时，应该使用mahalanobis距离，其定义为：



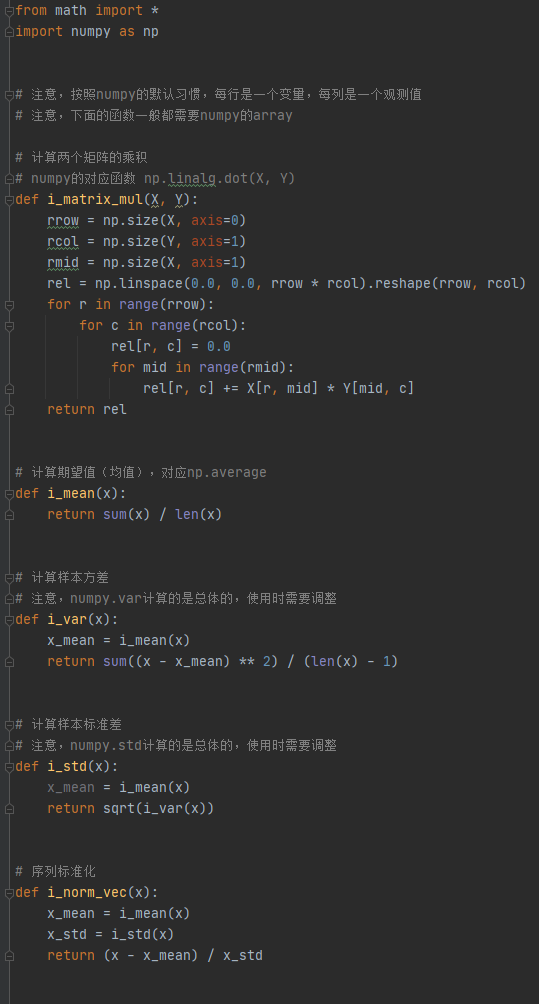


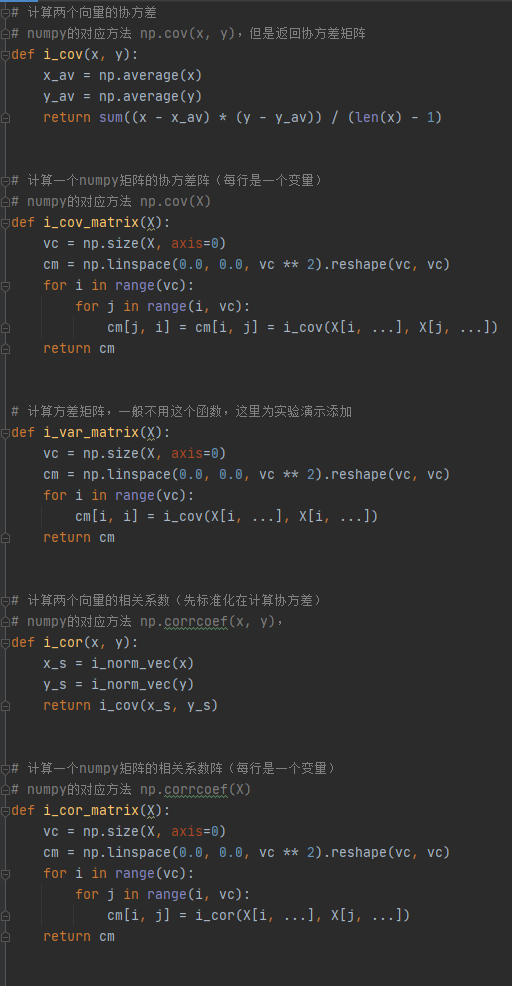
属性间协方差的定义如下所示：

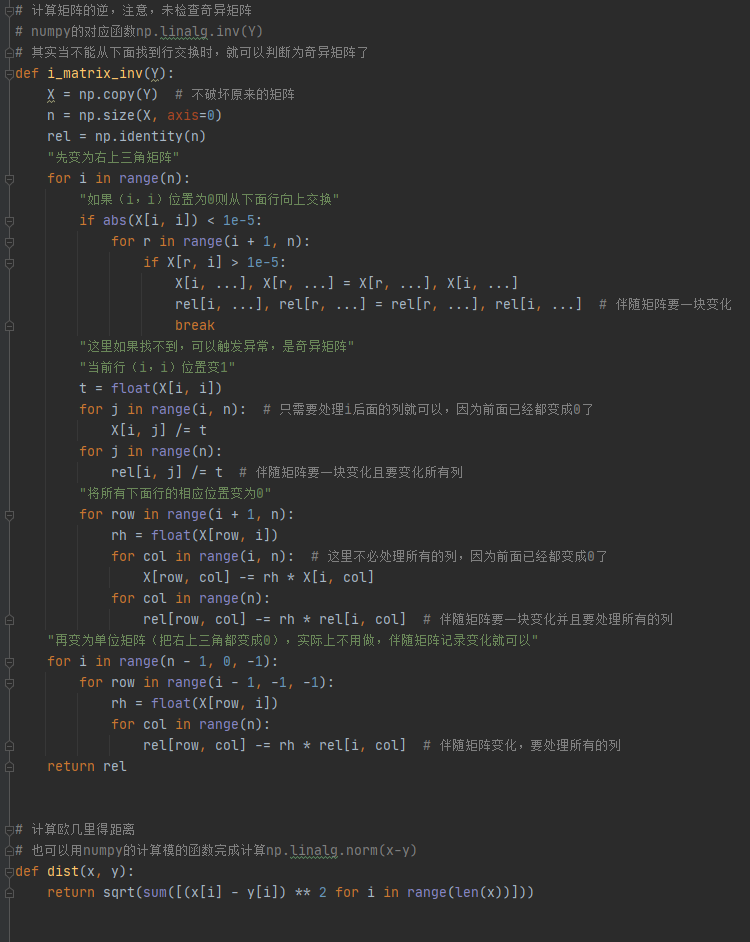


实现欧几里得距离、使用方差逆矩阵调整的距离、使用协方差逆矩阵调整的距离三类距离的计算函数，并构造两个随机向量验证三种距离的不同。

**参考答案：**









**参考输出结果：**

