**实验PCA**

姓名: 李昀哲 学号： 20123101

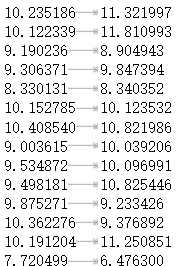
**1 题目和数据：**

题目：PCA

* 通过对给定数据进行主成分分析来了解PCA算法

数据描述:

一组生成的二维数据；



**2 算法：**

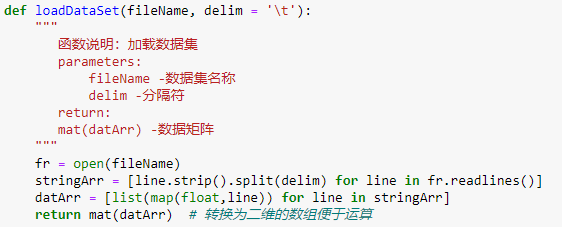
主成分分析是设法将原来众多具有一定相关性（比如P个指标），重新组合成一组新的互相无关的[综合指标](https://baike.baidu.com/item/%E7%BB%BC%E5%90%88%E6%8C%87%E6%A0%87?fromModule=lemma_inlink" \t "https://baike.baidu.com/item/%E4%B8%BB%E6%88%90%E5%88%86%E5%88%86%E6%9E%90/_blank)来代替原来的指标。

主成分分析，是考察多个变量间相关性一种多元统计方法，研究如何通过少数几个主成分来揭示多个变量间的内部结构，即从原始变量中导出少数几个主成分，使它们尽可能多地保留原始变量的信息，且彼此间互不相关.通常数学上的处理就是将原来P个指标作[线性组合](https://baike.baidu.com/item/%E7%BA%BF%E6%80%A7%E7%BB%84%E5%90%88?fromModule=lemma_inlink" \t "https://baike.baidu.com/item/%E4%B8%BB%E6%88%90%E5%88%86%E5%88%86%E6%9E%90/_blank)，作为新的综合指标。

主成分分析的原理是设法将原来变量重新组合成一组新的相互无关的几个综合变量，同时根据实际需要从中可以取出几个较少的总和变量尽可能多地反映原来变量的信息的统计方法叫做主成分分析或称主分量分析，也是数学上处理[降维](https://baike.baidu.com/item/%E9%99%8D%E7%BB%B4?fromModule=lemma_inlink" \t "https://baike.baidu.com/item/%E4%B8%BB%E6%88%90%E5%88%86%E5%88%86%E6%9E%90/_blank)的一种方法。主成分分析是设法将原来众多具有一定相关性（比如P个指标），重新组合成一组新的互相无关的综合指标来代替原来的指标。通常数学上的处理就是将原来P个指标作线性组合，作为新的综合指标。最经典的做法就是用F1（选取的第一个线性组合，即第一个综合指标）的方差来表达，即Va（rF1）越大，表示F1包含的信息越多。因此在所有的线性组合中选取的F1应该是方差最大的，故称F1为第一主成分。如果第一主成分不足以代表原来P个指标的信息，再考虑选取F2即选第二个线性组合，为了有效地反映原来信息，F1已有的信息就不需要再出现再F2中，用数学语言表达就是要求Cov（F1,F2）=0，则称F2为第二主成分，依此类推可以构造出第三、第四，……，第P个主成分

**3 代码及结果**

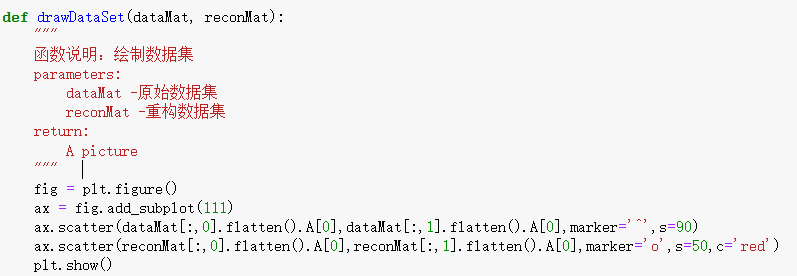
**数据集读入**



**PCA原理实现**



**画图**



**测试结果**

降维前的结果以及降维后的结果

