**Sources：**

1.《R语言实战》(第二版)

2. 《机器学习与R语言实战》丘祐玮

3. 《机器学习》 周志华

4. <https://www.jianshu.com/p/75e805ff247c>

5. <https://chenrudan.github.io/blog/2016/04/01/dimensionalityreduction.html>

6. <http://geopku.com/%E9%99%8D%E7%BB%B4%E4%B8%8E%E5%BA%A6%E9%87%8F%E5%AD%A6%E4%B9%A0/>

7. 球体积随维度增长图[https://book.51cto.com/art/201710/555075.htm](https://book.51cto.com/art/201710/555075.htm" \t "_blank)

8. <https://blog.csdn.net/pxhdky/article/details/85248575>

**前言：**

主要是从理解降维和用R实现降维这两个层面上来阐述，具体的算法还需要感兴趣的小伙伴另外了解。

**目录：**

1. 降维的简介

2. 降维的分类

3. 降维的方法

4. 降维的具体方法和实现

**线性：**

（1）PCA主成分分析

（2）MDS多维尺度分析

（3）SVD奇异值分解法

**非线性：**

（1）ISOMAP

（2）LLE局部线性嵌入法

**1. 降维的简介**

在现实生活中，大多数数据集都有高维度的变量或特征，而算法在计算这些高冗余特征时，需要大量的时间，为了降低计算成本，需要对特征进行无损归约，数学上称为降维。

**维度灾难**

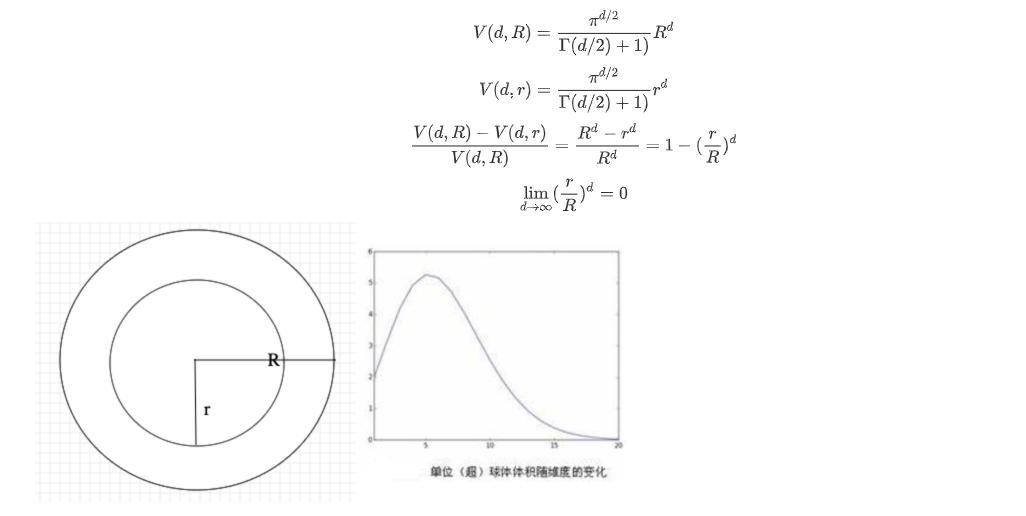
高维情形下，数据样本稀疏，距离计算困难（许多学习方法都涉及距离计算）等问题，会导致数据的过拟合。降维是其中一种缓解维度灾难的途径，当然数据量足够多的话过拟合也可以得到缓解。

Q：为什么能够进行降维？

A：因为虽然数据是高维的，但是和学习任务相关的也许仅仅是低维分布。

Q：如何理解随着维度增高，数据样本稀疏？

A：

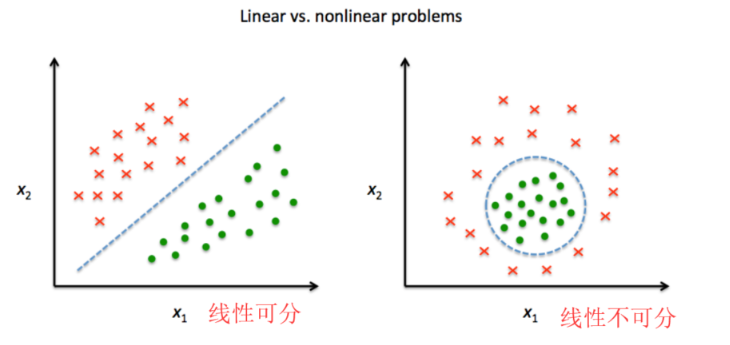
****

**2. 降维的分类**

特征提取和特征筛选。

特征提取是指将高维数据投影到低维数据空间中，可以分成线性提取和非线性提取两种方法：线性抽取试图找到一个仿射空间能够最好地说明数据分布的变化，非线性方法对高维非线性曲线平面分布的数据很有效；特征筛选是指用特征子集来替代原始特征集，这样实现仅利用一组相关特征构建数据模型的目的。

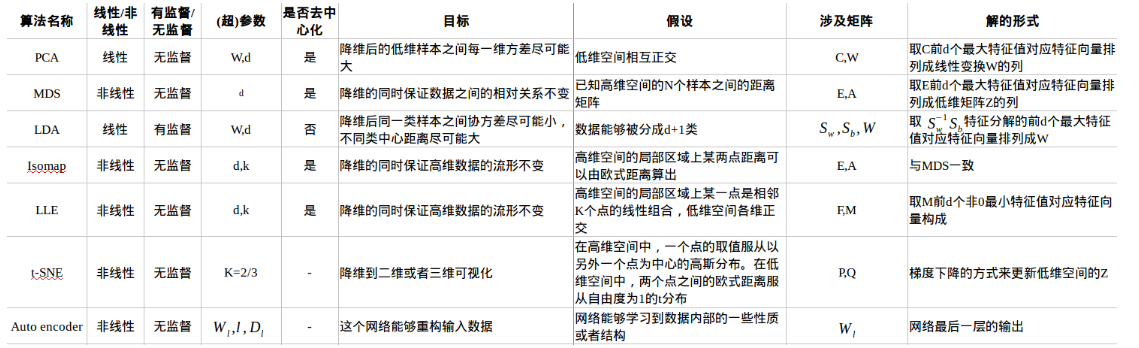
线性和非线性降维：



**3. 降维的主要方法**

投影（projection）和流形学习（Manifold Learning），非线性降维一类是基于核的，如 KPCA（Kernel PCA），另一类是流形学习的。

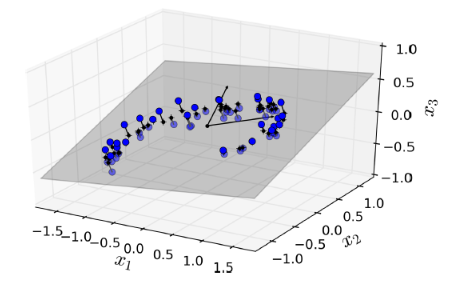
下图是七种降维算法的列表：



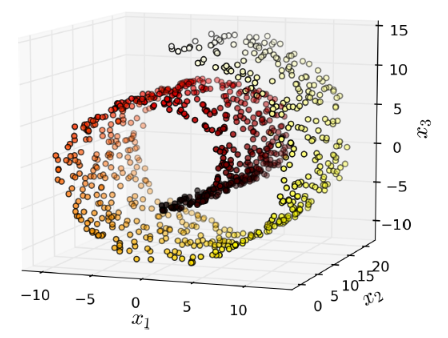
**3.1 投影**

训练样例不是均匀的分散在所有的维度，部分特征是固定的，部分特征是相关的。因此所有的训练样例实际上可以投影在高维空间中的低维子空间中。

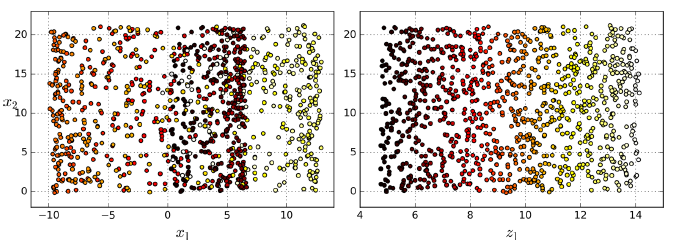
如下图中：3维空间中的训练样例都分布在同一个2维平面中，因此所有的样例都能够投影在2维平面上。



投影并不总是降维最好的方法，在很多情况下空间可以扭转，如下图：瑞士卷（Swiss roll）数据。



直接通过压平第3维，会成为下图左，不同类别样例混在了一起，而实际想要的降维结果是下图右。



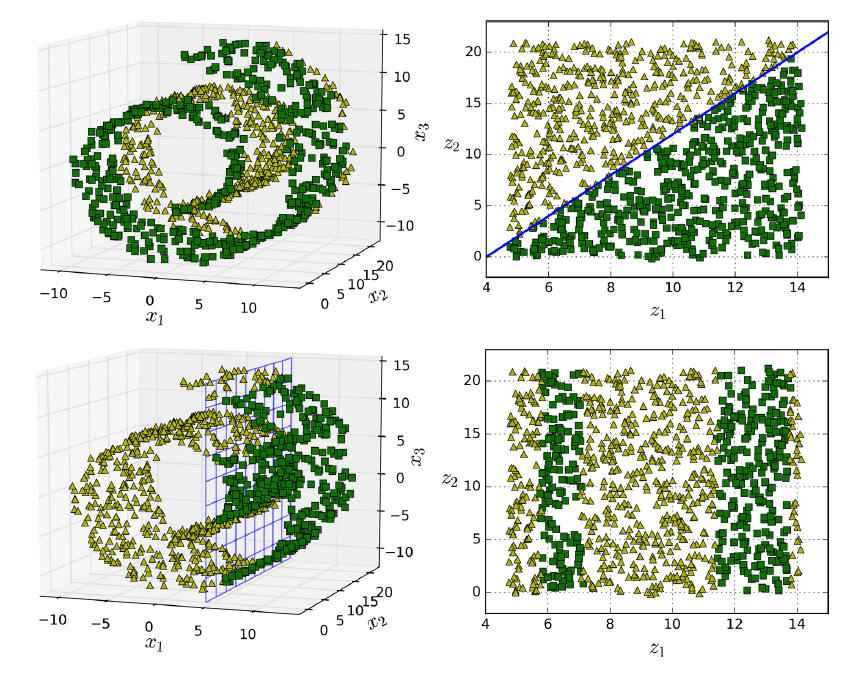
**3.2 流形学习**

流形学习自2000年在《Science》上提出，主要思想是从高维采样数据中恢复低维流形结构，即找到高维空间中的低维流形，并求出相应的嵌入映射，以实现维数约简或者数据可视化。它是从观测到的现象中去寻找事物的本质，找到产生数据的内在规律。

瑞士卷（Swiss roll）是二维流形的例子。它可以在高维空间中弯曲。更一般地，一个d维流形在n维空间弯曲（其中d＜n）。在瑞士卷的情况下，D＝2和n＝3。基于流行数据进行建模的降维算法称为流形学习（Manifold Learning）。它假设大多数现实世界的高维数据集接近于一个低维流形。

流行假设通常隐含着另一个假设：通过流形在低维空间中表达，任务（例如分类或回归）应该变得简单。如下图第一行，Swiss roll分为两类，在3D的空间看起来很复杂，但通过流行假设到2D就能变得简单。

但是这个假设并不总是能成立，比如下图第二行，决策线为x=5，2D的的决策线明显比3D的要复杂。因此在训练模型之前先降维能够加快训练速度，但是效果可能会又增有减，这取决于数据的形式。

****

**4. 降维的具体方法及实现**

**4.1. 主成分分析（PCA）**

特点：是最常用的线性降维方法，期望在所投影的维度上数据的方差最大，这样数据点分散开来能保留更多的信息，PCA是丢失原始数据信息最少的一种线性降维方式，但是它不试图去探索数据内在结构。

**4.1.1 含义**

推导所得的变量成为主成分，是观测变量的线性组合。如第一主成分为：

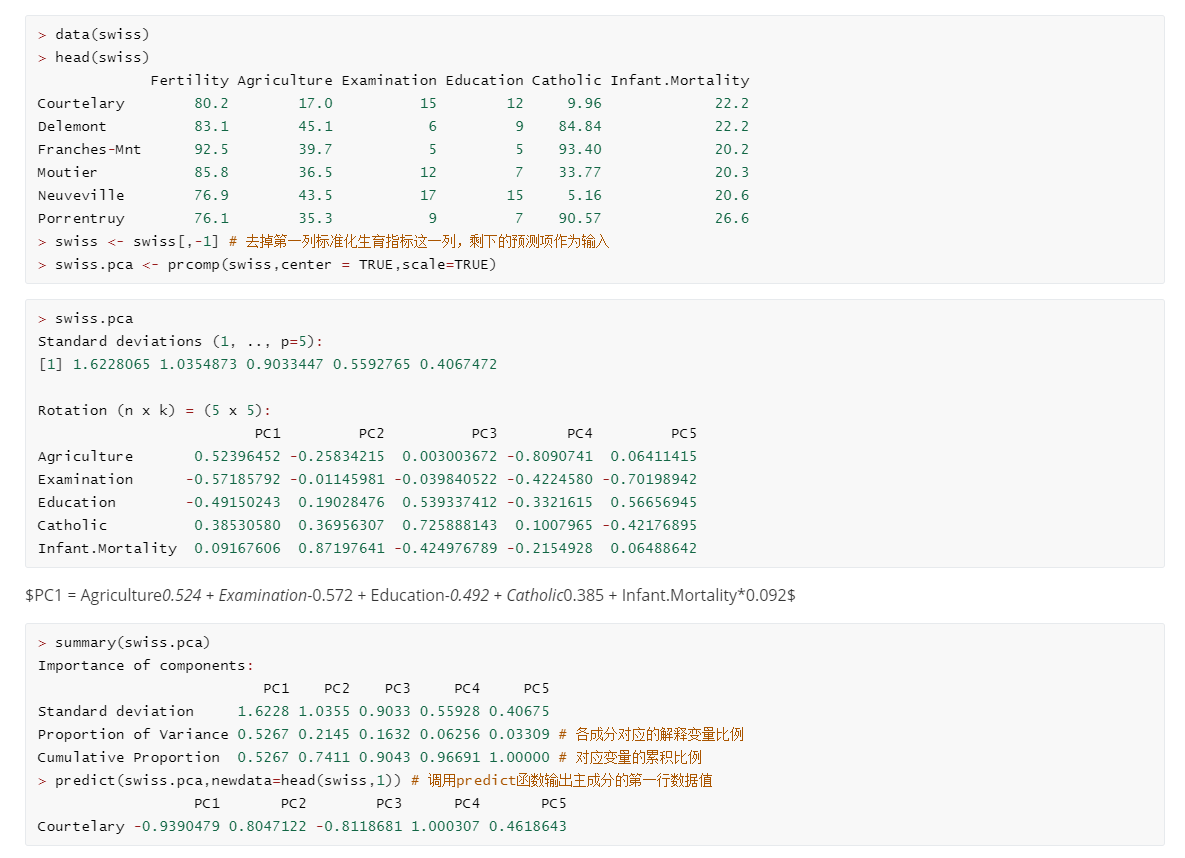


是k个观测变量的加权组合，对初识变量集的方差解释性最大，第二主成分也是初始变量的线性组合，对方差的解释性排第二，所有的主成分都和之前所有的主成分正交，由于解释程度越来越差，因此要用较少的主成分来近似全变量集。

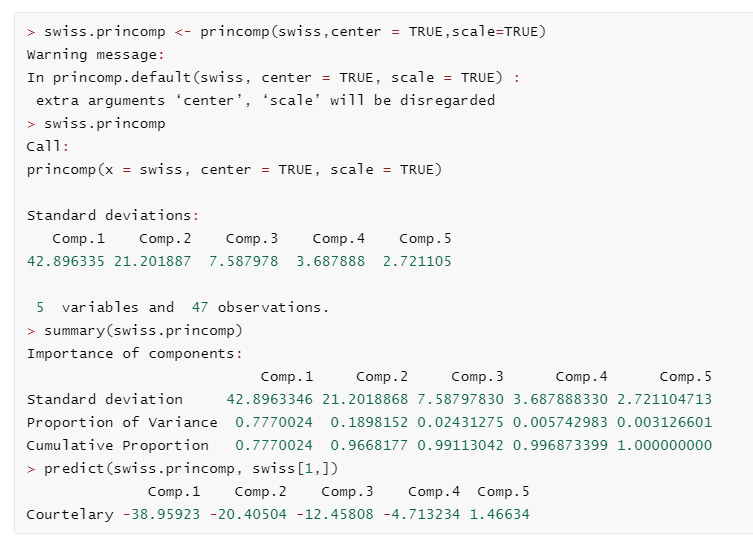
**4.1.2 用R实现**

第一种实现：使用stats包中的prcomp()函数对swiss数据进行PCA处理 【奇异值分解方法】

swiss数据集收集了1888年瑞士47个法语省份的标准化生育指标以及社会经济指数。



第二种实现：使用stats包中的princomp()函数对swiss数据进行PCA处理 【相关矩阵或协方差矩阵的特征值计算方法】



第三种实现：使用psych包中principal() 函数对swiss数据进行PCA处理

R包：library(psych)

函数：principal()

步骤：

1. 数据预处理

PCA是根据观测变量间的相关性来推导结果。输入的可以是原始矩阵或者相关系数矩阵，输入初始数据后相关系数矩阵会被自动计算，**计算前确保数据中没有缺失值**。

1. 选择因子模型

选择PCA（数据降维）还是EFA（潜在结构的发现）。

1. 判断要选择的主成分数目

三种特征值判别准则：

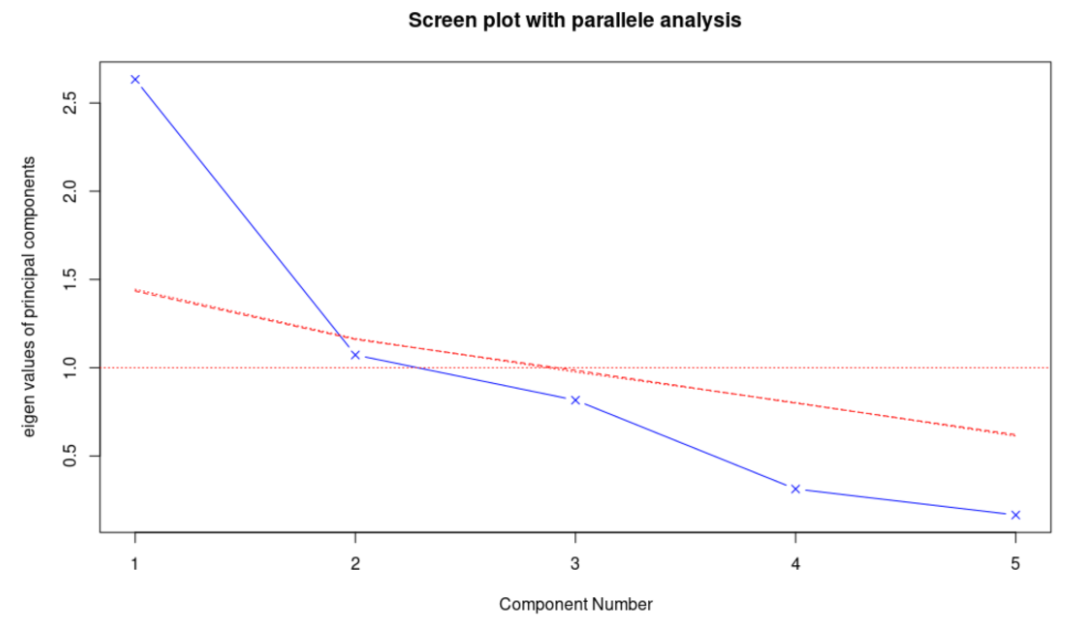
（1）Kaiser-Harris准则：建议保留特征值大于1的主成分

（2）Cattell碎石验则：保留图形变化最大处之上的主成分。

（3）平行分析：基于真实数据的某个特征值大于一组随机数据矩阵相应的平均特征值，该主成分可以保留。



基于观测特征值的碎石检验（线段和x符号组成）、根据100个随机数据矩阵推导出来的特征值均值（虚线）、大于1的特征值准则（y = 1的水平线）。



根据保留方差大于1的主成分选择成分1和成分2，根据特征值准则选择也选择成分1和成分2，根据均值则只选成分1.

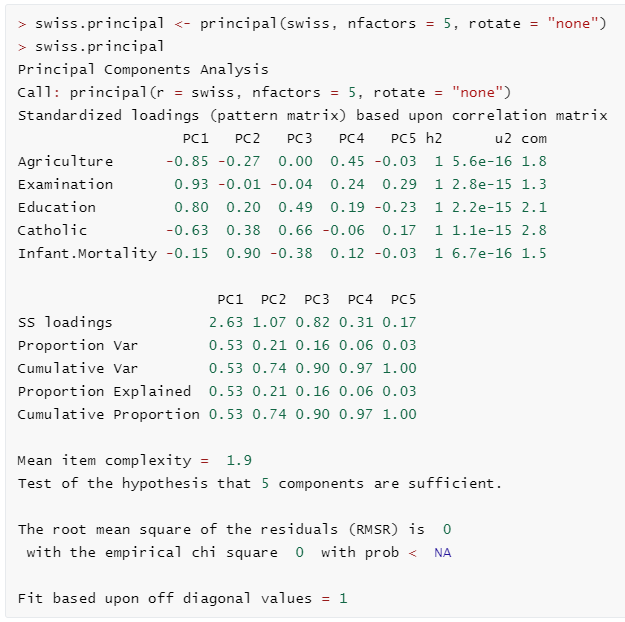
1. 选择主成分



r：相关系数矩阵或原始数据矩阵

nfactors：设定主成分数（默认为1）

rotate：制定旋转方法（默认最大方差旋转（varimax））

解读一下数据结果：

PC1：这一列表示PC1和每个变量的相关性，可以看到都是高度正相关。

h2：成分公因子方差，即主成分对每个变量的方差解释度。

u2：成分唯一性（1-h2），方差无法被主成分解释的比例，PHYS是被PC1解释最差的变量。

proportion var：第一主成分解释了53%的方差。

1. 旋转矩阵

为什么要旋转矩阵？

旋转时为了重新分配各个因子所解释方差的比例，并不改变模型对数据的拟合程度。因子分析需要旋转，当险要解释主成分时，主成分分析分析也可以旋转矩阵。

1. 解释结果

7. 计算主成分或因子得分

**4.1.4 和其他的概念的对比**

主成分分析（PCA）和线性判别分析（LDA）

两者本质都是将原始的样本映射到维度更低的样本空间，但PCA是为了让映射后的样本具有最大的**发散性**，LDA是为了让映射后的样本有最好的**分类性能**。

主成分分析（PCA）和探索性因子分析（EFA）

当多变量数据的变量很多时，了解变量之间的所有交互关系很复杂，假设有20个变量，理解各变量和其他变量的关系时则需要考虑190对两两相互关系。

PCA和EFA都是用来探索和简化多变量复杂关系常用的方法，两者有联系和区别。PCA是数据降维技巧，将大量的 **相关** 变量转化为一组很少的 **不相关** 变量，这些无关变量就是主成分。EFA是用来发现一组变量的潜在结构的方法。这个潜在结构怎么理解呢？举个例子：R包Harman74.cor中有24个心理检测，这些检测的观测得分是根据4个潜在心理学因素（语言能力、反应速度、推理能力和记忆能力）的加权能力值组合成的，这四个因子是观测变量的结构基础或者“原因”。

**4.2 多维尺度分析（MDS）**

**4.2.1 含义**

MDS多维尺度分析法是把高维的数据点映射到一个低维的流形上；同时也是一种可视化方法，利用2D或3D的MDS结果投影后的分布和聚集来研究数据的性质，展示多个对象的之间的相似或相异程度。

MDS分成计量和非计量两类，计量考虑如何保证降维以后的各对象之间的距离尽可能接近它们在原始空间中的距离，先将距离转换成相似度，在相似度上进行非线性投影；非计量则假设两个空间中对象的距离排名已知，且变换后排名不变。

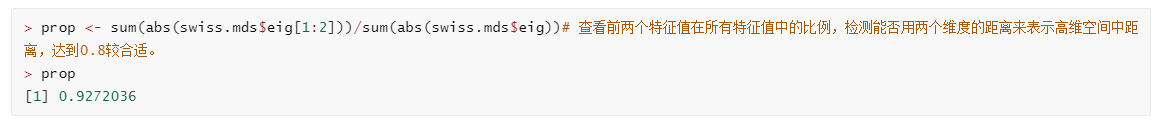
通常情况下选择二维欧氏空间作为目标空间，不过也可以选择其他距离或者更高维的空间。

**4.2.2 用R实现**

使用计量方法完成

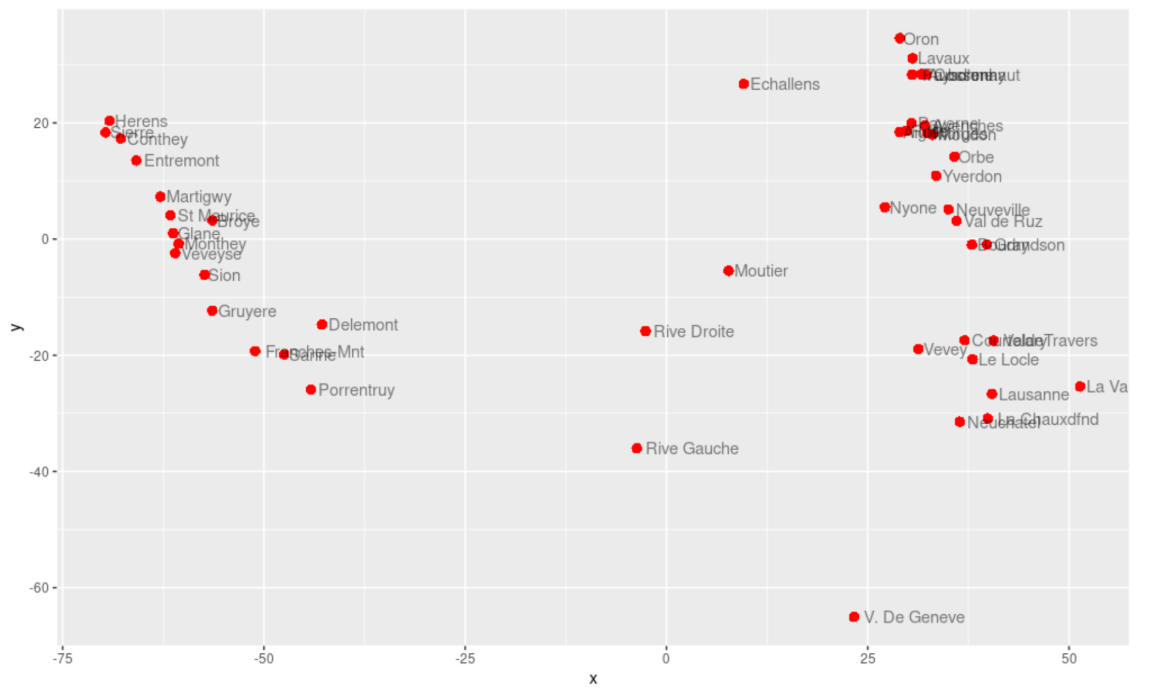
（1）选择二维空间最大值计量MDS分析

（2）查看特征值比例

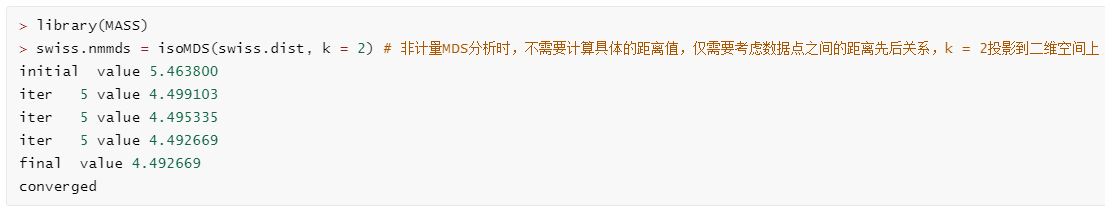
（3）将数据以二维散点图方式实现

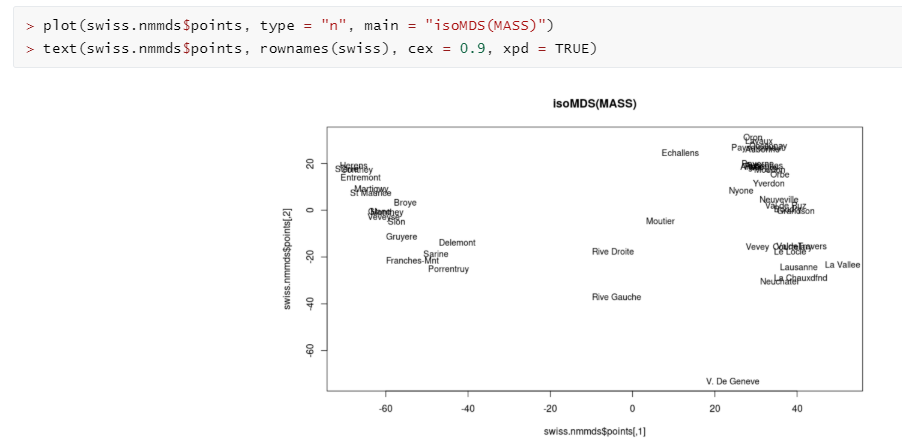


2D可视化结果表示原始高维数据集

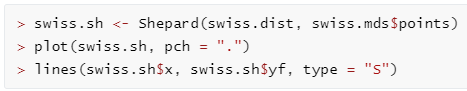


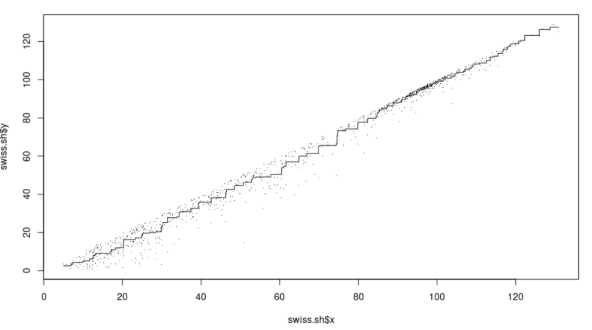
（4）使用isoMDS进行非计量MDS分析

分析结果按照二维散点图展现：



用Shepard曲线展现投影后的距离与原始距离之间的匹配程度，该示例中吻合度很高





**4.3 奇异值分解（SVD）**

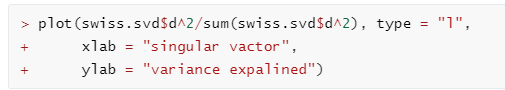
**4.3.1 含义**

是矩阵分解的一种形式，通过奇异值分解，将原始矩阵分解成两个正交矩阵和一个对角矩阵，帮助去除从线性代数角度观察存在线性相关的冗余数据，常被应用在特征筛选、图像处理和聚类等很多领域。

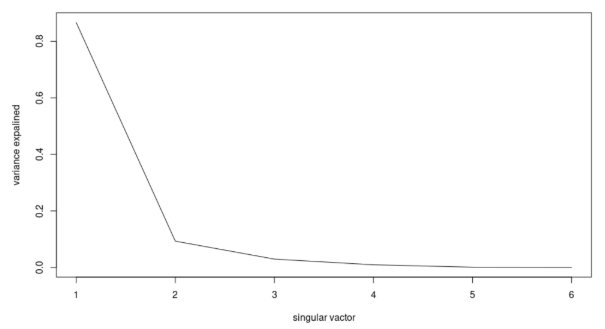
**4.3.2 用R实现**

（1）数据集进行SVD处理，获得分解矩阵

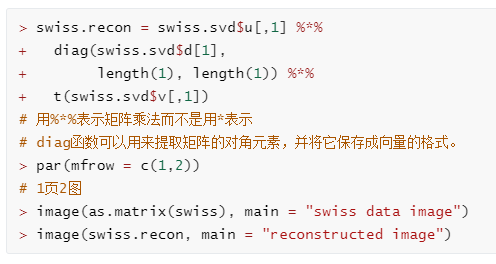
（2）在图中展现SVD列能够解释的变量百分比

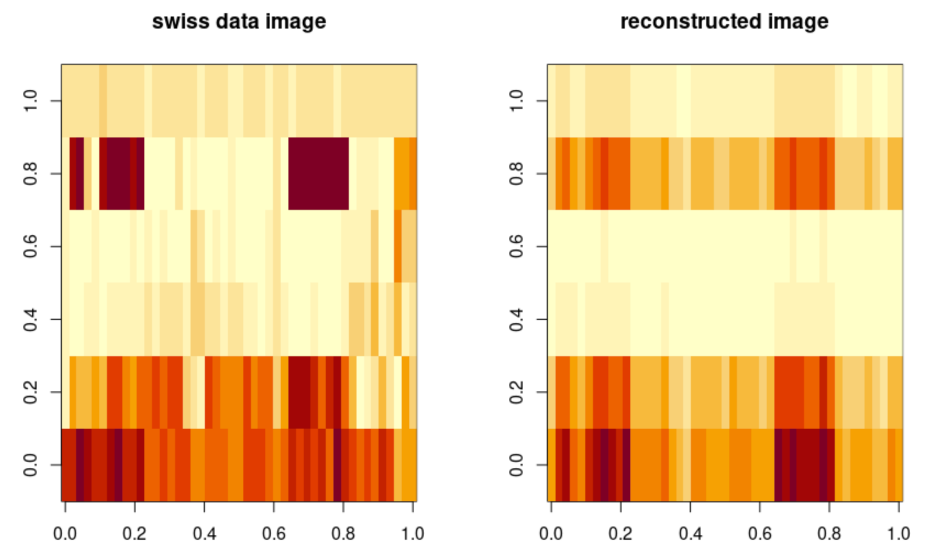


从图中可以看到第一奇异值能够解释80%的变量



（3）比较原始数据集以及奇异值重构后的数据集之间的差别



****

**4.4. ISOMAP进行非线性降维**

**4.4.1 含义**

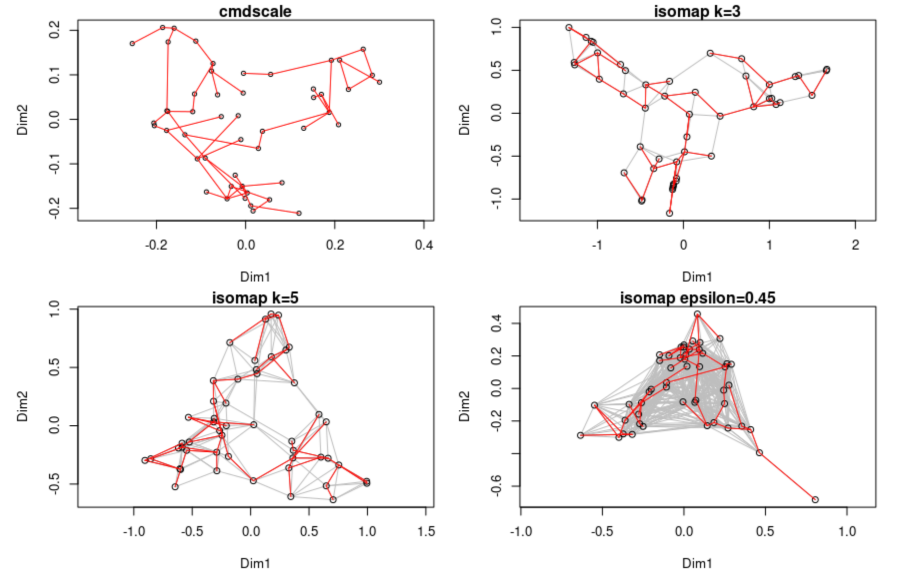
是流形学习方法，支持线性空间到非线性数据结构的转换，由于数据采用非线性结构，在ISOMAP方法中使用几何距离代替了MDS中使用的欧式距离度量，是非线性降维的方式。

可分为四步：

1. 确定每个点的临近点
2. 构建临接图
3. 计算两节点间的最短路径

4. 通过MDS分析找到数据间的低维嵌入

**4.4.2 用R实现**红色是最小生成树



**4.5 局部线性嵌入（LLE）进行非线性降维**

**4.5.1 含义**

假设数据中每个点可以由其近邻的几个点重构出来。降到低维，使样本仍能保持原来的重构关系，且重构系数也一样。

步骤：

第一步是求K近邻，使用KNN等算法通过欧式距离或其它规则找出与自身最邻近的K个点。（在此步骤中，唯一可以控制的参数将是临近点个数K）

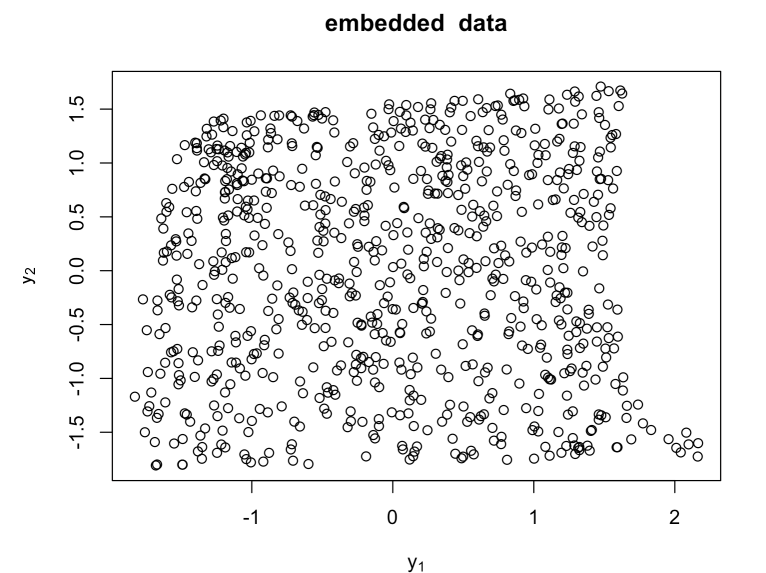
第二步，对每一个样本利用它在邻域的线性关系求权重系数矩阵W。

第三步，利用权重系数在低维空间重构样本数据。

**4.5.2 用R实现**







**其他的降维方法：**

1. 拉普拉斯特征映射（Laplacian Eigenmaps,LE）
2. Sammon映射
3. CCA
4. t-sne（常用于识别肿瘤亚群等等）

......