# SVD(奇异值分解)

## 简介

对于矩阵，r(A)=r，

A的正奇异值为：

奇异值分解：

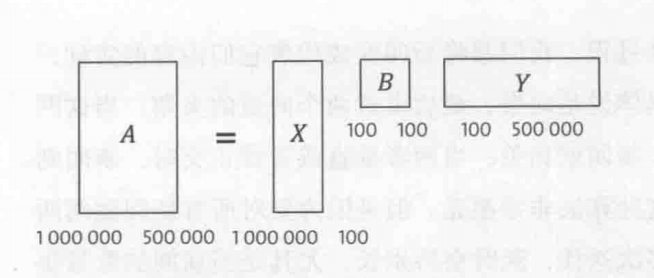
存在酉阵和酉阵，使得

简奇异值分解：

存在半酉阵和半酉阵()，使得

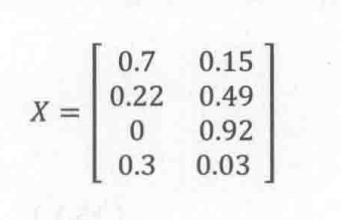
## 一般的SVD分解应用

文本分类

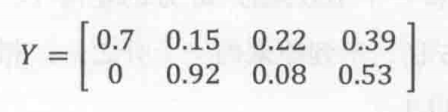


A是文本矩阵，50万个文本，每个文本包含100万个词的频率。

X是近义词分类结果，比如一个4\*2的矩阵X，代表着4个词，2个语义类，代表第i个词与第j个语义类的相关性。



Y是文本分类结果，比如一个2\*4的矩阵Y，代表2个主题，4个文本，代表第j个文本与第i个主题的相关性。



## 局限性

降维维度无法控制

效果不稳定，对于一些数据，降维后的效果特别差

没法扩展，每批数据的特征值都不同

# 随机投影

## 随机投影简介

随机投影是一种数据降维方法，该方法会在原始数据的高维空间中随机选择一组单位向量，这组向量不必相互正交，然后将高维空间的数据映射到以这组单位向量为基的低维空间中。

这种随机选择低维空间基底的方法如何保证降维效果呢？

## Johnson–Lindenstrauss定理

，，，

对

转换一下：

这条定理说明，对于任意的投影方法，当投影维度确定之后，投影前后任意两点之间的距离差都是有上下限的。所以使用随机投影的方式效果即不会太好也不会太差，而且实现也必将简单。

## 随机投影步骤

随机投影流程一般可以分为4步：

1. 确认投影矩阵的维度
2. 使用随机数填充矩阵
3. 归一化矩阵
4. 对数据进行降维

# 实验设计

## 整体实验设计

## 结果

## SVD分析

数值太小，除非两个向量相同，不然任意两个向量之间的距离都是同一个数

31

001 12

01 233