<https://blog.csdn.net/program_developer/article/details/80632779>

降维就是一种对高维度特征数据预处理方法。降维是将高维度的数据保留下最重要的一些特征，去除噪声和不重要的特征，从而实现提升数据处理速度的目的。在实际的生产和应用中，降维在一定的信息损失范围内，可以为我们节省大量的时间和成本。降维也成为应用非常广泛的数据预处理方法。

降维具有如下一些优点：

* 1) 使得数据集更易使用。
* 2) 降低算法的计算开销。
* 3) 去除噪声。
* 4) 使得结果容易理解。

降维的算法有很多，比如[奇异值分解(SVD)](https://link.zhihu.com/?target=https%3A//mp.weixin.qq.com/s/Dv51K8JETakIKe5dPBAPVg)、主成分分析(PCA)、因子分析(FA)、独立成分分析(ICA)。

**PCA原理详解**

PCA(Principal Component Analysis)，即主成分分析方法，是一种使用最广泛的数据降维算法。PCA的主要思想是将n维特征映射到k维上，这k维是全新的正交特征也被称为主成分，是在原有n维特征的基础上重新构造出来的k维特征。

**思考：**我们如何得到这些包含最大差异性的主成分方向呢？

事实上，通过计算数据矩阵的协方差矩阵，然后得到协方差矩阵的特征值特征向量，选择特征值最大(即方差最大)的k个特征所对应的特征向量组成的矩阵。这样就可以将数据矩阵转换到新的空间当中，实现数据特征的降维。

由于得到协方差矩阵的特征值特征向量有两种方法：

特征值分解协方差矩阵、奇异值分解协方差矩阵，所以PCA算法有两种实现方法：

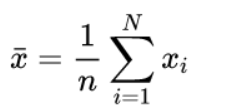
基于特征值分解协方差矩阵实现PCA算法、基于SVD分解协方差矩阵实现PCA算法

## 协方差矩阵的应用与理解

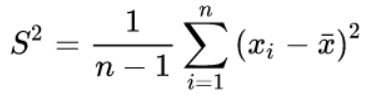
协方差矩阵的计算 https://blog.csdn.net/sh199210/article/details/51778028/

1 单个向量的性质特性

EX 向量的期望：主要是表达这个向量的平均值



DX 向量的方差：主要表达的是这个向量的的每一个变量与其均值的偏离程度



[8,9,10],[2,9,16] 两个向量的期望一样，方差却偏差很大。

2 两个向量之间的关系 二维数据

cov(x,y) 协方差： 表示Xy之间相互关系的数字特征 cov(x,y) = (x-Ex)(y-Ey)

当Cov(X, Y) > 0时，X与Y正相关；

当Cov(X, Y) < 0时，X与Y负相关；

当Cov(X, Y) = 0时，X与Y不相关，独立；

p(x,y) 除以两变量标准差之积以标准化，即相关系数 皮尔逊相关系数

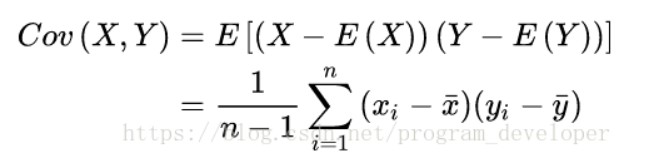
3 多个向量之间的特性

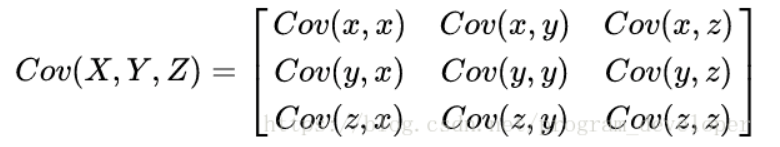
协方差矩阵： 就是多个变量两两间协方差值，按顺序排成的矩阵

比如二维矩阵的协方差矩阵的计算过程就是

[ cov(x,x),cov(x,y) ]

[ cov(y,x),cov(y,y) ] 所以说协方差矩阵是一个对称矩阵，主对角线就是c(x,x) = (x-Ex)(x-Ex) = DX





4 协方差矩阵与原矩阵的关系

# 对于机器学习领域的PCA来说，如果遇到的矩阵不是方阵，需要计算他的协方差矩阵来进行下一步计算，

# 因为协方差矩阵一定是方阵，而特征值分解针对的必须是方阵，svd针对的可以是非方阵情况。

# 协方差矩阵其实保留了原矩阵的主要数据特性，每一个向量本身的性质DX，以及各个向量之间的性质cov（x, y）

# 也就是说，每一个向量的数据之间的离散程度我们保留了，而且各个向量之间的相关性我们也保留下来了，接下来可以去进行特征值分解。

## 特征值分解矩阵原理

(1) 特征值与特征向量

如果一个向量v是矩阵A的特征向量，将一定可以表示成下面的形式：



其中，λ是特征向量v对应的特征值，一个矩阵的一组特征向量是一组正交向量。

(2) 特征值分解矩阵

对于矩阵A，有一组特征向量v，将这组向量进行正交化单位化，就能得到一组正交单位向量。**特征值分解**，就是将矩阵A分解为如下式：



其中，Q是矩阵A的特征向量组成的矩阵，则是一个对角阵，对角线上的元素就是特征值。

## PCA算法两种实现方法

1. 去平均值(即去中心化)，即每一位特征减去各自的平均值。
2. 计算协方差矩阵
3. 用特征值分解方法求协方差矩阵的特征值与特征向量。
4. 对特征值从大到小排序，选择其中最大的k个。然后将其对应的k个特征向量分别作为行向量组成特征向量矩阵P
5. 将数据转换到k个特征向量构建的新空间中，即Y=PX