**主成分分析（PCA）**

https://mp.weixin.qq.com/s?\_\_biz=MzA5ODUxOTA5Mg==&mid=2652551576&idx=1&sn=17a125bb29001b3d8d5e3964dcc599a3&chksm=8b7e48c3bc09c1d55dbab168011cba2a853af5623a24a499a2ae110a4facb07c2a4bd033da36&mpshare=1&scene=2&srcid=0123V8FT7YEhcDXZco9gT4Vf&from=timeline&key=e4aa053ffd46a2720096fdf7de2840d48e8716d03c5dd165e9247c7dcd05d35f32c5a56481c26829d4d3e5dbef395c4877e96528fe3a518bd34d91906f6403e0d63776163f6d172ca17cfb6ba5ea8ad2&ascene=2&uin=MTgwOTU2NjU0MQ%3D%3D&devicetype=android-24&version=26050430&nettype=WIFI&abtest\_cookie=AQABAAgAAQBChh4AAAA%3D&pass\_ticket=sNCtedDpTEPBC88xUKXM3tzXl%2F606nFOdOXtXdJOeRZjV7St1JAOwsumIcivZwOd&wx\_header=1

https://mp.weixin.qq.com/s?\_\_biz=MzI2MzAxMTA1Ng==&mid=2649499654&idx=2&sn=23f750ea4c67ac366e60067d8eb448ec&chksm=f25ae799c52d6e8fb2caf0c43e0b47cb4b0df55b1c627186baefad97b3a6deee190400d992de&mpshare=1&scene=2&srcid=0927Ddf4Tr7vrTgZCKdUVyF2&from=timeline&isappinstalled=0#wechat\_redirect

1. 特征值

<https://www.zhihu.com/question/21874816>

2、降维

假如某学籍数据有两列M和F，其中M列的取值是如何此学生为男性取值1，为女性取值0；而F列是学生为女性取值1，男性取值0。此时如果我们统计全部学籍数据，会发现对于任何一条记录来说，当M为1时F必定为0，反之当M为0时F必定为1。在这种情况下，我们将M或F去掉实际上没有任何信息的损失，因为只要保留一列就可以完全还原另一列。

1. 内积

设向量B的模为1，则A与B的内积值等于A向B所在直线投影的矢量长度。

123

要准确描述向量，首先要确定一组基，然后给出在基所在的各个直线上的投影值，就可以了。

1. 基变换的矩阵表示

一般的，如果我们有M个N维向量，想将其变换为由R个N维向量表示的新空间中，那么首先将R个基按行组成矩阵A，然后将向量按列组成矩阵B，那么两矩阵的乘积AB就是变换结果，其中AB的第m列为A中第m列变换后的结果。

两个矩阵相乘的意义是将右边矩阵中的每一列列向量变换到左边矩阵中每一行行向量为基所表示的空间中去。

1. 协方差矩阵及其优化目标

对于上面二维降成一维的问题来说，找到那个使得方差最大的方向就可以了。不过对于更高维，还有一个问题需要解决。考虑三维降到二维问题。与之前相同，首先我们希望找到一个方向使得投影后方差最大，这样就完成了第一个方向的选择，继而我们选择第二个投影方向。

如果我们还是单纯只选择方差最大的方向，很明显，这个方向与第一个方向应该是“几乎重合在一起”，显然这样的维度是没有用的，因此，应该有其他约束条件。从直观上说，让两个字段（指的是由3维变换到2维的两个维度）尽可能表示更多的原始信息，我们是不希望它们之间存在（线性）相关性的，因为相关性意味着两个字段不是完全独立，必然存在重复表示的信息。

降维问题的优化目标：将一组N维向量降为K维（K大于0，小于N），其目标是选择K个单位（模为1）正交基，使得原始数据变换到这组基上后，各字段两两间协方差为0，而字段的方差则尽可能大（在正交的约束下，取最大的K个方差）。

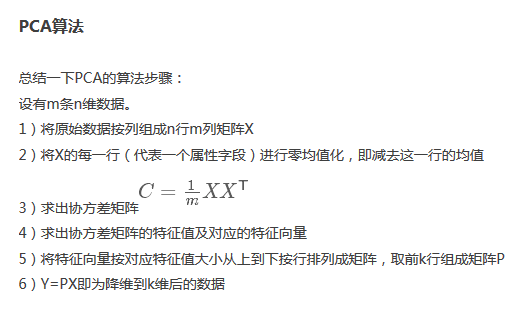
设我们有m个n维数据记录，将其按列排成n乘m的矩阵X，设

IMG_256

则C是一个对称矩阵，其对角线分别个各个字段的方差，而第i行j列和j行i列元素相同，表示i和j两个字段的协方差。

1. 协方差矩阵对角化

寻找一个矩阵P，IMG_256满足 是一个对角矩阵，并且对角元素按从大到小依次排列，那么P的前K行就是要寻找的基，用P的前K行组成的矩阵乘以X就使得X从N维降到了K维并满足上述优化条件。



Scikit-Learn：PCA

<https://www.cnblogs.com/pinard/p/6243025.html>

