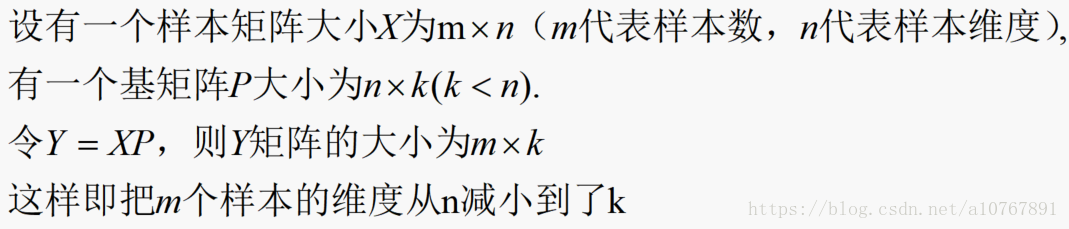
上文中提到过，PCA变换其实就是一种降维技术。

什么是降维？降维就是指通过矩阵乘法运算后，把原来的矩阵维度减少。比如



维数减少了，虽然可以大大减少算法的计算量，但是若对基矩阵P选择不当的话就很有可能会导致信息量的缺失。

因此我们要选择哪K个基(这里还不知道是特征向量)才能保证降维后能最大程度保留原有的信息，是进行设计的主方向。

什么是信息呢？根据信息论的定义，我们可以知道信息来源于未知。也就是说如果不同样本的同一维度的值差异特别大，那该维度带给我们的信息量就是极大的。转换成数学语言，也就是说某维度的方差越大，它的信息量越大。

举个例子：假如你有3个人的人脸特征数据，他们均有3个维度：眼睛、鼻子、嘴巴。如果他们鼻子这一维度的数据都是一样的，如下图中，三个人都是大鼻子(方差=0)。那么我们从鼻子这一维度获得的信息量就是为零，因为无法从该维度得知该人脸图像到底属于谁的。



那如果他们鼻子这一维度的数据各不相同，如下图中，三个人分别是大、中、小鼻子(方差很大)。那么我们从鼻子这一维度获得的信息量就很大了，甚至直接用这一个维度就可以识别出是谁的人脸图像。



**综上，我们就可以很容易联想到 第一个优化目标：降维后各维度的方差尽可能大。**

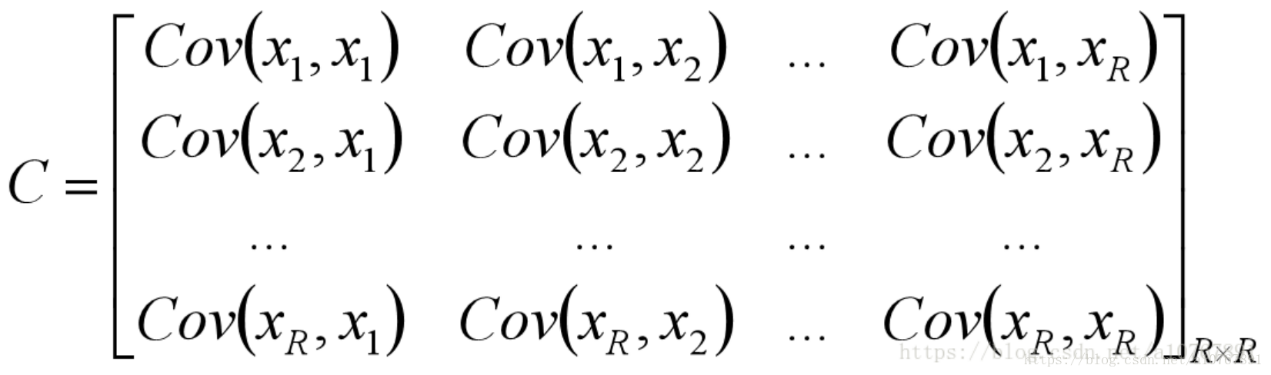
**既然有了第一个，当然就会有第二个，我们的第二个优化目标便是保证不同维度之间的相关性为0(其实就是让基向量互相正交)。**

因为如果某2个维度间存在相关性，就说明从一个维度的值可以推测出另一个维度的值。说明该维度中有一个是多余的，对我们识别特征帮助不大，那自然可以把其中一个维度舍去。

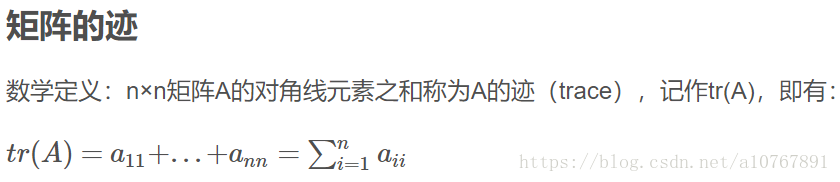
PCA算法的优化目标就是: ① 降维后同一维度的方差最大

② 不同维度之间的相关性为0

根据线性代数，我们可以知道同一元素的协方差就表示该元素的方差，不同元素之间的协方差就表示它们的相关性。因此这两个优化目标可以用协方差矩阵来表示，如下图：



如下图，根据矩阵的迹的定义：

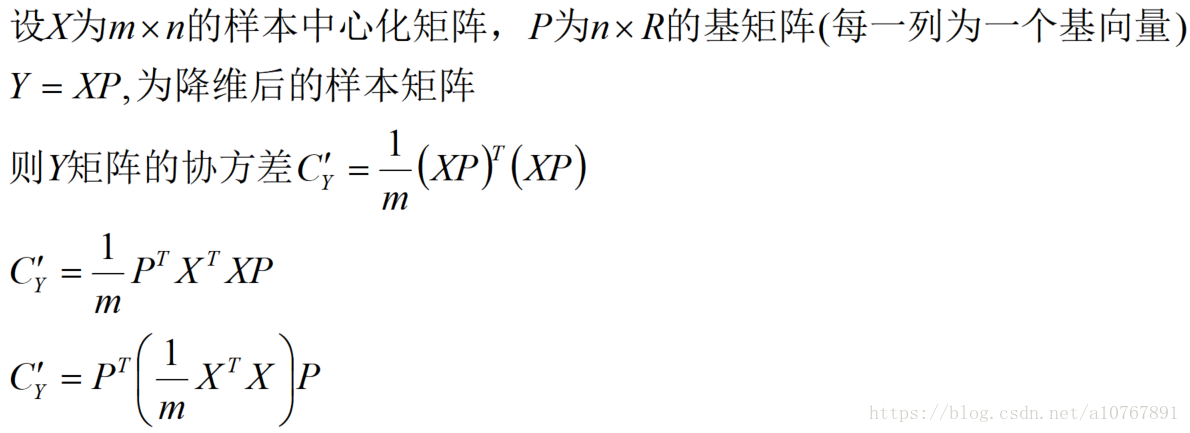


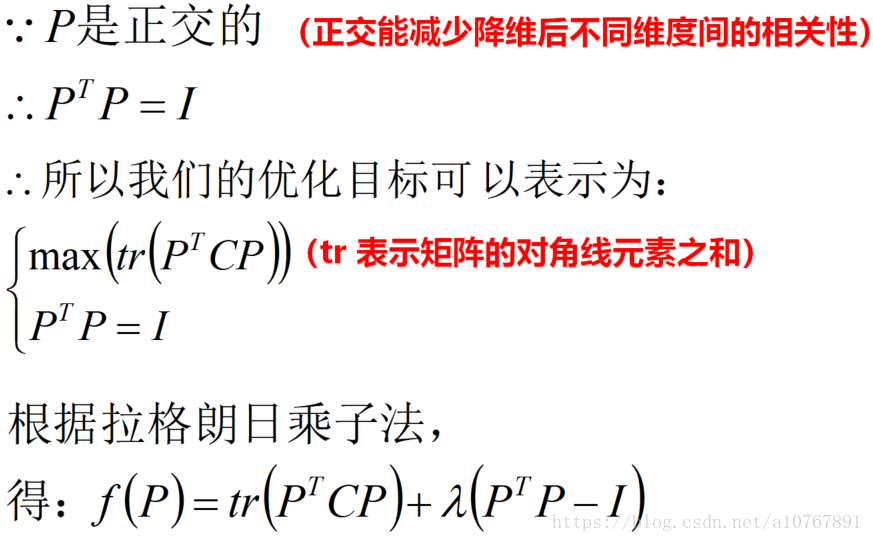
可以知道优化目标一便是令Cy矩阵的对角线元素之和最大。即max tr(Cy)

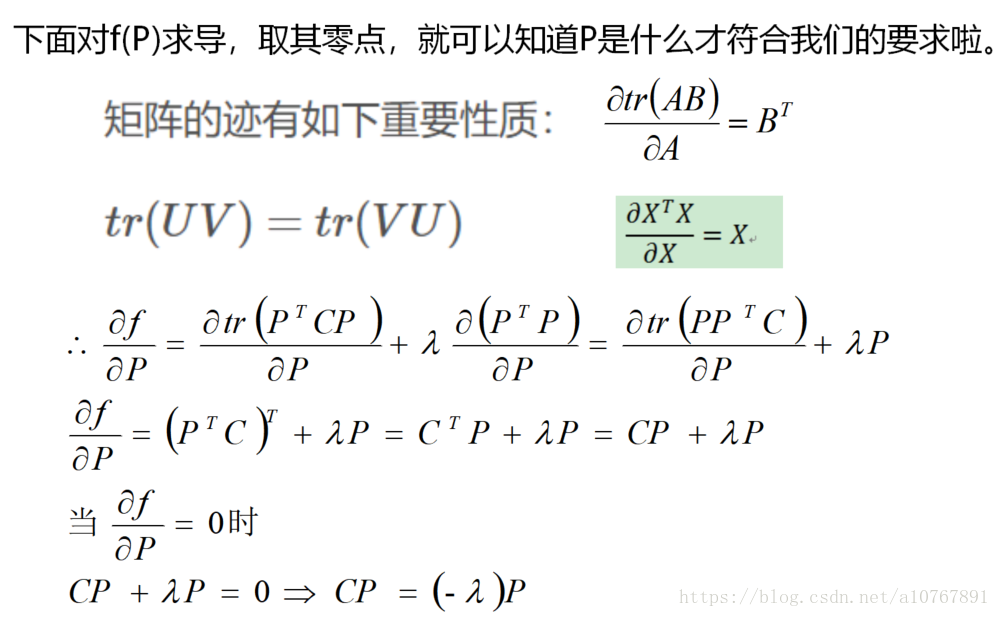
-------------------------------------------------------------------------------------------------------------

知道了目标，接下来就好办，我们要做的就是想尽一切办法去达到我们的优化目标。

下面便是推导过程：







可以看到，最终求得的结果满足特征向量的关系式，因此由C矩阵的特征向量所构成的基矩阵，就是我们要求的变换矩阵。

由该矩阵降维得到的新样本矩阵可以最大程度保留原样本的信息。