主元分析(PCA)理论分析及应用

（主要基于外文教程翻译）

什么是PCA?

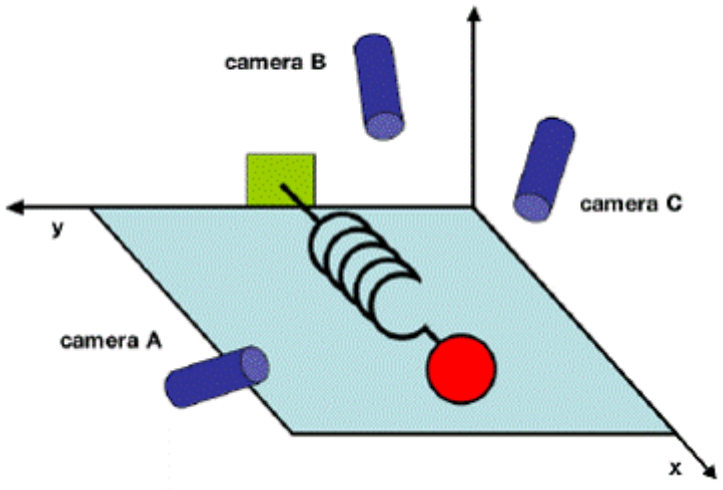
PCA是Principal component analysis的缩写，中文翻译为主元分析。它是一种对数据进行分析的技术，最重要的应用是对原有数据进行简化。正如它的名字：主元分析，这种方法可以有效的找出数据中最“主要”的元素和结构，去除噪音和冗余，将原有的复杂数据降维，揭示隐藏在复杂数据背后的简单结构。它的优点是简单，而且无参数限制，可以方便的应用与各个场合。因此应用极其广泛，从神经科学到计算机图形学都有它的用武之地。被誉为应用线形代数最价值的结果之一。

在以下的章节中，不仅有对PCA的比较直观的解释，同时也配有较为深入的分析。首先将从一个简单的例子开始说明PCA应用的场合以及想法的由来，进行一个比较直观的解释；然后加入数学的严格推导，引入线形代数，进行问题的求解。随后将揭示PCA与SVD(Singular Value Decomposition)之间的联系以及如何将之应用于真实世界。最后将分析PCA理论模型的假设条件以及针对这些条件可能进行的改进。

一个简单的模型

在实验科学中我常遇到的情况是，使用大量的变量代表可能变化的因素，例如光谱、电压、速度等等。但是由于实验环境和观测手段的限制，实验数据往往变得极其的复杂、混乱和冗余的。如何对数据进行分析，取得隐藏在数据背后的变量关系，是一个很困难的问题。在神经科学、气象学、海洋学等等学科实验中，假设的变量个数可能非常之多，但是真正的影响因素以及它们之间的关系可能又是非常之简单的。

下面的模型取自一个物理学中的实验。它看上去比较简单，但足以说明问题。如图表 1所示。这是一个理想弹簧运动规律的测定实验。假设球是连接在一个无质量无摩擦的弹簧之上，从平衡位置沿 x 轴拉开一定的距离然后释放。



图表 1

对于一个具有先验知识的实验者来说，这个实验是非常容易的。球的运动只是在x轴向上发生，只需要记录下 x 轴向上的运动序列并加以分析即可。但是，在真实世界中，对于第一次实验的探索者来说（这也是实验科学中最常遇到的一种情况），是不可能进行这样的假设的。那么，一般来说，必须记录下球的三维位置 (x_0, y_0, z_0) 。这一点可以通过在不同角度放置三个摄像机实现（如图所示），假设以 $200Hz$ 的频率拍摄画面，就可以得到球在空间中的运动序列。但是，由于实验的限制，这三台摄像机的角度可能比较任意，并不是正交的。事实上，在真实世界中也并没有所谓的 (x, y, z) 轴，每个摄像机记录下的都是一幅二维的图像，有其自己的空间坐标系，球的空间位置是由一组二维坐标记录的：$[(x_A, y_A), (x_B, y_B), (x_C, y_C)]$ 。经过实验，系统产生了

几分钟内球的位置序列。怎样从这些数据中得到球是沿着某个x轴运动的规律呢？怎样将实验数据中的冗余变量剔除，化归到这个潜在的x轴上呢？

这是一个真实的实验场景，数据的噪音是必须面对的因素。在这个实验中噪音可能来自空气、摩擦、摄像机的误差以及非理想化的弹簧等等。噪音使数据变得混乱，掩盖了变量间的真实关系。如何去除噪音是实验者每天所要面对的巨大考验。

上面提出的两个问题就是PCA方法的目标。PCA主元分析方法是解决此类问题的一个有力的武器。下文将结合以上的例子提出解决方案，逐步叙述PCA方法的思想和求解过程。

线形代数：基变换

从线形代数的角度来看，PCA的目标就是使用另一组基去重新描述得到的数据空间。而新的基要能尽量揭示原有的数据间的关系。在这个例子中，沿着某x轴上的运动是最重要的。这个维度即最重要的“主元”。PCA的目标就是找到这样的“主元”，最大程度的去除冗余和噪音的干扰。

A. 标准正交基

为了引入推导，需要将上文的数据进行明确的定义。在上面描述的实验过程中，在每一个采样时间点上，每个摄像机记录了一组二维坐标(x_A,y_A)，综合三台摄像机数据，在每一个时间点上得到的位置数据对应于一个六维列向量。

$$\vec{X}=\begin{bmatrix}x_A\\y_A\\x_B\\y_B\\x_C\\y_C\end{bmatrix}$$

如果以$200Hz$的频率拍摄10分钟，将得到$10\times60\times200=120000$个这样的向量数据。

抽象一点来说，每一个采样点数据$\vec{X}$都是在m维向量空间（此例中$m=6$）内的一个向量，这里的m是牵涉的变量个数。由线形代数我们知道，在m维向量空间中的每一个向量都是一组正交基的线形组合。最普通的一组正交基是标准正交基，实验采样的结果通常可以看作是在标准正交基下表示的。举例来说，上例中每个摄像机记录的数据坐标为(x_A,y_A)，这样的基便是$\{(1,0),(0,1)\}$。那为什么不取$\{(\frac{\sqrt{2}}{2},\frac{\sqrt{2}}{2}),(\frac{-\sqrt{2}}{2},\frac{-\sqrt{2}}{2})\}$或是其他任意的基呢？原因是，这样的标准正交基反映了数据的采集方式。假设采集数据点是$(2,2)$，一般并不会记录$(2\sqrt{2},0)$（在$\{(\frac{\sqrt{2}}{2},\frac{\sqrt{2}}{2}),(\frac{-\sqrt{2}}{2},\frac{-\sqrt{2}}{2})\}$基下），因为一般的观测者都是习惯于取摄像机的屏幕坐标，即向上和向右的方向作为观测的基准。也就是说，标准正交基表现了数据观测的一般方式。

在线形代数中，这组基表示为行列向量线形无关的单位矩阵。

$$B=\begin{bmatrix}b_1\\b_2\\\vdots\\b_m\end{bmatrix}=\begin{bmatrix}1&0&\cdots&0\\0&1&\cdots&0\\\vdots&\vdots&\ddots&\vdots\\0&0&\cdots&1\end{bmatrix}=I$$

B. 基变换

从更严格的数学定义上来说，PCA回答的问题是：如何寻找到另一组正交基，它们是标准正交基的线性组合，而且能够最好的表示数据集？

这里提出了PCA方法的一个最关键的假设：线性。这是一个非常强的假设条件。它使问题得到了很大程度的简化：1）数据被限制在一个向量空间中，能被一组基表示；2）隐含的假设了数据之间的连续性关系。

这样一来数据就可以被表示为各种基的线性组合。令X表示原数据集。X是一个$m\times n$的矩阵，它的每一个列向量都表示一个时间采样点上的数据$\vec{X}$，在上面的例子中，$m=6,n=120000$。Y表示转换以后的新的

数据集表示。P 是他们之间的线性转换。

$$PX = Y \tag{1}$$

有如下定义：

- p_i 表示 P 的行向量。
- x_i 表示 X 的列向量（或者 $\vec{X}$）。
- y_i 表示 Y 的列向量。

公式(1)表示不同基之间的转换，在线性代数中，它有如下的含义：

Ø P 是从 X 到 Y 的转换矩阵。

Ø 几何上来说，P 对 X 进行旋转和拉伸得到 Y。

Ø P 的行向量，$\{p_1,\cdots,p_m\}$ 是一组新的基，而 Y 是原数据 X 在这组新的基表示下得到的重新表示。

下面是对最后一个含义的显式说明：

$$PX = \begin{bmatrix} p_1 \\ \vdots \\ p_m \end{bmatrix} \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}$$

$$Y = \begin{bmatrix} p_1 \cdot x_1 & \cdots & p_1 \cdot x_n \\ \vdots & \ddots & \vdots \\ p_m \cdot x_1 & \cdots & p_m \cdot x_n \end{bmatrix}$$

注意到 Y 的列向量：

$$y_i = \begin{bmatrix} p_1 \cdot x_i \\ \vdots \\ p_m \cdot x_i \end{bmatrix}$$

可见 y_i 表示的是 x_i 与 P 中对应列的点积，也就是相当于是在对应向量上的投影。所以，P 的行向量事实上就是一组新的基。它对原数据 X 进行重新表示。在一些文献中，将数据 X 成为“源”，而将变换后的 Y 称为“信号”。这是由于变换后的数据更能体现信号成分的原因。

C. 问题

在线性的假设条件下，问题转化为寻找一组变换后的基，也就是 P 的行向量 $(p_1,\cdots,p_m)$，这些向量就是PCA中所谓的“主元”。问题转化为如下的形式：

- 怎样才能最好的表示原数据 X？
- P 的基怎样选择才是最好的？

解决问题的关键是如何体现数据的特征。那么，什么是数据的特征，如何体现呢？

方差和目标

“最好的表示”是什么意思呢？下面的章节将给出一个较为直观的解释，并增加一些额外的假设条件。在线性系统中，所谓的“混乱数据”通常包含以下的三种成分：噪音、旋转以及冗余。下面将对这三种成分做出数学上的描述并针对目标作出分析。

A. 噪音和旋转

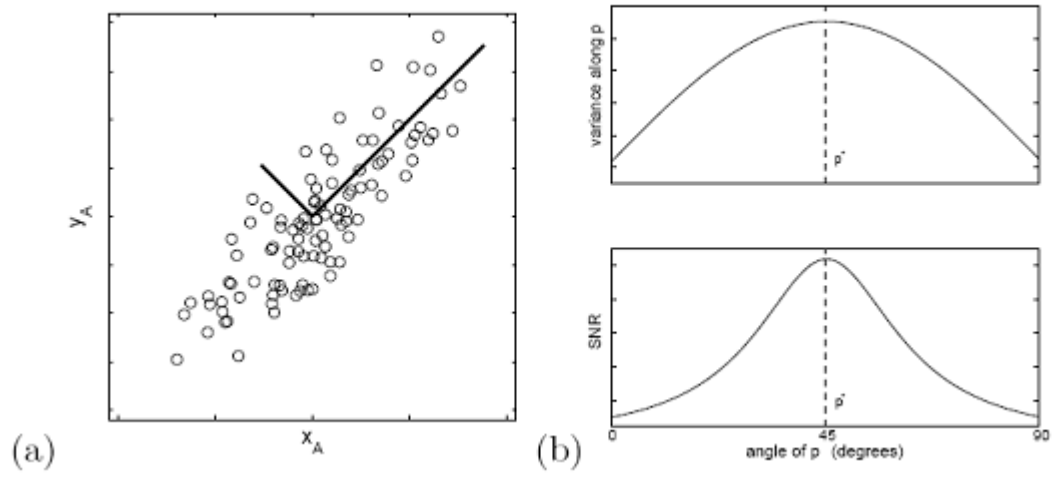
噪音对数据的影响是巨大的，如果不能对噪音进行区分，就不可能抽取数据中有用的信息。噪音的横梁有多种方式，最常见的定义是信噪比 SNR (*signal-to-noise ratio*)，或是方差比 σ^2：

$$SNR = \frac{\sigma^2_{signal}}{\sigma^2_{noise}} \tag{2}$$

比较大的信噪比表示数据的准确度高，而信噪比低则说明数据中的噪音成分比较多。那么怎样区分什么是信号，什么是噪音呢？这里假设，变化较大的信息被认为是信号，变化较小的则是噪音。事实上，这个标准等价于一个低通的滤波器，是一种标准的去噪准则。而变化的大小则是由方差来描述的。

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

它表示了采样点在平均值两侧的分布，对应于图表 2(a)就是采样点云的“胖瘦”。显然的，方差较大，也就是较“宽”较“胖”的分布，表示了采样点的主要分布趋势，是主信号或主要分量；而方差较小的分布则被认为是噪音或次要分量。

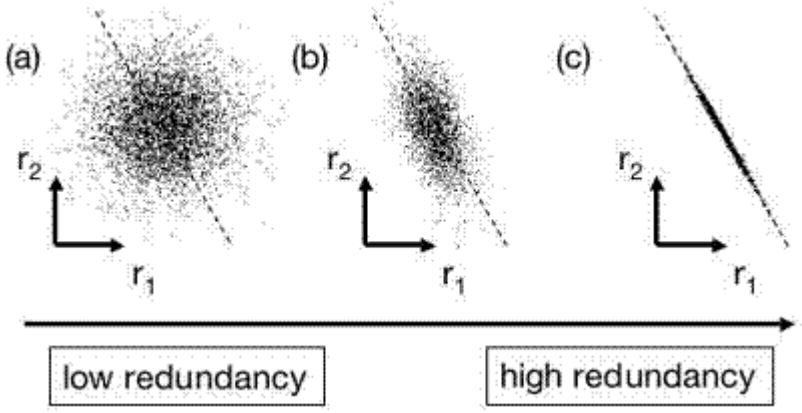


图表 2：(a)摄像机A的采集数据。图中黑色垂直直线表示一组正交基的方向。σ^2_{signal} 是采样点云在长线方向上分布的方差，而 σ^2_{noise} 是数据点在短线方向上分布的方差。(b)对 P 的基向量进行旋转使SNR和方差最大。

假设摄像机A拍摄到的数据如图表 2(a)所示，圆圈代表采样点，因为运动理论上是只存在于一条直线上，所以偏离直线的分布都属于噪音。此时 SNR 描述的就是采样点云在某对垂直方向上的概率分布的比值。那么，最大限度的揭示原数据的结构和关系，找出某条潜在的，最优的 x 轴，事实上等价寻找一对空间内的垂直直线（图中黑线表示，也对应于此空间的一组基），使得信噪比尽可能大的方向。容易看出，本例中潜在的 x 轴就是图上的较长黑线方向。那么怎样寻找这样一组方向呢？直接的想法是对基向量进行旋转。如图表 2(b)所示，随着这对直线的转动 SNR 以及方差的变化情况。应于 SNR 最大值的一组基 P^*，就是最优的“主元”方向。在进行数学中求取这组基的推导之前，先介绍另一个影响因素。

B. 冗余

有时在实验中引入了一些不必要的变量。可能会使两种情况：1）该变量对结果没有影响；2）该变量可以用其它变量表示，从而造成数据冗余。下面对这样的冗余情况进行分析和分类。



图表 3：可能冗余数据的频谱图表示。r_1 和 r_2 分别是两个不同的观测变量。（比如例子中的 x_A，y_B）。最佳拟合线 $r_2 = kr_1$ 用虚线表示。

如图表 3所示，它揭示了两个观测变量之间的关系。(a)图所示的情况是低冗余的，从统计学上说，这两个观测变量是相互独立的，它们之间的信息没有冗余。而相反的极端情况如(c)，r_1 和 r_2 高度相关，r_2 完全可以用 r_1

表示。一般来说，这种情况发生可能是因为摄像机A和摄像机B放置的位置太近或是数据被重复记录了，也可能是由于实验设计的不合理所造成的。那么对于观测者而言，这个变量的观测数据就是完全冗余的，应当去除，只用一个变量就可以表示了。这也就是PCA中“降维”思想的本源。

C. 协方差矩阵

对于上面的简单情况，可以通过简单的线性拟合的方法来判断各观测变量之间是否出现冗余的情况，而对于复杂的情况，需要借助协方差来进行衡量和判断：

$$\sigma_{AB}^2 = \frac{\sum_{i=1}^{n}(a_i-\bar{a})(b_i-\bar{b})}{n-1}$$

A，B 分别表示不同的观测变量所记录的一组值，在统计学中，由协方差的性质可以得到：

- $\sigma_{AB}^2 \geq 0$，且 $\sigma_{AB}^2 = 0$ 当且仅当观测变量 A，B 相互独立。
- $\sigma_{AB}^2 = \sigma_A^2$，当 $A = B$。

等价的，将 A，B 写成行向量的形式：

$$A = [a_1 \quad a_2 \quad \cdots \quad a_n], \quad B = [b_1 \quad b_2 \quad \cdots \quad b_n]$$

协方差可以表示为：

$$\sigma_{AB}^2 \equiv \frac{1}{n-1} AB^T \tag{3}$$

那么，对于一组具有 m 个观测变量，n 个采样时间点的采样数据 X，将每个观测变量的值写为行向量，可以得到一个 $m \times n$ 的矩阵：

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \tag{4}$$

接下来定义协方差矩阵如下：

$$C_X \equiv \frac{1}{n-1} XX^T \tag{5}$$

容易发现协方差矩阵 C_X 性质如下：

- C_X 是一个 $m \times m$ 的平方对称矩阵。
- C_X 对角线上的元素是对应的观测变量的方差。
- 非对角线上的元素是对应的观测变量之间的协方差。

$$C_X \equiv \begin{bmatrix} \sigma_{x_1x_1}^2 & \sigma_{x_1x_2}^2 & \cdots & \sigma_{x_1x_m}^2 \\ \sigma_{x_2x_1}^2 & \sigma_{x_2x_2}^2 & \cdots & \sigma_{x_2x_m}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{x_mx_1}^2 & \sigma_{x_mx_2}^2 & \cdots & \sigma_{x_mx_m}^2 \end{bmatrix}$$

协方差矩阵 C_X 包含了所有观测变量之间的相关性度量。更重要的是，根据前两节的说明，这些相关性度量反映了数据的噪音和冗余的程度。

- 在对角线上的元素越大，表明信号越强，变量的重要性越高；元素越小则表明可能是存在的噪音或是次要变量。
- 在非对角线上的元素大小则对应于相关观测变量对之间冗余程度的大小。

一般情况下，初始数据的协方差矩阵总是不太好的，表现为信噪比不高且变量间相关度大。PCA的目标就是通过基变换对协方差矩阵进行优化，找到相关“主元”。那么，如何进行优化？矩阵的那些性质是需要注意的呢？

D. 协方差矩阵的对角化

总结上面的章节，主元分析以及协方差矩阵优化的原则是：1）最小化变量冗余，对应于协方差矩阵的非对

角元素要尽量小；2）最大化信号，对应于要使协方差矩阵的对角线上的元素尽可能的大。因为协方差矩阵的每一项都是正值，最小值为0，所以优化的目标矩阵C_Y的非对角元素应该都是0，对应于冗余最小。所以优化的目标矩阵C_Y应该是一个对角阵。即只有对角线上的元素可能是非零值。同时，PCA假设P所对应的一组变换基$\{p_1,\ldots,p_m\}$必须是标准正交的，而优化矩阵C_Y对角线上的元素越大，就说明信号的成分越大，换句话就是对应于越重要的“主元”。

对于协方差矩阵进行对角化的方法很多。根据上面的分析，最简单最直接的算法就是在多维空间内进行搜索。和图表 2(a)的例子中旋转P的方法类似：

1） 在m维空间中进行遍历，找到一个方差最大的向量，令作p_1。

2） 在与p_1垂直的向量空间中进行遍历，找出次大的方差对应的向量，记作p_2。

3） 对以上过程循环，直到找出全部m的向量。它们生成的顺序也就是“主元”的排序。

这个理论上成立的算法说明了PCA的主要思想和过程。在这中间，牵涉到两个重要的特性：a)转换基是一组标准正交基。这给PCA的求解带来了很大的好处，它可以运用线性代数的相关理论进行快速有效的分解。这些方法将在后面提到。b）在PCA的过程中，可以同时得到新的基向量所对应的“主元排序”，利用这个重要性排序可以方便的对数据进行光顺、简化处理或是压缩。

A. PCA的假设和局限

PCA的模型中存在诸多的假设条件，决定了它存在一定的限制，在有些场合可能会造成效果不好甚至失效。对于学习和掌握PCA来说，理解这些内容是非常重要的，同时也有利于理解基于改进这些限制条件的PCA的一些扩展技术。

PCA的假设条件包括：

1. 线形性假设。

如同文章开始的例子，PCA的内部模型是线性的。这也就决定了它能进行的主元分析之间的关系也是线性的。现在比较流行的*kernel*-PCA的一类方法就是使用非线性的权值对原有PCA技术的拓展。

2. 使用中值和方差进行充分统计。

使用中值和方差进行充分的概率分布描述的模型只限于指数型概率分布模型。（例如高斯分布），也就是说，如果我们考察的数据的概率分布并不满足高斯分布或是指数型的概率分布，那么PCA将会失效。在这种模型下，不能使用方差和协方差来很好的描述噪音和冗余，对教化之后的协方差矩阵并不能得到很合适的结果。

事实上，去除冗余的最基础的方程是：

$$P(y_1,y_2)=P(y_1)P(y_2)$$

其中$P(\cdot)$代表概率分布的密度函数。基于这个方程进行冗余去除的方法被称作独立主元分析(ICA)方法(*Independent Component Analysis*)。不过，所幸的是，根据中央极限定理，现实生活中所遇到的大部分采样数据的概率分布都是遵从高斯分布的。所以PCA仍然是一个使用于绝大部分领域的稳定且有效的算法。

3. 大方差向量具有较大重要性。

PCA方法隐含了这样的假设：数据本身具有较高的信噪比，所以具有最高方差的一维向量就可以被看作是主元，而方差较小的变化则被认为是噪音。这是由于低通滤波器的选择决定的。

4. 主元正交。

PCA方法假设主元向量之间都是正交的，从而可以利用线形代数的一系列有效的数学工具进行求解，大大提高了效率和应用的范围。

PCA求解：特征根分解

在线形代数中，PCA问题可以描述成以下形式：

寻找一组正交基组成的矩阵P，有$Y=PX$，使得$C_Y\equiv\frac{1}{n-1}YY^T$是对角阵。则P的行向量（也就是一组正交基），就是数据X的主元向量。

对C_Y进行推导：

$$
\begin{aligned}
C_Y &= \frac{1}{n-1} YY^T \\
&= \frac{1}{n-1} (PX)(PX)^T \\
&= \frac{1}{n-1} PXX^T P^T \\
&= \frac{1}{n-1} P(XX^T)P^T \\
C_Y &= \frac{1}{n-1} PAP^T
\end{aligned}
$$

定义 $A \equiv XX^T$，则 A 是一个对称阵。对 A 进行对角化求取特征向量得：

$$A = EDE^T$$

则 D 是一个对角阵而 E 则是对称阵 A 的特征向量排成的矩阵。

这里要提出的一点是，A 是一个 $m \times m$ 的矩阵，而它将有 $r\ (r \le m)$ 个特征向量。其中 r 是矩阵 A 的秩。如果 $r < m$，则 A 即为退化阵。此时分解出的特征向量不能覆盖整个 m 空间。此时只需要在保证基的正交性的前提下，在剩余的空间中任意取得 $m - r$ 维正交向量填充 E 的空格即可。它们将不对结果造成影响。因为此时对应于这些特征向量的特征值，也就是方差值为零。

求出特征向量矩阵后我们取 $P = E^T$，则 $A = P^T DP$，由线形代数可知矩阵 P 有性质 $P^{-1} = P^T$，从而进行如下计算：

$$
\begin{aligned}
C_Y &= \frac{1}{n-1} PAP^T \\
&= \frac{1}{n-1} P(P^T DP)P^T \\
&= \frac{1}{n-1} (PP^T) D (PP^T) \\
&= \frac{1}{n-1} (PP^{-1}) D (PP^{-1}) \\
C_Y &= \frac{1}{n-1} D
\end{aligned}
$$

可知此时的 P 就是我们需要求得变换基。至此我们可以得到PCA的结果：

- X 的主元即是 XX^T 的特征向量，也就是矩阵 P 的行向量。
- 矩阵 C_Y 对角线上第*i*个元素是数据 X 在方向 p_i 的方差。

我们可以得到PCA求解的一般步骤：

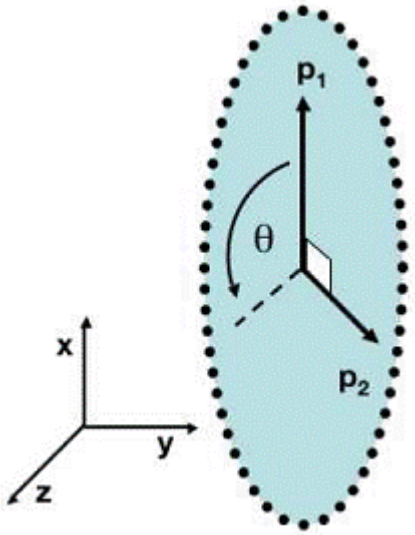
1）采集数据形成 $m \times n$ 的矩阵。m 为观测变量个数，n 为采样点个数。

2）在每个观测变量（矩阵行向量）上减去该观测变量的平均值得到矩阵 X。

3）对 XX^T 进行特征分解，求取特征向量以及所对应的特征根。

总结和讨论

- PCA技术的一大好处是对数据进行降维的处理。我们可以对新求出的“主元”向量的重要性进行排序，根据需要取前面最重要的部分，将后面的维数省去，可以达到降维从而简化模型或是对数据进行压缩的效果。同时最大程度的保持了原有数据的信息。
  在前文的例子中，经过PCA处理后的数据只剩下了一维，也就是弹簧运动的那一维，从而去除了冗余的变量，揭示了实验数据背后的物理原理。
- PCA技术的一个很大的优点是，它是完全无参数限制的。在PCA的计算过程中完全不需要人为的设定参数或是根据任何经验模型对计算进行干预，最后的结果只与数据相关，与用户是独立的。
  但是，这一点同时也可以看作是缺点。如果用户对观测对象有一定的先验知识，掌握了数据的一些特征，却无法通过参数化等方法对处理过程进行干预，可能会得不到预期的效果，效率也不高。



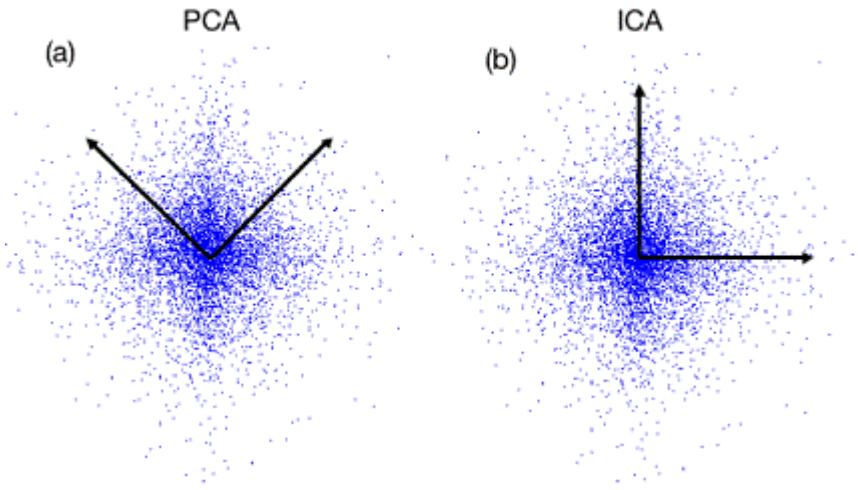
图表 4：黑色点表示采样数据，排列成转盘的形状。
容易想象，该数据的主元是(P_1, P_2)或是旋转角θ。

如图表 4中的例子，PCA找出的主元将是(P_1, P_2)。但是这显然不是最优和最简化的主元。(P_1, P_2)之间存在着非线性的关系。根据先验的知识可知旋转角θ是最优的主元。则在这种情况下，PCA就会失效。但是，如果加入先验的知识，对数据进行某种划归，就可以将数据转化为以θ为线性的空间中。这类根据先验知识对数据预先进行非线性转换的方法就成为*kernel*-PCA，它扩展了PCA能够处理的问题的范围，又可以结合一些先验约束，是比较流行的方法。

- 有时数据的分布并不是满足高斯分布。如图表 5所示，在非高斯分布的情况下，PCA方法得出的主元可能并不是最优的。在寻找主元时不能将方差作为衡量重要性的标准。要根据数据的分布情况选择合适的描述完全分布的变量，然后根据概率分布式

$$P(y_1, y_2) = P(y_1)P(y_2)$$

来计算两个向量上数据分布的相关性。等价的，保持主元间的正交假设，寻找的主元同样要使$P(y_1, y_2) = 0$。这一类方法被称为独立主元分解(ICA)。



图表 5：数据的分布并不满足高斯分布，呈明显的十字星状。
这种情况下，方差最大的方向并不是最优主元方向。

- PCA方法和线形代数中的奇异值分解(SVD)方法有内在的联系，一定意义上来说，PCA的解法是SVD的一种变形和弱化。对于$m \times n$的矩阵X，通过奇异值分解可以直接得到如下形式：

$$X = U \sum V^T$$

其中U是一个$m \times m$的矩阵，V是一个$n \times n$的矩阵，而Σ是$m \times n$的对角阵。Σ形式如下：

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & \mathbf{0} \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ \mathbf{0} & & & & \ddots \\ & & & & & 0 \end{bmatrix}$$

其中$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r$，是原矩阵的奇异值。由简单推导可知，如果对奇异值分解加以约束：U的向量必须正交，则矩阵U即为PCA的特征值分解中的E，则说明PCA并不一定需要求取XX^T，也可以直接对原数据矩阵X进行SVD奇异值分解即可得到特征向量矩阵，也就是主元向量。

计算机视学领域的应用

PCA方法是一个具有很高普适性的方法，被广泛应用于多个领域。这里要特别介绍的是它在计算机视觉领域的应用，包括如何对图像进行处理以及在人脸识别方面的特别作用。

A. 数据表示

如果要将PCA方法应用于视觉领域，最基本的问题就是图像的表达。如果是一幅$N \times N$大小的图像，它的数据将被表达为一个N^2维的向量：

$$X = (x_1 \quad x_2 \quad \ldots \quad x_{N^2})^T$$

在这里图像的结构将被打乱，每一个像素点被看作是一维，最直接的方法就是将图像的像素一行行的头尾相接成一个一维向量。还必须要注意的是，每一维上的数据对应于对应像素的亮度、灰度或是色彩值，但是需要划归到同一纬度上。

B. 模式识别

假设数据源是一系列的20幅图像，每幅图像都是$N \times N$大小，那么它们都可以表示为一个N^2维的向量。将它们排成一个矩阵：

$$ImagesMatrix = (ImageVec1 \quad ImageVec2 \quad \ldots \quad ImageVec20)$$

然后对它们进行PCA处理，找出主元。

为什么这样做呢？据人脸识别的例子来说，数据源是20幅不同的人脸图像，PCA方法的实质是寻找这些图像中的相似的维度，因为人脸的结构有极大的相似性（特别是同一个人的人脸图像），则使用PCA方法就可以很容易的提取出人脸的内在结构，也及时所谓“模式”，如果有新的图像需要与原有图像比较，就可以在变换后的主元维度上进行比较，则可衡量新图与原有数据集的相似度如何。

对这样的一组人脸图像进行处理，提取其中最重要的主元，即可大致描述人脸的结构信息，称作“特征脸”(*EigenFace*)。这就是人脸识别中的重要方法“特征脸方法”的理论根据。近些年来，基于对一般PCA方法的改进，结合ICA、*kernel*-PCA等方法，在主元分析中加入关于人脸图像的先验知识，则能得到更好的效果。

C. 图像信息压缩

使用PCA方法进行图像压缩，又被称为*Hotelling*算法，或者*Karhunen and Leove(KL)*变换。这是视觉领域内图像处理的经典算法之一。具体算法与上述过程相同，使用PCA方法处理一个图像序列，提取其中的主元。然后根据主元的排序去除其中次要的分量，然后变换回原空间，则图像序列因为维数降低得到很大的压缩。例如上例中取出次要的5个维度，则图像就被压缩了1/4。但是这种有损的压缩方法同时又保持了其中最“重要”的信息，是一种非常重要且有效的算法。

参考文献

[1] Lindsay I Smith. (2002) “A tutorial on Principal Components Analysis”

[2] Jonathon Shlens. (2005) “A Tutorial on Principal Component Analysis”

[3] Will, Todd (1999) “Introduction to the Singular Value Decomposition” Davidson College. http://www.davidson.edu/academic/math/will/svd/index.html

[4] Bell, Anthony and Sejnowski, Terry. (1997) “The Independent Components of Natural Scenes are EdgeFilters.” Vision Research 37(23), 3327-3338.

[5] T.F. Cootes and C.J.Taylor (2004) “Statistical Models of Appearance for Computer Vision”

http://www.isbe.man.ac.uk/~bim/Models/app_models.pdf
[6] 张翠平 苏光大 (2000)“人脸识别技术综述”《中国图像图形学报》第五卷A版第11期
[7] 何国辉 甘俊英 (2006)“PCA类内平均脸法在人脸识别中的应用研究”《计算机应用研究》2006年第三期
[8] 牛丽平 付仲良 魏文利 (2006)“人脸识别技术研究”《电脑开发与应用》2006年第五期
[9] Wikipedia “principal components analysis”词条解释 From Answers.com