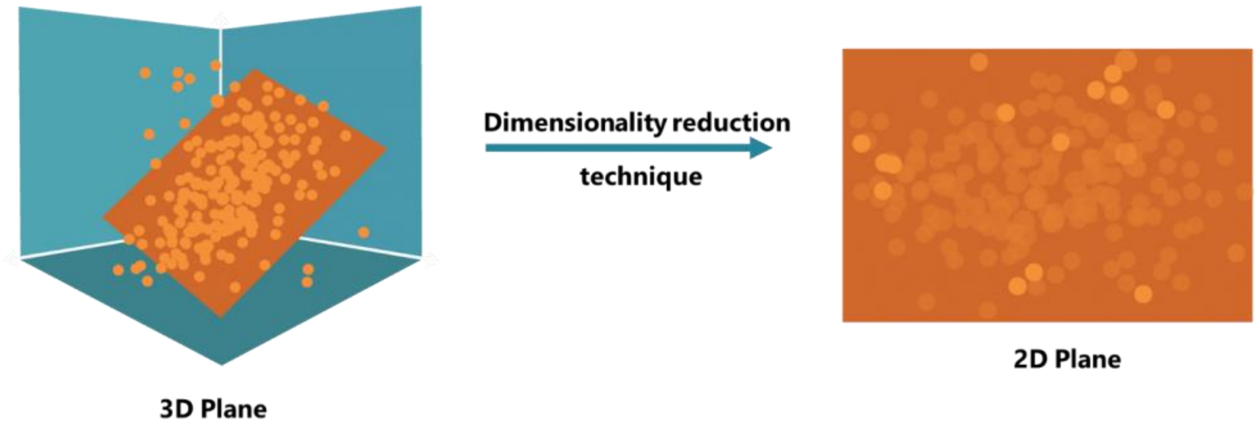
主成分分析

Principal Component Analysis



赵海涛

haitaozhao@ecust.edu.cn

大纲

1 主成分分析介绍

2 经典主成分分析

3 主成分个数的选择

4 高维数据下的PCA

5 PCA与SVD的关系

6 主成分分析案例分析

1 主成分分析介绍

多个变量之间往往具有一定的相关性。当变量个数较多而且变量之间存在相关性时，会显著增加分析问题的复杂性和解决问题的运算量

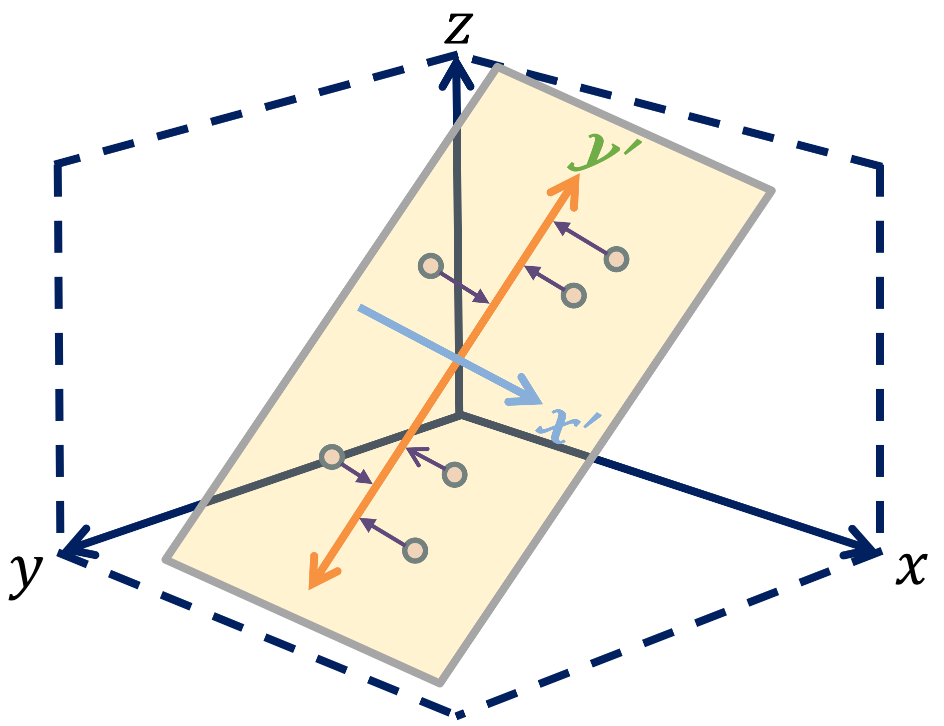
如果有一种方法可以将多个变量综合为少数几个主要特征，使这些特征既可以保留原始数据的主要信息，则可以有效提升特征的表达能力，降低训练的复杂度。在这其中，最经典的方法就是主成分分析法（Principal components analysis，PCA）

1.主成分分析介绍

如图所示，在三维空间中有一系列数据点，如果我们用自然坐标系x,y,z三个轴来表示数据，就需要使用三个维度。

而实际上，这些点只出现在一个二维平面上，那么我们就可以通过x',y'这两个维度表达原始数据，并且没有任何损失。而x',y'两个轴所包含的信息就是我们所要找的主成分。

PCA旨在找到数据中的主成分，并利用这些主成分表征原始数据，从而达到降维目的



2 经典主成分分析

那么在实际应用中，如何将多个变量综合为少数几个代表性的特征，使这些特征既可以代表原始变量的主要信息又互不相关？

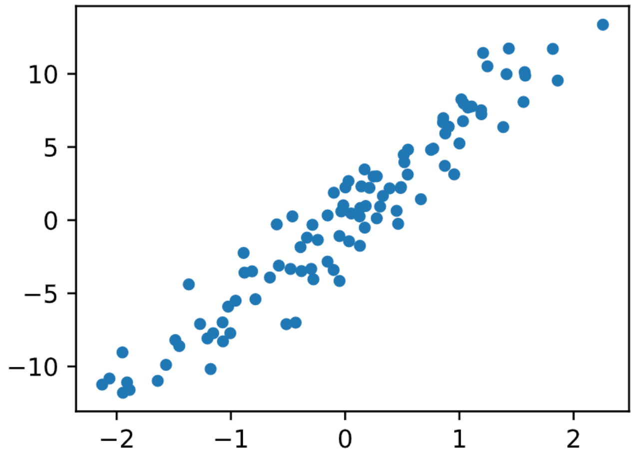
主要从以下两个角度来分析主成分的求解方式，并对两种求解方式的关系进行分析。学习时应掌握两种解法中主成分求解的过程以及两种解法的关系。

最大方差准则

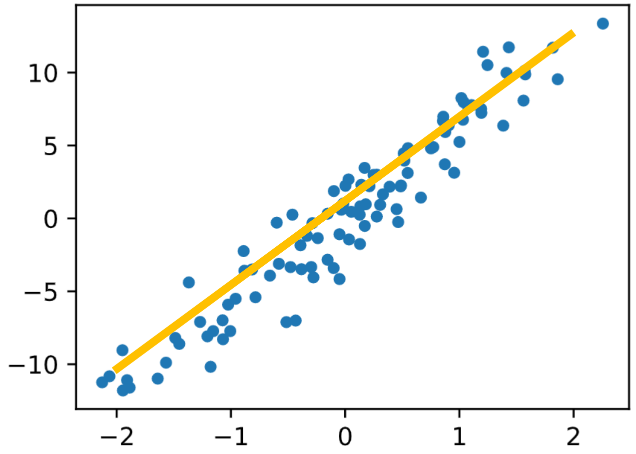
最小平方误差准则

2.1经典主成分分析—最大方差准则

图(a)是二维空间中经过中心化的一组数据，我们很容易看出主成分所在的轴(以下称为主轴)的大致方向，即图(b)中黄线所处的轴。因为在黄线所处的轴上，数据分布得更为分散，这也意味着数据在这个方向上方差更大。在信号处理中认为信号具有较大的方差，噪声有较小的方差。由此我们不难引出PCA的目标，即最大化投影方差，也就是让数据在主轴上投影的方差最大。我们称这种想法为PCA最大方差准则。



(a)



(b)

2.1经典主成分分析—最大方差准则

在最大方差准则中，求解PCA中主成分的问题可归结如下：

假设数据$\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n] \in \mathcal{R}^{m\times n}$，由$n$个$m$维向量组成（为不失一般性，假设数据已中心化，即$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i = 0$，**不然PCA的第一步应该先对数据进行中心化**），PCA希望找到一个单位投影向量$\boldsymbol{u} \in \mathcal{R}^{m\times 1}$（$\boldsymbol{u}^T\boldsymbol{u} = 1$），使得所有的$\boldsymbol{x}_i$在这个方向上的正交投影方差达到最大。换言之，希望数据在这个方向上有最大的散布。

2.1经典主成分分析—最大方差准则

由于数据已经中心化处理，故数据均值为0，即有：

$$\mu' = \frac{1}{\mathrm{n}}\sum_{\mathrm{i=1}}^{\mathrm{n}} \boldsymbol{x_i}^T \boldsymbol{u} = \left(\frac{1}{\mathrm{n}}\sum_{\mathrm{i=1}}^{\mathrm{n}} \boldsymbol{x_i}^T\right)\boldsymbol{u} = 0$$

投影后的方差可以表示为：

$$D(\boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{u}^T\boldsymbol{x}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{u}^T\boldsymbol{x}_i)(\boldsymbol{x}_i{}^T\boldsymbol{u}) = \boldsymbol{u}^T\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i{}^T\right)\boldsymbol{u}$$

2.1经典主成分分析—最大方差准则

因此，PCA问题的目标函数如下：

$$\max_{\boldsymbol{u}^T\boldsymbol{u}=1} \boldsymbol{u}^T\boldsymbol{C}\boldsymbol{u}$$

其中：$\boldsymbol{C} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^T = \frac{1}{n}\boldsymbol{X}\boldsymbol{X}^T \in \mathcal{R}^{m\times m}$

2.1经典主成分分析—最大方差准则（解法1）

对于目标函数：$\max\limits_{\boldsymbol{u}^T\boldsymbol{u}=1} \boldsymbol{u}^T\boldsymbol{C}\boldsymbol{u}$

- 令协方差矩阵$\boldsymbol{C}$的特征分解为：$\boldsymbol{C} = \boldsymbol{W}\boldsymbol{\Lambda}\boldsymbol{W}^T$

其中W为正交阵$W = [\boldsymbol{w_1}, \boldsymbol{w_2}, \cdots, \boldsymbol{w_m}]$, $\Lambda = diag(\lambda_1, \lambda_2, \cdots, \lambda_m)$, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$

- 令单位向量$\boldsymbol{u}$的旋转变换为：$\boldsymbol{u} = \boldsymbol{W}\boldsymbol{a} = \sum_{i=1}^{n} \boldsymbol{a}_i\boldsymbol{w}_i$, $\boldsymbol{a} = [a_1, a_2, \cdots, a_m]^T$

由 $\boldsymbol{u}^T\boldsymbol{u} = 1$($\boldsymbol{u}$是单位向量)，可得 $\boldsymbol{a}^T\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{a} = 1$

由 $\boldsymbol{W}^T\boldsymbol{W} = I$ ($\boldsymbol{W}$是正交阵)，可得$\boldsymbol{a}^T\boldsymbol{a} = \sum_{i=1}^{m} a_i{}^2 = 1$

2.1经典主成分分析—最大方差准则（解法１）

因此目标函数可以转变为

$$\boldsymbol{u}^T\boldsymbol{C}\boldsymbol{u} = \boldsymbol{a}^T\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{\Lambda}\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{a} = \boldsymbol{a}^T\boldsymbol{\Lambda}\boldsymbol{a} = \sum_{i=1}^{m} {a_i}^2\lambda_i \leq \lambda_1\sum_{i=1}^{m} {a_i}^2 = \lambda_1$$

当$a_1 = 1$，$a_1 = a_2 = \cdots = a_m = 0$时，等号成立。此时$\boldsymbol{u} = \boldsymbol{w}_1$

因此，主成分可取为对应于$\boldsymbol{C}$矩阵最大特征值的特征向量。

2.1经典主成分分析—最大方差准则（解法1）

$\boldsymbol{u} = \boldsymbol{w}_1$是我们求得的第一个主方向，如果我们希望求取更多的方向用来刻画数据，那么要对目标函数进行约束（否则每次的解都一样），为此我们引入正交约束。现用归纳法证明第r个主方向$\boldsymbol{u}_r$是特征值λ_r所对应的特征向量$\boldsymbol{w}_r$。

1）$n = 1$时：由之前的主成分求解的过程可知$\boldsymbol{u}_1 = \boldsymbol{w}_1$。

2）$n \geq 2$时，假设$n = r - 1$时成立，即有$\boldsymbol{u}_1 = \boldsymbol{w}_1, \boldsymbol{u}_2 = \boldsymbol{w}_2, \cdots, \boldsymbol{u}_{r-1} = \boldsymbol{w}_{r-1}$，现需证明$n = r$时命题依旧成立，即证明$\boldsymbol{u}_r = \boldsymbol{w}_r$（其中$\boldsymbol{u}_r^T\boldsymbol{u}_i = 0$，$i = 1, \cdots, r-1$，$\boldsymbol{u}_r^T\boldsymbol{u}_r = 1$，即$\boldsymbol{u}_r$与$\boldsymbol{u}_i$正交且$\boldsymbol{u}_r$的模为1）。

2.1经典主成分分析—最大方差准则（解法1）

对于目标函数: $\max_{\substack{\boldsymbol{u}_r^T\boldsymbol{u}_r=1\\ \boldsymbol{u}_r^T\boldsymbol{u}_i=0\\ i=1,\cdots,r-1}} \boldsymbol{u}_r{}^T\boldsymbol{C}\boldsymbol{u}_r$

- 令$\boldsymbol{C} = \boldsymbol{W}\boldsymbol{\Lambda}\boldsymbol{W}^T$
- 令$\boldsymbol{u}_r = \boldsymbol{W}\boldsymbol{b} = \sum_{i=1}^{m} b_i\boldsymbol{w}_i$

由 $\boldsymbol{u}_r{}^T\boldsymbol{u}_r = 1$, 可得$\boldsymbol{b}^T\boldsymbol{b} = \sum_{i=1}^{m} b_i{}^2 = 1$

由 $\boldsymbol{u}_r{}^T\boldsymbol{u}_i = 0$, 可得$(\sum_{j=1}^{m} b_j\,\boldsymbol{w}_j)^T\boldsymbol{w}_i = 0$；因为$\boldsymbol{W}$为正交阵，有$\boldsymbol{w}_i\boldsymbol{w}_j = 0, \boldsymbol{w}_i\boldsymbol{w}_i = 1$,

因此 $\left(\sum_{j=1}^{m} b_j\,\boldsymbol{w}_j\right)^T\boldsymbol{w}_i = b_i\boldsymbol{w}_i\boldsymbol{w}_i = b_i = 0, (i = 1,\cdots,r-1)$

故结合$\boldsymbol{u}_r{}^T\boldsymbol{u}_r = 1$，$\boldsymbol{u}_r{}^T\boldsymbol{u}_i = 0$，可得$\displaystyle\sum_{i=r}^{m} b_i{}^2 = 1$

2.1经典主成分分析—最大方差准则（解法1）

因此目标函数可以转变为

$$\boldsymbol{u}_r{}^T\boldsymbol{C}\boldsymbol{u}_r=\boldsymbol{b}^T\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{\Lambda}\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{b}=\boldsymbol{b}^T\boldsymbol{\Lambda}\boldsymbol{b}=\sum_{i=1}^{m}b_i{}^2\lambda_i\sum_{i=r}^{m}b_i{}^2\lambda_i\le\lambda_r\sum_{i=\mathrm{r}}^{m}b_i{}^2=\lambda_r$$

$b_r=1$，$b_1=b_2=\cdots b_{r-1}=b_{r+1}=\cdots b_m=0$时，等号成立。此时$\boldsymbol{u}_r=\boldsymbol{w}_r$。

综合1）2），结合数据归纳法可知，第r个主方向$\boldsymbol{u}_r$是特征值λ_r所对应的特征向量$\boldsymbol{w}_r$。

2.1经典主成分分析—最大方差准则（解法2）

对于目标函数：

$$\max_{\boldsymbol{u}^T\boldsymbol{u}=1} \boldsymbol{u}^T\boldsymbol{C}\boldsymbol{u}$$

引入拉格朗日乘子λ：

$$L = \boldsymbol{u}^T\boldsymbol{C}\boldsymbol{u} + \lambda(1 - \boldsymbol{u}^T\boldsymbol{u})$$

对u求导并令导数等于0，有：

$$\frac{\partial L}{\partial \boldsymbol{u}} = 2\boldsymbol{C}\boldsymbol{u} - 2\lambda\boldsymbol{u} = 0$$

即：

$$\boldsymbol{C}\boldsymbol{u} = \lambda\boldsymbol{u}$$

由线性代数相关知识可知，λ为矩阵$\boldsymbol{C}$的特征值，$\boldsymbol{u}$为特征值λ对应的特征向量

2.1经典主成分分析—最大方差准则（解法2）

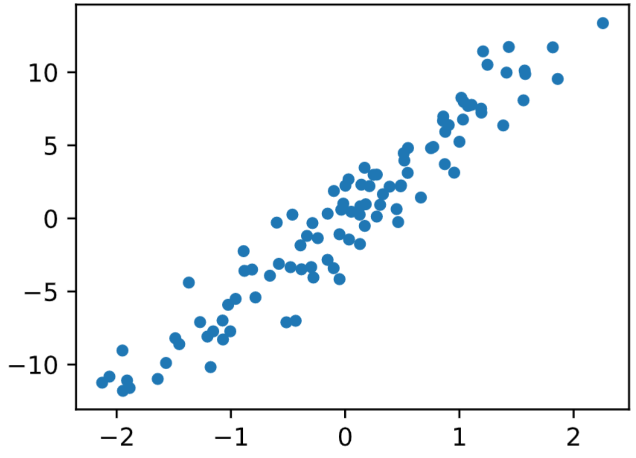
此时目标函数为：

$$D(x) = \boldsymbol{u}^T \boldsymbol{C} \boldsymbol{u} = \lambda \boldsymbol{u}^T \boldsymbol{u} = \lambda$$

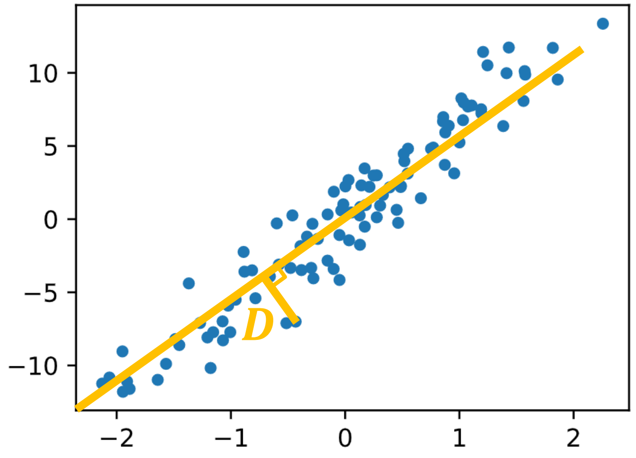
由此可知，要使$\boldsymbol{u}^T\boldsymbol{C}\boldsymbol{u}$在$\boldsymbol{u}^T\boldsymbol{u} = 1$条件下取得最大值，当且仅当$\lambda$取$C$最大的特征值，主成分$\boldsymbol{u}$取最大特征值$\lambda$对应的特征向量。且目标函数的最大值（数据投影方差最大值）为协方差矩阵C的最大特征值。第二个主成分（次佳投影方向）取第二大特征值对应的特征向量，依次类推。

2.2经典主成分分析—最小平方误差准则

重新看(a)中数据的主成分，从求解直线的思路出发，容易联想到线性回归问题。我们可以把其看成求解一个线性函数，使得对应直线能够更好拟合样本数据点。顺着这个思路，在高维空间中，我们实际上是要找到一个d维超平面，使得数据点到这个超平面的距离平方和最小。若$d = 1$，超平面退化为直线，即把样本点投影到最佳直线，最小化的就是所有点到直线的距离平方之和，如图(c)所示。我们称这种想法为PCA最小平方误差准则



(a)



(c)

2.2经典主成分分析—最小平方误差准则

在最小平方误差理论中，求解PCA中主成分的问题可归结如下：

假设数据$\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n] \in \mathcal{R}^{m\times n}$，并已中心化，PCA的目的是想找到一个单位向量$\boldsymbol{u} \in \mathcal{R}^{m\times 1}$ $(\boldsymbol{u}^T\boldsymbol{u} = 1)$，使得数据集中的每个点$\boldsymbol{x}_i$到单位向量$\boldsymbol{u}$的距离的平方均值最小，即：

$$\min \frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{x}_i - \boldsymbol{P}\boldsymbol{x}_i\|^2$$

- $\boldsymbol{u}^T\boldsymbol{x}_i$或者$\boldsymbol{x}_i{}^T\boldsymbol{u}$可以表示$\boldsymbol{x}_i$在$\boldsymbol{u}$方向上的投影长度，
- $\boldsymbol{P}\boldsymbol{x}_i = (\boldsymbol{u}^T\boldsymbol{x}_i)\boldsymbol{u}$，注：$\boldsymbol{u}^T\boldsymbol{x}_i$代表向量的长度，$\boldsymbol{u}$为单位向量，代表向量的方向）

2.2经典主成分分析—最小平方误差准则

PCA问题的目标函数如下：

$$\min_{\boldsymbol{u}^T\boldsymbol{u}=1} \frac{1}{n}\sum_{i=1}^{n} \|\boldsymbol{x}_i - \boldsymbol{x}_i{}'\|_2^2$$

其中：$\boldsymbol{x}_i{}' = (\boldsymbol{u}^T\boldsymbol{x}_i)\boldsymbol{u}$

由于$\boldsymbol{x}_i$，$\boldsymbol{x}_i{}' \in \mathcal{R}^{m\times 1}$，故$\boldsymbol{x}_i{}^T\boldsymbol{x}_i{}' = \boldsymbol{x}_i{}'^T\boldsymbol{x}_i \in \mathcal{R}^1$；对于目标函数：

$$\begin{aligned}\|\boldsymbol{x}_i - \boldsymbol{x}_i{}'\|_2^2 &= (\boldsymbol{x}_i - \boldsymbol{x}_i{}')^T(\boldsymbol{x}_i - \boldsymbol{x}_i{}') \\ &= \boldsymbol{x}_i{}^T\boldsymbol{x}_i - \boldsymbol{x}_i{}^T\boldsymbol{x}_i{}' - \boldsymbol{x}_i{}'^T\boldsymbol{x}_i \\ &= \boldsymbol{x}_i{}^T\boldsymbol{x}_i - 2\boldsymbol{x}_i{}^T\boldsymbol{x}_i{}' + \boldsymbol{x}_i{}'^T\boldsymbol{x}_i{}'\end{aligned}$$

2.2经典主成分分析—最小平方误差准则

$$\|\boldsymbol{x}_i - \boldsymbol{x}_i'\|_2^2 = \boldsymbol{x}_i{}^T\boldsymbol{x}_i - 2\boldsymbol{x}_i{}^T\boldsymbol{x}_i' + \boldsymbol{x}_i'^T\boldsymbol{x}_i'$$

- $\boldsymbol{x}_i{}^T\boldsymbol{x}_i$由于仅与数据样本有关，故可以先不作考虑。
- $\boldsymbol{x}_i{}^T\boldsymbol{x}_i' = \boldsymbol{x}_i{}^T(\boldsymbol{u}^T\boldsymbol{x}_i)\boldsymbol{u} = (\boldsymbol{u}^T\boldsymbol{x}_i)\boldsymbol{x}_i{}^T\boldsymbol{u} = \boldsymbol{u}^T\boldsymbol{x}_i\boldsymbol{x}_i{}^T\boldsymbol{u}$
- 由于$\boldsymbol{u}^T\boldsymbol{x}_i$表示投影长度，是一个数；$\boldsymbol{u}$为单位向量，故

$$\boldsymbol{x}_i'^T\boldsymbol{x}_i' = \left((\boldsymbol{u}^T\boldsymbol{x}_i)\boldsymbol{u}\right)^T(\boldsymbol{u}^T\boldsymbol{x}_i)\boldsymbol{u} = (\boldsymbol{u}^T\boldsymbol{x}_i)(\boldsymbol{u}^T\boldsymbol{x}_i) = (\boldsymbol{u}^T\boldsymbol{x}_i)(\boldsymbol{x}_i{}^T\boldsymbol{u}) = \boldsymbol{u}^T\boldsymbol{x}_i\boldsymbol{x}_i{}^T\boldsymbol{u}$$

因此有： $$\frac{1}{n}\|\boldsymbol{x}_i - \boldsymbol{x}_i'\|_2^2 = \frac{1}{n}\boldsymbol{x}_i{}^T\boldsymbol{x}_i - \frac{1}{n}\boldsymbol{u}^T\boldsymbol{x}_i\boldsymbol{x}_i{}^T\boldsymbol{u}$$

2.2经典主成分分析—最小平方误差准则

故原目标函数可优化成：

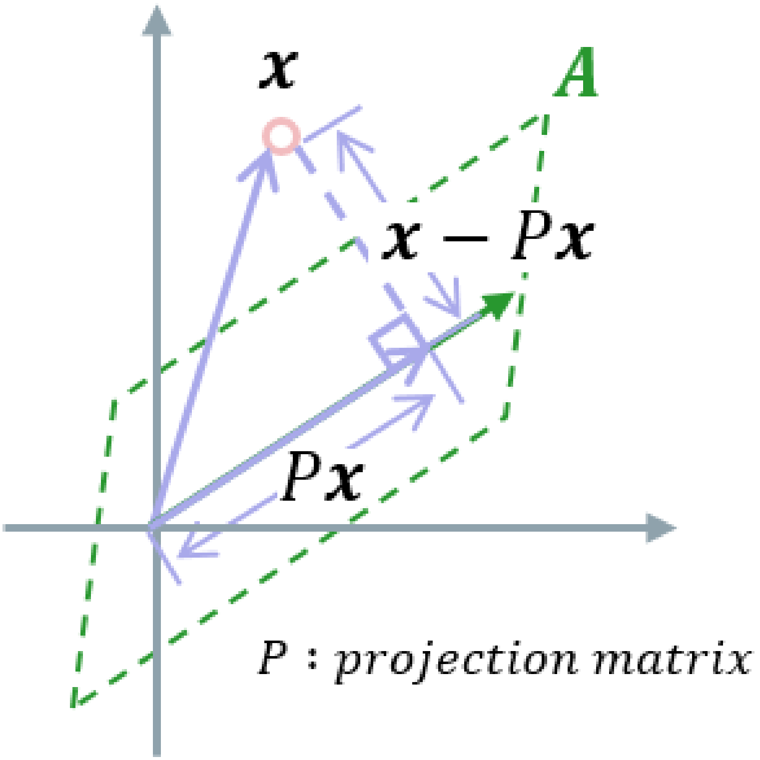
$$\min_{\boldsymbol{u}^T\boldsymbol{u}=1}\sum_{i=1}^{n}-\frac{1}{n}\boldsymbol{u}^T\boldsymbol{x}_i{\boldsymbol{x}_i}^T\boldsymbol{u}=\max_{\boldsymbol{u}^T\boldsymbol{u}=1}\sum_{i=1}^{n}\frac{1}{n}\boldsymbol{u}^T\boldsymbol{x}_i{\boldsymbol{x}_i}^T\boldsymbol{u}$$

$$=\max_{\boldsymbol{u}^T\boldsymbol{u}=1}\boldsymbol{u}^T\left(\sum_{i=1}^{n}\frac{1}{n}\boldsymbol{x}_i{\boldsymbol{x}_i}^T\right)\boldsymbol{u}$$

可以发现，最小平方误差准则优化完的目标函数等同于最大方差准则的目标函数，因此两者求解的主成分（最佳投影方向）一致，即为协方差矩阵的最大特征值所对应的特征向量。

2.3经典主成分分析—两种解法的关系

如图所示，最大方差理论考虑数据点投影在绿色轴线上的长度（$\boldsymbol{P x}$），最小均方误差理论考虑数据点与绿色轴线的距离（$\boldsymbol{x}-\boldsymbol{P x}$）。



由三角形勾股定理可知：

$$\|\boldsymbol{x}\|^2=\|\boldsymbol{P x}\|^2+\|\boldsymbol{x}-\boldsymbol{P x}\|^2$$

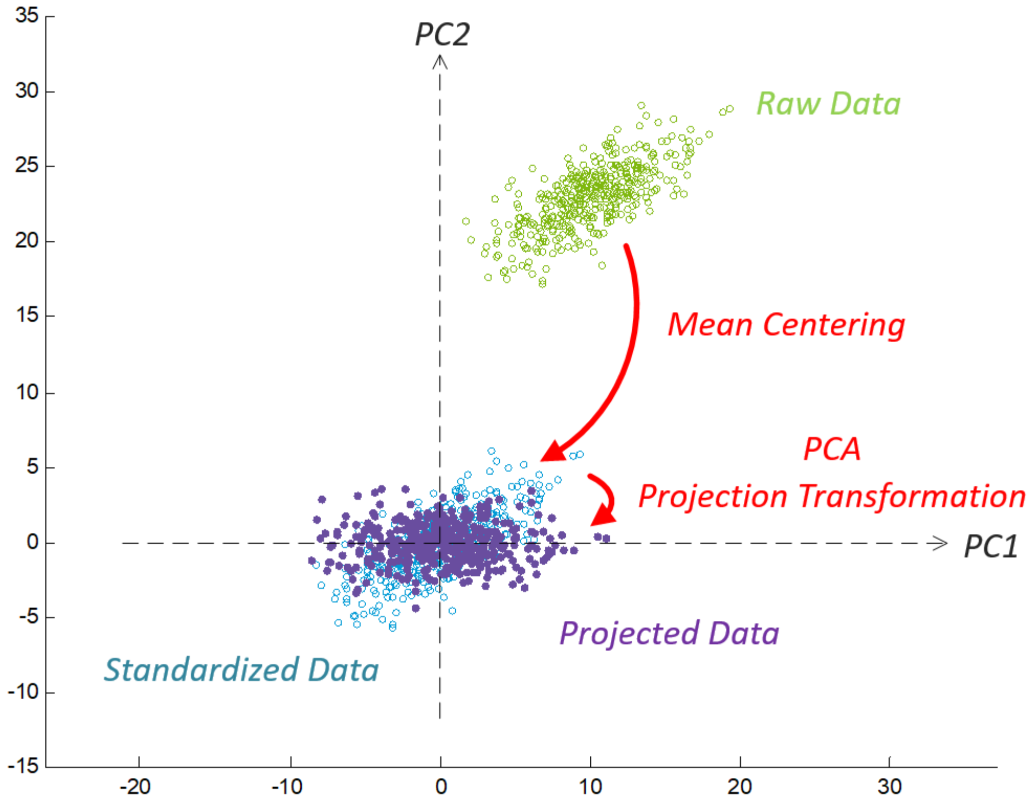
更一般地，对于数据$\boldsymbol{X}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n]\in\mathcal{R}^{m\times n}$：

$$\frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{x}_i\|^2=\frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{P x}_i\|^2+\frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{x}_i-\boldsymbol{P x}_i\|^2$$

2.4PCA算法总结

输入：数据$\boldsymbol{X}$，降维维数d，数据个数n

输出：$\boldsymbol{Y}$，$\boldsymbol{W}_{PCA}$



1. 求解X的均值$m = \frac{1}{n}\sum_{i=1}^{n} x_i$

2. 求解矩阵$\boldsymbol{C} = \frac{1}{n}(\boldsymbol{X} - m)(\boldsymbol{X} - m)^T$

3. 对矩阵$\boldsymbol{C}$进行特征分解$\boldsymbol{C} = \boldsymbol{W\Lambda W}^T$，$\boldsymbol{W} = [\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_m]$，

   $\boldsymbol{\Lambda} = diag(\lambda_1, \lambda_2, \cdots, \lambda_m)$，$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$

4. 令$\boldsymbol{W}_{PCA} = [\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_d]$，即为我们要求的变换矩阵

5. $\boldsymbol{Y} = \boldsymbol{W}_{PCA}{}^T\boldsymbol{X}$，$\boldsymbol{Y} \in \mathcal{R}^{d\times n}$，$d < m$

3 主成分个数的选择

上述介绍了主成分的计算方式，但是在实际问题中，我们还应思考应该选择多少个主成分作为最终的代表结果

在经典的PCA算法中，样本协方差矩阵的前k个特征值所对应的特征向量是特征空间的正交基。一般用方差贡献率$R_i = \frac{\lambda_i}{\sum_{j=1}^{m} \lambda_j}\ (i = 1,2,\cdots,k)$来解释第$i$个主成分所反映的信息量。我们选择的前$k$个主成分所构成的累积贡献率通常应该在80%~90%之间。其中，前k个主成分所构成的累积贡献率为：

$$G(k) = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{j=1}^{m} \lambda_j}$$

4 高维数据下的PCA

在主成分的计算中，需要对协方差矩阵进行特征值分解。对于数据$\boldsymbol{X}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n]\in\mathcal{R}^{m\times n}$，如果数据点的数量远远小于数据空间的尺寸，即$n\ll m$，则协方差矩阵$\boldsymbol{C}=\frac{1}{n}\boldsymbol{X}\boldsymbol{X}^T\in\mathcal{R}^{m\times m}$ 的维数就会变得很大。当m特别大时，C的特征分解在计算上就变得不太可行。

通常的解决方式：通过对$\frac{1}{n}\boldsymbol{X}^T\boldsymbol{X}\in\mathcal{R}^{n\times n}$进行特征值分解来得到$\frac{1}{n}\boldsymbol{X}\boldsymbol{X}^T$的非零特征值和对应的特征向量

下面先来看看$\frac{1}{n}\boldsymbol{X}^T\boldsymbol{X}$和$\frac{1}{n}\boldsymbol{X}\boldsymbol{X}^T$特征值和特征向量的关系

4 高维数据下的PCA

若λ和$\boldsymbol{u}$是$\frac{1}{n}\boldsymbol{X}\boldsymbol{X}^T$的非零特征根和对应的特征向量，则有： $\frac{1}{n}\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{u}=\lambda\boldsymbol{u}$

等式两边左乘$\boldsymbol{X}^T$： $\frac{1}{n}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{u}=\boldsymbol{X}^T\lambda\boldsymbol{u}$

即： $\left(\frac{1}{n}\boldsymbol{X}^T\boldsymbol{X}\right)(\boldsymbol{X}^T\boldsymbol{u})=\lambda(\boldsymbol{X}^T\boldsymbol{u})$

从中我们可以看到$\boldsymbol{X}^T\boldsymbol{u}$是$\frac{1}{n}\boldsymbol{X}^T\boldsymbol{X}$的一个特征向量，对应的特征值为$\lambda$。因此理论上对于高维数据，在对$\frac{1}{n}\boldsymbol{X}\boldsymbol{X}^T$进行特征分解时，可以转化成对$\frac{1}{n}\boldsymbol{X}^T\boldsymbol{X}$进行特征分解

4 高维数据下的PCA

若λ 和$\boldsymbol{v}$是$\frac{1}{n}\boldsymbol{X}^T\boldsymbol{X}$的非零特征根和对应的特征向量，根据前页PPT所示，则有：

$$\left(\frac{1}{n}\boldsymbol{X}\boldsymbol{X}^T\right)(\boldsymbol{X}\boldsymbol{v}) = \lambda(\boldsymbol{X}\boldsymbol{v})$$

但此时的$\boldsymbol{X}\boldsymbol{v}$的长度未必等于1。为了确定合适的归一化，令$\boldsymbol{u} = \alpha\boldsymbol{X}\boldsymbol{v}$，其中$\alpha$ 是使$\boldsymbol{X}\boldsymbol{v}$进行归一化的常数。则有：

$$\begin{aligned}\|\boldsymbol{u}\|^2 &= \boldsymbol{u}^T\boldsymbol{u} = \alpha^2\boldsymbol{v}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{v} = n\alpha^2\boldsymbol{v}^T\frac{1}{n}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{v} \\ &= n\alpha^2\boldsymbol{v}^T\left(\frac{1}{n}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{v}\right) = n\alpha^2\boldsymbol{v}^T(\boldsymbol{\lambda}\boldsymbol{v}) \\ &= \boldsymbol{\lambda}\alpha^2\boldsymbol{v}^T\boldsymbol{v} = n\boldsymbol{\lambda}\alpha^2 = 1\end{aligned}$$

$$n\boldsymbol{\lambda}\alpha^2 = n\boldsymbol{\lambda}(\boldsymbol{u}/\boldsymbol{X}\boldsymbol{v})^2 = 1$$

$$\boldsymbol{u} = \left(\frac{1}{n\lambda}\right)^{1/2}\boldsymbol{X}\boldsymbol{v}$$

4 高维数据下的PCA

结论：

对于数据$\boldsymbol{X}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n]\in\mathcal{R}^{m\times n}$，若$n\ll m$， 会造成$\frac{1}{n}\boldsymbol{X}\boldsymbol{X}^T\in\mathcal{R}^{m\times m}$维数过高，则可先计算$\frac{1}{n}\boldsymbol{X}^T\boldsymbol{X}\in\mathcal{R}^{n\times n}$的特征值和对应的特征向量。

若λ 和$\boldsymbol{v}$是$\frac{1}{n}\boldsymbol{X}^T\boldsymbol{X}$的非零特征根和对应的特征向量，则$\lambda$ 和$\left(\frac{1}{n\lambda}\right)^{1/2}\boldsymbol{X}\boldsymbol{v}$ 是$\frac{1}{n}\boldsymbol{X}\boldsymbol{X}^T$的非零特征根和对应的特征向量。

5 PCA与SVD的关系

奇异值分解(SVD)和主成分分析(PCA)是将高维数据点转化为低维特征的两种特征值方法。很明显，PCA和SVD是密切相关的。

在PCA中，数据样本需要满足$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i=\boldsymbol{0}$。样本协方差矩阵$\boldsymbol{C}=\frac{1}{n}\boldsymbol{X}\boldsymbol{X}^T$可以对角化为$\boldsymbol{C}=\boldsymbol{W}\boldsymbol{\Lambda}\boldsymbol{W}^T$。其中$\boldsymbol{W}$是特征向量矩阵（$\boldsymbol{W}=[\boldsymbol{w}_1,\boldsymbol{w}_2,\cdots,\boldsymbol{w}_m]$，$\boldsymbol{w}_i$是一个特征向量），$\boldsymbol{\Lambda}$是一个对角矩阵（$\boldsymbol{\Lambda}=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_m)$，$\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_m$）。

在SVD中，数据矩阵$\boldsymbol{X}$可以分解为$\boldsymbol{X}=\boldsymbol{P}\boldsymbol{\Sigma}\boldsymbol{Q}^T$。其中$\boldsymbol{P}$是一个标准正交矩阵，$\boldsymbol{\Sigma}$是奇异值$\sigma_i$的对角矩阵。

5 PCA与SVD的关系

在SVD中，如果数据$\boldsymbol{X}$为归一化的数据，即$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i = 0$，有

$$\boldsymbol{C} = \frac{1}{n}\boldsymbol{X}\boldsymbol{X}^T = \frac{\mathbf{1}}{\boldsymbol{n}}\boldsymbol{P}\boldsymbol{\Sigma}\boldsymbol{Q}^T\boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{P}^T = \boldsymbol{P}\frac{\boldsymbol{\Sigma}^2}{n}\boldsymbol{P}^T$$

对比式$\boldsymbol{C} = \boldsymbol{W}\boldsymbol{\Lambda}\boldsymbol{W}^T$，我们很容易发现在$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i = 0$前提下，SVD中的标准正交矩阵$\boldsymbol{P}$可以被认为是PCA中的特征向量$\boldsymbol{W}$的矩阵。并且，奇异值与样本协方差矩阵的特征值通过$\lambda_i = \frac{{\sigma_i}^2}{n}$联系起来。

5 PCA与SVD的关系

结论:

如果数据矩阵$\boldsymbol{X}$满足$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i = 0$(这是PCA算法所要求的，但SVD中不要求)， PCA 与SVD 的计算可以进行转化。主要的区别是SVD不计算样本协方差矩阵。而PCA则需要计算样本协方差矩阵，然后进行特征分解。

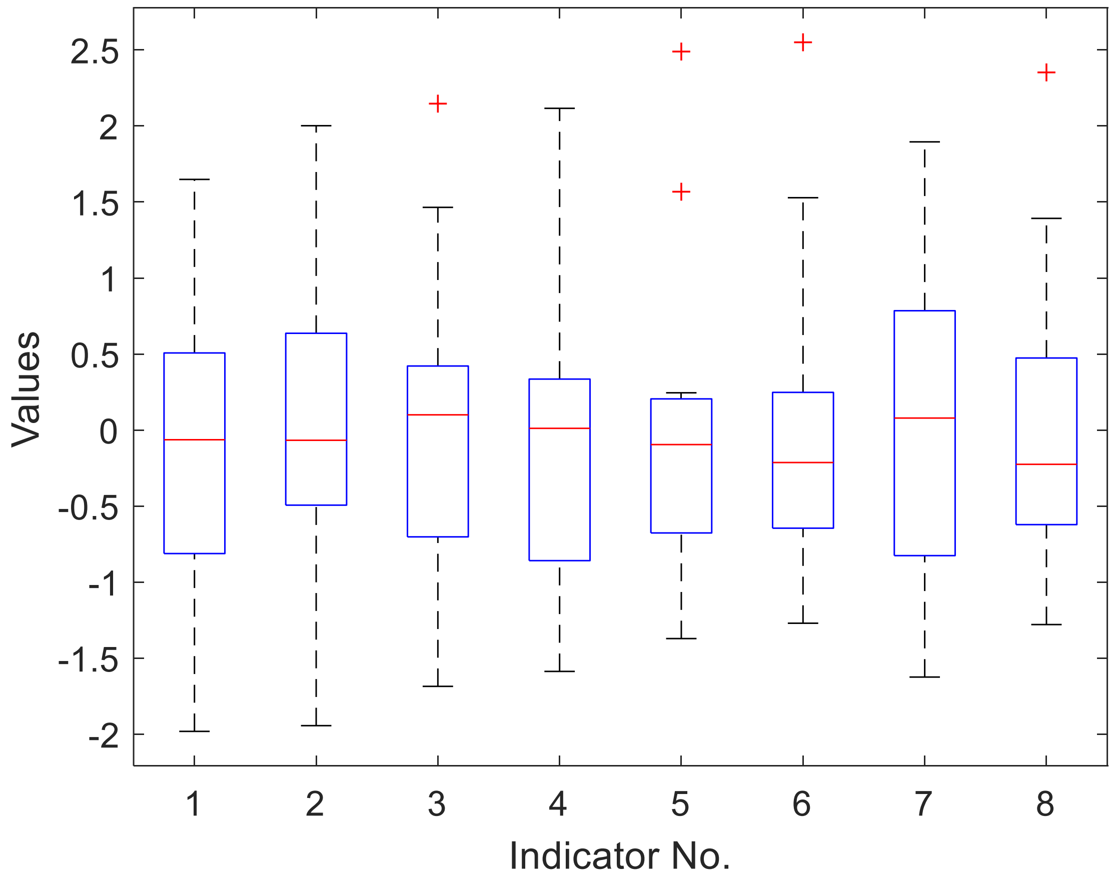
6 主成分分析案例分析

目的：根据8个不同的指标，分析比较14家公司的综合实力

序号	净利润率	固定资产利润率	毛利率	销售收入利润率	产品成本利润率	实物利润率	人均利润率	流动资产利润率
1	20.4	7.4	7.2	6.1	8.3	8.7	2.442	20
2	25	12.7	11.2	11	12.9	10.2	3.542	9.1
3	13.2	3.3	3.9	4.3	4.4	5.5	0.578	3.6
4	22.3	6.7	5.6	3.7	6	7.4	0.176	7.3
5	14.3	11.8	7.1	7.1	8	8.9	1.726	7.5
6	35.6	12.5	16.4	16.7	22.8	29.3	3.017	26.6
7	22	7.8	9.9	10.2	12.6	17.6	0.847	10.6
8	28.4	13.4	10.9	9.9	10.9	13.9	1.772	17.8
9	40.6	19.1	19.8	19	29.7	39.6	2.449	35.8
10	24.8	8	9.8	8.9	11.9	16.2	0.789	13.7
11	12.5	9.7	4.2	4.2	4.6	6.5	0.874	3.9
12	1.8	0.6	0.7	0.7	0.8	1.1	0.056	1
13	22.3	13.9	9.4	8.3	9.8	13.3	2.126	17.1
14	38.5	9.1	11.3	9.5	12.2	16.4	1.327	11.6

6 主成分分析案例分析

我们先利用箱线图探究这八个指标之间的关系。可以发现每个公司的人均利润率分布变化最大，产品成本利润率的差异最小。不同指标的变化范围不同。因此如果使用PCA 需要对数据进行归一化处理



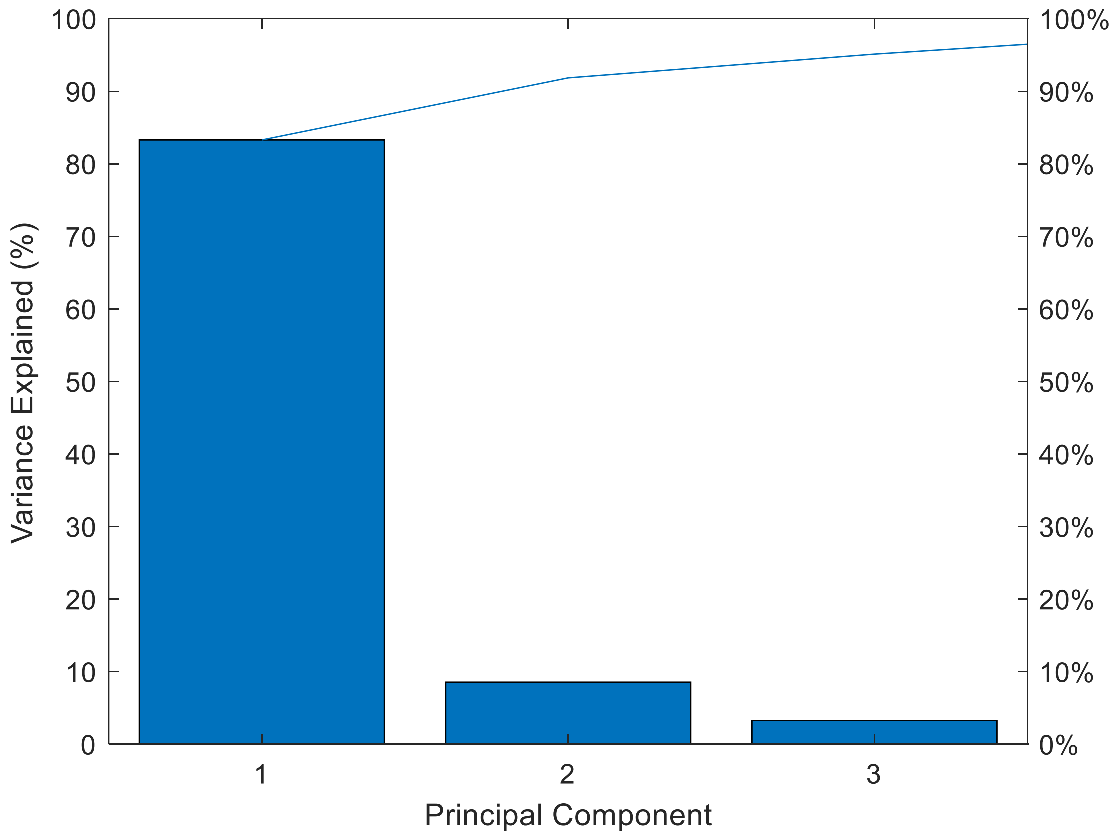
6 主成分分析案例分析

表中给出了数据之间的相关矩阵，从中可以发现数据之间的相关性特别高。毛利率与销售收入利润率的相关性甚至达到了0.98。说明8个指标之间存在较高的相关性和冗余度。因此，利用主成分分析(PCA)将原始数据投影到主方向上以表示原始数据是比较合适的。

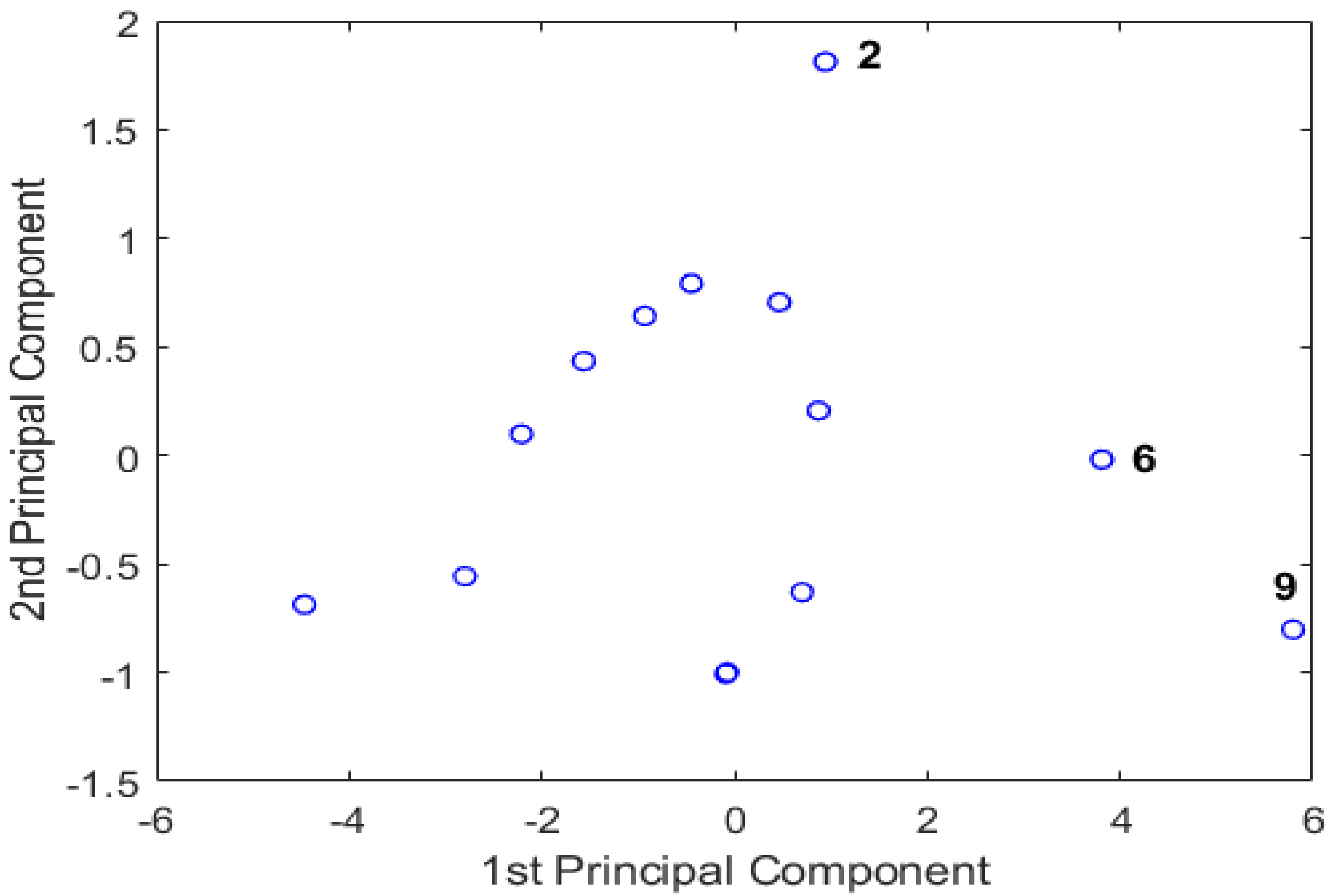
1.00	0.70	0.88	0.83	0.83	0.80	0.50	0.76
0.70	1.00	0.81	0.76	0.77	0.68	0.63	0.72
0.88	0.81	1.00	0.84	0.98	0.78	0.66	0.76
0.83	0.76	0.84	1.00	0.87	0.96	0.41	0.88
0.83	0.77	0.98	0.87	1.00	0.85	0.59	0.80
0.80	0.68	0.78	0.96	0.85	1.00	0.23	0.91
0.50	0.63	0.66	0.41	0.59	0.23	1.00	0.40
0.76	0.72	0.76	0.88	0.80	0.91	0.40	1.00

6 主成分分析案例分析

我们标准化了X的每一个变量。然后利用PCA算法得到主成分。我们可以得到每个分量解释方差的百分比，如图所示，图中表明了前三个分量可以解释了总方差的95%，即利用PCA得到的前三个分量可以代表原八个变量95%的内容。我们注意到通过PCA计算得到的前两个主成分分量可以解释总方差的90%以上。因此我们可以将这8个指标简化为两个特征。



6 主成分分析案例分析

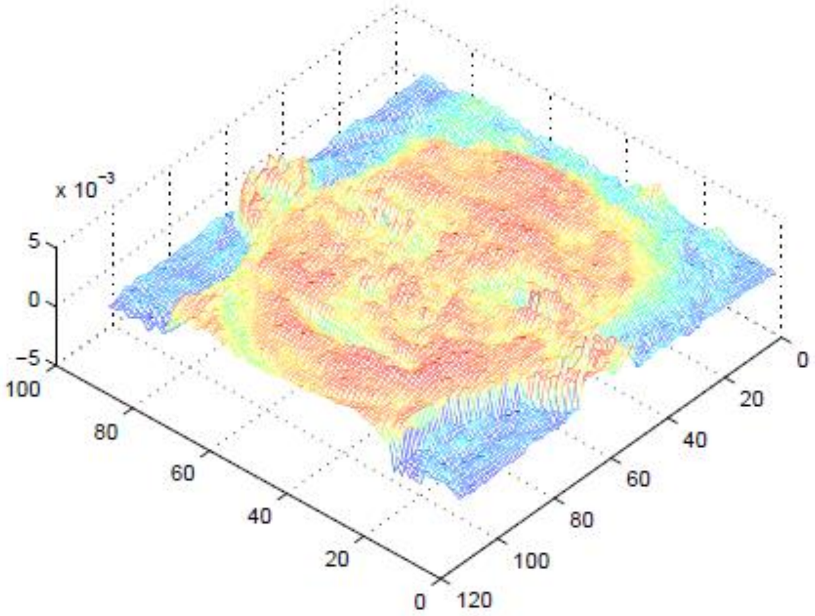

2
1.5
1
0.5
0
-0.5
-1
-1.5
-6
-4
-2
0
2
4
6
2nd Principal Component
1st Principal Component
2
6
9

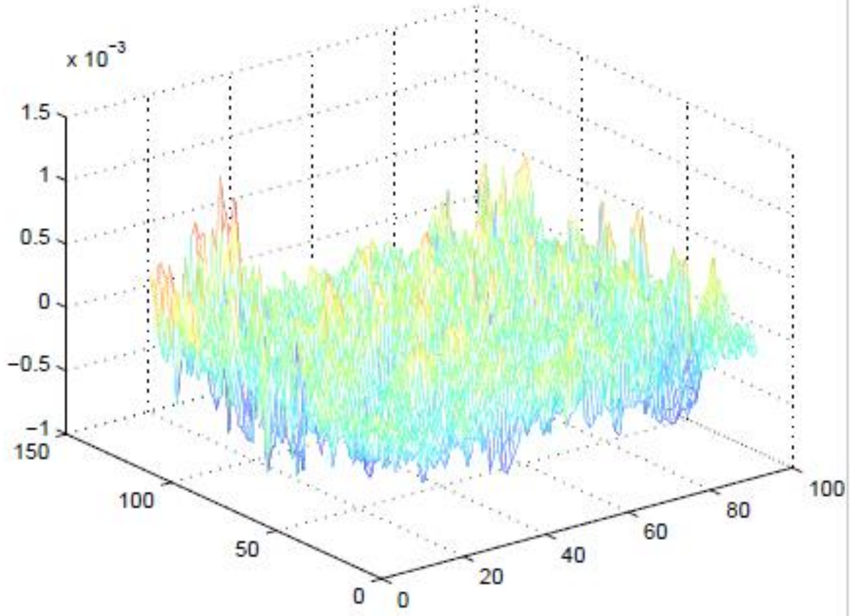
7 Face Recognition Technology

-Eigenfaces

7 Eigenfaces

 $\approx 0.1439\times$  $+0.1216\times$  $-0.0375\times$ 





谢谢各位同学！