主元分析

授课教师:　赵春晖

联系方式:

Email: chhzhao@zju.edu.cn

Phone: 13588312064

Room: 工控新楼308室

内容

1. 主元分析简介
2. 数学模型与几何意义
3. 算法求解与推导
4. 主元分析的应用

§0 绪论

- 希望从数据中验证某种推断或提炼某种特征：矛盾的统一体
- **一方面**人们为了避免遗漏重要的信息而考虑尽可能多的指标
- **另一方面**随着考虑指标的增多增加了问题的复杂性，同时由于各指标均是对同一事物的反映，不可避免地造成信息的大量重叠，这种信息的重叠有时甚至会抹杀事物的真正特征与内在规律。人们希望从中能抓住主要信息

§0 绪论

- 在地区或企业经济效益的评价中，涉及的指标往往很多。如给定30个地区的经济发展8项指标： GDP，居民消费水平、固定资产投资、职工平均工资、货物周转量、居民消费价格指数、商品零售价格指数、工业总产值。如何研究经济发展状况和地区差异？
- 大的化学和药品公司生产过程要测量100多个过程变量，包括不同场合下的温度、压力及重量等。如何形象化显示重要变量又能够灵敏检测变异的发生？

§ 0 绪论

- 变量太多增加问题的复杂性，也给合理分析问题和解决问题带来困难；虽然每个变量都提供了一定的信息，但其重要性有所不同，在很多情况下，变量间有一定的相关性，从而使得这些变量所提供的信息有一定的重叠。

所有这些应用背景归结为：

- 研究中经常会遇到多指标的问题，这些指标间往往存在一定的相关，直接纳入分析不仅复杂，变量间难以取舍，而且可能因多元共线性而无法得出正确结论。
- 问题实质均为数据化简、信息浓缩或者说降维，即将分散在多个变量中的同类信息集中、提纯，从而便于分析、解释和利用。

§ 0 绪论

高维数据如何降维且尽可能少损失信息？

- 降维最简单的方法就是保留一个变量，舍弃其余的变量；
- 对所有变量平均加权；

  ---除非所有变量具有同样方程，否则不合理
- 基于某种标准做加权平均；

  ---何种标准？

主成分分析（PCA）是解决这些问题一种有效途径

---主成分分析的目的就是通过线性变换，将原来的多个指标组合成相互独立的少数几个能充分反映总体信息的指标，便于进一步分析尽可能保留原始变量的信息，且彼此不相关。

先人的智慧

3、方法论要求:

射人先射马，擒贼先擒王

牵牛要牵牛鼻子

打蛇要打七寸

善于抓重点，集中力量解决主要矛盾

“成大事者
以识为主
以才为辅”

—— 曾国藩

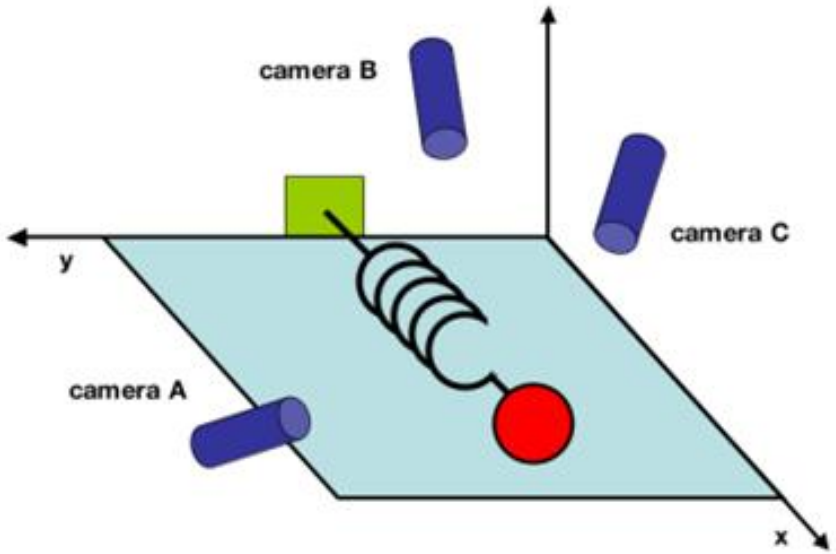
想一想

哲学感悟

做任何事情一要善于抓重点，集中力量解决主要矛盾。二要统筹兼顾，恰当处理次要矛盾

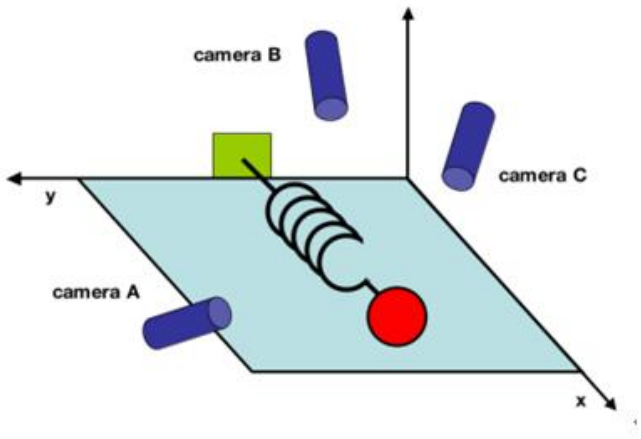
§1 主元分析简介——引例

- 下面的模型看上去比较简单，但足以说明问题。这是一个理想弹簧运动规律的测定实验，红球连接在弹簧之上，从平衡位置沿x轴拉开一定的距离然后释放。



- 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轴方向上发生，只需要记录下x轴方向上的运动序列并加以分析即可。

§1 主元分析简介——引例



◆ 但是：对于第一次实验的探索者来说，是不可能进行这样的假设的。因此在三个角度摆放了三台摄像机，三台摄像机之间并不正交。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，将三台摄像机记录的坐标合并可得到：$[(x_A, y_A), (x_B, y_B), (x_C, y_C)]$

经过实验，系统产生了几分钟内球的12000个位置序列，组成（12000*6）的矩阵。

●**问题：怎样从这些数据中得到球是沿着某个轴运动的规律呢？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？**

§1 主元分析简介

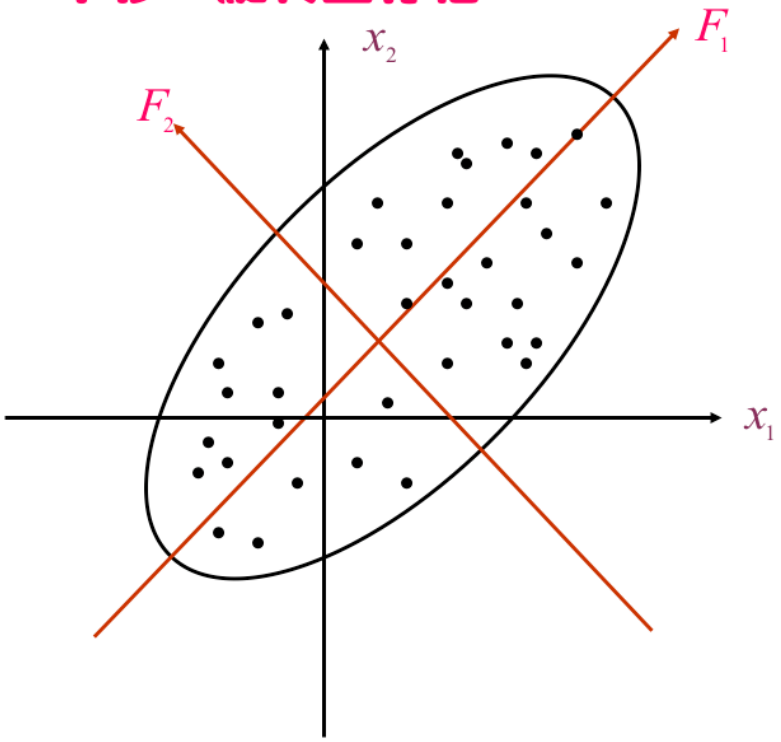
◆PCA就是解决此类问题的有力武器。在这个有关弹簧的例子中，PCA会将观测变量经过一系列的投影变换，帮我们寻找到有意义的变量仅仅是x轴变量而已。而这个x轴变量就是所谓的“主元”。

§2 数学模型与几何意义

为了方便，我们在二维空间中讨论主成分的几何意义。 设有n个样品，每个样品有两个观测变量x_1和x_2，在由变量x_1和x_2 所确定的二维平面中，n个样本点所散布的情况如椭圆状。

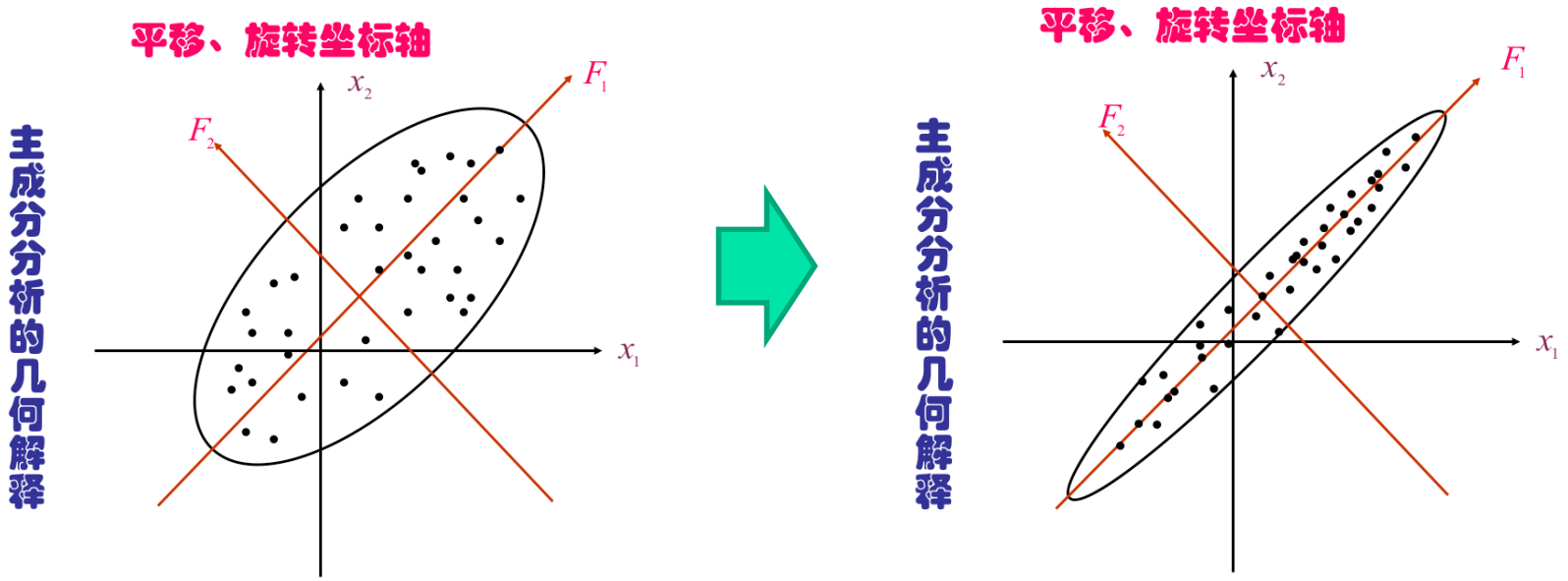
主成分分析的几何解释

平移、旋转坐标轴



§2 数学模型与几何意义

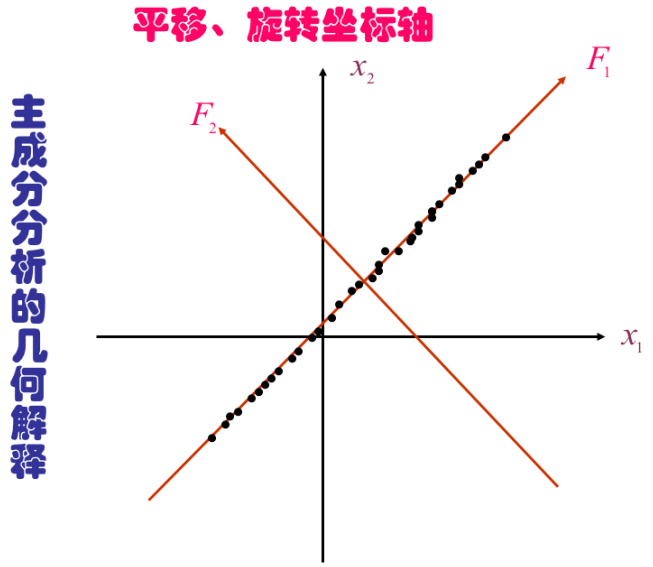
如果我们将xl 轴和x2轴先平移，再同时按逆时针方向旋转，得到新坐标轴Fl和F2。Fl和F2是两个新变量。经过这样的旋转变换原始数据的大部分信息集中到F_l轴上，对数据中包含的信息起到了浓缩作用。F_l，F_2具有不相关的性质，这就使得在研究复杂的问题时避免了信息重叠所带来的虚假性。



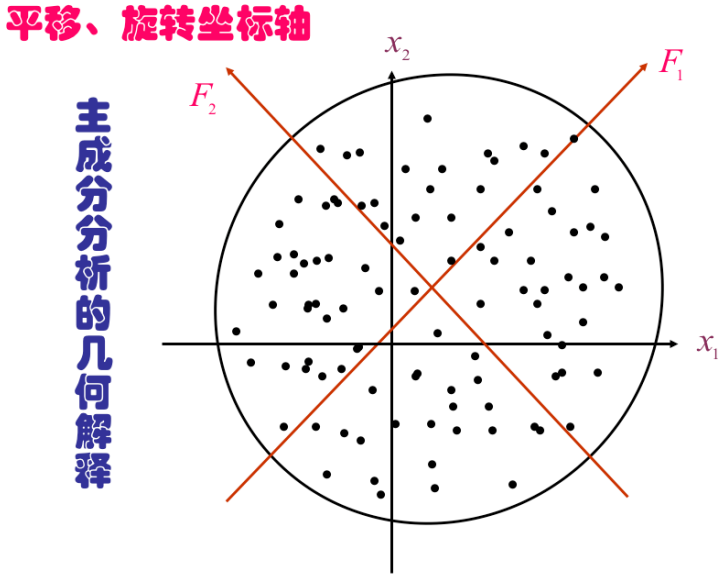
§2 数学模型与几何意义

- 考虑两种极端的情形：

1.椭圆扁平到了极限，变成y_1轴上的一条线



2. 椭圆的长轴与短轴的长度相等，即椭圆变成圆



§2 数学模型与几何意义

两种极端的情形：

1.椭圆扁平到了极限，变成y_1轴上的一条线

第一主成分包含有二维空间点的全部信息，仅用这一个综合变量代替原始数据不会有任何的信息损失
原始变量X_1和X_2信息完全重叠

2.椭圆的长轴与短轴的长度相等，即椭圆变成圆

第一主成分只含有二维空间点的约一半信息，若仅用这一个综合变量，则将损失约50％的信息，这显然是不可取的。

原始变量X_1和X_2的相关程度几乎为零，也就是说，它们所包含的信息几乎不重迭，因此无法用一个一维的综合变量来代替

§2 数学模型与几何意义

这种由讨论多个指标降为少数几个综合指标的过程在数学上就叫做降维。主成分分析通常的做法是，寻求原指标的线性组合F_i？要求？

$$F_1 = u_{11}X_1 + u_{21}X_2 + \cdots + u_{p1}X_p$$

$$F_2 = u_{12}X_1 + u_{22}X_2 + \cdots + u_{p2}X_p$$

$$\cdots\cdots$$

$$F_p = u_{1p}X_1 + u_{2p}X_2 + \cdots + u_{pp}X_p$$

§2 数学模型与几何意义

主成分分析满足如下的条件：

1. 每个主成分的系数平方和为1。即

$$u_{1i}^2 + u_{2i}^2 + \cdots + u_{pi}^2 = 1$$

2.主成分之间相互独立，即无重叠的信息。即

$$Cov(F_i, F_j) = 0, \quad i \neq j, \quad i, j = 1, 2, \cdots, p$$

3.主成分的方差依次递减，重要性依次递减，即

$$Var(F_1) \geq Var(F_2) \geq \cdots \geq Var(F_p)$$

- 统计学中，方差越大，包含的信息量越大，降维后的主成分向量应包含尽可能大的方差。
- 当主成分变量个数大于 1 个时，为了避免信息重叠（相关），要求各主成分向量之间不相关。

§3 算法求解与推导

我们现在以广泛用于机器学习的葡萄酒数据集为例，这份数据集包含来自3种不同起源的葡萄酒的共178条记录。13个属性是葡萄酒的13种化学成分，如下表所示。

1.Alcohol	8.Nonflavanoid phenols
2.Malic acid	9.Proanthocyanins
3.Ash	10.Color intensity
4.Alcalinity of ash	11.Hue
5.Magnesium	12.OD280/OD315 of diluted wines
6.Total phenols	13.Proline
7.Flavanoids	

葡萄酒数据集的13个指标

该问题中有p个指标(p=13)，我们把这13个指标看作13个随机变量，记为X_1，X_2，…，X_{13}，主成分分析就是要把这13个指标的问题，转变为讨论13个指标的线性组合的问题，而这些新的指标F_1，F_2，…，F_k($k \leqslant p$)，按照保留主要信息量的原则充分反映原指标的信息，并且相互独立。

§3 算法求解与推导

- 这份数据集包含来自3种不同起源的葡萄酒的共178条记录。13个属性是葡萄酒的13种化学成分。

- 记为原始数据矩阵X（m*n），m为样本数，本例中为178；n为特征数，本例中为13。下面将以该数据为例，介绍PCA是如何完成数据降维的。

§3 算法求解与推导

- 首先对数据进行标准化处理，标准化处理的好处会在后面看到。

$$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad ；\ i=1,2,\ldots,m；\ j=1,2,\ldots,n$$

$$\bar{x}_j = \frac{1}{m}\sum_{i=1}^{m} x_{ij} \qquad s_j = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(x_{ij} - \bar{v}_j\right)^2}$$

➢标准化处理后的X的每个变量（每一列）的均值为0，方差为1。

- PCA的目标是寻找负载矩阵 P 使得 T ＝XU，T为降维后的矩阵，称为主元矩阵。要求：降维后的**各个特征向量（ T 的列向量）要满足相互正交（线性无关）且尽可能多地表示原始数据信息**。我们要做的工作就是寻找这个满足要求的负载矩阵U。

§3 算法求解与推导

方差和协方差

- 统计学中，方差越大，包含的信息量越大，因此我们希望降维后的主元矩阵 T 中的主成分向量包含尽可能大的方差。
- 当主成分变量个数大于 1 个时，为了避免信息重叠（相关），要求各主成分向量之间不相关。
- 线性代数的知识给出协方差为 0 的两个变量是不相关的。因此我们的求解问题就转化为了寻找矩阵U，使得主成分矩阵 T 的各个列向量方差尽可能大，且协方差为 0 。

方差 $$D(\mathrm{X}) = \mathrm{E}\left\{\left[X - E(\mathrm{X})\right]\right\}^2$$

协方差 $$Cov(\mathrm{X},\mathrm{Y}) = E[\mathrm{XY}] - \mathrm{E}[\mathrm{X}]\mathrm{E}[\mathrm{Y}]$$

§3 算法求解与推导

协方差矩阵

- 记主元矩阵 T 的协方差矩阵为 D，$D = T^T T / (m-1)$，由于已经对原始数据X做了标准化处理，容易发现D的对角线元素恰好是每个主成分变量的方差，而其余位置的元素D_{ij}则为第 i 个变量与第 j 个变量的协方差。
- 显然，我们的目标变得很清晰：求解U矩阵，使得 T 的协方差矩阵为一个对角矩阵，且对角线元素按从大到小排列。

存在性

线性代数的结论

若A是p阶实对称阵，则一定可以找到正交阵U，使

$$\mathbf{U}^{-1}\mathbf{A}\mathbf{U}=\begin{bmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \lambda_p\end{bmatrix}_{p\times p}$$

其中$\lambda_i, i=1.2.\cdots p$ 是A的特征根。

- 推导过程

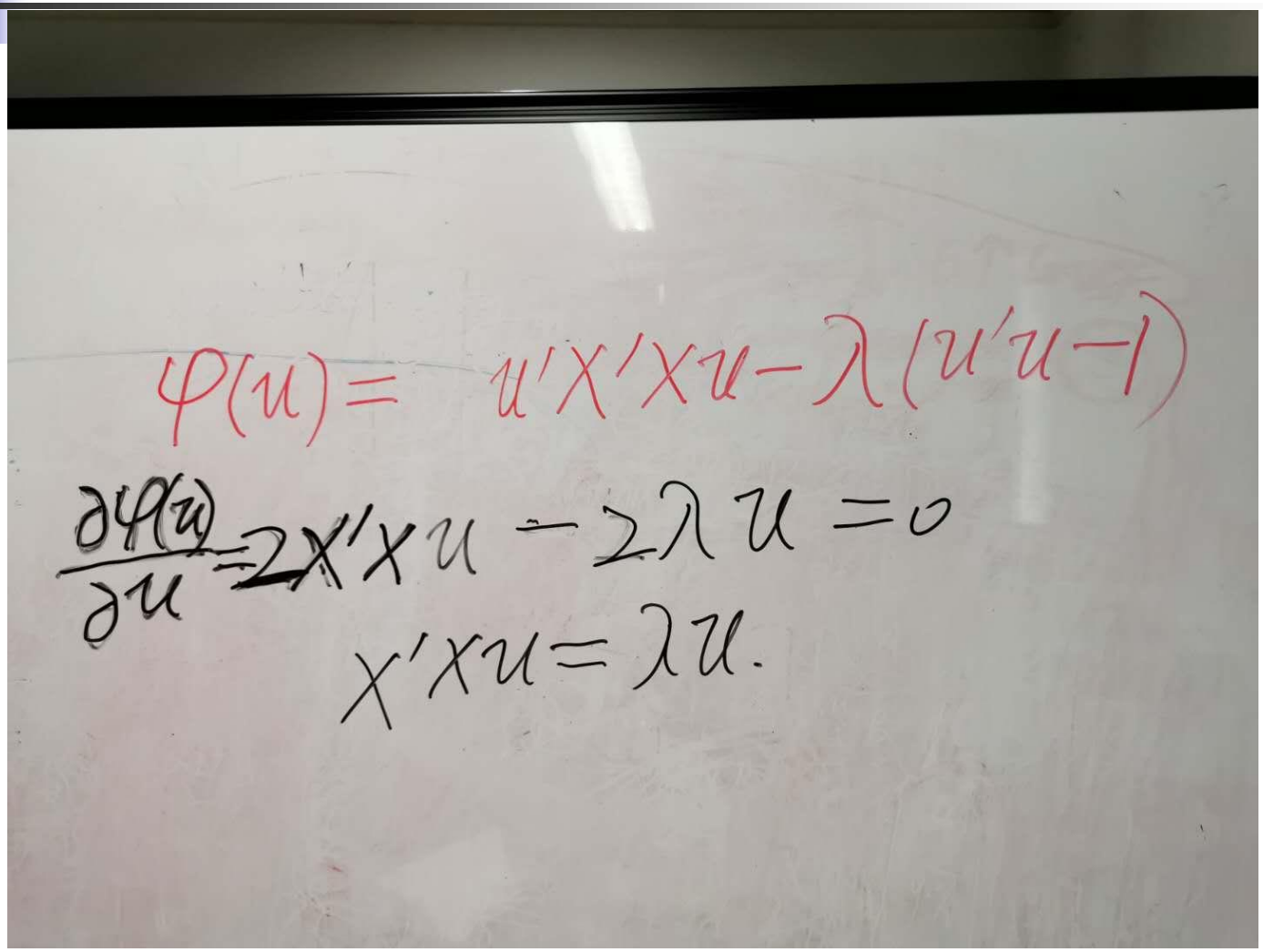
目标函数 $t=Xu$; $\max(t'*t)=\max(u'*X'*Xu)$

约束条件： $u'*u=1$

我们在利用样本数据求解主成分的过程实际上就转化为求相关阵或协方差阵的特征值和特征向量的过程。

$$(X'*X)u=\lambda u$$

拉格朗日乘数法


$\varphi(u) = u'X'Xu - \lambda(u'u-1)$
$\frac{\partial \varphi(u)}{\partial u} = 2X'Xu - 2\lambda u = 0$
$X'Xu = \lambda u.$

§3 算法求解与推导

- 求解U矩阵
- 负载矩阵U的求解转化成了对原始数据的协方差矩阵S的特征值分解，S的特征值对应的特征向量即为U的列向量。

$$Su=\lambda u$$

$$SU=U\Lambda$$

$$X=TU'$$

特征根分解得到的特征向量都是互相垂直的，得到的主元也都是互相垂直的，恰好满足了之前提到的要求。

算法小结

- 主成分分析(Principal Component Analysis， PCA)

解法：采用特征根分解 **$(X'*X)u=\lambda u$**

几个重要关系公式：

$T=XU$

$T'*T=\Lambda$; $U'*U=I$

$X=T*U'+E$

利用的几个定理：

- 若A是p阶实对称阵，则一定可以找到正交阵U，使其转换为对角矩阵$U'AU=\Lambda$。
- 实对称矩阵对应不同特征值的特征向量互相垂直。

§3 算法求解与推导

- 主元个数的选取

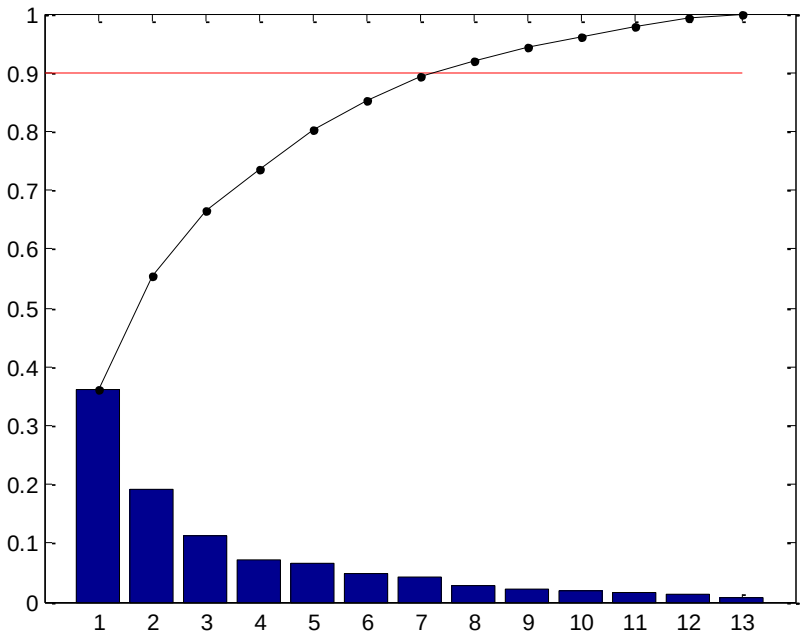
➢ 矩阵X的协方差矩阵S的特征值按从大到小排列后，数值越大的特征值所对应的方差越大，主元向量所包含的信息越多。

➢ 累积方差百分比（Cumulative Percent Variance，简称CPV）是一种通过计算前k个主元的累积贡献率来确定主元个数的方法。

$$CPV_k = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{n} \lambda_n}$$

§3 算法求解与推导

- 主元个数的选取（葡萄酒数据集）



- 由图可知，用累积方差贡献率的方法，当解释率阈值设为90%时，求得主元个数k为8，将葡萄酒数据集数据集从13维降到了8维，后续可以利用降维后的数据对3种酒进行分类。

$$T = XU$$

(m*k) (m*n) (n*k)

本例中m=178，n=13，k=8

§4 PCA应用（人脸识别）

2015年3月15日汉诺威IT博览会(CeBIT)在德国开幕，阿里巴巴创始人马云作为唯一受邀的企业家代表，在会上发表了演讲，并为德国总理默克尔与中国副总理马凯演示了蚂蚁金服的Smile to Pay扫脸技术，并当场刷自己的脸给嘉宾买礼物。马云选择的礼物是淘宝网上一枚1948年的汉诺威纪念邮票。

- 人脸识别的方法有很多，基于PCA的人脸识别是很重要的一种方法。

§4 PCA应用（人脸识别）

- Step 1 创建人脸数据库

(a)

(b)

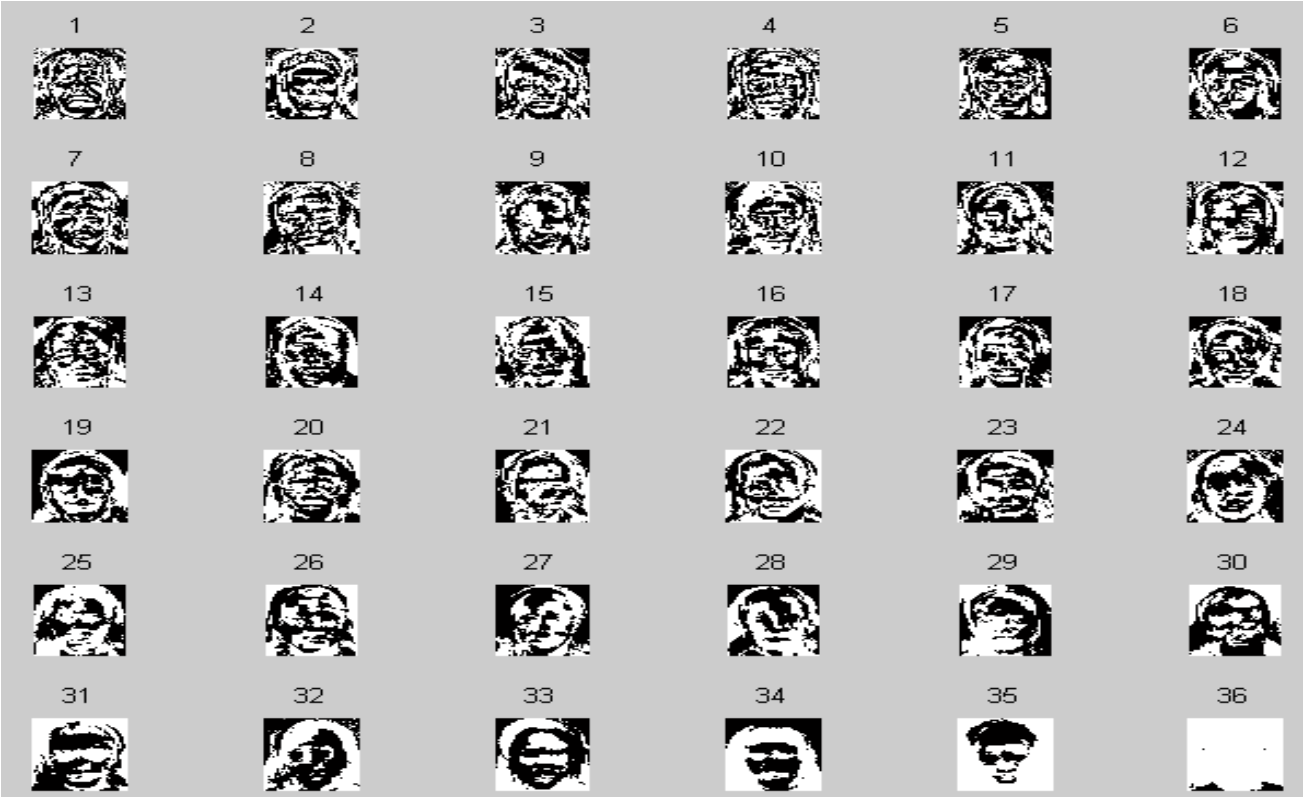
图(a)为系统人脸数据库中的原始人脸图像，图(b)为经过灰度转换后的人脸图像

读入系统人脸数据库，将图像变换为相应的二维人脸灰度图像。将变换后的二维人脸灰度图像变换为一维人脸向量矩阵，一个大小为M*N的二维人脸图像可以看成长度为MN的人脸图像向量。人脸图像通常维数很高，如128*128的矩阵可以展开为16384维的向量。

§4 PCA应用（人脸识别）

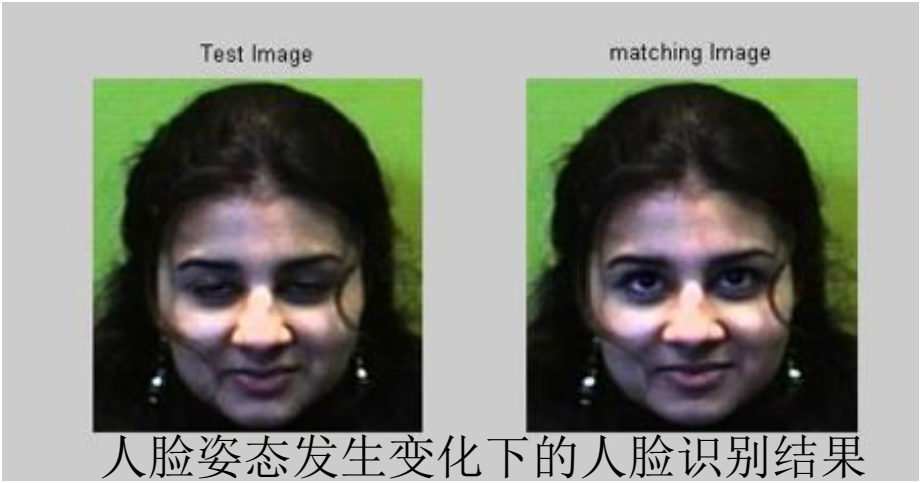
- Step 2 计算特征脸，形成特征库。如此高的向量维度，各维度间相关性很强。这时PCA就派上了用场，使用PCA对原始向量进行特征提取，大大降低了向量维度。特征向量构成特征脸。

特征脸（特征值最大的前36个特征向量）

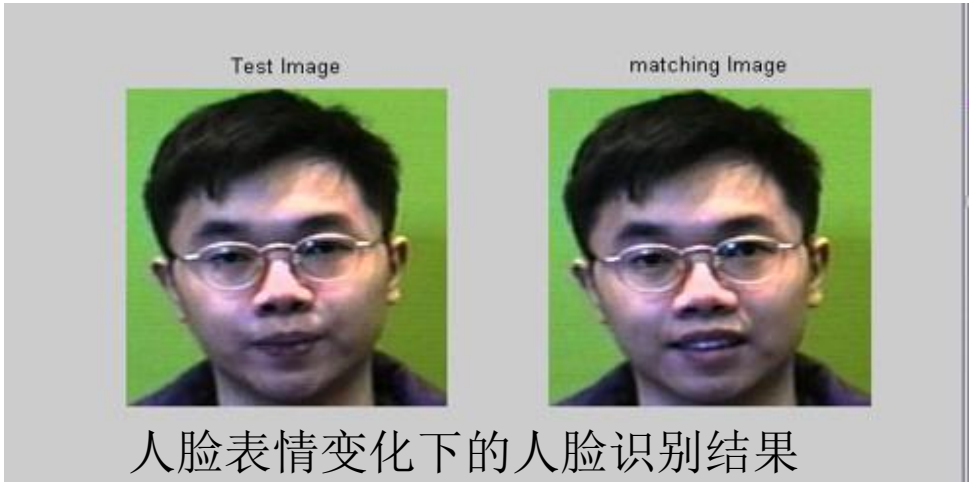


§4 PCA应用（人脸识别）

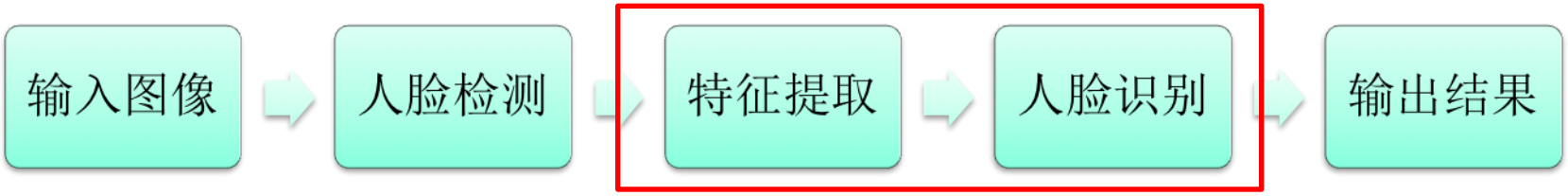
- Step 3 对人脸进行识别。对测试的人脸图像，按上述方法转化为特征向量，与特征库内的特征向量进行比对，使用欧氏距离等方法进行判决分类。



人脸姿态发生变化下的人脸识别结果

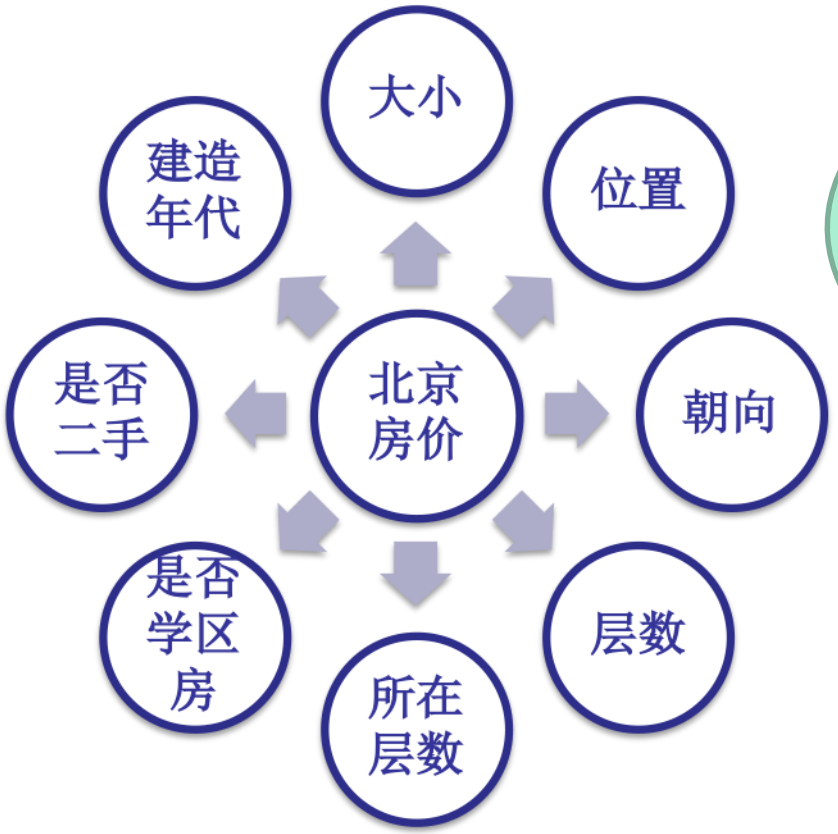


人脸表情变化下的人脸识别结果



§4 PCA应用（房价）

- 拿到一个样本，特征非常多，而样例特别少，这样用回归去直接拟合非常困难，容易过度拟合。



- 比如北京的房价：假设房子的特征如左图所示。这么多特征，结果只有不到十个房子的样例。要拟合房子特征->房价。这么多特征就会造成过度拟合。PCA可以解决这个问题。

§4 PCA应用（法官评分）

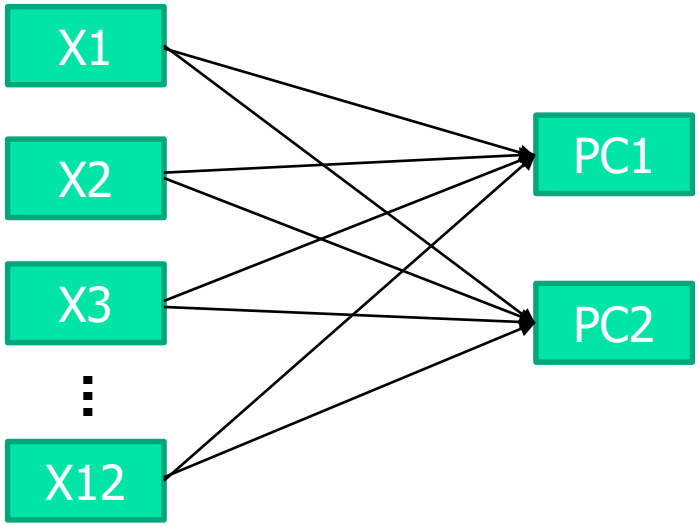
- 数据集USJudgeRatings包含了律师对美国高等法院法官的评分。数据集包含43个样本，12个变量：

变　量	描　述	变　量	描　述
CONT	律师与法官的接触次数	PREP	审理前的准备工作
INTG	法官正直程度	FAMI	对法律的熟稔程度
DMNR	风度	ORAL	口头裁决的可靠度
DILG	勤勉度	WRIT	书面裁决的可靠度
CFMG	案例流程管理水平	PHYS	体能
DECI	决策效率	RTEN	是否值得保留

- 那么问题来了：是否能够用较少的变量来总结这12个变量评估的信息呢？如果可以，需要多少个？

§4 PCA应用（法官评分）

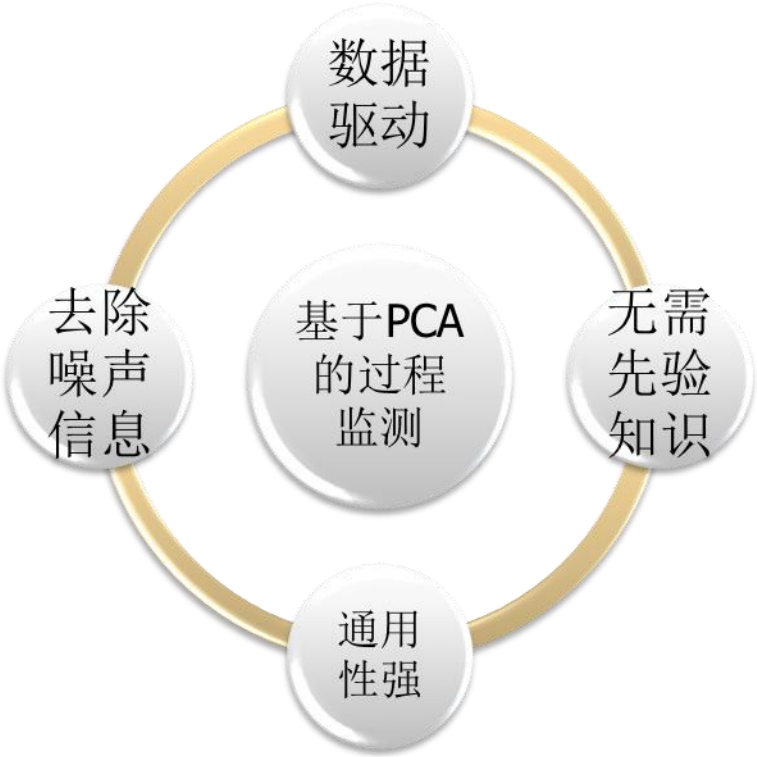
用PCA来解决该问题



主成分分析模型，变量（X1到X12）映射为主成分（PC1,PC2）

- PC1(84.4%)和PC2(9.2%)共可以解释这12个变量的93.6%的程度

§4 PCA应用（过程监测）

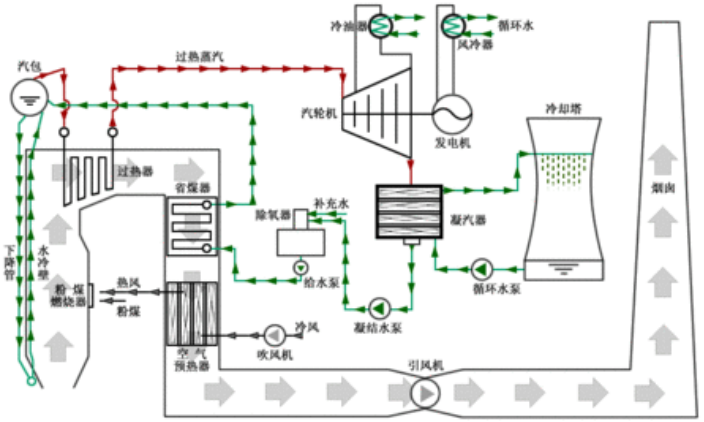


- 生产过程常常有许多测量变量，这些测量变量可能存在大量耦合，同时伴有噪声干扰，基于PCA的过程监测是一种十分有效的方法。

实战练习——PCA处理大规模工业数据集

在大型流程工业，比如火电生产过程中，往往收集数据的监测点有成千上万个之多，高维的数据会对分析和处理带来巨大的困难，而PCA就是我们实现降低数据维度，实现数据压缩的有效工具

在这一小节，我们的任务是**从一组数据中辨别出哪些采样点是汽轮机正常运行时的数据，哪些采样点是汽轮机振动异常时的数据**，即这是一个二分类的问题。

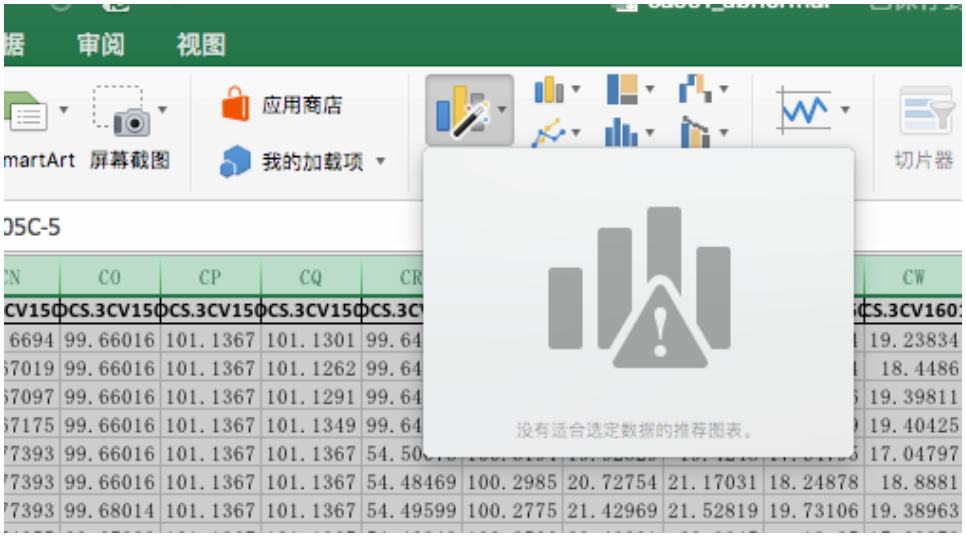


我们要面临的数据维度达2179维，样本点共16560个，样本点分为正常和异常两类部分展示如下：

	CK	CL	CM	CN	CO	CP	CQ	CR	CS	CT	CU	CV	CW	CX	CY	CZ	DA	DB	DC	DD	DE	DF	DG	DH
5	DCS.3CV15	DCS.3CV15	DCS.3CV15	DCS.3CV15	DCS.3CV15	DCS.3CV15	DCS.3CV15	DCS.3CV15	DCS.3CV15	DCS.3CV16	CS.3CV160	DCS.3CV16	CS.3CV160	DCS.3CV16	DCS.3CV16	DCS.3CV16	DCS.3CV16	DCS.3CV16	DCS.3CV16	DCS.3CV16	DCS.3CV16	DCS.3CV16	DCS.3CV16	DCS.3CV16
3	100.9395	99.84151	100.6675	99.6694	99.66016	101.1367	101.1301	99.64381	100.2186	21.93066	22.34386	19.12744	19.23834	60.57718	60.69783	60.39798	58.57381	60.32164	60.91201	60.28087	78.03424	60.62275	60.5314	59.98381
1	100.9656	99.84342	100.6649	99.67019	99.66016	101.1367	101.1262	99.64465	100.2322	21.93066	21.59228	19.12744	18.4486	60.55759	60.72316	60.39186	58.30405	60.34565	60.90518	60.3007	77.0835	60.62762	60.54728	59.99224
1	100.9395	99.84534	100.6688	99.67097	99.66016	101.1367	101.1291	99.64548	100.2173	22.6335	22.58988	19.05596	19.39811	60.56397	60.69146	60.37529	58.39713	60.32038	60.89834	60.32603	82.3125	60.62667	60.51727	60.00066
1	100.9466	99.84727	100.6727	99.67175	99.66016	101.1367	101.1349	99.64632	100.2281	22.60791	21.93182	19.72949	19.40425	60.57618	60.67871	60.35871	58.48347	60.31202	60.89151	60.32456	74.4668	60.62363	60.51198	59.99676
1	100.9651	54.82584	100.6766	24.77393	99.66016	101.1367	101.1367	54.50073	100.3194	19.92529	19.4248	17.64795	17.04797	60.53729	60.67871	60.37269	58.58574	60.31985	60.85125	60.31799	75.09375	60.62058	60.52303	59.98831
1	100.9638	54.81952	100.6805	24.77393	99.66016	101.1367	101.1367	54.48469	100.2985	20.72754	21.17031	18.24878	18.8881	60.54959	60.67871	60.3607	58.57721	60.3119	60.89956	60.31142	78.87891	60.61754	60.53408	59.97986
1	100.9453	54.81319	100.6844	24.77393	99.68014	101.1367	101.1367	54.49599	100.2775	21.42969	21.52819	19.73106	19.38963	60.53353	60.64828	60.37261	58.55993	60.32574	60.883	60.30486	78.55219	60.6145	60.51948	59.99588
1	100.9395	54.80687	100.6855	24.74955	99.67889	101.1367	101.1367	54.49842	100.2566	20.42901	20.3245	18.25	17.98278	60.54946	60.64381	60.36091	58.28223	60.31866	60.88763	60.30485	78.91904	60.61146	60.52466	59.98783
	100.9679	54.80536	100.6855	24.76222	99.67043	101.1367	101.1367	54.49029	100.2996	18.64988	18.71875	16.14404	16.28711	60.54284	60.65208	60.37237	58.27441	60.31158	60.87391	60.30623	80.08203	60.60841	60.52482	59.96688

实战练习——PCA处理大规模工业数据集

很明显，面对这样高维的情况，如果我们直接使用分类算法，会给电脑带来很大的运行负担，同时，分类器也很有可能无法从这么多维数据中找到有效的信息。甚至当我们想要画出这些变量的波形时，我们会发现



CO	CP	CQ	CR	CW
99.66016	101.1367	101.1301	99.64	19.23834
99.66016	101.1367	101.1262	99.64	18.4486
99.66016	101.1367	101.1291	99.64	19.39811
99.66016	101.1367	101.1349	99.64	19.40425
99.66016	101.1367	101.1367	54.50	17.04797
99.66016	101.1367	101.1367	54.48469	100.2985	20.72754	21.17031	18.24878	18.8881
99.68014	101.1367	101.1367	54.49599	100.2775	21.42969	21.52819	19.73106	19.38963

总而言之，直接对这样高维的数据进行分析是不明智的，接下来让我们看一看PCA处理这些高维数据时的表现，这里我们使用python语言处理数据

实战练习——PCA处理大规模工业数据集

实际上，大规模数据的清洗和导入处理本身就是一项非常专业的问题，为了简化问题我们现在假设在python环境中已经有处理和划分好的数据（其实经过了一番犀利的操作）如下：

```
shape of traindata:    (11591, 2179)
shape of trainlabel:   (11591, 1)

shape of testdata:     (4969, 2179)
shape of testlabel:    (4969, 1)
```

将原始数据划分为训练数据和测试数据（7：3），标签设置故障为0，正常为1

我们使用决策树分类器，对数据进行处理，如果不使用PCA算法的话，我们看看结果如何：

```
start time:   2018-03-10 12:24:59
accuarcy:     0.91306
end time   :  2018-03-10 12:44:12
time cost :   00:19:13
```

在这个二分类问题上，模型的运行时间约为20分钟，精度为91%，再来看看使用PCA算法后的效果如何

实战练习——PCA处理大规模工业数据集

Python中的PCA模型使用如下：

```
67 from sklearn.decomposition import PCA #导入PCA模块
68 pca = PCA(0.90)            #建立PCA模型，设置保留原始数据90%的方差信息
69 traindata_pca = pca.fit_transform(traindata)  #PCA处理训练数据
70 testdata_pca = pca.transform(testdata)        #PCA处理测试数据
```

处理结果：

```
shape of traindata:    (11591, 9)
shape of trainlabel:   (11591, 1)

shape of testdata:     (4969, 9)
shape of testlabel:    (4969, 1)
```

从方差的角度来看,我们只需要使用9个维度的属性就可以代替之前2100多个维度的属性

再对使用PCA处理后的数据进行决策树分类

```
start time:   2018-03-10 1:33:02
accuarcy:     0.90518
end time   :  2018-03-10 1:33:02
time cost :   00:00:00
```

可以发现，运行时间已经到了毫秒级，以致程序显示时间为0，而分类精度也仅仅是从0.91降到了0.90

参考书目

- 多变量统计过程控制(SPC) 张杰 阳宪惠
- An Introduction to Multivariate Statistical Analysis

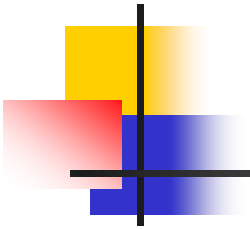
  (Anderson T W, 1984), A JOHN WILEY &SONS, INC., PUBLICATION

小结

- 主元分析，即PCA（Principal Component Analysis）是一种数据降维的方法。PCA把原始变量通过线性组合的方式，导出几个主成分，使它们尽可能多地保留原始变量的信息，且彼此间不相关。从高维空间到低维空间的映射去除了噪声，捕获了数据中的固有变异性。

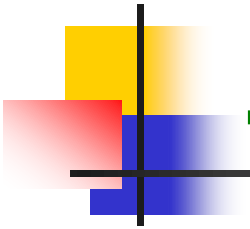
- 目的：既可以降低数据“维数”又保留了原数据的大部分信息。

- 应用：PCA被广泛应用于过程监测、图像处理、人脸识别等众多领域。



小结

- 主成分分析(Principal Component Analysis， PCA)
  - 这仅仅是开始？距离我们的目标是这么近，却又那么远



主元分析

END

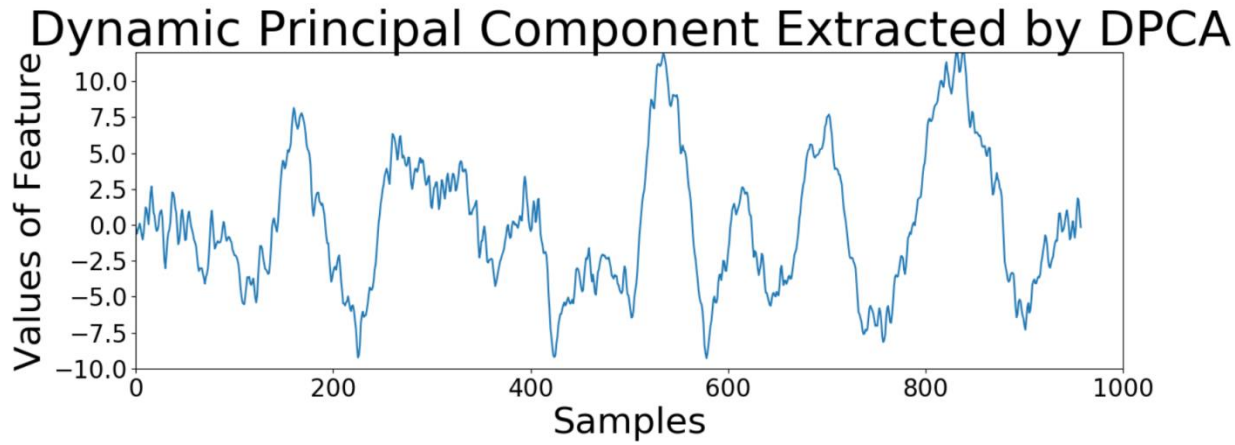
扩展方法-动态PCA

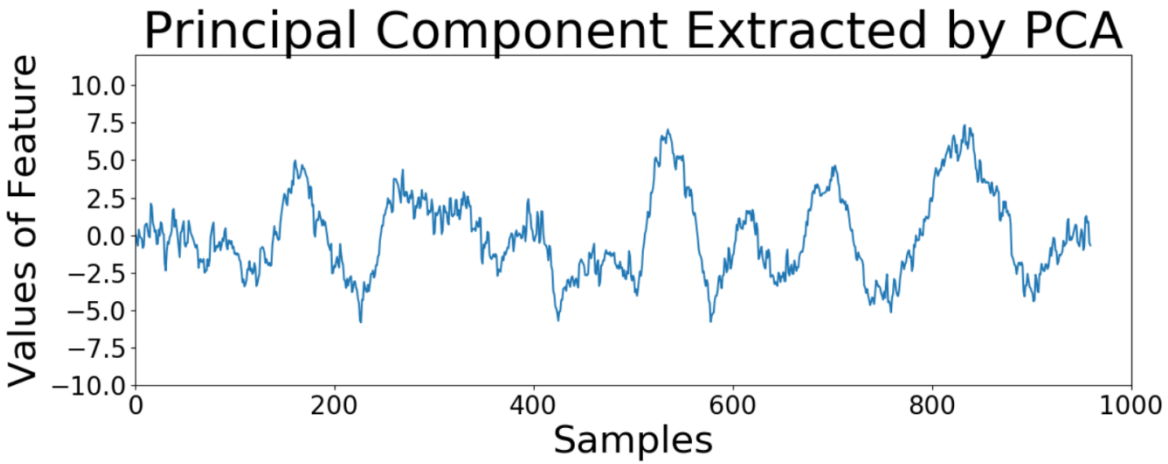
时序相关耦合性的考虑：对分析对象数据矩阵的时序扩展

$$\mathbf{X}_g = \left[\mathbf{X}_1, \mathbf{X}_{2,\ldots,} \mathbf{X}_k\right]$$

(Xg'*Xg)u=λu

扩展方法-动态PCA


Dynamic Principal Component Extracted by DPCA
Values of Feature
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
0
200
400
600
800
1000
Samples


Principal Component Extracted by PCA
Values of Feature
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
0
200
400
600
800
1000
Samples

扩展方法-SFA(slow feature analysis)

慢特征分析方法：变化速度的考虑

SFA的目标是寻找一组投影方向**w**，使得输出信号的变化尽可能缓慢：

$$\min \left\langle \dot{s}_i^2 \right\rangle$$

$$\left\langle s_i \right\rangle = 0 \quad \text{(zero mean)}$$

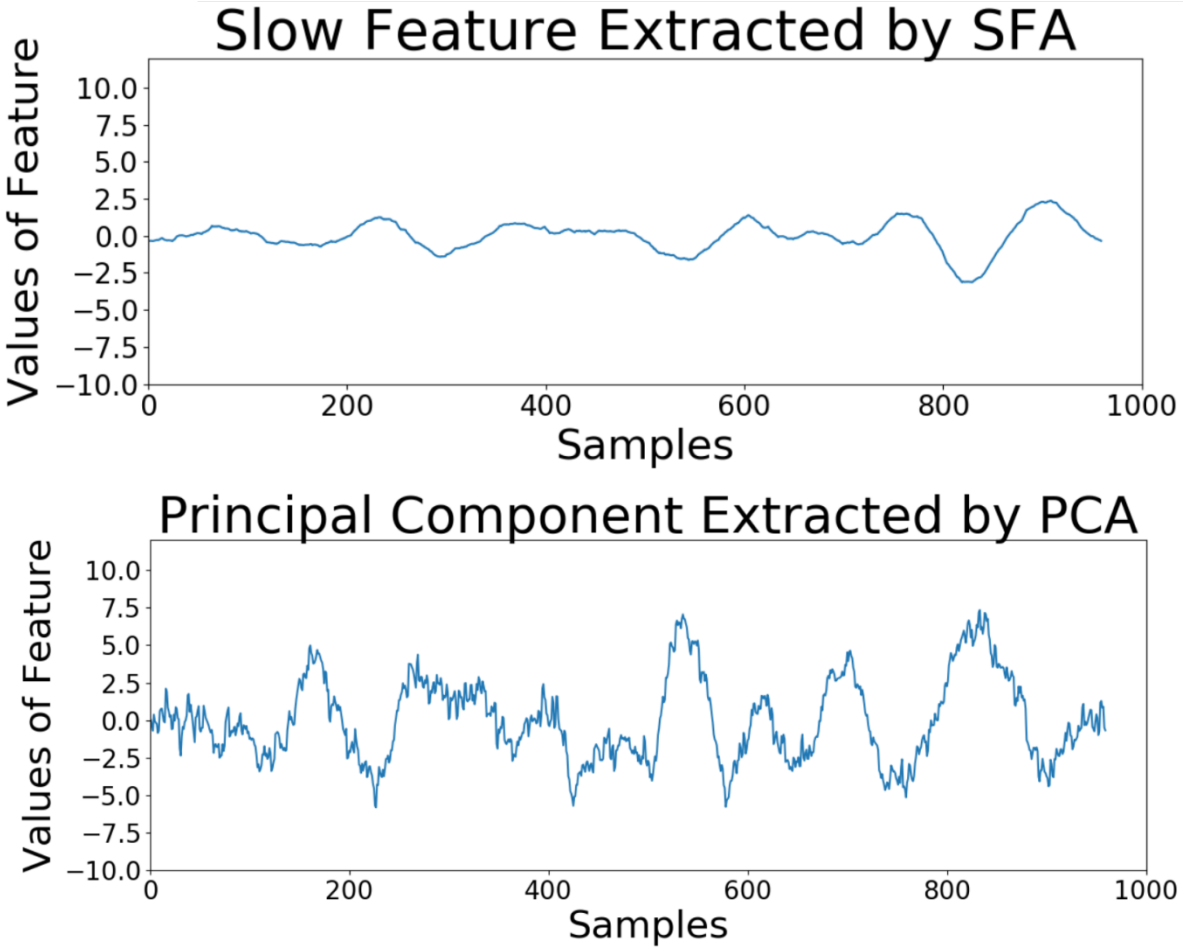
$$\left\langle s_i^2 \right\rangle = 1 \quad \text{(unit deviation)}$$

$$\forall i \neq j : \left\langle s_i, s_j \right\rangle = 0 \quad \text{(decorrelation)}$$

其中，特征信号

$$s_i(t) = \mathbf{w}_i^T \mathbf{x}(t)$$

扩展方法-SFA(slow feature analysis)



问题的改变带来主要信息定义以及提取方式的改变