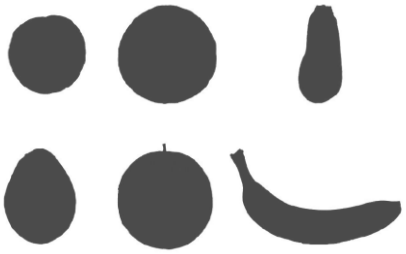
第1-10讲： 微积分
第11-20+讲: 线性代数

第十一讲 主成分分析

2025.4.9



方差=长度^2=信息

Recap

假设样本$\mathbf{y}_i \in R^q, \mathbf{x}_i \in R^p$满足多元线性回归模型：

$$Y_{n\times q} = X B_{p\times q} + \mathcal{E} \Leftrightarrow \mathbf{y}_i = B^{\top}\mathbf{x}_i + \boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_i, i = 1, \dots, n \; iid\sim(\mathbf{0}, \Sigma)$$

横向排

$$Y = \begin{pmatrix} \mathbf{y}_1^{\top} \\ \vdots \\ \mathbf{y}_n^{\top} \end{pmatrix} = \begin{pmatrix} (B^{\top}\mathbf{x}_1+\boldsymbol{\varepsilon}_1)^{\top} \\ \vdots \\ (B^{\top}\mathbf{x}_n+\boldsymbol{\varepsilon}_n)^{\top} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^{\top} \\ \vdots \\ \mathbf{x}_n^{\top} \end{pmatrix} B + \begin{pmatrix} \boldsymbol{\varepsilon}_1^{\top} \\ \vdots \\ \boldsymbol{\varepsilon}_n^{\top} \end{pmatrix} = XB + \mathcal{E}$$

竖向看

$$Y = XB + \mathcal{E} \Leftrightarrow Y = \left(\mathbf{y}_{(1)}, \dots, \mathbf{y}_{(q)}\right) = X\left(\boldsymbol{\beta}_{(1)}, \dots, \boldsymbol{\beta}_{(q)}\right) + \left(\boldsymbol{\varepsilon}_{(1)}, \dots, \boldsymbol{\varepsilon}_{(q)}\right),$$

$$\Leftrightarrow \mathbf{y}_{(k)} = X\boldsymbol{\beta}_{(k)} + \boldsymbol{\varepsilon}_{(k)}, \quad \boldsymbol{\varepsilon}_{(k)}\sim(\mathbf{0}, \sigma_{kk} I_n)$$

$$\text{LS: } \min \|\mathcal{E}\|^2 \Leftrightarrow \min \sum_{k=1}^{q} \left\|\boldsymbol{\varepsilon}_{(k)}\right\|^2 \Leftrightarrow \min \left\|\mathbf{y}_{(k)} - X\boldsymbol{\beta}_{(k)}\right\|^2, k = 1, \dots, q,$$

$$\Rightarrow \hat{\boldsymbol{\beta}}_{(k)} = (X^{\top}X)^{-1}X^{\top}\mathbf{y}_{(k)}, \; \hat{B} = \left(\hat{\boldsymbol{\beta}}_{(1)}, \dots, \hat{\boldsymbol{\beta}}_{(q)}\right) = (X^{\top}X)^{-1}X^{\top}Y,$$

每个响应分别回归：$\mathbf{y}_{(k)}\sim X$，利用得到的q个残差向量$\hat{\boldsymbol{\varepsilon}}_{(1)}, \dots, \hat{\boldsymbol{\varepsilon}}_{(q)}$估计$\Sigma$：$\hat{\sigma}_{ij} = \hat{\boldsymbol{\varepsilon}}_{(i)}^{\top}\hat{\boldsymbol{\varepsilon}}_{(j)}/(n-p), \hat{\Sigma} = \hat{\mathcal{E}}^{\top}\hat{\mathcal{E}}/(n-p), \hat{\mathcal{E}} = \left(\hat{\boldsymbol{\varepsilon}}_{(1)}, \dots, \hat{\boldsymbol{\varepsilon}}_{(q)}\right).$

Y

$\hat{B} = \text{argmin}\|\mathcal{E}\|^2 = \text{argmin} \left\|\mathcal{E}\Sigma^{-1/2}\right\|^2$ (命题3), 矩阵$\mathcal{E}$或Y的每一行可看作是一个“标量”，其方差矩阵Σ看作是“常数”。

主成分分析（PCA）

主成分分析（PCA， principal component analysis） 是由K. Pearson、Hotelling独立发展出来的降维方法。以后会看到，PCA实际上是奇异值分解（SVD）的一个特殊应用。

PCA把多个相关变量线性组合成新的“变量”，如果少数几个组合能包含原来所有变量的大部分**方差信息**，那么我们可以用它们替代原来较多的变量。这些变量的组合称为主成分(PC, principal component)。

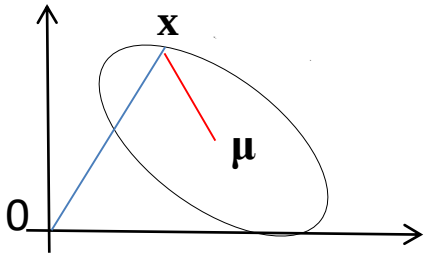
方差: 随机变量长度^2

物体的长度

物体的长度以远端边界点之间的距离，或边界点与中心的距离度量。

随机变量的长度^2：方差

随机变量作为ω的函数，是有“形状”的数学对象，表现为其分布形状。我们以方差$E(x-\mu)^2$而不是$E(x)^2$度量随机变量的“长度”（后者不代表几何形状的长度，参见下图）。



随机向量的各向“长度”

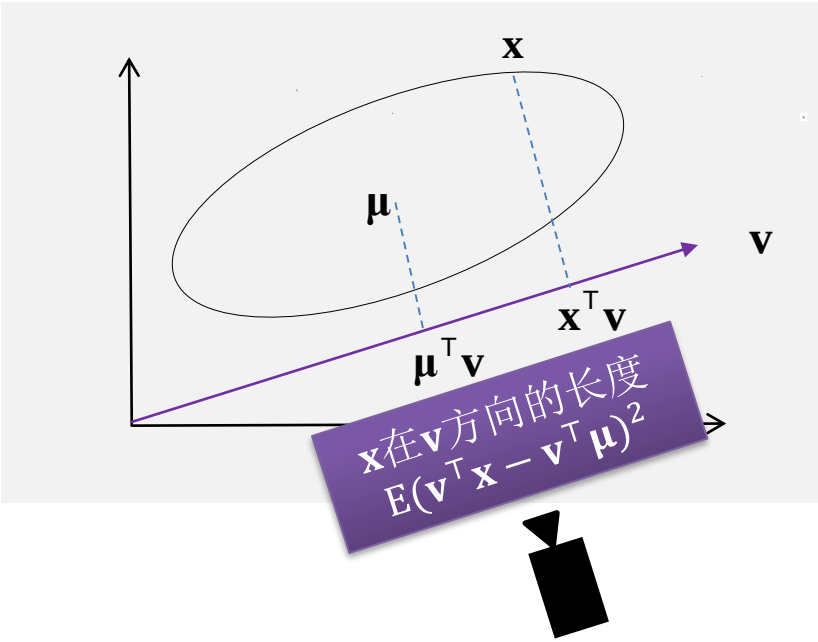
随机向量$\mathbf{x}$的空间轮廓在各个方向上都有长度。对任何方向$\mathbf{v}$，$\|\mathbf{v}\| = 1$，

$\mathbf{x}$在$\mathbf{v}$上的投影坐标为$\mathbf{v}^{\mathsf{T}}\mathbf{x}$，中心$\boldsymbol{\mu}$的投影坐标为$\mathbf{v}^{\mathsf{T}}\boldsymbol{\mu}$，

随机向量在该方向上的方差或“长度^2”：

$$\mathrm{E}(\mathbf{v}^{\mathsf{T}}\mathbf{x} - \mathbf{v}^{\mathsf{T}}\boldsymbol{\mu})^2 = \mathbf{v}^{\mathsf{T}}\,\Sigma\,\mathbf{v},\quad \|\mathbf{v}\| = 1,$$

$$E(\mathbf{x}) = \boldsymbol{\mu},\quad \mathrm{var}(\mathbf{x}) = \Sigma$$



PCA求解随机向量投影坐标方差最大的方向，此时投影坐标称为主成分。

$\mathbf{x}$与中心$\boldsymbol{\mu}$的平均距离$E\|\mathbf{x} - \boldsymbol{\mu}\|^2$ 不能体现出长度的方向性及分量之间的相依性。

总体PCA

我们首先考虑总体（单个随机向量）的主成分分析。

PCA寻找一个低维空间（特别地，一组正交基），使得随机向量在其上的投影最大可能地保留原始数据的信息，信息以投影坐标的方差（长度^2）代表。

这些正交基称为主成分方向（或载荷），而随机向量在这些方向上的投影坐标称为主成分。

投影坐标

$\mathbf{x} \in R^p$在$\mathbf{v} \in R^p$ 上的投影为$P_{\mathbf{v}}\mathbf{x} = \mathbf{v}(\mathbf{v}^\top\mathbf{v})^{-1}\mathbf{v}^\top\mathbf{x}$, 投影坐标为$(\mathbf{v}^\top\mathbf{v})^{-1}\mathbf{v}^\top\mathbf{x}$。特别地，若$\mathbf{v} \in S^{p-1}$, 即$\|\mathbf{v}\| = 1$， $P_{\mathbf{v}}\mathbf{x} = \mathbf{v}(\mathbf{v}^\top\mathbf{x})$, 投影坐标为$\mathbf{v}^\top\mathbf{x}$。

投影坐标的方差

假设随机向量 $\mathbf{x} \in R^p$ 的方差矩阵为$\Sigma = \mathrm{var}(\mathbf{x})$, $\mathbf{x}$在在常数向量$\mathbf{v} \in S^{p-1}$上的投影坐标$\mathbf{v}^\top\mathbf{x}$是一个随机变量，其方差为

$$\mathrm{var}(\mathbf{v}^\top\mathbf{x}) = \mathbf{v}^\top\Sigma\mathbf{v}.$$

哪个方向上投影方差/长度最大？

二次型的极值

定理1. 若$p \times p$ 正定矩阵Σ的特征根为$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$，对应的正交、单位长特征向量为$\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$), 则

$$\max_{\|\mathbf{v}\|=1} \mathbf{v}^{\mathsf{T}}\Sigma\mathbf{v} = \lambda_1$$

最大值在特征向量$\mathbf{v}_1$达到，且对$\forall\ 2 \leq k \leq p$

$$\max_{\|\mathbf{v}\|=1, \mathbf{v}\perp\mathbf{v}_1,\ldots,\mathbf{v}_{k-1}} \mathbf{v}^{\mathsf{T}}\Sigma\mathbf{v} = \lambda_k$$

最大值在特征向量$\mathbf{v}_k$达到.

证明：记$\Lambda = \mathrm{diag}(\lambda_1,\ldots,\lambda_p), V = (\mathbf{v}_1,\ldots,\mathbf{v}_p), V^{\mathsf{T}}V = V^{\mathsf{T}}V = I_p$,即$\Sigma$有谱分解$\Sigma = V\Lambda V^{\mathsf{T}}$。对任何向量$\mathbf{v} \in S^{p-1}$,令$\mathbf{y} = V^{\mathsf{T}}\mathbf{v}$，则$\mathbf{y}^{\mathsf{T}}\mathbf{y} = 1$，$\mathbf{v}^{\mathsf{T}}\Sigma\mathbf{v} = \mathbf{y}^{\mathsf{T}}\Lambda\mathbf{y}$，故

$$\mathbf{v}^{\mathsf{T}}\Sigma\mathbf{v} = \mathbf{y}^{\mathsf{T}}\Lambda\mathbf{y} = \sum_{i=1}^{p}\lambda_1 y_i^2 \leq \lambda_1\sum_{i=1}^{p} y_i^2 = \lambda_1$$

当$y_2 = \ldots = y_p = 0$时等号成立，因为$y_k = \mathbf{v}_k^{\mathsf{T}}\mathbf{v}$，故最大值在$\mathbf{v} \perp \mathbf{v}_2,\ldots,\mathbf{v}_p$时，即$\mathbf{v} = \mathbf{v}_1$时达到。类似地，在$\mathbf{v} \perp \mathbf{v}_1$约束下，$y_1 = \mathbf{v}_1^{\mathsf{T}}\mathbf{x} = 0$，所以

$$\mathbf{v}^{\mathsf{T}}\Sigma\mathbf{v} = \mathbf{y}^{\mathsf{T}}\Lambda\mathbf{y} = \sum_{i=2}^{p}\lambda_i y_i^2 \leq \lambda_2.$$

最大值在$y_3 = \ldots = y_p = 0$，即$\mathbf{v}_3^{\mathsf{T}}\mathbf{v} = \ldots = \mathbf{v}_p^{\mathsf{T}}\mathbf{v} = 0$时达到，故$\mathbf{v} = \mathbf{v}_2$时达到最大。以此类推。

约定：

- 随机向量$\mathbf{x} \in R^p$, 方差矩阵$\Sigma = \mathrm{var}(\mathbf{x}) > 0$的谱分解

$$\Sigma = V\Lambda V^{\mathsf{T}},\ V^{\mathsf{T}}V = VV^{\mathsf{T}} = I_p,\ \Lambda = \mathrm{diag}(\lambda_1,\ldots,\lambda_p),$$

其中$V = (\mathbf{v}_1, \ldots, \mathbf{v}_p)$, $\lambda_1 \geq \cdots \geq \lambda_p > 0$为$\Sigma$的特征根，
$\mathbf{v}_1, \ldots, \mathbf{v}_p$为对应的正交、单位长特征向量。

- 因为PCA只考虑方差，我们不妨假定$\boldsymbol{\mu} = E(\mathbf{x}) = 0$。

主成分定义

最大特征根的特征向量$\mathbf{v}_1 = \arg\max_{\|\mathbf{v}\|=1} \mathbf{v}^{\mathsf{T}}\Sigma\mathbf{v}$ 称为第一主成分方向，$\mathbf{x}$ 在$\mathbf{v}_1$上的投影坐标$y_1 \triangleq \mathbf{v}_1^{\mathsf{T}}\mathbf{x}$称为第一主成分。

在$\mathbf{v}_1$的正交补空间，继续寻找投影长度最大的方向.

$\mathbf{v}_k = \arg\max_{\|\mathbf{v}\|=1, \mathbf{v}\perp\mathbf{v}_1,\ldots,\mathbf{v}_{k-1}} \mathbf{v}^{\mathsf{T}}\Sigma\mathbf{v}$称为第$k$主成分方向， $\mathbf{x}$ 在$\mathbf{v}_k$上的投影坐标$y_k \triangleq \mathbf{v}_k^{\mathsf{T}}\mathbf{x}$称为第$k$个主成分, $k = 2, \ldots, p$。

所有主成分 (主成分变换): $\mathbf{y} = \begin{pmatrix} \mathbf{v}_1^{\mathsf{T}}\mathbf{x} \\ \vdots \\ \mathbf{v}_p^{\mathsf{T}}\mathbf{x} \end{pmatrix} = \begin{pmatrix} \mathbf{v}_1^{\mathsf{T}} \\ \vdots \\ \mathbf{v}_p^{\mathsf{T}} \end{pmatrix} \mathbf{x} = V^{\mathsf{T}}\mathbf{x}.$

V称为载荷矩阵。

$V = (\mathbf{v}_1, \ldots, \mathbf{v}_p) = (v_{ij})$,
行下标 i ： 变量。列下标 j： 主成分

主成分的方差

主成分 $\mathbf{y} = V^{\mathsf{T}}\mathbf{x}$ 的方差 $\mathrm{var}(\mathbf{y})=V^{\mathsf{T}}\Sigma V = \Lambda$, 对角，各个主成分不相关， $\mathrm{var}(y_i) = \lambda_i, i = 1, \dots, p$, 第一主成分的方差最大，第二主成分次之, ...

显然所有主成分的方差之和等于所有$x_1, \dots, x_p$的方差之和：

$$\mathrm{tr}\big(\mathrm{var}(\mathbf{y})\big) = \mathrm{tr}(\Lambda) = \mathrm{tr}(\Sigma) = \mathrm{tr}\big(\mathrm{var}(\mathbf{x})\big) = \lambda_1 + \cdots + \lambda_p$$

载荷矩阵

载荷V作为PC的组合系数，其大小、符号代表了PC的含义。
注意：若 $\mathbf{v}$ 是特征向量/载荷，那么 $-\mathbf{v}$ 也是，所以解释PC含义时，正负号只有相对的含义。

命题1. 假设$\mathrm{var}(\mathbf{x}) = \Sigma > 0$，其谱分解$\Sigma = V\Lambda V^{\mathsf{T}}, V = \big(\mathbf{v}_1, \dots, \mathbf{v}_p\big)$, 主成分$\mathbf{y} = V^{\mathsf{T}}\mathbf{x}$， 则 $\mathrm{cov}\big(\mathbf{x}, y_j\big) = \mathbf{v}_j\lambda_j$。

证明：因为 $y_j = \mathbf{v}_j^{\mathsf{T}}\mathbf{x}$，所以

$$\mathrm{cov}\big(\mathbf{x}, y_j\big) = \mathrm{cov}\big(\mathbf{x}, \mathbf{v}_j^{\mathsf{T}}\mathbf{x}\big) = \mathrm{cov}(\mathbf{x}, \mathbf{x})\mathbf{v}_j = \Sigma\mathbf{v}_j = \mathbf{v}_j\lambda_j$$

注：载荷（主成分方向）$\mathbf{v}_j = \mathrm{cov}(\mathbf{x}, y_j)/\lambda_j \propto \mathrm{cov}(\mathbf{x}, y_j)$，代表第$j$主成分与$\mathbf{x}$的相关程度。

$$\mathrm{cov}(\mathbf{x}, y_j) = \mathbf{v}_j\lambda_j \Rightarrow \mathrm{cov}(x_i, y_j) = \boldsymbol{v}_{ij}\lambda_j$$

v_{ij}代表x_i与y_j的相关程度。

例1. 若$\Sigma = \sigma^2 I_p$,所有特征根都是σ^2，特征向量可取为任意一组正交基，各个主成分重要性相同。若$\Sigma = \mathrm{diag}(\sigma_{11},....,\sigma_{pp})$,所有特征根$\sigma_{11} \geq ... \geq \sigma_{pp}$，特征向量$\mathbf{v}_k = (0,...,1,...,0)^{\mathsf{T}}$,主成分$y_k = x_k, k = 1,...., p.$

例2. 若$\mathbf{x} = (x_1, x_2)^{\mathsf{T}}$的方差矩阵$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$，则其特征根为$1+\rho$和$1-\rho$，对应的特征向量分别是$\mathbf{v}_1 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$，$\mathbf{v}_2 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$

两个主成分别为$y_1 = \mathbf{x}^{\mathsf{T}}\mathbf{v}_1 = (x_1 + x_2)/\sqrt{2}, \quad y_2 = \mathbf{x}^{\mathsf{T}}\mathbf{v}_2 = (x_1 - x_2)/\sqrt{2},$

若$\rho > 0$,则第一主成分$y_1 = (x_1 + x_2)/\sqrt{2}$；若$\rho < 0$,则第一主成分$y_2 = (x_1 - x_2)/\sqrt{2}$

例3. 假设$p\times 1$向量$\mathbf{x}$已经标准化，其协方差矩阵（相关系数矩阵）

$$\mathrm{cov}(\mathbf{x})=\Sigma=\begin{pmatrix}1 & \rho & \cdots & \rho\\ \rho & 1 & \cdots & \rho\\ \vdots & \vdots & \ddots & \vdots\\ \rho & \rho & \cdots & 1\end{pmatrix}=\rho\mathbf{1}\mathbf{1}^{\mathrm{T}}+(1-\rho)I_p,\quad \rho>-\frac{1}{p-1}$$

正定性要求 $\rho>-\dfrac{1}{p-1}$

因为

$$\Sigma\mathbf{1}=\rho\mathbf{1}\mathbf{1}^{\mathrm{T}}\mathbf{1}+(1-\rho)\mathbf{1}=(1+(p-1)\rho)\mathbf{1},$$

所以$1+(p-1)\rho>0$是一个特征根，对应的特征向量为$\mathbf{1}/\sqrt{p}$.

另外，对任何$\mathbf{v}\perp\mathbf{1}$,

$$\Sigma\mathbf{v}=\rho\mathbf{1}\mathbf{1}^{\mathrm{T}}\mathbf{v}+(1-\rho)\mathbf{v}=(1-\rho)\mathbf{v},$$

所以其它$p-1$个特征根都是$1-\rho$，任何$\mathbf{v}\perp\mathbf{1}$都是特征向量。

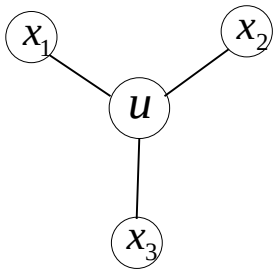
(1) $\rho>0$的情形：最大特征根为$\lambda_1=1+(p-1)\rho$,

第一主成分：$y_1=\mathbf{1}^{\mathrm{T}}\mathbf{x}/\sqrt{p}=(x_1+\ldots+x_p)/\sqrt{p}$, 方差占比：$\dfrac{1+(p-1)\rho}{p}$.

$\rho > 0$时，由$\Sigma = \rho \mathbf{1}\mathbf{1}^{\mathrm{T}} + (1-\rho) I_p$的形式知$\mathbf{x}$可以表示为生成模型

$$\mathbf{x} = \mathbf{1}u + \boldsymbol{\varepsilon}, \quad \mathrm{var}(u) = \rho, \quad \mathrm{var}(\varepsilon_i) = 1-\rho,$$

其中$u, \varepsilon_1, ..., \varepsilon_p$独立，即$x's$由一个共同的变量$u$和若干独立的误差$\varepsilon's$决定。诸$x's$的对称性说明第一主成分取为$\bar{x}$是合理的。



这将被称为
单因子模型

(2) $-\dfrac{1}{p-1} < \rho < 0$的情形：

最大特征根为$1-\rho$(重数$1-p$),最小特征根$1+(p-1)\rho$,类似于$\Sigma = \sigma^2 I_p$情形,少数几个主成分难以替代原始变量$\mathbf{x}$。

问题：两两负相关且相关性相同的生成模型？

例3. $\Sigma = \begin{pmatrix} 1 & \rho & 0 \\ \rho & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \begin{matrix} x_1 \\ x_2 \\ x_3 \end{matrix}$，$\rho < 0$。$\lambda_1 = 2 > \lambda_2 = 1-\rho > \lambda_3 = 1+\rho$,

$$\mathbf{v}_1 = (0,\ 0,\ 1)^{\mathsf{T}},\ \ \mathbf{v}_2 = (1,-1,\ 0)^{\mathsf{T}}/\sqrt{2},\ \ \mathbf{v}_3 = (1,\ \ 1,\ 0)^{\mathsf{T}}/\sqrt{2}$$

主成分：

$$y_1 = \mathbf{v}_1{}^{\mathsf{T}}\mathbf{x} = x_3,\quad y_2 = \mathbf{v}_2{}^{\mathsf{T}}\mathbf{x} = (x_1 - x_2)/\sqrt{2},\quad y_3 = \mathbf{v}_3{}^{\mathsf{T}}\mathbf{x} = (x_1 + x_2)/\sqrt{2},$$

三个主成分解释总方差的比例分别是$\frac{2}{4}, \frac{1-\rho}{4}, \frac{1+\rho}{4}$.

我们可根据载荷解释PC的含义：

载荷矩阵$V = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$:

$$\begin{matrix} & \text{PC1} & \text{PC2} & \text{PC3} \\ x_1 & 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ x_2 & 0 & -1/\sqrt{2} & 1/\sqrt{2} \\ x_3 & 1 & 0 & 0 \end{matrix} \quad V$$

⇨

第一主成分PC1的载荷

$$\mathbf{v}_1 = (0,0,1)^{\mathsf{T}} = \mathrm{cov}(\mathbf{x}, y_1)/\lambda_1$$

这表明第一主成分只与x_3有关，
与x_1, x_2无关。

样本 PCA

样本PCA与总体PCA基本相同，只需将总体协方差矩阵Σ替换为样本协方差矩阵S. 每个样本点都需做主成分变换。

样本：$\mathbf{x}_1, \dots, \mathbf{x}_n \in R^p,\ var(\mathbf{x}_i) = \Sigma;$

数据矩阵: $X_{n\times p} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top;$

样本均值: $\bar{\mathbf{x}} = \frac{\mathbf{x}_1+\cdots+\mathbf{x}_n}{n} = X^\top \mathbb{1}/n$

中心化：$X_c = (\mathbf{x}_1 - \bar{\mathbf{x}}, \dots, \mathbf{x}_n - \bar{\mathbf{x}})^\top$

样本方差矩阵：$S = \frac{1}{n-1}\sum(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top = \frac{1}{n-1}X_c^\top X_c$

样本PCA

样本$\mathbf{x}_1, \dots, \mathbf{x}_n \in R^p$，样本方差矩阵$S$，以S替代总体PCA中的$\Sigma$：

- 假设S的特征根$\lambda_1 \geq \cdots \geq \lambda_p$, 对应的特征向量$\mathbf{v}_1, \dots, \mathbf{v}_p$;

  $S = V\Lambda V^\top, V^\top V = VV^\top = I_p,\ \Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_p), V = (\mathbf{v}_1, \dots, \mathbf{v}_p).$

- $\mathbf{x}_i$的所有主成分$\mathbf{y}_i = V^\top(\mathbf{x}_i - \bar{\mathbf{x}})$, 所有样本的所有主成分：

  $$Y = (y_{ij}) = (\mathbf{y}_1, \dots, \mathbf{y}_n)^\top = X_c V,$$

  其中$y_{ij} = \mathbf{v}_j^\top(\mathbf{x}_i - \bar{\mathbf{x}})$为样本 i 的第 j 个主成分。

- 主成分的样本方差矩阵为

  $$\frac{1}{n-1}Y^\top Y = \frac{1}{n-1}V^\top X_c^\top X_c V = V^\top S V = \Lambda,$$

主成分的选取

主成分变换是1-1变换，主成分矩阵Y与原始数据矩阵X含同样多的信息（同样的总方差：$tr(S) = tr(\Lambda)$）。PCA选取少量主成分代表原来的所有变量，从而达到降维的目的。

样本各分量的方差总和等于所有主成分的方差总和：

$$tr(S) = tr(V\Lambda V^{\mathsf{T}}) = tr(\Lambda) = \lambda_1 + \cdots + \lambda_p,$$

第k个主成分的方差贡献率$\lambda_k/(\lambda_1 + \cdots + \lambda_p)$,

主成分分析（PCA）选择选择部分主成分使得它们能保留原始随机向量总方差的大部分：选择方差最大的前k个主成分$y_1,\ldots, y_k$，使得累计方差贡献率

$$\frac{\lambda_1 + \cdots + \lambda_k}{\lambda_1 + \cdots + \lambda_p} > 80\%$$

我们认为k个主成分保留了原始随机向量的大部分信息，能近似地替代原始随机向量。

用PC可视化数据时, 一般取前两或三个主成分。

```
# PCA函数： princomp （n > p）
mypca = princomp(covmat=R) # R: 协方差矩阵，总体PCA
mypca = princomp(x) #x: 数据矩阵，样本PCA
summary(mypca) #特征根/方差、贡献率
mypca$loadings #载荷
mypca$scores #PC
```

例4. 112个孩子进行了6项测试

general:综合/智商，picture:绘画，blocks:积木，
maze:迷宫， reading:阅 读, vocab:词汇量

其中general是智商测试。各项测试的成绩总分不同，故我们考虑标准化数据的协方差矩阵，即原始数据的相关系数矩阵：

$R =$

	general	picture	blocks	maze	reading	vocab
general	1	0.47	0.55	0.34	0.58	0.51
picture	0.47	1	0.57	0.19	0.26	0.24
blocks	0.55	0.57	1	0.45	0.35	0.36
maze	0.34	0.19	0.45	1	0.18	0.22
reading	0.58	0.26	0.35	0.18	1	0.79
vocab	0.51	0.24	0.36	0.22	0.79	1

相关系数阵的特征根：

主成分	1	2	3	4	5	6
特征根λ_j/方差	3.07	1.14	0.82	0.41	0.35	0.20
累积方差贡献率	0.51	0.70	0.84	0.91	0.97	1.00

总方差$=\lambda_1+\ldots+\lambda_6=tr(R)=6$

第一主成分能解释总方差的$3.07/6=51\%$,

第二主成分解释$1.14/6=19\%$, 累积70%

第三主成分解释$0.82/6=14\%$, 累积84%.

各个主成分的含义可参看载荷矩阵（下页）

载荷矩阵V

	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_3$	$\mathbf{v}_4$	$\mathbf{v}_5$	$\mathbf{v}_6$
	pc1	pc2	pc3	pc4	pc5	pc6
general	0.47	0	0	0.86	0	-0.19
picture	0.36	0.4	0.6	-0.24	0.54	0
blocks	0.43	0.4	0	-0.25	-0.76	0
maze	0.29	0.41	-0.79	0	0.35	0
reading	0.44	-0.51	0	0	0	0.73
vocab	0.43	-0.5	0	-0.37	0	-0.65

$\mathbf{v}_1$：各个分量符号相同，取值近似，故PC1代表平均成绩,综合能力。

$\mathbf{v}_2 = (0,0.4,0.4,0.4,-0.5,-0.5)^{\mathsf{T}}$：general载荷0，PC2与*general*无关，与*picture*, *maze*, *blocks*正相关，与*reading*, *vocab*负相关：

$$\mathrm{PC2} \approx 0.4*(\mathit{picture}+\mathit{maze}+\mathit{blocks})-0.5*(\mathit{reading}+\mathit{vocab})$$

故PC2代表了专业特殊能力（文理差异）而不是一般综合能力。

注意0.4和-0.5的正负号只有相对含义，PC2也可表示为

$$\mathrm{PC2} \approx -0.4*(\mathit{picture}+\mathit{maze}+\mathit{blocks})+0.5*(\mathit{reading}+\mathit{vocab})$$

PC3与general、阅读、词汇量、积木无关，与picture正相关，与maze负相关，PC3代表什么特殊能力？

如果对成绩的协方差矩阵（而不是相关系数矩阵）进行PCA分析，前两个主成分贡献89%, 但载荷的含义不容易解释（如下）。所以，**当随机向量的各个分量尺度(scale)不同时，最好基于相关系数矩阵进行PCA分析。**

协方差矩阵S =

	general	picture	blocks	maze	reading	vocab
general	24.641	5.991	33.52	6.023	20.755	29.701
picture	5.991	6.7	18.137	1.782	4.936	7.204
blocks	33.52	18.137	149.831	19.424	31.43	50.753
maze	6.023	1.782	19.424	12.711	4.757	9.075
reading	20.755	4.936	31.43	4.757	52.604	66.762
vocab	29.701	7.204	50.753	9.075	66.762	135.292

S的特征根

主成分	1	2	3	4	5	6
特征根λ_j / 方差	237.09	102.04	17.57	11.66	9.39	4.02
累积方差贡献率	0.62	0.89	0.93	0.96	0.99	1.00

S的特征向量 / 载荷矩阵V =

	pc1	pc2	pc3	pc4	pc5	pc6
general	0.23	0	0.55	0.65	-0.42	-0.19
picture	0	0	0	0	-0.22	0.96
blocks	0.63	-0.74	-0.13	-0.18	0	0
maze	0	0	0	0.55	0.82	0.14
reading	0.37	0.3	0.67	-0.47	0.32	0.03
vocab	0.63	0.59	-0.47	0.13	-0.11	0

例5.课本Table8.4给出的是5个公司股票的周收益率数据（103周），5家公司为JP Morgan, Citibank, Wells Fargo, Royal Dutch Shell, Exxon Mobil, 各周之间的数据是独立的。5个股票的协方差矩阵 ($x10^{-4}$) 如下

	JP Morgan	Citibank	Wells Fargo	Royal Dutch Shell	Exxon Mobil
JP Morgan	4.33	2.76	1.59	0.64	0.89
Citibank	2.76	4.39	1.8	1.81	1.23
Wells Fargo	1.59	1.8	2.24	0.73	0.61
Royal Dutch Shell	0.64	1.81	0.73	7.22	5.08
Exxon Mobil	0.89	1.23	0.61	5.08	7.66

```
## R 函数princomp
Weekly.return = read.table("http://staff.ustc.edu.cn/~ynyang/vector/data/T8-4.DAT")
dimnames(Weekly.return)[[2]]=c("JP Morgan","Citibank", "Wells Fargo", "Royal Dutch Shell","Exxon Mobil")

cov(weekly.return)->v  ##计算协方差矩阵
mypca = princomp(covmat =v)  ##总体PCA：基于协方差矩阵求解主成分
mypca =princomp(weekly.return) ## 样本PCA：基于原始数据，与上一行结果相同
summary(mypca,loading=T) ##结果汇总
```

> summary(mypca, loading=T) #R 函数princomp分析结果输出：

Importance of components:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Standard deviation	0.037	0.026	0.016	0.012	0.011
Proportion of Variance	0.53	0.27	0.10	0.06	0.05
Cumulative Proportion	0.53	0.80	0.90	0.95	1.00

$\leftarrow \sqrt{\lambda_j}$

$\leftarrow \lambda_j / (\lambda_1 + \ldots + \lambda_5)$

← 两个PC贡献率达到80%

Loadings:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
JP Morgan	0.223	+0.625	0.326	0.663	0.118
Citibank	0.307	+0.570	-0.250	-0.414	-0.589
Wells Fargo	0.155	+0.345	0	-0.497	0.780
Royal Dutch Shell	0.639	-0.248	-0.642	0.309	0.148
Exxon Mobil	0.651	-0.322	0.646	-0.216	

PC1可称为“**市场成分**”：

载荷$\mathbf{v}_1$的各分量符号相同，PC1是各股票的加权平均，该加权平均具有最大的方差，反映了市场波动性。其中最后2个股票对PC1(市场波动性)影响最大.

PC2可称为“**行业成分**”：

载荷$\mathbf{v}_2$的最后2个股票为负(石油)，前三个为正(为银行)，故这两类股票的差距具有第二大的方差或变化，