第十一讲 奇异值分解与主成分分析简介

2025年5月

PART I 奇异值分解

概要

对非零矩阵 $A \in \mathbf{R}^{m \times n}$, 其秩 $\text{rank}(A) = r$

- $r \leq \min\{m, n\}$.
- A的四个基本子空间
  - A的值域:

$$R(A) = \{z \in \mathbf{R}^m | \exists x \in \mathbf{R}^n, z = Ax\} = A \text{ 的列空间}.$$

  - A的零空间:

$$N(A) = \{x \in \mathbf{R}^n | Ax = 0\}.$$

  - A^T的值域:

$$R(A^T) = \{y \in \mathbf{R}^n | \exists x \in \mathbf{R}^m, y = A^T x\} = A \text{ 的行空间}.$$

  - A^T的零空间:

$$N(A^T) = \{x \in \mathbf{R}^m | A^T x = 0\}.$$

对非零矩阵 $A \in \mathbf{R}^{m\times n}$, $\mathrm{rank}(A) = r$.
A的四个基本子空间:

- $R(A), N(A^T) \subseteq \mathbf{R}^m$.
  - $R(A)^\perp = N(A^T)$.
  - $\dim(R(A)) = \mathrm{rank}(A) = r$.
  - $\dim(N(A^T)) = m - r$.
- $R(A^T), N(A) \subseteq \mathbf{R}^n$.
  - $R(A^T)^\perp = N(A)$.
  - $\dim(R(A^T)) = \mathrm{rank}(A) = r$.
  - $\dim(N(A)) = n - r$.

对非零矩阵 $A \in \mathbf{R}^{m\times n}$, 其秩 $\mathrm{rank}(A) = r$ 且不妨设$m \geq n$, 考虑实对称矩阵 $A^TA_{n\times n}$:

- 不妨设 A^TA的特征值按降序排列为$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, 则$\lambda_n \geq 0$.
- $\mathrm{rank}(A^TA) = \mathrm{rank}(A) = r$.
- A^TA可对角化：即

$$V^TA^TAV = \Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n),$$

  这里$V = [v_1, v_2, \cdots, v_n]$为正交矩阵.
- $A^TA = V\Lambda V^T$.

$$A^T A = V \Lambda V^T.$$

能否构造A的类似因子乘积的表示形式?

- $\mathrm{rank}(\Lambda) = \mathrm{rank}(A^T A) = r$, 则
  - $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$,
  - $\lambda_{r+1} = \cdots = \lambda_n = 0$.
- 我们将V分成两部分$V = [V_1, V_2]$,其中 $V_1 = [v_1, \cdots, v_r]$, $V_2 = [v_{r+1}, \cdots, v_n]$.
  - $v_{r+1}, \cdots, v_n$正好构成$A^T A$的零空间$\mathrm{N}(A^T A)$(也是$\mathrm{N}(A)$)的一组标准正交基.
  - $AV_2 = 0$.
  - $I = VV^T = V_1 V_1^T + V_2 V_2^T$.
  - $A = AI = A(V_1 V_1^T + V_2 V_2^T) = AV_1 V_1^T$.

- 从$V_1=[v_1,v_2,\cdots,v_r]$出发， 构造$U_1=[u_1,\cdots,u_r]$如下：

$$u_i=\frac{1}{\sqrt{\lambda_i}}Av_i,\ i=1,\cdots,r.$$

则

$$u_i^Tu_j=\frac{1}{\sqrt{\lambda_i\lambda_j}}v_i^TA^TAv_j=\frac{1}{\sqrt{\lambda_i\lambda_j}}v_i^T\lambda_jv_j=\delta_{ij}.$$

这说明

$$u_1,\cdots,u_r$$

是A的值域空间$R(A)$的一组标准正交基.

- $R(A)$的正交补空间为 $N(A^T)$.
- 设$u_{r+1},\cdots,u_m$为$N(A^T)$的一组标准正交集.
- 令$U_2=[u_{r+1},\cdots,u_m]$, 构造正交矩阵

$$U=[U_1,U_2].$$

- 令$\sigma_i=\sqrt{\lambda_i}$, $i=1,2,\cdots,r$. 构造对角线元素降序排列的对角矩阵

$$\Sigma_1=\mathrm{diag}(\sigma_1,\cdots,\sigma_r).$$

由前面U_1构造可知$U_1\Sigma_1=AV_1$.
- 进一步构造$m\times n$矩形对角矩阵

$$\Sigma=\begin{bmatrix}\Sigma_1 & 0\\ 0 & 0\end{bmatrix}$$

定理1

$$A = U\Sigma V^T.$$

证明: 我们先计算等式右边

$$\begin{aligned} U\Sigma V^T &= [U_1, U_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \\ &= U_1 \Sigma_1 V_1^T \\ &= AV_1 V_1^T \\ &= A. \qquad \square \end{aligned}$$

奇异值分解

矩阵的奇异值分解运算就是将一个非零的$m\times n$实矩阵$A\in\mathbf{R}^{m\times n}$，表示为满足如下特性的三个实矩阵乘积形式的因子分解运算：

$$A=U\Sigma V^T,$$

其中

- U 是m阶正交矩阵，$U^TU=I$;
- V 是n阶正交矩阵，$V^TV=I$;
- Σ是由降序排列的非负的对角线元素组成的$m\times n$矩形对角矩阵：

  $$\Sigma=\mathrm{diag}(\sigma_1,\cdots,\sigma_p),$$

  这里$\sigma_1\geq\cdots\geq\sigma_p\geq 0$, 且$p=\min(m,n)$.

$U\Sigma V^T$ 称为A的奇异值分解，σ_i称为A的奇异值, U和 V 的列向量分别称为A的左、右奇异向量.

- 由定理1可知，对任一非零的$m \times n$实矩阵 $A \in \mathbf{R}^{m\times n}$，其奇异值分解存在.
- 由$A = U\Sigma V^T$可推出
  - $A^TA = V(\Sigma^T\Sigma)V^T$.
  - 右奇异向量为A^TA的特征向量.
  - $AA^T = U(\Sigma\Sigma^T)U^T$.
  - 左奇异向量为AA^T的特征向量.
  - $AV = U\Sigma$.
  - $A^TU = V\Sigma^T$.
- A的奇异值$\sigma_1, \cdots, \sigma_n$是唯一的，但$U$和 V 不唯一.
- 矩阵的奇异值分解可以看作将其对应的线性变换分解为旋转变换(V^T)、缩放变换(Σ)及旋转变换(U)的组合.

紧奇异值分解与截断奇异值分解

一个非零的秩为r的$m \times n$实矩阵 $A \in \mathbf{R}^{m \times n}$奇异值分解为$U \Sigma V^T$，其中

- $U = [u_1, u_2, \cdots, u_m]$；
- $V = [v_1, v_2, \cdots, v_n]$；
- $\Sigma = \mathrm{diag}(\sigma_1, \cdots, \sigma_p)$, $p = \min(m, n)$.

则$A = U_r \Sigma_r V_r^T$ 称为A的紧奇异值分解， 其中

- $U_r = [u_1, u_2, \cdots, u_r]$；
- $V_r = [v_1, v_2, \cdots, v_r]$；
- $\Sigma_r = \mathrm{diag}(\sigma_1, \cdots, \sigma_r)$.

紧奇异值分解与截断奇异值分解

对$0 < k < r$，$U_k\Sigma_k V_k^T$ 称为A的截断奇异值分解， 其中

- $U_k = [u_1, u_2, \cdots, u_k]$；
- $V_k = [v_1, v_2, \cdots, v_k]$；
- $\Sigma_k = \mathrm{diag}(\sigma_1, \cdots, \sigma_k)$.

问题: $A \approx U_k\Sigma_k V_k^T$ 的程度如何度量?

概要

费罗贝尼乌斯范数

设$A \in \mathbf{R}^{m\times n}$且$A = [a_{ij}]_{m\times n}$, A的费罗贝尼乌斯范数 (Frobenius norm) $\| A \|_F$定义如下:

$$\| A \|_F = \left[\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 \right]^{\frac{1}{2}}$$

- 若Q是 m阶正交矩阵, 则$\| QA \|_F = \| A \|_F$.
- 若P是 n阶正交矩阵, 则$\| AP \|_F = \| A \|_F$.
- 若Q是 m阶正交矩阵, P是 n阶正交矩阵, 则

$$\| QAP \|_F = \| A \|_F \, .$$

- 若Q是 m阶正交矩阵, 则$\| QA \|_F = \| A \|_F$.
  证明: 不妨设$A = [a_1, \cdots, a_n]$, 则$QA = [Qa_1, \cdots, Qa_n]$. 故

$$
\begin{aligned}
\| QA \|_F^2 = \| [Qa_1, \cdots, Qa_n] \|_F^2 &= \sum_{i=1}^{n} \| Qa_i \|_2^2 \\
= \sum_{i=1}^{n} (Qa_i)^T Qa_i &= \sum_{i=1}^{n} a_i^T Q^T Qa_i \\
= \sum_{i=1}^{n} \| a_i \|_2^2 &= \| A \|_F^2 . \qquad \square
\end{aligned}
$$

引理1

设 $A \in \mathbf{R}^{m\times n}$的奇异值分解为$U\Sigma V^T$, 其中$\Sigma = \mathrm{diag}(\sigma_1, \cdots, \sigma_n)$,

$$\| A \|_F = \sqrt{\sum_{i=1}^{n} \sigma_i^2}$$

证明：$\| A \|_F = \| U\Sigma V^T \|_F = \| \Sigma \|_F = \sqrt{\sum_{i=1}^{n} \sigma_i^2}$. $\square$

引理2

设 $A \in \mathbf{R}^{m\times n}$ 且 $\mathrm{rank}(A) = r$, 并设$\mathcal{M}$为$\mathbf{R}^{m\times n}$中所有秩不超过k的矩阵集合,$0 < k < r$, 则存在一个秩为k的矩阵X, 使得

$$\| A - X \|_F = \min_{S \in \mathcal{M}} \| A - S \|_F \, .$$

称矩阵X为矩阵A在费罗贝尼乌斯范数意义下的最优近似.

定理2

设 $A \in \mathbf{R}^{m\times n}$ 且 $\mathrm{rank}(A) = r$, 其奇异值分解为$U\Sigma V^T$,并设$\mathcal{M}$为$\mathbf{R}^{m\times n}$中所有秩不超过k的矩阵集合,$0 < k < r$,

- 若秩为k的矩阵$X \in \mathcal{M}$满足 $\| A - X \|_F = \min\limits_{S\in\mathcal{M}} \| A - S \|_F$,

  则 $\| A - X \|_F = \sqrt{\sum\limits_{l=k+1}^{n} \sigma_l^2}$.

- 特别地，若$A' = U\Sigma' V^T$,其中$\Sigma'_{m\times n} = \begin{bmatrix} \Sigma_k & 0 \\ 0 & 0 \end{bmatrix}$, 这里$\Sigma_k = \mathrm{diag}(\sigma_1, \cdots, \sigma_k)$, 则

$$\| A - A' \|_F = \sqrt{\sum_{l=k+1}^{n} \sigma_l^2} = \min_{S\in\mathcal{M}} \| A - S \|_F .$$

证明：设X为A的最优近似，则

$$\| A - X \|_F \leq \| A - A' \|_F = \sqrt{\sum_{l=k+1}^{n} \sigma_l^2}.$$

下面证明

$$\| A - X \|_F \geq \sqrt{\sum_{l=k+1}^{n} \sigma_l^2}.$$

设X的奇异值分解为$Q\Omega P^T$，其中

$$\Omega = \begin{bmatrix} \Omega_1 & 0 \\ 0 & 0 \end{bmatrix},$$

这里 $\Omega_1 = \mathrm{diag}(\omega_1, \cdots, \omega_k)$.

令$B=Q^TAP$, 则$A=QBP^T$. 按照Ω的分块方法对B进行分块

$$B=\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix},$$

则

$$\begin{aligned}\| A-X \|_F^2 &= \| Q(B-\Omega)P^T \|_F^2 = \| B-\Omega \|_F^2 \\ &= \| B_{11}-\Omega_1 \|_F^2 + \| B_{12} \|_F^2 + \| B_{21} \|_F^2 + \| B_{22} \|_F^2 .\end{aligned}$$

下面我们证明$B_{12}=0$、 $B_{21}=0$ 、$B_{11}=\Omega_k$.

- 若$B_{12}\neq 0$, 则构造 $Y=Q\begin{bmatrix} B_{11} & B_{12} \\ 0 & 0 \end{bmatrix}P^T$, 则$Y\in\mathcal{M}$, 且

$$\| A-Y \|_F^2 = \| B_{21} \|_F^2 + \| B_{22} \|_F^2 < \| A-X \|_F^2 .$$

这与X是最优近似相矛盾.

- 同理可证$B_{21}=0$.
- 若$B_{11}\neq\Omega_1$, 则构造 $Z=Q\begin{bmatrix} B_{11} & 0 \\ 0 & 0 \end{bmatrix}P^T$, 则$Z\in\mathcal{M}$, 且

$$\| A-Z \|_F^2=\| B_{22} \|_F^2<\| B_{11}-\Omega_1 \|_F^2+\| B_{22} \|_F^2=\| A-X \|_F^2 .$$

这与X是最优近似相矛盾,故$B_{11}=\Omega_k$.

基于上面已经证明的结论, 则

$$\| A-X \|_F=\| B_{22} \|_F .$$

设B_{22}的奇异值分解为$U_1\Lambda V_1^T$, 则

$$\| A-X \|_F=\| B_{22} \|_F=\| \Lambda \|_F .$$

令$U_2 = \begin{bmatrix} I_k & 0 \\ 0 & U_1 \end{bmatrix}$, $V_2 = \begin{bmatrix} I_k & 0 \\ 0 & V_1 \end{bmatrix}$, 则

$$U_2^T Q^T A P V_2 = \begin{bmatrix} \Omega_1 & 0 \\ 0 & \Lambda \end{bmatrix},$$

即

$$A = Q U_2 \begin{bmatrix} \Omega_1 & 0 \\ 0 & \Lambda \end{bmatrix} (P V_2)^T.$$

这意味着Λ的对角线元素为A的奇异值，故有

$$\| A - X \|_F = \| \Lambda \|_F \geq \sqrt{\sum_{l=k+1}^{n} \sigma_l^2}.$$

因此 $\| A - X \|_F = \sqrt{\sum_{l=k+1}^{n} \sigma_l^2} = \| A - A' \|_F$. $\square$

- 定理2说明在秩不超过k的$m \times n$矩阵集合中， 存在矩阵A的最优近似矩阵X; $A' = U\Sigma' V^T$就是这样的最优近似矩阵.
- A的紧奇异值分解是在费罗贝尼乌斯范数意义下A的无损压缩.
- A的秩为k的截断奇异值分解是A的有损压缩, 通常k远小于r， 因此是由低秩矩阵实现了对A的压缩。

- 设$A \in \mathbf{R}^{m\times n}$ 且 其奇异值分解为$U\Sigma V^T$, 则

$$\begin{aligned} A &= U\Sigma V^T \\ &= [\sigma_1 u_1, \cdots, \sigma_n u_n] \begin{bmatrix} v_1^T \\ \vdots \\ v_n^T \end{bmatrix} \\ &= \sigma_1 u_1 v_1^T + \cdots + \sigma_n u_n v_n^T. \end{aligned}$$

- 称$A = \sigma_1 u_1 v_1^T + \cdots + \sigma_n u_n v_n^T$为$A$的外积展开式.
- A的外积展开式也可以看作矩阵的有序加权和$\sum_{k=1}^{n} \sigma_k u_k v_k^T$.

- 若A的秩为n，则

$$A = \sigma_1 u_1 v_1^T + \cdots + \sigma_n u_n v_n^T.$$

- 令

$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T,$$

则A_k的秩为k，是A的截断奇异值分解.

- A_k是秩为k的矩阵中在费罗贝尼乌斯范数意义下A的最优近似矩阵.
- 由于通常奇异值σ_i递减很快，所以k很小时，A_k对A也可有很好的近似.

奇异值分解小结

- 矩阵的基本子空间
- 矩阵的奇异值分解
- 紧奇异值分解与截断奇异值分解
- 矩阵的费罗贝尼乌斯范数
- 截断奇异值分解与最优近似矩阵
- 矩阵的外积表示
- 奇异值分解与谱分解

PART II 主成分分析

概要

3 总体主成分分析
- 标准化随机变量
- 方差贡献率
- 因子负荷量

主成分分析

- 如果数据的一些特征之间存在相关性，处理起来不太方便;
- 如果数据维数过高，影响算法性能.

我们希望能构造一组新的相互不相关的特征来表示数据:

- 通常用原来特征的线性组合来构造新特征.
- 希望特征变换的过程中损失的信息尽可能少.
- 构造出的新特征个数比原来的特征数少很多，达到降维的目的.

标准线性组合

设$\mathbf{x} = (x_1, x_2, \cdots, x_m)^T$是$m$维随机向量, $\alpha \in \mathbf{R}^m$ 且$\alpha^T\alpha = 1$, 则称

$$y = \alpha^T\mathbf{x}$$

为标准线性组合.

本讲主要考虑标准线性组合.

设$\mathbf{x} = (x_1, x_2, \cdots, x_m)^T$是$m$维随机向量，其均值为$\mu$，协方差矩阵

$$\Sigma = E[(\mathbf{x} - \mu)(\mathbf{x} - \mu)^T] = [\sigma_{ij}]_{m\times m},$$

则

- $E[(\mathbf{x} - \mu)] = 0$.
- $\mathrm{trace}(\Sigma) = \sum_{i=1}^{m} \mathrm{Var}(x_i) = \sum_{i=1}^{m} \sigma_{ii}$.
- Σ是半正定的.
- 不妨设Σ的特征值按照降序排列为

$$\lambda_1 \geq \lambda_2 \cdots \geq \lambda_m \geq 0$$

- Σ可对角化: 即

$$A^T \Sigma A = \Lambda = diag(\lambda_1, \lambda_2, \cdots, \lambda_m),$$

这里$A = [\alpha_1, \alpha_2, \cdots, \alpha_m]$为正交矩阵,
$\alpha_i = (\alpha_{1i}, \alpha_{2i}, \cdots, \alpha_{mi})^T$ 且α_i是Σ的属于λ_i的特征向量.

- $\alpha_1, \alpha_2, \cdots, \alpha_m$正好构成了$\mathbf{R}^m$的一组标准正交基, 即
对$\forall \alpha \in \mathbf{R}^m$, $\exists c_1, c_2, \cdots, c_m \in \mathbf{R}$ s.t. $\alpha = \sum_{i=1}^{m} c_i \alpha_i$.

主成分变换

设$\mathbf{x} = (x_1, x_2, \cdots, x_m)^T$是均值为$\mu$, 协方差矩阵为$\Sigma$的$m$维随机向量, 则如下线性变换被称为主成分变换:

$$\mathbf{y} = A^T(\mathbf{x} - \mu).$$

并称$\mathbf{y}$的第i个分量

$$y_i = \alpha_i^T(\mathbf{x} - \mu)$$

为$\mathbf{x}$的第i主成分, 这里α_i为A的第i个列向量.

考虑$\mathbf{x} = (x_1, x_2)^T$, 其中

- $E[x_1] = E[x_2] = 0$,
- $\mathrm{Var}(x_1) = \mathrm{Var}(x_2) = 1$,
- $\mathrm{Cov}(x_1, x_2) = \rho > 0$,

协方差矩阵为

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

其特征值为

$$\lambda_1 = 1 + \rho, \quad \lambda_2 = 1 - \rho$$

相应的特征向量为

$$\alpha_1 = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})^T, \quad \alpha_2 = (\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})^T$$

由此得到**x**的第一、第二主成分分别为

$$y_1 = \frac{1}{\sqrt{2}}(x_1 + x_2);$$
$$y_2 = \frac{1}{\sqrt{2}}(x_1 - x_2)$$

相应的方差为

$$\mathrm{Var}(y_1) = 1 + \rho = \lambda_1;$$
$$\mathrm{Var}(y_2) = 1 - \rho = \lambda_2.$$
$$\mathrm{Var}(y_1) + \mathrm{Var}(y_2) = \mathrm{Var}(x_1) + \mathrm{Var}(x_2) = 2.$$

主成分的性质

TH1. 设$\mathbf{x} \sim (\mu, \Sigma)$， 则$\mathbf{y} = A^T(\mathbf{x} - \mu)$满足

(1) $E[\mathbf{y}] = \mathbf{0}$.

(2) $\mathrm{Var}(y_i) = \lambda_i,\ i = 1, 2, \cdots, m$.

(3) $\mathrm{Cov}(y_i, y_j) = 0,\ i \neq j,\ i, j = 1, 2, \cdots, m$..

(4) $\mathrm{Var}(y_1) \geq \mathrm{Var}(y_2) \geq \cdots \geq \mathrm{Var}(y_m) \geq 0$.

(5) $\sum_{i=1}^{m} \mathrm{Var}(y_i) = \mathrm{trace}(\Sigma) = \sum_{i=1}^{m} \mathrm{Var}(x_i)$.

(6) $\prod_{i=1}^{m} \mathrm{Var}(y_i) = |\Sigma|$.

(1) 的证明： $E[\mathbf{y}] = E[A^T(\mathbf{x}-\mu)] = A^T E[(\mathbf{x}-\mu)] = \mathbf{0}.$

(2)的证明：

$$\begin{aligned}\mathrm{Var}(y_i) = \mathrm{Var}(\alpha_i^T(\mathbf{x}-\mu)) &= \alpha_i^T E[(\mathbf{x}-\mu)(\mathbf{x}-\mu)^T]\alpha_i \\ &= \alpha_i^T \Sigma \alpha_i = \lambda_i \alpha_i^T \alpha_i = \lambda_i\end{aligned}$$

(3)的证明：如果$i \neq j$,

$$\begin{aligned}\mathrm{Cov}(y_i, y_j) &= \mathrm{Cov}(\alpha_i^T(\mathbf{x}-\mu), \alpha_j^T(\mathbf{x}-\mu)) \\ &= \alpha_i^T E[(\mathbf{x}-\mu)(\mathbf{x}-\mu)^T]\alpha_j \\ &= \alpha_i^T \Sigma \alpha_j = \lambda_j \alpha_i^T \alpha_j = 0.\end{aligned}$$

由(2)可知， (4), (5),(6)都成立. □

TH2. 不存在方差比λ_1更大的标准线性组合$y = \alpha^T \mathbf{x}$.

证明：考虑标准线性组合$y = \alpha^T \mathbf{x}$, 其中$\alpha \in \mathbf{R}^m$ 且$\alpha^T \alpha = 1$. 由于$\alpha_1, \alpha_2, \cdots, \alpha_m$正好构成了$\mathbf{R}^m$的一组标准正交基, 则

$$\exists c_1, c_2, \cdots, c_m \in \mathbf{R} \ s.t. \ \alpha = \sum_{i=1}^{m} c_i \alpha_i.$$

对此线性组合来说,

$$\begin{aligned}\mathrm{Var}(y) = \alpha^T \Sigma \alpha &= \left[\sum_{i=1}^{m} c_i \alpha_i^T\right] \Sigma \left[\sum_{i=1}^{m} c_i \alpha_i\right] \\ &= \sum_{i=1}^{m} c_i^2 \lambda_i \alpha_i^T \alpha_i = \sum_{i=1}^{m} c_i^2 \lambda_i.\end{aligned}$$

另一方面$\alpha^T\alpha = 1$意味着$\sum_{i=1}^{m} c_i^2 = 1$.

考虑如下最优化问题：

$$\max_{c_1,\cdots,c_m} \sum_{i=1}^{m} c_i^2 \lambda_i$$

$$s.t. \quad \sum_{i=1}^{m} c_i^2 = 1.$$

则此问题的最优解为

$$c_1 = 1, c_2 = \cdots = c_m = 0$$

此时

$$\lambda_1 = \max_{c_1,\cdots,c_m} \sum_{i=1}^{m} c_i^2 \lambda_i$$

对应的标准线性组合

$$y_1 = \alpha_1^T \mathbf{x}$$

正好是的第一主成分. □

TH3. 如果标准线性组合$y=\alpha^T\mathbf{x}$与$\mathbf{x}$的前k个主成分都不相关，则y的方差当y是第$k+1$主成分时达到最大.

证明：设$y=\alpha^T\mathbf{x}$, 其中$\alpha^T\alpha=1$ 且$\alpha=\sum_{i=1}^m c_i\alpha_i$. 对此线性组合来说，

$$\mathrm{Var}(y)=\sum_{i=1}^m c_i^2\lambda_i.$$

对$1\le j<k$来说，

$$\begin{aligned}\mathrm{Cov}(y,y_j)=\mathrm{Cov}(\alpha^T\mathbf{x},\alpha_j^T\mathbf{x})&=\left[\sum_{i=1}^m c_i\alpha_i^T\right]\Sigma\alpha_j\\&=c_j\lambda_j\alpha_j^T\alpha_j=c_j\lambda_j=0.\end{aligned}$$

这意味着对$1 \le j < k$来说，$c_j^2\lambda_j = 0$. 故

$$\mathrm{Var}(y) = \sum_{i=k+1}^{m} c_i^2\lambda_i.$$

和前面证明类似， 我们可得

$$\max_{c_1,\cdots,c_m} \mathrm{Var}(y) = \lambda_{k+1}.$$

对应的标准线性组合

$$y = \alpha_{k+1}^T \mathbf{x}$$

正好是的第$k+1$主成分. □

设$\mathbf{x} = (x_1, x_2, \cdots, x_m)^T$是均值为$\mu$, 协方差矩阵为$\Sigma$的$m$维随机向量, 对每个$x_i$定义

$$x'_i = \frac{x_i - \mu_i}{\sqrt{\sigma_{ii}}},$$

则

$$E[x'_i] = 0, \qquad \mathrm{Var}(x'_i) = 1.$$

对$\mathbf{x}' = (x'_1, x'_2, \cdots, x'_m)^T$来说,

$$\begin{aligned} \mathrm{Cov}(x'_i, x'_j) &= E\left[\left(\frac{x_i - \mu_i}{\sqrt{\sigma_{ii}}}\right)\left(\frac{x_j - \mu_j}{\sqrt{\sigma_{jj}}}\right) \right] \\ &= \frac{E[(x_i - \mu_i)(x_j - \mu_j)]}{\sqrt{\sigma_{ii}\sigma_{jj}}} = \rho_{ij} \end{aligned}$$

即$\mathbf{x}'$的协方差矩阵Σ'为$\mathbf{x}$的相关矩阵: $\Sigma' = [\rho_{ij}]_{m \times m}$. 此时$\mathrm{trace}(\Sigma') = m$.

$\mathbf{x}$的第 k主成分 y_k的方差贡献率 η_k定义为 $\eta_k = \frac{\lambda_k}{\sum\limits_{i=0}^{m} \lambda_i}$.

$\mathbf{x}$的前k个主成分 $y_1, \cdots, y_k$的累计方差贡献率 $\eta_{1\to k}$定义为

$$\eta_{1\to k} = \sum_{i=1}^{k} \eta_i = \frac{\sum\limits_{i=1}^{k} \lambda_i}{\sum\limits_{j=0}^{m} \lambda_j}.$$

累计方差贡献率 $\eta_{1\to k}$反映了$\mathbf{x}$的前k个主成分保留原有变量方差信息的比例，可以作为k的选择标准， 比如选择k使得 $\eta_{1\to k}$达到规定的百分比(如80%)以上.

- 主成分变换$\mathbf{y} = A^T\mathbf{x}$的逆变换为 $\mathbf{x} = A\mathbf{y}$.
- $x_i = \sum_{j=1}^{m} \alpha_{ij} y_j, \ i = 1, 2, \cdots, m.$
- 因子负荷量：第k主成分与变量x_i的相关系数，即y_k对x_i的贡献程度：

$$\rho(y_k, x_i) = \frac{\mathrm{Cov}(\sum_{j=1}^{m} \alpha_{ij} y_j, y_k)}{\sqrt{\lambda_k \sigma_{ii}}} = \frac{\alpha_{ik} \mathrm{Var}(y_k)}{\sqrt{\lambda_k \sigma_{ii}}} = \frac{\sqrt{\lambda_k} \alpha_{ik}}{\sqrt{\sigma_{ii}}}$$

- 因子负荷量满足如下性质：
  1. $\sum_{i=1}^{m} \sigma_{ii} \rho^2(y_k, x_i) = \lambda_k (\sum_{i=1}^{m} \alpha_{ik}^2) = \lambda_k.$
  2. $\sum_{k=1}^{m} \rho^2(y_k, x_i) = 1.$

$\mathbf{x}$的前k个主成分 $y_1, \cdots, y_k$的对原有变量x_i的贡献率 $\nu_{1\to k}(i)$定义为

$$\nu_{1\to k}(i) = \sum_{j=1}^{k} \rho^2(y_j, x_i) = \sum_{j=1}^{k} \frac{\lambda_j \alpha_{ij}^2}{\sigma_{ii}}$$

概要

3 总体主成分分析
- 标准化随机变量
- 方差贡献率
- 因子负荷量

4 样本主成分分析

- 设$\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$是对$m$维随机向量$\mathbf{x} = (x_1, x_2, \cdots, x_m)^T$进行$n$次独立观测的样本，其中$\mathbf{x}_j = (x_{1j}, x_{2j}, \cdots, x_{mj})^T$表示第$j$个观测样本，则观测数据矩阵

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n] = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \vdots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix}$$

- 样本均值向量为$\bar{\mathbf{x}} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j = (\bar{x}_1, \cdots, \bar{x}_m)^T$.
- 样本协方差矩阵为$S = [s_{ij}]_{m\times m}$，其中$s_{ij} = \frac{1}{n-1}\sum_{k=1}^{n}(x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j), i, j = 1, 2, \cdots, m$.
- 样本相关矩阵为 $R = [r_{ij}]_{m\times m}$，其中 $r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}}$.

- 定义m维随机向量$\mathbf{x}=(x_1,x_2,\cdots,x_m)^T$ 到 $\mathbf{y}=(y_1,y_2,\cdots,y_m)^T$ 的线性变换

$$\mathbf{y}=A^T\mathbf{x},$$

其中 $A=[\alpha_1,\alpha_2,\cdots,\alpha_m]=\begin{bmatrix}\alpha_{11} & \cdots & \alpha_{1m}\\ \vdots & \vdots & \vdots\\ \alpha_{m1} & \cdots & \alpha_{mm}\end{bmatrix}.$

- 对每个分量来说，$y_i=\alpha_i^T\mathbf{x}$.
- 对每个观测数据$\mathbf{x}_j$来说，$\mathbf{y}_j=A^T\mathbf{x}_j$.
- y_i对应于$\mathbf{X}$的样本均值$\bar{y}_i=\frac{1}{n}\sum_{j=1}^{n}\alpha_i^T\mathbf{x}_j=\alpha_i^T\bar{\mathbf{x}}$.

- y_i对应于$\mathbf{X}$的样本方差$\mathrm{Var}(y_i) = \alpha_i^T S \alpha_i$.
- y_i, y_j对应于$\mathbf{X}$的样本协方差$\mathrm{Cov}(y_i, y_j) = \alpha_i^T S \alpha_j$.

- 我们首先对数据进行规范化:

$$x'_{ik} = \frac{x_{ik} - \bar{x}_i}{\sqrt{s_{ii}}}, \quad i = 1, 2, \cdots, m; k = 1, 2, \cdots, n.$$

- 我们仍以x_{ik}表示规范化的x'_{ik}, 并将规范的样本矩阵仍然记为$\mathbf{X}$, 此时样本协方差矩阵

$$S = R = \frac{1}{n-1}\mathbf{X}\mathbf{X}^T$$

- 我们设R的特征值为$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$, α_i 为R的属于λ_i的单位特征向量, 则样本主成分变换为

$$\mathbf{y} = A^T\mathbf{x},$$

这里$A = [\alpha_1, \alpha_2, \cdots, \alpha_m]$.

- n个样本$\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$的第$j$主成分的行向量为

$$(y_{j1}, \cdots, y_{jn}) = \alpha_j^T \mathbf{X}$$

- 样本前k主成分矩阵为

$$\mathbf{Y}_{k\times n} = [\alpha_1, \cdots, \alpha_k]^T \mathbf{X}$$

基于以上的分析，我们可以

- 求出R的特征值为$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$;
- 确定使得累计方差贡献率达到预定值的主成分个数k;
- 求出前k个特征值 λ_i对应的单位特征向量 α_i;
- 求k个样本主成分
$$y_i = \alpha_i^T x$$
- 计算$\rho(y_j, x_i)$以及 $\nu_{1\to k}(i)$;
- 计算样本前k主成分矩阵.

奇异值分解与主成分分析

- 定义一个新的数据矩阵

$$\mathbf{X}' = \frac{1}{\sqrt{n-1}}\mathbf{X}^T$$

则

$$\left(\mathbf{X}'\right)^T\mathbf{X}' = \frac{1}{n-1}\mathbf{X}\mathbf{X}^T = R = S$$

- $\mathbf{X}'$的右奇异向量就是$S = \left(\mathbf{X}'\right)^T\mathbf{X}'$的单位特征向量.

主成分分析算法

输入： 规范化的样本矩阵$\mathbf{X}_{m\times n}$
输出： 样本主成分矩阵$\mathbf{Y}_{k\times n}$
参数： 主成分个数k

(1) 构造

$$\mathbf{X}' = \frac{1}{\sqrt{n-1}}\mathbf{X}^T$$

(2) 求$\mathbf{X}'$的k秩截断奇异值分解

$$U\Sigma V^T$$

(3) 样本前k主成分矩阵为

$$\mathbf{Y}_{k\times n} = V^T\mathbf{X}$$

主成分分析小结

- 标准线性组合
- 总体主成分变换
- 总体主成分性质
- 方差贡献率
- 因子负荷量
- 规范化数据
- 样本主成分分析