十月算法班第二讲：线性代数与矩阵论

管老师

七月在线

Oct, 2016

主要内容

线性代数的描述性理论

- 线性空间与基
  - 举例说明线性空间的基本概念
- 线性映射与矩阵
  - 什么是矩阵?
  - 矩阵作为线性映射的代数表达方式
  - 线性方程的几何意义
- 线性回归
  - 线性回归作为方程求解问题
  - 线性回归作为几何逼近问题
  - 最小二乘法

主要内容

线性代数的分类理论

- 矩阵标准型
  - 把矩阵看做线性映射：相似变换
  - 把矩阵看做度量：相合变换
  - 正交相似变换
  - 奇异值分解
- 应用选讲
  - 主成分分析
  - SVD 在推荐系统中的应用
- 参考材料与作业

记号

- 本节课常用数学记号

V, W	向量空间
v, w	向量
$\mathbb{R}^n, \mathbb{R}^m$	实坐标空间
α, β	V 和 W 的基
$T : V \to W$	向量空间 V 到 W 的线性映射
$A_{\alpha,\beta}(T)$	线性映射 T 在 α 和 β 这两组基下的矩阵
$G(v_1, v_2)$	内积空间 V 上的内积
H_α	G 在基 α 下的矩阵形式

线性空间与基

实系数线性空间是一个由向量组成的集合，向量之间可以做加减法，向量与实数之间可以做乘法，而且这些加，减，乘运算要求满足常见的交换律和结合律. 我们也可以类似地定义其他系数的线性空间.

Example (线性空间)

有原点的平面。

- 如果平面有一个原点 O, 那么平面上任何一个点 P, 都对应着一个向量 $\overrightarrow{OP}$。
- 这些向量以及他们的运算结构放在一起，就组成一个向量空间。

- 原点 O 在空间中引入了线性结构。(向量之间的加法，以及向量与实数的乘法)

线性空间与基

基是线性空间里的一组向量，使得任何一个向量都可以唯一的表示成这组基的线性组合.

Example (坐标空间)

有原点的平面，加上一组基 $\{\overrightarrow{X}, \overrightarrow{Y}\}$。

- 任何一个向量 $\overrightarrow{OP}$, 都可以唯一表达成 $\overrightarrow{OP} = a\overrightarrow{X} + b\overrightarrow{Y}$ 的形式。
- (a, b) 就是 P 点的坐标。

- 基给出了定量描述线性结构的方法——坐标系。

线性空间与基

Example (坐标系的选择)

考虑纽约中城区的地图。街道用数字编号，但不是正南正北。在这个地图上以中心为原点，如何选取基?

- 基的选择取决于要解决的问题。
- 没有十全十美的基，只有适合解决问题的基。

小结 (线性空间与基)

- 线性空间是一种结构 *(加法及乘法运算结构)*
- 基使得我们可以用坐标描述线性结构
- 基的选择取决于要研究的问题

线性映射与矩阵

Definition (线性映射)

V 和 W 是两个实线性空间，$T : V \to W$ 如果满足如下条件就是一个线性映射。

$$
\begin{array}{lll}
(i) & T(v_1 + v_2) = T(v_1) + T(v_2), & \forall v_1, v_2 \in V \\
(ii) & T(\lambda v) = \lambda T(v), & \forall \lambda \in \mathbb{R}, v \in V
\end{array}
$$

- 线性映射的本质就是保持线性结构的映射
- 到自身的线性映射 $T : V \to V$ 叫做线性变换

线性变换的矩阵描述

V, W 分别为 n, m 维的线性空间，
$\alpha = \{\alpha_1, ..., \alpha_n\}, \beta = \{\beta_1, ..., \beta_m\}$ 分别为 V, W 的一组基。
$T : V \rightarrow W$ 是一个线性映射。于是 T, α, β 唯一决定一个矩阵 $A_{\alpha,\beta}(T) = [A_{ij}]_{m \times n}$，使得

$$T(\alpha_j) = \sum_{i=1}^{m} A_{ij} * \beta_i, \forall j \in 1, ..., n \tag{1}$$

(1) 等价于

$$T(\alpha_1, ..., \alpha_n) = (\beta_1, ..., \beta_m) \cdot A_{\alpha,\beta}(T) \tag{2}$$

简记为

$$T(\alpha) = \beta \cdot A_{\alpha,\beta}(T) \tag{3}$$

如果我们选取 V, W 的另外一组基，$\tilde{\alpha} = \alpha \cdot P$, $\tilde{\beta} = \beta \cdot Q$. 那么存在矩阵 $A_{\tilde{\alpha},\tilde{\beta}}(T)$ 使得,

$$T(\tilde{\alpha}) = \tilde{\beta} \cdot A_{\tilde{\alpha},\tilde{\beta}}(T)$$

两边分别代入 $\tilde{\alpha}$ 与 $\tilde{\beta}$ 得到,

$$T(\alpha) \cdot P = T(\alpha \cdot P) = \beta \cdot Q \cdot A_{\tilde{\alpha},\tilde{\beta}}(T)$$

与(3)比较我们得到矩阵变换公式:

$$Q \cdot A_{\tilde{\alpha},\tilde{\beta}}(T) \cdot P^{-1} = A_{\alpha,\beta}(T) \tag{4}$$

线性映射与矩阵

小结 (线性映射与矩阵)

- 矩阵是线性映射在特定基下的一种定量描述

$$T(\alpha) = \beta \cdot A_{\alpha,\beta}(T)$$

- 基变换下的矩阵变换公式的推导方法

$$Q \cdot A_{\tilde{\alpha},\tilde{\beta}}(T) \cdot P^{-1} = A_{\alpha,\beta}(T)$$

例题 (几何变换：拉伸, 反转与旋转)

实数平面 $\mathbb{R}^2$ 到自身的映射：

- 拉伸 $T\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1/2 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$
- 反转 $T\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$
- 旋转 $T\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \cos(\pi/6) & -\sin(\pi/6) \\ \sin(\pi/6) & \cos(\pi/6) \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

Example (算法问题：如何计算斐波那契数列)

斐波那契数列：$a_1 = 1, a_2 = 1$，并且满足 $a_{n+2} = a_n + a_{n+1}$.

Solution (1. 简单递归)

我们可以使用定义进行简单递归

```
def arrayComputer(n):
    if n <= 2:
        return 1
    else:
        prevA = arrayComputer(n-1)
        prevprevA = arrayComputer(n-2)
        return prevA + prevprevA
```

Solution (2. 建立线性模型)

把 $[a_n, a_{n+1}]^T$ 看成一个向量，于是递归公式变成一个线性模型：

$$\begin{bmatrix} a_{n+1} \\ a_{n+2} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} a_n \\ a_{n+1} \end{bmatrix}$$

于是：

$$\begin{bmatrix} a_n \\ a_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}^{n-1} \cdot \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

线性模型算法的 Python code:

```
def linearComputer(n):
    if n <= 2:
        return 1
    current2A = [1, 1]
    prev2A = copy.deepcopy(current2A)
    for ind in range(n-2):
        current2A[0] = 0*prev2A[0] + 1*prev2A[1]
        current2A[1] = 1*prev2A[0] + 1*prev2A[1]
        prev2A = copy.deepcopy(current2A)
    return prev2A[1]
```

思考题

如何计算类似的数列 $a_1 = 1, a_2 = 1, a_3 = 1$, 并且 $a_{n+3} = a_n + 2a_{n+1} + 3a_{n+2}$?

线性回归

线性回归

- 模型:$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon$, 余项 ϵ 服从正态分布 $N(0, \sigma)$
- 数据: 样本 $(Y_i, X_{i1}, \cdots, X_{ik})$,$1 \leq i \leq n$
- 目的: 估计参数 $\beta_1, \cdots, \beta_k$

线性回归: 矩阵模型

可以把整个样本用矩阵表示。令 $x_0 = 1$, 原模型变成

$$y = \beta_0 x_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon$$

用矩阵 Y, X, β 定义如下

$$\begin{aligned} Y &= [Y_1, \cdots, Y_n]^T \\ X &= [X_{ij}]_{1 \le i \le n, 0 \le j \le k} \\ \beta &= [\beta_0, \cdots, \beta_k]^T \end{aligned}$$

于是整个模型写成:

$$Y = X \cdot \beta + \epsilon$$

线性回归: 线性方程（代数）

如果我们省略掉最后一个误差项，问题变为解线性方程的问题

$$X \cdot \beta = Y \tag{5}$$

一般来讲，样本个数大于自参数个数。所以方程个数大于这个方程的未知数个数，于是方程通常是没有解，长方形矩阵也一定没有逆矩阵。但是如果 $X^T X$ 是可逆矩阵 (一般是满足的)，那么代数上可以用如下方法求一个近似的解答。

$$X^T X \cdot \beta = X^T Y$$
$$\beta = (X^T X)^{-1} X^T Y$$

所以如若(5)有解，就一定是这个 $\beta = (X^T X)^{-1} X^T Y$. 而如果(5)没有解，这个 β 也是一个合理的估计.

线性回归: 几何逼近（几何）

我们从线性映射的角度重新来审视这个方程

$$X \cdot \beta = Y \qquad (6)$$

等式的左边是矩阵 X 的列向量的一个线性组合，所以这个方程的几何意义是希望把 Y 当成 X 列空间中的一个点，然后求出这个点在列向量这组基下的坐标.

——————————但是!! ——————————

Y 不见得是 X 的列空间中的点啊，怎么办? 几何学家的想法是找到 Y 在 X 列空间里的投影 Y^*. 然后用 Y^* 在列空间上的坐标来估计 β。换句话说希望找到 β 使得 $X\beta - Y$ 与 X 垂直，也就是说 $X^T \cdot (X\beta - Y) = 0$, 于是乎

$$X^T X\beta = X^T Y$$
$$\beta = (X^T X)^{-1} X^T Y$$

线性回归: 最小二乘（统计）

统计学家采取更加简单直接的想法，线性模型最终是用来做预测的，那么预测的准确程度才是这个模型优良的最终度量. 所以统计学家决定最小化误差项 $Y - X\beta$. 也就是 $(Y - X\beta)^T \cdot (Y - X\beta)$. 这是一个关于 β 的二次型，关于二次型的更加深入内容我们下次来讲解，不过这里我们先用它来解决一下线性回归的最小二乘方法。为了求极值，我们对 β 求导数（梯度）。

$$\nabla (Y - X\beta)^T \cdot (Y - X\beta) = -2X^T Y + 2X^T X\beta$$

极值条件为 $\nabla (Y - X\beta)^T \cdot (Y - X\beta) = 0$, 于是我们得到方程

$$-2X^T Y + 2X^T X = 0$$

再一次我们得到

$$\beta = (X^T X)^{-1} X^T Y$$

线性回归: 极大似然估计（统计）

统计学家也可以采取极大似然估计的方法[1]. 将模型转化为概率分布的参数估计问题.y 服从 $N(\sum_{j=o}^{k} x_j\beta_j,\sigma)$，利用我们的样本进行极大似然估计. 似然函数为

$$L(\beta)=\prod_{i=1}^{n} exp(\frac{(\sum_{j=0}^{k} x_j\beta_j-y)^2}{2\sigma^2})$$

$$\begin{aligned} l(\beta) &= \sum_{i=1}^{n}\frac{(\sum_{j=0}^{k} x_j\beta_j-y)^2}{2\sigma^2} \\ &= \frac{1}{2\sigma^2}(Y-X\beta)^T(Y-X\beta) \end{aligned}$$

于是得到了和前面最小二乘法一样的优化函数. 值得注意的是，这种方法更具备一般性，即使不是线性模型也可以使用.

---

[1]感谢 @Zhui 同学提供的材料: 程序员的数学 2：概率统计，平冈和幸堀玄(著)，陈筱烟 (译)

矩阵的标准型: 概述

矩阵的变换

- 标准型用来表示矩阵在变换下不变的性质
- 矩阵变换本质上是基的转换
- 相似变换：线性映射
- 相合变换：二次型（度量）

矩阵的标准型: 相似变换 (把矩阵看做线性映射)

如果 $T: V \to V$ 是一个线性变换, 那么对于 V 的两组基 α 与 $\tilde{\alpha} = \alpha \cdot P$，线性变换 T 的矩阵分别为

$$A_\alpha(T) \text{ and } A_{\tilde{\alpha}}(T) = P^{-1} \cdot A_\alpha(T) \cdot P$$

方阵的相似变换

- 如果两个方阵 A 和 $\tilde{A}$ 满足，$\tilde{A} = P^{-1}AP$. 那么这两个方阵就互为相似矩阵
- 相似矩阵的几何意义是同一个线性变换在不同的基下的表达形式
- 当研究对象是线性变换的时候，我们只关心矩阵在相似变换下不变的几何性质。

矩阵的标准型: 相似变换 (把矩阵看做线性映射)

矩阵的标准型: 相似变换 (把矩阵看做线性映射)

相似变换下不变的性质

- 行列式 (det)

$$\begin{aligned}\det(P^{-1}AP) &= \det(P^{-1})\det(A)\det(P)\\ &= \det(P^{-1})\det(P)\det(A)\\ &= \det(A)\end{aligned}$$

- 迹 (trace),$\mathrm{tr}(AB) = \mathrm{tr}(BA)$

$$\mathrm{tr}(P^{-1}AP) = \mathrm{tr}(APP^{-1}) = \mathrm{tr}(A\cdot I) = \mathrm{tr}(A)$$

- 秩 (rank)

矩阵的标准型: 相似不变量

相似变换下不变的性质

- 特征值：特征方程 $\det(A-\lambda I)=0$ 的根。
  如果 $\det(A-\lambda I)=0$，那么 $\det(P^{-1}(A-\lambda I)P)=0$, 于是 $\det(P^{-1}AP-\lambda I)=0$

- 特征值是最重要的相似不变量，利用这个相似不变量可以方便的得出上面所有的不变量。

矩阵的标准型：相似不变量

Theorem (矩阵的相似标准型)

任何一个实系数方阵 A, 都存在一个可逆实系数方阵 P, 使得 $P^{-1}AP$ 是一个分块对角矩阵 *diag*$(J_1, ..., J_k)$. 且每一个对角块 *(约当块)*J_k 是如下四种情况之一：

- i 1×1 伸缩变换矩阵块 $\left[\lambda\right]$
- ii 2×2 伸缩旋转变换矩阵块 $R_{(\mu,\theta)} = \mu \cdot \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}$

矩阵的标准型：相似标准型

iii $m \times m$ 循环伸缩矩阵块 $\begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \lambda \end{bmatrix}$

iv $m \times m$ 循环伸缩旋转矩阵块

$$\begin{bmatrix} R_{(\mu,\theta)} & I_{2\times 2} & 0 & \cdots & 0 \\ 0 & R_{(\mu,\theta)} & I_{2\times 2} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & R_{(\mu,\theta)} \end{bmatrix}$$

如果对角块中只有 [i] 和 [ii] 两种，那么这个矩阵称作可复对角化矩阵.

Proof.

I have discovered a truly marvellous proof of this, which this margin is too narrow to contain.– Fermat □

Theorem (几乎所有方阵都可复对角化)

对于任何一个实系数方阵 A, 都存在一个可复对角化的矩阵序列 $\{A_i\}_{i=1}^{\infty}$, 使得

$$\lim_{i\to\infty} A_i = A$$

Proof.

证明思路:

1. 如果一个矩阵的特征多项式没有重跟，那么这个矩阵一定可以复对角化.
2. 几乎所有的矩阵特征多项式都没有重跟.

□

例题 (行列式 (det)：线性映射的体积膨胀系数)

如果 A 是线性变换 $T: V \to V$ 的矩阵，C 是 V 里边的立方体，那么：

$$\mathit{Volume}(T(C)) = \det(A) \cdot \mathit{Volume}(C)$$

Proof

因为行列式是相似不变量，不失一般性，可以假定 A 就是约当标准型.
证明分两步：先对可以复对角化的矩阵 A 进行证明，然后证明一般情况.
第一步：如果 A 是复对角化的矩阵，那么 $A = \mathrm{diag}(J_1, ..., J_k)$，其中 J_i 要么是一维的伸缩变换矩阵块，要么是二维的旋转变换矩阵块. 所以可以进一步把 A 写成

$$A = \mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_p, R_{(\mu_1,\theta_1)}, ..., R_{(\mu_q,\theta_q)})$$

于是 $\det(A) = \lambda_1 \cdot \lambda_2 \cdots \lambda_p \cdot \mu_1 \cdots \mu_q = \prod_{i=1}^{p} \lambda_i \cdot \prod_{j=1}^{q} \mu_j^2.$

Continue Proof

另一方面 $T(C)$ 是由 C 在前 p 个维度上进行拉伸，而在后面的维度上进行二维拉伸及旋转得到的. 所以

$$\text{Volume}(T(C)) = \prod_{i=1}^{p} \lambda_i \cdot \prod_{j=1}^{q} \mu_j^2 \cdot \text{Volume}(C)$$

所以

$$\text{Volume}(T(C)) = \det(A) \cdot \text{Volume}(C)$$

Continue Proof

第二步：对一般的 A，存在一个可复对角化的矩阵序列 $\{A_i\}_{i=1}^{\infty}$，及其所对应的线性变换序列 $\{T_i\}_{i=1}^{\infty}$，使得：

$$\lim_{i\to\infty} A_i = A \text{ 而且 } \lim_{i\to\infty} T_i = T$$

根据第一步的结论，对于任何 i 我们有，

$$\mathsf{Volume}(T_i(C)) = \det(A_i)\cdot \mathsf{Volume}(C).$$

因为体积与行列式都是连续函数，我们得到

$$\begin{aligned}\mathsf{Volume}(T(C)) &= \lim_{i\to\infty} \mathsf{Volume}(T_i(C)) \\ &= \lim_{i\to\infty} \det(A_i)\cdot \mathsf{Volume}(C) \\ &= \det(A)\cdot \mathsf{Volume}(C)\end{aligned}$$

于是证毕.

例题 (迹 (tr): exp(tr) 是线性映射 $exp(A)$ 的体积膨胀系数)

$$\exp(tr(A)) = \det(\exp(A))$$

Proof

首先对可复对角化的矩阵进行证明. 为了简化步骤我们不加证明的使用

Proposition

如果 A 是可复对角化矩阵，则存在一个复系数可逆方阵 Q, 使得 $\tilde{A} = Q^{-1}AQ$ 是复系数对角矩阵.

$$\tilde{A} = diag(\lambda_1, ..., \lambda_n)$$

Continue Proof

于是 $\exp(\tilde{A}) = \mathsf{diag}(\exp(\lambda_1), ..., \exp(\lambda_n))$

$$\begin{aligned}
\exp(\mathsf{tr}(A)) &= \exp(\mathsf{tr}(\tilde{A})) \\
&= \exp(\sum_{i=1}^{n} \lambda_i) \\
&= \prod_{i=1}^{n} \exp(\lambda_i) \\
&= \det(\tilde{A}) \\
&= \det(A)
\end{aligned}$$

对于一般的 A, 可以模仿上一个证明中的第二步，留作思考题。

例题 (秩 (rank)：像空间的维数)

如果 A 是线性变换 $T: V \to V$ 在基 α 下的矩阵，那么

$$\mathit{rank}(A) = \dim T(V)$$

Proof

注意矩阵的秩并不是一个连续函数，所以对于这个问题我们不能仅仅考虑可复对角化的矩阵。我们必须直接考虑一般情况。
假定 $A = \mathrm{diag}(J_1, ..., J_K)$, 其中每一个对角块都是约当标准型四中对角块里面的一种。于是我们得到空间 V 以及变换 T 的直和分解

$$V = V_1 \oplus \cdots \oplus V_k$$
$$T_i = T|_{V_i} : V_i \to V_i, \quad \text{for all } i \text{ from } 1 \text{ to } K$$

矩阵的标准型：相似不变量的几何意义

Continue Proof

所以

$$\dim T(V) = \sum_{i=1}^{K} \dim T_i(V_i)$$

而且

$$\mathsf{rank}(A) = \sum_{i=1}^{K} \mathsf{rank}(J_i)$$

所以要证明

$$\mathsf{rank}(A) = \dim T(V)$$

只需要对所有的约当块 J_i, 证明

$$\mathsf{rank}(J_i) = \dim T_i(V_i)$$

对于每一个约当块的证明留作思考题。

矩阵的标准型

小结 (矩阵的标准型)

- 任何一个方阵总存在约当标准型（线性变换的相关问题总是存在一组好基）
- 如果问题具有相似不变性，可以假定矩阵是约当标准型从而简化问题
- 如果问题具有相似不变性以及连续性，可以假定矩阵是可复对角化的约当标准型，从而进一步简化问题

矩阵的标准型: 相合变换 (二次型)

假设 V 是一个实系数线性空间，那么线性空间上的度量指的是空间中向量的内积关系 $G(v_1, v_2)$. 如果 $\alpha\{\alpha_1, \cdots, \alpha_k\}$ 是空间 V 的一组基，那么这个内积一般可以用一个对称矩阵
$H_\alpha = [h_{ij}]_{n\times n}$ 来表示.

$$h_{ij} = G(\alpha_i, \alpha_j)$$

这时候对于任意两个向量 v_1, v_2, 如果 $v_1 = \alpha \cdot x_1, v_2 = \alpha \cdot x_2$，那么

$$G(v_1, v_2) = x_1^T H_\alpha x_2$$

矩阵的标准型: 相合变换 (二次型)

方阵的相合变换

- 如果两个对称方阵 A 和 $\tilde{A}$ 满足，$\tilde{A} = P^T AP$. 那么这两个方阵就互为相合矩阵
- 相似矩阵的几何意义是同一个内积结构在不同基下的表示形式

矩阵的标准型: 相合变换 (二次型)

相合不变量

- 矩阵的正定性（正定，负定）
- 矩阵的正负特征值的个数（Signature）
- 相合变换下矩阵保持对称性

我们涉及到的不定矩阵不多，所以关于相合不变量只需做大概了解

矩阵的标准型: 相合变换 (二次型)

方阵的正交相似变换

正交相似变换同时满足相似与相合变换的条件，也就是说它同时保持了矩阵的相似与相合不变量。

- 如果两个对称方阵 A 和 $\tilde{A}$ 满足，$\tilde{A} = P^T AP$. 而且 P 是正交矩阵:$P^T = P^{-1}$. 那么这 A 与 $\tilde{A}$ 就互为正交相似.

矩阵的标准型: 相合变换 (二次型)

矩阵的标准型: 相合变换 (二次型)

方阵的正交相似标准型

任何一个对称矩阵 A 都可以正交相似于一个对角矩阵 D.

总存在一个正交矩阵 P 使得，$A = P^T D P$.

方阵的正交相似标准型的几何意义

矩阵的标准型: 相合变换 (二次型)

主成分分析 (PCA)

PCA 的主要目的是降维，也可以起到分类的作用

- 当数据维度很大的时候，如果相信大部分变量之间存在线性关系，那么我们就希望降低维数，用较少的变量来抓住大部分的信息.
- 一般来讲做 PCA 之前要做 normalization 使得变量中心为 0，而且方差为 1.

比较广泛应用于图像识别，文档处理，推荐系统

矩阵的标准型: 相合变换 (二次型)

主成分分析 (PCA)

- 首先计算变量之间的协方差矩阵 Σ(利用样本)
- 找到 Σ 的正交相似标准型

正交相似标准性的求解由计算机完成，我们主要关心他的几何意义

矩阵的标准型:PCA 例子

推荐系统

如果一个旅游网站里面有 10000000 个注册用户，以及 40000 个注册酒店. 网站有用户通过本网站点击酒店页面的记录信息.$A = [A_{ij}]_{10000000\times 40000}$,$A_{ij}$ 表示第 i 个用户点击 j 酒店的次数.

- 如何评价酒店之间的相似度?
- 给定一个酒店，请找出与它最相似的其他几个酒店?
- 如果要给酒店分类，有什么办法?

矩阵的标准型:PCA 例子

长方矩阵的奇异值分解 (SVD)

对于任何一个矩阵 $B_{m\times n}$, 存在正交矩阵 $P_{m\times m}$,$Q_{n\times n}$. 使得

$$B = PDQ$$

其中 $D_{m\times n}$ 是一个只有对角元素不为零的矩阵.

矩阵的标准型: 相合变换 (二次型)

证明

考虑 B^TB 与 BB^T 这两个对称矩阵

- B^TB 与 BB^T 拥有相同的特征多项式，所以拥有几乎相同的正交相似标准型
- $P_1^TB^TBP_1 = D_P$ 是 B^TB 的标准型
- $Q_1^TBB^TQ_1 = D_Q$ 是 BB^T 的标准型

那么考虑 $\tilde{B} = Q_1^TBP_1$, 我们知道

$$\tilde{B}^T\tilde{B} = P_1^TB^TQ_1Q_1^TBP_1 = P_1^TB^TBP_1 = D_P$$

另一方面

$$\tilde{B}\tilde{B}^T = Q_2^TBP_1P_1^TB^TQ_1 = Q_1^TBB^TQ_1 = D_Q$$

矩阵的标准型: 相合变换 (二次型)

继续证明

如果 $\tilde{B}$ 列数多于行数 $(m > n)$, 那么 $\tilde{B} = [B_1, O_{(n-m)\times m}]$. 而 B_1 的列向量彼此正交，长度平方为 D_Q 的对角元素. 于是存在正交矩阵 Q_2 使得 $Q_2^T B_1 = \sqrt{D_Q}$.

令 $D = Q_2^T \tilde{B} = Q_2^T Q_1^T B P_1$. 那么因为 $Q_2^T Q_1^T = (Q_1 Q_2)^T$ 仍然是正交矩阵，所以原命题得证.

矩阵的标准型:SVD 例子

推荐系统

如果一个旅游网站里面有 10000000 个注册用户，以及 40000 个注册酒店. 网站有用户通过本网站点击酒店页面的记录信息.$A = [A_{ij}]_{10000000\times 40000}$,$A_{ij}$ 表示第 i 个用户点击 j 酒店的次数.

- 如何评价用户的相似度?
- 给定一个用户的访问历史，请问他下一次最可能访问的酒店是哪一家?

矩阵的标准型:PCA 例子

线性代数：参考资料

参考资料

- 陶哲轩的讲义
  http://www.math.ucla.edu/~tao/resource/general/115a.3.02f/
- 入门教材：线性代数及其应用, 莱 (Lay D.C.) (作者), 刘深泉 (译者)
- 艰深教材：线性代数，李炯生、查建国

作业

- 入门教材：线性代数及其应用, 莱 (Lay D.C.) (作者), 刘深泉 (译者)
  p69:8,13,15,16; p79:7,14
  p185:18,30; p186:9,14,15
  p242:6,17,18; p292:5,6
  p293:19,20,22,25,26; p99:25,26,27,35
  p413:9,11; p423:17,18,21
  p431:1

谢谢大家!