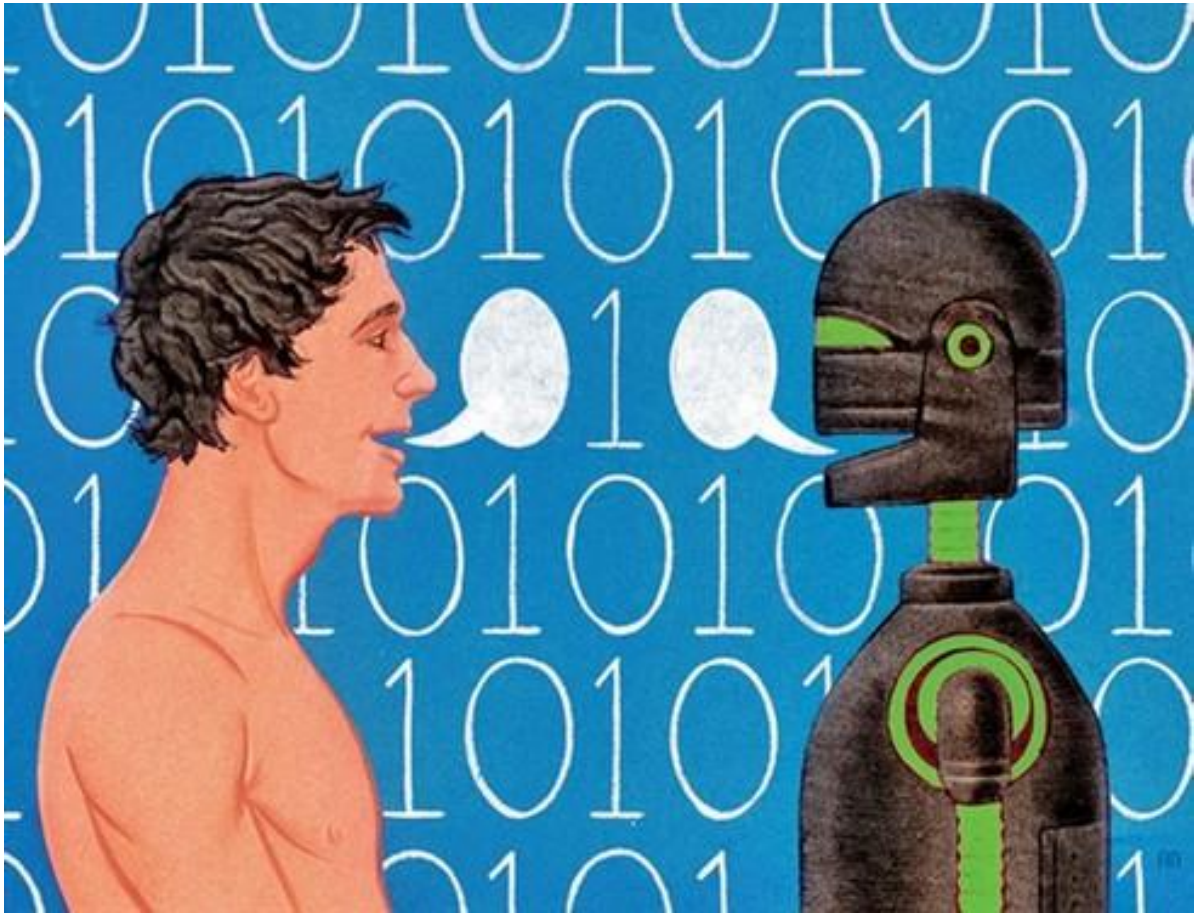


大数据的矩阵计算基础——第12周

DATAGURU专业数据分析社区

法律声明

【声明】本视频和幻灯片为炼数成金网络课程的教学资料，所有资料只能在课程内使用，不得在课程以外范围散播，违者将可能被追究法律和经济责任。

关注炼数成金企业微信

■提供全面的数据价值资讯，涵盖商业智能与数据分析、大数据、企业信息化、数字化技术等，各种高性价比课程信息，赶紧掏出您的手机关注吧！

本周内容

- 矩阵技术在机器学习中的应用
  - 广义逆矩阵与多元线性回归
  - 奇异值分解与主成分分析
  - 因子分析

Moore-Penrose广义逆

- 对于任意复数m*n阶矩阵A，如果存在n*m阶复矩阵G，满足
  - 1. AGA=A
  - 2. GAG=G
  - 3. $(AG)^H = AG$
  - 4. $(GA)^H = GA$
- 称G为A的一个Moore-Penrose广义逆，上述四个方程称为M-P方程
- 若G满足M-P方程的全部或其中一部分，则称G为A的广义逆

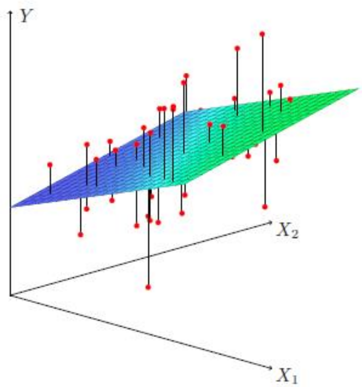
极小范数最小二乘解

- 不一致方程AX=b的极小范数最小二乘解为$X = A^{+}b$
- 对于线性方程组AX=b，其解为$X = A^{+}b$
- 在多元统计中的应用：求解多元线性回归方程的参数

多元线性回归模型

- 当Y值的影响因素不唯一时，采用多元线性回归模型

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots + \beta_m X_m + \varepsilon$$



- 例如商品的销售额可能与电视广告投入，收音机广告投入，报纸
以有

$$\text{sales} = \beta_0 + \beta_1 \times \text{TV} + \beta_2 \times \text{radio} + \beta_m \times \text{newspaper} + \varepsilon$$

参数估计

- 最小二乘法：
- 与一元回归方程的算法相似
- $\mathrm{RSS} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ 是关于βi的函数。分别对βi求偏导并令偏导等于0，可以解出相应的βi的值
- $\hat{\beta} = (X^T X)^{-1} X^T Y$

广义逆在多元线性回归中的应用

- 实际上，$Y = X\beta$是一个非一致方程，可通过直接求X的加号逆得到β的最小范数最小二乘解

- 对于$X^T X$不可逆的情况同样适用

例子

- 对于数据

$$y = 1,4,3,2,1; x_1 = 1,2,1,4,3;$$
$$x_2 = 1,-1,2,-6,-1; \quad x_3 = 2,2,0,4,-2$$

- X= $\begin{pmatrix} 1 & 1 & 1 & 2 \\ 1 & 2 & -1 & 2 \\ 1 & 1 & 2 & 0 \\ 1 & 4 & -6 & 4 \\ 1 & 3 & -1 & -2 \end{pmatrix}$; $y=(1 \quad 4 \quad 3 \quad 2 \quad 1)^T$

例子

◆ 由于$X^T X$不可逆，直接求出X的加号逆为

$$\begin{bmatrix} 0.0701 & 0.0483 & 0.0837 & -0.0081 & 0.0538 \\ 0.0458 & 0.0537 & 0.0959 & 0.0194 & 0.1617 \\ 0.1247 & 0.0508 & 0.1288 & -0.1011 & -0.0149 \\ 0.1282 & 0.0703 & 0.0285 & 0.0542 & -0.1871 \end{bmatrix}$$

◆ 从而根据$\beta = X^+ Y$，得

$$y = 0.552 + 0.7488x_1 + 0.4972x_2 + 0.4162x_3$$

主成分分析PCA

- 对于p维随机变量$X=(x_1,x_2,\ldots\ldots,x_p)^T$,数学期望$E(X)=\boldsymbol{\mu}$,

协方差矩阵$V(X)=\Sigma=\frac{1}{N}(X-\mu)(X-\mu)^T$

- 考虑这样的线性变换

$$\begin{cases} Z_1=a_1^TX \\ Z_2=a_2^TX \\ \quad\vdots \\ Z_p=a_p^TX \end{cases},$$

- 显然

$$\begin{aligned} \mathrm{Var}(Z_i) &= a_i^T\Sigma a_i, \quad i=1,2,\cdots,p, \\ \mathrm{Cov}(Z_i,Z_j) &= a_i^T\Sigma a_j, \quad i,j=1,2,\cdots,p, \quad i\neq j. \end{aligned}$$

主成分分析PCA

- 我们希望寻找合适的a_1使得Z_1方差最大，即a_1是约束优化问题

$$\begin{array}{ll} \max & a^T\Sigma a \\ \text{s.t.} & a^Ta = 1 \end{array}$$

的解 → 二次型的条件优化问题

$$m = \min\{\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x} : \|\boldsymbol{x}\| = 1\}\ ,\ M = \max\{\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x} : \|\boldsymbol{x}\| = 1\} \qquad (2)$$

定理 6 设 A 是对称矩阵，且 m 和 M 的定义如(2)式所示，那么 M 是 A 的最大特征值 λ_1，m 是 A 的最小特征值，如果 $\boldsymbol{x}$ 是对应 M 的单位特征向量 $\boldsymbol{u}_1$，那么 $\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x}$ 的值等于 M，如果 $\boldsymbol{x}$ 是对应 m 的单位特征向量，$\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x}$ 的值等于 m.

故a_1是协方差矩阵最大的特征值对应的单位特征向量。称$Z_1 = a_1^TX$为第一主成分。

主成分分析PCA

- 类似的，可以求出与第一主成分正交的第二主成分

定理 7 设 A, λ_1 和 u_1 如定理 6 所示. 在如下条件限制下

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}=1,\ \boldsymbol{x}^{\mathrm{T}}\boldsymbol{u}_1=0$$

$\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x}$ 的最大值是第二大特征值 λ_2，且这个最大值，可以在 $\boldsymbol{x}$ 是对应 λ_2 的特征向量 $\boldsymbol{u}_2$ 处达到.

- 还有第三主成分，第四主成分……

定理 8 设 A 是一个 $n\times n$ 对称矩阵，且其正交对角化为 $A=PDP^{-1}$，将对角矩阵 D 上的元素重新排列，使得 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$，且 P 的列是其对应的单位特征向量 $\boldsymbol{u}_1,\cdots,\boldsymbol{u}_n$. 那么对 $k=2,\cdots,n$ 时，在以下限制条件下

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}=1,\ \boldsymbol{x}^{\mathrm{T}}\boldsymbol{u}_1=0,\ \cdots,\boldsymbol{x}^{\mathrm{T}}\boldsymbol{u}_{k-1}=0$$

$\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x}$ 的最大值是特征值 λ_k，且这个最大值在 $\boldsymbol{x}=\boldsymbol{u}_k$ 处可以达到.

主成分分析PCA

- 对于变量代换$Z = Q^T X$
- 协方差矩阵可正交对角化

$$Q^T \Sigma Q = \Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p \end{bmatrix}, \qquad (9.4)$$

且 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$. 则矩阵 Q 的第 i 列就对应于 a_i, 相应的 Z_i 为第 i 主成分.

SVD与PCA

- 实际应用中，SVD分解是PCA的主要工具
- SVD分解的迭代计算比特征值分解更快更准确

- 若B是中心化后的p*n阶观测矩阵，$A=\frac{1}{\sqrt{N-1}}B^T$，A的SVD分解等价于B的协方差阵特征值分解

因子分析

- 因子分析是主成分分析的推广和发展，降维方法的一种

例 9.4 为了解学生的学习能力，观测了 n 个学生的 p 个科目的成绩 (分数)，用 $X_1, X_2, \cdots, X_p$ 表示 p 个科目 (例如代数、几何、语文、英语、政治，$\cdots$)，$X_{(i)} = (x_{i1}, x_{i2}, \cdots, x_{ip})^T$，$(i = 1, 2, \cdots, n)$ 表示第 i 个学生的 p 科目的成绩．现要分析主要由哪些因素决定学生的学习能力．

$$X_i = a_i f + \varepsilon_i, \quad i = 1, 2, \cdots, p,$$

因子分析

例 9.5 *Linden* 对二次大战以来奥林匹克十项全能的得分作研究，他收集了 160 组数据，以 $X_1, X_2, \cdots, X_{10}$ 分别表示十项全能的标准得分，这里十项全能依次是： 100 米短跑、跳远、跳高、 400 米跑、 110 米跨栏、铁饼、撑杆跳高、标枪、 1500 米跑．现要分析主要由哪些因素决定十项全能的成绩，以此可用来指导运动员的选拔.

例 9.6 考察人体的五项生理指标：收缩压 (X_1) 、舒张压 (X_2) 、心跳间隔 (X_3) 、呼吸间隔 (X_4) 和舌下温度 (X_5). 从这些指标考察人体的健康状况.

因子分析

◆ 数学模型

设 $X=(X_1,X_2,\cdots,X_p)^T$ 是可观测的随机向量，且

$$E(X)=\mu=(\mu_1,\mu_2,\cdots,\mu_p)^T,\qquad \mathrm{Var}(X)=\Sigma=(\sigma_{ij})_{p\times p}.$$

因子分析的一般模型为

$$\begin{cases} X_1-\mu_1=a_{11}f_1+a_{12}f_2+\cdots+a_{1m}f_m+\varepsilon_1 \\ X_2-\mu_2=a_{21}f_1+a_{22}f_2+\cdots+a_{2m}f_m+\varepsilon_2 \\ \qquad\vdots \\ X_p-\mu_p=a_{p1}f_1+a_{p2}f_2+\cdots+a_{pm}f_m+\varepsilon_p \end{cases}$$

$$X = \mu + AF + \varepsilon,$$

$$\begin{aligned} E(F) &= 0, & \mathrm{Var}(F) &= I_m, \\ E(\varepsilon) &= 0, & \mathrm{Var}(\varepsilon) &= D = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \cdots, \sigma_p^2), \\ \mathrm{Cov}(F, \varepsilon) &= 0. \end{aligned}$$

2. 因子模型的性质

(1) Σ 的分解

$$\Sigma = AA^T + D. \tag{9.26}$$

(2) 模型不受单位的影响. 若 $X^* = CX$, 则有

$$X^* = \mu^* + A^* F^* + \varepsilon^*,$$

其中 $\mu^* = C\mu$, $A^* = CA$, $F^* = F$, $\varepsilon^* = C\varepsilon$.

(3) 因子载荷不是惟一的. 设 T 是一 m 阶正交矩阵, 令 $A^* = AT$, $F^* = T^T F$, 则模型 (9.22) 可表示为

$$X = \mu + A^* F^* + \varepsilon. \tag{9.27}$$

因子载荷矩阵不惟一对实际应用是有好处的，通常利用这一点，通过因子旋转，使得新因子有更好的实际意义.

因子分析

$\mathrm{Cov}(X, F) = A$ 或 $\mathrm{Cov}(X_i, f_i) = a_{ij}$.

令 $h_i^2 = \sum_{j=1}^{m} a_{ij}^2$, 则有

$$\sigma_{ii} = h_i^2 + \sigma_i^2, \quad i = 1, 2, \cdots, p.$$

令 $g_j^2 = \sum_{i=1}^{p} a_{ij}^2$, 则有

$$\sum_{i=1}^{p} \mathrm{Var}(X_i) = \sum_{j=1}^{m} g_j^2 + \sum_{i=1}^{p} \sigma_i^2.$$

因子分析

- 求解方法
  - 主成分法
  - 主因子法
  - 极大似然法

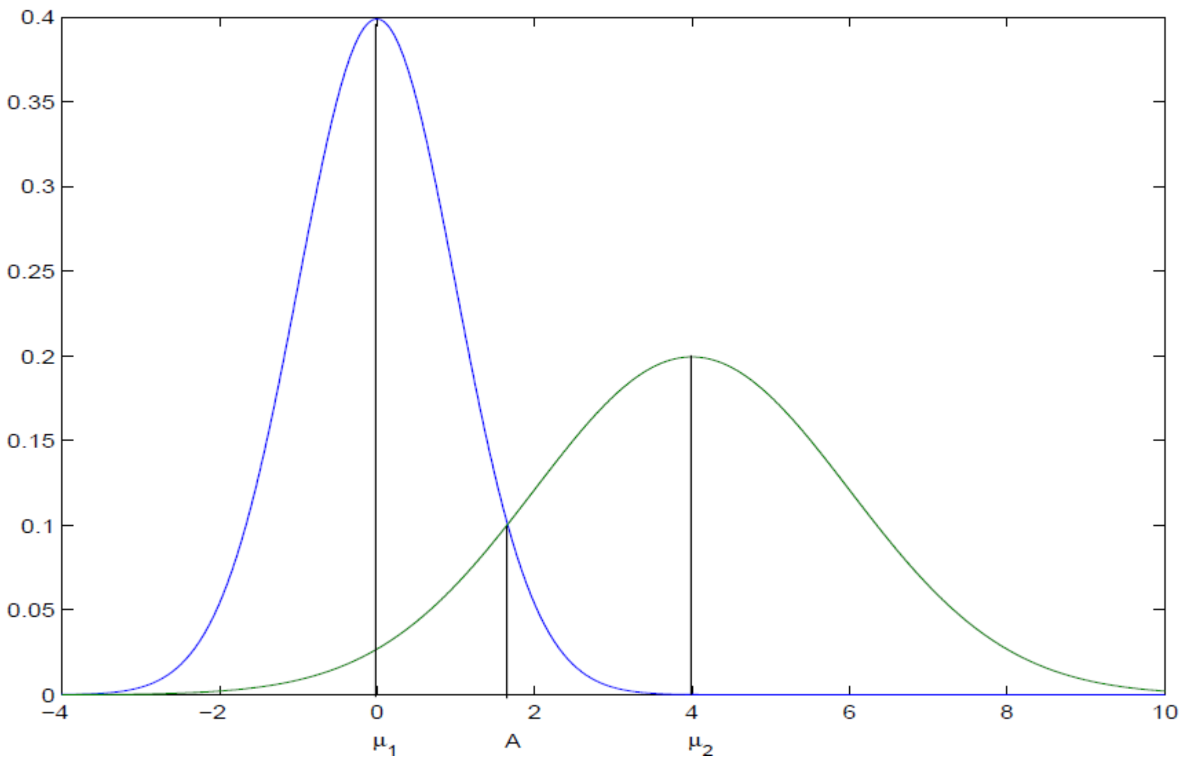
因子分析

- 由于因子载荷矩阵不是唯一，有时因子的实际意义会变得难以解释。
- 因子载荷矩阵的正交旋转
- 因子载荷方差
- 载荷值趋于1或趋于0，公共因子具有简单化的结构

判别分析

- 欧氏距离

$$d(x,y)=\|x-y\|_2=\sqrt{(x-y)^T(x-y)}.$$



判别分析

定义 8.1 设 x, y 是服从均值为 μ, 协方差阵为 Σ 的总体 X 中抽取的样本，则总体 X 内两点 x 与 y 的 *Mahalanobis* 距离（简称马氏距离）定义为

$$d(x,y)=\sqrt{(x-y)^T\Sigma^{-1}(x-y)}. \tag{8.1}$$

定义样本 x 与总体 X 的 *Mahalanobis* 距离为

$$d(x,X)=\sqrt{(x-\mu)^T\Sigma^{-1}(x-\mu)}. \tag{8.2}$$

距离判别法

- 协方差阵相等时

$$d^2(x, X_2) - d^2(x, X_1) = 2(x - \overline{\mu})^T \Sigma^{-1}(\mu_1 - \mu_2),$$

- 样本

$$\hat{\mu}_i = \overline{x^{(i)}} = \frac{1}{n_i}\sum_{j=1}^{n_i} x_j^{(i)}, \quad i = 1, 2,$$

$$\widehat{\Sigma} = \frac{1}{n_1 + n_2 - 2}\sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(x_j^{(i)} - \overline{x^{(i)}}\right)\left(x_j^{(i)} - \overline{x^{(i)}}\right)^T$$

$$= \frac{1}{n_1 + n_2 - 2}(S_1 + S_2),$$

$$S_i = \sum_{j=1}^{n_i}\left(x_j^{(i)} - \overline{x^{(i)}}\right)\left(x_j^{(i)} - \overline{x^{(i)}}\right)^T, \quad i = 1, 2.$$

对于待测样本 x, 其判别函数定义为

$$\hat{w}(x) = (x - \overline{x})^T \widehat{\Sigma}^{-1}(\overline{x^{(1)}} - \overline{x^{(2)}}),$$

距离判别法

- 协方差阵不等时

$$w(x)=(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)-(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1).$$

$$\hat{w}(x)=(x-\overline{x^{(2)}})^T\widehat{\Sigma}_2^{-1}(x-\overline{x^{(2)}})-(x-\overline{x^{(1)}})^T\widehat{\Sigma}_1^{-1}(x-\overline{x^{(1)}}),$$

$$\begin{aligned}\widehat{\Sigma}_i &= \frac{1}{n_i-1}\sum_{j=1}^{n_i}\left(x_j^{(i)}-\overline{x^{(i)}}\right)\left(x_j^{(i)}-\overline{x^{(i)}}\right)^T \\ &= \frac{1}{n_i-1}S_i,\quad i=1,2.\end{aligned}$$

炼数成金逆向收费式网络课程

- Dataguru（炼数成金）是专业数据分析网站，提供教育，媒体，内容，社区，出版，数据分析业务等服务。我们的课程采用新兴的互联网教育形式，独创地发展了逆向收费式网络培训课程模式。既继承传统教育重学习氛围，重竞争压力的特点，同时又发挥互联网的威力打破时空限制，把天南地北志同道合的朋友组织在一起交流学习，使到原先孤立的学习个体组合成有组织的探索力量。并且把原先动辄成千上万的学习成本，直线下降至百元范围，造福大众。我们的目标是：低成本传播高价值知识，构架中国第一的网上知识流转阵地。
- 关于逆向收费式网络的详情，请看我们的培训网站 http://edu.dataguru.cn

Thanks

FAQ时间