

模式识别

第1讲 绪论、数学基础

2018~2019学年



课程说明

- **模式识别**：是研究分类识别理论与方法的一门学科，其研究目标是，建造能**自动进行事物识别与分类的智能机器系统**，以代替人工分类与识别工作。
- **学科性质**：既是人工智能领域的重要研究分支，也是实现机器智能所必不可少技术手段；**综合性、交叉性**很强。
- **理论基础**：**线性代数、概率论、数理统计、矩阵计算、最优化方法**、信号处理、计算机科学，乃至神经物理学等。
- **应用领域**：**字符识别**（手写、印刷）、**语音识别**、**图像识别**（指纹、人脸、虹膜、车牌）、**信息推荐**、**文本分析**（邮件过滤）、**工业检测**（次品）、**故障诊断**、**医疗诊断**（心电图）、**网络安全**、**信用卡欺诈检测**等。

内容安排

一、绪论、数学基础（第1讲）

二、聚类分析（第2讲）

三、判别函数分类法（几何分类法）（第3、4讲）

四、统计决策分类法（概率分类法）（第5、6讲）

五、特征提取与选择（第7讲）

六、模糊模式识别（第8讲）

七、神经网络模式识别（第9讲）

参考教材

- 齐敏等，**模式识别导论**，清华大学出版社，2009年
简明扼要，条理清晰。适合教学型阅读。**本课程教材**。
- 孙即祥等，**现代模式识别（第二版）**，国防科大出版社，2003年
分析详尽，推导完备。适合研究型阅读。**本课程主要参考**。
- 边肇祺，张学工，**模式识别（第二版）**，清华大学出版社，2000年
比较系统整理了相关理论与方法。**本课程学习参考**。根据需要选择阅读。
- R. O. Duda等，**模式分类（第二版）**，李宏东等译，机械工业出版社，2012年（原书2001年出版）
本学科的**教材经典**，部分内容更适合**研究型参考**。叙述逻辑连贯，有思想深度，内容也很多。如果希望彻底弄懂一些重点难点概念缘由或算法原理（而不满足于仅仅了解算法步骤），也包括开展后续相关研究，**学有余力或有较高学习目标的同学可参阅**。
- 模式识别工具（互联网）
- 模式识别文献（互联网）

学习要求

- 补足**矩阵计算、概率统计、最优化**相关知识；
- 牢固掌握**基本概念、重要原理**及其**算法步骤**；
- 运用重要算法对简单实例进行**手工演算求解**；
- 熟记重点算法与指标的**计算公式和推导过程**；
- 熟记重点概念与算法的**原理图**及其丰富含义；
- 能为后续科研工作打下必要的**专业理论基础**。

分级学习目标

- I. **拿学分**：掌握课堂所讲内容，做好作业与考试。
- II. **拿学分+学概念**：（I）+多读教材，多做笔记。
- III. **拿学分+学概念+懂原理**：（II）+多做习题、多做推导。

模式识别是一门综合性交叉性很强的学科，每一讲内容都有很多专著、都可单独开课，不同基础的同学其学习难度也不同。因此，课堂学习只能作为一个内容引导、初步了解，真正学好需要课后完整认真地阅读、理解和解题，以求真正熟记概念、弄懂原理、熟悉步骤，形成自己的知识框架。

一、绪论

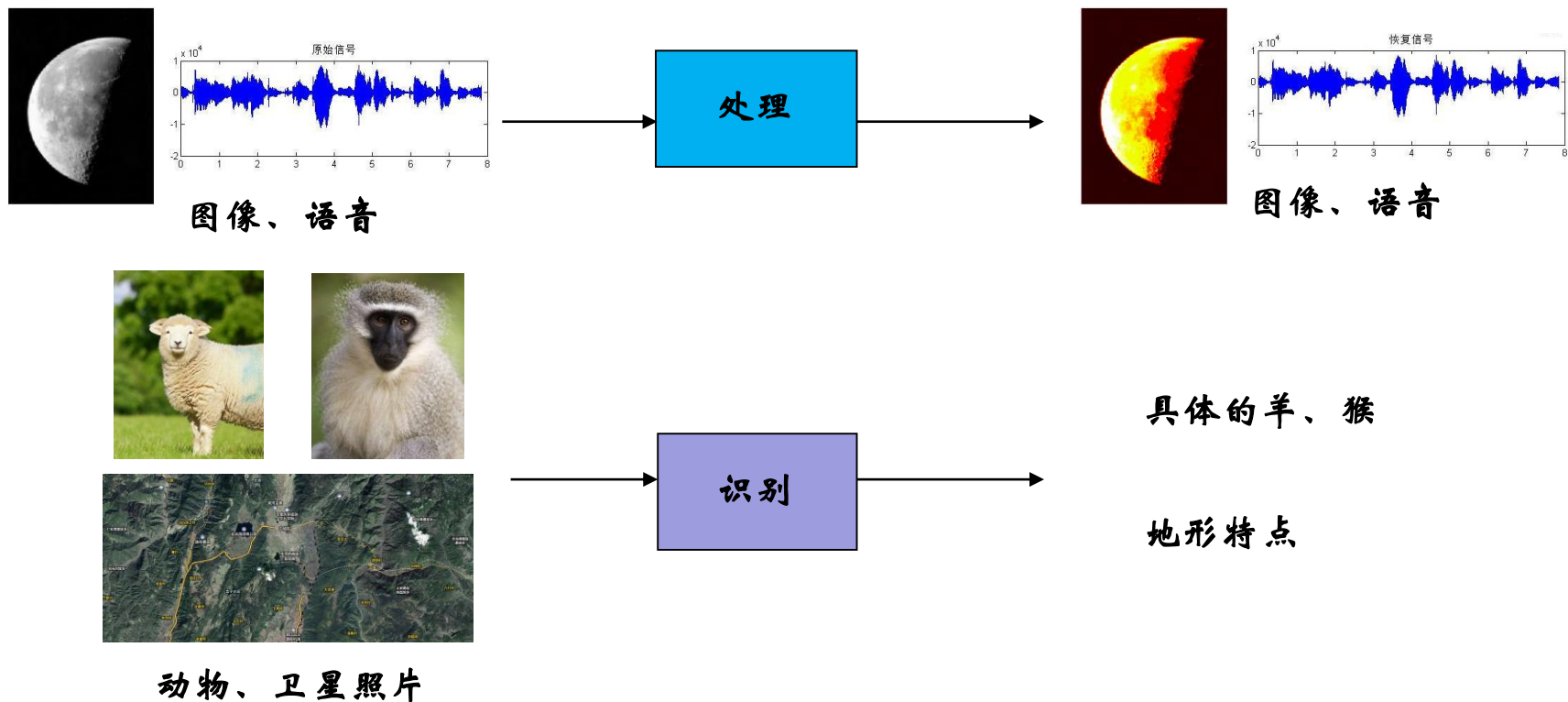
- **样本** (sample)：一个具体的研究对象，具有一个或多个可观测量。
- **特征** (features)：能从某个方面对样本进行描述、刻画或表达的可观测量（数值型、结构型）。多个特征通常用矢量表示——特征矢量。
- **模式** (pattern)：样本特征矢量的观测值，是抽象样本的数值代表。在这个意义上，“样本”与“模式”说的是同一件事情。
- **类别、类属、模式类** (class)：指在一定合理颗粒度下、有实际区分意义的基础上，主观或客观地被归属于“同一类”的客观对象（样本、模式）的**类别代号**。数学上一般处理为整数。
- **样本集** (sample set)：多个样本的集合。训练集、验证集、测试集。
- **已知样本** (known samples, example)：事先知道类别标号的样本。
- **未知样本** (unknown samples)：指特征已知、但类别标号未知的样本，也称：待分样本、待识样本。
- ——机器学习：称已知样本为**样例**，未知样本为**实例**。
- **模式识别** (pattern recognition)：基于样本和应用目的设计分类器；并对性能进行验证，进行必要优化；最后用分类器对未知类属样本进行类属判定。**识别**，不是宽泛的认识或理解，而是对类别判定。

模式识别系统的组成

- **训练集**：样本的集合，用它来设计开发模式分类器。在有监督分类中，已知类别样本的集合，在无监督分类中，是未知类别样本的集合。但都被用来提取“关于各类之间的分类知识”，即使这些样本本身没有类别信息，仍然隐含着某种分类信息！
- **验证集**：既可以认为是宽泛的训练集中的一部分，也可以理解为独立于训练集的有监督样本集。在一次特定的模型训练过程中，两者不重合；但是在多次模型训练（伴随不同模型超参数设置），两者之间是可以交换部分数据的。其作用在于：通过对特定超参数或特定子模型的设置后，对训练后的模型进行“交付前测试”，以便从多个交付模型中选择泛化性能最优者。
- **测试集**：在设计与调优分类系统时没有用过的独立样本集。
- **系统评价原则**：为了更好地对模式识别系统性能进行评价，必须使用一组独立于训练集的测试集对系统进行测试与评价。

模式识别系统的组成

“处理”与“识别”两个概念的区别

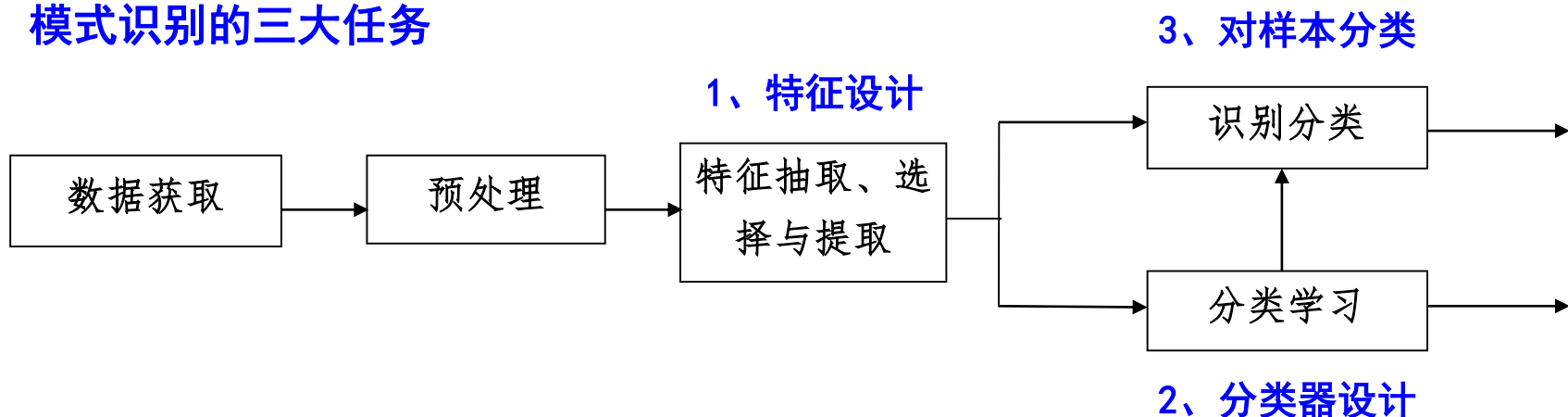


处理：输入与输出是同样性质的数值或特征。

识别：输入的是对象特征，输出的是它的分类和描述。

模式识别系统的组成

模式识别的三大任务



模式识别系统概念框图

模式识别系统的主要环节：

- **特征抽取**：特征数据的采集。如长度测量、波形测量、...
- **特征选择**：选择尽可能少、且更有利于分类的特征。含特征变换（提取）
- **分类学习**：利用样本集建立分类和识别规则，即分类器设计
- **识别分类**：对所获得样本按建立的分类规则进行分类识别

模式识别方法的分类

■ 模式识别：按理论基础划分

- 统计模式识别：以模式特征的概率分布为基础进行统计分析
- 结构模式识别：以形式语言理论对模式的结构特征进行分析
- 模糊模式识别：以模糊数学理论与方法进行的模式分类
- 神经网络模式识别：以人工神经网络理论与方法为基础
- 人工智能模式识别：以逻辑推理与专家系统理论为基础

■ 统计模式识别：按任务类型划分

- 聚类分析（Clustering Analysis）——简称：聚类
 - 简单聚类方法：最大最小距离法
 - 层次聚类方法：分裂式、凝聚式
 - 动态聚类方法：C-均值，ISODATA
- 判别分析（Discriminatory Analysis）——简称：分类
 - 几何分类法（判别函数分类法）：线性、分段线性、二次、支持向量机
 - 概率分类法（统计决策分类法）：判别式Discriminative、生成式Generative
 - 近邻分类法（几何分类法和概率分类法的一种融合方法）

模式识别方法的分类

■ 模式识别：按实现方式划分

- **有监督分类**：有**已知类别的**样本集用于分类器设计。需要有足够的先验知识。一般就指**判别分析**。
- **无监督分类**：用**不知类别的**样本集进行分类器设计。需要某种意义的先验知识和人工判断。主要指**聚类分析**，还包括降维、流形学习、**MDS**等探索性数据分析方法。

模式识别简要历史

- 1950年代Noam Chomsky 提出形式语言理论，美籍华人**乔姆斯基**提出句法结构模式识别。
- 1960年代L.A. Zadeh提出了模糊集理论，模糊模式识别理论得到了较广泛的应用。
- 1980年代Hopfield提出神经元网络模型理论。近些年人工神经网络**NN (Neural Networks)**在模式识别和人工智能上得到较广泛的应用。
- 1990年代Vapnik等人提出的小样本学习理论，支持向量机**SVM (Support Vector Machines)**得到了很大的重视。

模式识别的几个学派

- (1) 符号学派，又称为逻辑主义、心理学派或计算机学派，其原理主要为物理符号系统假设和有限合理性原理。决策树方法，专家系统。
- (2) 连接学派，又称为仿生学派或生理学派，其主要原理为神经网络及神经网络间的连接机制与学习算法。ANN。
- (3) 进化学派。
- (4) 贝叶斯学派：HMM，概率图模型。
- (5) 类推学派：kNN，支持向量机。

二、聚类分析

2.1 相似性聚类的概念

2.2 相似性测度和聚类准则

2.3 基于距离阈值的聚类算法

2.4 层次聚类法

2.5 动态聚类法

2.6 聚类结果评价

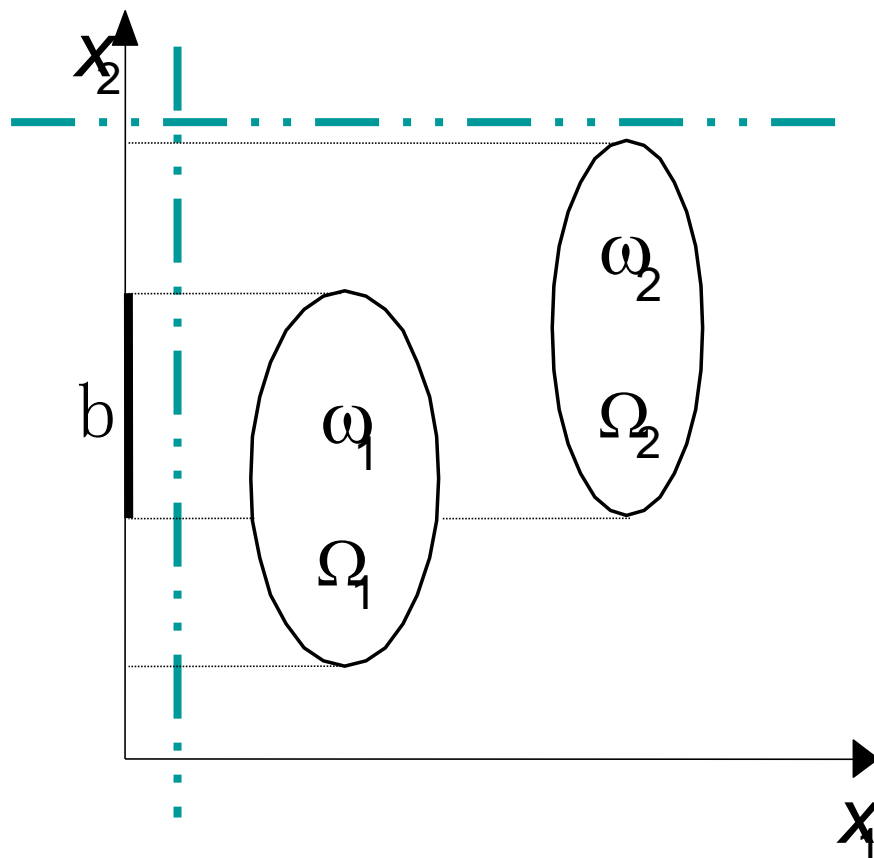
2.1 相似性聚类的概念

- **无监督学习**：使用**不知类别**的样本集进行分类器设计
 - 基于概率密度函数估计的方法（不讲）
 - 基于样本间相似性度量的方法（**聚类分析**）
- **聚类分析**：是指在没有太多先验知识的情况下，按“**物以类聚**”思想，根据模式间的**某种相似性**，对样本进行分类。因此也称为**相似性聚类**
 - 训练前，甚至没有确切的类别数目和类别定义，需要根据待分类样本集的实际特征分布情况与分类活动的应用目的，通过训练样本来**学习出类别数目和“类别的操作定义”**，同时为训练样本分配类别
 - 一般需要迭代多次才会得出有意义的结果
 - “物以类聚”原则展开说就是：
“**同类样本间的相似性 大于 不同类样本间的相似性**”。
 - 聚类方法的有效性：取决于**分类算法**与样本**特征分布**的**匹配**
 - 从方法丰富性与成熟度看，与判别分析（分类）方法相比，聚类方法更多利用直观思想和启发式方法，方法较多，但是缺乏完整稳定的聚类理论基础

特征设计对聚类分析的影响

(1) 特征选取不当使分类无效

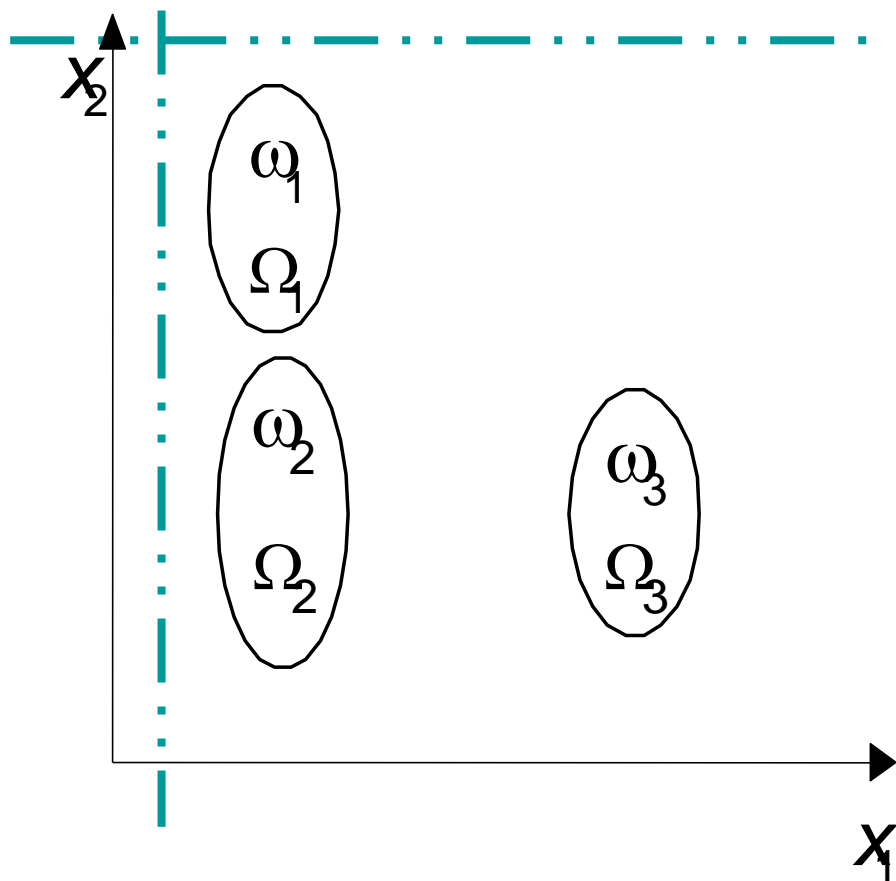
- 如果选择 x_2 作为分类特征，则在重叠区无法区分 ω_1 和 ω_2 这两类
- 如果选择 x_1 作为分类特征，则可以很容易区分 ω_1 和 ω_2 这两类



特征设计对聚类分析的影响

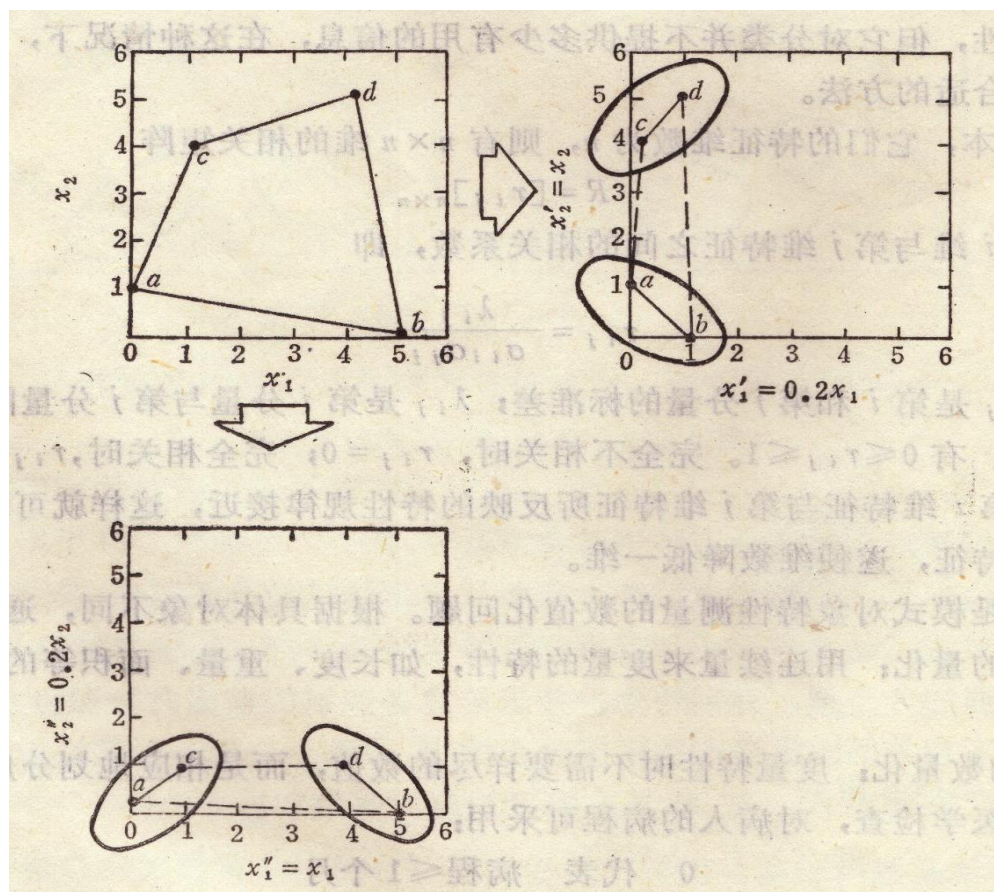
(2) 特征选取不足使分类无效

- 不管选择 x_1 还是 x_2 作为分类特征，都无法单独区分这三类
- 但是，如果选择 (x_1, x_2) 作为分类特征，则可以区分这三类

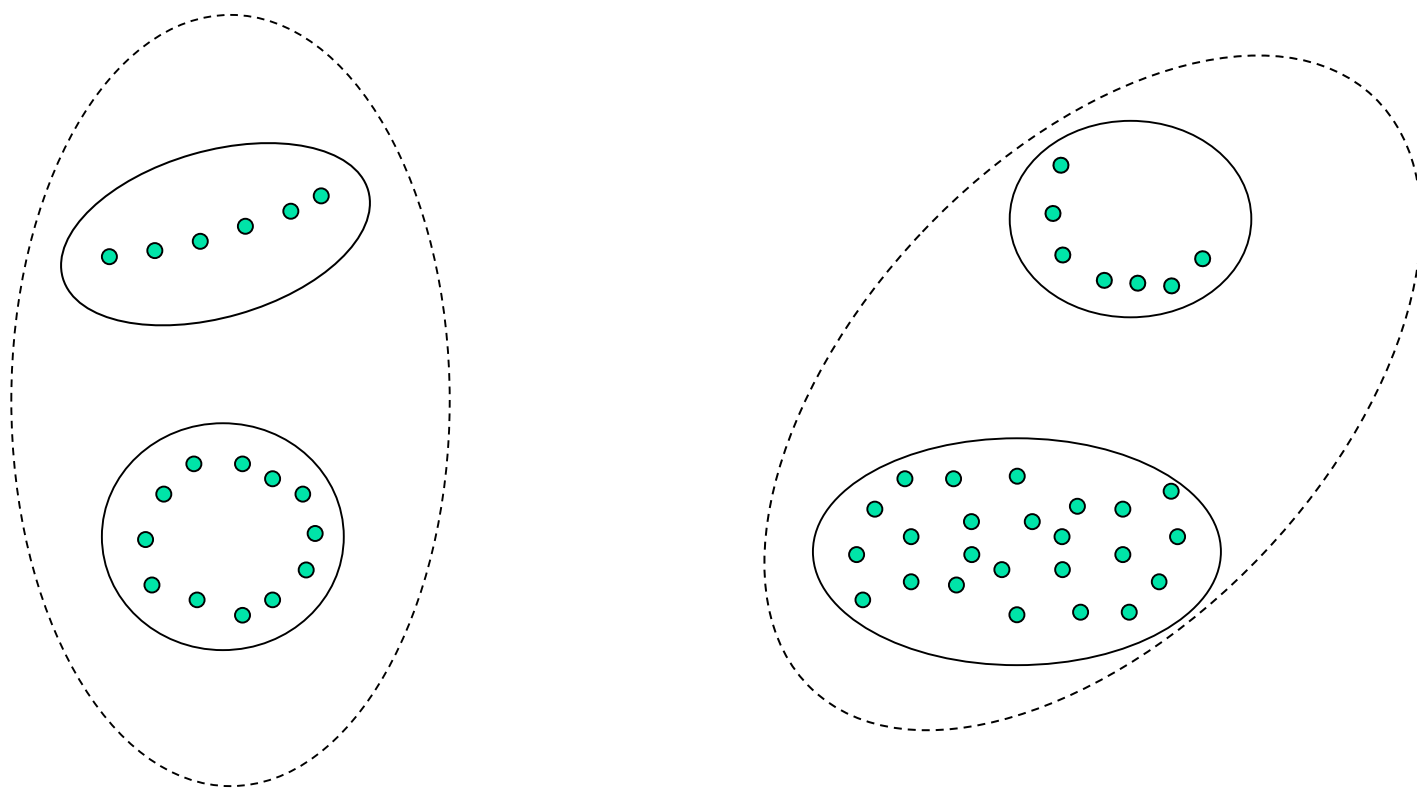


量纲不同，聚类结果可能不同

- 初始: a, b, c, d 属于4个不同的类（左上角）。
- 当 x_1 的量纲变大导致数值变小时（右上角）， a, b 成为一类， c, d 成为一类。
- 当 x_2 的量纲变大导致数值变小时（左下角）， a, c 成为一类， b, d 成为一类。
- 当然，如果 x_1 和 x_2 的量纲按照相同尺度同时变化，则不影响聚类结果



相似性（距离）测度不同，聚类结果也不同



粗聚类则为2类，细聚类则为4类

影响聚类方法有效性的几个因素

综上所述可见：

- 选择什么特征？
- 选择多少个特征？
- 选择什么样的量纲？
- 选择什么样的距离测度？

对分类结果都会产生极大影响。

2.2 相似性测度和聚类准则

- **相似性测度**：衡量模式之间相似性的一种量度。
- 由 n 个特征值组成的 n 维向量 $\mathbf{X} = [x_1, x_2, \dots, x_n]^T$ ，称为该模式（样本）的**特征矢量**。它相当于特征空间中的一个点，以特征空间中的**点间距离**作为**模式相似性的测量**，作为模式归类依据，**距离越小，越“相似”**。
- 复习一下：

$$\mathbf{Y}\mathbf{Y}^T = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix} = \begin{bmatrix} y_1^2 & y_1 y_2 & y_1 y_3 \\ y_2 y_1 & y_2^2 & y_2 y_3 \\ y_3 y_1 & y_3 y_2 & y_3^2 \end{bmatrix}$$

$$\mathbf{Y}^T \mathbf{Y} = \begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = y_1^2 + y_2^2 + y_3^2 = \|\mathbf{Y}\|^2$$

1. 欧氏距离（Euclidean, 欧几里德）

设 \mathbf{X}_1 、 \mathbf{X}_2 为两个 n 维模式样本

$$\mathbf{X}_1 = [x_{11}, x_{12}, \dots, x_{1n}]^T \quad \mathbf{X}_2 = [x_{21}, x_{22}, \dots, x_{2n}]^T$$

欧氏距离定义为：

$$\begin{aligned} D(\mathbf{X}_1, \mathbf{X}_2) &= \|\mathbf{X}_1 - \mathbf{X}_2\| = \sqrt{(\mathbf{X}_1 - \mathbf{X}_2)^T (\mathbf{X}_1 - \mathbf{X}_2)} \\ &= \sqrt{(x_{11} - x_{21})^2 + \dots + (x_{1n} - x_{2n})^2} \end{aligned}$$

距离越小，越相似。

注意：

- 各特征维上应当是相同的物理量；
- 注意同类物理量的量纲应该一样。

2. 马氏距离 (Mahalanobis, 马哈拉诺比斯)

平方表达式: $D^2 = (\mathbf{X} - \mathbf{M})^T \mathbf{C}^{-1} (\mathbf{X} - \mathbf{M})$

式中, \mathbf{X} : 模式向量; \mathbf{M} : 均值向量;

\mathbf{C} : 该类模式总体的协方差矩阵。

对 n 维向量: $\mathbf{X} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{M} = \begin{bmatrix} m_1 \\ \vdots \\ m_n \end{bmatrix}$

$$\begin{aligned} \mathbf{C} &= E\{(\mathbf{X} - \mathbf{M})(\mathbf{X} - \mathbf{M})^T\} \\ &= E\left\{\begin{bmatrix} (x_1 - m_1) \\ (x_2 - m_2) \\ \vdots \\ (x_n - m_n) \end{bmatrix} \begin{bmatrix} (x_1 - m_1) & (x_2 - m_2) & \cdots & (x_n - m_n) \end{bmatrix}\right\} \end{aligned}$$

2. 马氏距离

$$\begin{aligned} &= \begin{bmatrix} E(x_1 - m_1)(x_1 - m_1) & E(x_1 - m_1)(x_2 - m_2) & \cdots & E(x_1 - m_1)(x_n - m_n) \\ E(x_2 - m_2)(x_1 - m_1) & E(x_2 - m_2)(x_2 - m_2) & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ E(x_n - m_n)(x_1 - m_1) & \cdots & \cdots & E(x_n - m_n)(x_n - m_n) \end{bmatrix} \\ &= \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \cdots & \sigma_{1n}^2 \\ \sigma_{21}^2 & \ddots & \sigma_{jk}^2 & \vdots \\ \vdots & \vdots & \sigma_{kk}^2 & \vdots \\ \sigma_{n1}^2 & \cdots & \cdots & \sigma_{nn}^2 \end{bmatrix} \end{aligned}$$

马氏距离：在各分量特征维度计算样本模式与均值模式的距离上，剔除了该维度上模式类的方差影响——方差大，说明模式取值变化大，因此计算距离时要用方差归一化，这样得出的分量模式与均值模式的距离才具有比较意义。

优点：排除了模式样本之间的相关性影响。

特例：当 $\mathbf{C} = \mathbf{I}$ 时，马氏距离退化为欧氏距离。

3. 明氏距离(Minkowski , 闵可夫斯基)

n 维模式向量 X_i 、 X_j 间的明氏距离表示为：

$$D_m(X_i, X_j) = \left[\sum_{k=1}^n |x_{ik} - x_{jk}|^m \right]^{1/m}$$

式中， x_{ik} 、 x_{jk} 分别表示 X_i 和 X_j 的第 k 个分量。也称为 **m-范数**。

当 $m=2$ 时，明氏距离为欧氏距离。

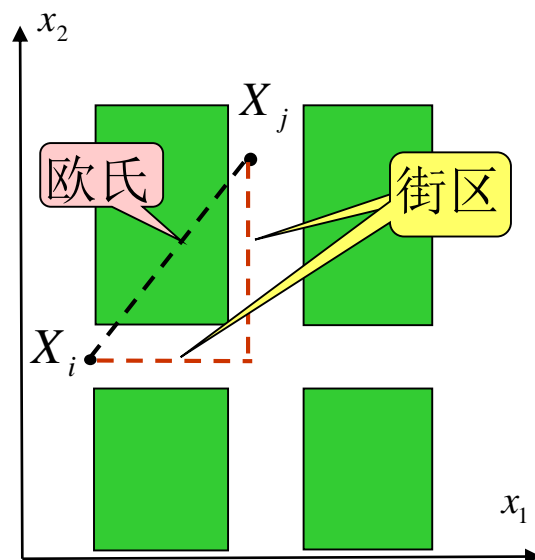
当 $m=1$ 时：

$$D_1(X_i, X_j) = \sum_{k=1}^n |x_{ik} - x_{jk}|$$

称为“**街区**”距离 (“City block” distance)、**曼哈顿距离**。

示例：当 $k=2$ 时：图中

$$D_1(X_i, X_j) = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}|$$



4. 汉明距离 (Hamming)

设 \mathbf{X}_i 、 \mathbf{X}_j 为 n 维二值模式（分量取值1或-1）样本向量，则

汉明距离：
$$D_h(\mathbf{X}_i, \mathbf{X}_j) = \frac{1}{2} \left(n - \sum_{k=1}^n x_{ik} \cdot x_{jk} \right)$$

式中， x_{ik} 、 x_{jk} 分别表示 \mathbf{X}_i 和 \mathbf{X}_j 的第 k 个分量。

说明：汉明距离中的**求和式**，表达的是两个二值向量之间，**同值分量数与不同值分量数之差**；显然，这个求和式表示的**差值**越大，表示两个向量越相似，于是**求和式的负值也就越小**，因此**求和式的负值可表示距离**，取值于 **$(-n, n)$** 。为更直观表达，用最大值 **n** 加上这个差值的负值并除以**2**，取值于 **$(0, n)$** 。

——0：表示两模式向量的各个分量都相同；

—— n ：表达两模式向量的各个分量都不同；

—— $n/4$ ：表示两模式向量中，取值相同的分量数比取值不同的分量数多 **$n/2$** 。因为此时：取值相同的分量数为 **$(3/4)n$** ，取值相同的分量数为 **$(1/4)n$** 。

5. 角度相似性函数

$$S(\mathbf{X}_i, \mathbf{X}_j) = \frac{\mathbf{X}_i^T \mathbf{X}_j}{\|\mathbf{X}_i\| \cdot \|\mathbf{X}_j\|}$$

定义为：模式向量 \mathbf{X}_i ， \mathbf{X}_j 之间夹角的余弦。

测度基础：以两矢量的方向是否相近作为考虑的基础，矢量长度并不重要。
设

$$\bar{x} = (x_1, x_2, \dots, x_n)', \bar{y} = (y_1, y_2, \dots, y_n)'$$

注意：该测度对坐标系的旋转和尺度的缩放是不变的，但对一般的线形变换和坐标系的平移不具有不变性。

6. Tanimoto测度（广义Jaccard系数）

用于0-1型二值特征情况，定义：

$$S(\mathbf{X}_i, \mathbf{X}_j) = \frac{\mathbf{X}_i^T \mathbf{X}_j}{\mathbf{X}_i^T \mathbf{X}_i + \mathbf{X}_j^T \mathbf{X}_j - \mathbf{X}_i^T \mathbf{X}_j}$$
$$= \frac{\mathbf{X}_i, \mathbf{X}_j \text{ 之间均具有某特征的数目}}{\mathbf{X}_i \text{ 和 } \mathbf{X}_j \text{ 中有某种特征的分量总数}}$$

如果把**记1**为“有某种特征”，**记0**为“无某种特征”，则Tanimoto测度的分子表达的是两特征向量之间“都具有某种特征”的分量数，分母表达的则是，在两个向量中“至少有一个向量的分量有某种特征”的分量数。

相似性测度函数的共同点都涉及到把两个相比较的向量 \mathbf{X}_i ， \mathbf{X}_j 的分量值组合起来，但怎样组合并无普遍有效的方法，对具体的模式分类，需视情况作适当选择。

相关数学基础

M1 齐次线性方程组的求解(基础解系)

M2 向量的内积、方阵的特征分解

M3 向量范数与矩阵范数

M4 多元函数对矩阵变量的求导

M5 矩阵的QR分解与奇异值分解

M6 投影矩阵、广义逆矩阵、正定矩阵

(概率论与数理统计：自行复习)

M1-线性方程组求解（一）齐次线性方程组解的性质

1. 解向量的概念

设有齐次线性方程组

[illegible]

若记

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

则上述方程组 (1) 可写成向量方程

$$Ax = \mathbf{0}.$$

若 $x_1 = \xi_{11}, x_2 = \xi_{21}, \cdots, x_n = \xi_{n1}$ 为方程 $Ax = \mathbf{0}$ 的解, 则

$$x = \xi_1 = \begin{pmatrix} \xi_{11} \\ \xi_{21} \\ \vdots \\ \xi_{n1} \end{pmatrix}$$

称为方程组(1)的**解向量**，它也就是向量方程(2)的解.

2. 齐次线性方程组解的性质

(1) 若 $x = \xi_1, x = \xi_2$ 为 $Ax = 0$ 的解, 则

$$x = \xi_1 + \xi_2$$

也是 $Ax = 0$ 的解.

(2) 若 $x = \xi_1$ 为 $Ax = 0$ 的解, k 为实数, 则 $x = k\xi_1$ 也是 $Ax = 0$ 的解.

由以上两个性质可知, 方程组的全体解向量所组成的集合, 对于加法和数乘运算是封闭的, 因此构成一个向量空间, 称此向量空间为齐次线性方程组 $Ax = 0$ 的解空间.

M1-线性方程组求解（二）基础解系及其求法

1. 基础解系的定义

$\eta_1, \eta_2, \dots, \eta_t$ 称为齐次线性方程组 $Ax = 0$ 的基础解系, 如果

(1) $\eta_1, \eta_2, \dots, \eta_t$ 是 $Ax = 0$ 的一组线性无关的解;

(2) $Ax = 0$ 的任一解都可由 $\eta_1, \eta_2, \dots, \eta_t$ 线性表出.

如果 $\eta_1, \eta_2, \dots, \eta_t$ 为齐次线性方程组 $Ax = 0$ 的一组基础解系,那么, $Ax = 0$ 的通解可表示为

$$x = k_1\eta_1 + k_2\eta_2 + \dots + k_t\eta_t$$

其中 k_1, k_2, \dots, k_{n-r} 是任意常数.

2. 齐次线性方程组基础解系的求法

设齐次线性方程组的系数矩阵为 A ，并不妨设 A 的前 r 个列向量线性无关. 于是 A 可化为

$$A \sim \begin{pmatrix} \mathbf{1} & \cdots & \mathbf{0} & b_{11} & \cdots & b_{1,n-r} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \cdots & \mathbf{1} & b_{r1} & \cdots & b_{r,n-r} \\ \mathbf{0} & \cdots & \cdots & \cdots & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \cdots & \cdots & \cdots & \cdots & \mathbf{0} \end{pmatrix}$$

$$Ax = \mathbf{0} \Leftrightarrow \begin{pmatrix} \mathbf{1} & \dots & \mathbf{0} & b_{11} & \dots & b_{1,n-r} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \mathbf{0} & \dots & \mathbf{1} & b_{r1} & \dots & b_{r,n-r} \\ \mathbf{0} & \dots & \dots & \dots & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \mathbf{0} & \dots & \dots & \dots & \dots & \mathbf{0} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ \vdots \\ x_n \end{pmatrix} = \mathbf{0}$$

[illegible]

现对 x_{r+1}, \cdots, x_n 取下列 $n-r$ 组数:

$$\begin{pmatrix} \mathbf{x}_{r+1} \\ \mathbf{x}_{r+2} \\ \vdots \\ \mathbf{x}_n \end{pmatrix} = \begin{pmatrix} \mathbf{1} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{0} \\ \mathbf{1} \\ \vdots \\ \mathbf{0} \end{pmatrix}, \dots, \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{1} \end{pmatrix}.$$

[illegible]

依次得
$$\begin{pmatrix} x_1 \\ \vdots \\ x_r \end{pmatrix} = \begin{pmatrix} -b_{11} \\ \vdots \\ -b_{r1} \end{pmatrix}, \begin{pmatrix} -b_{12} \\ \vdots \\ -b_{r2} \end{pmatrix}, \dots, \begin{pmatrix} -b_{1,n-r} \\ \vdots \\ -b_{r,n-r} \end{pmatrix}.$$

从而求得原方程组的 $n-r$ 个解:

$$\xi_1 = \begin{pmatrix} -b_{11} \\ \vdots \\ -b_{r1} \\ \mathbf{1} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}, \quad \xi_2 = \begin{pmatrix} -b_{12} \\ \vdots \\ -b_{r2} \\ \mathbf{0} \\ \mathbf{1} \\ \vdots \\ \mathbf{0} \end{pmatrix}, \quad \dots, \quad \xi_{n-r} = \begin{pmatrix} -b_{1,n-r} \\ \vdots \\ -b_{r,n-r} \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{1} \end{pmatrix}.$$

M2-向量内积与方阵的特征分解（一）向量的内积

1 向量内积的定义

定义 设有 n 维向量

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix},$$

令 $[x, y] = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$, $[x, y]$ 称为向量 x 与 y 的内积.

内积的矩阵表示

$$[x, y] = x^T y,$$

其中 x, y 都是列向量.

内积满足下列运算规律(其中 x, y, z 为 n 维向量, λ 为实数):

$$(1) [x, y] = [y, x];$$

$$(2) [\lambda x, y] = \lambda [x, y];$$

$$(3) [x + y, z] = [x, z] + [y, z].$$

2 向量的长度

定义 令

$$\|x\| = \sqrt{[x, x]} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2},$$

$\|x\|$ 称为 n 维向量 x 的长度(或范数).

向量的长度具有下列性质:

- (1)非负性 当 $x \neq 0$ 时, $\|x\| > 0$; 当 $x = 0$ 时, $\|x\| = 0$;
- (2)齐次性 $\|\lambda x\| = |\lambda| \|x\|$;
- (3)三角不等式 $\|x + y\| \leq \|x\| + \|y\|$.

当 $\|x\| = 1$ 时,称 x 为单位向量.

向量的内积满足施瓦茨不等式

$$[x, y]^2 \leq [x, x][y, y],$$

从而有 $\left| \frac{[x, y]}{\|x\| \|y\|} \right| \leq 1, \quad (\text{当} \|x\| \|y\| \neq 0 \text{时}).$

3 向量的夹角

定义 当 $\|x\| \neq 0, \|y\| \neq 0$ 时,

$$\theta = \arccos \frac{[x, y]}{\|x\| \|y\|}$$

称为 n 维向量 x 与 y 的夹角.

当 $[x, y] = 0$ 时,称向量 x 与 y 正交.

若 $x = 0$,则 x 与任何向量都正交.

4 正交向量组的性质

所谓正交向量组，是指一组两两正交的非零向量。向量空间的基若是正交向量组，就称为正交基。

定理 若 n 维向量 a_1, a_2, \dots, a_r 是一组两两正交的非零向量, 则 a_1, a_2, \dots, a_r 线性无关.

定义 设 n 维向量 e_1, e_2, \dots, e_r 是向量空间 $V (V \subset R^n)$ 的一个基, 如果 e_1, e_2, \dots, e_r 两两正交, 则称 e_1, e_2, \dots, e_r 是 V 的一个规范正交基

若 e_1, e_2, \dots, e_r 是 V 的一个规范正交基, 那么 V 中任一向量 a 都可表为

$$a = \lambda_1 e_1 + \lambda_2 e_2 + \dots + \lambda_r e_r,$$

其中 $\lambda_i = e_i^T a = [a, e_i], (i = 1, 2, \dots, r).$

施密特正交化方法

设 a_1, a_2, \dots, a_r 是向量空间 V 的一个基, 要求 V 的一个规范正交基, 只需把 a_1, a_2, \dots, a_r 这个基规范正交化.

第一步 正交化

取 $b_1 = a_1$;

$$b_2 = a_2 - \frac{[b_1, a_2]}{[b_1, b_1]} b_1;$$

.....

$$b_r = a_r - \frac{[b_1, a_r]}{[b_1, b_1]} b_1 - \frac{[b_2, a_r]}{[b_2, b_2]} b_2 - \cdots - \frac{[b_{r-1}, a_r]}{[b_{r-1}, b_{r-1}]} b_{r-1}.$$

则 b_1, b_2, \cdots, b_r 两两正交, 且与 a_1, a_2, \cdots, a_r 等价.

第二步 单位化

取 $e_1 = \frac{1}{\|b_1\|} b_1, e_2 = \frac{1}{\|b_2\|} b_2, \dots, e_r = \frac{1}{\|b_r\|} b_r$, 就得

V 的一个规范正交基

5 正交矩阵与正交变换

定义 如果 n 阶矩阵 A 满足

$$A^T A = E \quad (\text{即 } A^{-1} = A^T),$$

那么称 A 为正交矩阵.

方阵 A 为正交矩阵的充分必要条件是 A 的行(列)向量都是单位向量, 且两两正交.

正交矩阵 A 的 n 个列(行)向量构成向量空间 R^n 的一个规范正交基

定义 若 P 为正交矩阵, 则线性变换 $y = Px$ 称为正交变换.

正交变换的特性在于保持线段的长度不变.

设 $y = Px$ 为正交变换, 则有

$$\|y\| = \sqrt{y^T y} = \sqrt{x^T P^T P x} = \sqrt{x^T x} = \|x\|.$$

M2-向量内积与方阵的特征分解

(二) 方阵的特征值和特征向量

定义 设 A 是 n 阶矩阵,如果数 λ 和 n 维非零列向量 x 使关系式

$$Ax = \lambda x$$

成立,那么,这样的数 λ 称为方阵 A 的特征值,非零向量 x 称为 A 的对应于特征值 λ 的特征向量.

$|A - \lambda E| = 0$ 称为方阵 A 的特征方程.

$f(\lambda) = |A - \lambda E|$ 称为方阵 A 的特征多项式.

n 阶方阵 A 有 n 个特征值.若 $A = (a_{ij})$ 的特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_n$,则有

(1) $\lambda_1 + \lambda_2 + \cdots + \lambda_n = a_{11} + a_{22} + \cdots + a_{nn}$;

(2) $\lambda_1 \lambda_2 \cdots \lambda_n = |A|$.

1、有关特征值的一些结论

设 λ 是 $A = (a_{ij})_{n \times n}$ 的特征值,则

(1) λ 也是 A^T 的特征值;

(2) λ^k 是 A^k 的特征值(k 为任意自然数); $\varphi(\lambda)$ 是 $\varphi(A)$ 的特征值.其中 $\varphi(\lambda) = a_0 + a_1 \lambda + \cdots + a_m \lambda^m$,
 $\varphi(A) = a_0 E + a_1 A + \cdots + a_m A^m$.

(3)当 A 可逆时, $\frac{1}{\lambda}$ 是 A^{-1} 的特征值; $\frac{1}{\lambda} \cdot |A|$ 是 A^* 的特征值.

2、有关特征向量的一些结论

定理 设 $\lambda_1, \lambda_2, \dots, \lambda_m$ 是方阵 A 的 m 个特征值, p_1, p_2, \dots, p_m 依次是与之对应的特征向量, 如果 $\lambda_1, \lambda_2, \dots, \lambda_m$ 各不相等, 则 p_1, p_2, \dots, p_m 线性无关. 即属于不同特征值的特征向量是线性无关的

定理 属于同一个特征值的特征向量的**非零线性组合**仍是属于这个特征值的特征向量.

3、特征值与特征向量的求法

第一步 计算 A 的特征多项式;

第二步 求出特征方程的全部根, 即得 A 的全部特征值;

第三步 将每一个特征值代入相应的线性方程组, 求出基础解系, 即得该特征值的特征向量.

例 计算3阶实矩阵 $A = \begin{pmatrix} 3 & 2 & 4 \\ 2 & 0 & 2 \\ 4 & 2 & 3 \end{pmatrix}$ 的全部特征值

和特征向量.

解 第一步 计算 A 的特征多项式

$$\begin{aligned} f(\lambda) &= |\lambda E - A| = \begin{vmatrix} \lambda - 3 & -2 & -4 \\ -2 & \lambda & -2 \\ -4 & -2 & \lambda - 3 \end{vmatrix} \\ &= (\lambda - 8)(\lambda + 1)^2. \end{aligned}$$

第二步 求出特征多项式 $f(\lambda)$ 的全部根,即 A 的全部特征值.

令 $f(\lambda) = 0$,解之得 $\lambda_1 = 8, \lambda_2 = \lambda_3 = -1$,为 A 的全部特征值.

第三步 求出 A 的全部特征向量

对 $\lambda_1 = 8$,求相应线性方程组 $(\lambda_1 E - A)x = 0$ 的一个基础解系

$$\begin{cases} 5x_1 - 2x_2 - 4x_3 = 0, \\ -2x_1 + 8x_2 - 2x_3 = 0, \\ -4x_1 - 2x_2 + 5x_3 = 0, \end{cases}$$

化简求得此方程组的一个基础解系

$$\alpha_1 = \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}.$$

属于 $\lambda_1 = 8$ 的全部特征向量为 $k_1 \alpha_1$ ($k_1 \neq 0$ 为实数).

同理对 $\lambda_2 = \lambda_3 = -1$, 求相应线性方程组 $(\lambda_2 E - A)x = 0$ 的一个基础解系:

$$\begin{cases} -4x_1 - 2x_2 - 4x_3 = 0, \\ -2x_1 - x_2 - 2x_3 = 0, \\ -4x_1 - 2x_2 - 4x_3 = 0, \end{cases}$$

求解得此方程组的一个基础解系:

$$\alpha_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \quad \alpha_3 = \begin{pmatrix} 1 \\ -2 \\ 0 \end{pmatrix}.$$

于是A的属于 $\lambda_2 = \lambda_3 = -1$ 的全部特征向量为

$$k_2\alpha_2 + k_3\alpha_3,$$

k_2, k_3 是不全为零的实数.

从而A的全部特征向量为 $k_1\alpha_1; k_2\alpha_2 + k_3\alpha_3$, 这里 $k_1 \neq 0$ 为实数, k_2, k_3 是不全为零的实数.

M2-向量内积与方阵的特征分解（三）相似矩阵

定义 设 A, B 都是 n 阶矩阵,若有可逆矩阵 P ,使

$$P^{-1}AP = B,$$

则称 B 是 A 的相似矩阵,或说矩阵 A 与 B 相似.

对 A 进行运算 $P^{-1}AP$ 称为对 A 进行相似变换,
可逆矩阵 P 称为把 A 变成 B 的相似变换矩阵.

矩阵之间的相似具有(1)自反性; (2)对称性;
(3)传递性.

1、有关相似矩阵的性质

(1)若 A 与 B 相似，则 A 与 B 的特征多项式相同，从而 A 与 B 的特征值亦相同。

(2)若 A 与对角矩阵

$$\Lambda = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$$

相似,则 $\lambda_1, \lambda_2, \dots, \lambda_n$ 是 A 的 n 个特征值.

(3) 若 $A = P B P^{-1}$, 则 $A^k = P B^k P^{-1}$,

$$\varphi(A) = P \varphi(B) P^{-1}.$$

特别地, 若有可逆阵 P , 使 $P^{-1} A P = \Lambda$ 为对角阵,
则有 $A^k = P \Lambda^k P^{-1}$, $\varphi(A) = P \varphi(\Lambda) P^{-1}$.

(4) A 能对角化的充分必要条件是 A 有 n 个线性无关的特征向量.

(5) A 有 n 个互异的特征值, 则 A 与对角阵相似.

2、实对称矩阵的相似矩阵

(1)实对称矩阵的特征值为实数.

(2)实对称矩阵的属于不同特征值的特征向量必正交.

(3)若 λ 是实对称矩阵 A 的 r 重特征值,则对应 λ 的必有 r 个线性无关的特征向量

(4)实对称矩阵必可对角化即若 A 为 n 阶实对称阵,则必有正交阵 P ,使得 $P^{-1}AP = \Lambda$,其中 Λ 是以 A 的 n 个特征值为对角元素的对角阵.

M2-向量内积与方阵的特征分解（四）二次型

定义 含有 n 个变量 x_1, x_2, \dots, x_n 的二次齐次函数

$$\begin{aligned} f(x_1, x_2, \dots, x_n) = & a_{11}x_1^2 + a_{22}x_2^2 + \dots \\ & + a_{nn}x_n^2 + 2a_{12}x_1x_2 + 2a_{13}x_1x_3 + \dots \\ & + 2a_{n-1,n}x_{n-1}x_n \end{aligned}$$

称为二次型.

二次型可记作 $f = x^T A x$, 其中 $A^T = A$. A 称为二次型 f 的矩阵, f 称为对称阵 A 的二次型, 对称阵 A 的秩称为二次型 f 的秩.

二次型与它的矩阵是一一对应的.

当 a_{ij} 是复数时, f 称为复二次型; 当 a_{ij} 是实数时, f 称为实二次型.

正定二次型

定义 设有实二次型 $f(x) = x^T A x$, 如果对任何 $x \neq 0$, 都有 $f(x) > 0$ (显然 $f(0) = 0$), 则称 f 为正定二次型, 并称对称矩阵 A 是正定的; 如果对任何 $x \neq 0$, 都有 $f(x) < 0$, 则称 f 为负定二次型, 并称对称矩阵 A 是负定的.

M3-向量范数与矩阵范数（一）向量范数

定义： 如果 V 是数域 K 上的线性空间，且对于 V 的任一向量 x ，对应一个实数值 $\|x\|$ ，满足以下三个条件

1) 非负性： $\|x\| \geq 0$ ， 且 $\|x\| = 0 \Leftrightarrow x = 0$

2) 齐次性： $\|kx\| = |k| \cdot \|x\|$ ， $\forall k \in K$

3) 三角不等式： $\|x + y\| \leq \|x\| + \|y\|$

则称 $\|x\|$ 为 V 上向量 x 的范数，简称为向量范数。

注意： 2)中 $|k|$ 当 K 为实数时为绝对值，
当 K 为复数域时为复数的模。

向量的范数具有下列简单性质：

$$(1) \text{ 当 } \|x\| \neq 0 \text{ 时, } \left\| \frac{1}{\|x\|} x \right\| = 1 \quad \because \left\| \frac{1}{\|x\|} x \right\| = \frac{1}{\|x\|} \|x\| = 1$$

$$(2) \forall x \in V, \quad \|-x\| = \|x\| \quad \because \|-x\| = |-1| \|x\| = \|x\|$$

$$(3) \forall x, y \in V, \quad \|x\| - \|y\| \leq \|x - y\|$$

$$\because \|x\| = \|(x - y) + y\| \leq \|x - y\| + \|y\| \Rightarrow \|x\| - \|y\| \leq \|x - y\|$$

$$(4) \forall x, y \in V, \quad \left| \|x\| - \|y\| \right| \leq \|x - y\|$$

$$\|x\| = \|(x - y) + y\| \leq \|x - y\| + \|y\| \Rightarrow \|x\| - \|y\| \leq \|x - y\|$$

$$\text{同样} \quad \|y\| - \|x\| \leq \|y - x\| = \|x - y\|$$

例1: 线性空间 C^n , 设 $x = (x_1, x_2, \dots, x_n)^T \in C^n$

1: $\|x\|_1 = \sum |\xi_i|$ 是一种向量范数, 记为**1-范数**

2: $\|x\| = \sqrt{(x, x)}$ 是一种向量范数, 记**2-范数**

3: $\|x\| = \max_i |x_i|$ 是一种向量范数, 记为 ∞ -**范数**

4: $\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} \quad (1 \leq p < \infty)$

是一种向量范数, 记为**p-范数**或 l_p 范数

例2: 线性空间 V^n 中, 任取它的一组基 x_1, \dots, x_n

则对于任意向量 x , 它可以表示为

$$x = \xi_1 x_1 + \dots + \xi_n x_n$$

与 $\alpha = (\xi_1, \dots, \xi_n)^T \in C^n$ 是同构的

所以 $\|x\|_p = \|\alpha\|_p$ 是 V^n 中元素 x 的 p -范数

例3: $C[a, b]$ 为闭区间 $[a, b]$ 上的所有实连续函数所成线性空间, 可以验证以下定义式均满足范数条件

$$\|f(x)\|_1 = \int_a^b f(x) dt \quad \|f(x)\|_\infty = \max_{t \in [a, b]} |f(x)|$$

$$\|f(x)\|_p = \left(\int_a^b |f(x)|^p dt \right)^{\frac{1}{p}}, 1 < p < \infty$$

例4: 设A为n阶实对称正定矩阵, 对 $\mathbf{x} \in \mathbb{R}^n$,

定义 $\|\mathbf{x}\|_A = (\mathbf{x}^T A \mathbf{x})^{1/2}$ 称为**加权范数**或**椭圆范数**

由正定矩阵定义可知 $\|\mathbf{x}\|_A = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$; $\|\mathbf{x}\|_A \neq 0 \Leftrightarrow \mathbf{x} \neq \mathbf{0}$

对任意数 $\alpha \in \mathbb{R}$, 有

$$\|\alpha \mathbf{x}\|_A = \sqrt{(\alpha \mathbf{x})^T A \alpha \mathbf{x}} = \sqrt{\alpha^2 \mathbf{x}^T A \mathbf{x}} = |\alpha| \sqrt{\mathbf{x}^T A \mathbf{x}} = |\alpha| \|\mathbf{x}\|_A$$

由A正定且实对称 $\Rightarrow \exists$ 正交矩阵Q, 使得

$$Q^T A Q = \text{diag}(\lambda_1, \dots, \lambda_n), \quad \lambda_i > 0, i = 1, \dots, n$$

定义 $B = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}) Q^T$ 可得 $A = B^T B$

$$\because \|\mathbf{x}\|_A = (\mathbf{x}^T B^T B \mathbf{x})^{1/2} = \left[(B\mathbf{x})^T (B\mathbf{x}) \right]^{1/2} = \|B\mathbf{x}\|_2$$

$$\therefore \|\mathbf{x} + \mathbf{y}\|_A = \|B(\mathbf{x} + \mathbf{y})\|_2 \leq \|B\mathbf{x}\|_2 + \|B\mathbf{y}\|_2 = \|\mathbf{x}\|_A + \|\mathbf{y}\|_A$$

M3-向量范数与矩阵范数（二） 向量范数的等价性

定义：有限维线性空间 V^n 中任意两个向量范数

$\|\mathbf{x}\|_\alpha$ 和 $\|\mathbf{x}\|_\beta$ ，如果存在着正常数 c_1 和 c_2 ，

使得 $c_1 \|\mathbf{x}\|_\beta \leq \|\mathbf{x}\|_\alpha \leq c_2 \|\mathbf{x}\|_\beta \quad (\forall \mathbf{x} \in V^n)$

则称范数 $\|\mathbf{x}\|_\alpha$ 与 $\|\mathbf{x}\|_\beta$ 等价

(1) 自反性： $1 \cdot \|\mathbf{x}\|_\alpha \leq \|\mathbf{x}\|_\alpha \leq 1 \cdot \|\mathbf{x}\|_\alpha, \forall \mathbf{x} \in V^n$

(2) 对称性： $\frac{1}{c_2} \|\mathbf{x}\|_\alpha \leq \|\mathbf{x}\|_\beta \leq \frac{1}{c_1} \|\mathbf{x}\|_\alpha, \forall \mathbf{x} \in V^n$

(3) 传递性： $\left. \begin{array}{l} c_1 \|\mathbf{x}\|_\beta \leq \|\mathbf{x}\|_\alpha \leq c_2 \|\mathbf{x}\|_\beta \\ c_3 \|\mathbf{x}\|_\gamma \leq \|\mathbf{x}\|_\beta \leq c_4 \|\mathbf{x}\|_\gamma \end{array} \right\} \quad \forall \mathbf{x} \in V^n$
 $\Rightarrow c_5 \|\mathbf{x}\|_\gamma \leq \|\mathbf{x}\|_\alpha \leq c_6 \|\mathbf{x}\|_\gamma$

定理：有限维线性空间中任意两个向量范数都等价。

证明思路

- 1) 范数等价关系, 满足传递性;
- 2) 任意范数为坐标函数的连续函数;
- 3) 在单位超球面上有大于零的极大极小值, 与2-范数等价。

M3-向量范数与矩阵范数（三）矩阵范数

定义：矩阵空间 $C^{m \times n}$ 中， $\forall A \in C^{m \times n}$ ，

定义实数值 $\|A\|$ ，且满足以下条件

1)正定条件： $\|A\| \geq 0$ ，且 $\|A\| = 0 \Leftrightarrow A = 0_{m \times n}$

2)齐次条件： $\|kA\| = |k| \cdot \|A\|$ ， $\forall k \in K$

3)三角不等式： $\|A + B\| \leq \|A\| + \|B\|$ ， $B \in C^{m \times n}$

则称 $\|A\|$ 为 A 的广义范数。

若对于 $C^{m \times n}$, $C^{n \times l}$ 及 $C^{m \times n}$ 上的同类广义矩阵范数有

4)相容条件： $\|AB\| \leq \|A\| \|B\|$ ， $B \in C^{n \times l}$

则称 $\|A\|$ 为 A 的范数。

定义：设 $C^{m \times n}$ 的矩阵函数 $\|A\|_M$, C^m 与 C^n 中的

同类范数 $\|x\|_V$, 若 $\|Ax\|_V \leq \|A\|_M \|x\|_V$

则称矩阵范数 $\|A\|_M$ 与向量范数 $\|x\|_V$ 相容

M3-向量范数与矩阵范数（四）矩阵的从属范数

定理：对 C^m 与 C^n 上的同类向量范数 $\|\mathbf{x}\|_V$ ，定义

$$\|A\| = \max_{\|\mathbf{x}\|_V=1} \|A\mathbf{x}\|_V \quad \left(\forall A_{m \times n}, \mathbf{x} \in C^n \right)$$

则 $\|A\|$ 是 $C^{m \times n}$ 中矩阵 A 的范数，且 $\|A\|$ 与 $\|\mathbf{x}\|_V$ 相容
 $\|A\|$ 称为由 $\|\mathbf{x}\|_V$ 导出的矩阵范数（或称为从属范数）

等价定义：

$$\max_{\|\mathbf{x}\|_V=1} \|A\mathbf{x}\|_V = \max_{\mathbf{x} \neq \theta} \frac{\|A\mathbf{x}\|_V}{\|\mathbf{x}\|_V}$$

注:

(1) 一般的矩阵范数: $\because I = I \cdot I$

$$\|I\| \leq \|I\| \cdot \|I\| \quad \therefore \|I\| \geq 1$$

$$\text{例如: } \|I\|_{m1} = n, \quad \|I\|_F = \sqrt{n}$$

(2) 矩阵的从属范数: $\|I\| = \max_{\|x\|_V=1} \|Ix\|_V = 1$

(3) 常用的从属范数:

$\ x\ _V$	$\ x\ _1$	$\ x\ _2$	$\ x\ _\infty$
$\ A\ _M$	$\ A\ _1$	$\ A\ _2$	$\ A\ _\infty$

定理： 设 $A = (a_{ij})_{m \times n}$ ，则

(1) 列和范数： $\|A\|_1 = \max_j \left\{ \sum_{i=1}^m |a_{ij}| \right\}$

(2) 谱范数： $\|A\|_2 = \sqrt{\lambda_1}, \lambda_1 = \max \left\{ \lambda(A^H A) \right\}$

(3) 行和范数： $\|A\|_\infty = \max_i \left\{ \sum_{j=1}^n |a_{ij}| \right\}$

M4-多元函数对矩阵变量的求导

(一) 数量函数关于向量（矩阵）的微分

定义1 设 $y = f(X) = f(x_1, x_2, \dots, x_n)$

对 x_1, x_2, \dots, x_n 有偏导数, 定义 $y = f(X)$

对向量 $X = (x_1, x_2, \dots, x_n)^T$ 的导数为

$$\frac{df}{dX} = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right)^T = \text{grad} f,$$

而数量函数 $y = f(X)$ 对向量 $X^T = (x_1, x_2, \dots, x_n)$ 的导数定义为

$$\frac{df}{dX^T} = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right). \quad \text{显然} \left(\frac{df}{dX^T} \right)^T = \frac{df}{dX}.$$

一般地，若 $y = f(X) =$

$$f(x_{11}, x_{12}, \dots, x_{1n}, x_{21}, x_{22}, \dots, x_{2n}, \dots, x_{m1}, x_{m2}, \dots, x_{mn})$$

对每个 x_{ij} 有偏导数，则定义 $y = f(X)$

对矩阵 $X = (x_{ij})_{m \times n}$ 的导数为

$$\frac{df}{dX} = \begin{pmatrix} \frac{\partial f}{\partial x_{11}} & \frac{\partial f}{\partial x_{12}} & \dots & \frac{\partial f}{\partial x_{1n}} \\ \frac{\partial f}{\partial x_{21}} & \frac{\partial f}{\partial x_{22}} & \dots & \frac{\partial f}{\partial x_{2n}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f}{\partial x_{m1}} & \frac{\partial f}{\partial x_{m2}} & \dots & \frac{\partial f}{\partial x_{mn}} \end{pmatrix} \circ$$

性质1 设 $f(X)$, $g(X)$ 为矩阵 X 的

数量函数, 如果 $\frac{df}{dX}$ 及 $\frac{dg}{dX}$ 存在, 则

$$(1) \frac{d}{dX} [f(X) + g(X)] = \frac{df(X)}{dX} + \frac{dg(X)}{dX};$$

$$(2) \frac{d}{dX} [f(X)g(X)] = f(X) \frac{dg(X)}{dX} + g(X) \frac{df(X)}{dX}$$

例1 设自变量矩阵及其数量函数分别为

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix},$$

$$y = f(X) = x_{11}^2 + x_{12}^2 + \cdots + x_{1n}^2 + x_{21}^2 + x_{22}^2 + \cdots + x_{2n}^2 + \cdots + x_{m1}^2 + x_{m2}^2 + \cdots + x_{mn}^2$$

$$= \sum_{i=1}^m \sum_{j=1}^n x_{ij}^2, \text{求} \frac{dy}{dX}.$$

解 由于 $\frac{\partial f}{\partial x_{ij}} = 2x_{ij}$

$(i = 1, 2, \dots, m; j = 1, 2, \dots, n)$

$$\frac{dy}{dX} = \left(\frac{\partial f}{\partial x_{ij}} \right)_{m \times n} = (2x_{ij})_{m \times n} = 2X$$

性质2 设 $X = (x_{ij})_{m \times n}$, $B = (b_{ij})_{n \times m}$,

$A = (a_{ij})_{m \times m}$, 则

$$(1) \frac{d}{dX} \text{tr}(XX^T) = 2X;$$

$$(2) \frac{d}{dX} \text{tr}(BX) = \frac{d}{dX} \text{tr}(X^T B^T) = B^T;$$

$$(3) \frac{d}{dX} \text{tr}(XAX^T) = (A + A^T)X。$$

例2 设 $y = f(X) = k^T X = \sum_{j=1}^n k_j x_j$, 其中,

$$X = (x_1, x_2, \dots, x_n)^T, \quad k = (k_1, k_2, \dots, k_n)^T,$$

求 $\frac{dy}{dX}$ 。

解:
$$\frac{dy}{dx_j} = \frac{d(\sum_{j=1}^n k_j x_j)}{dx_j} = k_j$$

$$\frac{dy}{dX} = \left(\frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2}, \dots, \frac{\partial y}{\partial x_n} \right)^T = k$$

例3 设二次型 $y = f(X) =$

$$X^T A X = \sum_{i,j=1}^n a_{ij} x_i x_j, \text{ 其中,}$$

$$X = (x_1, x_2, \dots, x_n)^T, \quad A = (a_{ij})_{n \times n},$$

$$\text{且 } A^T = A, \text{ 求 } \frac{dy}{dX}.$$

解 因为 $\frac{dy}{dX} = \left(\frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2}, \dots, \frac{\partial y}{\partial x_n} \right)^T$

故只需求出 $\frac{\partial y}{\partial x_k}, k = 1, 2, \dots, n$ 即可。

由于 $y = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j = \sum_{i \neq k} \sum_{j \neq k} a_{ij} x_i x_j +$

$2 \sum_{i \neq k} a_{ik} x_i + a_{kk} x_k^2$, 故 $\frac{\partial y}{\partial x_k} = 2 \sum_{i \neq k} a_{ik} x_i + 2a_{kk} x_k$

$= 2 \sum_{i=1}^n a_{ik} x_i = 2 \sum_{j=1}^n a_{kj} x_j$, 从而 $\frac{dy}{dX} = 2AX$ 。

例14: $A \in R^{m \times n}, b \in R^m$, 若 $x \in R^n$ 使得 $\|Ax - b\|_2 = \min$,

则 $A^T Ax = A^T b$

解: $f(x) = \|Ax - b\|_2^2 = (Ax - b)^T (Ax - b)$

$$= x^T A^T Ax - 2b^T Ax + b^T b$$

$$g(x) = b^T Ax = b_1 \sum_{j=1}^n a_{1j} \xi_j + \cdots + b_m \sum_{j=1}^n a_{mj} \xi_j$$

$$\frac{dg}{dx} = \begin{bmatrix} \frac{\partial g}{\partial \xi_1} \\ \vdots \\ \frac{\partial g}{\partial \xi_n} \end{bmatrix} = \begin{bmatrix} b_1 a_{11} + \cdots + b_m a_{m1} \\ \vdots \\ b_1 a_{1n} + \cdots + b_m a_{mn} \end{bmatrix} = A^T b$$

$$\frac{df}{dx} = 2A^T Ax - 2A^T b = 0 \Rightarrow A^T Ax = A^T b$$

【注】 $r(A^T A) = r(A) \Rightarrow r(A^T A | A^T b) = r(A^T A) \Rightarrow A^T Ax = A^T b$ 有解

M4-多元函数对矩阵变量的求导

(二) 向量函数对向量（矩阵）的导数

定义2 设 $a_1(X), a_2(X), \dots, a_m(X)$ 对 $x_i (i = 1, 2, \dots, n)$ 的偏导数都存在, 定义向量函数 $a^T(X)$ 对 X 的导数为一个 $n \times m$ 阶矩阵, 它的第 i 列向量就是 $a_i(X)$ 对 X 的导数, 即

$$\frac{da^T(X)}{dX} = \begin{pmatrix} \frac{\partial a_1(X)}{\partial x_1} & \frac{\partial a_2(X)}{\partial x_1} & \dots & \frac{\partial a_m(X)}{\partial x_1} \\ \frac{\partial a_1(X)}{\partial x_2} & \frac{\partial a_2(X)}{\partial x_2} & \dots & \frac{\partial a_m(X)}{\partial x_2} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial a_1(X)}{\partial x_n} & \frac{\partial a_2(X)}{\partial x_n} & \dots & \frac{\partial a_m(X)}{\partial x_n} \end{pmatrix}_{n \times m},$$

同理， $a(X)$ 对 X^T 的导数定义为

$$\frac{da(X)}{dX^T} = \begin{pmatrix} \frac{\partial a_1(X)}{\partial x_1} & \frac{\partial a_1(X)}{\partial x_2} & \dots & \frac{\partial a_1(X)}{\partial x_n} \\ \frac{\partial a_2(X)}{\partial x_1} & \frac{\partial a_2(X)}{\partial x_2} & \dots & \frac{\partial a_2(X)}{\partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial a_m(X)}{\partial x_1} & \frac{\partial a_m(X)}{\partial x_2} & \dots & \frac{\partial a_m(X)}{\partial x_n} \end{pmatrix}_{m \times n}$$

显然有：(1) $\frac{da^T(X)}{dX} = \left[\frac{da(X)}{dX^T} \right]^T$, (2) $\frac{dX}{dX^T} = I_n = \frac{dX^T}{dX}$

性质2 设 A 为 $s \times m$ 阶常数矩阵, $f(X)$ 为向量 X 的数量函数, $a(X)$, $b(X)$ 为 X 的 m 维列向量函数, 则

$$(1) \frac{d}{dX} [a^T(X) + b^T(X)] = \frac{da^T(X)}{dX} + \frac{db^T(X)}{dX};$$

$$(2) \frac{d}{dX} [f(X) a^T(X)] = \frac{df(X)}{dX} a^T(X) + f(X) \frac{da^T(X)}{dX};$$

$$(3) \frac{d}{dX^T} [Aa(X)] = A \frac{da(X)}{dX^T};$$

$$(4) \frac{d}{dX} [a^T(X) b(X)] = \frac{da^T(X)}{dX} b(X) + \frac{db^T(X)}{dX} a(X);$$

$$(5) \frac{d}{dX^T} [a^T(X) b(X)] = b^T(X) \frac{da(X)}{dX^T} + a^T(X) \frac{db(X)}{dX^T}.$$

$$\text{例15: } \mathbf{x} = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix}, \quad F(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_l(\mathbf{x})]$$

$$\frac{dF}{d\mathbf{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial \xi_1} & \dots & \frac{\partial f_l}{\partial \xi_1} \\ \vdots & & \vdots \\ \frac{\partial f_1}{\partial \xi_n} & \dots & \frac{\partial f_l}{\partial \xi_n} \end{bmatrix}$$

$$\text{例16: } A = (a_{ij})_{n \times n}, \quad \mathbf{x} = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix}$$

$$A\mathbf{x} = \begin{bmatrix} \sum_{j=1}^n a_{1j} \xi_j \\ \vdots \\ \sum_{j=1}^n a_{nj} \xi_j \end{bmatrix}$$

$$\frac{d(A\mathbf{x})}{d\mathbf{x}^T} = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} = A$$

M5-矩阵的QR分解与奇异值分解

(一) 方阵的正交三角分解 (QR分解)

定理6: $A_{n \times n}$ 可逆 $\Rightarrow \exists$ 正交矩阵 Q , 可逆上三角矩阵 R , 使得 $A=QR$ 。

证明: $A=(a_1, a_2, \dots, a_n)$ 可逆 $\Rightarrow a_1, a_2, \dots, a_n$ 线性无关,
正交化后可得:

$$\begin{cases} b_1 = a_1 \\ b_2 = a_2 - k_{21}b_1 \\ \dots\dots\dots \\ b_n = a_n - k_{n,n-1}b_{n-1} - \dots - k_{n1}b_1 \end{cases} \quad \begin{cases} a_1 = b_1 \\ a_2 = k_{21}b_1 + b_2 \\ \dots\dots\dots \\ a_n = k_{n1}b_1 + \dots + k_{n,n-1}b_{n-1} + b_n \end{cases}$$

$$\begin{aligned}
 (a_1, a_2, \cdots, a_n) &= (b_1, b_2, \cdots, b_n)K \\
 &= (q_1, q_2, \cdots, q_n) \begin{bmatrix} |b_1| & & & \\ & |b_2| & & \\ & & \ddots & \\ & & & |b_n| \end{bmatrix} \begin{bmatrix} 1 & k_{21} & \cdots & k_{n1} \\ & 1 & \cdots & k_{n2} \\ & & \ddots & \vdots \\ & & & 1 \end{bmatrix}
 \end{aligned}$$

$$\text{令 } Q = (q_1, q_2, \cdots, q_n), R = \begin{bmatrix} |b_1| & & & \\ & |b_2| & & \\ & & \ddots & \\ & & & |b_n| \end{bmatrix} \begin{bmatrix} 1 & k_{21} & \cdots & k_{n1} \\ & 1 & \cdots & k_{n2} \\ & & \ddots & \vdots \\ & & & 1 \end{bmatrix}$$

$$\text{则 } A = QR, \text{ 其中 } q_i = \frac{b_i}{|b_i|} \quad (i = 1, 2, \cdots, n)$$

定理7: $A_{m \times n}$ 列满秩 $\Rightarrow \exists$ 矩阵 $Q_{m \times n}$ 满足 $Q^H Q = I$,
可逆上三角矩阵 $R_{n \times n}$, 使得 $A = QR$ 。

M5-矩阵的QR分解与奇异值分解

(二) 奇异值分解 (SVD)

一、预备知识

(1) $\forall A_{m \times n}, (A^H A)_{n \times n}$ 是 **Hermite** (半) 正定矩阵.

$$\forall x \neq 0, x^H A^H A x = (Ax)^H (Ax) = |Ax|^2 \geq 0$$

(2) 齐次方程组 $Ax = 0$ 与 $A^H Ax = 0$ 同解

若 $Ax = 0$, 则 $A^H Ax = 0$;

反之, $A^H Ax = 0 \Rightarrow |Ax|^2 = (Ax)^H (Ax) = x^H (A^H Ax) = 0$
 $\Rightarrow Ax = 0$

$$(3) \quad \mathbf{rank} \, A = \mathbf{rank}(A^{\mathbf{H}} A)$$

$$S_1 = \{x \mid Ax = 0\}, \quad S_2 = \{x \mid A^{\mathbf{H}} Ax = 0\}$$

$$S_1 = S_2 \Rightarrow \dim S_1 = \dim S_2 \Rightarrow n - r_A = n - r_{A^{\mathbf{H}} A}$$

$$\Rightarrow r_A = r_{A^{\mathbf{H}} A}$$

$$(4) \quad A = \mathbf{O}_{m \times n} \Leftrightarrow A^{\mathbf{H}} A = \mathbf{O}_{n \times n}$$

必要性. 左乘即得;

$$\text{充分性} \quad r_A = r_{A^{\mathbf{H}} A} = 0 \Rightarrow A = \mathbf{O}$$

二、正交对角分解

定理15: $A_{n \times n}$ 可逆 $\Rightarrow \exists$ 酉矩阵 $U_{n \times n}, V_{n \times n}$, 使得

$$U^H A V = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} \triangleq D \quad (\sigma_i > 0)$$

三、奇异值分解

$A_{m \times n} \in C_r^{m \times n} (r \geq 1) \Rightarrow A^H A \in C_r^{n \times n}$ 半正定

$A^H A$ 的特征值: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r \geq \lambda_{r+1} = \cdots = \lambda_n = 0$

A 的奇异值: $\sigma_i = \sqrt{\lambda_i}, \quad i = 1, 2, \cdots, n$

特点: (1) A 的奇异值个数等于 A 的列数

(2) A 的非零奇异值个数等于 $\text{rank } A$

定理16: $A_{m \times n} \in C_r^{m \times n} (r \geq 1), \Sigma_r = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \Rightarrow$

存在酉矩阵 $U_{m \times m}$ 及 $V_{n \times n}$ ，使得 $U^H A V = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}_{m \times n} \triangleq D$

[注]: 称 $A = U D V^H$ 为 A 的奇异值分解

- (1) U 与 V 不唯一;
- (2) U 的列为 $A A^H$ 的特征向量, V 的列为 $A^H A$ 的特征向量
- (3) 称 U 的列为 A 的左奇异向量, 称 V 的列为 A 的右奇异向量.

定理17: $A_{m \times n} \in C_r^{m \times n} (r \geq 0)$ 的奇异值分解 $A = U \begin{bmatrix} \Sigma & O \\ O & O \end{bmatrix} V^H$

中, 划分 $U = (u_1, u_2, \dots, u_m), V = (v_1, v_2, \dots, v_n)$, 则有

(1) $N(A) = \text{span}\{v_{r+1}, v_{r+2}, \dots, v_n\};$

(2) $R(A) = \text{span}\{u_1, u_2, \dots, u_r\};$

(3) $A = \sigma_1 u_1 v_1^H + \sigma_2 u_2 v_2^H + \dots + \sigma_r u_r v_r^H$

M6-投影矩阵、广义逆矩阵（一）投影矩阵

1、投影变换

定义：向量空间 C^n 中，子空间 L 与 M 满足 $C^n = L \oplus M$ ，
对 $\forall x \in C^n$ ，分解式 $x = y + z, y \in L, z \in M$ 唯一。

称变换 $T_{L,M}(x) = y$ 为沿着 M 到 L 的投影

性质(1): $T_{L,M}$ 是线性变换

性质(2): $R(T_{L,M}) = L, N(T_{L,M}) = M$

性质(3): $\forall x \in L \Rightarrow T_{L,M}(x) = x \quad \forall x \in M \Rightarrow T_{L,M}(x) = \theta$

[注] $T_{L,M}$ 是 L 中的单位变换

$T_{L,M}$ 是 M 中的零变换

2、投影矩阵

定义：取线性空间 C^n 的基为 e_1, e_2, \dots, e_n 时，元素 x 与它的坐标“形式一致”。称 $T_{L,M}$ 在该基下的矩阵记为投影矩阵 $P_{L,M}$

性质(4): $T_{L,M}(x) = y \Leftrightarrow P_{L,M}x = y$

$$x \in L \Rightarrow T_{L,M}(x) = x \Rightarrow P_{L,M}x = x$$

$$x \in M \Rightarrow T_{L,M}(x) = \theta \Rightarrow P_{L,M}x = \theta$$

4、正交投影变换

欧氏空间 C^n 中, 子空间 L 给定, 取 $M = L^\perp$,

则 $C^n = L \oplus M$

正交投影变换 $T_L = T_{L,M}$; 正交投影矩阵 $P_L = P_{L,M}$

定理2: 方阵 $P = P_L \Leftrightarrow P^2 = P, P^H = P$

4、正交投影矩阵的确定办法

$$\left. \begin{array}{l} L \text{ 的基为 } \mathbf{x}_1, \cdots, \mathbf{x}_r : \mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_r) \\ L^\perp \text{ 的基为 } \mathbf{y}_1, \cdots, \mathbf{y}_{n-r} : \mathbf{Y} = (\mathbf{y}_1, \cdots, \mathbf{y}_{n-r}) \end{array} \right\} \Rightarrow \begin{cases} \mathbf{X}^H \mathbf{Y} = \mathbf{O} \\ \mathbf{Y}^H \mathbf{X} = \mathbf{O} \end{cases}$$

已求得 $\mathbf{P}_L = \mathbf{P}_{L, L^\perp} = (\mathbf{X} | \mathbf{O}) \cdot (\mathbf{X} | \mathbf{Y})^{-1}$

因为 $(\mathbf{X} | \mathbf{Y})^H \cdot (\mathbf{X} | \mathbf{Y}) = \begin{pmatrix} \mathbf{X}^H \\ \mathbf{Y}^H \end{pmatrix} \cdot (\mathbf{X} | \mathbf{Y}) = \begin{bmatrix} \mathbf{X}^H \mathbf{X} & \mathbf{O} \\ \mathbf{O} & \mathbf{Y}^H \mathbf{Y} \end{bmatrix}$

所以 $(\mathbf{X} | \mathbf{Y})^{-1} = \begin{bmatrix} (\mathbf{X}^H \mathbf{X})^{-1} & \mathbf{O} \\ \mathbf{O} & (\mathbf{Y}^H \mathbf{Y})^{-1} \end{bmatrix} \cdot (\mathbf{X} | \mathbf{Y})^H = \begin{bmatrix} (\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H \\ (\mathbf{Y}^H \mathbf{Y})^{-1} \mathbf{Y}^H \end{bmatrix}$

于是 $\mathbf{P}_L = (\mathbf{X} | \mathbf{O}) \cdot \begin{bmatrix} (\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H \\ (\mathbf{Y}^H \mathbf{Y})^{-1} \mathbf{Y}^H \end{bmatrix} = \mathbf{X} \cdot (\mathbf{X}^H \mathbf{X})^{-1} \cdot \mathbf{X}^H$

M6-投影矩阵、广义逆矩阵（二） 广义逆矩阵

定义：对 $A_{m \times n}$ 若有 $X_{n \times m}$ 满足Penrose方程

$$(1) \quad AXA = A$$

$$(2) \quad XAX = X$$

$$(3) \quad (AX)^H = AX$$

$$(4) \quad (XA)^H = XA$$

称 X 为 A 的M-P逆，记作 A^+ .(Moore 1920, Penrose1955)

例如 $A_{m \times n}$ 可逆, $X = A^{-1}$ 满足P-方程: $A^+ = A^{-1}$

$$A = O_{m \times n}, X = O_{n \times m} \quad \text{满足P-方程: } O_{m \times n}^+ = O_{n \times m}$$

$$A = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, X = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix} \quad \text{满足P-方程: } A^+ = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

例5：设 $A \in \mathbb{C}^{m \times n}$ 的奇异值分解为 $A = U \begin{bmatrix} \Sigma_r & O \\ O & O \end{bmatrix}_{m \times n} V^H$

$$\text{则 } A^+ = V \begin{bmatrix} \Sigma_r^{-1} & O \\ O & O \end{bmatrix}_{n \times m} U^H$$

直接验证即可。

进一步的有

$$\begin{aligned} A &= (U_s, U_n) \begin{bmatrix} \Sigma_r & O \\ O & O \end{bmatrix}_{m \times n} (V_s, V_n)^H \\ &= U_s \Sigma_r V_s^H \end{aligned}$$

式子中，下标 s 改为 r ， n 改为 $n-r$ 。

$$A^+ = V_s \Sigma_r^{-1} U_s^H$$

定理10: (1) $r_{A^+} = r_A$ (2) $(A^+)^+ = A$

(3) $(A^H)^+ = (A^+)^H$ $(A^T)^+ = (A^+)^T$

(4) $(A^H A)^+ = A^+ (A^H)^+$ $(A A^H)^+ = (A^H)^+ A^+$

(5) $A^+ = (A^H A)^+ A^H = A^H (A A^H)^+$

(6) $R(A^+) = R(A^H), N(A^+) = N(A^H)$

M6-投影矩阵、广义逆矩阵（三） 正定矩阵

正定矩阵

定义：令 A 是 n 阶Hermite矩阵,若对任意复 n 维列向量

$$\mathbf{x} \neq \mathbf{0} \quad , \quad \text{都有} \quad \mathbf{x}^H A \mathbf{x} \geq 0$$

则称 A 为**Hermite非负定矩阵**，简称为非负定矩阵。对任意复 n 维列向量 $\mathbf{x} \neq \mathbf{0}$ ， 都有

$$\mathbf{x}^H A \mathbf{x} > 0$$

称为**Hermite正定矩阵**，简称为正定矩阵

性质(1)：若 A 正定，且 k 为正常数，则 kA 正定

性质(2)：若 A 、 B 正定，则 $A + B$ 正定

性质(3)：若 A 正定，则 A^{-1} 正定

正定矩阵的判断准则

Hermite矩阵 A 是正定矩阵 \longleftrightarrow A 的特征值都为正数
因为如果 $Ax = \lambda x$, 则有

$$0 < x^H Ax = x^H \lambda x = \lambda x^H x = \lambda \|x\|^2$$

一定有 $\lambda > 0$ 。反之, 必然存在酉矩阵 U , 使得

$$A = U\Lambda U^H, \quad \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$$

因为 $\lambda_i > 0$, 且对任意 $x \neq 0$ 有 $Ux = y \neq 0$

$$\text{从而 } x^H Ax = (U^H y)^H A (U^H y) = y^H U A U^H y = y^H \Lambda y > 0$$

所以 A 是正定的

Hermite矩阵 A 是非负定矩阵 \longleftrightarrow A 的特征值都为非负数

Hermite矩阵 A 是正定矩阵（非负定）

↔ A 的特征值都为正数（非负数）

↔ $\text{tr } A > \lambda_i(A)$, (or $\text{tr } A \geq \lambda_i(A)$) $i = 1, \dots, n$

↔ $A = P^H P$ P 是 n 阶非奇异矩阵(矩阵)

↔ $A = B^2$ B 是 n 阶正定矩阵(非负定矩阵)

↔ A 的 n 个顺序主子式为正数(非负数)

- A, C 是正定矩阵, 且 $AC=CA$, 则 AC 为正定矩阵
- A 是正定矩阵, C 是非负定矩阵, 且 $AC=CA$,
- A, B 是 n 阶 Hermite 矩阵, 若 $A-B$ 为正定矩阵(非负定矩阵), 则称 A 大于 B , 记为 $A > B$. (A 不小于 B 记 $A \geq B$)
也可记为 $B < A$. ($B \leq A$)

$A \geq B$ 的充要条件是, 对任意复 n 维列向量 x 都有

$$x^H A x \geq x^H B x$$

对 A, B 是实对角阵时, $A = \text{diag}(a_1, a_2, \cdots, a_n)$

$B = \text{diag}(b_1, b_2, \cdots, b_n)$, 则有 $a_i \geq b_i$

课后作业

- 见另文。
- 下次上课前提交。
- 最好使用电子档。

End of This Part