矩阵论

2024年秋学期

第三讲

2024年9月14日

矩阵代数基础

向量的范数

常用的向量范数

(1) L_0 范数（也称0范数），在稀疏表示中常用

$$\|\boldsymbol{x}\|_0 \overset{\text{def}}{=} \text{非零元素的个数}$$

(2) L_1 范数（也称1范数），将0范数松弛为1范数

$$\|\boldsymbol{x}\|_1 \overset{\text{def}}{=} \sum_{i=1}^{m} |x_i| = |x_1| + \cdots + |x_m|$$

(3) L_2 范数（常称Euclidean范数，有时也称Frobenius范数）

$$\|\boldsymbol{x}\|_2 = \left(\sum_{i=1}^{m} |x_i|^2 \right)^{1/2} = \left(|x_1|^2 + \cdots + |x_m|^2 \right)^{1/2}$$

也可缩写为 $\|\boldsymbol{x}\|$

向量的范数

常用的向量范数

(4) L_∞ 范数（也称无穷0范数）或极大范数

$$\|\boldsymbol{x}\|_\infty = \max\left\{|x_1|, \cdots, |x_m|\right\}$$

(5) L_p 范数（也称Hölder范数）

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{m} |x_i|^p\right)^{1/p} = \left(|x_1|^p + \cdots + |x_m|^p\right)^{1/p}, \quad p \geq 1$$

向量的内积与范数

两个连续函数内积 $\langle x(t), y(t)\rangle = \int_a^b x^*(t) y(t)\mathrm{d}t$

两个函数向量的内积

$$\langle \boldsymbol{x}(t), \boldsymbol{y}(t)\rangle \overset{\text{def}}{=} \int_a^b \boldsymbol{x}^{\mathrm{H}}(t)\boldsymbol{y}(t)\mathrm{d}t$$

两个函数向量的夹角定义

$$\cos\theta \overset{\text{def}}{=} \frac{\langle \boldsymbol{x}, \boldsymbol{y}\rangle}{\sqrt{\langle \boldsymbol{x}, \boldsymbol{x}\rangle}\sqrt{\langle \boldsymbol{y}, \boldsymbol{y}\rangle}} = \frac{\int_a^b \boldsymbol{x}^{\mathrm{H}}(t)\boldsymbol{y}(t)\mathrm{d}t}{\|\boldsymbol{x}(t)\|\,\|\boldsymbol{y}(t)\|}$$

函数向量的范数 $\|\boldsymbol{x}(t)\| \overset{\text{def}}{=} \left(\int_a^b \boldsymbol{x}^{\mathrm{H}}(t)\boldsymbol{x}(t)\mathrm{d}t\right)^{1/2}$

机器学习中向量的相似比较

距离衡量：未知模式向量x与样本模式向量s_1更相似

$$D(\boldsymbol{x},\boldsymbol{s}_1) \le D(\boldsymbol{x},\boldsymbol{s}_2)$$

一种比较的方法：Euclidean距离

$$D_{\mathrm{E}}(\boldsymbol{x},\boldsymbol{s}_i) = \|\boldsymbol{x}-\boldsymbol{s}_i\|_2 = \sqrt{(\boldsymbol{x}-\boldsymbol{s}_i)^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{s}_i)}$$

$$D_{\mathrm{E}}(\boldsymbol{x},\boldsymbol{y}) = 0 \quad \Leftrightarrow \quad \boldsymbol{x}=\boldsymbol{y}$$

$$D_{\mathrm{E}}(\boldsymbol{x},\boldsymbol{s}_i) = \min_k D_{\mathrm{E}}(\boldsymbol{x},\boldsymbol{s}_k), \quad k=1,2,\cdots,M$$

最近的邻居

- K最近邻(k-Nearest Neighbor，KNN)
- K均值聚类（clustering）：不同于分类（classification），将无标签的数据点按照某种相似度来进行归类

随机向量

随机向量的内积与范数

随机向量的内积涉及到概率论和统计学中的概念，特别是在处理具有随机性的向量时。随机向量是由随机变量构成的向量，其内积的计算和理解，对于数据分析、信号处理、机器学习等领域非常重要

期望

内积 $$\left\langle \boldsymbol{x}(\xi), \boldsymbol{y}(\xi) \right\rangle \stackrel{\text{def}}{=} \mathrm{E}\left\{ \boldsymbol{x}^{\mathrm{H}}(\xi)\, \boldsymbol{y}(\xi) \right\}$$

范数 $$\left\| \boldsymbol{x}(\xi) \right\|_2^2 \stackrel{\text{def}}{=} \mathrm{E}\left\{ \boldsymbol{x}^{\mathrm{H}}(\xi)\, \boldsymbol{x}(\xi) \right\}$$

正交性：两个向量之间在统计意义上正交（随机向量的统计特性描述）

$$\mathrm{E}\left\{ \boldsymbol{x}(\xi)\, \boldsymbol{y}^{\mathrm{H}}(\xi) \right\} = \boldsymbol{O}_{m\times n}$$

随机向量

概率密度函数

随机向量 $\boldsymbol{x}(\xi)=\left[x_1(\xi),\cdots,x_m(\xi)\right]^{\mathrm{T}}$

ξ **：时间，频率，位置信息等（空、时、频）**

可由联合累积分布函数（CDF: Culmulative Distribution Function）完全描述

$$F(\boldsymbol{x})\overset{\text{def}}{=}P\left\{\xi:x_1(\xi)\leqslant x_1,\cdots,x_m(\xi)\leqslant x_m\right\}$$

也可以由概率密度函数（PDF: Probability Density Function）完全描述

$$f(\boldsymbol{x})\overset{\text{def}}{=}\lim_{\Delta x_1\to 0,\cdots,\Delta x_m\to 0}\frac{P\left\{\xi:x_1<x_1(\xi)\leqslant x_1+\Delta x_1,\cdots,x_m<x_m(\xi)\leqslant x_m+\Delta x_m\right\}}{\Delta x_1\cdots\Delta x_m}$$

$$=\frac{\partial^m}{\partial x_1\cdots\partial x_m}F_x\left(x_1,\cdots,x_m\right)$$

随机向量

概率密度函数

复随机向量

$$\boldsymbol{x}(\xi)=\boldsymbol{x}_{\mathrm{R}}(\xi)+\mathrm{j}\boldsymbol{x}_{\mathrm{I}}(\xi)=\begin{bmatrix} x_{\mathrm{R}1}(\xi) \\ \vdots \\ x_{\mathrm{R}m}(\xi) \end{bmatrix}+\mathrm{j}\begin{bmatrix} x_{\mathrm{I}1}(\xi) \\ \vdots \\ x_{\mathrm{I}m}(\xi) \end{bmatrix}$$

可由联合累积分布函数完全描述

$$F(\boldsymbol{x}) \stackrel{\text{def}}{=} P\{\boldsymbol{x}(\xi)\leqslant \boldsymbol{x}\} \stackrel{\text{def}}{=} P\left\{\boldsymbol{x}_{\mathrm{R}}(\xi)\leqslant \boldsymbol{x}_{\mathrm{R}},\boldsymbol{x}_{\mathrm{I}}(\xi)\leqslant \boldsymbol{x}_{\mathrm{I}}\right\}$$

也可以由概率密度函数完全描述

$$f(\boldsymbol{x}) \stackrel{\text{def}}{=} \frac{\partial^{2m}F(\boldsymbol{x})}{\partial x_{\mathrm{R}1}\partial x_{\mathrm{I}1}\cdots\partial x_{\mathrm{R}m}\partial x_{\mathrm{I}m}}$$

随机向量

随机向量的统计描述

均值向量

$$\boldsymbol{\mu}_x = \mathrm{E}\{\boldsymbol{x}(\xi)\} = \begin{bmatrix} \mathrm{E}\{x_1(\xi)\} \\ \vdots \\ \mathrm{E}\{x_m(\xi)\} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_m \end{bmatrix}$$

随机变量的期望

连续随机变量 $\mathrm{E}\{x(\xi)\} \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} x f(x) \mathrm{d}x$ $f(x)$ 为概率密度函数

离散随机变量 $E[X] = \sum_{i=1}^{n} x_i P(X = x_i)$

对于一个离散随机变量 X ，其可能的取值为 $x_1, x_2, \ldots, x_n$

与这些取值相对应的概率分别为 $P(X = x_1), P(X = x_2), \ldots, P(X = x_n)$

随机向量

随机向量的统计描述

相关矩阵

$$\boldsymbol{R}_x \overset{\text{def}}{=} \mathrm{E}\left\{\boldsymbol{x}(\xi)\boldsymbol{x}^{\mathrm{H}}(\xi)\right\} = \begin{bmatrix} r_{11} & \cdots & r_{1m} \\ \vdots & \ddots & \vdots \\ r_{m1} & \cdots & r_{mm} \end{bmatrix}$$

对角：自相关函数 $r_{ii} \overset{\text{def}}{=} \mathrm{E}\left\{\left|x_i(\xi)\right|^2\right\}, \quad i=1,\cdots,m$

非对角：互相关函数 $r_{ij} \overset{\text{def}}{=} \mathrm{E}\left\{x_i(\xi)x_j^*(\xi)\right\}, \quad i,j=1,\cdots,m, i\neq j$

$\boldsymbol{R}_x$ **：复共轭对称矩阵，即Hermitian矩阵**

随机向量

随机向量的统计描述

自协方差矩阵

$$\boldsymbol{C}_x = \mathrm{Cov}(\boldsymbol{x},\boldsymbol{x}) \overset{\mathrm{def}}{=} \mathrm{E}\left\{\left[\boldsymbol{x}(\xi)-\boldsymbol{\mu}_x\right]\left[\boldsymbol{x}(\xi)-\boldsymbol{\mu}_x\right]^{\mathrm{H}}\right\} = \begin{bmatrix} c_{11} & \cdots & c_{1m} \\ \vdots & \ddots & \vdots \\ c_{m1} & \cdots & c_{mm} \end{bmatrix}$$

自相关矩阵与自协方差矩阵之间的关系

$$\boldsymbol{C}_x = \boldsymbol{R}_x - \boldsymbol{\mu}_x\boldsymbol{\mu}_x^{\mathrm{H}}$$

推广：互相关矩阵

$$\boldsymbol{R}_{xy} \overset{\mathrm{def}}{=} \mathrm{E}\left\{\boldsymbol{x}(\xi)\boldsymbol{y}^{\mathrm{H}}(\xi)\right\} = \begin{bmatrix} r_{x_1,y_1} & \cdots & r_{x_1,y_m} \\ \vdots & \ddots & \vdots \\ r_{x_m,y_1} & \cdots & r_{x_m,y_m} \end{bmatrix}$$

互协方差矩阵

$$\boldsymbol{C}_{xy} \overset{\mathrm{def}}{=} \mathrm{E}\left\{\left[\boldsymbol{x}(\xi)-\boldsymbol{\mu}_x\right]\left[\boldsymbol{y}(\xi)-\boldsymbol{\mu}_y\right]^{\mathrm{H}}\right\} = \begin{bmatrix} c_{x_1,y_1} & \cdots & c_{x_1,y_m} \\ \vdots & \ddots & \vdots \\ c_{x_m,y_1} & \cdots & c_{x_m,y_m} \end{bmatrix}$$

$$\boldsymbol{C}_{xy} = \boldsymbol{R}_{xy} - \boldsymbol{\mu}_x\boldsymbol{\mu}_y^{\mathrm{H}}$$

随机向量

两个随机变量之间的相关系数定义

$$\rho_{xy} \overset{\text{def}}{=} \frac{c_{xy}}{\sqrt{\mathrm{E}\left\{|x(\xi)-\mu_x|^2\right\}\mathrm{E}\left\{|y(\xi)-\mu_y|^2\right\}}} = \frac{c_{xy}}{\sigma_x\sigma_y}$$

由Cauchy-Schwartz $0 \leqslant \left|\rho_{xy}\right| \leqslant 1$

统计特性可由均值和协方差矩阵表征

实高斯白噪声向量 各个元素独立同分布

$$\mathrm{E}\{\boldsymbol{x}(t)\} = \boldsymbol{0} \qquad \mathrm{E}\left\{\boldsymbol{x}(t)\boldsymbol{x}^{\mathrm{T}}(t)\right\} = \sigma^2\boldsymbol{I}$$

复高斯白噪声向量 $\mathrm{E}\{\boldsymbol{x}(t)\} = \boldsymbol{0}$

$$\mathrm{E}\left\{\boldsymbol{x}(t)\boldsymbol{x}^{\mathrm{H}}(t)\right\} = \sigma^2\boldsymbol{I}$$

$$\mathrm{E}\left\{\boldsymbol{x}(t)\boldsymbol{x}^{\mathrm{T}}(t)\right\} = \boldsymbol{O}$$

矩阵的内积与范数

向量内积与范数的推广

$mn\times 1$ **向量**

拉长 $\quad \boldsymbol{a}=\mathrm{vec}(\boldsymbol{A})=\begin{bmatrix}\boldsymbol{a}_1\\ \vdots\\ \boldsymbol{a}_n\end{bmatrix},\qquad b=\mathrm{vec}(\boldsymbol{B})=\begin{bmatrix}\boldsymbol{b}_1\\ \vdots\\ \boldsymbol{b}_n\end{bmatrix},$

$$\langle \boldsymbol{A},\boldsymbol{B}\rangle : C^{m\times n}\times C^{m\times n}\to C$$

$$\langle \boldsymbol{A},\boldsymbol{B}\rangle=\langle \mathrm{vec}(\boldsymbol{A}),\mathrm{vec}(\boldsymbol{B})\rangle=\sum_{i=1}^{n}\boldsymbol{a}_i^{\mathrm{H}}\boldsymbol{b}_i=\sum_{i=1}^{n}\langle \boldsymbol{a}_i,\boldsymbol{b}_i\rangle$$

等价于 $\quad \langle \boldsymbol{A},\boldsymbol{B}\rangle=\mathrm{vec}(\boldsymbol{A})^{\mathrm{H}}\mathrm{vec}(\boldsymbol{B})=\mathrm{tr}\left(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{B}\right)$

矩阵的内积与范数

线性放大算子

诱导范数

$$\|A\| = \max\left\{\|Ax\| : x \in \mathbb{k}^n, \|x\| = 1\right\}$$

$$= \max\left\{\frac{\|Ax\|}{\|x\|} : x \in \mathbb{k}^n, x \neq 0\right\}$$

常用的诱导范数-*p*范数

$$\|A\|_p \overset{\text{def}}{=} \max_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p}$$

_p_取不同的数

$$\|A\|_1 = \max_{1 \le j \le n} \sum_{i=1}^{m} |a_{ij}|$$ **绝对列和范数**

$$\|A\|_{\text{spec}} = \|A\|_2$$ **矩阵的最大奇异值**

$$\|A\|_1 = \max_{1 \le i \le m} \sum_{j=1}^{n} |a_{ij}|$$ **绝对行和范数**

矩阵的内积与范数

诱导范数

例题

$$A=\begin{bmatrix} 1 & -2 & 3 \\ -4 & 5 & -6 \\ 7 & -8 & -9 \\ -10 & 11 & 12 \end{bmatrix}$$

绝对列和范数

$$\|A\|_1=\max\{22,26,30\}=30$$

绝对行和范数

$$\|A\|_\infty=\max\{6,15,24,33\}=33$$

矩阵的内积与范数

“元素形式”范数

$$\|\boldsymbol{A}\|_p \overset{\text{def}}{=} \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^p \right)^{1/p}$$

(1) L_1 范数（和范数）$(p=1)$

$$\|\boldsymbol{A}\|_1 \overset{\text{def}}{=} \sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|$$

(2) Frobenius范数 $(p=2)$

$$\|\boldsymbol{A}\|_{\mathrm{F}} \overset{\text{def}}{=} \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2}$$

(3) 最大范数（max norm）即 $p=\infty$ 的 p 范数，定义为

$$||\boldsymbol{A}||_\infty = \max_{i=1,\cdots,m;\, j=1,\cdots,n} \{|a_{ij}|\}$$

矩阵的性能指标

矩阵的二次型

任意一个正方矩阵 $\boldsymbol{A}$ 的二次型定义为 $\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}$，其中 $\boldsymbol{x}$ 可以是任意的非零复向量。

对于任何一个二次型函数

$$f\left(x_1,\cdots,x_n\right)=\sum_{i=1}^{n}\alpha_{ii}x_i^2+\sum_{i=1,i\neq j}^{n}\sum_{j=1}^{n}\alpha_{ij}x_ix_j$$

存在许多矩阵 $\boldsymbol{A}$，它们的二次型 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=f(x_1,\cdots,x_n)$ 相同。

只有实对称矩阵或复共轭矩阵满足唯一性

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=f\left(x_1,\cdots,x_n\right)$$

为保证唯一性，在讨论矩阵的二次型时，有必要假定矩阵为实对称矩阵或复共轭对称矩阵。

二次型函数一定是实值函数

矩阵的性能指标

矩阵的二次型

任意一个正方矩阵 $\boldsymbol{A}$ 的二次型定义为 $\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}$，其中 $\boldsymbol{x}$ 可以是任意的非零复向量。

对于任何一个二次型函数

$$f\left(x_1,\cdots,x_n\right)=\sum_{i=1}^{n}\alpha_{ii}x_i^2+\sum_{i=1,i\neq j}^{n}\sum_{j=1}^{n}\alpha_{ij}x_ix_j$$

存在许多矩阵 $\boldsymbol{A}$，它们的二次型 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=f(x_1,\cdots,x_n)$ 相同。

只有实对称矩阵或复共轭矩阵满足唯一性

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=f\left(x_1,\cdots,x_n\right)$$

为保证唯一性，在讨论矩阵的二次型时，有必要假定矩阵为实对称矩阵或复共轭对称矩阵。

二次型函数一定是实值函数 $\left(\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}\right)^{*}=\left(\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}\right)^{\mathrm{H}}=\boldsymbol{x}^{\mathrm{H}}\underline{\boldsymbol{A}}^{\mathrm{H}}\boldsymbol{x}=\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}$

矩阵的性能指标

矩阵的二次型

定义 1.6.1 一个复共轭对称矩阵 $\boldsymbol{A}$ 称为：

(1) 正定矩阵，记作 $\boldsymbol{A} \succ 0$，若 二次型 $\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x} > 0, \quad \forall \boldsymbol{x} \neq \boldsymbol{0}$;

(2) 半正定矩阵，记作 $\boldsymbol{A} \succeq 0$，若 二次型 $\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x} \geqslant 0, \quad \forall \boldsymbol{x} \neq \boldsymbol{0}$ (也称非负定的)；

(3) 负定矩阵，记作 $\boldsymbol{A} \prec 0$，若 二次型 $\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x} < 0, \quad \forall \boldsymbol{x} \neq \boldsymbol{0}$;

(4) 半负定矩阵，记作 $\boldsymbol{A} \preceq 0$，若 二次型 $\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x} \leqslant 0, \quad \forall \boldsymbol{x} \neq \boldsymbol{0}$ (也称非正定的)；

(5) 不定矩阵，若二次型 $\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}$ 既可能取正值，也可能取负值。

例：

$$\boldsymbol{R} = \begin{bmatrix} 3 & -1 & 0 \\ -1 & 3 & -1 \\ 0 & -1 & 3 \end{bmatrix}$$

$$\boldsymbol{x}^{\mathrm{H}}\boldsymbol{R}\boldsymbol{x} = 2x_1^2 + x_2^2 + 2x_3^2 + (x_1 - x_2)^2 + (x_2 - x_3)^2 > 0$$

矩阵的二次型刻画矩阵的正定性：特征值分布

矩阵的性能指标

矩阵的特征值

定义 $$Au = \lambda u$$

第二定义公式 $$\det(A - \lambda I) = 0$$

用特征值描述矩阵的正定性与非正定性

(1) 正定矩阵：所有特征值取正实数的矩阵。

(2) 半正定矩阵：各个特征值取非负实数的矩阵。

(3) 负定矩阵：全部特征值为负实数的矩阵。

(4) 半负定矩阵：每个特征值取非正实数的矩阵。

(5) 不定矩阵：特征值有些取正实数，另一些取负实数的矩阵。

矩阵的特征值可描述正定性、奇异性及对角元素的特殊结构

$$\det(A) = \lambda_1 \times \lambda_2 \times \ldots \times \lambda_n$$

矩阵的性能指标

矩阵的迹

矩阵的迹反映所有特征值之和

矩阵的秩

矩阵中线性无关的行或列的数目

(1) 适定方程：若 $m = n$，并且 $\mathrm{rank}(\boldsymbol{A}) = n$，即矩阵 $\boldsymbol{A}$ 非奇异，则称矩阵方程 $\boldsymbol{Ax} = \boldsymbol{b}$ 为适定 (well-determined) 方程。

(2) 欠定方程：若独立的方程个数小于独立的未知参数个数，则称矩阵方程 $\boldsymbol{Ax} = \boldsymbol{b}$ 为欠定 (under-determined) 方程。

(3) 超定方程：若独立的方程个数大于独立的未知参数个数，则称矩阵方程 $\boldsymbol{Ax} = \boldsymbol{b}$ 为超定 (over-determined) 方程。

矩阵的性能指标

表 1.6.1　矩阵的性能指标

性能指标	描述的矩阵性能
二次型	矩阵的正定性与负定性
行列式	矩阵的奇异性
特征值	矩阵的奇异性、正定性和对角元素的结构
迹	矩阵对角元素之和、特征值之和
秩	行(或列)之间的线性无关性；矩阵方程的适定性

第一章习题

见学在浙大
作业版块
Homework1

9.28 交