线性空间

一、线性空间的定义

二、线性空间的基

三、线性空间的范数

线性空间的定义

设V是一个非空集合，R为实数域。如果对于任意两个元素α，$\beta \in V$，总有唯一的元素$\gamma \in V$与之对应，成为α和β的和，记为$\gamma = \alpha + \beta$；又对于任一实数$\lambda \in R$和任一元素$\alpha \in V$，总有唯一的元素$\delta \in V$与之对应，称为λ与α的积，记作$\delta = \lambda\alpha$；并且这两种运算满足以下八条运算规律：

（设α，β，$\gamma \in V$，λ，$\mu \in R$）

（i）$\alpha+\beta=\beta+\alpha$

（ii）$(\alpha+\beta)+\gamma=\alpha+(\beta+\gamma)$

（iii）在V中存在零元素$\theta\in V$，对任意的$\alpha\in V$，都有$\alpha+\theta=\alpha$

（iv）对于任何的$\alpha\in V$，都有α的负元素δ，使得$\alpha+\delta=\theta$

（v）$1\alpha=\alpha$

（vi）$\lambda(\mu\alpha)=(\lambda\mu)\alpha$

（vii）$(\lambda+\mu)\alpha=\lambda\alpha+\mu\alpha$

（viii）$\lambda(\alpha+\beta)=\lambda\alpha+\lambda\beta$

那么，集合V就称为（实数域上的）线性空间（或向量空间），满足上述八条性质的加法和数乘运算，称之为线性运算。

本质：对于任意的α，$\beta\in V$，λ，$\mu\in R$，都有$\lambda\alpha+\mu\beta\in V$

例：设$A_{m\times n}$是一个固定矩阵，则集合$V=\{\boldsymbol{y}\in R^m|\boldsymbol{y}=A\boldsymbol{x},\boldsymbol{x}\in R^n\}$是线性空间

证明：（目标：任意$\boldsymbol{y_1},\boldsymbol{y_2}\in V$，$a,b\in \boldsymbol{R}$，证明$a\boldsymbol{y_1}+b\boldsymbol{y_2}\in V$）

设 $\boldsymbol{y}_1=A\boldsymbol{x}_1,\boldsymbol{y}_2=A\boldsymbol{x}_2$，$\boldsymbol{x}_1,\boldsymbol{x}_2\in R^n$

则$a\boldsymbol{y_1}+b\boldsymbol{y_2}=aA\boldsymbol{x}_1+bA\boldsymbol{x}_2=A(a\boldsymbol{x}_1+b\boldsymbol{x}_2)$

由于$\boldsymbol{x}_1,\boldsymbol{x}_2\in R^n$，所以 $\boldsymbol{x}_3=a\boldsymbol{x}_1+b\boldsymbol{x}_2\in R^n$

所以 $a\boldsymbol{y_1}+b\boldsymbol{y_2}=A\boldsymbol{x_3}\in V$。 所以，$V$是一个线性空间

我们成这样的V为矩阵A的值域

例：设$A_{m\times n}$是一个固定矩阵，则集合$V=\{\boldsymbol{x}\in R^n|A\boldsymbol{x}=0\}$是线性空间

（目标：任意$\boldsymbol{x_1},\boldsymbol{x_2}\in V$，$a,b\in \boldsymbol{R}$，证明$a\boldsymbol{x_1}+b\boldsymbol{x_2}\in V$）

证明： $A(a\boldsymbol{x_1}+b\boldsymbol{x_2})=A(a\boldsymbol{x}_1)+A(b\boldsymbol{x}_2)$

$=aA\boldsymbol{x}_1+bA\boldsymbol{x}_2$

$=a\boldsymbol{0}+b\boldsymbol{0}=\boldsymbol{0}$

我们成这样的V为矩阵A的核（方程组$A\boldsymbol{x}$=0的解空间）

线性空间的基

线性表示： $\alpha_1, \alpha_2, \dots, \alpha_n \ \beta \in V$,若存在一组实数$k_1, k_2, \dots, k_n \in R$,满足

$k_1\alpha_1 + k_2\alpha_2 + \cdots + k_n\alpha_n = \beta$， 则称$\beta$可以由$\alpha_1, \alpha_2, \dots, \alpha_n$ 线性表示

线性相关： $\alpha_1, \alpha_2, \dots, \alpha_n \in V$,若存在一组不全为0的实数$k_1, k_2, \dots, k_n$,满足

$k_1\alpha_1 + k_2\alpha_2 + \cdots + k_n\alpha_n = \theta$，则$\alpha_1, \alpha_2, \dots, \alpha_n$ 线性相关

线性相关⇔至少存在一个向量可以被其余向量线性表示

线性无关： $\alpha_1, \alpha_2, \dots, \alpha_n \in V$,若满足 $k_1\alpha_1 + k_2\alpha_2 + \cdots + k_n\alpha_n = \theta$， 则必有

$k_1 = k_2 = \cdots = k_n = 0$，则$\alpha_1, \alpha_2, \dots, \alpha_n$ 线性无关。

例： 向量$\boldsymbol{x_1}=(3,4),\boldsymbol{x_2}=(1,0),\boldsymbol{x_3}=(0,1)$，我们有如下关系：

$$\boldsymbol{x_1}+(-3)\boldsymbol{x_2}+(-4)\boldsymbol{x_3}=(0,0)$$

所以$\boldsymbol{x_1},\boldsymbol{x_2},\boldsymbol{x_3}$线性相关。

例： 证明向量$\boldsymbol{x_1}=(3,4),\boldsymbol{x_2}=(1,0)$线性无关

证： 设$k_1\boldsymbol{x}_1+k_2\boldsymbol{x}_2=\theta$，我们有

$$(3k_1,4k_1)+(k_2,0)=(0,0)$$

$$\begin{cases}3k_1+k_2=0\\ \quad 4k_1=0\end{cases} \quad => \begin{cases}k_1=0\\ k_2=0\end{cases}$$

所以，$\boldsymbol{x_1},\boldsymbol{x_2}$线性无关

线性空间基的定义

在线性空间V中，如果存在n个元素$\alpha_1, \alpha_2, \dots, \alpha_n$，满足：

（i）$\alpha_1, \alpha_2, \dots, \alpha_n$线性无关

（ii）V中任一元素α都可以由$\alpha_1, \alpha_2, \dots, \alpha_n$线性表示

那么，$\alpha_1, \alpha_2, \dots, \alpha_n$称为线性空间V的一个基，$n$（基的个数）称为线性空间$V$的维数

空间V称为由基$\{\alpha_1, \alpha_2, \dots, \alpha_n\}$张成的线性空间，记作$V$=span$\{\alpha_1, \alpha_2, \dots, \alpha_n\}$。

性质： $V=\{x | x = c_1\alpha_1 + c_2\alpha_2 + \cdots + c_n\alpha_n, c_i$为任意实数，$i = 1,2, \dots, n\}$

例： 方程组$A\boldsymbol{x}=0$的基础解系$\boldsymbol{x}_i (i = 1, \dots n - r)$为其解空间的一组基

方程的解：$\boldsymbol{x}=c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 + \cdots + c_{n-r}\boldsymbol{x}_{n-r}$ 其中c_i为任意实数

坐标：　若V是一个线性空间，$\{\alpha_1, \alpha_2, \dots \alpha_n\}$是线性空间V的一组基，

对于$\alpha \in V$，如果有$\alpha = x_1\alpha_1 + x_2\alpha_2 + \cdots + x_n\alpha_n$，那么其表示系数所构成的$n$为实向量$(x_1, x_2, \dots, x_n)$称为$\alpha$在基$\{\alpha_1, \alpha_2, \dots, \alpha_n\}$下的坐标。所以，线性空间的元素称为向量

例：　二位实向量空间$R^2 = \{(x, y) | x, y \in R\}$是一个线性空间，向量组

$\{(1,0), (0,1)\}$，$\{(1,0), (1,1)\}$分别都是R^2的一组基。向量(5,3)可以表示成

$(5,3) = 5(1,0) + 3(0,1)$ 或者$(5,3) = 2(1,0) + 3(1,1)$。 那么(5,3)在基

$\{(1,0), (0,1)\}$下的坐标是(5,3)，在基$\{(1,0), (1,1)\}$下的坐标是(2,3)。

线性空间的范数（模）

在线性空间V中定义一种运算$\|.\|:V \to \boldsymbol{R}$，对于任意的$\alpha, \beta \in V$，$\lambda \in R$满足如下性质：

（i）$\|\alpha\| \geq 0$，若$\|\alpha\|=0 \Leftrightarrow \alpha = \theta$（零向量）

（ii）$\|\lambda\alpha\| = |\lambda|\|\alpha\|$

（iii）$\|\alpha + \beta\| \leq \|\alpha\| + \|\beta\|$（三角不等式）

则称$\|.\|$为线性空间V的一个范数，这样的V称为赋范线性空间。

赋范线性空间中的元素$\alpha, \beta \in V$，定义$\|\alpha - \beta\|$为α, β之间的距离

例：n维向量空间R^n是一线性空间，对于任意向量$\boldsymbol{x}$= $(x_1, x_2, \ldots, x_n) \in R^n$,

定义欧几里得范数$\|\boldsymbol{x}\|$=$\sqrt{(x_1)^2 + (x_2)^2 + \cdots + (x_n)^2}$

例：当$1 \le p < \infty$时，定义$\boldsymbol{R}^n$空间的L^p范数如下：

$$\|\boldsymbol{x}\|_{L^p}(\|\boldsymbol{x}\|_p) = \left(\sum_i x_i^p\right)^{\frac{1}{p}}$$

当$p = \infty$时，R^n空间的L^∞范数定义为$\|\boldsymbol{x}\|_{L^\infty}(\|\boldsymbol{x}\|_\infty)$=$\max\limits_i |x_i|$

在机器学习中，矩阵最常用的范数是Frobenius范数，定义如下：

$$\|A\|_F = \sqrt{\sum_{i,j} a_{i,j}^2}$$

矩阵分解

一、方阵的正交分解

二、矩阵的奇异值分解（SVD）

三、应用举例：主成分分析（PCA）

特征值与特征向量定义

定义：设 $A_{n\times n}$, 如果有数 λ_i 和 n 维非零列向量 $\boldsymbol{x}$ ，使

$$A\boldsymbol{x}=\lambda_i\boldsymbol{x}\quad i=1,2,\ldots,n$$

则称 λ_i 为 A 的特征值 , 非零列向量 $\boldsymbol{x}$ 为 A 的对应于特征值 λ_i 的特征向量

注： (1) A 是方阵 ;

(2) 特征向量 $\boldsymbol{x}$ 是非零列向量 ;

(3) 属于每一个特征值 λ_i 的特征向量不唯一，有无数个

(4) 一个特征向量只能属于一个特征值 .

特征值：λ_i是关于λ的多项式 $|A-\lambda I_n|=0$的根，记作$\lambda_1,\lambda_2,\ldots,\lambda_n$

特征向量：属于λ_i的特征向量是线性方程组$(A-\lambda_i I_n)\boldsymbol{x}=0$的解

矩阵分解：

设$\{\boldsymbol{x}_{i1},\boldsymbol{x}_{i2},\ldots,\boldsymbol{x}_{im}\}$是方程组$(A-\lambda_i I_n)\boldsymbol{x}=0$解空间的基，定义矩阵

$P_{n\times n}=[\boldsymbol{x}_{11},\boldsymbol{x}_{12},\ldots,\boldsymbol{x}_{1m},\boldsymbol{x}_{21},\boldsymbol{x}_{22},\ldots]_{n\times n}$ 那么，我们可以把方阵A分解成

$$P^{-1}AP=\begin{bmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_1 & \cdots & 0\\ \vdots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & \lambda_n\end{bmatrix}\quad \text{或者} \quad A=P\begin{bmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_1 & \cdots & 0\\ \vdots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & \lambda_n\end{bmatrix}P^{-1}$$

我们称这样的分解为特征分解（相似对角化）

由 $P^{-1}AP=\begin{pmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{pmatrix}$，可以得到 $AP=P\begin{pmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{pmatrix}$.

令$P=[\boldsymbol{x}_1,\boldsymbol{x}_2,\dots,\boldsymbol{x}_n]$，有

$$AP=A[\boldsymbol{x}_1,\boldsymbol{x}_2,\dots,\boldsymbol{x}_n]=[A\boldsymbol{x}_1,A\boldsymbol{x}_2,\dots,A\boldsymbol{x}_n]=(\lambda_1\boldsymbol{x}_1,\lambda_2\boldsymbol{x}_2,\dots,\lambda_n\boldsymbol{x}_n)$$

即

$$A\boldsymbol{x}_i=\lambda_i\boldsymbol{x}_i\quad(i=1,2,\dots,n)$$

可见，λ_i 是 A 的特征值，$\boldsymbol{x}_i$ 为其特征值 λ_i 的特征向量 .

正交矩阵：满足$AA^T = I_n (A^{-1} = A^T)$的n阶方阵称为正交矩阵

标准正交基：n个n维向量$\{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n \in R^n\}$，若满足如下性质

$$\begin{cases} \boldsymbol{x}_i \boldsymbol{x}_j = 1 & \text{当} i = j \text{时} \\ \boldsymbol{x}_i \boldsymbol{x}_j = 0 & \text{当} i \neq j \text{时} \end{cases}$$

则称$\{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n \in R^n\}$为一组标准正交基

例：$\{(1,0,0), (0,1,0), (0,0,1)\}$为三维向量空间$R^3$的一组标准正交基

性质：若A=$[\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n]$为n阶正交矩阵⇔$\{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n \in R^n\}$为一组标准正交基

正交矩阵的性质:

1、若A是正交矩阵，$\{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n \in R^n\}$是一组标准正交基，那么

$\{A\boldsymbol{x}_1, A\boldsymbol{x}_2, \dots, A\boldsymbol{x}_n \in R^n\}$也是一组标准正交基

2、对直角坐标系做旋转变换，新旧坐标之间关系是一个正交变换，即存在一个正交矩阵A，使得原来坐标的点$\boldsymbol{x} = (x_1, x_2, \dots, x_n)$在新的坐标系下的坐标为$\boldsymbol{y} = A\boldsymbol{x}$

3、正交变换保持内积、长度、距离不变，即任意的n维向量$\boldsymbol{x}, \boldsymbol{y} \in R^n$,有

内积不变$(\boldsymbol{x}, \boldsymbol{y}) = (A\boldsymbol{x}, A\boldsymbol{y})$

长度不变$\|\boldsymbol{x}\|_2 = \|A\boldsymbol{x}\|_2$（取$\boldsymbol{y} = \boldsymbol{x}$）

距离不变$\|\boldsymbol{x} - \boldsymbol{y}\|_2 = \|A\boldsymbol{x} - A\boldsymbol{y}\|_2$（取$\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{y}$）

正交分解：

若n阶方阵A可进行特征分解（相似对角化），即存在n阶可逆矩阵P，使得

$$P^{-1}AP = diag(\lambda_1, \lambda_2, \dots, \lambda_n)$$

其中λ_i为A的特征值，P=$[\boldsymbol{x_{11}}, \boldsymbol{x_{12}}, \dots, \boldsymbol{x_{1m}}, \boldsymbol{x_{21}}, \boldsymbol{x_{22}}, \dots]$列向量为$\lambda_i$对应的特征向量。那么，一定存在另一组属于$\lambda_i$的特征向量Q=$[\boldsymbol{y}_{11}, \boldsymbol{y}_{12}, \dots, \boldsymbol{y}_{1m}, \boldsymbol{y}_{21}, \boldsymbol{y}_{22}, \dots]$，满足向量组$\{\boldsymbol{y}_{11}, \boldsymbol{y}_{12}, \dots, \boldsymbol{y}_{1m}, \boldsymbol{y}_{21}, \boldsymbol{y}_{22}, \dots.\}$是一组n维标准正交基，即Q是n阶正交矩阵

$$Q^{-1}AQ = Q^TAQ = diag(\lambda_1, \lambda_2, \dots, \lambda_n)$$

我们称该分解为正交分解。

注：正交分解是一种特殊的特征分解

正交分解的几何意义：

设A是一个实对称矩阵，即$A^T = A, \boldsymbol{x} = (x_1, x_2, \dots, x_n)$。称

$$f(x_1, x_2, \dots, x_n) = \boldsymbol{x}A\boldsymbol{x}^T$$

为二次型。

对矩阵A进行正交变换，$A = Q\Lambda Q^T$。那么$f(\boldsymbol{x}) = \boldsymbol{x}Q\Lambda Q^T\boldsymbol{x}^T = (\boldsymbol{x}Q)\Lambda(\boldsymbol{x}Q)^T$

令$\boldsymbol{x}Q = \boldsymbol{y} = (y_1, y_2, \dots, y_n)$,有$f(\boldsymbol{x})$=$\lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$

正交变换的几何意义：把一般的二次曲线经过坐标旋转变成标准的二次曲线，将其主轴（长轴、短轴）转到坐标轴上。

矩阵的SVD（奇异值分解）

非退化方阵的SVD：设A是n阶非退化方阵，即$r(A) = n$。那么存在正交矩阵P和Q，使得

$$P^T AQ = \mathrm{diag}(\alpha_1, \alpha_2, \dots, \alpha_n)$$

其中 $\alpha_i > 0$（$i = 1,2, \dots, n$）

1、不一定每个方阵都可以正交分解，只有实对称矩阵($A = A^T$)一定可以。但是每个方阵都可以有SVD

2、正交分解是同一个正交矩阵Q，SVD分解是两个正交矩阵P、Q

3、正交分解对角线是特征值，SVD对角线不是特征值，但是都大于0

一般矩阵的SVD

设A是秩为$r(r>0)$的$m \times n$阶实矩阵，则存在m阶正交矩阵U和n阶正交矩阵V，使得

$$U^T AV=\begin{bmatrix} \Lambda_r & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix}$$

其中Λ_r=diag$(\sigma_1, \sigma_2, \ldots, \sigma_r)$

$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$为矩阵$A$的全部奇异值

$$A = U\begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} V^T = \sigma_1 U_1 V_1^T + \sigma_2 U_2 V_2^T + \cdots + \sigma_r U_r V_r^T$$

U_i, V_i为矩阵U, V的列向量。

Moore−*Penrose*伪逆:

设 $A \in R^{m\times n}$ ，若 $A^+ \in R^{n\times m}$，满足 ：

$$AA^+A = A \qquad A^+AA^+ = A^+$$

$$(AA^+)^T = AA^+ \qquad (A^+A)^T = A^+A$$

则称A^+为矩阵A的伪逆，上述四个方程称为Moore−Penrose方程。

不相容线性方程组的解

定义：设$A\in R^{m\times n}$,$\boldsymbol{b}\in R^m$，$A\boldsymbol{x}=\boldsymbol{b}$是不相容线性方程组。若存在向量$x_0\in R^n$

使得对于任何$\boldsymbol{x}\in R^n$，都有$\|Ax_0-b\|\le\|Ax-b\|$

则称x_0为方程组$A\boldsymbol{x}=\boldsymbol{b}$的最小二乘解

若u是方程组$A\boldsymbol{x}=\boldsymbol{b}$的最小二乘解，如果对于任意一个最小二乘$x_0$，都有

$$\|u\|\le\|x_0\|$$

则称u是最佳最小二乘解。

定理： 设$A \in R^{m\times n}$,$\boldsymbol{b} \in R^m$，则向量$\boldsymbol{x} = A^+\boldsymbol{b}$是方程组$A\boldsymbol{x} = \boldsymbol{b}$的最佳最小二乘解。

定理： 如果矩阵A的SVD为$A = U\Lambda V^T$,那么A的伪逆为 $A^+ = V\Lambda^+U^T$,
其中Λ^+是Λ的伪逆，是将Λ主对角线上非零元素σ_i
取倒数变成$\frac{1}{\sigma_i}$之后再取转置

PCA（主成分分析）的数学原理

1、计算样品数据的协方差矩阵$A=(s_{ij})_{n\times n}$，其中

$$s_{ij}=\frac{1}{n-1}\sum_{k=1}^{n}(x_{ki}-\overline{x_i})(x_{kj}-\overline{x_j})$$

2、对矩阵A进行正交分解，并对特征值进行排序

$$Q^TAQ=diag(\lambda_1,\dots,\lambda_r,0,\dots,0)\,, \qquad \lambda_1\ge\lambda_2\ge\cdots\ge\lambda_r>0$$

3、确定最小的m，使得贡献率$G(m)=\frac{\sum_1^m\lambda_i}{\sum_1^r\lambda_i}>85\%$

4、取$F_i=Q_i\boldsymbol{x}$（i=1…m）为主成分变量，其中

$\boldsymbol{x}=(x_1,\dots,x_n)^T, Q_i$为正交矩阵$Q$的第$i$列向量

THANK YOU