线性代数与向量微积分

1. 向量与空间

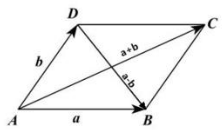
1.1 什么是向量

向量：具有大小和方向的量

向量$\vec{a}$

$|\vec{a}|$ 向量的模

1.2 向量运算



＋：$\vec{a}+\vec{b}=\vec{c}$

－：$\vec{c}-\vec{a}=\vec{b}$

×：$|\vec{a}||\vec{c}|\cdot\cos\theta$

＝：方向相同，模相等

向量没有除法。为什么没有除法？

1.3 内积/范数/夹角

范数：

$\|\mathbf{x}\|=\sqrt{x_1^2+x_2^2+\cdots+x_n^2}$ （Norm2）

$\|\mathbf{x}\|+\|\mathbf{y}\|\geq\|\mathbf{x}+\mathbf{y}\|$。三角不等式

$\|\mathbf{x}\|=1$ 单位向量

$\|\mathbf{x}\|_1=|\mathbf{x}|+|\mathbf{x_2}|+|\mathbf{x_n}|$

$\|\mathbf{x}\|_\infty=\max|x_i|$

$\|\mathbf{x}\|_p=(|x_1|^p+|x_2|^p+|x_n|^p)^{\frac{1}{p}}$ （1≤p≤n）

夹角：

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

正交向量：$[\vec{x},\vec{y}]=0$，则 $\vec{x},\vec{y}$ 正交，$\vec{0}$ 与任何向量都正交

正交向量组：两两正交的向量组。

内积：

$\vec{x}\cdot\vec{y}=\sum_{i=1}^{n}x_i y_i$

运算：$[\vec{x},\vec{y}]=[\vec{y},\vec{x}]$　　　　　　$\vec{x}\cdot\vec{y}=|x|\cdot|y|\cdot\cos\theta$

$[\lambda\vec{x},\vec{y}]=\lambda[\vec{x},\vec{y}]$

$[\vec{x}+\vec{y},\vec{z}]=[\vec{x},\vec{z}]+[\vec{y},\vec{z}]$

定理：$[\vec{x},\vec{x}]\geq 0$；当且仅当 $x\neq 0$，$[\vec{x},\vec{x}]>0$；

$[\vec{x},\vec{y}]^2\leq[\vec{x}\cdot\vec{x}][\vec{y}\cdot\vec{y}]$

2. 从向量空间到矩阵

2.1 向量组，向量空间

$\vec{a}$，$\vec{b}$，$\vec{c}$，$\vec{d}$

向量：$\mathbf{a_1}\quad \mathbf{a_2}\quad \cdots\quad \mathbf{a_n}$

$$\mathbf{A}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$ —行向量组 丨列向量组

①线性组合：$\vec{a}_1,\vec{a}_2,\vec{a}_3,\cdots,\vec{a}_m$ 对于任意的实数 $k_1,k_2,\cdots,k_m$，向量 $k_1\vec{a}_1+k_2\vec{a}_2+...+k_m\vec{a}_m$ 称为向量组的一个线性组合。

②线性表示：$\vec{b}$，存在 $k_1,k_2,\cdots,k_m$，使 $\vec{b}=\sum_{i=1}^{m}k_i\vec{\alpha}_i$，$\vec{b}$ 能被 A 线性表示。

③线性相关：存在不全为零的数 $k_1,k_2,\cdots,k_m$，使得 $\sum_{i=1}^{m}k_i\vec{\alpha}_i=\vec{0}$，称 A 线性相关，否则线性无关。即当且仅当 k_i 全为 0 时，$\sum_{i=1}^{m}k_i\vec{\alpha}_i=\vec{0}$ 成立。

2.2 向量空间与基

定义 1：V 为 n 维向量的集合，如果集合 V 非空，且集合 V 对加法及数乘两种运算封闭，那么称集合 V 为向量空间。

① $\vec{\alpha} \in V$， $\vec{\beta} \in V$， $(\vec{\alpha} + \vec{\beta}) \in V$

② $\vec{\alpha} \in V$， $\lambda \vec{\alpha} \in V$

定义 2：

子空间：V_1, V_2 为向量空间，若 $V_1 \in V_2$，则 V_1 是 V_2 的子空间。

定义 3：V 是向量空间，如果 r 个向量 $\vec{\alpha}_1, \vec{\alpha}_2, \cdots \vec{\alpha}_r \in V$

且 $\vec{\alpha}_1, \vec{\alpha}_2, \cdots \vec{\alpha}_r$ 线性无关

V 中任一向量可由 $\vec{\alpha}_1, \vec{\alpha}_2, \cdots \vec{\alpha}_r$ 线性表示

那么称 $\vec{\alpha}_1, \vec{\alpha}_2, \cdots \vec{\alpha}_r$ 为 V 的一组基，r 称为向量空间 V 的维数，即 r 维向量空间。

2.3 欧氏空间

V 是实数域上的线性空间，对于 V 中任意两个向量 $\boldsymbol{\alpha}$，$\boldsymbol{\beta}$，定义一个二元实函数，记为（$\boldsymbol{\alpha}$，$\boldsymbol{\beta}$），满足若（$\boldsymbol{\alpha}$，$\boldsymbol{\beta}$），$\forall \boldsymbol{\alpha}$，$\boldsymbol{\beta}$，$\mathbf{y} \in V$，$\forall k \in R$，

① $(\boldsymbol{\alpha}, \boldsymbol{\beta}) = (\boldsymbol{\beta}, \boldsymbol{\alpha})$ 对称性

② $(k\boldsymbol{\alpha}, \boldsymbol{\beta}) = k(\boldsymbol{\alpha}, \boldsymbol{\beta})$ 数乘

③ $(\boldsymbol{\alpha} + \boldsymbol{\beta}, \mathbf{y}) = (\boldsymbol{\alpha}, \mathbf{y}) + (\boldsymbol{\beta}, \mathbf{y})$ 可加性

④ $(\boldsymbol{\alpha}, \boldsymbol{\alpha}) \geq 0$。当且仅当 $\boldsymbol{\alpha} = 0$，$(\boldsymbol{\alpha}, \boldsymbol{\alpha}) = 0$ 正交性

称（$\boldsymbol{\alpha}$，$\boldsymbol{\beta}$）为内积运算，并称定义了这种内积的实数域 R 上的线性空间 V 为欧氏空间。

3. 矩阵

3.1 矩阵的运算（＋，－，×，数乘）

det.　T.　-1. 矩阵的相似，特征值，特征向量

1、A＋B　　**A×B**（不满足交换律）

A－b　　$\mathbf{A}^T$　　$(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$（转置）

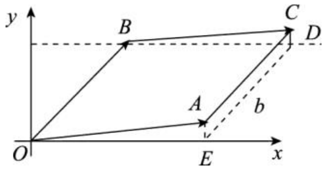
$\mathbf{A}^{-1}\mathbf{B} = \mathbf{A} \backslash \mathbf{B}$（左除）

$AB^{-1}=A/B$（右除）

2、逆矩阵（方阵才有逆矩阵）。

$\mathbf{A}\cdot\mathbf{A}^{-1}=\mathbf{I}$（则 **A** 为满秩），则称 $\mathbf{A}^{-1}$ 为 **A** 的逆矩阵。

3、det A 的物理意义：



$$\begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} \qquad \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$$

$$\begin{aligned} S_{OACB} &= S_{OEDB}+S_{CDB}-S_{AED}-S_{AEDC} \\ &= a_1b_2-a_2b_1 \end{aligned}$$

3.2 矩阵的轶

最大线性无关组（向量组）**A**：$\vec{\alpha}_i, i=1,2,\cdots m$，如果，$\exists r$，满足

①$\vec{\alpha}_i, i=1,2,\cdots r$，线性无关

②任意 $r+1$ 个向量都线性相关

则称向量组$\vec{\alpha}_1\cdots\vec{\alpha}_r$是向量组 **A** 的一个最大线性无关组，包含的向量个数称为向量组 A 的轶，记为 R_A

矩阵的轶等于行向量组的轶，也等于列向量组的轶。

3.3 特征值（向量）

定义：设 A 是 n 阶矩阵，若存在数入和非零向量 $\vec{x}$，使 $\mathbf{A}\vec{x}=\lambda\vec{x}$，成立，则称数$\lambda$为 **A** 的特征值。并称 $\vec{x}$ 为矩阵 **A** 的属于特征值λ的特征向量。

特征向量的物理意义：如果将矩阵 A 视为一个线性变换，线性变换作用在向量 $\vec{x}$ 上相当于对向量 $\vec{x}$ 进行了线性拉伸。

特征向量的应用：PCA 中主成分方向就是协方差矩阵的特征向量的方向。参考《Eigenproblems in Pattern Recognition》。

3.4 相似矩阵

定义：设 **A**、**B** 是 n 阶方阵，若存在 n 阶可逆矩阵 **P**，使得 $\mathbf{P}^{-1}\mathbf{AP}=\mathbf{B}$，则称 **A** 与 **B** 相似，**B** 称为 **A** 相似矩阵。

例：**A** 有 n 个特征向量 $\mathbf{P}_1$，$\mathbf{P}_2$，…，$\mathbf{P}_n$。则 $\mathbf{AP}=\mathbf{P}\Lambda$

取 $\mathbf{\Lambda}=\begin{bmatrix}\lambda_1 & 0 & 0 & \cdots & 0\\ 0 & \lambda_2 & 0 & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & 0 & \cdots & \lambda_n\end{bmatrix}$

$\therefore \mathbf{P}^{-1}\mathbf{A}\mathbf{P}=\mathbf{\Lambda}$，$\mathbf{A}$ 与 $\mathbf{\Lambda}$ 相似

3.5 矩阵逆的物理意义

基变换与坐标变换，R^3 为例

1）$\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，$\boldsymbol{\alpha}_3$ 为一组基，$\mathbf{b}_1$，$\mathbf{b}_2$，$\mathbf{b}_3$ 为一组新基

（$\mathbf{b}_1$，$\mathbf{b}_2$，$\mathbf{b}_3$）＝（$\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，$\boldsymbol{\alpha}_3$）$\mathbf{P}$，$\mathbf{P}=\mathbf{A}^{-1}\mathbf{B}$（旧基到新基）

2）坐标变换

旧基为（y_1，y_2，y_3）新基为（z_1，z_2，z_3），则

$$\mathbf{X}=\underset{\substack{\downarrow\\ [\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha}_2,\ \boldsymbol{\alpha}_3]}}{\mathbf{A}}\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\\ \mathbf{y}_3\end{bmatrix}=\mathbf{B}\begin{bmatrix}\mathbf{z}_1\\ \mathbf{z}_2\\ \mathbf{z}_3\end{bmatrix}\Rightarrow\begin{bmatrix}\mathbf{z}_1\\ \mathbf{z}_2\\ \mathbf{z}_3\end{bmatrix}=\underset{\substack{\downarrow\\ \mathbf{P}^{-1}}}{\mathbf{B}^{-1}\mathbf{A}}\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\\ \mathbf{y}_3\end{bmatrix}$$

3.6 线性方程组的物理意义

$$\begin{cases} a_{11}x_1 + a_{12}x_2 \cdots + a_{1n}x_n = b, 0 \\ \vdots \qquad \vdots \qquad \vdots \qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 \cdots + a_{mn}x_n = b_m, 0 \end{cases} \Rightarrow \mathbf{Ax} = \mathbf{b} / \mathbf{0}$$

$$\sum_{i=1}^{n} x_i \vec{\alpha}_i = \vec{b} \qquad \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \qquad \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$\mathbf{Ax} = \vec{0}$ 齐次
$\mathbf{Ax} = \mathbf{b}$ 非齐次
方程组

$\mathbf{Ax} = \vec{0}$ 的解空间 Rs＝n－r

R(**A**)＝r＜n　　方程组有一个含 n－r 个向量的基础解系 S_1，S_2，…，S_{n-r}

$x = \sum_{k_i} k_i s_i$ 为方程组的解。

引申：子空间方法的物理意义

4. 向量微积分

4.1 Jacobian 矩阵

$\mathbf{y} = \varphi(\mathbf{x})$　　**y** 是 m×1 维矩阵　　**x** 是 n×1 维矩阵

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial \mathbf{y_1}}{\partial \mathbf{x_1}} & \frac{\partial \mathbf{y_2}}{\partial \mathbf{x_1}} & \mathbf{L} & \frac{\partial \mathbf{y_m}}{\partial \mathbf{x_1}} \\ \mathbf{M} & \mathbf{M} & & \mathbf{M} \\ \frac{\partial \mathbf{y_1}}{\partial \mathbf{x_n}} & \frac{\partial \mathbf{y_2}}{\partial \mathbf{x_n}} & \mathbf{L} & \frac{\partial \mathbf{y_m}}{\partial \mathbf{x_n}} \end{bmatrix}$$

$$\because \mathbf{y_i} = \sum_{\mathbf{k=1}}^{\mathbf{n}} \mathbf{w_{ik}} \times \mathbf{x_k}, \quad \therefore \frac{\partial \mathbf{y_i}}{\partial \mathbf{x_j}} = \mathbf{w_{ij}}$$

即 $\mathbf{y} = \mathbf{wx}$　　　　$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \mathbf{w^T}$

mx1　mxn　nx1　　　　nxm

4.2 向量微积分常见形式

$$
\begin{array}{cc} \mathbf{y} & \dfrac{\partial \mathbf{y}}{\partial \mathbf{x}} \end{array}
$$

$$
\mathbf{y} = \boxed{\begin{array}{cc} \mathbf{Ax} & \mathbf{A}^{\mathrm{T}} \\ \mathbf{x}^{\mathrm{T}}\mathbf{A} & \mathbf{A} \\ \mathbf{x}^{\mathrm{T}}\mathbf{x} & \mathbf{2x} \\ \mathbf{x}^{\mathrm{T}}\mathbf{Ax} & \mathbf{Ax} + \mathbf{A}^{\mathrm{T}}\mathbf{x} \end{array}}
$$

$\mathbf{y} = \mathbf{x}^{\mathrm{T}}\mathbf{Ax}$ 如果 $\mathbf{A}$ 是对称矩阵

则 $\dfrac{\partial \mathbf{y}}{\partial \mathbf{x}} = \mathbf{2Ax}$

形如 $\mathbf{x}^{\mathrm{T}}\mathbf{Ax}$，可用复合函数求导证明

求得 $\mathbf{y}' = \mathbf{Ax} + \mathbf{A}^{\mathrm{T}}\mathbf{x}$

4.3 应用：ridge regression 岭回归

$$
loss = \sum_{i=1}^{n} (\mathbf{y_i} - \mathbf{x_i^T}\boldsymbol{\beta})^2 + \lambda \sum_{j=1}^{p} \boldsymbol{\beta}_j^2
$$

$$
\boldsymbol{\beta} = \arg\min \|\mathbf{y} - \mathbf{x}\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_2^2 \qquad (\boldsymbol{\beta} \in \mathbf{R}^p, \mathbf{x} \in \mathbf{R}^{n\times p}, \mathbf{y} \in \mathbf{R}^n)
$$

$$
L = (\mathbf{y} - \mathbf{x}\boldsymbol{\beta})^{\mathrm{T}}(\mathbf{y} - \mathbf{x}\boldsymbol{\beta}) + \lambda \|\boldsymbol{\beta}\|_{\mathbf{2}}^{\mathbf{2}}
$$

$$
\frac{\partial L}{\partial \boldsymbol{\beta}} = \mathbf{0} \rightarrow \mathbf{-2x}^{\mathrm{T}}(\mathbf{y} - \mathbf{x}\boldsymbol{\beta}) + 2\lambda\boldsymbol{\beta} = \mathbf{0}
$$

半正定矩阵 $\mathbf{x}^{\mathrm{T}}\mathbf{Ax} \geq 0$，则 $\mathbf{A}$ 为半正定

$$
(\mathbf{x}^{\mathrm{T}}\mathbf{x} + \lambda)\boldsymbol{\beta} = \mathbf{x}^{\mathrm{T}}\mathbf{y}
$$

$$
\boldsymbol{\beta} = (\mathbf{x}^{\mathrm{T}}\mathbf{x} + \underset{\substack{\downarrow \\ \lambda\mathbf{I}}}{\lambda})^{-1}\mathbf{x}^{\mathrm{T}}\mathbf{y}
$$

求导过程：$\dfrac{\partial L}{\partial \boldsymbol{\beta}} = \mathbf{y}^{\mathrm{T}}\mathbf{y} - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}^{\mathrm{T}}\mathbf{y} + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}^{\mathrm{T}}\mathbf{x}\boldsymbol{\beta} - \mathbf{x}\boldsymbol{\beta}\mathbf{y}^{\mathrm{T}} + \lambda\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\beta}$

$$
\rightarrow \mathbf{-x}^{\mathrm{T}}\mathbf{y} + \mathbf{2x}^{\mathrm{T}}\mathbf{x}\boldsymbol{\beta} - \mathbf{x}^{\mathrm{T}}\mathbf{y} + 2\lambda\boldsymbol{\beta}
$$

$$
\rightarrow \mathbf{2x}^{\mathrm{T}}(\mathbf{y} - \mathbf{x}\boldsymbol{\beta}) + 2\lambda\boldsymbol{\beta}
$$

线性代数习题

1、证明：$\mathbf{y} = \mathbf{w}^{\mathrm{T}}\mathbf{x} + \mathbf{b}$ 每条映射的权重，$\mathbf{w}_{ij}$ 还是 $\mathbf{w}_{ji}$ 。

2、说明向量运算为什么没有除法。

3、证明矩阵的铁等于行向量的铁，也等于列向量的铁。

4、多维 det 的物理意义。

5、$\begin{bmatrix} \cos\theta, & -\sin\theta \\ \sin\theta, & -\cos\theta \end{bmatrix}$求特征值。

6、证明：$\mathbf{y} = \mathbf{x}^{\mathrm{T}}\mathbf{A}$，$\dfrac{\partial \mathbf{y}}{\partial \mathbf{x}} = \mathbf{A}$ 。