2018/10/12 1 algebra

线性代数

一、基本知识

1. 本书中所有的向量都是列向量的形式:

$$ec{\mathbf{x}} = egin{bmatrix} x_1 \ x_2 \ dots \ x_n \end{bmatrix}$$

2. 矩阵的 \mathbf{F} 范数: 设 $\mathbf{A} = (a_{i,j})_{m \times n}$

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

它是向量的 L_2 范数的推广。

3. 矩阵的迹 $tr(\mathbf{A}) = \sum_i a_{i,i}$ 。 其性质有:

$$\circ \ ||\mathbf{A}||_F = \sqrt{tr(\mathbf{A}\mathbf{A}^T)}$$

$$\circ tr(\mathbf{A}) = tr(\mathbf{A}^T)$$

 \circ 假设 $\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{B} \in \mathbb{R}^{n \times m}$,则有:

$$tr(\mathbf{AB}) = tr(\mathbf{BA})$$

$$\circ tr(\mathbf{ABC}) = tr(\mathbf{CAB}) = tr(\mathbf{BCA})$$

二、向量操作

1. 一组向量 $\vec{\mathbf{v}}_1, \vec{\mathbf{v}}_2, \cdots, \vec{\mathbf{v}}_n$ 是线性相关的:指存在一组不全为零的实数 a_1, a_2, \cdots, a_n ,使得:

$$\sum_{i=1}^n a_i ec{\mathbf{v}}_i = ec{\mathbf{0}}$$

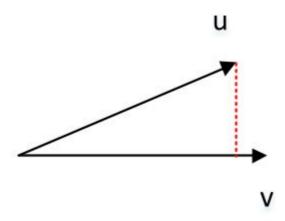
一组向量 $\vec{\mathbf{v}}_1, \vec{\mathbf{v}}_2, \cdots, \vec{\mathbf{v}}_n$ 是线性无关的,当且仅当 $a_i=0, i=1,2,\cdots,n$ 时,才有

$$\sum_{i=1}^n a_i ec{\mathbf{v}}_i = ec{\mathbf{0}}_i$$

- 2. 一个向量空间所包含的最大线性无关向量的数目,称作该向量空间的维数。
- 3. 三维向量的点积:

$$ec{\mathbf{u}}\cdotec{\mathbf{v}}=u_xv_x+u_yv_y+u_zv_z=|ec{\mathbf{u}}||ec{\mathbf{v}}|\cos(ec{\mathbf{u}},ec{\mathbf{v}})$$

2018/10/12 1 algebra



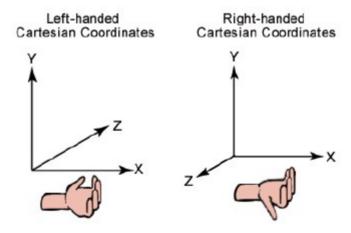
4. 三维向量的叉积:

$$ec{\mathbf{w}} = ec{\mathbf{u}} imes ec{\mathbf{v}} = egin{bmatrix} ec{\mathbf{i}} & ec{\mathbf{j}} & ec{\mathbf{k}} \ u_x & u_y & u_z \ v_x & v_y & v_z \ \end{bmatrix}$$

其中 \vec{i} , \vec{j} , \vec{k} 分别为 x, y, z 轴的单位向量。

$$ec{\mathbf{u}} = u_x ec{\mathbf{i}} + u_y ec{\mathbf{j}} + u_z ec{\mathbf{k}}, \quad ec{\mathbf{v}} = v_x ec{\mathbf{i}} + v_y ec{\mathbf{j}} + v_z ec{\mathbf{k}}$$

- \circ $\vec{\mathbf{u}}$ 和 $\vec{\mathbf{v}}$ 的叉积垂直于 $\vec{\mathbf{u}}$, $\vec{\mathbf{v}}$ 构成的平面,其方向符合右手规则。
- \circ 叉积的模等于 $\vec{\mathbf{u}}$, $\vec{\mathbf{v}}$ 构成的平行四边形的面积
- $\circ \ \vec{\mathbf{u}} \times \vec{\mathbf{v}} = -\vec{\mathbf{v}} \times \vec{\mathbf{u}}$
- $\circ \ \vec{\mathbf{u}} \times (\vec{\mathbf{v}} \times \vec{\mathbf{w}}) = (\vec{\mathbf{u}} \cdot \vec{\mathbf{w}})\vec{\mathbf{v}} (\vec{\mathbf{u}} \cdot \vec{\mathbf{v}})\vec{\mathbf{w}}$



5. 三维向量的混合积:

$$egin{aligned} \left[ec{\mathbf{u}} \ ec{\mathbf{v}} \ ec{\mathbf{w}}
ight] &= \left(ec{\mathbf{u}} imes ec{\mathbf{v}}
ight) \cdot ec{\mathbf{w}} = ec{\mathbf{u}} \cdot \left(ec{\mathbf{v}} imes ec{\mathbf{w}}
ight) \ &= egin{aligned} \left| u_x & u_y & u_z \ v_x & v_y & v_y \ w_x & w_y & w_z \ \end{array}
ight| &= egin{aligned} \left| u_x & v_x & w_x \ u_y & v_y & w_y \ u_z & v_z & w_z \ \end{array}
ight| \end{aligned}$$

- 。 其物理意义为:以 $\vec{\mathbf{u}}$, $\vec{\mathbf{v}}$, $\vec{\mathbf{w}}$ 为三个棱边所围成的平行六面体的体积。 当 $\vec{\mathbf{u}}$, $\vec{\mathbf{v}}$, $\vec{\mathbf{w}}$ 构成右手系时,该平行六面体的体积为正号。
- 6. 两个向量的并矢: 给定两个向量 $\vec{\mathbf{x}}=(x_1,x_2,\cdots,x_n)^T, \vec{\mathbf{y}}=(y_1,y_2,\cdots,y_m)^T$,则向量的并矢记作:

2018/10/12 1_algebra

$$ec{\mathbf{x}} ec{\mathbf{y}} = egin{bmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_m \ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_m \ dots & dots & \ddots & dots \ x_n y_1 & x_n y_2 & \cdots & x_n y_m \end{bmatrix}$$

也记作 $\vec{\mathbf{x}} \otimes \vec{\mathbf{y}}$ 或者 $\vec{\mathbf{x}}\vec{\mathbf{y}}^T$

三、矩阵运算

- 1. 给定两个矩阵 $\mathbf{A}=(a_{i,j})\in\mathbb{R}^{m imes n}, \mathbf{B}=(b_{i,j})\in\mathbb{R}^{m imes n}$, 定义:
 - 阿达马积 Hadamard product (又称作逐元素积)

$$\mathbf{A} \circ \mathbf{B} = egin{bmatrix} a_{1,1}b_{1,1} & a_{1,2}b_{1,2} & \cdots & a_{1,n}b_{1,n} \ a_{2,1}b_{2,1} & a_{2,2}b_{2,2} & \cdots & a_{2,n}b_{2,n} \ dots & dots & \ddots & dots \ a_{m,1}b_{m,1} & a_{m,2}b_{m,2} & \cdots & a_{m,n}b_{m,n} \end{bmatrix}$$

○ 克罗内积 Kronnecker product :

$$\mathbf{A}\otimes\mathbf{B}=egin{bmatrix} a_{1,1}\mathbf{B} & a_{1,2}\mathbf{B} & \cdots & a_{1,n}\mathbf{B} \ a_{2,1}\mathbf{B} & a_{2,2}\mathbf{B} & \cdots & a_{2,n}\mathbf{B} \ dots & dots & \ddots & dots \ a_{m,1}\mathbf{B} & a_{m,2}\mathbf{B} & \cdots & a_{m,n}\mathbf{B} \end{bmatrix}$$

2. 设 \vec{x} , \vec{a} , \vec{b} , \vec{c} 为 n 阶向量, A, B, C, X 为 n 阶方阵, 则

$$\frac{\partial (\vec{\mathbf{a}}^T\vec{\mathbf{x}})}{\partial \vec{\mathbf{x}}} = \frac{\partial (\vec{\mathbf{x}}^T\vec{\mathbf{a}})}{\partial \vec{\mathbf{x}}} = \vec{\mathbf{a}}$$

$$\frac{\partial (\vec{\mathbf{a}}^T \mathbf{X} \vec{\mathbf{b}})}{\partial \mathbf{X}} = \vec{\mathbf{a}} \vec{\mathbf{b}}^T = \vec{\mathbf{a}} \otimes \vec{\mathbf{b}} \in \mathbb{R}^{n \times n}$$

$$rac{\partial (ec{\mathbf{a}}^T \mathbf{X}^T ec{\mathbf{b}})}{\partial \mathbf{X}} = ec{\mathbf{b}} ec{\mathbf{a}}^T = ec{\mathbf{b}} \otimes ec{\mathbf{a}} \in \mathbb{R}^{n imes n}$$

$$rac{\partial (ec{\mathbf{a}}^T \mathbf{X} ec{\mathbf{a}})}{\partial \mathbf{X}} = rac{\partial (ec{\mathbf{a}}^T \mathbf{X}^T ec{\mathbf{a}})}{\partial \mathbf{X}} = ec{\mathbf{a}} \otimes ec{\mathbf{a}}$$

$$\frac{\partial (\vec{\mathbf{a}}^T \mathbf{X}^T \mathbf{X} \vec{\mathbf{b}})}{\partial \mathbf{X}} = \mathbf{X} (\vec{\mathbf{a}} \otimes \vec{\mathbf{b}} + \vec{\mathbf{b}} \otimes \vec{\mathbf{a}})$$

$$\frac{\partial [(\mathbf{A}\vec{\mathbf{x}} + \vec{\mathbf{a}})^T \mathbf{C} (\mathbf{B}\vec{\mathbf{x}} + \vec{\mathbf{b}})]}{\partial \vec{\mathbf{x}}} = \mathbf{A}^T \mathbf{C} (\mathbf{B}\vec{\mathbf{x}} + \vec{\mathbf{b}}) + \mathbf{B}^T \mathbf{C} (\mathbf{A}\vec{\mathbf{x}} + \vec{\mathbf{a}})$$

2018/10/12 1 algebra

$$rac{\partial (ec{\mathbf{x}}^T \mathbf{A} ec{\mathbf{x}})}{\partial ec{\mathbf{x}}} = (\mathbf{A} + \mathbf{A}^T) ec{\mathbf{x}}$$

$$\frac{\partial [(\mathbf{X}\vec{\mathbf{b}} + \vec{\mathbf{c}})^T \mathbf{A} (\mathbf{X}\vec{\mathbf{b}} + \vec{\mathbf{c}})]}{\partial \mathbf{X}} = (\mathbf{A} + \mathbf{A}^T)(\mathbf{X}\vec{\mathbf{b}} + \vec{\mathbf{c}})\vec{\mathbf{b}}^T$$

$$\frac{\partial (\vec{\mathbf{b}}^T \mathbf{X}^T \mathbf{A} \mathbf{X} \vec{\mathbf{c}})}{\partial \mathbf{X}} = \mathbf{A}^T \mathbf{X} \vec{\mathbf{b}} \vec{\mathbf{c}}^T + \mathbf{A} \mathbf{X} \vec{\mathbf{c}} \vec{\mathbf{b}}^T$$

- 3. 如果 *f* 是一元函数,则:
 - 。 其逐元向量函数为:

$$f(\vec{\mathbf{x}}) = (f(x_1), f(x_2), \cdots, f(x_n))^T$$

。 其逐矩阵函数为:

$$f(\mathbf{X}) = [f(x_{i,j})]$$

。 其逐元导数分别为:

$$f'(\vec{\mathbf{x}}) = (f'(x1), f'(x2), \cdots, f'(x_n))^T$$

 $f'(\mathbf{X}) = [f'(x_{i,j})]$

- 4. 各种类型的偏导数:
 - 标量对标量的偏导数

$$\frac{\partial u}{\partial v}$$

○ 标量对向量 (n 维向量) 的偏导数

$$\frac{\partial u}{\partial \vec{\mathbf{v}}} = (\frac{\partial u}{\partial v_1}, \frac{\partial u}{\partial v_2}, \cdots, \frac{\partial u}{\partial v_n})^T$$

 \circ 标量对矩阵($m \times n$ 阶矩阵)的偏导数

$$\frac{\partial u}{\partial \mathbf{V}} = \begin{bmatrix} \frac{\partial u}{\partial V_{1,1}} & \frac{\partial u}{\partial V_{1,2}} & \cdots & \frac{\partial u}{\partial V_{1,n}} \\ \frac{\partial u}{\partial V_{2,1}} & \frac{\partial u}{\partial V_{2,2}} & \cdots & \frac{\partial u}{\partial V_{2,n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial u}{\partial V_{m,1}} & \frac{\partial u}{\partial V_{m,2}} & \cdots & \frac{\partial u}{\partial V_{m,n}} \end{bmatrix}$$

○ 向量 (m 维向量) 对标量的偏导数

$$\frac{\partial \vec{\mathbf{u}}}{\partial v} = (\frac{\partial u_1}{\partial v}, \frac{\partial u_2}{\partial v}, \cdots, \frac{\partial u_m}{\partial v})^T$$

2018/10/12 1_algebra

○ 向量 (*m* 维向量) 对向量 (*n* 维向量) 的偏导数 (雅可比矩阵, 行优先)

$$egin{aligned} rac{\partial ec{\mathbf{u}}}{\partial ec{\mathbf{v}}} = egin{bmatrix} rac{\partial u_1}{\partial v_1} & rac{\partial u_1}{\partial v_2} & \cdots & rac{\partial u_1}{\partial v_n} \ rac{\partial u_2}{\partial v_1} & rac{\partial u_2}{\partial v_2} & \cdots & rac{\partial u_2}{\partial v_n} \ dots & dots & \ddots & dots \ rac{\partial u_m}{\partial v_1} & rac{\partial u_m}{\partial v_2} & \cdots & rac{\partial u_m}{\partial v_n} \ \end{bmatrix}$$

如果为列优先,则为上面矩阵的转置

• 矩阵(m×n) 阶矩阵)对标量的偏导数

$$rac{\partial \mathbf{U}}{\partial v} = egin{bmatrix} rac{\partial U_{1,1}}{\partial v} & rac{\partial U_{1,2}}{\partial v} & \cdots & rac{\partial U_{1,n}}{\partial v} \\ rac{\partial U_{2,1}}{\partial v} & rac{\partial U_{2,2}}{\partial v} & \cdots & rac{\partial U_{2,n}}{\partial v} \\ dots & dots & \ddots & dots \\ rac{\partial U_{m,1}}{\partial v} & rac{\partial U_{m,2}}{\partial v} & \cdots & rac{\partial U_{m,n}}{\partial v} \end{bmatrix}$$

- 更复杂的情况依次类推。对于 $\frac{\partial \mathbf{u}}{\partial \mathbf{v}}$ 。根据 numpy 的术语:
 - \circ 假设 \mathbf{u} 的 ndim (维度) 为 d_u

对于标量, ndim 为 0;对于向量, ndim 为1;对于矩阵, ndim 为 2

- 假设 v 的 ndim 为 d_r
- \circ 则 $rac{\partial \mathbf{u}}{\partial \mathbf{v}}$ 的 ndim 为 $d_u + d_v$
- 5. 对于矩阵的迹,有下列偏导数成立:

$$rac{\partial [tr(f(\mathbf{X}))]}{\partial \mathbf{X}} = (f'(\mathbf{X}))^T$$

$$\frac{\partial [tr(\mathbf{AXB})]}{\partial \mathbf{X}} = \mathbf{A}^T \mathbf{B}^T$$

$$rac{\partial [tr(\mathbf{A}\mathbf{X}^T\mathbf{B})]}{\partial \mathbf{X}} = \mathbf{B}\mathbf{A}$$

$$\frac{\partial [tr(\mathbf{A} \otimes \mathbf{X})]}{\partial \mathbf{X}} = tr(\mathbf{A})\mathbf{I}$$

$$\frac{\partial [tr(\mathbf{A}\mathbf{X}\mathbf{B}\mathbf{X})]}{\partial \mathbf{X}} = \mathbf{A}^T\mathbf{X}^T\mathbf{B}^T + \mathbf{B}^T\mathbf{X}\mathbf{A}^T$$

$$\frac{\partial [tr(\mathbf{X}^T\mathbf{B}\mathbf{X}\mathbf{C})]}{\partial \mathbf{X}} = (\mathbf{B}^T + \mathbf{B})\mathbf{X}\mathbf{C}\mathbf{C}^T$$

2018/10/12 1_algebra

$$\frac{\partial [tr(\mathbf{C}^T\mathbf{X}^T\mathbf{B}\mathbf{X}\mathbf{C})]}{\partial \mathbf{X}} = \mathbf{B}\mathbf{X}\mathbf{C} + \mathbf{B}^T\mathbf{X}\mathbf{C}^T$$

$$\frac{\partial [tr(\mathbf{A}\mathbf{X}\mathbf{B}\mathbf{X}^T\mathbf{C})]}{\partial \mathbf{X}} = \mathbf{A}^T\mathbf{C}^T\mathbf{X}\mathbf{B}^T + \mathbf{C}\mathbf{A}\mathbf{X}\mathbf{B}$$

$$\frac{\partial [tr((\mathbf{A}\mathbf{X}\mathbf{B}+\mathbf{C})(\mathbf{A}\mathbf{X}\mathbf{B}+\mathbf{C}))]}{\partial \mathbf{X}} = 2\mathbf{A}^T(\mathbf{A}\mathbf{X}\mathbf{B}+\mathbf{C})\mathbf{B}^T$$

6. 假设 $\mathbf{U} = \mathbf{f}(\mathbf{X})$ 是关于 \mathbf{X} 的矩阵值函数 $(f: \mathbb{R}^{m \times n} \to \mathbb{R}^{m \times n})$,且 $g(\mathbf{U})$ 是关于 \mathbf{U} 的实值函数 ($g: \mathbb{R}^{m \times n} \to \mathbb{R}$) ,则下面链式法则成立:

$$egin{aligned} rac{\partial g(\mathbf{U})}{\partial \mathbf{X}} &= \left(rac{\partial g(\mathbf{U})}{\partial x_{1,1}}
ight) = egin{bmatrix} rac{\partial g(\mathbf{U})}{\partial x_{1,1}} & rac{\partial g(\mathbf{U})}{\partial x_{1,2}} & \cdots & rac{\partial g(\mathbf{U})}{\partial x_{1,n}} \ rac{\partial g(\mathbf{U})}{\partial x_{2,1}} & rac{\partial g(\mathbf{U})}{\partial x_{2,2}} & \cdots & rac{\partial g(\mathbf{U})}{\partial x_{2,n}} \ dots & dots & \ddots & dots \ rac{\partial g(\mathbf{U})}{\partial x_{m,1}} & rac{\partial g(\mathbf{U})}{\partial x_{m,2}} & \cdots & rac{\partial g(\mathbf{U})}{\partial x_{m,n}} \ \end{bmatrix} \ &= \left(\sum_k \sum_l rac{\partial g(\mathbf{U})}{\partial u_{k,l}} rac{\partial g(\mathbf{U})}{\partial x_{i,j}}
ight) \ &= tr\left[\left(rac{\partial g(\mathbf{U})}{\partial \mathbf{U}}
ight)^T rac{\partial \mathbf{U}}{\partial x_{i,j}}
ight] \end{aligned}$$