线性代数

机器学习基础知识

一、基本知识

1.本书中所有的向量都是列向量的形式：

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix}$$

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

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

它是向量的 L_2 范数的推广。

3.矩阵的迹 $tr(A) = \sum_i a_{i,i}$.其性质有：

- $||A||_F = \sqrt{tr(AA^T)}$
- $tr(A) = tr(A^T)$
- 假设 $A \in R^{m\times n}, B \in R^{n\times m}$ ，则有：

$$tr(AB) = tr(BA)$$

- $tr(ABC) = tr(CAB) = tr(BCA)$

二、向量操作

1.一组向量 $\overrightarrow{V_1}$ 、$\overrightarrow{V_2}$ 、... 、$\overrightarrow{V_n}$ 是**线性相关**的：是指存在一组不全为零的实数 a_1 、a_2 、... 、a_n 、使得：

$$\sum_{i=1}^{n} a_i \overrightarrow{V_i} = \vec{0}$$

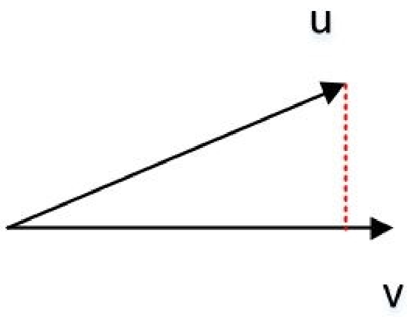
一组向量 $\overrightarrow{V_1}$ 、$\overrightarrow{V_2}$ 、... 、$\overrightarrow{V_n}$ 是**线性无关**的：当且仅当 $a_i = 0, i = 1, 2, \ldots, n$ 时，才有：

$$\sum_{i=1}^{n} a_i \overrightarrow{V_i} = \vec{0}$$

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。**

3.三维向量的点集：

$$\vec{u} \cdot \vec{v} = u_x v_x + u_y v_y + u_z v_z = |\vec{u}||\vec{v}|cos(\vec{u}, \vec{v})$$



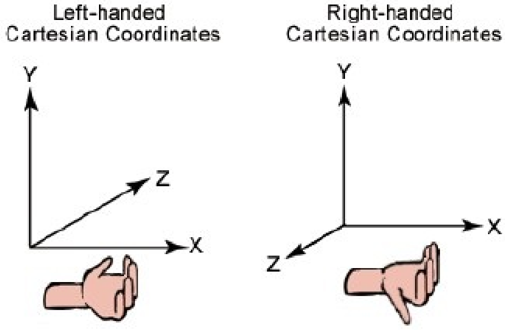
4.三维向量的叉积：

$$\vec{W} = \vec{u} \times \vec{v} = \begin{bmatrix} \vec{i} & \vec{j} & \vec{k} \\ u_x & u_y & u_z \\ v_x & v_y & v_z \end{bmatrix}$$

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

$$\vec{u} = u_x \vec{i} + u_y \vec{j} + u_z \vec{k}$$
$$\vec{v} = v_x \vec{i} + u_y \vec{j} + u_z \vec{k}$$

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



5.三维向量的混合积：

$$[\vec{u}\vec{v}\vec{w}] = (\vec{u} \times \vec{v}) \cdot \vec{w} = \vec{u} \cdot (\vec{v} \times \vec{w})$$

$$= \begin{vmatrix} u_x & u_y & u_z \\ v_x & v_y & v_z \\ w_x & w_y & w_z \end{vmatrix}$$
$$= \begin{vmatrix} u_x & v_x & w_x \\ u_y & v_y & w_y \\ u_z & v_z & w_z \end{vmatrix}$$

- 其物理意义为：以 $\vec{u}$ ，$\vec{v}$ ，$\vec{w}$ 为三个棱边所围成的平行六面体的体积。当 $\vec{u}$ ，$\vec{v}$ ，$\vec{w}$ 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 $\vec{x} = (x_1, x_2, \ldots, x_n)^T, \vec{y} = (y_1, y_2, \ldots, y_n)^T$ ，则向量的并矢记作：

$$\vec{x}\vec{y} = \begin{bmatrix} x_1y_1 & x_1y_2 & \ldots & x_1y_m \\ x_2y_1 & x_2y_2 & \ldots & x_2y_m \\ \ldots & \ldots & \ldots & \ldots \\ x_ny_1 & x_ny_2 & \ldots & x_ny_m \end{bmatrix}$$

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

三、矩阵运算

1.给定两个矩阵 $A = (a_{i,j}) \in R^{m\times n}$, $B = (b_{i,j}) \in R^{m\times n}$,定义：

- 阿达马积（Hadamard Product）(又称作逐元素积)：

$$A \circ B = \begin{bmatrix} a_{1,1}b_{1,1} & a_{1,2}b_{1,2} & \ldots & a_{1,n}b_{1,n} \\ a_{2,1}b_{2,1} & a_{2,2}b_{2,2} & \ldots & a_{2,n}b_{2,n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m,1}b_{m,1} & a_{m,2}b_{m,2} & \ldots & a_{m,n}b_{m,n} \end{bmatrix}$$

- 克罗内积（Kronnecker Product）：

$$A \otimes B = \begin{bmatrix} a_{1,1}B & a_{1,2}B & \ldots & a_{1,n}B \\ a_{2,1}B & a_{2,2}B & \ldots & a_{2,n}B \\ \ldots & \ldots & \ldots & \ldots \\ a_{m,1}B & a_{m,2}B & \ldots & a_{m,n}B \end{bmatrix}$$

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：

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

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

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

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

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

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

$$\frac{\partial(\vec{x}^T A \vec{x})}{\partial X} = (A + A^T)\vec{x}$$

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

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

3.如果 f 是一元函数，则：

- 其逐元向量函数为：

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

- 其逐矩阵函数为：

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

- 其逐元导数分别为：

$$f'(\vec{x}) = (f'(x_1), f'(x_2), \ldots, f'(x_n))^T$$

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

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：

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

- 标量对矩阵（$m \times n$ 阶矩阵）的偏导数：

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

- 向量（ m 维向量）对标量的偏导数：

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

- 向量（m 维向量）对向量（n 维向量）的偏导数（雅克比矩阵，行优先）：

$$\frac{\partial \vec{u}}{\partial \vec{v}}=\begin{bmatrix}\frac{\partial u_1}{\partial v_1} & \frac{\partial u_1}{\partial v_2} & \cdots & \frac{\partial u_1}{\partial v_n} \\ \frac{\partial u_2}{\partial v_1} & \frac{\partial u_2}{\partial v_2} & \cdots & \frac{\partial u_2}{\partial v_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial u_m}{\partial v_1} & \frac{\partial u_m}{\partial v_2} & \cdots & \frac{\partial u_m}{\partial v_n}\end{bmatrix}$$

如果为列优先，则为上面矩阵的转置。

- 矩阵（$m\times n$ 阶矩阵）对标量的偏导数：

$$\frac{\partial U}{\partial v}=\begin{bmatrix}\frac{\partial U_{1,1}}{\partial v} & \frac{\partial U_{1,2}}{\partial v} & \cdots & \frac{\partial U_{1,n}}{\partial v} \\ \frac{\partial U_{2,1}}{\partial v} & \frac{\partial U_{2,2}}{\partial v} & \cdots & \frac{\partial U_{2,n}}{\partial v} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial U_{m,1}}{\partial v} & \frac{\partial U_{m,2}}{\partial v} & \cdots & \frac{\partial U_{m,n}}{\partial v}\end{bmatrix}$$

- 更复杂的情况依次类推。对于 $\frac{\partial u}{\partial v}$ 。根据 $numpy$ 的术语：
  -假设 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_v:
  则 $\frac{\partial u}{\partial v}$ 的 $ndim$ 为 d_u+d_v.

5.对于矩阵的迹，有下列偏导数成立：

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

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

$$\frac{\partial[tr(AX^TB)]}{\partial X}=BA$$

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

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

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

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

$$\frac{\partial[tr(AXBX^T C)]}{\partial X} = A^T C^T XB^T + CAXB$$

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

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

$$\frac{\partial g(U))}{\partial X} = (\frac{\partial g(U)}{\partial x_{i,j}}) = \begin{bmatrix} \frac{\partial g(U)}{\partial x_{1,1}} & \frac{\partial g(U)}{\partial x_{1,2}} & \cdots & \frac{\partial g(U)}{\partial x_{1,n}} \\ \frac{\partial g(U)}{\partial x_{2,j}} & \frac{\partial g(U)}{\partial x_{2,2}} & \cdots & \frac{\partial g(U)}{\partial x_{2,n}} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial g(U)}{\partial x_{m,1}} & \frac{\partial g(U)}{\partial x_{m,2}} & \cdots & \frac{\partial g(U)}{\partial x_{m,n}} \end{bmatrix} = (\sum_k \sum_l \frac{\partial g(U)}{\partial u_{k,l}} \frac{\partial u_{k,l}}{\partial x_{i,j}}) = tr[(\frac{\partial g(U)}{\partial U})^T \frac{\partial U}{\partial x_{i,j}}$$