作者华校专，曾任阿里巴巴资深算法工程师，现任智易科技首席算法研究员，《Python 大战机器学习》的作者。

这是作者多年以来学习总结的笔记，经整理之后开源于世。目前还有约一半的内容在陆续整理中，已经整理好的内容放置在此。 曾有出版社约稿，但是考虑到出版时间周期较长，而且书本购买成本高不利于技术广泛传播，因此作者采取开源的形式。 笔记内容仅供个人学习使用，非本人同意不得应用于商业领域。

笔记内容较多，可能有些总结的不到位的地方，欢迎大家探讨。联系方式:huaxz1986@163.com

数学基础

统计学习

深度学习

自然语言处理

计算机视觉

工具

CRF

lightgbm

xgboost

scikit-learn

spark

numpy

- [numpy 使用指南]
    - 一、 ndarray
    - 二、 ufunc 函数
    - 三、 函数库
    - 四、数组的存储和加载

scipy

- [scipy 使用指南]
    - 一、 常数和特殊函数
    - 二、 拟合与优化
    - 三、线性代数
    - 四、 统计
    - 五、数值积分
    - 六、 稀疏矩阵

matplotlib

- [matplotlib 使用指南]
    - 一、matplotlib配置
    - 二、 matplotlib Artist
    - 三、基本概念
    - 四、布局
    - 五、 Path
    - 六、 path effect
    - 七、坐标变换
    - 八、 3D 绘图
    - 九、技巧

pandas

- [pandas 使用指南]
    - 一、基本数据结构
    - 二、 内部数据结构
    - 三、 下标存取
    - 四、 运算
    - 五、变换
    - 六、数据清洗
    - 七、 字符串操作
    - 八、 聚合与分组
    - 九、时间序列
    - 十、 DataFrame 绘图
    - 十一、 移动窗口函数
    - 十二、 数据加载和保存

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：

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

2. 矩阵的`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：
   - $||\mathbf{A}||_F=\sqrt{tr(\mathbf{A}\mathbf{A}^T)}$
   - $tr(\mathbf{A})=tr(\mathbf{A}^T)$
   - 假设 $\mathbf{A}\in\mathbb{R}^{m\times n},\mathbf{B}\in\mathbb{R}^{n\times m}$ ，则有：

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

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

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：

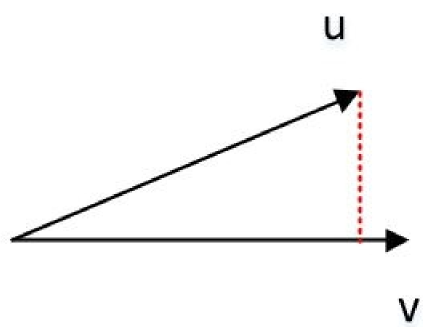
$$\sum_{i=1}^n a_i\vec{\mathbf{v}}_i=\vec{\mathbf{0}}$$

   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

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

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. 三维向量的点积：

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



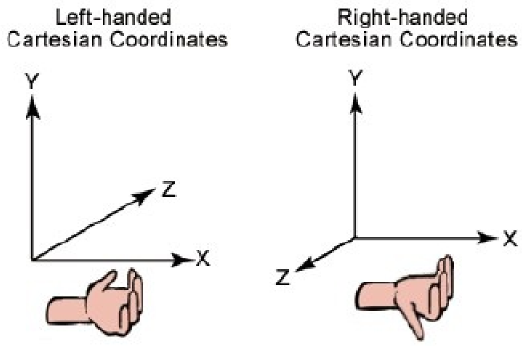
4. 三维向量的叉积：

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

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

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

- $\vec{\mathbf{u}}$ 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。
- 叉积的模等于 $\vec{\mathbf{u}},\vec{\mathbf{v}}$ 构成的平行四边形的面积
- $\vec{\mathbf{u}}\times\vec{\mathbf{v}}=-\vec{\mathbf{v}}\times\vec{\mathbf{u}}$
- $\vec{\mathbf{u}}\times(\vec{\mathbf{v}}\times\vec{\mathbf{w}})=(\vec{\mathbf{u}}\cdot\vec{\mathbf{w}})\vec{\mathbf{v}}-(\vec{\mathbf{u}}\cdot\vec{\mathbf{v}})\vec{\mathbf{w}}$



5. 三维向量的混合积：

$$[\vec{\mathbf{u}}\;\vec{\mathbf{v}}\;\vec{\mathbf{w}}]=(\vec{\mathbf{u}}\times\vec{\mathbf{v}})\cdot\vec{\mathbf{w}}=\vec{\mathbf{u}}\cdot(\vec{\mathbf{v}}\times\vec{\mathbf{w}})$$
$$=\begin{vmatrix}u_x & u_y & u_z\\ v_x & v_y & v_z\\ w_x & w_y & w_z\end{vmatrix}=\begin{vmatrix}u_x & v_x & w_x\\ u_y & v_y & w_y\\ u_z & v_z & w_z\end{vmatrix}$$

   - 其物理意义为：以 $\vec{\mathbf{u}},\vec{\mathbf{v}},\vec{\mathbf{w}}$ 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：

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

   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 $\mathbf{A}=(a_{i,j})\in\mathbb{R}^{m\times n},\mathbf{B}=(b_{i,j})\in\mathbb{R}^{m\times n}$ ，定义：
   - 阿达马积`Hadamard product`（又称作逐元素积）：

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

   - 克罗内积`Kronnecker product`：

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

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

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

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

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

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

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

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

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

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

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

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

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

    - 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

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

    - 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）

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

> 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

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

- 更复杂的情况依次类推。对于 $\frac{\partial \mathbf{u}}{\partial \mathbf{v}}$。根据`numpy`的术语：
  - 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 $\mathbf{v}$ 的 `ndim`为 d_v

  - 则 $\frac{\partial \mathbf{u}}{\partial \mathbf{v}}$ 的 `ndim`为 $d_u + d_v$

5. 对于矩阵的迹，有下列偏导数成立：

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

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

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

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

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

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

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

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

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

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

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