Preface

写在前面的话

此笔记是基于 某个笔记 增删该而来，如有侵权，请联系我哦! 当然有错误也可以联系我哦！

目前此笔记尚未开源，如有用于商业用途请联系我！

联系方式：发送邮件

预备知识

向量Vector

- 线性相关

存在一组 $a_1, a_2, ...a_k$ 不全为零的数，使得 $a_1\vec{x}_1 + a_2\vec{x}_2 + ... + a_k\vec{x}_k = \vec{0}$ ，那么可以称这组向量 $\vec{x}_1, \vec{x}_2, ..., \vec{x}_k$ 是线性相关的。

- 线性无关

当且仅当 $a_1, a_2, ...a_k$ 全都为0时， $a_1\vec{x}_1 + a_2\vec{x}_2 + ... + a_k\vec{x}_k = \vec{0}$ 才成立，，那么可以称这组向量 $\vec{x}_1, \vec{x}_2, ..., \vec{x}_k$ 是线性无关的。

- 极大线性无关组

如果线性无关的 $\vec{x}_1, \vec{x}_2, ..., \vec{x}_k$ 是向量组 $\vec{x}$ 部分组，且 $\vec{x}$ 中任一向量都可以用 $\vec{x}_1, \vec{x}_2, ..., \vec{x}_k$ 表示，那么 $\vec{x}_1, \vec{x}_2, ..., \vec{x}_k$ 就是一个极大线性无关组或最大线性无关组。

- 向量运算

内积 $\vec{a} \cdot \vec{b} = \sum a_i b_i$

叉积 $\vec{a} \times \vec{b} = \begin{bmatrix} \vec{i} & \vec{j} & \vec{k} \\ a_x & a_y & a_z \\ b_x & b_y & b_z \end{bmatrix}$

- 范数
  - $\|x\|_1 = \sum |x_i|$
  - $\|x\|_2 = \sqrt{\sum_i x_i^2}$
  - $\|x\|_\infty = \max\{|x_i|\}$
  - $\|x\|_p = (\sum |x_i|^p)^{\frac{1}{p}}$
  - 向量范数的性质：1) 非负性；2) 齐次性；3) 三角不等性

矩阵Matrix

- 矩阵转置

$$A^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{bmatrix}$$

$$(AB)^T = B^T A^T$$

$$(A+B)^T = A^T + B^T$$

$$(kA)^T = kA^T$$

- 共轭转置

$$A^H = (\bar{A})^T$$

eg.

$$\begin{pmatrix} 1 & 2+i \\ 1-i & 2 \end{pmatrix}^H = \begin{pmatrix} 1 & 1+i \\ 2-i & 2 \end{pmatrix}$$

- 伴随矩阵

$$A^* = \begin{bmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{bmatrix}$$

$A_{ij} = (-1)^{i+j} M_{ij}$

- 矩阵的迹

$$tr(A) = \sum_i^n a_{ii} = \sum_i^n \lambda_i$$

$$a = tr(a)$$

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

$$tr(A+B) = tr(A) + tr(B)$$

$$tr(A) = tr(A^T)$$

$$tr(A^T B) = \sum_{i,j} A_{ij} B_{ij}$$

$$tr(A^T (B \odot C)) = tr((A \odot B)^T C)$$

- 范数

$$\|A\|_F = \sqrt{\sum_{i,j} a_{ij}^2} = tr(AA^T)$$

$$\|A\|_2 = \sqrt{\lambda_{max}(A^TA)} = \delta_{max}(A)$$

$$\|A\|_1 = \max_j \sum_{i=1}^{n} |a_{ij}|$$

$$\|A\|_\infty = \max_i \sum_{j=1}^{n} |a_{ij}|$$

$$\|A\|_* = \sum \delta(A)$$

$$\|A\|_p = \max_{\|x\|_p=1} \|Ax\|_p$$

- 标准正交矩阵$U = [\alpha_1, \alpha_2, \cdots, \alpha_n]$

$$\alpha_i^T \alpha_j = \begin{cases} 1, & i \neq j \\ 0, & i = j \end{cases}$$

满足如下性质：

1. $U^{-1} = U^T$
2. $rank(U) = n$
3. $U^TU = UU^T = E$
4. $\|U \cdot A\| = \|A\|$

- 部分列正交矩阵$U = [U_1, U_2]$

$U_1 = [u_1, u_2, \cdots, u_r] \in R^{n\times r}$

$u_i^T u_j = \begin{cases} 1, & i \neq j \\ 0, & i = j \end{cases}$

$U_1^T U_1 = E_{r\times r}$

$UU^T = [U_1, U_2] \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = U_1U_1^T + U_2U_2^T = E$

U_2是正交补

- 正交化

有一组向量$a_1, a_2, \cdots, a_n$ 寻找$q_1, q_2, \cdots, q_n$使得

$$span\{a_1, a_2, \cdots, a_n\} = span\{q_1, q_2, \cdots, q_n\}$$

且$Q=[q_1,q_2,\cdots,q_n]$是标准正交矩阵

`Gram-Schmidt`

1. $span\{q_1\}=span\{a_1\}$，则$q_1=\frac{a_1}{\|a_1\|}$
2. 假设$span\{a_1,a_2,\cdots,a_k\}=span\{q_1,q_2,\cdots,q_k\}$且满足$q_i\perp q_j,i\neq j,\|q_i\|=1$
3. 如何构造q_{k+1}使其满足：$\begin{cases} span\{q_1,q_2,\cdots,q_k\}\oplus span\{q_{k+1}\}=span\{a_1,a_2,\cdots,a_{k+1}\} \\ q_{k+1}\perp q_i,i=1,2,\cdots,k \\ \|q_{k+1}\|=1 \end{cases}$

   $a_{k+1}=\sum_{i=1}^{k+1}r_{i,k+1}q_i\Rightarrow q_i^Ta_{k+1}=r_{i,k+1}q_i^Tq_i$

   $q_{k+1}=\frac{a_{k+1}-\sum_{i=1}^{k}r_{i,k+1}q_i}{\|a_{k+1}-\sum_{i=1}^{k}r_{i,k+1}q_i\|}$，其中$r_{i,k+1}=q_i^Ta_{k+1},i=1,2,3,\cdots,k$

从 `QR Decomposition` 看 `Gram-Schmidt` ：

$$[a_1,a_2,\cdots,a_n]=[q_1,q_2,\cdots,q_n]\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_{nn} \end{bmatrix}$$

$$\Rightarrow\begin{cases} a_1=r_{11}q_1 \\ a_2=r_{12}q_1+r_{22}q_2 \\ \vdots \end{cases}\Rightarrow\begin{cases} q_1=\dfrac{a_1}{\|a_1\|} \\ q_2=\dfrac{a_2-r_{12}q_1}{\|a_2-r_{12}q_1\|} \\ \vdots \end{cases}$$

`Arnoldi` 分解：在$\mathrm{K}_k(A,r_0)=span\{r_0,Ar_0,\cdots,A^{k-1}r_0\}$上运用 `Gram-Schmidt`

1. 令$v_1=\frac{r_0}{\|r_0\|}$
2. 假设已构造$\{v_1,v_2,\cdots,v_k\},v_k\in\mathrm{K}_k(A,r_0),q_i\perp q_j,i\neq j,\|v_i\|=1$
3. 如何构造v_{k+1}使其满足：$\begin{cases} v_{k+1}\in\mathrm{K}_{k+1}(A,r_0) \\ v_{k+1}\perp v_i,i=1,2,\cdots,k \\ \|v_{k+1}\|=1 \end{cases}$

   $v_{k+1}\in\mathrm{K}_{k+1}(A,r_0)=\mathrm{K}_k(A,r_0)\oplus span\{Av_k\}=span\{v_1,v_2,\cdots,v_{k+1}\}$
   $Av_k=\sum_{i=1}^{k+1}h_{i,k}v_i\Rightarrow v_i^TAv_k=h_{i,k}v_i^Tv_i$
   $v_{k+1}=\frac{Av_k-\sum_{i=1}^{k}h_{i,k}v_i}{\|Av_k-\sum_{i=1}^{k}h_{i,k}v_i\|}$，其中$h_{i,k}=v_i^TAv_k,i=1,2,3,\cdots,k$

- A-正交
  $$

矩阵分解

矩阵微分

导数

- 标量对向量的求导

对于标量$f, \vec{x}_{(n\times 1)}$ 有：

$$\frac{\partial f}{\partial \vec{x}} = \left[\frac{\partial f}{\partial x_i}\right]$$

- 标量对矩阵的求导

对于标量$f, X_{(m\times n)}$ 有：

$$\frac{\partial f}{\partial X} = \left[\frac{\partial f}{\partial x_{ij}}\right]$$

我们知道标量对标量的梯度gradient和微分differentiation有这样的关系：

$$df = f'(x)dx$$

$$df = \sum_i \frac{\partial f}{\partial x_i} dx_i = \frac{\partial f}{\partial \vec{x}}^T d\vec{x}$$

那么标量对矩阵也存在：

$$df = \sum_{ij} \frac{\partial f}{\partial x_{ij}} dx_{ij} = tr(\frac{\partial f}{\partial X}^T dX)$$

例子1：已知$f = |X|$，求df？
我们知道

$$|X| = \sum_i x_{ij} A_{ij} (A_{ij}\text{代数余子式})$$

将上式代入得：

$$\frac{\partial f}{\partial X} = \left[\frac{\partial \sum_k x_{kj} A_{kj}}{\partial x_{ij}}\right] = [A_{ij}] = (X^*)^T$$

因此有

$$df = tr(\frac{\partial f}{\partial X}^T dX) = tr(X^* dX) = |X| tr(X^{-1} dX)$$

例子2：求dX^{-1}？

我们知道

$$XX^{-1} = E$$

对等式两边微分有

$$dXX^{-1} = dE$$

$$XdX^{-1} = -X^{-1}dX$$

因此有

$$dX^{-1} = -X^{-1}dXX^{-1}$$

例子3：$f = \vec{a}^T X \vec{b}$，求 $\frac{\partial f}{\partial X}$?

$$df = \vec{a}^T dX \vec{b} = tr(\vec{a}^T dX \vec{b}) = tr(\vec{b}\vec{a}^T dX) = tr(\frac{\partial f}{\partial X}^T dX)$$

因此 $\frac{\partial f}{\partial X} = \vec{a}\vec{b}^T$

- 复合法则

已知$f = g(Y)$，且 $Y = h(X)$，怎么求 $\frac{\partial f}{\partial X}$?(其中g和h都是逐元素的函数)

$$df = tr(\frac{\partial f}{\partial Y}^T dY) = tr(\frac{\partial f}{\partial Y}^T (h'(X) \odot dX)) = tr((\frac{\partial f}{\partial Y} \odot h'(X))^T dX)$$

例子4：$loss = -\vec{y}^T \log\ softmax(W\vec{x})$，求 $\frac{\partial\, loss}{\partial W}$。$\vec{y}$ 是只有一个元素为1其余元素为0的向量。

$$softmax(\vec{x}) = \frac{e^{\vec{x}}}{\vec{1}^T e^{\vec{x}}}$$

$$loss = -\vec{y}^T W\vec{x} + (\vec{y}^T\vec{1}) \log(\vec{1}^T e^{W\vec{x}}) \tag{*}$$

$$d\, loss = -\vec{y}^T dW\vec{x} + \frac{\vec{1}^T (e^{W\vec{x}} \odot dW\vec{x})}{\vec{1}^T e^{W\vec{x}}} \tag{**}$$

$$d\,loss = -\vec{y}^T dW\vec{x} + \frac{(e^{W\vec{x}})^T dW\vec{x}}{\vec{1}^T e^{W\vec{x}}}$$

$$d\,loss = tr(-\vec{y}^T dW\vec{x} + \frac{(e^{W\vec{x}})^T dW\vec{x}}{\vec{1}^T e^{W\vec{x}}})$$

$$d\,loss = tr(\vec{x}(softmax(W\vec{x}) - \vec{y})^T dW)$$

$$\frac{\partial\,loss}{\partial W} = (softmax(W\vec{x}) - \vec{y})\vec{x}^T$$

注意:

(*)式 $\log(\frac{\vec{b}}{c}) = \log\vec{b} - \vec{1}\log c$，且 $\vec{y}^T\vec{1} = 1$

(**)式 $\log(\vec{1}^T e^{W\vec{x}})$ 是标量，$e^{W\vec{x}}$ 是逐元素函数，因此$d\log(\vec{1}^T e^{W\vec{x}}) = \frac{1}{\vec{1}^T e^{W\vec{x}}} \cdot \vec{1}^T(e^{W\vec{x}} \odot dW\vec{x})$

- 向量对向量求导

对于 $\vec{f}_{(m\times 1)}, \vec{x}_{(n\times 1)}$ 有：

$$\frac{\partial\vec{f}}{\partial\vec{x}} = \left[\frac{\partial f_i}{\partial x_j}\right]_{(n\times m)}$$

$$d\vec{f} = \left[\frac{\partial f_i}{\partial\vec{x}}^T\right] d\vec{x} = \frac{\partial\vec{f}}{\partial\vec{x}}^T d\vec{x}$$

- 矩阵对矩阵求导

对于矩阵$F_{(m\times n)}, X_{(p\times q)}$ 有：

$$\frac{\partial F}{\partial X} = \left[\frac{\partial F_{ij}}{\partial x_{kl}}\right]_{(pq\times mn)}$$

矩阵向量化：

对于矩阵$X_{(p\times q)}$，其矩阵向量化$vec(X)_{(pq\times 1)} = \left[X_1^T, X_2^T, ..., X_q^T\right]^T$, X_i是X的列向量。

$$vec(A+B) = vec(A) + vec(B)$$

$$vec(\vec{a}\vec{b}^T) = \vec{b}\otimes\vec{a}$$

$$X = \sum_i X_i {e_i}^T$$

$$vec((AB)\otimes(CD)) = vec((A\otimes C)(B\otimes D))$$

$$vec(AXB) = vec(\sum_i AX_i {e_i}^T B) = \sum_i vec((AX_i)(B^T e_i)^T) = \sum_i (B^T e_i) \otimes (AX_i) = (B^T \otimes A) vec(X)$$

因此有：

$$\frac{\partial F}{\partial X} = \frac{\partial vec(F)}{\partial vec(X)}_{(pq \times mn)}$$

$$vec(dF) = \frac{\partial F}{\partial X}^T vec(dX)$$

求导时矩阵被向量化，弊端是这在一定程度破坏了矩阵的结构，会导致结果变得形式复杂；好处是多元微积分中关于 `Gradient` 、`Hessian` 矩阵的结论可以沿用过来，只需将矩阵向量化。