

Definition 1

对于单变量线性函数 $f(x) = ax$, $f'(x) = a$, 为了前后的连续性我们使用偏导符号

$$\frac{\partial f}{\partial x} = a$$

对于多变量线性函数 $f(x) = a^T x = \sum_i a_i x_i$, a, x 为(列)向量, 有:

$$\frac{\partial f}{\partial x_k} = \frac{\partial \sum_i a_i x_i}{\partial x_k} = a_k \forall k = 1, \dots, n$$

我们定义标量函数对向量的求导为下:

$$\frac{\partial f}{\partial x} = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right]^T = [a_1, a_2, \dots, a_n]^T = a$$

广义:

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f}{\partial x_{11}} & \dots & \frac{\partial f}{\partial x_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial x_{m1}} & \dots & \frac{\partial f}{\partial x_{mn}} \end{bmatrix}$$

Theorem 1

对于多元标量函数 $f(x) = a^T x$, $\frac{\partial f}{\partial x} = a$.

Proposition: 由于标量的值等于它的迹。对于 $g(x) = Tr[a^T x]$, $\frac{\partial g}{\partial x} = a$.

Properties 1

Definition of Trace: $Tr[A] = \sum_i A_{ii}$, 根据定义, 矩阵的迹有以下的性质:

1. $Tr[A + B] = Tr[A] + Tr[B]$
2. $Tr[cA] = cTr[A]$
3. $Tr[A] = Tr[A^T]$
4. $Tr[A_1 A_2 \dots A_n] = Tr[A_n A_1 \dots A_{n-1}]$
5. $Tr[A^T B] = \sum_i \sum_j A_{ij} B_{ij}$

Theorem 2

对于多元标量函数 $f(x) = Tr[A^T x]$, 有 $\frac{\partial f}{\partial x} = A \quad \forall A, x \in \mathbb{R}^{m \times n}$.

利用Properties 1(5)易证。

Tips

对于任意的标量函数 $f: \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$, 有 $\frac{\partial f}{\partial x} = \frac{\partial Tr[x]}{\partial x}$ 。然后可以利用Properties 1中的性质, 进行变换, 化成Theorem 2的形式。

Definition 2

定义矩阵微分 (differential) 如下:

$$dA = \begin{bmatrix} dA_{11} & \cdots & dA_{1n} \\ \vdots & \ddots & \vdots \\ dA_{m1} & \cdots & dA_{mn} \end{bmatrix}$$

Theorem 3

根据迹的定义和Definition 2可得:

$$dTr[A] = Tr[dA]$$

Theorem 4

对于标量函数 $f: \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$, $df = Tr[(\frac{\partial f}{\partial x})^T dx]$. (建立了矩阵微分和矩阵求导之间的关系)

证明: $LHS = df = \sum_{ij} \frac{\partial f}{\partial x_{ij}} dx_{ij}$

$$\begin{aligned} \text{依次利用 Properties 1、Definition 2、Definition 1: } RHS &= \sum_{ij} (\frac{\partial f}{\partial x})_{ij} (dx)_{ij} = \\ \sum_{ij} (\frac{\partial f}{\partial x})_{ij} dx_{ij} &= \sum_{ij} \frac{\partial f}{\partial x_{ij}} dx_{ij} = LHS \end{aligned}$$

Properties 2

利用Definition 2, 可以得到矩阵微分的性质:

1. $d(cA) = cdA$
2. $d(A+B) = dA + dB$
3. $d(AB) = dAB + AdB$

Conclusion 1

自此, 对于标量函数 $f: \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$, 我们能够通过以下流程轻易对其求导:

1. i.e., $df = dTr[x] = Tr[dx]$
2. 利用迹的性质Properties 1对 df 进行化简, 化简成 $Tr[A^T dx]$ 形式的线性相加.
3. 利用Theorem 4, 得到 $\frac{\partial f}{\partial x}$.

Examples

1. $f(x) = x^T Ax, A \in \mathbb{R}^{n \times n}, x \in \mathbb{R}^n$

$$df = dTr[x^T Ax] = Tr[d(x^T Ax)] = Tr[d(x^T)Ax + x^T Adx] = Tr[d(x^T)Ax] + Tr[x^T Adx] = Tr[x^T A^T dx] + Tr[x^T A dx]$$

Hence,

$$\begin{aligned} \frac{\partial f}{\partial x} &= (x^T A^T + x^T A)^T = (A + A^T)x \\ 2. XX^{-1} &= I \rightarrow d(XX^{-1}) = dI \rightarrow dXX^{-1} + XdX^{-1} \rightarrow dX^{-1} = \\ &= -X^{-1}dXX^{-1} \end{aligned}$$

Definition 3

上面总结了scalar函数对x的求导。下面定义vector函数对向量x的求导：

对于vector函数 $f = [f_1, f_2, \dots, f_n]^T$, $f_i = f_i(x)$, $x = [x_1, x_2, \dots, x_m]^T$, 我们定义:

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} & \dots & \frac{\partial f_n}{\partial x_1} \\ \frac{\partial f_1}{\partial x_2} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_n}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial x_m} & \frac{\partial f_2}{\partial x_m} & \dots & \frac{\partial f_n}{\partial x_m} \end{bmatrix}$$

Hessian Matrix and Jacobian Matrix

假设 $f: \mathbb{R}^m \rightarrow \mathbb{R}^n$, f 的Jacobian Matrix为

$$J(f) = [\frac{\partial f}{\partial x_1} \dots \frac{\partial f}{\partial x_m}] = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \dots & \frac{\partial f_n}{\partial x_m} \end{bmatrix}$$

假设 $f: \mathbb{R}^m \rightarrow \mathbb{R}$, f 的Hessian Matrix为

$$H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_m} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \dots & \frac{\partial^2 f}{\partial x_2 \partial x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_m \partial x_1} & \frac{\partial^2 f}{\partial x_m \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_m^2} \end{bmatrix} = J(\nabla f)$$

根据Definition 3, 我们可以把Hessian Matrix和Jacobian Matrix重写为:

$$J(f) = (\frac{\partial f}{\partial x})^T H(f) = \frac{\partial}{\partial x} (\frac{\partial f}{\partial x})$$

Theorem 5

对于 $f: \mathbb{R}^m \rightarrow \mathbb{R}^n$ $df = (\frac{\partial f}{\partial x})^T dx$.

证明 $df_j = ((\frac{\partial f}{\partial x})^T dx)_j \quad \forall j$.

Conclusion 2

对于vector函数的求导, 整体流程同Conclusion 1, 除了不能随便使用trace。

Frequently Used Formula

$$\frac{\partial \text{Tr}[A]}{\partial A} = I \frac{\partial x^T A x}{\partial x} = (A + A^T) x \frac{\partial x^T x}{\partial x} = 2x \frac{\partial A x}{\partial x} = A^T d(X^{-1}) = -X^{-1} dX X^{-1} \frac{\partial \det(A)}{\partial A} = \det(A) A^{-T} \frac{\partial \log \det(A)}{\partial A}$$