常用矩阵微分公式

1. 函数相对于实值向量的梯度

函数以实值向量为变元。

1.1 实值函数相对向量的梯度矩阵

实值函数 $f(x)$ 相对于 $n\times 1$ 行向量 x 的梯度为 $n\times 1$ 的行向量，定义为

$$\frac{\partial f(x)}{\partial x}=\left[\frac{\partial f(x)}{\partial x_1},\frac{\partial f(x)}{\partial x_2},\cdots,\frac{\partial f(x)}{\partial x_n}\right]^T=\nabla_x f(x)$$

向量梯度算子： $\nabla_x=\left[\frac{\partial}{\partial x_1},\frac{\partial}{\partial x_2},\cdots,\frac{\partial}{\partial x_n}\right]$

m **维行实值向量函数** $f(x)=\left[f_1(x),f_2(x),\cdots,f_m(x)\right]$

$$\frac{\partial f(x)}{\partial x}=\begin{bmatrix}\frac{\partial f_1(x)}{\partial x_1},\frac{\partial f_2(x)}{\partial x_1},\cdots,\frac{\partial f_m(x)}{\partial x_1}\\ \frac{\partial f_1(x)}{\partial x_2},\frac{\partial f_2(x)}{\partial x_2},\cdots,\frac{\partial f_m(x)}{\partial x_2}\\ \cdots \\ \frac{\partial f_1(x)}{\partial x_n},\frac{\partial f_2(x)}{\partial x_n},\cdots,\frac{\partial f_m(x)}{\partial x_n}\end{bmatrix}=\nabla_x f(x)$$

1.2 运算法则

(1)线性法则：若 $f(x)$ 和 $g(x)$ 分别是向量 x 的实值函数，c_1 和 c_2 为实常数，则

$$\frac{\partial\left[c_1 f(x)+c_2 g(x)\right]}{\partial x}=c_1\frac{\partial f(x)}{\partial x}+c_2\frac{\partial g(x)}{\partial x}$$

(2)乘积法则：若 $f(x)$ 和 $g(x)$ 分别是向量 x 的实值函数，则

$$\frac{\partial\left[f(x)g(x)\right]}{\partial x}=g(x)\frac{\partial f(x)}{\partial x}+f(x)\frac{\partial g(x)}{\partial x}$$

(3)商法则： $g(x)\neq 0$

$$\frac{\partial\left[f(x)/g(x)\right]}{\partial x}=\frac{1}{g^2(x)}\left[g(x)\frac{\partial f(x)}{\partial x}-f(x)\frac{\partial g(x)}{\partial x}\right]$$

(4)链式法则：若 $y(x)$ 是 x 的向量值函数，则

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

1.3 基本公式

x,y 为向量，$x=[x_1,x_2,\cdots,x_n]$, $y=[y_1,y_2,\cdots,y_n]$。

A 和 y 为与 x 无关，A 为矩阵，I 为单位矩阵。

(1) $\frac{\partial c}{\partial x}=0$，$c$ 为常数。

(2) $\frac{\partial x^T}{\partial x}=I$

(3) $\frac{\partial x^T x}{\partial x}=2x$（自已证的，不一定对）

证明：$\frac{\partial x^T x}{\partial x}=\frac{\partial x^T y}{\partial x}+\frac{\partial y^T x}{\partial x}=y+y=2y=2x$

注：这里 y 是一个中间代换量。

注：$\frac{\partial x^T x}{\partial x}=\frac{\partial x^T Ix}{\partial x}$

(4) $\frac{\partial Ax}{\partial x}=A^T$

证明：$\frac{\partial Ax}{\partial x}=\frac{\partial\left\langle A^T,x\right\rangle}{\partial x}=\frac{\partial\left\langle x,A^T\right\rangle}{\partial x}=\frac{\partial x^T A^T}{\partial x}=A^T$

注：Ax 可以被认为是一个向量函数。

(5) $\frac{\partial x^T Ay}{\partial x}=\frac{\partial x^T}{\partial x}Ay=Ay$

(3) $\frac{\partial y^T Ax}{\partial x}=\frac{\partial x^T A^T y}{\partial x}=A^T y$

证明：

$$y^T Ax=\left(A^T y\right)^T x=\left\langle A^T y,x\right\rangle=\left\langle x,A^T y\right\rangle=y^T A^T x$$

(4) $\frac{\partial\left(x^T Ax\right)}{\partial x}=Ax+A^T x$

当 A 为对称阵时，有 $\frac{\partial\left(x^T Ax\right)}{\partial x}=2Ax$

证明：$x^T Ax$ 相当于复合函数的微分。x^T 和 x 是与同一变量有关但不同的函数，对一个变量求微分时另一个变量保持不变，可以将保持不变的 x 替换成与无关的向量 y，由公式(2)和(3)即可得证。

注：x^T 不是相对于 x 的函数，而是相对于 x^T 的函数。

(5)若 $n\times 1$ 向量 a 是与 x 无关的常数向量，则

$$\frac{\partial a^T x}{\partial x}=a,\quad \frac{\partial x^T a}{\partial x}=a$$

2. 实值函数相对于实值矩阵的梯度

函数以实值矩阵为变元。

2.1 实值函数相对实值矩阵的梯度矩阵

实值函数 $f(A)$ 相对于 $m\times n$ 实矩阵 A 的梯度为一 $m\times n$ 矩阵，简称梯度矩阵，定义为

$$\frac{\partial f(A)}{\partial A}=\begin{bmatrix} \frac{\partial f(A)}{\partial A_{11}},\frac{\partial f(A)}{\partial A_{12}},\cdots,\frac{\partial f(A)}{\partial A_{1n}} \\ \frac{\partial f(A)}{\partial A_{21}},\frac{\partial f(A)}{\partial A_{22}},\cdots,\frac{\partial f(A)}{\partial A_{2n}} \\ \cdots \\ \frac{\partial f(A)}{\partial A_{m1}},\frac{\partial f(A)}{\partial A_{m2}},\cdots,\frac{\partial f(A)}{\partial A_{mn}} \end{bmatrix}$$

2.2 运算法则

(1)线性法则：若 $f(A)$ 和 $g(A)$ 分别是矩阵 A 的实值函数，c_1 和 c_2 为实常数，则

$$\frac{\partial\left[c_1 f(A)+c_2 g(A)\right]}{\partial A}=c_1\frac{\partial f(A)}{\partial A}+c_2\frac{\partial g(A)}{\partial A}$$

(2)乘积法则：若 $f(A)$ 和 $g(A)$ 分别是矩阵 A 的实值函数，则

$$\frac{\partial\left[f(A)g(A)\right]}{\partial A}=g(A)\frac{\partial f(A)}{\partial A}+f(A)\frac{\partial g(A)}{\partial A}$$

(3)商法则：$g(A)\neq 0$

$$\frac{\partial\left[f(A)/g(A)\right]}{\partial A}=\frac{1}{g^2(A)}\left[g(A)\frac{\partial f(A)}{\partial A}-f(A)\frac{\partial g(A)}{\partial A}\right]$$

(4)链式法则：若 $y(A)$ 是 A 的矩阵值函数，则

$$\frac{\partial f(y(A))}{\partial A}=\frac{\partial y^T(A)}{\partial A}\frac{\partial f(A)}{\partial A}$$

2.3 基本公式

(1) c 为常数，则 $\frac{\partial c}{\partial A}=\mathbf{0}_{m\times n}$。

(2)若 $A\in R^{m\times n},x\in R^{m\times 1},y\in R^{n\times 1}$，则

$$\frac{\partial x^T Ay}{\partial A}=xy^T$$

(3) 若 $A\in R^{m\times n}$ 非奇异，$x\in R^{m\times 1},y\in R^{n\times 1}$，则

$$\frac{\partial x^T A^{-1}y}{\partial A}=-A^{-T}xy^T A^{-T},\quad A^{-T}=(A^{-1})^T$$

(4) 若 $A\in R^{m\times n}$，$x,y\in R^{n\times 1}$，则

$$\frac{\partial x^T A^T Ay}{\partial A}=A(xy^T+yx^T)$$

(5) 若 $A\in R^{m\times n}$，$x,y\in R^{m\times 1}$，则

$$\frac{\partial x^T AA^T y}{\partial A}=(xy^T+yx^T)A$$

(6)指数函数的梯度

$$\frac{\partial\,\mathbf{exp}(x^T Ay)}{\partial A}=xy^T\,\mathbf{exp}(x^T Ay)$$

3. 迹函数的梯度矩阵

2.1 迹和矩阵、向量的关系

1. 二次型

$$f(x)=x^T Ax=tr(x^T Ax)=tr(Axx^T)$$

说明：二次型目标函数 $x^T Ax$ 等于其核矩阵 A 和向量外积 xx^T 的乘积的迹($tr(Axx^T)$)。

2. 矩阵和迹

$$\langle A,A\rangle=A^T A$$
$$\left|\langle A,A\rangle\right|=\|A\|_2^2=tr(A^T A)=tr(AA^T)$$

3. 向量和迹

$$x^T y = tr(xy^T) = tr(yx^T)$$

说明：**这个关系很重要**，可以简单推导过程。

2.2 迹的梯度矩阵

1. $W \in R^{m\times m}$

$$\frac{\partial tr(W)}{\partial W} = I_m$$

2. $W \in R^{m\times m}$ 可逆

$$\frac{\partial tr(W^{-1})}{\partial W} = -(W^{-2})^T$$

3. $x, y \in R^m$ 的外积

$$\frac{\partial tr(xy^T)}{\partial x} = \frac{\partial tr(yx^T)}{\partial x} = y$$

4. $W \in R^{m\times n}, A \in R^{n\times m}$

$$\frac{\partial tr(WA)}{\partial W} = \frac{\partial tr(AW)}{\partial W} = A^T$$

特别，W 为对称矩阵

$$\frac{\partial tr(WA)}{\partial W} = \frac{\partial tr(AW)}{\partial W} = A + A^T - diag(A)$$

5. $W \in R^{m\times n}, A \in R^{m\times n}$

$$\frac{\partial tr(W^T A)}{\partial W} = \frac{\partial tr(AW^T)}{\partial W} = A$$

6. $W \in R^{m\times n}$

$$\frac{\partial tr(WW^T)}{\partial W} = \frac{\partial tr(W^T W)}{\partial W} = 2W$$

7. $W \in R^{m\times m}$

$$\frac{\partial tr(W^2)}{\partial W} = \frac{\partial tr(WW)}{\partial W} = 2W^T$$

略，关于迹还有好多公式，祥见张贤达的矩阵分析与应用。