

粒子物理与核物理实验中的 数据分析

陈少敏
清华大学

补充讲义：数学基础：
——向量微分

本讲要点

- 向量微分定义
- 标量函数对向量的梯度
- 向量函数对向量的梯度
- 向量微分常用公式
- 举例

简介

- 本课程很多问题涉及到向量微分。处理这些问题，可以按分量展开解决，但是直接进行向量微分会方便很多。
- 函数对向量的微分形式与普通微分类似，只是微分后的结果是向量或者矩阵，不论被微分的函数本身是标量函数还是矢量函数(或者矩阵)
- 本讲义以实函数为例

行、列向量定义

行向量:

$$\mathbf{x}^T = [x_1, x_2, \dots, x_n]$$

列向量:

$$\mathbf{x} = [x_1, x_2, \dots, x_n]^T$$

向量微分的定义(1)

■ 对列向量和行向量的微分（梯度）定义

对列向量的梯度：

相对于 $n \times 1$ 向量(列向量) $\mathbf{x}$ 的梯度算子记作 $\nabla_{\mathbf{x}}$ ，定义为

$$\nabla_{\mathbf{x}} \equiv \left[\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \dots, \frac{\partial}{\partial x_n} \right]^T = \frac{\partial}{\partial \mathbf{x}}$$

对行向量的梯度：

相对于 $1 \times n$ 向量(行向量) $\mathbf{x}^T$ 的梯度算子记作 $\nabla_{\mathbf{x}^T}$ ，定义为

$$\nabla_{\mathbf{x}^T} \equiv \left[\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \dots, \frac{\partial}{\partial x_n} \right] = \frac{\partial}{\partial \mathbf{x}^T}$$

标量函数对向量的梯度

所以，假设标量函数 $f(\mathbf{x})$ 以 n 维列向量 $\mathbf{x}$ 为变元，其相对于 $\mathbf{x}$ 的梯度是 n 维列向量，

$$\nabla_{\mathbf{x}} f(\mathbf{x}) \equiv \left[\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \dots, \frac{\partial f(\mathbf{x})}{\partial x_n} \right]^T = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}$$

而 $f(\mathbf{x})$ 相对于行向量 $\mathbf{x}^T$ 的梯度是 n 维行向量，

$$\nabla_{\mathbf{x}^T} f(\mathbf{x}) \equiv \left[\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \dots, \frac{\partial f(\mathbf{x})}{\partial x_n} \right] = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}^T}$$

- 1) 梯度的分量 $\frac{\partial f(\mathbf{x})}{\partial x_i}$ 给出了 $f(\mathbf{x})$ 在第 i 个方向上的变化率。
- 2) 标量函数对列向量的微分结果为列向量
标量函数对行向量的微分结果为行向量

举例

标量函数 $f(\mathbf{x}) = \mathbf{x}^T \mathbf{x} (= \sum_{i=1}^n x_i^2)$, 求 $\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}$ 和 $\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}^T}$

解：根据对列向量和行向量微分的定义，

$$\begin{aligned}\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} &= \left[\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \dots, \frac{\partial f(\mathbf{x})}{\partial x_n} \right]^T \\ &= [2x_1, 2x_2, \dots, 2x_n]^T = 2\mathbf{x} \text{ (列向量)}\end{aligned}$$

$$\begin{aligned}\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}^T} &= \left[\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \dots, \frac{\partial f(\mathbf{x})}{\partial x_n} \right] \\ &= [2x_1, 2x_2, \dots, 2x_n] = 2\mathbf{x}^T \text{ (行向量)}\end{aligned}$$

推广到矩阵函数(1)

m 维行向量函数 $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x})]$

相对于 n 维列向量 $\mathbf{x}$ 的梯度是一个 $n \times m$ 矩阵,

$$\frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} \equiv \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1}, \frac{\partial f_2(\mathbf{x})}{\partial x_1}, \dots, \frac{\partial f_m(\mathbf{x})}{\partial x_1} \\ \frac{\partial f_1(\mathbf{x})}{\partial x_2}, \frac{\partial f_2(\mathbf{x})}{\partial x_2}, \dots, \frac{\partial f_m(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f_1(\mathbf{x})}{\partial x_n}, \frac{\partial f_2(\mathbf{x})}{\partial x_n}, \dots, \frac{\partial f_m(\mathbf{x})}{\partial x_n} \end{bmatrix} = \nabla_{\mathbf{x}} \mathbf{f}(\mathbf{x})$$

推广到矩阵函数(2)

m 维列向量函数 $\mathbf{f}(\mathbf{x}) = \mathbf{y} = [y_1(\mathbf{x}), y_2(\mathbf{x}), \dots, y_m(\mathbf{x})]^T$

相对于 n 维行向量 $\mathbf{x}^T$ 的梯度是一个 $m \times n$ 矩阵,

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}^T} \equiv \begin{bmatrix} \frac{\partial y_1(\mathbf{x})}{\partial x_1}, \frac{\partial y_1(\mathbf{x})}{\partial x_2}, \dots, \frac{\partial y_1(\mathbf{x})}{\partial x_n} \\ \frac{\partial y_2(\mathbf{x})}{\partial x_1}, \frac{\partial y_2(\mathbf{x})}{\partial x_2}, \dots, \frac{\partial y_2(\mathbf{x})}{\partial x_n} \\ \vdots \\ \frac{\partial y_m(\mathbf{x})}{\partial x_1}, \frac{\partial y_m(\mathbf{x})}{\partial x_2}, \dots, \frac{\partial y_m(\mathbf{x})}{\partial x_n} \end{bmatrix} = \nabla_{\mathbf{x}^T} \mathbf{y}$$

注：这正是 **Jacobi** 矩阵

再举例(1)

(1) 行向量函数 $f(\mathbf{x}) = \mathbf{x}^T$, 求 $\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}$

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial \mathbf{x}^T}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial x_1}{\partial x_1} & \frac{\partial x_2}{\partial x_1} & \cdots & \frac{\partial x_n}{\partial x_1} \\ \frac{\partial x_1}{\partial x_2} & \frac{\partial x_2}{\partial x_2} & \cdots & \frac{\partial x_n}{\partial x_2} \\ \vdots & \vdots & & \vdots \\ \frac{\partial x_1}{\partial x_n} & \frac{\partial x_2}{\partial x_n} & \cdots & \frac{\partial x_n}{\partial x_n} \end{bmatrix} = \begin{bmatrix} 1, 0, \cdots, 0 \\ 0, 1, \cdots, 0 \\ \vdots & \vdots & & \vdots \\ 0, 0, \cdots, 1 \end{bmatrix} = I$$

再举例(2)

(2) 列向量函数 $\mathbf{g}(\mathbf{x}) = \mathbf{x}$, 求 $\frac{\partial \mathbf{g}(\mathbf{x})}{\partial \mathbf{x}^T}$

$$\frac{\partial \mathbf{g}(\mathbf{x})}{\partial \mathbf{x}^T} = \frac{\partial \mathbf{x}}{\partial \mathbf{x}^T} = \begin{bmatrix} \frac{\partial x_1}{\partial x_1}, \frac{\partial x_1}{\partial x_2}, \dots, \frac{\partial x_1}{\partial x_n} \\ \frac{\partial x_2}{\partial x_1}, \frac{\partial x_2}{\partial x_2}, \dots, \frac{\partial x_2}{\partial x_n} \\ \vdots \\ \frac{\partial x_n}{\partial x_1}, \frac{\partial x_n}{\partial x_2}, \dots, \frac{\partial x_n}{\partial x_n} \end{bmatrix} = \begin{bmatrix} 1, 0, \dots, 0 \\ 0, 1, \dots, 0 \\ \vdots \\ 0, 0, \dots, 1 \end{bmatrix} = I$$

一些常用公式(1)

$$(1) \frac{\partial \mathbf{x}^T}{\partial \mathbf{x}} = I, \quad \frac{\partial \mathbf{x}}{\partial \mathbf{x}^T} = I$$

若 $\mathbf{A}$ 为 n 维方矩阵, $\mathbf{y}$ 为 n 维列向量, 且都与 $\mathbf{x}$ 无关:

$$(2) \frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}^T} = \mathbf{A}, \quad \frac{\partial \mathbf{x}^T \mathbf{A}}{\partial \mathbf{x}} = \mathbf{A},$$

$$(3) \frac{\partial \mathbf{x}^T \mathbf{A}\mathbf{y}}{\partial \mathbf{x}} = \mathbf{A}\mathbf{y}, \quad \frac{\partial \mathbf{y}^T \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \frac{\partial \mathbf{x}^T \mathbf{A}^T \mathbf{y}}{\partial \mathbf{x}} = \mathbf{A}^T \mathbf{y}(!!!!)$$

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

一些常用公式(2)

形式上，普通微分的一些法则都适用于向量微分，如

$$(1) f(\mathbf{x}) = c \Rightarrow \frac{\partial c}{\partial \mathbf{x}} = 0$$

$$(2) \frac{\partial [c_1 f(\mathbf{x}) + c_2 g(\mathbf{x})]}{\partial \mathbf{x}} = c_1 \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} + c_2 \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}}$$

$$(3) \frac{\partial f(\mathbf{x})g(\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} g(\mathbf{x}) + f(\mathbf{x}) \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}}$$

$$(4) \frac{\partial f(\mathbf{x})/g(\mathbf{x})}{\partial \mathbf{x}} = \frac{1}{g^2(\mathbf{x})} \left[g(\mathbf{x}) \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} - f(\mathbf{x}) \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}} \right]$$

$$(5) \frac{\partial f(\mathbf{y}(\mathbf{x}))}{\partial \mathbf{x}} = \frac{\partial \mathbf{y}^T(\mathbf{x})}{\partial \mathbf{x}} \frac{\partial f(\mathbf{y})}{\partial \mathbf{y}}$$

例子(Exercise 1.8)

假设 n 维随机变量 $\mathbf{x} = (x_1, \dots, x_n)^T$ 的联合p.d.f.为 $f(\mathbf{x})$, 变量 $\mathbf{y} = (y_1, \dots, y_n)^T$ 是 $\mathbf{x}$ 的线性变换, $\mathbf{y} = A\mathbf{x}$, 即

$$y_i = \sum_{j=1}^n A_{ij} x_j \circ$$

假设逆变换 $\mathbf{x} = A^{-1}\mathbf{y}$ 存在。

(a)证明 $\mathbf{y}$ 的联合p.d.f.为 $g(\mathbf{y}) = f(A^{-1}\mathbf{y})|\det(A^{-1})|$ 。

(b)当 A 是正交矩阵, 即 $A^{-1} = A^T$ 时, 求 $g(\mathbf{y})$ 。

续上

解:(a)根据公式(1.37), $g(a_1, \dots, a_n) = f(x_1, \dots, x_n)|J|$,

很容易得到本题所求为

$$g(\mathbf{y}) = f(\mathbf{x})|J| = f(A^{-1}\mathbf{y})|J|,$$

其中 J 为Jacobian矩阵的行列式, 即

$$J = \det\left(\frac{\partial \mathbf{x}}{\partial \mathbf{y}^T}\right) = \det\left(\frac{\partial A^{-1}\mathbf{y}}{\partial \mathbf{y}^T}\right) = \det(A^{-1})$$

所以, $g(\mathbf{y}) = f(A^{-1}\mathbf{y})|J| = f(A^{-1}\mathbf{y})|\det(A^{-1})|$ 得证。

(b)如果 A 是正交矩阵, 则 $A^{-1} = A^T$ 。

由 $\det(AB) = \det(A)\det(B)$, 以及 $\det(A^T) = \det(A)$ 得到

$$1 = \det(I) = \det(A^{-1}A) = \det(A^T A) = \det(A^T)\det(A) = (\det(A^T))^2.$$

所以, $|\det(A^{-1})| = |\det(A^T)| = 1$, 即

$$g(\mathbf{y}) = f(A^{-1}\mathbf{y})$$

续上

如果用分量方式，由于Jacobian行列式 $J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \dots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \dots & \frac{\partial x_2}{\partial y_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \dots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$

则要计算出 $\frac{\partial x_i}{\partial y_j}$ ，证明 $\frac{\partial x_i}{\partial y_j} = (A^{-1})_{ij}$

由 $x_i = \sum_{k=1}^n (A^{-1})_{ik} y_k$ ，显然可得

$$\frac{\partial x_i}{\partial y_j} = \sum_{k=1}^n (A^{-1})_{ik} \frac{\partial y_k}{\partial y_j} = \sum_{k=1}^n (A^{-1})_{ik} \frac{\partial y_k}{\partial y_j} = \sum_{k=1}^n (A^{-1})_{ik} \delta_{kj} = (A^{-1})_{ij}.$$

得证。但不如直接向量微分简洁。

参考资料

1. 《矩阵分析与应用》(2004), 张贤达, 清华大学出版社
(第5.1.2节)