西北工業大學

矩阵微积分与自动微分

Matrix Calculus and Automatic Differentiation

软件学院　**钱锋**　编

2024 年 4 月 18 日

矩阵微积分与自动微分

钱锋[1,2]

2024 年 4 月 18 日

[1]Email: strik0r.qf@gmail.com
[2]西北工业大学软件学院, School of Software, Northwestern Polytechnical University, 西安 710072

目录

第 1 章　概述

矩阵微积分主要讨论的是涉及向量、矩阵的函数的导数.

1.1　定义与例子

表 1.1: 一阶导数

	标量 f	**向量** $\boldsymbol{f}$	**矩阵** $\boldsymbol{F}$
标量 x	标量 $\partial f/\partial x$	向量 $\partial \boldsymbol{f}/\partial x$	矩阵 $\partial \boldsymbol{F}/\partial x$
向量 $\boldsymbol{x}$	向量 $\partial f/\partial \boldsymbol{x}$	Jacobi 矩阵	更高维的数组
	梯度 (gradiant)	$\partial \boldsymbol{f}/\partial \boldsymbol{x}$	
矩阵 $\boldsymbol{X}$	矩阵 $\partial \boldsymbol{f}/\partial \boldsymbol{X}$	更高维的数组	更高维的数组

表 1.1 总结了不同形式输入和输出的函数的一阶导数的的形式, 其中, 每一行代表输入参数的一种可能性: 标量、向量或者矩阵, 而每一列则代表着返回值的一种可能性.

本章中我们只研究表 1.1 中提及的六种情况, 其中, 单变量标量值函数 $f(x)$ 的导数已经在微积分课程中讨论过了.

本章中的矩阵均采用分子布局.

1.1.1　一元向量值函数的导数

函数 $\boldsymbol{f}: x \mapsto [f_0(x), f_1(x), \cdots, f_{m-1}(x)]^{\mathrm{T}}$ 接受一个标量输入 $x \in \mathbb{R}$, 并返回一个向量 $\boldsymbol{f}(x) \in \mathbb{R}^m$, 它的 m 个分量分别为 $f_0(x), f_1(x), \cdots, f_{m-1}(x)$[1]. 它的导数为

$$\frac{\partial \boldsymbol{f}}{\partial x} = \begin{bmatrix} \partial f_0/\partial x \\ \partial f_1/\partial x \\ \vdots \\ \partial f_{m-1}/\partial x \end{bmatrix},$$

这就是说, *一元向量值函数的导数是一个向量, 它的每一个分量等于它对应的分量函数对自变量的导数*. 我们在处理一元向量值函数的导数的时候, 只需要把 $\partial/\partial x$ 放到向量符号当中就可以了.

例 1.1.1. 设向量值函数 $\boldsymbol{f}(x) = \begin{bmatrix} 2x^2 - 3x + 2 \\ x^3 - 3 \end{bmatrix}$, 则 $\dfrac{\partial \boldsymbol{f}}{\partial x} = \begin{bmatrix} 4x - 3 \\ 3x^2 \end{bmatrix}$.

[1] 为了方便在后期将我们的公式转换为代码, 我们的数组索引是从 0 开始的.

1.1.2　多元标量值函数的导数

函数 $f:[x_0,x_1,\cdots,x_{m-1}]\mapsto f(\boldsymbol{x})$ 接受一个矢量输入 $\boldsymbol{x}\in\mathbb{R}^m$, 并返回一个标量值 $f(\boldsymbol{x})\in\mathbb{R}$. 它的导数为

$$\frac{\partial f}{\partial \boldsymbol{x}}=\left[\frac{\partial f}{\partial x_0},\frac{\partial f}{\partial x_1},\cdots,\frac{\partial f}{\partial x_{m-1}}\right],$$

为了在我们的公式中采用分子布局, 我们将 $\partial f/\partial\boldsymbol{x}$ 转置以使其成为一个列向量, 并将其定义为多元标量值函数的**梯度** (gradient), 用 Hamilton 算子 ∇ 表示, 即

$$\nabla f(\boldsymbol{x})=\begin{bmatrix}\dfrac{\partial f}{\partial x_0}\\ \dfrac{\partial f}{\partial x_1}\\ \vdots\\ \dfrac{\partial f}{\partial x_{m-1}}\end{bmatrix}$$

性质 1. (齐次性). $\dfrac{\partial}{\partial \boldsymbol{x}}(af)=a\dfrac{\partial f}{\partial \boldsymbol{x}}$.

证明. $\dfrac{\partial}{\partial \boldsymbol{x}}(af)=\left[\dfrac{\partial (af)}{\partial x_0},\dfrac{\partial (af)}{\partial x_1},\cdots,\dfrac{\partial (af)}{\partial x_{m-1}}\right]=a\dfrac{\partial f}{\partial \boldsymbol{x}}$. □

性质 2. (叠加性). $\dfrac{\partial}{\partial \boldsymbol{x}}(f+g)=\dfrac{\partial f}{\partial \boldsymbol{x}}+\dfrac{\partial g}{\partial \boldsymbol{x}}$.

证明. 由定义可知,

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{x}}(f+g)&=\left[\frac{\partial (f+g)}{\partial x_0},\frac{\partial (f+g)}{\partial x_1},\cdots,\frac{\partial (f+g)}{\partial x_{m-1}}\right]\\&=\left[\frac{\partial f}{\partial x_0}+\frac{\partial g}{\partial x_0},\frac{\partial f}{\partial x_1}+\frac{\partial g}{\partial x_1},\cdots,\frac{\partial f}{\partial x_{m-1}}+\frac{\partial g}{\partial x_{m-1}}\right],\end{aligned}$$

根据一元函数的导数的可加性可知结论成立. □

性质 3. (乘积法则). $\dfrac{\partial}{\partial \boldsymbol{x}}(fg)=f\dfrac{\partial g}{\partial \boldsymbol{x}}+g\dfrac{\partial x}{\partial \boldsymbol{x}}$.

证明. 这可以由一元函数的乘积法则直接导出. □

性质 4. (链式法则). 如果 $f:\mathbb{R}\to\mathbb{R}$, $g:\mathbb{R}^m\to\mathbb{R}$. 则 $\dfrac{\partial}{\partial \boldsymbol{x}}f(g(\boldsymbol{x}))=\dfrac{\partial f}{\partial g}\dfrac{\partial g}{\partial \boldsymbol{x}}$.

1.1.3　多元向量值函数的导数

函数 $\boldsymbol{f}:[x_0,x_1,\cdots,x_{m-1}]^{\mathrm{T}}\mapsto[\boldsymbol{f}_0(x_0),\boldsymbol{f}_1(x_1),\cdots,\boldsymbol{f}_{n-1}(x_{m-1})]^{\mathrm{T}}$ 接受一个向量 $\boldsymbol{x}\in\mathbb{R}^m$ 作为输入, 并返回一个向量 $\boldsymbol{f}(\boldsymbol{x})\in\mathbb{R}^n$ 作为输出. 事实上, 我们可以把 $\boldsymbol{f}$ 分解成若干个多元标量值函数, 即

$$\boldsymbol{f}(\boldsymbol{x})=\begin{bmatrix}f_0(\boldsymbol{x})\\ f_1(\boldsymbol{x})\\ \vdots\\ f_{n-1}(\boldsymbol{x})\end{bmatrix},$$

为了求解它关于 $\boldsymbol{x}$ 的导数, 我们把 $\partial/\partial\boldsymbol{x}$ 放进向量符号中,

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} = \begin{bmatrix} \partial f_0/\partial\boldsymbol{x} \\ \partial f_1/\partial\boldsymbol{x} \\ \vdots \\ \partial f_{n-1}/\partial\boldsymbol{x} \end{bmatrix}$$

而这时 $\partial f_i/\partial\boldsymbol{x}$ 就是多元标量值函数的导数, 因此

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} = \begin{bmatrix} \dfrac{\partial f_0}{\partial x_0} & \dfrac{\partial f_0}{\partial x_1} & \cdots & \dfrac{\partial f_0}{\partial x_{m-1}} \\ \dfrac{\partial f_1}{\partial x_0} & \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_{m-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial f_{n-1}}{\partial x_0} & \dfrac{\partial f_{n-1}}{\partial x_1} & \cdots & \dfrac{\partial f_{n-1}}{\partial x_{m-1}} \end{bmatrix},$$

它的每一行都是 $(\nabla f_i(\boldsymbol{x}))^{\mathrm{T}}$.

1.1.4 一元矩阵值函数的导数

函数 $\boldsymbol{F}:\mathbb{R}\to\mathbb{R}^{n\times m}$ 接受一个标量输入, 返回一个矩阵

$$\boldsymbol{F} = \begin{bmatrix} f_{00}(x) & f_{01}(x) & \cdots & f_{1,m-1}(x) \\ f_{10}(x) & f_{11}(x) & \cdots & f_{1,m-1}(x) \\ \vdots & \vdots & \ddots & \vdots \\ f_{n-1,0}(x) & f_{n-1,1}(x) & \cdots & f_{n-1,m-1}(x) \end{bmatrix},$$

它的导数为

$$\frac{\partial \boldsymbol{F}}{\partial x} = \begin{bmatrix} \dfrac{\partial f_{00}}{\partial x} & \dfrac{\partial f_{01}}{\partial x} & \cdots & \dfrac{\partial f_{0,m-1}}{\partial x} \\ \dfrac{\partial f_{10}}{\partial x} & \dfrac{\partial f_{11}}{\partial x} & \cdots & \dfrac{\partial f_{1,m-1}}{\partial x} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial f_{n-1,0}}{\partial x} & \dfrac{\partial f_{n-1,1}}{\partial x} & \cdots & \dfrac{\partial f_{n-1,m-1}}{\partial x} \end{bmatrix}.$$

1.1.5 标量值矩阵函数的导数

函数 $f:\mathbb{R}^{n\times m}\to\mathbb{R}$ 接受一个矩阵 $\boldsymbol{X}$ 作为输入, 并返回一个标量值. 它的导数为

$$\frac{\partial f}{\partial \boldsymbol{X}}=\begin{bmatrix} \frac{\partial f}{\partial x_{00}} & \frac{\partial f}{\partial x_{10}} & \cdots & \frac{\partial f}{\partial x_{n-1,0}} \\ \frac{\partial f}{\partial x_{01}} & \frac{\partial f}{\partial x_{11}} & \cdots & \frac{\partial f}{\partial x_{n-1,1}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial x_{0,m-1}} & \frac{\partial f}{\partial x_{1,m-1}} & \cdots & \frac{\partial f}{\partial x_{n-1,m-1}} \end{bmatrix}.$$

参考文献

I ♥ NPU

公诚勇毅　永矢毋忘
中华灿烂　工大无疆

2023 年　西北工业大学　中国西安