Matrix Derivative矩阵求导

LittleSasuke

Welcome To My Blog
学习机器学习算法时总碰见矩阵求导,现学习一波,主要总结下
注意:这里只涉及实数的求导,研究通信的人可能接触的往往是负数求导
矩阵可以写成列向量(column vectors)或行向量(row vectors)的形式,这两种不同的形式把矩阵求导分成了两种不同的情况

求导类型

Types of Matrix Derivatives

Types	Scalar	Vector	Matrix
Scalar	$\frac{\partial y}{\partial x}$	$\frac{\partial \mathbf{y}}{\partial x}$	$\frac{\partial \mathbf{Y}}{\partial x}$
Vector	$\frac{\partial y}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$	
Matrix	$\frac{\partial y}{\partial \mathbf{X}}$		

1.jpg

表格列举了六种不同的矩阵求导类型,粗体代表向量或者矩阵(其实标量和向量也可以看作矩阵).
表格中还有三个空格没写出,实际上也是存在,但暂时先不讨论,因为这三种情况的求导结果大部分都是高于二阶的张量(tensor)形式,与常见的二维矩阵形式不同.

布局约定Layout conventions

机器学习中,以线性回归为例,每个输入都有多个属性,在表示属性时可以采用列向量或者行向量的形式,这两种形式会造成求导结果形式的不同.
注意是形式上的不同,因为本质上形式的不同不会影响求导结果,只不过将结果按照不同的方式组织起来,

方便进一步运算

布局决定(Layout conventions)就是为了将不同形式的求导分类.分为两种布局:分子布局(numerator layout)和分母布局(denominator layout)

通俗解释,现规定**向量或者矩阵**分为原始形式和转置形式两种,比如在线性回归中我们把列向量作为属性值的原始形式,其转置形式就是行向量

- 对于分子布局(numerator layout),求导结果中分子保持原始形式,分母为转置形式
- 对于分母布局(denominator layout),求导结果中分子为转置形式,分母保持原始形式

  下图展示各种类型求导与两种布局之间的关系

Result of differentiating various kinds of aggregates with other kinds of aggregates

	Scalar y		Vector y (size m)		Matrix Y (size $m \times n$)	
	Notation	Type	Notation	Type	Notation	Type
Scalar x	$\frac{\partial y}{\partial x}$	scalar	$\frac{\partial \mathbf{y}}{\partial x}$	(numerator layout) size-m column vector (denominator layout) size-m row vector	$\frac{\partial \mathbf{Y}}{\partial x}$	(numerator layout) $m \times n$ matrix
Vector x (size n)	$\frac{\partial y}{\partial \mathbf{x}}$	(numerator layout) size-n row vector (denominator layout) size-n column vector	$\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$	(numerator layout) $m \times n$ matrix (denominator layout) $n \times m$ matrix	$\frac{\partial \mathbf{Y}}{\partial \mathbf{x}}$	
Matrix X (size $p \times q$)	$\frac{\partial y}{\partial \mathbf{X}}$	(numerator layout) $q \times p$ matrix (denominator layout) $p \times q$ matrix	$\frac{\partial \mathbf{y}}{\partial \mathbf{X}}$		$\frac{\partial \mathbf{Y}}{\partial \mathbf{X}}$	

2.png

numerator layout

将上述表格中的分子布局单独拿出来,求导结果如下

$$\frac{\partial y}{\partial \mathbf{x}} = \left[\frac{\partial y}{\partial x_1} \frac{\partial y}{\partial x_2} \cdots \frac{\partial y}{\partial x_n} \right].$$

$$\frac{\partial \mathbf{y}}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x} \\ \frac{\partial y_2}{\partial x} \\ \vdots \\ \frac{\partial y_m}{\partial x} \end{bmatrix}.$$

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \end{bmatrix}.$$

3.png

下面的两种定义只在分子布局中有意义

$$\frac{\partial \mathbf{Y}}{\partial x} = \begin{bmatrix} \frac{\partial y_{11}}{\partial x} & \frac{\partial y_{12}}{\partial x} & \cdots & \frac{\partial y_{1n}}{\partial x} \\ \frac{\partial y_{21}}{\partial x} & \frac{\partial y_{22}}{\partial x} & \cdots & \frac{\partial y_{2n}}{\partial x} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_{m1}}{\partial x} & \frac{\partial y_{m2}}{\partial x} & \cdots & \frac{\partial y_{mn}}{\partial x} \end{bmatrix}.$$

$$d\mathbf{X} = \begin{bmatrix} dx_{11} & dx_{12} & \cdots & dx_{1n} \\ dx_{21} & dx_{22} & \cdots & dx_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ dx_{m1} & dx_{m2} & \cdots & dx_{mn} \end{bmatrix}.$$

4.png

denominator layout

将上述表格中的分母布局单独拿出来,求导结果如下

$$\frac{\partial y}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix}.$$

$$\frac{\partial \mathbf{y}}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x} & \frac{\partial y_2}{\partial x} & \cdots & \frac{\partial y_m}{\partial x} \end{bmatrix}.$$

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_n} & \frac{\partial y_2}{\partial x_n} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}.$$

$$\frac{\partial y}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{12}} & \cdots & \frac{\partial y}{\partial x_{1q}} \\ \frac{\partial y}{\partial x_{21}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{2q}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial x_{p1}} & \frac{\partial y}{\partial x_{p2}} & \cdots & \frac{\partial y}{\partial x_{pq}} \end{bmatrix}.$$

5.png

常见求导结果

现给出常见的求导结果,推导相关公式时可以查表
求导有链式法则(Chain Rule),但是矩阵乘积不满足交换律,所以链式法则对于matrix-by-scalar derivatives和scalar-by-matrix derivatives这两种情况不适用
下面贴出三种求导结果

Vector-by-vector

之所以先展示vector-by-vector的表格,是因为所有适用于vector-by-vector求导的操作也直接适用于vector-by-scalar or scalar-by-vector这两种情况

Condition	Expression	Numerator layout,	Denominator layout, i.e. by $\mathbf{y}^{T}$

6.png

Scalar-by-vector

Condition	Expression	Numerator layout, i.e. by $\mathbf{x}^{T}$; result is row vector	Denominator layout, i.e. by $\mathbf{x}$; result is column vector
a is not a function of $\mathbf{x}$	$\frac{\partial a}{\partial \mathbf{x}} =$	$\mathbf{0}^{\top}$ [4]	$\mathbf{0}$ [4]
a is not a function of $\mathbf{x}$, $u = u(\mathbf{x})$	$\frac{\partial au}{\partial \mathbf{x}} =$	$a\frac{\partial u}{\partial \mathbf{x}}$	
$u = u(\mathbf{x})$, $v = v(\mathbf{x})$	$\frac{\partial (u+v)}{\partial \mathbf{x}} =$	$\frac{\partial u}{\partial \mathbf{x}} + \frac{\partial v}{\partial \mathbf{x}}$	
$u = u(\mathbf{x})$, $v = v(\mathbf{x})$	$\frac{\partial uv}{\partial \mathbf{x}} =$	$u\frac{\partial v}{\partial \mathbf{x}} + v\frac{\partial u}{\partial \mathbf{x}}$	
$u = u(\mathbf{x})$	$\frac{\partial g(u)}{\partial \mathbf{x}} =$	$\frac{\partial g(u)}{\partial u}\frac{\partial u}{\partial \mathbf{x}}$	
$u = u(\mathbf{x})$	$\frac{\partial f(g(u))}{\partial \mathbf{x}} =$	$\frac{\partial f(g)}{\partial g}\frac{\partial g(u)}{\partial u}\frac{\partial u}{\partial \mathbf{x}}$	
$\mathbf{u} = \mathbf{u}(\mathbf{x})$, $\mathbf{v} = \mathbf{v}(\mathbf{x})$	$\frac{\partial (\mathbf{u}\cdot\mathbf{v})}{\partial \mathbf{x}} = \frac{\partial \mathbf{u}^{\top}\mathbf{v}}{\partial \mathbf{x}} =$	$\mathbf{u}^{\top}\frac{\partial \mathbf{v}}{\partial \mathbf{x}} + \mathbf{v}^{\top}\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$ • assumes numerator layout of $\frac{\partial \mathbf{u}}{\partial \mathbf{x}}, \frac{\partial \mathbf{v}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}}\mathbf{v} + \frac{\partial \mathbf{v}}{\partial \mathbf{x}}\mathbf{u}$ • assumes denominator layout of $\frac{\partial \mathbf{u}}{\partial \mathbf{x}}, \frac{\partial \mathbf{v}}{\partial \mathbf{x}}$
$\mathbf{u} = \mathbf{u}(\mathbf{x})$, $\mathbf{v} = \mathbf{v}(\mathbf{x})$, $\mathbf{A}$ is not a function of $\mathbf{x}$	$\frac{\partial (\mathbf{u}\cdot\mathbf{A}\mathbf{v})}{\partial \mathbf{x}} = \frac{\partial \mathbf{u}^{\top}\mathbf{A}\mathbf{v}}{\partial \mathbf{x}} =$	$\mathbf{u}^{\top}\mathbf{A}\frac{\partial \mathbf{v}}{\partial \mathbf{x}} + \mathbf{v}^{\top}\mathbf{A}^{\top}\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$ • assumes numerator layout of $\frac{\partial \mathbf{u}}{\partial \mathbf{x}}, \frac{\partial \mathbf{v}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}}\mathbf{A}\mathbf{v} + \frac{\partial \mathbf{v}}{\partial \mathbf{x}}\mathbf{A}^{\top}\mathbf{u}$ • assumes denominator layout of $\frac{\partial \mathbf{u}}{\partial \mathbf{x}}, \frac{\partial \mathbf{v}}{\partial \mathbf{x}}$
	$\frac{\partial^2 f}{\partial \mathbf{x}\partial \mathbf{x}^T} =$		$\mathbf{H}$, the Hessian matrix[5]

7.png

a is not a function of **x**	$\frac{\partial(\mathbf{a}\cdot\mathbf{x})}{\partial\mathbf{x}} = \frac{\partial(\mathbf{x}\cdot\mathbf{a})}{\partial\mathbf{x}} =$ $\frac{\partial\mathbf{a}^\top\mathbf{x}}{\partial\mathbf{x}} = \frac{\partial\mathbf{x}^\top\mathbf{a}}{\partial\mathbf{x}} =$	$\mathbf{a}^\top$	$\mathbf{a}$

8.png

Vector-by-scalar

Condition	Expression	Numerator layout, i.e. by **y**, result is column vector	Denominator layout, i.e. by **y**$^{\mathbf{T}}$, result is row vector
a is not a function of *x*	$\frac{\partial\mathbf{a}}{\partial x} =$	$\mathbf{0}$[4]	
a is not a function of *x*, **u** = **u**(*x*)	$\frac{\partial a\mathbf{u}}{\partial x} =$	$a\frac{\partial\mathbf{u}}{\partial x}$	
A is not a function of *x*, **u** = **u**(*x*)	$\frac{\partial\mathbf{A}\mathbf{u}}{\partial x} =$	$\mathbf{A}\frac{\partial\mathbf{u}}{\partial x}$	$\frac{\partial\mathbf{u}}{\partial x}\mathbf{A}^\top$
u = **u**(*x*)	$\frac{\partial\mathbf{u}^\top}{\partial x} =$	$\left(\frac{\partial\mathbf{u}}{\partial x}\right)^\top$	
u = **u**(*x*), **v** = **v**(*x*)	$\frac{\partial(\mathbf{u}+\mathbf{v})}{\partial x} =$	$\frac{\partial\mathbf{u}}{\partial x} + \frac{\partial\mathbf{v}}{\partial x}$	
u = **u**(*x*), **v** = **v**(*x*)	$\frac{\partial(\mathbf{u}^\top\times\mathbf{v})}{\partial x} =$	$\left(\frac{\partial\mathbf{u}}{\partial x}\right)^\top\times\mathbf{v}+\mathbf{u}^\top\times\frac{\partial\mathbf{v}}{\partial x}$	$\frac{\partial\mathbf{u}}{\partial x}\times\mathbf{v}+\mathbf{u}^\top\times\left(\frac{\partial\mathbf{v}}{\partial x}\right)^\top$
u = **u**(*x*)	$\frac{\partial\mathbf{g}(\mathbf{u})}{\partial x} =$	$\frac{\partial\mathbf{g}(\mathbf{u})}{\partial\mathbf{u}}\frac{\partial\mathbf{u}}{\partial x}$ Assumes consistent matrix layout; see below.	$\frac{\partial\mathbf{u}}{\partial x}\frac{\partial\mathbf{g}(\mathbf{u})}{\partial\mathbf{u}}$
u = **u**(*x*)	$\frac{\partial\mathbf{f}(\mathbf{g}(\mathbf{u}))}{\partial x} =$	$\frac{\partial\mathbf{f}(\mathbf{g})}{\partial\mathbf{g}}\frac{\partial\mathbf{g}(\mathbf{u})}{\partial\mathbf{u}}\frac{\partial\mathbf{u}}{\partial x}$ Assumes consistent matrix layout; see below.	$\frac{\partial\mathbf{u}}{\partial x}\frac{\partial\mathbf{g}(\mathbf{u})}{\partial\mathbf{u}}\frac{\partial\mathbf{f}(\mathbf{g})}{\partial\mathbf{g}}$
U = **U**(*x*), **v** = **v**(*x*)	$\frac{\partial(\mathbf{U}\times\mathbf{v})}{\partial x} =$	$\frac{\partial\mathbf{U}}{\partial x}\times\mathbf{v}+\mathbf{U}\times\frac{\partial\mathbf{v}}{\partial x}$	$\mathbf{v}^\top\times\left(\frac{\partial\mathbf{U}}{\partial x}\right)^\top+\frac{\partial\mathbf{v}}{\partial x}\times\mathbf{U}^\top$

9.png

参考:

[Matrix calculus]