矩阵求导总结

kissjz (/users/2mvvqvhuvrmiq) 2018-01-30 01:12:15 浏览5812

云栖社区 (/tags/type_blog-tagid_1/) 布局 (/tags/type_blog-tagid_687/)
换装攻略 (/tags/type_blog-tagid_25731/) 矩阵求导 (/tags/type_blog-tagid_25947/)

> 在看"多元线性回归"(multivariate linear regression)的时候又碰到了矩阵求导，决心把矩阵求导学一学了，网上搜了一些，也看了一下维基，简单整理一下，以后再碰到矩阵求导就方便查看了d=====(￣▽￣*)b

矩阵求导，没什么可怕的，**无非就是 一个矩阵里的元素对另一个矩阵里的元素进行求导，只是符合一些规则而已。**

和朋友的白水东城博客链接-矩阵求导总结 (https://yq.aliyun.com/go/articleRenderRedirect?spm=a2c4e.11153940.0.0.43d750427uusMo&url=http%3A%2F%2Fwww.baishuidongcheng.com%2F1000.html)

先从简单的来说(⊙ ▽ ⊙)

向量（矩阵的特例）对一个标量进行求导，就是向量的每个元素对这个标量进行求导，结果可以定为行向量也可定为列向量（这个牵扯到后文要提到的分布）

标量对一个向量求导，就是标量分别对这个向量的每个元素求导，结果也一样，可以为行向量也可以为列向量。

维基上提到了一个布局（layout）的定义，大概意思是这样的：
分子布局（Numerator layout）：分子为$\mathbf{y}$或者分母为$\mathbf{x}^T$（即分子为列向量或者分母为行向量）

分母布局（Denominator layout）：分子为$\mathbf{y}^T$或者分母为$\mathbf{x}$（即分子为行向量或者分母为列向量）

注：对于矩阵也是一样的，可以这样理解，（列）矩阵和（行）矩阵分别对应着列向量和行向量（如果现在这句话不能理解，不妨完下文的分布记法再来想想这句话的意思）。

用一个例子来解释一下这两个布局的区别和作用：

用$\frac{\partial \vec{y}}{\partial \vec{x}}$为例，假设$\vec{x}$和$\vec{y}$都是列向量，那么分子布局的意思就是$\frac{\partial \vec{y}}{\partial \vec{x}^{\mathrm{T}}}$,分母布局就是$\frac{\partial \vec{y}^{\mathrm{T}}}{\partial \vec{x}}$。这样就有明显的区分了，分子布局可以看做是求偏导的分子项保持不变，而分母项则需要转置，而分母布局正好相反。

再举个例子：

向量$\vec{y} = \begin{pmatrix} y_1 & y_2 & \cdots & y_m \end{pmatrix}^{\mathrm{T}}$对标量$x$的偏导，如下：

$$\frac{\partial \vec{y}}{\partial x} = \begin{pmatrix} \frac{\partial y_1}{\partial x} & \frac{\partial y_2}{\partial x} & \cdots & \frac{\partial y_m}{\partial x} \end{pmatrix}^{\mathrm{T}}$$

标量y对向量$\vec{x} = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix}^{\mathrm{T}}$求偏导，如下：

$$\frac{\partial y}{\partial \vec{x}} = \begin{pmatrix} \frac{\partial y}{\partial x_1} & \frac{\partial y}{\partial x_2} & \cdots & \frac{\partial y}{\partial x_n} \end{pmatrix}$$

第一个例子，结果还是列向量，但在第二个例子中结果却变成了行向量。注意这些都是分子布局的形式，如果换成是分母布局，那就要将上面的结果都转置过来才行（这也是之前所说的为什么有些资料上会把矩阵对标量的求导需要转置的原因）。同理，分子布局中向量
$\vec{y} = \begin{pmatrix} y_1 & y_2 & \cdots & y_m \end{pmatrix}^{\mathrm{T}}$对向量$\vec{x} = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix}^{\mathrm{T}}$求偏导的结果：

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

这就是著名的**雅可比（Jacobian）矩阵**。同样的，如果在分母布局中结果就是雅可比矩阵的转置。

又比如，对于分子布局一个$m \times n$行的矩阵$\mathbf{Y}$对于标量x求导的结果：

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

而标量y对矩阵$\mathbf{X}$的求导结果：

$$\frac{\partial y}{\partial X}=\begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{21}} & \cdots & \frac{\partial y}{\partial x_{m1}} \\ \frac{\partial y}{\partial x_{12}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{m2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial x_{1n}} & \frac{\partial y}{\partial x_{2n}} & \cdots & \frac{\partial y}{\partial x_{mn}} \end{bmatrix}$$

我们根据经验（虽然这里不是列向量和行向量，但可以这么记忆），分母布局就是 上述结果的转置形式。

可能会疑惑，为啥要这样区分呢？

由于我们一般将向量默许为行向量或列向量，而这两种形式都客观存在，也没有孰优孰劣。

所以在看一些数学公式 的推导中，比如求矩阵对标量的倒数，直觉上是觉得是矩阵的每个元素对标量求导，但却有时候要进行转置，原因就是因为矩阵求导的两种分布了。

下面整理了一些维基上这两个分布的各自记法以及求导的**参照表格**。

维基上的矩阵求导参照表格

分子布局记法（Numerator-layout notation）

- 标量/行向量：$\frac{\partial y}{\partial \mathbf{x}}=\left[\frac{\partial y}{\partial x_1}\frac{\partial y}{\partial x_2}\cdots\frac{\partial y}{\partial x_n}\right].$
- 列向量/标量：$\frac{\partial \mathbf{y}}{\partial x}=\begin{bmatrix} \frac{\partial y_1}{\partial x} \\ \frac{\partial y_2}{\partial x} \\ \vdots \\ \frac{\partial y_m}{\partial x} \end{bmatrix}.$
- 列向量/行向量：$\frac{\partial \mathbf{y}}{\partial \mathbf{x}}=\begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}.$

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

下面是分子布局中的特例：

- 矩阵/标量：$\frac{\partial \mathbf{Y}}{\partial x} = \begin{bmatrix} \frac{\partial y_{11}}{\partial x} & \frac{\partial y_{12}}{\partial x} & \cdots & \frac{\partial y_{1n}}{\partial x} \\ \frac{\partial y_{21}}{\partial x} & \frac{\partial y_{22}}{\partial x} & \cdots & \frac{\partial y_{2n}}{\partial x} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_{m1}}{\partial x} & \frac{\partial y_{m2}}{\partial x} & \cdots & \frac{\partial y_{mn}}{\partial x} \end{bmatrix}.$
- 对矩阵(X)每个元素求导：$d\mathbf{X} = \begin{bmatrix} dx_{11} & dx_{12} & \cdots & dx_{1n} \\ dx_{21} & dx_{22} & \cdots & dx_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ dx_{m1} & dx_{m2} & \cdots & dx_{mn} \end{bmatrix}.$

分母布局标记（Denominator-layout notation）

- 标量/列向量：$\frac{\partial y}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix}.$
- 行向量/标量：$\frac{\partial \mathbf{y}}{\partial x} = \left[\frac{\partial y_1}{\partial x} \frac{\partial y_2}{\partial x} \cdots \frac{\partial y_m}{\partial x} \right].$
- 行向量/列向量：$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_n} & \frac{\partial y_2}{\partial x_n} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}.$
- 标量/矩阵：$\frac{\partial y}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{12}} & \cdots & \frac{\partial y}{\partial x_{1q}} \\ \frac{\partial y}{\partial x_{21}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{2q}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial x_{p1}} & \frac{\partial y}{\partial x_{p2}} & \cdots & \frac{\partial y}{\partial x_{pq}} \end{bmatrix}.$

好了，知道啥叫分子布局和分母布局了，然后对于要求导的式子，就可以对照着快速查阅下面的表格了=￣ω￣=

1.**向量—向量** Vector-by-vector identities

所谓的向量的计算，就是矩阵计算中最简单的一种，是把矩阵的行向量设为1的情况.

Identities: vector-by-vector $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$

Condition	Expression	Numerator layout, i.e. by y and x$^\mathsf{T}$	Denominator layout, i.e. by y$^\mathsf{T}$ and x
a is not a function of **x**	$\frac{\partial \mathbf{a}}{\partial \mathbf{x}} =$	$\mathbf{0}$	
	$\frac{\partial \mathbf{x}}{\partial \mathbf{x}} =$	$\mathbf{I}$	
A is not a function of **x**	$\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} =$	$\mathbf{A}$	$\mathbf{A}^\top$
A is not a function of **x**	$\frac{\partial \mathbf{x}^\top \mathbf{A}}{\partial \mathbf{x}} =$	$\mathbf{A}^\top$	$\mathbf{A}$
a is not a function of **x**, **u** = **u**(**x**)	$\frac{\partial a\mathbf{u}}{\partial \mathbf{x}} =$	$a\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	
a = *a*(**x**), **u** = **u**(**x**)	$\frac{\partial a\mathbf{u}}{\partial \mathbf{x}} =$	$a\frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \mathbf{u}\frac{\partial a}{\partial \mathbf{x}}$	$a\frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \frac{\partial a}{\partial \mathbf{x}}\mathbf{u}^\top$
A is not a function of **x**, **u** = **u**(**x**)	$\frac{\partial \mathbf{A}\mathbf{u}}{\partial \mathbf{x}} =$	$\mathbf{A}\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}}\mathbf{A}^\top$
u = **u**(**x**), **v** = **v**(**x**)	$\frac{\partial (\mathbf{u}+\mathbf{v})}{\partial \mathbf{x}} =$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \frac{\partial \mathbf{v}}{\partial \mathbf{x}}$	
u = **u**(**x**)	$\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{x}} =$	$\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}}\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}$
u = **u**(**x**)	$\frac{\partial \mathbf{f}(\mathbf{g}(\mathbf{u}))}{\partial \mathbf{x}} =$	$\frac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}}\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}}\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\frac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}}$

2.**标量—向量** Scalar-by-vector identities

其实只要在上面的"向量—向量"运算中把分子或分母中的向量该称标量就得到"标量—向量"了。

Identities: scalar-by-vector $\frac{\partial y}{\partial \mathbf{x}} = \nabla_\mathbf{x} y$

Condition	Expression	Numerator layout, i.e. by x$^\mathsf{T}$; result is row vector	Denominator layout, i.e. by x; result is column vector
a is not a function of **x**	$\frac{\partial a}{\partial \mathbf{x}} =$	$\mathbf{0}^\top$ [4]	$\mathbf{0}$ [4]
a is not a function of **x**, *u* = *u*(**x**)	$\frac{\partial au}{\partial \mathbf{x}} =$	$a\frac{\partial u}{\partial \mathbf{x}}$	
u = *u*(**x**), *v* = *v*(**x**)	$\frac{\partial (u+v)}{\partial \mathbf{x}} =$	$\frac{\partial u}{\partial \mathbf{x}} + \frac{\partial v}{\partial \mathbf{x}}$	
u = *u*(**x**), *v* = *v*(**x**)	$\frac{\partial uv}{\partial \mathbf{x}} =$	$u\frac{\partial v}{\partial \mathbf{x}} + v\frac{\partial u}{\partial \mathbf{x}}$	
u = *u*(**x**)	$\frac{\partial g(u)}{\partial \mathbf{x}} =$	$\frac{\partial g(u)}{\partial u}\frac{\partial u}{\partial \mathbf{x}}$	
u = *u*(**x**)	$\frac{\partial f(g(u))}{\partial \mathbf{x}} =$	$\frac{\partial f(g)}{\partial g}\frac{\partial g(u)}{\partial u}\frac{\partial u}{\partial \mathbf{x}}$	
u = **u**(**x**), **v** = **v**(**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 $\frac{\partial \mathbf{u}}{\partial \mathbf{x}}, \frac{\partial \mathbf{v}}{\partial \mathbf{x}}$
u = **u**(**x**), **v** = **v**(**x**), **A** is not a function of **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]

$\mathbf{a}$ is not a function of $\mathbf{x}$	$\frac{\partial(\mathbf{a}\cdot\mathbf{x})}{\partial\mathbf{x}} = \frac{\partial(\mathbf{x}\cdot\mathbf{a})}{\partial\mathbf{x}} = \frac{\partial\mathbf{a}^\top\mathbf{x}}{\partial\mathbf{x}} = \frac{\partial\mathbf{x}^\top\mathbf{a}}{\partial\mathbf{x}} =$	$\mathbf{a}^\top$	$\mathbf{a}$
$\mathbf{A}$ is not a function of $\mathbf{x}$ $\mathbf{b}$ is not a function of $\mathbf{x}$	$\frac{\partial\mathbf{b}^\top\mathbf{A}\mathbf{x}}{\partial\mathbf{x}} =$	$\mathbf{b}^\top\mathbf{A}$	$\mathbf{A}^\top\mathbf{b}$
$\mathbf{A}$ is not a function of $\mathbf{x}$	$\frac{\partial\mathbf{x}^\top\mathbf{A}\mathbf{x}}{\partial\mathbf{x}} =$	$\mathbf{x}^\top(\mathbf{A}+\mathbf{A}^\top)$	$(\mathbf{A}+\mathbf{A}^\top)\mathbf{x}$
$\mathbf{A}$ is not a function of $\mathbf{x}$ $\mathbf{A}$ is symmetric	$\frac{\partial\mathbf{x}^\top\mathbf{A}\mathbf{x}}{\partial\mathbf{x}} =$	$2\mathbf{x}^\top\mathbf{A}$	$2\mathbf{A}\mathbf{x}$
$\mathbf{A}$ is not a function of $\mathbf{x}$	$\frac{\partial^2\mathbf{x}^\top\mathbf{A}\mathbf{x}}{\partial\mathbf{x}^2} =$	$\mathbf{A}+\mathbf{A}^\top$	
$\mathbf{A}$ is not a function of $\mathbf{x}$ $\mathbf{A}$ is symmetric	$\frac{\partial^2\mathbf{x}^\top\mathbf{A}\mathbf{x}}{\partial\mathbf{x}^2} =$	$2\mathbf{A}$	
	$\frac{\partial(\mathbf{x}\cdot\mathbf{x})}{\partial\mathbf{x}} = \frac{\partial\mathbf{x}^\top\mathbf{x}}{\partial\mathbf{x}} =$	$2\mathbf{x}^\top$	$2\mathbf{x}$
$\mathbf{a}$ is not a function of $\mathbf{x}$, $\mathbf{u}=\mathbf{u}(\mathbf{x})$	$\frac{\partial(\mathbf{a}\cdot\mathbf{u})}{\partial\mathbf{x}} = \frac{\partial\mathbf{a}^\top\mathbf{u}}{\partial\mathbf{x}} =$	$\mathbf{a}^\top\frac{\partial\mathbf{u}}{\partial\mathbf{x}}$ ▪ assumes numerator layout of $\frac{\partial\mathbf{u}}{\partial\mathbf{x}}$	$\frac{\partial\mathbf{u}}{\partial\mathbf{x}}\mathbf{a}$ ▪ assumes denominator layout of $\frac{\partial\mathbf{u}}{\partial\mathbf{x}}$
$\mathbf{a}$, $\mathbf{b}$ are not functions of $\mathbf{x}$	$\frac{\partial\,\mathbf{a}^\top\mathbf{x}\mathbf{x}^\top\mathbf{b}}{\partial\,\mathbf{x}} =$	$\mathbf{x}^\top(\mathbf{a}\mathbf{b}^\top+\mathbf{b}\mathbf{a}^\top)$	$(\mathbf{a}\mathbf{b}^\top+\mathbf{b}\mathbf{a}^\top)\mathbf{x}$
$\mathbf{A}$, $\mathbf{b}$, $\mathbf{C}$, $\mathbf{D}$, $\mathbf{e}$ are not functions of $\mathbf{x}$	$\frac{\partial\,(\mathbf{A}\mathbf{x}+\mathbf{b})^\top\mathbf{C}(\mathbf{D}\mathbf{x}+\mathbf{e})}{\partial\,\mathbf{x}} =$	$(\mathbf{D}\mathbf{x}+\mathbf{e})^\top\mathbf{C}^\top\mathbf{A}+(\mathbf{A}\mathbf{x}+\mathbf{b})^\top\mathbf{C}\mathbf{D}$	$\mathbf{D}^\top\mathbf{C}^\top(\mathbf{A}\mathbf{x}+\mathbf{b})+\mathbf{A}^\top\mathbf{C}(\mathbf{D}\mathbf{x}+\mathbf{e})$
$\mathbf{a}$ is not a function of $\mathbf{x}$	$\frac{\partial\,\lVert\mathbf{x}-\mathbf{a}\rVert}{\partial\,\mathbf{x}} =$	$\frac{(\mathbf{x}-\mathbf{a})^\top}{\lVert\mathbf{x}-\mathbf{a}\rVert}$	$\frac{\mathbf{x}-\mathbf{a}}{\lVert\mathbf{x}-\mathbf{a}\rVert}$

3.**向量—标量** Vector-by-scalar identities

Identities: vector-by-scalar $\frac{\partial\mathbf{y}}{\partial x}$

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

4.**标量—矩阵** Scalar-by-matrix identities

Identities: scalar-by-matrix $\frac{\partial y}{\partial \mathbf{X}}$

Condition	Expression	Numerator layout, i.e. by $\mathbf{X}^\top$	Denominator layout, i.e. by $\mathbf{X}$
a is not a function of $\mathbf{X}$	$\frac{\partial a}{\partial \mathbf{X}} =$	$\mathbf{0}^\top$ [6]	$\mathbf{0}$ [6]
a is not a function of $\mathbf{X}$, $u = u(\mathbf{X})$	$\frac{\partial au}{\partial \mathbf{X}} =$	$a\frac{\partial u}{\partial \mathbf{X}}$	
$u = u(\mathbf{X})$, $v = v(\mathbf{X})$	$\frac{\partial (u+v)}{\partial \mathbf{X}} =$	$\frac{\partial u}{\partial \mathbf{X}} + \frac{\partial v}{\partial \mathbf{X}}$	
$u = u(\mathbf{X})$, $v = v(\mathbf{X})$	$\frac{\partial uv}{\partial \mathbf{X}} =$	$u\frac{\partial v}{\partial \mathbf{X}} + v\frac{\partial u}{\partial \mathbf{X}}$	
$u = u(\mathbf{X})$	$\frac{\partial g(u)}{\partial \mathbf{X}} =$	$\frac{\partial g(u)}{\partial u}\frac{\partial u}{\partial \mathbf{X}}$	
$u = u(\mathbf{X})$	$\frac{\partial f(g(u))}{\partial \mathbf{X}} =$	$\frac{\partial f(g)}{\partial g}\frac{\partial g(u)}{\partial u}\frac{\partial u}{\partial \mathbf{X}}$	
$\mathbf{U} = \mathbf{U}(\mathbf{X})$	[5] $\frac{\partial g(\mathbf{U})}{\partial X_{ij}} =$	$\operatorname{tr}\left(\frac{\partial g(\mathbf{U})}{\partial \mathbf{U}}\frac{\partial \mathbf{U}}{\partial X_{ij}}\right)$	$\operatorname{tr}\left(\left(\frac{\partial g(\mathbf{U})}{\partial \mathbf{U}}\right)^\top\frac{\partial \mathbf{U}}{\partial X_{ij}}\right)$
		Both forms assume *numerator* layout for $\frac{\partial \mathbf{U}}{\partial X_{ij}}$, i.e. mixed layout if denominator layout for $\mathbf{X}$ is being used.	

5.**矩阵—标量** Matrix-by-scalar identities

Identities: matrix-by-scalar $\frac{\partial \mathbf{Y}}{\partial x}$

Condition	Expression	Numerator layout, i.e. by Y
$\mathbf{U} = \mathbf{U}(x)$	$\frac{\partial a\mathbf{U}}{\partial x} =$	$a\frac{\partial \mathbf{U}}{\partial x}$
$\mathbf{A}$, $\mathbf{B}$ are not functions of x, $\mathbf{U} = \mathbf{U}(x)$	$\frac{\partial \mathbf{AUB}}{\partial x} =$	$\mathbf{A}\frac{\partial \mathbf{U}}{\partial x}\mathbf{B}$
$\mathbf{U} = \mathbf{U}(x)$, $\mathbf{V} = \mathbf{V}(x)$	$\frac{\partial (\mathbf{U}+\mathbf{V})}{\partial x} =$	$\frac{\partial \mathbf{U}}{\partial x} + \frac{\partial \mathbf{V}}{\partial x}$
$\mathbf{U} = \mathbf{U}(x)$, $\mathbf{V} = \mathbf{V}(x)$	$\frac{\partial (\mathbf{UV})}{\partial x} =$	$\mathbf{U}\frac{\partial \mathbf{V}}{\partial x} + \frac{\partial \mathbf{U}}{\partial x}\mathbf{V}$
$\mathbf{U} = \mathbf{U}(x)$, $\mathbf{V} = \mathbf{V}(x)$	$\frac{\partial (\mathbf{U}\otimes\mathbf{V})}{\partial x} =$	$\mathbf{U}\otimes\frac{\partial \mathbf{V}}{\partial x} + \frac{\partial \mathbf{U}}{\partial x}\otimes\mathbf{V}$
$\mathbf{U} = \mathbf{U}(x)$, $\mathbf{V} = \mathbf{V}(x)$	$\frac{\partial (\mathbf{U}\circ\mathbf{V})}{\partial x} =$	$\mathbf{U}\circ\frac{\partial \mathbf{V}}{\partial x} + \frac{\partial \mathbf{U}}{\partial x}\circ\mathbf{V}$
$\mathbf{U} = \mathbf{U}(x)$	$\frac{\partial \mathbf{U}^{-1}}{\partial x} =$	$-\mathbf{U}^{-1}\frac{\partial \mathbf{U}}{\partial x}\mathbf{U}^{-1}$
$\mathbf{U} = \mathbf{U}(x,y)$	$\frac{\partial^2 \mathbf{U}^{-1}}{\partial x\partial y} =$	$\mathbf{U}^{-1}\left(\frac{\partial \mathbf{U}}{\partial x}\mathbf{U}^{-1}\frac{\partial \mathbf{U}}{\partial y} - \frac{\partial^2 \mathbf{U}}{\partial x\partial y} + \frac{\partial \mathbf{U}}{\partial y}\mathbf{U}^{-1}\frac{\partial \mathbf{U}}{\partial x}\right)\mathbf{U}^{-1}$
$\mathbf{A}$ is not a function of x, $\mathbf{g}(\mathbf{X})$ is any polynomial with scalar coefficients, or any matrix function defined by an infinite polynomial series (e.g. $e^{\mathbf{X}}$, $\sin(\mathbf{X})$, $\cos(\mathbf{X})$, $\ln(\mathbf{X})$, etc.); $g(x)$ is the equivalent scalar function, $g'(x)$ is its derivative, and $\mathbf{g}'(\mathbf{X})$ is the corresponding matrix function	$\frac{\partial\, \mathbf{g}(x\mathbf{A})}{\partial x} =$	$\mathbf{A}\mathbf{g}'(x\mathbf{A}) = \mathbf{g}'(x\mathbf{A})\mathbf{A}$
$\mathbf{A}$ is not a function of x	$\frac{\partial e^{x\mathbf{A}}}{\partial x} =$	$\mathbf{A}e^{x\mathbf{A}} = e^{x\mathbf{A}}\mathbf{A}$

6.**标量—标量** Scalar-by-scalar identities

就当复习

Identities: scalar-by-scalar, with vectors involved

Condition	Expression	Any layout (assumes dot product ignores row vs. column layout)
$\mathbf{u} = \mathbf{u}(x)$	$\frac{\partial g(\mathbf{u})}{\partial x} =$	$\frac{\partial g(\mathbf{u})}{\partial \mathbf{u}}\cdot\frac{\partial \mathbf{u}}{\partial x}$
$\mathbf{u} = \mathbf{u}(x)$, $\mathbf{v} = \mathbf{v}(x)$	$\frac{\partial (\mathbf{u}\cdot\mathbf{v})}{\partial x} =$	$\mathbf{u}\cdot\frac{\partial \mathbf{v}}{\partial x} + \frac{\partial \mathbf{u}}{\partial x}\cdot\mathbf{v}$

Identities: scalar-by-scalar, with matrices involved[5]

Condition	Expression	Consistent numerator layout, i.e. by Y and X^T	Mixed layout, i.e. by Y and X
U = **U**(x)	$\frac{\partial \lvert\mathbf{U}\rvert}{\partial x} =$	$\lvert\mathbf{U}\rvert \operatorname{tr}\left(\mathbf{U}^{-1}\frac{\partial \mathbf{U}}{\partial x}\right)$	
U = **U**(x)	$\frac{\partial \ln \lvert\mathbf{U}\rvert}{\partial x} =$	$\operatorname{tr}\left(\mathbf{U}^{-1}\frac{\partial \mathbf{U}}{\partial x}\right)$	
U = **U**(x)	$\frac{\partial^2 \lvert\mathbf{U}\rvert}{\partial x^2} =$	$\lvert\mathbf{U}\rvert\left[\operatorname{tr}\left(\mathbf{U}^{-1}\frac{\partial^2 \mathbf{U}}{\partial x^2}\right)+\left(\operatorname{tr}\left(\mathbf{U}^{-1}\frac{\partial \mathbf{U}}{\partial x}\right)\right)^2-\operatorname{tr}\left(\left(\mathbf{U}^{-1}\frac{\partial \mathbf{U}}{\partial x}\right)\left(\mathbf{U}^{-1}\frac{\partial \mathbf{U}}{\partial x}\right)\right)\right]$	
U = **U**(x)	$\frac{\partial g(\mathbf{U})}{\partial x} =$	$\operatorname{tr}\left(\frac{\partial g(\mathbf{U})}{\partial \mathbf{U}}\frac{\partial \mathbf{U}}{\partial x}\right)$	$\operatorname{tr}\left(\left(\frac{\partial g(\mathbf{U})}{\partial \mathbf{U}}\right)^\top\frac{\partial \mathbf{U}}{\partial x}\right)$
A is not a function of x, **g**(**X**) is any polynomial with scalar coefficients, or any matrix function defined by an infinite polynomial series (e.g. $e^{\mathbf{X}}$, sin(**X**), cos(**X**), ln (**X**), etc.); $g(x)$ is the equivalent scalar function, $g'(x)$ is its derivative, and **g**′(**X**) is the corresponding matrix function.	$\frac{\partial \operatorname{tr}(\mathbf{g}(x\mathbf{A}))}{\partial x} =$	$\operatorname{tr}(\mathbf{A}\mathbf{g}'(x\mathbf{A}))$	
A is not a function of x	$\frac{\partial \operatorname{tr}(e^{x\mathbf{A}})}{\partial x} =$	$\operatorname{tr}(\mathbf{A}e^{x\mathbf{A}})$	

其实，这些求导的表格需要记忆，只要在遇到求导的地方对照着查阅一下就OK了。

最后，来解决我遇到的矩阵求导：

$$\frac{\partial (y-Xw)^T(y-Xw)}{\partial w}$$

注:ω是$(d+1)\times 1$的列向量,X是$m\times(d+1)$的矩阵,y是$m\times 1$的列向量.

首先,将式子展开:

$$\frac{\partial (y^Ty-y^TXw-w^TX^Ty+w^TX^TXw)}{\partial w}$$

$$\Rightarrow$$

$$\frac{\partial y^Ty}{\partial w}-\frac{\partial y^TXw}{\partial w}-\frac{\partial w^TX^Ty}{\partial w}+\frac{\partial w^TX^TXw}{\partial w}$$

分别对每一项求导:

- $$\frac{\partial y^Ty}{\partial w}:\frac{\partial y^Ty}{\partial w}=0$$

解释:分子为标量,分母为向量。找到上文2.**标量—向量**表格中的第一行

Condition	Expression	Numerator layout, i.e. by $\mathbf{x}^\top$; result is row vector	Denominator layout, i.e. by x; result is column vector
a is not a function of **x**	$\frac{\partial a}{\partial \mathbf{x}} =$	$\mathbf{0}^\top$ [4]	$\mathbf{0}$ [4]

又因为分母是列向量，所以是分母布局，结果为0.

- $$\frac{\partial y^TXw}{\partial w}:\frac{\partial y^TXw}{\partial w}=X^Ty$$

解释：分子为标量，分母为列向量。找到2.**标量—向量**表格中第11行

A is not a function of **x** **b** is not a function of **x**	$\frac{\partial \mathbf{b}^\top\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} =$	$\mathbf{b}^\top\mathbf{A}$	$\mathbf{A}^\top\mathbf{b}$

又因为分母是列向量，所以结果对应为分母布局，应该是第二个X^Ty.

- $$\frac{\partial w^T X^T y}{\partial w}:\frac{\partial w^T X^T y}{\partial w}=\frac{\partial (w^T X^T y)^T}{\partial w}$$
$$=\frac{\partial y^T Xw}{\partial w}=X^T y$$

解释：分子为标量，分母为向量。因为$\omega^T X^T y$是标量，而标量的转置还是等于本身，所以进行转置后再查表，和上一条一样。（为啥要转置？因为表格中没有一模一样的啊(o_ _)ﾉ）

- $$\frac{\partial w^T X^T Xw}{\partial w}:\frac{\partial w^T X^T Xw}{\partial w}=2X^T Xw$$

解释：分子为标量，分母为向量。在2.**标量—向量**表格中第13行找到

A is not a function of **x** **A** is symmetric	$\frac{\partial \mathbf{x}^\top \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} =$	$2\mathbf{x}^\top \mathbf{A}$	$2\mathbf{A}\mathbf{x}$

注：（行向量（$1\times(d+1)$）×矩阵（$(d+1)\times(m)$）× 矩阵（$m\times(d+1)$）×列向量（$(d+1)\times 1$=标量）

又因为分母为列向量，所以符合分母布局，选第二个，结果为$2X^T Xw$.

最终，我们得到结果：

$$\frac{\partial y^T y}{\partial w}-\frac{\partial y^T Xw}{\partial w}-\frac{\partial w^T X^T y}{\partial w}+\frac{\partial w^T X^T Xw}{\partial w}=$$
$$0-X^T y-X^T y+2X^T Xw=2X^T(Xw-y)$$

令上式结果为零，不考虑矩阵逆的计算等，简单的可以得到：

$$\omega=(X^T X)^{-1}X^T y$$

> 哈哈哈，以后再也不怕矩阵求导了(•̀ω •́)y

参考：
Matrix calculus-Wikipedia (https://yq.aliyun.com/go/articleRenderRedirect?url=https%3A%2F%2Fen.wikipedia.org%2Fwiki%2FMatrix_calculus)
通过一个例子快速上手矩阵求导-CSDN (https://yq.aliyun.com/go/articleRenderRedirect?url=http%3A%2F%2Fblog.csdn.net%2Fnomadlx53%2Farticle%2Fdetails%2F50849941)