矩阵求导（一）

矩阵求导（一）

thinkando 关注

- 矩阵求导（Matrix Derivative）也称作矩阵微分（Matrix Differential），在机器学习、图像处理、最优化等领域的公式推导中经常用到。

1. 布局约定（Layout conventions）

- 布局（Layout）：在矩阵求导中有两种布局，分别为分母布局(denominator layout)和分子布局(numerator layout)。这两种不同布局的求导规则是不一样的。

布局（Layout）：在矩阵求导中有两种布局，分别为**分母布局(denominator layout)**和**分子布局(numerator layout)**。这两种不同布局的求导规则是不一样的。

向量 $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$，关于标量$x$ 的求导，

在**分子布局**下，为：

$$\frac{\partial \mathbf{y}}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x} \\ \frac{\partial y_2}{\partial x} \\ \vdots \\ \frac{\partial y_n}{\partial x} \end{bmatrix} \tag{1}$$

而在**分母布局**下，为：

$$\frac{\partial \mathbf{y}}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x} & \frac{\partial y_2}{\partial x} & \cdots & \frac{\partial y_n}{\partial x} \end{bmatrix} \tag{2}$$

通过观察和推导我们可以知道，分子布局和分母布局之间刚好差一个转置，即在分子布局下与原来$\mathbf{Y}$相同，而在分母布局下差一个转置。

对于正切矩阵$\frac{\partial y}{\partial x}$采用分母布局，即$\mathbf{Y}^\top$，很不符合表达的习惯，所以本文中我们采用的是**分子布局**。

image.png

2. 关于标量的导数

2.1 标量与标量X的求导

这中情况就是我们平时的代数求导，直接就是$\frac{\partial y}{\partial x}$

image.png

2.2 向量与标量X的求导

矩阵求导（一）

向量 $\mathbf{y}=\begin{bmatrix} y_2 \\ \vdots \end{bmatrix}$，关于标量$x$ 的求导就是 $\mathbf{y}$ 的每一个元素分别对x求导，可以表示为

image.png

2.3 矩阵与标量X的求导

矩阵对标量的求导类似于向量关于标量的求导，也就是矩阵的每个元素分别对标量x求导，矩阵 $\mathbf{Y}=\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nn} \end{bmatrix}$ 对标量

x的导数为

$$\frac{\partial \mathbf{Y}}{\partial x}=\begin{bmatrix} \frac{\partial y_{11}}{\partial x} & \frac{\partial y_{12}}{\partial x} & \cdots & \frac{\partial y_{1n}}{\partial x} \\ \frac{\partial y_{21}}{\partial x} & \frac{\partial y_{22}}{\partial x} & \cdots & \frac{\partial y_{2n}}{\partial x} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_{n1}}{\partial x} & \frac{\partial y_{n2}}{\partial x} & \cdots & \frac{\partial y_{nn}}{\partial x} \end{bmatrix} \quad (4$$

image.png

3. 关于向量的导数

3.1 标量关于向量 x 的导数

标量y 关于向量 $\mathbf{x}=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ 的求导可以表示为

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

此时的向量叫做**梯度向量**。 $\frac{\partial y}{\partial \mathbf{x}}$ 为标量y 在空间 $\mathbb{R}^n$ 的梯度，该空间以x 为基。

image.png

3.2 向量关于向量 x 的导数

向量函数（即函数组成的向量）$\mathbf{y}=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ 关于向量$\mathbf{x}=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ 的导数记作

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

此时获得的矩阵 $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$ 叫做**Jacobian 矩阵**。

image.png

矩阵 $\mathbf{Y} = \begin{bmatrix} \cdots \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nn} \end{bmatrix}$ 对向量$\mathbf{x} = \begin{bmatrix} \cdots \\ \vdots \\ x_n \end{bmatrix}$ 的导数是推导中最复杂的一种，我们可以表示为

$$\frac{\partial \mathbf{Y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_{11}}{\partial x_1} & \frac{\partial y_{1n}}{\partial x_2} & \cdots & \frac{\partial y_{1n}}{\partial x_n} \\ \frac{\partial y_{21}}{\partial x_1} & \frac{\partial y_{22}}{\partial x_2} & \cdots & \frac{\partial y_{2n}}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_{n1}}{\partial x_1} & \frac{\partial y_{n2}}{\partial x_2} & \cdots & \frac{\partial y_{nn}}{\partial x_n} \end{bmatrix}$$

image.png

4. 关于矩阵的导数

我们一般只考虑标量关于矩阵的导数（因为矩阵对向量和矩阵的导数与前面2.3节的内容一致或相似），即标量y 对矩阵 $\mathbf{X}$ 的导数为 $\frac{\partial y}{\partial \mathbf{X}}$，此时的导数是**梯度矩阵**，可以表示为下式：

$$\frac{\partial y}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{21}} & \cdots & \frac{\partial y}{\partial x_{n1}} \\ \frac{\partial y}{\partial x_{12}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{n2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial x_{1n}} & \frac{\partial y}{\partial x_{2n}} & \cdots & \frac{\partial y}{\partial x_{nn}} \end{bmatrix} \tag{8}$$

image.png

5. 维度分析（没怎么看懂）

- 当我们对一些复杂的矩阵乘积求偏导的时候，直接求很难直接求出，这时候我们可以通过分析矩阵的维度来得到结果。例如:

考虑以下导数 $\frac{\partial \mathbf{Au}}{\partial \mathbf{x}}$，其中 $\mathbf{A}$ 与$\mathbf{x}$ 无关 且有 $\mathbf{A} \in \mathbb{R}^{m\times n}$， $\mathbf{u} \in BbbR^{n\times 1}$，$\mathbf{x} \in \mathbb{R}^{p\times 1}$，我们知道结果肯定和$\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$ 有关，于是先把 $\mathbf{A}$ 提出求导式，至于到了哪暂时不知道，接着我们知道 $\frac{\partial \mathbf{u}}{\partial \mathbf{x}} \in \mathbb{R}^{p\times n}$，于是 $\mathbf{A}$ 只能转置后添加到后面。因此有

$$\frac{\partial \mathbf{Au}}{\partial \mathbf{x}} = \$\frac{\partial \mathbf{u}}{\partial \mathbf{x}}\mathbf{A}^\top \tag{9}$$

再考虑问题 $\frac{\partial \mathbf{x}^\top \mathbf{Ax}}{\partial \mathbf{x}}$，其中 $\mathbf{A} \in \mathbb{R}^{n\times n}$， $\mathbf{x} \in \mathbb{R}^{n\times n}$，
为了分析这个问题我们考虑一个更一半的问题

$$\frac{\partial \mathbf{x}^\top \mathbf{Ax}}{\partial \mathbf{x}} \tag{10}$$

其中 $\mathbf{A} \in \mathbb{R}^{n\times n}$， $\mathbf{x} \in BbbR^{n\times n}$，且 $\mathbf{A}$ 与$\mathbf{x}$ 和 $\mathbf{y}$ 无关 。于是我们利用维度分析，采用非精确的乘积法则，可以将它分为两个部分

$$\frac{\partial (\mathbf{x}^\top \mathbf{A})\mathbf{y}}{\partial \mathbf{x}} \tag{11}$$

于是结果与两部分相关，一个是

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} \in \mathbb{R}^{m\times n} \tag{12}$$

另一个是

$$\frac{\partial \mathbf{x}^\top \mathbf{A}}{\partial \mathbf{x}} = \mathbf{A} \in \mathbb{R}^{m\times n} \tag{13}$$

同样通过维度分析，我们可以得到

$$\frac{\partial (\mathbf{x}^\top \mathbf{A})\mathbf{y}}{\partial \mathbf{x}} = \frac{\partial \mathbf{y}}{\partial \mathbf{x}}\mathbf{A}^\top \mathbf{x} + \mathbf{A}\mathbf{y} \tag{14}$$

因此经过维度的比较我们可以得到

$$\frac{\partial \mathbf{x}^\top \mathbf{Ax}}{\partial \mathbf{x}} = (\mathbf{A}^\top + \mathbf{A})\mathbf{x} \tag{14}$$

通过以上两个示例的学习，我们可以知道在求解复杂矩阵的求导问题时，通过维度来判断矩阵的导数形式很简便同时也不容易出错。下图是机器学习中常见的矩阵求导形式，可供参考：

image.png

赞 回复

穆新星
3楼 03.27 10:16

3.3好像不对

2 回复

矩阵求导（一）

$$\frac{\partial \beta^T X}{X} = \beta$$

$$\frac{\partial X^T X}{\partial X} = X$$

$$\frac{\partial X^T A X}{\partial X} = (A + A^T)X$$

Andrew Ng 推荐的使用矩阵的迹的相关公式：

$$\mathrm{tr}(a) = a$$

$$\mathrm{tr}(AB) = \mathrm{tr}(BA)$$

$$\mathrm{tr}(ABC) = \mathrm{tr}(CAB) = \mathrm{tr}(BCA)$$

$$\frac{\partial tr(AB)}{\partial A} = B^T$$

$$\mathrm{tr}(A) = \mathrm{tr}(A^T)$$

$$\partial tr(ABA^T C)$$

image.png

参考文献

1. https://blog.csdn.net/u010976453/article/details/54381248