矩阵微分

目录

在实际计算过程中，按照定义求解矩阵梯度往往不现实。如果还要计算Hessian 矩阵就更复杂了。那么有没有简单的办法呢？有。矩阵微分为标量函数，向量函数和矩阵函数的梯度矩阵与Hessian矩阵的计算提供了便捷的算法。

矩阵微分是实函数微分对矩阵函数的推广。在微积分中，函数$\phi(x)$ 的导数$\phi^{'}(x)$ 定义为：

$$\phi^{'}(x) = \lim_{\Delta x \to 0} \frac{\phi(x+\Delta x) - \phi(x)}{\Delta x} \tag{1}$$

式 (1) 又可以写为：

$$\phi(x+\Delta x) = \phi(x) + \Delta x \phi^{'}(x) + r_x(\Delta) \tag{2}$$

式 (2) 中 $r_x(\Delta x)$是$\phi(x)$的留数，有$\lim_{\Delta x \to 0} \frac{r_x(\Delta x)}{\Delta x} = 0$ 。 $\phi(x+\Delta x) - \phi(x)$由两部分构成：与导数成正比的分量$\Delta x \phi^{'}(x)$ 和误差分量。如果误差分量可以忽略不计则函数$\phi(x+\Delta x) = \phi(x) + \Delta x \phi^{'}(x)$

把上述定义推广到向量，即可得到向量函数的微分。令$\mathbf{c}$ 和 $\mathbf{u}$是两个$(n \times 1)$向量，其中$\mathbf{c}$是 $m \times 1$ 向量函数$\mathbf{f}(x)$ 的向量变元，而$\mathbf{u}$是 R^n中的一个点，并且$\|\mathbf{u}\|_2 < r$ 。若存在一个$m \times n$的矩阵$\mathbf{A}$，它与$\mathbf{c}$有关，但与 $\mathbf{u}$无关，并且使得：

$$f(\mathbf{c}+\mathbf{u}) = f(\mathbf{c}) + \mathbf{A}(\mathbf{c})\mathbf{u} + r_{\mathbf{c}}(\mathbf{u}) \tag{3}$$

对所有具有$\|\mathbf{u}\|_2 < r$的$n \times 1$实向量 $\mathbf{u}$恒成立，以及：

$$\lim_{\mathbf{u} \to \mathbf{0}} \frac{r_c(\mathbf{u})}{\|\mathbf{u}\|_2} = 0 \tag{4}$$

则称函数f在向量点$\mathbf{c}$是可微分的。称$m\times n$矩阵 $\mathbf{A}(\mathbf{c})$是函数f 在$\mathbf{c}$的一阶导数，并称 $m\times 1$向量

$$df(\mathbf{c};\mathbf{u}) = \mathbf{A}(\mathbf{c})\mathbf{u} \tag{5}$$

是f在c的一阶微分向量。

现在考虑另外一个向量点$\mathbf{c}+t\mathbf{e}_j$，其中e_j 是一个仅第j个元素为1，其他元素为0的$n\times 1$的基本向量。t 是非零标量。换句话说$\mathbf{c}$与$\mathbf{c}+t\mathbf{e}_j$只第j个元素不同。当t足够小时，极限：

$$D_j\mathbf{f}_i(c) = \lim_{t\to 0}\frac{f_i(c+t\mathbf{e}_j)-f_i(\mathbf{c})}{t} \tag{6}$$

称为向量函数$\mathbf{f}$ 的第i个元素相对于向量变元$\mathbf{c}$ 的第j个元素的偏导。

根据以上定义，n个偏导的排列$[D_1\mathbf{f}_i(\mathbf{c}), D_2\mathbf{f}_i(\mathbf{c}), \ldots, D_n\mathbf{f}_i(\mathbf{c})]$。定义了向量函数$f(\mathbf{c})$的第$i$个元素 $\mathbf{f}_i(\mathbf{c})$ 相对于$1\times n$行向量 $\mathbf{c}^T$的偏导向量，即有：

$$\frac{\partial f_i(\mathbf{c})}{\partial \mathbf{c}^T} = [D_1\mathbf{f}_i(\mathbf{c}), D_2\mathbf{f}_i(\mathbf{c}), \ldots, D_n\mathbf{f}_i(\mathbf{c})] \tag{7}$$

于是，列向量$\mathbf{f}(\mathbf{c}) = [\mathbf{f}_1(\mathbf{c}), \mathbf{f}_2(\mathbf{c}), \ldots, \mathbf{f}_m(\mathbf{c})]^T$相对于行向量$\mathbf{c}^T = [c_1, c_2, \ldots, c_n]$ 的偏导数矩阵为$m\times n$矩阵。定义为：

$$D[\mathbf{f}(\mathbf{c})] = \begin{bmatrix} \frac{\partial f_1(c)}{\partial \mathbf{c}^T} \\ \frac{\partial f_2(c)}{\partial \mathbf{c}^T} \\ \vdots \\ \frac{\partial f_m(c)}{\partial \mathbf{c}^T} \end{bmatrix} = \begin{bmatrix} D_1\mathbf{f}_1(\mathbf{c}) & D_2\mathbf{f}_1(\mathbf{c}) & \ldots & D_n\mathbf{f}_1(\mathbf{c}) \\ D_1\mathbf{f}_2(\mathbf{c}) & D_2\mathbf{f}_2(\mathbf{c}) & \ldots & D_n\mathbf{f}_2(\mathbf{c}) \\ \vdots & & & \\ D_1\mathbf{f}_m(\mathbf{c}) & D_2\mathbf{f}_m(\mathbf{c}) & \ldots & D_n\mathbf{f}_m(\mathbf{c}) \end{bmatrix} \tag{8}$$

矩阵微分有很多性质和单变量函数微分性质类似，此处不再一一列举。只待用到的时候再去查阅工具书。