第 2 章补充材料

● 矩阵和矢量的导数

令 $f(x_1,x_2,\cdots,x_d)$ 是一个 d 元标量函数，如果这 d 个自变量刚好是矢量 $\mathbf{x}=(x_1,x_2,\cdots,x_d)^t$ 的元素，那么这个函数可以表示为以矢量 $\mathbf{x}$ 为自变量的函数 $f(\mathbf{x})$。多元函数 f 可以对它的每一个自变量求偏导数 $\partial f/\partial x_i$，所有这些偏导数可以表示为一个 d 维的列矢量，称为函数 $f(\mathbf{x})$ 关于矢量 $\mathbf{x}$ 的导数，也称为函数 $f(\mathbf{x})$ 的梯度矢量：

$$\nabla f(\mathbf{x})=\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}=\begin{pmatrix}\partial f/\partial x_1\\ \partial f/\partial x_2\\ \vdots\\ \partial f/\partial x_d\end{pmatrix} \qquad (1)$$

同样，如果 nm 元函数 $f(w_{11},\cdots,w_{1m},w_{21},\cdots,w_{nm})$ 的自变量是 $n\times m$ 矩阵 $\mathbf{W}$ 的元素，那么函数也可以表示为以矩阵 $\mathbf{W}$ 为自变量的函数 $f(\mathbf{W})$，函数关于矩阵 $\mathbf{W}$ 的导数可以表示为一个 $n\times m$ 的矩阵：

$$\frac{\partial f(\mathbf{W})}{\partial \mathbf{W}}=\begin{pmatrix}\partial f/\partial w_{11} & \partial f/\partial w_{12} & \cdots & \partial f/\partial w_{1m}\\ \partial f/\partial w_{21} & \partial f/\partial w_{22} & \cdots & \partial f/\partial w_{2m}\\ \vdots & \vdots & \ddots & \vdots\\ \partial f/\partial w_{n1} & \partial f/\partial w_{n2} & \cdots & \partial f/\partial w_{nm}\end{pmatrix} \qquad (2)$$

再来看更复杂一些的情况。令 $\mathbf{f}(\mathbf{x})=\left(f_1(\mathbf{x}),f_2(\mathbf{x}),\cdots,f_n(\mathbf{x})\right)^t$ 是一个定义在矢量 $\mathbf{x}$ 上的矢量函数，也就是说 $\mathbf{f}(\mathbf{x})$ 的每一维元素都是以矢量 $\mathbf{x}$ 为自变量的标量函数。$\mathbf{f}(\mathbf{x})$ 通常也称为由 d 维空间到 n 维空间的映射，记作 $\mathbf{f}(\mathbf{x})$：$R^d\to R^n$。$\mathbf{f}(\mathbf{x})$ 关于自变量 $\mathbf{x}$ 的导数是一个 $n\times d$ 的矩阵，称为雅克比（Jacobi）矩阵：

$$\mathbf{J}(\mathbf{x})=\frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}}=\begin{pmatrix}\partial f_1/\partial x_1 & \partial f_1/\partial x_2 & \cdots & \partial f_1/\partial x_d\\ \partial f_2/\partial x_1 & \partial f_2/\partial x_2 & \cdots & \partial f_2/\partial x_d\\ \vdots & \vdots & \ddots & \vdots\\ \partial f_n/\partial x_1 & \partial f_n/\partial x_2 & \cdots & \partial f_n/\partial x_d\end{pmatrix} \qquad (3)$$

观察（1）式我们会发现，标量函数 $f(\mathbf{x})$ 关于 $\mathbf{x}$ 的梯度 $\nabla f(\mathbf{x})$ 就是一个定义在矢量 $\mathbf{x}$ 上的 d 维矢量函数，梯度矢量关于 $\mathbf{x}$ 的导数是一个 $d\times d$ 的矩阵，一般称为海森（Hessian）矩阵：

$$\mathbf{H}(\mathbf{x})=\frac{\partial \nabla f(\mathbf{x})}{\partial \mathbf{x}}=\begin{pmatrix}\partial^2 f/\partial x_1\partial x_1 & \partial^2 f/\partial x_1\partial x_2 & \cdots & \partial^2 f/\partial x_1\partial x_d\\ \partial^2 f/\partial x_2\partial x_1 & \partial^2 f/\partial x_2\partial x_2 & \cdots & \partial^2 f/\partial x_2\partial x_d\\ \vdots & \vdots & \ddots & \vdots\\ \partial^2 f/\partial x_n\partial x_1 & \partial^2 f/\partial x_n\partial x_2 & \cdots & \partial^2 f/\partial x_d\partial x_d\end{pmatrix} \qquad (4)$$

因为 $\partial^2 f/\partial x_i\partial x_j=\partial^2 f/\partial x_j\partial x_i$，所以海森矩阵 $\mathbf{H}(\mathbf{x})$ 是对称矩阵。

下面给出一些常用的关于矢量和矩阵导数的公式，其中 a 为标量，$\mathbf{x}$ 和 $\mathbf{y}$ 为 d 维矢量，$\mathbf{A}$ 和 $\mathbf{W}$ 为 $d\times d$ 的矩阵，$\mathbf{1}$ 为元素均为 1 的 d 维矢量：

1. $f(\mathbf{x})=\mathbf{x}^t\mathbf{y}=\mathbf{y}^t\mathbf{x}$： $\dfrac{\partial f(\mathbf{x})}{\partial \mathbf{x}}=\mathbf{y}$ （5）

2. $f(\mathbf{x})=\mathbf{x}^t\mathbf{A}\mathbf{x}$： $\dfrac{\partial f(\mathbf{x})}{\partial \mathbf{x}}=\left(\mathbf{A}+\mathbf{A}^t\right)\mathbf{x}$ （6）

   当 $\mathbf{A}$ 为对称矩阵时：$\dfrac{\partial f(\mathbf{x})}{\partial \mathbf{x}}=2\mathbf{A}\mathbf{x}$

3. $f(\mathbf{W})=\mathbf{x}^t\mathbf{W}\mathbf{y}$： $\dfrac{\partial f(\mathbf{W})}{\partial \mathbf{W}}=\mathbf{x}\mathbf{y}^t$ （7）

4. $f(\mathbf{x})=g(\mathbf{x})h(\mathbf{x})$： $\dfrac{\partial f(\mathbf{x})}{\partial \mathbf{x}}=\dfrac{\partial g(\mathbf{x})}{\partial \mathbf{x}}h(\mathbf{x})+\dfrac{\partial h(\mathbf{x})}{\partial \mathbf{x}}g(\mathbf{x})$ （8）

5. $f(\mathbf{x})=g(u)$，$u=h(\mathbf{x})$： $\dfrac{\partial f(\mathbf{x})}{\partial \mathbf{x}}=g'(u)\dfrac{\partial h(\mathbf{x})}{\partial \mathbf{x}}$ （9）