人工智能基础与进阶

数学基础

周 越

上海交通大学

目录 content

第一节 向量与特征表达

向量的基本定义

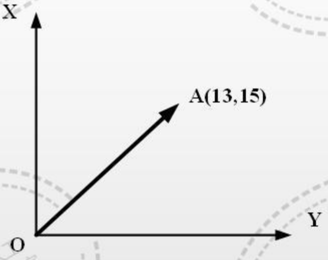
在数学中，向量（也称为欧几里得向量、几何向量、矢量），指具有大小（magnitude）和方向的量。它可以形象化地表示为带箭头的线段。箭头所指：代表向量的方向；线段长度：代表向量的大小。与向量对应的量叫做数量（物理学中称标量），数量（或标量）只有大小，没有方向。

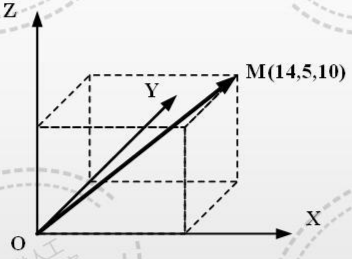
向量的记法：印刷体记作黑体（粗体）的字母（如**a**、**b**、**u**、**v**），书写时在字母顶上加一小箭头“→”。如：$\vec{a}, \vec{b}$

如果给定向量的起点（A）和终点（B），可将向量记作**AB**（并于顶上加→）。在空间直角坐标系中，也能把向量以数对形式表示，如**XOY**平面中(2, 3)是一向量。如：$\overrightarrow{\mathbf{AB}}$，(2,3)

向量的基本定义

常见的二维空间里的向量和三维空间中的向量：





N维空间里的向量？

向量的基本定义

把若干个数字并列放在一起，彼此之间用刻个分隔，再用括号围起来，这种样子的数组就可以称为向量。如果有n个数字，称为n维向量。

比如（2, 35, 72, 24）就是一个4维向量。

用行的方式排列的向量叫作行向量，用列的方式排列的向量叫作列向量。如果是行向量，加上一个上标符号T，T代表转置

$$\vec{\mathbf{a}} = [a_1 \quad a_2 \quad \cdots \quad a_n], \qquad \vec{\mathbf{a}}^T = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

向量的基本定义

向量的模

向量的大小，也就是向量的长度(或称模)。向量$\vec{\mathbf{a}}$的模记作$|\vec{\mathbf{a}}|$。

向量的模是非负实数，向量的模是可以比较大小的。向量 $\vec{\mathbf{a}} = (\mathbf{x}, \mathbf{y})$ ，

$$\|\vec{\mathbf{a}}\| = \sqrt{x^2 + y^2}$$

因为方向不能比较大小，所以向量也就不能比较大小。对于向量来说“大于”和“小于”的概念是没有意义的。例如: $\overrightarrow{AB} > \overrightarrow{CD}$ 是没有意义的。

若向量 $\vec{\mathbf{a}} = (a_1, a_2, a_3, \cdots a_n)$ ，模为:

$$\|\vec{\mathbf{a}}\| = \sqrt{a_1^2 + a_2^2 + a_3^2 + \cdots + a_n^2}$$

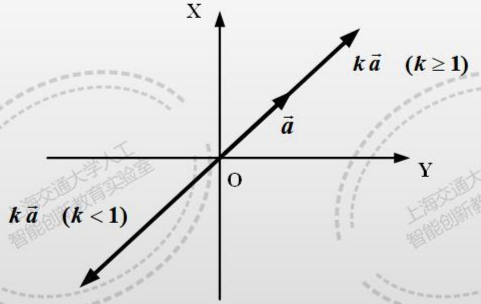
向量的基本运算

向量的数乘

向量的数乘就是向量和一个数字相乘，等于向量的各个分量都乘以相同的系数。向量数乘的几何意义就是把向量拉伸了k倍。

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \times k = \begin{bmatrix} ka_1 \\ ka_2 \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix} \times 7 = \begin{bmatrix} 7 \\ 21 \\ 28 \end{bmatrix}$$



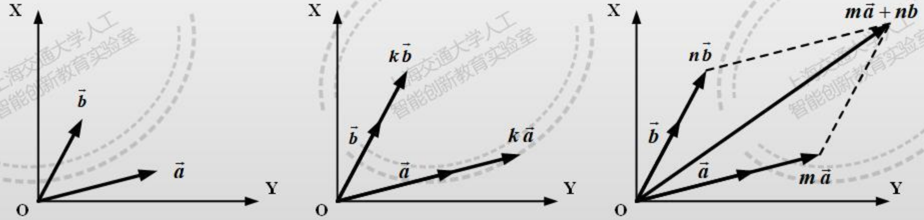
向量的基本运算

向量的线性组合

把向量的加法和数乘组合起来就得到了向量的线性组合，之所以叫线性是因为加法和乘法都是线性运算。

$$\vec{a}=\begin{bmatrix}1\\3\\4\end{bmatrix},\vec{b}=\begin{bmatrix}2\\11\\6\end{bmatrix} \qquad 4\vec{a}-6\vec{b}=\begin{bmatrix}4\times1-6\times2\\4\times3-6\times11\\4\times7-6\times6\end{bmatrix}=\begin{bmatrix}-8\\-54\\-8\end{bmatrix}$$

$$\vec{a}=\begin{bmatrix}a_1\\a_2\end{bmatrix},\ \vec{b}=\begin{bmatrix}b_1\\b_2\end{bmatrix} \qquad m\vec{a}+n\vec{b}=\begin{bmatrix}ma_1+nb_1\\ma_2+nb_2\end{bmatrix}$$



向量的基本运算

向量的乘法（点积）

两个向量的乘法（点积）等于两个向量对应元素之积的和。点积运算的公式如下：

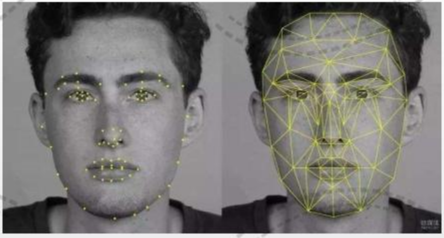
$$\vec{a}=\begin{bmatrix}a_1\\a_2\\\vdots\\a_n\end{bmatrix},\ \vec{b}=\begin{bmatrix}b_1\\b_2\\\vdots\\b_n\end{bmatrix}\qquad \vec{a}\cdot\vec{b}=\begin{bmatrix}a_1\\a_2\\\vdots\\a_n\end{bmatrix}\cdot\begin{bmatrix}b_1\\b_2\\\vdots\\b_n\end{bmatrix}=a_1b_1+a_2b_2+\cdots a_nb_n$$

两个向量的点积得到的是一个数字，在数值上等于两个向量的长度和夹角余弦之积。因此两个向量的夹角余弦可以用一下的公式计算：

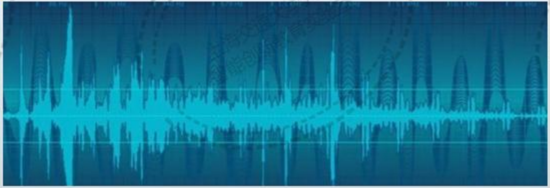
$$\cos\theta=\frac{\vec{a}\cdot\vec{b}}{\left\|\vec{a}\right\|\left\|\vec{b}\right\|}$$

数据特征的向量化

数据分析中最难解决的问题是数据的向量化，比如人脸识别和自然语言处理中，如何将人脸图像数据和语音的时序数据用一个向量表示。



$$\vec{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

$$\vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

第二节 行列式

行列式的基本知识

行列式是线性代数的一个重要组成部分. 它是研究矩阵、线性方程组、特征多项式的重要工具. 本章介绍了n阶行列式的定义、性质及计算方法，最后给出了它的一个简单应用

——克莱姆法则.

行列式的基本知识

记号: $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$ 称其为**二阶行列式**.

它表示数: $a_{11}a_{22} - a_{12}a_{21}$

即 $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$

对任何行列式，横排称为行，竖排称为列

a_{ij} 中 i 称为行标，j 称为列标。a_{ij} 表示第 i 行第 j 列的元素。

左上角到右下角表示**主对角线**

行列式的基本知识

例 1 $\begin{vmatrix} 5 & -1 \\ 3 & 2 \end{vmatrix} = 5\times 2-(-1)\times 3=13$

例 2 设 $D=\begin{vmatrix} \lambda^2 & \lambda \\ 3 & 1 \end{vmatrix}$,

(1)当 λ 为何值时， $D=0$,

(2)当 λ 为何值时， $D\neq 0$.

解 $\lambda^2-3\lambda=0 \Leftarrow\Rightarrow \lambda=0$, 或 $\lambda=3$

$D=\begin{vmatrix} \lambda^2 & \lambda \\ 3 & 1 \end{vmatrix} = \lambda^2$

行列式的基本知识

记号 $\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$ 称为**三阶行列式**.

它表示数

$$a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32}$$

即

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32}$$

行列式的基本知识

可以用**对角线法则**来记忆如下.

主对角线法

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32}$$

$$- a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32}$$

$$D = \begin{vmatrix} 1 & 2 & -4 \\ -2 & 2 & 1 \\ -3 & 4 & -2 \end{vmatrix}$$

$$D = 1\times2\times(-2) + 2\times1\times(-3) + (-4)\times(-2)\times4$$

$$-(-4)\times2\times(-3) - 2\times(-2)\times(-2) - 1\times1\times4$$

$$= -4 - 6 + 32 - 24 - 8 - 4$$

$$= -14$$

第三节　　矩　阵

矩阵的基本定义

矩阵就是n行m列的二维表格。如果行数和列数相等，即n=m，这样的矩阵就是方阵，否则就是非方阵。一个n行n列的方阵习惯称之为n阶方阵。

$$A=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} \quad \mathrm{B}=\begin{pmatrix} 2 & 5 & -3 \\ -1 & 7 & 14 \end{pmatrix} \quad \mathrm{C}=\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

如果一个矩阵只有一行，则称之为行向量；只有一列元素的矩阵则称之为列向量。B可以看作是由三个2×1的列向量构成，也可以看作是由两个1×3的行向量构成。

矩阵的基本定义

对于方阵，有种特殊的形态：对角矩阵D、单位矩阵E、对称矩阵F。

$$D=\begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{mm} \end{pmatrix}$$ 除了对角线之外的元素都是0

$$E=\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$ 对角线元素全为1，其他全为0；对角线元素为1的对焦矩阵

$$F=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}$$ 且$a_{ij} = a_{ji}$, 对称矩阵的特点是$A = A^T$。

$$A_1=\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix} \quad A_2=\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad A_3=\begin{pmatrix} 10 & 8 & 16 & 64 \\ 8 & 20 & 32 & 128 \\ 16 & 32 & 30 & 256 \\ 64 & 128 & 256 & 40 \end{pmatrix}$$

矩阵的基本运算

矩阵加法与数乘

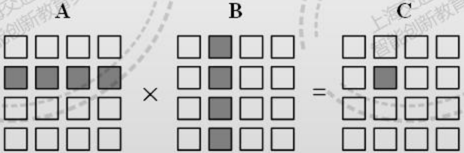
矩阵的加法与数乘运算是向量对应运算的延伸：

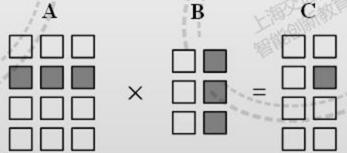
- 两个大小相同的矩阵的加法就是把其对应元素相加；
- 数字乘以矩阵就是把矩阵中的每一个元素与这个数字相乘。

矩阵乘法

矩阵之间定义的乘法是：如果矩阵A为$m\times n$大小的矩阵，矩阵B为$n\times p$大小的矩阵，那么矩阵AB为$m\times p$大小矩阵，其元素为

$$(AB)_{ij}=\sum_{k=1}^{n}a_{ik}b_{kj}=a_{i1}b_{1j}+a_{i2}b_{2j}+\cdots+a_{in}b_{nj}$$





矩阵的基本运算

矩阵乘法满足结合律

$$(AB)C = A(BC) = ABC$$

矩阵乘法-左右分配律

$$(A+B)C = AB + BC$$

$$C(A+B) = CA + CB$$

矩阵的乘法一般不满足交换律，即使AB和BA都有定义，通常

$$AB \neq BA$$

$$\begin{pmatrix}1 & 4\\ 2 & 3\end{pmatrix}\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}=\begin{pmatrix}0 & 4\\ 0 & 3\end{pmatrix} \quad \begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}1 & 4\\ 2 & 3\end{pmatrix}=\begin{pmatrix}0 & 0\\ 2 & 3\end{pmatrix}$$

矩阵乘法中需要明确相乘的顺序，AB称为B右乘A，或A左乘B。

向量的点积可以用矩阵乘法表示，两个n维向量r和v的点积，可以看做是行向量r^T与列向量v的矩阵乘法。

第四节　导数与梯度

导数基本概念

平均变化率

如果两个不同的变量x_1和x_2，分别对应不同的因变量y_1 y_2，要表示$y=f(x)$在x_1和x_2之间变化的快慢，只需将因变量的变化$(\Delta y=y_2-y_1=f(x_2)-f(x_1))$除以自变量的变化$(\Delta x=x_2-x_1)$，即

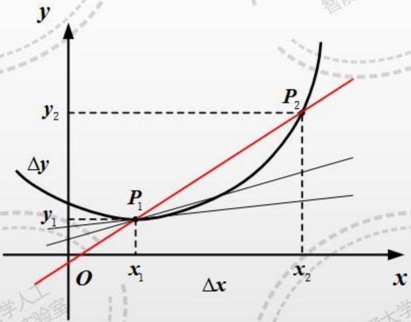
$$\frac{\Delta y}{\Delta x}=\frac{y_2-y_1}{x_2-x_1}=\frac{f(x_2)-f(x_1)}{x_2-x_1}$$

上式称为函数$y=f(x)$从x_1到x_2的平均变化率，也可以写成如下形式：

$$\frac{f(x_1+\Delta x)-f(x_1)}{\Delta x}$$

导数基本概念

平均变化率



对于一次线性函数:

$$y = f(x) = ax + b$$

计算它在 $x = x_0$ 的平均变化率:

$$\frac{\Delta y}{\Delta x} = \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

$$= \frac{a(x_0 + \Delta x) + b - (ax_0 + b)}{\Delta x}$$

$$= \frac{a\Delta x}{\Delta x} = a$$

线性函数的平均变化率就是它的斜率

平均变化率在函数图像上的意义为x_1和x_2对应的P_1和P_2之间连线（函数的割线）的斜率。

导数基本概念

瞬时变化率

平均变换率是在Δx内对应函数$y = f(x)$的变化速度。当Δx越取越小，不断逼近0，x_1和x_2也会越来越近，直至几乎变为同一个点。在这种情况下，原先定义的平均变化率也逐渐变为$y = f(x)$在(x_1, y_2)这一点处瞬间所具有的变化速度，称为瞬间变化率，记作：

$$\lim_{\Delta x \to 0} \frac{f(x_1 + \Delta x) - f(x_1)}{\Delta x}$$

极限符号$\lim\limits_{x \to c} f(x)$表示$x$趋于$c$时函数$f(x)$的值。计算表明，在$\Delta x$趋于0时，上式的值趋于一个定值，即为函数在这一点的导数。

导数基本概念

瞬时变化率

一个二次函数$y = f(x) = x^2$在$x = 1$附近的变化率，有

$$\frac{f(1+\Delta x) - f(1)}{\Delta x} = \frac{(1+\Delta x)^2 - 1^2}{\Delta x} = \frac{2\Delta x + (\Delta x)^2}{\Delta x} = 2 + \Delta x$$

而当Δx趋于0时，不难看出，这个式子的值趋于一个定值2，即

$$\lim_{\Delta x \to 0} \frac{f(1+\Delta x) - f(1)}{\Delta x} = 2$$

导数基本概念

导数的定义

假设函数$y=f(x)$在某个区间上的导数存在，则在此区间上的某个点$(x_1,f(x_1))$处的导数定义为：

$$f^{'}(x_1)=\lim_{\Delta x\to 0}\frac{f(x_1+\Delta x)-f(x_1)}{\Delta x}$$

此区间上所有点的导数构成以x为自变量的函数，称为导函数（有时也简称为导数），记为$f^{'}(x_1)\left(或y^{'},\frac{dy}{dx},\frac{df}{dx}\right)$。

寻找已知的函数在某点的导数或者其导函数的过程称之为求导。不难看出，当Δx趋于0时，点$x_1+\Delta x$与x_1无限接近，原本的割线变为函数图像的切线。因此导数的几何意义为函数$y=f(x)$的图像在点$(x_1,f(x_1))$处切线的斜率。

导数基本概念

导数的定义

常见的导数公式：

$C' = 0 \quad (C\text{为常数})$；

$(x^{\mu})' = \mu x^{\mu-1}$;

$(a^{x})' = a^{x}\ln a$;

$(e^{x})' = e^{x}$;

$(\log_a x)' = \frac{1}{x\ln a}$;

$(\ln x)' = \frac{1}{x}$;

$(\sin x)' = \cos x$;

$(\cos x)' = -\sin x$;

导数计算的一些主要性质如下

两个函数的和差：

$(u \pm v)' = u' \pm v'$;

两个函数的积

$(uv)' = u'v + uv'$;

两个函数的商

$$\left(\frac{u}{v}\right)' = \frac{u'v + uv'}{v^2}$$

复合函数：

$\{f[\varphi(x)]\}' = f'[\varphi(x)]\varphi'(x)$

或 $\frac{df}{dx} = \frac{df}{d\varphi}\cdot\frac{d\varphi}{dx}$（链式法则）

高阶导数和偏导数

导数$f'(x)$本身也可以视作是自变量x的函数，因此可以再导数的基础上再次求导，得到高阶导数。例如二阶导数：$y = f(x)$的一阶导数为$y' = f'(x)$, $f'(x)$的导数就是$f(x)$的二阶导数，记作y''，或者$f''(x)$，$\frac{d^2y}{dx^2}$。

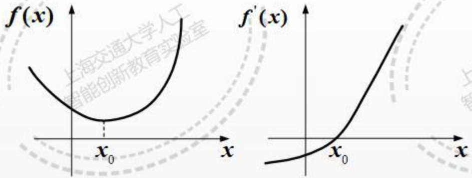
上述讨论的函数默认为是一个自变量。若函数的自变量多于一个自变量，为一多元函数$y = f(x_1, x_2, \cdots)$，研究它的变化率也是一个有意义的问题，这里就引入偏导数的概念。

在数学中，一个多变量函数的偏导数，是它关于其中一个变量的导数，而保持其他变量不变。例如：$z = x^2 + 3xy + 2y^2$,x和y的偏导数：

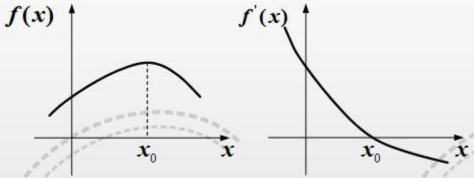
$$\frac{\partial z}{\partial x} = 2x + 3y \qquad \frac{\partial z}{\partial y} = 3x + 4y$$

导数与函数极值

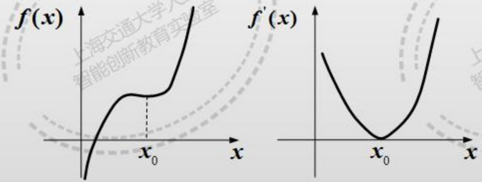
$f''(x_0) > 0$



$f''(x_0) < 0$



$f''(x_0) = 0$



梯度

方向导数

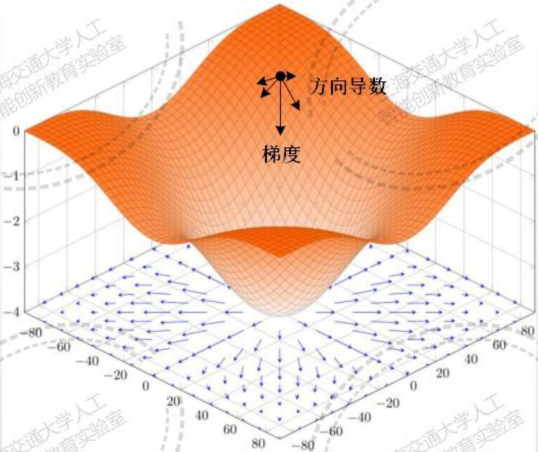
方向导数是偏导数的概念的推广，偏导数研究的是指定方向 (坐标轴方向) 的变化率。函数在某一点处的方向导数在其梯度方向上达到最大值，此最大值即梯度的范数。

梯 度

函数在某一点处的方向导数在其梯度方向上达到最大值。

这就是说，沿梯度方向，函数值增加最快。同样可知，方向导数的最小值在梯度的相反方向取得，此最小值为最大值的相反数，从而沿梯度相反方向函数值的减少最快。

梯度



比如在一个三维直角坐标系，该函数的梯度就可以表示为公式：

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right)$$

梯度的本意是一个向量（矢量），表示某一函数在该点处的方向导数沿着该方向取得最大值，即函数在该点处沿着该方向（此梯度的方向）变化最快，变化率最大（为该梯度的模）。