微积分

变化速率与累积效应

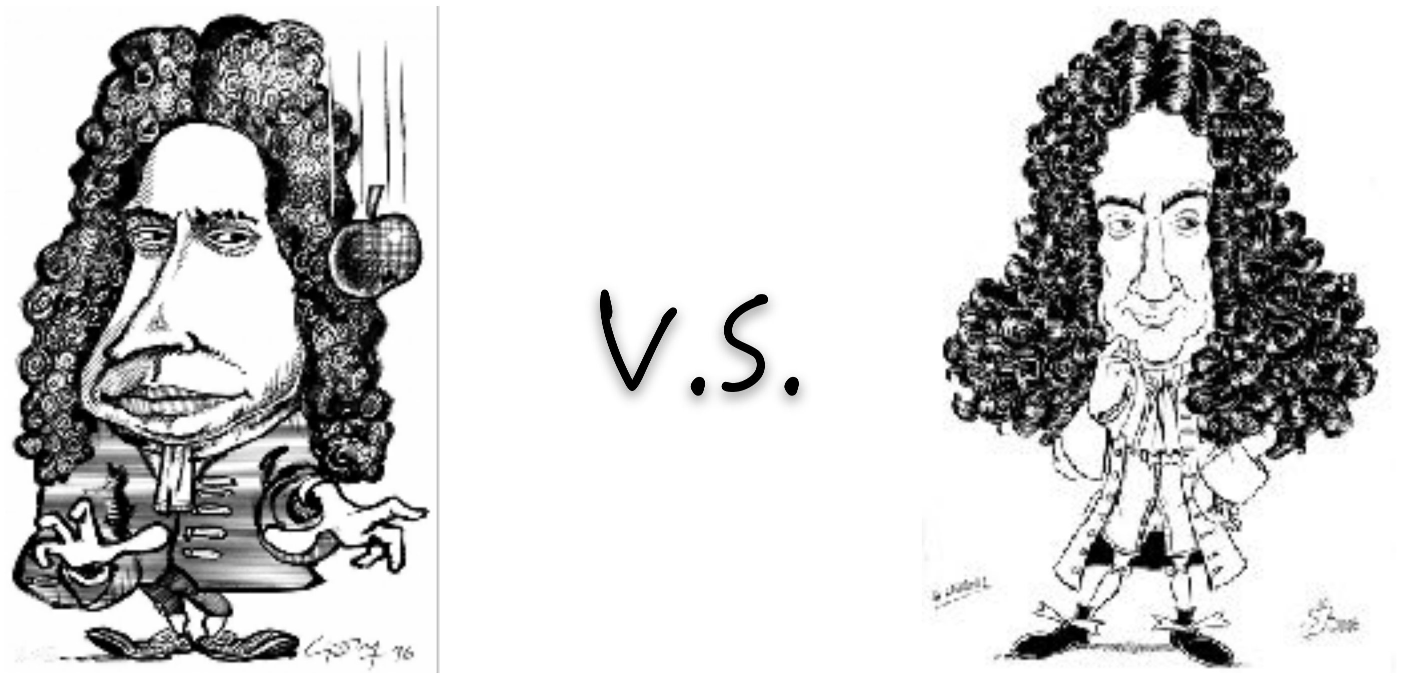
小胖

目录

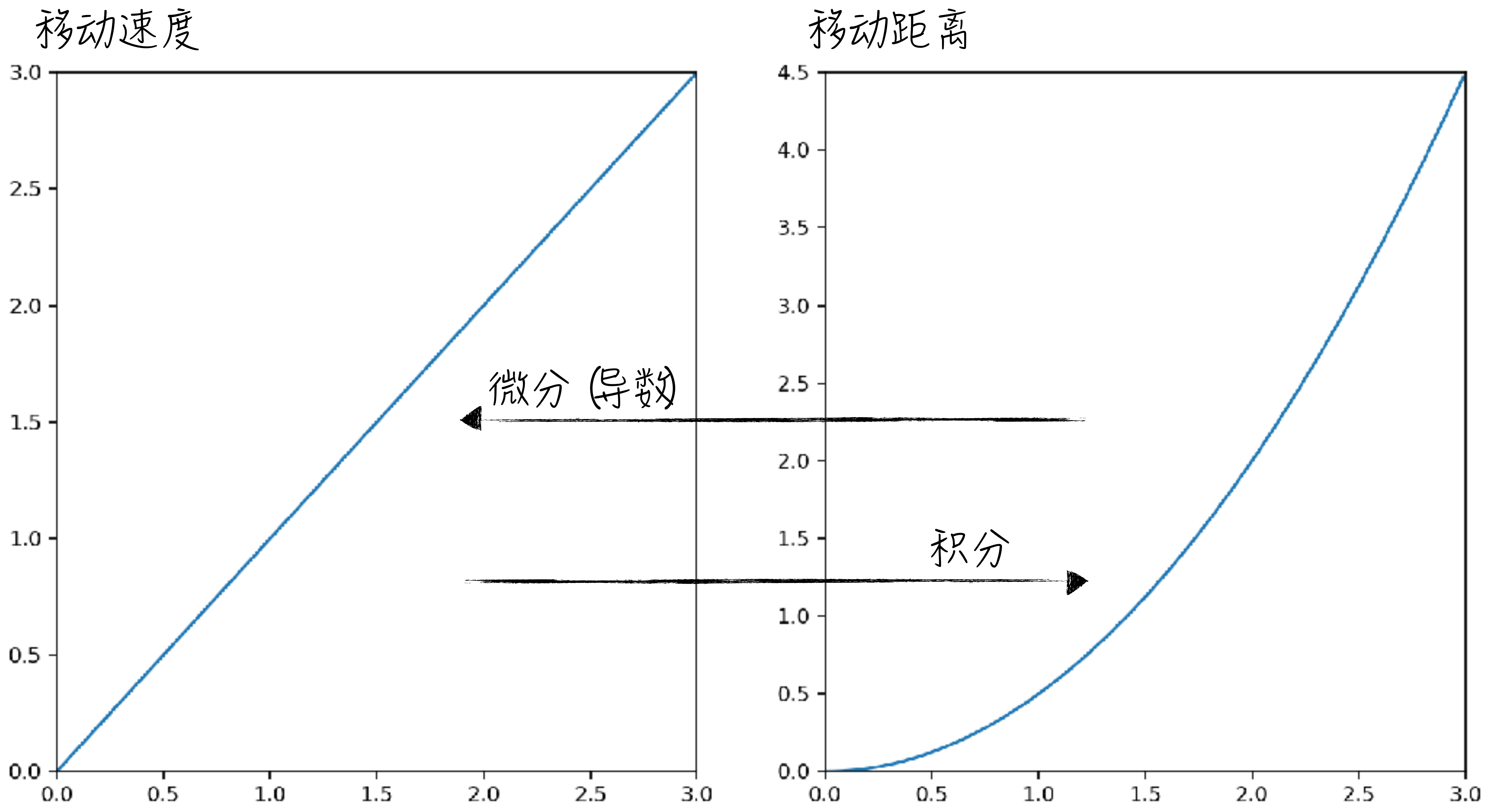
微分与积分

位置与速度

- 微分主要研究函数在局部的变化速率
- 积分主要研究函数在一段范围内的累积效果



微积分包含两个互补的方面：
微分（导数）和积分

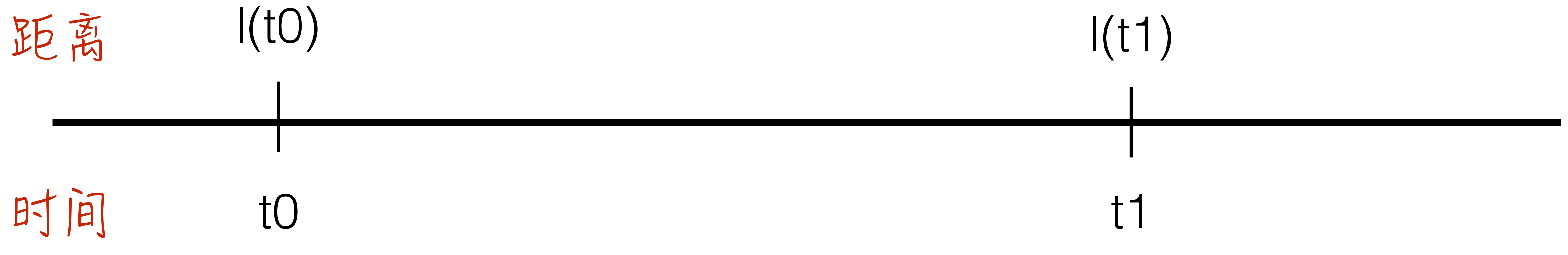


微分与积分

导数、微分

假设小明在一条直线上行走

$$\bar{v} = \frac{l(t1) - l(t0)}{t1 - t0} \longrightarrow v(t0)$$



翻译成数学语言就是，平均速度的极限就是起点的速度。记为

$$\lim_{t1->t0} \frac{l(t1) - l(t0)}{t1 - t0} = v(t0)$$

在数学上，我们称速度v(t)是距离l(t)的**导数**：$l'(t) = v(t)$

数学领域使用字母d表示极小的量（或者无穷小量）；则成立下面的公式，这个公式就是微分方程

$$\mathrm{d}l(t) = v(t)\mathrm{d}t$$

事实上，我们可以继续对速度v(t)求导数，得到加速度a(t)

a(t)也是距离l(t)的二阶导数：$a(t) = v'(t) = l''(t)$

以此类推，可以得到距离函数l(t)的n阶导数

第一位用三阶导数证明执政才能的领导人

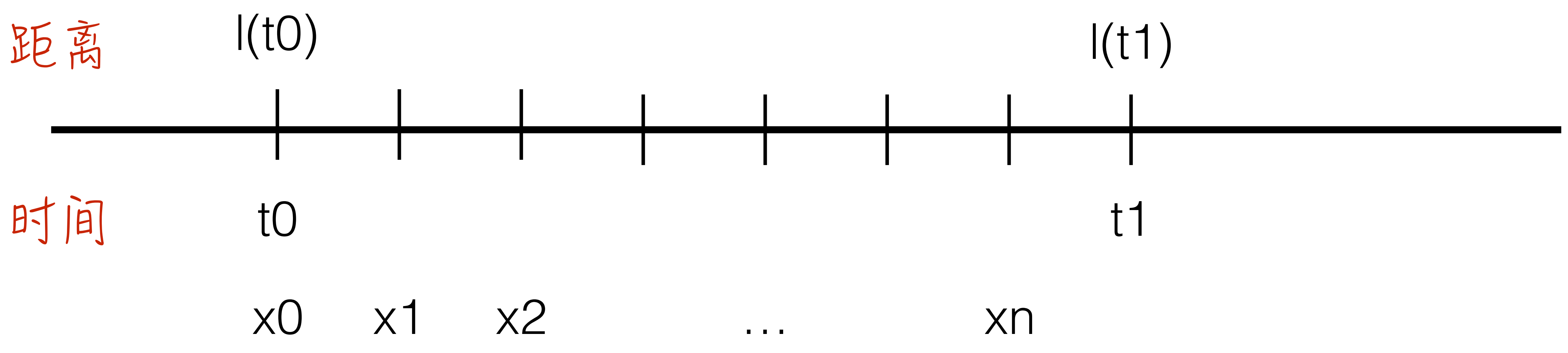
微分与积分

积分

假设小明在一条直线上行走

将t1和t0等分成n + 1段，记每一段的起点为xi

$$l(t0) + \sum_{i=0}^{n} v(x_i)\frac{(t1 - t0)}{n} \longrightarrow l(t1)$$



翻译成数学语言就是，匀速运动的累加极限就是距离的变动，记为

$$l(t1) - l(t0) = \lim_{n->\infty} \sum_{i=0}^{n} v(x_i)\frac{(t1 - t0)}{n}$$

在数学上，我们称距离l(t)是速度v(t)的**积分**：$l(t) = \int v(t)\mathrm{d}t$

根据之前的讨论：$\mathrm{d}l(t) = v(t)\mathrm{d}t$

就像加减法一样，微分和积分可以相互推导，互为逆运算

并不是所有的函数都存在微分和积分

存在微分的函数
称为**可微函数**

存在积分的函数
称为**可积函数**

数据科学领域，常用的函数都是可微和可积的

微分与积分

极限的定义

微分和积分的定义都涉及“无穷小量”这个数学概念

但“无穷小量”并不好理解，甚至曾引发了第二次数学危机

贝克莱悖论：Δx等于0吗？

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

计算过程中，Δx不等于0

在计算的最后一步，Δx等于0

为了解决这个危机，以柯西为首的数学家建立了严格的实数极限理论

- **将无穷小量理解为一个过程，而非一个确定的量**

函数f(x)在x0处的导数为f’(x0)：

- 对于任意的 $\varepsilon > 0$
- 都存在一个 $\delta > 0$，使得下面的公式成立

$$\left|\frac{f(x)-f(x_0)}{x-x_0}-f'(x_0)\right|<\varepsilon$$

$$x \in (x_0-\delta, x_0+\delta), \quad x! = x_0$$

这就是著名的 $\varepsilon - \delta$ 语言

目录

微分计算进阶

四则运算法则与链式法则

导数的计算涉及函数的极限，直接计算比较困难

常用函数的导数公式

$$(x^n)' = nx^{n-1}$$

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

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

$$(e^x)' = e^x$$

$$(lnx)' = \frac{1}{x}$$

导数的四则运算法则

$$(f \pm g)' = f' \pm g'$$

$$(fg)' = f'g + fg'$$

$$(\frac{f}{g})' = \frac{f'g - fg'}{g^2}$$

导数的链式法则

$$f(x) = h(g(x)) = h \circ g$$

$$f'(x) = (h' \circ g)g'$$

例如

$$f(x) = (x^2 + 1)^2$$

$$h(t) = t^2; g(x) = x^2 + 1$$

$$f'(x) = 2g(x) \times g'(x) = 2(x^2 + 1) \times 2x$$

微分计算进阶

偏导数

之前讨论的微分（导数）都是针对单变量函数的

但在实际当中，我们常面对的是多变量函数，也就是有多个自变量的函数

$$f(x, y) = xy + xy^2$$

对于多变量函数，需要使用偏导数来研究函数的局部变化情况

x当成变量，其他变量当成常数；
对x求导，得到函数对x的偏导数

$$\frac{\partial f}{\partial x} = y + y^2$$

y当成变量，其他变量当成常数；
对y求导，得到函数对y的偏导数

$$\frac{\partial f}{\partial y} = x + 2xy$$

在一定条件下，多元函数的微分和偏导数有如下的关系

$$\mathrm{d}f(x_1, \ldots, x_n) = \sum \frac{\partial f}{\partial x_i}\mathrm{d}x_i$$

向量 $\nabla f = (\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n})$ 被称为梯度

目录

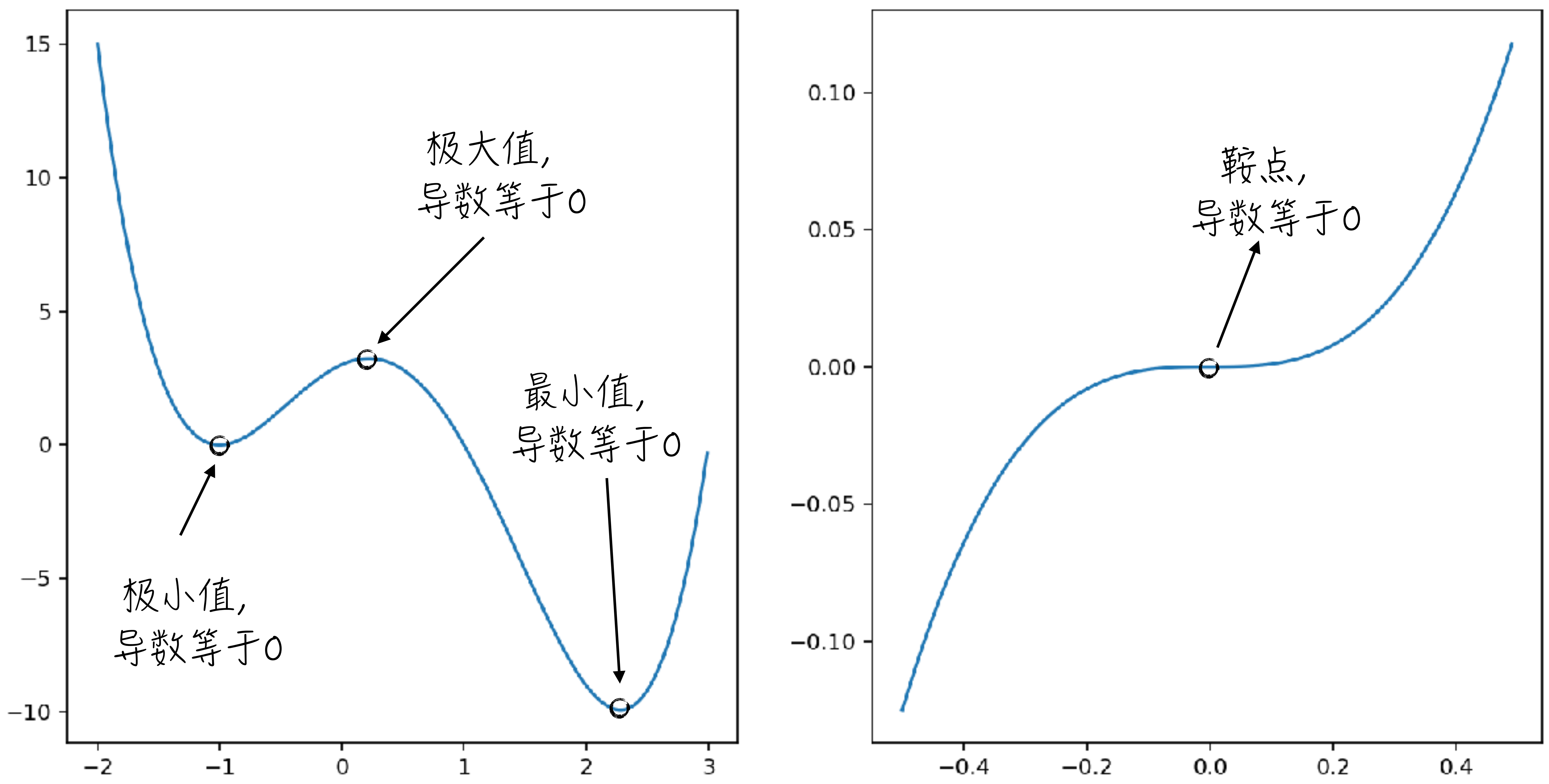
极值与最值

极值点、最值点以及鞍点

在数据科学领域，常常遇到寻找函数最值点的问题

最值点的导数值等于0，但导数值等于0并不是最值点的充要条件

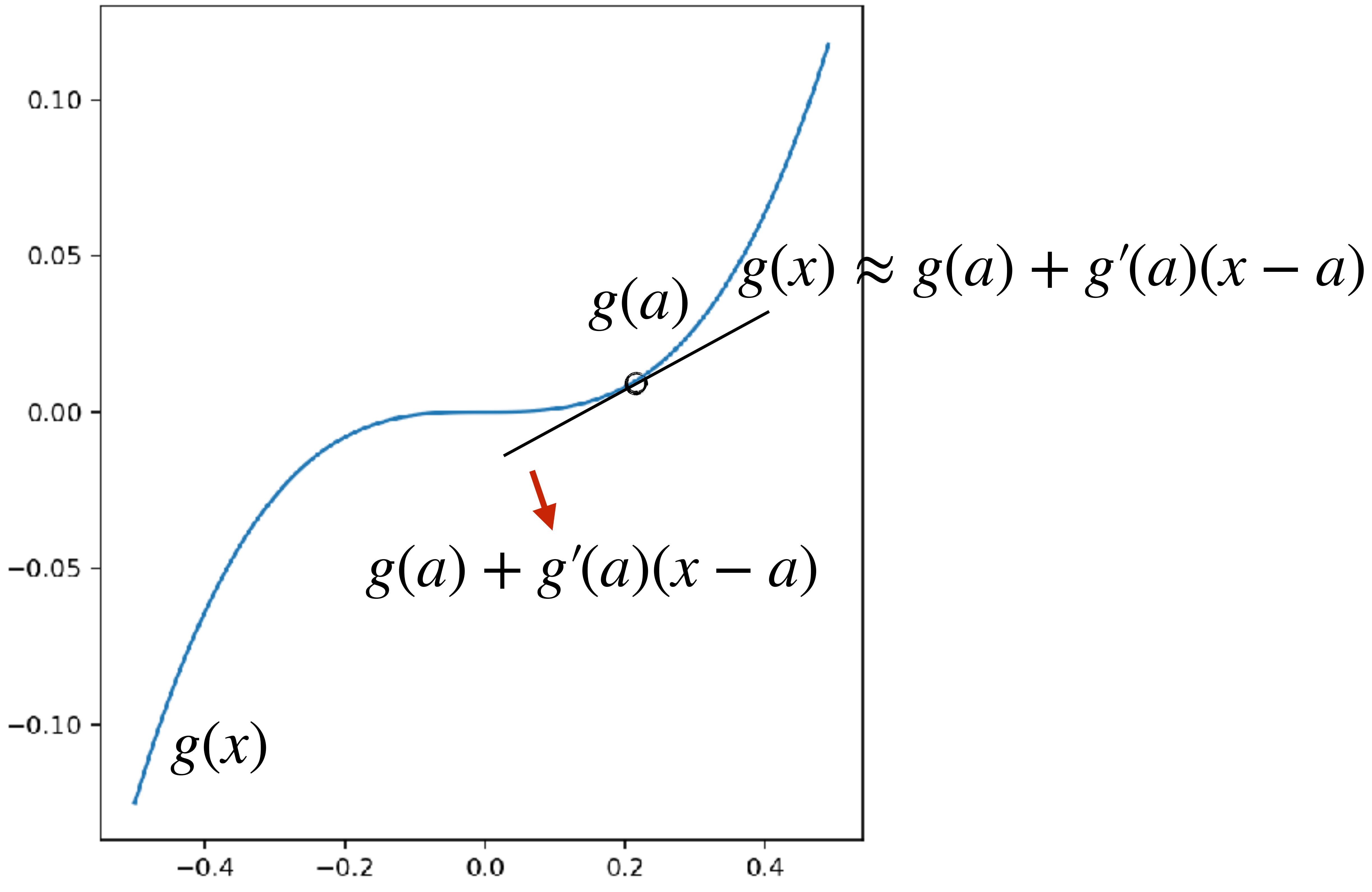
- 极值点、最值点以及鞍点的共同点是**导数值等于0**
- 实际中，常用导数值等于0这个条件来得到备选最值点



极值与最值

泰勒级数

为了更方便地研究函数的局部性质，数学上常用泰勒级数将函数局部近似为多项式的形式。



推广到多变量函数 $f(x_1, \ldots, x_n) \approx f(a_0, \ldots, a_n) + \sum \frac{\partial f(a_1, \ldots, a_n)}{x_i}(x_i - a_i)$

这个公式是梯度下降法的数学基础

THANK YOU