微积分与逼近论

邓明格

七月在线

mingge_deng@brown.edu

April 7, 2017

主要内容

1. 微分学回顾
   - 极限与导数
   - 泰勒级数
   - 牛顿法与梯度下降法
2. 积分学回顾
   - 黎曼积分
   - 勒贝格积分与概率
3. 概率论简单回顾
   - 两大基本定理
   - 参数估计

$\epsilon - \delta$ 语言

极限

对于任意的正数 $\epsilon > 0$，存在正数 δ，使得任何满足 $|x - x_0| < \delta$ 的 x，都有 $|f(x) - L| < \delta$，称函数 f 在 x_0 处的极限为 L，记为

$$\lim_{x \to x_0} f(x) = L \tag{1}$$

$\epsilon - \delta$ 语言

连续

- 通俗描述：函数 $f(x)$ 在某一点 x_0 的一个邻域上有定义，则函数 f 在 x_0 点连续当且仅当 f 在 x 趋于 x_0 时的极限等于 $f(x_0)$，i.e.，$\lim_{x \to x_0} f(x) = f(x_0)$。
- $\epsilon - \delta$ 描述：函数 $f : \mathbb{D} \subset \mathbb{R} \to \mathbb{R}$。$x_0$ 是 $\mathbb{D}$ 中一点，并且 f 在 x_0 的一个邻域上有定义。如果对任意的正实数 ϵ，都存在正实数 δ，使得对任意 $x \in \mathbb{D}$ ，只要 $|x - x_0| < \delta$，就有 $|f(x) - f(c)| < \epsilon$，那么就说 f 在 x_0 处连续。

无穷小（大）

无穷小（大）：$f \to 0$ （$f \to \infty$）

例：比较 $sin(x)$ 与 $\tan(x)$ 在 $x \to 0$ 处的极限

$$\lim_{x \to 0} \frac{\sin(x)}{\tan(x)} = \lim_{x \to 0} \cos(x) = \cos(0) = 1 \tag{2}$$

故 $sin(x)$ 与 $\tan(x)$ 在 $x \to 0$ 处的极限为同阶无穷小。

无穷小（大）阶数

无穷小（大）的阶数：f 趋于 $0(\infty)$ 的速度

o 记号

$x \to 0$, 如果 $\lim_{x\to 0} f(x) = 0$ 且 $\lim_{x\to 0} f(x)/x^n = 0$, 那么 $f(x)$ 为 n 阶以上无穷小，记为 $f(x) = \mathbf{o}(x^n)$.

O 记号

$x \to 0$, 如果 $\lim_{x\to 0} f(x) = 0$ 且 $\lim_{x\to 0} f(x)/x^n$ 存在且不为零, 那么 $f(x)$ 为 n 阶无穷小，记为 $f(x) = \mathbf{O}(x^n)$.

重要极限

- $\lim_{x \to 0} \sin(x)/x = 1$
- $\lim_{x \to \infty} x^\alpha / e^x = 0, \quad \forall \alpha > 0$
- $\lim_{x \to \infty} ln(x)/x^\alpha = 0, \quad \forall \alpha > 0$
- $\lim_{x \to \infty} (1 + 1/x)^x = e$

作业 1：证明以上极限
作业 2: 比较常见函数在零点处的无穷小阶数（泰勒级数）

导数定义及线性逼近

导数定义

如果一个函数 $f(x)$ 在 x_0 附近有定义，且存在极限，

$$L = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} \tag{3}$$

那么 $f(x)$ 在 x_0 处可导且导数 $f'(x_0) = L$.

导数线性逼近

如果存在一个实数 L 使得 $f(x)$ 满足，

$$f(x) = f(x_0) + L(x - x_0) + \mathbf{o}(x - x_0), \quad x \to x_0 \tag{4}$$

那么 $f(x)$ 在 x_0 处可导且导数 $f'(x_0) = L$.

求导法则

求导法则

- 链式法则：$\frac{d}{dx}(g \circ f) = \frac{dg}{dx}(f) \cdot \frac{df}{dx}$
- 加法法则：$\frac{d}{dx}(g + f) = \frac{dg}{dx} + \frac{df}{dx}$
- 乘法法则：$\frac{d}{dx}(g \cdot f) = \frac{dg}{dx} \cdot f + g \cdot \frac{df}{dx}$
- 除法法则：$\frac{d}{dx}(\frac{g}{f}) = \frac{\frac{dg}{dx} \cdot f - \frac{df}{dx} \cdot g}{f^2}$

作业 1：证明链式法则（导数定义）
作业 2：证明其他求导法则（链式法则）
作业 3：求 x^x 的导数 (链式法则)

高阶导数

- 函数的导数函数仍然可导，那么导数函数的导数是二阶导数，二阶导数函数的导数是三阶导数，记为

$$f^{(n)}(x) = \frac{d}{dx} f^{(n-1)}(x) = \frac{d^n}{dx^n} f(x) \tag{5}$$

- 一阶导数是对函数的线性逼近，高阶导数是对导数函数的线性逼近。
- 一阶导数告诉函数的单调性，二阶导数告诉函数的凹凸性。

二元函数的全微分与偏导数

如果 $f(x,y)$ 是一个二元函数，而且存在 L_x 和 L_y 使得:

$$f(x_0+\Delta_x, y_0+\Delta_y) = f(x_0, y_0) + L_x\Delta_x + L_y\Delta_y + \mathbf{o}(|\Delta_x| + |\Delta_y|) \tag{6}$$

那么 $f(x,y)$ 在 (x_0, y_0) 点处可微，且 L_x,L_y 分别是 f 在 x,y 方向上的的偏导数，记为

$$\frac{\partial}{\partial x} f(x_0, y_0) = L_x \quad \frac{\partial}{\partial y} f(x_0, y_0) = L_y \tag{7}$$

二元函数的二阶偏导数

如果 $f(x,y)$ 是一个二元函数，而且存在 L_x,L_y,L_{xy}，L_{x^2}，L_{y^2} 使得:

$$\begin{aligned} f(x_0+\Delta_x, y_0+\Delta_y) = f(x_0, y_0) &+ L_x\Delta_x + L_y\Delta_y \\ &+ L_{xy}\Delta_x\Delta_y + \frac{L_{x^2}}{2}\Delta_x^2 + \frac{L_{y^2}}{2}\Delta_y^2 \\ &+ \mathbf{o}(|\Delta_x|^2 + |\Delta_y|^2) \end{aligned} \tag{8}$$

那么 $f(x,y)$ 在 (x_0, y_0) 点处二阶可微，且二阶偏导数为

$$\frac{\partial^2}{\partial x^2} f(x_0, y_0) = L_{x^2} \quad \frac{\partial^2}{\partial x \partial y} f(x_0, y_0) = L_{xy} \quad \frac{\partial^2}{\partial y^2} f(x_0, y_0) = L_{y^2} \tag{9}$$

泰勒级数定义

$f(x)$ 是一个无限次可导的函数，那么在任何一点 x_0 附近可以对 $f(x)$ 做泰勒级数展开

$$f(x_0+\Delta_x) = f(x_0) + f'(x_0)\Delta_x + \frac{f''(x_0)}{2}\Delta_x^2 + \dots + \frac{f^{(n)}(x_0)}{n!}\Delta_x^n + \mathbf{o}(\Delta_x^n) \quad (10)$$

- 泰勒级数是对函数的多项式逼近 (或逐次线性逼近)。
- 泰勒级数是微分学的巅峰或精髓。
- 麦克劳林级数 $(x_0 = 0)$

泰勒级数应用

常见函数的泰勒级数

- $e^x = 1 + x + x^2/2 + \cdots + x^n/n! + o(x^n)$
- $\ln(1+x) = x - x^2/2 + x^3/3 + \cdot + (-1)^{n-1}x^n/n + o(x^n)$
- $\sin(x) = x - x^3/6 + \cdots + (-1)^n x^{2n+1}/(2n+1)! + o(x^{2n+1})$
- $\cos(x) = 1 - x^2/2 + x^4/24 + \cdots + (-1)^n x^{2n}/(2n)! + o(x^{2n})$

作业：证明欧拉公式 $e^{i\theta} = \cos(\theta) + i\sin(\theta)$ (泰勒级数)

洛必达法则

f, g 为无穷阶可导的函数，且 $f(x_0) = g(x_0) = 0$, $g'(x_0) \neq 0$, 那么

$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = \lim_{x \to x_0} \frac{f'(x)}{g'(x)} \tag{11}$$

泰勒级数应用

求解简单微分方程

求满足如下条件的解析函数 $f(x)$:

$$f''(x) = -f(x), \quad f(0) = 0, \quad f'(0) = 1 \tag{12}$$

提示: $f(x) = \sum_{0}^{\infty} f^{(n)}(0)x^n/n!, \quad f''(x) = \sum_{0}^{\infty} f^{(n+2)}(0)x^n/n!$

泰勒级数应用

微分方程数值求解

基本思路：用差分方程来逼近微分方程，用邻点函数的值来表示当前的导数

- 向前差分：$f'(x_0) = \dfrac{f(x_0+h) - f(x_0)}{h} + \mathbf{O}(h)$
- 向后差分：$f'(x_0) = \dfrac{f(x_0) - f(x_0-h)}{h} + \mathbf{O}(h)$
- 中心差分：$f'(x_0) = \dfrac{f(x_0+h) - f(x_0-h)}{2h} + \mathbf{O}(h^2)$
- 二阶中心差分 $f''(x_0) = \dfrac{f(x_0+h) - 2f(x_0) + f(x_0-h)}{h^2} + \mathbf{O}(h^2)$

微分方程 $Lf(x) = g(x)$，其中 L 为微分算子，最终可以写成 $Af = g$ 的矩阵方程形式来求解 $f(x)$ 在格点上的值。

极值点条件

- 全局极小值: 如果对于任何 x, 都有 $f(x_*) \leq f(x)$, 那么 x_* 就是全局极小值点.
- 局部极小值: 如果存在一个正数 δ 使得, 对于任何满足 $|x - x_*| < \delta$ 的 x, 都有 $f(x_*) \leq f(x)$, 那么 x_* 就是局部极小值点.(方圆 δ 内的极小值点).
- 不论是全局极小值还是局部极小值一定满足一阶导数/梯度为零, $f' = 0$ 或者 $\nabla f = 0$.

梯度下降法回顾

如果实值函数 $F(\mathbf{x})$ 在点 $\mathbf{x}'$ 处可微且有定义，那么函数 $F(\mathbf{x})$ 在 $\mathbf{x}'$ 点沿着梯度相反的方向 $-\nabla F(\mathbf{x}')$ 下降最快。

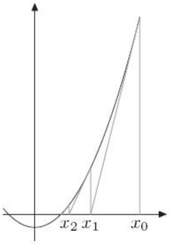
考虑到这一点，我们可以从函数 $F(\mathbf{x})$ 的局部极小值的初始估计 $\mathbf{x}_0$ 出发，并考虑如下序列 $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots$，使得

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \gamma \nabla F(\mathbf{x}_n) \tag{13}$$

因此可得到 $F(\mathbf{x}_0) \geq F(\mathbf{x}_1) \geq F(\mathbf{x}_2) \geq \ldots$，那么 $(\mathbf{x}_n)$ 将收敛到期望的极值。

牛顿法理解思路一：函数零点问题

求 x_0 附近函数 $f(x)$ 的零点。



$$0 = (x - x_0)f'(x_0) + f(x0) \tag{14}$$

新求得的点的命名为 x_1，通常 x_1 会比 x_0 更接近方程 $f(x) = 0$ 的解，可以利用 x_1 开始下一轮迭代

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \tag{15}$$

对于极值问题，相当于求解其一阶导数函数的零点问题。

牛顿法理解思路二：函数的二次逼近

在初始点 x0 处，二阶泰勒级数逼近：

$$f(x_0+\Delta_x) = f(x_0) + f'(x_0)\Delta_x + \frac{f''(x_0)}{2}\Delta_x^2 + \mathbf{o}(\Delta_x^2) \approx g(\Delta_x) \tag{16}$$

关于 Δ_x 的二次函数 $g(\Delta_x)$ 的极值点为 $-\dfrac{f'(x_0)}{f''(x_0)}$，故逼近点

$x_1 = x_0 - \dfrac{f'(x_0)}{f''(x_0)}$ 更接近极值点，重复此步骤得到序列 x_n

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)} \tag{17}$$

$\lim_{n\to\infty} x_n$ 将收敛于 x_0 附近的极值点。
对于多元函数，更新序列为

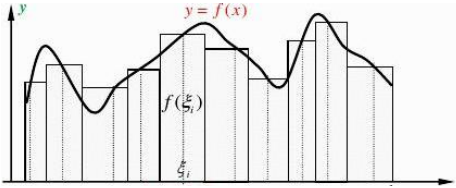
$$x_{n+1} = x_n - (\mathcal{H}f(x_n))^{-1} \cdot \nabla f(x_n) \tag{18}$$

其中 $\mathcal{H}f$ 为 Hessian 矩阵。

梯度下降法与牛顿法的比较

- 都是对连续函数的局部逼近，梯度下降法是一阶逼近，而牛顿法为二阶逼近。
- 由于为局部逼近，得到的解均为局部极值，故 x_0 的选取极其重要。
- 梯度下降法为一阶收敛，牛顿法为二阶收敛，但需要求解 Hessian 矩阵。

一维曲线面积：无穷求和



黎曼积分的定义

$f(x)$ 为开区间 (a, b) 上的一个连续函数，对于任意一个对定义域区间的分割 T,

$$\int_a^b f(x)dx = \lim_{||T|| \to 0} \sum_{i=1}^{n} f(\xi_i)\Delta x_i, \quad \Delta x_i = x_i - x_{i-1}, \quad x_{i-1} \le \xi_i \le x_i \quad (19)$$

1. 黎曼积分的几何意义为非负函数图像下方的图像面积。
2. 黎曼积分的值与分割，介点集的取法无关。

黎曼积分可积的充要条件

黎曼可积的充要条件

$$\int_a^b f(x)dx = \lim_{||T||\to 0}\sum_{i=1}^n M_i\Delta x_i = \lim_{||T||\to 0}\sum_{i=1}^n m_i\Delta x_i \tag{20}$$

其中 $M_i = \sup\{f(x) : x_{i-1} \le x_i\}, \quad m_i = \inf\{f(x) : x_{i-1} \le x_i\}$

或者等价于：$\forall\epsilon > 0$, 存在划分 T, 使得 $\sum_i = 1^n\omega_i\Delta x_i \le \epsilon$, 其中 $\omega_i = M_i - m_i$.

黎曼不可积函数：δ 函数（狄拉克函数），上积分为 1，下积分为 0.

牛顿 -莱布尼茨公式

如果 $f(x)$ 是定义在闭区间 $[a, b]$ 上的可微函数，那么就有

$$\int_a^b f'(x)dx = f(b) - f(a) \tag{21}$$

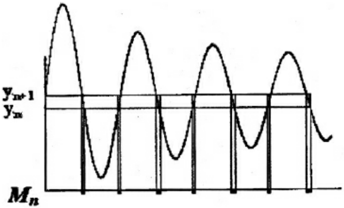
不定积分表示为

$$\int f'(t)dt = f(x) + C \tag{22}$$

牛顿 -莱布尼茨公式展示了微分与积分的基本关系：在一定程度上微分与积分互为逆运算.

勒贝格积分

基本思想：不从分割定义域入手，从分割值域入手。



勒贝格积分

$$(L)\int_a^b f(x)dx = \lim_{\delta\to 0}\sum_{i=1}^{n}\xi_i m(M_i) \tag{23}$$

其中 $M_i = \{x : y_{i-1} \le f(x) \le y_i\}, \quad y_{i-1} \le \xi \le y_i$.
用 $m(M_i)$ 表示集合 M_i 的“长度”或度量。

勒贝格积分与概率

勒贝格自己的比喻:

1. 勒贝格积分：欠人一笔钱，现在要还，按照钞票的面值分类，清点每类面额总值，再相加。
2. 黎曼积分：不按面额分类，按从口袋掏出的先后次序相加。

测度概念和概率概念的联系:

测度————概率
积分————期望

大数定理与中心极限定理

大数定理

X 是随机变量，μ 是 X 的期望，σ 是 X 的方差. $\{X_k\}_{k=1}^{\infty}$ 为 i.i.d. 随机变量, 那么 $\bar{X}_n = \frac{\sum_{k=1}^{n} X_k}{n}$ 依概率收敛于 μ, 也就是说对于任何 $\epsilon > 0$ 有

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| > \epsilon) = 0 \tag{24}$$

中心极限定理

X 是随机变量,μ 是 X 的期望，σ 是 X 的方差. $\{X_k\}_{k=1}^{\infty}$ 为 i.i.d. 随机变量,$\bar{X}_n = \frac{\sum_{k=1}^{n} X_k}{n}$, 那么

$$Z_n = \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \tag{25}$$

依分布收敛于标准正态分布 $N(0,1)$.

参数估计问题

已知一个随机变量 X 的分布函数 $f_\theta(X)$, 其中 $\theta = (\theta_1, ..., \theta_k)$ 为未知参数，利用已有样本 $X_1, ..., X_n$ 对参数 θ 或者 θ 的函数 $g(\theta)$ 作出估计。

1. 点估计：用样本的一个函数 $T(X_1, ..., X_n)$ 估计 $g(\theta)$

2. 区间估计：用一个置信区间去估计 $g(\theta)$

矩估计

基本原理：大数定理，对于任何 i.i.d 变量 X，当 $n \to \infty$，$\frac{1}{n}\sum_i^n X_i$ 收敛于 $E(X)$，同理其 k 阶矩也满足大数定理，$\frac{1}{n}\sum_i^n X_i^k$ 收敛于 $E(X^k)$，可以构造 k 组方程求救。

两点分布的矩估计

X 服从两点分布取值为 $\{-1, 1\}$，$P(-1) = 1 - \theta$，$P(1) = \theta$，用样本 $X_1, ...X_n$ 估计参数 θ.

能用低阶矩不用高阶矩!!!
作业：正态分布的参数矩估计。

极大似然估计

给定随机变量的分布与未知参数，利用观测到的样本计算似然函数，选择最大化似然函数的参数作为参数估计量.

似然函数

假设 $X = (X_1, ..., X_n$ 样本的观测值. 那么整个样本的似然函数就是

$$L(\theta) = \prod_{i=1}^{n} f_\theta(X_i) \tag{26}$$

这是一个关于 θ 的函数，选取使得 $L(\theta)$ 最大化的 $(\hat{\theta})$ 作为 θ 估计量.

作业：正态分布的参数极大似然估计。

点估计的评判

- 相合性 (consistency): 当样本数量趋于无穷时，估计量收敛于参数真实值.
- 无偏性 (bias): 对于有限的样本，估计量所符合的分布之期望等于参数真实值.
- 有效性 (efficiency): 估计值所满足的分布方差越小越好.
- 渐进正态性 (asymptotic normality): 当样本趋于无穷时，去中心化去量纲化的估计量符合标准正态分布.

作业：正态分布的参数矩估计以及极大似然估计是否无偏?

谢谢大家！！