

高等数学中梯度教学案例研究

1 问题的提出

在新时代,新工科的背景下,数学思维、逻辑思维在各个领域有广泛应用,是这些领域创新发展的强有力的工具。高等数学课程是工科学生必修课程。提升高等数学课程的教学质量,对新工科的人才培养质量有着十分重要的作用。数学工作者积极探索高等数学课程的教学改革方法。高等数学课程是一门理论课,基础性强,应用领域广泛,将高等数学的知识点与实际问题相结合做成的教学案例,对激发学生学习兴趣,更好掌握本课程教学知识点,培养学生用高等数学知识解决实际问题的思维方式都起到了十分重要的作用。关于高等数学教学案例的研究已经引起数学教育工作者的广泛关注^[1-4]。

高等数学教学案例从内容方面看,通用型案例是所有专业的学生都可以使用的,比如速度加速度问题,面积体积问题,重心形心问题,最值应用问题等。而专业型案例通常是与学生专业相关性较强的案例。比如气象相关专业的温度梯度可以刻画温度的扩散方向,计算机等工科专业的梯度下降法在机器学习中的应用;管理等专业的随机梯度估计在系统管理中的应用等。针对方向导数和梯度的教学研究,数学教育工作者做了很多工作^[5-11],但在梯度的教学案例方面成果很少,只有两篇^[11-12]。

温度梯度是指空间中温度分布的变化率,通俗讲是单位长度或单位面积内的温度变化量。例如,一块热的金属板从一端加热,另一端为常温,那么板中心温度与板两端温度的差值就是温度梯度。在大气科学的气候研究中,山区气温的垂直梯度是山区气候考察和研究的重要问题之一。

在机器学习的核心内容就是把数据喂给一个人工设计的模型,然后让模型自动的“学习”,从而优化模型自身的各种参数,最终使得在某一组参数下该模型能够最佳的匹配该学习任务。那么这个“学习”的过程就是机器学习算法的关键。梯度下降法就是实现该“学习”过程的一种最常见的方式,尤其是在深度学习(神经网络)模型中,BP反向传播方法的核心就是对每层的权重参数不断使用梯度下降来进行优化。梯度下降法是一种常见的一阶优化算法,是求解无约束优化问题的最基本也是最简单的方法之一。我们考虑连续可微函数 $f(x)$,那么怎么找到 $f(x)$ 图像某段的局部极小点,就成为解决优化问题的核心。

通过蒙特卡洛仿真估计的随机系统中目标函数关于参数的梯度称为随机梯度估计。随机活动网络是系统管理中的一个系统,项目管理人员采用随机活动网络表示项目中计划完成的各项生产活动的先后顺序和逻辑关系。那么,如何选取合适的估计方法对随机活动网络进行随机梯度估计成为一个重点问题。

本文给出了三个梯度应用案例,分别是温度梯度的应用,梯度下降法的应用和随机梯度估计法的应用,针对问题进行了数值模拟和分析。

2 预备知识

2.1 温度梯度

温度梯度分为垂直温度梯度和水平温度梯度,其计算公式为: $\frac{\partial T}{\partial x}$,其中 $\frac{\partial T}{\partial x}$ 表示函

数 $T(x)$ 的导数,即温度场梯度。在传热学中,不同的温度梯度会对热传导过程产生不同的影响。例如,在一个稳态热传导过程中,传热速率正比于温度梯度。当温度梯度较大时,热传导速率也会变大。

2.1.1 图示理解

如图 1 所示，方向 n 为法向方向， Δn 为两等温面的法向距离， t 为温度，两等温面的温差 Δt 与其法向距离 Δn 的比值的极限，称为温度梯度，用 $gradt$ 表示，温度梯度的大小可

记为 $gradt = \lim_{\Delta n \rightarrow 0} \frac{\Delta t}{\Delta n} = \frac{\partial t}{\partial n}$ 。式中 $gradt$ 为温度梯度， $\frac{\partial t}{\partial n}$ 为等温面法向方向的温度变化率。

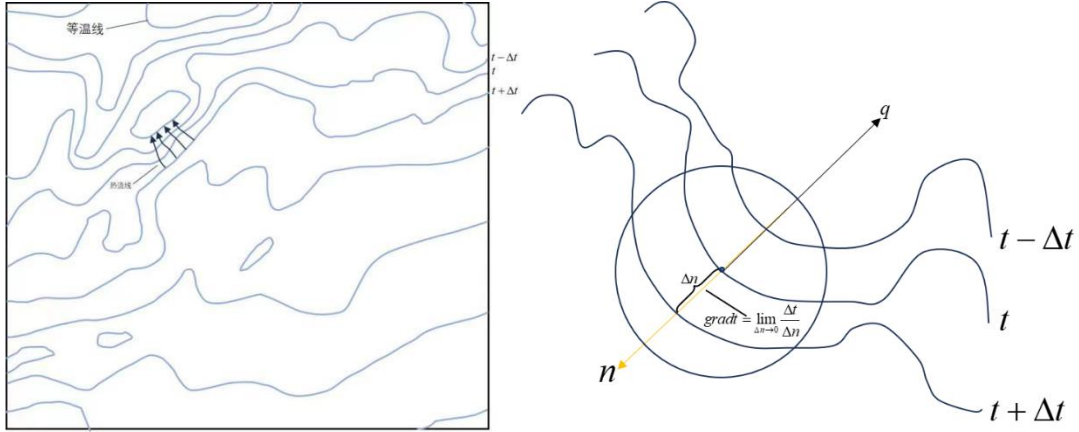


图 1 等温面、温度梯度与热流线示意图

2.1.2 计算公式

一般来说，我们可以通过温度场的梯度来计算温度梯度。温度场是指物体内部各点的温度分布情况，通常可以用一个 T 函数来表示。温度梯度是在等温面法线方向 n 上单位长度的温度增量，它是一个矢量，指向温度增大的方向，可表示为：

$$GradT = \lim_{\Delta n \rightarrow 0} \left(\frac{\Delta T}{\Delta n} \right) \vec{n} = \frac{\partial T}{\partial n} \vec{n} \quad (1)$$

式 (1) 中， $\frac{\partial T}{\partial n}$ 指对 T 函数求关于 n 的偏导数， \vec{n} 为法向方向单位矢量。注意，在本文中， $gradt$ 表示温度梯度的大小，而 $GradT$ 表示温度梯度这个矢量。

2.2 梯度下降法

梯度下降法，是一种常见的一阶优化算法，是求解无约束优化问题的最基本也是最简单的方法之一。梯度下降法是机器学习中一种常用到的算法，但其本身不是机器学习算法，而是一种求解的最优化算法。主要解决求最小值问题，其基本思想在于不断地逼近最优点，每一步的优化方向就是梯度的方向。

2.2.1 基本思路

解决问题的基本思路就是，构造出一个序列 $\{x_0, x_1, \dots\}$ ，使其中的元素满足如下条件：

$$f(x_{t+1}) < f(x_t), t = 0, 1, 2, \dots \quad (2)$$

只要我们能不断地找到序列的下一个点，那么我们最终就能找到局部最小点。图 2 展示了寻找最小点的过程。

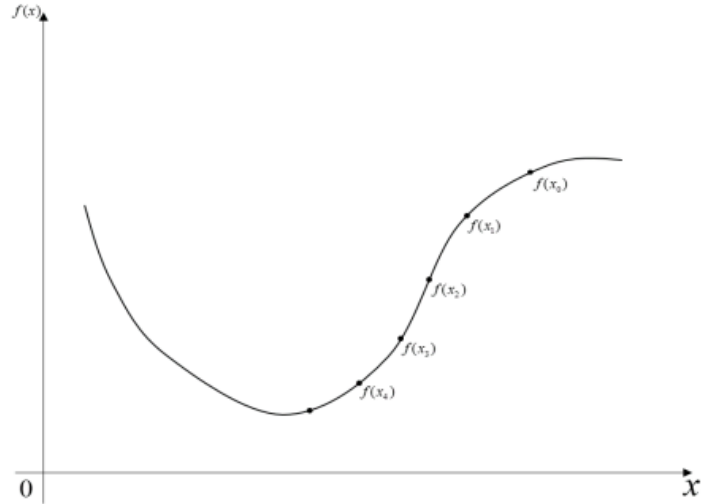


图 2 寻找局部最小点的过程

2.2.2 基本原理

怎么保证每次找到的下一个点 x_{t+1} ，一定满足 $x_{t+1} < x_t$ 呢？假设我们当前函数 $f(x)$ 如图 2 所示的形状，现在我们随机取得初始点 x_1 ，对于一元函数来说，函数值只会随 x 的变化而变，那么我们就可以认为 x_{t+1} 是上一个时刻的点沿某一方向走一个小步长 Δx 得到的。那么现在只需确定方向即可。

对于一元函数来说， x 只存在两个方向，即 $\Delta x > 0$ 的正方向和 $\Delta x < 0$ 的负方向，考虑泰勒展式：

$$f(x + \Delta x) = f(x) + \Delta x \nabla f(x) \quad (3)$$

展式的左边是当前 x 的下一个点位，要使 $f(x + \Delta x) < f(x)$ ，就要 $\Delta x \nabla f(x) < 0$ 。令 $\Delta x = -\alpha \nabla f(x)$ ，($\alpha > 0$) 其中步长 α 是一个较小的正数，从而就有

$$\Delta x \nabla f(x) = -\alpha (\nabla f(x))^2 < 0 \quad (4)$$

将公式 (4) 带入公式 (3)，有

$$f(x + \Delta x) = f(x - \alpha \nabla f(x)) < f(x) \quad (5)$$

这样，我们就得到了 $x_{t+1} = x_t - \alpha \nabla f(x)$ ，也即，沿负梯度方向的一小步。

2.2.3 常用的梯度下降法

机器学习的本质就是“喂”给模型数据，让模型不断地去学习，而这个学习的过程就是利用梯度下降法不断去优化的过程，目前最为常见的神经网络便是利用梯度的反向传播，反复更新模型参数直至收敛，从而达到优化模型的目的。

常见的梯度下降法包括随机梯度下降法 (SGD)、批梯度下降法、Momentum 梯度下降法、Nesterov Momentum 梯度下降法、AdaGrad 梯度下降法、RMSprop 梯度下降法、Adam 梯度下降法。每种算法都有自己的优缺点，对应解决不同的问题。

2.3 随机梯度估计及随机活动网络

通过蒙特卡洛仿真估计的随机系统中目标函数关于参数的梯度称为随机梯度估计。随机活动网络是系统管理中的一个系统，项目管理人员采用随机活动网络表示项目中计划完成的各项生产活动的先后顺序和逻辑关系。

2.3.1 随机梯度估计方法

随机梯度估计方法分为间接的梯度估计方法和直接的梯度估计方法。间接估计方法分为有限差分法和同时扰动法,直接的梯度估计方法分为扰动分析法(常用无穷小扰动分析法)、似然比法,弱导数法和广义似然比法。

2.3.2 随机活动网络

随机活动网络由有向非循环图表示,生产活动时间为随机变量 X_i , X_i 相互独立,其分布为 $F_i, i=1,2,\dots,N$ 。 p 为所有路径的集合, $p^* \in p$ 为最优路径,最优路径的时间可以表示为: $Y(X) = \sum_{j \in p^*} X_j$ 。问题的目标是估计 $dE[Y]/d\theta$, 其中, θ 为 X_i 的分布参数。

将各种估计方法应用在随机活动网络上,得到各估计量的数值(平均值±标准误差),再与精确的均值作比较即可知道哪种估计量为合适的估计量。

3 应用实例

以下我们将介绍前两节的理论和方法如何应用于三个实例。

3.1 温度梯度应用实例

3.1.1 温度扩散方向

已知 $T(x) = 3x^2 - 5x + 2$, 求点(1,0)处的温度扩散方向。

求温度扩散方向, 只需求出 $\frac{\partial T}{\partial x}$ 即可。

由 $T(x) = 3x^2 - 5x + 2$, 有 $\frac{\partial T}{\partial x} = 6x - 5$ 。

在点(1,0)处 $\frac{\partial T}{\partial x} = 1$ 。

即点(1,0)处的温度扩散方向与 x 正轴夹角为 45° 。

3.1.2 温度梯度影响因素

如表 1 所示, 该表是根据数据计算出的不同季节、晴雨天气下的温度梯度, 试分析各季节和晴雨天气对温度梯度的影响。

表 1 各季晴雨天的温度梯度 ($^\circ\text{C}/100$ 米)

	春	夏	秋	冬	平均值
晴	0.632	0.731	0.595	0.474	0.6080
雨	0.459	0.506	0.467	0.278	0.4275
平均值	0.5455	0.6185	0.5310	0.3760	

从表 1 可见, 在该地区, 全年的温度梯度以夏季为最大, 春秋次之, 冬季最小。晴天的温度梯度比雨天的大。不论在哪个季节, 晴天的温度梯度总比雨天的要大; 不论是晴天还是雨天, 夏季的温度梯度总是最大的, 而冬季的温度梯度总是最小的。其中, 还可以看出冬

季雨天气温垂直梯度最小, 仅 $0.278^{\circ}\text{C}/100$ 米, 而夏季晴天气温垂直梯度最大, 达到了 $0.731^{\circ}\text{C}/100$ 米。

因此, 可推断出, 夏季和晴天会导致该山区的温度梯度增大, 而冬季和雨天会导致该山区的温度梯度减小。

3.2 梯度下降法应用实例

给定初始点 $X^{(1)} = (0,0)^T$, $\min f(x) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$, 利用梯度下降法求出 $X^{(2)}$ 。

$$\text{目标函数梯度 } \nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 1 + 4x_1 + 2x_2 \\ -1 + 2x_1 + 2x_2 \end{bmatrix}。$$

令搜索方向 $d^{(1)} = -\nabla f(X^{(1)}) = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, 从 $X^{(1)}$ 出发, 沿 $d^{(1)}$ 方向寻优。

令步长变量为 λ , 最优步长 λ_1 , 有 $X^{(1)} + \lambda d^{(1)} = \begin{bmatrix} -\lambda \\ \lambda \end{bmatrix}$,

则 $f(X^{(1)} + \lambda d^{(1)}) = \lambda^2 - 2\lambda = \varphi_1(\lambda)$, 令 $\varphi_1(\lambda) = 0$, 得 $\lambda_1 = 1$,

得到点 $X^{(2)} = X^{(1)} + \lambda_1 d^{(1)} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ 。

3.3 随机梯度估计法应用实例

3.3.1 随机梯度估计在随机活动网络中的应用

如图 3 所示, 考虑包含 5 个节点、6 条边的随机活动网络。节点 5 为任务终止点, 所有路径的集合为 $p = \{(1 \rightarrow 2 \rightarrow 4 \rightarrow 5), (1 \rightarrow 3 \rightarrow 4 \rightarrow 5), (1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5)\}$ 。生产活动时间 X_i 相互独立, 服从均值为 i 的指数分布。现在需要选取合适的估计方法对其进行随机梯度估计。

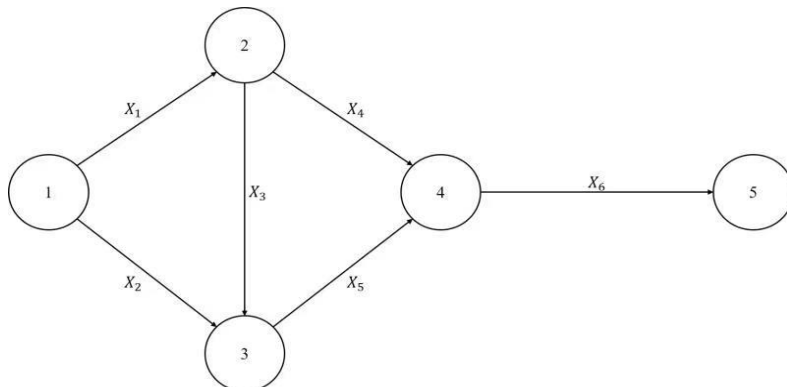


图 3 随机活动网络

生产活动时间 X_i 服从均值为 i 的指数分布。令 $\theta = \theta_1 = 1$ 为 X_1 中的分布参数, p^* 为使生产活动时间最短的最优路径。根据研究, 在具有 5 个节点、6 条边的随机活动网络中, 当 $\theta = \theta_1 = 1$ 时, 无穷小扰动分析估计量可表示为 $(\frac{dY}{d\theta}|_{\theta=1})_{IPA} = X_1 \mathbf{1}\{(1 \rightarrow 2) \in P^*\}$, 其中, $1 \rightarrow 2$ 表示路径, $\mathbf{1}\{\bullet\}$ 是示性函数, 括号内式子成立则函数为 1, 反之则为 0; 似然比估计量为 $(\frac{dY}{d\theta}|_{\theta=1})_{LR} = Y(X)(X_1 - 1)$; 弱导数估计量 $(\frac{dY}{d\theta}|_{\theta=1})_{WD} = Y(X_1^{(+)}, \dots, X_6) - Y(X)$ 。其中, $X_i^{(+)}$ 服从形状参数为 2、尺度参数为 1 的爱尔朗分布。各估计量表达式如公式 (6) 所示。

$$\begin{aligned} (\frac{dY}{d\theta}|_{\theta=1})_{IPA} &= X_1 \mathbf{1}\{(1 \rightarrow 2) \in P^*\} \\ (\frac{dY}{d\theta}|_{\theta=1})_{LR} &= Y(X)(X_1 - 1) \\ (\frac{dY}{d\theta}|_{\theta=1})_{WD} &= Y(X_1^{(+)}, \dots, X_6) - Y(X) \end{aligned} \quad (6)$$

实验在指数分布的设定下, 比较 3 种估计量和单边前向有限差分估计量 $FD(c)$, 当 c 为 0.001 时, $FD(0.001)$ 的均值最为精确。实验可得到各估计量的数值(平均值 \pm 标准误差)。

如表 2 所示, 当 c 为 0.1 时, $FD(0.1)$ 的均值较小; 当 c 为 0.01 时, $FD(0.01)$ 的方差较大。基于 158 次独立的仿真实验, 在指数分布下, $FD(0.001)$ 的梯度估计量的均值为 0.591。这表明, 无穷小扰动分析估计量, 似然比估计量和弱导数估计量具有无偏性。

表 2 随机活动网络最优路径时间的梯度估计量(平均值 \pm 标准误差)

	IPA	LR	WD	FD(0.1)	FD(0.01)
指数分布	0.591 \pm 0.0006	0.591 \pm 0.0119	0.591 \pm 0.0018	0.441 \pm 0.0120	0.594 \pm 0.1167

建议在该问题中, 选择无穷小扰动分析估计量, 似然比估计量和弱导数估计量进行随机梯度估计来实现总工期最短的目标。

3.3.2 问题引申

该例缺少了广义似然比法, 如果用广义似然比估计量会怎么样?

实际上, 广义似然比法是在似然比法的基础上产生的, 因为似然比法不适用样本表现带有结构参数的问题, 广义似然比法便在似然比法的基础上解决了这一问题。由于在该例中, 样本表现 $Y(X)$ 不包含结构参数, 广义似然比估计量与似然比估计量是相同的。

因此, 如果在该例中使用广义似然比估计量进行计算, 最后的结果也会跟似然比估计的

结果相同。如果样本表现 $Y(X)$ 包含了结构参数,那似然比法将不再适用。

3.4 结果分析

本文给出了梯度的三个应用实例,分别关于温度梯度,梯度下降法和随机梯度估计法。结果表明,在不同的领域,均可以利用梯度来解决实际问题。下一步的工作可以寻找梯度在其他领域中的应用,使得结果更加丰富,进一步完善本课题。

参考文献

- [1] 董昌州,张玉平. 微课程背景下“梯度概念”课堂教学的创新设计[J]. 科技视界, 2016(03):22+25. DOI:10.19694/j.cnki.issn2095-2457.2016.03.013.
- [2] 刘新国,姚增善. 梯度与等高线的教学体会[J]. 高等数学研究, 2006(02):50-51.
- [3] 崔丽鸿,伍毅. 梯度教学设计一例[J]. 中州大学学报, 2000(04):94-95.
- [4] 邱高. 电势梯度教学的新尝试[J]. 纺织高校基础科学学报, 1998(01):86-89.
- [5] 张京良. 基于能力导向的工科高等数学教学——以方向导数与梯度为例[J]. 高教学刊, 2023, 9(02):142-145. DOI:10.19980/j.cn23-1593/G4.2023.02.035.
- [6] 林丽芳,曾月迪,陈梅香. “互联网+”背景下《高等数学》智慧课堂教学模式设计——以方向导数与梯度为例[J]. 创新创业理论研究与实践, 2020, 3(24):127-129.
- [7] 赵金玲,王萍,胡筱,等. 从方向导数向量到梯度的直观讲析教学设计[J]. 高师理科学刊, 2019, 39(05):47-52.
- [8] 詹环,陈平,张雪飞. 梯度概念的案例式教学研究[J]. 信息系统工程, 2019(05):162-163.
- [9] 赵文才,刘洪霞,包云霞. 基于混合式教学的“梯度”教学案例研究[J]. 大学教育, 2018(03):61-63+70.