

反向传播算法中负梯度在参数更新中的作用：方向与步长的详析

ChenZhuoWen

2025 年 3 月 14 日

摘要

本报告深入探讨了反向传播算法中负梯度在参数更新过程中的核心作用。我们从数学定义出发，详细分析了梯度和方向导数的关系，并阐明了为何负梯度不仅指示了更新方向，还通过其模长决定了更新步长。通过三维图示，我们直观展示了损失函数曲面及负梯度的作用。希望本报告能帮助读者更深入地理解深度学习中梯度下降法的基本原理。

1 引言

反向传播算法是训练神经网络的基础，其核心在于通过计算损失函数对参数的梯度来指导参数更新。梯度下降法的更新公式为

$$\theta_{\text{new}} = \theta_{\text{old}} - \eta \nabla L(\theta), \quad (1)$$

其中 L 为损失函数， θ 为参数向量， η 为学习率。由于梯度 ∇L 指向损失增加最快的方向，所以取其负值 $-\nabla L$ 则能使损失下降最快。本文将详细论述这一原理，并结合图示说明其直观意义。

2 理论背景

2.1 梯度与方向导数

设多元函数 $f(\theta_1, \theta_2, \dots, \theta_n)$ 在点 $\theta = (\theta_1, \theta_2, \dots, \theta_n)$ 处的梯度定义为

$$\nabla f(\theta) = \left(\frac{\partial f}{\partial \theta_1}, \frac{\partial f}{\partial \theta_2}, \dots, \frac{\partial f}{\partial \theta_n} \right). \quad (2)$$

梯度向量不仅反映了函数在各方向上的变化率，其方向正是函数增值最快的方向。对于任一单位向量 \mathbf{v} ，函数沿该方向的方向导数为

$$D_{\mathbf{v}}f(\theta) = \nabla f(\theta) \cdot \mathbf{v} = \|\nabla f(\theta)\| \cos \theta, \quad (3)$$

其中 θ 为 $\nabla f(\theta)$ 与 \mathbf{v} 的夹角。

2.2 负梯度与参数更新

在梯度下降法中，参数更新公式为

$$\Delta \theta = -\eta \nabla L(\theta). \quad (4)$$

这里：

- **更新方向：**由于梯度 ∇L 指向损失函数上升最快的方向，因此负梯度 $-\nabla L$ 指向损失下降最快的方向；

- **更新步长：**梯度的模长 $\|\nabla L\|$ 表示函数变化的速率，乘以学习率 η 后决定了步长大小。

因此，负梯度在一次更新中既给出了参数改变的方向，又决定了改变的幅度。

3 三维图示与直观分析

为便于直观理解，我们以简单的二元函数为例：

$$L(\theta_1, \theta_2) = \theta_1^2 + \theta_2^2, \quad (5)$$

其梯度为

$$\nabla L = (2\theta_1, 2\theta_2). \quad (6)$$

在此例中，梯度的方向与参数 (θ_1, θ_2) 的方向一致，而负梯度则指向原点（损失值较低的区域）。

图 ?? 展示了函数 $L(\theta_1, \theta_2)$ 的三维曲面；图 ?? 中，我们在点 $(1, 1)$ 处标出了负梯度，并用箭头展示了参数更新的方向和相对步长。

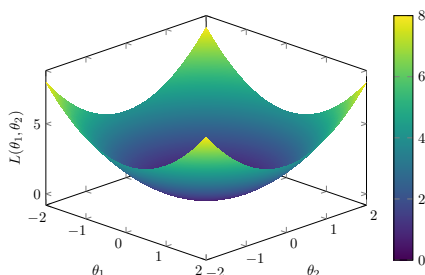


图 1: 损失函数 $L(\theta_1, \theta_2) = \theta_1^2 + \theta_2^2$ 的三维曲面图示。

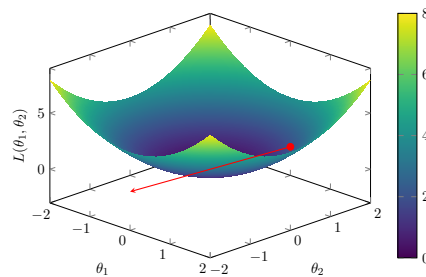


图 2: 在点 $(1, 1)$ 处，红色箭头表示负梯度 $-\nabla L$ ，指示了参数更新的方向与步长（由梯度模长决定）。

4 实验与讨论

在实际神经网络训练中，损失函数往往具有更为复杂的形状（可能是非凸的），但梯度下降法仍然依靠负梯度来引导参数向局部或全局最优解更新。实际情况中，我们可能遇到以下现象：

- 当损失函数变化剧烈时，梯度模长较大，更新步长也大，可能会引起震荡；
- 学习率的选取对平衡更新步长起到关键作用；
- 为提高稳定性和收敛速度，常采用动量、AdaGrad、Adam等改进的优化算法，这些方法在原始负梯度的基础上对更新进行调节。

上述讨论表明，理解负梯度在更新中的双重作用对于设计高效的优化算法至关重要。

5 结论

本文从数学定义和直观图示两方面详细阐述了反向传播算法中负梯度的作用。负梯度不仅为参数更新提供了明确的方向（即

损失下降最快的方向)，还通过梯度模长决定了更新步长。这一原理构成了梯度下降法及其改进算法的理论基础，对深度学习模型的训练起着决定性作用。

6 未来工作

未来的研究可以进一步探讨复杂损失函数下负梯度信息的利用，特别是在自适应优化和非凸优化问题中的应用。同时，通过更高维度的可视化方法，帮助我们更直观地理解深度网络中参数空间的结构和优化路径。