

深度学习中 GD 算法的学习过程研究

□ 李杰骏 罗耀祖 王宇坤 罗庆佳 唐灏

摘要: 梯度下降算法 (Gradient Descent) 是一个用于求函数最小值的优化算法, 其原理是通过向负梯度最陡的方向不断地调整参数值。GD 算法是梯度下降算法的一种优化, 本文从数学原理的角度总结分析 GD 算法的参数学习过程, 总结梯度下降计算对比传统函数求解的参数计算方法的优劣。

关键词: 机器学习; 神经网络; 深度学习; 梯度下降; SGD

从 2006 年开始深度学习已经开始迅速地称为越来越热门的研究方向, 成为多种领域中最前沿的人工智能技术, 例如: 实体识别、图像识别、语言识别和语言翻译等等^[1]。^[2] 使用深度学习进行医疗检测, 通过低成本传感器和物联网分析处理医疗检测数据, 达到健康评估、损伤定位、损伤类型识别等目的。其中梯度下降算法是深度学习的其中的重要一环, 通过该算法实现了高效的大量未知参数的计算。

一、损失函数

本来通过手写字识别的任务中, 假设 x 代表训练样本的输入, 若 x 是一个 28×28 的灰度图, 那么 x 则使用一个 784 维度的向量表示。把期望的输出用 $y = y(x)$ 表示, 其中 y 是一个 10 维度的向量。例如, 一个手写数字样本为 4, 那么

$$y(x) = (0, 0, 0, 1, 0, 0, 0, 0, 0, 0)^T$$

深度学习的学习过程也就是参数寻找的过程, 这里通过定义损失函数求需要学习的参数:

$$C(w, b) = \frac{1}{2n} \sum_x \|y(x) - a\|^2 \quad (1)$$

其中 w 代表所有的权值, b 代表偏差值, n 代表训练样本数量, a 是输入样本真实值的向量。 $C(w, b)$ 称为损失函数, 从公式 (1) 看出 $C(w, b)$ 是非负的。当 $C(w, b) \approx 0$ 时候损失最小, 因此需要通过一个优化算法找到最优的 w 和 b 值。而这里使用的是一种梯度下降的算法。

二、SGD 算法的参数学习过程

由于上述损失函数的需要未知参数过多, 通过解方程

的方式求解并不实际, 因此替代的是通过一种优化算法来近似地求方程的最优解。这里的优化算法过程可以形象地看成是一个“球”在一定形状的“山谷”中向下慢慢滚动, 当“球”滚到“山谷”的最低端, 参数即是最优的。而该“山谷”的坡度可以用基于各个未知参数的偏导数代表, 例如 w 基于 C 的偏导数则是 $\frac{\partial C}{\partial w}$, w 的变化量是 Δw , C 的变化量是 ΔC 。那么有:

$$\Delta C \approx \frac{\partial C}{\partial w_1} \Delta w_1 + \frac{\partial C}{\partial w_2} \Delta w_2 + \dots + \frac{\partial C}{\partial w_n} \Delta w_n + \frac{\partial C}{\partial b} \Delta b \quad (2)$$

若把公式 (2) 看作是两个向量的点积, 把偏导的系数向量看作 ∇C , 把参数 $\Delta w_1 \dots \Delta w_n$ w 和 b 组成向量 Δv 。最终把公式 (2) 简化为: $\Delta C \approx \nabla C \cdot \Delta v$

为了保证 C 的变化是向下因此需要保证 ΔC 为负值, 则 $\Delta v = -\eta \nabla C$, 因为 $(\nabla C)^2$ 必为非负, 其中 η 称为学习速率。而对于所有参数 v 梯度下降的过程用公式代表为:

$$v \rightarrow v' = v - \eta \nabla C \quad (3)$$

通过多次迭代, C 越趋近最小同时参数 v 越趋近最优。

三、结论

在机器学习上, 梯度下降算法用于更新神经网络模型的参数。由于 SGD 算法每次只需要一个小的样本子集, 算法所需要的内存相对批量梯度下降算法更小, 而且拟合速度比批量训练快并且可以更新模型参数更加频繁。但由于更新频繁也往往导致 SGD 算法往损失最小值拟合的过程产生较多噪声, 导致拟合方向出错。

参考文献

- [1] Schmidhuber, Jürgen. “Deep learning in neural networks: An overview.” *Neural networks* 61(2015): 85-117.
- [2] Zhao, Rui, et al. “Deep learning and its applications to machine health monitoring.” *Mechanical Systems and Signal Processing* 115 (2019): 213-237.

(作者单位: 江门职业技术学院)

(上接第 205 页)

实际特点。高职院校平面设计教学中, 教师应该重点培养学生的个性化思维, 让学生在创作平面设计作品的时候能够拥有自己的风格, 不可千篇一律。只有学生平面设计作品具有自己的风格才能将个性充分展示出来, 并带给人一定的视觉冲击性。教师在平面设计教学中要激发学生的想象思维, 让学生大胆创作, 大胆设计。高职院校教师队伍建设能提升平面设计专业教学质量, 高职院校可以引进一些具有丰富经验的教师, 并在校内开展定期的培训讲座。想要提升实践教学水平就要向企业引进实践型人才, 还可以将校内教师送到企业进行实践学习, 完善的教师队伍建设可以为高职院校平面设计专业奠定基础, 只有教师自身素养得到了提升, 才能够为平面设计教学的顺利进行提供辅助。双师型教师队伍可以满足目前教育改革需求也能够满足企业对于新型平面设计人才的需求^[3]。

四、结语

综上所述, 高职院校平面设计类课程传统教学方式无法激

发学生的学习积极性和学习兴趣, 并阻碍了高职院校平面设计类人才的培养。

对此, 高职院校只有使用新型的教学模式和教学理念, 改善传统教学方式的束缚才能更好的培养出具有专业素养的平面设计类人才。师生间需要相互交流, 共同进步, 引导学生大胆实践, 进而培养出高素质的平面设计类人才。

参考文献

- [1] 孙湛, 杨兴胜. 浅谈高职院校平面设计类专业摄影课程教学改革 [J]. 包装世界, 2018(8): 64-69.
- [2] 李会芬. 高职院校平面设计类课程教学改革探索 [J]. 电脑知识与技术, 2019, 15(8): 114-117.
- [3] 胡晓. 基于激进建构主义的高职课程多元评价策略——以计算机平面设计类课程为例 [J]. 电脑迷, 2018, 97(6): 186-187.

(作者单位: 江西财经职业学院)

作者简介: 于昊 (1990~), 男, 硕士研究生, 助教, 研究方向为计算机应用技术。