

先谈谈对目前《动手学深度学习》中前四章涉及到的关键词进行梳理和解释，有助于我对整体学习做总结。

机器学习(machine learning, ML): 一类强大的可以从经验中学习的技术。

数据(data): 每个数据集由一个个样本(example, sample)组成, 样本有时也叫做数据点(data point)或者数据实例(data instance), 通常每个样本由一组称为特征(features, 或协变量(covariates))的属性组成。

过拟合(Overfitting): 指的是一个模型在训练数据上表现得非常好, 但在新数据(测试数据或验证数据)上的表现却很差。可能是因为学习了噪声。

批量大小(batch size): 指每次用于计算损失和梯度更新的样本数量。它决定了在一次参数更新中使用的训练数据量。

学习率(learning rate): 用于控制在每次更新过程中, 模型参数变化的幅度。学习率决定了你在梯度方向上移动的速度。

断言(assert): 用于验证训练过程中某些条件是否满足, 通常用于调试和验证代码的正确性。

泛化(generalization): 指的是机器学习模型在未见数据上的预测能力。一个具有良好泛化能力的模型在训练集上表现良好的同时, 也能够测试集或其他未见样本上保持较高的性能。

权重衰减(weight decay): 一种正则化技术, 通过在损失函数中引入权重惩罚项来限制模型的复杂度, 从而提高其泛化能力。它通常表现在添加权重的平方和到损失函数中。

分布偏移(distribution shift): 指训练集数据和测试集数据的分布有所不同。在深度学习和机器学习中, 这种偏移可能会导致模型在测试集上的性能下降, 尽管在训练集上的表现良好。

SGD 中的随机: 在每个 epoch 中随机打乱整个数据集, 在训练集中随机选择样本构建小批量, 保证每一个 batch 中的数据是随机取出来的, 增加模型的泛化能力。

数值上溢和下溢: 在计算结果超出了计算机可以表示的最大数值范围或小于计算机可以表达的最小正数(接近于 0 的数值)。减去 max 为解决办法。

监督学习(supervised learning): 擅长在“给定输入特征”的情况下预测标签, 许多重要的任务可以清晰地描述为“在给定一组特定的可用数据的情况下, 估计未知事物的概率”, 常见的监督学习有回归问题、分类问题、标记问题、搜索问题、推荐系统和序列学习。

回归(regression): 当标签取任意数值时, 我们称之为回归问题, 此时的目标是生成一个模型, 使它的预测非常接近实际标签值(任何有关“有多少”的问题很可能就是回归问题)

分类(classification): “哪一个”的问题, 希望模型能够预测样本属于哪个类别, 有二项分类、多项分类、层次分类和多标签分类等等

无监督学习(unsupervised learning): 数据不含有“目标”的机器学习问题, 常见的无监督学习有聚类问题-没标签来分类、主成分分析问题-少参数捕捉、因果关系和概率图模型问题-根本原因 和 生成对抗性网络-合成数据方法。

离线学习(offline learning): 预先获取大量数据, 然后启动模型, 不再与环境交互, 包含监督学习和无监督学习。

强化学习(reinforcement learning): 一类明确考虑与环境交互的问题。强化学习智能体选择的"动作"受策略控制,即一个从环境观察映射到行动的功能;强化学习者必须处理学分分配(credit assignment)问题:决定哪些行为是值得奖励的,哪些行为是需要惩罚的;强化学习智能体必须不断地做出选择:是应该利用当前最好的策略,还是探索新的策略空间(放弃一些短期回报来换取知识)。

全连接层(fully-connected layer): 每个输入和每个输出都相连。

广播机制(Broadcasting): 会自动调整某些操作的数组形状,使它们彼此兼容并进行计算,而不需要显式地调整数组的尺寸。

Softmax 函数作用: 能够将未规范化的预测变换为非负数(exp)并且总和为1(归一化),同时让模型保持可导的性质。