随机优化的算法及其理论

王小玉 *

摘要

机器学习是数据科学和人工智能的重要基石. 近些年, 随着互联网, 信息技术和计算机技术等领域的快速发展, 机器学习迎来了新的热潮, 引起了社会各界的广泛关注. 随机优化算法作为求解机器学习问题一类重要的优化算法, 发挥了重要的作用. 在本文中, 我们主要从算法及其理论的角度, 系统地研究了随机优化算法求解机器学习问题.

首先我们系统介绍了求解机器学习问题最经典的随机梯度下降算法. 从算法的计算代价与收敛速率等方面, 详细介绍了随机梯度下降算法的优势以及难点. 我们还分析了随机梯度下降算法在强凸和光滑的情形下的理论性质. 并进一步介绍了它的收敛性与随机梯度的方差、步长之间的关系. 紧接着, 我们从算法是一阶, 还是二阶 (或者更高阶) 两个大的方面, 详细介绍了改进随机梯度下降算法的一些重要的变体算法.

接着我们考虑了机器学习中常见的正则化经验风险极小化问题, 其中正则项是非光滑的. 针对此类问题, 我们提出了一个随机邻近点拟牛顿算法框架. 我们引入了新的判定收敛性的准则, 不同于常用的广义投影算子. 在非凸且 L-光滑情况下, 我们分析了该算法框架的非渐进收敛性和复杂度. 在 CP-PL 不等式的条件下, 我们证明了算法的全局线性收敛性. 理论结果表明, 相比广义投影算子, 新算子可以得到更好的误差界. 此外, 我们提出了一种修正的自调节对称秩 1 算法 (MSSR1), 更新拟牛顿矩阵, 使得子问题很容易求解. 最后, 我们将 MSSR1 算法应用到提出的框架中, 设计数值实验验证了算法的有效性.

然后, 我们研究了随机信赖域算法求解机器学习问题. 我们的动机是通过改进信赖域半径, 来提升随机信赖域算法 (STORM) 的数值表现. 我们考虑了信赖域半径直接和当前的随机模型相关. 特别地, 我们给出一种信赖域半径更新策略 (称为 STRME 算法), 其中信赖域半径与当前计算的随机梯度线性相关. 我们分析了 STRME 算法在非凸、一般凸以及强凸情况下的复杂度. 该结果与现有的随机信赖域算法的结果是一致的. 最重要的是, 我们设计数值实验测试了正则化的逻辑回归问题和简单的神经网络问题. 数值结果表明了: 对非凸问题, 我们提出的 STRME 算法, 比 STORM 算法表现更好.

最后, 我们探究了随机梯度下降算法在有带宽的 (区间) 步长下收敛性. 在这种步长假设下, 我们拓宽了随机梯度下降算法的最优的收敛条件, 并为迭代初期大步长提供了一些理论保障. 当步长所在区间的上下界函数是同阶时, 我们的理论分析显式地给出了 SGD 算法的误差界上界, 其结果涵盖了大多数常见的步长策略. 至今, 关于随机梯度下降算法的理论分析大多数都是基于单调的步长. 而我们提出的有带宽的步长策略允许步长有一定的扰动, 可以是非单调的. 据我们所知, 非单调的步长对非凸问题, 特别是深度学习的一些问题是非常有益的. 此外, 我们还设计了一种非单调的步长. 数值实验表明我们提出的步长对非凸问题有明显的优势, 也说明基于有带宽的步长来设计非单调步长是非常有潜力的.

关键词: 机器学习, 随机梯度下降算法, 随机邻近点拟牛顿算法, 随机信赖域算法, 有带宽的步长

*wxy@lsec.cc.ac.cn,