摘要

近年来以深度学习为代表的机器学习算法在许多领域取得了极大成功，这与大数据时代涌现出来的诸多大规模标注数据集高度相关。可是很多数据集并不完美，数据甚至标签中都可能含有一些噪声或离群点，显著影响学习效果，因此学习算法的噪声鲁棒性至关重要。我们尝试从集成学习/随机森林、模型假设/损失函数、离群点检测/自步学习，以及深度学习模型训练等多个角度研究机器学习算法的噪声鲁棒性，给出一些近期结果并讨论进一步的工作设想。