Mathematical Characterizations of Accuracy Loss in Differentially Private Machine Learning Models

Avni Jain
Mathematics, Jesus and Mary College, New Delhi, India

Abstract

Differential Privacy is a theoretical framework for ensuring data privacy when conducting statistical analysis. It was introduced to quantify privacy in statistical analyses. However, as we progress and machine learning takes over, it is essential to assess our progress in privacy as well. Initially, DP algorithms focused on sanitising simple statistics, but in recent years, it has made its way to machine learning. There have been several advancements in the basic differential privacy models over the years to combat with changes. While Rényi Differential Privacy (RDP) has been widely adopted due to its composability properties, Bayesian Differential has emerged as an alternative that adapts noise based on posterior distributions. In this study, I analyse the mathematical properties of accuracy loss under these two DP frameworks and explore their implications for differentially private machine learning models.

I theoretically analyse the mathematical bounds for accuracy tests and compare how RDP and Bayesian DP affect model performance. To support this analysis, I conduct empirical experiments on differentially private logistic regression models and evaluate accuracy loss across different privacy budgets. To verify the theorem, I implement and test the bounds on MATLAB, providing a visual and intuitive understanding. This research provides insights into the privacy-utility trade-offs of these DP mechanisms & a look into the better algorithm supported by theoretical bounds.

Keywords

Bayesian Differential Privacy; Differential Privacy; Privacy-Preserving Machine Learning; Rényi Differential Privacy.