World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:18, No:12, 2024

Optimizing Privacy, Accuracy and Calibration in Deep Learning Models

Authors : Rizwan Rizwan

Abstract : Differentially private ({DP}) training preserves the data privacy but often leads to slower convergence and lower accuracy, along with notable mis-calibration compared to non-private training. Analyzing {DP} training through a continuous-time approach with the neural tangent kernel ({NTK}). The {NTK} helps characterize per sample {(PS)} gradient clipping and the incorporation of noise during {DP} training across arbitrary network architectures as well as loss functions. Our analysis reveals that noise addition impacts privacy risk exclusively, leaving convergence and calibration unaffected. In contrast, {PS} gradient clipping (flat styles, layerwise styles) influences convergence as well as calibration but not privacy risk. Models with a small clipping norm generally achieve optimal accuracy but exhibit poor calibration, making them less reliable. Conversely, {DP} models that are trained with a large clipping norm maintain the similar accuracy and same privacy guarantee, yet they demonstrate notably improved calibration.

Keywords : deep learning, convergence, differential privacy, calibration
Conference Title : ICM 2024 : International Conference on Mathematics
Conference Location : Quebec City, Canada
Conference Dates : December 23-24, 2024