Proceedings of the 8th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'22)
Prague, Czech Republic - July 28- 30, 2022
DOI: 10.11159/mvml22.003

How to Provably Generate Differentially-Private Synthetic Data

Gerhard Wunder
Freie Universität Berlin, Germany

Abstract

In this talk we will explore the generation of synthetic data using variational auto encoders under privacy constraints. Synthetic data is seen as a major tool to meet the privacy concerns in, e.g., health applications. The relevant frameworks and information-theoretic metrics, such as various flavors of differential privacy, will be discussed in the context of privacy preserving data generation. Specifically, membership inference attack on synthetic data will be investigated as a hypothesis test. Gaussian DP will be used to interpret the guarantees of privacy preserving ML in this setting. The privacy/utility trade-off of generative models with and without privacy guarantees will be theoretically analyzed and evaluated empirically. New method based on the inherent mechanisms of variational autoencoders is proposed and analysed. We will also comment on operational implications for standard tools in establishing privacy.