Dhruvil Doshi

STUDENT, MASTER OF SCIENCE IN INFORMATION TECHNOLOGY

A comparative study of secure privacy-preserving machine learning algorithms

Machine learning (ML) is increasingly being adopted in a wide variety of application domains. Usually, a well-performing ML model relies on a large volume of training data and high-powered computational resources. Such a need for and the use of huge volumes of data raise serious privacy concerns because of the potential risks of leakage of highly privacy-sensitive information; further, the evolving regulatory environments that increasingly restrict access to and use of privacy-sensitive data add significant challenges to fully benefiting from the power of ML for data-driven applications.

There are three major techniques through which privacy in ML can be achieved: Homomorphic encryption (HE), Federated Learning (FL), and Split Learning (SL). HE is a public key cryptographic scheme. HE can perform inference on encrypted data, so the model owner never sees the client's private data and therefore cannot leak it or misuse it. HE is computationally expensive and restricted to certain kinds of calculations. FL is a collaborative machine learning method with decentralized data and multiple client devices. During the FL process, each client trains a model on their dataset and then each client sends a model to the server, where a model is aggregated to one global model and then again distributed over clients. SL is a distributed and private deep learning technique that can be used to train deep neural networks over multiple data sources while mitigating the need to share raw labeled data directly. SL provides higher accuracy, improved security, and less computation bandwidth compared to FL. This research will provide an insight on the trade-off between performance and security for HE, SL, and FL over various ML and DL algorithms. This research will unfold many applications that have sensitive data like healthcare, and finance. It will allow an organization to use and share data between business segments and jurisdictions without violating data privacy constraints.

Research Advisor: Dr. Baidya Saha, Assistant Professor