Generative Adversarial Privacy: A Data-driven Approach

Lalitha Sankar, Associate Professor, Arizona State University

Abstract:
Privacy is the problem of ensuring limited leakage of information about sensitive features while sharing information (utility) about non-private features to legitimate data users. Even as differential privacy has emerged as a strong desideratum for privacy, there is also an equally strong need for context-aware utility-guaranteeing approaches in many data sharing settings. To this end, we present Generative Adversarial Privacy and Fairness (GAPF), a data-driven framework for learning private and fair representations of large-scale datasets. GAPF leverages recent advances in generative adversarial networks (GANs) to allow a data holder to learn "universal" data representations that decouple a set of sensitive attributes from the rest of the dataset. Under GAPF, finding the optimal privacy/fairness mechanism is formulated as a constrained minimax game
between a private/fair encoder and an adversary. We show that for appropriately chosen adversarial loss functions, GAPF provides privacy guarantees against information-theoretic adversaries and enforces demographic parity. We also evaluate the performance of GAPF on the GENKI face dataset and the Human Activity Recognition (HAR) dataset.

In the second half of the talk, we seek to understand if it is meaningful to use finite capacity adversarial models, defined as those with access to finite samples and one hidden layer neural network, to design privacy/fairness mechanisms that restrict learning sensitive features when publishing a given dataset. We present probabilistic bounds on the discrepancy in the risk performance of such a finite capacity adversary relative to an infinite capacity adversary for the squared and log-losses, where an infinite-capacity adversary is one with full statistical knowledge and expressiveness capabilities. Our bounds quantify both the generalization error resulting from limited samples and the function approximation limits resulting from finite expressiveness. We illustrate our results for both scalar and multi-dimensional Gaussian mixture models.
This work is done jointly with Mario Diaz (ASU/CiMAT), C. Huang (ASU), and P. Kairouz (Google).

Biography
Lalitha Sankar is an Associate Professor in the School of Electrical, Computer, and Energy Engineering at Arizona State University. Prior to this, she was an Associate Research Scholar at Princeton University. Sankar was a recipient of a three year Science and Technology Teaching Postdoctoral Fellowship from the Council on Science and Technology at Princeton University. Prior to her doctoral studies, she was a Senior Member of Technical Staff at AT&T Shannon Laboratories. She received the B.Tech degree from the Indian Institute of Technology, Bombay, the M.S. degree from the University of Maryland, and the Ph.D degree from Rutgers University. Her research interests include applying information and learning theoretic methods to a variety of problems including privacy and cyber-security. She received the NSF CAREER award in 2014. She received the IEEE Globecom 2011 Best Paper Award for her work on privacy

of side information in multi-user data systems. For her doctoral work, she received the 2007-2008 Electrical Engineering Academic Achievement Award from Rutgers University.