Carnegie Mellon University

Can we Generalize and Distribute Private Representation Learning?

Sheikh Shams Azam[1], Taejin Kim[2], Seyyedali Hosseinalipour[1],

Carlee Joe-Wong[2], Saurabh Bagchi[1], Christopher Brinton[1]

[1]Purdue University, [2]Carnegie Mellon University

25th International Conference on Artificial Intelligence and Statistics (AISTATS), 2022

Outline

- Motivation
- Background: Private Representation Learning
- Our Methodology: EIGAN and Distributed-EIGAN
- Theoretical and Experimental Results
- Summary

Motivation

Why privacy in Machine Learning?

- Machine Learning is a data intensive learning process.
- Building datasets requires aggregating data from various sources including individual users; e.g., photo sharing on Instagram is used for weakly labeled learning, text from google keyboard is used for next word prediction.
- (Mis)use of such data might result privacy leakages, e.g., gender or racial profiling, sensitive attribute leaks, etc.

Anonymized data released for open competition by Netflix was shown to expose user identity by using auxiliary data from IMDB.

Source: https://www.wired.com/2010/03/netflix-cancels-contest/

ML algorithm trained on open-source code reveals usernames, passwords and other secret keys

Source: https://spectralops.io/blog/ai-ml-in-code-github-copilot-leak/

Background: Private Representation Learning

Context-agnostic Transformations

Differential privacy methods such as Laplace mechanism or Gaussian mechanism add independent noise to the data to prevent attribute identification.

Source: https://blog.openmined.org/maintaining-privacy-in-medical-data-with-differential-privacy/

Context-aware Private Representation Learning

Using the knowledge of sensitive attributes in training data, PRL techniques learn a transformation that hides only the private attributes in data.

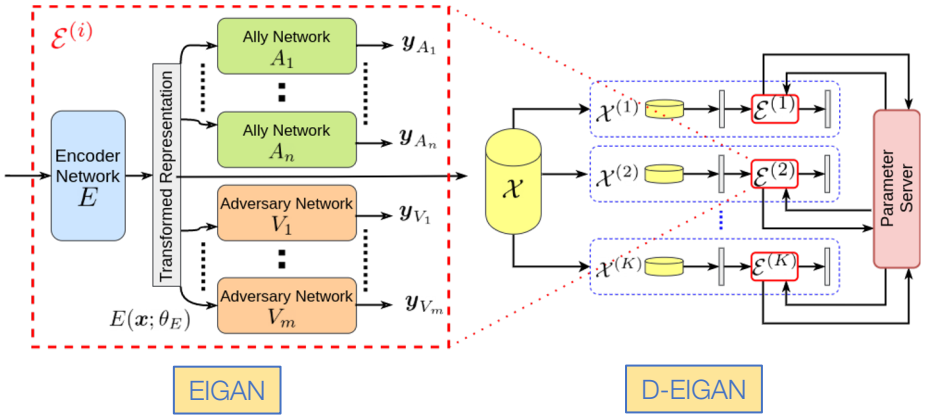
Source: Bertran, Martin, et al. "Adversarially learned representations for information obfuscation and inference." *ICML*, 2019.

Our Methodology: EIGAN and Distributed-EIGAN

Contributions of EIGAN

1. EIGAN can account for multiple ally and adversary attributes
2. D-EIGAN can train on a distributed dataset while ensuring differentially private parameter sharing.
3. Optimization function naturally pushes adversary inference to uniform distribution.

'Adversarial training to preserve **ally/utility attributes** but obfuscate **adversary /sensitive attributes**'



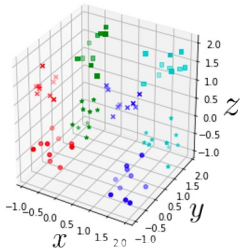
Exclusion-Inclusion Generative Adversarial Network

Theoretical and Experimental Results

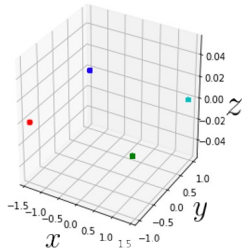
Theoretical Results

Theorem 1: EIGAN's adversarial loss formulation naturally pushes the inference on sensitive attributes towards uniform distribution.

Visualization:



Original Data



EIGAN Transformation

Experimental Results

Objective	Adult Dataset		Facescrub Dataset	
	Ally (identity)	**Adversary** (gender)	**Ally** (income)	**Adversary** (gender)
Unencoded	0.85	0.85	0.98	0.99
Linear-ARL	**0.84**	0.67	-	-
Kernel-ARL	**0.84**	0.67	-	-
Bertran-PRL	0.82	0.67	**0.56**	0.68
EIGAN	**0.84**	0.67	**0.82**	0.68
% Improv.	Matches closed form solution	Controlled to be equal	47.01%	Controlled to be equal

Table 1: Performance comparison between EIGAN, [1] (Linear-ARL, Kernel-ARL), and [2] (Bertran-PRL) on the Adult & FaceScrub datasets considered in those works.

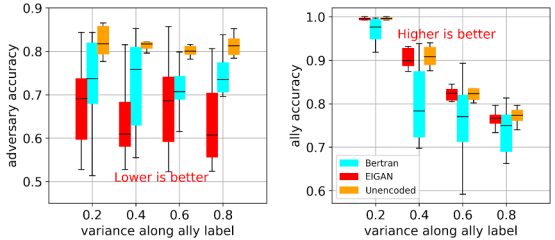


Figure 1: Effect of varying the ally class overlap (by changing the variances of synthetic Gaussian data) on the performance of EIGAN, [2] and the unencoded data.

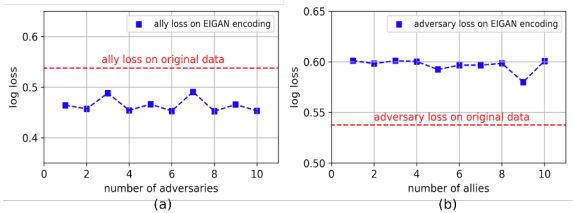


Figure 2: EIGAN's effect of the number of (a) adversaries, and (b) allies on the testing loss for MIMIC-III dataset.

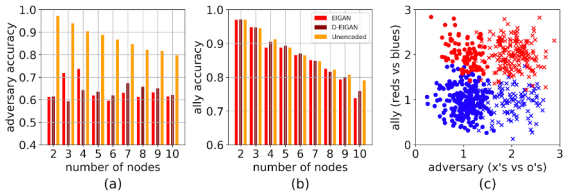


Figure 3: Performance comparison between EIGAN, D-EIGAN, and unencoded data. We observe that D-EIGAN performs as well as EIGAN under the distributed constraints.

[1] Sadeghi et al. Imparting fairness to pre-trained biased representations. In IEEE CVPRW, 2020.
[2] Bertran et al. Adversarially Learned Representations for Information Obfuscation and Inference. In ICML, 2019.

Summary

- Why privacy in machine learning?
- Context-agnostic privacy methods vs private representation learning
- Our methodology: EIGAN, D-EIGAN, and its contributions
- Theoretical and experimental results.

Thank you!

Questions?

Contact: azam1@purdue.edu

Or refer to our paper:

https://arxiv.org/pdf/2010.01792.pdf