# DP-SGD vs PATE: Which Has Less Disparate Impact on GANs?

**Georgi Ganev**

UCL & Hazy

## Problem description

**Goal:** Empirically measure the disparate effect of Differential Privacy (DP-SGD vs PATE) on GANs in terms of 1) size and 2) classification accuracy on different subgroups of synthetic data.

## Experimental settings

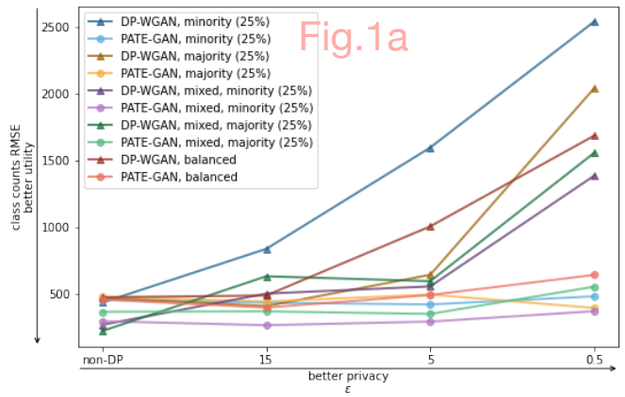
On MNIST:

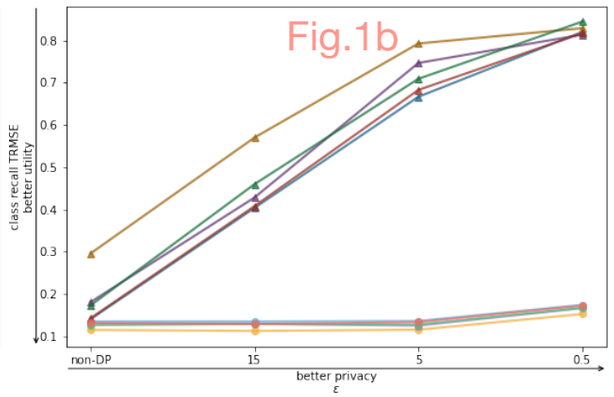
1. *Minority* – undersample "8"
2. *Majority* – undersample all classes except "8"
3. *Mixed* – undersample all classes in uniformly decreasing manner turning "8" into minority/majority

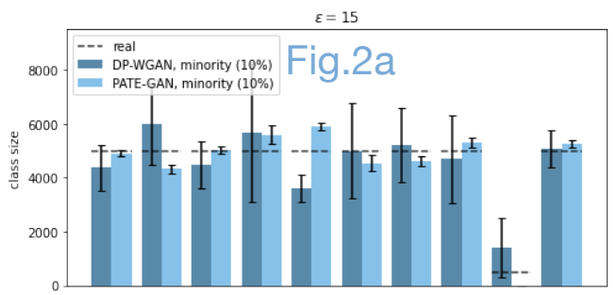
Compare DP-WGAN vs PATE-GAN for various ε.

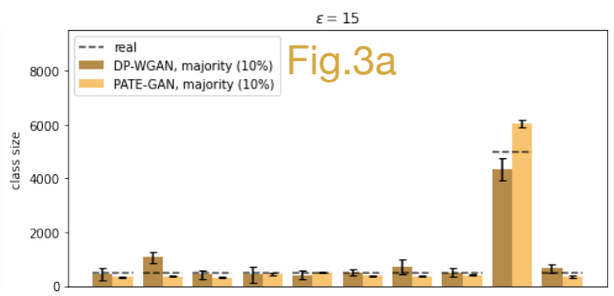
## Main findings

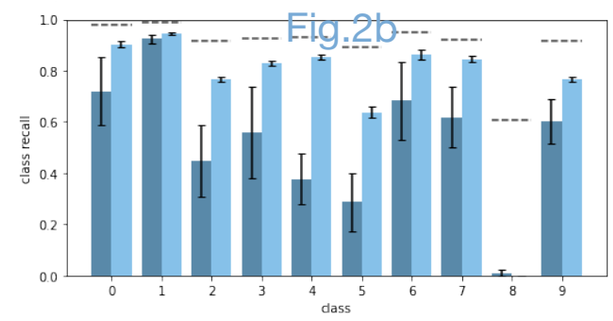
- PATE-GAN exhibits a *milder* disparity and much *better* privacy-utility trade-off (Fig.1, Fig.2b & Fig.3b).
- In terms of size, the two models behave in opposite directions – DP-WGAN "*evens*" the classes while PATE-GAN *increases* the imbalance (Fig.2a & Fig.3a).
- PATE-GAN, unlike DP-WGAN, *fails* to learn whole subpopulation with highly imbalanced class (Fig.2a).
- PATE-GAN, unlike DP-WGAN, *benefits* from some degree of privacy (serves as regularization) (Fig.1).

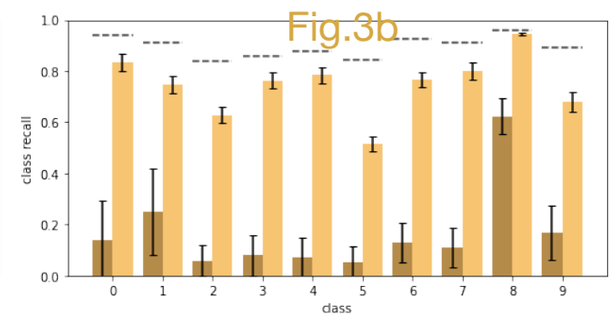



## So what?

Analyzing/training models on DP synthetic data could result in:

- treating different subpopulations unevenly
- unreliable/unfair conclusions with real societal costs

Full paper (+ further analysis on mixed/ balanced settings and # of teachers in PATE-GAN):

georgi.ganev.16@ucl.ac.uk

AAAI Workshop on Privacy-Preserving Artificial Intelligence (PPAI-22)