Mitigating Statistical Bias within Differentially Private Synthetic Data

Sahra Ghalebikesabi

Our Team


Harrison Wilde


Jack Jewson


Arnaud Doucet


Sebastian Vollmer


Chris Holmes

Our Mission

The right of the people to
useful private data
shall not be infringed.

Our Mission

The right of the people to
useful private data
shall not be infringed.

Our Mission

The right of the people to
useful private data
shall not be infringed.

Our Mission

kaggle

NEW

The right of the people to
useful private data
shall not be infringed.

Our Mission

kaggle

The right of the people to
useful private data
shall not be infringed.

NEW

Our Mission

The right of the people to
useful private data
shall not be infringed.

NEW

Our Road Map

1. Train Differentially Private Synthetic Data Generator

Our Road Map

1. Train Differentially Private Synthetic Data Generator
2. Generate Synthetic Data Samples

Our Road Map

1. Train Differentially Private Synthetic Data Generator
2. Generate Synthetic Data Samples
3. Estimate Differentially Private Importance Weights

Our Road Map

1. Train Differentially Private Synthetic Data Generator
2. Generate Synthetic Data Samples
3. Estimate Differentially Private Importance Weights
4. Perform downstream tasks on importance weighted synthetic data

Our Road Map

1 Train **Differentially Private** Synthetic Data Generator

Differential Privacy

A randomised algorithm $g : \mathcal{M} \to \mathcal{R}$ satisfies (ϵ, δ)-**differential privacy** for $\epsilon, \delta \geq 0$ if and only if for all neighbouring datasets $\mathcal{D}, \mathcal{D}'$ and all subsets $S \subseteq \mathcal{R}$, we have

$$\Pr(g(\mathcal{D}) \in S) \leq \delta + e^{\epsilon}\Pr(g(\mathcal{D}') \in S).$$

Differential Privacy by Noising

The **sensitivity** of g w.r.t a norm $|\cdot|$ is defined by the smallest number $S(g)$ such that for any two neighbouring datasets $\mathcal{D}$ and $\mathcal{D}'$ it holds that

$$|g(\mathcal{D}) - g(\mathcal{D}')| \leq S(g).$$

To ensure the $(\epsilon, 0)$-differential privacy of g, it suffices to add Laplacian noise with standard deviation $S(g)/\epsilon$ to g.

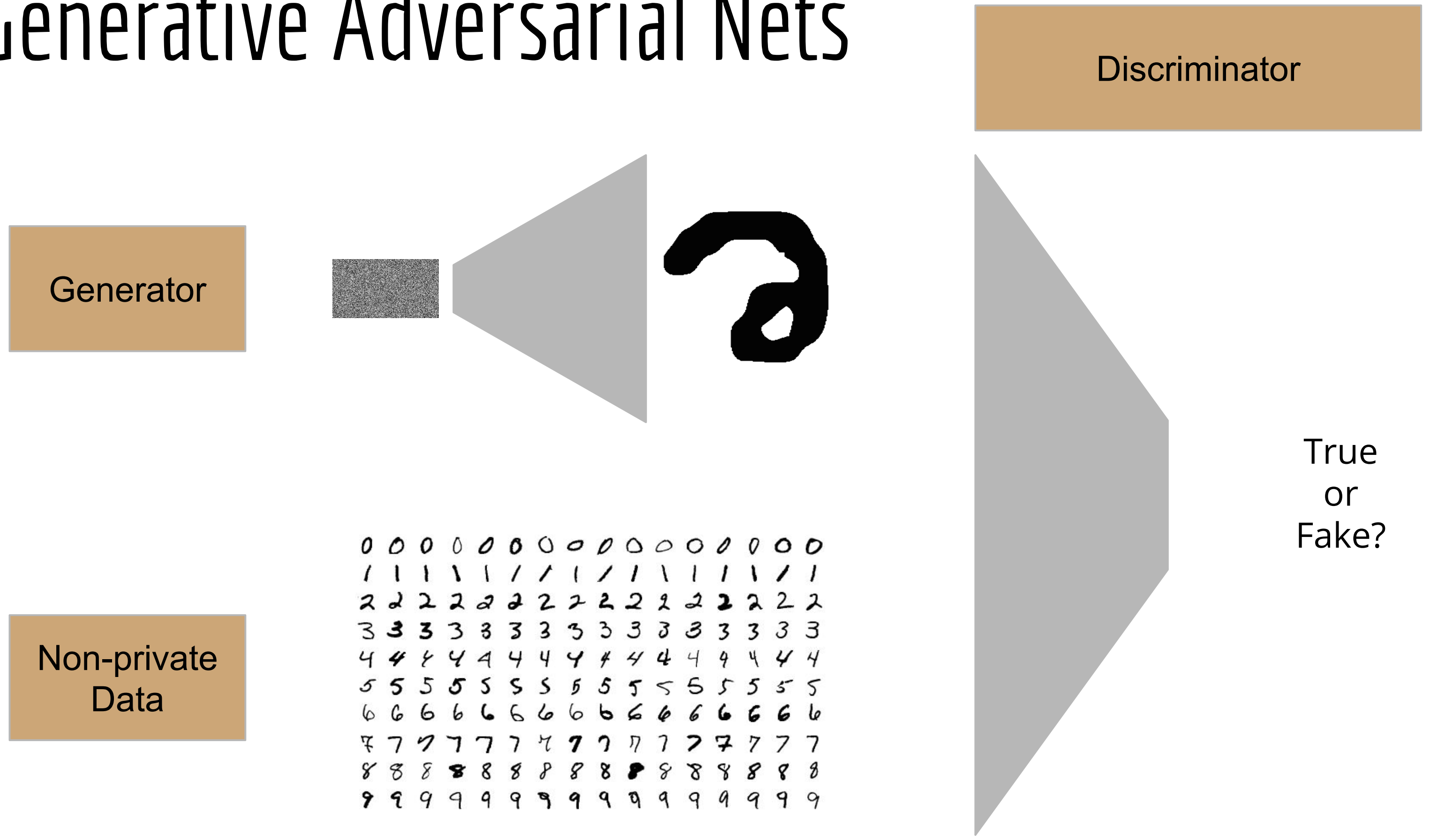
Our Road Map

1. Train Differentially Private **Synthetic Data Generator**

Our Road Map

1. Train Differentially Private **Synthetic Data Generator**

Generative Adversarial Nets



Our Road Map

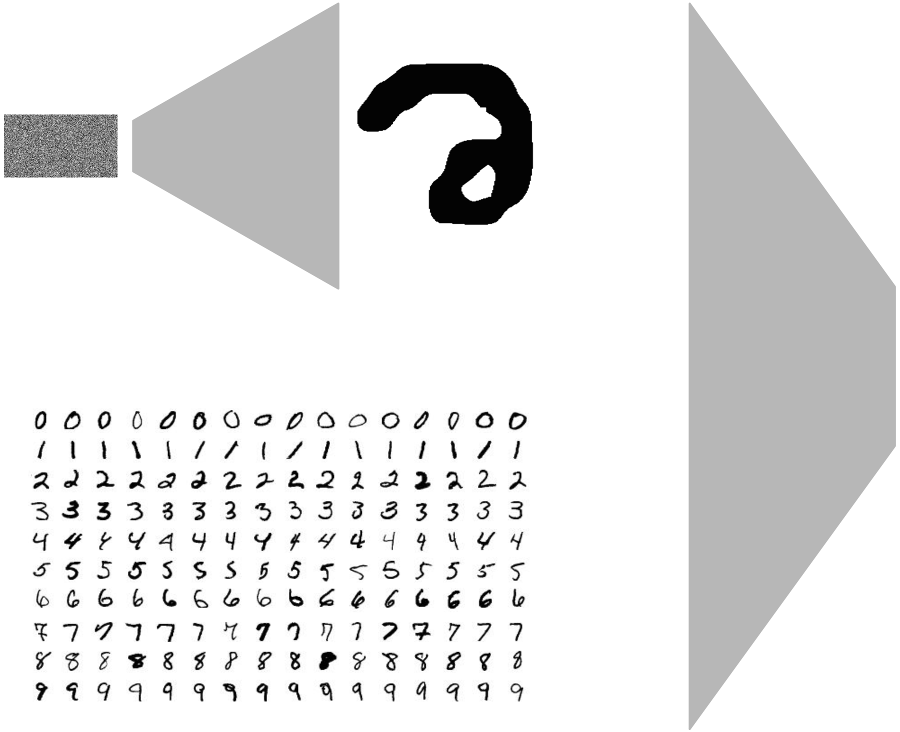
1. Train **Differentially Private Synthetic Data Generator**

DP-GANs

Discriminator

Generator

Non-private Data

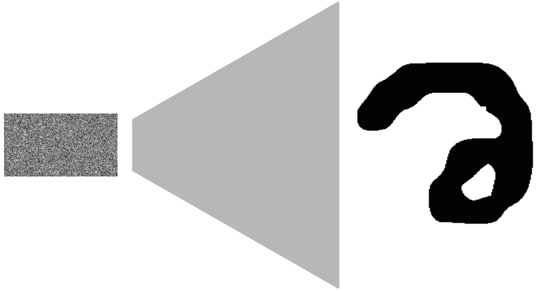


True or Fake?

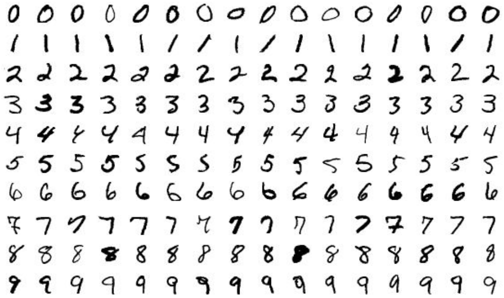
DP-GANs

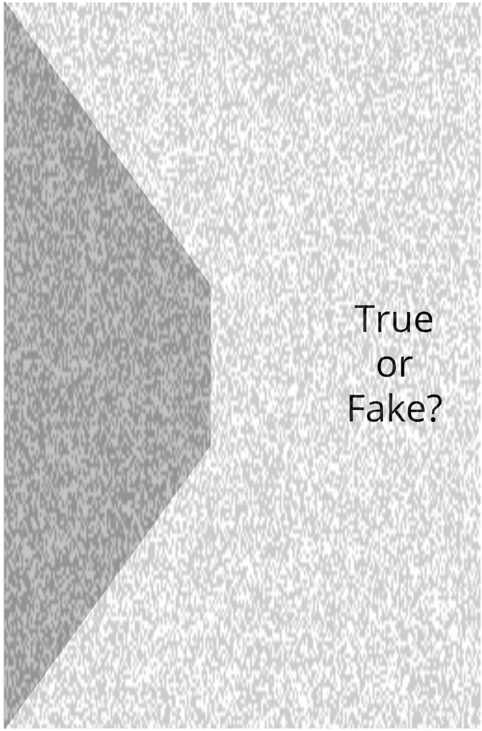
Discriminator

Generator



Non-private Data





Our Road Map

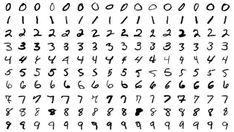
1. Train Differentially Private Synthetic Data Generator
2. **Generate Synthetic Data Samples**

Our Road Map

1. Train Differentially Private Synthetic Data Generator
2. Generate Synthetic Data Samples
3. **Estimate Differentially Private Importance Weights**

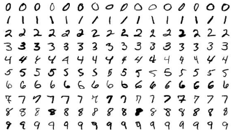
Importance Weighting

 $x_{1:N_G} \overset{\text{i.i.d.}}{\sim} p_G$

 $x'_{1:N_D} \overset{\text{i.i.d.}}{\sim} p_D$

Importance Weighting

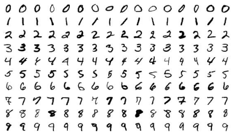
 $x_{1:N_G} \overset{\text{i.i.d.}}{\sim} p_G$

 $x'_{1:N_D} \overset{\text{i.i.d.}}{\sim} p_D$

$p_G(\cdot) > 0$ whenever $h(\cdot)p_D(\cdot) > 0$

Importance Weighting

 $x_{1:N_G} \overset{\text{i.i.d.}}{\sim} p_G$

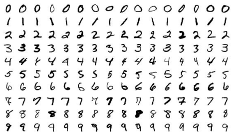
 $x'_{1:N_D} \overset{\text{i.i.d.}}{\sim} p_D$

$p_G(\cdot) > 0$ whenever $h(\cdot)p_D(\cdot) > 0$

$$w(x) := \frac{p_D(x)}{p_G(x)}$$

Importance Weighting

 $x_{1:N_G} \overset{\text{i.i.d.}}{\sim} p_G$

 $x'_{1:N_D} \overset{\text{i.i.d.}}{\sim} p_D$

$p_G(\cdot) > 0$ **whenever** $h(\cdot)p_D(\cdot) > 0$

$$w(x) := \frac{p_D(x)}{p_G(x)}$$

$$\boxed{\mathbb{E}_{x\sim p_D}[h(x)] = \mathbb{E}_{x\sim p_G}[w(x)h(x)]}$$

1) GAN discriminator weights

$$\frac{\widehat{p}(y=1|x)}{\widehat{p}(y=0|x)}$$

1) GAN discriminator weights

$$\frac{\widehat{p}(y=1|x)}{\widehat{p}(y=0|x)}$$

$$=\frac{\widehat{p}(y=1|x)}{\widehat{p}(y=0|x)}\frac{\widehat{p}(y=0)}{\widehat{p}(y=1)}$$

1) GAN discriminator weights

$$\frac{\widehat{p}(y=1|x)}{\widehat{p}(y=0|x)}$$

$$=\frac{\widehat{p}(y=1|x)}{\widehat{p}(y=0|x)}\frac{\widehat{p}(y=0)}{\widehat{p}(y=1)}$$

$$=\frac{\widehat{p}(x|y=1)}{\widehat{p}(x|y=0)}=\frac{\widehat{p}_D(x)}{\widehat{p}_G(x)}$$

2) Differentially Private Logistic Regression

If the data is scaled to a range from 0 to 1 such that $X \subset [0,1]^d$, Chaudhuri et al. (2021) show that the L_2 sensitivity of the optimal coefficient vector estimated by $\widehat{\beta}$ in a regularised logistic regression with model

$$\widehat{p}(y=1|x_i) = \sigma(\widehat{\beta}^T x_i) = \left(1 + e^{-\widehat{\beta}^T x_i}\right)^{-1}$$

is $S(\widehat{\beta}) = 2\sqrt{d}/(N_D \lambda)$ where λ is the coefficient of the L_2 regularisation term added to the loss during training.

2) Differentially Private Logistic Regression

Ji and Elkan (2013)

$$\overline{\beta} = \widehat{\beta} + \zeta$$

$$\log \overline{w}(x_i) = \overline{\beta}^T x_i = \widehat{\beta}^T x_i + \zeta x_i$$

$I_N(h|\overline{w})$ *is biased.*

2) Differentially Private Logistic Regression

Proposition 2 (Supplement B.2). *Let $\overline{w}$ denote the importance weights computed by noise perturbing the regression coefficients as in Equation (8) (Ji and Elkan, 2013, Algorithm 1) where ζ can be sampled from any noise distribution that ensures (ϵ, δ)-differential privacy of $\overline{\beta}$. Define*

$$b(x_i) := 1/\mathbb{E}_{p_\zeta}[\exp\left(\zeta^T x_i\right)],$$

and adjusted importance weight

$$\overline{w}^*(x_i) = \overline{w}(x_i) b(x_i) = \widehat{w}(x_i) \exp\left(\zeta^T x_i\right) b(x_i). \quad (9)$$

The importance sampling estimator $I_N(h|\overline{w}^*)$ is unbiased and (ϵ, δ)-DP for $\mathbb{E}_{p_\zeta}[\exp\left(\zeta^T x_i\right)] > 0$.

2) Differentially Private Logistic Regression

SDGP	data	$\epsilon = 1$		$\epsilon = 6$	
		BetaNoised	BetaDebiased	BetaNoised	BetaDebiased
CGAN	Breast	$1.4833_{\pm 0.9603}$	$\mathbf{0.0775}_{\pm \mathbf{0.0197}}$	$0.0024_{\pm 0.0006}$	$\mathbf{0.0020}_{\pm \mathbf{0.0004}}$
	Banknote	$0.0420_{\pm 0.0211}$	$\mathbf{0.0413}_{\pm \mathbf{0.0196}}$	$\mathbf{0.0014}_{\pm \mathbf{0.0007}}$	$\mathbf{0.0014}_{\pm \mathbf{0.0007}}$
	Iris	$8.7522_{\pm 4.9893}$	$\mathbf{3.4687}_{\pm \mathbf{1.3044}}$	$\mathbf{0.1160}_{\pm \mathbf{0.0240}}$	$0.1290_{\pm 0.0311}$
GAN	Housing	$8.2081_{\pm 7.7702}$	$\mathbf{1.4406}_{\pm \mathbf{0.8314}}$	$3.7916_{\pm 3.3246}$	$\mathbf{1.5479}_{\pm \mathbf{1.0430}}$
DPCGAN	Breast	$0.0582_{\pm 0.0165}$	$\mathbf{0.0445}_{\pm \mathbf{0.0162}}$	$0.0015_{\pm 0.0003}$	$\mathbf{0.0014}_{\pm \mathbf{0.0003}}$
	Banknote	$0.0420_{\pm 0.0211}$	$\mathbf{0.0413}_{\pm \mathbf{0.0196}}$	$0.0022_{\pm 0.0013}$	$\mathbf{0.0021}_{\pm \mathbf{0.0012}}$
	Iris	$\mathbf{0.7834}_{\pm \mathbf{0.2341}}$	$1.2300_{\pm 0.7050}$	$\mathbf{0.2502}_{\pm \mathbf{0.1627}}$	$0.2806_{\pm 0.1760}$
DPGAN	Breast	$6.0487_{\pm 3.7927}$	$\mathbf{3.7629}_{\pm \mathbf{2.2881}}$	$0.0251_{\pm 0.0245}$	$\mathbf{0.0238}_{\pm \mathbf{0.0234}}$
	Banknote	$\mathbf{0.0582}_{\pm \mathbf{0.0353}}$	$0.0610_{\pm 0.0397}$	$0.0062_{\pm 0.0057}$	$\mathbf{0.0061}_{\pm \mathbf{0.0056}}$
	Iris	$2.6486_{\pm 1.3518}$	$\mathbf{1.3698}_{\pm \mathbf{1.1554}}$	$\mathbf{0.0741}_{\pm \mathbf{0.0228}}$	$0.0864_{\pm 0.0274}$
	Housing	$5.9175_{\pm 2.8546}$	$\mathbf{0.8398}_{\pm \mathbf{0.6328}}$	$\mathbf{1.9044}_{\pm \mathbf{1.1426}}$	$2.1111_{\pm 1.3450}$

Table 6: Mean squared error of the privatised log importance weights $\log \overline{w}$ resp. $\log \overline{w}^*$ averaged over 10 runs with standard errors reported in brackets for $(\epsilon = 1, \delta = 10^{-5})$ and $(\epsilon = 6, \delta = 10^{-5})$ where $\epsilon_{IW} = 0.1\epsilon$.

3) Differentially Private Deep Learning

Algorithm 1: Relaxed DP SGD

Input: Examples $x_{1:N_D}, y_{1:N_D}$ from the DGP and $x_{N_D+1:N_D+N_G}, y_{N_D+1:N_D+N_G}$ from the SDGP, loss function $\mathcal{L}(\theta) = \frac{1}{N_G+N_D}\sum_i \mathcal{L}(\theta, x_i, y_i)$. Parameters: learning rate η_t, noise scale σ, expected lot size L, gradient norm bound C.

Initialise θ_0 randomly
for $t \in [T]$ **do**
  Construct a random subset $L_t \subset \{1, \ldots, N_D + N_G\}$ by including each index independently at random with probability $\frac{L}{N_D+N_G}$
  Compute gradient
  For each $i \in L_t$, compute $g_t(x_i, y_i) \leftarrow \Delta_{\theta_t}\mathcal{L}(\theta_t, x_i, y_i)$
  Clip gradient
  $\overline{g}_t(x_i, y_i) \leftarrow g_t(x_i, y_i)/\max(1, \frac{||g_t(x_i,y_i)||_2}{C})$
  Add noise
  $\tilde{g}_t \leftarrow \frac{1}{L}\sum_{i \in L_t}(\overline{g}_t(x_i, y_i) + N(0, \sigma^2 C^2 \mathbf{I})\mathbb{1}_{(y_i=1)})$, where $\mathbb{1}_{(y_i=1)}$ is 1 if $y_i = 1$ and 0 otherwise
  Descent
  $\theta_{t+1} \leftarrow \theta_t + \eta_t \tilde{g}_t$

Output: θ_T and the overall privacy cost (ϵ, δ) using the moment's accountant of Abadi et al. (2016) with sampling probability $q = \frac{L}{N_D+N_G}$.

Our Road Map

1. Train Differentially Private Synthetic Data Generator
2. Generate Synthetic Data Samples
3. Estimate Differentially Private Importance Weights
4. **Perform downstream tasks on importance weighted synthetic data**

Versatility of Importance Weighting

- Empirical Risk Minimisation

$$\frac{1}{N_D}\sum_{i=1}^{N_D} h(f(\cdot), x_i') \approx \mathbb{E}_{x\sim p_D}\left[h(f(\cdot), x)\right]$$

Versatility of Importance Weighting

- Empirical Risk Minimisation

$$\frac{1}{N_D}\sum_{i=1}^{N_D} h(f(\cdot), x_i') \approx \mathbb{E}_{x\sim p_D}\left[h(f(\cdot), x)\right]$$

- Bayesian Updating

$$\pi_{IW}(\theta|\tilde{x}) \propto \pi(\theta)\exp\left(\sum_{i=1}^{N_G} w(x_i)\log f(x_i|\theta)\right)$$

Versatility of Importance Weighting

- Empirical Risk Minimisation

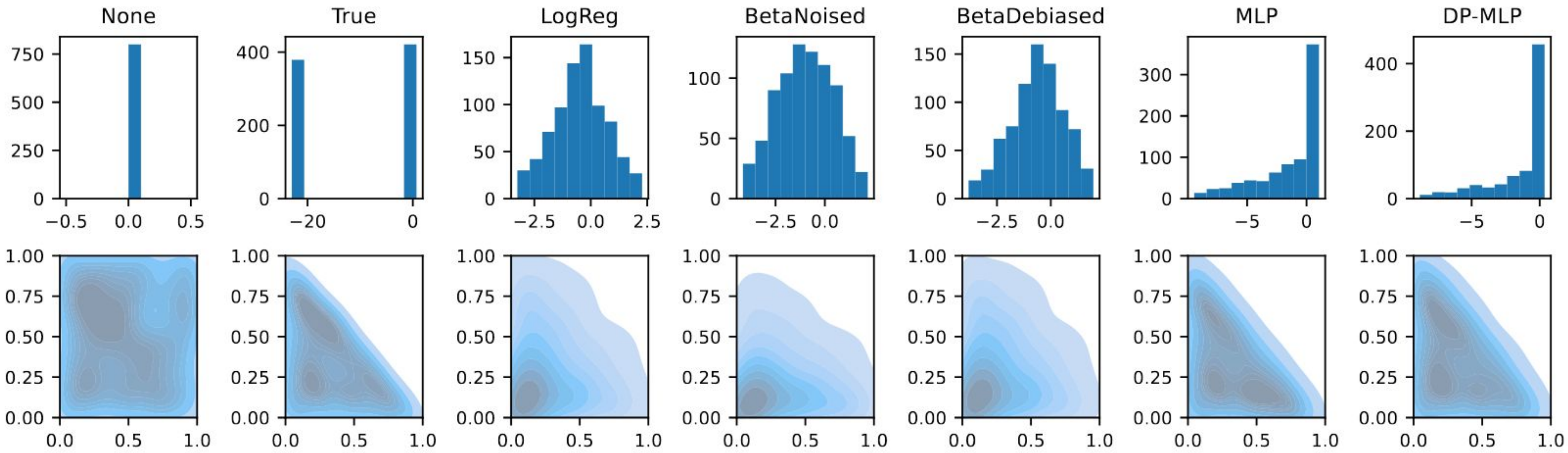
$$\frac{1}{N_D}\sum_{i=1}^{N_D} h(f(\cdot), x'_i) \approx \mathbb{E}_{x\sim p_D}\left[h(f(\cdot), x)\right]$$

- Bayesian Updating

$$\pi_{IW}(\theta|\tilde{x}) \propto \pi(\theta)\exp\left(\sum_{i=1}^{N_G} w(x_i)\log f(x_i|\theta)\right)$$

- Sampling

Data Visualisation



References

Ghalebikesabi, Sahra, et al. "Bias Mitigated Learning from Differentially Private Synthetic Data: A Cautionary Tale." *UAI* (2022).

Dwork, Cynthia, and Aaron Roth. "The algorithmic foundations of differential privacy." *Foundations and Trends® in Theoretical Computer Science* 9.3–4 (2014): 211-407.

Chaudhuri, Kamalika, Claire Monteleoni, and Anand D. Sarwate. "Differentially private empirical risk minimization." *Journal of Machine Learning Research* 12.3 (2011).

Grover, Aditya, et al. "Bias correction of learned generative models using likelihood-free importance weighting." *Advances in neural information processing systems* 32 (2019).

Ji, Zhanglong, and Charles Elkan. "Differential privacy based on importance weighting." *Machine learning* 93.1 (2013): 163-183.

Abadi, Martin, et al. "Deep learning with differential privacy." *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*. 2016.

Useful Links

https://sghalebikesabi.github.io

Website with
contact information

Paper