Differentially Private Sliced Wasserstein Distance

Alain Rakotomamonjy & Liva Ralaivola
Criteo AI Lab

Objective of the work

Privacy Preserving Learning

- Learn from data while guaranteeing privacy..
- ... in the context of domain adaptation and generative models
- we propose a

Differentially Private Distribution Distance

How ?

- Exploit the privacy property of
$$\mathcal{M}(\mathbf{X}) = \mathbf{X}\mathbf{U} + \mathbf{V},$$
- Make clear the ink between $\mathcal{M}(\mathbf{X})$ and Sliced Wasserstein Distance
- Introduce Differential Private SWD and its metric properties

Differential Privacy

Definition

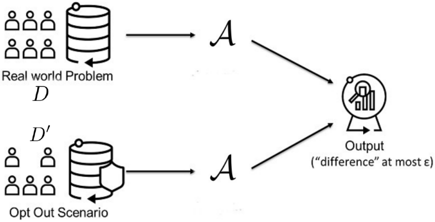
[Dwork, 2008]

Let $\varepsilon, \delta > 0$. Let $\mathcal{A} : \mathcal{D} \rightarrow \text{Im } \mathcal{A}$ be a *randomized* algorithm, where Im $\mathcal{A}$ is the image of $\mathcal{D}$ through $\mathcal{A}$. $\mathcal{A}$ is (ε, δ)-differentially private, or (ε, δ)-DP, if for all neighboring datasets $D, D' \in \mathcal{D}$ and for all sets of outputs $\mathcal{O} \in \text{Im } \mathcal{A}$, the following inequality holds :

$$\mathbb{P}[\mathcal{A}(D) \in \mathcal{O}] \leq e^{\varepsilon}\mathbb{P}[\mathcal{A}(D') \in \mathcal{O}] + \delta$$

where the probability relates to the randomness of $\mathcal{A}$.

Illustration



Renyi DP and Gaussian Mechanism

Renyi DP

[Mironov, 2017]

Let $\varepsilon > 0$ and $\alpha > 1$. A randomized algorithm $\mathcal{A}$ is (α, ε)-Rényi differential private or (α, ε)-RDP, if for any neighboring datasets $D, D' \in \mathcal{D}$,

$$\mathbb{D}_\alpha \left(\mathcal{A}(D) \| \mathcal{A}(D')\right) \leq \varepsilon$$

where $\mathbb{D}_\alpha(\cdot\|\cdot)$ is the Rényi α-divergence between two distributions.

How to easily make a function DP ?

Given a function $f : \mathcal{X} \to \mathbb{R}^d$, the *Gaussian mechanism* $\mathcal{M}_\sigma$ defined as follows :

$$\mathcal{M}_\sigma f(\cdot) = f(\cdot) + \mathbf{v}$$

where $\mathbf{v} \sim \mathcal{N}(0, \sigma^2 I_d)$. If f has Δ_2- (or ℓ_2-) *sensitivity*

$$\Delta_2 f \doteq \max_{D, D' \text{neighbors}} \|f(D) - f(D')\|_2,$$

then $\mathcal{M}_\sigma$ is $\left(\alpha, \frac{\alpha \Delta_2^2 f}{2\sigma^2}\right)$-RDP.

From Wasserstein Distance ...

Definition

- Given two probability distributions μ_s, μ_t on space Ω with metric $c(\cdot,\cdot)$
- For empirical distributions, the q-Wasserstein distance is When $\mu_s = \sum_{i=1}^{n} a_i \delta_{\mathbf{x}_i^s}$ and $\mu_t = \sum_{i=1}^{n} b_i \delta_{\mathbf{x}_i^t}$

$$W_q^q = \arg\min_{\mathbf{G}\in\mathbf{P}} \left\{ \langle \mathbf{G}, \mathbf{C}_q \rangle_F = \sum_{i,j} \gamma_{i,j} c_{i,j}^q \right\}$$

where $\mathbf{C}_q$ is a cost matrix with $c_{i,j} = c(\mathbf{x}_i^s, \mathbf{x}_j^t)^q$ and the marginals constraints are

$$\mathbf{P} = \left\{ \mathbf{G} \in (\mathbb{R}^+)^{n_s \times n_t} \,\middle|\, \mathbf{G}\mathbf{1}_{n_t} = \mathbf{a}, \mathbf{G}^T \mathbf{1}_{n_s} = \mathbf{b} \right\}$$

... through Sliced Wasserstein Distance ...

Statement

1D Wasserstein distance is cheap to compute

computing Sliced Wasserstein distance

- some sample random directions $\mathbf{u} \in \mathbb{S}^{d-1}$ uniformly
- project data on each random direction
- compute all 1d Wasserstein distance and average them

$$\text{SWD}_q^q = \frac{1}{k}\sum_{j=1}^{k} W_q^q\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{\mathbf{x}_i^{s\top}\mathbf{u}_j}, \frac{1}{m}\sum_{i=1}^{m}\delta_{\mathbf{x}_i^{t\top}\mathbf{u}_j}\right) \tag{1}$$

Features

- Still a distance, efficient to compute
- Randomized algorithm through $\mathbf{x}^\top\mathbf{u}$ (and $\mathbf{XU}$)

... to Differentially Private SWD

How to make SWD differentially private ?

- add Gaussian noise to the random projection $\mathbf{XU}$
- exploit post-processing DP property

Definition for empirical distributions

$$\text{DP-SWD}_q^q = \frac{1}{k}\sum_{j=1}^{k} W_q^q\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{\mathbf{x}_i^\top \mathbf{u}_j + v_j}, \frac{1}{m}\sum_{i=1}^{m}\delta_{\mathbf{x}_i'^\top \mathbf{u}_j + v_j'}\right) \tag{2}$$

Questions

- What DP guarantee do we get ?
- Do we preserve any metric structure ?

(ε, δ)-DP Property

Sensitivity of $\mathbf{XU}$

- Assume $\mathbf{X}$, $\mathbf{X}'$ neighbouring datasets differing only at row i, $\mathbf{z} = \|\mathbf{x}_i - \mathbf{x}'_i\|_2 \leq 1$
- with $\mathbf{u} \in \mathbb{S}^{d-1}$, $\mathbf{z}^\top \mathbf{u} \sim B(1/2), (d-1)/2)$
- With prob $1 - \delta$

$$\|\mathbf{XU} - \mathbf{X}'\mathbf{U}\|_F^2 \leq w(k, \delta) \doteq \begin{cases} \frac{k}{d} + \frac{2}{3}\ln\frac{1}{\delta} + \frac{2}{d}\sqrt{k\frac{d-1}{d+2}\ln\frac{1}{\delta}} & \text{Bernstein} \\ \frac{k}{d} + \frac{z_{1-\delta}}{d}\sqrt{\frac{2k(d-1)}{d+2}} & \text{CLT} \end{cases}$$

(ε, δ)-DPness of $\mathbf{XU} + \mathbf{V}$

Assume $\mathbf{V}$ is a Gaussian matrix in $\mathbb{R}^{n \times k}$ with entries drawn from $\mathcal{N}(0, \sigma^2)$, for $\alpha > 1$,

$\mathbf{XU} + \mathbf{V}$ is $\left(\frac{\alpha w(k, \delta/2)}{2\sigma^2} + \frac{log(2/\delta)}{\alpha - 1}, \delta\right)$-DP.

Analyses of $w(k, \delta)$

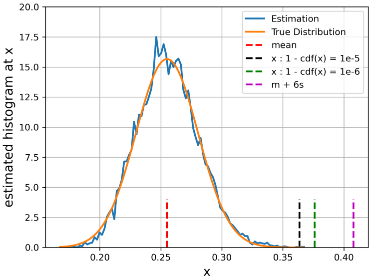
$$w(k,\delta) \doteq \begin{cases} \frac{k}{d} + \frac{2}{3}\ln\frac{1}{\delta} + \frac{2}{d}\sqrt{k\frac{d-1}{d+2}\ln\frac{1}{\delta}} & \text{Bernstein} \\ \frac{k}{d} + \frac{z_{1-\delta}}{d}\sqrt{\frac{2k(d-1)}{d+2}} & \text{CLT} \end{cases}$$

Looking at the equation

- from term $\frac{k}{d}$: the higher the dimension, the smaller the sensitivity
- the smaller the number of projection, the smaller the sensitivity
- there is an imcompressible term in $\frac{1}{\delta}$ for the Bernstein bound
- the CLT bound is tighter

Simulation

- $\|\mathbf{Uz}\|_2^2$ with fixed $\mathbf{z} \in \mathbb{S}^{d-1}$
- $d = 784$, $k = 200$, and 10000 draws of $\mathbf{U}$
- $\delta = 10^{-5}$
- Bernstein bound > 1
- CLT bound $< \frac{k}{d} + 6\sigma$



Metric properties of DP-SWD

DP-SWD is a distance

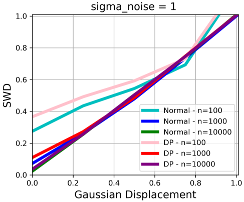
- Formalization

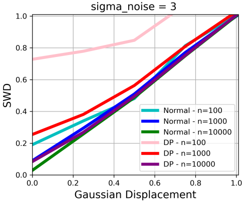
$$\mathsf{DP}_\sigma\mathsf{SWD}_q^q(\mu,\nu) \doteq \int_{\mathbb{S}^{d-1}} W_q^q(\mathcal{R}_{\mathbf{u}}\mu * \mathcal{N}_\sigma, \mathcal{R}_{\mathbf{u}}\nu * \mathcal{N}_\sigma) u_d(\mathbf{u}) d\mathbf{u}$$

  projection $\mathcal{R}_{\mathbf{u}}$, adding Gaussian noise is convolution with Gaussian
- All properties of a distance are preserved
- Gaussian smoothed version of original projected distributions

Simulation

Comparing two Gaussians, one with varying mean





Application of DP-SWD

Distribution matching in ML problems

- Generative modelling
$$\min_f \mathcal{D}(\mathbf{X}_t, f(z))$$
- Unsupervised domain adaptation
$$\min_{g,h} L_c(h(g(\mathbf{X}_s)), \mathbf{y}_s) + \mathcal{D}(g(\mathbf{X}_s), g(\mathbf{X}_t))$$

  with $\mathbf{X}_s, \mathbf{y}_s$, public labeled data from source domain, $\mathbf{X}_t$ unlabeled private data from target domain. $h(\cdot)$ the representation mapping, $g(\cdot)$ the classifier.

How-to make them privacy-preserving

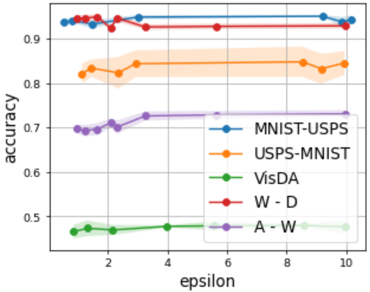
- Clip the input space so that $\|\mathbf{x}_i - \mathbf{x}_i'\|_2 \leq 1$
- In domain adaptation [Lee et al., 2019] or generative model [Deshpande et al., 2018], plug-in DP-SWD in place of SWD as distance $\mathcal{D}$.

Experiments on domain adaptation

Settings

- Computer Vision dataset (MNIST $\rightarrow$ USPS, VisDA, Office)
- UDA : learning representation + classifier
- Baselines : DANN, DA, using SWD and DP-DANN (with gradient clipping)
- Outcome : small loss of accuracy wrt SWD, robustness of the model accross large range of ε-DP guarantee.

Data	Methods DANN	SWD	DP-DANN	DP-SWD
M-U	93.9 $\pm$ 0	95.5 $\pm$ 1	87.1 $\pm$ 2	**94.0$\pm$ 0**
U-M	86.2 $\pm$ 2	84.8$\pm$2	73.5 $\pm$ 2	**83.4$\pm$ 2**
VisDA	57.4 $\pm$ 1	53.8$\pm$1	**49.0 $\pm$ 1**	47.0$\pm$ 1
D - W	90.9$\pm$ 1	90.7$\pm$ 1	88.0$\pm$ 1	**90.9$\pm$ 1**
D - A	58.6$\pm$ 1	59.4$\pm$ 1	**56.5$\pm$ 1**	55.2$\pm$ 2
A - W	70.4$\pm$ 3	74.5$\pm$ 1	68.7$\pm$ 1	**72.6$\pm$ 1**
A - D	78.6$\pm$ 2	78.5$\pm$ 1	73.7$\pm$ 1	**79.8$\pm$ 1**
W - A	54.7$\pm$ 3	59.1$\pm$ 0	56.0$\pm$ 1	**59.0$\pm$ 1**
W - D	91.1$\pm$ 0	95.7$\pm$ 1	63.4$\pm$ 3	**92.6$\pm$ 1**



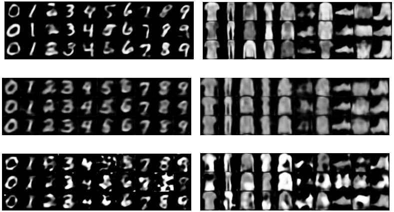
Experiments on generative modelling

Settings

- generate MNIST and FashionMNIST samples from private data
- Evaluate quality of the generated data on classification task
- same experimental setting as in DP-MERF [Harder et al., 2020].

	MNIST		FashionMNIST	
Method	MLP	LogReg	MLP	LogReg
SWD	87	82	77	76
GS-WGAN	79	79	65	68
DP-CGAN	60	60	50	51
DP-MERF	76	75	**72**	**71**
DP-SWD-c	**77**	**78**	67	66
DP-SWD-b	76	77	67	66

DP-SWD DP-MERF

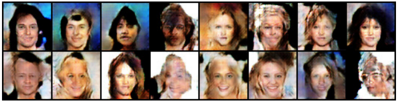


Experiments on generative modelling

Setting

- CelebA dataset. original input $64 \times 64 \times 3$. first application of DP generative model on this dataset
- architecture and optimizer as in [Nguyen et al., 2020]. Latent space of distributions to be compared 8192.
- plugged-in DP-SWD instead of SWD.
- bound choice $w(k, \delta)$ strongly impacts visual quality

first row : SWD, second row DP-SWD with (left) CLT and (right) Berstein bound.



Conclusion

What we proposed

- a differentially private distance on distributions
- DP-SWD exploits random projection + Gaussian mechanism
- Seamless plug into learning models
- but ...
  - introduce smoothness

On-going extension

- theoretical analysis of the Gaussian smoothed SWD
- better post-processing for generative modelling

References I

- Deshpande, I., Zhang, Z., and Schwing, A. G. (2018).
  Generative modeling using the sliced wasserstein distance.
  In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3483–3491.
- Dwork, C. (2008).
  Differential privacy : A survey of results.
  In *International conference on theory and applications of models of computation*, pages 1–19. Springer.
- Harder, F., Adamczewski, K., and Park, M. (2020).
  Differentially private mean embeddings with random features (dp-merf) for simple & practical synthetic data generation.
  arXiv preprint arXiv :2002.11603.
- Lee, C.-Y., Batra, T., Baig, M. H., and Ulbricht, D. (2019).
  Sliced wasserstein discrepancy for unsupervised domain adaptation.
  In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10285–10295.
- Mironov, I. (2017).
  Rényi differential privacy.
  In *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, pages 263–275.
- Nguyen, K., Ho, N., Pham, T., and Bui, H. (2020).
  Distributional sliced-wasserstein and applications to generative modeling.
  arXiv preprint arXiv :2002.07367.