Comparative study between Generative Models based on Sinkhorn loss and GAN

Louis Geist, Valentin Gerard

21/12/2023

- Generative task : unsupervised learning that learns the probability distribution of the images
- Loss between empirical probability measures :
  - Use a distance on probability measures $\rightarrow$ Optimal transport
  - Test whether the generated samples and the samples from the training dataset can be separated by a neural network $\rightarrow$ GAN approach

Sommaire

Entropic relaxation of Optimal Transport problem

Let μ and ν be two probability measures.
For $\varepsilon \in \mathbb{R}_+$:

$$(\mathcal{P}_\varepsilon) : \min_{\pi \in \Pi(\mu,\nu)} \int c(x,y)\mathrm{d}\pi(x,y) + \varepsilon \mathrm{KL}(\pi \mid \mu \otimes \nu)$$

with :

$$\mathrm{KL}(\pi \mid \xi) = \int_{\mathcal{X}\times\mathcal{X}} \log\left(\frac{\mathrm{d}\pi}{\mathrm{d}\xi}(x,y)\right)\mathrm{d}\pi(x,y)$$

Benefits of the Sinkhorn loss

- Fast computations with Sinkhorn algorithm
- Interpolation between Optimal Transport distance and Maximum Mean Discrepancy (MMD)

Generative Adversial Networks (GAN)

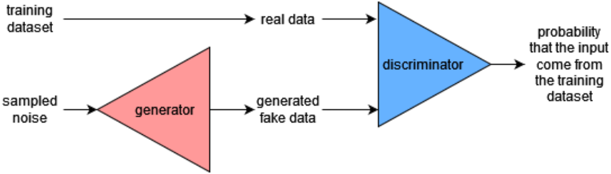

Figure: GAN architecture

- Adversial Training :
  - Discriminator is trained to classify correctly fake and real data
  - Generator is trained to make the discriminator classify generated data as real data

Sommaire

Architecture and training parameters

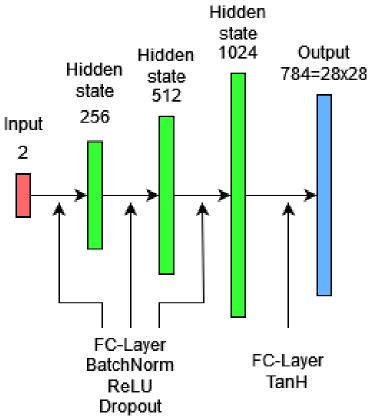

Figure: Generators architecture

Parameter	Value
Nb Epochs	40
Learning Rate	10^{-3}
Batch Size	200
Optimizer	Adam

Table: Training Configuration

Dataset	Sizes
MNIST	60000x28x28
FashionMNIST	60000x28x28

Table: Training Datasets

Sommaire

Qualitative results

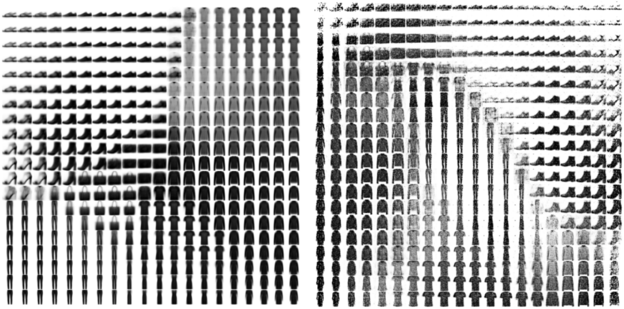

Figure: Images generated using a grid of the latent space (Sink1 left and GAN right)

- Geometric considerations of Sinkhorn loss give to the latent space interpolation properties.

Quantitative results on Fashion MNIST

FashionMNIST	GSink_1	GSink_10	GSink_100	GAN
OT distance	49	63	75	82
Inception score	6.8	7.7	7.5	8.3

Table: Quantitative results of generative models learned with Sinkhorn loss with ε (GSink_ε) and the GAN on the FashionMNIST dataset.

Sinkhorn based models achieve :

- Similar performance to GAN
- Better geometric properties

Further works :

- Confirm the results of the Inception Score and try to understand them better
- Validity of our conclusions for larger datasets?

[1] Shane Barratt and Rishi Sharma. *A Note on the Inception Score*. 2018. arXiv: 1801.01973 [stat.ML].

[2] Eyal Betzalel et al. *A Study on the Evaluation of Generative Models*. 2022. arXiv: 2206.10935 [cs.LG].

[3] Jia Deng et al. "ImageNet: A large-scale hierarchical image database". In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 2009, pp. 248–255. DOI: 10.1109/CVPR.2009.5206848.

[4] Li Deng. "The mnist database of handwritten digit images for machine learning research". In: *IEEE Signal Processing Magazine* 29.6 (2012), pp. 141–142.

[5] Jean Feydy. "Analyse de données géométriques, au delà des convolutions". Theses. Université Paris-Saclay, July 2020. URL: https://theses.hal.science/tel-02945979.

[6] Jean Feydy et al. "Interpolating between Optimal Transport and MMD using Sinkhorn Divergences". In:

The 22nd International Conference on Artificial Intelligence and Statistics. 2019, pp. 2681–2690.

[7] Chuan Guo et al. "On calibration of modern neural networks". In: *International conference on machine learning*. PMLR. 2017, pp. 1321–1330.

[8] L. Kantorovitch. "On the Translocation of Masses". In: *Management Science* 5.1 (1958), pp. 1–4. ISSN: 00251909, 15265501. URL: http://www.jstor.org/stable/2626967 (visited on 12/10/2023).

[9] Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton. "CIFAR-10 (Canadian Institute for Advanced Research)". In: (). URL: http://www.cs.toronto.edu/~kriz/cifar.html.

[10] Y. Lecun et al. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324. DOI: 10.1109/5.726791.

[11] Gabriel Peyré and Marco Cuturi. *Computational Optimal Transport*. 2020. arXiv: 1803.00567 [stat.ML].

[12] Richard Sinkhorn. "A relationship between arbitrary positive matrices and doubly stochastic matrices". In: *The annals of mathematical statistics* 35.2 (1964), pp. 876–879.

[13] Christian Szegedy et al. *Going Deeper with Convolutions*. 2014. arXiv: 1409.4842 [cs.CV].

[14] Han Xiao, Kashif Rasul, and Roland Vollgraf. *Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms*. 2017. arXiv: 1708.07747 [cs.LG].

Kantorovitch formulation of OT

For μ and ν two probability measures on $\mathcal{X}$:

$$(\mathcal{P}) : \min_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} c(x,y)\mathrm{d}\pi(x,y)$$

with the set of admissible couplings :

$$\Pi(\mu,\nu) \overset{\text{def.}}{=} \{\pi \in \mathcal{M}_+^1(\mathcal{X}\times\mathcal{X}); P_{1\sharp}\pi = \mu, P_{2\sharp}\pi = \nu\}$$

Sinkhorn loss between μ and ν

By noting $\mathcal{W}_{\boldsymbol{c},\varepsilon}(\mu, \nu)$ the value of $(\mathcal{P}_\varepsilon)$ for μ and ν :

$$\overline{\mathcal{W}}_{\boldsymbol{c},\varepsilon}(\mu, \nu) = 2\mathcal{W}_{\boldsymbol{c},\varepsilon}(\mu, \nu) - \mathcal{W}_{\boldsymbol{c},\varepsilon}(\mu, \mu) - \mathcal{W}_{\boldsymbol{c},\varepsilon}(\nu, \nu)$$