Louis Geist. Valentin Gerard

Comparative study between Generative Models based on Sinkhorn loss and GAN

Louis Geist. Valentin Gerard

21/12/2023

Louis Geist, Valentin Gerard

Introduction

Learning Generative Models with Sinkhorn Divergences

Architecture an

Qualitative results Quantitative results

Quantitutive resur

References

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

Louis Geist. Valentin Gerard

Introduction

- 1 Related works
- 2 Method

Results

Louis Geist, Valentin Gerard

Introduction

Related works

Models with
Sinkhorn Divergence

Networks

Method

Architecture and

parameters

Qualitative result

Quantitative resul

Conclusion

References

Sommaire

1 Related works

Learning Generative Models with Sinkhorn Divergences Generative Adversial Networks

- 2 Method
- 3 Results

Louis Geist. Valentin Gerard

Learning Generative Models with Sinkhorn Divergences

Entropic relaxation of Optimal Transport problem

Let μ and ν be two probability measures.

For $\varepsilon \in \mathbb{R}_+$:

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

with:

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

Louis Geist, Valentin Gerard

Introduction

Related work

Learning Generative Models with Sinkhorn Divergences

Generative Adversia

IVetworks

Method

Architecture ar parameters

Qualitative results

C . . . I

References

Benefits of the Sinkhorn loss

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

Louis Geist, Valentin Gerard

Introduction

Related wor

Learning Generative Models with

Generative Adversial

Method

Architecture an

parameters

Results

Qualitative result

Quantitative result

Conclusion

Reference:

Generative Adversial Networks (GAN)

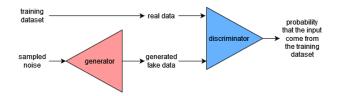


Figure: GAN architecture

- 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

Louis Geist, Valentin Gerard

Introduction

Related work

Models with
Sinkhorn Divergences

Networks

Method

Architecture a parameters

Results

Quantitative results

_ . . .

Reference:

Sommaire

1 Related works

2 Method Architecture and parameters

3 Results

Louis Geist, Valentin Gerard

Introduction

Learning Generative Models with Sinkhorn Divergence

Architecture and parameters

<u>.</u>

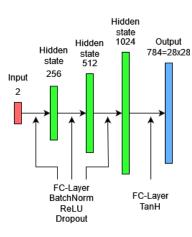
Results

Qualitative results

Conclusio

References

Architecture and training parameters



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

Table: Training Configuration

Dataset	Sizes		
MNIST	60000×28×28		
FashionMNIST	60000×28×28		

Figure: Generators architecture

Table: Training Datasets

Louis Geist. Valentin Gerard

Results

Sommaire

3 Results

Qualitative results Quantitative results

Louis Geist, Valentin Gerard

Introduction

Related works

Learning Generative
Models with
Sinkhorn Divergence
Generative Adversia

Madhaal

Architecture a

Recult

Qualitative results

Conclusion

References

Qualitative results

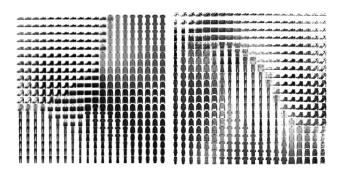


Figure: Images generated using a grid of the latent space (Sink1 left and GAN right)

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

Louis Geist, Valentin Gerard

Introduction

Related wor

Learning Generative Models with Sinkhorn Divergence

Generative Networks

NA selection

Architecture an

parameters

Qualitative result

Qualitative results

Ouantitative results

Conclusion

Reference

Quantitative results on Fashion MNIST

${\sf FashionMNIST}$	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 $_{\varepsilon}$) and the GAN on the FashionMNIST dataset.

Louis Geist, Valentin Gerard

Introduction

Learning Generative Models with Sinkhorn Divergences

Generative Adversia

IVELWORKS

Method

Architecture an parameters

Qualitative results

Quantitative results

Conclusion

References

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?

based models Inception Score. 2018. arXiv: 1801.01973 [stat.ML]. and GANs [2] Louis Geist. Valentin Gerard

[1]

Eyal Betzalel et al. A Study on the Evaluation of Generative Models, 2022, arXiv: 2206,10935 [cs.LG].

Shane Barratt and Rishi Sharma. A Note on the

Sinkhorn

[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.

References

[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:

Louis Geist. Valentin Gerard

References

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.

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

Louis Geist, Valentin Gerard

- [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

Louis Geist, Valentin Gerard

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

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

with the set of admissible couplings :

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

Sinkhorn loss between μ and ν

Louis Geist, Valentin Gerard

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

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