Learning distributions : Generative Adversarial Networks and Autoencoder

Titouan Vayer

September 12, 2018

Overview

Introduction : learning a probability distribution

Introduction and motivations

Not classif but Unsupervised learning , Oversampling, Learning words,Text translation into images

semi-supervised learning: features from the Discriminator or inference net could improve performance of classifiers when limited labeled data is available;

a conditional generative model $p(x|c)$ can be obtained by adding c as the input to both the Generator and the Discriminator.

https://blog.statsbot.co/generative-adversarial-networks-gans-engine-and-applications-f96291965b

Blotzman machine,log likehood

sinon ça peut aussi faire des chats

Learning a probability measure

The aim of generative models is to learn a data distribution $p(x)$, $x \sim \mathcal{D}$ where $\mathcal{D}$ is the "data". The underlying paradigm is Unsupervised Learning.

Some applications of generative models :

- Text to image synthesis [RAY$^+$16]
- Text to speech, Image and content generation [GDG$^+$15]
- Data augmentation [ASE17]
- Future simulation, DeepFake [CGZE18]

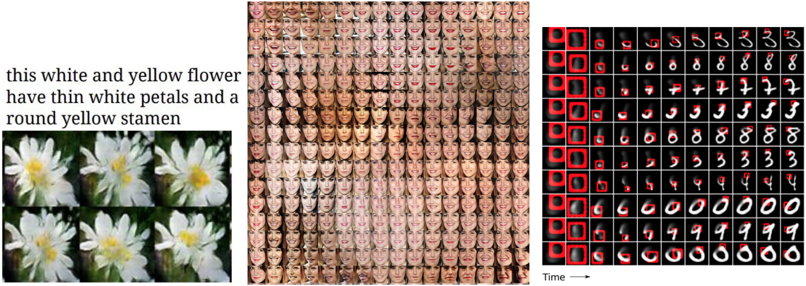


Figure 1: [RAY$^+$16, ASE17, GDG$^+$15]

Original GAN [GPM+14]

The adverserial Setting

Faire schema de deux réseaux

Problem setting

$$\max_D \min_G \mathop{\mathbb{E}}_{x \sim P_r}[\log(D(x))] + \mathop{\mathbb{E}}_{z \sim P_Z}[\log(1 - D(G(z)))]$$

Implementation of GAN

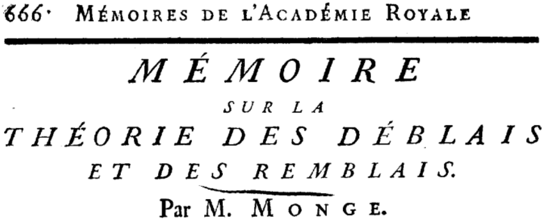
Downside

Multimodal

Some enlightening response [AB17]

Wasserstein GAN's

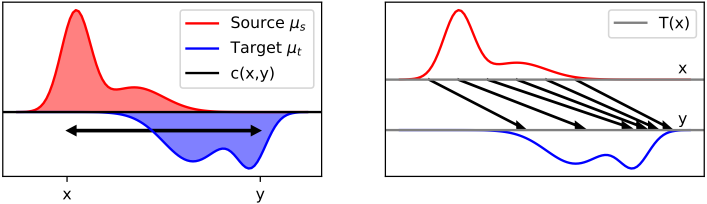
The origins of optimal transport


666· MÉMOIRES DE L'ACADÉMIE ROYALE

MÉMOIRE
SUR LA
THÉORIE DES DÉBLAIS
ET DES REMBLAIS.
Par M. MONGE.

Problem

- How to move dirt from one place (déblais) to another (remblais) while minimizing the effort ?
- Find a mapping T between the two distributions of mass (transport).
- Optimize with respect to a displacement cost $c(x, y)$ (optimal).

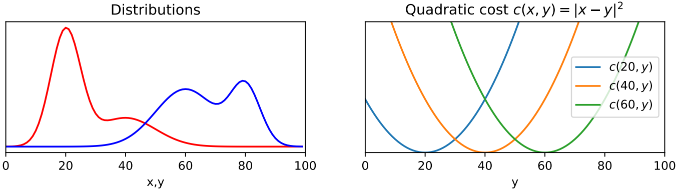


Problem

- How to move dirt from one place (déblais) to another (remblais) while minimizing the effort ?
- Find a mapping T between the two distributions of mass (transport).
- Optimize with respect to a displacement cost $c(x, y)$ (optimal).

Optimal transport (Monge formulation)

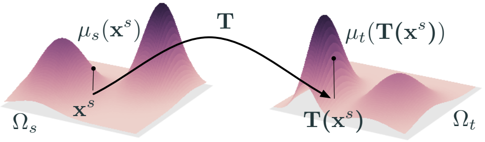
- Mathematical tools aiming at comparing distributions



- Probability measures μ_s and μ_t on Ω_s, Ω_t with a cost function $d : \Omega_s \times \Omega_t \to \mathbb{R}^+$.
- The Monge formulation aim at finding a mapping $T : \Omega_s \to \Omega_t$

$$\inf_{T\#\mu_s=\mu_t} \int_{\Omega_s} d(x, T(x))\mu_s(x)dx \qquad (1)$$

What is $T\#\mu_s = \mu_t$?



- $T\#$ is the so called push forward operator
- it transfers measures from one space Ω_s to another space Ω_t
- it is equivalent to:

$$\mu_t(A) = \mu_s(T^{-1}(A))$$

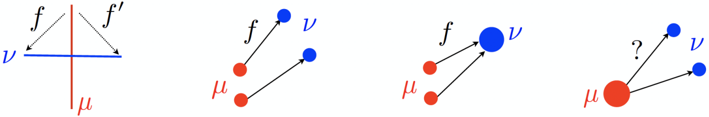
- consists simply in moving the positions of all the points in the support of the measure : for $\mu_s = \sum_{i=1}^n a_i \delta_{x_i}$

$$T\#\mu_s = \sum_{i=1}^n a_i \delta_{T(x_i)}$$

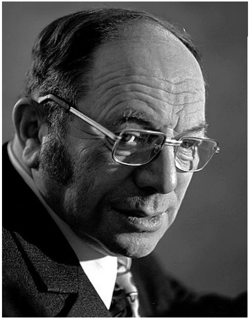
Non-existence Non-uniqueness

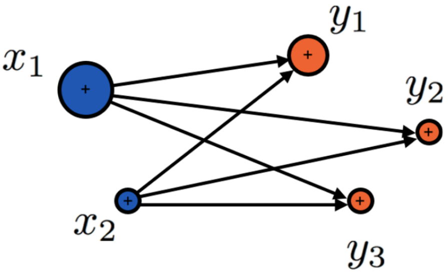
Solving for this push-forward operator is a non-convex optimization problem,

- for which existence is not guaranteed,
- nor unicity



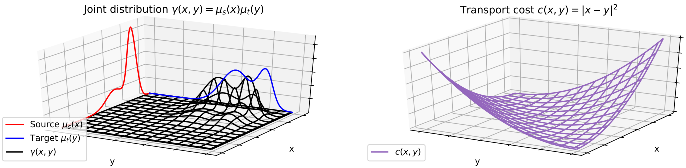
Kantorovich relaxation





- Leonid Kantorovich (1912–1986), Economy nobelist in 1975, proposed a different formulation of the problem
- with applications mainly for ressource allocation problems

Kantorovich relaxation

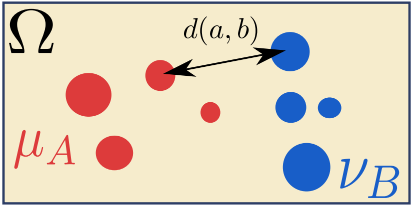


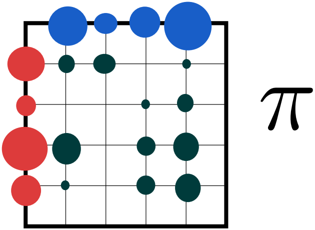
$\mu_s = \sum_{i=1}^{n} a_i \delta_{x_i}$ and $\mu_t = \sum_{j=1}^{m} b_j \delta_{y_j}$ on a commun ground space equipped with a distance

- The Kantorovich formulation seeks for a probabilistic coupling $\pi \in \Pi(\mu_s \times \mu_t)$ between μ_s and μ_t.
- π is a joint probability measure with prescribed marginals μ_s and μ_t.
- Computes the Wasserstein distance :

$$\mathcal{W}_p(\mu_s, \mu_t) = \left(\min_{\pi \in \Pi(\mu_s, \mu_t)} \sum_{i,j} d(x_i, y_j)^q \pi_{i,j} \right)^{\frac{1}{p}} \tag{2}$$

Probabilistic couplings





Properties of Wasserstein distance

Optimal transport in Machine Learning

Numerous applications for the Wasserstein distance :

- Learning with Wasserstein Loss [FZM+15]
- Wasserstein GAN's [ACB17]
- Domain Adaptation [CFTR17]
- Image colorization [FPPA14], Dictionary Learning [RCP16] ...

Problem Setting [ACB17]

$$\min_{\theta \in \Theta} \mathcal{W}_1(\mathbb{P}_r, \mathbb{P}_\theta)$$

idea : $\mathbb{P}_\theta = g_\theta(z), z \sim Z$, g_θ is a neural network.

$$\leftrightarrow \max_{w \in \mathcal{W}} \min_{\theta \in \Theta} \mathop{\mathbb{E}}_{x \sim P_r}[f_w(x)] - \mathop{\mathbb{E}}_{z \sim P_Z}[f_w(g_\theta(z))]$$

$f_w = D, g_\theta = G$

Implementation of WGAN

What's more ? [GAA+17]

Autoencoder

Variationnal autoencoder

- Martín Arjovsky and Léon Bottou, Towards principled methods for training generative adversarial networks, CoRR **abs/1701.04862** (2017).
- Martín Arjovsky, Soumith Chintala, and Léon Bottou, Wasserstein GAN, CoRR **abs/1701.07875** (2017).
- A. Antoniou, A. Storkey, and H. Edwards, Data Augmentation Generative Adversarial Networks, ArXiv e-prints (2017).
- Nicolas Courty, Rémi Flamary, Devis Tuia, and Alain Rakotomamonjy, Optimal transport for domain adaptation, IEEETPAMI **39** (2017), no. 9, 1853–1865.
- C. Chan, S. Ginosar, T. Zhou, and A. A. Efros, Everybody Dance Now, ArXiv e-prints (2018).

References ii

Sira Ferradans, Nicolas Papadakis, Gabriel Peyré, and Jean-François Aujol, Regularized discrete optimal transport, SIAM Journal on Imaging Sciences **7** (2014), no. 3, 1853–1882.

C. Frogner, C. Zhang, H. Mobahi, M. Araya-Polo, and T. Poggio, Learning with a Wasserstein Loss, ArXiv e-prints (2015).

Ishaan Gulrajani, Faruk Ahmed, Martín Arjovsky, Vincent Dumoulin, and Aaron C. Courville, Improved training of wasserstein gans, CoRR **abs/1704.00028** (2017).

K. Gregor, I. Danihelka, A. Graves, D. Jimenez Rezende, and D. Wierstra, DRAW: A Recurrent Neural Network For Image Generation, ArXiv e-prints (2015).

- Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio, Generative adversarial nets, Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada, 2014, pp. 2672–2680.
- Scott E. Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran, Bernt Schiele, and Honglak Lee, Generative adversarial text to image synthesis, CoRR **abs/1605.05396** (2016).
- Antoine Rolet, Marco Cuturi, and Gabriel Peyré, Fast dictionary learning with a smoothed wasserstein loss, Proceedings of the 19th International Conference on Artificial Intelligence and Statistics (Cadiz, Spain) (Arthur Gretton and Christian C. Robert, eds.), Proceedings of Machine Learning Research, vol. 51, PMLR, 09–11 May 2016, pp. 630–638.