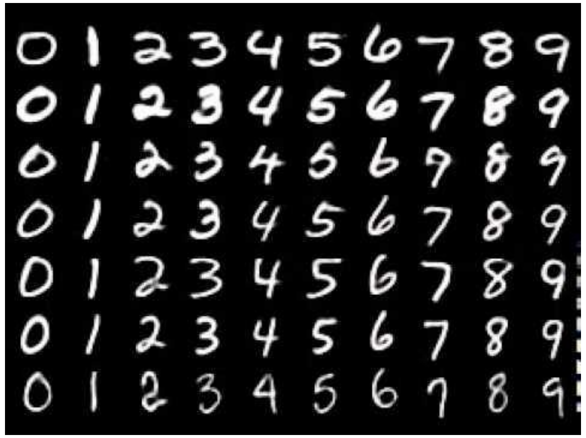
Wasserstein GAN

Objectif d'un GAN



$\Longleftrightarrow$

$$P_r : R^{100} \longmapsto R^{28\times 28}$$

P_r suit une distribution de probabilité D

Le but d'un GAN est d'approcher la distribution D

La distance de Wasserstein

Soit μ et ν deux distributions de probabilités sur deux ensembles X et Y

On pose :
$c : X \times Y \longmapsto [0, 1]$ une fonction donnant le coût de déplacement du point x au point y
$\gamma : X \times Y \longmapsto [0, 1]$ la quantité à transférer du point x au point y

γ doit respecter les propriétés suivantes :

$$\int_X \gamma(x, y)\mathrm{d}x = \nu(y)$$
$$\int_Y \gamma(x, y)\mathrm{d}y = \mu(x)$$

Autrement dit, γ est une loi de marginales μ et ν

La distance de Wasserstein

On comprend alors que le coût total de déplacement en suivant la stratégie γ est :

$$\int_{X\times Y} c(x,y)\gamma(x,y)\mathrm{d}x\mathrm{d}y = \int_{X\times Y} c(x,y)\mathrm{d}\gamma(x,y)$$

Ainsi, on peut définir la distance d'une distribution à une autre par

$$\boxed{\inf_{\gamma\in\Pi(\mu,\nu)} \int_{X\times Y} c(x,y)\mathrm{d}\gamma(x,y)}$$

La distance de Wasserstein adaptée pour les GANs

Le générateur du GAN suit une distribution de probabilité notée P_θ (où θ représente l'ensemble des poids du générateur)

On choisit $c : x, y \longmapsto |x - y|$

On peut donc définir la distance utilisée par le GAN :

$$\int_{X \times Y} c(x,y) \mathrm{d}\gamma(x,y) = \inf_{\gamma \in \Pi(P_r, P_\theta)} E_{(x,y)\sim\gamma}(|x-y|)$$

Le théorème de Kantorovich-Rubinstein

$$inf_{\gamma \in \Pi(P_r, P_\theta)} E_{(x,y)\sim\gamma}(|x-y|) = sup_{||f|| \leq 1} E_{x\sim P_r}(f(x)) - E_{x\sim P_\theta}(f(x))$$

Fonctionnement du WGAN

Algorithm 1 WGAN, our proposed algorithm. All experiments in the paper used the default values $\alpha = 0.00005$, $c = 0.01$, $m = 64$, $n_{\text{critic}} = 5$.

Require: : α, the learning rate. c, the clipping parameter. m, the batch size. n_{critic}, the number of iterations of the critic per generator iteration.

Require: : w_0, initial critic parameters. θ_0, initial generator's parameters.

while θ has not converged **do**

  for $t = 0, ..., n_{\text{critic}}$ **do**

    Sample $\{x^{(i)}\}_{i=1}^m \sim \mathbb{P}_r$ a batch from the real data.

    Sample $\{z^{(i)}\}_{i=1}^m \sim p(z)$ a batch of prior samples.

    $g_w \leftarrow \nabla_w \left[\frac{1}{m}\sum_{i=1}^m f_w(x^{(i)}) - \frac{1}{m}\sum_{i=1}^m f_w(g_\theta(z^{(i)}))\right]$

    $w \leftarrow w + \alpha \cdot \text{RMSProp}(w, g_w)$

    $w \leftarrow \text{clip}(w, -c, c)$

  end for

  Sample $\{z^{(i)}\}_{i=1}^m \sim p(z)$ a batch of prior samples.

  $g_\theta \leftarrow -\nabla_\theta \frac{1}{m}\sum_{i=1}^m f_w(g_\theta(z^{(i)}))$

  $\theta \leftarrow \theta - \alpha \cdot \text{RMSProp}(\theta, g_\theta)$

end while