High quality vs Diversity using GANs on MNIST

Evan Azoulay, Alejandro Jorba, Aziz Agrebi

Introduction ici

Un peu de texte

- Un item

Encore un peu de texte, suivi d'une équation

$$E = mc^2$$

1ere partie ici

Un peu de texte

- Un item

Encore un peu de texte, suivi d'une équation

$$E = mc^2$$

How WGAN Improves Vanilla GANs

Wasserstein GAN (WGAN)

- **Wasserstein Distance** : Replaces JS divergence with Wasserstein distance, providing a more stable measure of distribution difference.
- **Training Stability** : Smoother gradients from Wasserstein distance improve stability during training.

Wasserstein Distance Formula

$$W(P,Q) = \inf_{\gamma \in \Pi(P,Q)} \mathbb{E}_{(x,y)\sim\gamma}[\|x-y\|]$$

- Represents the minimum effort required to align the generated data distribution with the real data distribution.

Using the Kantorovich-Rubinstein duality :

$$W(P_r, P_\theta) = \sup_{\|f\|_L \leq 1} \mathbb{E}_{x\sim P_r}[f(x)] - \mathbb{E}_{x\sim P_\theta}[f(x)]$$

WGAN Loss Function and Architecture Changes

WGAN Loss Functions

- **Discriminator Loss** : Approximates the Wasserstein distance by maximizing the difference in the discriminator's output between real and generated samples.

$$\mathcal{L}_D = -\mathbb{E}_{x \sim p_{\text{data}}} [D(x)] + \mathbb{E}_{z \sim p_z} [D(G(z))]$$

- **Generator Loss** : Minimizes the discriminator's score for generated samples, encouraging the generator to produce data closer to the real distribution.

$$\mathcal{L}_G = -\mathbb{E}_{z \sim p_z} [D(G(z))]$$

Architecture Changes

- Sigmoid activation in the discriminator is removed resulting in outputs values in $(-\infty, \infty)$ instead of binary classification.
- Weight clipping is applied to enforce the 1-Lipschitz constraint.

WGAN-GP : Adding Gradient Penalty

Gradient Penalty

- To address the limitations of weight clipping in WGAN, a gradient penalty term is introduced to enforce the 1-Lipschitz constraint more effectively.
- This penalty improves convergence stability without the downsides of clipping.

Final WGAN-GP Loss Function :

$$L = \underbrace{\mathbb{E}_{\tilde{x}\sim P_G}[D(\tilde{x})] - \mathbb{E}_{x\sim P_r}[D(x)]}_{\text{Original WGAN Discriminator loss}} + \lambda \underbrace{\mathbb{E}_{\hat{x}\sim P_{\hat{x}}}\left[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2\right]}_{\text{Gradient penalty}}$$

3eme partie ici

Un peu de texte

- Un item

Encore un peu de texte, suivi d'une équation

$$E = mc^2$$

Conclusion ici

Un peu de texte

- Un item

Encore un peu de texte, suivi d'une équation

$$E = mc^2$$

References I

- M. Arjovsky, S. Chintala, and L. Bottou, *Wasserstein GAN*, arXiv preprint arXiv :1701.07875, 2017.
- I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, *Improved Training of Wasserstein GANs*, arXiv preprint arXiv :1704.00028, 2017.
- T. Issenhuth, U. Tanielian, D. Picard, and J. Mary, *Latent reweighting, an almost free improvement for GANs*, arXiv preprint arXiv :2110.09803, 2021. https://arxiv.org/abs/2110.09803