Your GAN is secretly an Energy-Based Model

Principles of Applied Statistics project, Skoltech

Alsu Vakhitova   Saveliy Galochkin   Anton Zubekhin

21.12.21

Problem statement

How can one improve the quality of GAN samples?

- Discriminator Rejection Sampling
- Metropolis-Hastings GAN
- Discriminator Optimal Transport

Drawbacks: inefficient or implying mode collapse
Idea: Consider GAN as Energy-Based Model and sample from latent space

GAN as an Energy-Based Model

GAN

Original data: $(X_1, \ldots, X_N) \in \mathbb{R}^{N \times d}, \ X_i \sim p_d$
Generator $G: \ \mathcal{Z} \to X, \, z \in \mathcal{Z} : z \sim p_z$
Discriminator $D: \ X \to \mathbb{R}$

EBM

State space: χ
Energy: $E: \ \chi \to \mathbb{R}$
Boltzmann distribution: $p(x) = e^{-E(x)}/Z, \ x \in \chi$
Normalizing constant: $Z = \int e^{-E(x)} dx$

GAN as an Energy-Based Model

Trained GAN with generator distribution p_g, we assume $G(z)$ is imperfect

Discriminator is almost optimal: $D(x) \approx \dfrac{p_d(x)}{p_d(x) + p_g(x)}$

Logit of $D(x)$: $d(x)$, $D(x) = \sigma(d(x))$

$$D(x) \approx \frac{1}{1 + \exp(-d(x))}$$

Energy-Based Model: $p_d^* = p_g(x)e^{d(x)}/Z_0$

If $D = D^*$, D^* - optimal discriminator, then $p_d^* = p_d$

GAN as an Energy-Based Model

Main Theorem

For data generating distribution p_d, generator distribution p_g with generator $G : \mathcal{Z} \to X$, where latent space $\mathcal{Z}$ has prior distribution $p_0(z)$, we define Boltzmann distribution $p_d^* = e^{\log p_d(x) + d(x)}/Z_0$.
For discriminator $D(x)$ its logit is $d(x) : \ D(x) = \sigma(d(x))$. If p_g and p_d have the same support, then defining energy function $E(z) = -\log p_0(z) - d(G(z))$ and its Boltzmann distribution $p_t(z) = e^{-E(z)}/Z$ we have:

- $p_d^* = p_d$ when D is the optimal discriminator
- If we sample $z \sim p_t$, and $x = G(z)$, then $x \sim p_d^*$

Note: now we can sample from the above Boltzmann distribution with some MCMC sampler

GAN as an Energy-Based Model

Sampling for WGAN

Wasserstein GANs (WGANs) target Kantorovich-Wasserstein distance, so its objectives are

$$L_D = \mathbb{E}_{p_g}\left[D(x)\right] - \mathbb{E}_{p_d}\left[D(x)\right]$$
$$L_G = -\mathbb{E}_{p_0}\left[D\left(G(x)\right)\right]$$

where D has to be K-Lipshitz function

Note: WGAN leads to EBM with $E(z) = -\log p_0(z) - D\left(G(z)\right)$

Mode collapse issue

If G cannot recover some of the modes in p_d, p_d^* also cannot
Idea: add some Gaussian noise $z' \sim \mathcal{N}(0,1)$
New generator: $G(z,z') = G(z) + \varepsilon z'$

Sampling algorithm

Algorithm Discriminator Driven Langevin Sampling

Input: $N \in \mathbb{N}_+,\ \varepsilon > 0$

Output: Latent code $z_N \sim p_t(z)$

Sample $z_0 \sim p_0(z)$

while i < N **do**

$n_i \sim \mathcal{N}(0,1)$

$z_{i+1} = z_i - \frac{\varepsilon}{2} \nabla_z E(z)\Big|_{z_i} + \sqrt{\varepsilon} n_i$

$i = i + 1$

end

Experiments

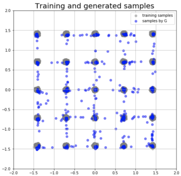
For experiments we took

- 2d mixture of 25 Gaussians
- Swiss roll dataset
- CIFAR10

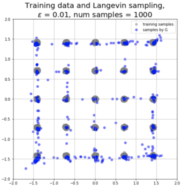
For synthetic data we used custom WGAN, for image dataset a pretrained SNGAN was used

Experiments with synthetic data

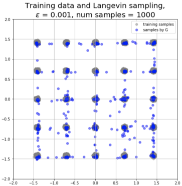
2d mixture of 25 Gaussians generation



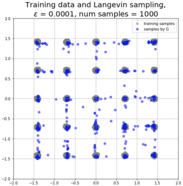
Vanilla GAN
EMD =
0.04412



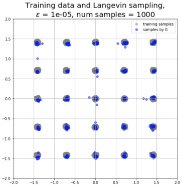
$\varepsilon = 10^{-2}$
EMD =
0.24396



$\varepsilon = 10^{-3}$
EMD =
0.07591



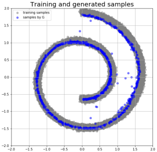
$\varepsilon = 10^{-4}$
EMD =
0.06467



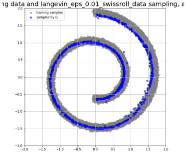
$\varepsilon = 10^{-5}$
EMD =
0.03502

Experiments with synthetic data

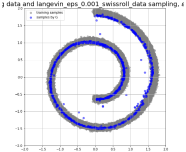
Swiss Roll generation



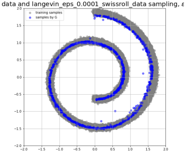
Vanilla GAN EMD = 0.076409



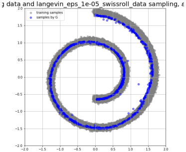
$\varepsilon = 10^{-2}$ EMD = 0.066459



$\varepsilon = 10^{-3}$ EMD = 0.056704



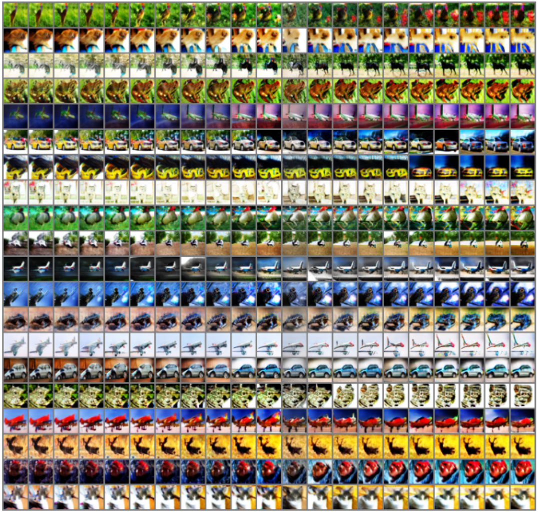
$\varepsilon = 10^{-4}$ EMD = 0.077774



$\varepsilon = 10^{-5}$ EMD = 0.06915

Experiments on CIFAR10

$\varepsilon = 0.01$, noise $\sim \mathcal{N}(0, 0.1)$, vanilla FID: 19.31, DDLS FID: 33.13



DDLS sampling for 500 iterations, plot each 25th

Conclusions

- On synthetic data the quality of DDLS samples is significantly better
- Tuning Langevin step size is important
- DDLS tends to oversaturate colors in images
- A more detailed study of DDLS image generation is required
- DDLS sampling is much less time consuming then retraining GAN

Contribution:
Alsu Vakhitova: Gaussian mixture experiments
Saveliy Galochkin: Swiss roll experiments
Anton Zubekhin: CIFAR10 experiments, presentation preparation

References

Your GAN is Secretly an Energy-based Model and You Should Use Discriminator Driven Latent Sampling:
arXiv:2003.06060

Thank you for your attention!