

1. Inherent Problems with GAN Models.

GAN 모델은 Generator 와 Discriminator 간의 min-max game 을 통해 학습이 이루어진다. 이때 Generator 은 실제 데이터의 분포와 '유사한' (똑같은 이 아니라) 분포의 sample 들을 출력하게 만드는 것이 목표입니다. 하지만 그러기 위해서 두 모델이 적절히 서로의 결과에 따라 학습을 해야하는데 하나라도 성능이 크게 뛰거나 무너지면 모델 자체가 실패로 돌아간다.

이러한 GAN 모델의 특징 때문에 내재적인 문제점, 또는 한계가 여러 존재하는데 대표적인 2 가지가 'Mode Collapse' 와 'Non-convergence' 이다. 조금 일반화하자면 Discriminator 가 실패하면 mode collapse 가 이루어지고 generator 이 실패하면 non-convergence 로 이루어진다.

1. Mode Collapse

Mode collapse 가 GAN 모델의 대표적인 한계점으로 꼽힌다. Mode collapse 에서 mode 란 generator 의 output distribution 을 말하며 collapse 했다는 것은 output 들이 일정하거나 매우 유사한 모습만 보일때를 말합니다. MNIST 와 같은 Balanced 0-9 dataset 의 데이터셋을 학습시키면 0-9 가 골고루 나타나야하는데 1 만 계속 출력되는 상태가 mode collapse 이다.

이러한 문제는 Discriminator 가 여러 번 분간을 제대로 못해 (50% 의 확률로 찍기) Generator 가 이러한 특징을 잡아내고 계속 그 특징들만을 내기 때문이다. 그러면 Discriminator 은 이미 너무 앞서간 Generator 을 못 따라가게 되고 Generator 은 자기의 loss 은 최소이기 때문에 똑 같은 sample 들만 내게 된다.

이러한 문제를 해결하는데 여러 방법이 있는데 learning rate 을 감소시키거나, encoder 들의 latent dimension 을 늘리는 등의 방법 등이 있다. 또한 GAN 모델을 학습시키면서 sample output 을 지속적으로 출력해 확인해 보는것도 중요하다. 이미 collapse 된 모델이 복구될 가능성이 희박하기 때문에 output 이 전부 다 일정하다면 학습을 멈추고 수정을 해야한다.



Image source :

<https://machinelearningmastery.com/practical-guide-to-gan-failure-modes/>

2. Non-convergence

Non-convergence 은 discriminator 와 generator 이 서로의 학습에 equilibrium 을 못 찾았기 때문에 나타난다. Discriminator 의 classifying neural network 가 초반에 generator 이 내는 가짜들의 특징을 너무 빨리 파악해 generator 가 제대로 학습하기도 전에 다 분간해서 맞추는 상황을 말합니다. 결국에는 generator 은 trash output 만 내게 되고 discriminator 은 다 맞추는 상황이 되고 GAN 모델은 실패로 돌아갑니다.

이러한 문제가 생기는 데에는 latent dimension 이 너무 작거나, learning rate 가 너무 높거나, kernel size 가 너무 낮아서와 같은 이유로 생깁니다. Non-convergence 또한 mode collapse 와 같이 회복될 가능성이 매우 낮기 때문에 중간에 output 들을 출력해 아래와 같은 결과들이 너무 오래 나타나면 학습을 중단하고 수정을 해야한다. 학습 초창기에는 아래와 같은 결과가 나오는 것이 당연한데 중반 또는 후반까지 가서도 아래와 같이 나오면 non-convergence 가 일어난 것이다.

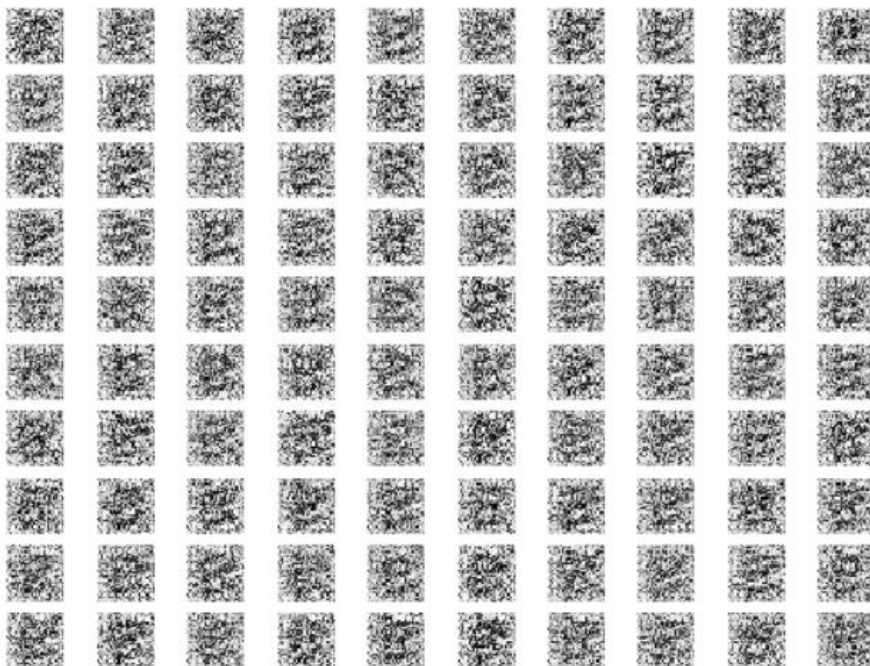


Image source :

<https://machinelearningmastery.com/practical-guide-to-gan-failure-modes/>

3. GAN Evaluation Metrics

GAN evaluation metric 으로 대표적인 것은 FID(Fréchet Inception Distance), IS(Inception Score), JSD(Jensen-Shannon distance) 등이 있습니다.

GAN 의 기초 모델들은 학습하는데 KL Divergence 을 이용했습니다. (물론 요즘은 KL Divergence 보다 Wasserstein distance 와 같은 더 유동적인 function 을 이용하곤 합니다). 이 KL Divergence 은 두 분포의 유사도를 측정하는 것인데 GAN 의 목표인 '유사한 분포를 생성한다' 와 일치합니다. JSD 또한 두 확률분포의 유사성을 판별하는 척도이기 때문에 JSD 를 통해 GAN 이 얼마나 잘 학습했는지 알 수 있습니다.

Resources :

<https://machinelearningmastery.com/practical-guide-to-gan-failure-modes/>

<https://wandb.ai/authors/DCGAN-ndb-test/reports/Measuring-Mode-Collapse-in-GANs--VmlldzoxNzg5MDk#:~:text=Mode%20collapse%20happens%20when%20the,keeps%20generating%20that%20one%20type.>

<https://github.com/yhlleo/GAN-Metrics>

https://en.wikipedia.org/wiki/Jensen%E2%80%93Shannon_divergence

<https://arxiv.org/pdf/1701.07875.pdf>