Generative Adversarial Networks

2022.02.24 김은희

Supervised learning vs Unsupervised learning

Supervised Learning

- Input : (data, label)
- Goal : learn a function to map x → y

Unsupervised Learning

- Input : data
- Goal : learn some underlying hidden structure of the data

Generative Model이란?

- Unsupervised learning의 일종
- Training data의 distribution 학습 $p_{data}(x)$ → 모델이 real data의 distribution 추정 $p_{model}(x)$
  - 예를 들어 '금발'을 나타내는 vector x_1이 많이 포함되어 있으면 $p_{data}(x_1)$의 값이 클 것
- $p_{data}(x)$와 비슷한 $p_{model}(x)$의 분포를 만든 다음, train data에는 존재하지 않지만 존재할 법한 데이터를 만들어내어 sampling할 수 있게 하자

Taxonomy of Generative Models

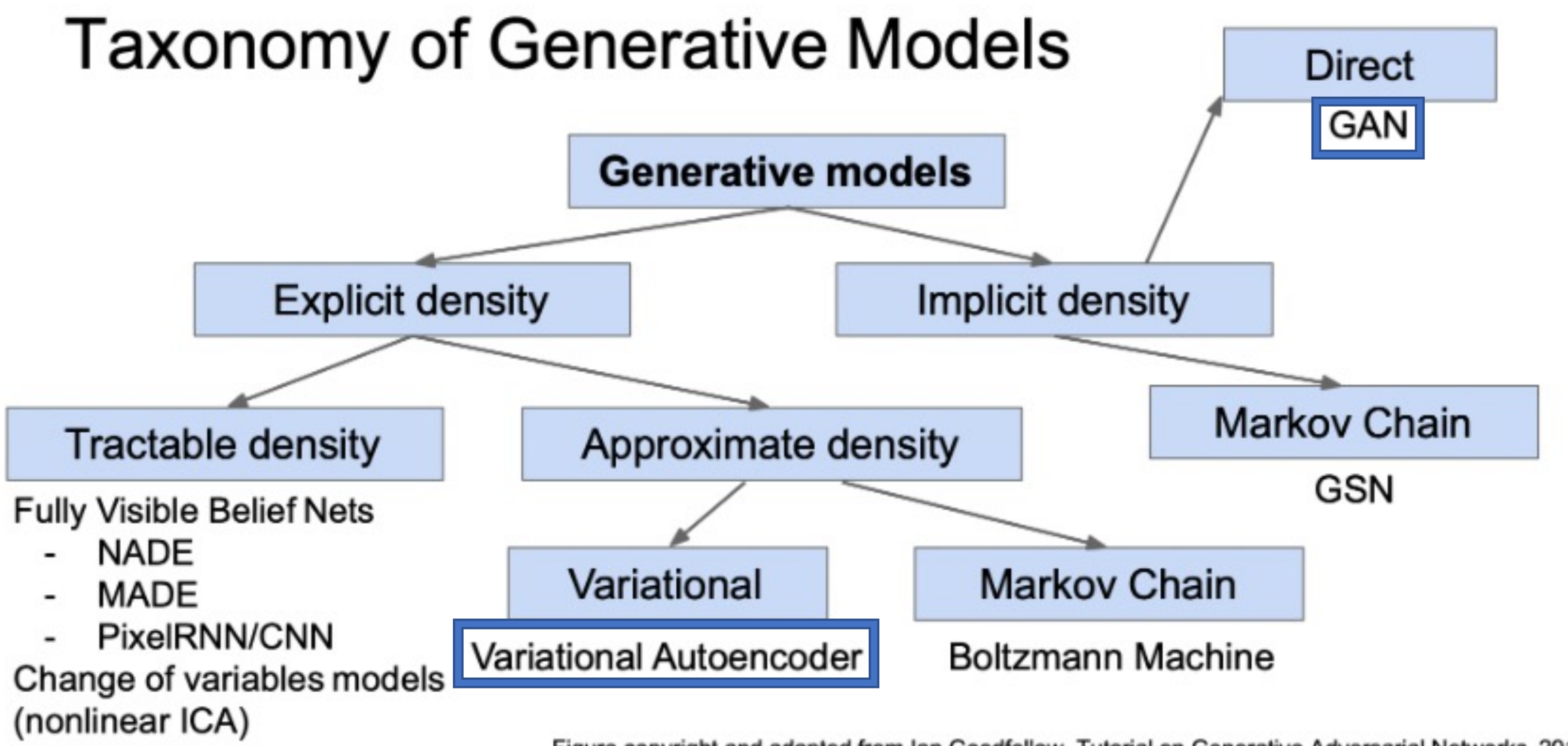


Figure copyright and adapted from Ian Goodfellow, Tutorial on Generative Adversarial Networks, 2017.

Explicit density vs Implicit density

VAE's Explicit density

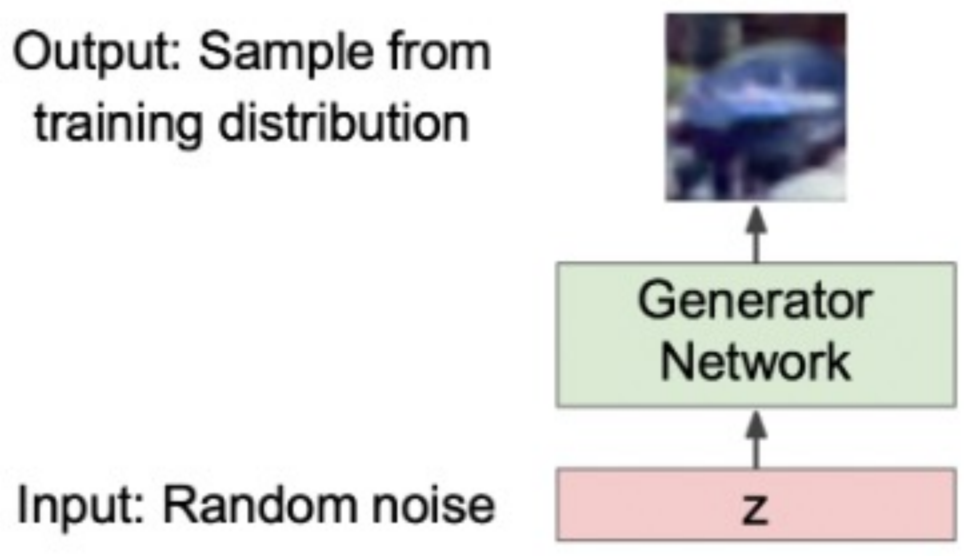
- Latent vector z를 이용해 $p_\theta(x)$의 probability density function을 정의
- $p_\theta(x) = \int p_\theta(z)p_\theta(x|z)dz$
- 직접적으로 maximize하는 것은 불가능하지만, 간접적으로 maximize할 수 있음

GAN's Implicit density

- $p_\theta(\mathrm{x})$를 명시적으로 모델링하지 않고 sampling할 수 없을까?
- adversarial한 두 player 간의 mini-max game 이라는 아이디어를 도입

Generative Adversarial Networks

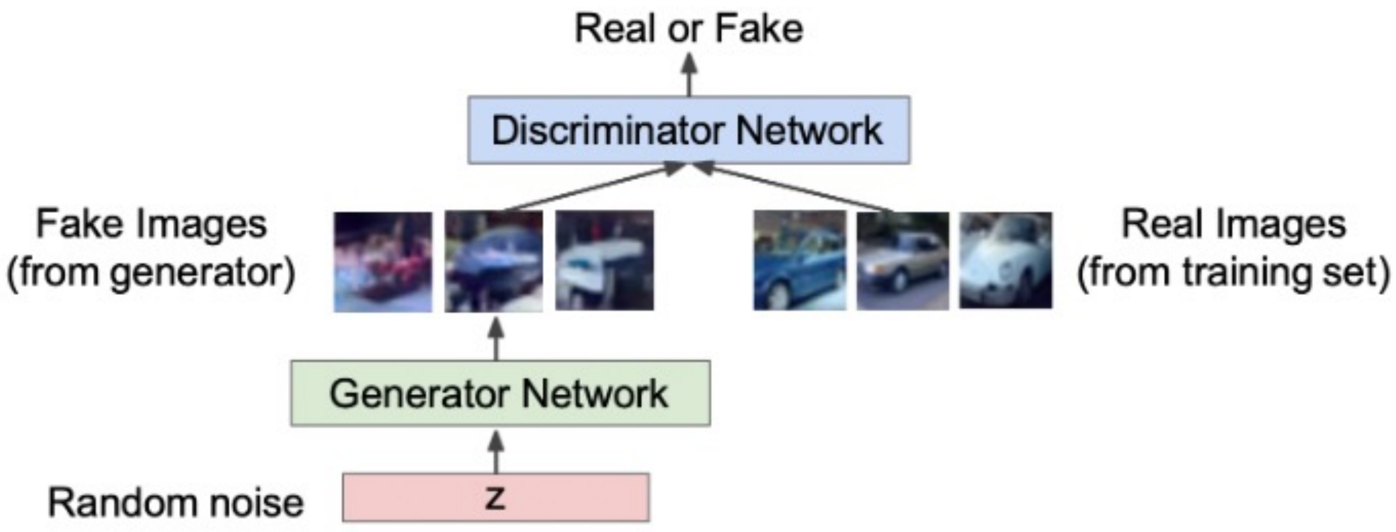
- Sample distribution을 따르는 sample → neural network → complex training distribution



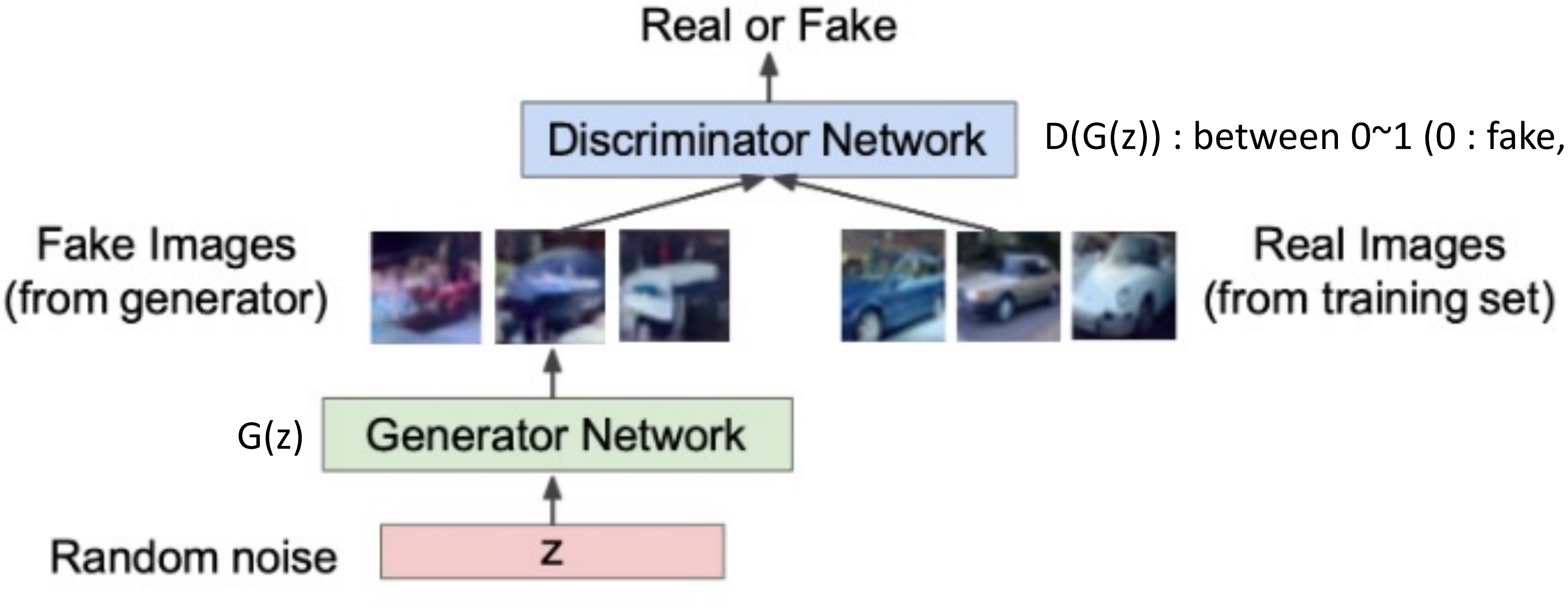
이 neural network의 parameter update를 위한 loss function은?

Intuition in GAN

- Generator : random latent vector를 input으로 받아 discriminator를 속일 정도로 정교한 image를 생성하는 역할
- Discriminator : real image와 generator가 만들어낸 fake image를 비교하는 역할



Intuition in GAN



G(z)

D(G(z)) : between 0~1 (0 : fake, 1 : real)

Training GANs: minimax game

Minimax objective function:

Discriminator outputs likelihood in (0,1) of real image

Gaussian에서 sampling한 random vector z

$$\min_{\theta_g}\max_{\theta_d}\left[\mathbb{E}_{x\sim p_{data}}\log \underbrace{D_{\theta_d}(x)}_{\text{Discriminator output for real data x}} + \mathbb{E}_{z\sim p(z)}\log(1-\underbrace{D_{\theta_d}(G_{\theta_g}(z))}_{\text{Discriminator output for generated fake data G(z)}})\right]$$

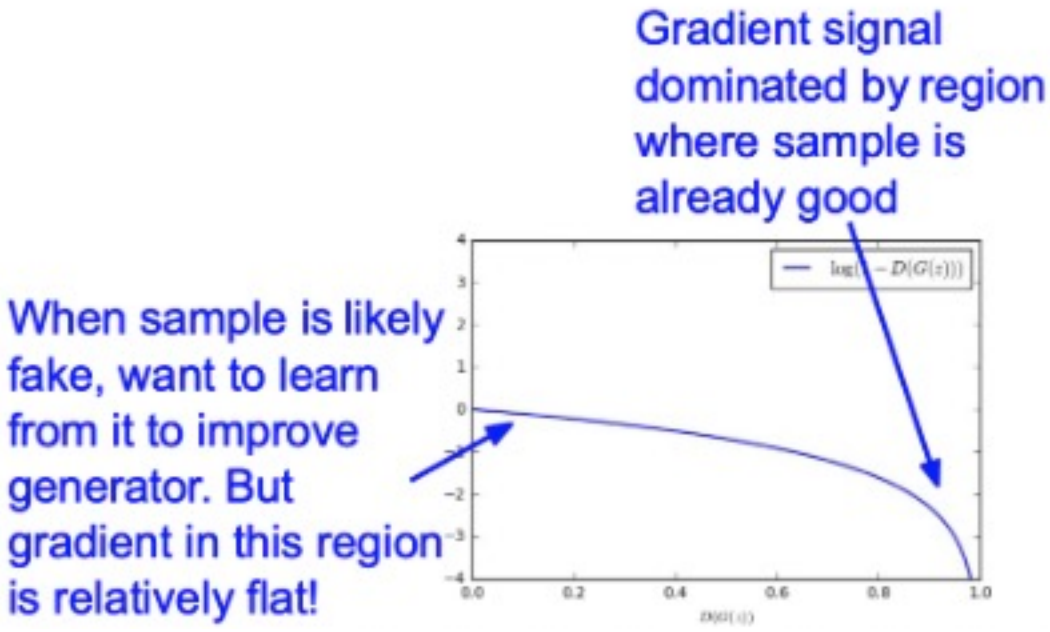
Generator : discriminator가 가짜를 진짜로 classification D(G(z))=1이 좋음 → minimize됨
Discriminator : 진짜를 진짜로, 가짜를 가짜로 classification D(x)=1, D(G(z))=0이 좋음 → maximize됨

같은 object function에 대해 generator는 이를 minimize, discriminator는 maximize하는 목표를 가짐

Training GANs

- Generator : gradient descent

$$\min_{\theta_g} \mathbb{E}_{z \sim p(z)} \log(1 - D_{\theta_d}(G_{\theta_g}(z)))$$



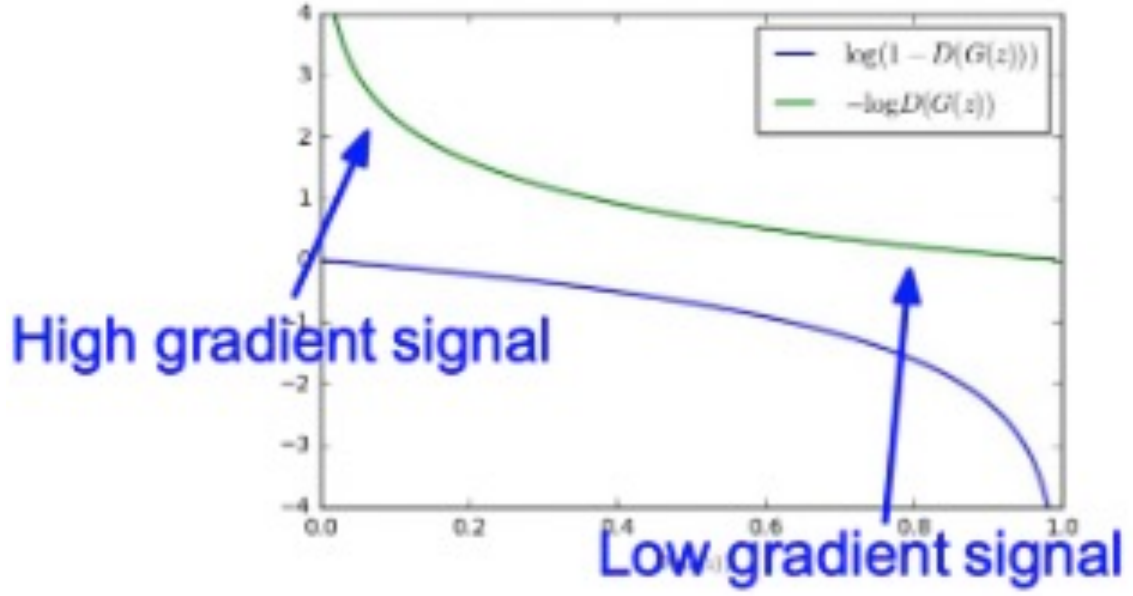
- Discriminator : gradient ascent

$$\max_{\theta_d} \left[\mathbb{E}_{x \sim p_{data}} \log D_{\theta_d}(x) + \mathbb{E}_{z \sim p(z)} \log(1 - D_{\theta_d}(G_{\theta_g}(z))) \right]$$

Training GAN’s

- 따라서, 실제로 generator를 학습시키기 위해서는 다음과 같이 objective function을 수정한 다음 gradient ascent를 적용한다.

$$\max_{\theta_g} \mathbb{E}_{z \sim p(z)} \log(D_{\theta_d}(G_{\theta_g}(z)))$$



Sources

- [#33.Lec] Basic of GAN - 딥러닝 홀로서기
- https://github.com/yunjey/pytorch-tutorial/blob/master/tutorials/03-advanced/generative_adversarial_network/main.py - L41-L57