

Generative Adversarial Networks(2014)

적대적 (adversarial) process를 통해 생성 모델을 추정하는 새로운 프레임워크를 제안

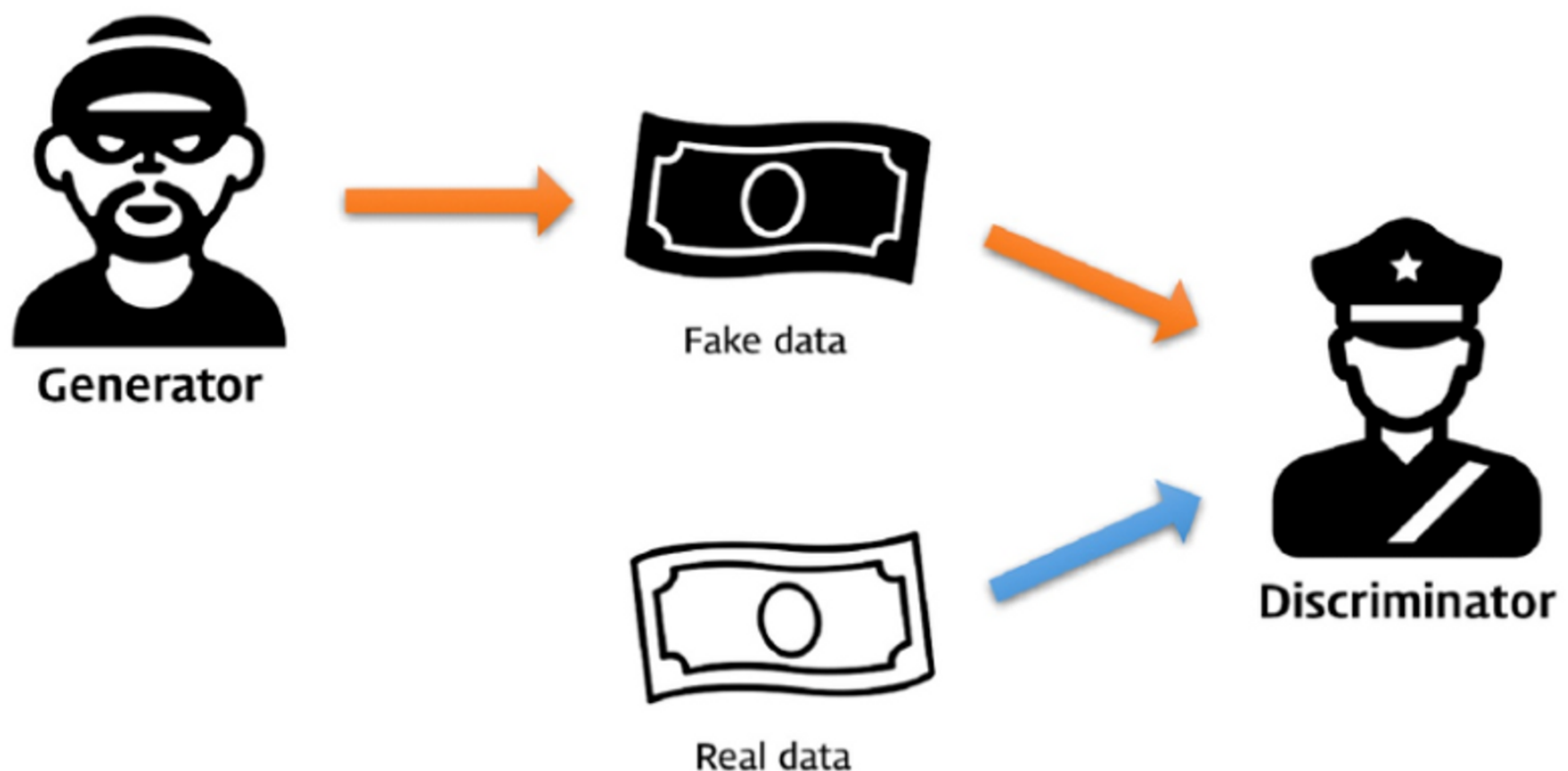
- G (generative model) : 데이터 분포를 포착하는 생성 모델 G
 - G의 훈련 절차 : D가 sample 데이터가 G로부터 나온 가짜 데이터와 실제 training 데이터를 판별하는데 실수를 할 확률을 최소화하는 것
 - D (discriminative model) : G가 아닌 훈련 데이터에서 샘플이 나올 확률을 추정하는 판별 모델 D
- G와 D를 동시 훈련

기존 모델들과의 차이?

- **discriminative model** : 고차원의 방대한 sensory input을 클래스 레이블에 mapping해서 판별하는 모델
 - maximum likelihood estimation에서의 많은 난해한 확률 연산들
 - generative context 에서 piecewise linear units의 이점 활용에 대한 어려움

제안된 모델

: 이 논문에서 소개되는 adversarial nets 프레임워크의 컨셉은 '경쟁'으로, discriminative model은 sample data가 G model이 생성해 낸 sample data인지, 실제 training data distribution인지 판별하는 것을 학습



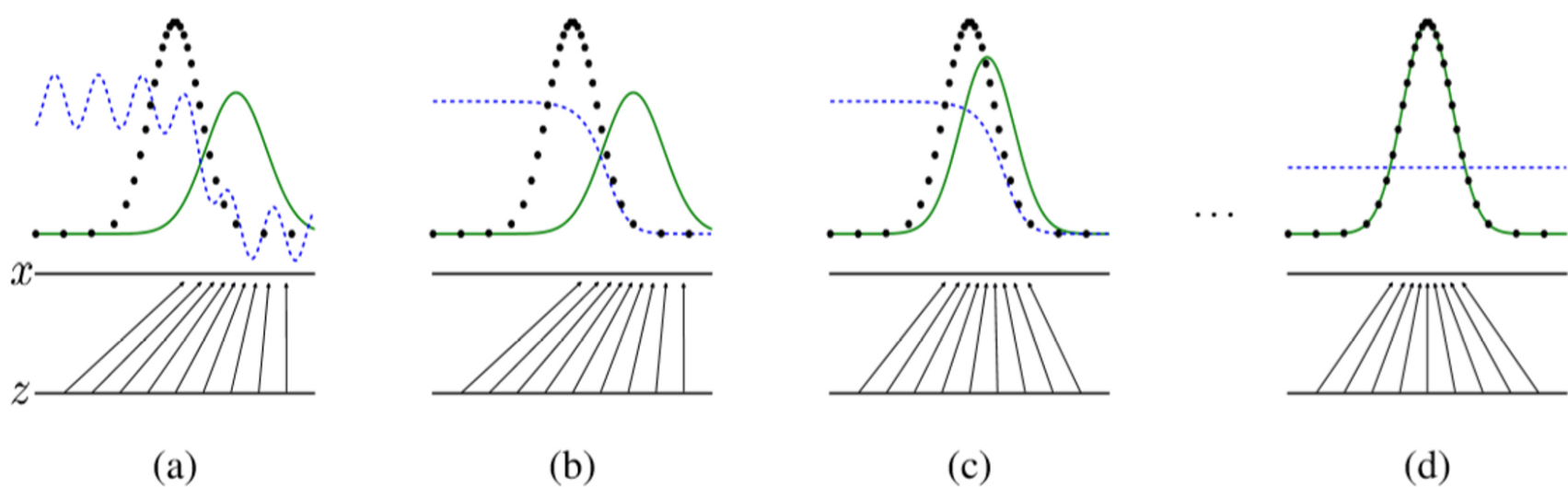
- Point : 적대적 학습을 하면서 '진짜같은 가짜'를 생성해내는 능력을 키우는 것
 - 경찰(분류 모델, 판별자) ↔ 위조지폐범(생성 모델, 생성자)
 - 경쟁을 반복하며 어느 순간 위조지폐범이 진짜와 다를 바 없는 위조지폐를 만들 수 있고 경찰이 위조지폐를 구별할 수 있는 확률 역시 50%로 수렴하게 됨으로써 경찰이 위조지폐와 실제 화폐를 구분할 수 없는 상태가 됨

Adversarial nets

1. G정의 : 입력 데이터 X에 대한 generator의 분포 p_g 를 학습하기 전에 input noise 변수 $p_z(x)$ 를 정의하고 $p_z(x)$ 를 data space에 매핑하는 $G(z; \theta_g)$
→ D가 correct label을 제대로 할당할 수 없을 정도로 학습 데이터셋과 유사한 데이터를 만드는 것
2. D정의 : 입력 데이터 x가 p_g 가 아니고 진짜 데이터 분포일 확률을 single scalar로 나타내는 함수로 multilayer perceptron $D(x; \theta_d)$
→ 학습 데이터셋과 G에서 만든 데이터에 correct label을 할당할 확률을 최대값으로 만드는 것

$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))]$$

- 첫번째 항: real data x를 discriminator에 넣었을 때 나오는 결과를 log취했을 때 얻는 기댓값
- 두번째 항: fake data z를 generator에 넣었을 때 나오는 결과를 discriminator에 넣었을 때 그 결과를 log(1-결과)했을 때 얻는 기댓값

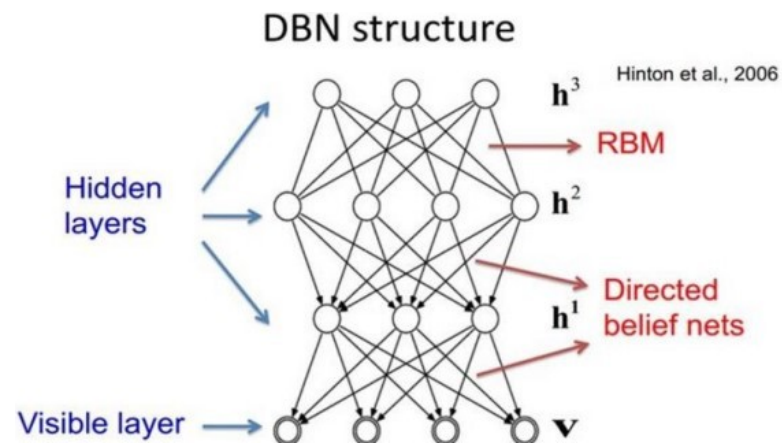


파란색 점선: discriminative distribution, 검은색 점선: data generating distribution(real), 녹색 실선: generative distribution(fake)

- (d): 이 과정의 반복의 결과로 real과 fake의 분포가 거의 비슷해져 구분할 수 없을 만큼 G가 학습을 하게 되고 결국, D가 이 둘을 구분할 수 없게 되어 확률을 1/2로 계산하게 됨
→ > 이 과정으로 진짜와 가짜 이미지를 구별할 수 없을 만한 데이터를 G가 생성해내고 이게 GAN의 최종 결과

Experiments

: DBN / Stacked CAE / Deep GSN / Adversarial Nets 비교



Evaluation

💡 generative 모델들은 실제 확률을 다루기 어려워 **마땅한 평가지표 X**

→ 생성된 sample에 Gaussian Parzen window 기반의 log-likelihood를 측정함으로써 P_g 에 따른 test set data 추정

- p_g 에 있는 test set data들의 확률은 G 로 생성된 데이터들에 Gaussian Parzen window를 맞추는 방식으로 측정하며 측정된 확률은 p_g 에 속한 log 확률로 변환
- Gaussian의 표준편차 σ 는 검증 데이터셋의 교차 검증을 통해 값 선정
- 확률이란?
 - 진짜 데이터인지, 생성된 데이터인지 구분하는 확률

Model	MNIST	TFD
DBN [3]	138 ± 2	1909 ± 66
Stacked CAE [3]	121 ± 1.6	2110 ± 50
Deep GSN [5]	214 ± 1.1	1890 ± 29
Adversarial nets	225 ± 2	2057 ± 26

- G 가 생성해낸 sample이 기존 존재 방법으로 만든 sample보다 좋다고 주장할 수 는 없지만, 더 나은 generative 모델과 경쟁할 수 있다 생각하며, adversarial framework의 잠재력 강조