생성적 적대 네트

Ian J. Goodfellow, Jean Pouget-Abadie* , Mehdi Mirza, Bing Xu, David Warde-Farley,
´ 셰르질 오자르 †, Aaron Courville, 요슈아 벤지오 ‡
컴퓨터 과학 및 운영 연구 대학 몬트리올 몬트리올 대학, QC H3C3J7
´

추상적인

우리는 적대적 프로세스를 통해 생성 모델을 추정하기 위한 새로운 프레임워크를 제안합니다. 여기서 생성 모델 G
데이터 분포를 포착하고 추정하는 판별 모델 D
G가 아닌 훈련 데이터에서 샘플이 나올 확률. G에 대한 훈련 절차는 D가 실수할 확률을 최대화하는 것입니다. 이것
프레임워크는 minimax 2인용 게임에 해당합니다. 임의의 공간에서
기능 G 및 D, 고유 솔루션이 존재하며 G는 훈련 데이터 복구
분포 및 D는 모든 곳에서 동일합니다. G와 D가 정의된 경우
다층 퍼셉트론에 의해 전체 시스템은 역전파로 훈련될 수 있습니다.
샘플을 훈련하거나 생성하는 동안 Markov 체인이나 펼쳐진 근사 추론 네트워크가 필요하지 않습니다. 실험 시연

의 질적 및 양적 평가를 통한 프레임워크의 잠재력
생성된 샘플.

1. 소개

딥 러닝의 약속은 확률을 나타내는 풍부한 계층적 모델[2]을 발견하는 것입니다.
자연 지능과 같은 인공 지능 응용 프로그램에서 발생하는 데이터 종류에 대한 분포
이미지, 음성을 포함하는 오디오 파형 및 자연어 말뭉치의 기호. 지금까지,
딥 러닝에서 가장 눈에 띄는 성공은 판별 모델과 관련되어 있습니다. 일반적으로
고차원의 풍부한 감각 입력을 클래스 레이블에 매핑합니다[14, 22]. 이러한 놀라운 성공은
주로 조각별 선형 단위를 사용하여 역전파 및 드롭아웃 알고리즘을 기반으로 합니다.
[19, 9, 10] 특히 잘 작동하는 기울기가 있습니다. 딥 생성 모델은
많은 다루기 힘든 확률적 계산을 근사화하는 것이 어렵기 때문에
최대 가능성 추정 및 관련 전략에서 발생하며 활용의 어려움으로 인해
생성 컨텍스트에서 조각별 선형 단위의 이점. 새로운 생성 모델을 제안합니다.[1]
이러한 어려움을 회피하는 추정 절차.

제안된 adversarial nets 프레임워크에서 생성 모델은 다음과 같이 적대적입니다.
표본이 모형 분포 또는 모형 분포에서 나온 것인지 판별하는 방법을 학습하는 판별 모형
데이터 배포. 생성 모델은 위조 팀과 유사하다고 생각할 수 있습니다.
위조 화폐를 만들어 들키지 않고 사용하려고 하는 반면, 차별적 모델은
위조 화폐를 탐지하는 경찰과 유사합니다. 이 게임의 경쟁 드라이브
위조품과 진품을 구별할 수 없을 때까지 두 팀 모두 방법을 개선합니다.
조항.

---

*´ †Jean Pouget-Abadie가 Ecole Polytechnique에서 Universite de Montreal을 방문하고 있습니다.
Sherjil Ozair는 인도 델리 공과 대학에서 Universite de Montreal을 방문 중입니다 . ‡ Yoshua Bengio는 CIFAR 선임 연구원
입니다.
1모든 코드 및 하이퍼파라미터는 http://www.github.com/goodfeli/adversarial에서 사용 가능

이 프레임워크는 다양한 모델 및 최적화 알고리즘에 대한 특정 훈련 알고리즘을 생성할 수 있습니다. 이 기사에서는 생성 모델이 다층 퍼셉트론을 통해 무작위 노이즈를 전달하여 샘플을 생성하고 판별 모델도 다층 퍼셉트론인 특수한 경우를 살펴봅니다. 우리는 이 특별한 경우를 적대적 네트라고 합니다. 이 경우 매우 성공적인 역전파 및 드롭아웃 알고리즘[17]만을 사용하여 두 모델을 모두 훈련할 수 있고 순방향 전파만 사용하여 생성 모델에서 샘플링할 수 있습니다. 근사 추론이나 마르코프 체인이 필요하지 않습니다.

2 관련 작업

잠재 변수가 있는 유향 그래픽 모델의 대안은 제한된 볼츠만 기계(RBM)[27, 16], 깊은 볼츠만 기계(DBM)[26] 및 다양한 변형과 같은 잠재 변수가 있는 무방향 그래픽 모델입니다. 이러한 모델 내의 상호 작용은 확률 변수의 모든 상태에 대한 전역 합계/적분에 의해 정규화된 비정규화된 잠재적 기능의 제품으로 표시됩니다. 이 양(분할 함수)과 그 기울기는 가장 사소한 경우를 제외하고는 모두 다루기 힘들지만 Markov Chain Monte Carlo(MCMC) 방법으로 추정할 수 있습니다. 믹싱은 MCMC에 의존하는 학습 알고리즘에 중요한 문제를 제기합니다[3, 5].

DBN(Deep Belief Networks)[16]은 단일 무방향 계층과 여러 방향 계층을 포함하는 하이브리드 모델입니다. 빠른 근사 계층별 훈련 기준이 존재하지만 DBN은 무향 및 유향 모델과 관련된 계산상의 어려움을 초래합니다.

점수 일치[18] 및 잡음 대비 추정(NCE)[13]과 같이 로그 가능성을 근사하거나 제한하지 않는 대체 기준도 제안되었습니다. 이 두 가지 모두 학습된 확률 밀도가 정규화 상수까지 분석적으로 지정되어야 합니다. 여러 층의 잠재 변수(DBN 및 DBM과 같은)가 있는 많은 흥미로운 생성 모델에서는 다루기 쉬운 비정규화 확률 밀도를 유도하는 것조차 불가능합니다. 잡음 제거 자동 인코더[30] 및 축소 자동 인코더와 같은 일부 모델에는 RBM에 적용된 점수 일치와 매우 유사한 학습 규칙이 있습니다. NCE에서는 이 작업에서와 같이 생성 모델에 적합하도록 판별 훈련 기준을 사용합니다. 그러나 별도의 판별 모델을 맞추는 대신 생성 모델 자체를 사용하여 고정된 노이즈 분포 샘플에서 생성된 데이터를 판별합니다.

NCE는 고정된 잡음 분포를 사용하기 때문에 모델이 관찰된 변수의 작은 하위 집합에 대해 대략적으로 정확한 분포를 학습한 후 학습 속도가 급격히 느려집니다.

마지막으로, 일부 기술은 확률 분포를 명시적으로 정의하는 것이 아니라 원하는 분포에서 샘플을 추출하도록 생성 기계를 훈련시키는 것을 포함합니다. 이 접근 방식은 이러한 기계가 역전파에 의해 훈련되도록 설계할 수 있다는 장점이 있습니다. 이 분야의 저명한 최근 작업에는 일반화된 잡음 제거 자동 인코더[4]를 확장하는 GSN(Generative Stochastic Network) 프레임워크[5]가 있습니다. 둘 다 매개변수화된 Markov 체인을 정의하는 것으로 볼 수 있습니다. 즉, 기계의 매개변수를 학습합니다 생성적 마르코프 사슬의 한 단계를 수행합니다. GSN과 비교하여 adversarial nets 프레임워크는 샘플링을 위해 Markov 체인이 필요하지 않습니다. adversarial nets는 생성 중에 피드백 루프가 필요하지 않기 때문에 조각별 선형 단위[19, 9, 10]를 더 잘 활용할 수 있어 역전파의 성능을 개선하지만 피드백 루프에서 사용할 때 무제한 활성화 문제가 있습니다. 역전파를 통해 생성 기계를 훈련시키는 보다 최근의 예는 자동 인코딩 변이 베이즈[20] 및 확률적 역전파[24]에 대한 최근 작업을 포함합니다.

3 적의 그물

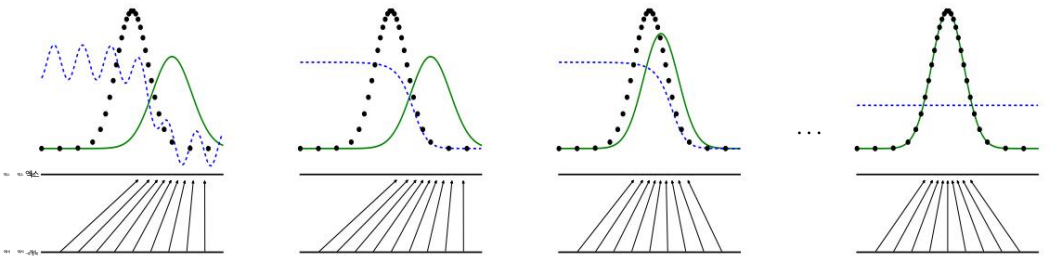
적대적 모델링 프레임워크는 모델이 모두 다층 퍼셉트론일 때 적용하기 가장 간단합니다. 데이터 x에 대한 생성기의 분포 pg를 학습하기 위해 입력 잡음 변수 pz(z)에 대한 사전 정의를 정의한 다음 데이터 공간에 대한 매핑을 G(z; θg)로 나타냅니다. 여기서 G는 매개변수 θg. 또한 단일 스칼라를 출력하는 두 번째 다층 퍼셉트론 D(x; θd) 를 정의합니다. D(x)는 x가 pg 가 아닌 데이터에서 나온 확률을 나타냅니다 . 훈련 예제와 G의 샘플 모두에 올바른 레이블을 할당할 확률을 최대화하기 위해 D를 훈련합니다. 동시에 log(1 D(G(z)))를 최소화하도록 G를 훈련합니다.

즉, D와 G는 값 함수 V(G, D)를 사용하여 다음 2인 미니맥스 게임을 합니다.

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))]. \quad (1)$$

다음 섹션에서, 우리는 기본적으로 훈련 기준이 G와 D에 충분한 용량, 즉 비모수적 한계가 주어졌을 때 데이터 생성 분포를 복구할 수 있음을 보여주는 adversarial nets의 이론적 분석을 제시합니다. 접근 방식에 대한 덜 형식적이고 교육적인 설명은 그림 1을 참조하십시오. 실제로, 우리는 반복적이고 수치적인 접근을 사용하여 게임을 구현해야 합니다. 훈련의 내부 루프에서 완료까지 D를 최적화하는 것은 계산적으로 금지되어 있으며 유한 데이터 세트에서는 과적합이 발생할 수 있습니다. 대신, 우리는 D를 최적화하는 k 단계와 G를 최적화하는 한 단계를 번갈아 가며 수행합니다. G가 충분히 천천히 변하는 한 D는 최적 솔루션 근처에서 유지됩니다. 이 전략은 SML/PCD[31, 29] 학습이 학습의 내부 루프의 일부로 Markov 체인에서 소진되는 것을 피하기 위해 한 학습 단계에서 다음 학습 단계로 Markov 체인의 샘플을 유지 관리하는 방식과 유사합니다. 절차는 공식적으로 알고리즘 1에 나와 있습니다.

실제로 방정식 1은 G가 잘 학습하기에 충분한 기울기를 제공하지 않을 수 있습니다. 학습 초기에 G가 좋지 않을 때 D는 훈련 데이터와 분명히 다르기 때문에 높은 신뢰도로 샘플을 거부할 수 있습니다. 이 경우 log(1 D(G(z)))가 포화됩니다. log(1 D(G(z)))를 최소화하기 위해 G를 훈련하는 대신 log D(G(z))를 최대화하기 위해 G를 훈련할 수 있습니다. 이 목적 함수는 G와 D의 역학의 동일한 고정 소수점을 가져오지만 학습 초기에 훨씬 더 강한 기울기를 제공합니다.



(ㅏ) (비) (씨) (디)

그림 1: 생성적 적대적 네트워크는 데이터 생성 분포(검정색, 점선)의 샘플을 생성 분포 pg(G)의 샘플과 구별하도록 판별 분포(D, 파란색, 파선)를 동시에 업데이트 하여 학습 됩니다 . (녹색, 실선). 아래쪽 수평선은 z 가 이 경우 균일하게 샘플링되는 영역입니다. 위의 수평선은 x 도메인의 일부입니다. 위쪽 화살표는 매핑 x = G(z) 가 변환된 샘플에 균일하지 않은 분포 pg 를 부과하는 방법을 보여줍니다. G는 밀도가 높은 영역에서 수축하고 pg 의 밀도가 낮은 영역에서 확장 합니다. (ㅏ)

수렴 근처에 있는 적대적 쌍을 고려하십시오. pg 는 pdata 와 유사 하고 D는 부분적으로 정확한 분류기입니다. (x) = (b) 알의 내부 루프에서 D는 데이터에서 샘플을 구별하도록 훈련되어 D pdata(x)+pg(x) 로 수렴합니다 . (c) G로 업데이트한 후 D의 기울기는 G(z) 가 데이터로 분류될 가능성이 더 높은 영역으로 흐르도록 안내했습니다. (d) 여러 단계의 훈련 후 G와 D에 충분한 용량이 있으면 pg = pdata 이기 때문에 둘 다 향상될 수 없는 지점에 도달합니다 . 판별자는 두 분포를 구별할 수 없습니다. 즉, D(x) = $\frac{1}{2}$.

4 이론적 결과

생성기 G는 z ~ pz 일 때 얻은 샘플 G(z)의 분포로 확률 분포 pg 를 암시적으로 정의합니다. 따라서 충분한 용량과 훈련 시간이 주어진다면 알고리즘 1이 pdata 의 좋은 추정기로 수렴되기를 바랍니다. 이 섹션의 결과는 비모수 설정에서 수행됩니다. 예를 들어 확률 밀도 함수 공간에서 수렴을 연구하여 무한 용량을 가진 모델을 나타냅니다.

이 minimax 게임이 pg = pdata 에 대한 전역 최적값을 갖는다는 것을 섹션 4.1에서 보여줄 것입니다 . 그런 다음 섹션 4.2에서 알고리즘 1이 식 1을 최적화하여 원하는 결과를 얻는다는 것을 보여줄 것입니다.

알고리즘 1 생성적 적대 네트워크의 Minibatch 확률적 경사 하강 훈련. 판별자에 적용할 단계 수 k는 초매개변수입니다. 우리는 실험에서 가장 저렴한 옵션인 k = 1을 사용했습니다. 훈련 반복 횟수에 대해 k 단계 에 대해 수행 • m 노이즈 샘플의 샘플 미니배치 {z (1) • m개 예의 샘플 미니배치 {x (1) pdata(x). • 확률적 기울기를 높여 판별자를 업데이트합니다.

, . . . , 노이즈 사전 pg(z) 에서 z (m)} . x
, . . . , (m)} 데이터 생성 분포

$$\nabla\theta d \frac{1}{\text{중}} \text{엑스엠 나는=1} \text{ h 로그 D x } ^{(나)} \text{ + 로그 1 - DG z } \qquad \text{(i)} \qquad .$$

끝내다

• m 노이즈 샘플의 샘플 미니배치{z (1) • 확률적 기울기를 , . . . , 노이즈 사전 pg(z) 에서 z (m)} . 내림으로써 생성기를 업데이트합니다.

$$\nabla\theta g \frac{1}{\text{중}} \text{엑스엠 나는=1} \text{ 로그 1 - DG z } \qquad \text{(i)} \qquad .$$

그라디언

트 기반 업데이트의 끝은 모든 표준 그라디언트 기반 학습 규칙을 사용할 수 있습니다. 우리는 실험에서 운동량을 사용했습니다.

4.1 pg = pdata 의 전역 최적성

우리는 먼저 주어진 생성기 G에 대한 최적의 판별자 D를 고려합니다.

명제 1. 고정된 G의 경우 최적 판별자 D는 다음과 같습니다.

$$D^* G(x) = \frac{\text{pdata(x)}}{\text{pdata(x) + pg(x)}} \qquad (2)$$

증거. 생성자 G가 주어지면 판별자 D에 대한 훈련 기준은 수량 V(G, D)를 최대화하는 것입니다.

$$V(G, D) = Z \text{ 엑스 pdata(x) log(D(x))dx + Z } \text{ pz(z) 로그(1 - D(g(z))) div}$$

$$= Z \text{ 엑스 pdata(x) log(D(x)) + pg(x) log(1 D(x))dx} \qquad \text{(삼)}$$

임의의 (a, b) ∈ R \ {0, 0}에 대해 함수 y → a log(y) + b log(1 y)는 [0, 1]에서 최대값을 얻습니다. $\frac{\text{ㅏ}}{\text{a+b}}$. 판별자는 Supp(pdata) ∪ Supp(pg) 외부에서 정의할 필요가 없습니다 . 증명을 마칩니다. □

D에 대한 훈련 목표는 조건부 확률 P(Y = y|x)를 추정하기 위한 로그 가능성을 최대화하는 것으로 해석될 수 있습니다. 여기서 Y는 x가 pdata (y = 1 사용)에서 또는 pg (사용 y = 0). Eq.의 미니맥스 게임. 1은 이제 다음과 같이 다시 공식화될 수 있습니다.

$$C(G) = \text{최대 } _{\text{디}} \text{ V(지, 디)}$$

$$= Ex \text{ pdata [log } D^* G(x)] + Ez \text{ pz [log(1 } D^* G(G(z)))] \qquad (4)$$

$$= Ex \text{ pdata [log } D^* G(x)] + Ex \text{ pg [log(1 } D^* G(x))]$$

$$= Ex \text{ pdata 로그 } \frac{\text{pdata(x)}}{\text{Pdata(x) + pg(x)}} + Ex \text{ pg log } \frac{\text{페이지(x)}}{\text{pdata(x) + pg(x)}}$$

정리 1. 가상 훈련 기준 C(G)의 전역 최소값은 pg = pdata 인 경우에만 달성됩니다 . 그 시점에서 C(G)는 값 - log 4를 달성합니다.

증거. pg = pdata의 경우, D * G(x) = 찾 $\frac{1}{2}$, (식 2를 고려). 따라서 Eq. 4 at D* G(x) = = log 4. 이것이 가능한 최상의 C(G) 값인지 확인하기 $\frac{1}{2}$ 우리 2 ,
기 C(G) = log + log는 pg에서만 pdata를 얻을 수 있습니다 $\frac{1}{2}$
관찰 하십시오.

$$Ex \quad pdata [\quad \log 2] + Ex \quad pg [\quad \log 2] = \quad \log 4$$

C(G) = V(D*에서 이 식을 빼면 G, G), 우리는 다음을 얻습니다:

$$C(G) = \quad \log(4) + KL\ pdata \quad \frac{pdata + 페이지}{2} \quad + KL\ 페이지 \quad \frac{pdata + 페이지}{2} \tag{5}$$

여기서 KL은 Kullback-Leibler 발산입니다. 우리는 이전 표현에서 모델의 분포와 데이터 생성 프로세스 사이의 Jensen-Shannon 다이버전스를 인식합니다.

$$C(G) = \quad \log(4) + 2\ JSD\ (pdata\ kpg\ ) \tag{6}$$

두 분포 사이의 Jensen-Shannon divergence는 항상 음이 아니고 0 = log(4)가 C(G)의 전역 최소값이고 두 분포가 pdata, 즉 데이터 생성 과정을 완벽하게 복제하는 생성 모델. 같을 때만 C * 가 유일한 해가 pg 임을 보여주었습니다. =

□

4.2 알고리즘 1의 수렴

명제 2. G와 D의 용량이 충분하고 알고리즘 1의 각 단계에서 판별자는 주어진 G에 대해 최적의 값에 도달하도록 허용 하고 기준을 개선하기 위해 pg 를 업데이트합니다.

$$Ex \quad pdata [\log D^* G(x)] + Ex \quad pg [\log(1 \quad D^*$$

G(x))] 그러면 pg 는 pdata 로 수렴

증거. V(G, D) = U(pg, D)를 위의 기준에서와 같이 pg 의 함수로 고려하십시오. U(pg, D)는 pg에서 볼록합니다. 볼록 함수의 최상부의 도함수는 최대값에 도달하는 지점에서의 함수의 도함수를 포함합니다. 즉, f(x) = supα∈A fα(x)이고 fα(x)가 모든 α에 대해 x에서 볼록한 경우 β = arg인 경우 ∂fβ(x) ∈ ∂f supα∈A fα(x) .

이것은 해당 G가 주어졌을 때 최적 D에서 pg 에 대한 경사하강법 업데이트를 계산하는 것과 같습니다. supD U(pg, D)는 Thm 1에서 입증된 고유한 전역 최적값으로 pg 에서 볼록하므로 pg 의 업데이트가 충분히 적습니다. , pg 는 px로 수렴 하여 증명을 마무리합니다. □

실제로, adversarial nets는 함수 G(z; θg) 를 통해 제한된 pg 분포 패밀리를 나타내며, pg 자체 보다는 θg 를 최적화합니다. 다층 퍼셉트론을 사용하여 G를 정의하면 매개변수 공간에 여러 임계점이 도입됩니다. 그러나 실제로 셉트론당 다층의 우수한 성능은 이론적 보장이 부족함에도 불구하고 사용하기에 합리적인 모델임을 시사합니다.

5가지 실험

우리는 MNIST[23], TFD(Toronto Face Database)[28], CIFAR-10[21]을 포함한 다양한 데이터 세트에서 적대적 네트워크를 훈련했습니다. 제너레이터 네트는 정류기 선형 활성화[19, 9]와 시그모이드 활성화를 혼합하여 사용하는 반면 판별자 네트는 최대값 활성화[10]를 사용합니다. Dropout[17]은 discriminator net 훈련에 적용되었습니다. 우리의 이론적 프레임워크는 생성기의 중간 레이어에서 드롭아웃 및 기타 노이즈의 사용을 허용하지만 노이즈를 생성기 네트워크의 맨 아래 레이어에만 입력으로 사용했습니다.

G로 생성된 샘플에 가우스 파젠 창을 맞추고 이 분포에서 로그 가능성을 보고하여 pg 에서 테스트 세트 데이터의 확률을 추정합니다 . σ 매개변수

모델	MNIST	TFD
DBN [3]	138 ± 2	1909 ± 66
누적 CAE [3]	121 ± 1.6	2110 ± 50
깊은 GSN [6]	214 ± 1.1	1890 ± 29
적대적 네트	225 ± 2	2057 ± 26

표 1: Parzen 창 기반 로그 가능성 추정치. MNIST에 보고된 숫자는 테스트 세트에서 샘플의 평균 로그 가능도이며, 평균의 표준 오차는 여러 예제에서 계산됩니다. TFD에서 우리는
검증 세트를 사용하여 선택된 다른 σ를 사용하여 데이터 세트의 접힌 부분에 대해 표준 오차를 계산했습니다.
각 접기. TFD에서 σ는 각 폴드에 대해 교차 검증되었고 각 폴드에 대한 평균 로그 가능성이 계산되었습니다.
MNIST의 경우 데이터 세트의 실제 값(바이너리가 아닌) 버전의 다른 모델과 비교합니다.

검증 세트에 대한 교차 검증을 통해 가우시안의 값을 얻었습니다. 이 절차는 Breuleux et al. [8] 정확한 가능성이 있는 다양한 생성 모델에 사용됩니다.
다루기 어렵다[25, 3, 5]. 결과는 표 1에 보고됩니다. 이 가능성 추정 방법
분산이 다소 높고 고차원 공간에서는 잘 수행되지 않지만 가장 좋습니다.
우리가 알고 있는 방법. 샘플링할 수 있지만 추정할 수 없는 생성 모델의 발전
가능성은 그러한 모델을 평가하는 방법에 대한 추가 연구에 직접적인 동기를 부여합니다.

그림 2와 3에서는 훈련 후 생성기 네트에서 추출한 샘플을 보여줍니다. 우리가 하지 않는 동안
이러한 샘플이 기존 방법으로 생성된 샘플보다 낫다고 주장하지만, 우리는 이러한 샘플이
샘플은 문헌에서 더 나은 생성 모델과 적어도 경쟁적이며 강조 표시됩니다.
적대적 프레임워크의 가능성.

그림 2: 모델의 샘플 시각화. 가장 오른쪽 열은 가장 가까운 훈련 예를 보여줍니다.
모델이 훈련 세트를 기억하지 않았음을 입증하기 위해 이웃 샘플. 시료
체리 피킹이 아닌 공정한 무작위 추첨입니다. 심층 생성 모델의 다른 대부분의 시각화와 달리 이러한
이미지는 숨겨진 단위의 샘플이 주어진 조건부 평균이 아니라 모델 분포의 실제 샘플을 보여줍니다.
또한 샘플링 프로세스가 Markov 체인에 의존하지 않기 때문에 이러한 샘플은 상관 관계가 없습니다.
혼입. a) MNIST b) TFD c) CIFAR-10(완전 연결 모델) d) CIFAR-10(convolutional discriminator)
및 "디콘볼루션" 생성기)

그림 3: 전체 모델의 z 공간 에서 좌표 사이를 선형으로 보간하여 얻은 자릿수 .

	딥 디렉티드 그래픽 모델	깊은 무방향 그래픽 모델 훈련 중 추론이 필요합니다.	제너레이티브 카 코더	적대적 모델
훈련	훈련 중 추론이 필요합니다.	MCMC는 분할 함수 기울기를 근사화하는 데 필요했습니다.	혼합과 재건 생성의 힘 사이의 강제 절충	판별자와 생성자를 동기화합니다. 헬베티카.
추론	학습된 대략적인 추론	변이 추론	MCMC 기반 추론	학습된 대략적인 추론
견본 추출	어려움 없음	마르코프 사슬 필요	마르코프 사슬 필요	어려움 없음
p(x) 평가	다루기 힘든, 다음과 같이 근사될 수 있습니다. AIS	다루기 힘든, 다음과 같이 근사될 수 있습니다. AIS	명시적으로 표시되지 않으며 다음과 같이 근사될 수 있습니다. 파젠 밀도 추정	명시적으로 표시되지 않으며 다음과 같이 근사될 수 있습니다. 파젠 밀도 추정
모델 디자인	거의 모든 모델은 극도의 어려움을 겪습니다.	여러 속성을 보장하기 위해 세심한 설계 필요	미분 가능한 함수는 이론적으로 허용됩니다.	미분 가능한 함수는 이론적으로 허용됩니다.

표 2: 생성 모델링의 과제: 모델과 관련된 각 주요 작업에 대한 심층 생성 모델링에 대한 다양한 접근 방식이 직면하는 어려움에 대한 요약입니다.

6 장점과 단점

이 새로운 프레임워크는 이전 모델링 프레임 작업에 비해 장점과 단점이 있습니다. 단점은 주로 pg(x)의 명시적 표현이 없고 훈련 중에 D가 G와 잘 동기화되어야 한다는 것입니다(특히 G는 "Helvetica 시나리오"를 피하기 위해 D를 업데이트하지 않고 너무 많이 훈련되어서는 안 됩니다. "에서 G는 너무 많은 z 값을 x의 동일한 값으로 축소하여 pdata 를 모델링하기에 충분한 다양성을 갖 습니다. 장점은 Markov 체인이 절대 필요하지 않고, 기울기를 얻기 위해 backprop만 사용되며, 학습 중에 추론이 필요하지 않으며, 다양한 기능을 모델에 통합할 수 있다는 것입니다. 표 2는 생성적 적대 네트워크와 다른 생성적 모델링 접근 방식의 비교를 요약한 것입니다.

앞서 언급한 장점은 주로 계산적입니다. 적대적 모델은 또한 데이터 예제로 직접 업데이트되지 않고 판별자를 통해 흐르는 그라디언트만 사용하여 생성기 네트워크에서 통계적 이점을 얻을 수 있습니다. 이것은 입력의 구성 요소가 생성기의 매개변수에 직접 복사되지 않음을 의미합니다. 적대적 네트워크의 또 다른 장점은 매우 날카롭고 심지어 퇴화된 분포를 나타낼 수 있다는 것입니다. 반면 Markov 체인을 기반으로 하는 방법은 체인이 모드 간에 혼합될 수 있도록 분포가 다소 흐릿해야 합니다.

7 결론 및 향후 과제

이 프레임워크는 많은 간단한 확장을 허용합니다.

1. 조건부 생성 모델 p(x | c)는 G와 D 모두에 입력으로 c를 추가하여 얻을 수 있습니다.
2. 학습된 근사 추론은 x가 주어졌을 때 z를 예측하도록 보조 네트워크를 훈련하여 수행할 수 있습니다. 이것은 wake-sleep 알고리즘[15]에 의해 훈련된 추론망과 유사하지만, 발전기망이 훈련을 마친 후에 고정된 발전기망에 대해 추론망을 훈련할 수 있다는 장점이 있습니다.

3. 매개변수를 공유하는 조건부 모델군을 훈련함으로써 모든 조건부 p(xS | x6S) 를 대략적으로 모델링할 수 있습니다 . 여기서 S는 x 인덱스의 하위 집합입니다. 기본적으로 적대적 네트를 사용하여 결정론적 MP-DBM의 확률적 확장을 구현할 수 있습니다[11].
4. 반 지도 학습: 판별기 또는 추론 네트워크의 기능이 성능을 향상시킬 수 있습니다.
   제한된 레이블이 지정된 데이터를 사용할 수 있는 경우 분류기의 범위.
5. 효율성 개선: G와 D를 조정하는 더 나은 방법을 나누거나 훈련 중에 샘플 z에 대한 더 나은 분포를 결정함으로써 훈련을 크게 가속화할 수 있습니다.

이 문서는 적대적 모델링 프레임워크의 실행 가능성을 보여주었으며 이러한 연구 방향이 유용할 수 있음을 시사했습니다.

감사의 말

도움이 되는 토론에 대해 Patrice Marcotte, Olivier Delalleau, 조경현, Guillaume Alain 및 Jason Yosinski에게 감사드립니다. Yann Dauphin은 Parzen 창 평가 코드를 우리와 공유했습니다. 우리는 Pylearn2 [12] 및 Theano [7, 1] 개발자, 특히 이 프로젝트에 도움이 되도록 Theano 기능을 서두른 Fred' eric Bastien에게 감사드립니다. Arnaud Bergeron은 LATEX 조판에 필요한 지원을 제공했습니다. 또한 자금을 제공한 CIFAR 및 Canada Research Chairs, 컴퓨팅 리소스를 제공한 Compute Canada 및 Calcul Quebec에 감사드립니다. Ian Goodfellow는 2013 Google Fellowship in Deep Learning의 지원을 받습니다. 마지막으로 우리의 창의성을 자극해 준 Les Trois Brasseurs에게 감사드립니다.

참고문헌

[1] Bastien, F., Lamblin, P., Pascanu, R., Bergstra, J., Goodfellow, IJ, Bergeron, A., Bouchard, N. 및 Bengio, Y.(2012). ano: 새로운 기능 및 속도 개선. 딥 러닝 및 비지도 기능 학습 NIPS 2012 워크샵.

[2] Bengio, Y. (2009). AI를 위한 심층 아키텍처 학습. 이제 퍼블리셔.

[3] Bengio, Y., Mesnil, G., Dauphin, Y. 및 Rifai, S.(2013a). 깊은 표현을 통해 더 나은 믹싱. 에 ICML'13.

[4] Bengio, Y., Yao, L., Alain, G. 및 Vincent, P. (2013b). 제너레이티브로 일반화된 노이즈 제거 자동 인코더 모델. NIPS26에서. 닙스 재단.

[5] Bengio, Y., Thibodeau-Laufer, E. 및 Yosinski, J. (2014a). 심층 생성 확률적 네트워크 학습 가능 백프롭으로. ICML'14에서.

[6] Bengio, Y., Thibodeau-Laufer, E., Alain, G. 및 Yosinski, J.(2014b). backprop으로 학습할 수 있는 Deep Generative stochastic net works. 기계 학습에 관한 제30차 국제 회의(ICML'14)의 회보에서.

[7] Bergstra, J., Breuleux, O., Bastien, F., Lamblin, P., Pascanu, R., Desjardins, G., Turian, J., Warde-Farley, D. 및 Bengio, Y. (2010). Theano: CPU 및 GPU 수학 표현식 컴파일러. Python for Scientific Computing Conference(SciPy) 회보에서. 구두 발표.

[8] Breuleux, O., Bengio, Y. 및 Vincent, P. (2011). 에서 대표 샘플을 빠르게 생성 RBM 파생 프로세스. 신경 계산, 23(8), 2053–2073.

[9] Glorot, X., Bordes, A. 및 Bengio, Y.(2011). 깊은 희소 정류기 신경망. AISTATS'2011.

[10] Goodfellow, IJ, Warde-Farley, D., Mirza, M., Courville, A. 및 Bengio, Y.(2013a). Maxout 네트워크. ICML'2013.

[11] Goodfellow, IJ, Mirza, M., Courville, A. 및 Bengio, Y.(2013b). 다중 예측 심층 볼츠만 기계. NIPS'2013.

[12] Goodfellow, IJ, Warde-Farley, D., Lamblin, P., Dumoulin, V., Mirza, M., Pascanu, R., Bergstra, J., Bastien, F. 및 Bengio, Y.(2013c). Pylearn2: 기계 학습 연구 라이브러리. arXiv 사전 인쇄 arXiv:1308.4214.

[13] Gutmann, M. 및 Hyvarinen, A. (2010). 잡음 대비 추정: 비정규화된 통계 모델에 대한 새로운 추정 원칙. AISTATS'2010에서.

[14] Hinton, G., Deng, L., Dahl, GE, Mohamed, A., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T. 및 Kingsbury, 나.(2012a). 음성 인식에서 음향 모델링을 위한 심층 신경망. IEEE Signal Processing Magazine, 29(6), 82–97.

[15] Hinton, GE, Dayan, P., Frey, BJ 및 Neal, RM(1995). 비지도자를 위한 웨이크-슬립 알고리즘 신경망. 과학, 268, 1558-1161.

[16] Hinton, GE, Osindero, S. 및 Teh, Y.(2006). 깊은 신뢰망을 위한 빠른 학습 알고리즘. 신경
계산, 18, 1527-1554.

[17] Hinton, GE, Srivastava, N., Krizhevsky, A., Sutskever, I. 및 Salakhutdinov, R.(2012b). 특징 탐지기의 공동 적응을 방지하여 신경망
을 개선합니다. 기술 보고서, arXiv:1207.0580.

[18] Hyvarinen, A. (2005). 점수 매칭을 사용한 비정규화 통계 모델의 추정.
학습 해상도, 6.

[19] Jarrett, K., Kavukcuoglu, K., Ranzato, M. 및 LeCun, Y. (2009). 물체 인식을 위한 최고의 다단계 아키텍처는 무엇입니까? 프로시저에서 컴
퓨터 비전에 관한 국제 회의(ICCV'09), 2146–2153페이지.
IEEE.

[20] Kingma, DP 및 Welling, M.(2014). 자동 인코딩 변형 베이.
학습 표현에 관한 회의(ICLR).

[21] Krizhevsky, A. 및 Hinton, G. (2009). 작은 이미지에서 여러 계층의 기능을 학습합니다. 전문인
보고, 토론토 대학.

[22] Krizhevsky, A., Sutskever, I. 및 Hinton, G. (2012). 심층 컨볼루션 신경망을 사용한 ImageNet 분류. NIPS'2012.

[23] LeCun, Y., Bottou, L., Bengio, Y. 및 Haffner, P. (1998). 문서에 적용된 기울기 기반 학습
인식. IEEE, 86(11), 2278–2324의 절차.

[24] Rezende, DJ, Mohamed, S. 및 Wierstra, D.(2014). 확률적 역전파 및 근사
심층 생성 모델의 추론 기술 보고서, arXiv:1401.4082.

[25] Rifai, S., Bengio, Y., Dauphin, Y. 및 Vincent, P. (2012). 수축 표본 추출을 위한 생성 과정
자동 인코더. ICML'12에서.

[26] Salakhutdinov, R. 및 Hinton, GE(2009). 딥 볼츠만 머신. AISTATS'2009, 448페이지–
455.

[27] Smolensky, P. (1986). 역학 시스템의 정보 처리: 조화 이론의 기초. DE Rumelhart 및 JL McClelland 편집자, 병렬 분산 처리, 1권, 6장,
194–281페이지. MIT 프레스, 케임브리지.

[28] Susskind, J., Anderson, A. 및 Hinton, GE(2010). 토론토 얼굴 데이터셋. 기술 보고서 UTML
TR 2010-001, U. 토론토.

[29] Tieleman, T. (2008). 우도 기울기에 대한 근사를 사용하여 제한된 Boltzmann 기계를 훈련합니다. WW Cohen, A. McCallum 및 ST
Roweis, 편집자, ICML 2008, 페이지 1064–1071. ACM.

[30] Vincent, P., Larochelle, H., Bengio, Y. 및 Manzagol, P.-A. (2008). 노이즈 제거 자동 인코더로 강력한 기능을 추출하고 구성합니다.
ICML 2008에서.

[31] Younes, L.(1999). ergodicity 비율이 급격히 감소하는 Markovian stochastic 알고리즘의 수렴. 확률론 및 확률론적 보고서, 65(3), 177–
228.