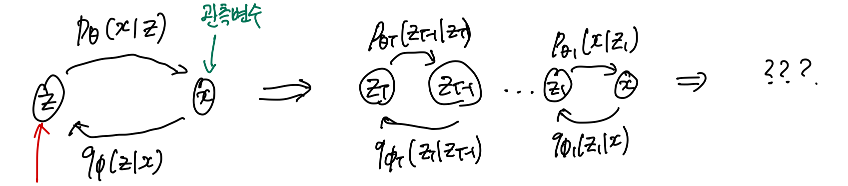
Ch 08. 확산 모델 이론

1. VAE에서 확산 모델로
2. 확산 과정에서 역확산 과정
3. ELBO 계산 ①
4. ELBO 계산 ②
5. ELBO 계산 ③
6. 확산 모델의 학습 (알고리즘)

8.1 VAE에서 확산 모델로

DDPM? → 확산모델

(Denoising Diffusion Probabilistic Models,
노이즈 제거 확산 확률 모델)



VAE 모형도

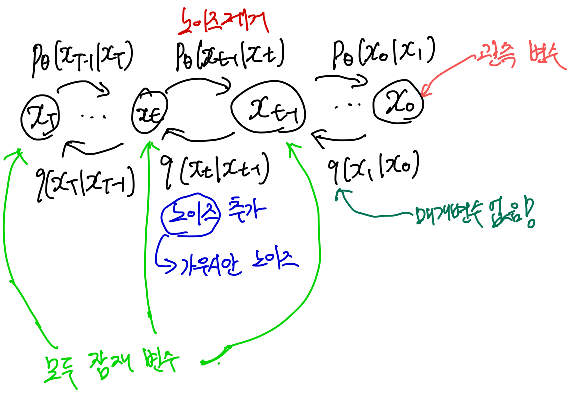
계층형 VAE 모형도

확산 모델

확산모델은?

계층형 VAE + ① 관측 변수와 잠재 변수의 차원 일치? ② 고정된 정규 분포를 따르는 노이즈를 인코더에 추가? ⇒ 확산모델?

즉? 확산모델 모형도? ⇒



8.2 확산 과정과 역확산 과정

확산 모델 — 확산 과정 : 노이즈를 추가하는 과정
— 역확산 과정 : 노이즈를 제거하는 과정

확산 과정(diffusion process)

$$x_T \leftarrow \cdots \leftarrow x_t \leftarrow x_{t-1} \leftarrow \cdots \leftarrow x_0$$

완전한 노이즈

$q(x_t|x_{t-1})$

미리 설정한 값!

즉! $q(x_t|x_{t-1}) = N(x_t; \sqrt{1-\beta_t}\, x_{t-1}, \beta_t I)$

그럼! $p(x_T) \approx N(x_T; 0, I)$!

이때! 재매개변수화 트릭 발동!

$$\epsilon \sim N(\epsilon; 0, I)$$

$$\therefore x_t = \sqrt{1-\beta_t}\, x_{t-1} + \sqrt{\beta_t}\, \epsilon$$

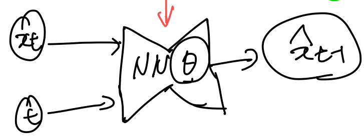
역확산 과정(reverse diffusion process)

$$x_T \to \cdots \to x_t \to x_{t-1} \to \cdots \to x_0$$

$p_\theta(x_{t-1}|x_t)$

요건 주목해!

이는!



즉! $\hat{x}_{t-1} = NeuralNet(x_t, t; \theta)$

$$\therefore p_\theta(x_{t-1}|x_t) = N(x_{t-1}; \hat{x}_{t-1}, I)$$

요건 그냥 예시(그렇지)

정확히는! 공분산항!!

8.3 ELBO 계산 ①

확산 모델 $\longrightarrow$ ELBO $\longrightarrow$ 근삿값 $\xrightarrow{\text{우선!}}$ 최종 시각 $=T$ $\Rightarrow$ T개 샘플 데이터 $x_1, x_2 \cdots x_T$ 근삿값!

3가지 변경
1. $x \to x_0$
2. $z \to x_{1:T}$
3. ϕ 삭제

확산 모델의 ELBO

$$\text{VAE의 ELBO}(x;\theta,\phi) = \int q_\phi(z|x)\log\frac{p_\theta(x,z)}{q_\phi(z|x)}dz$$

$$= E_{q_\phi(z|x)}\left[\log\frac{p_\theta(x,z)}{q_\phi(z|x)}\right]$$

$$\Rightarrow \text{확산 모델의 ELBO}(x_0;\theta) = E_{q(x_{1:T}|x_0)}\left[\log\frac{p_\theta(x_{0:T})}{q(x_{1:T}|x_0)}\right]$$

알아두자!

$$p_\theta(x_{0:T}) = p_\theta(x_0|x_1)\,p(x_1|x_2)\cdots p_\theta(x_{T-1}|x_T)\,p(x_T)$$

$$= p(x_T)\prod_{t=1}^{T}p_\theta(x_{t-1}|x_t)$$

$$q(x_{1:T}|x_0) = \prod_{t=1}^{T}q(x_t|x_{t-1})$$

$\xrightarrow{\text{즉!}}$

$$\text{ELBO}(x_0;\theta)$$

$$= E_{q(x_{1:T}|x_0)}\left[\log\frac{p_\theta(x_{0:T})}{q(x_{1:T}|x_0)}\right]$$

$$= E_{q(x_{1:T}|x_0)}\left[\log\frac{p(x_T)\prod_{t=1}^{T}p_\theta(x_{t-1}|x_t)}{\prod_{t=1}^{T}q(x_t|x_{t-1})}\right]$$

$$= E_{q(x_{1:T}|x_0)}\left[\underbrace{\log\prod_{t=1}^{T}p_\theta(x_{t-1}|x_t)}_{\theta\text{ 포함!}}+\underbrace{\log\frac{p(x_T)}{\prod_{t=1}^{T}q(x_t|x_{t-1})}}_{\theta\text{ 미포함}}\right]$$

즉!

$$J(\theta) = E_{q(x_{1:T}|x_0)}\left[\log\prod_{t=1}^{T}p_\theta(x_{t-1}|x_t)\right]$$

$$= E_{q(x_{1:T}|x_0)}\left[\sum_{t=1}^{T}\log p_\theta(x_{t-1}|x_t)\right]$$

따라서! 몬테 카를로 사용!

$\Downarrow$

여러 $x_{1:T}$ 생성

$\Downarrow$

$\sum_{t=1}^{T}\log p_\theta(x_{t-1}|x_t)$ 평균 계산

$\xrightarrow{\text{샘플 크기가 1이라면}}$

$$x_{1:T} \sim q(x_{1:T}|x_0)$$

$$J(\theta) \approx \sum_{t=1}^{T}\log p_\theta(x_{t-1}|x_t)$$

즉!

원본 $x_0 \xrightarrow{\text{생성}} x_{1:T} \xrightarrow{\text{각 시각에서 계산}} \log p_\theta(x_{t-1}|x_t)$

$$\hat{x}_{t-1} = \text{NeuralNet}(x_t, t; \theta)$$

$$p_\theta(x_{t-1}|x_t) = N(x_{t-1}; \hat{x}_{t-1}, I)$$

또한!

$$J(\theta) \approx \sum_{t=1}^{T} \log p_\theta(x_{t-1}|x_t)$$

$$= \sum_{t=1}^{T} \log N(x_{t-1}; \hat{x}_{t-1}, I)$$

t 시각 번호 변경, 정규 분포 식 대입

$$= \sum_{t=0}^{T-1} \log \frac{1}{\sqrt{(2\pi)^D |I|}} \exp\left\{-\frac{1}{2}(x_t - \hat{x}_t)^T I^{-1} (x_t - \hat{x}_t)\right\}$$

$$= -\frac{1}{2} \sum_{t=0}^{T-1} (x_t - \hat{x}_t)^T (x_t - \hat{x}_t) + \underbrace{T \log \frac{1}{\sqrt{(2\pi)^D}}}_{\text{상수} \rightarrow \text{무시!}}$$

$$= -\frac{1}{2} \sum_{t=0}^{T-1} \|x_t - \hat{x}_t\|^2$$

즉! 목적 함수 $J(\theta)$는!

1. 확산 과정을 통해 T개 샘플 얻기!
2. 신경망을 T번 적용해서 노이즈 제거!
3. 각 시각의 제곱오차 $\|x_t - \hat{x}_t\|^2$ 계산!

⇒ 'T'번이나...?. 너무 많을 것 같아...

8.4 ELBO 계산 ②

역시? T개는 너무 많아? $\xrightarrow{\text{따라서}}$ 2개로 해보자?

즉? $q(x_t|x_0) = N(x_t; \sqrt{\bar{\alpha}_t}\,x_0, (1-\bar{\alpha}_t)I)$ $\xrightarrow{\text{기호?}}$ $\alpha_t = 1-\beta_t$

$\bar{\alpha}_t = \alpha_t\alpha_{t-1}\cdots\alpha_1$

우선 J(θ)를 축약!

$$J(\theta) = E_{q(x_{1:T}|x_0)}\left[\sum_{t=1}^{T} \log p_\theta(x_{t-1}|x_t)\right]$$

기댓값의 선형성

$$= \sum_{t=1}^{T} E_{q(x_{1:T}|x_0)}\left[\log p_\theta(x_{t-1}|x_t)\right]$$

관련 변수의 기댓값

$$= \sum_{t=1}^{T} E_{q(x_{t-1},x_t|x_0)}\left[\log p_\theta(x_{t-1}|x_t)\right]$$

즉 ⇒ 2변의 샘플링만 해도 됨?

요기에? 관련X 따라서? 가능 O

그리고?

이것도 너무 커? 즉? 균등분포 + 몬테카를로! 즉? $J(\theta) = \sum_{t=1}^{T} E_{q(x_{t-1},x_t|x_0)}\left[\log p_\theta(x_{t-1}|x_t)\right]$

$= T\,E_{u(t)}\left[E_{q(x_{t-1},x_t|x_0)}\left[\log p_\theta(x_{t-1}|x_t)\right]\right]$

⇓ 몬테카를로?

요거?

$$E_{u(t)}[f(t)] = \sum_{t=1}^{T} u(t)f(t) = \sum_{t=1}^{T} \frac{1}{T}f(t) = \frac{1}{T}\sum_{t=1}^{T} f(t)$$

$$\therefore T\,E_{u(t)}[f(t)] = \sum_{t=1}^{T} f(t)$$

if 샘플 = 1개

$t \sim U\{1,T\}$

$x_{t-1} \sim q(x_{t-1}|x_t)$

$x_t \sim q(x_t|x_{t-1})$

$\Rightarrow J(\theta) \approx T \log p_\theta(x_{t-1}|x_t)$ 즉?

⇓

$\hat{x}_{t-1} = \mathrm{NeuralNet}(x_t, t; \theta)$

$p_\theta(x_{t-1}|x_t) = N(x_{t-1}; \hat{x}_{t-1}, I)$

⇓ J(θ)는?!?!

J(θ) 계산은? ⟸ $J(\theta) \approx -\frac{T}{2}\|x_{t-1} - \hat{x}_{t-1}\|^2$

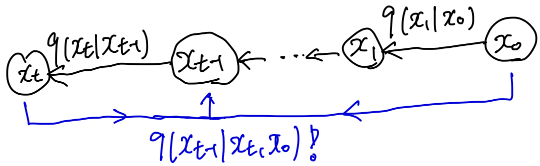
1. 균등분포 U{1, T}에서 시각 t를 샘플링
2. $q(x_{t-1}|x_0)$에서 x_{t-1}을 샘플링
3. $q(x_t|x_{t-1})$에서 x_t를 샘플링
4. 신경망에 x_t를 입력? $\hat{x}_{t-1}$를 출력?
5. 제곱오차 $\|x_{t-1} - \hat{x}_{t-1}\|^2$ 계산?

8.5 ELBO 계산 ③

2단계! 이제는 단 하나의 샘플 데이터만 사용해서 ELBO을 근사!

핵심은! 정규 분포!

$q(x_{t-1}|x_t, x_0)$ 뜻! x_0와 x_t가 주어졌을 때 x_{t-1}의 확률! $\xrightarrow{\text{중요!}}$ 해석적 풀이가 가능



$$q(x_{t-1}|x_t, x_0) = N(x_{t-1}; \mu_q(x_t, x_0), \sigma_q^2(t) I)$$

$$\sigma_q^2(t) = \frac{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}$$

$1-\beta_t$

$\bar{\alpha}_t = \alpha_t \alpha_{t-1} \cdots \alpha_1$

$$\mu_q(x_t, x_0) = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1}) x_t + \sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t) x_0}{1-\bar{\alpha}_t}$$

중요점!

이건 두 개 ELBO $q(x_{t-1}|x_t, x_0)$ 형태! → 한 개 ELBO

즉! $J(\theta) = \sum_{t=1}^{T} E_{q(x_{t-1}, x_t|x_0)}[\log p_\theta(x_{t-1}|x_t)]$

$= T E_{u(t)}\Big[\underbrace{E_{q(x_{t-1}, x_t|x_0)}[\log p_\theta(x_{t-1}|x_t)]}_{J_0}\Big] \Rightarrow T E_{u(t)}[J_0]$

여기에 $q(x_{t-1}|x_t, x_0)$을 추가!

따라서!

$$\underset{\theta}{\arg\max}\, J_0 = \underset{\theta}{\arg\max} \Big(J_0 - \overbrace{E_{q(x_{t-1}, x_t|x_0)}[\log q(x_{t-1}|x_t, x_0)]}^{\text{고정! 상수!}}\Big)$$

$$= \underset{\theta}{\arg\max}\, E_{q(x_{t-1}, x_t|x_0)}[\log p_\theta(x_{t-1}|x_t) - \log q(x_{t-1}|x_t, x_0)]$$

$$= \underset{\theta}{\arg\max}\, \underbrace{E_{q(x_{t-1}, x_t|x_0)}\left[\log \frac{p_\theta(x_{t-1}|x_t)}{q(x_{t-1}|x_t, x_0)}\right]}_{J_1} \xrightarrow{\text{새로운 } J(\theta)} J(\theta) = T E_{u(t)}[J_1]$$

New!

그럼 J_1은? 어떤 식 전개?

$$J_1 = \int q(x_{t-1}, x_t | x_0) \log \frac{p_\theta(x_{t-1}|x_t)}{q(x_{t-1}|x_t, x_0)} \, dx_{t-1} \, dx_t$$

$$q(x_{t-1}, x_t | x_0) = q(x_t|x_0)\, q(x_{t-1}|x_t, x_0)$$

$$= -\int q(x_t|x_0) \underbrace{\int q(x_{t-1}|x_t, x_0) \log \frac{q(x_{t-1}|x_t, x_0)}{p_\theta(x_{t-1}|x_t)} \, dx_{t-1}}_{\text{KL발산}} \, dx_t$$

정규분포!

$$= -E_{q(x_t|x_0)}\left[\underbrace{D_{KL}\left(q(x_{t-1}|x_t, x_0) \,\|\, p_\theta(x_{t-1}|x_t)\right)}_{\frac{1}{2\sigma_q^2(t)} \|\mu_\theta(x_t, t) - \mu_q(x_t, x_0)\|^2}\right] \xRightarrow{\text{즉!}} q(x_{t-1}|x_t, x_0) = p_\theta(x_{t-1}|x_t) \Rightarrow \text{최대!}$$

즉!

신경망을 통해

$$\hat{x}_{t-1} = \mathrm{NeuralNet}(x_t, t; \theta)$$

$q(x_{t-1}|x_t, x_0)$과 같은 값!

$$p_\theta(x_{t-1}|x_t) = N(x_{t-1}; \hat{x}_{t-1}, \sigma_q^2(t) I)$$

New!

$$J(\theta) = -T\, E_{u(t)}\left[E_{q(x_t|x_0)}\left[\frac{1}{2\sigma_q^2(t)} \|\mu_\theta(x_t, t) - \mu_q(x_t, t)\|^2\right]\right]$$

$$\mathrm{Loss}(x_0; \theta) = E_{u(t)}\left[E_{q(x_t|x_0)}\left[\frac{1}{2\sigma_q^2(t)} \|\mu_\theta(x_t, t) - \mu_q(x_t, t)\|^2\right]\right]$$

$t \sim U\{1, T\}$

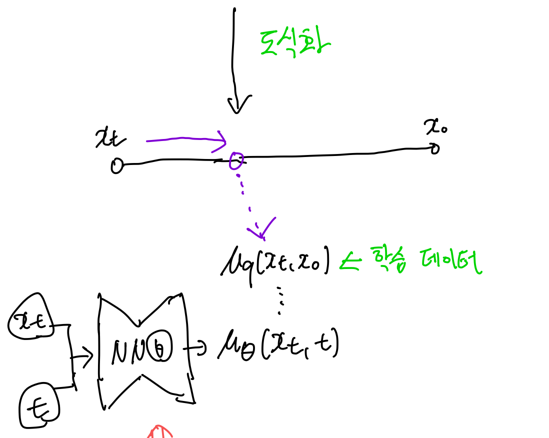
$x_t \sim q(x_t|x_0)$

$$= \frac{1}{\sigma_q^2(t)} \|\mu_\theta(x_t, t) - \mu_q(x_t, t)\|^2$$

8.6 확산 모델의 학습 (알고리즘)

확산 모델의 학습 알고리즘

1. 반복:
2. x_0을 학습 데이터에서 무작위로 가져옴
3. $t \sim U\{1, T\}$, $\epsilon \sim N(0, I)$
4. $x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \epsilon$ ← $q(x_t|x_0)$에서 샘플링
5. $\mu_q(x_t, x_0) = \dfrac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})x_t + \sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)x_0}{1-\bar{\alpha}_t}$
6. $\sigma_q^2(t) = \dfrac{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}$
7. $Loss(x_0;\theta) = \dfrac{1}{\sigma_q^2(t)} \left\| \mu_\theta(x_t, t) - \mu_q(x_t, x_0) \right\|^2$
8. $\dfrac{\partial}{\partial\theta} Loss(x_0;\theta)$ 갱신



$\mu_q(x_t, x_0)$를 학습 데이터로 하는 확산 모델!

원본 데이터를 복원하는 신경망

$$\mu_q(x_t, x_0) = \frac{\sqrt{\alpha_t}\,(1-\bar{\alpha}_{t-1})\,x_t + \bar{\alpha}_{t-1}(1-\alpha_t)\,x_0}{1-\bar{\alpha}_t}$$

대응!

$$\mu_\theta(x_t, t) = \frac{\sqrt{\alpha_t}\,(1-\bar{\alpha}_{t-1})\,x_t + \bar{\alpha}_{t-1}(1-\alpha_t)\,\boxed{\hat{x}_\theta(x_t, t)}}{1-\bar{\alpha}_t}$$

신경망이 추정한 값

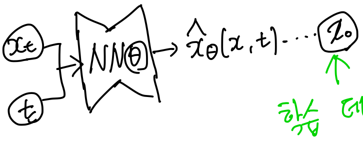
즉! $\Longrightarrow$

$$D_{KL}\left(q(x_{t-1}|x_t, x_0) \,\|\, p_\theta(x_{t-1}|x_t)\right)$$

$$= \frac{1}{2\sigma_q^2(t)} \left\| \mu_\theta(x_t, t) - \mu_q(x_t, x_0) \right\|^2$$

$$= \frac{1}{2\sigma_q^2(t)} \left(\frac{\sqrt{\bar{\alpha}_{t-1}}\,(1-\alpha_t)^2}{1-\bar{\alpha}_t} \right) \underbrace{\left\| \hat{x}_\theta(x_t, t) - x_0 \right\|^2}_{\text{손실 함수로!}}$$

따라서!



즉! $\longrightarrow$ x_0를 학습데이터로 하는 확산 모델!

노이즈를 예측하는 신경망

$$q(x_t|x_0) = N\left(x_t; \sqrt{\bar{\alpha}_t}\,x_0, (1-\bar{\alpha}_t)I\right)$$

$\Downarrow$ 재매개

$$\varepsilon \sim N(0, I) \Longrightarrow x_t = \sqrt{\bar{\alpha}_t}\,x_0 + \sqrt{1-\bar{\alpha}_t}\,\varepsilon \xrightarrow{\text{따라서}} x_0 = \frac{x_t - \sqrt{1-\bar{\alpha}_t}\,\varepsilon}{\sqrt{\bar{\alpha}_t}}$$

즉!

$$\mu_q(x_t, x_0) = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\,\varepsilon \right)$$

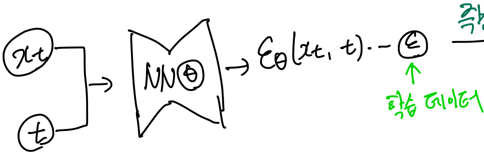
$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\,\varepsilon_\theta(x_t, t) \right)$$

즉! $\Longrightarrow$

$$D_{KL}\left(q(x_{t-1}|x_t, x_0) \,\|\, p_\theta(x_{t-1}|x_t)\right)$$

$$= \frac{1}{2\sigma_q^2(t)} \frac{(1-\alpha_t)^2}{(1-\bar{\alpha}_t)\alpha_t} \underbrace{\left\| \varepsilon_\theta(x_t, t) - \varepsilon \right\|^2}_{\text{손실 함수}}$$

따라서!



즉! $\longrightarrow$ ε을 학습 데이터로 하는 확산 모델!

새로운 데이터 샘플링

$$p_\theta(x_{t-1} \mid x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \sigma_q^2(t) I)$$

$\downarrow$ 재매개화

$$\varepsilon \sim \mathcal{N}(0, I) \xrightarrow{\text{샘플링}} x_{t-1} = \mu_\theta(x_t, t) + \sigma_q(t)\varepsilon$$

$$\sigma_q(t) = \sqrt{\frac{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}}$$

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\varepsilon_\theta(x_t, t)\right)$$

즉, 확산 모델의 데이터 생성 방법

1. $x_T \sim \mathcal{N}(0, I)$
2. for t in $[T \ldots 1]$:
3. $\quad \varepsilon \sim \mathcal{N}(0, I)$
4. $\quad$ if $t=1$ then $\varepsilon = 0$
5. $\quad \sigma_q(t) = \sqrt{\frac{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}}$
6. $\quad x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\varepsilon_\theta(x_t, t)\right) + \sigma_q(t)\varepsilon$
7. return x_0