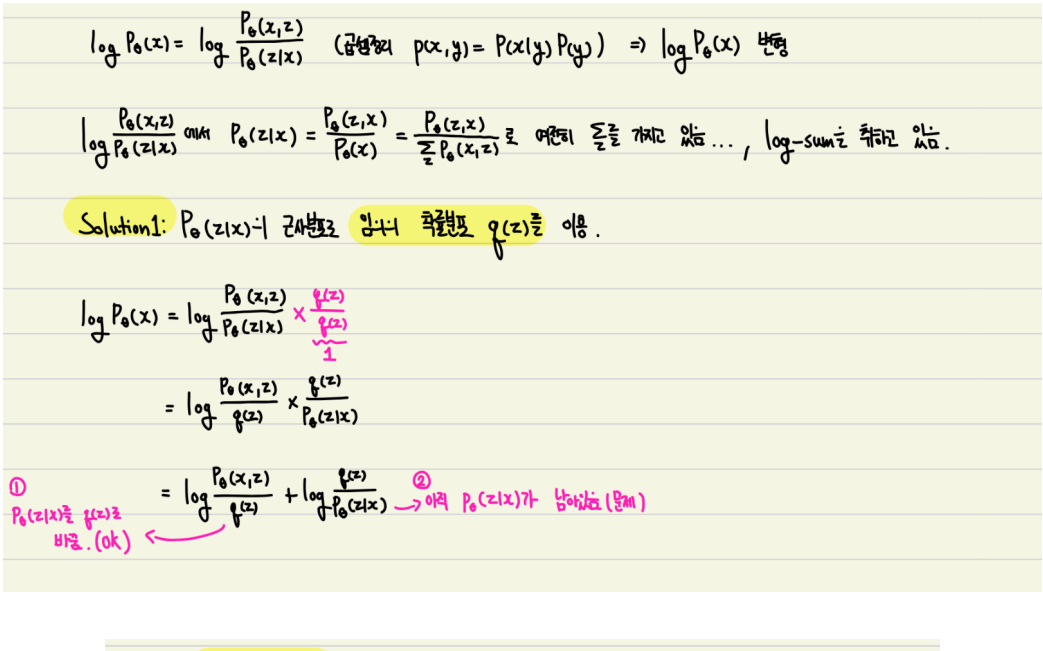
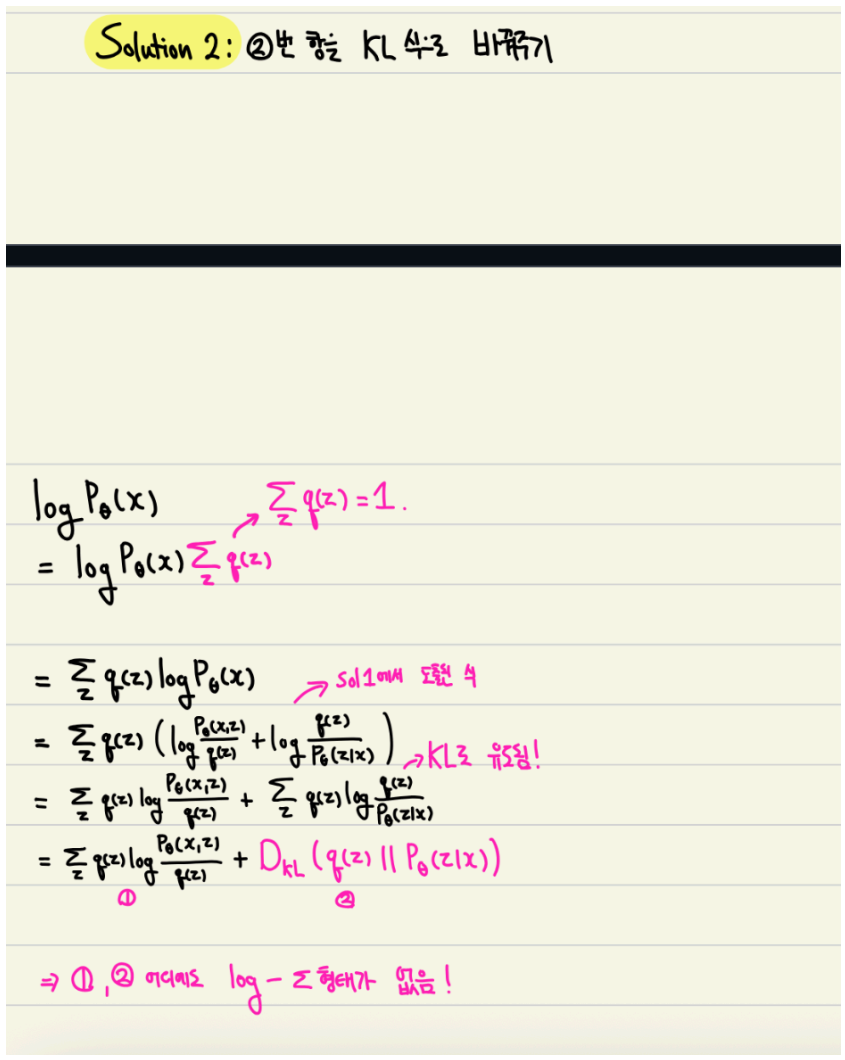
VAE 정리

VAE는 잠재 변수를 가지는 생성 모델이다.

생성 방식 : z가 표준 정규 분포를 따른다는 가정(prior)을 기반으로 p(z)에서 z를 추출하고, 추출된 z를 기반으로 x를 생성하는 모델이다.





앞서 다뤘듯이 잠재 변수를 가지는 생성 모델에서 하나의 데이터 x에 대한 log-likelihood는 공통 확률 분포에서의 잠재 변수 marginalization으로 인해 log-sum의 형태를 지니고 이는 해석적으로 최적화(최대화) 하기 어려운 문제점이 있다.

이 log-sum의 형태를 없에기 위해 원래 logPtheta(x)의 값을 근사하기 위해(더 구체적으로는 logPtheta(x)를 풀어서 구성한 ptheta(z|x)항을 근사하기 위해) q(z)를 도입한다. (VAE의 경우 q(z)가 관측 데이터 x를 기반으로 만들어지기에, q(z|x)로 표현함)

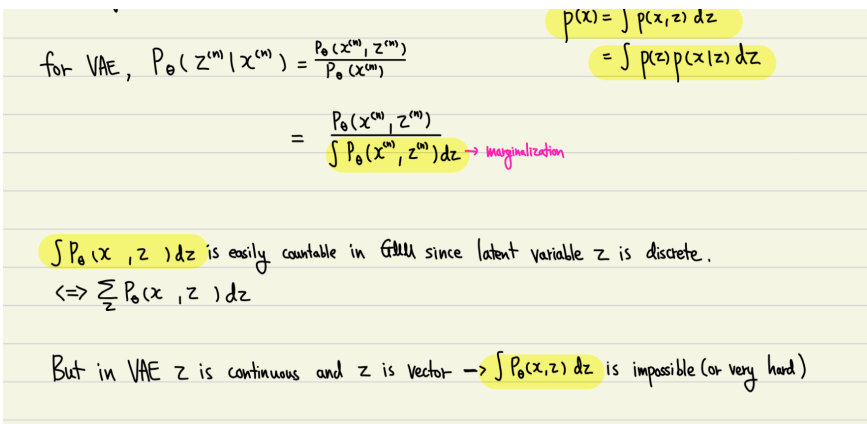
이렇게 q(z)를 도입하여 log-likelihood의 식을 풀어내면 원래의 log-likelihood와 식이 근사한 ELBO항과 KL항으로 풀어진다. 우리의 목표는 원래의 log-likelihood항을 최대화 하는 것인데 이 항이 log-sum의 형태를 지니고 있기에 해석적인 최적화가 어렵고 앞서 sum 형태를 없에기 위해 풀어진 두 항 중 log-likelihood항과 유사한 1) ELBO항을 간접적으로 최대화하면서도 2)이 ELBO항이 우리가 원래 최대화시키고자 했던 log-likelihood와 비슷하도록 만들어야한다. 1)은 ELBO를 구성하는 theta를 업데이트하면서 달성되고 2)는 q(z)를 ptheta(z|x)로 업데이트 함으로써 KL항을 0으로 만들며 달성된다.

결론적으로 도입하는 q(z)가 도입된 이유는 원래의 log-likelihood항에 있는 ptheta(z|x)를 근사하기 위함이다.

1)→ M step

2)→ E step

그러나 VAE의 경우 전통적인 EM 알고리즘 수행을 위한 E-step이 어렵다. q(z)를 ptheta(z|x)로 업데이트 하는 과정에서 뒤의 항을 구하는 것이 어렵기 때문이다. z가 연속적인 값을 가지는 백터이기에 주변화가 어려움.



q(z)를 ptheta(z|x)로 근사하기 위해(업데이트 하기 위해) 다른 방법을 이용하는 것이다.

그래서 q(z)를 파라미터로 조절 가능한 tracable한 분포로 바꾼다. 다시 말해 파라미터로 조절 가능한 분포인 가우시안 분포로 만든다. 이후에 ELBO를 q(z)를 구성하는 파라미터(pie)와 theta에 대해 최대화되도록 신경망을 training하면 q(z)는 ptheta(z|x)에 근사되고(EM에서의 E-step) x에 대한 likelihood도 최대화 된다(M-step)

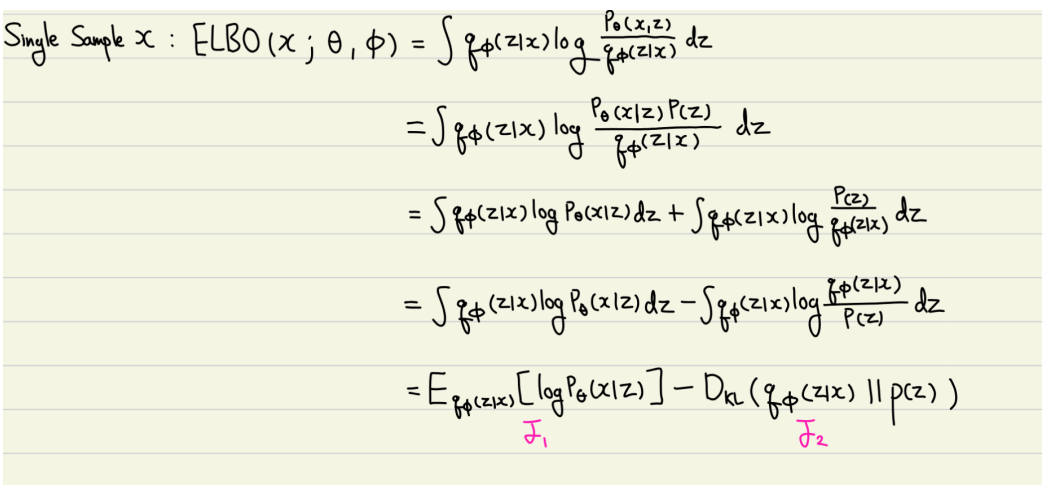
→ 이렇듯 계산이 어려운 복잡한 분포 ptheta(z|x)를 tracable한 분포로 근사하는 방식을 변분 추론/변분 근사라고 부른다.

→ 신경망 학습을 통해 동시에 두가지 종류의 파라미터를 학습시켜 ELBO가 최대화 되도록 업데이트 하면 ELBO는 원래의 log-likelihood에 가까워 지면서도(q(z)는 ptheta(z|x)에 근사되고(EM에서의 E-step)), 원래의 log-likelihood는 커진다.

그런데 매번 하나의 데이터 x에 대해 q(z)에 대한 파라미터를 일일이 준비할 수는 없음. 그래서 이 파라미터 만드는 작업을 x를 입력으로 받는 신경망을 통해 구현함.

신경망은 x를 입력으로 받아 q(z)의 평균 백터와 공분산 행렬의 대각 성분을 구성하는 백터를 출력함. VAE에서는 x로부터 z의 분산이 결정되는 것이기에, qpie(z|x)로 표현되고 이때 pie는 VAE에서 인코더 신경망의 파라미터임. 결론적으로 ELBO항의 q(z|x)를 표현하기 위해 인코더 신경망이 생겨난 것임.

아래는 VAE에서 하나의 데이터 x에 대한 ELBO항의 식이다.



ELBO항 또 위와 같이 두개의 항으로 쪼개짐.

J_1 는 $q_\phi(z|x)$를 따르는 z에 대한 $\log P_\theta(x|z)$의 Expectation

Monte carlo approximation으로 $q_\phi(z|x)$ 에서 z를 "1"개만 sampling해서 근사

$$u, \delta = \text{NeuralNet}(x; \phi)$$
$$z \sim N(z; u, \delta^2 I) \quad \text{(sample only 1)}$$
$$\hat{x} = \text{NeuralNet}(z; \theta)$$
$$J_1 \approx \log P_\theta(x|z)$$

$$\Leftrightarrow J_1 \approx \log N(x; \hat{x}, I)$$

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

$$= -\frac{1}{2}(x-\hat{x})^T(x-\hat{x}) + \log\frac{1}{\sqrt{(2\pi)^D}} \quad (I^{-1}=I,\ |I|=1)$$

$$= -\frac{1}{2}\sum_{d=1}^{D}(x_d-\hat{x}_d)^2 + \log\frac{1}{\sqrt{(2\pi)^D}}$$

($\hat{x}_d$: defined by θ, ϕ / $\log\frac{1}{\sqrt{(2\pi)^D}}$: constant)

$$\underset{\theta,\phi}{\text{argmax}}\ J_1 = \underset{\theta,\phi}{\text{argmax}} -\frac{1}{2}\sum_{d=1}^{D}(x_d-\hat{x}_d)^2$$
$$= \underset{\theta,\phi}{\text{argmin}} \sum_{d=1}^{D}(x_d-\hat{x}_d)^2 \quad : \text{reconstruction error}$$

- J_2
- Minimize J_2 for maximizing $\text{ELBO}(x; \phi, \theta)$

$q_\phi(z|x) : N(z; u, \delta^2 I)$
$p(z) : N(z; 0, I)$

why ④? : $D_{KL}(q_\phi(z|x) \| p(z))$ 값도 해석적으로 구해 놓음.
$q(z) = N(z; u_1, \delta_1^2 I)$, $p(z) = N(z; u_2, \delta_2^2 I)$ (두 확률분포가 Normal distribution일 때)
$D_{KL}(q\|p) = -\frac{1}{2}\sum_{i=1}^{M}\left(1+\log\frac{\delta_{1,i}^2}{\delta_{2,i}^2} - \frac{(u_{1,i}-u_{2,i})^2}{\delta_{2,i}^2} - \frac{\delta_{1,i}^2}{\delta_{2,i}^2}\right)$ ($D_{KL}(p\|q)$도 해석적으로 구할 수 있음) ····④

$$J_2 = D_{KL}(q_\phi(z|x) \| p(z))$$
$$= -\frac{1}{2}\sum_{h=1}^{H}(1+\log\delta_h^2 - u_h^2 - \delta_h^2)$$

Minimize J_2 ? $\Leftrightarrow q_\phi(z|x) = p(z)$ (consistency / regularization term)

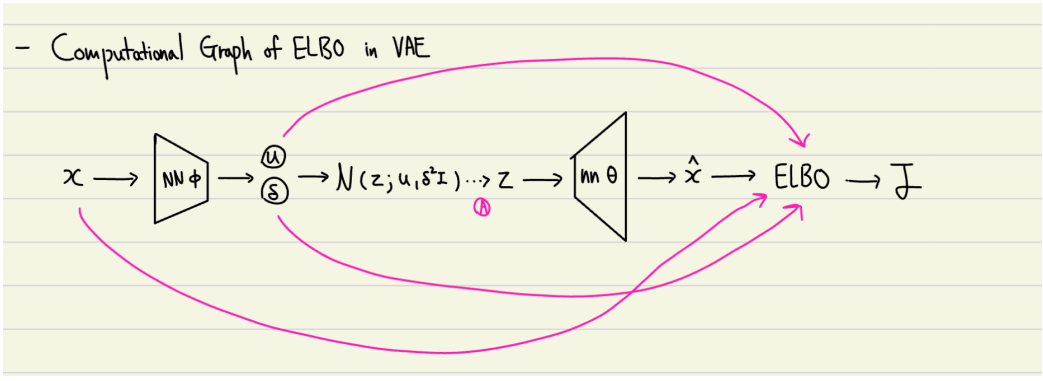
$$\therefore \text{ELBO}(x, \theta, \phi) \approx -\frac{1}{2}\sum_{d=1}^{D}(x_d-\hat{x}_d)^2 + \frac{1}{2}\sum_{h=1}^{H}(1+\log\delta_h^2 - u_h^2 - \delta_h^2) + \text{const}$$

(reconstruction loss / regularization term)

ELBO값을 최대화 시키기 위해서 원본 관측 데이터를 복원하면서도(J1 : reconstruction Term), q(z|x) 분포가 z의 prior인 p(z)에 가까워지도록(J2 : Regularization Term) 인코더와 디코더의 파라미터가 업데이트됨.

J1의 값은 Monte Carlo 근사를 이용했기에(주어진 x에 대해 z를 한개만 샘플링), 결론적으로는 ELBO의 근사값을 구한 것임

따라서 VAE의 모델의 ELBO값을 구하기 위해서는 인코더와 디코더 구조가 모두 필요한 것임. Training 시에 ELBO값 최적화를 위한 계산 그래프는 다음과 같음.



z sampling과정에서 gradient 연산이 끊기는 문제가 발생하기에 reparameterization trick을 도입함.