<탄생배경> **Reviewer : 김정학**

기본 상황

X: data z: latent vector

가정 : X는 어떠한 과정을 통해 z로부터 생성되었다.

① $z^{(i)}$는 어떤 사전확률 분포 $P_{\theta^*}(z)$로 부터 나왔다.

② $x^{(i)}$는 어떤 조건부 확률인 $P_{\theta^*}(x|z)$로 부터 생성되었다.

모든 PDF는 θ와 z에 대해서 미분가능하다.

하지만 θ와 z가 뭔지는 모른다.

기존에 존재했던 확률분포 추론은 두 가지 문제점이 있었다.

① Intratability

marginal probability 인 $P_\theta(x)$와 참 사후확률 $P_\theta(z|x)$

그리고 이를 위한 알고리즘이(적분) 모두 intractable (복잡)하다.

② A large dataset

기존의 배치 학습이 너무 저효율이다. (너무 느리다)

논문에서 제안하는 것들. (우리가 관심 있는 것들)

① 모수 θ 추정을 위한 효율적인 ML 이나 MAP 방법

② 사후확률을 효율적으로 추정하고 싶다 ($p(z|x)$)

③ marginal probability $p(x)$를 효율적으로 추정하고 싶다.

<Content>

x의 marginal likelihood $= \sum \log P_\theta(x^{(i)})$

다시쓰면 $\log P_\theta(x^{(i)}) = D_{KL}(q_\phi(z|x^{(i)}) \| P_\theta(z|x^{(i)})) + \mathcal{L}(\theta, \phi; x^{(i)})$

추정사후확률과 참사후확률의 분포차이 ≥ 0

따라서 lower bound를 구할 수 있음

$$\log p_\theta(x^{(i)}) \geq \mathcal{L}(\theta, \phi; x^{(i)})$$ 최대화해야 좋다 (Lower bound 효과)

① $\cdots = E_{q_\phi(z|x)}[-\log q_\phi(z|x) + \log p_\theta(x,z)]$

② $\cdots = -D_{KL}(q_\phi(z|x^{(i)}) \| p_\theta(z)) + E_{q_\phi(z|x^{(i)})}[\log p_\theta(x^{(i)}|z)]$

이 상태에서 Monte Carlo Method를 사용하면 분산이 커짐.

① $z \sim q_\phi(z|x)$를 따르는 분포에서 noise를 넣어서 z를 L개 추출한다. 다시 말하면 noise는 $\varepsilon \sim p(\varepsilon)$이고 이를 이용해 $g_\phi(\varepsilon, x)$를 추출 한다라고 생각하면 $z = g_\phi(\varepsilon, x)$이다. 그러면 $\mathcal{L}(\theta, \phi; x)$를 다음과 같이 근사할 수 있다. reparameterization trick

$$\tilde{\mathcal{L}}^A(\theta, \phi; x^{(i)}) = \frac{1}{L}\sum_{l=1}^{L}\left(\log p_\theta(x^{(i)}, z^{(i,l)}) - \log q_\phi(z^{(i,l)}|x^{(i)})\right)$$

② $D_{KL}(\cdot)$부분이 analytically integrate 할 수 있다면 (appendix B에 있음) $E[\cdot]$ 부분만 샘플링 과정을 거치면 된다. 식으로 쓰면 아래와 같다

$$\tilde{\mathcal{L}}^B(\theta, \phi; x^{(i)}) = \underbrace{-D_{KL}(q_\phi(z|x^{(i)}) \| p_\theta(z))}_{\text{Regularizatio}} + \underbrace{\frac{1}{L}\sum_{l=1}^{L}(\log p_\theta(x^{(i)}|z^{(i,l)}))}_{\text{Reconstruction Error}}$$

VAE관점.

※ reparameterization trick 사용조건

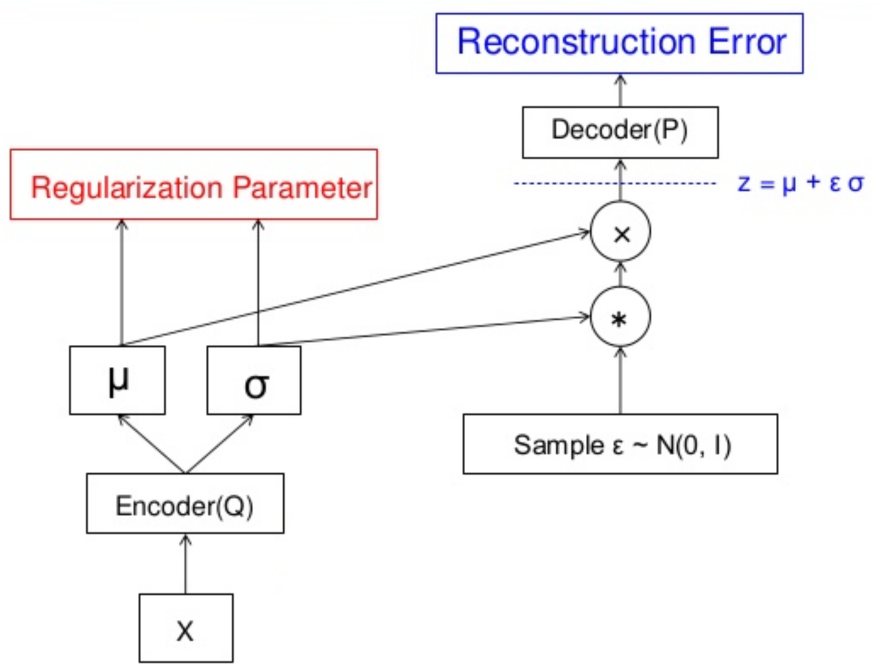
① CDF가 추적가능해야 한다

② 정규분포랑 비슷하게 "location-Scale"한 분포여야 한다. 즉 정규분포처럼 location → mean scale → std 가 될 수 있어야 한다.

③ 합성한 분포도 가능하다 (e.g. log-normal ..)

어렵네 ;;;

<VAE>



① forward

데이터 x가 Encoder를 거쳐서 latent vector인 μ와 σ를 추출한다. reparameterization trick을 사용해 $\varepsilon \sim N(0, I)$를 transform 한다. (이렇게해야 복원 loss로 부터 Encoder 학습가능) transformed 값을 Decoder를 통해 복원한다.

② backward

위에서 증명한 $\tilde{L}$을 loss값으로 해서
backpropagation을 진행한다.