

고려대학교  
빅데이터 연구회

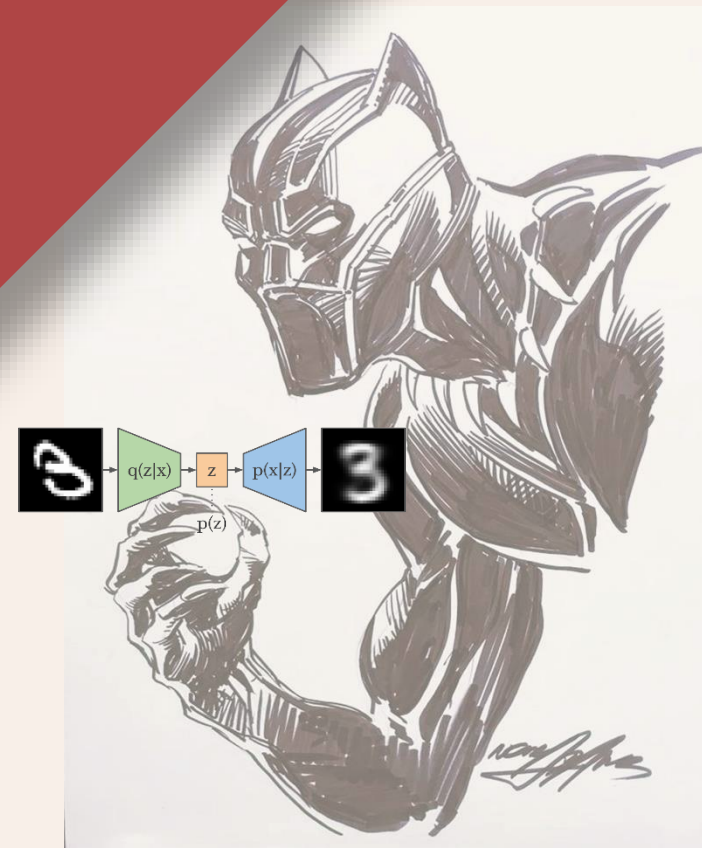
# KU-BIG

---

Credit Card Fraud Detection

- VAE algorithm

EBB



# 목 차

I

Autoencoder

II

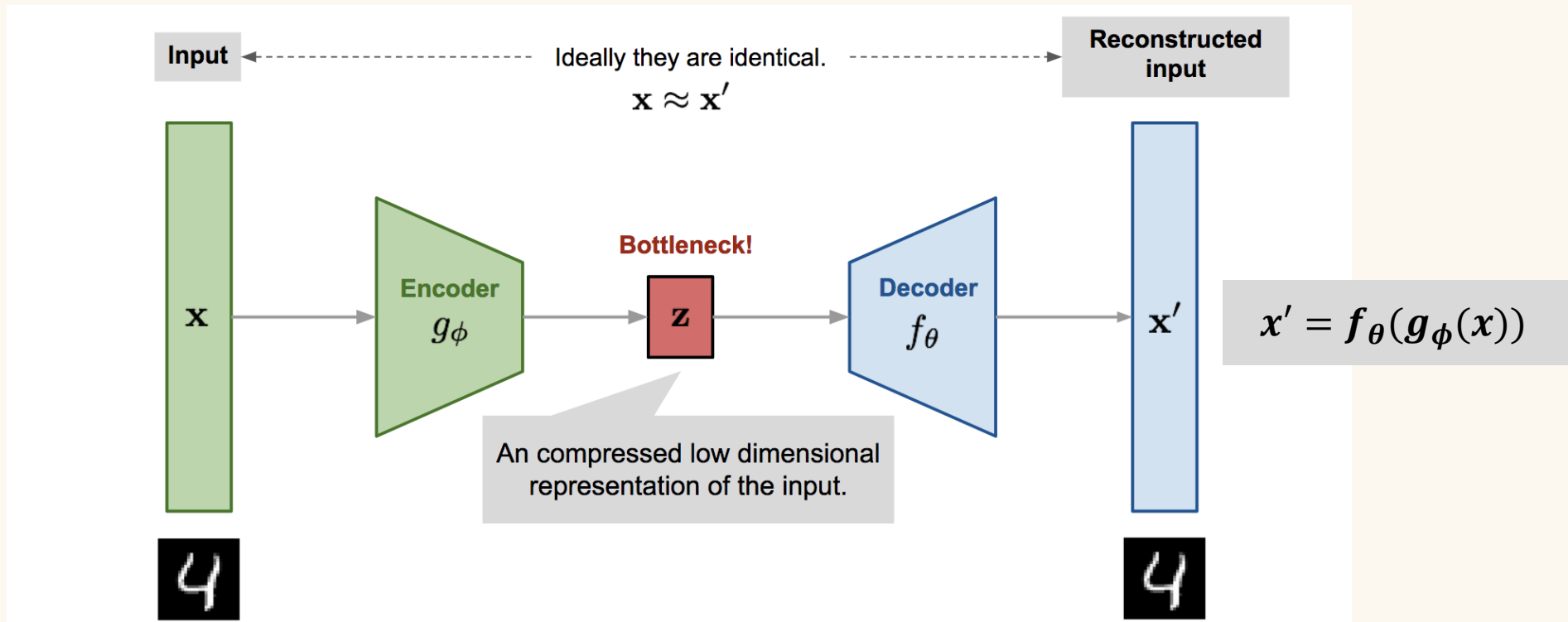
Variational AutoEncoder

III

프로젝트 현황

IV

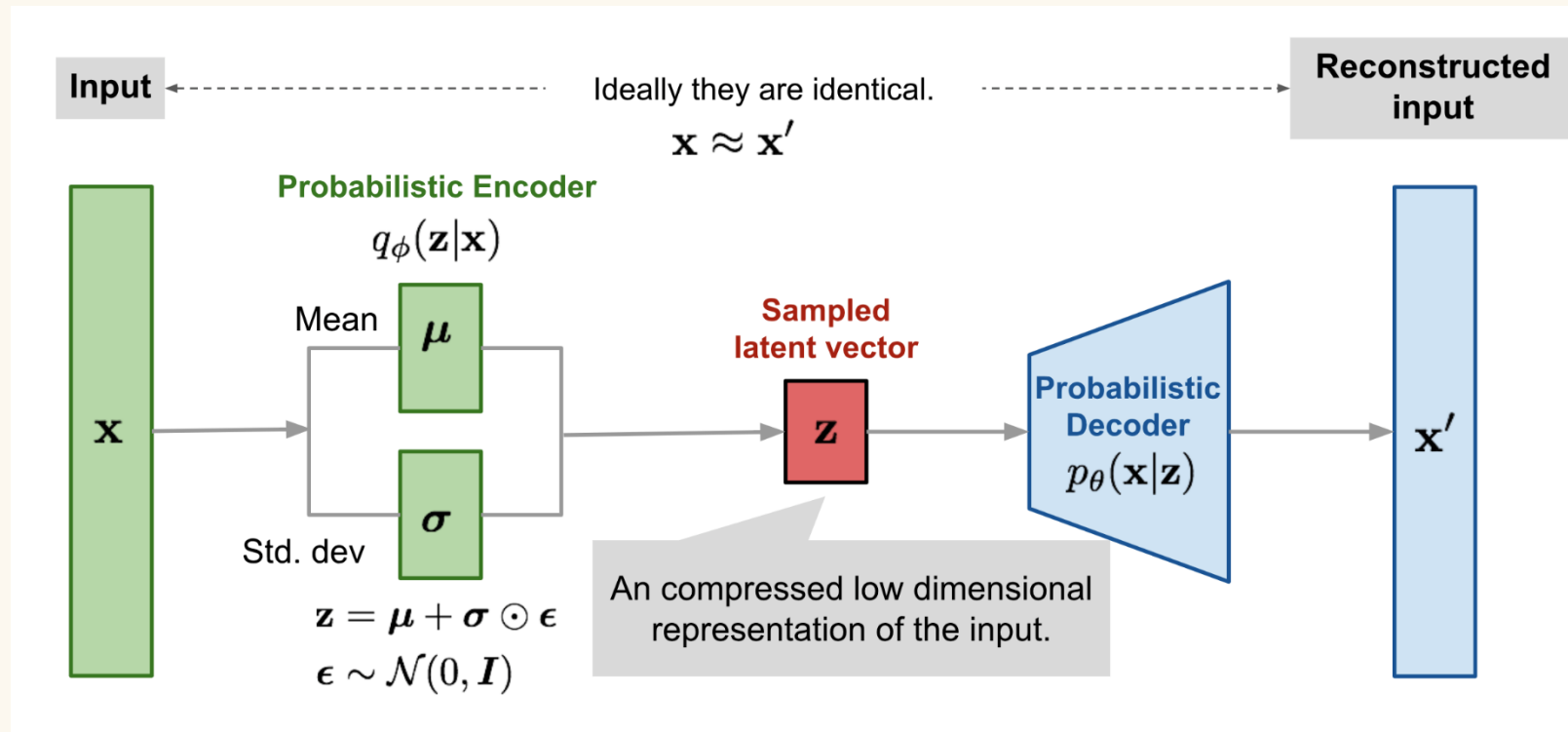
# Autoencoder?



<https://lilianweng.github.io/lil-log/2018/08/12/from-autoencoder-to-beta-vae.html>

$$\text{MSE loss function : } L_{AE}(\phi, \theta) = \frac{1}{n} \sum_{i=1}^n (x^{(i)} - f_\theta(g_\phi(x^{(i)})))^2$$

# Variational Autoencoder (VAE)?



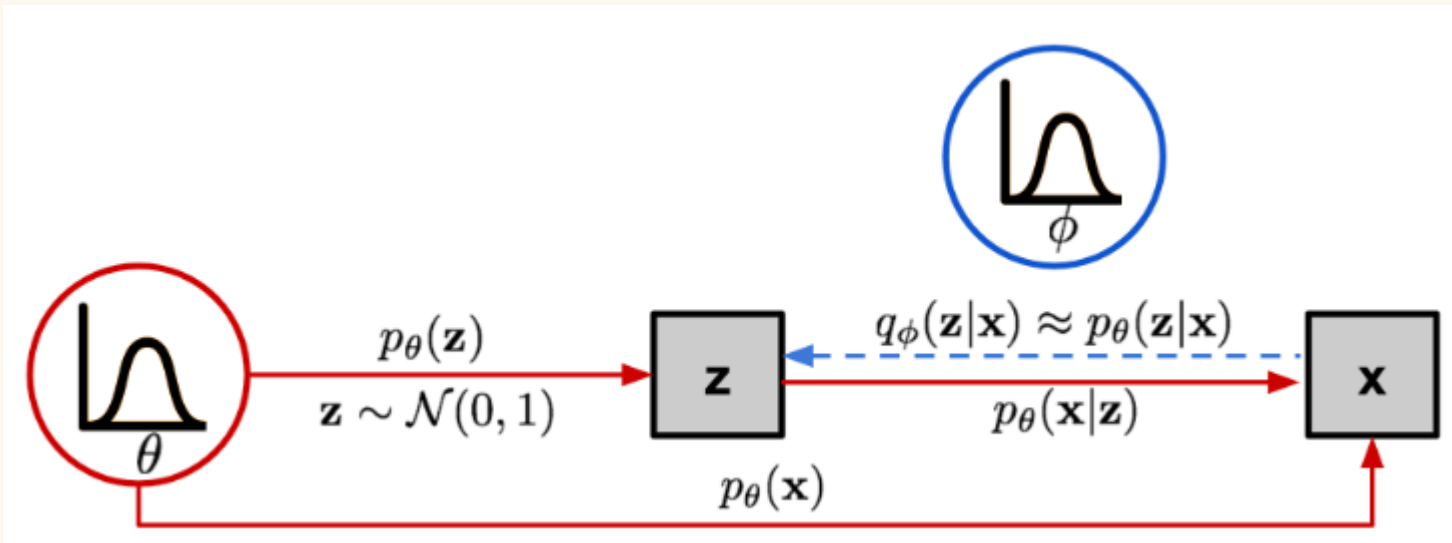
**<Decoder>**  
 $\theta^*$  : true parameter  
 1.  $p_{\theta^*}(\mathbf{z})$ 에서  $\mathbf{z}^{(i)}$  추출.  
 2.  $p_{\theta^*}(\mathbf{x}|\mathbf{z}^{(i)})$ 에서  $\mathbf{x}^{(i)}$  생성



Optimal parameter :  $\theta^* = \arg \max_{\theta} \sum_{i=1}^n \log p_{\theta}(\mathbf{x}^{(i)})$

**Intractable !!**

# Variational Autoencoder (VAE)?



$z$  : latent variable

$p_{\theta}(z)$  : prior

$p_{\theta}(x|z)$  : Likelihood

$p_{\theta}(z|x)$  : posterior

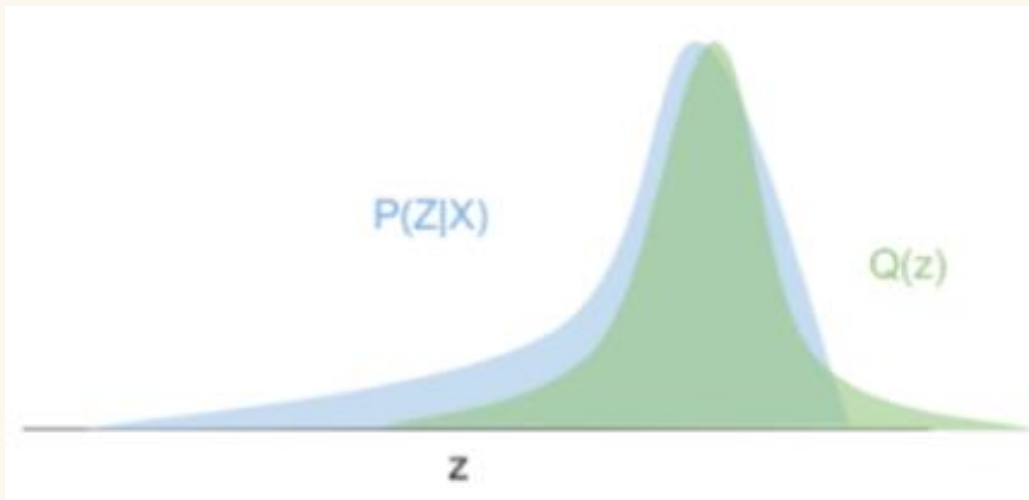
알려져 있는 분포 활용

$p_{\theta}(z)$ 를 추론하고자  $p_{\theta}(z|x)$ 를 이용하고,  $p_{\theta}(z|x)$ 를 추론하고자  $q_{\phi}(x|z)$ 를 활용한다.

# Variational Autoencoder (VAE)?

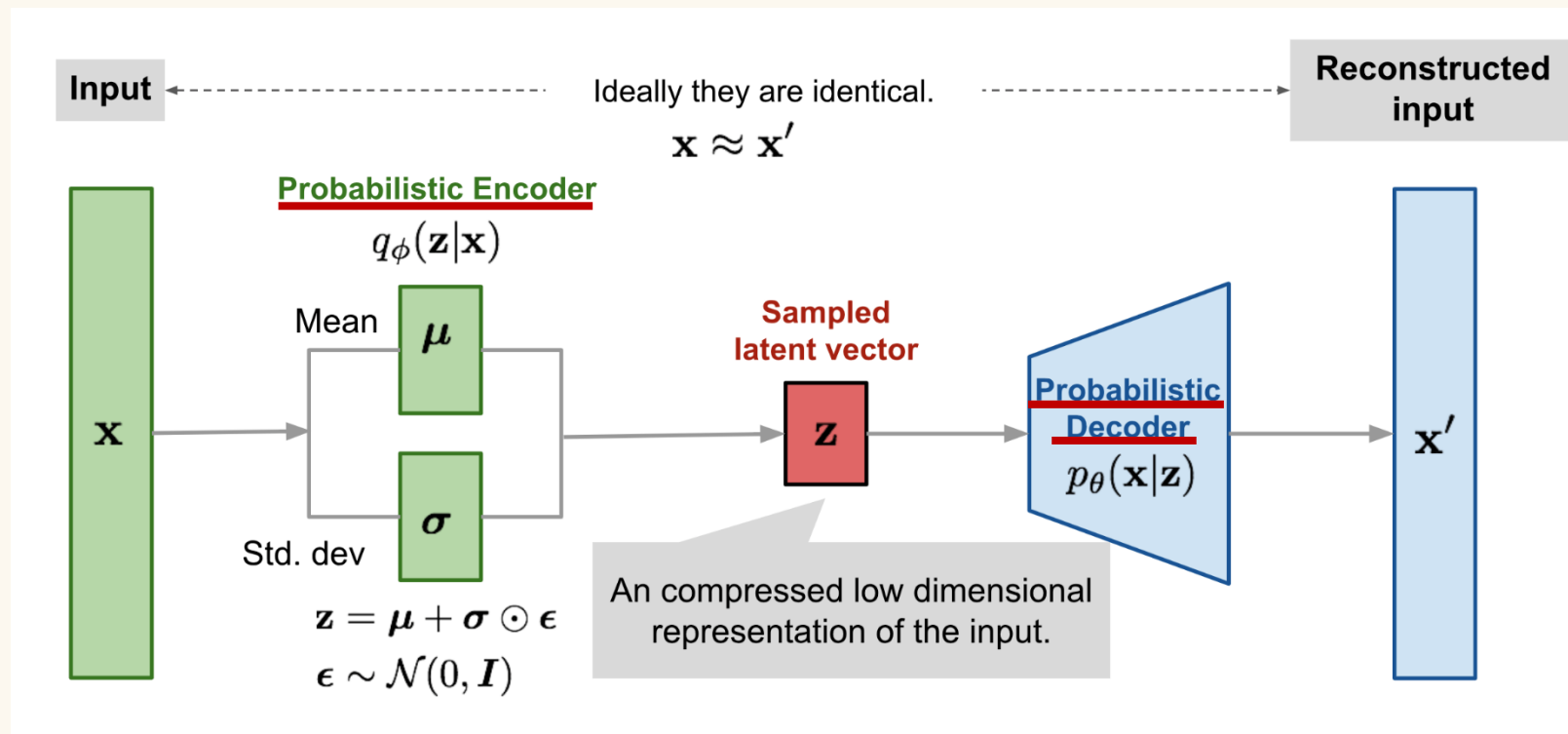
$p_{\theta}(\mathbf{z})$ 를 추론하고자  $p_{\theta}(\mathbf{z}|\mathbf{x})$ 를 이용하고,  $p_{\theta}(\mathbf{z}|\mathbf{x})$ 를 추론하고자  $q_{\phi}(\mathbf{x}|\mathbf{z})$ 를 활용한다.

Variational Inference



1.  $p_{\theta}(\mathbf{x})$ 를 계산하기 힘든 경우
2.  $p_{\theta}(\mathbf{z}), p_{\theta}(\mathbf{x}|\mathbf{z})$ 를 더 복잡하게 모델링하고 싶은 경우

# Variational Autoencoder (VAE)?



Loss function :

$$\begin{aligned}
 L_{\text{VAE}}(\theta, \phi) &= -\log p_{\theta}(\mathbf{x}) + D_{\text{KL}}(q_{\phi}(\mathbf{z}|\mathbf{x}) \| p_{\theta}(\mathbf{z}|\mathbf{x})) \\
 &= -\mathbb{E}_{\mathbf{z} \sim q_{\phi}(\mathbf{z}|\mathbf{x})} \log p_{\theta}(\mathbf{x}|\mathbf{z}) + D_{\text{KL}}(q_{\phi}(\mathbf{z}|\mathbf{x}) \| p_{\theta}(\mathbf{z})) \\
 \theta^*, \phi^* &= \arg \min_{\theta, \phi} L_{\text{VAE}}
 \end{aligned}$$

# $L_{VAE}(\theta, \phi)$ 뜯어보기.

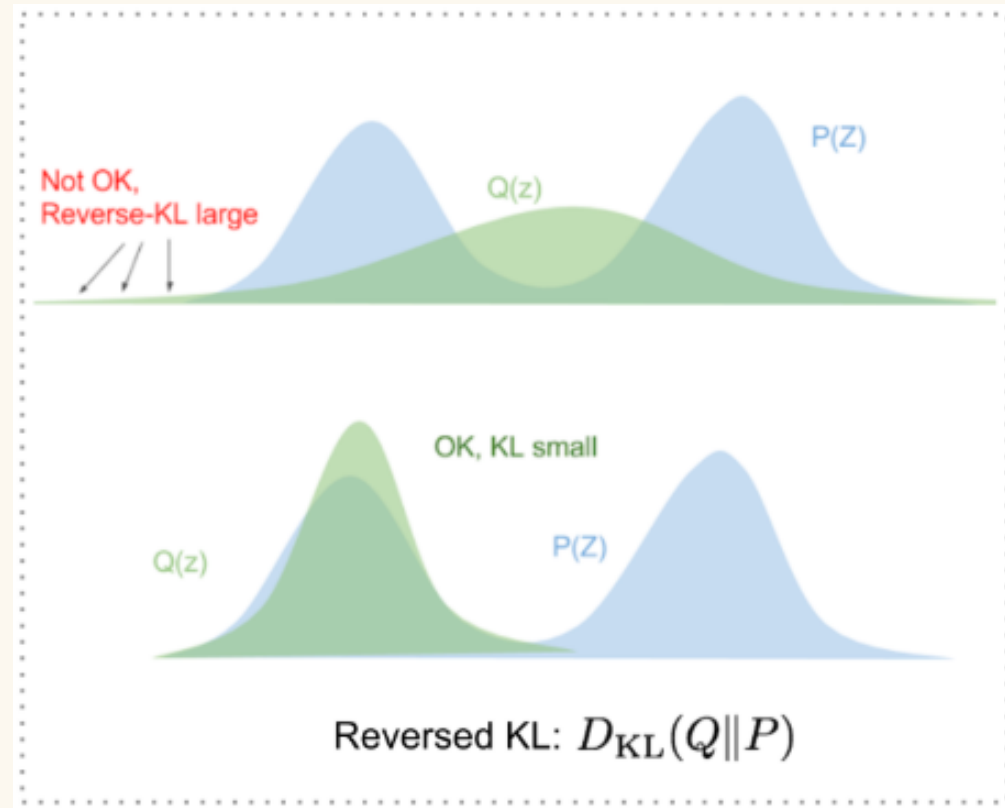
$$\begin{aligned} L_{VAE}(\theta, \phi) &= -\log p_{\theta}(\mathbf{x}) + \underline{D_{KL}(q_{\phi}(\mathbf{z}|\mathbf{x}) \| p_{\theta}(\mathbf{z}|\mathbf{x}))} \\ &= -\mathbb{E}_{\mathbf{z} \sim q_{\phi}(\mathbf{z}|\mathbf{x})} \log p_{\theta}(\mathbf{x}|\mathbf{z}) + D_{KL}(q_{\phi}(\mathbf{z}|\mathbf{x}) \| p_{\theta}(\mathbf{z})) \\ \theta^*, \phi^* &= \arg \min_{\theta, \phi} L_{VAE} \end{aligned}$$

$q_{\phi}(\mathbf{z}|\mathbf{x})$ 는  $p_{\theta}(\mathbf{z}|\mathbf{x})$ 에 최대한 비슷해야 한다.

→ **KL-Divergence**

$$D_{KL}(Q \| P) = \mathbb{E}_{z \sim Q(z)} \log \frac{Q(z)}{P(z)} \geq 0$$

→ **Minimize**  $D_{KL}(Q \| P)$





$$D_{\text{KL}}(q_{\phi}(\mathbf{z}|\mathbf{x})||p_{\theta}(\mathbf{z}|\mathbf{x}))$$

$$= \int q_{\phi}(\mathbf{z}|\mathbf{x}) \log \frac{q_{\phi}(\mathbf{z}|\mathbf{x})}{p_{\theta}(\mathbf{z}|\mathbf{x})} d\mathbf{z}$$

$$= \int q_{\phi}(\mathbf{z}|\mathbf{x}) \log \frac{q_{\phi}(\mathbf{z}|\mathbf{x})p_{\theta}(\mathbf{x})}{p_{\theta}(\mathbf{z}, \mathbf{x})} d\mathbf{z}$$

$$D_{\text{KL}}(Q||P) = \mathbb{E}_{z \sim Q(z)} \log \frac{Q(z)}{P(z)}$$

; Because  $p(z|x)=p(z,x)/p(x)$

$$= \int q_{\phi}(\mathbf{z}|\mathbf{x}) (\log p_{\theta}(\mathbf{x}) + \log \frac{q_{\phi}(\mathbf{z}|\mathbf{x})}{p_{\theta}(\mathbf{z}, \mathbf{x})}) d\mathbf{z}$$

$$= \log p_{\theta}(\mathbf{x}) + \int q_{\phi}(\mathbf{z}|\mathbf{x}) \log \frac{q_{\phi}(\mathbf{z}|\mathbf{x})}{p_{\theta}(\mathbf{z}, \mathbf{x})} d\mathbf{z}$$

; Because  $\int q(z|x)dz=1$

$$= \log p_{\theta}(\mathbf{x}) + \int q_{\phi}(\mathbf{z}|\mathbf{x}) \log \frac{q_{\phi}(\mathbf{z}|\mathbf{x})}{p_{\theta}(\mathbf{x}|\mathbf{z})p_{\theta}(\mathbf{z})} d\mathbf{z}$$

; Because  $p(z,x)=p(x|z)p(z)$

$$= \log p_{\theta}(\mathbf{x}) + \mathbb{E}_{\mathbf{z} \sim q_{\phi}(\mathbf{z}|\mathbf{x})} [\log \frac{q_{\phi}(\mathbf{z}|\mathbf{x})}{p_{\theta}(\mathbf{z})} - \log p_{\theta}(\mathbf{x}|\mathbf{z})]$$

$$= \log p_{\theta}(\mathbf{x}) + D_{\text{KL}}(q_{\phi}(\mathbf{z}|\mathbf{x})||p_{\theta}(\mathbf{z})) - \mathbb{E}_{\mathbf{z} \sim q_{\phi}(\mathbf{z}|\mathbf{x})} \log p_{\theta}(\mathbf{x}|\mathbf{z})$$

$g(\mathbf{z})$

$\mathbb{E}_{\mathbf{z} \sim q_{\phi}(\mathbf{z}|\mathbf{x})} [g(\mathbf{z})]$

$$D_{\text{KL}}(q_{\phi}(\mathbf{z}|\mathbf{x})||p_{\theta}(\mathbf{z}|\mathbf{x})) = \log p_{\theta}(\mathbf{x}) + D_{\text{KL}}(q_{\phi}(\mathbf{z}|\mathbf{x})||p_{\theta}(\mathbf{z})) - \mathbb{E}_{\mathbf{z} \sim q_{\phi}(\mathbf{z}|\mathbf{x})} \log p_{\theta}(\mathbf{x}|\mathbf{z})$$

$$\log p_{\theta}(\mathbf{x}) - \underbrace{D_{\text{KL}}(q_{\phi}(\mathbf{z}|\mathbf{x})||p_{\theta}(\mathbf{z}|\mathbf{x}))}_{\geq 0} = \mathbb{E}_{\mathbf{z} \sim q_{\phi}(\mathbf{z}|\mathbf{x})} \log p_{\theta}(\mathbf{x}|\mathbf{z}) - D_{\text{KL}}(q_{\phi}(\mathbf{z}|\mathbf{x})||p_{\theta}(\mathbf{z}))$$

$$\log p_{\theta}(\mathbf{x}) \geq \underbrace{\mathbb{E}_{\mathbf{z} \sim q_{\phi}(\mathbf{z}|\mathbf{x})} \log p_{\theta}(\mathbf{x}|\mathbf{z}) - D_{\text{KL}}(q_{\phi}(\mathbf{z}|\mathbf{x})||p_{\theta}(\mathbf{z}))}_{\text{Lower bound of } \log p_{\theta}(\mathbf{x})} = -L_{\text{VAE}}$$

Lower bound of  $\log p_{\theta}(\mathbf{x})$

Maximize

$$\begin{aligned} L_{\text{VAE}}(\theta, \phi) &= -\log p_{\theta}(\mathbf{x}) + \underbrace{D_{\text{KL}}(q_{\phi}(\mathbf{z}|\mathbf{x})||p_{\theta}(\mathbf{z}|\mathbf{x}))}_{\text{Reconstruction probability}} \\ &= -\mathbb{E}_{\mathbf{z} \sim q_{\phi}(\mathbf{z}|\mathbf{x})} \log p_{\theta}(\mathbf{x}|\mathbf{z}) + D_{\text{KL}}(q_{\phi}(\mathbf{z}|\mathbf{x})||p_{\theta}(\mathbf{z})) \end{aligned}$$

$$\theta^*, \phi^* = \arg \min_{\theta, \phi} L_{\text{VAE}}$$

Reconstruction probability

Minimize Loss → Maximize real data sample을 생성하는 확률의 lower bound

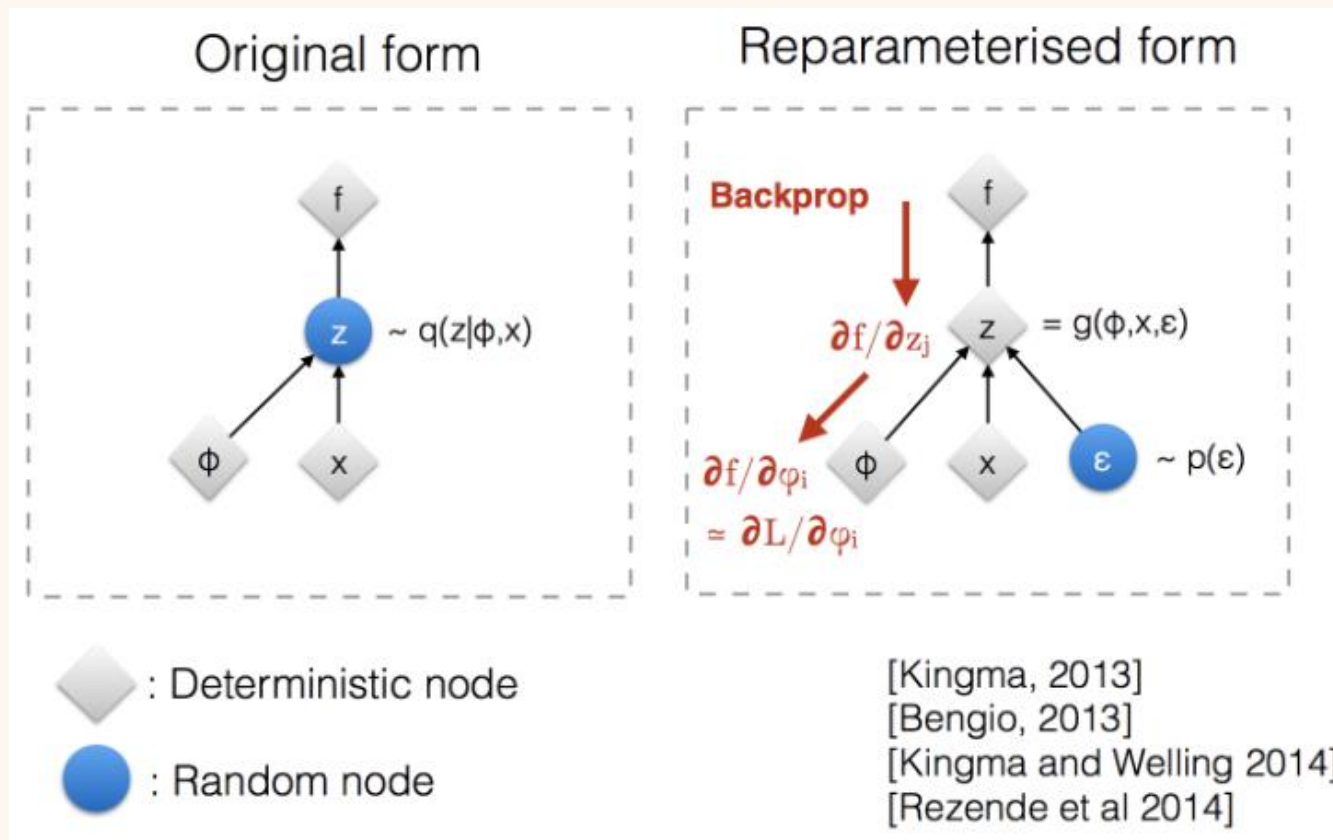
# Reparameterization trick

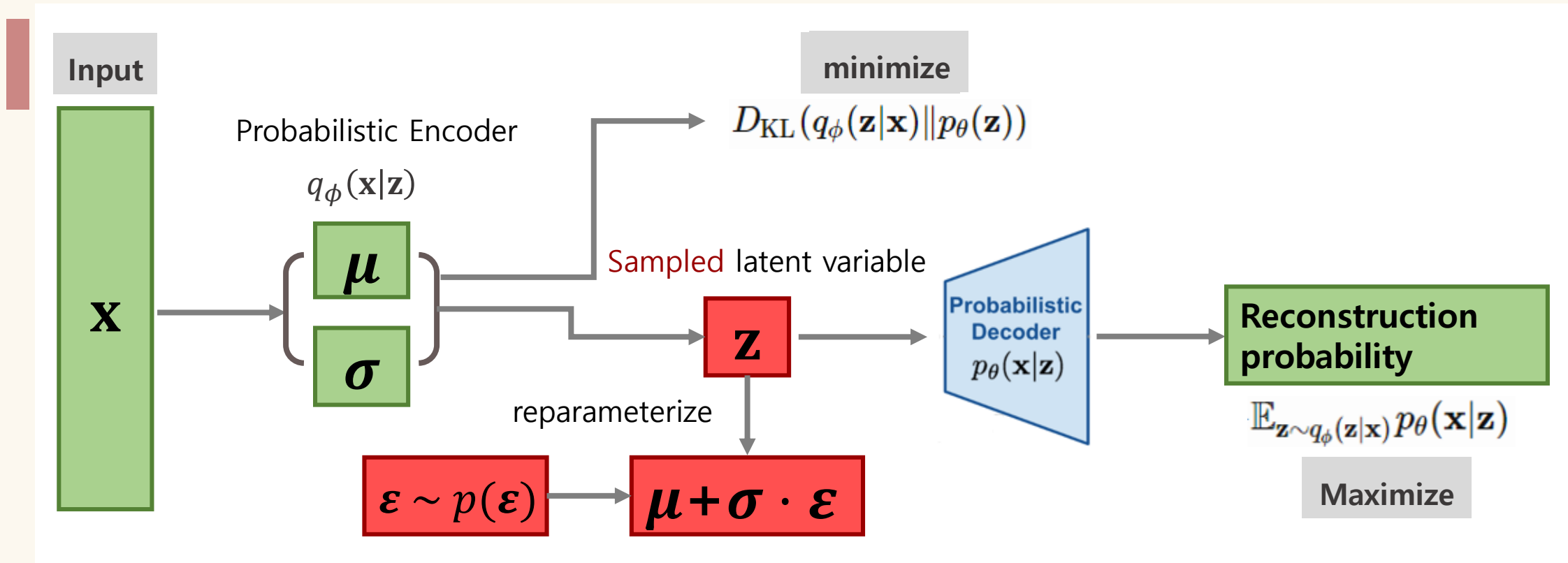
$$\mathbf{z} \sim q_{\phi}(\mathbf{z}|\mathbf{x}^{(i)}) = \mathcal{N}(\mathbf{z}; \boldsymbol{\mu}^{(i)}, \boldsymbol{\sigma}^{2(i)} \mathbf{I})$$

$$\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}, \text{ where } \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I})$$

$\odot$  refers to element-wise product.

Stochastic process(확률과정)라  
Backprop이 불가능한  $\mathbf{z}$ 를  
미분 가능하게 만들어서  
Backprop을 가능하게 만든다!!!!!!!!!!!!!!





$$\begin{aligned}
 L_{\text{VAE}}(\theta, \phi) &= -\log p_\theta(\mathbf{x}) + D_{\text{KL}}(q_\phi(\mathbf{z}|\mathbf{x}) \| p_\theta(\mathbf{z}|\mathbf{x})) \\
 &= -\mathbb{E}_{\mathbf{z} \sim q_\phi(\mathbf{z}|\mathbf{x})} \log p_\theta(\mathbf{x}|\mathbf{z}) + D_{\text{KL}}(q_\phi(\mathbf{z}|\mathbf{x}) \| p_\theta(\mathbf{z})) \\
 \theta^*, \phi^* &= \arg \min_{\theta, \phi} L_{\text{VAE}}
 \end{aligned}$$

**Algorithm 3** Variational autoencoder training algorithm

**INPUT:** Dataset  $x^{(1)}, \dots, x^{(N)}$

**OUTPUT:** probabilistic encoder  $f_\phi$ , probabilistic decoder  $g_\theta$

$\phi, \theta \leftarrow$  Initialize parameters

**repeat**

**for**  $i=1$  **to**  $N$  **do**

    Draw  $L$  samples from  $\epsilon \sim \mathcal{N}(0, 1)$

$z^{(i,l)} = h_\phi(\epsilon^{(i)}, x^{(i)}) \quad i = 1, \dots, N$

Reparameterization trick

**end for**

$E = \sum_{i=1}^N -D_{KL}(q_\phi(z|x^{(i)})||p_\theta(z)) + \frac{1}{L} \sum_{l=1}^L (\log p_\theta(x^{(i)}|z^{(i,l)}))$

loss

$\phi, \theta \leftarrow$  Update parameters using gradients of  $E$  (e.g. Stochastic Gradient Descent)

**until** convergence of parameters  $\phi, \theta$

**Algorithm 4** Variational autoencoder based anomaly detection algorithm

**INPUT:** Normal dataset  $X$ , Anomalous dataset  $x^{(i)} \quad i = 1, \dots, N$ , threshold  $\alpha$

**OUTPUT:** reconstruction probability  $p_{\theta}(x|\hat{x})$

$\phi, \theta \leftarrow$  train a variational autoencoder using the normal dataset  $X$

for  $i=1$  to  $N$  do

$\mu_{z^{(i)}}, \sigma_{z^{(i)}} = f_{\theta}(z|x^{(i)})$

draw  $L$  samples from  $z \sim \mathcal{N}(\mu_{z^{(i)}}, \sigma_{z^{(i)}})$

for  $l=1$  to  $L$  do

$\mu_{\hat{x}^{(i,l)}}, \sigma_{\hat{x}^{(i,l)}} = g_{\phi}(x|z^{(i,l)})$

end for

reconstruction probability $(i) = \frac{1}{L} \sum_{l=1}^L p_{\theta}(x^{(i)}|\mu_{\hat{x}^{(i,l)}}, \sigma_{\hat{x}^{(i,l)}})$

if reconstruction probability $(i) < \alpha$  then

$x^{(i)}$  is an anomaly

else

$x^{(i)}$  is not an anomaly

end if

end for

Encoder로부터  $z$ 분포에 대한  
parameter 추출

Decoder로부터  $\hat{x}$ 분포에 대한  
(reconstruction) parameter 추출

reconstruction parameter가 주어졌을 때,  
original input  $x$ 가 생성될 확률의 평균

```

class VAE(nn.Module):
    def __init__(self, image_size = 784, h_dim = 400, z_dim = 20):
        super(VAE, self).__init__()
        self.fc1 = nn.Linear(image_size, h_dim)
        self.fc2 = nn.Linear(h_dim, z_dim)
        self.fc3 = nn.Linear(h_dim, z_dim)
        self.fc4 = nn.Linear(z_dim, h_dim)
        self.fc5 = nn.Linear(h_dim, image_size)

    def encode(self, x):
        h = F.relu(self.fc1(x))
        return self.fc2(h), self.fc3(h)

    def reparameterize(self, mu, log_var):
        std = torch.exp(log_var/2)
        eps = torch.randn_like(std)
        return mu+eps*std

    def decode(self, z):
        h = F.relu(self.fc4(z))
        return self.fc5(h)

    def forward(self, x):
        mu, log_var = self.encode(x)
        z = self.reparameterize(mu, log_var)
        x_reconst = self.decode(z)
        return x_reconst, mu, log_var

model = VAE().to(device)
optimizer = torch.optim.Adam(model.parameters(), lr = learning_rate)

```

Encoder

Decoder

z\_dim : parameter쌍의 개수  
 fc2, fc3는  $\mu$ 와  $\sigma$ 를 도출함.  
 fc2, fc3 중에 어느 것이  $\mu$ 인지,  
 $\sigma$ 인지는 알 수 없음.

fc2가  $\mu$ 가 되도록, fc3이  $\sigma$ 가  
 되도록 최적화 시킬 뿐임.  
 (맞는지 모르겠음다  $\pi \pi$ )

빛희정 님의 은혜로운 코드. [https://github.com/KU-BIG/novelty-detection/blob/master/novelty\\_detection\\_pytorch.ipynb](https://github.com/KU-BIG/novelty-detection/blob/master/novelty_detection_pytorch.ipynb)

A stylized illustration of a person from the chest up, wearing a grey suit jacket, a white shirt, and a dark tie. The person's face is partially visible at the top, showing a red mouth and brown skin. A large, black-outlined speech bubble originates from the mouth area. Inside the speech bubble, the text "Q&A Time!" is written. The background is a solid light beige color.

Q&A  
Time!

Thank you  
for your attention.