Auto-Encoding Variational Bayes(VAE)

김기웅
kwkim02@g.skku.edu
Computer Vision Core

2024/09/10

TRAIN AND TEST

Contents

- Introduction
- Background
- Architecture
- Method
- Experiment

TRAIN AND TEST

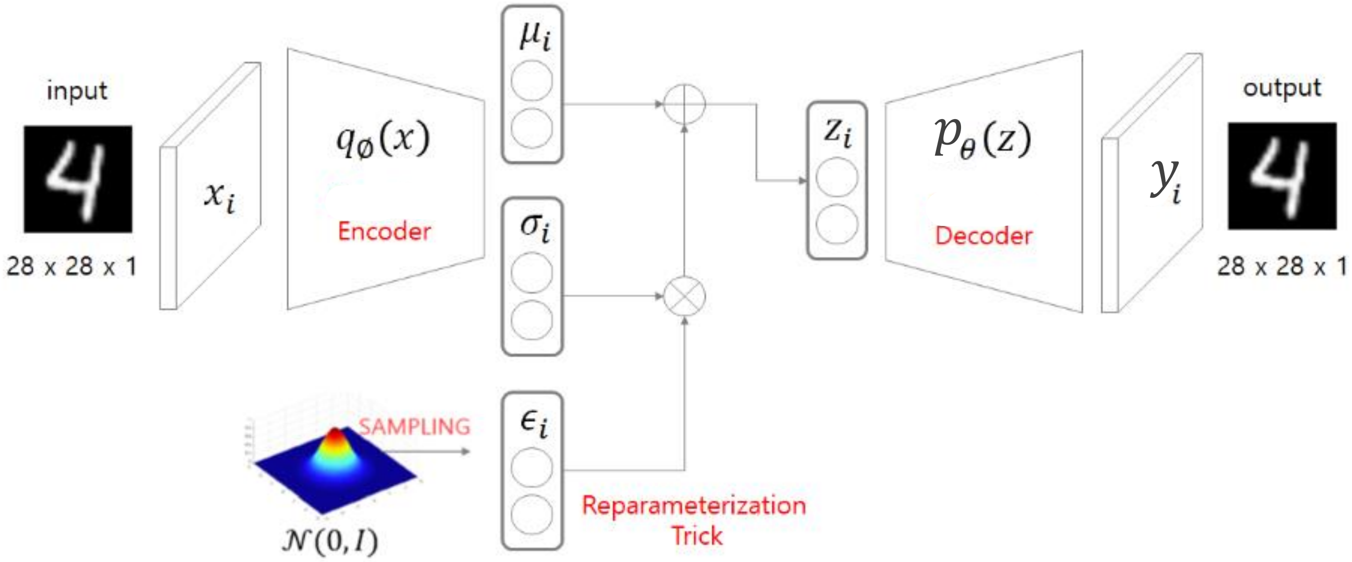
Introduction

VAE

- x에 대해 Likelihood를 최대화하는 p(x|z) 을 구하는 것이 목적
  p(x|z)의 분포를 알면, z를 샘플링하여 latent space 조작 가능
  학습 데이터 x와 유사한 새로운 데이터를 생성할 수 있다.

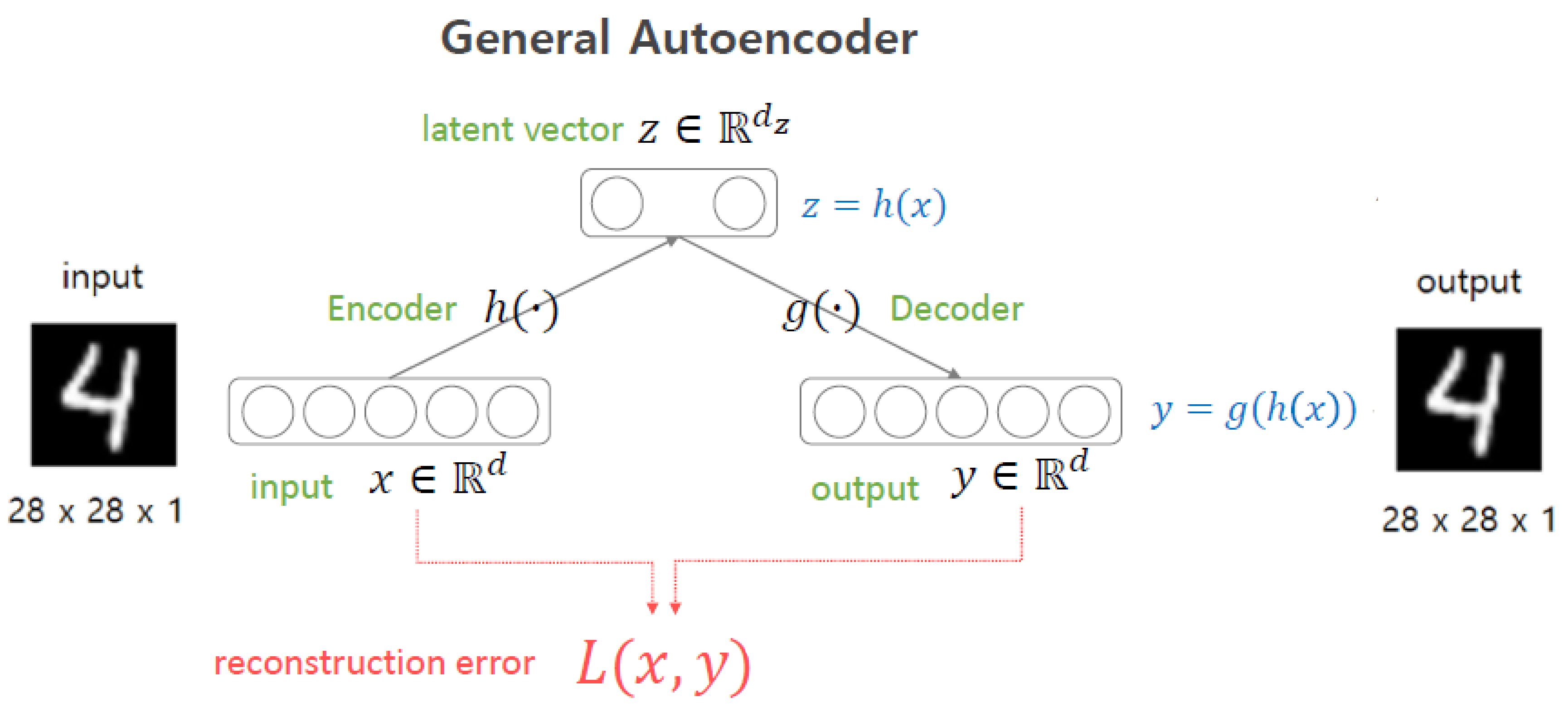
- Generative Model

- Unsupervised Learning



Background

AutoEncoder

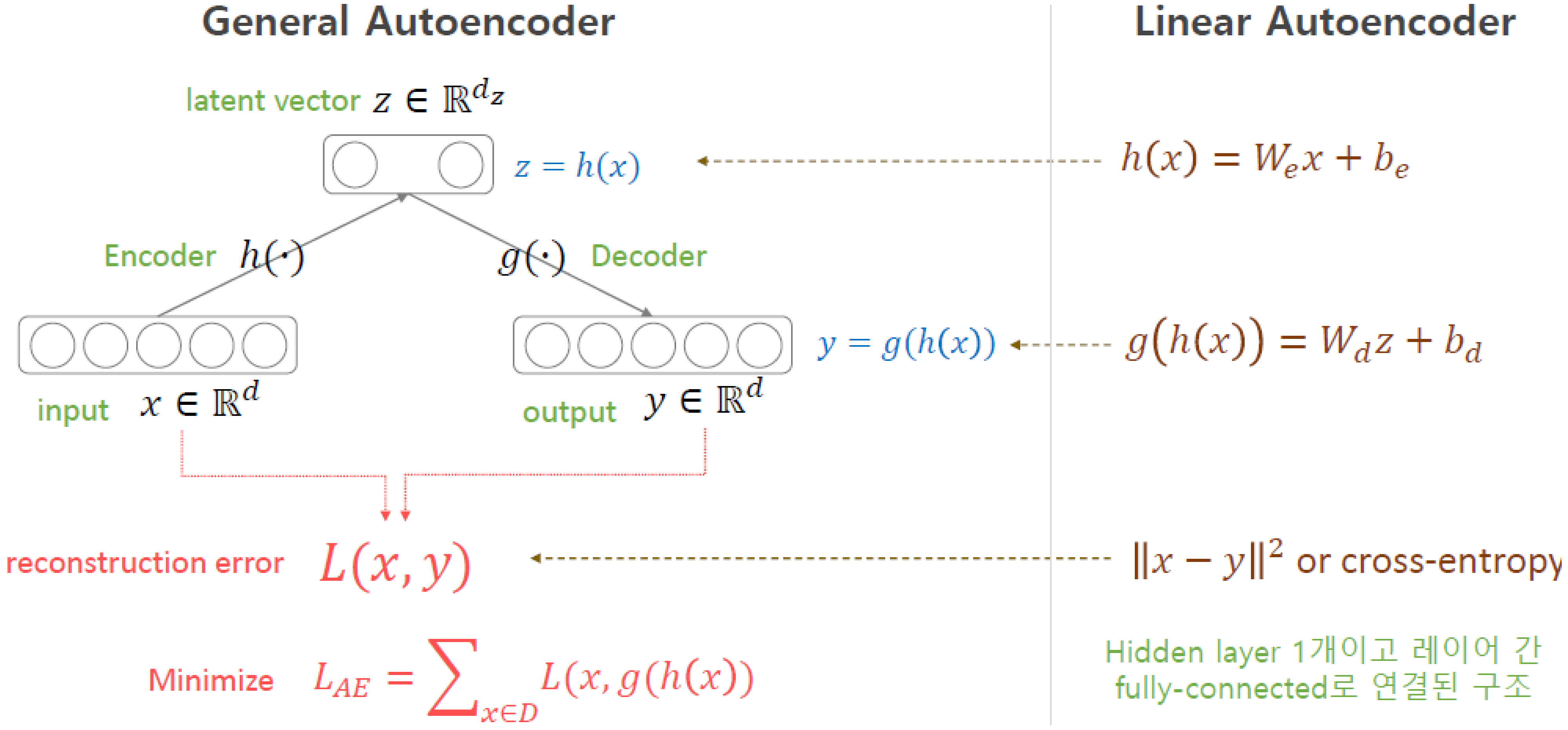


Minimize $L_{AE} = \sum_{x \in D} L(x, g(h(x)))$

Background

AutoEncoder

- Feature Learning
- Dimension Reduction
- Data Compression

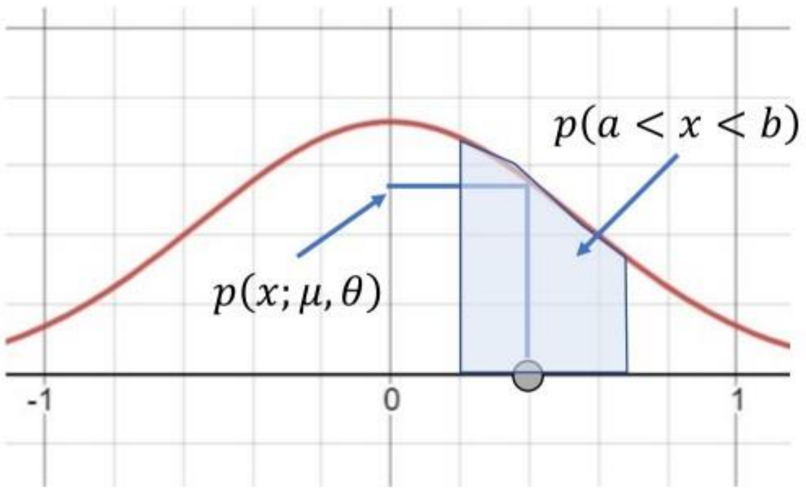


Background

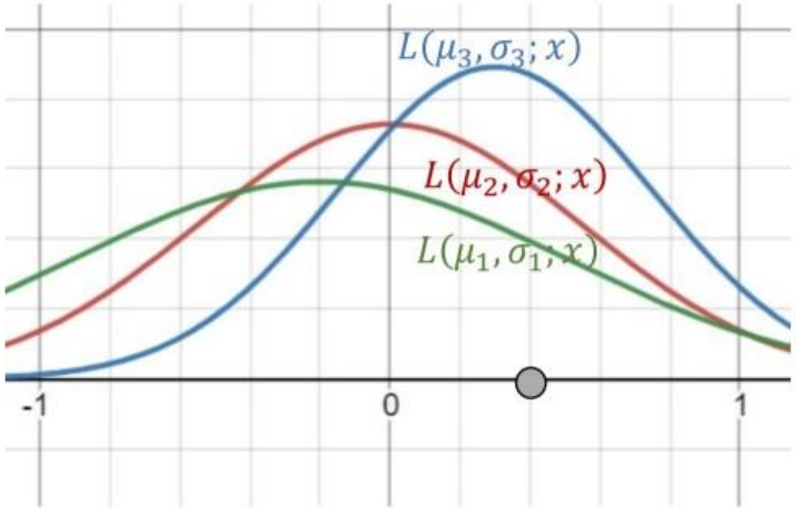
MLE
(Maximum Likelihood Estimation)

Likelihood: 데이터가 특정 분포(distribution)으로부터 만들어졌을 확률

Probability Density Function



Likelihood



Background

Bayes Rule

$$P(B \mid A) = \frac{\overset{\text{likelihood}}{P(A \mid B)}\, P(B)}{P(A)}$$

$$\underset{\substack{\textit{a priori} \\ \text{probability}}}{P(B)} \Rightarrow \underset{\substack{\textit{a posteriori} \\ \text{probability}}}{P(B \mid A)}$$

Ex)
한 공장에서 M1, M2, M3 기계
각 기계는 전체의 10%, 30%, 60%
각 기계를 사용했을 때 불량률은 1%, 2%, 3%
불량품이 발생했을 때, 기계 M1에서
발생했을 확률을 구하시오.

- $P(M_1|B) = \frac{P(M_1)P(B|M_1)}{P(M_1)P(B|M_1)+P(M_2)P(B|M_2)+P(M_3)P(B|M_3)}$
- $P(M_1|B) = \frac{(0.1)(0.01)}{(0.1)(0.01)+(0.3)(0.02)+(0.6)(0.03)}$

$$P(B) = \sum_{all\ A} P(B|A) \times P(A)$$

Background

KL Divergence

$$
\begin{aligned}
KL(P \parallel Q) &= \sum_{x \in \chi} P(x) \log_b \left(\frac{P(x)}{Q(x)} \right) \\
&\Rightarrow -\sum_{x \in \chi} P(x) \log_b \left(\frac{Q(x)}{P(x)} \right) \\
&= -\sum_{x \in \chi} P(x) \log_b Q(x) + \sum_{x \in \chi} P(x) \log_b P(x) \\
&\Rightarrow \underbrace{-E_P[\log_b Q(x)]}_{\text{P 기준에서 q의 cross entropy}} + \underbrace{E_P[\log_b P(x)]}_{\text{P의 entropy}}
\end{aligned}
$$

P 기준에서 q의 cross entropy

P의 entropy

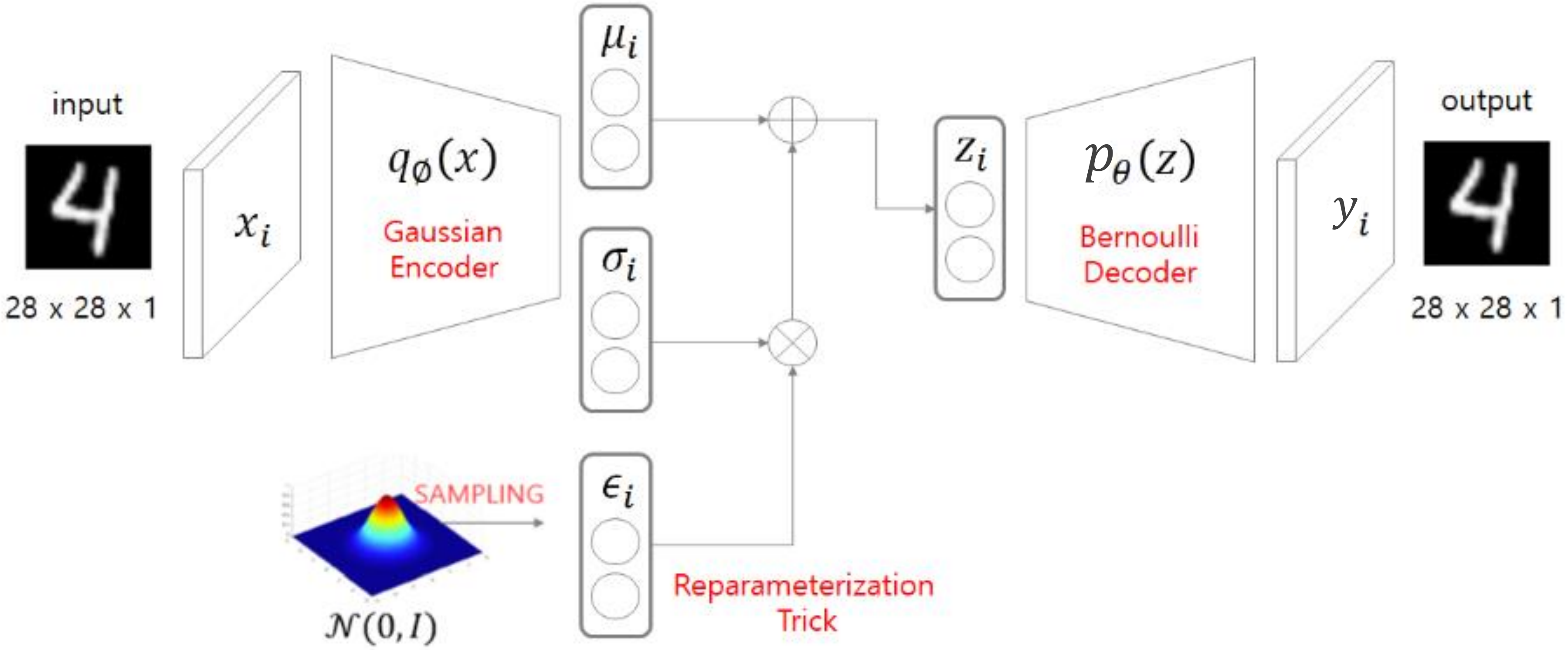
- 추정 분포 Q가 실제 분포 P와 얼마나 다른지
- Q로 P를 표현하는 데에 추가적으로 사용되는 bit
- P와 Q가 비슷한 분포일수록 KL Divergence 값이 작다
- 항상 0이상이다

https://angeloyeo.github.io/2020/10/27/KL_divergence.html

TRAIN AND TEST

Architecture

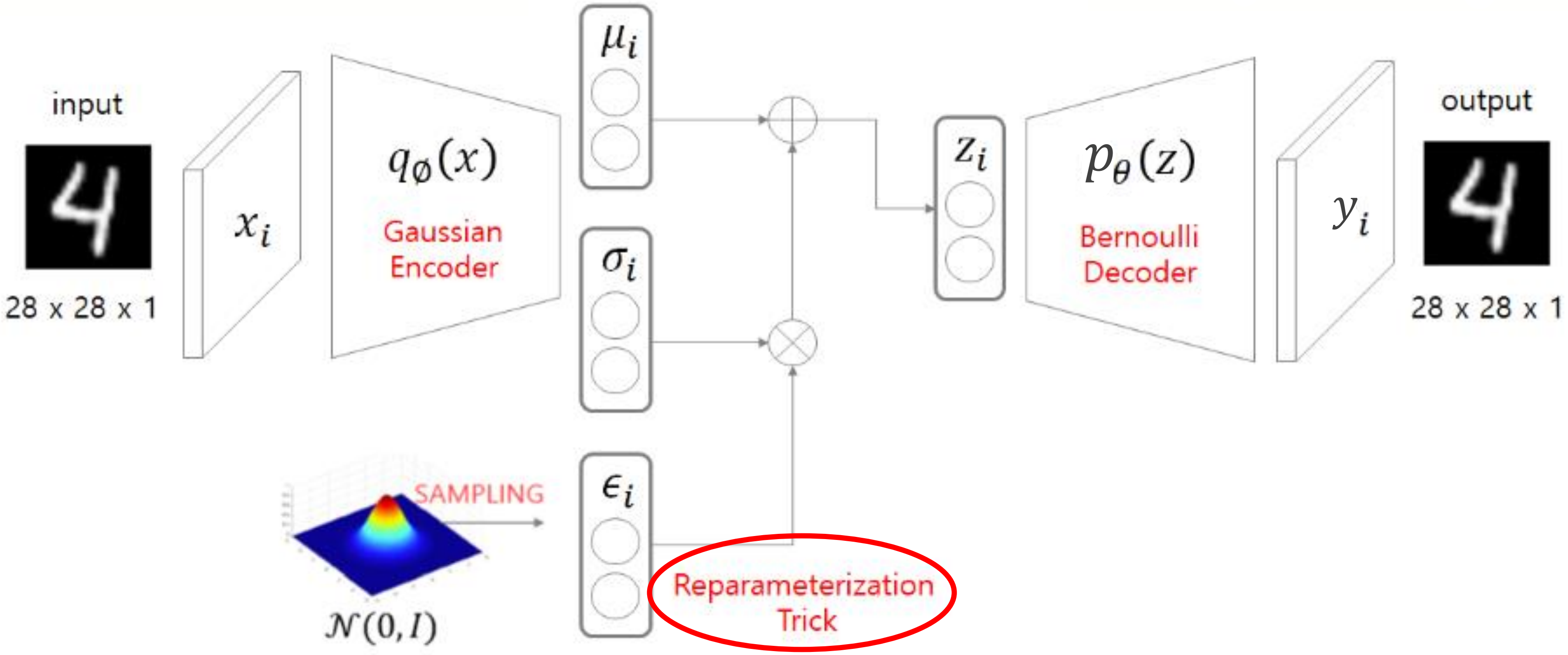
VAE



Architecture

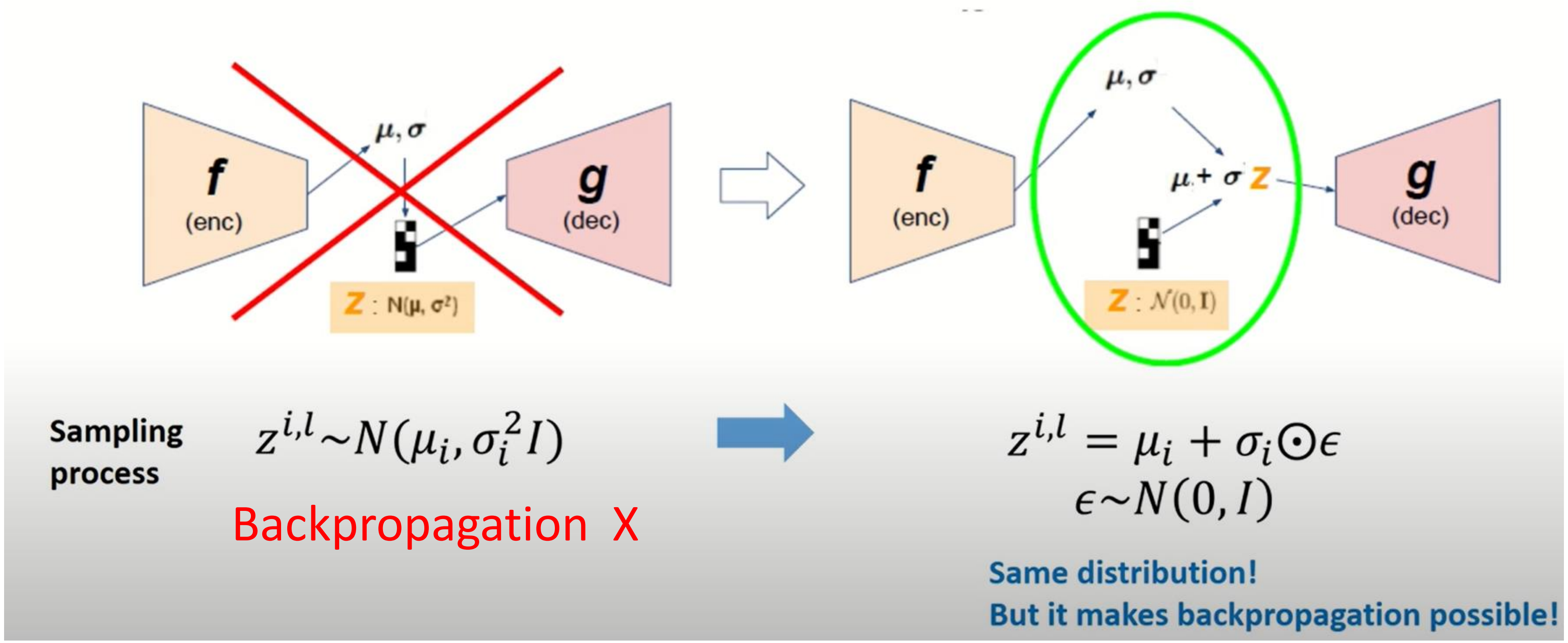
VAE

Input: $x_i \rightarrow q_\emptyset(x) \rightarrow \mu_i, \sigma_i$ $\mu_i, \sigma_i, \epsilon_i \rightarrow z_i$ $z_i \rightarrow p_\theta(z) \rightarrow y_i$:output



Method

Reparameterization trick



Method

Loss Function

Objective: $maximize \ \ p_\theta(x)$

모델이 입력 데이터 x를 잘 복원할 likelihood

$$
\begin{aligned}
\log p_\theta(x^{(i)}) &= \mathbf{E}_{z\sim q_\phi(z|x^{(i)})}\left[\log p_\theta(x^{(i)})\right] \qquad (p_\theta(x^{(i)}) \text{ Does not depend on } z) \\
&= \mathbf{E}_z\left[\log \frac{p_\theta(x^{(i)} \mid z)p_\theta(z)}{p_\theta(z \mid x^{(i)})}\right] \qquad (\text{Bayes' Rule}) \\
&= \mathbf{E}_z\left[\log \frac{p_\theta(x^{(i)} \mid z)p_\theta(z)}{p_\theta(z \mid x^{(i)})}\frac{q_\phi(z \mid x^{(i)})}{q_\phi(z \mid x^{(i)})}\right] \qquad (\text{Multiply by constant}) \\
&= \mathbf{E}_z\left[\log p_\theta(x^{(i)} \mid z)\right] - \mathbf{E}_z\left[\log \frac{q_\phi(z \mid x^{(i)})}{p_\theta(z)}\right] + \mathbf{E}_z\left[\log \frac{q_\phi(z \mid x^{(i)})}{p_\theta(z \mid x^{(i)})}\right] \qquad (\text{Logarithms}) \\
&= \mathbf{E}_z\left[\log p_\theta(x^{(i)} \mid z)\right] - D_{KL}(q_\phi(z \mid x^{(i)}) \,||\, p_\theta(z)) + D_{KL}(q_\phi(z \mid x^{(i)}) \,||\, p_\theta(z \mid x^{(i)}))
\end{aligned}
$$

intractable

tractable lower bound

≥ 0

$$
\geq \underbrace{\mathbf{E}_z\left[\log p_\theta(x^{(i)} \mid z)\right] - D_{KL}(q_\phi(z \mid x^{(i)}) \,||\, p_\theta(z))}_{\mathcal{L}(x^{(i)}, \theta, \phi)}
$$

TRAIN AND TEST

Method

Loss function

$$\log p_\theta(x^{(i)}) \geq \mathcal{L}(x^{(i)}, \theta, \phi) = \mathbf{E}_z\left[\log p_\theta(x^{(i)} \mid z)\right] - D_{KL}(q_\phi(z \mid x^{(i)}) \,||\, p_\theta(z)) \quad \text{(ELBO)}$$

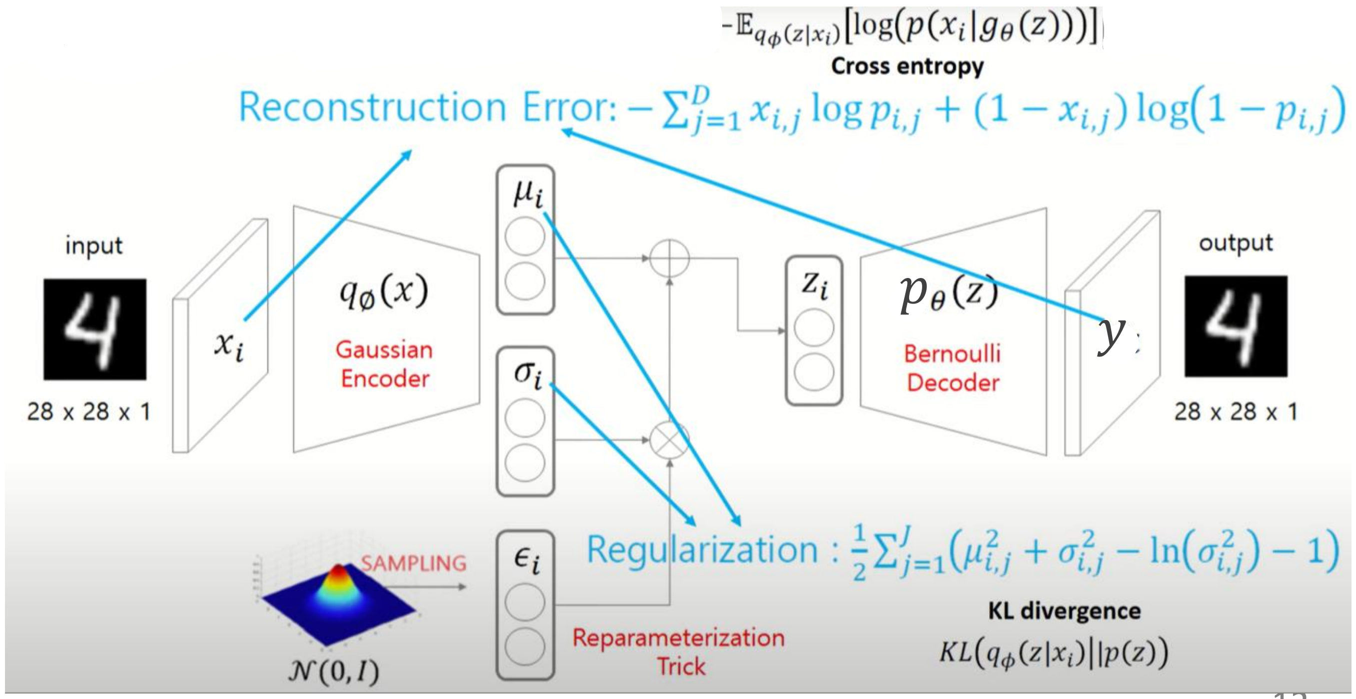
$$\boxed{\theta^*, \phi^* = \arg\max_{\theta,\phi} \sum_{i=1}^{N} \mathcal{L}(x^{(i)}, \theta, \phi)}$$

Training: Maximize lower bound

$$\underbrace{\mathbf{E}_z\left[\log p_\theta(x^{(i)} \mid z)\right]}_{\text{Reconstruction error}} - \underbrace{D_{KL}(q_\phi(z \mid x^{(i)}) \,||\, p_\theta(z))}_{\text{Regularization error}}$$

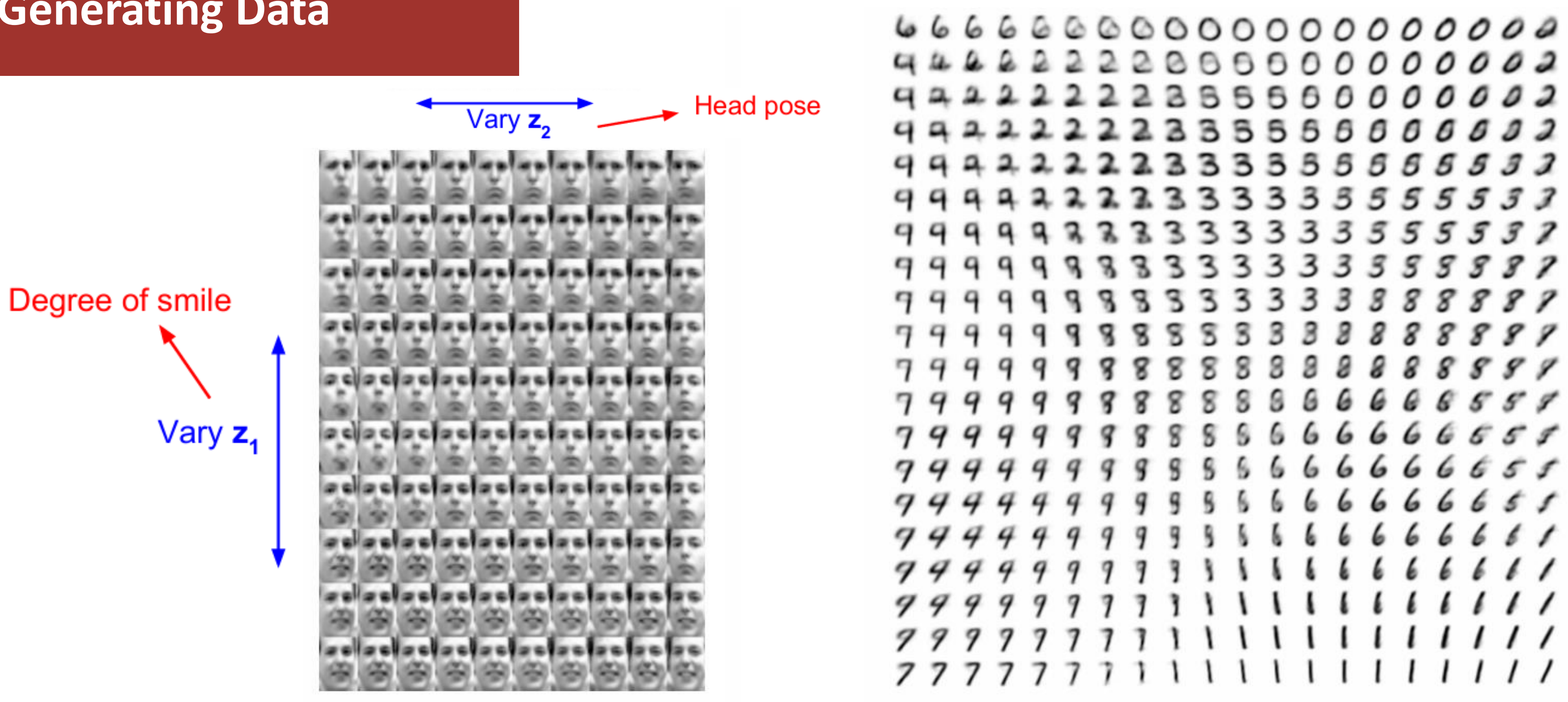
Reconstruction error: Maximize likelihood of original input being reconstructed

Regularization error: Make approximate posterior distribution close to pair



Experiment

Generating Data



(a) Learned Frey Face manifold

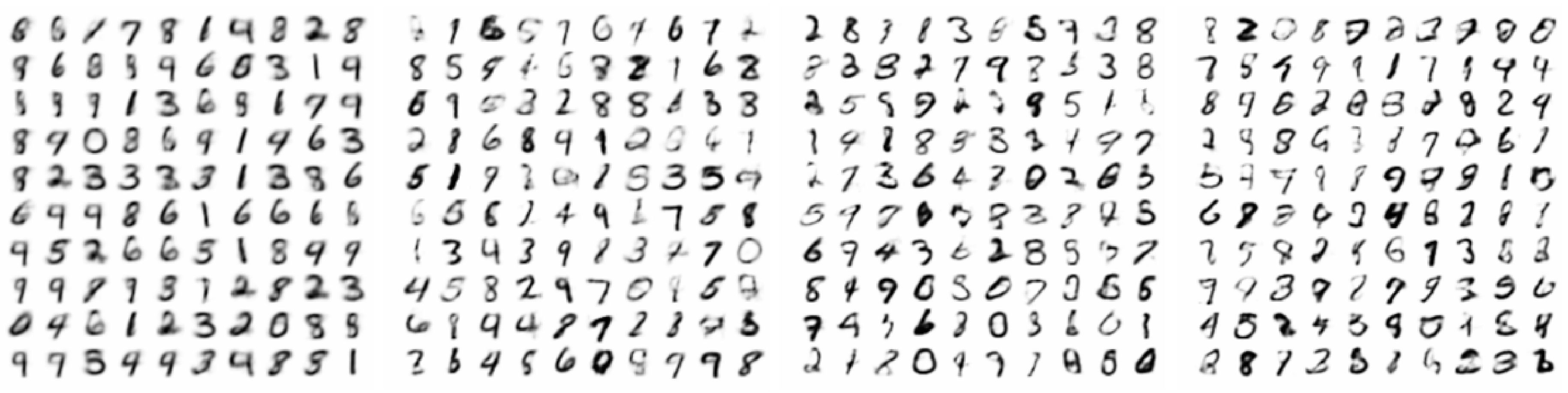
(b) Learned MNIST manifold

Figure 4: Visualisations of learned data manifold for generative models with two-dimensional latent space, learned with AEVB. Since the prior of the latent space is Gaussian, linearly spaced coordinates on the unit square were transformed through the inverse CDF of the Gaussian to produce values of the latent variables $\mathbf{z}$. For each of these values $\mathbf{z}$, we plotted the corresponding generative $p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})$ with the learned parameters $\boldsymbol{\theta}$.

TRAIN AND TEST

Experiment

Generating Data



(a) 2-D latent space (b) 5-D latent space (c) 10-D latent space (d) 20-D latent space

Figure 5: Random samples from learned generative models of MNIST for different dimensionalities of latent space.

TNT
TRAIN AND TEST

Reference

- Steve-Lee’s Deep Insight. [AutoEncoder의 모든것]. https://deepinsight.tistory.com/126
- 공돌이의 수학 정리노트. KL Divergence . https://angeloyeo.github.io/2020/10/27/KL_divergence.htmlBackground
- 유니의 공부. VAE 설명. https://process-mining.tistory.com/161
- govlKH.VAE 논문 리뷰. https://velog.io/@lee9843/VAE-Auto-Encoding-Variational-Bayes-%EB%85%BC%EB%AC%B8-%EB%A6%AC%EB%B7%B0
- Stanford University School of Engineering. CS231n Lecture 13|Generative Models. https://www.youtube.com/watch?v=5WoItGTWV54
- Smart Design Lab @ KAIST. 딥러닝 ch3.3 VAE. YouTube. https://www.youtube.com/watch?v=GbCAwVVKaHY&t=1221s

TRAIN AND TEST