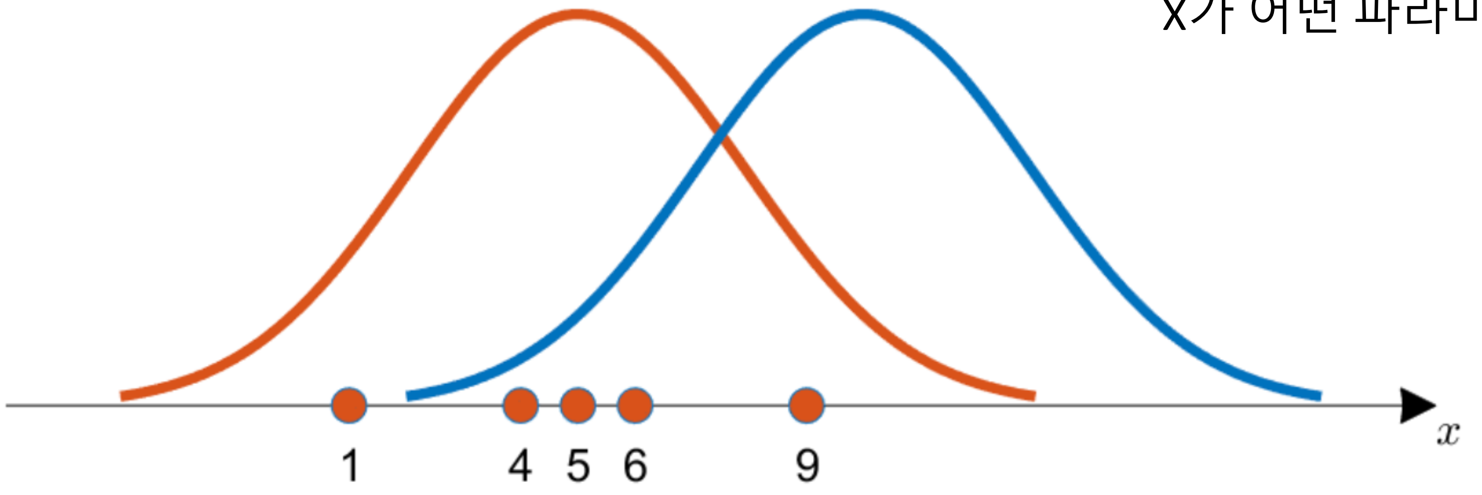
Auto-Encoding Variational Bayes

Maximum Likelihood Estimation

- 어떤 파라미터 θ로 이루어진 확률밀도함수 P($\boldsymbol{x}$| θ)에서 표본들의 집합을 $\boldsymbol{x}$라고 할 때 θ를 추정하는 방법

$\mathbf{X} \sim \mathcal{N}(\theta) = \mathcal{N}(m, \boldsymbol{\sigma}$^2)

x가 어떤 파라미터에서 나왔을까?



출처: https://angeloyeo.github.io/2020/07/17/MLE.html

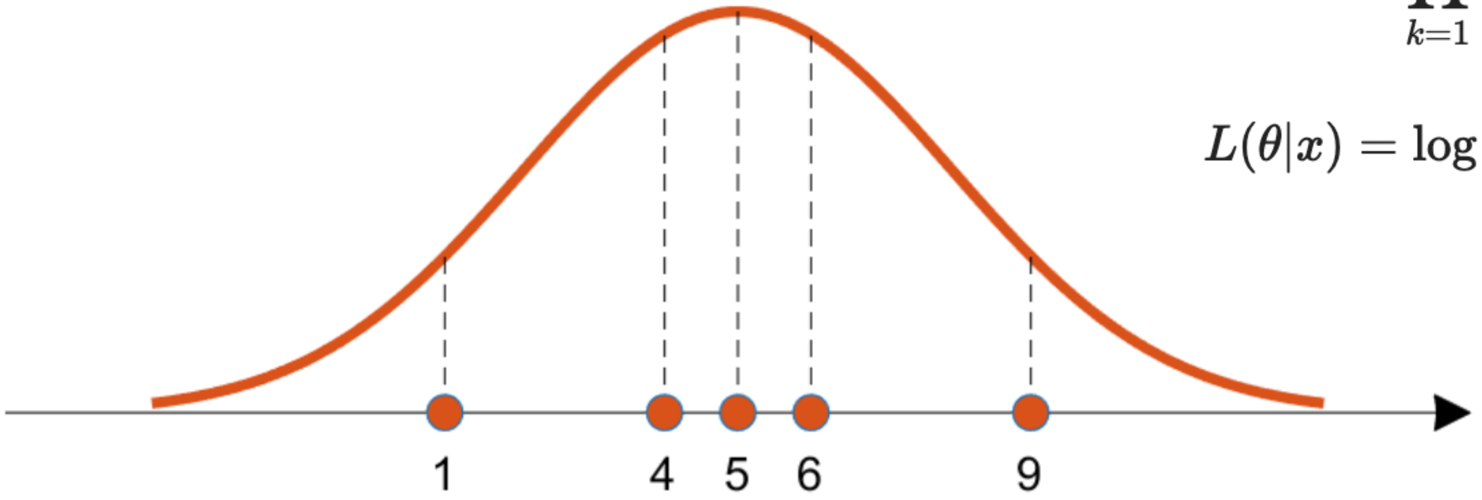
Maximum Likelihood Estimation

- Likelihood란 지금 얻은 데이터가 이 분포로부터 나왔을 가능도
- Likelihood는 각 샘플에서 후보 분포에 대한 높이를 다 곱하면 된다.

(모든 데이터가 독립적으로 추출됐다는 가정)

$$P(x|\theta) = \prod_{k=1}^{n} P(x_k|\theta)$$

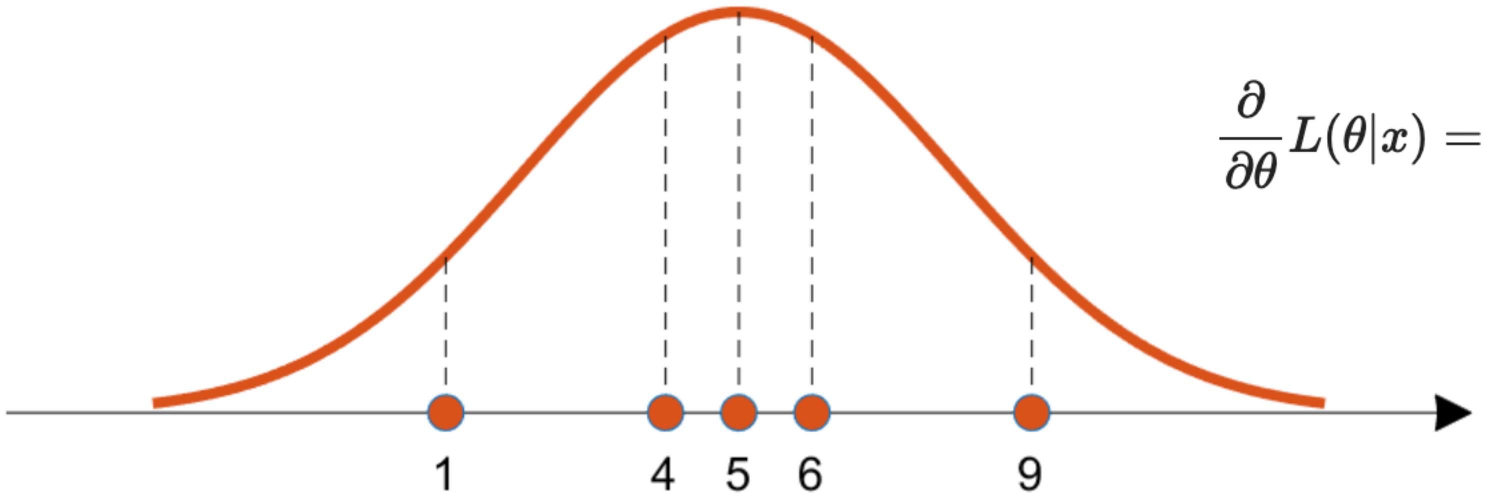
$$L(\theta|x) = \log P(x|\theta) = \sum_{i=1}^{n} \log P(x_i|\theta)$$



출처: https://angeloyeo.github.io/2020/07/17/MLE.html

Maximum Likelihood Estimation

- Θ에 대해서 미분할 수 있다면 Likelihood를 maximize하는 파라미터를 구할 수 있다.



$$\frac{\partial}{\partial \theta} L(\theta|x) = \frac{\partial}{\partial \theta} \log P(x|\theta) = \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log P(x_i|\theta) = 0$$

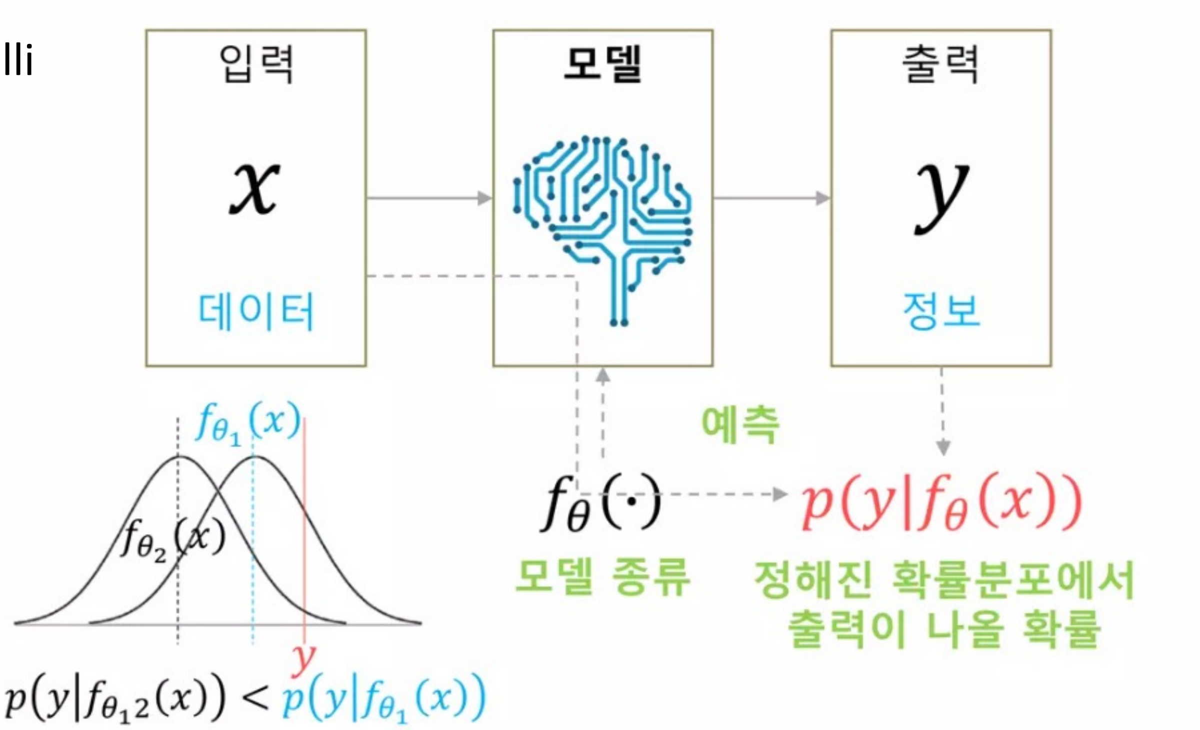
출처: https://angeloyeo.github.io/2020/07/17/MLE.html

DeepLearning의 Maximum Likelihood 관점 해석

P(Y|f(x))에 대한 분포를 가정 ex)Gaussian, Bernoulli

딥러닝 모델을 통한 output은 분포에 대한
Likelihood를 maximize하는 파라미터 θ를
추정하는 것 (가우시안 일때는 평균과 분산)

그렇다면
Loss = $-\log(p(y|f_\theta(x)))$



출처: 이활석 - 오토인코더의 모든것

DeepLearning의 Maximum Likelihood 관점 해석

Optimal θ는 negative log likelihood를
Minimize하게 만드는 θ

Y에 대한 분포가 있기 때문에
새로운 y를 generate 할 수 있다.

$$\theta^* = \underset{\theta}{\operatorname{argmin}}[-\log(p(y|f_\theta(x)))]$$

$$y_{new} \sim p(y|f_{\theta^*}(x_{new}))$$

다시 정리하면,

$$L(\theta|x) = \log P(x|\theta) = \sum_{i=1}^{n} \log P(x_i|\theta)$$

$$-\log\big(p(y|f_\theta(x))\big) = -\sum_i \log\big(p(y_i|f_\theta(x_i))\big)$$

출처: 이활석 - 오토인코더의 모든것

DeepLearning의 Maximum Likelihood 관점 해석

Maximum Likelihood관점의 해석에서

Y에 대한 분포를 가우시안으로 가정할 경우 MSE,

Bernoulli로 가정할 경우 cross-entropy가 Loss가 된다.

Univariate cases

$$-\log(p(y_i|f_\theta(x_i)))$$

Gaussian distribution

$$f_\theta(x_i) = \mu_i, \sigma_i = 1$$

$$p(y_i|\mu_i,\sigma_i) = \frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(-\frac{(y_i-\mu_i)^2}{2\sigma_i^2}\right)$$

$$\log(p(y_i|\mu_i,\sigma_i)) = \log\frac{1}{\sqrt{2\pi}\sigma_i} - \frac{(y_i-\mu_i)^2}{2\sigma_i^2}$$

$$-\log(p(y_i|\mu_i)) = -\log\frac{1}{\sqrt{2\pi}} + \frac{(y_i-\mu_i)^2}{2}$$

$$-\log(p(y_i|\mu_i)) \propto \frac{(y_i-\mu_i)^2}{2} = \frac{(y_i-f_\theta(x_i))^2}{2}$$

Mean Squared Error

Bernoulli distribution

$$f_\theta(x_i) = p_i$$

$$p(y_i|p_i) = p_i^{y_i}(1-p_i)^{1-y_i}$$

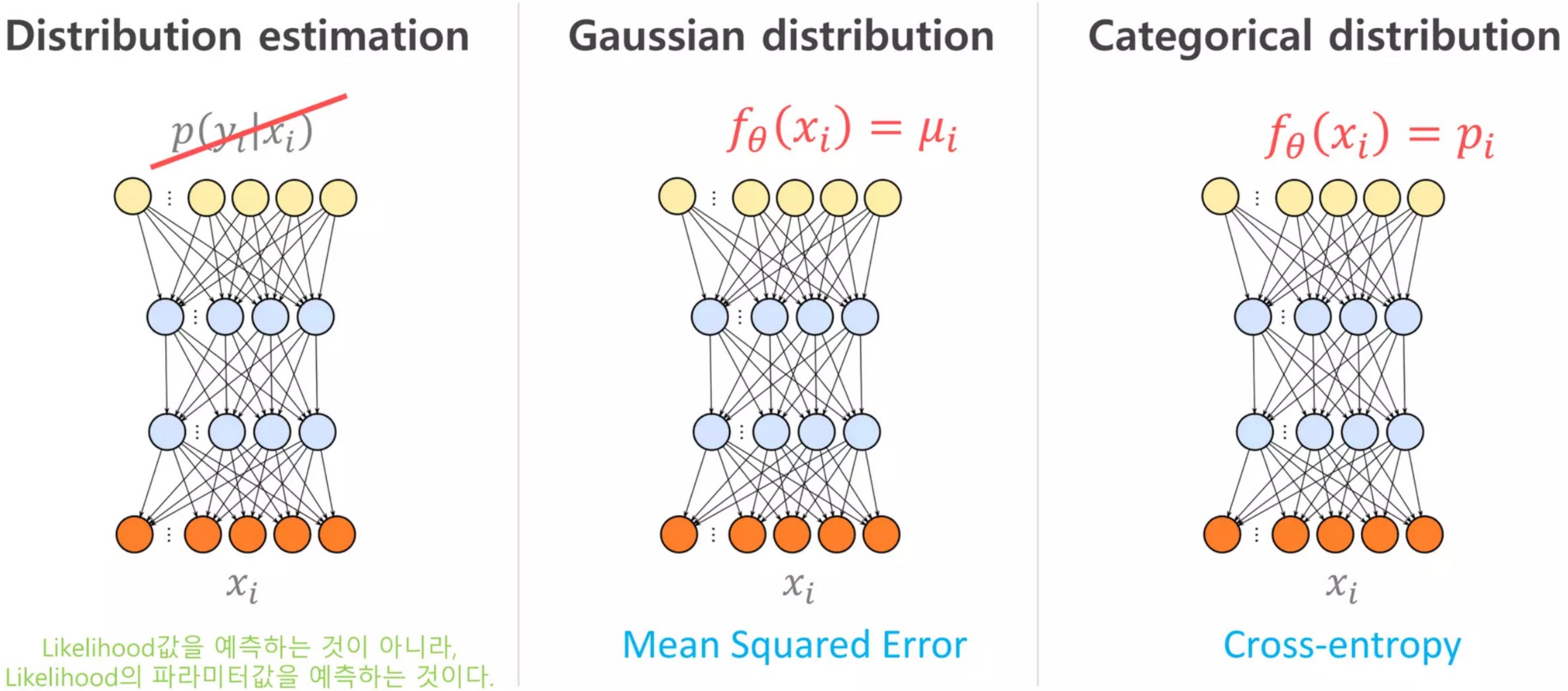
$$\log(p(y_i|p_i)) = y_i\log p_i + (1-y_i)\log(1-p_i)$$

$$-\log(p(y_i|p_i)) = -[y_i\log p_i + (1-y_i)\log(1-p_i)]$$

Cross-entropy

출처: 이활석 - 오토인코더의 모든것

DeepLearning의 Maximum Likelihood 관점 해석



출처: 이활석 - 오토인코더의 모든것

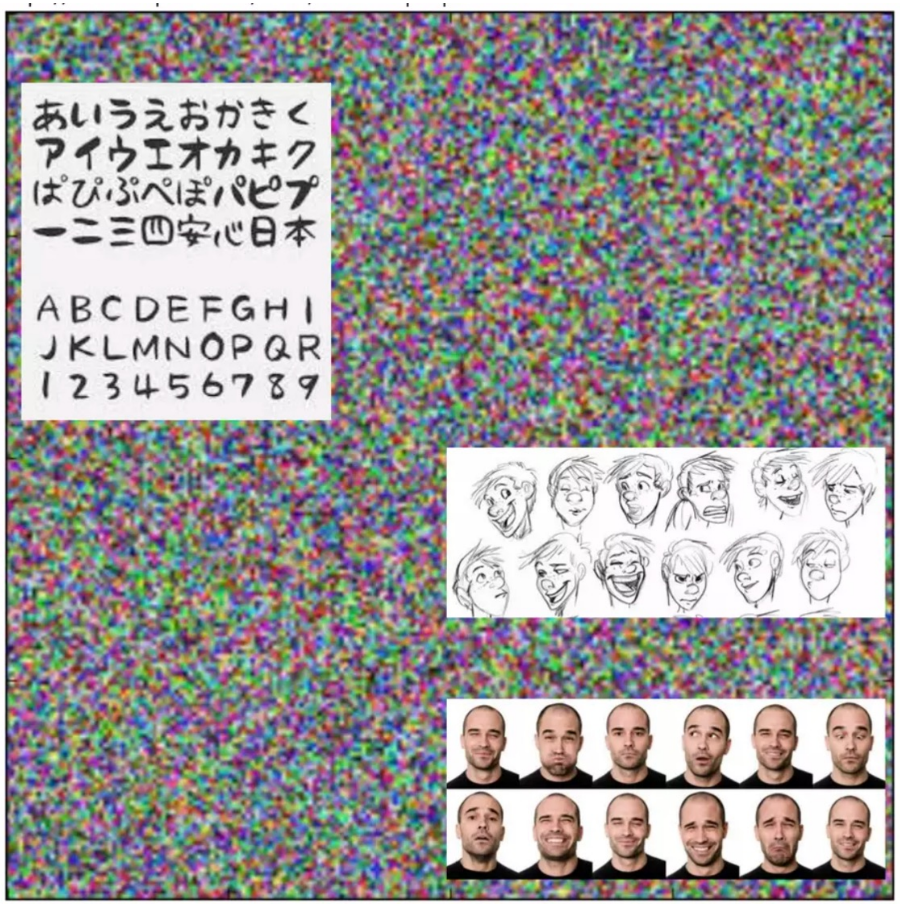
Manifold 가정

200x200 RGB 채널에서 random하게 값을 뽑아서 plot하면 오른쪽의 noise이미지만 계속 나온다.

얼굴이나 스케치, 글씨 같은 이미지도 나올 수 있으나 안나온다는 것은?

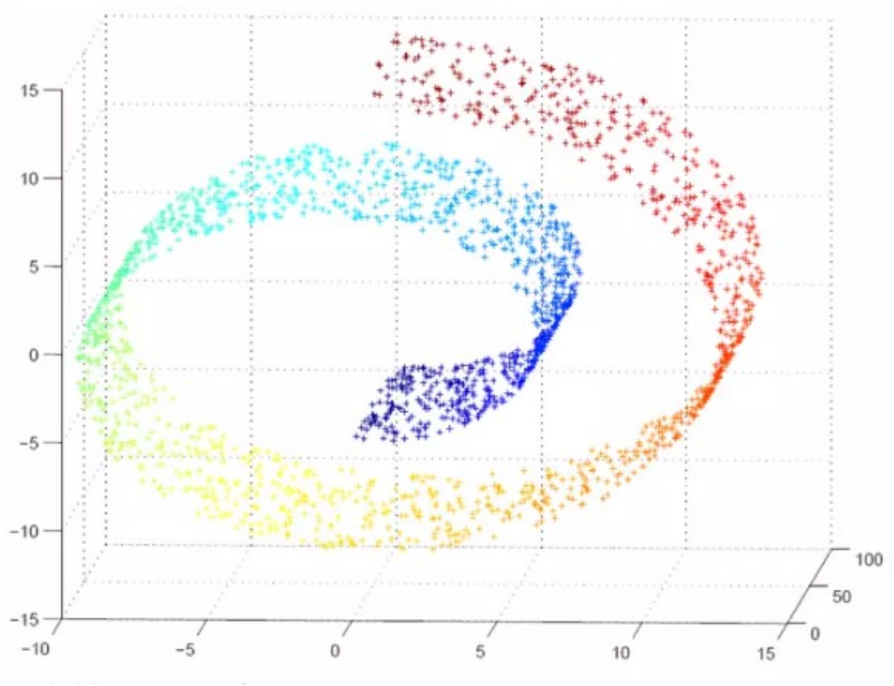
어떠한 특정한 공간에 데이터가 밀집해있다.

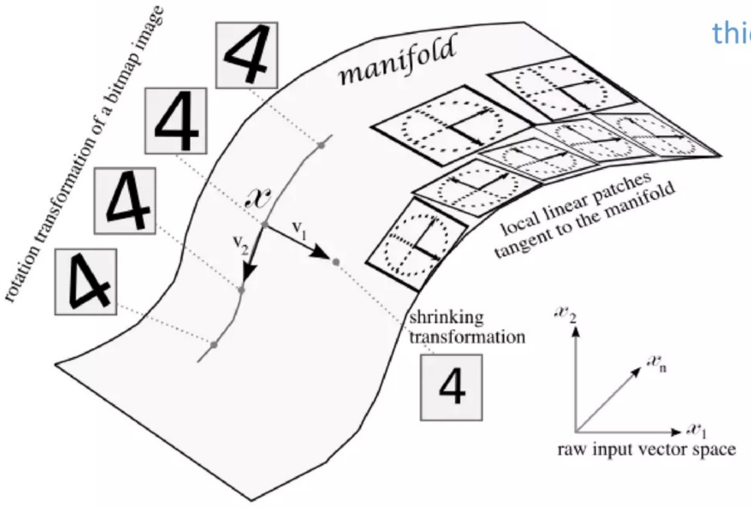
→ 그 공간을 잘 포함하는 저차원의 공간이 있다.

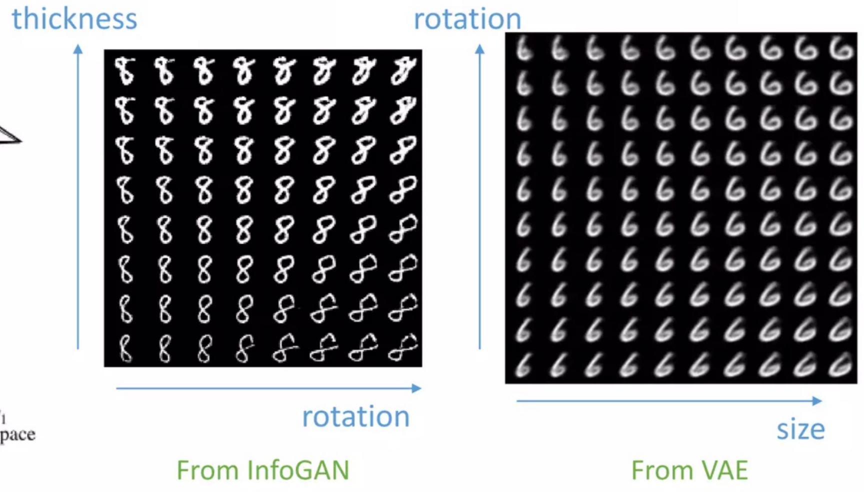


출처: 이활석 - 오토인코더의 모든것

Manifold 가정



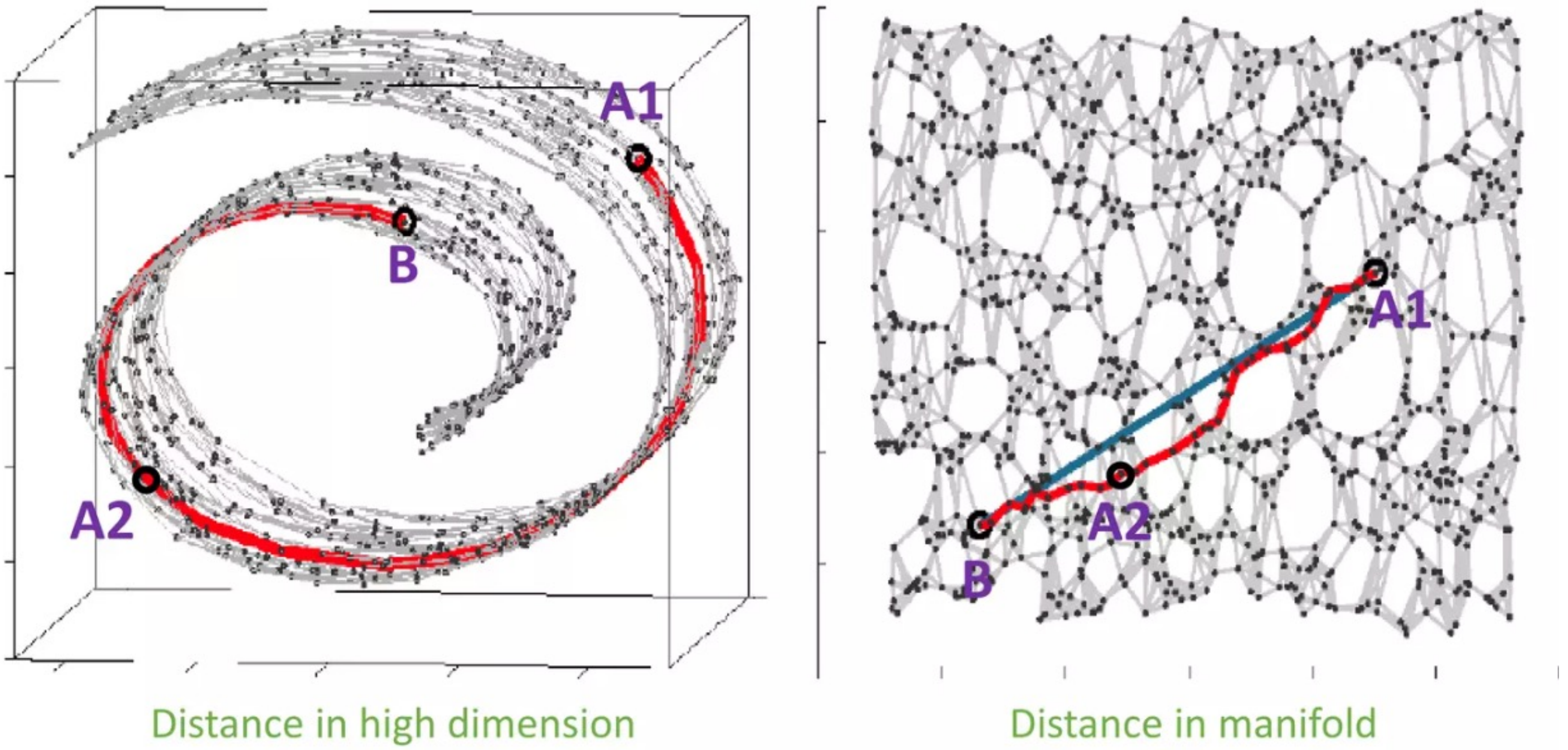




Manifold를 잘 찾았다면 manifold 좌표에 따라 이미지가 유의미하게 변한다.

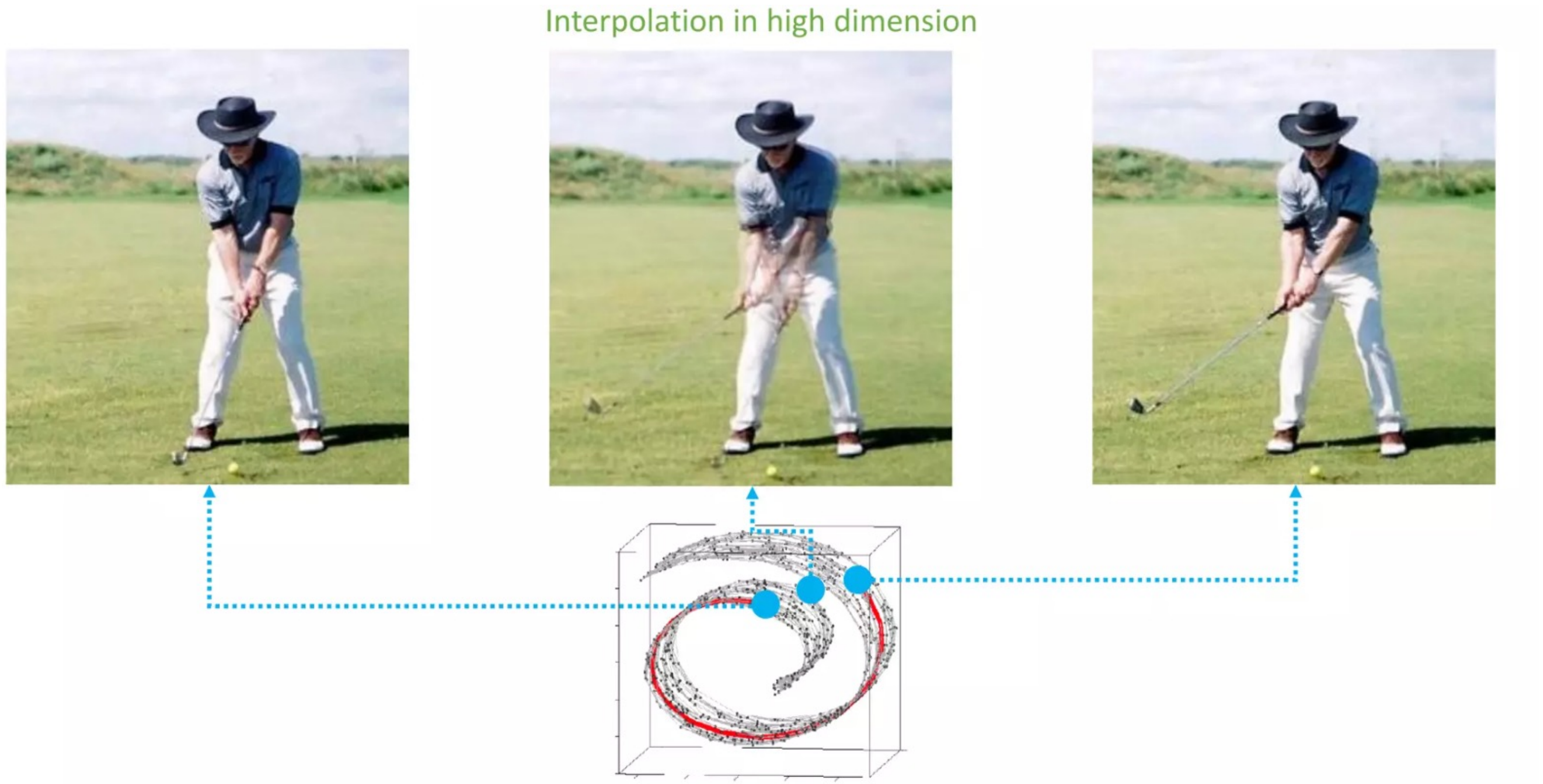
출처: 이활석 - 오토인코더의 모든것

Manifold 가정



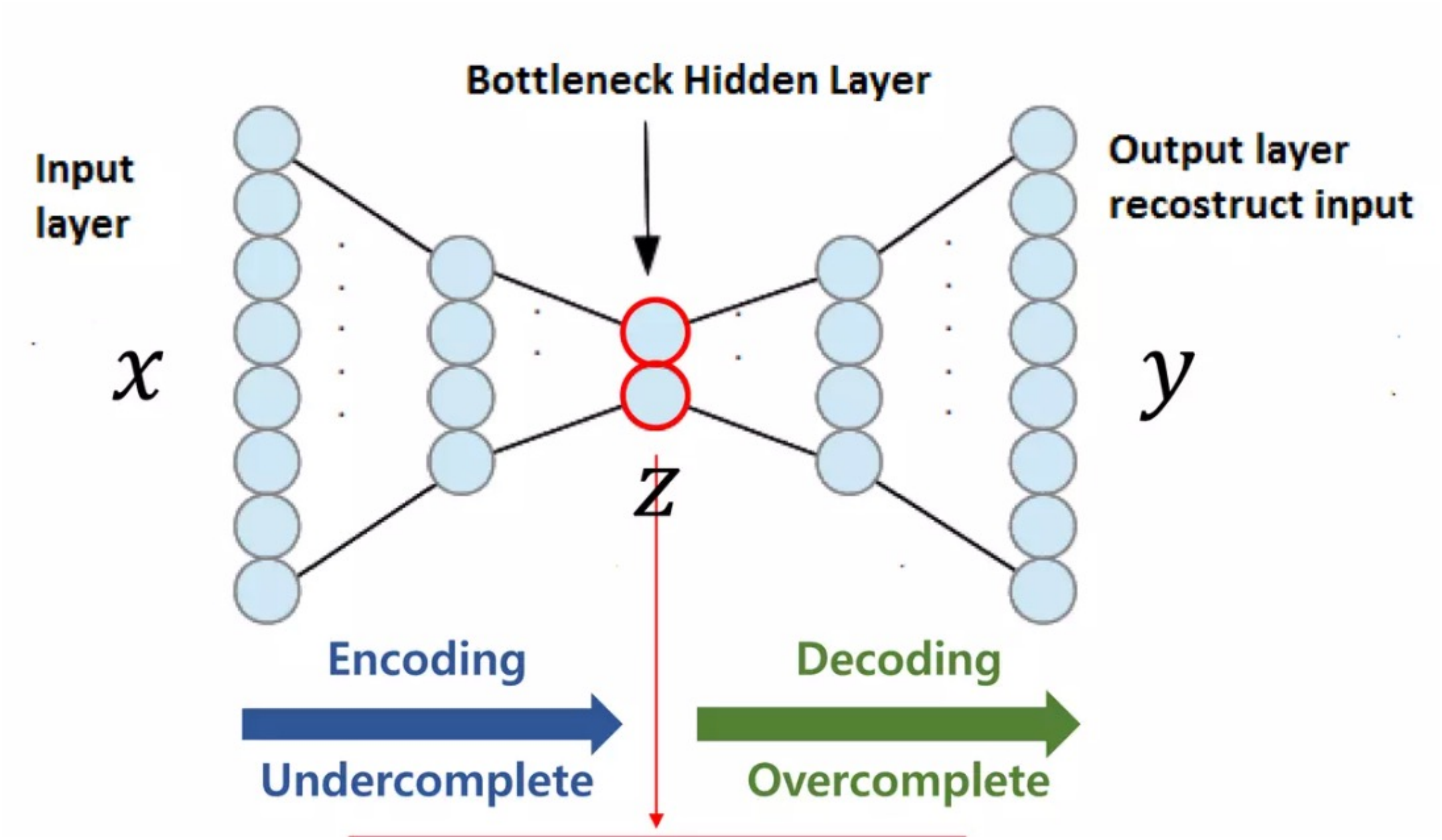
출처: 이활석 - 오토인코더의 모든것

Manifold 가정



출처: 이활석 - 오토인코더의 모든것

AutoEncoder



01. Data compression

02. Data visualization

03. Curse of dimensionality

Manifold Hypothesis

04. Discovering most important features

Reasonable distance metric
Needs disentagling the underlying explanatory factors
(making sense of the data)

출처: 이활석 - 오토인코더의 모든것

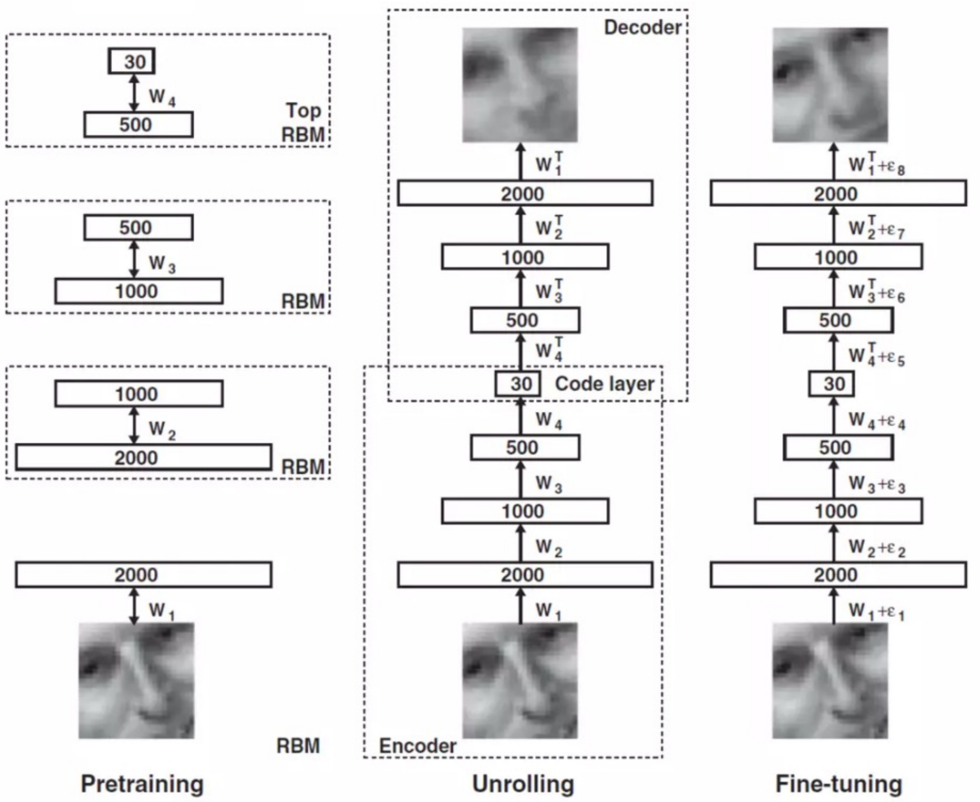
AutoEncoder의 여러 종류

Activation Function을 안쓰고 Linear 하게 만들면 PCA와 같다.

$h = f_\theta(x) = W(x - \mu)$ $with\ \theta = \{W, \mu\}$ in PCA Slide

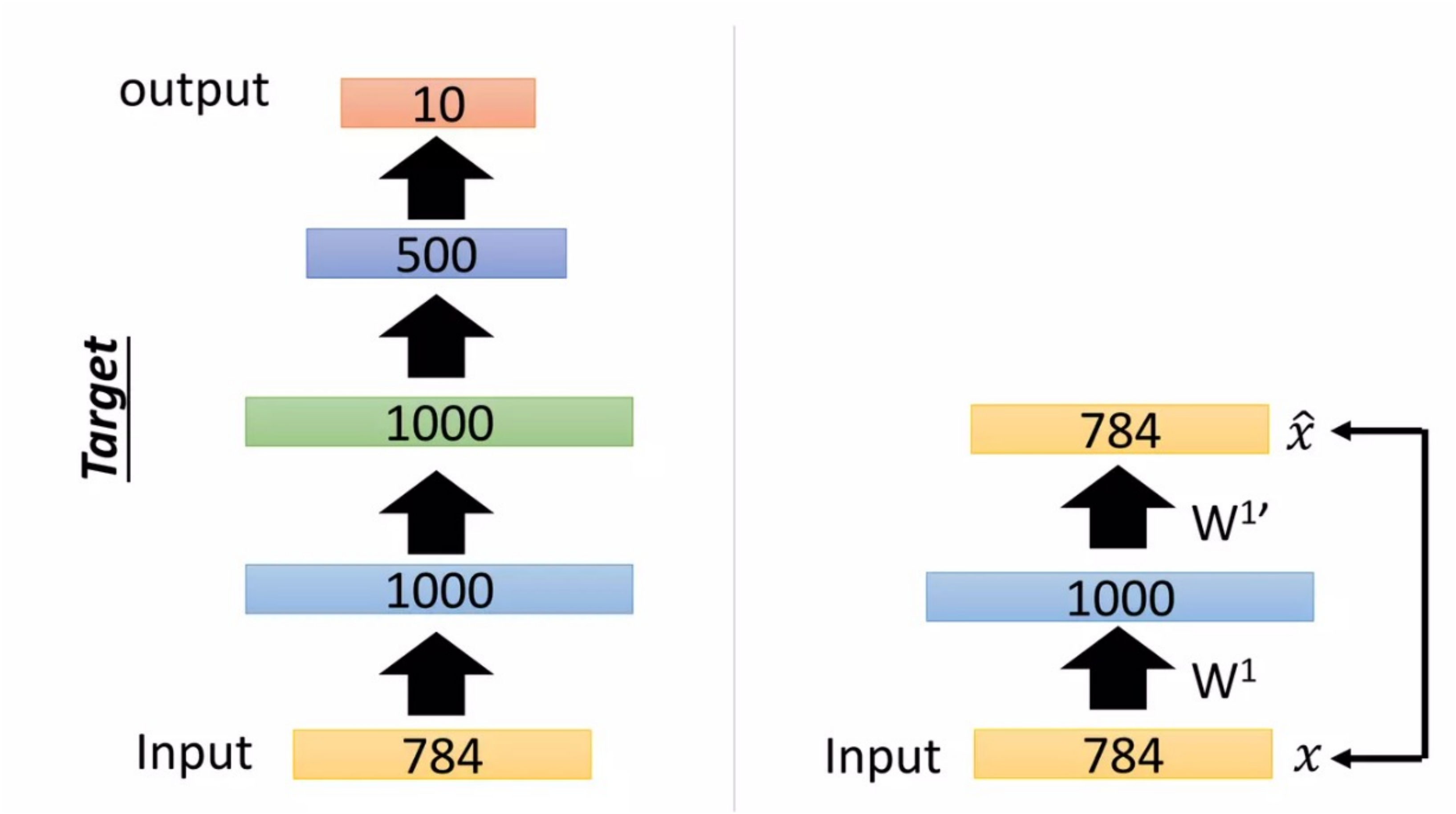
RBM과는 Stochastic하냐 deterministic하냐의 차이

Autoencoder	*RBM*
$z_i = \sigma(W_{ei}x + b_{ei})$	$P(h_i = 1 \vert v) = \sigma(W_{ei}v + b_{ei})$
$y_j = \sigma(W_{ej}^T z + b_{dj})$	$P(v_j = 1 \vert h) = \sigma(W_{ej}^T h + b_{dj})$
Determinisitc mapping z is a function x	Stochastic mapping z is a random variable



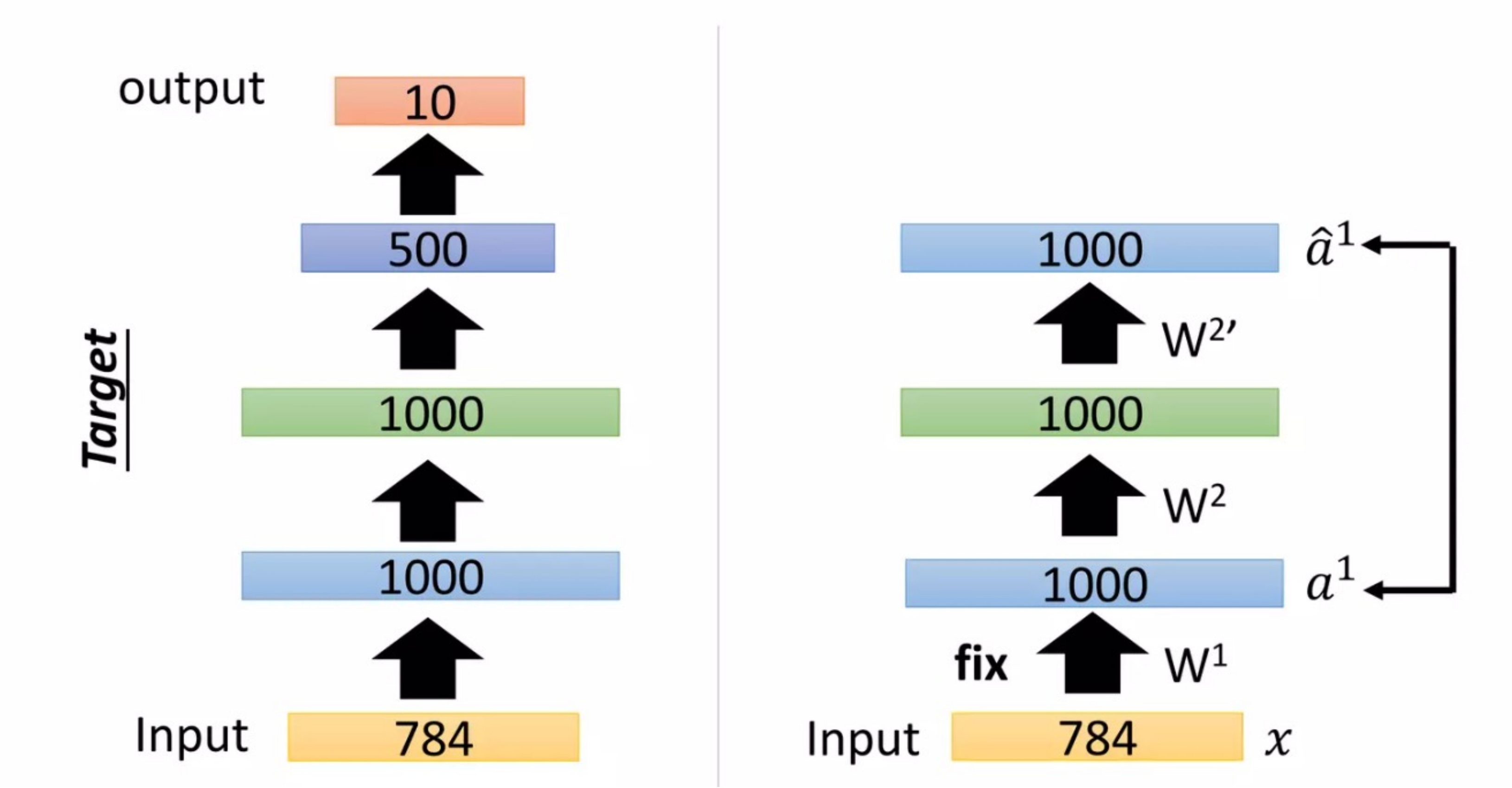
출처: 이활석 - 오토인코더의 모든것

AutoEncoder의 여러 종류



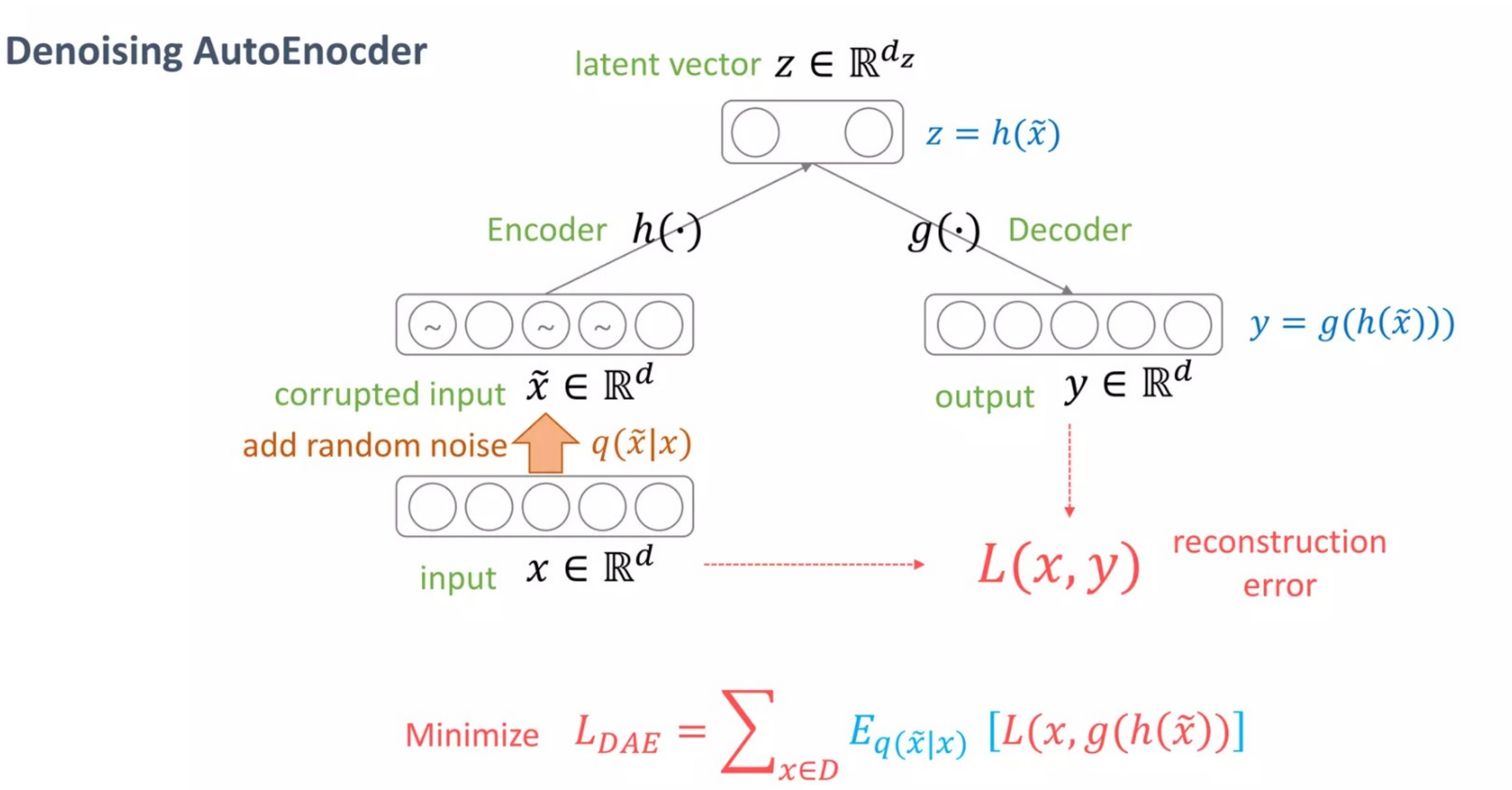
출처: 이활석 - 오토인코더의 모든것

AutoEncoder의 여러 종류



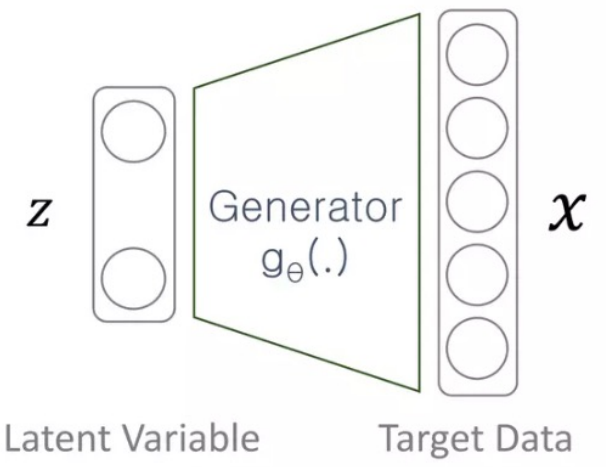
출처: 이활석 - 오토인코더의 모든것

AutoEncoder의 여러 종류



출처: 이활석 - 오토인코더의 모든것

Variational AutoEncoder



$z \sim p(z)$	Random variable
$g_\theta(\cdot)$	Deterministic function parameterized by θ
$x = g_\theta(z)$	Random variable

먼저 오토인코더를 생각하지 말고 어떤 latent space에서 Mnist와 같은 이미지를 생성하려고 한다면?

P(z)는 샘플링 하기 편한 분포여야 한다. (x를 만들어 내기 위해 Control하는 파라미터기 때문에)

$$p(x|g_\theta(z)\ ) = p_\theta(x|z)$$

결국 p(x)의 분포를 얻고 싶다!

$$\int p(x|g_\theta(z))p(z)dz = p(x)$$

출처: 이활석 - 오토인코더의 모든것

Variational AutoEncoder

적분은 어려우니 몬테카를로 방법으로 p(x)를 근사하면 되는거 아닌가요?

P(z)를 가우시안이라고 가정, p(x|g(z))도 가우시안이라고 가정

$$p(x) \approx \sum_i p(x|g_\theta(z_i))p(z_i)$$

(a) (b) (c)

Univariate cases

$$-\log(p(y_i|f_\theta(x_i)))$$

Gaussian distribution

$$f_\theta(x_i) = \mu_i, \sigma_i = 1$$

$$p(y_i|\mu_i,\sigma_i) = \frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(-\frac{(y_i-\mu_i)^2}{2\sigma_i^2}\right)$$

$$\log(p(y_i|\mu_i,\sigma_i)) = \log\frac{1}{\sqrt{2\pi}\sigma_i} - \frac{(y_i-\mu_i)^2}{2\sigma_i^2}$$

$$-\log(p(y_i|\mu_i)) = -\log\frac{1}{\sqrt{2\pi}} + \frac{(y_i-\mu_i)^2}{2}$$

$$-\log(p(y_i|\mu_i)) \propto \frac{(y_i-\mu_i)^2}{2} = \frac{(y_i-f_\theta(x_i))^2}{2}$$

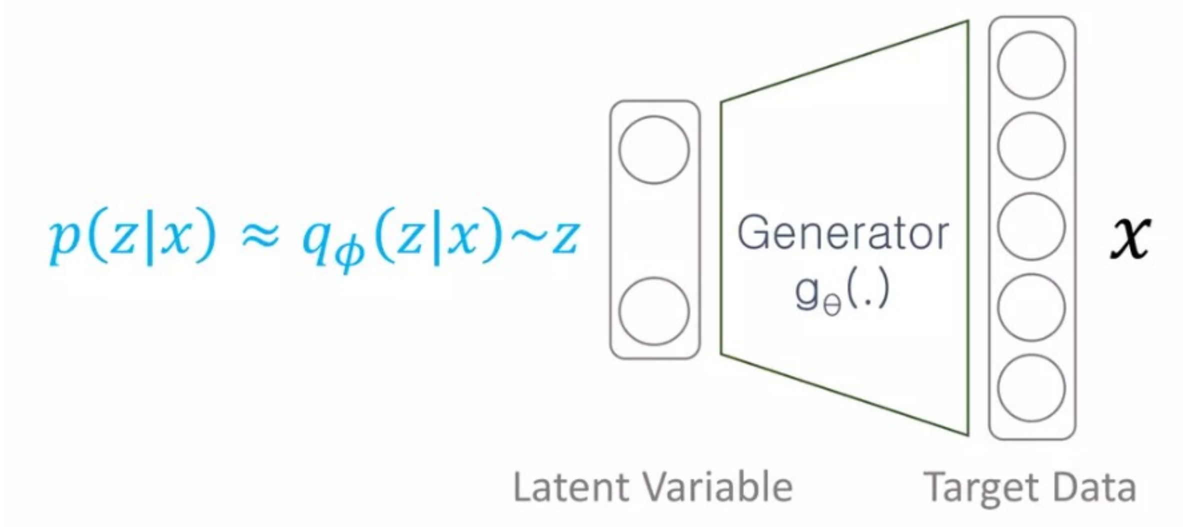
Mean Squared Error

출처: 이활석 - 오토인코더의 모든것

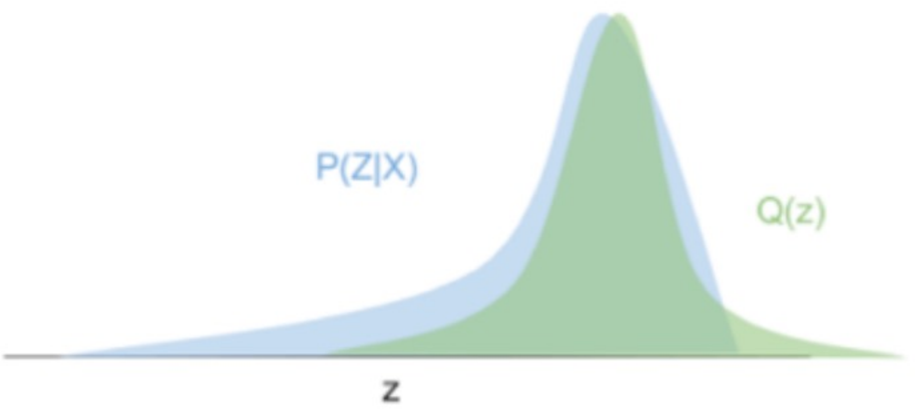
Variational AutoEncoder

Variational inference

Z를 정규분포에서 샘플링 하는게 아니라 x와 유사한 샘플이 나올 수 있는 분포 P(z|x)에서 샘플링 하면 된다.



But 우리는 P(z|x)를 알지 못하므로 확률분포를 골라서 P(z|x)와 유사하게 만든 $q_\phi(z|x)$ 를 이용한다.



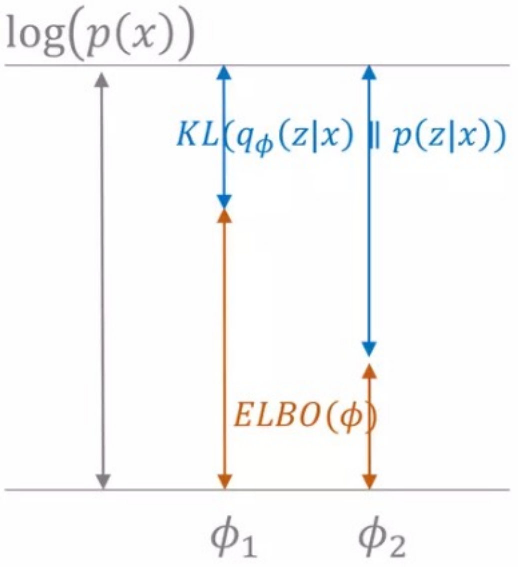
출처: 이활석 - 오토인코더의 모든것

Variational AutoEncoder

어떻게?

$$
\begin{aligned}
\log(p(x)) &= \int \log(p(x))q_\phi(z|x)dz \qquad \leftarrow \int q_\phi(z|x)dz = 1 \\
&= \int \log\left(\frac{p(x,z)}{p(z|x)}\right) q_\phi(z|x)dz \qquad \leftarrow p(x) = \frac{p(x,z)}{p(z|x)} \\
&= \int \log\left(\frac{p(x,z)}{q_\phi(z|x)} \cdot \frac{q_\phi(z|x)}{p(z|x)}\right) q_\phi(z|x)dz \\
&= \underbrace{\int \log\left(\frac{p(x,z)}{q_\phi(z|x)}\right) q_\phi(z|x)dz}_{ELBO(\phi)} + \underbrace{\int \log\left(\frac{q_\phi(z|x)}{p(z|x)}\right) q_\phi(z|x)dz}_{KL\left(q_\phi(z|x) \parallel p(z|x)\right)}
\end{aligned}
$$

두 확률분포 간의 거리≥0



출처: 이활석 - 오토인코더의 모든것

Variational AutoEncoder

어떻게?

$$\log(p(x)) = ELBO(\phi) + KL(q_\phi(z|x)|p(z|x))$$

$$q_{\phi^*}(z|x) = \underset{\phi}{\operatorname{argmax}}\, ELBO(\phi)$$

$$ELBO(\phi) = \int \log\left(\frac{p(x,z)}{q_\phi(z|x)}\right) q_\phi(z|x)dz$$

$$= \int \log\left(\frac{p(x|z)p(z)}{q_\phi(z|x)}\right) q_\phi(z|x)dz$$

$$= \int \log(p(x|z))q_\phi(z|x)dz - \int \log\left(\frac{q_\phi(z|x)}{p(z)}\right) q_\phi(z|x)dz$$

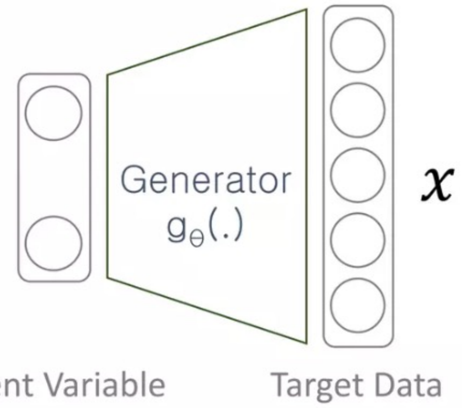
$$= \mathbb{E}_{q_\phi(z|x)}[\log(p(x|z))] - KL(q_\phi(z|x)||p(z))$$ 앞 슬라이드

출처: 이활석 - 오토인코더의 모든것

Variational AutoEncoder

어떻게?

$p(z|x) \approx q_\phi(z|x) \sim z$



Optimization Problem 1 on ϕ: Variational Inference

$$\log(p(x)) \geq \mathbb{E}_{q_\phi(z|x)}[\log(p(x|z))] - KL(q_\phi(z|x)||p(z)) = ELBO(\phi)$$

Optimization Problem 2 on θ: Maximum likelihood

$$-\sum_i \log(p(x_i)) \leq -\sum_i \{\mathbb{E}_{q_\phi(z|x_i)}[\log(p(x_i|g_\theta(z)))] - KL(q_\phi(z|x_i)||p(z))\}$$

$$p(x|g_\theta(z)\ ) = p_\theta(x|z)$$

Final Optimization Problem

$$\arg\min_{\phi,\theta} \sum_i -\mathbb{E}_{q_\phi(z|x_i)}[\log(p(x_i|g_\theta(z)))] + KL(q_\phi(z|x_i)||p(z))$$

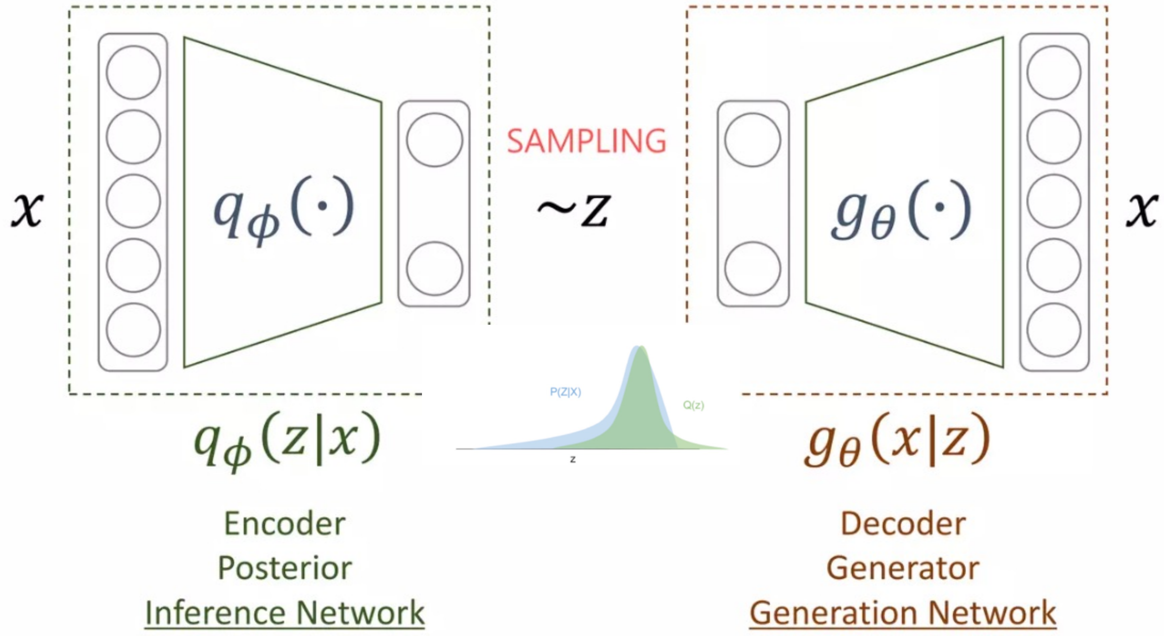
결국 p(x)의 분포를 얻고 싶다!

출처: 이활석 - 오토인코더의 모든것

Variational AutoEncoder

어떻게?

$$\arg\min_{\phi,\theta} \sum_i \underbrace{-\mathbb{E}_{q_\phi(z|x_i)}\left[\log\left(p(x_i|g_\theta(z))\right)\right] + KL\left(q_\phi(z|x_i)||p(z)\right)}_{L_i(\phi,\theta,x_i)}$$



$p(x|g_\theta(z)\ ) = p_\theta(x|z)$

결국 p(x)의 분포를 얻고 싶다!

The mathematical basis of VAEs actually has relatively little to do with classical autoencoders

출처: 이활석 - 오토인코더의 모든것

Variational AutoEncoder

$$\arg\min_{\phi,\theta}\sum_i \underbrace{-\mathbb{E}_{q_\phi(z|x_i)}\left[\log\left(p(x_i|g_\theta(z))\right)\right] + KL\left(q_\phi(z|x_i)||p(z)\right)}_{L_i(\phi,\theta,x_i)}$$

Variational inference를 위한 approximation class 중 선택

원 데이터에 대한 likelihood

다루기 쉬운 확률 분포 중 선택

$$L_i(\phi,\theta,x_i) = \underbrace{-\mathbb{E}_{q_\phi(z|x_i)}\left[\log\left(p(x_i|g_\theta(z))\right)\right]}_{\text{Reconstruction Error}} + \underbrace{KL\left(q_\phi(z|x_i)||p(z)\right)}_{\text{Regularization}}$$

Reconstruction Error

- 현재 샘플링 용 함수에 대한 negative log likelihood
- x_i에 대한 복원 오차 (AutoEncoder 관점)

Regularization

- 현재 샘플링 용 함수에 대한 추가 조건
- 샘플링의 용의성/생성 데이터에 대한 통제성을 위한 조건을 prior에 부여 하고 이와 유사해야 한다는 조건을 부여

출처: 이활석 - 오토인코더의 모든것

Variational AutoEncoder

$$L_i(\phi,\theta,x_i) = \underbrace{-\mathbb{E}_{q_\phi(z|x_i)}\left[\log\left(p(x_i|g_\theta(z))\right)\right]}_{\text{Reconstruction Error}} + \underbrace{KL\left(q_\phi(z|x_i)||p(z)\right)}_{\text{Regularization}}$$

$$KL\left(q_\phi(z|x_i)||p(z)\right) = \frac{1}{2}\left\{tr\left(\sigma_i^2 I\right) + \mu_i^T\mu_i - J + \ln\frac{1}{\prod_{j=1}^{J}\sigma_{i,j}^2}\right\}$$

$$= \frac{1}{2}\left\{\sum_{j=1}^{J}\sigma_{i,j}^2 + \sum_{j=1}^{J}\mu_{i,j}^2 - J - \sum_{j=1}^{J}\ln\left(\sigma_{i,j}^2\right)\right\}$$

$$= \frac{1}{2}\sum_{j=1}^{J}\left(\mu_{i,j}^2 + \sigma_{i,j}^2 - \ln\left(\sigma_{i,j}^2\right) - 1\right)$$

Easy to compute!!

출처: 이활석 - 오토인코더의 모든것

Variational AutoEncoder

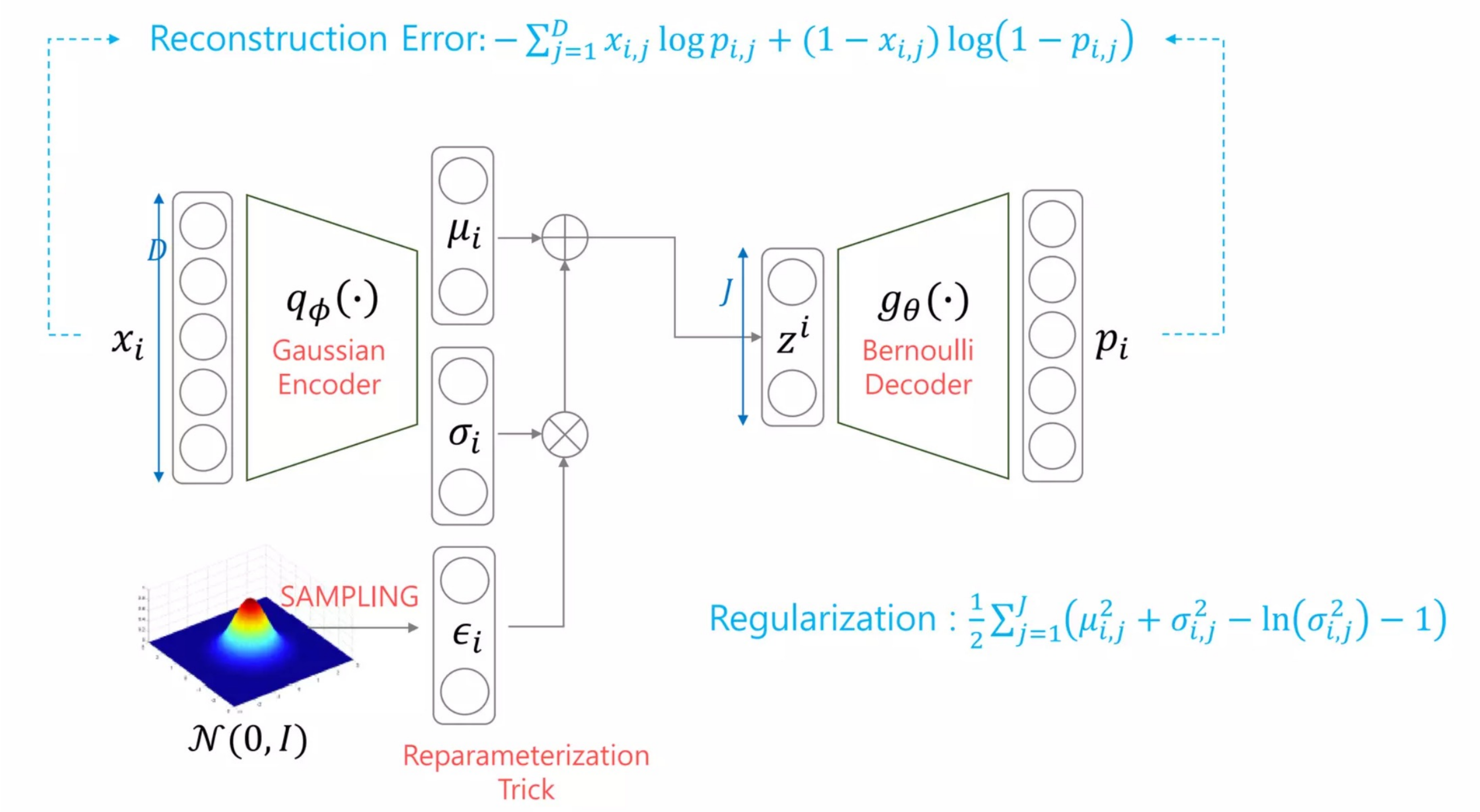
$$L_i(\phi, \theta, x_i) = \underbrace{-\mathbb{E}_{q_\phi(z|x_i)}\left[\log\left(p(x_i|g_\theta(z))\right)\right]}_{\text{Reconstruction Error}} + \underbrace{KL\left(q_\phi(z|x_i)||p(z)\right)}_{\text{Regularization}}$$

$$\mathbb{E}_{q_\phi(z|x_i)}\left[\log\left(p_\theta(x_i|z)\right)\right] = \int \log\left(p_\theta(x_i|z)\right)q_\phi(z|x_i)dz$$

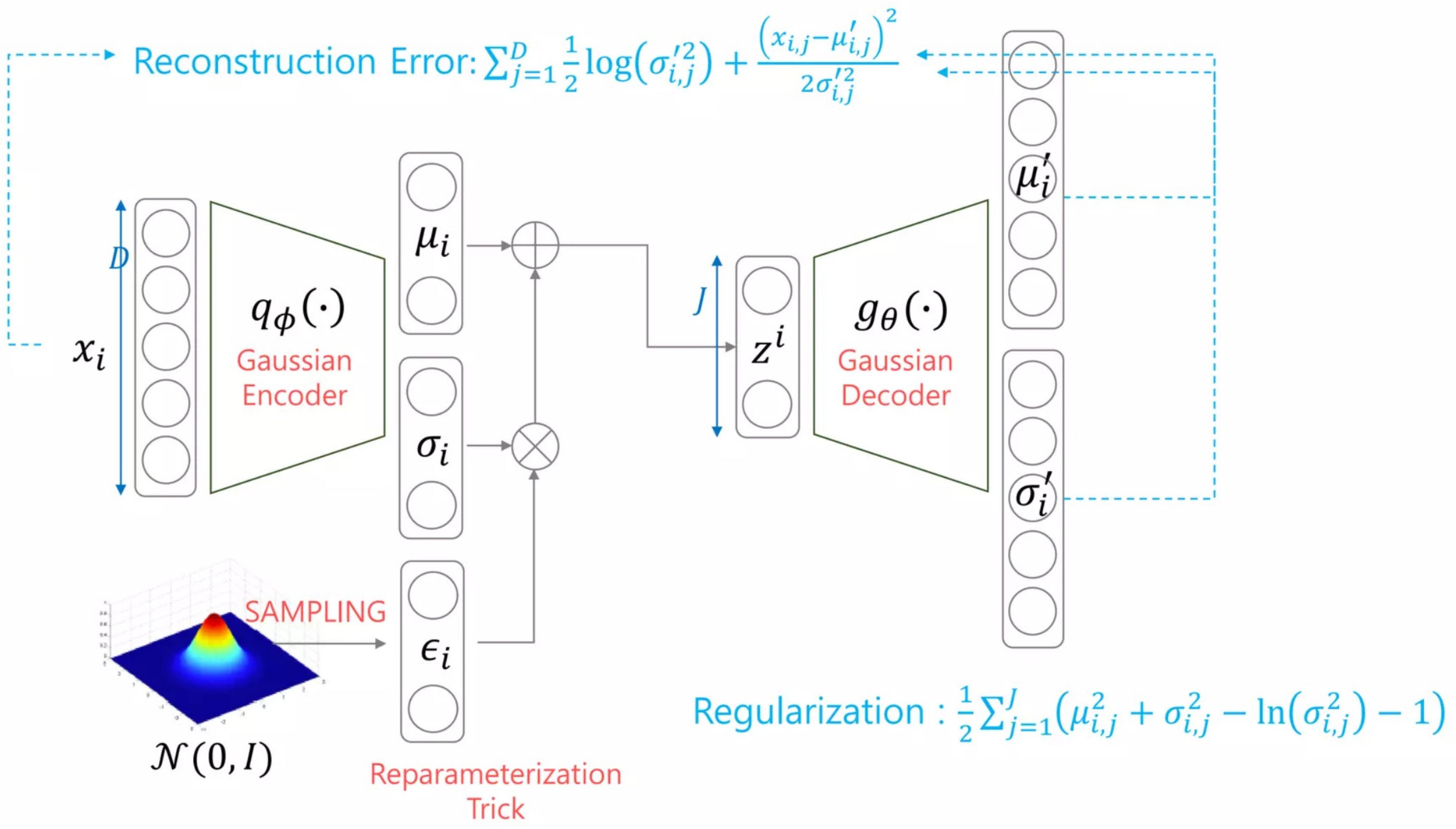
Monte-carlo technique → $\approx \frac{1}{L}\sum_{z^{i,l}} \log\left(p_\theta\left(x_i|z^{i,l}\right)\right)$

출처: 이활석 - 오토인코더의 모든것

Variational AutoEncoder



Variational AutoEncoder



Q&A

- 오토인코더의 모든것 - 이활석
- https://www.youtube.com/watch?v=o_peo6U7IRM