Variational Autoencoder

nonezero@kumoh.ac.kr

국립금오공과대학교

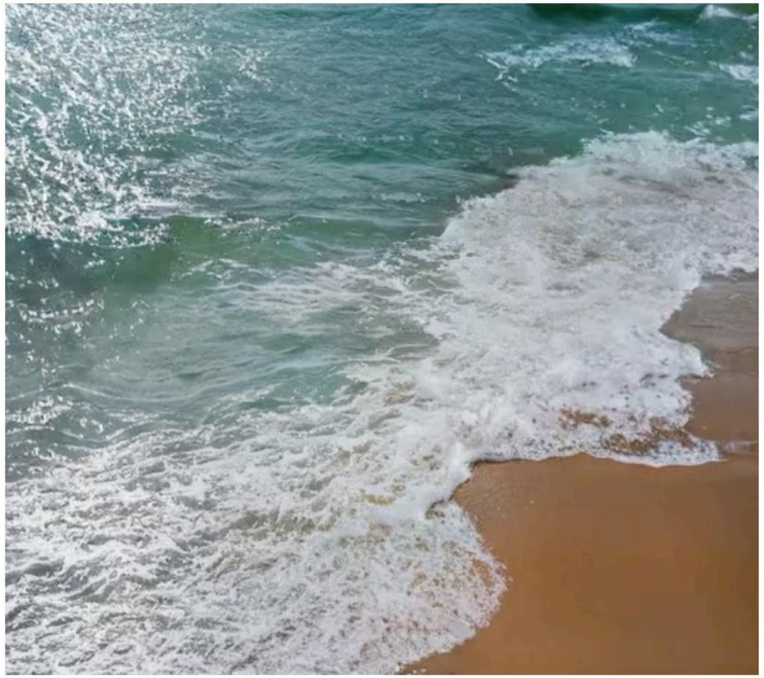
고 재 필

Which face is fake?

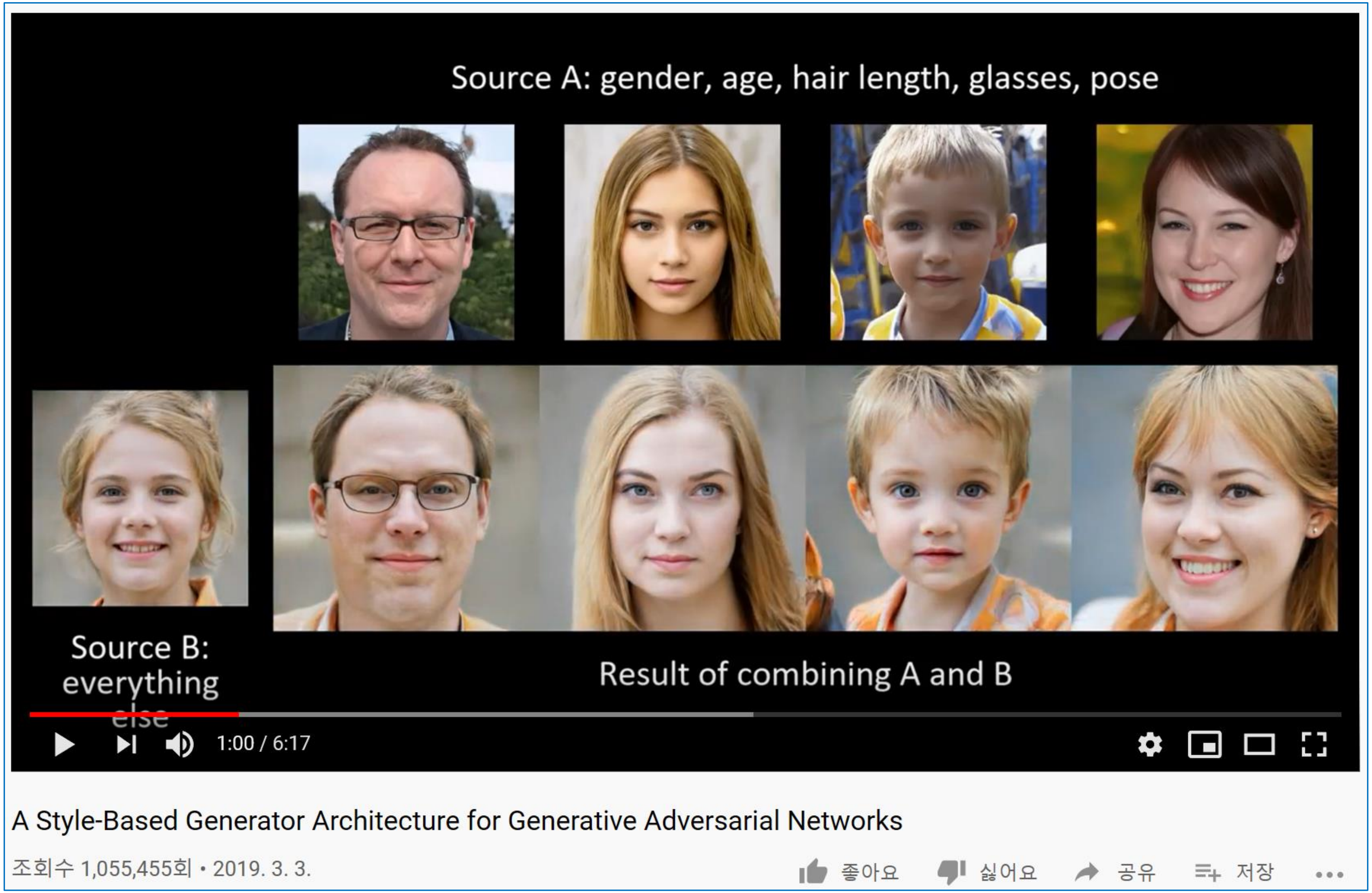
A B C

Application (1): Content Generation

StyleGAN3 example images



Karras et al. Alias-Free Generative Adversarial Networks, NeurIPS 2021



https://www.youtube.com/watch?v=kSLJriaOumA#action=share

Text-to-Image Generation

DALL·E 2

"a teddy bear on a skateboard in times square"

"Hierarchical Text-Conditional Image Generation with CLIP Latents" Ramesh et al., 2022

Imagen

A group of teddy bears in suit in a corporate office celebrating the birthday of their friend. There is a pizza cake on the desk.

"Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding", Saharia et al., 2022

(A.K.A. LDM & Stable Diffusion)

Robin Rombach[1,2], Andreas Blattmann[1,2], Dominik Lorenz[1,2], Patrick Esser[3],
Björn Ommer[1,2]

[1]LMU Munich, [2]IWR, Heidelberg University, [3]Runway

CVPR 2022 (ORAL)

LMU

IWR
Interdisciplinary Center
for Scientific Computing

Press Releases

LMU press release

Stability.AI press release

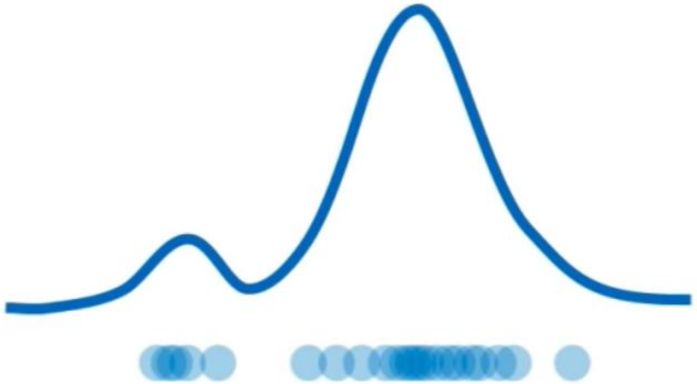
https://ommer-lab.com/research/latent-diffusion-models/

Generative modeling

Goal: Take as input training samples from some distribution and learn a model that represents that distribution

Density Estimation

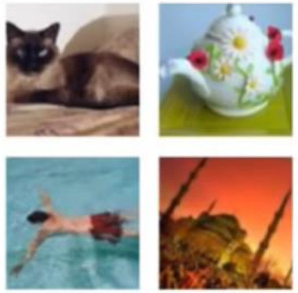
samples



Sample Generation

Input samples

Training data $\sim P_{data}(x)$



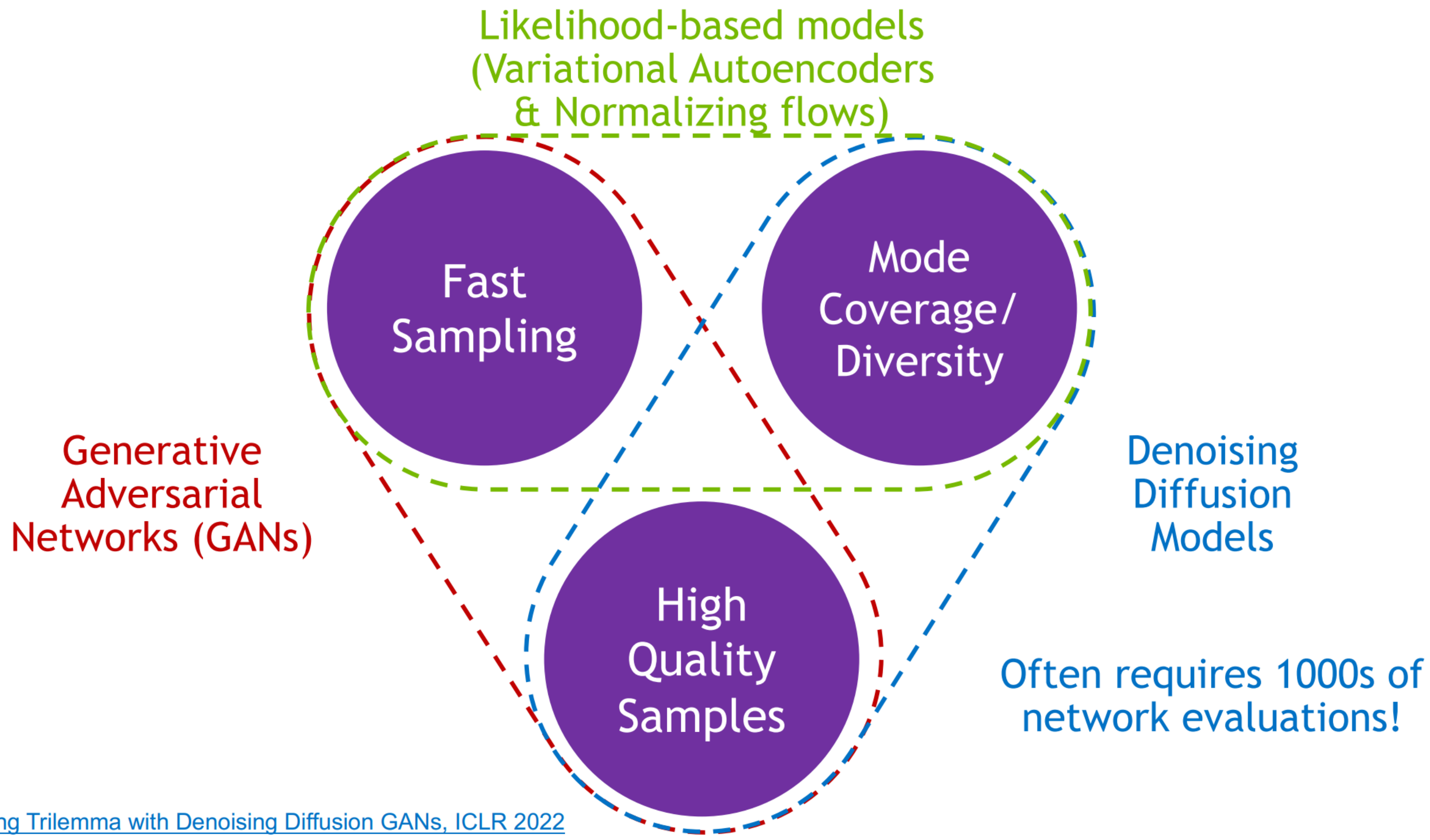
Generated samples

Generated $\sim P_{model}(x)$

How can we learn $P_{model}(x)$ similar to $P_{data}(x)$?

What makes a good generative model?

The generative learning trilemma



Tackling the Generative Learning Trilemma with Denoising Diffusion GANs, ICLR 2022

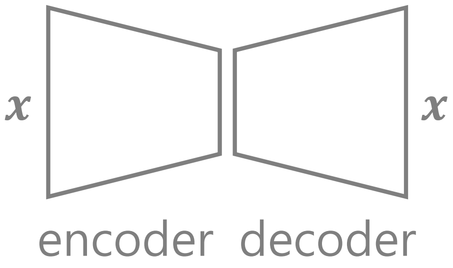
VAE

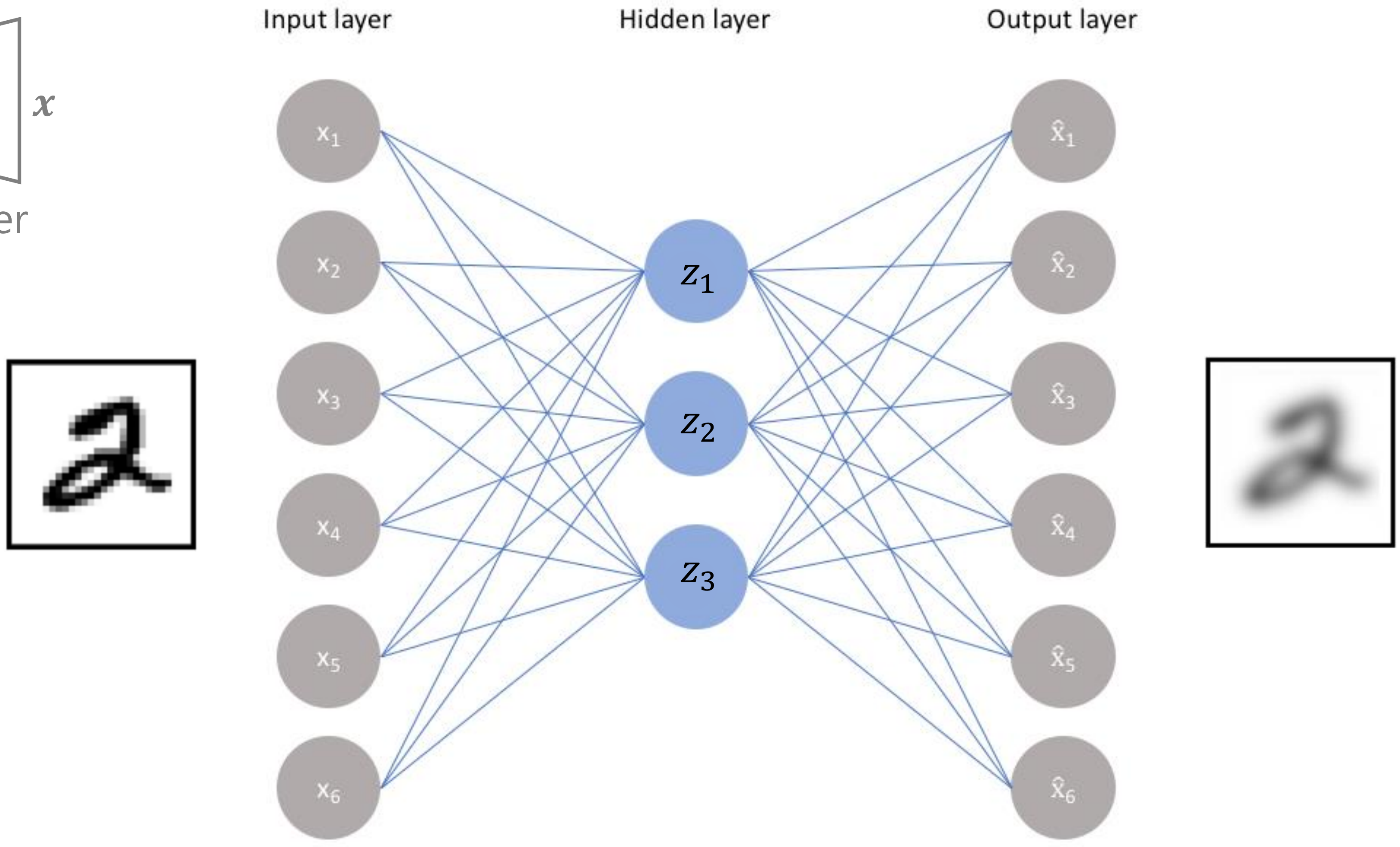
Variational Autoencoder

새로운 신경망?

신경망구조 + 손실함수

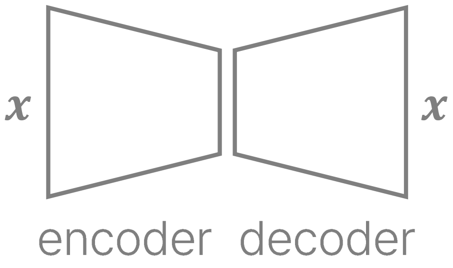
Autoencoder

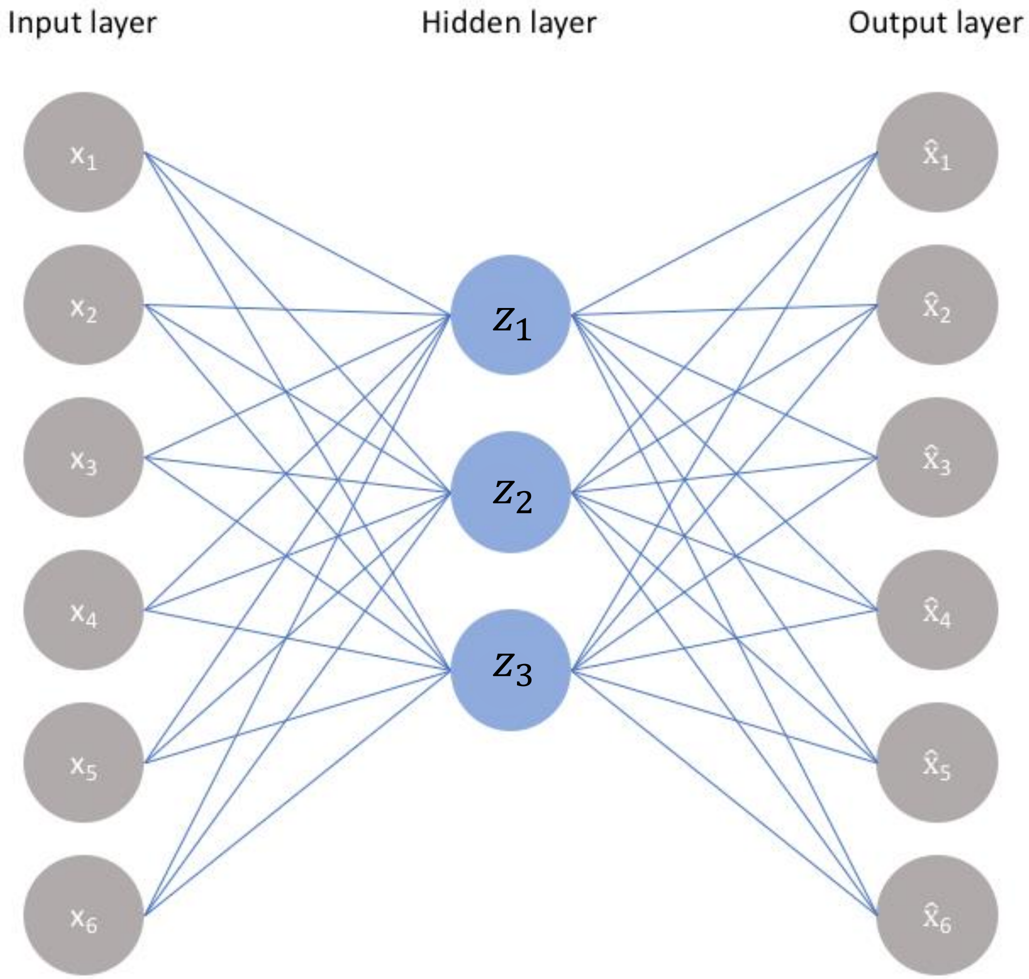




$f_1 : x \to z$ $\qquad$ $f_2 : z \to x$

Autoencoder





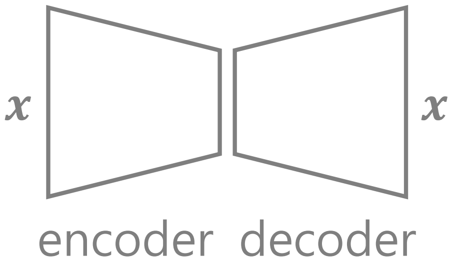
$f_1 : x \rightarrow z$ $f_2 : z \rightarrow x$

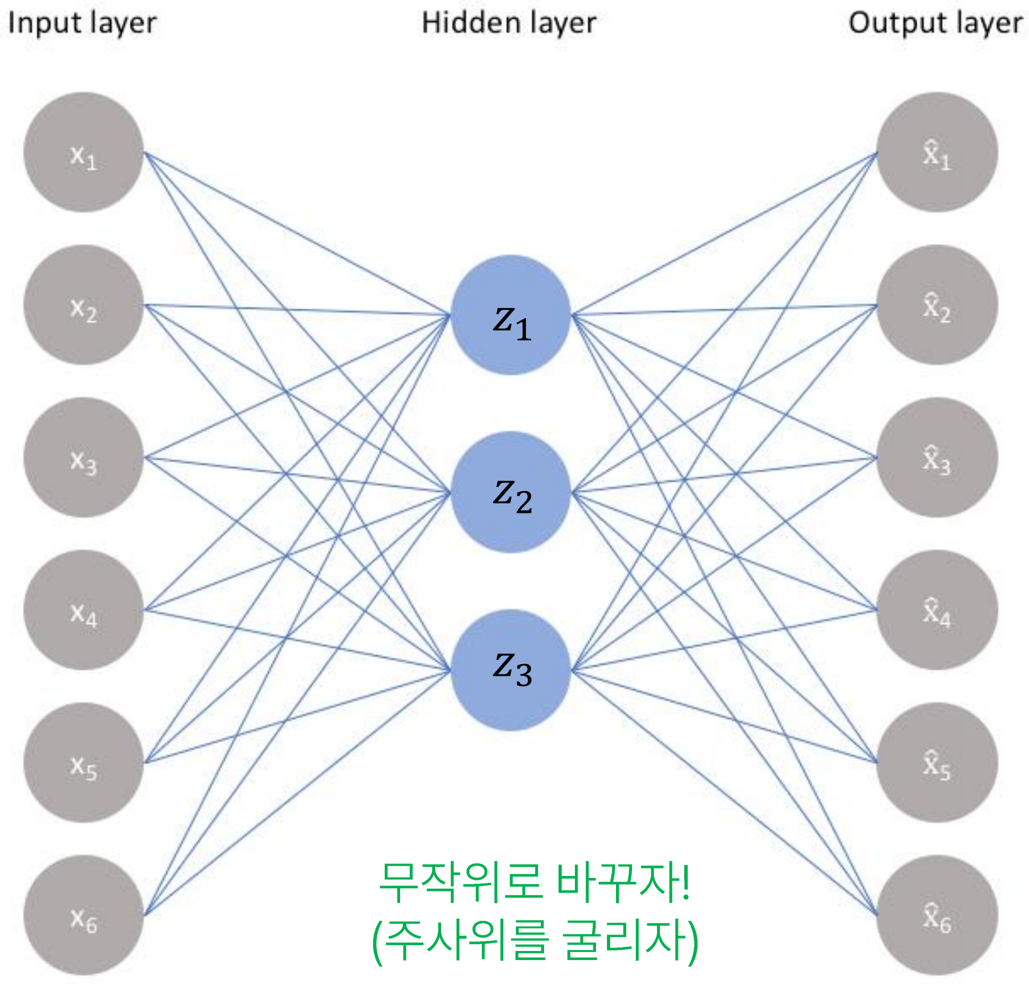
입력이 같으면, 출력도 같다.

deterministic function

non-deterministic하게 하고 싶다.

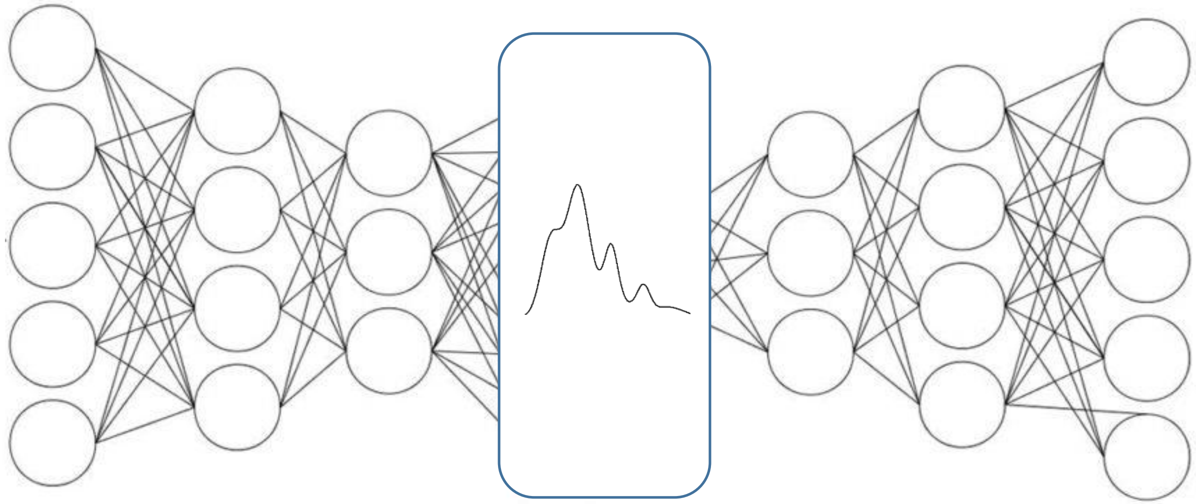
Autoencoder



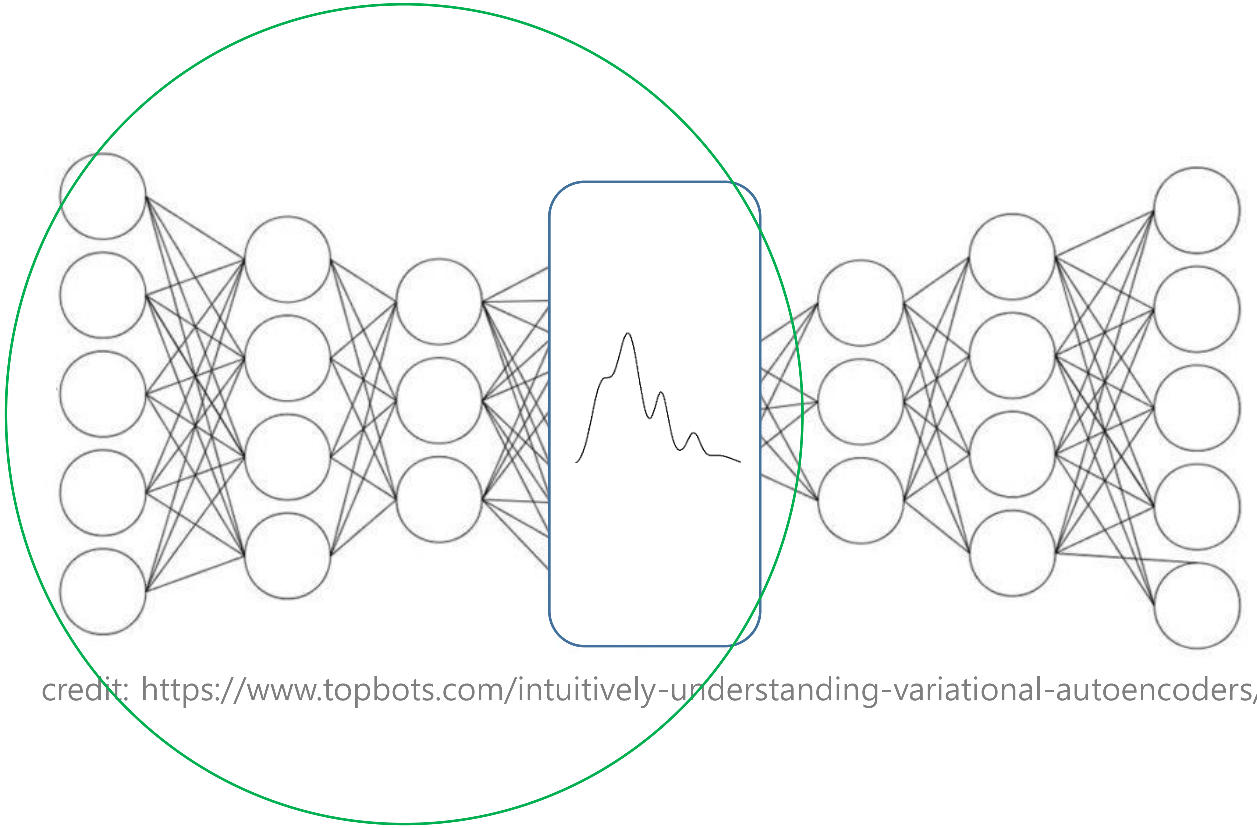


$f_1 : x \rightarrow z$ $f_2 : z' \rightarrow x$

확률 분포를 도입하자!



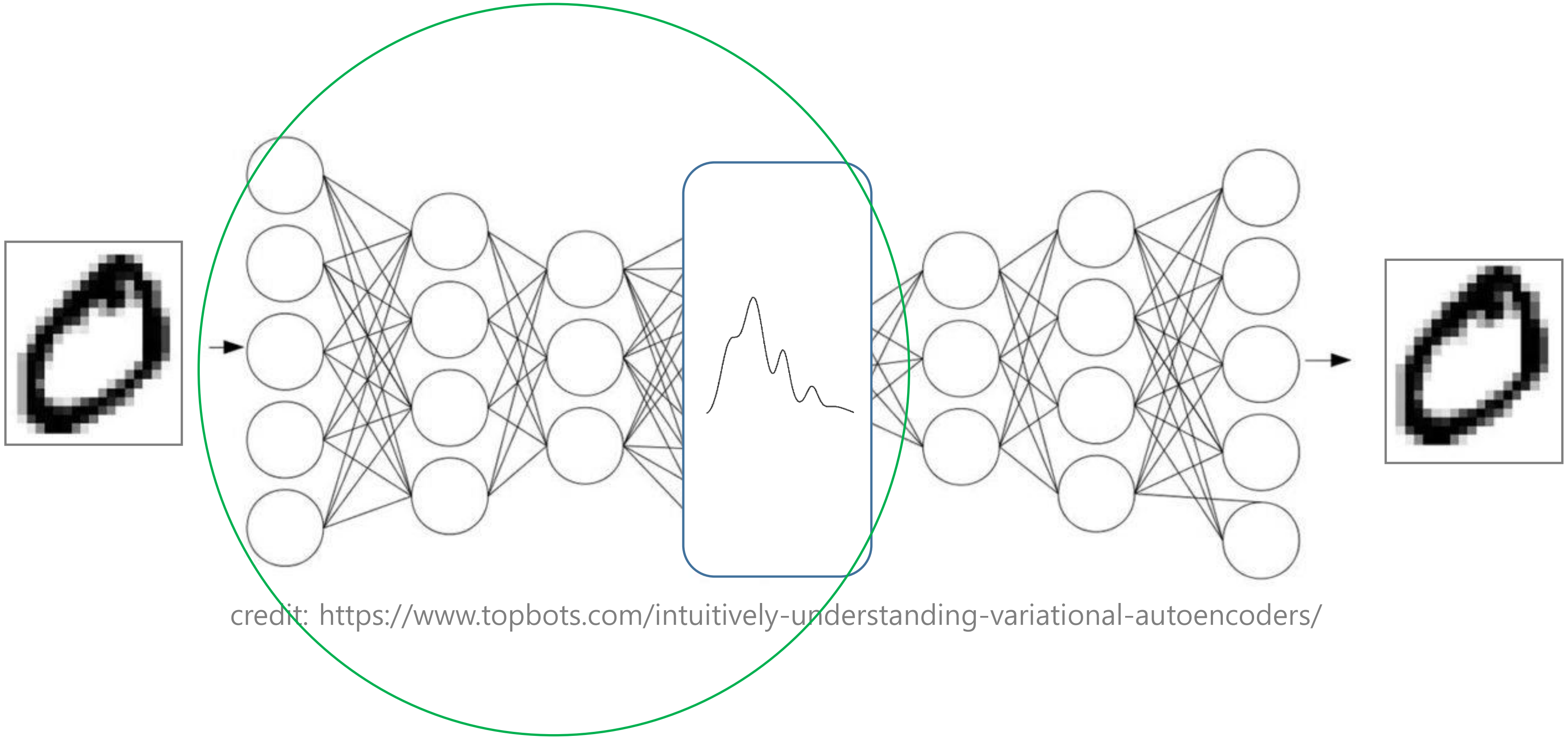
credit: https://www.topbots.com/intuitively-understanding-variational-autoencoders/



credit: https://www.topbots.com/intuitively-understanding-variational-autoencoders/

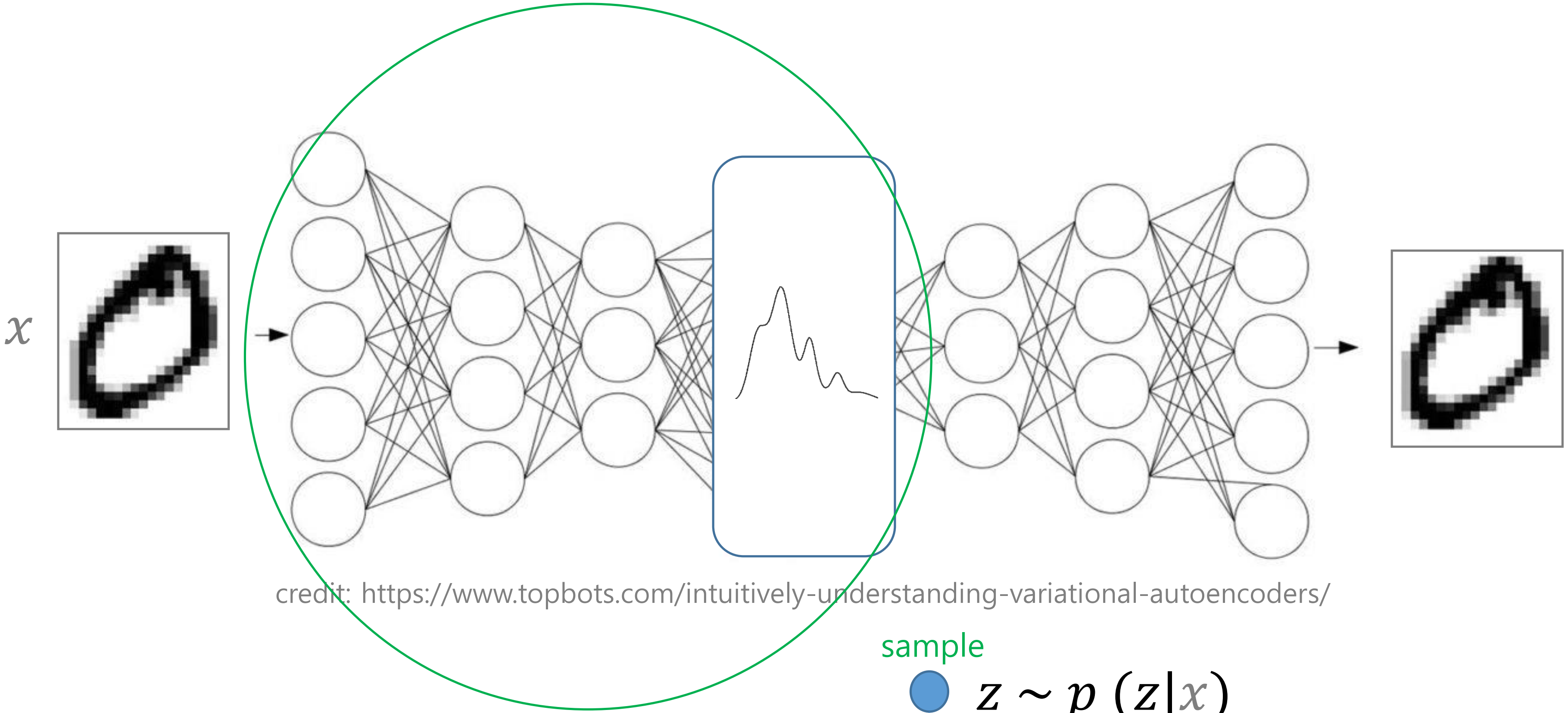
$p(z)$

x
is given



credit: https://www.topbots.com/intuitively-understanding-variational-autoencoders/

$p(z|x)$



credit: https://www.topbots.com/intuitively-understanding-variational-autoencoders/

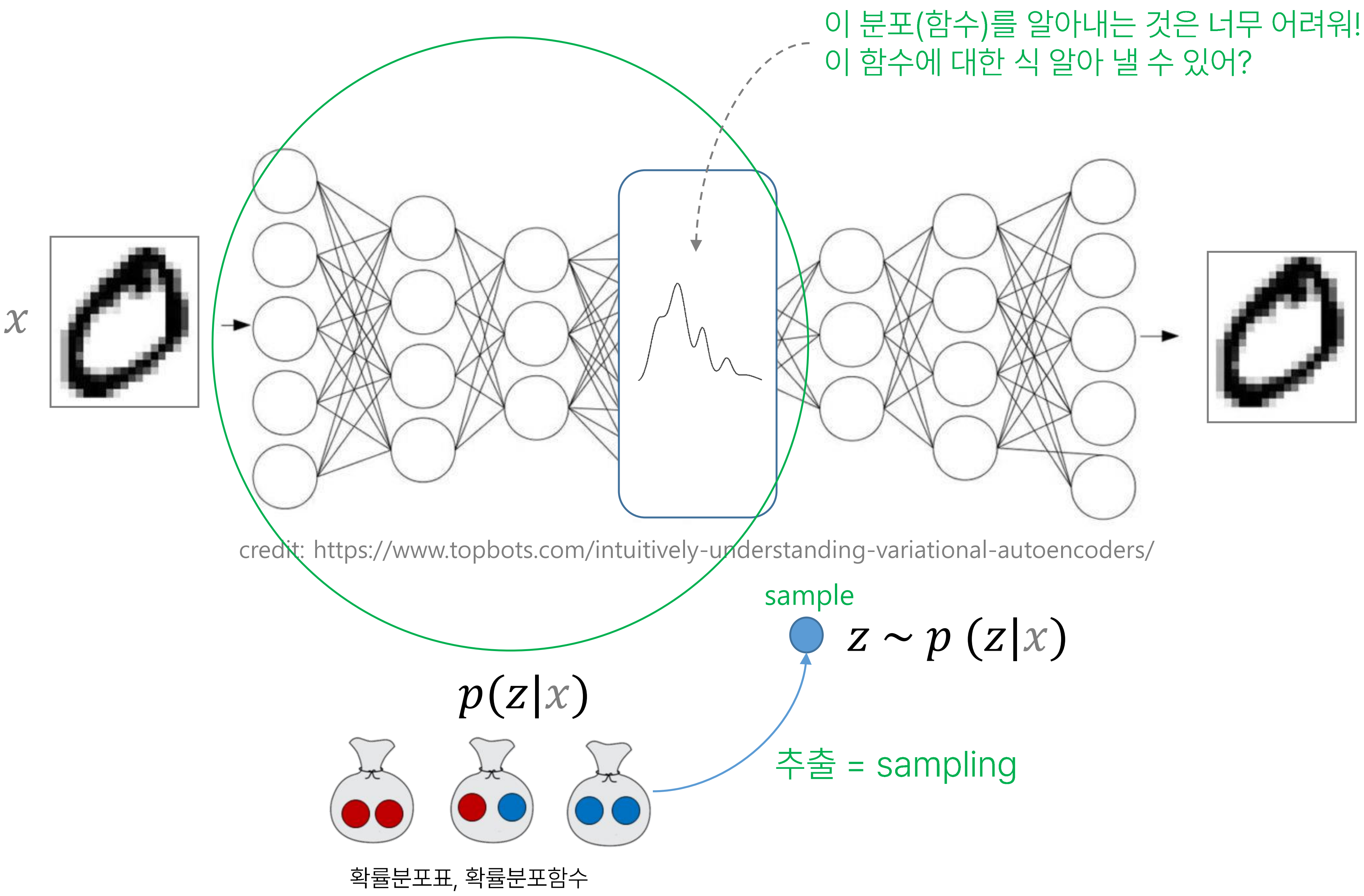
sample

$z \sim p\,(z|x)$

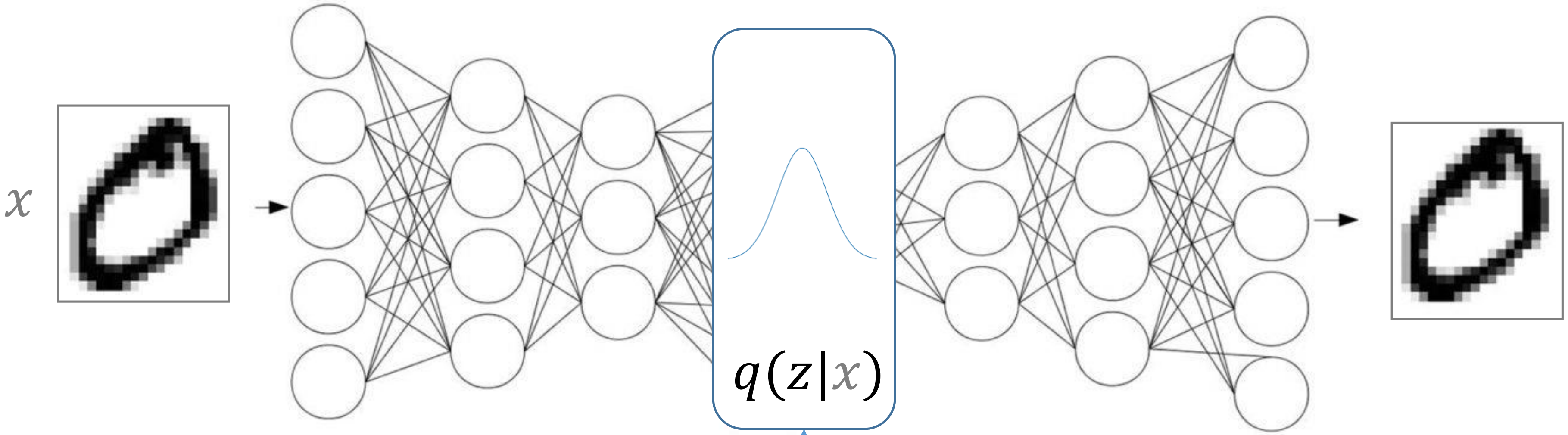
$p(z|x)$

추출 = sampling

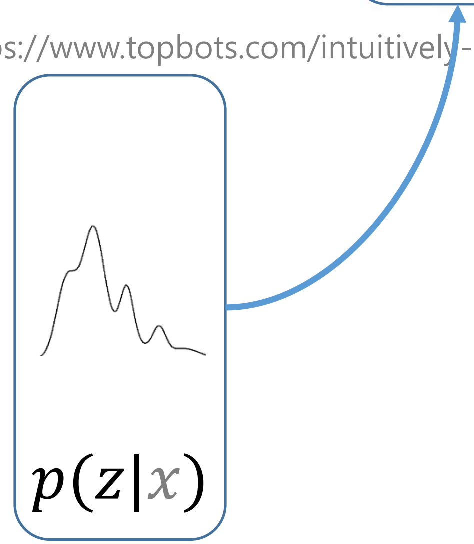
확률분포표, 확률분포함수


이 분포(함수)를 알아내는 것은 너무 어려워!
이 함수에 대한 식 알아 낼 수 있어?
x
credit: https://www.topbots.com/intuitively-understanding-variational-autoencoders/
sample
$z \sim p\,(z|x)$
$p(z|x)$
추출 = sampling
확률분포표, 확률분포함수

Variational Autoencoder (VAE)



credit: https://www.topbots.com/intuitively-understanding-variational-autoencoders/

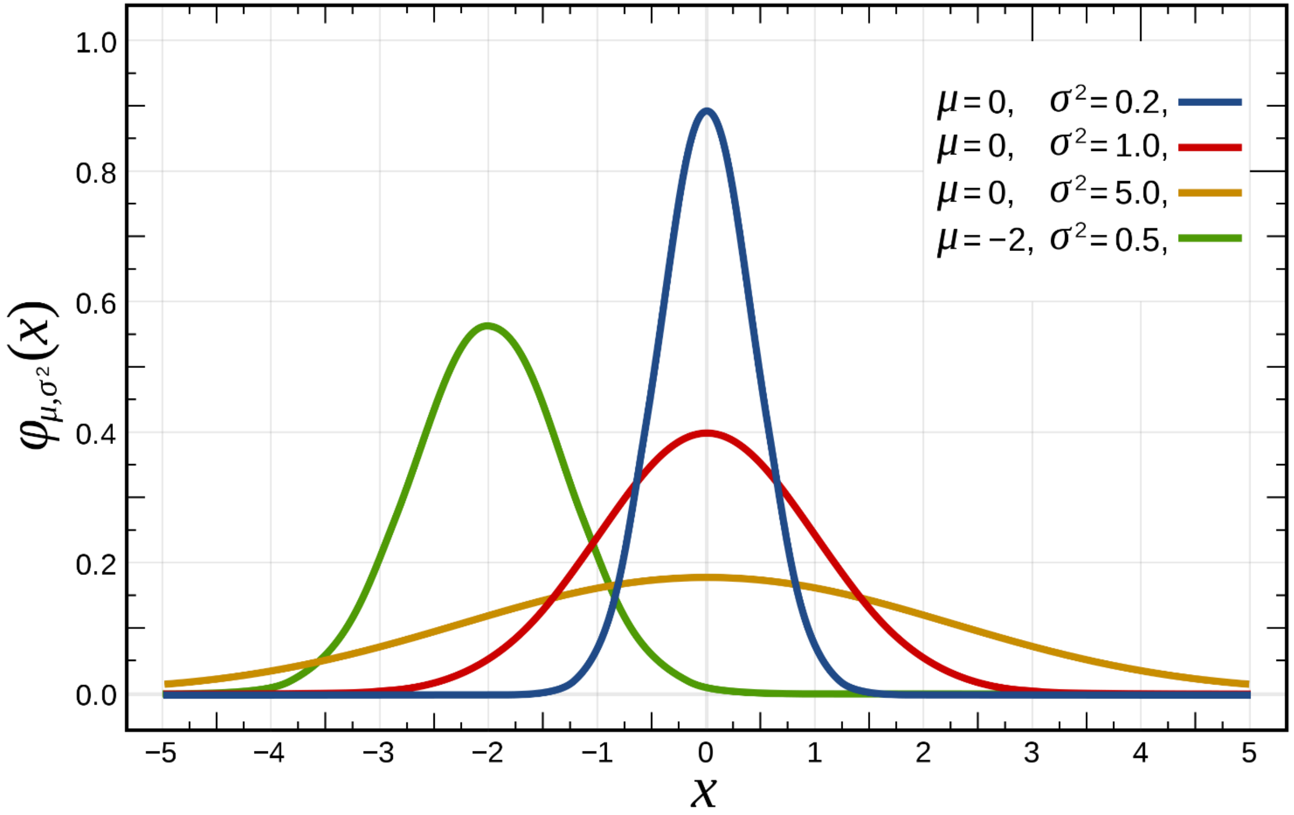


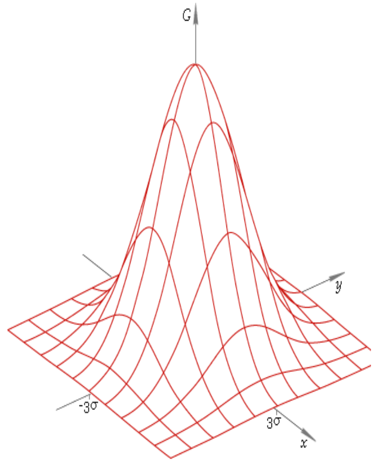
바꾸자!
대체하자!

꿩 **대신** 닭!

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

μ, σ

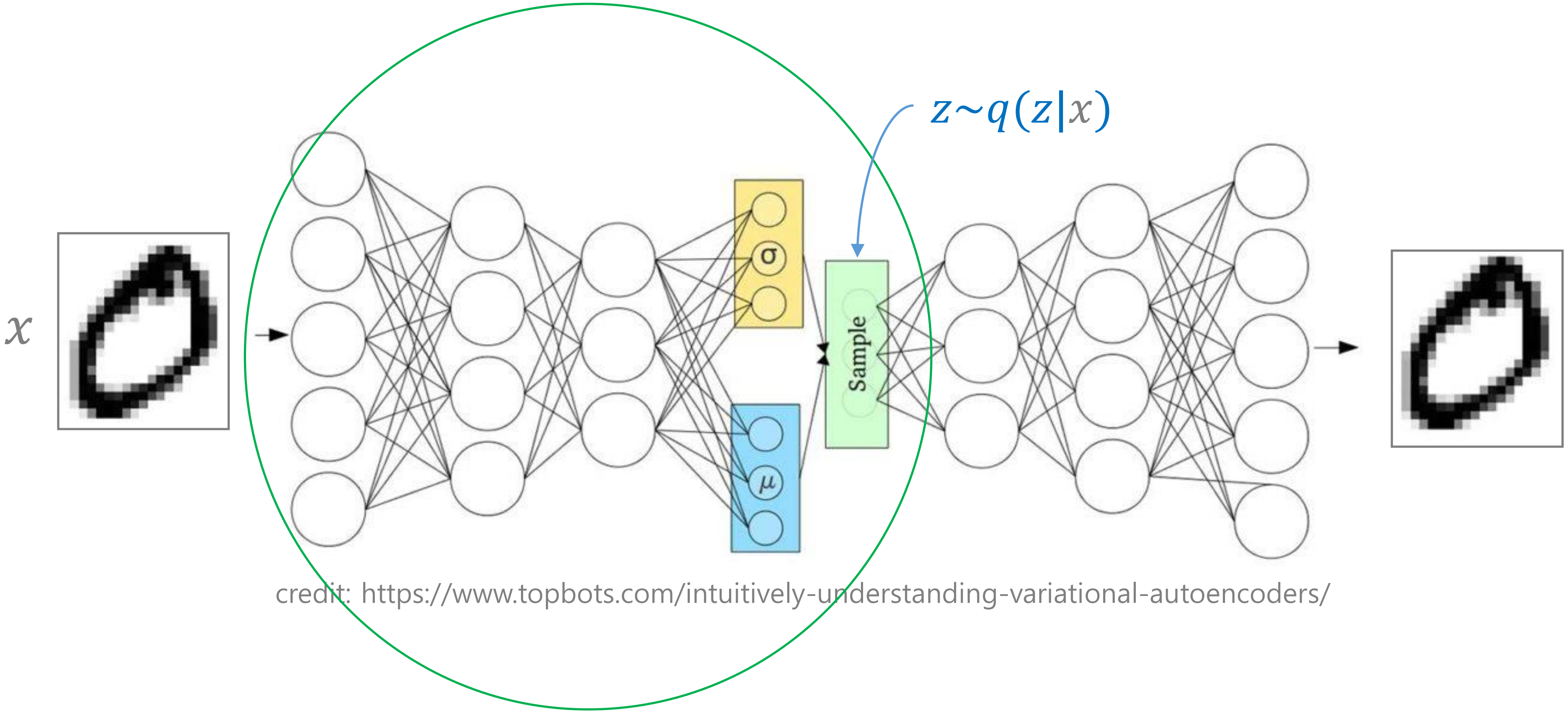




$$\frac{\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)}{\sqrt{(2\pi)^k|\boldsymbol{\Sigma}|}}$$

Gaussian (distribution) function

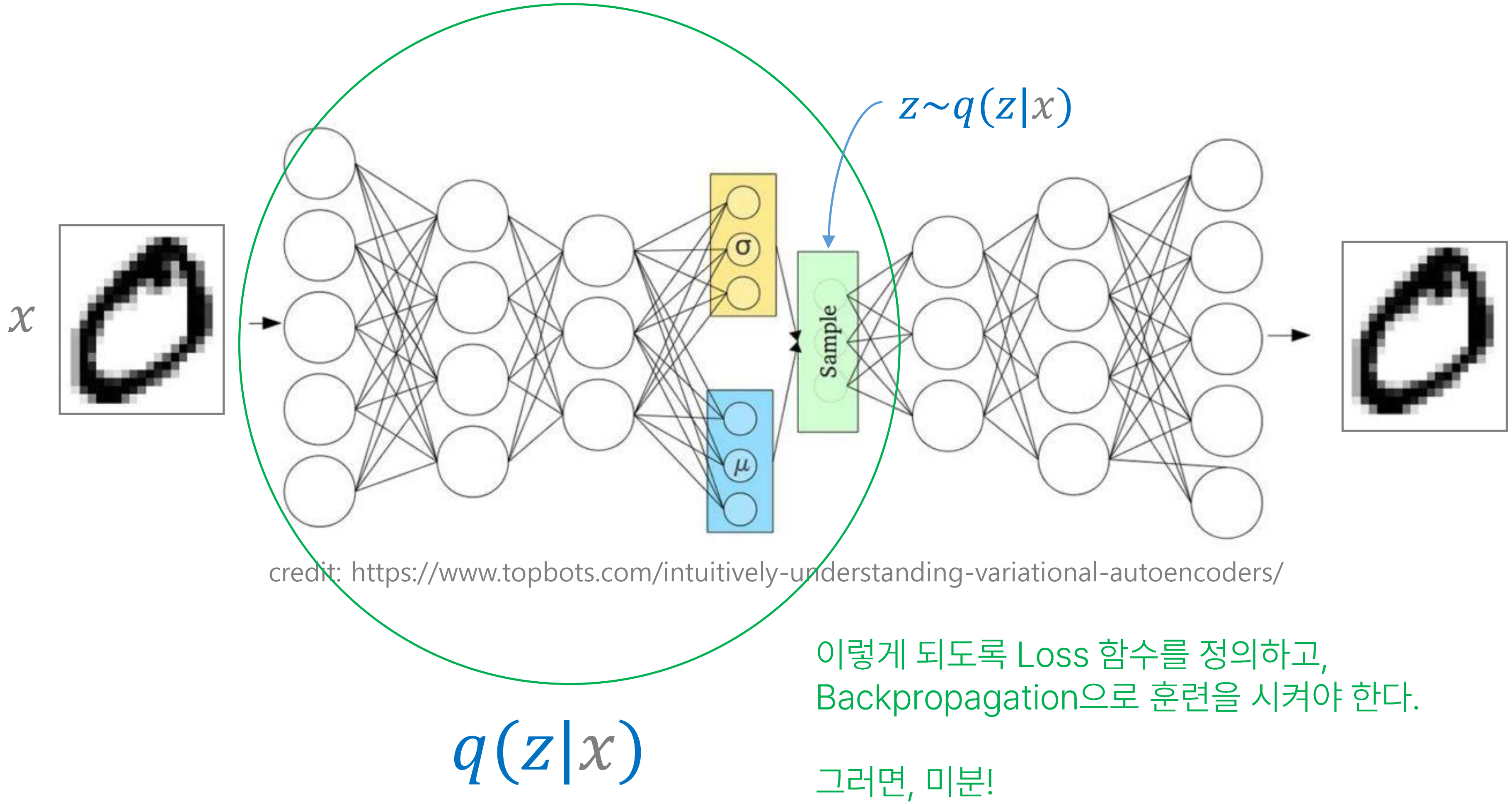
Variational Autoencoder (VAE)



credit: https://www.topbots.com/intuitively-understanding-variational-autoencoders/

$q(z|x)$

Variational Autoencoder (VAE)

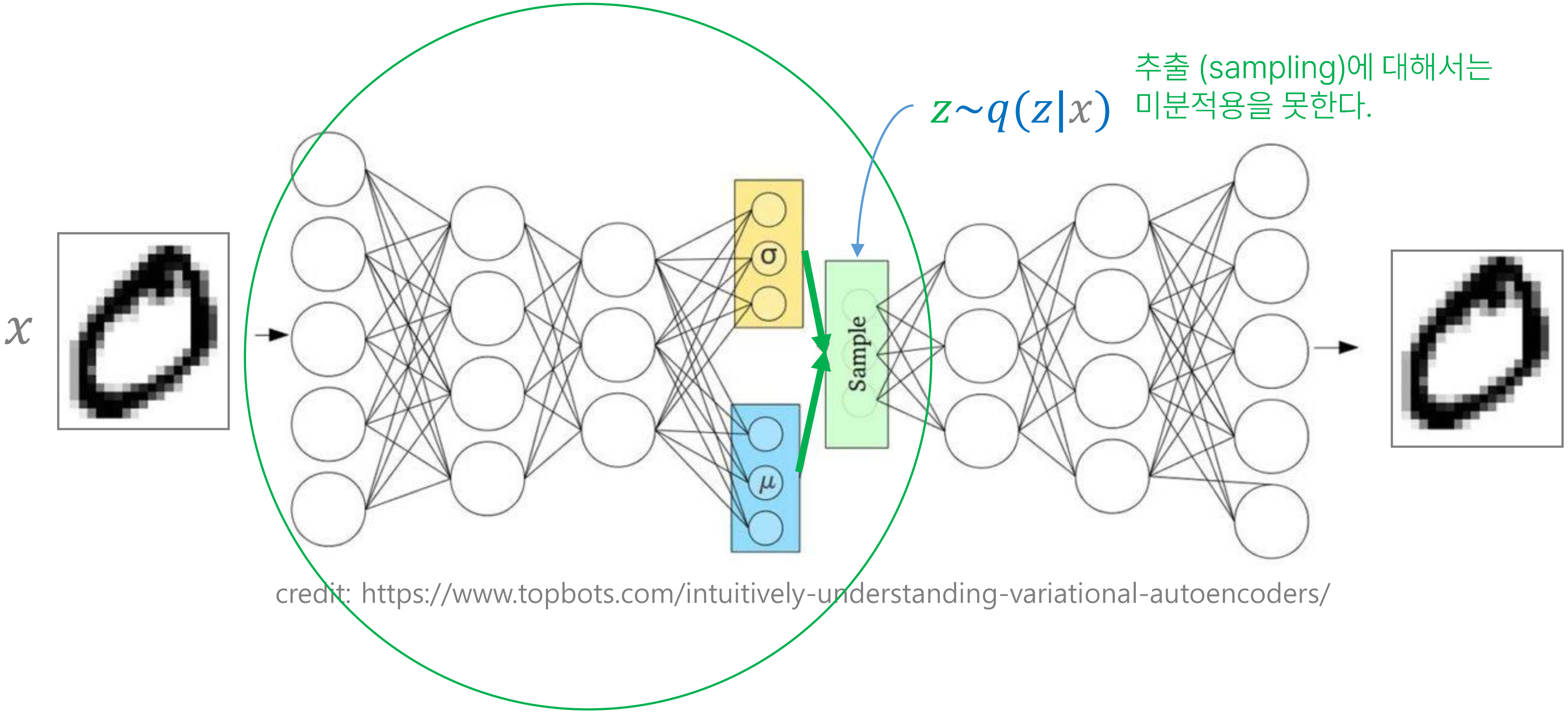


credit: https://www.topbots.com/intuitively-understanding-variational-autoencoders/

이렇게 되도록 Loss 함수를 정의하고,
Backpropagation으로 훈련을 시켜야 한다.

그러면, 미분!
이 그림에서 미분이 안되는 화살표가 있다.

Variational Autoencoder (VAE)



credit: https://www.topbots.com/intuitively-understanding-variational-autoencoders/

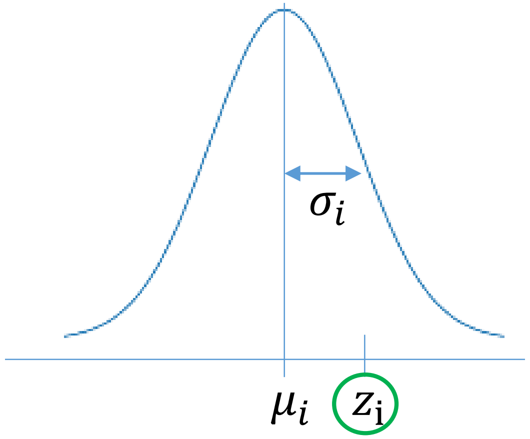
$q(z|x)$

곱하기, 더하기 꼴로 표현되어 있으면...

$$f(x) = wx + b \qquad \frac{\partial f(x)}{\partial w} = x, \frac{\partial f(x)}{\partial b} = 1$$

reparameterization trick

Sampling 은 미분 불가능하므로

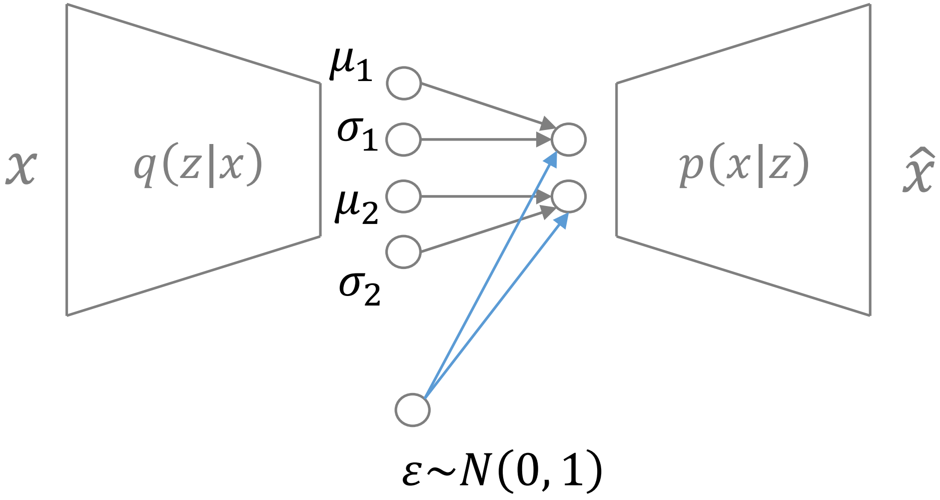


이것을 더하기 곱하기 식으로
표현하자!

$$z_i \sim N(\mu_i, \sigma_i^2)$$

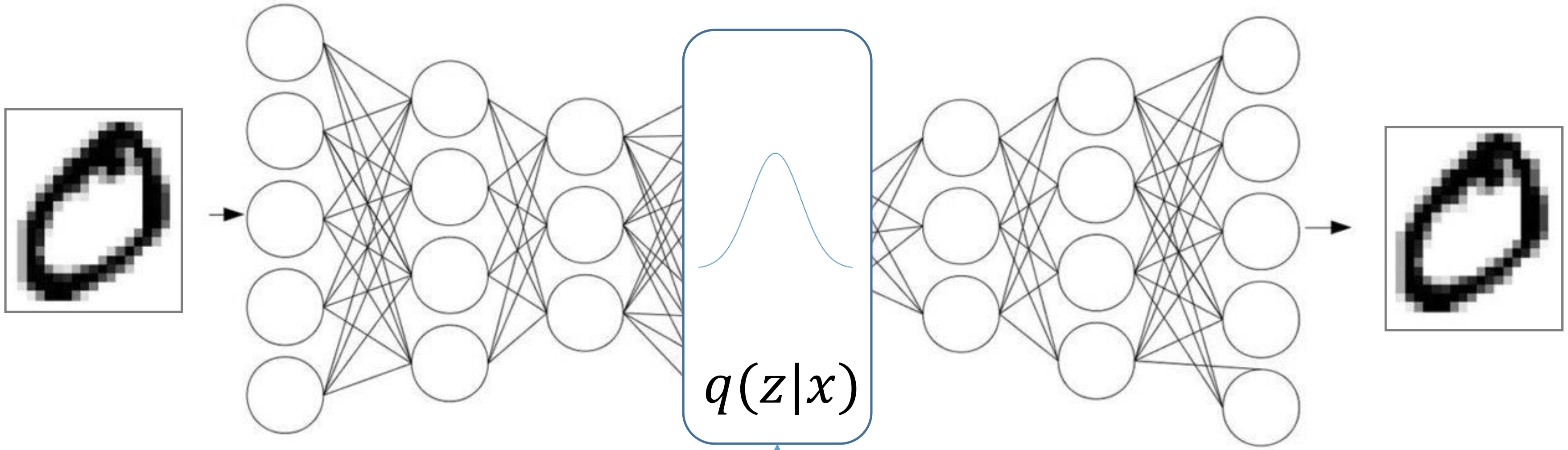
$$\Downarrow$$

$$z_i = \mu_i + \sigma_i \cdot \varepsilon \qquad \varepsilon \sim N(0,1)$$

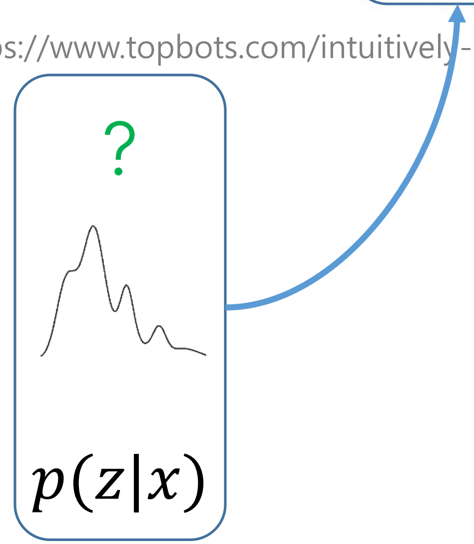


Variational Autoencoder (VAE)

우리가 원하는 목적대로 학습이 되도록 하려면,
손실함수를 어떻게 정해야 하나?



credit: https://www.topbots.com/intuitively-understanding-variational-autoencoders/

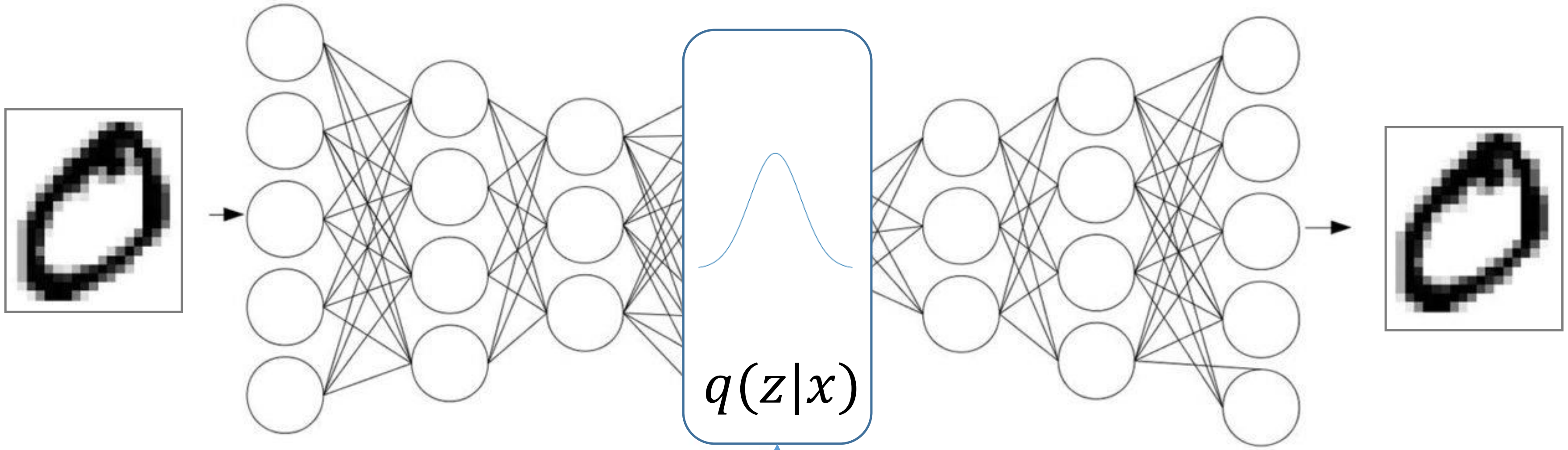


q는 p와 최대한 유사하도록 훈련한다.
즉, 두 분포의 차이가 작아지도록 훈련

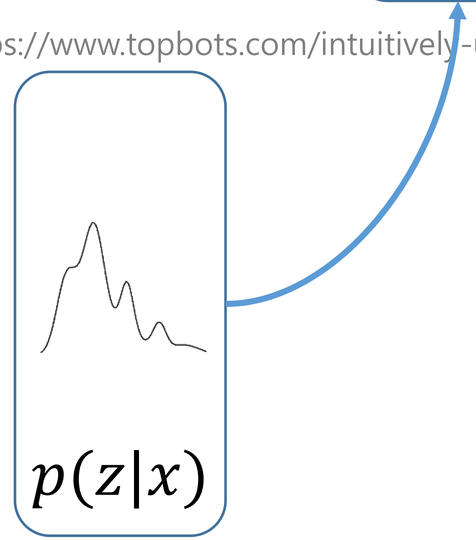
모르는 분포와의 **차이**를 어떻게 구해?

Variational Autoencoder (VAE)

우리가 원하는 목적대로 학습이 되도록 하려면,
손실함수를 어떻게 정해야 하나?



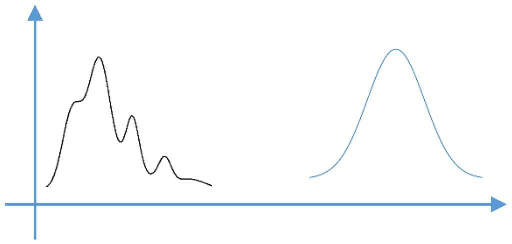
credit: https://www.topbots.com/intuitively-understanding-variational-autoencoders/



q는 p와 최대한 유사하도록 훈련한다.
즉, 두 분포의 차이가 작아지도록 훈련

모르는 분포와의 **차이**를 어떻게 구해?
그래서, 이론이 필요!
Bayesian Theorem(Rule), Information Theory

이런 모양인지 조차도 모르는 분포



$p(z|x)$ $q(z|x)$

모양은 이런데,
좌우 위치와 홀쭉인지 뚱뚱인지 아직 정해지지 않은 분포

μ σ

이 둘을 조정해서 왼쪽 분포와
그나마 유사하게 만들어 줄 수 있다.

이 둘의 유사도(차이)는 어떻게 계산하나?

KL divergence 라는 식이 있다.
(두 분포의 차이를 계산하는 식)

이 식은 엔트로피 항으로 이루어 진다.
엔트로피는 정보량의 평균이다.
정보량은 뉴스(확률사건)의 가치를 계산하는 식

KL divergence 식을 손실함수로 사용하면 된다!

선수지식

random variable, X

$p(X = x), p_X(x), p(x)$

joint pdf, marginal pdf, conditional pdf

$$p(x,y) = p(x)p(y|x) = p(y)p(x|y)$$

$$p(x) = \sum_y p(x,y)$$

$$E[X] = \sum_x p(x)x \qquad E[f(X)] = \sum_x p(x)f(x)$$

$$E_{p(x)}[X]$$

Information, Entropy

$$I(x) = log\frac{1}{p(x)}$$

$$H(X) = E\left[-log^{p(X)}\right] = -\sum_{x} p(x)log^{p(x)}$$

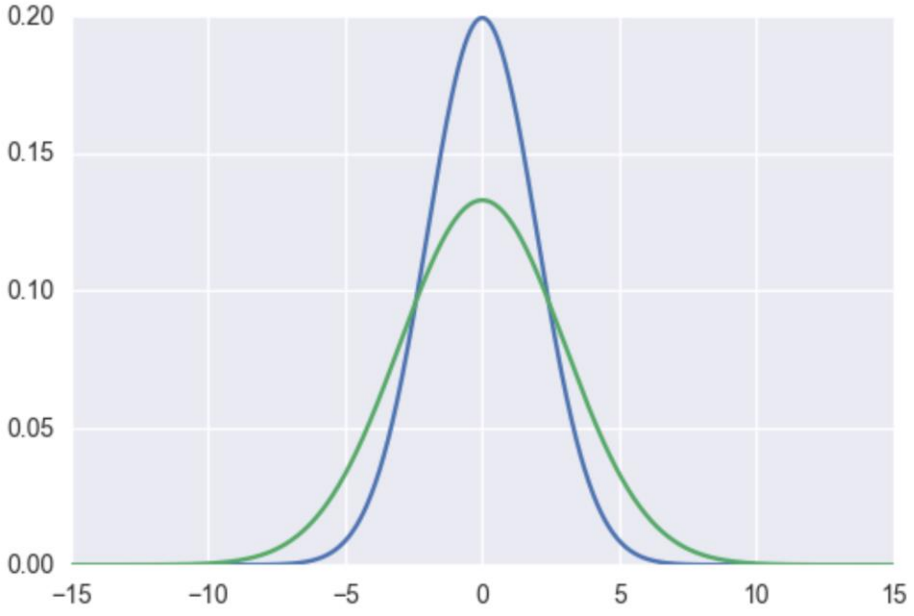
Expectation of Information

$$H(p(X))$$

KL-divergence

두 확률 분포의 어떤 차이

For example, if we have two gaussians, $P(X) = N(0, 2)$ and $Q(X) = N(0, 3)$, how different are those two gaussians?



두 분포의 차이를 어떤 식으로 계산하면 좋을까?

credit: https://wiseodd.github.io/techblog/2016/12/21/forward-reverse-kl/

엔트로피를 활용해 보자

$$H(X) = E\left[-log^{p(X)}\right] = -\sum_{x} p(x)log^{p(x)}$$

Expectation of Information

$$KL(p||q) = -\int p(x)log^{q(x)}dx - \left(-\int p(x)log^{p(x)}dx\right) = -\int p(x)\log\left(\frac{q(x)}{p(x)}\right)dx$$

with respect to

$p(x)$에 대한
$q(x)$의 엔트로피

$p(x)$에 대한
$p(x)$의 엔트로피

$$p(x) - q(x) \approx \frac{q(x)}{p(x)}$$

가중평균 구한 것

$$KL(p||q) \geq 0$$

$$KL(p||q) \neq KL(q||p)$$

$$KL(p\|q) = H(p, q) - H(p) > 0$$

두 분포의 엔트로피가,
한 분포의 엔트로피 보다는 크겠지

Bayes' Rule

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

$$p(z|x) = \frac{p(x, z)}{p(x)} = \frac{p(x|z)p(z)}{p(x)} = \frac{p(x|z)p(z)}{\sum_z p(x, z)} = \frac{p(x|z)p(z)}{\sum_z p(x|z)p(z)}$$

Bayesian Inference

Posterior(사후확률) 구하기

likelihood prior

$$p(\mathbf{z}|x) = \frac{p(x|\mathbf{z})p(\mathbf{z})}{p(x)}$$

posterior

다루기 쉬운 분포로 가정

관찰 값

c. f) classification $p(\mathbf{z}|x) \propto p(x|\mathbf{z})p(\mathbf{z})$

관찰 데이터에 대한 분포 구할 수 있나?

$$p(x) = \sum_{\mathbf{z}} p(x|\mathbf{z})p(\mathbf{z})$$

$$= \underbrace{\sum^{m} \sum \cdots \sum}_{n = |\mathbf{z}|} p(x|\mathbf{z})p(\mathbf{z})$$

계산량이 너무 많다.
(m^n)

intractable

Variational Inference

Posterior 구하기

꿩 **대신** 닭

변이

$p(z|x)$ 구하기 어렵다.

$q(z|x)$ 우리가 알고 있는 쉬운 분포; 정규분포로 대치하자.

대체(변이) 사후확률 구함

조건: $p(z|x) \approx q(z|x)$

$$\min_{q(z|x)} \mathrm{KL}\big(q(z|x)||p(z|x)\big)$$

KL 최소화

Reverse KL을 선호

x in black: a random variable
x in gray: a fixed value of the random variable x

꿩 대신 닭

$$KL(q(z|x)||p(z|x)) = -\sum q(z|x)\log\left(\frac{p(z|x)}{q(z|x)}\right)$$

$$= -\sum q(z|x)\log\left(\frac{\frac{p(x,z)}{p(x)}}{\frac{q(z|x)}{1}}\right)$$

$$= -\sum q(z|x)\log\left(\frac{p(x,z)}{q(z|x)}\frac{1}{p(x)}\right)$$

$$= -\sum q(z|x)\left(\log\left(\frac{p(x,z)}{q(z|x)}\right) + \log\left(\frac{1}{p(x)}\right)\right)$$

$$= -\sum q(z|x)\log\left(\frac{p(x,z)}{q(z|x)}\right) + \sum_z q(z|x)\log(p(x))$$

$$= -\sum q(z|x)\log\left(\frac{p(x,z)}{q(z|x)}\right) + \log(p(x))\sum_z q(z|x)$$ (→ 1)

$$= -\sum q(z|x)\log\left(\frac{p(x,z)}{q(z|x)}\right) + \log(p(x))$$

x is given, then it is fixed

$$\log(p(x)) = KL(q(z|x)||p(z|x)) + \sum q(z|x)\log\left(\frac{p(x,z)}{q(z|x)}\right)$$

≥ 0

최소화 해야함 → 최대화 하면 된다.

$$\log(p(x)) \geq \sum q(z|x)\log\left(\frac{p(x,z)}{q(z|x)}\right)$$ ↑

https://www.youtube.com/watch?v=uaaqyVS9-rM

x in black: a random variable
x in gray: a fixed value of the random variable x

꿩 대신 닭

$$KL(q(z|x)||p(z|x)) = -\sum q(z|x)log\left(\frac{p(z|x)}{q(z|x)}\right)$$

$$= -\sum q(z|x)log\left(\frac{\frac{p(x,z)}{p(x)}}{\frac{q(z|x)}{1}}\right)$$

$$= -\sum q(z|x)log\left(\frac{p(x,z)}{q(z|x)}\frac{1}{p(x)}\right)$$

$$= -\sum q(z|x)\left(log\left(\frac{p(x,z)}{q(z|x)}\right) + log\left(\frac{1}{p(x)}\right)\right)$$

$$= -\sum q(z|x)log\left(\frac{p(x,z)}{q(z|x)}\right) + \sum_z q(z|x)log(p(x))$$

$$= -\sum q(z|x)log\left(\frac{p(x,z)}{q(z|x)}\right) + log(p(x))\sum_z q(z|x)$$

($\sum_z q(z|x) \to 1$)

$$= -\sum q(z|x)log\left(\frac{p(x,z)}{q(z|x)}\right) + log(p(x))$$

x is given, then it is fixed

$$log(p(x)) = KL(q(z|x)||p(z|x)) + \sum q(z|x)log\left(\frac{p(x,z)}{q(z|x)}\right)$$

≥ 0

최소화 해야함 → 최대화 하면 된다.

$log(p(x)) \geq$ variational **lower bound**

Log-Likelihood 크게 하는 것

https://www.youtube.com/watch?v=uaaqyVS9-rM lower bound   when x is given

$$\log(p(x)) \geq \sum q(z|x)\log\left(\frac{p(x,z)}{q(z|x)}\right)$$

Log-Likelihood
크게 하는 것

$$= \sum q(z|x)\log\left(\frac{p(x|z)p(z)}{q(z|x)}\right) \qquad \sout{p(z|x)p(x)}$$

$$= \sum q(z|x)\left(\log(p(x|z)) + \log\left(\frac{p(z)}{q(z|x)}\right)\right)$$

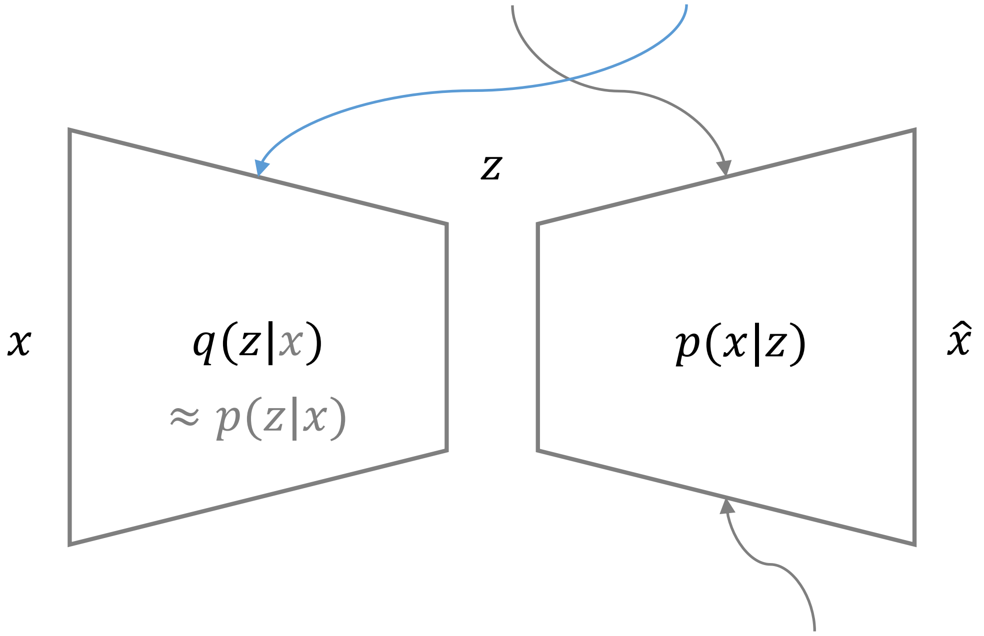
$$= \sum q(z|x)\log(p(x|z)) + \sum q(z|x)\log\left(\frac{p(z)}{q(z|x)}\right)$$

$$= E_{q(z|x)}[\log(p(x|z))] - KL(q(z|x)||p(z))$$

Evidence Lower BOund (ELBO)

Loss for a given x:

$$-E_{q(z|x)}\left[log\left(p(x|z)\right)\right] + KL\left(q(z|x)||p(z)\right)$$



그냥 신경망;
일반적인 복원 에러를 로스로 사용
Cross Entropy Loss

Loss 함수 식 정리

regularizer term

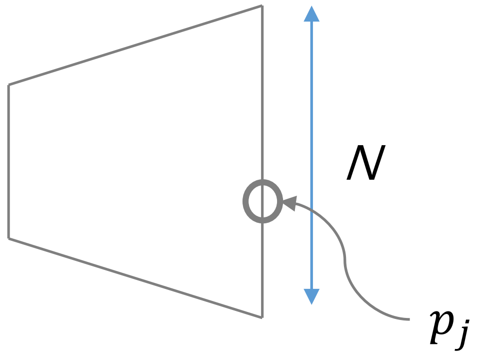
$$\sum\left(-E_{q(z|x_i)}\left[log\left(p(x_i|z)\right)\right] + KL\left(q(z|x_i)||p(z)\right)\right)$$

reconstruction term

두 분포 모두 가우시안 가정

베르누이 분포로 간주

$$\sum\left(-E_{q(z|x_i)}\left[log\left(p(x_i|z)\right)\right] + KL\left(q(z|x_i)||p(z)\right)\right)$$



j번째 노드 출력이 1일 확률

Monte Carlo **Sampling**

$$\approx \frac{1}{L}\sum log\left(p\left(x_i|z^l\right)\right)$$

경험적

$$\approx log\left(p(x_i|z^1)\right) = log\left(\prod_{j=1}^{N} p\left(x_{i,j}|z^1\right)\right) = \sum_{j=1}^{N} log\left(p\left(x_{i,j}|z^1\right)\right)$$

$$= \sum_{j=1}^{N} log\left(p_j^{x_{i,j}} \cdot (1-p_j)^{1-x_{i,j}}\right) = \sum_{j=1}^{N} x_{i,j}log(p_j) + (1-x_{i,j})log(1-p_j)$$

cross-entropy

network input (target)

network output

가우시안 분포로 간주

$$\sum\left(-E_{q(z|x_i)}\left[log\left(p(x_i|z)\right)\right] + KL\left(q(z|x_i)||p(z)\right)\right)$$

Monte Carlo **Sampling**

$$\approx \frac{1}{L}\sum log\left(p\left(x_i|z^l\right)\right)$$

경험적

$$\approx log\left(p(x_i|z^1)\right)$$

$$= log\left(N\left(\mu_i, \sigma_i^2 I\right)\right)$$

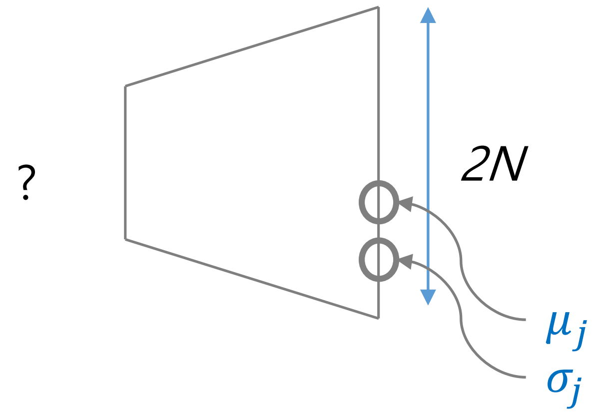
$$= -\sum_{j=1}^{N}\left(\frac{1}{2}log\left(\sigma_{i,j}^2\right) + \frac{\left(x_{i,j}-\mu_{i,j}\right)^2}{2\sigma_{i,j}^2}\right)$$

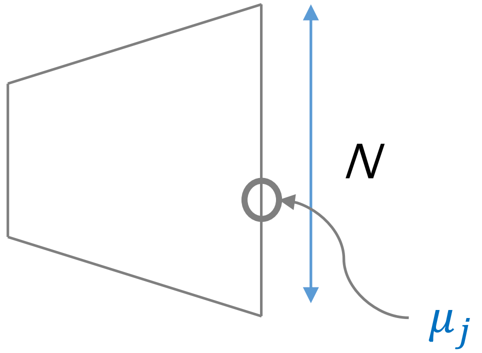
분산이 모두 같다고 가정

$$log\left(N\left(\mu_i, \sigma^2 I\right)\right)$$

$$\propto -\sum_{j=1}^{N}\left(x_{i,j}-\mu_{i,j}\right)^2$$

mean squared loss



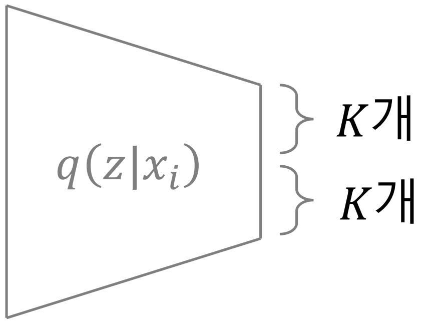


KL term

초간단 분포

$$\sum\left(-E_{q(z|x_i)}\left[log\left(p(x_i|z)\right)\right] + KL\left(q(z|x_i)||p(z)\right)\right)$$

x_i에 따라 달라지는 분포



covariance 0 가정,
u_i K개, σ_i^2 K개

$$KL\left(N\left(u_i, \sigma_i^2 I\right)||N(0, I)\right)$$

$$= \frac{1}{2}\left(tr\left(\sigma_i^2 I\right) + \mu_i^T \mu_i - K - log\left(\prod_{k=1}^{K} \sigma_{i,k}^2\right)\right)$$

$$= \frac{1}{2}\left(\sum_{k=1}^{K} \sigma_{i,k}^2 + \sum_{k=1}^{K} \mu_{i,k}^2 - \sum_{k=1}^{K} 1 - \sum_{k=1}^{K} log\left(\sigma_{i,k}^2\right)\right)$$

$$= \frac{1}{2}\sum_{k=1}^{K}\left(\sigma_{i,k}^2 + \mu_{i,k}^2 - log\left(\sigma_{i,k}^2\right) - 1\right)$$

Loss for a given x_i

$$L_i(\phi, \theta, x_i) = -\sum_{j=1}^{N} x_{i,j} log\left(x'_j\right) + \left(1 - x_{i,j}\right) log\left(1 - x'_j\right) + \frac{1}{2}\sum_{k=1}^{K}\left(\sigma_{i,k}^2 + \mu_{i,k}^2 - log\left(\sigma_{i,k}^2\right) - 1\right)$$

$$L_i(\phi, \theta, x_i) = \sum_{j=1}^{N}\left(\frac{1}{2} log\left(\sigma_{i,j}^2\right) + \frac{\left(x_{i,j} - \mu_{i,j}\right)^2}{2\sigma_{i,j}^2}\right) + \frac{1}{2}\sum_{k=1}^{K}\left(\sigma_{i,k}^2 + \mu_{i,k}^2 - log\left(\sigma_{i,k}^2\right) - 1\right)$$

분산이 모두 같다고 가정

$$L_i(\phi, \theta, x_i) = \sum_{j=1}^{N}\left(x_{i,j} - \mu_{i,j}\right)^2 + \frac{1}{2}\sum_{k=1}^{K}\left(\sigma_{i,k}^2 + \mu_{i,k}^2 - log\left(\sigma_{i,k}^2\right) - 1\right)$$

참고문헌

- Tutorial on Autoencoder, https://arxiv.org/pdf/1606.05908.pdf (2016)
- https://www.youtube.com/watch?v=uaaqyVS9-rM
- https://wiseodd.github.io/techblog/2016/12/10/variational-autoencoder/
- https://www.slideshare.net/NaverEngineering/ss-96581209

감사합니다.

1 0

nonezero@kumoh.ac.kr
cvpr.kumoh.ac.kr/nonezero