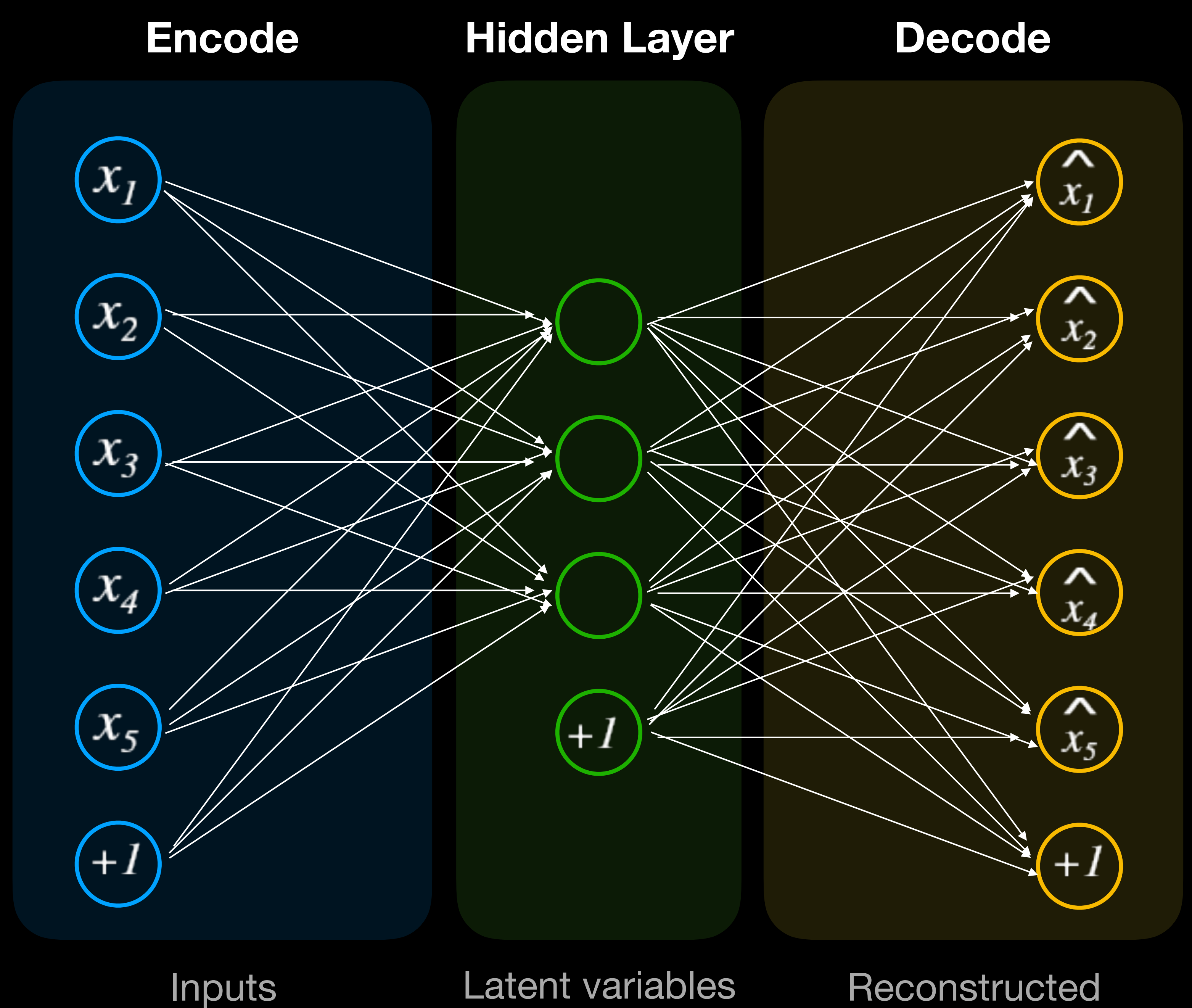
VAE / ELBO

POSCO ICT R&D center
김성욱

Index

- AE / VAE의 개념
- KL Divergence란
- Bayes Rule의 조건부 확률 개념
- Monte Carlo Approximation : 샘플을 통한 근사
- ELBO(evidence lower bound)

Auto Encoder

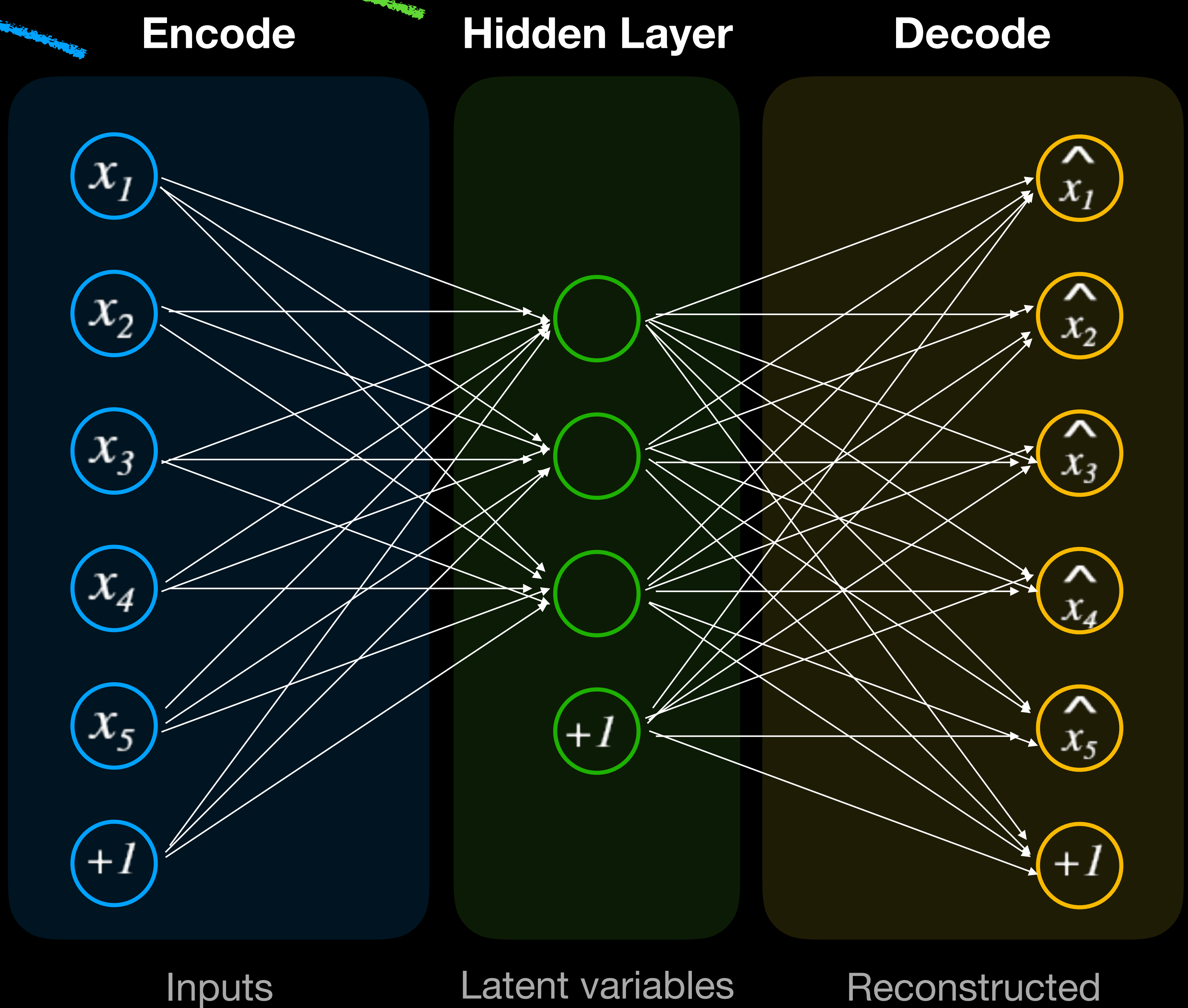


* 용도 : 정보 압축 및 복원, De-noising 등

수학적으로 PCA와 유사한 의미를 지님

* 고차원 정보 —> 저차원에 정보 압축 —> 고차원 공간으로 복원

Auto Encoder



* 이미지와 같은 고차원 데이터를 Encoding을 통해 저차원 Hidden 공간으로 표현한다면..

예) 사람의 얼굴 이미지(64*64 픽셀)

—> 특징 6가지(피부색, 성별, 눈, 코, 입, 귀)

: 4096 차원의 정보를 6차원으로 압축

@ 이러한 의미있는 정보를 담는 차원 공간을 Latent Space, 정보를 담는 변수를 Latent Variables 라고 부른다.

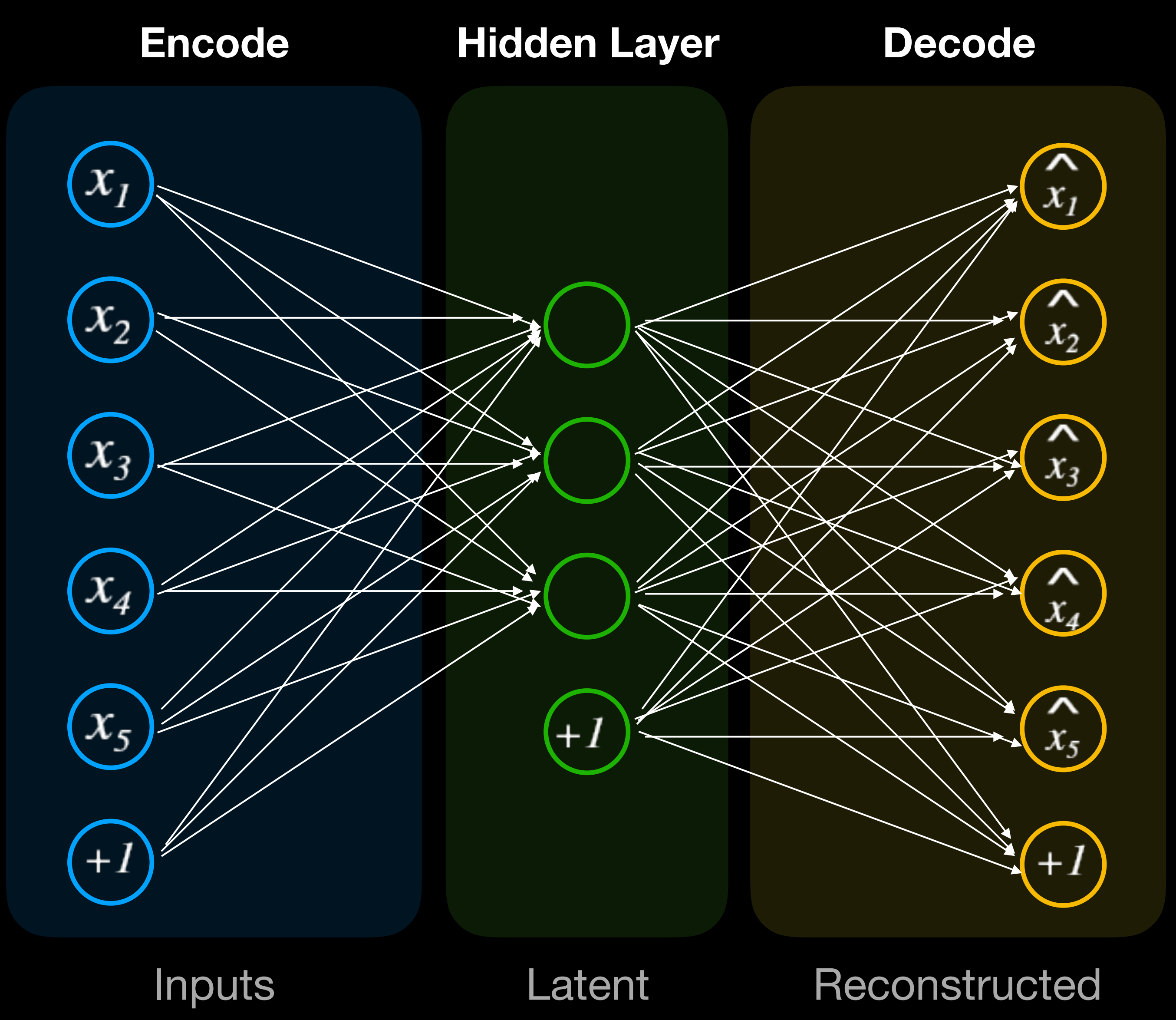
* Latent Variables의 정보를 기반으로 원래의 이미지를 복원(Reconstruction) 과정을 수행

예) 거무스름하며 남자 얼굴이고, 눈이 작고, 코가 크고, 입이 작으며, 귀가 큰 사람 얼굴은?

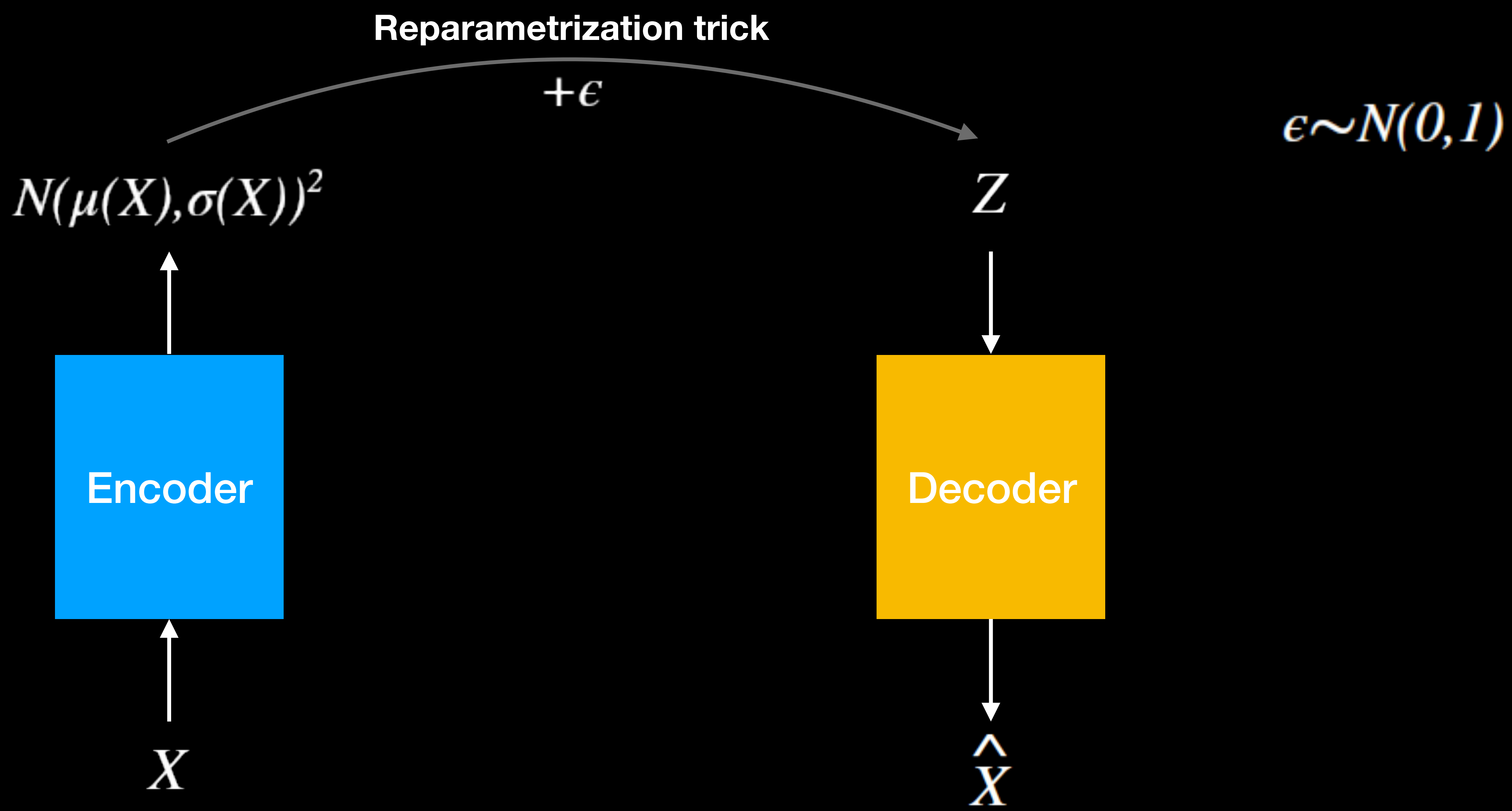
Auto Encoder + Generative model

* Auto-Encoder 모델을 기반으로 데이터를 생성하는 모델을 만들고자 한다면..?
* Latent Space에는 사람의 얼굴에 대한 특징들을 벡터로 표현해 두었으니까, 이 벡터 공간의 값들을 조정하면 새로운 얼굴을 만들어 낼 수 있지 않을까?

그럼, Latent Space에서 Sampling을 해보자



VAE(variational auto-encoder)
- Overview of Structure



주어진 샘플(X)들을 잠재 공간(latent space) 상에 잘 알려진 분포로 표현하고, 이 분포에서부터 샘플링을 통해 X를 추정하는 것
구해진 분포(mu, sigma)에서 부터 샘플링을 하게 된다면, 학습 시 샘플링된 z로부터 encoder의 파라미터와의 연결 관계를 찾아갈 수 없다.
따라서, reparametrization trick을 적용한다. 계산된 분포(mu, sigma)에 epsion을 더하는 방식으로 z를 샘플링하면, 추후 학습 시 z의
encoder의 파라미터가 영향을 미치는 관계를 찾아갈 수 있다. 즉, **미분이 가능한 모델**로 표현이 된다.

KL Divergence

- Kullback-Leibler divergence : 두 확률 분포의 다름의 정도 (=relative entropy)

$$D_{KL}(p||q) = \int p(x)\log\frac{p(x)}{q(x)}\,dx$$
$$= \int p(x)\log p(x)\,dx - \int p(x)\log q(x)\,dx$$

KL Divergence

- KL divergence의 특징 / 성질
  - 항상 0 이상의 값을 지님

$$D_{KL}(p||q) \geq 0$$

  - 두 분포가 같을 때 값이 0, 분포가 다를 수록 값이 커짐
  - 계산 순서가 바뀌면 값이 변할 수 있음

$$D_{KL}(p||q) \neq D_{KL}(q||p)$$

$$\int p(x) \log p(x)\, dx - \int p(x) \log q(x)\, dx \neq \int q(x) \log q(x)\, dx - \int q(x) \log p(x)\, dx$$

KL Divergence

- KL divergence의 특징 / 성질
  - **두 분포가 Gaussian Distribution을 따를 경우 간략하게 표현 가능함**

$$D_{KL}(p||q) = -\int p(x)\log q(x)\,dx + \int p(x)\log p(x)\,dx$$

$$= \frac{1}{2}\log(2\pi\sigma_2^2) + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} - \frac{1}{2}(1 + \log 2\pi\sigma_1^2)$$

$$= \log\frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} - \frac{1}{2}$$

$$where \quad p(x) = N(\mu_1, \sigma_1), \quad q(x) = N(\mu_2, \sigma_2), \quad N(\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Reference: https://stats.stackexchange.com/questions/7440/kl-divergence-between-two-univariate-gaussians

Bayes Rule

$$P(H|E)=\frac{P(H)*P(E|H)}{P(E)}$$

E : Evidence
H : Hypothesis

P(H) : Prior Probability - 사전에 알고 있는 H가 발생할 확률
P(E|H) : Likelihood of the evidence ‘E’ if the Hypothesis ‘H’ is true - 모든 사건 ‘H’에 대한 E가 발생할 우도
- How well H explains E : 각각의 H가 E를 얼마나 잘 나타내는가.
P(E) : Priori probability that the evidence itself is true - ‘E’에 대한 사전 확률. E가 발생할 확률
P(H|E) : Posterior Probability of ‘H’ given the evidence - E가 주어졌을 때, H가 발생할 사후 확률

“주어진 현 사건(E:evidence)에 대해서 H(hypothesis)가 발생할 믿음” 정도로 해석

Monte Carlo Approximation

$$E_{p(x)}[f(x)] = \int f(x)p(x)dx \approx \frac{1}{N} \sum_{i=1}^{N} f(x_i), \quad x_i \sim p(x)$$

확률 밀도 함수 p(x) 를 따르는 x에 대한 f(x)의 기대값은 p(x)를 따르는 샘플들로 근사할 수 있다.

ELBO: Evidence Lower Bound - step1, 2

Step 1.

$$\log P(x_i) = \log \frac{P(x_i|z)P(z)}{P(z|x_i)}$$
$$= \log P(x_i|z) + \log P(z) - \log P(z|x_i) - \log P(z|x_i)$$

1. Bayes Rule 조건부 확률 식을 이용하여, 샘플 x_i가 발생할 확률을 치환
2. log를 취함으로써, multiply 텀을 분해

해석 - 식의 목적은 "**주어진 샘플 x_i의 확률 분포를 잘 표현해보자**" 이다. 이를 위해 잠재 변수 z의 분포를 이용하고자 한다.

Step 2.

$$\log P(x_i) = \log P(x_i) \int q(z|x_i)dz$$
$$= \int q(z|x_i) \log P(x_i)\, dz$$

모든 샘플 x_i에 대한 샘플 z가 생성될 확률 밀도함수의 적분 $\int q(z|x_i)dz = 1$

해석 - 위 step1의 목적과 같으며, 여기에 확률 분포를 따르는 잠재변수 z의 확률 밀도함수의 성질을 이용하여 식을 변형한다.

ELBO: Evidence Lower Bound - step3

Step 3. step2의 식에 step 1의 결과를 대입

$$\log P(x_i) = \int q(z|x_i)[\log P(x_i|z) + \log P(z) - \log P(z|x_i)]dz$$

$$= E_{q(z|x_i)}[\log P(x_i|z)] + \int q(z|x_i)\log p(z)\,dz - \int q(z|x_i)\log P(z|x_i)\,dz \pm \int q(z|x_i)\log q(z|x_i)\,dz$$

$$= E_{q(z|x_i)}[\log P(x_i|z)]$$
$$- \int q(z|x_i)\log q(z|x_i)\,dz + \int q(z|x_i)\log p(z)\,dz$$
$$+ \int q(z|x_i)\log q(z|x_i)\,dz - \int q(z|x_i)\log P(z|x_i)\,dz$$

$$= E_{q(z|x_i)}[\log P(x_i|z)] - D_{KL}(q(z|x_i)||P(z)) + D_{KL}(q(z|x_i)||P(z|x_i))$$

부분 해석 - 잠재변수 z는 우리가 알고있는 확률 분포(Gaussian Distribution)을 따르기 때문에 P(z)의 계산이 용이하다.
즉, P(z) ~ N(mu, var)

부분 해석 - 알고있는 샘플 x_i의 분포는 실제 알지 못하는 매우 복잡한 분포를 나타내기 때문에 P(z|x_i)의 계산을 할 수 없다.

$$\geq E_{q(z|x_i)}[\log P(x_i|z)] - D_{KL}(q(z|x_i)||P(z))$$

부분 해석 - KL Divergence는 항상 0 이상의 값을 갖기 때문

해석 - step1과 step2의 식을 조합하여 표현하면, 잠재변수 z에 의한 샘플 x_i가 나타날 확률의 log 기대값과 KL Divergence로 표현이 가능하다. 이때, KL Divergence의 값은 항상 0 이상의 값을 갖는 성질을 이용하면 목적함수의 하한 경계를 나타낼 수 있다.

ELBO: Evidence Lower Bound - step4

Step 4. EBLO의 구현

$\log P(x_i)$ 가 최대가 되는 매개변수를 탐색!! : x_i의 분포를 가장 잘 표현하자!!

$$\boxed{\log P(x_i) \geq \underbrace{E_{q(z|x_i)}[\log P(x_i|z)]}_{\text{Reconstruction Error}} - \underbrace{D_{KL}(q(z|x_i)||P(z))}_{\text{Regularization Parameter}}}$$

$$z \sim N(0,1)$$
$$q(z|x_i) \sim N(\mu_{q(x_i)}, \sigma_{q(x_i)})$$

를 따르기 때문에 (둘 다 Gaussian Distribution을 따르기 때문에)

$$D_{KL}(q(z|x_i) \;||\; P(z)) = D_{KL}(N(\mu_{q(x_i)}, \sigma_{q(x_i)}{}^2) \;||\; N(0,1))$$

$$= \sum_i -\log \sigma_{q(x_i)} + \frac{1}{2}(\sigma^2_{q(x_i)} + \mu^2_{q(x_i)} - 1)$$

KL Divergence를
간략히 표현할 수 있다.

KL Divergence between two Gaussian distribution.

$$\because \; D_{KL}(N(\mu_1, \sigma_1)||N(\mu_2, \sigma_2)) = \log\frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} - \frac{1}{2}$$

두 가우시안 분포 간의 KL Divergence 는 다음과 같이 간략하게 표현이 가능하다.

ELBO: Evidence Lower Bound - step4

Step 4. EBLO의 구현

$$\log P(x_i) \geq \underbrace{E_{q(z|x_i)}[\log P(x_i|z)]}_{\text{Reconstruction Error}} - \underbrace{D_{KL}(q(z|x_i)||P(z))}_{\text{Regularization Parameter}}$$

주어진 x_i로부터 latent space에서 Sampling된 z가 있을 때, z로부터 x_i가 발생할 확률이 Likelihood 이다.
x_i —(encode)—> z —(decode)—> x_i^

z가 이미 x로부터 확률적으로 생성되었고, 다시 z로부터 x가 나타날 확률을 계산.
즉, x가 나타날 확률을 계산하는 것은 우리가 알고있는(가지고 있는) 샘플들(데이터)에 대한 확률 분포를 찾고자 하는 것
따라서, z로부터 x가 나타날 확률을 최대화 하는 것 : Maximum Likelihood Estimation (최우추정법)

MLE —> Cross-Entropy로 나타내면

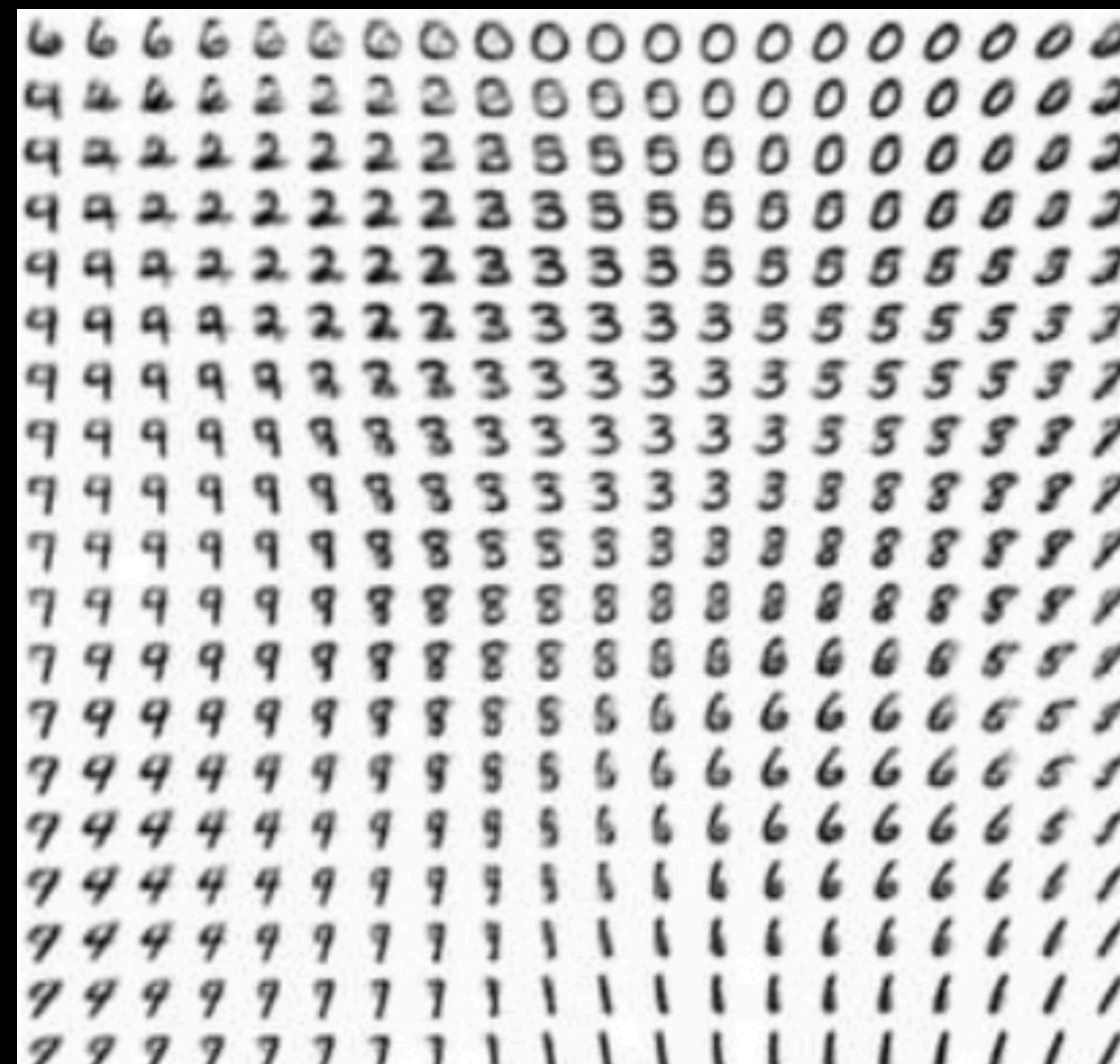
$$E_{q(z|x_i)}[\log P(x_i|z)] = \sum_i x_i \log Decode(Encode(x_i)) + (1-x_i)\log(1-Decode(Encode(x_i)))$$

Cross-Entropy

$$cross\text{-}entropy \rightarrow \sum_i x_i \log \hat{x_i} + (1-x_i)\log(1-\hat{x_i})$$

VAE Result: latent space에 대응하는 이미지 생성

다양한 표정의 이미지 생성



숫자 이미지 생성

Thank you.