# Deep generative models and inference

Data Mining & Quality Analytics Lab

Min-Jung Lee

# Contents

- Generative model
- Latent variable
- Inference
- Approximate inference
- Inference from sampling
- Inference as optimization
- Deep generative model

# Generative model

- DMQA 고등학교, 시험기간
- 수학시험범위 : 함수

민정

함수 단원에 대한 문제 n개를 풀기

현구

함수 단원을 이해하고
직접 출제 예상문제를 만들어 봄

# Generative model

- DMQA 고등학교, 시험기간
- 수학시험범위 : 함수

100점

민정

함수 단원에 대한 문제 n개를 풀기

현구

함수 단원을 이해하고
직접 출제 예상문제를 만들어 봄

**함수 단원을 더 잘 이해하고 있는 사람은 누구인가?**

# Generative model

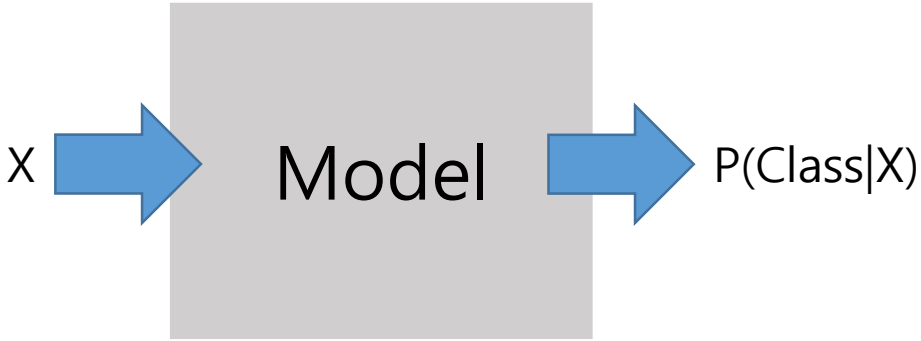
Bayes rule

$$P(Class|X) = \frac{P(X|Class)P(Class)}{P(X)}$$

- Supervised Learning, Classification, training data : (X, Class)
- Inference, Decision

Discriminative approach

P(Class|X)



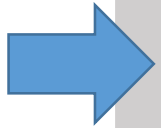

Generative approach

P(X|Class)P(Class)→ P(Class|X)

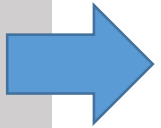
# Generative model

- Generative model

a training data set X



Generative Model



Estimate $P(X)$

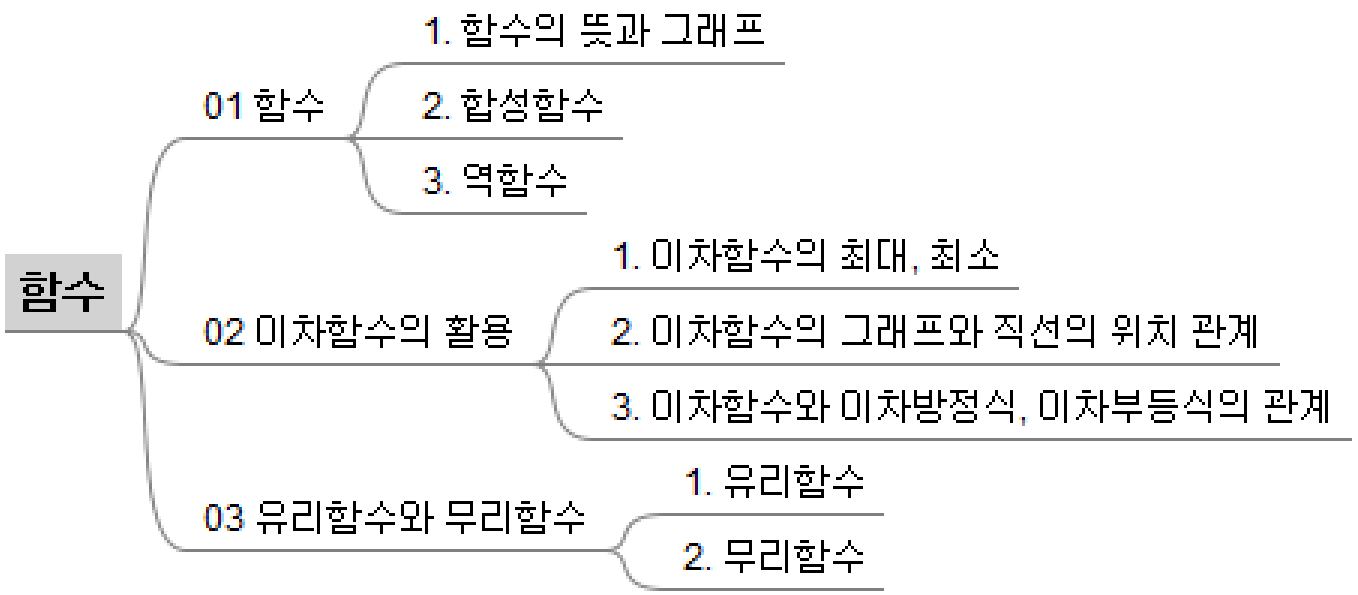
Generate samples $\sim P(X)$

# Latent variable

- Latent variable : 관측된 random variables이 아닌, 우리가 임의로 설정한 hidden variables(hidden causes/unobserved random variable)
- Latent variable 사용한 generative model이 좋은 점 : sampling이 쉬워짐

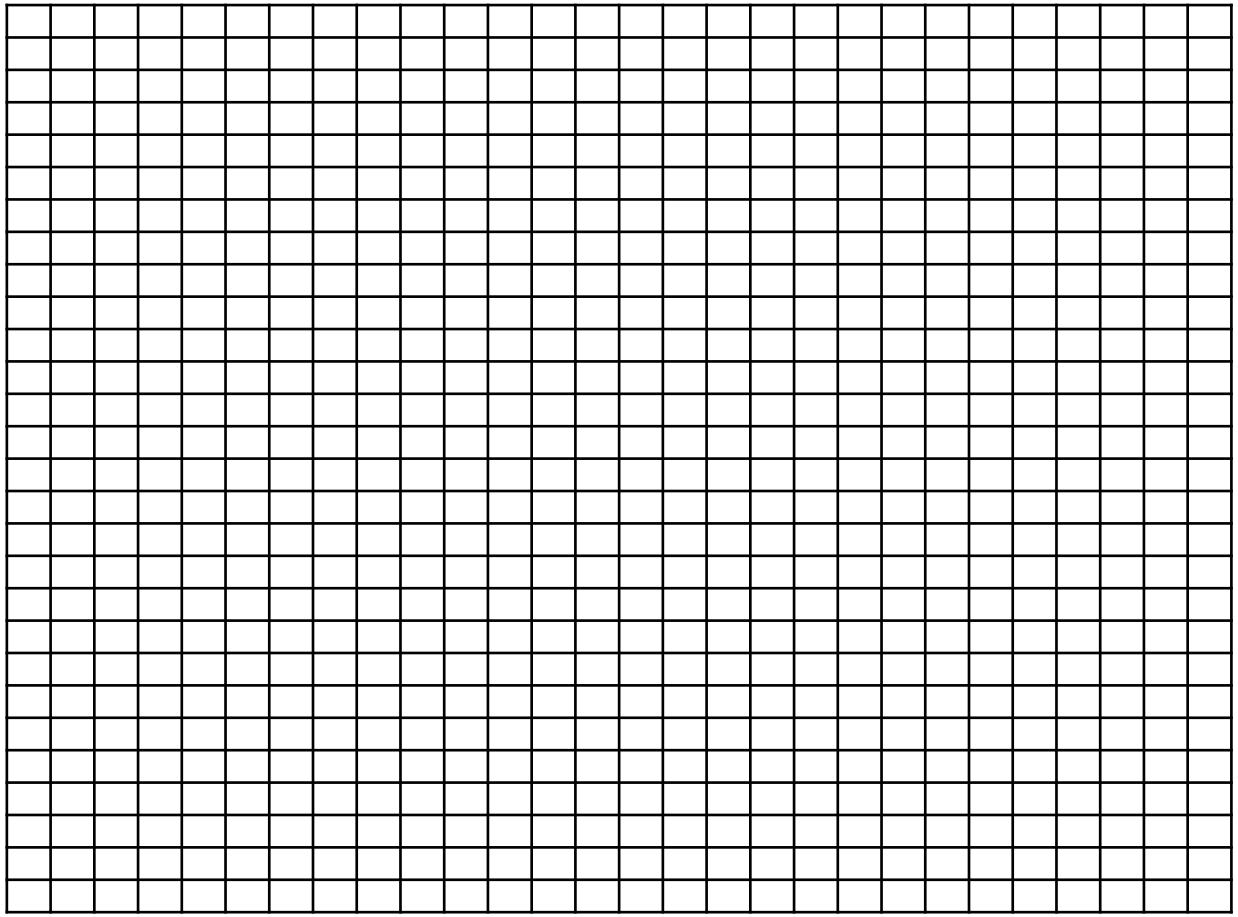
현구

함수 단원을 이해하고
직접 출제 예상문제를 만들어 봄



# Latent variable

- Latent variable 사용한 generative model이 좋은 점 : sampling이 쉬워짐
- The problem of generating images of handwritten characters





Doersch, C. (2016). Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908*.

# Latent variable

- Latent variable 사용한 generative model이 좋은 점 : sampling이 쉬워짐

Intuitively, it helps if **the model first decides which character to generate before it assigns a value to any specific pixel.**

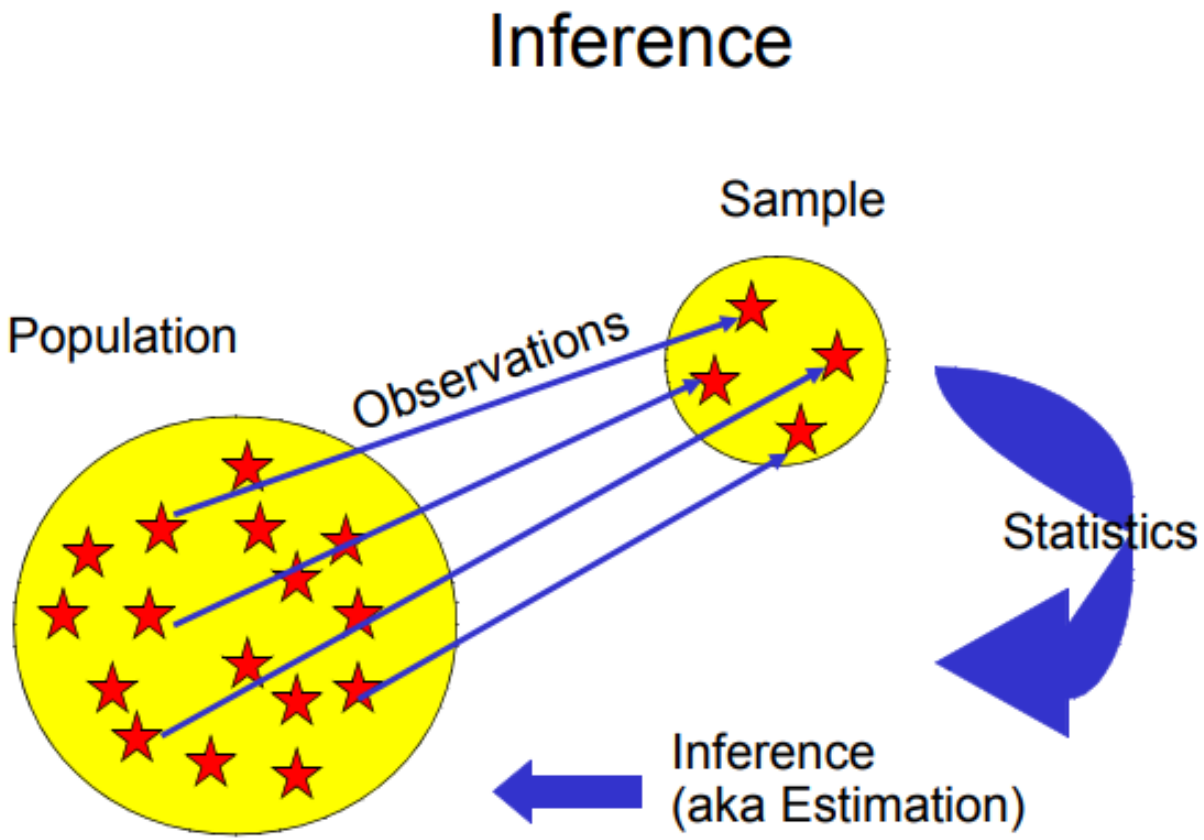
This kind of decision is formally called a **latent variable**.



Doersch, C. (2016). Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908*.

# Inference

- Statistical inference




https://www.ohio.edu/plantbio/staff/mccarthy/quantmet/lectures/lec5.pdf

# Inference

- Inference in machine learning : reason about (and compute) unknown probability distributions
- For most probabilistic models of practical interest, exact inference is intractable.

Bayesian Inference

$$P(Z|X) = \frac{P(X|Z)P(Z)}{P(X)} = \frac{P(X|Z)P(Z)}{\int P(x|z)P(z)dz}$$

$$P(X|Z) : Likelihood\ function\ of\ z$$
$$P(Z): Prior\ probability\ of\ z$$
$$P(Z|X): Posterior\ probability\ of\ z$$

https://www.ohio.edu/plantbio/staff/mccarthy/quantmet/lectures/lec5.pdf

# Inference

- Inference in machine learning : reason about (and compute) unknown probability distributions
- For most probabilistic models of practical interest, exact inference is intractable.
- So, we have to resort to some form of approximation

Generative model with latent variables +Inference

$$P(X) = \int P(x|z)P(z)dz$$

Probabilistic model + Bayesian Inference

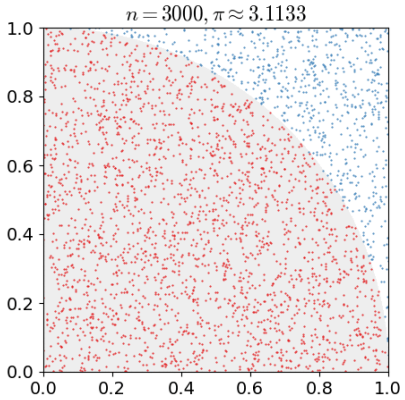
$$P(Z|X) = \frac{P(X|Z)P(Z)}{P(X)} = \frac{P(X|Z)P(Z)}{\int P(x|z)P(z)dz}$$

$$P(X|Z) : Likelihood\ function\ of\ z$$
$$P(Z): Prior\ probability\ of\ z$$
$$P(Z|X): Posterior\ probability\ of\ z$$

https://www.ohio.edu/plantbio/staff/mccarthy/quantmet/lectures/lec5.pdf

# Approximate Inference

## Stochastic approximations
### Approximate Inference from **sampling**




- Rejection sampling, Importance sampling
- Markov chain Monte Carlo (MCMC) : Gibbs sampling, Metropolis-Hastings

## Deterministic approximations
### Approximate Inference as **optimization**

$$KL[q(z)||p(z|x)]$$

$$\log P(X)$$

$$ELBO = \int \log\left(P(x|z)\frac{P(z)}{q(z)}\right)q(z)dz$$

- Variational inference

# Inference from sampling

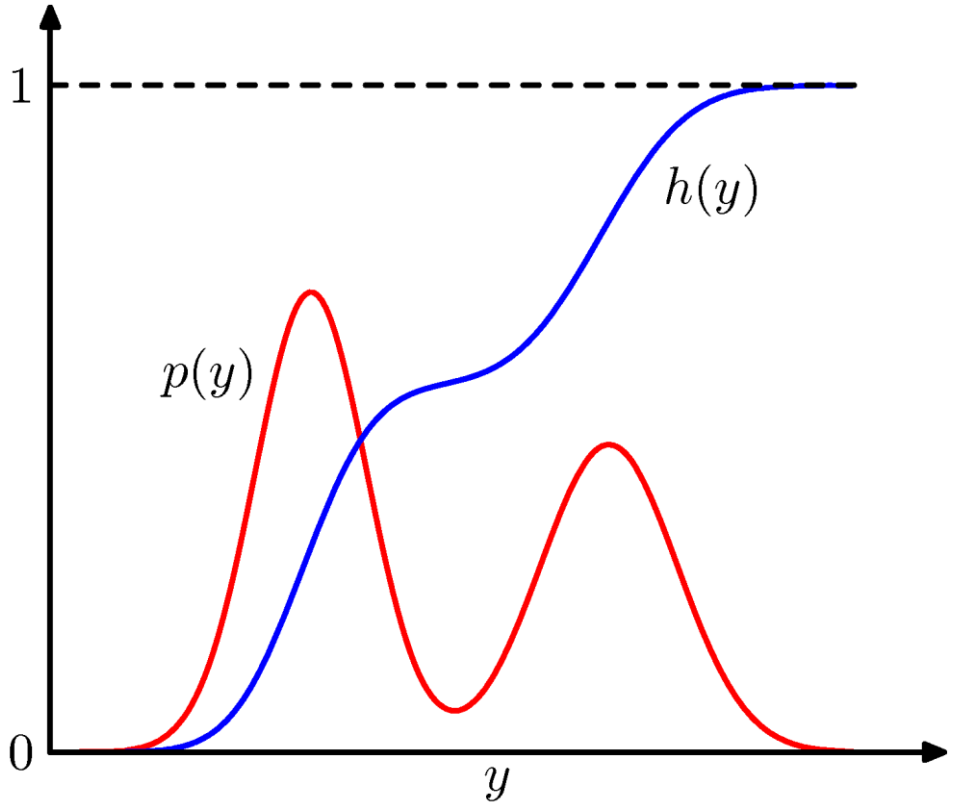
$$\underbrace{p(z|y)}_{\text{Posterior}} = \frac{\overbrace{p(y|z)}^{\text{Likelihood}}\ \overbrace{p(z)}^{\text{Prior}}}{\underbrace{\int p(y,z)dz}_{\text{Marginal likelihood/ Model evidence}}}$$

Most inference problems will be one of:

**Marginalisation** $$p(y) = \int p(y,\theta)d\theta$$

**Expectation** $$\mathbb{E}[f(y)|x] = \int f(y)p(y|x)dy$$

**Prediction** $$p(y_{t+1}) = \int p(y_{t+1}|y_t)p(y_t)dy_t$$



# Inference from sampling

- Importance sampling : Transform the **integral** into **an expectation over a simple, known distribution**

Integral problem

$$P(X) = \int P(x|z)P(z)dz$$

Proposal : easy to sample

$$P(X) = \int P(x|z)\frac{P(z)}{q(z)}q(z)dz$$

$$q(z) : proposal\ distribution$$

$$z^{(s)} \sim q(z), \qquad W^{(s)} = \frac{P(z)}{q(z)}$$

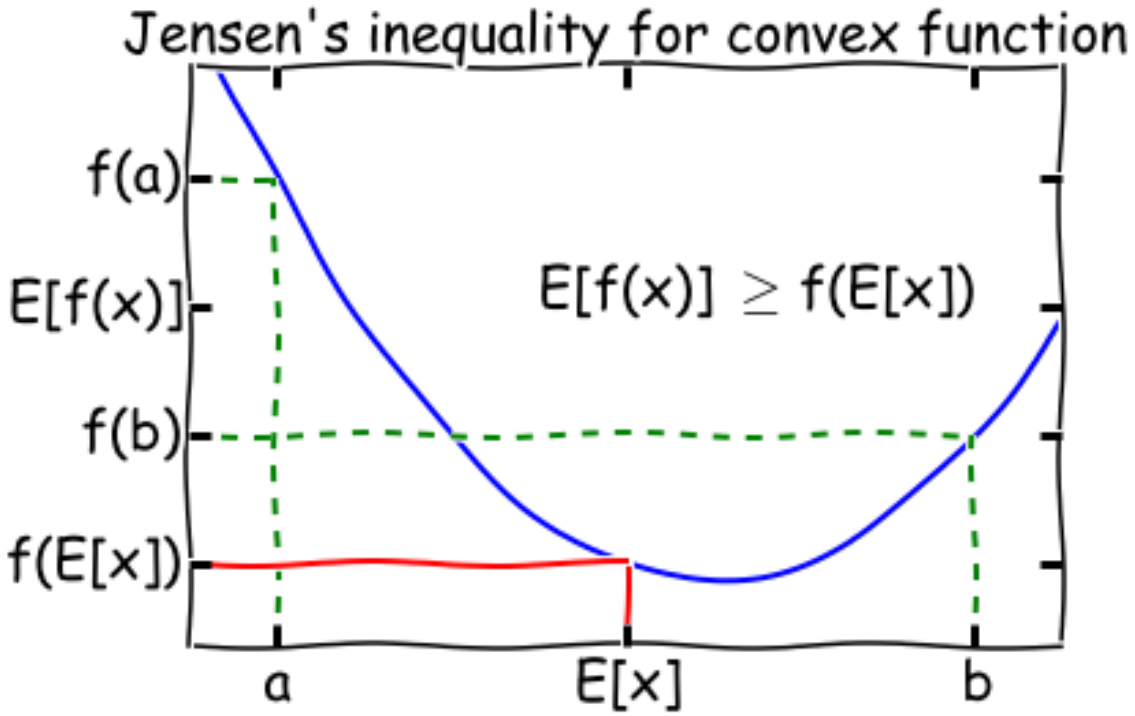
Importance Weight

Monte Carlo integration

$$P(X) = \frac{1}{S}\sum_{s} W^{(s)} P(X|\, z^{(s)})$$

# Inference as optimization

- Variational Inference
- Monte Carlo Integration 대신 Jensen's inequality 사용
- Monte Carlo Integration : 어려운 적분 문제들을 푸는데 도움을 줄 수 있지만 계산적으로 부담이 많이 되기때문



# Inference as optimization

- Variational Inference
- Monte Carlo Integration 대신 Jensen's inequality 사용

$$P(X) = \int P(x|z)\frac{P(z)}{q(z)}q(z)dz$$

$$\log P(X) = log\int P(x|z)\frac{P(z)}{q(z)}q(z)dz$$

$$= \log\left(E_{q(z)}\left[P(x|z)\frac{P(z)}{q(z)}\right]\right)$$

$$\log\left(E_{q(z)}\left[P(x|z)\frac{P(z)}{q(z)}\right]\right) \geq E_{q(z)}\left[\log\left(P(x|z)\frac{P(z)}{q(z)}\right)\right] = \int \log\left(P(x|z)\frac{P(z)}{q(z)}\right)q(z)dz$$

# Inference as optimization

- Variational Inference
- Monte Carlo Integration 대신 Jensen's inequality 사용

$$\log P(X) \geq \int \log\left(P(x|z)\frac{P(z)}{q(z)}\right)q(z)dz$$

$$= \int q(z)\log\big(P(x|z)\big)\,dz - \int q(z)\log\left(\frac{q(z)}{P(z)}\right)dz$$

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

**Lower Bound : L**

# Inference as optimization

- Variational Inference

$$\log P(X) - L(q)$$

$$= \log P(X) - \int \log\left(P(x|z)\frac{P(z)}{q(z)}\right) q(z)dz$$

$$= \log P(X) - \int \log\left(\frac{P(x,z)}{q(z)}\right) q(z)dz$$

$$= \log P(X) - \int \log\left(P(z|x)\frac{P(x)}{q(z)}\right) q(z)dz$$

$$= \log P(X) - \int \log\left(\frac{P(z|x)}{q(z)}\right) q(z)dz - \int \log\big(P(x)\big)\, q(z)dz$$

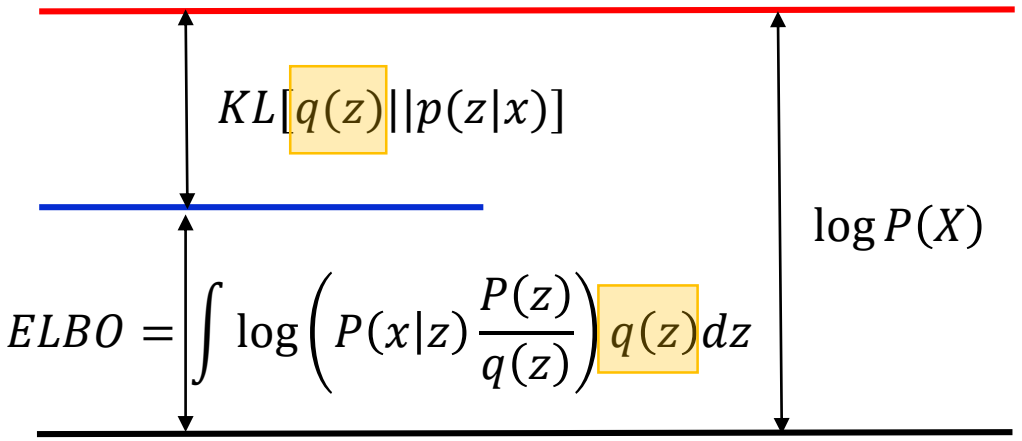
$$= -\int \log\left(\frac{P(z|x)}{q(z)}\right) q(z)dz = KL[q(z)||p(z|x)]$$

# Inference as optimization

- Variational Inference

$$\log P(X) = L(q) + KL[q(z)||p(z|x)]$$

$$KL[q(z)||p(z|x)] = 0\,, if\ q(z) = p(z|x)$$
$$KL[q(z)||p(z|x)] > 0\,, if\ q(z) \neq p(z|x)$$



# Inference as optimization

- Variational Inference

$$L(q) = \int \log\left(P(x|z)\frac{P(z)}{q(z)}\right) q(z)dz$$

$= q$에 *대한 범함수 (functional)*

**Functions:**

- Variables as input, output is a value.
- Full and partial derivatives $\frac{df}{dx}$
- E.g., Maximise likelihood p(x|$\theta$) w.r.t. parameters $\theta$

**Functionals:**

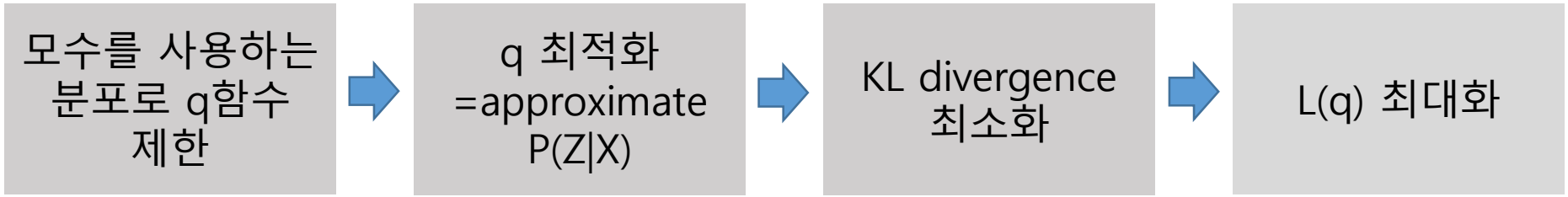
- Functions as input, output is a value.
- Functional derivatives $\frac{\delta F}{\delta f}$
- E.g., Maximise the entropy H[*p(x)*] w.r.t. *p(x)*

**Calculus of Variations**

# Inference as optimization

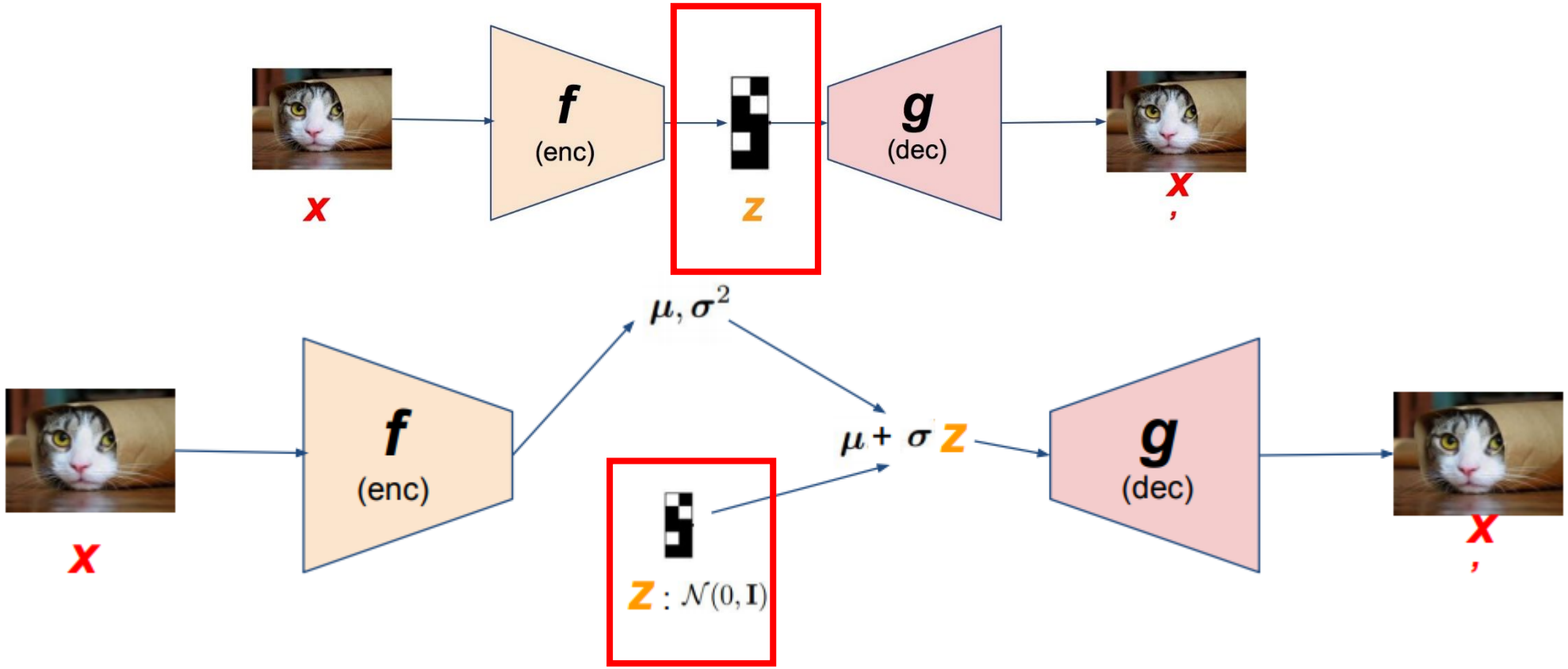
- Variational Inference

Maximize Likelihood $= argmax_\theta \log P_\theta(X) \geq L(q) = q$에 대한 *functional*



# Variational Autoencoders (VAE)

- Deep generative model, DNN=함수근사기
- Autoencoder 구조 사용
- Z 다른점 : Z: random variable, Z 를 정규분포로 가정

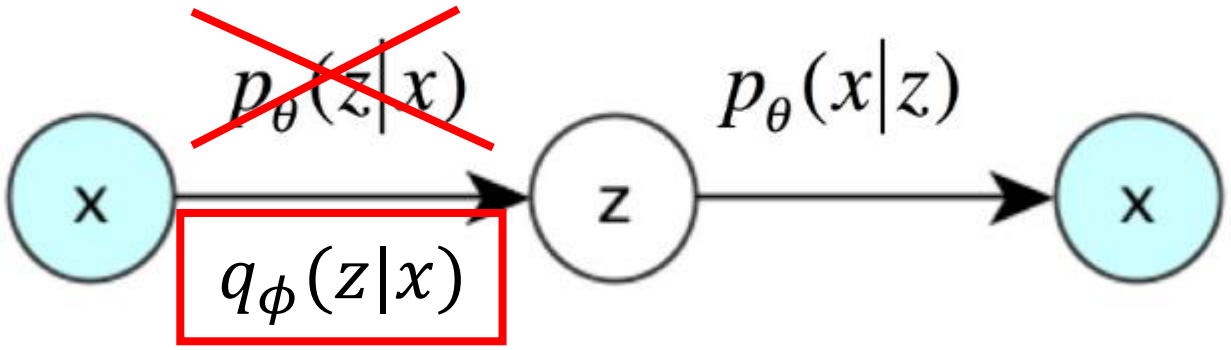


# Variational Autoencoders (VAE)

- Loss function 다름
- Variational Inference 이용

Maximize Likelihood $= argmax_\theta \log P_\theta(X) \geq L(q)$

모수를 사용하는 분포로 q함수 제한 → q 최적화 =approximate P(Z|X) → KL divergence 최소화 → L(q) 최대화



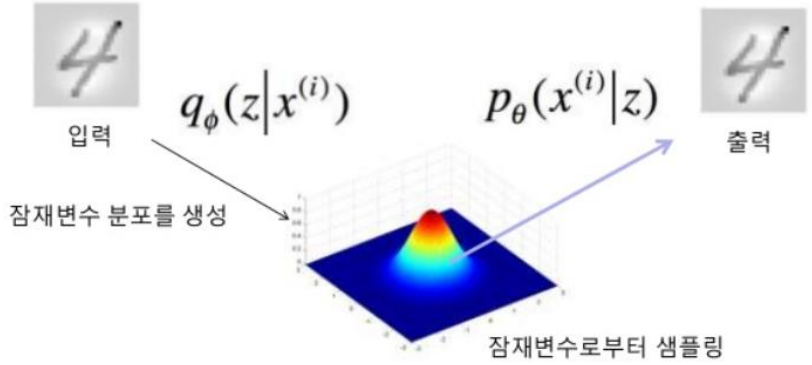
# Variational Autoencoders (VAE)

- Loss function 다름
- Variational Inference 이용

$$\text{Maximize Likelihood} = argmax_\theta \log P_\theta(X) \geq L(q)$$

모수를 사용하는 분포로 q함수 제한 → q 최적화 =approximate P(Z|X) → KL divergence 최소화 → L(q) 최대화

- 잠재변수를 다차원의 정규분포로 가정
- 입력 > 잠재변수 분포를 생성 > 잠재변수로부터 샘플링> 입력에 가까운 출력 생성



# Variational Autoencoders (VAE)

- Loss function 다름,
- Variational Inference

$$\text{Maximize Likelihood} = argmax_\theta \log P_\theta(X) \geq L(q)$$

$$= \int \log\left(P(x|z)\frac{P(z)}{q(z)}\right) q(z)dz$$

$$= \int q(z)\log\big(P(x|z)\big)\,dz - \int q(z)\log\left(\frac{q(z)}{P(z)}\right)dz$$

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

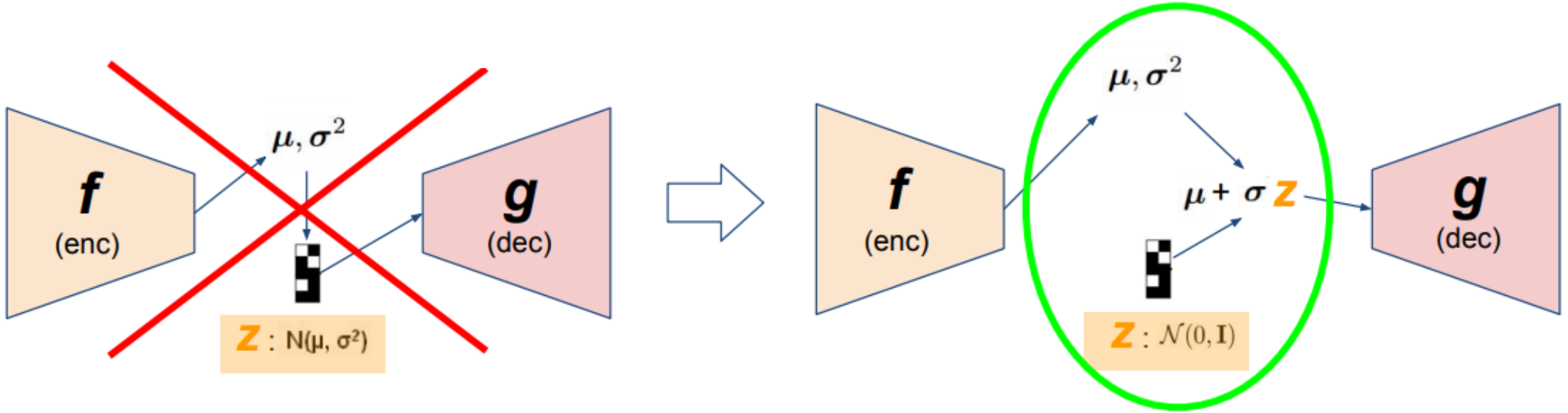
**Evidence lower bound (ELBO)**

**Reconstruction error**
**: 입력에 가까운 출력 생성**

**Proposal distribution**
**should be like Gaussian**

# Variational Autoencoders (VAE)

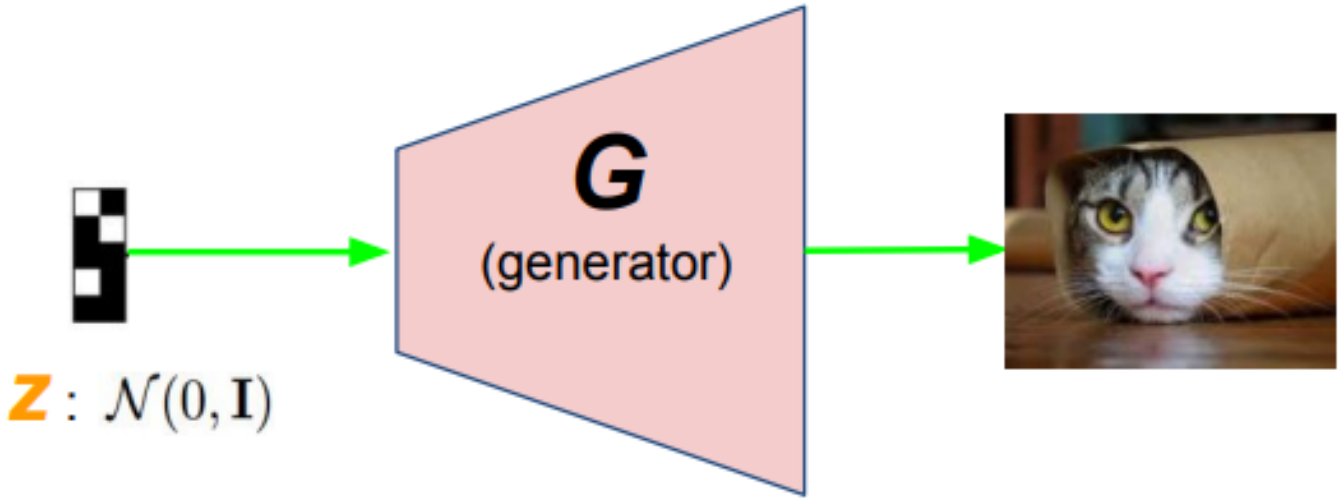
- Reparameterization trick (Enable back propagation)

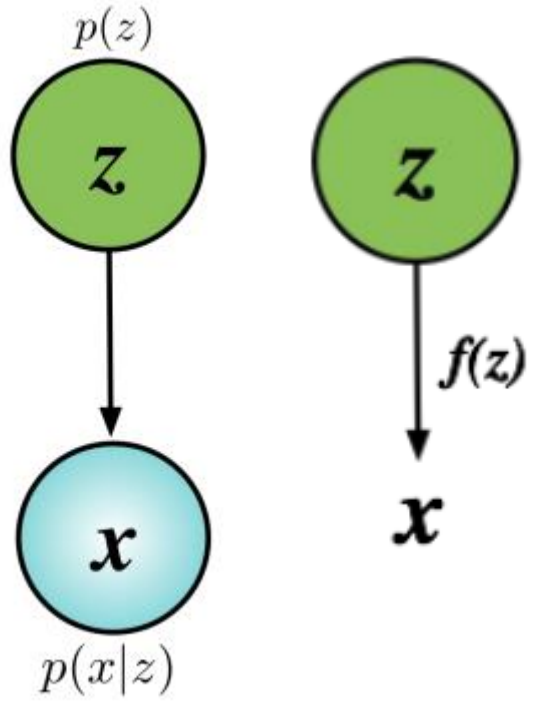


# GAN

- Deep generative model
- Latent variable model

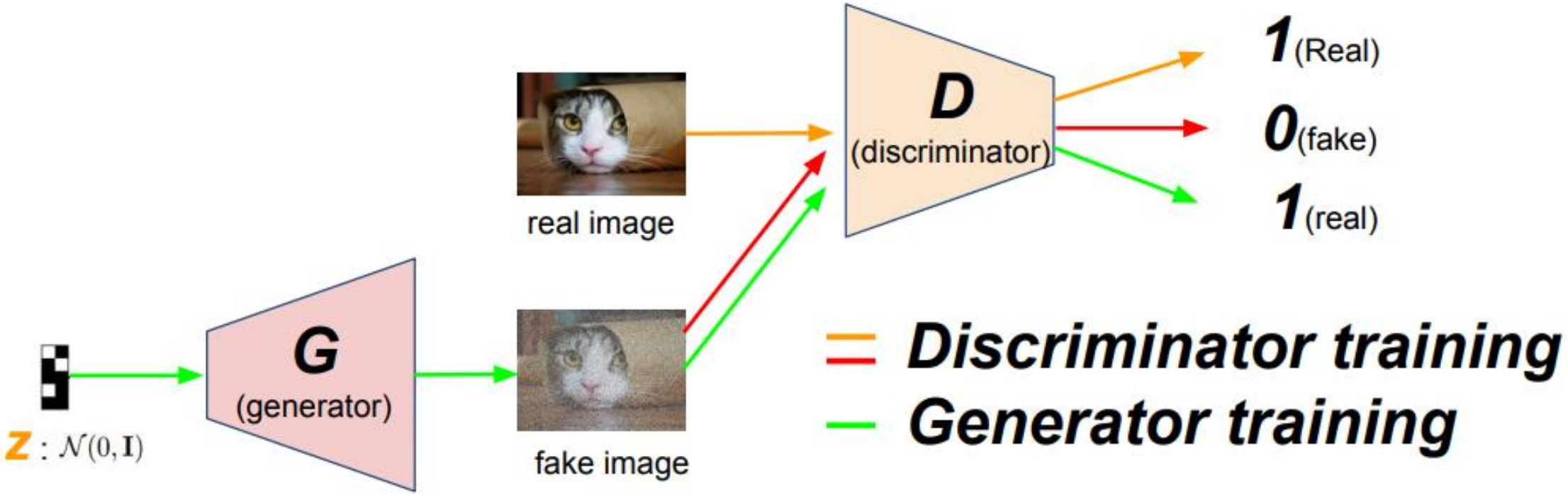
- VAE : Explicitly maximize likelihood
- GAN : Implicitly maximize likelihood



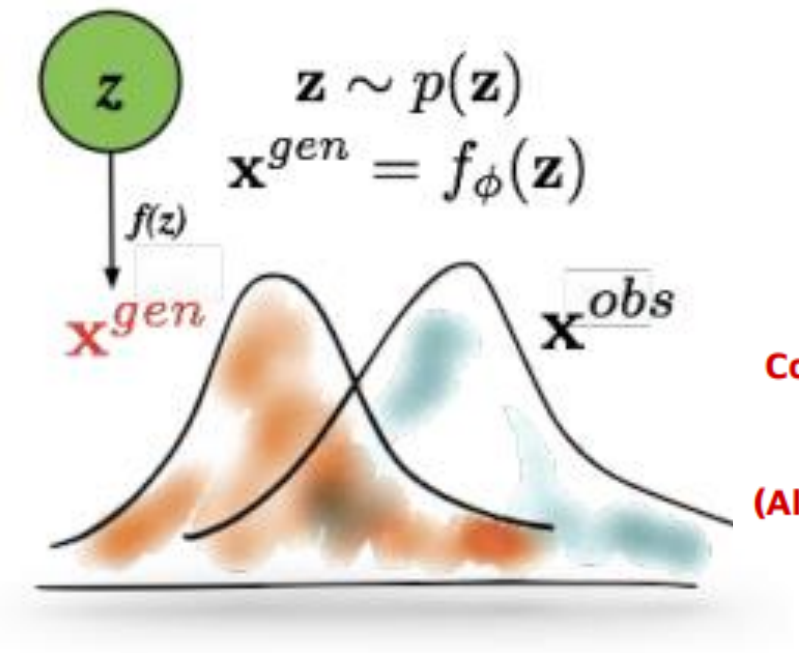
# GAN

- VAE : Inference, Generator
- GAN : Generator, Discriminator



# GAN

- Learning by comparison
- Comparing the estimated distribution q(x) to the true distribution p(x) using samples.




$$\mathcal{F}(\mathbf{x}, \theta, \phi) = \mathbb{E}_{p^*(x)}[\log D_\theta(\mathbf{x})] + \mathbb{E}_{q_\phi(x)}[\log(1 - D_\theta(\mathbf{x})]$$

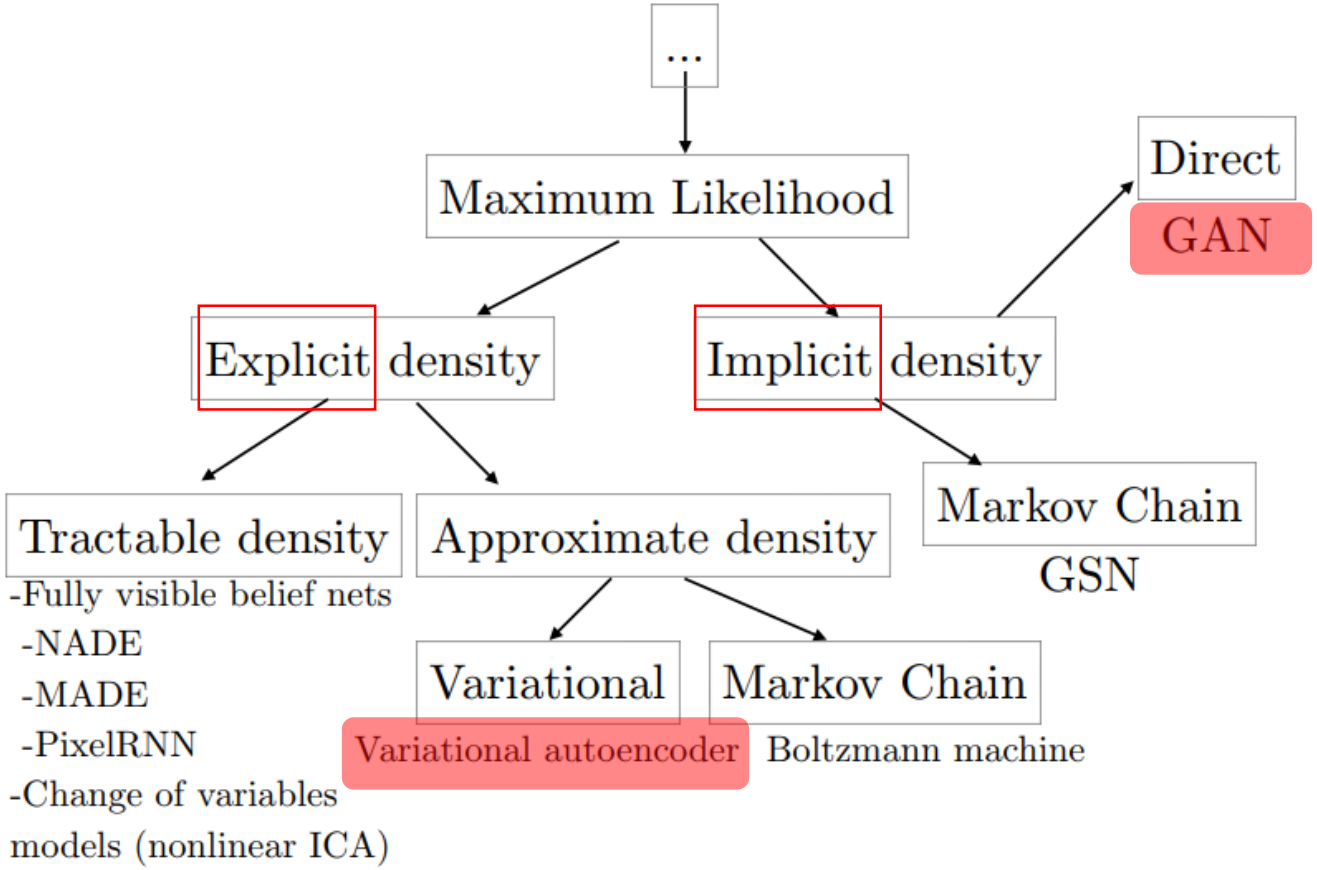
**Alternating optimisation** $\min_\phi \max_\theta \mathcal{F}(\mathbf{x}, \theta, \phi)$

**Comparison loss** $\theta \propto \nabla_\theta \mathbb{E}_{p^*(x)}[\log D_\theta(\mathbf{x})] + \nabla_\theta \mathbb{E}_{q_\phi(x)}[\log(1 - D_\theta(\mathbf{x})]$

**(Alt) Generative loss** $\phi \propto -\nabla_\phi \mathbb{E}_{q(z)}[\log D_\theta(f_\phi(\mathbf{z}))]$

# Deep generative models

- Maximum Likelihood
- Likelihood = the probability that the model assigns to the training data



# Conclusions

‘종구’는 ’일광’에게 왜 하필 자신의 딸이 이런 일을 당하는지를 물어봅니다.

‘자네는 낚시를 헐쩍에 뭣이 걸려 나올지 알고 허나? 그놈은 그냥 미끼를 던져분 것이고 자네 딸내미는 고것을 확 물어분 것이여’

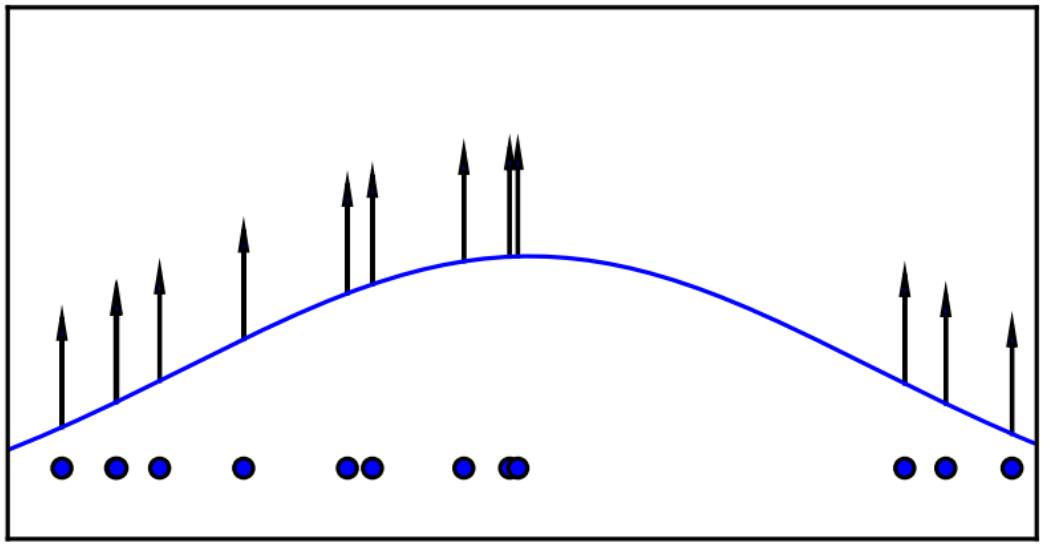
# Thank you

Data Mining & Quality Analytics Lab

Min-Jung Lee

# Appendix

- Maximum Likelihood
- Likelihood = the probability that the model assigns to the training data



$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} \mathbb{E}_{x \sim p_{\text{data}}} \log p_{\text{model}}(\boldsymbol{x} \mid \boldsymbol{\theta})$$

$$\theta^* = argmax_\theta \prod_{i=1}^{m} P_{model}(x^{(i)}; \theta)$$

$$= argmax_\theta \, log \prod_{i=1}^{m} P_{model}(x^{(i)}; \theta)$$

$$= argmax_\theta \sum_{i=1}^{m} log P_{model}(x^{(i)}; \theta)$$

$$= argmax_\theta \, E_{x \sim P_{data}} log P_{model}(x|\theta)$$

# Appendix

- Kullback-Leibler divergence
- 두 확률분포의 차이를 계산
- 어떤 이상적인 분포에 대해, 그 분포를 근사하는 다른 분포를 사용해 샘플링을 한다면 발생할 수 있는 정보 엔트로피 차이를 계산함
- 근사방법을 사용했을 때 '얼마나 많은 정보가 손실되었는가'
- 엔트로피는 '정보의 단위'

$$H = -\sum_{i=1}^{N} p(x_i) \cdot log\, p(x_i)$$

이산확률변수의 경우: $D_{\mathrm{KL}}(P\|Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$

연속확률변수의 경우: $D_{\mathrm{KL}}(P\|Q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} dx$

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