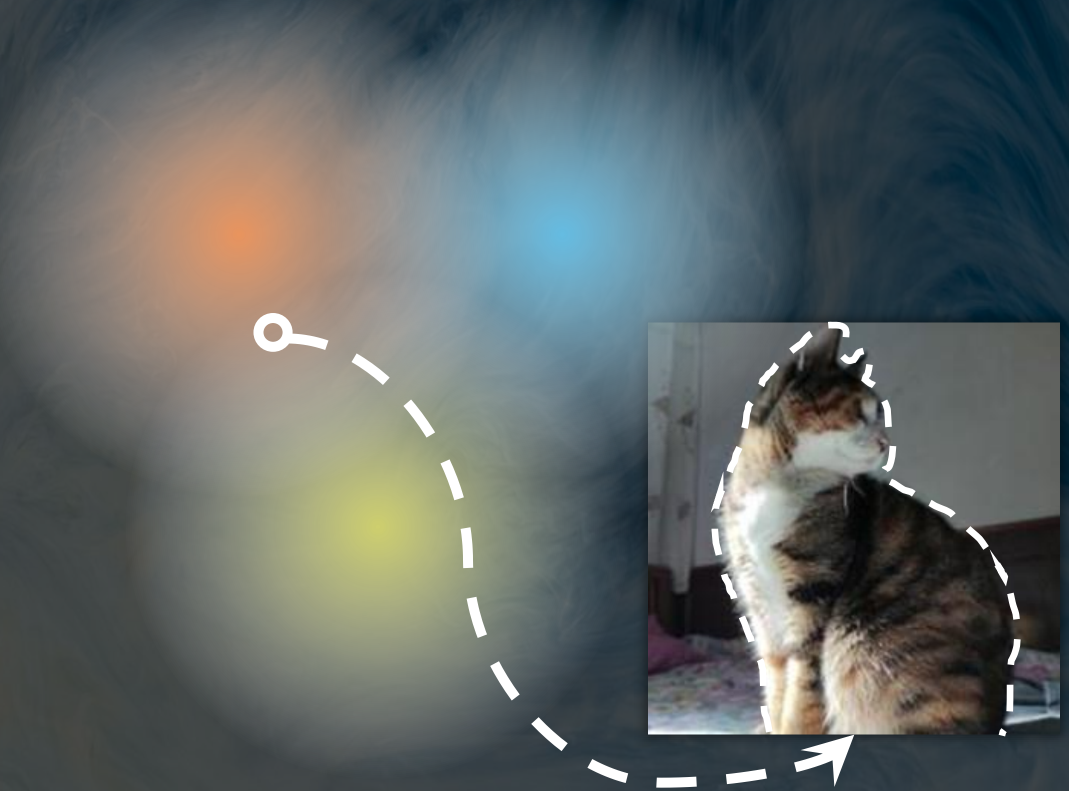
Variational Auto Encoder

原理图解

ELBO

公式推导



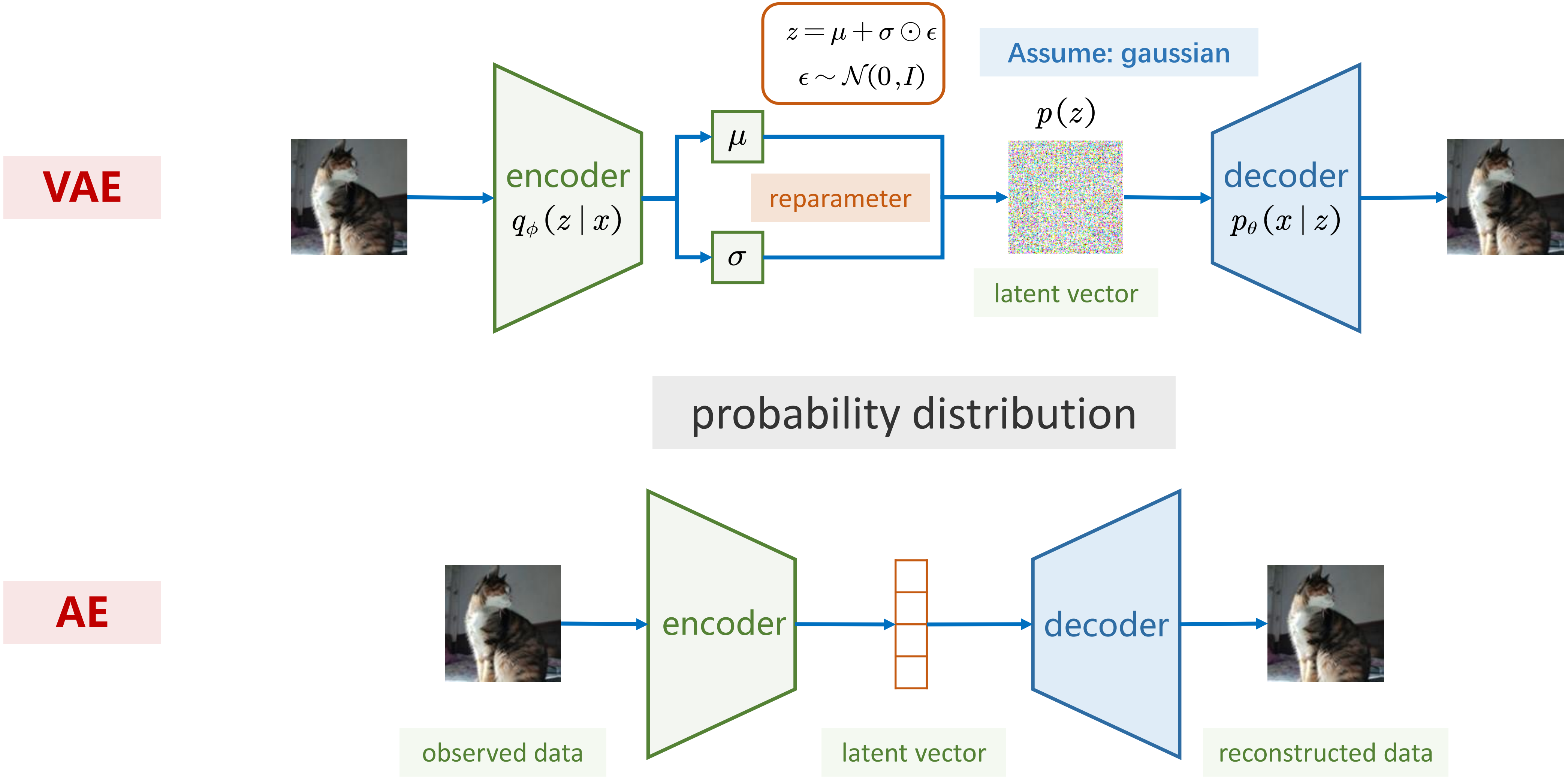
Variational Auto Encoder

intuition reason and relationship with diffusion

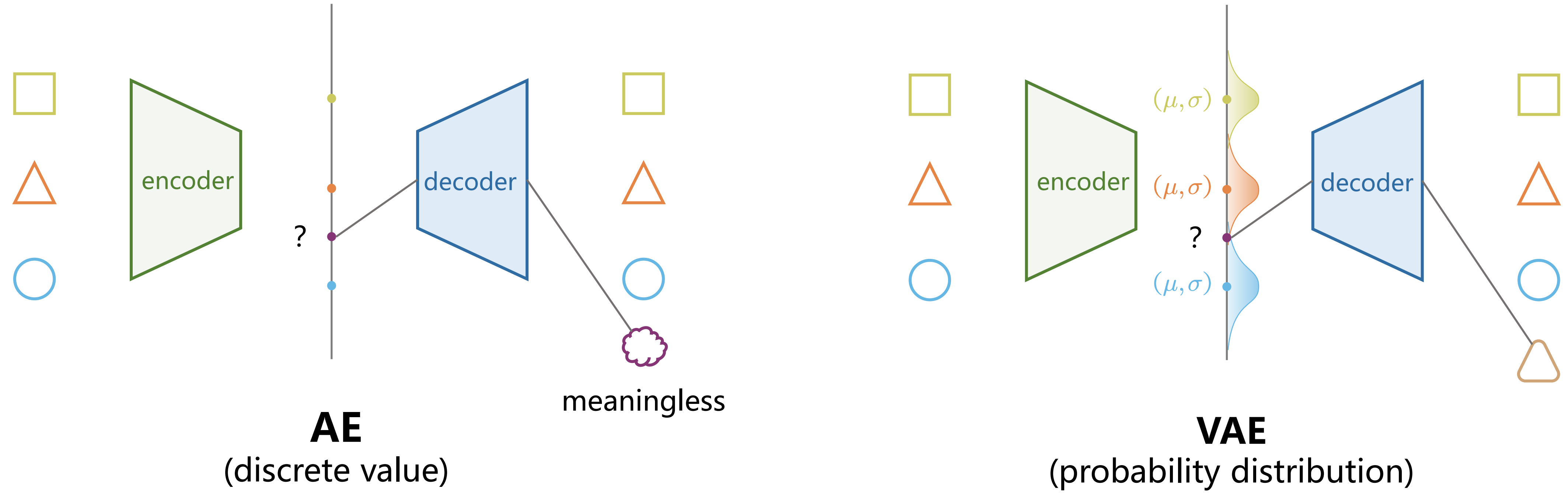
- What is the variational auto encoder?
- What is image distribution?
- What is variational inference and ELBO?
- What is the relationship between VAE and Diffusion?

Xin Zhang

What is Variational Auto Encoder



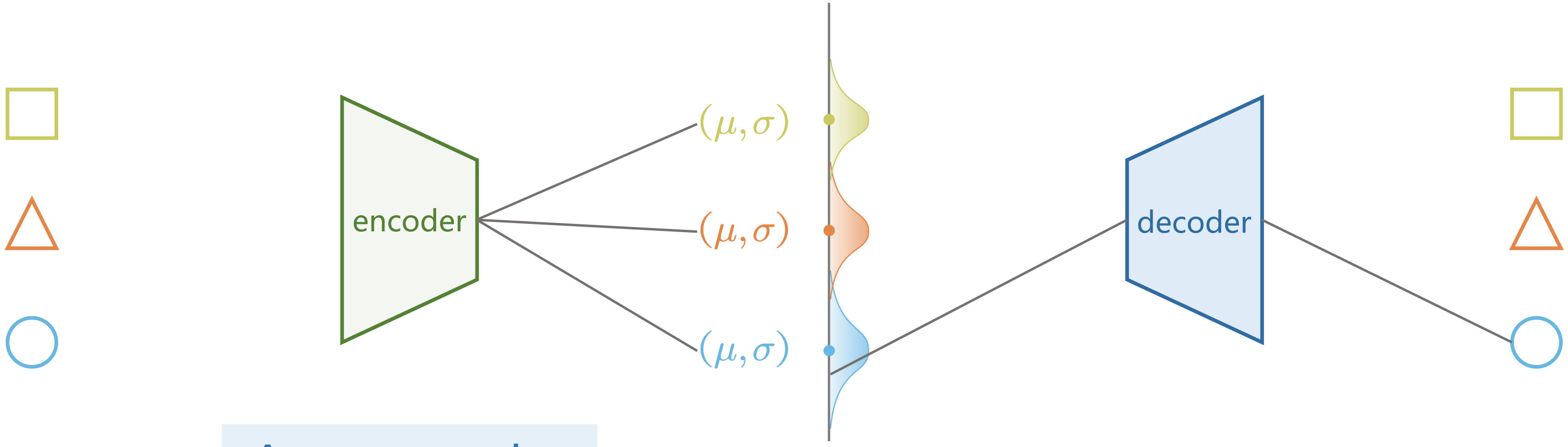
Intuition reason



> A variational autoencoder can be defined as being an autoencoder whose training is **regularized** to **avoid overfitting** and ensure that the latent space has **good properties** that enable generative process.

[1] Understanding Variational Autoencoders (VAEs), https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73

Intuition reason



Assume: gaussian

$$\boldsymbol{x} \longrightarrow p(z \mid \boldsymbol{x}) \longrightarrow z \sim p(z \mid \boldsymbol{x}) \longrightarrow p(\boldsymbol{x} \mid z) \longrightarrow \boldsymbol{x}$$

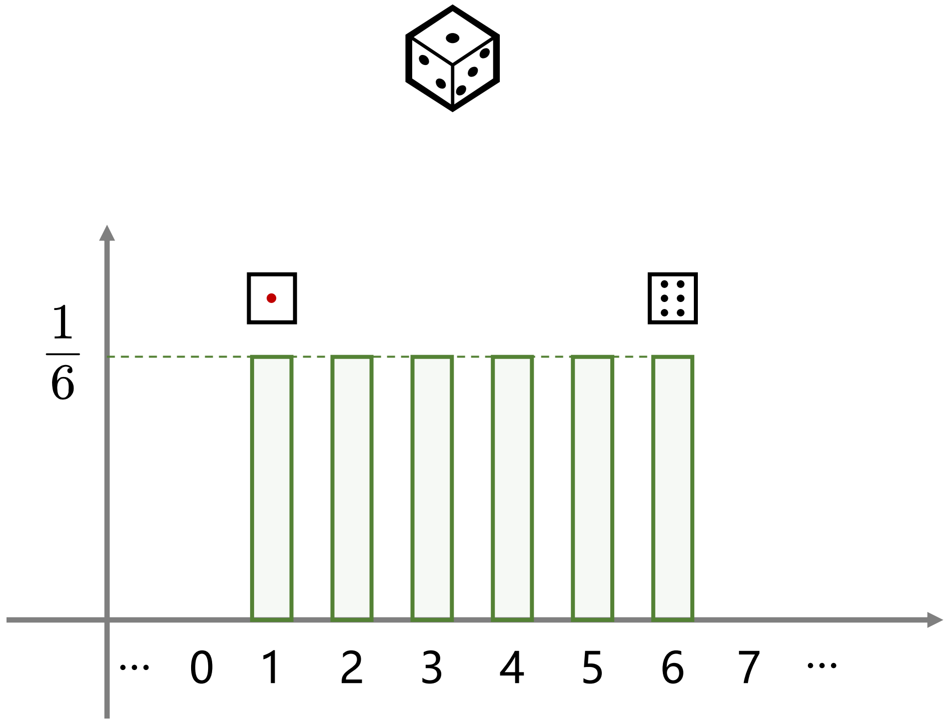
observed data | latent distribution | sampled latent representation | decoding process | reconstructed data

our target: $q_\phi(z \mid \boldsymbol{x})$

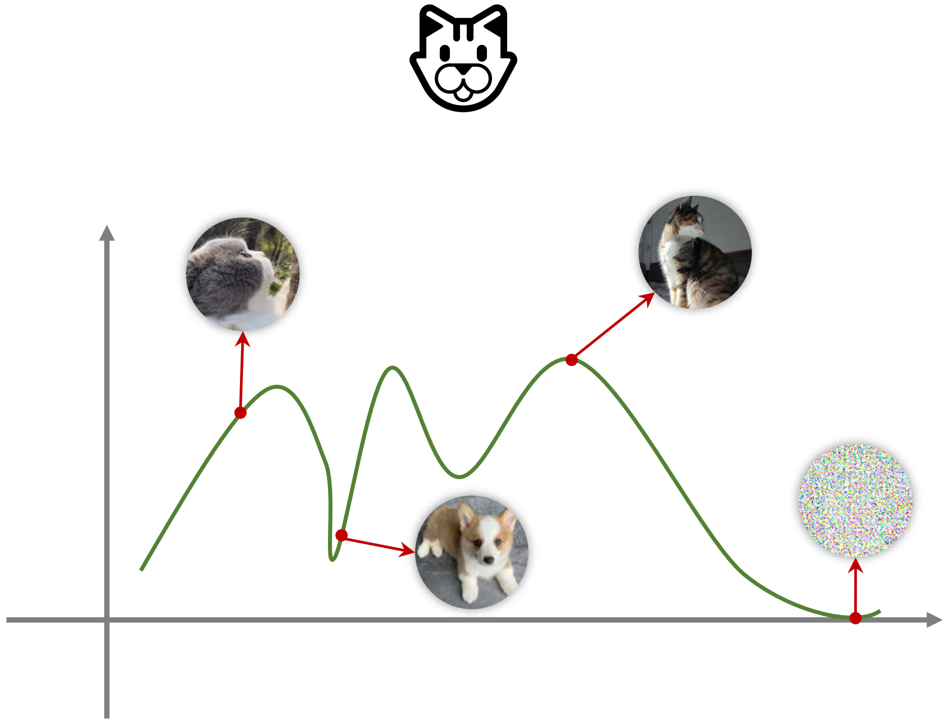
$$p(z \mid \boldsymbol{x}) = \frac{p(z, \boldsymbol{x})}{p(\boldsymbol{x})} = \frac{p(\boldsymbol{x} \mid z)\, p(z)}{\int_z p(\boldsymbol{x} \mid z)\, p(z)\, dz}$$

hard

Image distribution



$$P(x=1)=\frac{1}{6} \qquad P(x=0)=0$$



$$\operatorname{argmax}\ \log p_{\theta}(\boldsymbol{x})$$

Evidence Lower Bound

The view of the posterior

$$KL\left(q_\phi(z \mid \boldsymbol{x}) \| p(z \mid \boldsymbol{x})\right)$$

$$= \int_z q_\phi(z \mid \boldsymbol{x}) \log\left[\frac{q_\phi(z \mid \boldsymbol{x})}{p(z \mid \boldsymbol{x})}\right] dz$$

definition of KL Divergence

$$= \int_z q_\phi(z \mid \boldsymbol{x}) \log\left[\frac{q_\phi(z \mid \boldsymbol{x})\, p(\boldsymbol{x})}{p(z, \boldsymbol{x})}\right] dz$$

chain rule of probability

$$= \int_z q_\phi(z \mid \boldsymbol{x}) \log\left[\frac{q_\phi(z \mid \boldsymbol{x})}{p(z, \boldsymbol{x})}\right] dz + \int_z q_\phi(z \mid \boldsymbol{x}) \log p(\boldsymbol{x})\, dz$$

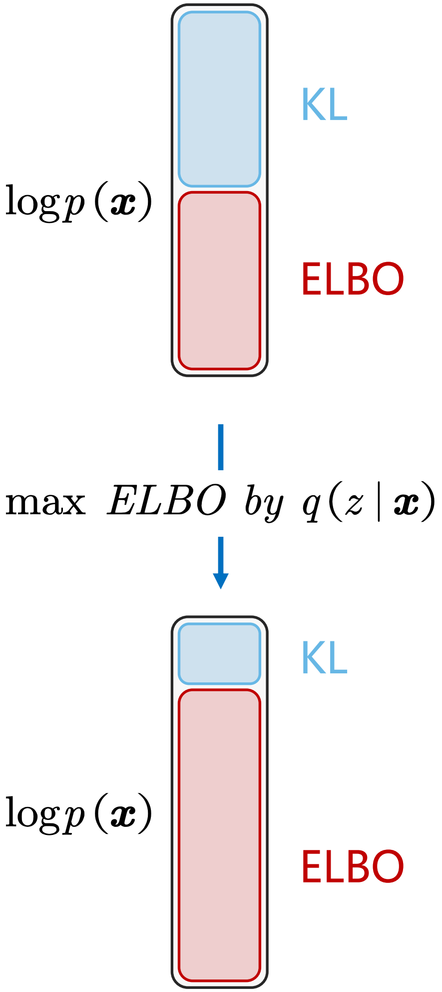
split the integral equation

$$= \mathbb{E}_{q_\phi(z \mid \boldsymbol{x})} \log\left[\frac{q_\phi(z \mid \boldsymbol{x})}{p(z, \boldsymbol{x})}\right] + \log p(\boldsymbol{x})$$

definition of Expectation

$$KL\left(q_\phi(z \mid x) \| p(z \mid x)\right) + \mathbb{E}_{q_\phi(z \mid \boldsymbol{x})} \log\left[\frac{p(z, \boldsymbol{x})}{q_\phi(z \mid \boldsymbol{x})}\right] = \log p(\boldsymbol{x})$$

⑥ ELBO



Evidence Lower Bound

⑦ VAE optimization

$$\mathbb{E}_{q_\phi(z|\boldsymbol{x})}\log\left[\frac{p(z,\boldsymbol{x})}{q_\phi(z\mid\boldsymbol{x})}\right]$$

$$=\mathbb{E}_{q_\phi(z|\boldsymbol{x})}\left[\log\frac{p_\theta(\boldsymbol{x}\mid z)\,p(z)}{q_\phi(z\mid\boldsymbol{x})}\right]$$

Chain Rule of Probability

$$=\mathbb{E}_{q_\phi(z\mid\boldsymbol{x})}[\log p_\theta(\boldsymbol{x}\mid z)]+\mathbb{E}_{q_\phi(z\mid\boldsymbol{x})}\left[\log\frac{p(z)}{q_\phi(z\mid\boldsymbol{x})}\right]$$

Split the Expectation

$$=\mathbb{E}_{q_\phi(z\mid\boldsymbol{x})}[\log p_\theta(\boldsymbol{x}\mid z)]-KL(q_\phi(z\mid\boldsymbol{x})||p(z))$$

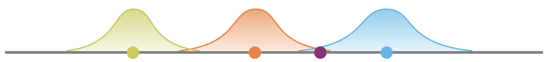
Definition of KL Divergence

reconstruction term

prior matching term

$$||\boldsymbol{x}-\hat{\boldsymbol{x}}||^2 \qquad \frac{1}{2}(\mu_\phi^2+\sigma_\phi^2-\log\sigma_\phi^2-1)$$

Assume: gaussian



avoid noise to be zero.

$$p(z\mid\boldsymbol{x})\to\mathcal{N}(0,\boldsymbol{I})$$

$$p(z)=\int_{\boldsymbol{x}}p(z\mid\boldsymbol{x})\,p(\boldsymbol{x})\,d\boldsymbol{x}$$

$$=\int_{\boldsymbol{x}}\mathcal{N}(0,\boldsymbol{I})\,p(\boldsymbol{x})\,d\boldsymbol{x}$$

$$=\mathcal{N}(0,\boldsymbol{I})\int_{\boldsymbol{x}}p(\boldsymbol{x})\,d\boldsymbol{x}$$

$$=\mathcal{N}(0,\boldsymbol{I})$$

$$q_\phi(z\mid\boldsymbol{x})=\mathcal{N}(z;\mu_\phi(\boldsymbol{x}),\sigma_\phi^2(\boldsymbol{x})\boldsymbol{I})$$

Evidence Lower Bound

The view of the likelihood

maximize the likelihood of observed x

$$\log p(\boldsymbol{x}) = p(\boldsymbol{x}) \int_z q_\phi(z \mid \boldsymbol{x}) dz = \int_z q_\phi(z \mid \boldsymbol{x}) \log p(\boldsymbol{x}) dz$$

multiply by $1 = \int_z q_\phi(z \mid \boldsymbol{x}) dz$

$$= \int_z q_\phi(z \mid \boldsymbol{x}) \log \left[\frac{p(z, \boldsymbol{x})}{p(z \mid \boldsymbol{x})} \right] dz$$

chain rule of probability

$$= \int_z q_\phi(z \mid \boldsymbol{x}) \log \left[\frac{p(z, \boldsymbol{x}) q_\phi(z \mid \boldsymbol{x})}{q_\phi(z \mid \boldsymbol{x}) p(z|\boldsymbol{x})} \right] dz$$

multiply by $1 = \frac{q_\phi(z \mid \boldsymbol{x})}{q_\phi(z \mid \boldsymbol{x})}$

$$= \int_z q_\phi(z \mid \boldsymbol{x}) \log \left[\frac{p(z, \boldsymbol{x})}{q_\phi(z \mid \boldsymbol{x})} \right] dz + KL(q_\phi(z \mid \boldsymbol{x}) || p(z \mid \boldsymbol{x}))$$

definition of KL Divergence

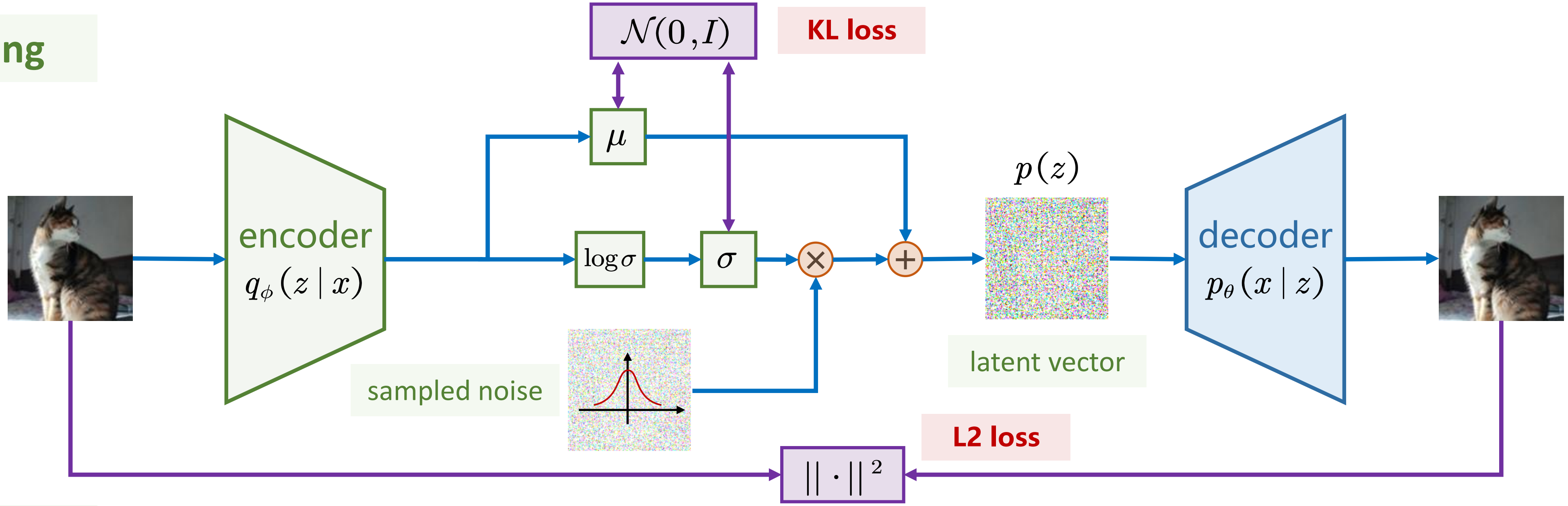
$$\geqslant \int_z q_\phi(z \mid \boldsymbol{x}) \log \left[\frac{p(z, \boldsymbol{x})}{q_\phi(z \mid \boldsymbol{x})} \right] dz = \mathbb{E}_{q_\phi(z|\boldsymbol{x})} \log \left[\frac{p(z, \boldsymbol{x})}{q_\phi(z \mid \boldsymbol{x})} \right]$$

KL Divergence always ≥ 0

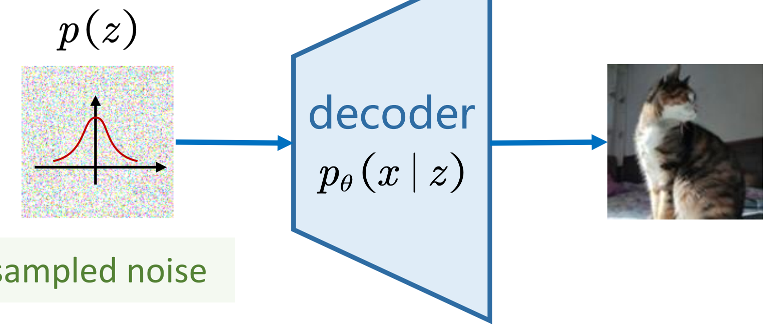
⑥ ELBO

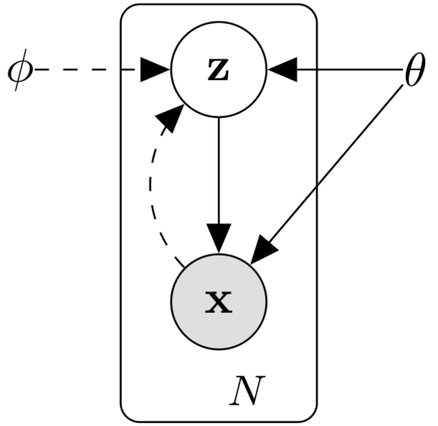
Review Variational Auto Encoder

1.Training



2.Sampling





The two loss in VAE

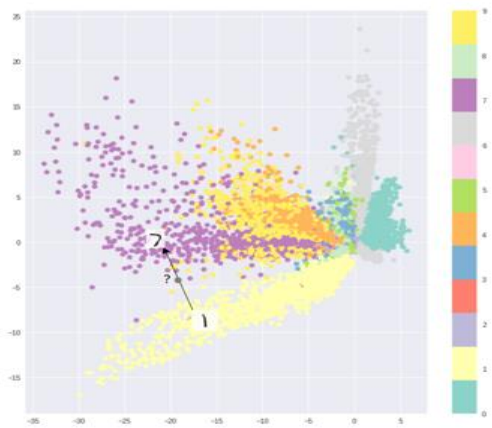
max | min

$$\mathbb{E}_{q_\phi(z \mid \boldsymbol{x})}[\log p_\theta(\boldsymbol{x} \mid z)] - KL(q_\phi(z \mid \boldsymbol{x}) || p(z))$$

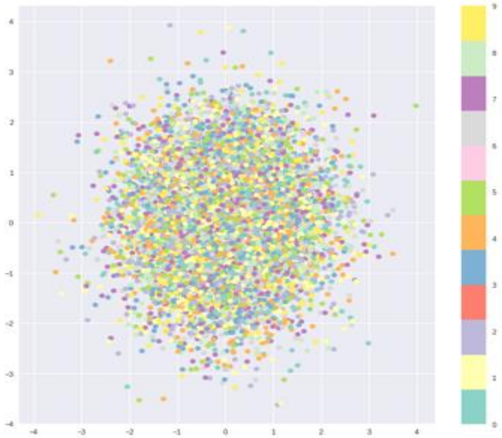
reconstruction term | prior matching term

$$||\boldsymbol{x} - \hat{\boldsymbol{x}}||^2 \qquad \frac{1}{2}(\mu_\phi^2 + \sigma_\phi^2 - \log \sigma_\phi^2 - 1)$$

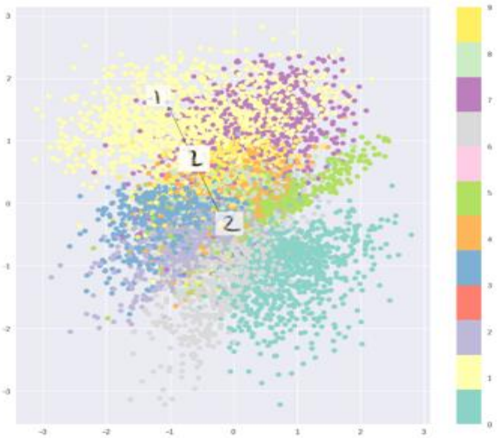
MINIST dataset



only reconstruction loss



only KL divergence

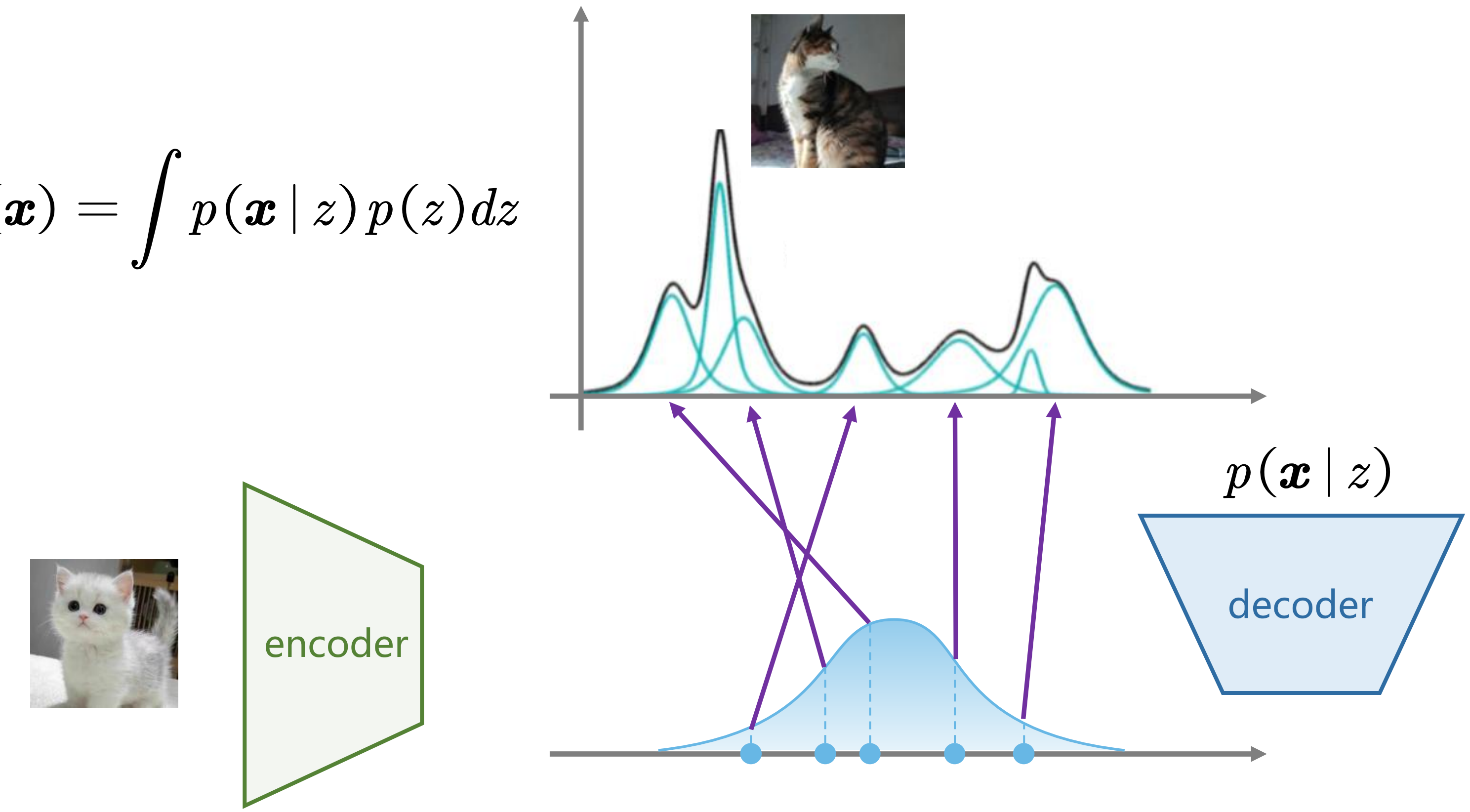


combination

Why does it work

The distribution of any data can be regarded as the superposition of several Gaussian distributions.

$$p(\boldsymbol{x}) = \int p(\boldsymbol{x} \mid z)\, p(z)\, dz$$



$$p(\boldsymbol{x} \mid z)$$

$$p(z \mid \boldsymbol{x}) = \mathcal{N}(\mu(z), \sigma(z))$$

$$p(z) = \mathcal{N}(0, \boldsymbol{I})$$

Posterior probability collapse

Independent(x and z)

$$p(z \mid \boldsymbol{x}) = \mathcal{N}(\mu(z), \sigma(z))$$

Incomplete approximation

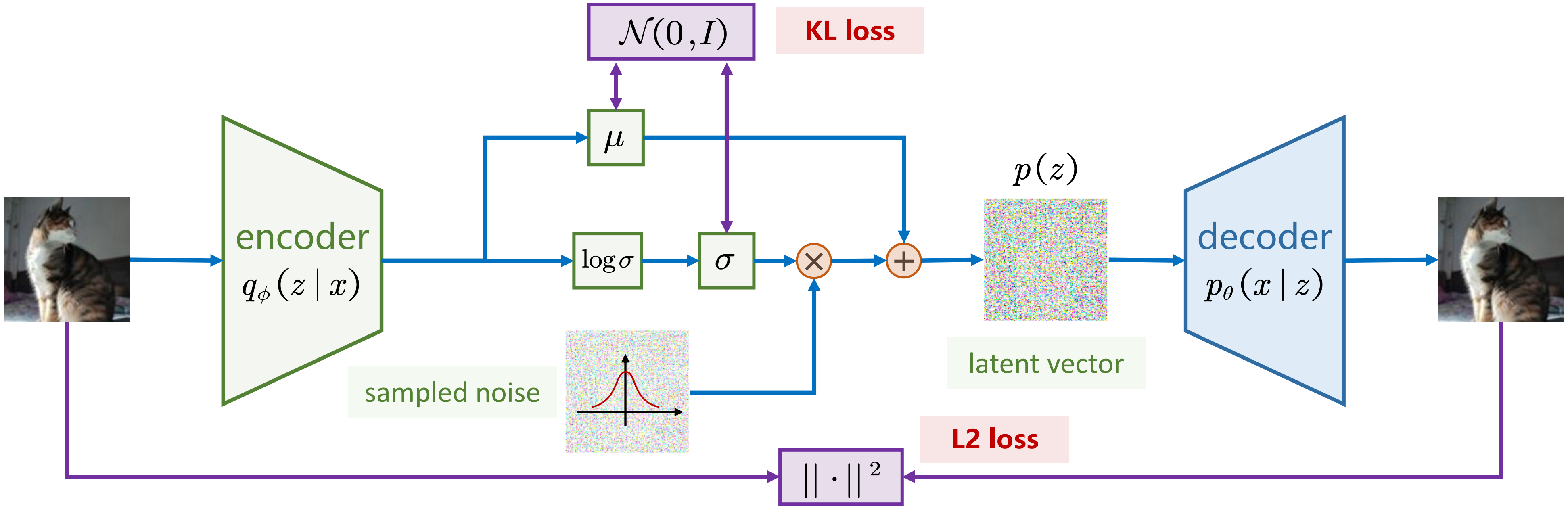
$$p(z) = \mathcal{N}(0, \boldsymbol{I})$$

Generative capacity

Reconstructive capacity

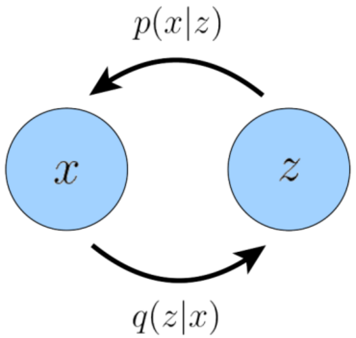
The limitation of VAE

Posterior probability and prior probability are not exactly equal. | LDM



MSE may cause the model to optimize toward a vague image. | GAN

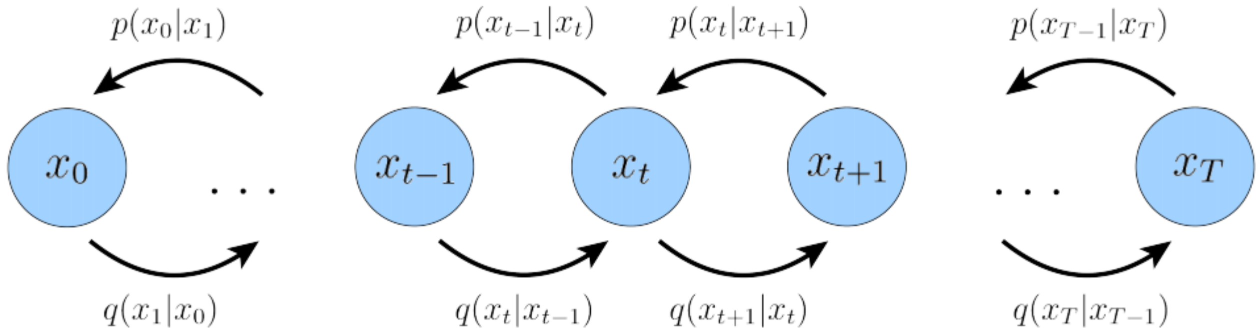
Relationship with diffusion



Markovian

Dimension equal

Gaussian step



VAE

Markovian Hierarchical VAE

⑥ ELBO

$$\log p(\boldsymbol{x}) \geqslant \mathbb{E}_{q_\phi(z|\boldsymbol{x})} \log \left[\frac{p(z, \boldsymbol{x})}{q_\phi(z \mid \boldsymbol{x})} \right]$$

② optimization (view 1)

$$\min \; -\log p_\theta(x_0) \leqslant -\log p_\theta(x_0) + D_{KL}(q(x_{1:T}|x_0) || p_\theta(x_{1:T}|x_0))$$

⑧ optimization (view 2)

$$\log p(\boldsymbol{x}) \geqslant \mathbb{E}_{q_\phi(z_{1:T}|\boldsymbol{x})} \log \left[\frac{p(z_{1:T}, \boldsymbol{x})}{q_\phi(z_{1:T} \mid \boldsymbol{x})} \right]$$

$$\min \; -\log p_\theta(x_0) \leqslant \mathbb{E}_{q(x_{1:T}|x_0)} \left[\log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})} \right]$$

$$\min \; -\log p_\theta(x_0) \leqslant \mathbb{E}_{q(x_{1:T}|x_0)} [D_{KL}(q(x_T|x_0) || p_\theta(x_T))] + \sum_{t=2}^{T} D_{KL}(q(x_{t-1}|x_t, x_0) || p_\theta(x_{t-1}|x_t)) - \log p_\theta(x_0|x_1)$$

prior matching term

consistency term

reconstruction term

Reference

1. Kingma D P, Welling M. Auto-encoding variational bayes. arXiv, 2013.
2. Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. NeuraIPS, 2020.
3. Luo C. Understanding diffusion models: A unified perspective. arXiv, 2022.
4. Rombach R, Blattmann A, Lorenz D, et al. High-resolution image synthesis with latent diffusion models. CVPR, 2022.
5. https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73
6. https://www.jeremyjordan.me/variational-autoencoders/
7. https://www.youtube.com/watch?v=8zomhgKrsmQ&t=2315s (李宏毅, Unsupervised Learning - Deep Generative Model (Part II))
8. https://spaces.ac.cn/archives/5253 (苏剑林, 变分自编码器（一）：原来是这么一回事)
9. https://kexue.fm/archives/5343 (苏剑林, 变分自编码器（二）：从贝叶斯观点出发)
10. https://www.bilibili.com/video/BV1xih7ecEMb (吃花椒的麦,【大白话01】一文理清 Diffusion Model 扩散模型 | 原理图解+公式推导)
11. https://www.bilibili.com/video/BV1KS421R7Yp (减论,【减论系列专栏：从分布到生成（二）】)

The end

非常感谢你能看到这，希望该课件对你有帮助，视频讲解版在B站。

课件中出现的是我家的猫咪的照片，她已经陪伴了我很多年了，感谢她的友情出镜。o(*￣▽￣*)ブ

Thank you so much for seeing this, I hope the slide is helpful, the video explanation version is on Bilibili.

The picture on the slide is of my cat, who has been with me for many years now, thanks for her friendly appearance! o(*￣▽￣*)ブ

> **“What I cannot create, I do not understand.”**
>
> **——Richard Feynman**