



变分自编码器

Variational Autoencoder, VAE

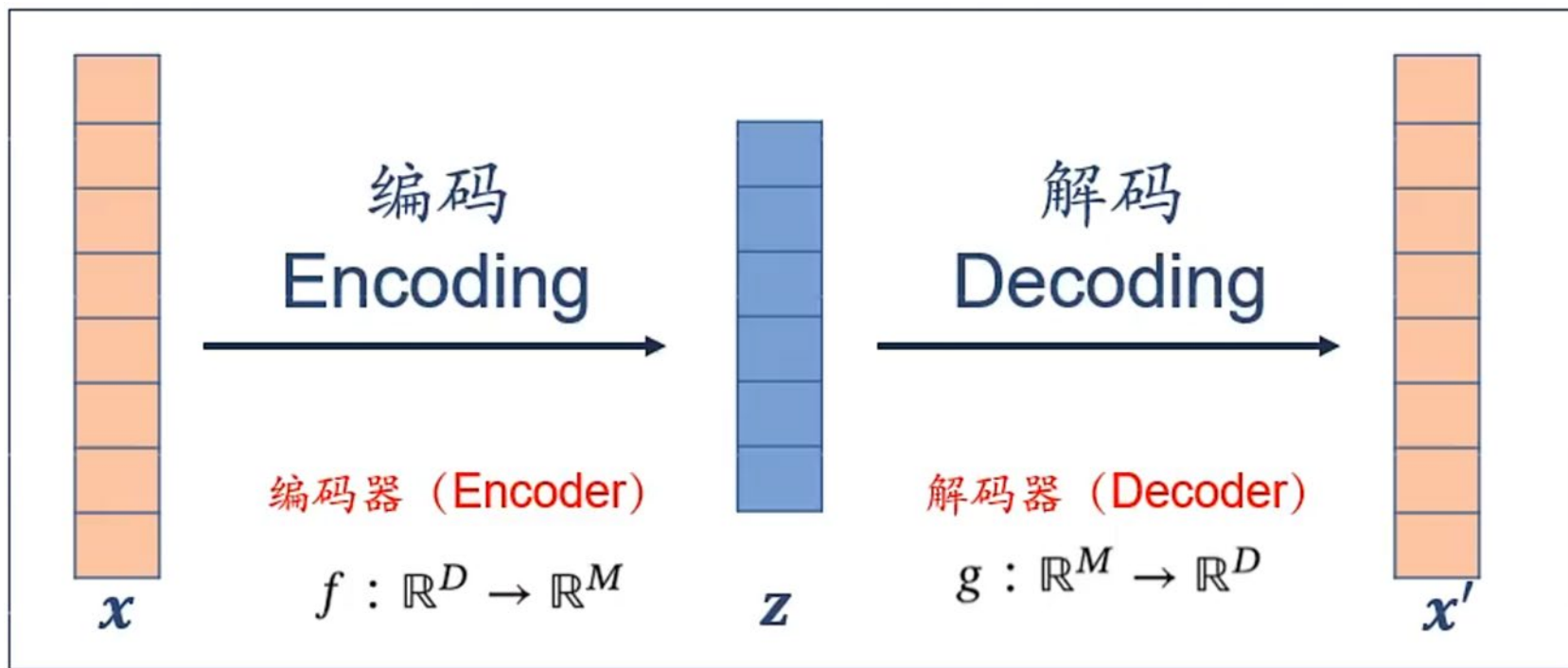
变分

- ▶ 泛函：从函数空间到数域的映射，即“函数的函数”
- ▶ 泛函的自变量称为“宗量”
- ▶ 变分：研究函数的“变形”对泛函变化 (δJ) 的影响
- ▶ 变分推理：用变分分布 q 去逼近复杂后验分布 p 。

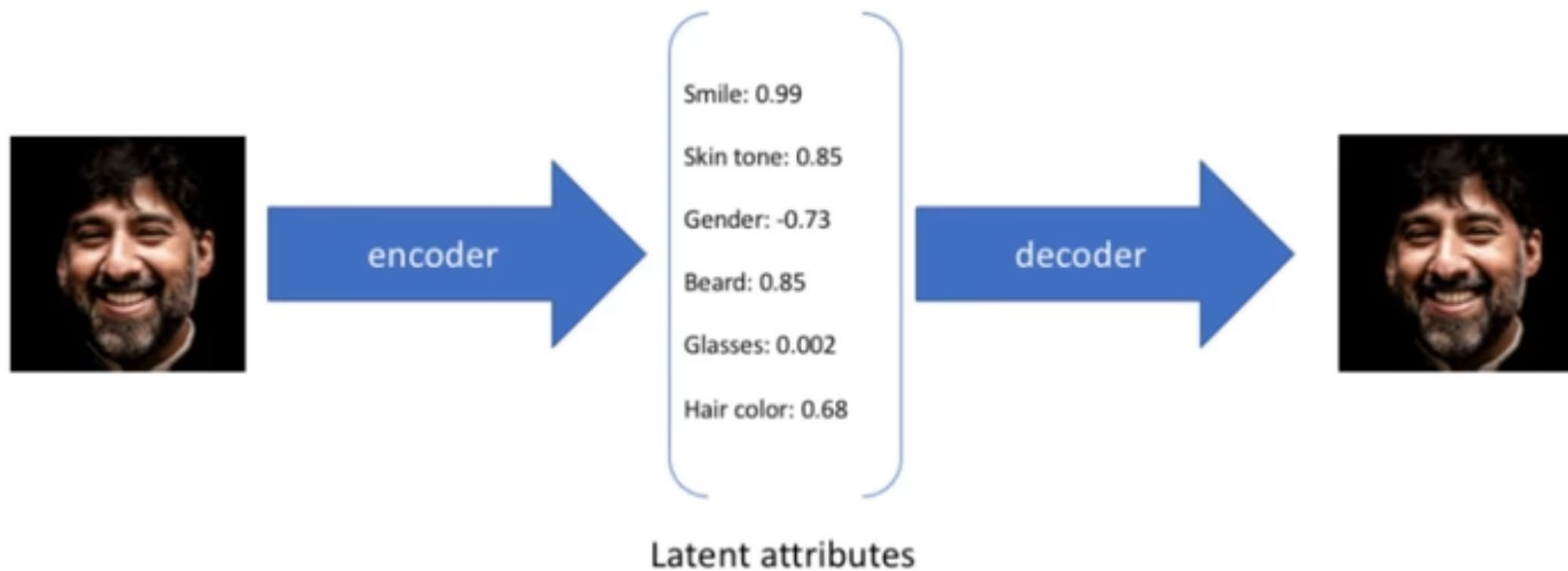
泛函 $J[\text{某个函数}] = A$

函数 $y[x = x_0] = A$

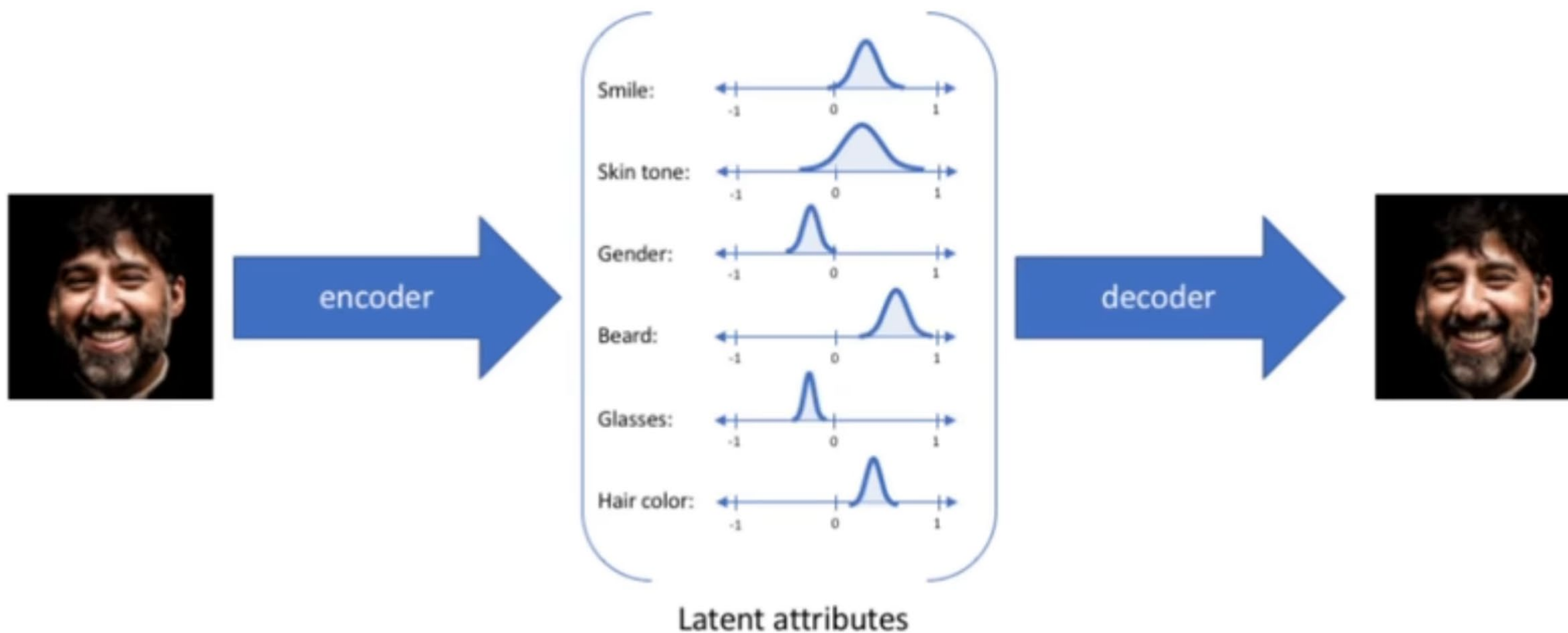
自编码器 (Autoencoder)



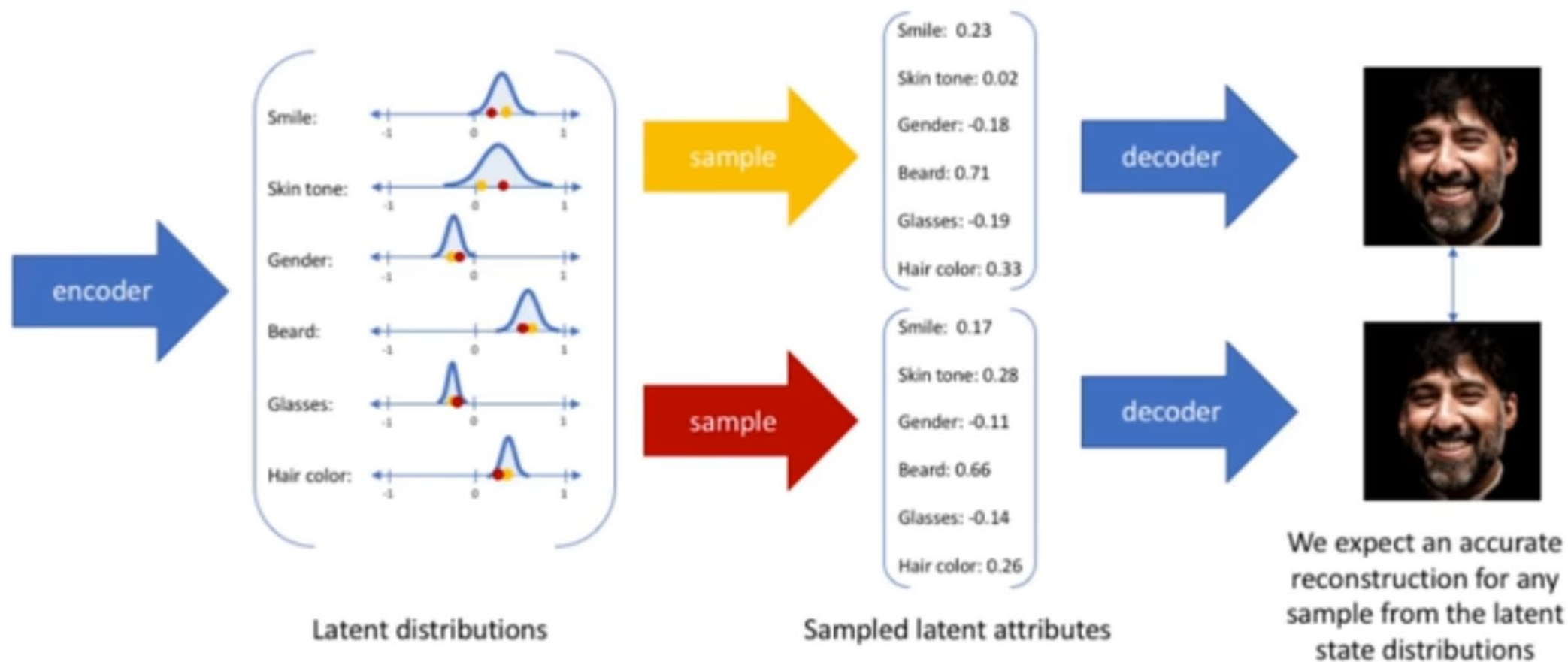
自编码器 (Autoencoder)



自编码器 (Autoencoder)



自编码器 (Autoencoder)



变分自编码器(VAE)

► 变分自编码器的模型结构可以分为两个部分：

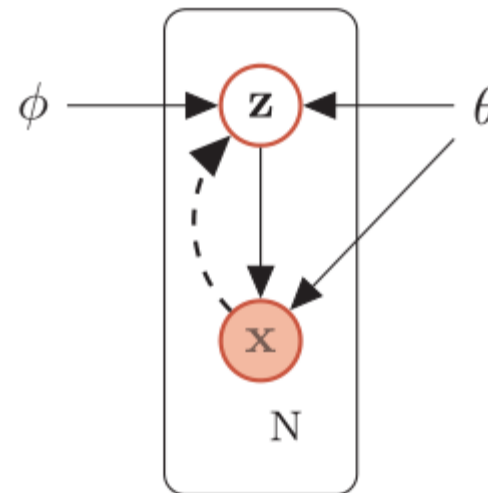
► 寻找后验分布 $p(z|x;\theta)$ 的变分近似

$q(z|x;\phi^*)$;

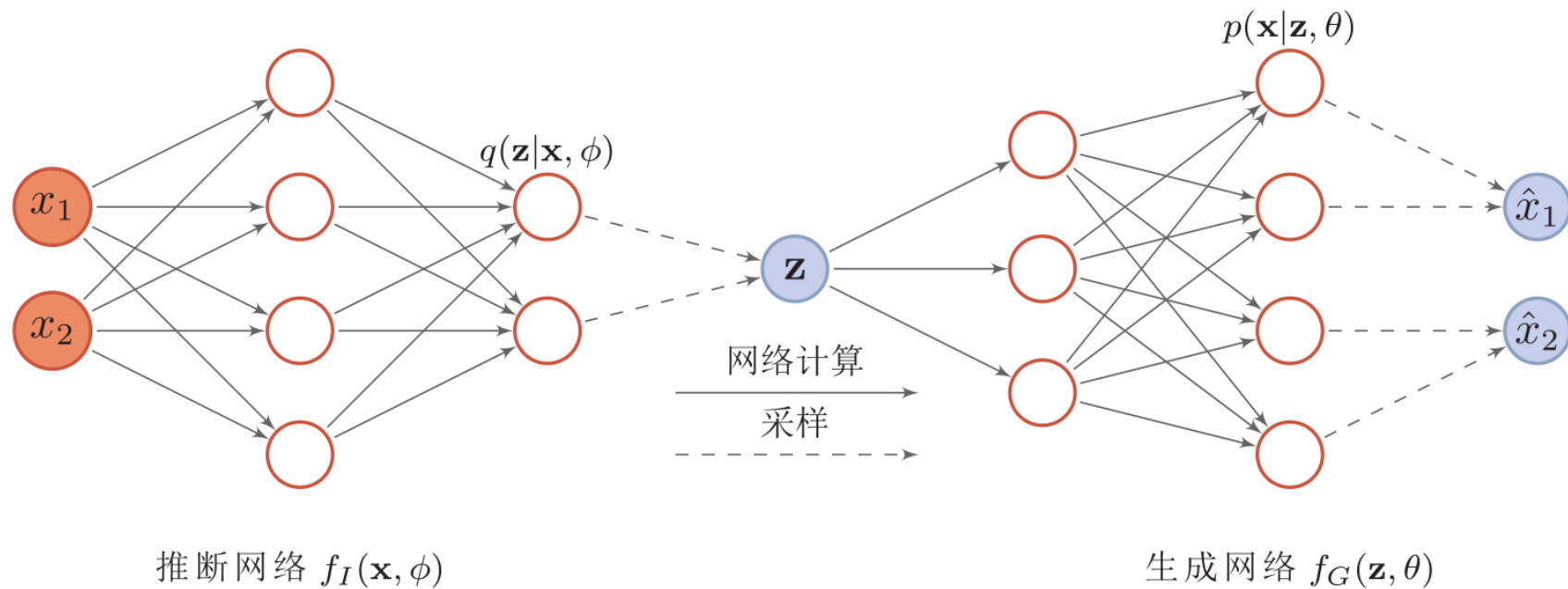
► 变分推断：用简单的分布 q 去近似复杂的分布 $p(z|x;\theta)$

► 在已知 $q(z|x;\phi^*)$ 的情况下，估计更好的分布 $p(x|z;\theta)$ 。

用神经网络来替代



变分自编码器



推断网络

► 推断网络

$$q(\mathbf{z}|\mathbf{x}, \phi) = \mathcal{N}(\mathbf{z}|\boldsymbol{\mu}_I, \boldsymbol{\sigma}_I^2 I)$$

$$\begin{bmatrix} \boldsymbol{\mu}_I \\ \boldsymbol{\sigma}_I \end{bmatrix} = f_I(\mathbf{x}, \phi)$$

► 目标

$$\begin{aligned} \phi^* &= \arg \min_{\phi} D_{\text{KL}}(q(\mathbf{z}|\mathbf{x}, \phi) || p(\mathbf{z}|\mathbf{x}, \theta)) \\ &= \arg \min_{\phi} \log p(\mathbf{x}|\theta) - ELBO(q, \mathbf{x}|\theta, \phi) \\ &= \arg \max_{\phi} ELBO(q, \mathbf{x}|\theta, \phi). \end{aligned}$$

生成网络

▶ 先验分布 $p(\mathbf{z}|\theta)$

- ▶ 一般假设隐变量 \mathbf{z} 的先验分布为各向同性的标准高斯分布 $\mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{I})$

▶ 条件概率分布 $p(\mathbf{x}|\mathbf{z}, \theta)$

- ▶ 假设 $p(\mathbf{x}|\mathbf{z}, \theta)$ 服从对角化协方差的高斯分布

$$p(\mathbf{x}|\mathbf{z}, \theta) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_G, \boldsymbol{\sigma}_G^2 \mathbf{I})$$

▶ 目标

$$\theta^* = \arg \max_{\theta} ELBO(q, \mathbf{x}|\theta, \phi)$$

模型汇总

$$\max_{\theta, \phi} ELBO(q, \mathbf{x}|\theta, \phi) = \max_{\theta, \phi} \mathbb{E}_{\mathbf{z} \sim q(\mathbf{z}|\phi)} \left[\log \frac{p(\mathbf{x}|\mathbf{z}, \theta)p(\mathbf{z}|\theta)}{q(\mathbf{z}|\phi)} \right]$$



$$\max_{\theta, \phi} \mathbb{E}_{\mathbf{z} \sim q(\mathbf{z}|\mathbf{x}, \phi)} \left[\log p(\mathbf{x}|\mathbf{z}, \theta) \right] - D_{\text{KL}} \left(q(\mathbf{z}|\mathbf{x}, \phi) || p(\mathbf{z}|\theta) \right)$$

再参数化



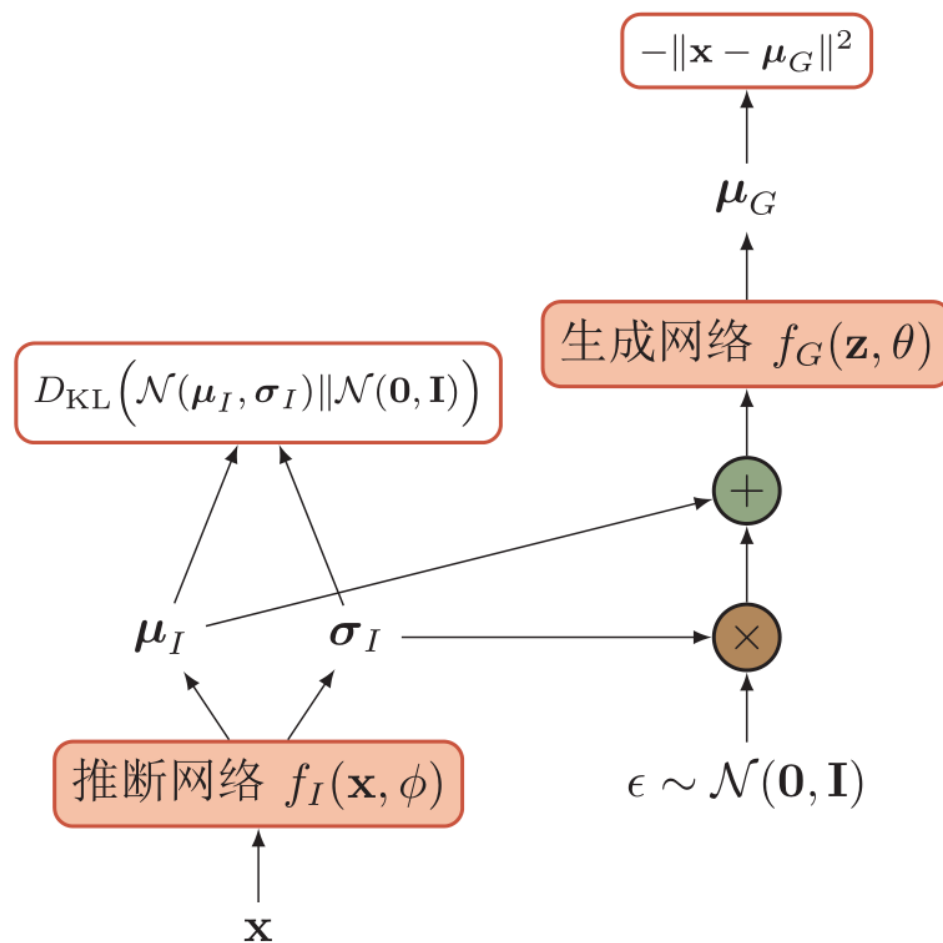
$$\sum_{n=1}^N \left(\frac{1}{M} \sum_{m=1}^M \log p(\mathbf{x}^{(n)}|\mathbf{z}^{(n,m)}, \theta) - D_{\text{KL}} \left(q(\mathbf{z}|\mathbf{x}^{(n)}, \phi) || \mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{I}) \right) \right)$$

再参数化

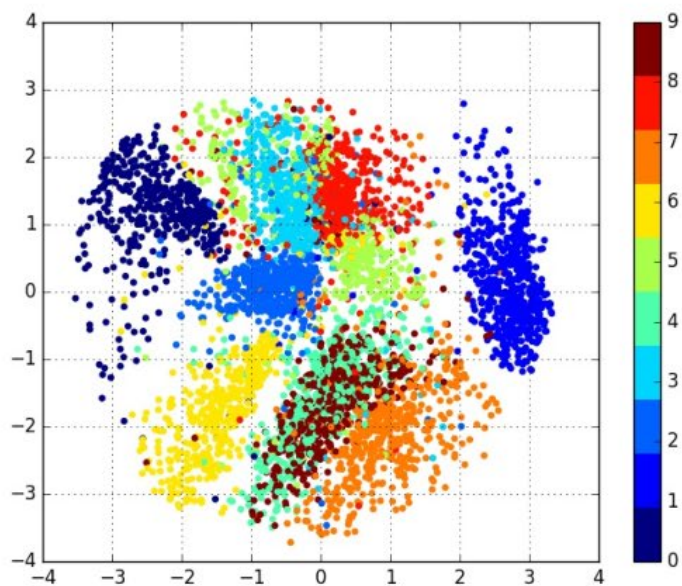
- ▶ 分布 $q(z|x,\phi)$ 依赖于参数 ϕ
- ▶ 再参数化 (reparameterization) 是实现通过随机变量实现反向传播的一种重要手段

$$z \sim N(\mu_I, \sigma_I^2 I) \quad \Rightarrow \quad \begin{aligned} \epsilon &\sim N(0, I) \\ z &= \mu_I + \sigma_I \odot \epsilon, \end{aligned}$$

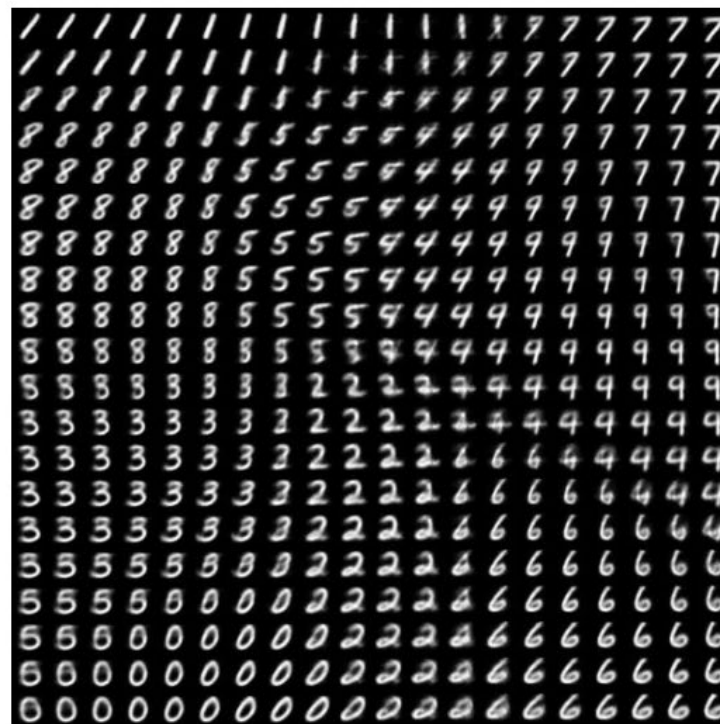
变分自编码器的训练过程



变分自编码器学习到的隐变量流形



(a) 训练集上所有样本在隐空间上的投影。



(b) 隐变量 \mathbf{z} 在图像空间的投影。



谢谢

<https://nndl.github.io/>