第7章：概率图模型和深度生成模型

中国科学技术大学
电子工程与信息科学系

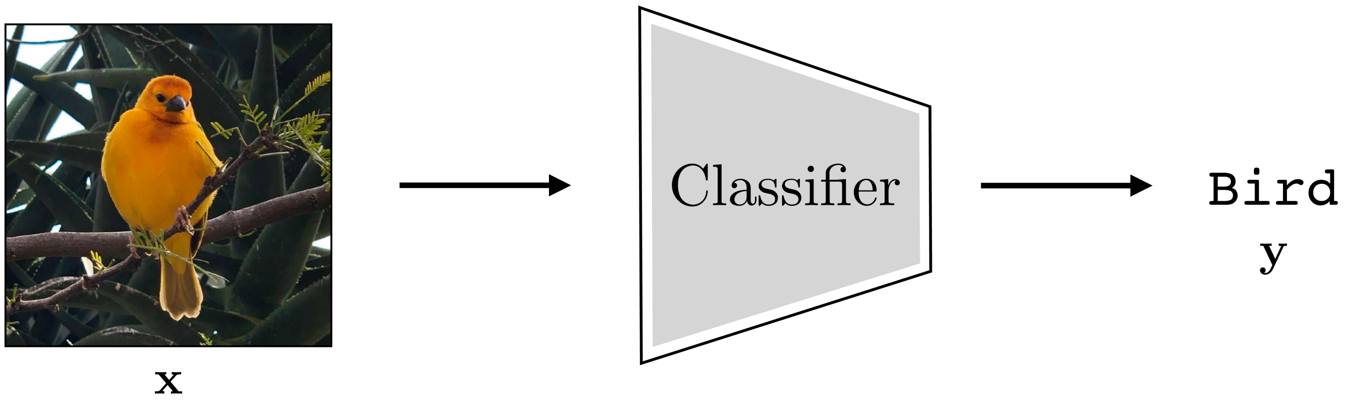
主讲教师：李厚强 (lihq@ustc.edu.cn)
周文罡 (zhwg@ustc.edu.cn)
李　礼 (lil1@ustc.edu.cn)
胡　洋 (eeyhu@ustc.edu.cn)

生成式模型（Generative Models）

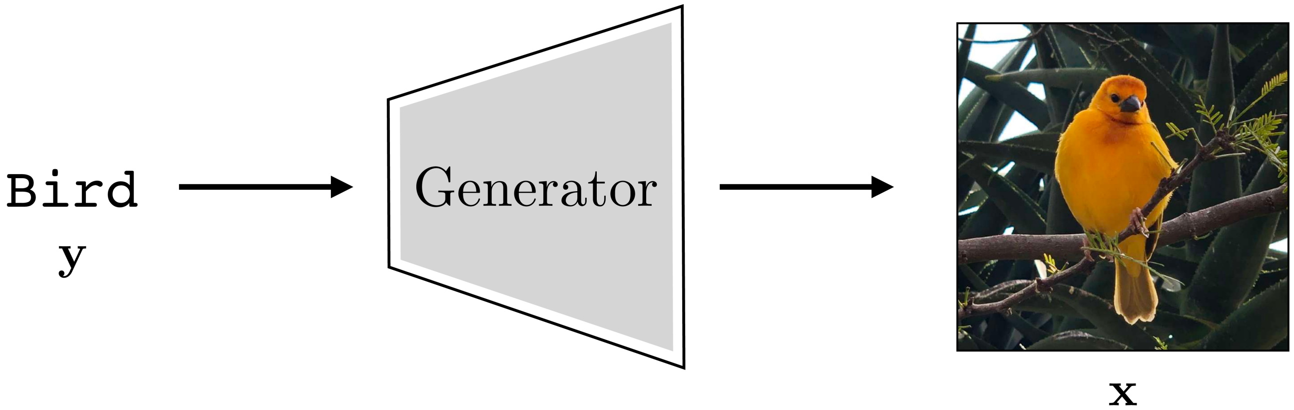
- 生成式模型概论
- 自回归生成模型
- 变分自编码器
- 扩散模型
- 生成对抗网络
- 条件生成式模型

生成式模型概论

- 判别式模型



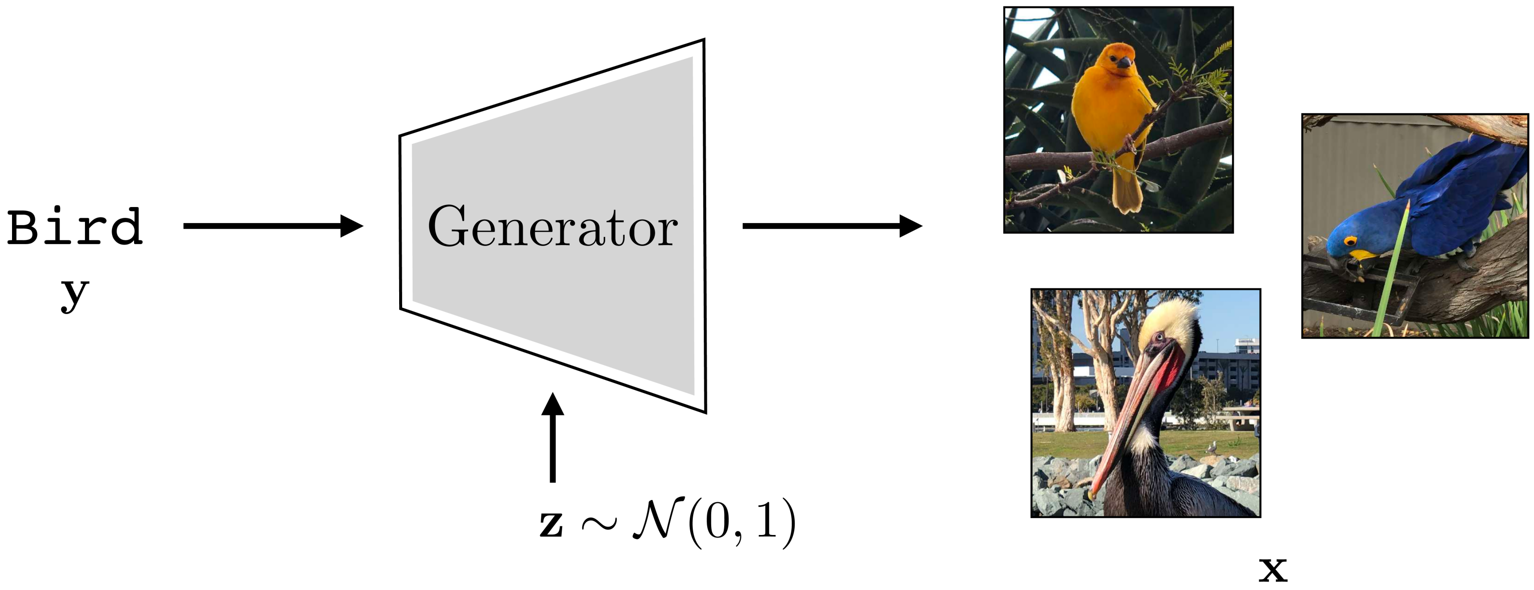
- 生成式模型



生成式模型概论

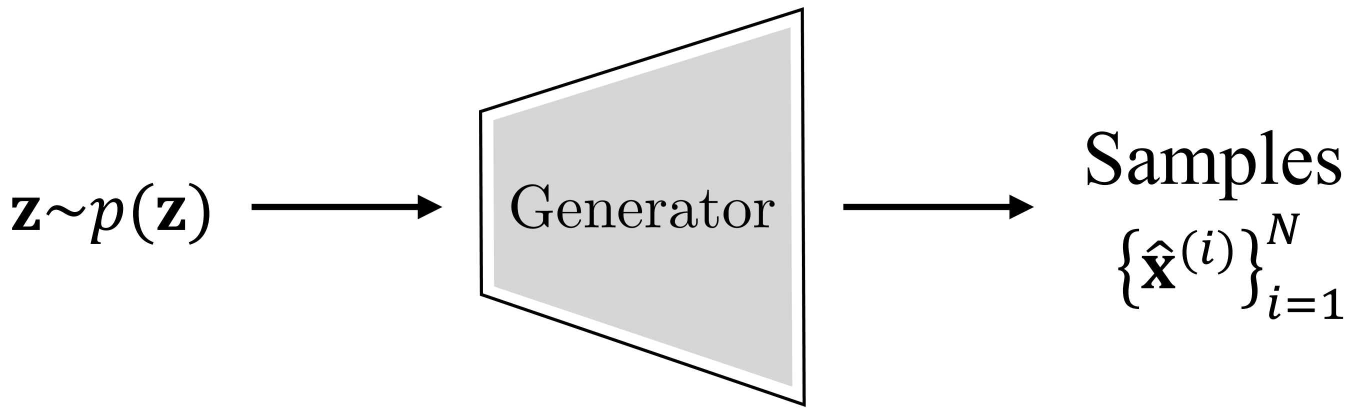
□ 生成式模型

- 一对多映射：引入随机性
- y：标签、文本描述、草图、。。。
- z：隐变量

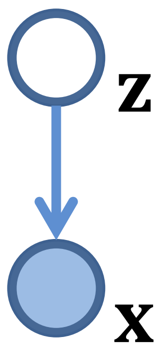


生成式模型概论

□ 无条件生成式模型



- 观测变量 $\mathbf{x}$ 服从某固定但未知的分布
- 隐变量 $\mathbf{z}$ 服从先验分布 $p(\mathbf{z})$
- 联合分布：$p(\mathbf{x},\mathbf{z}) = p(\mathbf{x}|\mathbf{z})p(\mathbf{z})$

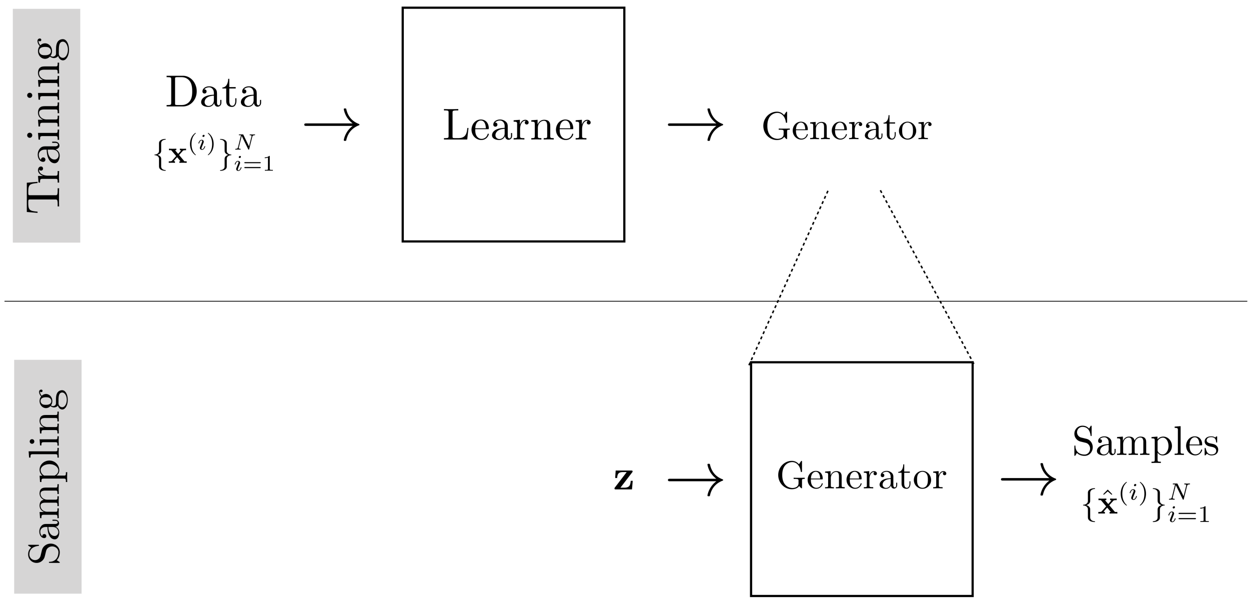


生成式模型概论

□ 生成器的学习和使用

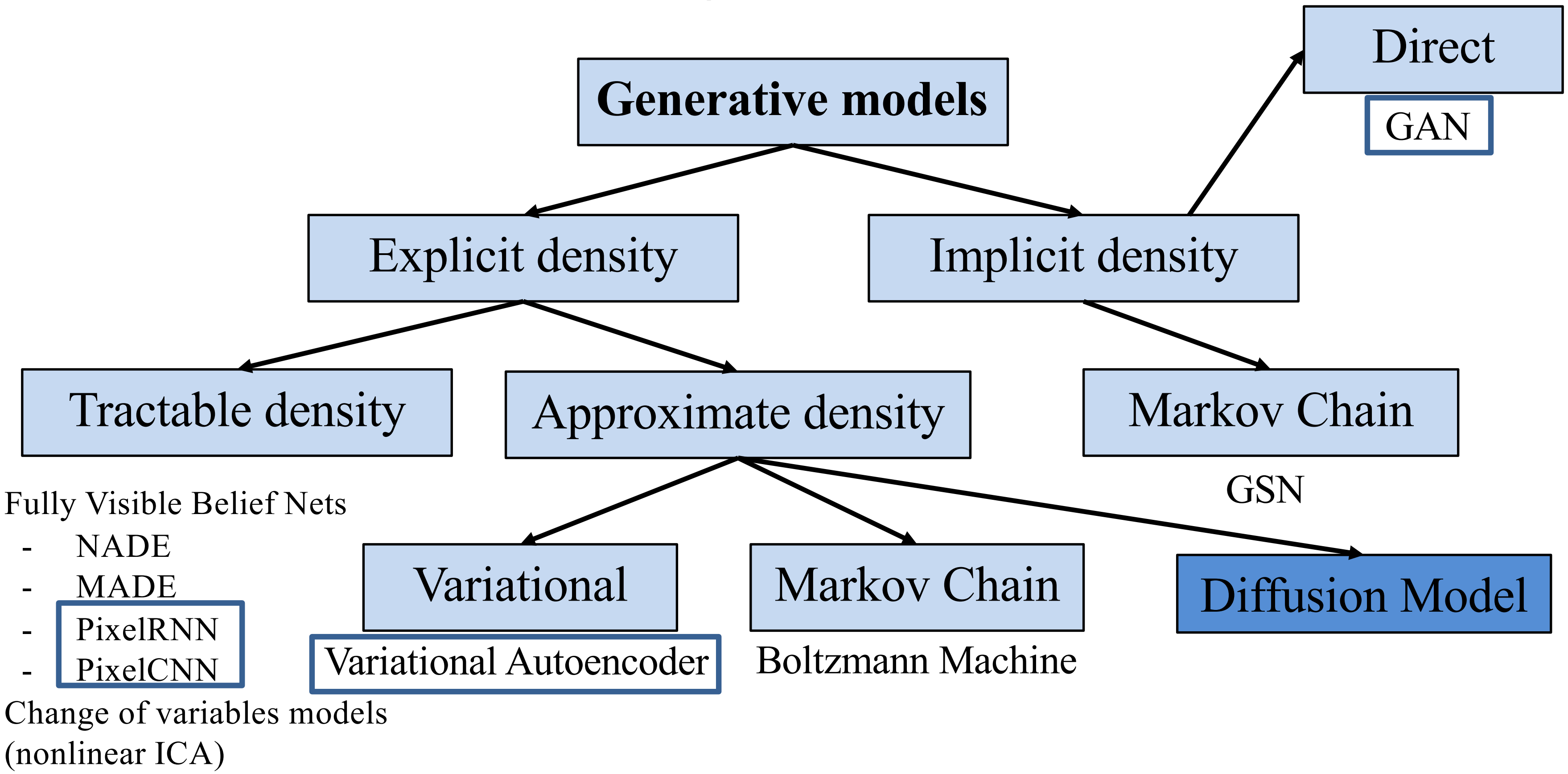
- 用观测数据 $\{\mathbf{x}^{(i)}\}_{i=1}^{N}$ 学习生成器
- 生成过程

$$\mathbf{z} \sim p(\mathbf{z})$$
$$\mathbf{x} = g(\mathbf{z}) \quad (\mathbf{x} \sim p(\mathbf{x}|\mathbf{z}))$$



生成式模型分类

- 根据对概率密度函数的表达，生成式模型可以分为两类：显式表达和隐式表达概率密度函数的生成式模型。
- 显式表达概率密度函数的生成式模型又可分为近似概率密度函数的生成式模型和解析概率密度函数的生成式模型。



生成式模型（Generative Models）

- 生成式模型概论
- 自回归生成模型
- 变分自编码器
- 扩散模型
- 生成对抗网络

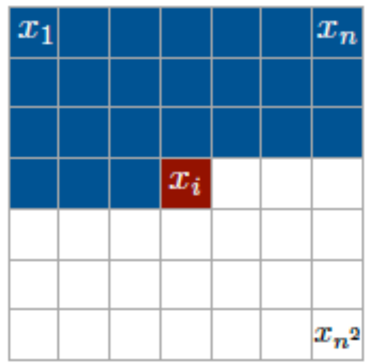
自回归生成模型

- PixelRNN和PixelCNN
  - 建模 $p(\mathrm{x})$

$$p(\mathbf{x}) = \prod_{i=1}^{n^2} p(x_i | x_1, ..., x_{i-1})$$

Likelihood of image x

Probability of the i-th pixel value given all previous pixels

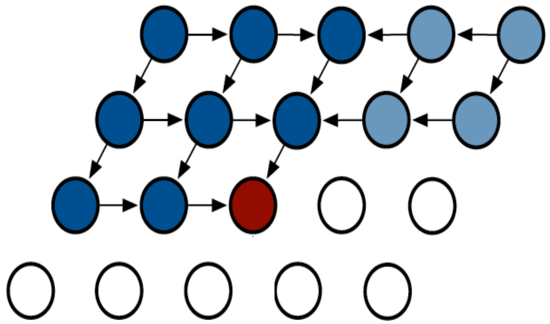


  - 利用训练数据进行最大似然估计

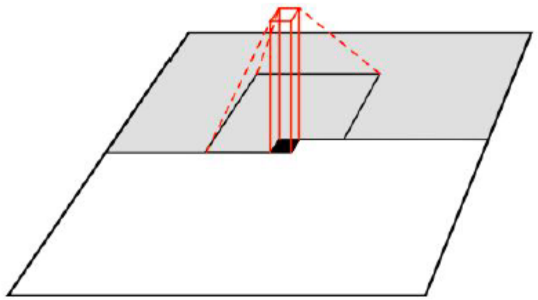
- A. van den Oord, N. Kalchbrenner and K. Kavukcuoglu. “Pixel Recurrent Neural Networks”. In ICML, 2016.

自回归生成模型

- PixelRNN
  - 利用RNN (BiLSTM)建模像素间的依赖关系
  - 缺点：训练、生成速度慢

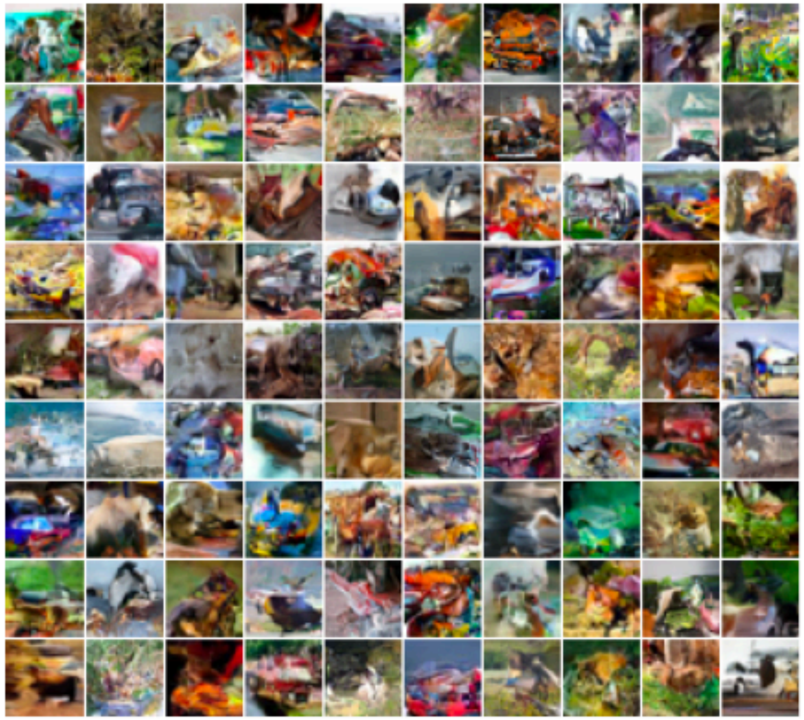


- PixelCNN
  - 利用CNN (masked conv)建模像间的依赖关系
  - 训练速度比PixelRNN快，生成速度依旧慢

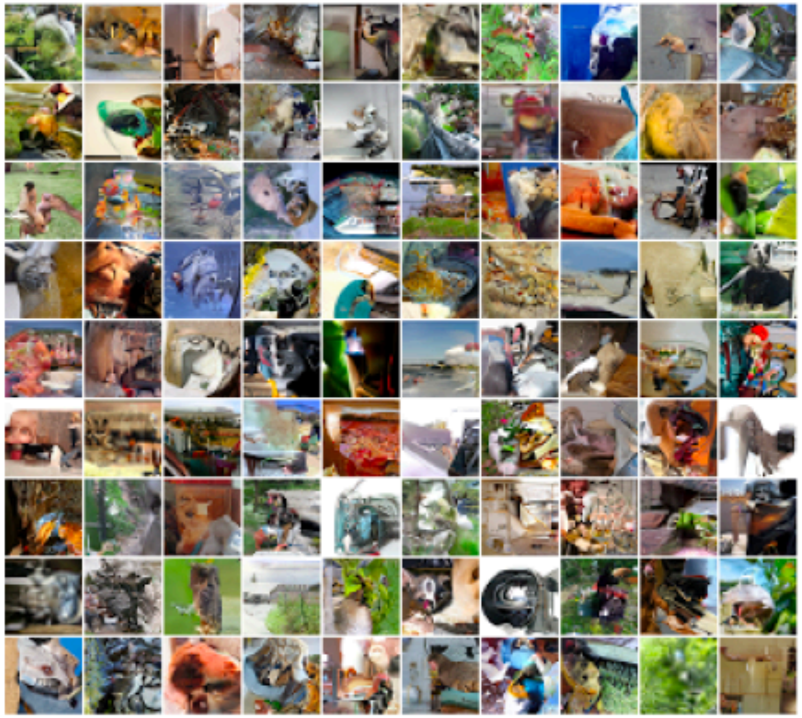


自回归生成模型

□ 生成结果



32×32 CIFAR-10



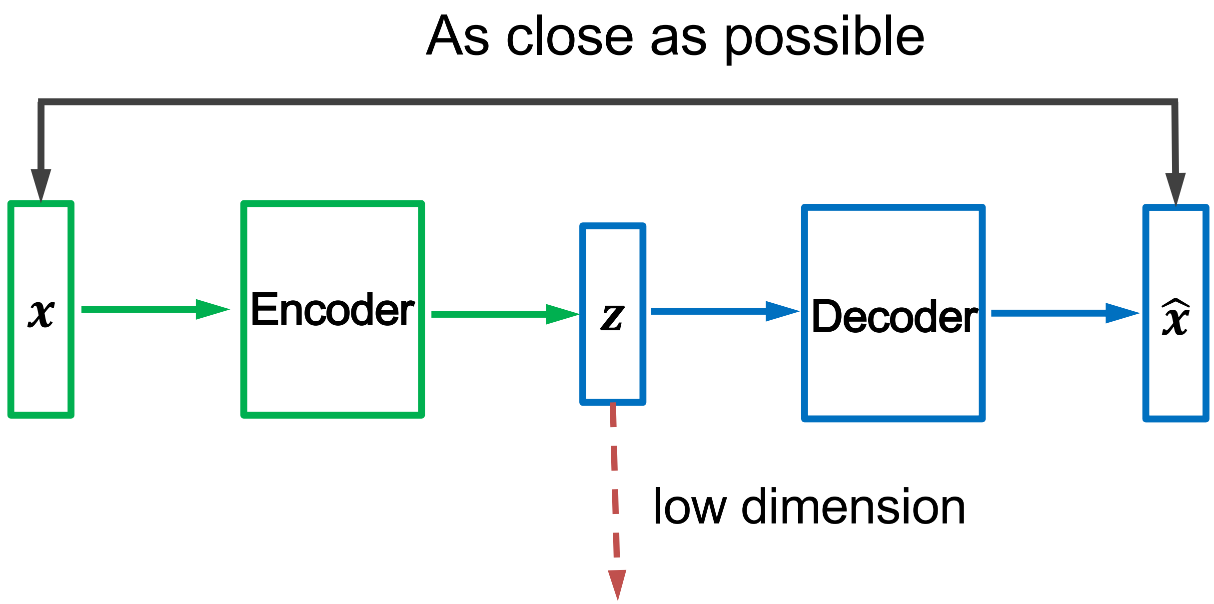
32×32 Imagenet

生成式模型（Generative Models）

- □ 生成式模型概论
- □ 自回归生成模型
- □ 变分自编码器
- □ 扩散模型
- □ 生成对抗网络
- □ 条件生成式模型

自编码器(Auto-Encoder)

- 自编码器(Auto-Encoder)

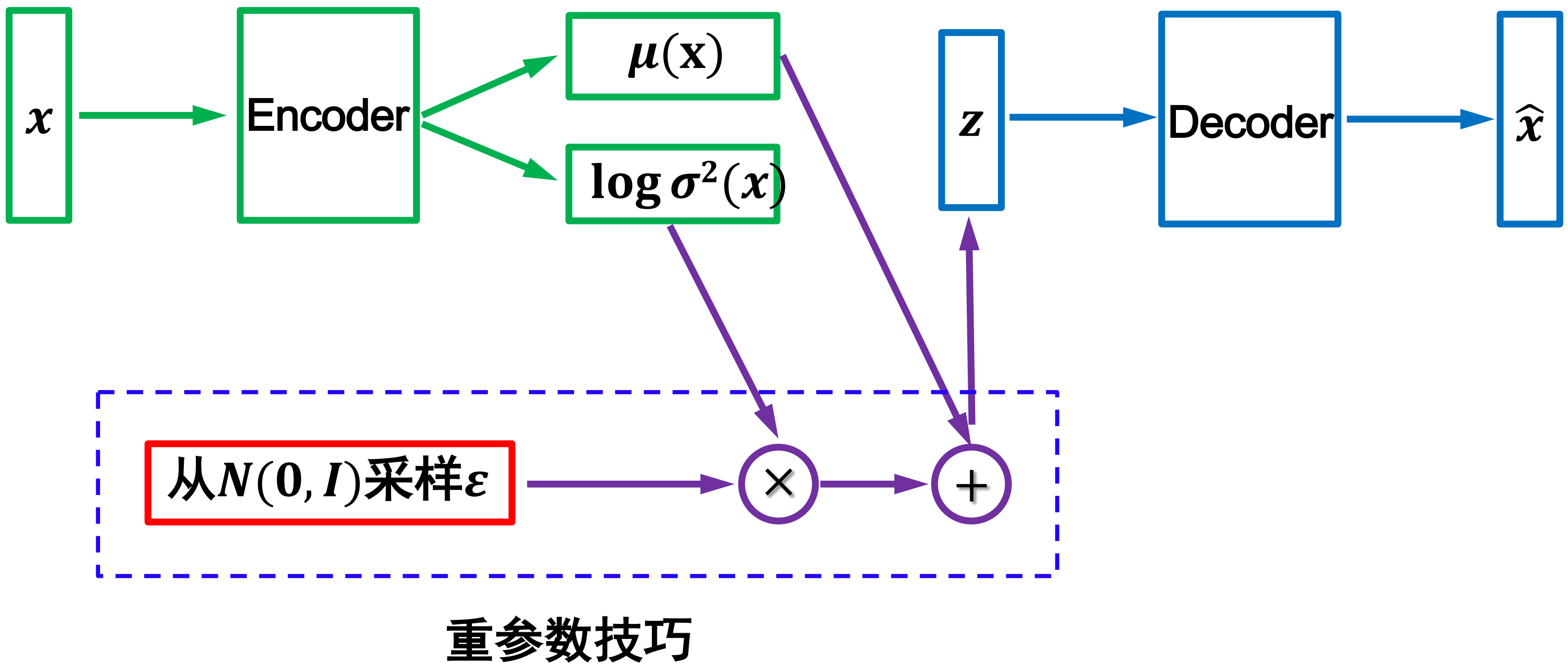


- Embedding, feature, code
- Used as feature for downstream tasks
- Cannot generate new sample

变分自编码器（VAE）

□ VAE的整体模型结构

- ■ 包含:编码器部分、解码器部分以及重参数化采样



- Diederik P. Kingma, and Max Welling. “Auto-Encoding Variational Bayes”. In arXiv:1312.6114, 2013.

变分自编码器（VAE）

- 观测样本与隐变量的联合分布

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

  - $\tilde{p}(x)$是根据样本$x_1,x_2,\ldots,x_n$确定的关于x的先验分布。尽管我们无法准确写出它的形式，但它是确定的、存在的。

- 设想用一个新的联合概率分布$q(x,z)$来逼近$p(x,z)$，并用KL散度来计算它们的距离：

$$KL(p(x,z)||q(x,z))=\iint p(x,z)\log\frac{p(x,z)}{q(x,z)}dzdx$$

- 我们希望两个分布越接近越好，所以KL散度越小越好。

变分自编码器（VAE）

将$p(x,z)=\tilde{p}(x)p(z|x)$ 带入

$$KL(p(x,z)||q(x,z)) \quad = \int \tilde{p}(x)[\int p(z|x)\log\frac{\tilde{p}(x)p(z|x)}{q(x,z)}dz]\,dx$$

$$= \mathbb{E}_{x\sim\tilde{p}(x)}[\int p(z|x)\log\frac{\tilde{p}(x)p(z|x)}{q(x,z)}dz]$$

进一步简化：

$$KL(p(x,z)||q(x,z)) = \mathbb{E}_{x\sim\tilde{p}(x)}[\int p(z|x)\log\tilde{p}(x)\,dz] + \mathbb{E}_{x\sim\tilde{p}(x)}[\int p(z|x)\log\frac{p(z|x)}{q(x,z)}dz)]$$

$$= \mathbb{E}_{x\sim\tilde{p}(x)}[\log\tilde{p}(x)\underbrace{\boxed{\int p(z|x)dz}}_{1}] + \mathbb{E}_{x\sim\tilde{p}(x)}[\int p(z|x)\log\frac{p(z|x)}{q(x,z)}dz)]$$

$$= \underbrace{\boxed{\mathbb{E}_{x\sim\tilde{p}(x)}[\log\tilde{p}(x)]}}_{\text{常量}C} + \mathbb{E}_{x\sim\tilde{p}(x)}[\int p(z|x)\log\frac{p(z|x)}{q(x,z)}dz)]$$

变分自编码器（VAE）

- 通过移项，我们可以令：

$$\mathcal{L} = KL(p(x,z)||q(x,z)) - C = \mathbb{E}_{x\sim\tilde{p}(x)}[\int p(z|x)\log\frac{p(z|x)}{q(x,z)}dz)]$$

- 最小化KL散度等价于最小化$\mathcal{L}$。为了得到生成模型，我们把$q(x,z)$写成$q(x|z)q(z)$，于是有：

$$\mathcal{L} = \mathbb{E}_{x\sim\tilde{p}(x)}[\int p(z|x)\log\frac{p(z|x)}{q(x|z)q(z)}dz)]$$

$$= \mathbb{E}_{x\sim\tilde{p}(x)}[-\int p(z|x)\log q(x|z)dz + \int p(z|x)\log\frac{p(z|x)}{q(z)}dz]$$

$$= \mathbb{E}_{x\sim\tilde{p}(x)}\underbrace{[\mathbb{E}_{z\sim p(Z|x)}[-\log q(x|z)] + KL(p(z|x)||q(z))]}_{\text{优化目标}}$$

变分自编码器（VAE）

$$\mathcal{L} = \mathbb{E}_{x\sim\tilde{p}(x)}[\mathbb{E}_{z\sim p(z|x)}[-\log q(x|z)] + KL(p(z|x)||q(z))]$$

- 为了方便采样，我们假设$z\sim N(0,I)$，即标准的多元正态分布
- 假设$p(z|x)$也是(各分量独立的)正态分布，其均值与方差由x来决定（用一个神经网络计算）：

$$p(z|x) = \frac{1}{\prod_{k=1}^{d}\sqrt{2\pi\sigma_{(k)}^2(x)}}\exp(-\frac{1}{2}\left\|\frac{z-\mu(x)}{\sigma(x)}\right\|^2)$$

- x是神经网络的输入，输出则是均值$\mu(x)$与方差$\sigma^2(x)$。这里的神经网络就起到了类似Encoder的作用。$\mathcal{L}$中的KL散度项可以解析计算：

$$KL(p(z|x)||q(z)) = \frac{1}{2}\sum_{k=1}^{d}(\mu_{(k)}^2(x) + \sigma_{(k)}^2(x) - \log\sigma_{(k)}^2(x) - 1)$$

变分自编码器（VAE）

$$\mathcal{L} = \mathbb{E}_{x\sim\tilde{p}(x)}\left[\mathbb{E}_{z\sim p(z|x)}[-\log q(x|z)] + KL(p(z|x)||q(z))\right]$$

- 若假设$q(x|z)$是正态分布：

$$q(x|z) = \frac{1}{\prod_{k=1}^{D}\sqrt{2\pi\sigma_{(k)}^2(z)}}\exp\left(-\frac{1}{2}\left\|\frac{x-\mu(z)}{\sigma(z)}\right\|^2\right)$$

- 用神经网络建模，输入是z，输出是$\mu(z)$与$\sigma^2(z)$。

$$-\log q(x|z) = \frac{1}{2}\left\|\frac{x-\mu(z)}{\sigma(z)}\right\|^2 + \frac{D}{2}\log 2\pi + \frac{1}{2}\sum_{k=1}^{D}\log\sigma_{(k)}^2(z)$$

- 若固定方差为一个常数σ^2，则有：

$$-\log q(x|z) \sim \frac{1}{2\sigma^2}\|x-\mu(z)\|^2$$

- 最小化负对数似然等价于最小化MSE损失，$\mu(z)$起到了Decoder的作用。

变分自编码器（VAE）

- 回看VAE的优化目标：

$$\mathcal{L} = \mathbb{E}_{x\sim\tilde{p}(x)}\left[\mathbb{E}_{z\sim p(z|x)}[-\log q(x|z)] + KL(p(z|x)||q(z))\right]$$

- 对于等号右侧的第二项， $p(z|x)$起到Encoder的作用，同时KL散度将Encoder输出的隐向量约束为标准的多元正态分布；

- 对于等号右侧的第一项， $q(x|z)$起到Decoder的作用。若我们用MSE作为损失函数，这对应于$q(x|z)$为固定方差的多元正态分布；

- 待训练完成后，Decoder就是我们的生成模型。生成过程就是从标准多元正态分布中采样得到隐变量z，再输入Decoder，就可以得到生成样本了。

变分自编码器（VAE）

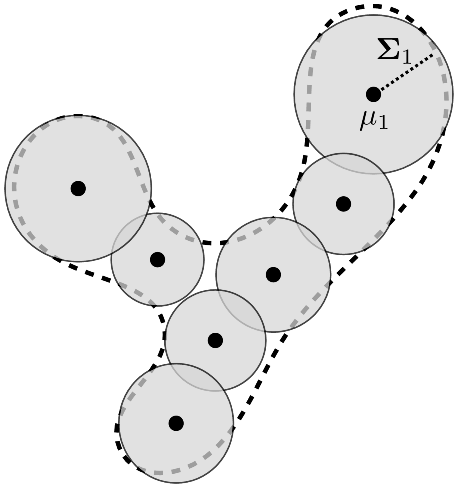
- VAE和自编码器
  - VAE首先通过Encoder将观测样本x编码（映射）为隐空间中的隐变量z，然后再将隐变量z重构回观测样本x。该训练过程与普通的自编码器类似，不同主要在于VAE多了隐变量采样和对隐变量的KL散度约束。
  - 一个训练好的自编码器，如果能够在它训练得到的隐空间采样，然后作为Decoder的输入，我们就得到了一个生成模型。但是普通自编码器的隐空间的分布是复杂且未知的，我们无法进行采样。VAE则对隐空间施加了KL散度约束，使其逼近简单的多元标准正态分布，这就使隐空间的采样变得可能，进而得到生成模型。
  - 所以，从训练过程来看，VAE就是隐空间受约束的自编码器。

变分自编码器（VAE）

□ VAE和高斯混合模型（GMM）

Finite mixture of Gaussians

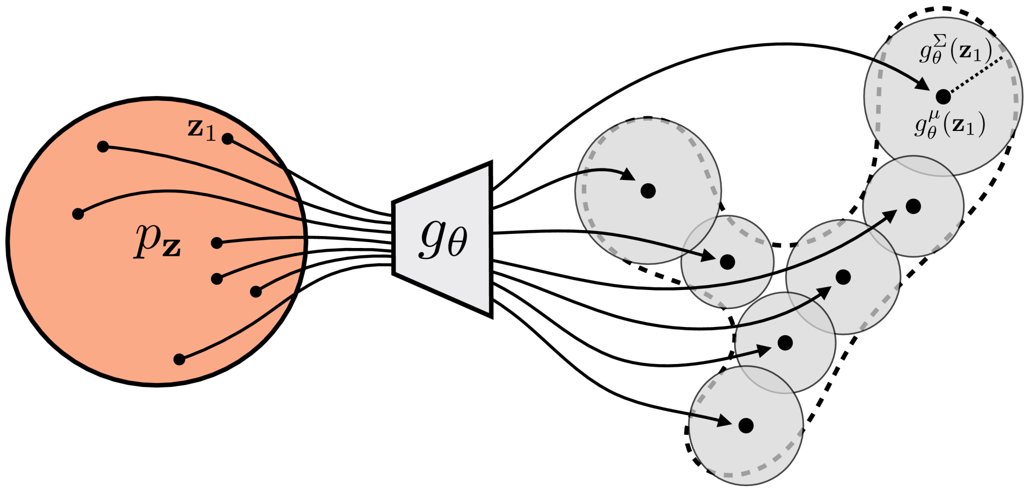
Parameters: $\{w_i, \mu_i, \boldsymbol{\Sigma}_i\}_{i=1}^{k}$



$$p_\theta(\mathbf{x}) = \sum_{i=1}^{k} w_i \mathcal{N}(\mathbf{x}; \mu_i, \boldsymbol{\Sigma}_i)$$

Infinite mixture of Gaussians (VAE)

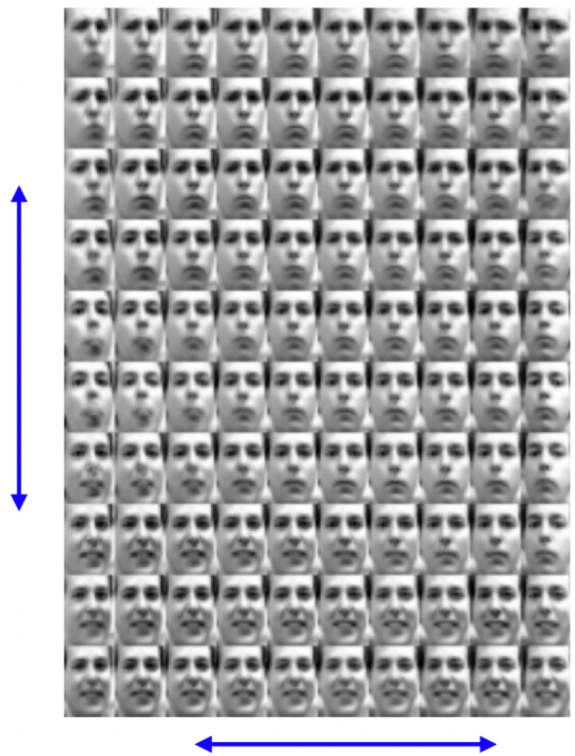
Parameters: θ



$$p_\theta(\mathbf{x}) = \int_{\mathbf{z}} p_{\mathbf{z}}(\mathbf{z}) \underbrace{\mathcal{N}(\mathbf{x}; g_\theta^\mu(\mathbf{z}), g_\theta^\Sigma(\mathbf{z}))}_{p(\mathbf{x}|\mathbf{z})} d\mathbf{z}$$

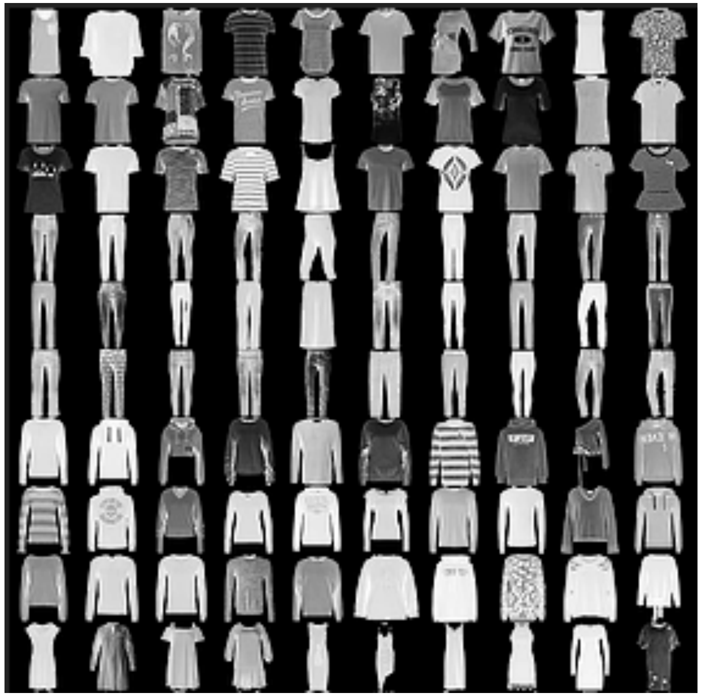
变分自编码器的优缺点

- VAE的优点
  - 得到 $p(z|x)$（编码器），可得到x的隐变量表示
  - 学习到具有一定可解释性的隐变量空间
  - VAE的训练过程比较稳定



变分自编码器的优缺点

- VAE的缺点
  - 变分法近似估计
  - VAE所生成的图片会比较模糊



训练图片

VAE随机生成