扩散模型 (Diffusion Model)

作者：Calvin
QQ: 179209347
Mail: 179209347@qq.com

介绍

笔记简介：
- 面向对象：深度学习初学者
- 依赖课程：**线性代数，统计概率**，优化理论，图论，离散数学，微积分，信息论

知乎专栏：
https://zhuanlan.zhihu.com/p/693738275

Github & Gitee 地址：
https://github.com/mymagicpower/AIAS/tree/main/deep_learning

https://gitee.com/mymagicpower/AIAS/tree/main/deep_learning

*** 版权声明：**
- 仅限用于个人学习
- 禁止用于任何商业用途

去噪扩散概率模型 (DDPM - Denoising Diffusion Probabilistic Models)

扩散模型的基本思想是通过逐步将数据分布从一个简单的分布（通常是高斯分布）逐渐变换为目标数据分布，从而生成新的数据样本。

基本原理

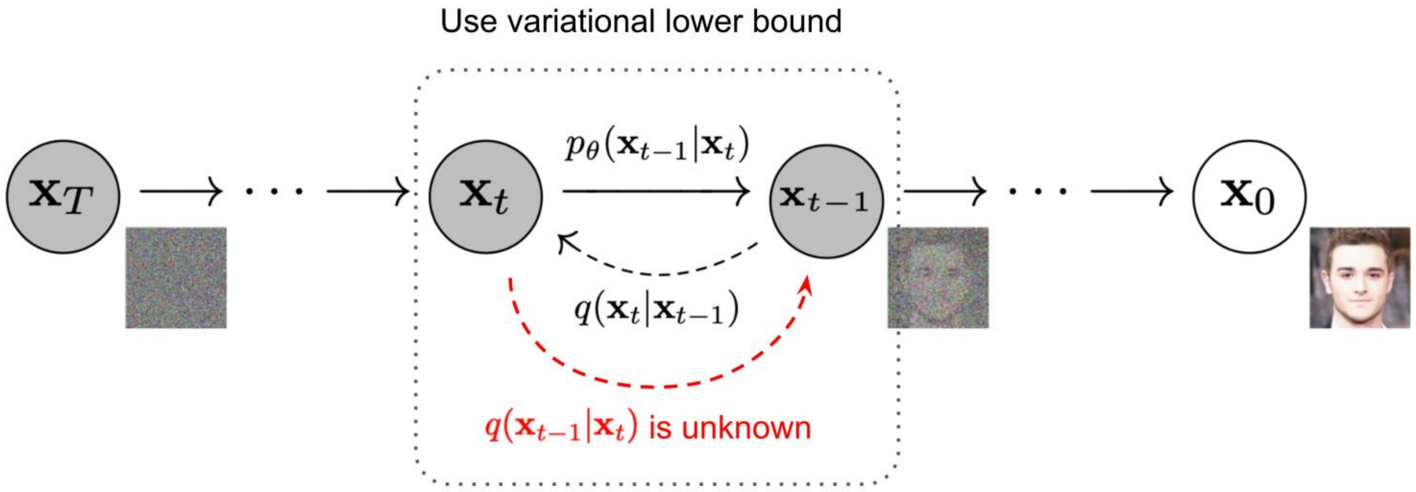
扩散模型的核心思想源于物理学中的扩散过程。在物理学中，扩散过程描述的是颗粒从高浓度区域向低浓度区域移动的现象。在深度学习的扩散模型中，这个过程被逆向应用，用于生成数据。

1.前向过程（Forward Process）：

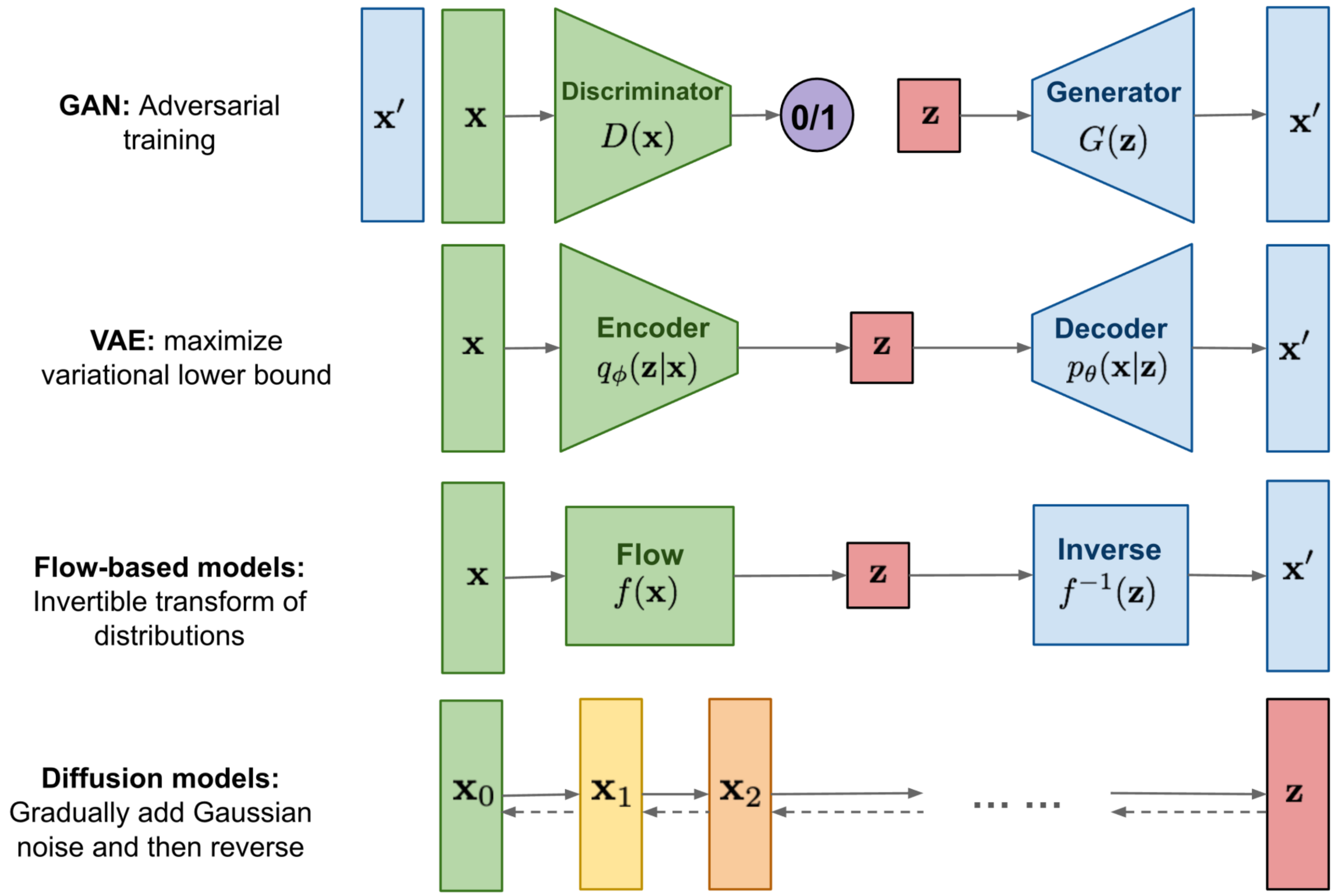
1. 在前向过程中，数据样本逐步加入噪声，逐渐变得更加模糊，直到最终变成纯噪声。这个过程通常用马尔可夫链来建模，定义为从数据分布 $q(x0)$ 开始，逐步加入噪声生成一系列中间状态 $x1,x2,...,xT$。
2. 每一步加入的噪声通常是高斯噪声，这样最终的 xT 接近于标准高斯分布。

2.反向过程（Reverse Process）：

1. 反向过程是前向过程的逆过程，即从纯噪声逐步去噪，最终生成目标数据样本。反向过程也可以通过马尔可夫链来建模，目标是在每一步逐渐减少噪声。
2. 训练反向过程的关键在于学会如何从噪声中逐步恢复出原始数据分布。通过学习一个神经网络（通常是U-Net结构），可以在每一步估计如何减少噪声。



生成模型对比



GAN: Adversarial training
$\mathbf{x}'$
$\mathbf{x}$
Discriminator
$D(\mathbf{x})$
0/1
$\mathbf{z}$
Generator
$G(\mathbf{z})$
$\mathbf{x}'$
VAE: maximize variational lower bound
$\mathbf{x}$
Encoder
$q_\phi(\mathbf{z}|\mathbf{x})$
$\mathbf{z}$
Decoder
$p_\theta(\mathbf{x}|\mathbf{z})$
$\mathbf{x}'$
Flow-based models: Invertible transform of distributions
$\mathbf{x}$
Flow
$f(\mathbf{x})$
$\mathbf{z}$
Inverse
$f^{-1}(\mathbf{z})$
$\mathbf{x}'$
Diffusion models: Gradually add Gaussian noise and then reverse
$\mathbf{x}_0$
$\mathbf{x}_1$
$\mathbf{x}_2$
… …
$\mathbf{z}$

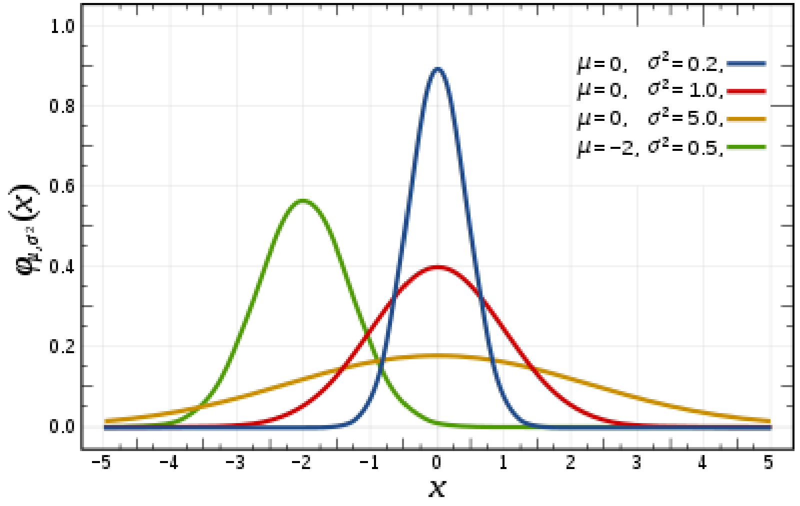
正态分布 – 一元正态分布

正态分布或**高斯分布**是实值随机变量的一种连续概率分布。其概率密度函数的一般形式为：

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

参数 μ 是分布的均值或期望，参数 σ 是标准差。方差是 σ^2 。

符号：$X \sim \mathcal{N}(\mu, \sigma^2)$



标准正态分布

最简单的正态分布称为标准正态分布或单位正态分布。这是特殊情况 μ=0 和 σ=1，它由概率密度函数（或密度） 描述：

$$\varphi(z) = \frac{e^{-z^2/2}}{\sqrt{2\pi}}$$

符号：$X \sim \mathcal{N}(0,1)$

正态分布 – 独立多元正态分布

假设n个变量互不相关，服从正态分布（维度不相关多元正态分布），根据联合概率密度公式有：

$$f(x)=p(x_1,x_2.\dots x_n)=p(x_1)p(x_2).\dots p(x_n)=\frac{1}{(\sqrt{2\pi})^n\sigma_1\sigma_2\cdots\sigma_n}e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}\dots-\frac{(x_n-\mu_n)^2}{2\sigma_n^2}}$$

令：$z^2=\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\cdots+\frac{(x_n-\mu_n)^2}{\sigma_n^2},\sigma_z=\sigma_1\sigma_2\cdots\sigma_n$

则有：

$$f(z)=\frac{1}{(\sqrt{2\pi})^n\sigma_z}e^{-\frac{z^2}{2}}$$

转换成矩阵形式：

$$z^2=z^{\mathrm{T}}z=[x_1-\mu_1,x_2-\mu_2,\cdots,x_n-\mu_n]\begin{bmatrix}\frac{1}{\sigma_1^2} & 0 & \cdots & 0\\ 0 & \frac{1}{\sigma_2^2} & \cdots & 0\\ \vdots & \cdots & \cdots & \vdots\\ 0 & 0 & \cdots & \frac{1}{\sigma_n^2}\end{bmatrix}[x_1-\mu_1,x_2-\mu_2,\cdots,x_n-\mu_n]^{\mathrm{T}}$$

正态分布 – 独立多元正态分布

各个维度的变量： $x=[x_1,x_2,\cdots,x_n]^{\mathrm{T}}$
各个维度的均值： $E(x)=\mu_x=[\mu_1,\mu_2,\cdots,\mu_n]^{\mathrm{T}}$,
各个维度的方差： $\sigma(x)=[\sigma_1,\sigma_2,\cdots,\sigma_n]^{\mathrm{T}}$

则有：

$$x-\mu_x=[x_1-\mu_1,x_2-\mu_2,\cdots,x_n-\mu_n]^{\mathrm{T}}$$

Σ 代表变量 X 的协方差矩阵， i行j列的元素值表示 x_i 与 x_j 的协方差。

$$\sum=\begin{bmatrix}\sigma_1^2 & 0 & \cdots & 0\\ 0 & \sigma_2^2 & \cdots & 0\\ \vdots & \cdots & \cdots & \vdots\\ 0 & 0 & \cdots & \sigma_n^2\end{bmatrix}$$

由于各变量之间是相互独立的，所以只有对角线上 (i = j) 存在元素，其他都为 0，即 **Σ** 是一个对角阵，且 x_i 与它本身的协方差就等于方差。根据对角矩阵的性质，它的逆矩阵：

$$\Sigma^{-1}=\begin{bmatrix}\frac{1}{\sigma_1^2} & 0 & \cdots & 0\\ 0 & \frac{1}{\sigma_2^2} & \cdots & 0\\ & \cdots & \cdots & \vdots\\ 0 & 0 & \cdots & \frac{1}{\sigma_n^2}\end{bmatrix}$$

正态分布 – 独立多元正态分布

由于对角矩阵的行列式等于对角元素的乘积，所以有：

$$\sigma_z = |\Sigma|^{\frac{1}{2}} = \sigma_1\sigma_2.\dots\dots\sigma_n$$

$$z^{\mathrm{T}}z = (x-\mu_x)^{\mathrm{T}}\Sigma^{-1}(x-\mu_x)$$

多元高斯正态分布形式为：

$$f(z) = \mathcal{N}(\mathbf{x}|\mu,\mathbf{\Sigma}) = \frac{1}{(\sqrt{2\pi})^n\sigma_z}e^{-\frac{z^2}{2}} = \frac{1}{(\sqrt{2\pi})^n|\Sigma|^{\frac{1}{2}}}e^{-\frac{(x-\mu_x)^{\mathrm{T}}(\Sigma)^{-1}(x-\mu_x)}{2}}$$

其中 μ 是一个 n 维均值向量， $\mathbf{\Sigma}$ 是 n x n 的协方差矩阵，并且 $|\mathbf{\Sigma}|$ 表示 $\mathbf{\Sigma}$ 的行列式。

符号：$X \sim \mathcal{N}(\mu,\mathbf{\Sigma})$

多元标准正态分布：

$$f(z) = \mathcal{N}(\mathbf{x}|\mu,\mathbf{\Sigma}) = \frac{1}{(\sqrt{2\pi})^n}e^{-\frac{x^{\mathrm{T}}x}{2}}$$

符号：$X \sim \mathcal{N}(\mathbf{0},\mathbf{I})$

马尔科夫链

马尔可夫链（Markov chain），又称离散时间马尔可夫链（discrete-time Markov chain），因俄国数学家安德烈·马尔可夫得名，为状态空间中经过从一个状态到另一个状态的转换的随机过程。该过程要求具备“无记忆”的性质：下一状态的概率分布只能由当前状态决定，在时间序列中它前面的事件均与之无关。这种特定类型的“无记忆性”称作**马尔可夫性质。**
用一句话来概括马尔科夫链的话，那就是某一时刻状态转移的概率只依赖于它的前一个状态。
例如，等红绿灯时，假设你每一秒看一次灯，有了当前这一秒看到红绿灯的状态，就不需要前面看到的所有状态了。

假设状态序列为

$$X_{t-2}, X_{t-1}, X_t, X_{t+1}, X_{t+2}, \cdots,$$

由马尔科夫链定义可知，时刻 t+1 的状态 X_{t+1} 只与 X_t 有关，用数学公式来描述就是：

$$P(X_{t+1}|X_t) = P(X_{t+1}|\cdots|X_{t-2}|X_{t-1}|X_t)$$

马尔科夫链 - 状态转移概率矩阵

对于一个马尔可夫状态 s 及其后续状态 s'，状态转移概率 定义为：

$$P_{ss'} = P[X_{t+1} = s' | X_t = s]$$

状态转移概率矩阵 P 定义为从所有的状态 s 到所有的后续状态 s' 的转移概率：

$$\boldsymbol{P} = \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1n} \\ P_{21} & P_{22} & \dots & P_{2n} \\ \dots & & & \\ \dots & & & \\ P_{n1} & P_{n2} & \dots & P_{nn} \end{bmatrix}$$

矩阵中行号表示当前状态 s，列号表示到达的后续状态 s'。由概率分布的性质可得，转移矩阵是一个正定矩阵，且每行元素之和等于1。

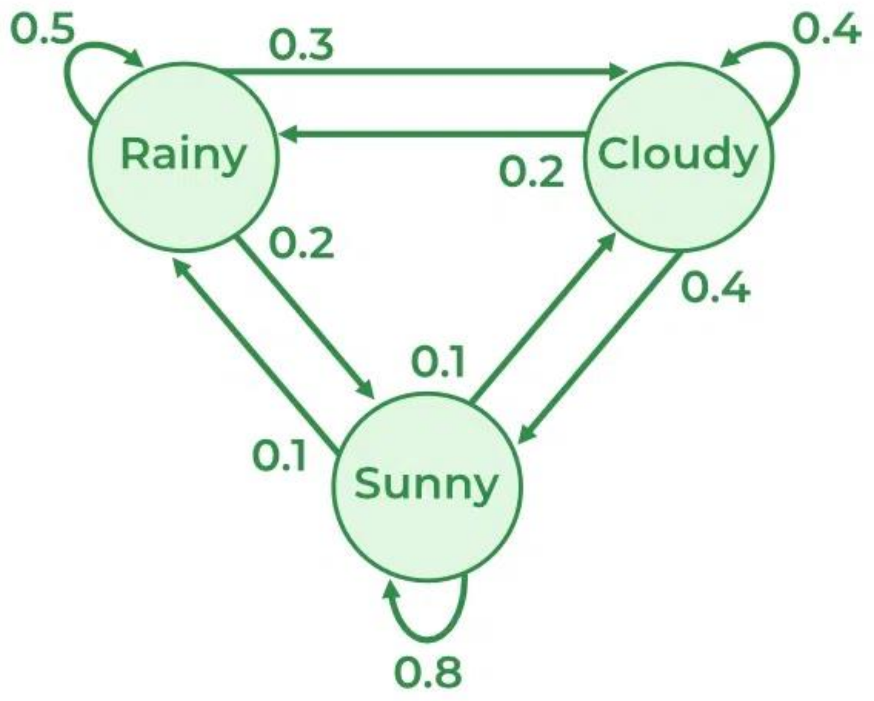
$$\forall i,j: P_{i,j} > 0, \forall i: \sum_j P_{i,j} = 1$$

按相同的方式也可定义n-步转移矩阵： 由n-步转移概率的性质（Chapman–Kolmogorov等式）可知，n-步转移矩阵是其之前所有转移矩阵的连续矩阵乘法：

$$\boldsymbol{P}^{(t)} = \boldsymbol{P}^{(t-1)} \dots \boldsymbol{P}^{(1)} \boldsymbol{I}$$

马尔科夫链 - 状态转移概率矩阵

假设我们有三种状态：晴天、多云和雨天。这个马尔可夫链的转移矩阵可能如下所示：



	sunny	cloudy	rainy
sunny	0.8	0.1	0.1
cloudy	0.4	0.4	0.2
rainy	0.2	0.3	0.5

重参数(reparameterization trick)

重参数化技巧（Reparameterization Trick）是生成模型中一种关键的方法，它允许我们在优化过程中将样本的随机性转移到一个确定性的模型中，使得我们可以通过标准的梯度下降方法来训练模型。

正态高斯分布的线性变换仍然是正态高斯分布：核心思想是将随机变量的采样过程重新参数化为一个确定性的操作，使得采样过程变得可导。以正态分布为例，我们可以**将从正态分布中采样的操作重新参数化为从一个固定的标准正态分布中采样**，**然后通过线性变换和平移**来获得我们所需的正态分布采样值。这样，整个采样过程就变成了可导的，可以直接应用于神经网络的反向传播算法中。

数学表示

如果从高斯分布 $z \sim \mathcal{N}(\mu, \sigma^2)$采样一个 z，我们可以将其重参数化为：

$$z = \mu + \sigma \odot \epsilon, \epsilon \sim \mathcal{N}(\mathbf{0}, 1)$$

其中，μ 和 σ 是学习到的参数，ϵ 是从标准正态分布 $\mathcal{N}(\mathbf{0}, \mathbf{I})$ 采样的噪声， $\odot$ 表示逐元素乘法。

N维数学表示

如果从高斯分布 $z \sim \mathcal{N}(\mu, \mathbf{\Sigma})$采样一个 z，我们可以将其重参数化为：

$$z = \mu + \Sigma^{\frac{1}{2}} \odot \epsilon, \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

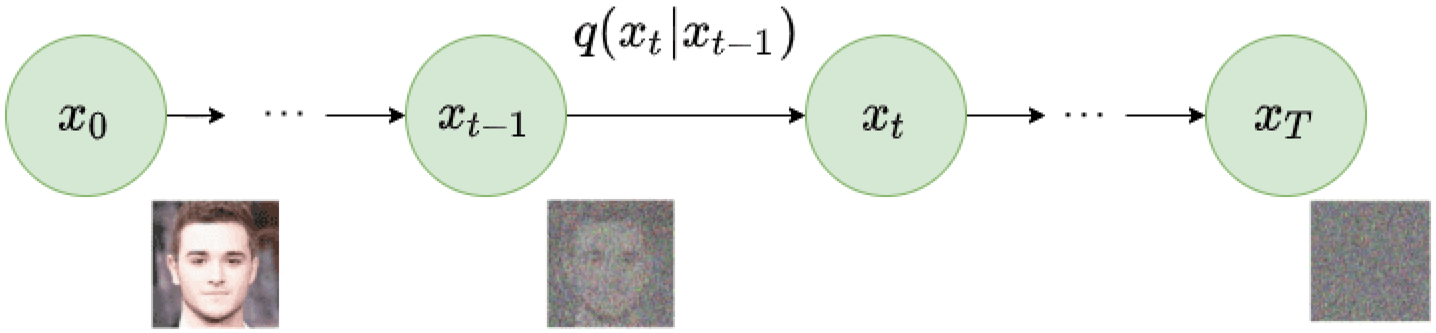
前向扩散过程

给定真实图片样本 $\mathbf{x}_0 \sim q(\mathbf{x})$，让我们定义一个前向扩散过程，我们向样本中添加少量高斯噪声时间步骤，产生一系列噪声样本 $\mathbf{x}_1, \dots, \mathbf{x}_T$。步长由高斯分布方差的超参数 $\{\beta_t \in (0,1)\}_{t=1}^T$ 控制。前向过程由于时刻 t 只与 t-1 时刻有关，所以可以看做是马尔科夫过程：

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \boldsymbol{\mu}_t = \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \boldsymbol{\Sigma}_t = \beta_t \mathbf{I})$$

$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) = \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1})$$

数据样本 $\mathbf{x}_0$ 随着t增大， $\mathbf{x}_t$ 越来越进阶纯噪声。当时间 T→无穷大， $\mathbf{x}_T$ 相当于各向同性高斯分布。



前向扩散过程

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \boldsymbol{\mu}_t = \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \boldsymbol{\Sigma}_t = \beta_t\mathbf{I})$$

根据重参数公式：

$$z = \boldsymbol{\mu}_t + {\boldsymbol{\Sigma}_t}^{\frac{1}{2}} \odot \epsilon, \epsilon \sim \mathcal{N}(0, \mathbf{I})$$

可得：

$$\mathbf{x}_t = q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \boldsymbol{\mu}_t + {\boldsymbol{\Sigma}_t}^{\frac{1}{2}} \odot \epsilon_{t-1}$$

$$= \sqrt{1-\beta_t}\mathbf{x}_{t-1} + \sqrt{\beta_t\mathbf{I}}\epsilon_{t-1}$$

$$= \sqrt{1-\beta_t}\mathbf{x}_{t-1} + \sqrt{\beta_t}\epsilon_{t-1}$$

令 $\alpha_t = 1 - \beta_t$，则有：

$$\mathbf{x}_t = \sqrt{\alpha_t}\mathbf{x}_{t-1} + \sqrt{1-\alpha_t}\epsilon_{t-1}$$

前向扩散过程

令 $\overline{\alpha_t} = \prod_{i=1}^{t} \alpha_i$，则有：

$$\mathbf{x}_t = \sqrt{\alpha_t}\mathbf{x}_{t-1} + \sqrt{1-\alpha_t}\epsilon_{t-1}$$

$$\mathbf{x}_{t-1} = \sqrt{\alpha_{t-1}}\mathbf{x}_{t-2} + \sqrt{1-\alpha_{t-1}}\epsilon_{t-2}$$

$$= \sqrt{\alpha_t}\left(\sqrt{\alpha_{t-1}}\mathbf{x}_{t-2} + \sqrt{1-\alpha_{t-1}}\epsilon_{t-2}\right) + \sqrt{1-\alpha_t}\epsilon_{t-1}$$

$$= \sqrt{\alpha_t\alpha_{t-1}}\mathbf{x}_{t-2} + \sqrt{\alpha_t}\sqrt{1-\alpha_{t-1}}\epsilon_{t-2} + \sqrt{1-\alpha_t}\epsilon_{t-1}$$

由于： $\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \cdots \sim \mathcal{N}(0, \mathbf{I})$，**且正态分布的线性变换仍然是正态分布**

所以：

- $\sqrt{\alpha_t}\sqrt{1-\alpha_{t-1}}\epsilon_{t-2}$ 满足正态分布 $\mathcal{N}(0, \boldsymbol{\Sigma}_1)$，其中 $\boldsymbol{\Sigma}_1 = (\sqrt{\alpha_t}\sqrt{1-\alpha_{t-1}})^2\mathbf{I}$
- $\sqrt{1-\alpha_t}\epsilon_{t-1}$ 满足正态分布 $\mathcal{N}(0, \boldsymbol{\Sigma}_2)$，其中 $\boldsymbol{\Sigma}_2 = (\sqrt{1-\alpha_t})^2\mathbf{I}$

由于两个独立高斯分布可加性，即： $\mathcal{N}(\mu, \boldsymbol{\Sigma}_1) + \mathcal{N}(\mu, \boldsymbol{\Sigma}_2) \sim \mathcal{N}(0, \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)$

所以有： $\sqrt{\alpha_t}\sqrt{1-\alpha_{t-1}}\epsilon_{t-2} + \sqrt{1-\alpha_t}\epsilon_{t-1}$ 满足正态分布 $\mathcal{N}(0, \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)$

其中： $\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 = \alpha_t(1-\alpha_{t-1})\mathbf{I} + (1-\alpha_t)\mathbf{I} = (1-\alpha_t\alpha_{t-1})\mathbf{I}$

令： $\bar{\boldsymbol{\epsilon}}_{t-2} \sim \mathcal{N}(0, (1-\alpha_t\alpha_{t-1})\mathbf{I})$

前向扩散过程

$$\mathbf{x}_t = \sqrt{\alpha_t\alpha_{t-1}}\mathbf{x}_{t-2} + \sqrt{\alpha_t}\sqrt{1-\alpha_{t-1}}\epsilon_{t-2} + \sqrt{1-\alpha_t}\epsilon_{t-1}$$

由于： $\bar{\boldsymbol{\epsilon}}_{t-2} \sim \mathcal{N}(0, (1-\alpha_t\alpha_{t-1})\mathbf{I})$

所以有： $\mathbf{x}_t = \sqrt{\alpha_t\alpha_{t-1}}\mathbf{x}_{t-2} + \sqrt{1-\alpha_t\alpha_{t-1}}\bar{\boldsymbol{\epsilon}}_{t-2}$

令 $\overline{\boldsymbol{\alpha}_t} = \prod_{i=1}^{t} \boldsymbol{\alpha}_i$，则有：

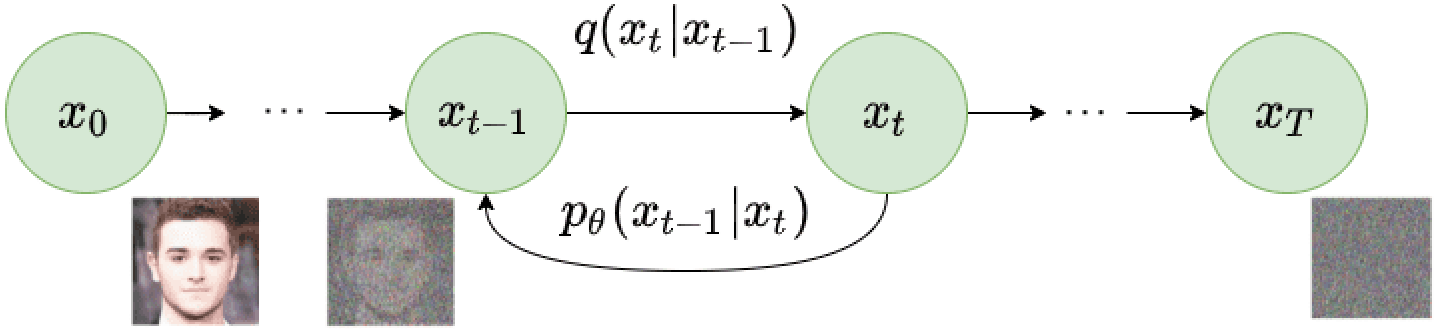
$$\begin{aligned}\mathbf{x}_t &= \sqrt{\alpha_t\alpha_{t-1}}\mathbf{x}_{t-2} + \sqrt{1-\alpha_t\alpha_{t-1}}\bar{\boldsymbol{\epsilon}}_{t-2} \\ &= \cdots \\ &= \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}\end{aligned}$$

因此任意时刻的 $\mathbf{x}_t$ 满足： $q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I})$　　$\boldsymbol{\alpha}_t = \mathbf{1} - \boldsymbol{\beta}_t$

β_t 随着 t 增大是递增的，即 $\beta_1 < \beta_2 < \cdots < \beta_t$，也就是 $\bar{\alpha}_1 > \cdots > \bar{\alpha}_T$。
在算法中， β_t 是由 0.0001 到 0.02 线性插值(以 T = 1000 为基准， T 增加， β_t 对应增大)。

逆扩散过程

逆扩散过程的目标是从最终的纯噪声状态恢复到初始的清晰数据。这个过程通过一个学习到的模型逐步去除噪声。形式上，从噪声数据 x_T 开始，通过一系列逆步骤逐渐移除噪声，生成更加清晰的中间状态 $x_{T-1}, x_{T-2}, \ldots, x_0$。



如果说前向过程 (forward) 是加噪的过程，那么逆向过程(reverse)就是diffusion 的去噪推断过程。
如果我们能够逆转上述过程，就可以从高斯噪声 $x_T \sim \mathcal{N}(0, \mathbf{I})$ 还原出原图分布$x_0 \sim q(x)$。
然而我们无法简单推断 $q(x_{t-1}|x_t)$，因此我们使用深度学习模型去预测这样的一个逆向的分布 p_θ:

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$$

$$p_\theta(X_{0:T}) = p(x_T) \prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)$$

逆扩散过程

贝叶斯公式：

$$P(A \mid B) = \frac{P(A)P(B \mid A)}{P(B)}$$

$$P(AB) = P(A)P(B|A)$$

$$P(ABC) = P(A)P(B|A)P(C|AB)$$

$$P(A|B) = \frac{P(AB)}{P(B)}$$

多变量贝叶斯条件概率公式：

$$P(A|B,C) = \frac{P(A)P(B|A)P(C|A,B)}{P(B)P(C|B)}$$

逆扩散过程

虽然我们无法简单推断 $q(x_{t-1}|x_t)$，但是如果知道 $x_0, q(x_{t-1}|x_t, x_0)$ 就可以直接写出：

$$q(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}; \tilde{\mu}(x_t, x_0), \tilde{\beta}_t \mathbf{I})$$

$$\begin{aligned}
q(x_{t-1}|x_t, x_0) &= \frac{q(x_t, x_0, x_{t-1})}{q(x_t, x_0)} \\
&= \frac{q(x_0)q(x_{t-1}|x_0)q(x_t|x_{t-1}|x_0)}{q(x_0)q(x_t|x_0)} \\
&= q(x_t|x_{t-1}|x_0)\frac{q(x_{t-1}|x_0)}{q(x_t|x_0)} \\
&\propto \exp(-\frac{1}{2}(\frac{(\mathbf{x}_t - \sqrt{\alpha_t}\mathbf{x}_{t-1})^2}{\beta_t} + \frac{(\mathbf{x}_{t-1} - \sqrt{\bar{\alpha}_{t-1}}\mathbf{x}_0)^2}{1 - \bar{\alpha}_{t-1}} - \frac{(\mathbf{x}_t - \sqrt{\bar{\alpha}_t}\mathbf{x}_0)^2}{1 - \bar{\alpha}_t})) \\
&= \exp(-\frac{1}{2}(\frac{\mathbf{x}_t^2 - 2\sqrt{\alpha_t}\mathbf{x}_t\mathbf{x}_{t-1} + \alpha_t\mathbf{x}_{t-1}^2}{\beta_t} + \frac{\mathbf{x}_{t-1}^2 - 2\sqrt{\bar{\alpha}_{t-1}}\mathbf{x}_0\mathbf{x}_{t-1} + \bar{\alpha}_{t-1}\mathbf{x}_0^2}{1 - \bar{\alpha}_{t-1}} - \frac{(\mathbf{x}_t - \sqrt{\bar{\alpha}_t}\mathbf{x}_0)^2}{1 - \bar{\alpha}_t})) \\
&= \exp\left(-\frac{1}{2}\left(\left(\frac{\alpha_t}{\beta_t} + \frac{1}{1 - \bar{\alpha}_{t-1}}\right)\mathbf{x}_{t-1}^2 - \left(\frac{2\sqrt{\alpha_t}}{\beta_t}\mathbf{x}_t + \frac{2\sqrt{\bar{\alpha}_{t-1}}}{1 - \bar{\alpha}_{t-1}}\mathbf{x}_0\right)\mathbf{x}_{t-1} + C(\mathbf{x}_t, \mathbf{x}_0)\right)\right)
\end{aligned}$$

逆扩散过程

概率密度函数的指数部分展开：

$$\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)=\exp\left(-\frac{1}{2}\left(\frac{1}{\sigma^2}x^2-\frac{2\mu}{\sigma^2}x+\frac{\mu^2}{\sigma^2}\right)\right)$$

根据公式：

$$\exp\left(-\frac{1}{2}\left(\left(\frac{\alpha_t}{\beta_t}+\frac{1}{1-\bar{\alpha}_{t-1}}\right)\mathbf{x}_{t-1}^2-\left(\frac{2\sqrt{\alpha_t}}{\beta_t}\mathbf{x}_t+\frac{2\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}\mathbf{x}_0\right)\mathbf{x}_{t-1}+C(\mathbf{x}_t,\mathbf{x}_0)\right)\right)$$

$$q(x_{t-1}|x_t,x_0)=\mathcal{N}(x_{t-1};\tilde{\mu}(x_t,x_0),\tilde{\beta}_t\mathbf{I})$$

得到：

$$\tilde{\beta}_t=1/\left(\frac{\alpha_t}{\beta_t}+\frac{1}{1-\bar{\alpha}_{t-1}}\right)=1/\left(\frac{\alpha_t-\bar{\alpha}_t+\beta_t}{\beta_t(1-\bar{\alpha}_{t-1})}\right)=\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\cdot\beta_t$$

$$\begin{aligned}\tilde{\mu}_t(\mathbf{x}_t,\mathbf{x}_0)&=\left(\frac{\sqrt{\alpha_t}}{\beta_t}\mathbf{x}_t+\frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}\mathbf{x}_0\right)/\left(\frac{\alpha_t}{\beta_t}+\frac{1}{1-\bar{\alpha}_{t-1}}\right)\\&=\left(\frac{\sqrt{\alpha_t}}{\beta_t}\mathbf{x}_t+\frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}\mathbf{x}_0\right)\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\cdot\beta_t\\&=\frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t+\frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\mathbf{x}_0\\&=\frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t-\frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_t\right)\end{aligned}$$

$$\mathbf{x}_0=\frac{1}{\sqrt{\bar{\alpha}_t}}(\mathbf{x}_t-\sqrt{1-\bar{\alpha}_t}\epsilon_t)$$


Thank
You