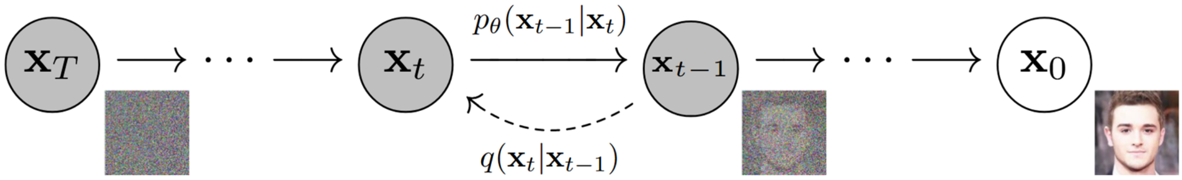
Diffusion Model 图示



扩散过程

1. 给定初始分布 $x_0 \sim q(x)$，添加噪声

$$q(x_t|x_{t-1}) := N(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I)$$

其中$\sqrt{1-\beta_t}x_{t-1}$为均值，β_t为固定方差（仅随时间步t变化）。

2. 令$\alpha_t = 1 - \beta_t$，$\bar{\alpha}_t = \pi_{i=1}^{T}\alpha_i$，取随机变量$z \sim N(0,1)$

运用**重参数化**技巧：

$$\begin{aligned} x_t &= \sqrt{\beta_t}z_{t-1} + \sqrt{1-\beta_t}x_{t-1} \\ &= \sqrt{1-\alpha_t}z_{t-1} + \sqrt{\alpha_t}x_{t-1} \end{aligned}$$

又 $x_{t-1} = \sqrt{\alpha_{t-1}}x_{t-2} + \sqrt{1-\alpha_{t-1}}z_{t-2}$

带入x_t中又：

$$x_t = \sqrt{\alpha_t\alpha_{t-1}}x_{t-2} + \sqrt{1-\alpha_t}z_{t-1} + \sqrt{\alpha_t - \alpha_t\alpha_{t-1}z_{t-2}}$$

其中$\sqrt{1-\alpha_t}z_{t-1} + \sqrt{\alpha_t - \alpha_t\alpha_{t-1}z_{t-2}}$可重参数化为$\bar{z}_{t-2} \sim N(0, \sqrt{1-\alpha_t\alpha_{t-1}})$

因此：

$$\begin{aligned} x_t &= \sqrt{\alpha_t\alpha_{t-1}}x_{t-2} + \sqrt{1-\alpha_t\alpha_{t-1}}\bar{z}_{t-2} \\ &= \ldots(\text{以此类推}) \\ &= \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}z_t \end{aligned}$$

所以有

$$q(x_t|x_0) := N(x_t; \sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)I)$$

反向过程

$p(x_{t-1}|x_t)$也是一个高斯分布，但很难拟合，因此构建一个参数分布来做估计。

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

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

其中$\Sigma_\theta(x_t, t)$设置为与参数θ无关的量σ_t^2

后验的扩散条件概率分布

尽管$p_\theta(x_{t-1}|x_t)$暂时不可求，但$q(x_{t-1}|x_t,x_0)$是可求的。

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

推导过程如下：

$$\begin{aligned} q(x_{t-1}|x_t,x_0) &= \frac{q(x_{t-1},x_t,x_0)}{q(x_t,x_0)} \\ &= \frac{q(x_t|x_{t-1},x_0)q(x_{t-1}|x_0)q(x_0)}{q(x_t|x_0)q(x_0)} \\ &= q(x_t|x_{t-1},x_0)\frac{q(x_{t-1}|x_0)}{q(x_t|x_0)} \end{aligned}$$

由于扩散过程是马尔可夫过程，所以x_t只与x_{t-1}有关，与x_0无关，所以

$$q(x_t|x_{t-1},x_0) = q(x_t|x_{t-1}) := N(x_t;\sqrt{\alpha_t}x_{t-1},\beta_t I)$$
$$q(x_{t-1}|x_0) := N(x_{t-1};\sqrt{\bar{\alpha}_{t-1}}x_0,(1-\bar{\alpha}_{t-1})I)$$
$$q(x_t|x_0) := N(x_t;\sqrt{\bar{\alpha}_t}x_0,(1-\bar{\alpha}_t)I)$$

且高斯分布的概率密度函数为$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

因此

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

所以有：

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

$$\tilde{\mu}_t(x_t,x_0) = (\frac{\sqrt{\alpha_t}}{\beta_t}x_t+\frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}x_0)/(\frac{\alpha_t}{\beta_t}+\frac{1}{1-\bar{\alpha}_{t-1}}) = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t+\frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0$$

将$x_0=\frac{1}{\sqrt{\bar{\alpha}_t}}(x_t-\sqrt{1-\bar{\alpha}_t}z_t)$带入到$\tilde{\mu}_t(x_t,x_0)$中：

$$\tilde{\mu}_t(x_t,x_0) = \frac{1}{\sqrt{\bar{\alpha}_t}}(x_t-\frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\cdot z_t)$$

目标数据分布的负对数似然函数与其上界

$$\begin{aligned} -log\,p_\theta(x_0) &\le -log\,p_\theta(x_0)+D_{KL}(q(x_{1:T}|x_0)||p_\theta(x_{1:T}|x_0) \\ &= -log\,p_\theta(x_0)+E_{x_{1:T}\sim q(x_{1:T}|x_0)}[log\frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})/p_\theta(x_0)}] \\ &= -log\,p_\theta(x_0)+E_{x_{1:T}\sim q(x_{1:T}|x_0)}[log\frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}+log\,p_\theta(x_0)] \end{aligned}$$

由于$log\,p_\theta(x_0)$与$q(x_{1:T}|x_0)$无关，可以移出后与$-log\,p_\theta(x_0)$抵消：

$$-log\,p_\theta(x_0) \le E_{x_{1:T}\sim q(x_{1:T}|x_0)}[log\frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}]$$

不等式两边同时做对于$q(x_0)$的期望，有：

$$L_{VLB} = E_{q(x_{0:T})}[log\frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}] \geq -E_{q(x_0)}log\,p_\theta(x_0)$$

化简:

$$\begin{aligned} L_{VLB} &= E_{q(x_{0:T})}[log\frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}] \\ &= E_q[log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p_\theta(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}] \\ &= E_q[-log\,p_\theta(x_T) + \sum_{t=1}^{T} log\frac{q(x_t|x_{t-1})}{p_\theta(x_{t-1}|x_t)}] \\ &= E_q[-log\,p_\theta(x_T) + \sum_{t=2}^{T} log\frac{q(x_t|x_{t-1})}{p_\theta(x_{t-1}|x_t)} + log\frac{q(x_1|x_0)}{p_\theta(x_0|x_1)}] \end{aligned}$$

由于

$q(x_t|x_{t-1}) = q(x_t|x_{t-1},x_0) = \frac{q(x_t,x_{t-1},x_0)}{q(x_{t-1},x_0)} = \frac{q(x_{t-1}|x_t,x_0)q(x_t|x_0)q(x_0)}{q(x_{t-1}|x_0)q(x_0)} = \frac{q(x_t,x_{t-1},x_0)}{q(x_{t-1},x_0)} = \frac{q(x_{t-1}|x_t,x_0)q(x_t|x_0)}{q(x_{t-1}|x_0)}$

所以:

$$\begin{aligned} L_{VLB} &= E_q[-log\,p_\theta(x_T) + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_t,x_0)}{p_\theta(x_{t-1}|x_t)} \cdot \frac{q(x_t|x_0)}{q(x_{t-1}|x_0)}) + log\frac{q(x_1|x_0)}{p_\theta(x_0|x_1)}] \\ &= E_q[-log\,p_\theta(x_T) + \sum_{t=2}^{T} log\frac{q(x_{t-1}|x_t,x_0)}{p_\theta(x_{t-1}|x_t)} + \sum_{t=2}^{T} log\frac{q(x_t|x_0)}{q(x_{t-1}|x_0)} + log\frac{q(x_1|x_0)}{p_\theta(x_0|x_1)}] \\ &= E_q[-log\,p_\theta(x_T) + \sum_{t=2}^{T} log\frac{q(x_{t-1}|x_t,x_0)}{p_\theta(x_{t-1}|x_t)} + log\frac{q(x_T|x_0)}{q(x_1|x_0)} + log\frac{q(x_1|x_0)}{p_\theta(x_0|x_1)}] \\ &= E_q[log\frac{q(x_T|x_0)}{p_\theta(x_T)} + \sum_{t=2}^{T} log\frac{q(x_{t-1}|x_t,x_0)}{p_\theta(x_{t-1}|x_t)} - log\,p_\theta(x_0|x_1)] \\ &= E_q[D_{KL}(q(x_T|x_0)||p_\theta(x_T) + \sum_{t=2}^{T} D_{KL}(q(x_{t-1}|x_t,x_0)||p_\theta(x_{t-1}|x_t)) - log\,p_\theta(x_0|x_1)] \end{aligned}$$

$L_T = D_{KL}(q(x_T|x_0)||p_\theta(x_T)$: 由于$q(x_T|x_0)$与$p_\theta(x_T)$都选用固定方差，所以$L_T$为常数。

$L_0 = -log\,p_\theta(x_0|x_1)$: 相当于从连续空间到离散空间的解码loss

由高斯分布的KL散度公式知:

对于两个单一变量的高斯分布p和q而言，它们的KL散度为

$$KL(p,q) = log\frac{\sigma_2}{\sigma_1} + \frac{\sigma^2+(\mu_1-\mu_2)^2}{2\sigma_2^2} - \frac{1}{2}$$

$$L_{t-1} = E_q[\frac{1}{2\sigma_t^2}||\tilde{\mu}_t(x_t,x_0) - \mu_\theta(x_t,t)||^2] + C$$

$$\begin{aligned} L_{t-1} - C &= E_{x_0,\epsilon}[\frac{1}{2\sigma_t^2}||\tilde{\mu}_t(x_t(x_0,\epsilon), \frac{1}{\sqrt{\bar{\alpha}_t}}(x_t(x_0,\epsilon) - \sqrt{1-\bar{\alpha}_t}\epsilon)) - \mu_\theta(x_t(x_0,\epsilon),t)||^2] \\ &= E_{x_0,\epsilon}[\frac{1}{2\sigma_t^2}||\frac{1}{\sqrt{\bar{\alpha}_t}}(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \cdot \epsilon) - \mu_\theta(x_t(x_0,\epsilon),t)||^2] \end{aligned}$$

> 依旧使用了重参数化技巧

优化目标为：

$$\mu_\theta(x_t,t)=\tilde{\mu}_t(x_t,\frac{1}{\sqrt{\bar{\alpha}_t}}(x_t-\sqrt{1-\bar{\alpha}_t}\epsilon_\theta(x_t)))=\frac{1}{\sqrt{\bar{\alpha}_t}}(x_t-\frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\cdot\epsilon_\theta(x_t,t))$$

因此

$$L_{simple}(\theta)=E_{t,x_0,\epsilon}[||\epsilon-\epsilon_\theta(\sqrt{\bar{\alpha}_t}x_0+\sqrt{1-\bar{\alpha}_t}\epsilon,t)||]$$