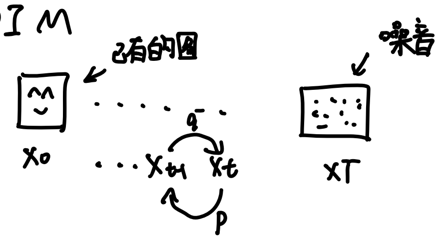
DDPM与DDIM

对于DDPM



★ $q(x_t|x_0) = N(x_t; \sqrt{\bar\alpha_t}\, x_0, (1-\bar\alpha_t) I)$

加噪

$$x_t = \sqrt{\bar\alpha_t}\, x_0 + \sqrt{1-\bar\alpha_t}\cdot\epsilon \qquad \epsilon \sim N(0,1)$$

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

马尔可夫 || 过程

★ $q(x_{t-1}|x_t, x_0) = \frac{q(x_t|x_{t-1})\cdot q(x_{t-1}|x_0)}{q(x_t|x_0)} = N(x_{t-1}; \mu(x_t, x_0), \sigma_t^2 I)$ 确定的

采样(生成)

x_0未知. 用$\hat{x}_0$估计

$q(x_t|x_{t-1}) = N(x_t; \sqrt{\alpha_t}\, x_{t-1}, (1-\alpha_t) I)$ => 可推导的. $\hat{x}_0 = \frac{1}{\sqrt{\bar\alpha_t}}(x_t - \sqrt{1-\bar\alpha_t}\,\epsilon_\theta)$

用Unet的估计值

其中, $\mu(x_t, x_0) = \frac{\sqrt{\alpha_t}(1-\bar\alpha_{t-1})}{1-\bar\alpha_t} x_t + \frac{\sqrt{\bar\alpha_{t-1}}\,\beta_t}{1-\bar\alpha_t} x_0$

$$\sigma_t^2 = \frac{1-\bar\alpha_{t-1}}{1-\bar\alpha_t}\cdot\beta_t$$

对于DDPM采样过程: ① $\hat{x}_{0|t} = \frac{1}{\sqrt{\bar\alpha_t}}(x_t - \sqrt{1-\bar\alpha_t}\,\epsilon_\theta(x_t, t))$

② $x_{t-1} = \mu(x_t, \hat{x}_{0|t}) + \sigma_t\cdot\epsilon \qquad \epsilon \sim N(0,1)$

对于 DDIM

引入： DDPM采样 $q(x_{t-1}|x_t, x_0) = \sqrt{\bar\alpha_{t-1}}\,\overset{未知}{x_0} + \sqrt{1-\bar\alpha_{t-1}}\cdot\varepsilon \quad \varepsilon \sim N(0,1)$

$$= \sqrt{\bar\alpha_{t-1}} \cdot \frac{1}{\sqrt{\bar\alpha_t}}\left(x_t - \sqrt{1-\bar\alpha_t}\,\varepsilon_\theta(x_t,t)\right) + \sqrt{1-\bar\alpha_{t-1}}\,\varepsilon$$

$$= \sqrt{\bar\alpha_{t-1}}\,\hat{x}_{0|t} + \sqrt{1-\bar\alpha_{t-1}}\,\varepsilon$$

$$= \sqrt{\bar\alpha_{t-1}}\,\hat{x}_{0|t} + \sqrt{1-\bar\alpha_{t-1}}\cdot\varepsilon_\theta(x_t,t)$$

为了方差相同

$$= \sqrt{\bar\alpha_{t-1}}\,\hat{x}_{0|t} + \sqrt{1-\bar\alpha_{t-1}-\sigma_t^2}\,\varepsilon_\theta + \sigma_t\cdot\varepsilon$$

但是，中间与 x_t 无关 即将所有 t 换为任一 τ，式子也成立

$q(x_s|x_k, x_0) \Rightarrow x_s = \sqrt{\bar\alpha_s}\,\hat{x}_{0|k} + \sqrt{1-\bar\alpha_s-\sigma_k^2}\,\varepsilon_\theta(x_k,k) + \sigma_k\cdot\varepsilon$

DDPM： T = 1000 999 ... 3 2 1 0

DDIM： T = 1000 900 800 .. 200 100 0

即 $q(x_t|x_{t-1})$ 有些多余 ⟹ 并无严格要求 要从 $t-1 \to t \to t+1$

例如：① 训练时

$N(x_t; \sqrt{\alpha_t}\,x_{t-1}, (1-\alpha_t)I)$ 但只是中间出现，作为推导过程

↓

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

② 采样时

$$\hat{x}_{0|t} = \frac{1}{\sqrt{\bar\alpha_t}}\left(x_t - \sqrt{1-\bar\alpha_t}\,\varepsilon_\theta(x_t,t)\right)$$

只用了 $q(x_t|x_0)$

$$\mathrm{DDIM} \underset{q(x_t|x_{t-1})}{\overset{q(x_t|x_{t-1})}{\rightleftarrows}} \mathrm{DDPM}$$

为什么DDIM能直接用DDPM训练好的模型

DDPM: $\bar{\alpha}_1$ $\bar{\alpha}_2$ $\cdots$ $\bar{\alpha}_t$ $\cdots$ $\bar{\alpha}_T$

$\downarrow$

$x_0 \to x_t$

$\downarrow$

训练目标 $\epsilon_\theta(x_1,1)$ $\epsilon_\theta(x_2,2)\cdots$ $\epsilon_\theta(x_t,t)\cdots$ $\epsilon_\theta(x_T,T)$

DDIM $\bar{\alpha}_{100}$, $\bar{\alpha}_{200}$, $\cdots$ $\bar{\alpha}_T$

$\epsilon_\theta(x_{100},100)$ $\epsilon_\theta(x_{200},200)$ $\epsilon_\theta(x_T,T)$

常微分方程

$$x_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\, x_0 + \sqrt{1-\bar{\alpha}_{t-1}-\sigma_t^2}\, \epsilon_\theta(x_t,t) + \sigma_t \epsilon$$