DDPM

1. Forward:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}\, x_{t-1}, \beta_t I) \qquad \beta_t \in (0,1) \qquad (1)$$

$$q(x_{1:T}|x_0) = \prod_{t=1}^{T} q(x_t|x_{t-1})$$

$\Rightarrow$ define $\alpha_t = 1-\beta_t$. $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$

$$x_t = \sqrt{\alpha_t}\, x_{t-1} + \sqrt{1-\alpha_t}\, z_{t-1}, \quad z_t \sim \mathcal{N}(0, I)$$

$$= \sqrt{\alpha_t \alpha_{t-1}}\, x_{t-2} + \sqrt{\alpha_t} \cdot \underbrace{\sqrt{1-\alpha_{t-1}}\, z_{t-2}}_{\text{standard normal}} + \underbrace{\sqrt{1-\alpha_t}\, z_{t-1}}_{\text{standard normal.}}$$

merge two gaussian

$$= \sqrt{\alpha_t \alpha_{t-1}}\, x_{t-2} + \sqrt{1-\alpha_t\alpha_{t-1}}\, \bar{z}_{t-2}, \quad \bar{z}_{t-2} \sim \mathcal{N}(0, I)$$

$$= \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, z$$

$$\Rightarrow q(x_t|x_0) = \mathcal{N}(x_0; \sqrt{\bar{\alpha}_t}, (I-\bar{\alpha}_t) I) \qquad (2)$$

also $0 < \beta_1 < \beta_2 < \cdots < \beta_T < 1$, hence $\bar{\alpha}_1 > \bar{\alpha}_2 > \cdots > \bar{\alpha}_T$

2. Reverse. (learn p_θ)

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

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

$$q(x_{t-1}|x_t, x_0) = q(x_t|x_{t-1}, x_0) \cdot \frac{q(x_{t-1}|x_0)}{q(x_t|x_0)} \quad \text{.... Bayes. ... known.}$$

given

$$\propto \exp\left(-\frac{1}{2}\left(\underbrace{\frac{(x_t-\sqrt{\alpha_t}\, x_{t-1})^2}{\beta_t}}_{x_t|x_{t-1}\ (1)} + \underbrace{\frac{(x_{t-1}-\sqrt{\bar{\alpha}_{t-1}}\, x_0)^2}{1-\bar{\alpha}_{t-1}}}_{x_t|x_0\ (2)} - \underbrace{\frac{(x_t-\sqrt{\bar{\alpha}_t}\, x_0)^2}{1-\bar{\alpha}_t}}_{x_t|x_0\ (2)}\right)\right)$$

$$= \exp\left(-\frac{1}{2}\left(\left(\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar{\alpha}_{t-1}}\right) x_{t-1}^2 - \left(\frac{2\sqrt{\alpha_t}}{\beta_t} x_t + \frac{2\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}} x_0\right) x_{t-1} + C\right)\right)$$

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

$f(x; \mu, 6) \propto \exp(-\frac{1}{2}\frac{(x-\mu)^2}{6^2})$

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

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

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

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

$$= \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} z_t\right) \qquad (3)$$

3. Loss.

Maximize $\log P(x_0)$ => Maximize ELBO

$\boxed{1}$. $$-\log P_\theta(x_0) \le -\log P_\theta(x_0) + KL(q(x_{1:T}|x_0)\,||\,P(x_{1:T}|x_0))$$

$$= -\log P_\theta(x_0) + E_q\left[\log \frac{q(x_{1:T}|x_0)}{p(x_{0:T})/p(x_0)}\right]$$

$$= -\log P_\theta(x_0) + E_q\left[\log \frac{q(x_{1:T}|x_0)}{p(x_{0:T})}\right] + \underbrace{E_{x_{1:T}\sim q}[\log P(x_0)]}_{=\log P(x_0)}$$

$$= E_q\left[\log \frac{q(x_{1:T}|x_0)}{p(x_{0:T})}\right]$$

let $L_{VLB} = E_{q(x_{1:T}|x_0)}\left[\log \frac{q(x_{1:T}|x_0)}{p(x_{0:T})}\right] \geq \log P_\theta(x_0)$

=> $\log P_\theta(x_0) \geq -L_{VLB}$. Lower Bound.

or $\boxed{2}$ same as $\boxed{1}$ $L_{CE} = -E_{q(x_0)} \log P_\theta(x_0)$ minimize cross entropy.

$$= -E_{q(x_0)} \log\left(\int P_\theta(x_{0:T})\,dx_{1:T}\right)$$

$$= -E_{q(x_0)} \log\left(\int q(x_{1:T}|x_0)\frac{P_\theta(x_{0:T})}{q(x_{1:T}|x_0)}\,dx_{1:T}\right)$$

Jensen's inequality $\longrightarrow$ $$\le -E_{q(x_0)} \int q(x_{1:T}|x_0) \log \frac{P_\theta(x_{0:T})}{q(x_{1:T}|x_0)}\,dx_{1:T}$$

$$= -E_{q(x_{0:T})}\left[\log \frac{P_\theta(x_{0:T})}{q(x_{1:T}|x_0)}\right]$$

$$= E_{q(x_{0:T})}\left[\log \frac{q(x_{1:T}|x_0)}{P_\theta(x_{0:T})}\right]$$

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

$$\Rightarrow L_{VLB} = E_q\left[\log \frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p_\theta(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}\right]$$

forward

reverse.

$$= E_q\left[-\log p_\theta(x_T) + \sum_{t=1}^{T} \log \frac{q(x_t|x_{t-1})}{p_\theta(x_{t-1}|x_t)}\right]$$

$$= E_q\left[-\log p_\theta(x_T) + \log \frac{q(x_1|x_0)}{p_\theta(x_0|x_1)} + \sum_{t=2}^{T} \log \frac{q}{p_\theta}\right]$$

Bayes

$$= E_q\left[-\log p_\theta(x_T) + \log \frac{q(x_1|x_0)}{p_\theta(x_0|x_1)} + \sum_{t=2}^{T} \log \frac{q(x_{t-1}|x_t,x_0)\frac{q(x_t|x_0)}{q(x_{t-1}|x_0)}}{p_\theta(x_{t-1}|x_t)}\right]$$

$$= E_q\left[-\log p_\theta(x_T) + \log \frac{q(x_1|x_0)}{p_\theta(x_0|x_1)} + \sum_{t=2}^{T} \log \frac{q}{p} + \boxed{\sum_{t=2}^{T} \log \frac{q(x_t|x_0)}{q(x_{t-1}|x_0)}}\right]$$

$= \log \frac{q(x_T|x_0)}{q(x_1|x_0)}$

$$= E_q\left[\log \frac{q(x_T|x_0)}{p_\theta(x_T)} + \sum_{t=2}^{T} \log \frac{q(x_{t-1}|x_t,x_0)}{p_\theta(x_{t-1}|x_t)} - \log p_\theta(x_0|x_1)\right]$$

$$= E_{q(x_T|x_0)}\left[\log \frac{q(x_T|x_0)}{p_\theta(x_T)}\right] + \sum_{t=2}^{T} E_{q(x_t,x_{t-1}|x_0)}\left[\log \frac{q(x_{t-1}|x_t,x_0)}{p_\theta(x_{t-1}|x_t)}\right] - E_{q(x_1|x_0)} \log p_\theta(x_0|x_1)$$

$$= KL(q(x_T|x_0)\,\|\,p_\theta(x_T)) + E_{q(x_t|x_0)}\left[\sum_{t=2}^{T} KL(q(x_{t-1}|x_t,x_0)\,\|\,p_\theta(x_{t-1}|x_t))\right] - E_{q(x_1|x_0)}\left[\log p_\theta(x_0|x_1)\right]$$

$\Rightarrow L_{VLB} = L_T + \cdots + L_0$ (4)

where $L_T = KL(q(x_T|x_0)\,\|\,p_\theta(x_T))$

denoise.

$L_t = E_{q(x_t|x_0)}\left[KL(q(x_t|x_{t+1},x_0)\,\|\,\boxed{p_\theta(x_t|x_{t+1})})\right]$, $t \in [1, T-1]$

$L_0 = E_{q(x_1|x_0)} - \log p_\theta(x_0|x_1)$ ground truth

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

$\hookrightarrow \mu_\theta(x_t,t) = \frac{1}{\sqrt{\bar\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar\alpha_t}} z_t\right)$

Reverse.
(3)

Due to the "simplification", let $\Sigma_\theta(x_t,t) = \beta_t I$.

Thus: $x_{t-1} = N\left(x_{t-1}; \frac{1}{\sqrt{\bar\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar\alpha_t}}\boxed{z_\theta(x_t,t)}\right), \beta_t I\right)$

model.

Note

$$L_t = KL(q(x_t|x_{t-1}, x_0) \| p_\theta(x_t|x_{t+1}))$$

(4) ELBO

$$KL(N(x|\mu_1,\Sigma_1) \| N(x|\mu_2,\Sigma_2)) = \frac{1}{2}\left[\log\frac{|\Sigma_2|}{|\Sigma_1|} - k + tr(\Sigma_2^{-1}\Sigma_1) + (\mu_1-\mu_2)^T\Sigma_2^{-1}(\mu_1-\mu_2)\right]$$

2 Gaussians KL Distance

$$\Rightarrow L_{simple} = E_{x_0, z_t}\left[\| z_t - z_\theta(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\, z_t, t)\|^2\right] + C.$$

Training.

1. Repeat.
2. $x_0 \sim q(x_0)$
3. $t \sim \text{Uniform}(\{1, \dots, T\})$
4. $\epsilon \sim N(0, I)$
5. Take gradient descent step on
   $\nabla_\theta \|\epsilon - z_\theta(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, t)\|^2$ ⟵ Loss update.
6. until Converge

Sampling

1. $x_T \sim N(0, I)$
2. for $t = T \dots 1$ do
3. $z \sim N(0, I)$ if $t > 1$. else $z = 0$
4. $x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} z_\theta(x_t, t)\right) + \sigma_t z$ ⟵ reverse diffusion
5. end for
6. return x_0.