Denoising Diffusion Probabilistic models (DDPM)

Given data $\mathcal{D} = \{x_0\} \sim p_{x_0}$ (data distribution)

Goal → To learn to sample from p_{x_0}

(Special case of Hierarchical VAE)

$$x \xrightarrow[q(z|x)]{\text{Encoder}} z \xrightarrow[p(x|z)]{\text{Decoder}} x$$

Data — latent space — data

— VAE

HVAE

$$x \rightarrow z_1 \rightarrow z_2 \rightarrow z_3 \rightarrow z_T \quad \text{(Encoding)}$$

$$z_T \rightarrow z_{T-1} \rightarrow z_{T-2} \rightarrow \text{- - - -} \rightarrow x \quad \text{(decoding)}$$

DDPM

- HVAE with 1) Multiple latent spaces
- Dimensionality of latent space same as data space
- encoding process is non-learnable / fixed

VAE → $q_\phi(z/x)$ → learnt ✓

Both encoding & decoding learnt.

DDPM — $q(z/x)$ — fixed / not learnt

only decoding is learned

$x_0(x)$ — Input Space

x_0

← latent space

$(x_1, x_2, x_3 - - - x_T)$ ← $(z_1, z_2 - - - - z_T)$

→ latent spaces/vectors

Forward Process

$x_0 \to x_1 \to x_2 \to x_3 - - - x_T$

input

latent vectors/spaces corresponding to x_0

input

noise

$$\boxed{x_1 = \sqrt{\alpha_1}\, x_0 + \sqrt{1-\alpha_1}\,\epsilon_1}$$

α = scalar

where $\epsilon_1 \sim N(0, I)$

$$x_2 = \sqrt{\alpha_2}\, x_1 + \sqrt{1-\alpha_2}\, \epsilon_2$$

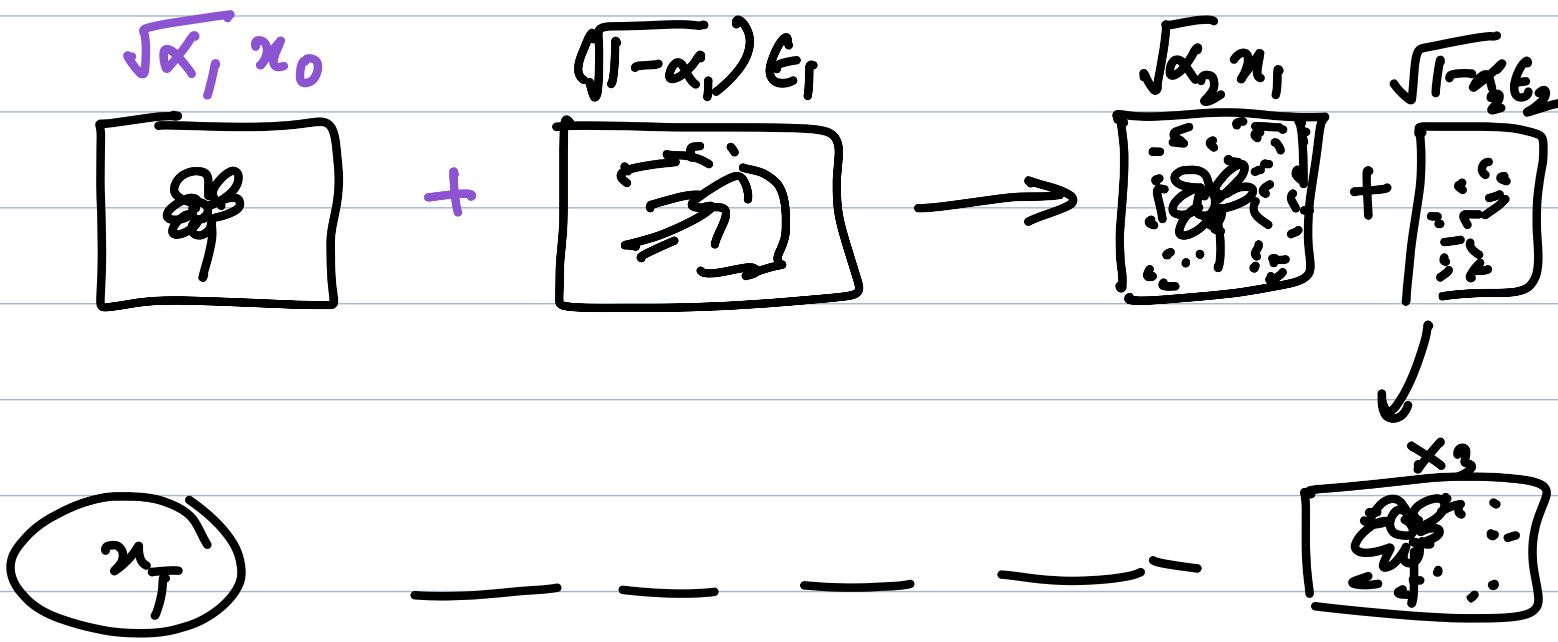
where $\epsilon_2 \sim N(0, I)$

$$x_3 = \sqrt{\alpha_3}\, x_2 + \sqrt{1-\alpha_3}\, \epsilon_3$$

$$x_t = \sqrt{\alpha_t}\, x_{t-1} + \sqrt{1-\alpha_t}\, \epsilon_t$$

$\alpha_1, \alpha_2, \alpha_3$ — — — α_T — are fixed scalars

$[0, 1]$



Forward Process

$$q(x_t \mid x_{t-1}) = \mathcal{N}\left(\underbrace{x_t}_{\text{output}} ;\ \underbrace{\sqrt{\alpha_t}\, x_{t-1}}_{\text{Mean}},\ \underbrace{(1-\alpha_t)I}_{\text{Variance}}\right)$$

Conditional forward encoding.

β_t : small variance schedule

$$\alpha_t + \beta_t = 1$$

$$\boxed{\beta_t = 1 - \alpha_t}$$

Define the Model —

$$p_\theta(\underbrace{x_0}_{\text{data}}, \underbrace{x_1, x_2 \text{ ------ } x_T}_{\text{latent variables}}) = p_\theta(x_T) \prod_{t=1}^{T} p_\theta(x_t \mid x_{t-1})$$

Reverse Process

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

Mean

where $\underbrace{p_\theta}\big(x_{t-1} \,|\, x_t\big) \doteq \mathcal{N}\big(x_{t-1}, \mu_\theta(x_t), \Sigma_\theta(x_t)\big)$

output

variance

ELBO optimization

↓ VAE

VAE

$$\boxed{\log p_\theta(x) = \log \int_z p_\theta(x, z)\, dz}$$

$$\text{loss} \approx J_\theta(q_\phi) = \mathop{E}_{q_\phi(z|x)} \log \frac{p_\theta(x,z)}{q_\theta(z|x)}$$

Evidence Lower Bound (ELBO)

In DDPM

$$\simeq J_\theta(q)^{DDPM} = \mathop{E}_{q(x_1, x_2, \ldots x_T | x_0)} \log \frac{p_\theta(x_0\, x_1 \ldots, x_T)}{q(x_1 x_2 \ldots x_T | x_0)}$$

$$x_{1:T} = (x_1, x_2 \dots x_T)$$

$$x_{0:T} = (x_0, x_1, x_2 \dots x_T)$$

$$\boxed{J_\theta(q)^{DDPM} = \underset{q(x_{1:T}|x_0)}{E} \log \frac{p_\theta(x_{0:T})}{q(x_{1:T}|x_0)}}$$

Optimizing the ELBO for DDPM

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

$$p(x_T) = \mathcal{N}(0, I)$$

Similarly

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

$$q(x_{1:T} \mid q_0) = q\left(\frac{x_1}{x_0}\right) q(x_2 \mid x_1, x_0)$$

$$q(x_3 \mid x_2, x_1, x_0)$$

$$= q(X_T \mid X_{T-1} X_{T-2} \cdots X_0)$$

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

encoding

Forward encoding