19/01/25

Computer Vision

- Introduction to Diffusion Models
  - Markov Chain
  - Forward (inference) trajectory
  - Reverse (generative) trajectory

$q(x^{(0 \cdots T)})$ is the forward distn.

$p_\theta(x^{(0 \cdots T)})$ is the reverse distn.

Ideally $\underbrace{p_\theta(x^{(0)})}_{\text{to be learned}} \sim \underbrace{q(x^{(0)})}_{\text{data dist}}$

- (Markov chain: Both forward and backward trajectories are Markov chains, Specifically first order.)

Goal:

$$\boxed{\max_{\theta \in \Theta} \mathcal{L}(\theta) = E_{x^{(0)} \sim q}\left[\log p_\theta(x^{(0)})\right]}, \quad (1)$$

i.e., we want to maximize the likelihood of generating samples $p_\theta(x^{(0)})$, coming from $q(x^{(0)})$

$$p_\theta(x^{(0)}) = \int \underbrace{p_\theta(x^{(0 \cdots T)})} \, dx^{(1 \cdots T)} \qquad \equiv p_\theta(x^{(0)}, x^{(1)}, \ldots, x^{(T)})$$

$$= \int p_\theta(x^{(T)}) \cdot \prod_{t=1}^{T} p_\theta(x^{(t-1)} \mid x^{(t)}) \, dx^{(1 \cdots T)}$$

(standard result for first order Markov chains) (2)

$$\left[p_\theta(x^{(0 \cdots T)}) = \underbrace{p_\theta(x^{(0)} \mid x^{(1 \cdots T)})}_{\text{(due to Markov property)}} \, p_\theta(x^{(1 \cdots T)}) \right.$$

$$= p_\theta(x^{(0)} \mid x^{(1)}) \cdot p_\theta(x^{(1 \cdots T)})$$

$$\left. = p_\theta(x^{(T)}) \cdot \prod_{t=1}^{T} p(x^{(t-1)} \mid x^{(t)}) \right]$$

Also, we assume $p_\theta(x^{(t-1)} \mid x^{(t)}) \sim \mathcal{N}(x^{(t-1)}; f_\mu(x^{(t)}, \theta), f_\Sigma(x^{(t)}, \theta))$ —③

$$p_\theta(x^{(0)}) = \int p_\theta(x^{(T)}) \cdot \prod_{t=1}^{T} p_\theta(x^{(t-1)} \mid x^{(t)}) \cdot dx^{(1 \cdots T)}$$

$$= \int p_\theta(x^{(T)}) \cdot \prod_{t=1}^{T} p_\theta(x^{(t-1)} \mid x^{(t)}) \cdot \frac{q(x^{(1 \cdots T)} \mid x^{(0)})}{q(x^{(1 \cdots T)} \mid x^{(0)})} \cdot dx^{(1 \cdots T)}$$

$$= \int p_\theta(x^{(T)}) \cdot q(x^{(1 \cdots T)} \mid x^{(0)}) \cdot \frac{\prod_{t=1}^{T} p_\theta(x^{(t-1)} \mid x^{(t)})}{\prod_{t=1}^{T} q(x^{(t)} \mid x^{(t-1)})} \, dx^{(1 \cdots T)}$$