27/01/25

Computer Vision

- Recap:

  - Maximizing log likelihood : $\mathcal{L}(\theta) = E_q[\log p_\theta(x^{(0)})]$ — ①

    Backward process:
  - $p_\theta(x^{(0)})$ evaluated as a marginal : $p_\theta(x^{(0)}) = \int p_\theta(x^{(0\cdots T)}) \cdot dx^{(1\cdots T)}$ — ②

  - Introduce forward process : $p_\theta(x^{(0)}) = \int q(x^{(1\cdots T)} | x^{(0)}) \cdot p_\theta(x^{(T)}) \cdot \frac{\prod p_\theta(x^{(t-1)}|x^{(t)})}{\prod q(x^{(t)}|x^{(t-1)})} \cdot dx^{(1\cdots T)}$

    (due to Markov chain assumption). — ③

  - Plug ③ into ① : $\mathcal{L}(\theta) = \int q(x^{(0)}) \left[\log \int q(x^{(1\cdots T)} | x^{(0)}) \cdot p_\theta(x^{(T)}) \cdot \frac{\prod p_\theta(x^{(t-1)}|x^{(t)})}{\prod q(x^{(t)}|x^{(t-1)})} dx^{(1\cdots T)}\right] \cdot dx^{(0)}$

    — ④

  - Lower bound $\mathcal{L}(\theta)$ in terms of $K(\theta)$: $\mathcal{L}(\theta) \geq K(\theta)$,

$$K(\theta) = \int q(x^{(0\cdots T)}) \cdot \log\left[p_\theta(x^{(T)}) \cdot \frac{\prod p_\theta(x^{(t-1)}|x^{(t)})}{\prod q(x^{(t)}|x^{(t-1)})}\right] dx^{(0\cdots T)}$$

$$= \int q(x^{(0\cdots T)}) \cdot \log\left[p_\theta(x^{(T)}) \cdot \frac{\prod p_\theta(x^{(t-1)}|x^{(t)})}{\prod q(x^{(t-1)}|x^{(t)}, x^{(0)})} \cdot \frac{q(x^{(t-1)}|x^{(t)})}{q(x^{(t)}|x^{(0)})}\right] \cdot dx^{(0\cdots T)}$$

— ⑤

  - After a bit of work we get:

$$K(\theta) = -\sum_{k=2}^{T} \int KL\left(q(x^{(t-1)}|x^{(t)}, x^{(0)}) \,\middle\|\, p_\theta(x^{(t-1)}|x^{(t)})\right) \cdot q(x^{(t)}, x^{(0)}) \cdot dx^{(t)} \cdot dx^{(0)}$$

$$+ H_q(x^{(T)}|x^{(0)}) - H_q(x^{(1)}|x^{(0)}) - H_p(x^{(T)}) .$$ — ⑥

- <u>Specifics</u>: $q(x^{(t)}|x^{(t-1)}) = \mathcal{N}(x^{(t)}; \sqrt{1-\beta_t}\cdot x^{(t-1)}, \beta_t I)$ — (7)

$$p_\theta(x^{(t-1)}|x^{(t)}) = \mathcal{N}(x^{(t-1)}; f_\mu(x^{(t)}, t; \theta_\mu), f_\Sigma(x^{(t)}, t; \theta_\Sigma)). \quad -(8)$$

(This is a consequence of the diffusion process)

- β_t is assumed to be deterministic.

- Further, the variance of the backward process is assumed to be deterministic as well. i.e., $f_\Sigma(x^{(t)}, t; \theta_\Sigma) = \sigma_t^2 I$

<u>$\Rightarrow$ We end up learning θ_μ</u>

- The lower bound $K(\theta) = \mathbb{E}_q\left[\sum_{t=2}^{T} L_{t-1}\right] + \mathbb{E}_q[L_T] - \mathbb{E}_q[L_0]$

$$L_{t-1} = KL\left(q(x^{(t-1)}|x^{(t)}, x^{(0)}) \,\|\, p_\theta(x^{(t-1)}|x^{(0)})\right)$$

$$L_{t-1} = \frac{1}{2\sigma_t^2}\left\|\tilde{\mu}_t(x^{(t)}, x^{(0)}) - f_\mu(x^{(t)}, t; \theta_\mu)\right\|^2 + \underbrace{C}_{\text{constant}} \quad -(8a)$$

Since the forward and backward trajectories follow (7) and (8)

and $q(x^{(t-1)}|x^{(t)}, x^{(0)}) \sim \mathcal{N}(x^{(t-1)}; \tilde{\mu}(x^{(t)}, x^{(0)}), \tilde{\beta}_t I)$. — (9) (HW)

$$p_\theta(x^{(t-1)}|x^{(t)}) \sim \mathcal{N}(x^{(t-1)}; f_\mu(x^{(t)}, t; \theta_\mu), \sigma_t^2 I). \quad -(10)$$

σ_t^2 is related to β_t.

Today: Denoising Diffusion Model.