17/11/23

Deep Learning

- Recap
- Bound on log likelihood K
- Tractable form of K.
- $q(x^{(t)} | x^{(t-1)}) \sim \mathcal{N}(\sqrt{1-\beta_t} \cdot x^{(t-1)}, \beta_t I)$ —(1)
- $p(x^{(t-1)} | x^{(t)}) \sim \mathcal{N}(f_\mu(x^{(t)}, t), f_\Sigma(x^{(t)}, t))$ —(2)

Markov assumptions

- Q: Why diffusion models? Read Appendix in paper, & paper itself. (6)
- $L = \int q(x^{(0)}) \cdot \log p(x^{(0)}) \, dx^{(0)}$ —(3)
- $p(x^{(0)}) = \int q(x^{(1\cdots T)} | x^{(0)}) \cdot \dfrac{\prod_{t=1}^{T} p(x^{(t-1)} | x^{(t)})}{\prod_{t=1}^{T} q(x^{(t)} | x^{(t-1)})} \, p(x^{(T)}) \, dx^{(1\cdots T)}$ —(4)
- Note: the above expression draws inspiration from nonequilibrium thermodynamics, specifically diffusion processes.
- $\theta = \left\{ \{\beta_t\}_{t=1}^{T} ; \{\theta_\mu^{(t)}\}_{t=1}^{T} \{\theta_\Sigma^{(t)}\}_{t=1}^{T} \right\}$
- Let's plug (4) into (3) and then apply Jensen's inequality
- $L \geqslant K$ where

$$K = \int q(x^{(0\cdots T)}) \cdot \log \left[p(x^{(T)}) \cdot \frac{\prod_{t=1}^{T} p(x^{(t-1)} | x^{(t)})}{\prod_{t=1}^{T} q(x^{(t)} | x^{(t-1)})} \right] dx^{(0\cdots T)} \quad —(5)$$

$$K = -\sum_{k=2}^{T} E_{q(x^{(t)}, x^{(0)})} \left[D\left(q(x^{(t-1)} | x^{(t)}, x^{(0)}) \| p(x^{(t-1)} | x^{(t)}) \right) \right]$$

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

- β_t and $\{\theta_\mu^t, \theta_\Sigma^t\}$ are found by optimizing / maximizing K.