16/11/23

Deep Learning

- Recap
- Diffusion Model Problem Setup
- Forward (Inference) Process:
  - $\underbrace{q(x^{(0)})}_{\text{Input/data distribution}} \Longrightarrow \underbrace{q(x^{(T)})}_{\mathcal{N}(0,I)}$ diffusion process
  - The diffusion process is Markov (first order) (Assumption)

    i.e. $q(x^{(t)} \mid x^{(t-1)}, x^{(t-2)}, \ldots, x^{(0)}) = q(x^{(t)} \mid x^{(t-1)})$ $\quad p(y|x) \sim \mathcal{N}(x, \sigma^2)$
  - $q(x^{(t)} \mid x^{(t-1)}) \sim \mathcal{N}\left(\sqrt{1-\beta_t} \cdot x^{(t-1)}, \beta_t \cdot I\right)$ $\quad y = x + \eta$

$q(x^{(t)} \mid x^{(t-1)})$ is a Gaussian RV whose mean depends on β_t & $x^{(t-1)}$, & whose covariance matrix depends on β_t.

  - β_t is a learnable parameter
  - $q(x^{(0\ldots T)})$ is the joint distribution of the RVs $x^{(0)} \ldots x^{(T)}$

$$q(x^{(0\ldots T)}) = q(x^{(T)} \mid x^{(0\ldots T-1)}) \cdot q(x^{(0\ldots T-1)})$$

$$= q(x^{(T)} \mid x^{(T-1)}) \cdot q(x^{(0\ldots (T-1))})$$

$$\vdots$$

$$q(x^{(0\ldots T)}) = q(x^{(0)}) \cdot \prod_{t=1}^{T} \underbrace{q(x^{(t)} \mid x^{(t-1)})}_{\mathcal{N}(\sqrt{1-\beta_t} \cdot x^{(t-1)}, \beta_t I)} \quad \text{— (1)}$$

- If the time steps a very small, then $q(x^{(t)} \mid x^{(t-1)})$ has an identical functional form as $q(x^{(t-1)} \mid x^{(t)})$ (Feller, 1948) — (*)
- Reverse / Generative process
  - $\underbrace{p(x^{(T)})}_{\mathcal{N}(0,I)} \Longrightarrow \underbrace{p(x^{(0)})}_{\text{Image/Data distribution}}$

- Again, a diffusion process that is first order Markov.
  - $p(x^{(t-1)} | x^{(t)}) \sim \mathcal{N}(f_\mu(x^{(t)}, t), f_\Sigma(x^{(t)}, t))$

- With this formulation, we want to maximize the following:

$$L = \int \underbrace{q(x^{(0)})}_{\text{data distribution}} \cdot \log \underbrace{p(x^{(0)})}_{\text{approx of the data distribution.}} \, dx^{(0)} \qquad \text{(effectively the log likelihood function)} \quad \text{— (A)}$$

- Find the parameters of the forward & reverse processes so that L is max.

- Challenge is that $q(x^{(0)})$ is not known explicitly.

- Let's look at $p(x^{(0)})$: $p(x^{(0)})$ can be found by marginalizing $p(x^{(0 \ldots T)})$ over $x^{(1 \ldots T)}$

$$p(x^{(0)}) = \int p(x^{(0 \ldots T)}) \, dx^{(1 \ldots T)} \quad \text{— (2)}$$

From (1) we can also write $p(x^{(0 \ldots T)})$ as (using the Markov assumption)

$$p(x^{(0 \ldots T)}) = p(x^{(T)}) \prod_{t=1}^{T} p(x^{(t-1)} | x^{(t)}) \quad \text{— (3)}$$

Plug (3) into (2)

$$p(x^{(0)}) = \int p(x^{(T)}) \cdot \prod_{t=1}^{T} p(x^{(t-1)} | x^{(t)}) \, dx^{(1 \ldots T)}$$

$$= \int p(x^{(T)}) \prod_{t=1}^{T} p(x^{(t-1)} | x^{(t)}) \cdot \frac{q(x^{(1 \ldots T)} | x^{(0)})}{q(x^{(1 \ldots T)} | x^{(0)})} \, dx^{(1 \ldots T)}$$

$$= \int p(x^{(T)}) \cdot \underbrace{q(x^{(1 \ldots T)} | x^{(0)}) \cdot \frac{\prod_{t=1}^{T} p(x^{(t-1)} | x^{(t)})}{\prod_{t=1}^{T} q(x^{(t)} | x^{(t-1)})}}_{\text{(follows from (1))}} \, dx^{(1 \ldots T)} \quad \text{— (4)}$$

Now plug (4) into (A)

$$L = \int q(x^{(0)}) \log p(x^{(0)}) \, dx^{(0)} =$$

$$\int q(x^{(0)}) \cdot \log\left[\quad\text{———}\quad ④ \quad\text{———}\quad \right] dx^{(1)}$$

- Jensen's inequality: Let $f : \mathbb{R} \to \mathbb{R}$ be a <u>convex</u> function and let $t \in [0,1]$, then for any $x, y \in \mathbb{R}$,

$$f(tx + (1-t)y) \underline{\leq} t \cdot f(x) + (1-t) \cdot f(y)$$