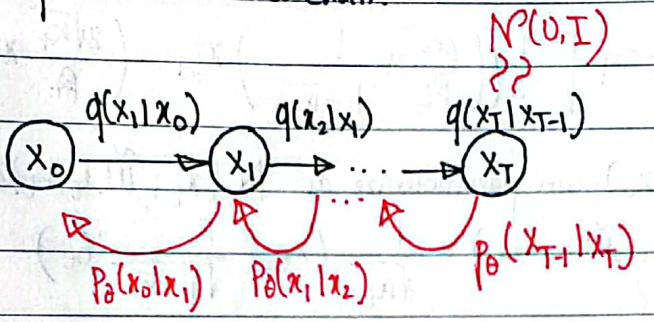
Thứ
Ngày . . No.

Personal Note #1: Diffusion Models for Hand-Writting generation.

(*) Diffusion model re-iteration.

↳ Diffusion model repeatedly distorts an input signal by adding noise after every step. We can define every step as a conditional data distribution $q(x_t|x_{t-1})$. Hence, this can be interpreted as a Markov Chain:



→ Our motive is to learn a model $p_\theta(x_{t-1}|x_t)$ to reverse the Forward Diffusion process to recover our signal.

↳ Define $q(x_t|x_{t-1}) = \mathcal{N}(x_{t-1}; \sqrt{1-\beta_t}\, x_{t-1}, \beta_t I)$

→ We can define $q(x_t|x_0)$ in closed form: Define $1-\beta_t = \alpha_t$

→ Using Re-parameterization we have:

$$\begin{aligned} x_t &= \sqrt{\alpha_t}\, x_{t-1} + \sqrt{1-\alpha_t}\, \varepsilon_{t-1} \\ &= \sqrt{\alpha_t}\left(\sqrt{\alpha_{t-1}}\, x_{t-2} + \sqrt{1-\alpha_{t-1}}\, \varepsilon_{t-2}\right) + \sqrt{1-\alpha_t}\, \varepsilon_{t-1} \\ &= \sqrt{\alpha_t\alpha_{t-1}}\, x_{t-2} + \sqrt{1-\alpha_t\alpha_{t-1}}\, \varepsilon_{t-2} \\ &= \dots \\ &= \sqrt{\alpha_t\alpha_{t-1}\cdots\alpha_1}\, x_0 + \sqrt{1-\alpha_t\alpha_{t-1}\cdots\alpha_1}\, \varepsilon_0 \\ &= \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \varepsilon_0 \quad (\varepsilon_0 \sim \mathcal{N}(0,I)) \end{aligned}$$

Where $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$

(*) $\Rightarrow$ $\boxed{q(x_{0:t}|x_0) = \mathcal{N}(x_0; \sqrt{\bar{\alpha}_t}\, x_0, (1-\bar{\alpha}_t) I)}$ (1)

↳ From (1) We can calculate $q(x_{t-1}|x_t, x_0)$. It's hard to calculate $q(x_{t-1}|x_t)$ directly because of intractibility.

Parameters (10)

Thứ
Ngày . . No.

$$\Rightarrow q(x_{t-1}|x_t,x_0) = q(x_t|x_{t-1},x_0)\frac{q(x_{t-1}|x_0)}{q(x_t|x_0)}$$

* $q(x_t|x_{t-1},x_0) \propto q(x_t|x_{t-1}) \propto q(x_t|x_{t-1})\frac{q(x_{t-1}|x_0)}{q(x_t|x_0)}$

because x_0 is fixed

* Not dependent on x_t & β_t

$$\propto \exp\left[-\frac{1}{2}\left(\left(\frac{\alpha_t}{\beta_t}+\frac{1}{1-\bar{\alpha}_{t-1}}\right)x_{t-1}^2 - \left(\frac{2\sqrt{\alpha_t}}{\beta_t}x_t + \frac{2\sqrt{\alpha_t}}{1-\bar{\alpha}_t}\right)x_{t-1} + C\right]\right]$$

$\Rightarrow q(x_{t-1}|x_t,x_0)$ can parameterize as $\mathcal{N}(x_t;\tilde{\mu}(x_t,\epsilon_t)x_t,\tilde{\beta}_t I)$ where:

$$\begin{cases}\tilde{\mu}(x_t,\epsilon_t) = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_t\right)\\ \tilde{\beta}_t = \beta_t\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\end{cases}$$

↳ Loss function: To optimize $p_\theta(\cdot)$, we use the variational Lower bound:

* We know this!

$$E_{VLB} = \mathbb{E}_q\left[L_T + \sum_{t>1} D_{KL}(q(x_{t-1}|x_t,x_0)\|p_\theta(x_{t-1}|x_t)) + L_0\right]$$

* Basically constant. * We are particularly interested in these KL divergence terms.

↳ The KL divergence terms can be simplified into:

$$L_t^{simple} = \mathbb{E}_{x_0,\epsilon_t}\left[\|\epsilon_t - \epsilon_\theta(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon_t, t)\|^2\right]$$

* Instead of modelling μ_θ directly we model ϵ_θ that predicts the noise ϵ_t.

Thứ
Ngày . . No.

(*) Diffusion Model for Hand Writting generation.

↳ The dataset used in this tuk will be a bit unique, it comprises of 3 components:

→ Handwriting pen-strokes in XML form.

→ Text-sequences describing what is written.

→ Images specifying handwritting style.

*) Pen-strokes:

→ Every datapoint $x_0 \in \mathbb{R}^2 \times \{0,1\}$ comprises of $\{x_1, \ldots, x_N\}$ vectors. Each x_t contains $[(x_i, y_i), d_i]$ ① where (x_i, y_i) specifies the offset coordinates w.r.t previous stroke. ② d_i is a binary value specifying if the pen was down ($d_i = 0$) when writing the stroke.

→ We cannot parameterize d_i as gaussians → We separate x_t into y_t and d_t. Where y_t specifies the strokes & d_t specifies if pen is down or nah.

→ We specify training loss for both pen strokes and pen lifts as:

① $L_{stroke}(\theta) = \| \epsilon_t - \epsilon_\theta(y_t; c, s, \sqrt{\bar{\alpha}_t}) \|_2^2$ → L_2 loss

② $L_{draw}(\theta) = -d_t \log(\hat{d}_t) - (1 - d_t) \log(1 - \hat{d}_t)$ → BLE.

↳ Where $\hat{d}_t = d_\theta(y_t; c, s, \sqrt{\bar{\alpha}_t})$ is the prediction whether pen is down or nah.

*) Model architecture: