DDPM

Mathmatical Foundation

① Markov chain: No memory

Current is only relavent to the last state

e.g. $N \rightarrow N-1 \rightarrow N-2$

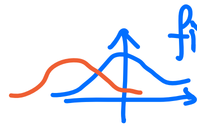
$x(t)$ is only dependent on $x(t-1)$ $P(x_t | x_{t-1})$

② Diffrentiable

$O \longrightarrow O$

↑ random sampling : no derivatives

$z \sim N(\mu, \sigma)$ use reparaparameterlization



find a num

shift $N(\mu, \sigma)$ to $N(0,1)$. throws the randomness outside

Becomes $z := N(z; \mu, \sigma I)$

$z = \mu + \sigma \cdot \varepsilon$ (dot product), where $\varepsilon \sim N(0,1)$

so z is diffrentiable, randomness gets thrown

Diffusion: time stamp $(0, 1, \dots t \dots T)$

$\overset{t-1}{O} \xrightarrow[q]{} \overset{t}{O}$ q: diffuse

Def: q: $x_t \sim N(\underbrace{\sqrt{1-\beta_t}\, x_{t-1}}_{\mu}, \underbrace{\sqrt{\beta_t}}_{\sigma})$

β_t is a hyper parameter
$0 < \beta_t < 1$

Use reparameterlization.

unit matrix to keep the shape

$x_t = N(x_t; \sqrt{1-\beta_t}\, x_{t-1}, \sqrt{\beta_t} \cdot I)$

$= \underbrace{\sqrt{1-\beta_t}}_{0<\cdot<1}\, x_{t-1} + \sqrt{\beta_t} \overset{\text{dot}}{\cdot}\, \varepsilon, \quad \varepsilon \sim N(0,1)$ 即，每一步加 N noise

spread out each step, increasing noise

$q(x_t \mid x_{t-1}) = N(x_t; \sqrt{1-\beta_t}\, x_{t-1}, \sqrt{\beta_t} I$

$$\boxed{N(0, \sigma_1^2 I) + N(0, \sigma_2^2 I) = N(0, (\sigma_1^2 + \sigma_2^2) I)}$$

$q(x_{1:T} \mid x_0) \qquad x_0 \to x_1 \to x_2 \cdots x_T$

$= \prod_{t=1}^{T} q(x_t \mid x_{t-1})$

Properties: ① $x_t \to x_0$, $\beta = \langle \beta_1, \beta_2, \dots \beta_t \rangle$

x_n only relavent to x_0 and β

Proof let $\alpha_t = 1-\beta_t \qquad \bar{\alpha}_t = \prod_{i=1}^{T} \alpha_i$

$$\begin{cases} x_t = \sqrt{\alpha_t}\, x_{t-1} + \sqrt{1-\alpha_t}\, \varepsilon_{t-1} \\ x_{t-1} = \sqrt{\alpha_{t-1}}\, x_{t-2} + \sqrt{1-\alpha_{t-1}}\, \varepsilon_{t-2} \end{cases}$$ substitute

$\therefore x_t = \sqrt{\alpha_t}(\sqrt{\alpha_{t-1}}\, x_{t-2}) + \sqrt{\alpha_t} \cdot \sqrt{1-\alpha_{t-1}}\, \varepsilon_{t-2} + \sqrt{1-\alpha_t}\, \varepsilon_{t-1}$

since $\underline{\varepsilon \sim N(0,1)}$ * not correlated

$= \sqrt{\alpha_t \cdot \alpha_{t-1}}\, x_{t-2} + \sqrt{\alpha_t \cdot (1-\alpha_{t-1}) + 1 - \alpha_t}\; \varepsilon_{t-2}$

$= \sqrt{\alpha_t \alpha_{t-1}}\, x_{t-2} + \sqrt{1 - \alpha_t \alpha_{t-1}}\; \bar{\varepsilon}_{t-2}$

Thus, $\sqrt{\alpha_t \alpha_{t-1} \alpha_{t-2} \cdots \alpha_1}\, x_0 + \sqrt{1 - \alpha_t \alpha_{t-1} \alpha_{t-2} \cdots \alpha_1}\; \bar{\varepsilon}_0$

$x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1 - \bar{\alpha}_t}\; \varepsilon, \quad \varepsilon \sim N(0,1)$

Namely, x_t only depends on x_0 and β (indeed Marcov chain)

$\beta \in (0,1)$

② $\lim_{T\to\infty} x_T = \sqrt{\bar{\alpha}_T}\, x_0 + \sqrt{1-\bar{\alpha}_T}\, \varepsilon = 0 + 1\cdot\varepsilon$

$x_T \sim N(0,1)$ when $T \to \infty$

Use random noise to generate an image

Then: differentiable Markov process

$\overset{x_0}{O} \quad O \quad O \quad \ldots O \quad \overset{x_T}{O} \sim z \sim N(0,1)$

smaller step to refine   larger step

$\beta = (0.0001, \ldots 0.02)$ step becomes larger

③ $\underset{x_{t-1}}{O} \longrightarrow \underset{x_t}{O}$

$q(x_t \mid x_{t-1})$

reverse: $q(x_{t-1} \mid x_t)$

is this also N?   (math problem)

(it is)   when β is small $\sim$ Normal

want to find $P_\theta(x_{t-1} \mid x_t)$ just a predict though

note that P_θ is based on x_0

$P_\theta(x_{t-1} \mid (x_t, x_0))$

only backward

determined by x_0 though?

$= N(x_{t-1};\ \tilde{\mu}(x_t, x_0),\ \tilde{\beta}_t I)$ use Bayesian

estimate

reverse

$q(x_{t-1} \mid (x_t, x_0)) \overset{4)}{=} \dfrac{q(x_t, x_0, x_{t-1})}{q(x_t, x_0)}$

1) $P(A \mid B) = \dfrac{P(A)P(B \mid A)}{P(B)}$

2) $P(AB) = P(A)P(B \mid A)$

3) $P(ABC) = P(A)P(B \mid A)P(C \mid AB)$

$$\overset{2)}{=}\frac{\cancel{q(x_0)}\,q(x_{t-1}|x_0)\,q(x_t|(x_{t-1},x_0))}{\underset{1)}{}\cancel{q(x_0)}\,q(x_t|x_0)}$$

4) $P(A|B)=\frac{P(AB)}{P(B)}$

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

reverse so contains θ

based on property ①, x_t only relavent to x_{t-1}

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

$$\boxed{x\sim N(\mu,\sigma)\propto e^{\frac{-(x-\mu)^2}{2\sigma^2}}=\exp\left(-\frac{1}{2}\left(\frac{\overset{2}{x^2}}{\sigma^2}-\frac{2\mu\overset{1}{x}}{\sigma^2}+\frac{\overset{0}{\mu^2}}{\sigma^2}\right)\right)}$$

$$=\exp\left(-\frac{1}{2}\left(\frac{(x_t-\sqrt{\alpha_t}x_{t-1})^2}{\beta_t}+\frac{(x_{t-1}-\sqrt{\bar\alpha_{t-1}}x_0)^2}{1-\bar\alpha_{t-1}}-\frac{(x_t-\sqrt{\bar\alpha_t}x_0)^2}{1-\bar\alpha_t}\right)\right)$$

$$=\exp\left(-\frac{1}{2}\left(\overset{2}{\left(\frac{\alpha_t}{\beta_t}+\frac{1}{1-\bar\alpha_{t-1}}\right)x_{t-1}^2}-\overset{1}{\left(\frac{2\sqrt{\alpha_t}}{\beta_t}x_t+\frac{2\sqrt{\bar\alpha_{t-1}}}{1-\bar\alpha_{t-1}}x_0\right)x_{t-1}}\right.\right.$$

$$\left.\left.\overset{0}{+\text{constant}\,x_{t-1}^0}\right)\right)$$

then, align the power term

$$\tilde\beta_t=\frac{1-\bar\alpha_{t-1}}{1-\bar\alpha_t}\beta_t \qquad \tilde\mu_t(x_t,x_0)=\frac{\sqrt{\alpha_t}(1-\bar\alpha_{t-1})}{1-\bar\alpha_t}x_t+\frac{\sqrt{\bar\alpha_{t-1}}\beta_t}{1-\bar\alpha_t}x_0$$

Because $x_t=\sqrt{\bar\alpha_t}x_0+\sqrt{1-\bar\alpha_t}\,\varepsilon$ Substitute... long... $x_0=?x_t$

$$\tilde\mu_t(x_t,x_0)=\frac{1}{\sqrt{\alpha_t}}\left(x_t-\frac{1-\alpha_t}{\sqrt{1-\bar\alpha_t}}\varepsilon_t\right)$$

estimation

solvable

$P_\theta(\neg q)$

Then, $x_t = N(x_{t-1}; \mu_t, \sigma_t) = N\left(x_{t-1}; \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\varepsilon_t\right)\right)$ *

$T \to \infty$, $\beta_t \ll 1$, and $\beta_t \in (0,1)$

Backward solvable

$$P(x_{t-1} \mid (x_t, x_0)) \sim N\left(\frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{1-\bar{\alpha}_t}x_0, \left(\frac{\sqrt{1-\alpha_t}\sqrt{1-\bar{\alpha}_{t-1}}}{\sqrt{1-\bar{\alpha}_t}}\right)^2\right).$$

* no beta

Namely, $\mu_{t-1} = \frac{x_t}{\sqrt{\alpha_t}} - \frac{1-\alpha_t}{\sqrt{\alpha_t}\sqrt{1-\bar{\alpha}_t}} \cdot \varepsilon$

$$\sigma^2 = \left(\frac{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\right) = \frac{\beta_t(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}$$