# Deep Learning

# Prove DDPM

311512064 鄧書桓

Problem 1:

$$q(x_T|x_0)\prod_{t=T}^{2} q(x_{t-1}|x_t, x_0)$$

$$= q(x_T|x_0)\,[\,q(x_{T-1}|x_T, x_0)\, q(x_{T-2}|x_{T-1}, x_0)\, q(x_{T-3}|x_{T-2}, x_0)\cdots q(x_2|x_3, x_0)\, q(x_1|x_2, x_0)\,]$$

$$= q(x_T|x_0)\,[\,q(x_T|x_{T-1}, x_0)\cdot \frac{\cancel{q(x_{T-1}|x_0)}}{q(x_T|x_0)} \cdot q(x_{T-1}|x_{T-2}, x_0)\frac{q(x_{T-2}|x_0)}{\cancel{q(x_{T-1}|x_0)}}\cdot\cdots\cdot q(x_2|x_1, x_0)\cdot\frac{q(x_1|x_0)}{\cancel{q(x_2|x_0)}}\,]$$

$$= \cancel{q(x_T|x_0)}\,[\,\prod_{t=2}^{T} q(x_t|x_{t-1})\frac{q(x_1|x_0)}{\cancel{q(x_T|x_0)}}\,]$$

$$= q(x_1|x_0)\prod_{t=2}^{T} q(x_t|x_{t-1})$$

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

$$= q(x_{1-T}|x_0)$$

Problem 2:

(a) prove Eq.4 $q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar\alpha_t}\, x_0, (1-\bar\alpha_t)I)$

suppose $\epsilon_i$ $\forall i \in \{t-1, t-2, \dots\} \sim \mathcal{N}(0,1)$

then $x_t = \sqrt{\alpha_t}\, x_{t-1} + \sqrt{1-\alpha_t}\, \epsilon_{t-1}$

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

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

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

$$= \cdots$$

$$= \sqrt{\bar\alpha_t}\, x_0 + \sqrt{1-\bar\alpha_t}\,\epsilon$$

In 3rd line, $\bar\epsilon_{t-2}$ merges two Gaussian matrics $\epsilon_{t-1}$ & $\epsilon_{t-2}$

$\therefore$ $q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar\alpha_t}\, x_0, (1-\bar\alpha_t)I)$

(b) prove Eq(6) $q(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}; \tilde{\mu}_t(x_t, x_0), \tilde{\beta}_t I)$

From Eq(4), we can know

$$q(x_t|x_{t-1}, x_0) = q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}\, x_{t-1}, \beta_t I)$$

$$q(x_{t-1}|x_0) = \mathcal{N}(x_{t-1}; \sqrt{\bar{\alpha}_{t-1}}\, x_0, (1-\bar{\alpha}_{t-1}) I)$$

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}\, x_0, (1-\bar{\alpha}_t) I)$$

Thus,

$$q(x_{t-1}|x_t, x_0)$$

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

$$\propto \exp\left\{-\frac{1}{2}\left(\frac{(x_t-\sqrt{\alpha_t}\, x_{t-1})^2}{\beta_t} + \frac{(x_{t-1}-\sqrt{\bar{\alpha}_{t-1}}\, x_0)^2}{1-\bar{\alpha}_{t-1}} + \frac{(x_t-\sqrt{\bar{\alpha}_t}\, x_0)^2}{1-\bar{\alpha}_t}\right)\right\}$$

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

$C(x_t, x_0)$ does not depend on $x_{t-1}$, can be omitted

Then,

$$\tilde{\beta}_t = 1/\left(\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar{\alpha}_{t-1}}\right) = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\cdot\beta_t$$

$$\tilde{\mu}(x_t, x_0) = \left(\frac{\sqrt{\alpha_t}}{\beta_t}x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}x_0\right) / \left(\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar{\alpha}_{t-1}}\right)$$

$$= \left(\frac{\sqrt{\alpha_t}}{\beta_t}x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}x_0\right)\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\cdot\beta_t$$

$$= \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\,\beta_t}{1-\bar{\alpha}_t}x_0$$

where $q(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}; \tilde{\mu}(x_t, x_0), \tilde{\beta}_t I)$

(c) prove Eq(8) $L_{t-1} = E_q\left[\frac{1}{2\sigma_t^2}\|\tilde{\mu}_t(x_t, x_0) - \mu_\theta(x_t, t)\|^2\right] + C$

from eq(16), $L = \underbrace{D_{KL}(q(x_T|x_0)\,\|\,p_\theta(x_T))}_{L_T} - \underbrace{E_q(x_1|x_0)\log p_\theta(x_0|x_1)}_{L_0}$

$$+ \underbrace{\sum_{t=2}^{T} E_{q(x_t|x_0)}\left[D_{KL}(q(x_{t-1}|x_t, x_0)\,\|\,p_\theta(x_{t-1}|x_t))\right]}_{L_{t-1}}$$

Fixed $\Sigma_\theta(x_t, t) = \sigma_t^2 I$, KL divergence of two Gaussian distribution $p_1$ & $p_2$ is

$$KL(p_1\|p_2) = \frac{1}{2}\left[tr(\Sigma_2^{-1}\Sigma_1) + (\mu_2-\mu_1)^T\Sigma_2^{-1}(\mu_2-\mu_1) - n + \log\frac{det(\Sigma_2)}{det(\Sigma_1)}\right]$$

Thus,

$$D_{KL}(q(x_{t-1}|x_t, x_0)\,\|\,p_\theta(x_{t-1}|x_t))$$

$$= D_{KL}(\mathcal{N}(x_{t-1}; \tilde{\mu}(x_t, x_0), \sigma_t^2 I)\,\|\,\mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \sigma_t^2 I))$$

$$= \frac{1}{2}\left(n + \frac{1}{\sigma_t^2}\|\tilde{\mu}_t(x_t, x_0) - \mu_\theta(x_t, t)\|^2 - n + \log 1\right)$$

$$= \frac{1}{2\sigma_t^2}\|\tilde{\mu}_t(x_t, x_0) - \mu_\theta(x_t, t)\|^2$$

$$\to L_{t-1} = E_{q(x_t|x_0)}\left[\frac{1}{2\sigma_t^2}\|\tilde{\mu}_t(x_t, x_0) - \mu_\theta(x_t, t)\|^2\right]$$