$$\mathcal{L}(\theta, \varphi) = \int q(z|x,\theta) \log \frac{p(x,z|\varphi)}{q(z|x,\theta)} dz =$$

$$= \{ p(x,z|\varphi) = p(x|z,\varphi)p(z) \} = \int q(\ ) \log \frac{p(x|z,\varphi)}{[- - -]}$$

вывод ем кри....

Лекция

Диффузионные модели

Denoising Diffusion Probabilistic Models (2020)

$\mathbb{R}^n$

время

$0, 1, \dots, T$

p_{data} $\qquad$ $\mathcal{N}(0, I)$

если T больше, то распред-е на шаге T близко к $\mathcal{N}(0, I)$

Forward process

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

transition prob dist.

помалого зашумляем картинку

$$q(x_{0:T}|x_0) = \prod^{T} q(x_t|x_{t-1})$$

Backward process

$$p_\theta(x_{t-1} \mid x_t) = \mathcal{N}(x_{t-1} \mid \mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$$

можно положить $\beta_t I$

$$p_\theta(x_{0:T}) := \underbrace{p(x_T)}_{\mathcal{N}(0, I)} \cdot \prod^{T} p_\theta(x_{t-1} \mid x_t)$$

как семплить?

1) $x_T \sim \mathcal{N}(0, I)$
2) for $t = T$ to 1:

   $x_{t-1} \sim p_\theta(x_{t-1} \mid x_t)$

3) x_0 - final object

хотим: $x_0 \sim p_{data}$

правдоп: $\log p_\theta(x_0) = \log \int p_\theta(x_{0:T}) dx_1 \dots dx_T =$

$$= \log \int q(x_{1:T} \mid x_0) \frac{p_\theta(x_{0:T})}{q(x_{1:T} \mid x_0)} dx_1 \dots dx_T =$$

$$= \log \mathbb{E}_{q(x_{1..T} \mid x_0)} \left[\frac{p_\theta(x_{0:T})}{q(\;)} \right] \geq \left\{ \begin{matrix} \text{Jensen} \\ \text{Ineq} \end{matrix} \right\} \geq$$

$$\geq \mathbb{E}_{q(x_{1:T} \mid x_0)} \left[\log \frac{p_\theta(x_{0:T})}{q(x_{1:T} \mid x_0)} \right] \longrightarrow \max_\theta$$

распишем

$$\mathbb{E}_{q(x_{1:T}|x_0)}\left[\log p(x_t) + \sum_{t=1}^{T} \log \frac{p_\theta(x_{t-1}|x_t)}{q(x_t|x_{t-1})}\right] = L$$

$$x_0 \to x_1 \to \dots\, x_T \to \nabla_\theta$$

Training:

1) Sample $x_1 \sim q(x_1|x_0) \dots\ x_T \sim q(x_T | x_{T-1})$

2) $\nabla_\theta \sum_t^T \log \frac{p_\theta(x_{t-1}|x_t)}{q(x_t|x_{t-1})}$

$$L = \mathbb{E}_{q(x_{1:T}|x_0)}\Big[-KL(q(x_T|x_1)\,||\,p(x_T)) -$$
$$- \sum^{T} KL(q(x_{T-1}|x_t, x_0)\,||\,p_\theta(x_{t-1}|x_t)) + \log p_\theta(x_0|x_1)\Big]$$

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

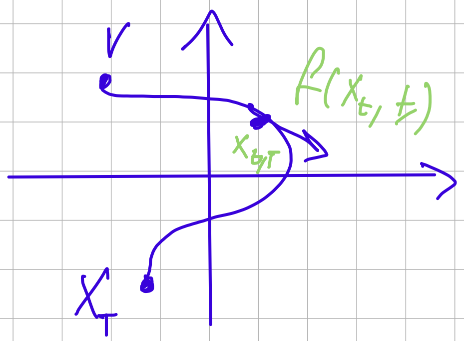
$$q(x_t | x_0) = \mathcal{N}\left(x_t \,\middle|\, \prod_i^t \sqrt{1-\beta_i}\, x_0, \left(1 - \prod_i^t (1-\beta_i)\right) I\right)$$

1) $t \sim U[1, \dots T]$ $x_t \sim q(x_t | x_0)$

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

Ordinary Differential Equation

$X_t \in \mathbb{R}^n \qquad t \in [0, T]$



$$\begin{cases} dX_t = f(X_t, t)\,dt \\ X_0 = V \in \mathbb{R}^n \end{cases}$$

Euler scheme:

$$X_{t+h} = X_t + h \cdot f(X_t, t) + \bar{o}(h) \approx X_t + h f(X_t, t)$$

$X(t)$ - ?

Stochastic Differential Equation (SDE)

$t \in [0, T]$

$$\begin{cases} dX_t = f(X_t, t)\,dt + g(t)\,\underbrace{dW_t}_{\text{стохастич. шум}} \\ X_0 \sim p_0 \end{cases}$$

dW_t ← Wiener Process

Euler scheme:

$$X_{t+h} = X_t + h \cdot f(X_t, t) + \sqrt{h}\, g(t)\, \varepsilon_t \qquad \varepsilon_t \sim \mathcal{N}(0, I)$$

$X_t \sim p_t$ – solution

зашумление = стохастич. ДУ!

$t: 0 \to T$

$$\begin{cases} dX_t = [f(X_t, t) - g^2(t) \nabla_x \log p_t(X_t)] dt + g(t) dW_t \\ X_T \sim p_T \end{cases}$$

$t: T \to 0$

$\nabla_x \log p_t(X_t)$ – score function

$$X_{t-h} = X_t - h f(X_t, t) + h g^2(t) \nabla_x \log p_t(X_t) + \sqrt{h}\, g(t) \varepsilon_t$$

SDE forward:

$$\begin{cases} dX_t = -\frac{1}{2} \beta(t) X_t dt + \sqrt{\beta_t}\, dW_t \\ X \sim p_{data} \end{cases}$$

$X_T \sim \mathcal{N}(0, I)$

SDE Backward:

$$\begin{cases} dX_t = (-\frac{1}{2} \beta(t) X_t dt - \beta_t \nabla_x \log(p_t(X_t)) dt + \sqrt{\beta(t)}\, dW_t \\ X_T \sim \mathcal{N}(0, I) \end{cases}$$

но у нас нет $\nabla_x \log p_t(X_t)$! – можно учить!

Denoising Score Matching

$$s_\theta(x,t) \approx \nabla_x \log p_t(x)$$

$$\int\int p_t(x) \| s_\theta(x,t) - \nabla_x \log p_t(x) \|^2 \, dx \, dt \to \min_\theta$$

$\Updownarrow$ эквив. задача

$$\mathbb{E} \| s_\theta(x_t, t) - \nabla_x \log p_t(x_t \mid x_0) \|^2 \to \min_\theta$$

уже умеем считать!

$t \sim U[0, T]$

$x_0 \sim p_{data}$

$x_t \sim p_t(x_t \mid x_0)$

$h \to$ any real value