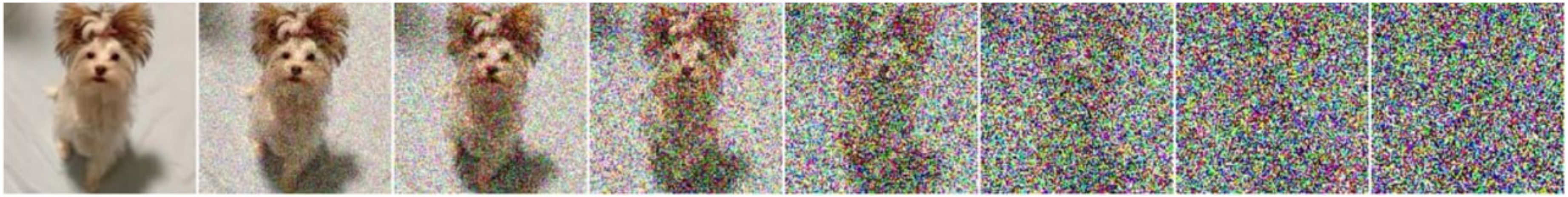
Diffusion models

Denis Rakitin

Diffusion

Forward $\longrightarrow$



p_{data} $\qquad\qquad$ $\mathcal{N}(0, \sigma^2)$

$\longleftarrow$ Backward

Idea: define forward noising process and try to reverse it.

https://arxiv.org/abs/2011.13456

$$q_{0\ldots T}(x_0 \ldots x_T) = p_{data}(x_0) \prod_{t=1}^{T} q_{t|t-1}(x_t | x_{t-1})$$

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

$$q_{t|0}(x_t | x_0) = \mathcal{N}(x_t | \sqrt{\alpha_t}\, x_0, (1-\alpha_t) I)$$

$$\alpha_t = \prod_{s=1}^{t} (1-\beta_s) \to 0$$

$$p^{\theta}_{0\ldots T}(x_0 \ldots x_T) = \mathcal{N}(x_T | 0, I) \prod_{t=1}^{T} \underbrace{p^{\theta}_{t-1|t}(x_{t-1} | x_t)}_{\mathcal{N}(x_{t-1} | A_t x_t + B_t \cdot D^{\theta}_t(x_t),\ C_t^2 I)}$$

$$KL(q_{0\ldots T}(x_0 \ldots x_T) \,\|\, p^{\theta}_{0\ldots T}(x_0 \ldots x_T)) \to \min_{\theta}$$

Ω

$$\text{const} + \sum_{t=2}^{T} \mathbb{E}_{q_{0,t}(x_0, x_t)} KL\left(q_{t-1|t,0}(x_{t-1}|x_t, x_0) \,\|\, p^{\theta}_{t-1|t}(x_{t-1}|x_t)\right)$$

$$= \sum_{t=2}^{T} w_t \, \mathbb{E}_{q_{0,t}(x_0, x_t)} \| D_t^{\theta}(x_t) - x_0 \|^2 \to \min_{\theta}$$

$$x_0^{(1)} \dots x_0^{(B)} \sim \text{dataset}$$

$$t^{(1)} \dots t^{(B)} \sim \text{распр.рав} \{1 \dots T\}$$

$$\varepsilon^{(1)} \dots \varepsilon^{(B)} \sim \mathcal{N}(0, I)$$

$$x_{noisy}^{(i)} = \sqrt{\alpha_{t^{(i)}}}\, x_0^{(i)} + \sqrt{1 - \alpha_{t^{(i)}}}\, \varepsilon^{(i)}$$

$$\frac{1}{B} \sum_{i=1}^{B} \| D^{\theta}_{t^{(i)}}(x_{noisy}^{(i)}) - x_0^{(i)} \|^2 \text{.backward()}$$

$$p^{\theta}_{t-1|t}(x_{t-1}|x_t) = \mathcal{N}(x_{t-1} | A_t x_t + B_t D^{\theta}_t(x_t), C^2_t)$$

Семплир.: $x_T \sim \mathcal{N}(0, I)$

$$x_{t-1} = A_t x_t + B_t D^{\theta}_t(x_t) + C_t \varepsilon_t$$

Фикс: T, $\beta_1, \dots, \beta_t, \dots, \beta_T \to \alpha_1, \dots, \alpha_t, \dots, \alpha_T$

$\beta_1 \dots \beta_T$ $\alpha_1 \dots \alpha_T$

$$\mathbb{E} \| D^{\theta}(\sqrt{\alpha_t} x_0 + \sqrt{1-\alpha_t}\,\varepsilon) - x_0 \|^2$$

$$\sqrt{\alpha_t} x_0 + \sqrt{1-\alpha_t}\,\varepsilon$$

$$x_0 + \sqrt{\frac{1-\alpha_t}{\alpha_t}}\,\varepsilon = 6$$

$\hat{\beta}_1 \dots \hat{\beta}_{\tilde{T}}$ $\hat{\alpha}_1 \dots \hat{\alpha}_{\tilde{T}}$

$$\mathbb{E} \| D^{\varphi}(\sqrt{\alpha_s} x_0 + \sqrt{1-\alpha_s}\,\varepsilon) - x_0 \|^2$$

$$x_0 + \sqrt{\frac{1-\alpha_s}{\alpha_s}}\,\varepsilon$$

Есть ли s: $\sigma = 6$?

Цель: $\sim q_0(x_0) \longrightarrow q_0(x_0|y)$

$$q_{0..T}(x_0..x_T|y) = q_0(x_0|y)\prod_t q_{t|t-1}(x_t|x_{t-1})$$

Обучение для усл. распр.:

$x \sim$ batch

$D_t(x_t)$

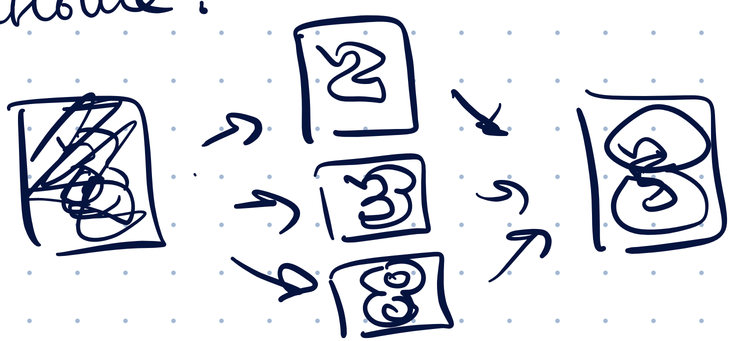
$$\sum_{t=2}^{T} \omega_t \underset{\substack{q_{0,t}(x_0,x_t|y) \\ \parallel \\ q_0(x_0|y)\,q_{t|0}(x_t|x_0)}}{\mathbb{E}} \|D_t^\theta(x_t|y) - x_0\|^2$$

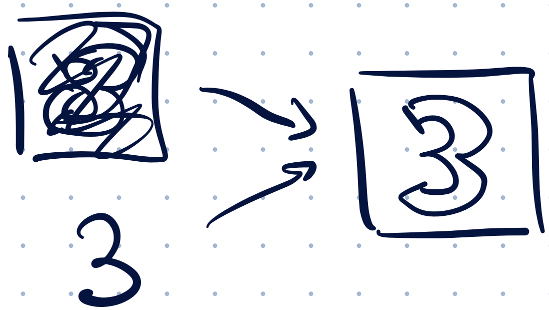
x, y - batch

$D_t(x_t|y)$

Семпл.: $x_T \sim \mathcal{N}(0, I)$ $\quad x_{t-1} = A_t x_t + B_t D_t^\theta(x_t|y) + C_t \varepsilon_t$

Раньше:





Хотелось бы: понять, как устроен $q_{t-1|t}$

просто

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

$$= q_{t|t-1}(x_t|x_{t-1}) \exp\left(\log q_{t-1}(x_{t-1}) - \log q_t(x_t)\right)$$

$$\approx q_{t|t-1}(x_t|x_{t-1}) \exp\left(\log q_t(x_{t-1}) - \log q_t(x_t)\right)$$

$$= q_{t|t-1}(x_t|x_{t-1}) \exp\left(\langle \underbrace{\nabla \log q_t(x_t)}_{\substack{\text{score функция} \\ \text{распр. } q_t(x_t)}}, x_{t-1} - x_t \rangle + \bar{o}(x_{t-1} - x_t)\right)$$

$$q_t(x_t) = \int q_{0,t}(x_0, x_t)\,dx_0 = \int q_{t|0}(x_t|x_0)\, q_0(x_0)\,dx_0$$

просто сложно

$$\nabla_{x_t}\log q_t(x_t) =$$

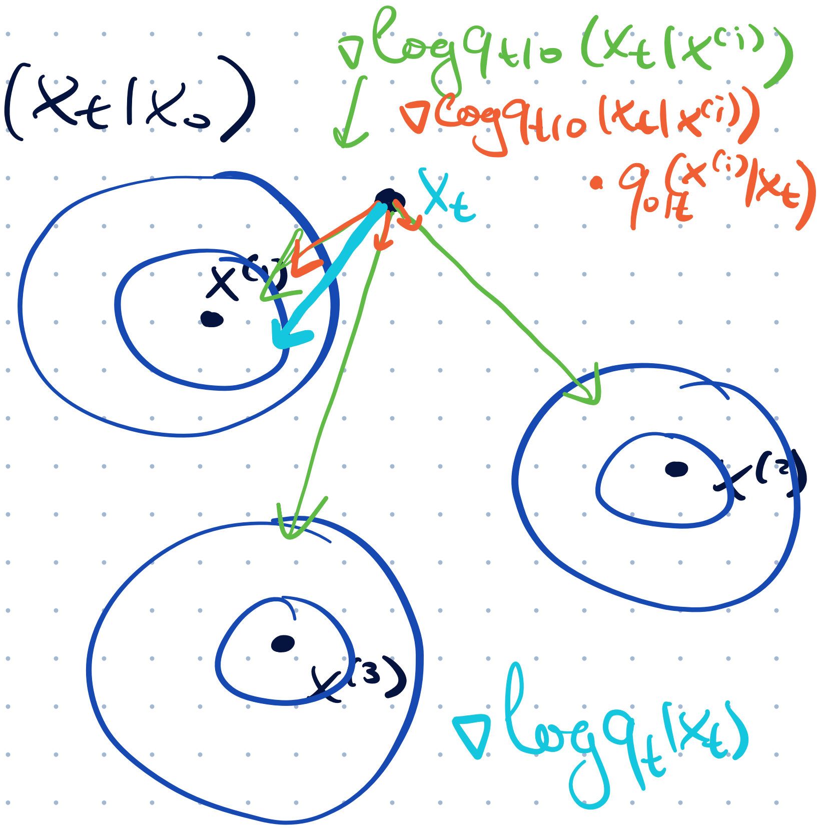
$$= \frac{\nabla_{x_t} q_t(x_t)}{q_t(x_t)} = \frac{\nabla_{x_t}\int q_{t|0}(x_t|x_0)\, q_0(x_0)\,dx_0}{q_t(x_t)} =$$

$$= \frac{\int \nabla_{x_t} q_{t|0}(x_t|x_0)\, q_0(x_0)\,dx_0}{q_t(x_t)}$$

$$= \frac{\int \nabla_{x_t}\log q_{t|0}(x_t|x_0)\, q_{t|0}(x_t|x_0)\, q_0(x_0)\,dx_0}{q_t(x_t)} =$$

$$\int \left(\nabla_{x_t}\log q_{t|0}(x_t|x_0)\right) q_{0|t}(x_0|x_t)\,dx_0 = \mathbb{E}_{q_{0|t}(x_0|x_t)} \nabla \log q_{t|0}(x_t|x_0)$$

$$\nabla_{x_t} \log q_t(x_t) = \mathbb{E}_{q_{0|t}(x_0|x_t)} \nabla \log q_{t|0}(x_t|x_0)$$



$$q_0(x_0) = \frac{1}{3}\sum_{i=1}^{3} \delta(x_0 = x^{(i)})$$

$$q_{t|0}(x_t|x_0) = \mathcal{N}(x_t|x_0, \sigma_t^2 I)$$

$$q_t(x_t) = \sum_{i=1}^{3} q_{t|0}(x_t|x^{(i)})\, q_0(x^{(i)}) =$$

$$= \sum_{i=1}^{3} \frac{1}{3} \mathcal{N}(x_t|x^{(i)}, \sigma_t^2 I)$$

$$\nabla_{x_t} \log q_t(x_t) = \sum_{i=1}^{3} q_{0|t}(x^{(i)}|x_t) \cdot \nabla \log q_{t|0}(x_t|x^{(i)})$$

$$q_{0|t}(x^{(i)}|x_t) \propto q_{t|0}(x_t|x^{(i)}) \cdot \frac{1}{3} \propto \frac{1}{3} \cdot \exp\left(-\frac{1}{2\sigma_t^2}\|x_t - x^{(i)}\|^2\right)$$

$$\nabla \log q_t(x_t) = \mathbb{E}_{q_{0|t}(x_0|x_t)} \nabla \log q_{t|0}(x_t|x_0)$$

$$q_{t|0}(x_t|x_0) = \mathcal{N}\left(x_t \mid \sqrt{\alpha_t}\, x_0, (1-\alpha_t) I\right)$$

$$= \text{const} \cdot \exp\left(\frac{-1}{2(1-\alpha_t)} \| x_t - \sqrt{\alpha_t}\, x_0 \|^2\right)$$

$$\nabla_{x_t} \log q_{t|0}(x_t|x_0) = \nabla_{x_t}\left(-\frac{1}{2(1-\alpha_t)} \| x_t - \sqrt{\alpha_t}\, x_0 \|^2\right) =$$

$$= -\frac{1}{2(1-\alpha_t)}\, 2(x_t - \sqrt{\alpha_t}\, x_0) = \frac{\sqrt{\alpha_t}\, x_0 - x_t}{1-\alpha_t}$$

$$\nabla_{x_t} \log q_t(x_t) = \mathbb{E}_{q_{0|t}(x_0|x_t)} \nabla \log q_{t|0}(x_t|x_0) =$$

$$= \mathbb{E}_{q_{0|t}(x_0|x_t)} \frac{\sqrt{\alpha_t}\, x_0 - x_t}{1-\alpha_t} = \frac{\sqrt{\alpha_t}\, \mathbb{E}_{q_{0|t}(x_0|x_t)} x_0 - x_t}{1-\alpha_t}$$

$$\sum_t \mathbb{E}_{q_t(x_t)} \left\| d_t^{\varphi}(x_t) - \mathbb{E}_{q_{0|t}(x_0|x_t)} x_0 \right\|^2 =$$

$$= \sum_t \mathbb{E}_{q_t(x_t)} \left(\| d_t^{\varphi}(x_t) \|^2 - 2 \left\langle d_t^{\varphi}(x_t), \mathbb{E}_{q_{0|t}(x_0|x_t)} x_0 \right\rangle \right) + \text{const}$$

$$= \text{const} + \sum_t \mathbb{E}_{q_t(x_t)} \| d_t^{\varphi}(x_t) \|^2 - 2 \sum_t \mathbb{E}_{q_t(x_t)} \left\langle d_t^{\varphi}(x_t), \mathbb{E}_{q_{0|t}(x_0|x_t)} x_0 \right\rangle =$$

$$\mathbb{E}_{q_t(x_t)} \left\langle d_t^{\varphi}(x_t), \mathbb{E}_{q_{0|t}(x_0|x_t)} x_0 \right\rangle = \mathbb{E}_{q_t(x_t)} \mathbb{E}_{q_{0|t}(x_0|x_t)} \left\langle d_t^{\varphi}(x_t), x_0 \right\rangle =$$

$$= \mathbb{E}_{q_{0,t}(x_0, x_t)} \left\langle d_t^{\varphi}(x_t), x_0 \right\rangle$$

$$\Leftrightarrow \text{const} + \sum_t \left(\mathbb{E}_{q_{0,t}(x_0,x_t)} \|d_t^{\varphi}(x_t)\|^2 - 2\,\mathbb{E}_{q_{0,t}(x_0,x_t)} \langle d_t^{\varphi}(x_t), x_0 \rangle \right)$$

$$= \text{const} + \sum_t \mathbb{E}_{q_{0,t}(x_0,x_t)} \| d_t^{\varphi}(x_t) - x_0 \|^2$$

$\Leftrightarrow$ обучение денойзера $D_t^{\theta}(x_t)$

$$\sum_t \mathbb{E}_{q_t(x_t)} \| d_t^{\varphi}(x_t) - \mathbb{E}_{q_{0|t}(x_0|x_t)} x_0 \|^2 = \text{const} + \sum_t \mathbb{E}_{q_{0,t}(x_0,x_t)} \| d_t^{\varphi}(x_t) - x_0 \|^2$$

$\Leftrightarrow$ в точке оптимума исходного денойзера $D_t^{\theta}(x_t)$

$$D_t^{*}(x_t) = \mathbb{E}_{q_{0|t}(x_0|x_t)} x_0$$

$$q_{t-1|t} \Leftrightarrow \nabla \log q_t(x_t) \Leftrightarrow \mathbb{E}_{q_{0|t}(x_0|x_t)}\, x_0 \Leftrightarrow \sum_t \mathbb{E}\|D_t^\theta(x_t) - x_0\|^2$$

$\hookrightarrow \min_\theta$

$$q_t \to q_t^\alpha$$

$$\log q_t^\alpha = \alpha \log q_t$$

$$\sum_t \mathbb{E}_{q_{0,t}(x_0, x_t)} \| D_t^\theta(x_t) - x_0 \|^2 \longrightarrow D_t^*(x_t) = \mathbb{E}_{q_{0|t}(x_0|x_t)} x_0$$

$$s_t^\theta(x_t) \qquad = \frac{-\varepsilon}{\sqrt{1-\alpha_t}} \qquad \nabla \log q_t(x_t) = \frac{\sqrt{\alpha_t}\, D_t^*(x_t) - x_t}{1-\alpha_t}$$

$$\sum_t \mathbb{E} \left\| s_t^\theta(x_t) - \frac{\sqrt{\alpha_t}\, x_0 - x_t}{1-\alpha_t} \right\|^2 \to s_t^*(x_t) = \nabla \log q_t(x_t)$$

$$x_t = \sqrt{\alpha_t}\, x_0 + \sqrt{1-\alpha_t}\, \varepsilon \qquad \sqrt{\alpha_t}\, x_0 - x_t = -\sqrt{1-\alpha_t}\, \varepsilon$$

$$s_t^\theta(x_t) = -\frac{\varepsilon_t^\theta(x_t)}{\sqrt{1-\alpha_t}}$$

$$\sum_t \mathbb{E} \left\| \varepsilon_t^\theta\left(\sqrt{\alpha_t}\, x_0 + \sqrt{1-\alpha_t}\, \varepsilon\right) - \varepsilon \right\|^2 \to \varepsilon_t^*(x_t) = \mathbb{E}_{q(\varepsilon|x_t)} \varepsilon$$

$$\nabla \log q_t(x_t) = \frac{\sqrt{\alpha_t}\, D_t^*(x_t) - x_t}{1-\alpha_t}$$

$$D_t^*(x_t) = \frac{(1-\alpha_t)\nabla \log q_t(x_t) + x_t}{\sqrt{\alpha_t}}$$

Как получить, зная $\nabla_{x_t} \log q_t(x_t)$?

$$\nabla_{x_t} \log q_t(x_t|y) = \nabla_{x_t} \log \frac{q_t(y|x_t)\, q_t(x_t)}{q(y)} =$$

$$= \nabla_{x_t} \log q_t(y|x_t) + \underbrace{\nabla_{x_t} \log q_t(x_t)}_{\text{предоб. диффуз.}}$$

$\approx \nabla_{x_t} c_t(y|x_t)$ $\qquad\qquad$ $\approx S_t^\theta(x_t)$

noisy classifier

$$\rightarrow \begin{array}{cc} \begin{bmatrix} 1/3 \\ 1/3 \\ \vdots \\ 1/3 \end{bmatrix} & \begin{array}{c} 2 \\ 3 \\ \\ 8 \end{array} \end{array}$$

$$(1+\gamma)\nabla_{x_t}\log q_t(y|x_t) + \nabla_{x_t}\log q_t(x_t)$$

$\gamma \geq 0$

Дано: $\nabla_{x_t}\log q_t(x_t)$

Цель:

Classifier guidance

Дано: $\nabla_{x_t}\log q_t(x_t|y)$

Цель: $\nabla_{x_t}\log q_t(x_t) + (1+\gamma)\nabla_{x_t}\log q_t(y|x_t) =$

$$= \left(\nabla_{x_t}\log q_t(x_t) + \nabla_{x_t}\log q_t(y|x_t)\right) + \gamma\nabla_{x_t}\log q_t(y|x_t)$$

$$= \nabla_{x_t}\log q_t(x_t|y) + \gamma\cdot\left(\nabla_{x_t}\log q_t(x_t|y) - \nabla_{x_t}\log q_t(x_t)\right)$$

$$\approx s_t^\theta(x_t|y) + \gamma\left(s_t^\theta(x_t|y) - s_t^\theta(x_t|\varnothing)\right)$$

Было: $y \in \{1..c\}$ Стало: $\hat{y} \in \{1...c, \varnothing\}$

$$(x_i, y_i) \longrightarrow (x_i, \hat{y}_i) \quad \hat{y}_i = \begin{cases} y_i, & \text{с вер-ю } p \\ \varnothing, & \text{с вер-ю } 1-p \end{cases}$$

Учим: $\mathbb{E} \sum_t \| D_t^\theta(x_t | \hat{y}) - x_0 \|^2$

classifier-free guidance

Effect of guidance

γ = Guidance scale

Summary

$$KL(q_{0..T} \,\|\, p^{\theta}_{0..T}) \to \min_{\theta}$$

$$\Leftrightarrow \sum_t \mathbb{E}_{q_{0,t}(x_0, x_t)} \| D^{\theta}_t(x_t) - x_0 \|^2$$

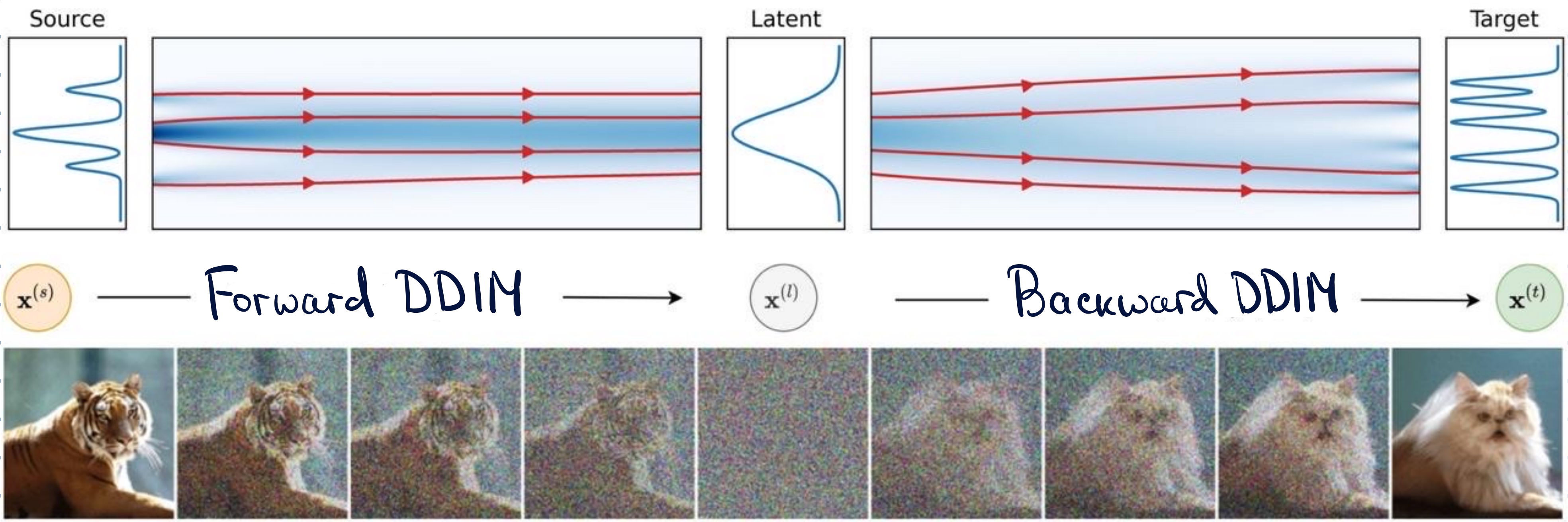
$\Leftrightarrow$ обучение score ф-и $\nabla \log q_t(x_t)$

$$\underset{\text{классиф.}}{\nabla \log q_t(x_t) \to} \nabla \log q_t(x_t|y)$$

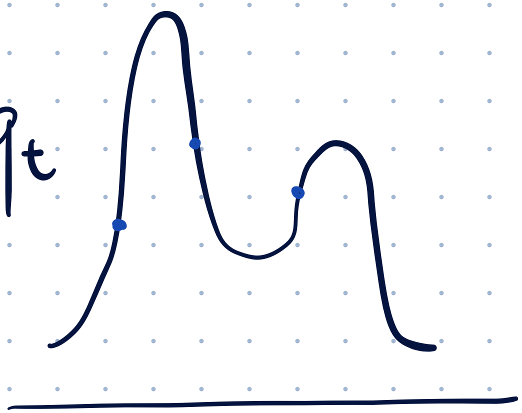
$$\to \nabla \log q_t(x_t|y)$$

$$+ \underset{\text{guidance scale}}{\gamma} \left(\nabla \log q_t(x_t|y) - \nabla \log q_t(x_t) \right)$$

Deterministic sampling



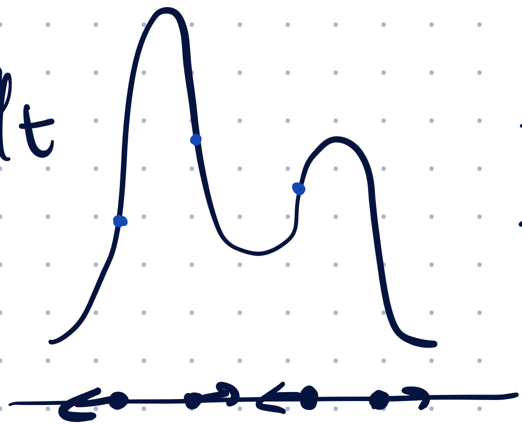
q_t



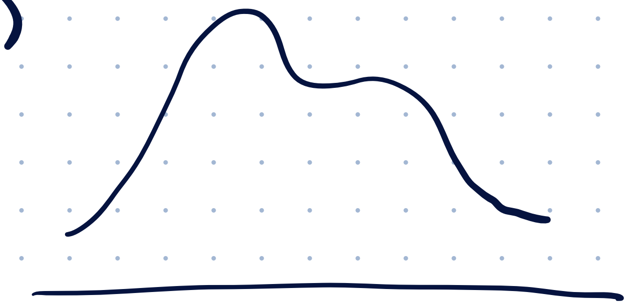
$\mathcal{N}(0, \gamma I)$

$\xrightarrow{+\,\varepsilon \cdot \gamma}$

q_t



$\xrightarrow{-\,\frac{\gamma^2}{2}\nabla \log q_t(x_t)}$



$$x_{t-1} = A_t x_t + B_t D_t^{\theta}(x_t) + \boxed{\cancel{C_t \varepsilon_t}}$$

$$- \frac{C_t^2}{2} \nabla \log q_t(x_t)$$

$$x \to x + \gamma \varepsilon = y \qquad x \sim p$$

$$\underline{x \to x - \frac{\gamma^2}{2} \nabla \log p(x) = z} \qquad z = x - \dots$$

$$z = z(x)$$

$$x = x(z)$$

$$p_z(z) = p_x(x(z)) \left| \det \frac{\partial x(z)}{\partial z} \right|$$

$$p_y(y) = \int p_x(x)\, p_{\gamma\varepsilon}(y - x)\, dx$$

$$p_0 \; p_1 \; p_2 \; \dots$$

$$p_t, \; t \in [0,1] \qquad \frac{\partial}{\partial t} p_t(x) = \dots \frac{\partial}{\partial x} p_t(x) \dots$$

$$x_{t+1} = x_t + \frac{\gamma^2}{2} \nabla \log p(x_t) + \gamma \cdot \varepsilon_t$$

$$- \frac{\gamma^2}{2} \nabla \log p(x_t)$$

Пусть $x_t \sim p$