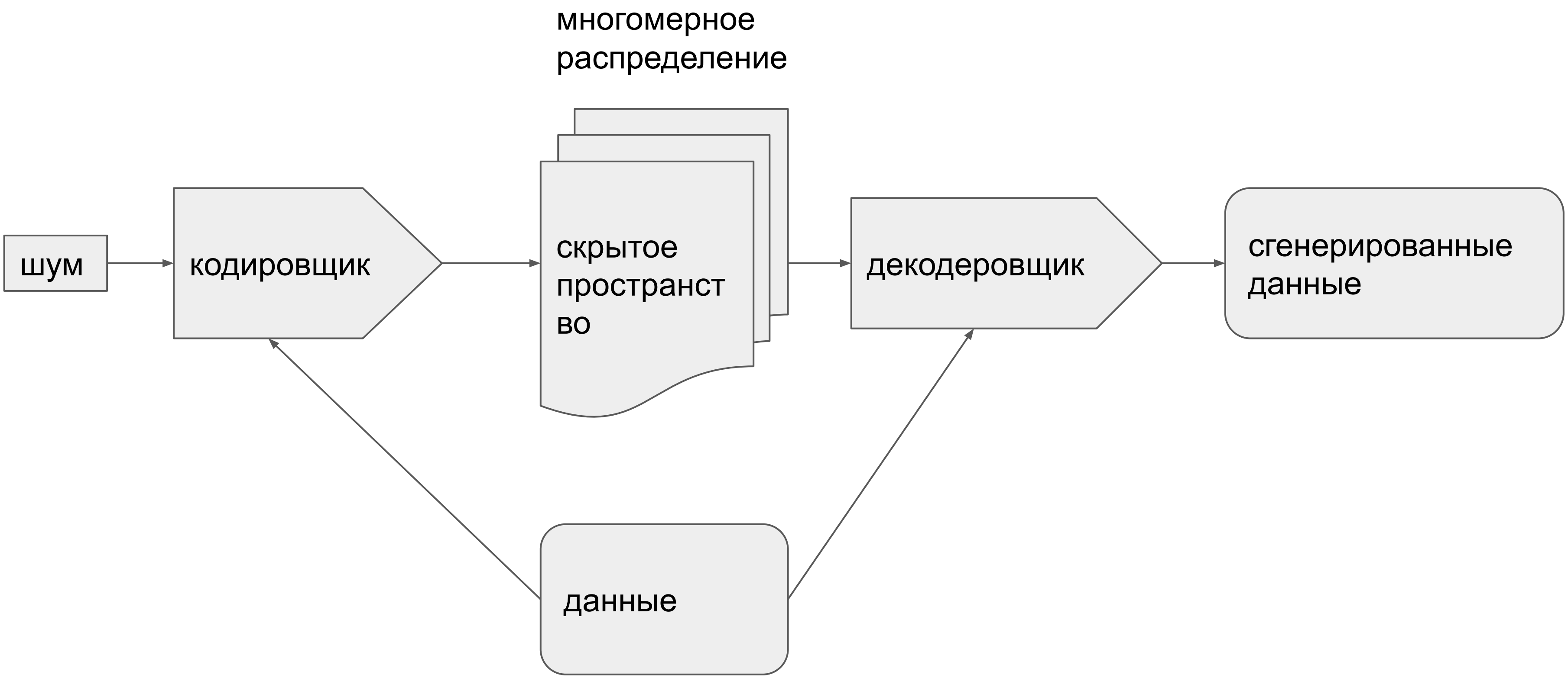
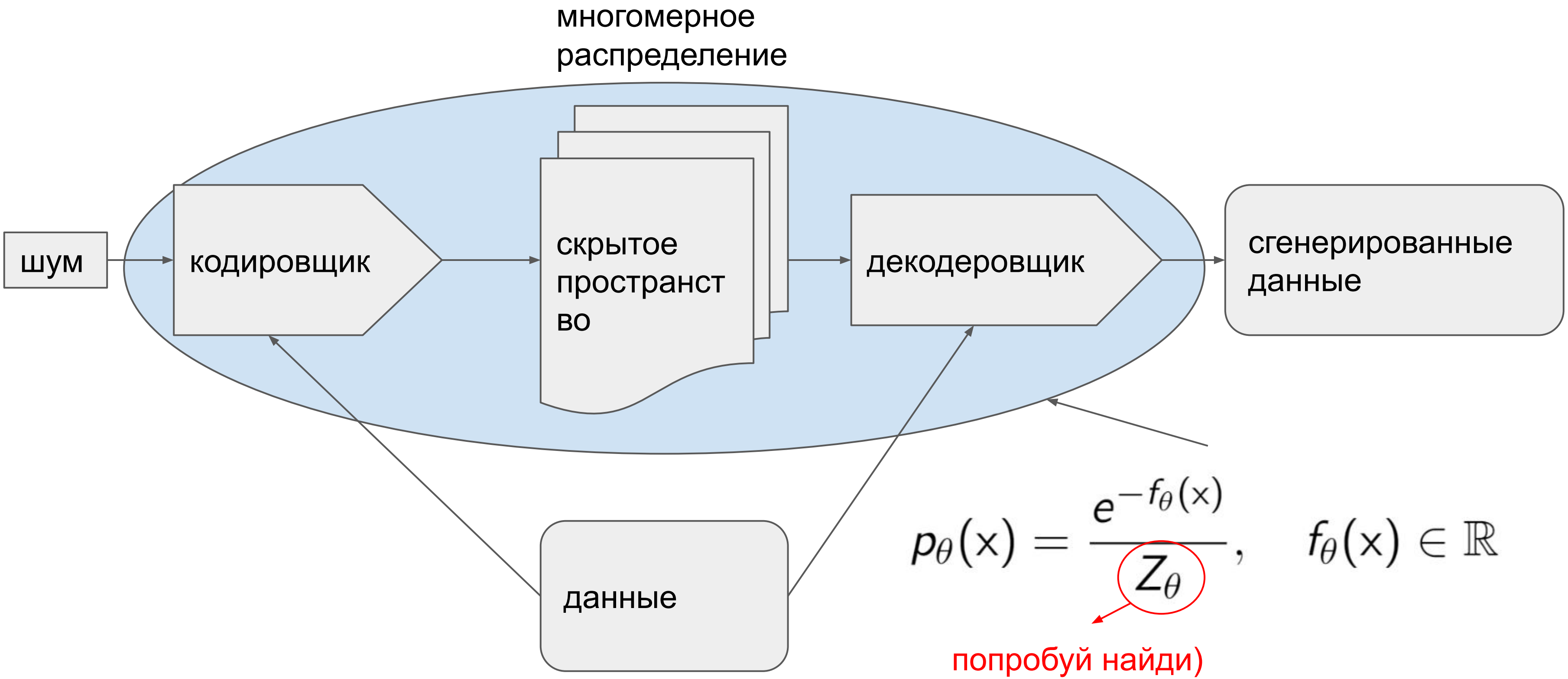
Дуффузионные модели

начало

Проблемы GANов и автоэнкодеров?



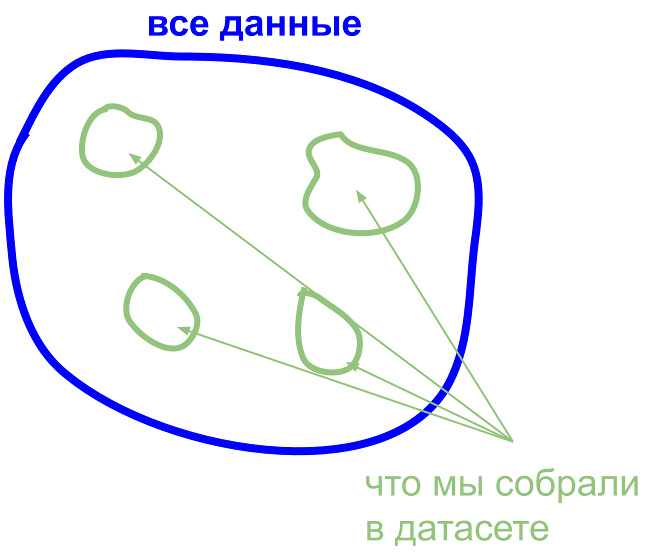
Нормализация

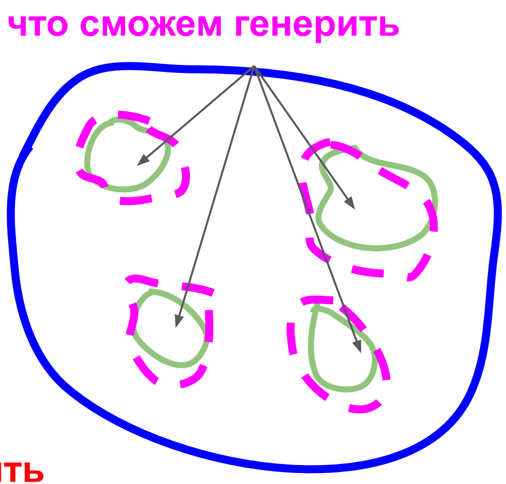


$$p_\theta(x) = \frac{e^{-f_\theta(x)}}{Z_\theta}, \quad f_\theta(x) \in \mathbb{R}$$

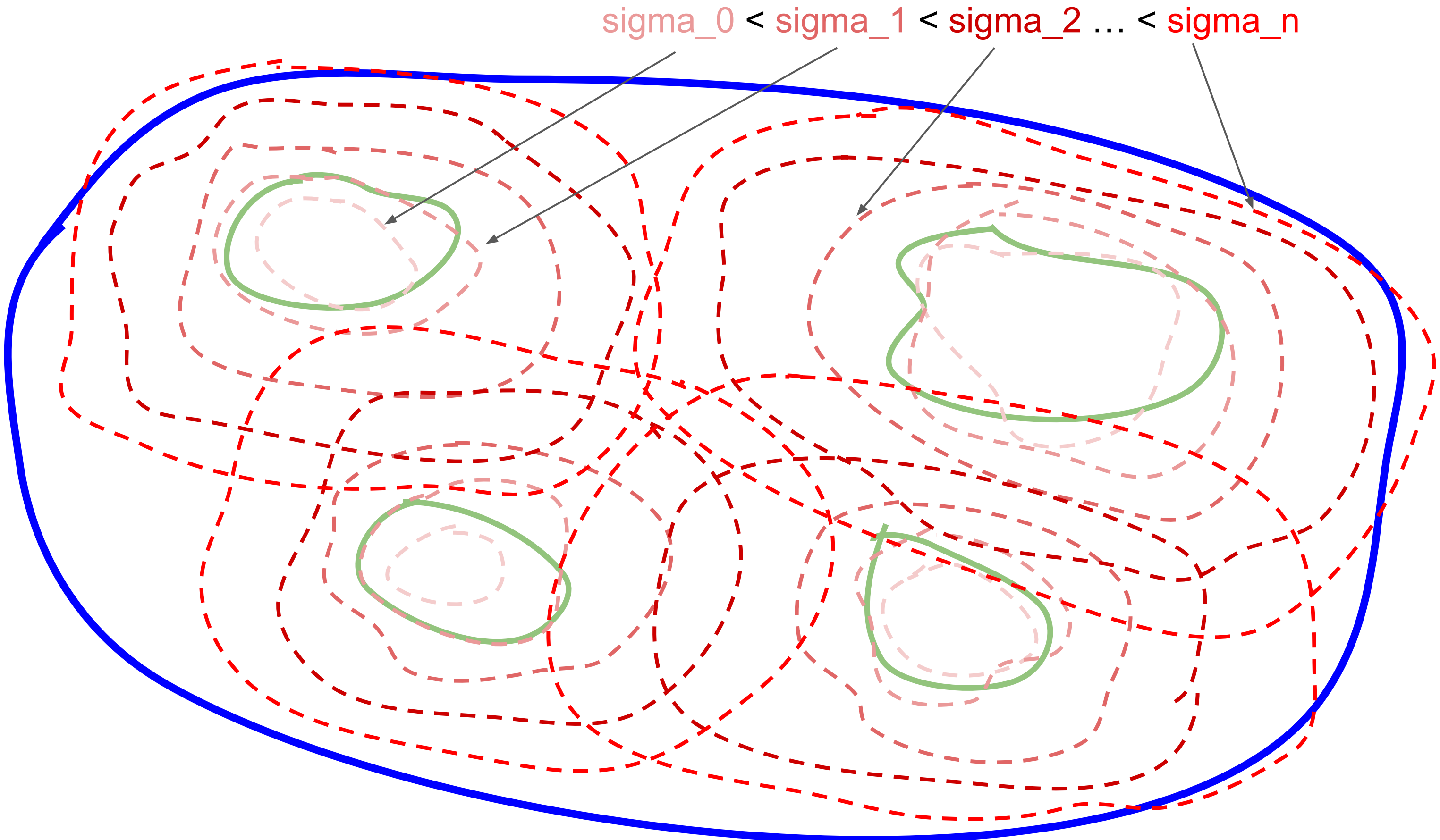
попробуй найди)

Хватает ли данных?





Пошумим, … ?



Лол… А как же учить? Ааа… Динамика Ланжевена!

датасет

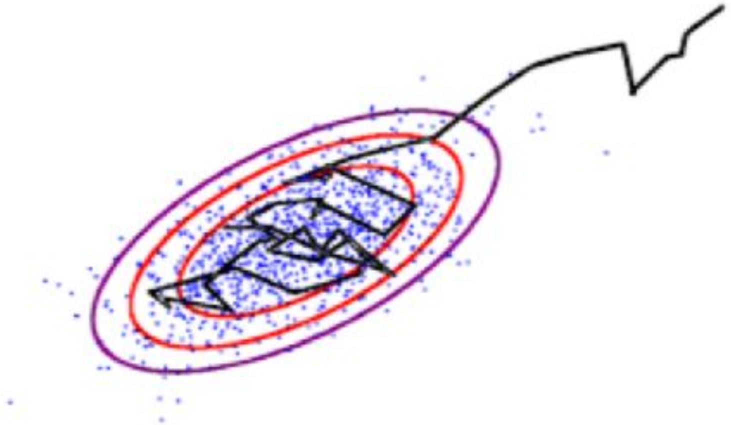
градиенты данных…
WHAAT?!?!?!

$$p(\mathrm{x}), \quad \nabla_{\mathrm{x}} \log p(\mathrm{x})$$

$$\mathrm{x}_{i+1} \leftarrow \mathrm{x}_i + \epsilon \nabla_{\mathrm{x}} \log p(\mathrm{x}) + \sqrt{2\epsilon}\, \mathrm{z}_i, \quad i = 0, 1, \cdots, K,$$

Где $\mathrm{z}_i \sim \mathcal{N}(0, I)$ и $K \to \infty$.

то, чем шумим

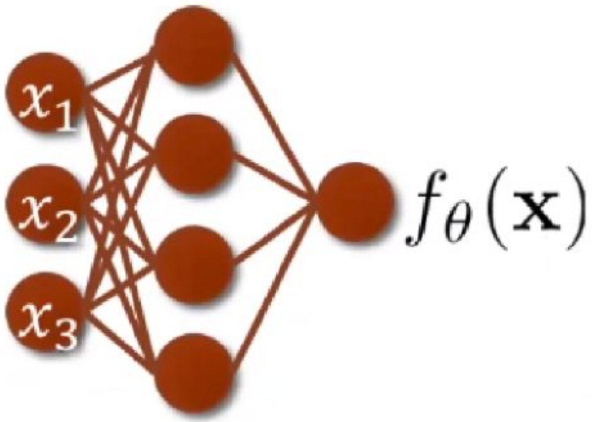


EBMs

- Energy-Based Models (EBMs)

$$p_\theta(\mathrm{x}) = \frac{e^{-f_\theta(\mathrm{x})}}{Z_\theta}, \quad f_\theta(\mathrm{x}) \in \mathbb{R}$$

$$Z_\theta = \int e^{-f_\theta(\mathrm{x})} \mathrm{dx}$$



- **Плюсы** Простая модель
- **Минусы** Сложно обучать θ через максимизацию правдоподобия

$$\mathbb{E}_{p_{\mathrm{data}}(\mathrm{x})}[-\log p_\theta(\mathrm{x})] = \mathbb{E}_{p_{\mathrm{data}}(\mathrm{x})}[f_\theta(\mathrm{x}) + \log Z_\theta]$$

Как же научить EBMs? Score matching !!!

- Score function

$$s(x) = \nabla_x \log p(x)$$

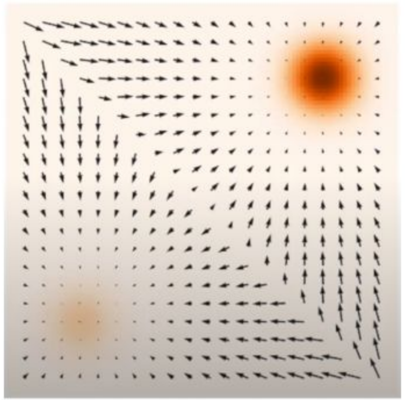
- Score function для EBM модели

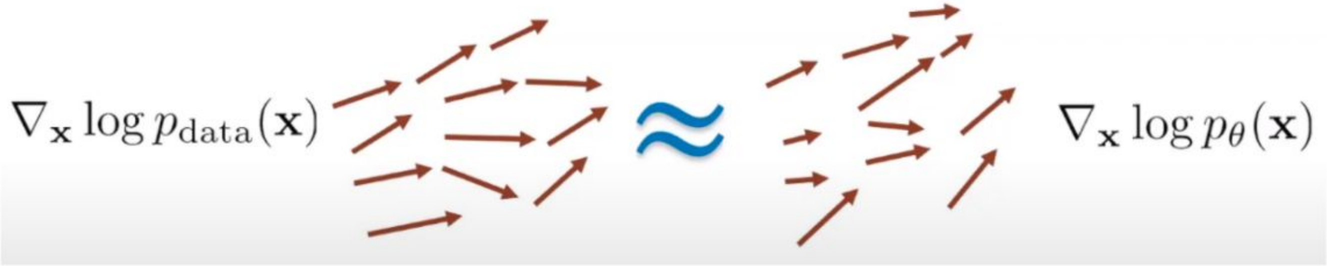
$$s_\theta(x) = \nabla_x \log p_\theta(x) = -\nabla_x f_\theta(x) - \underbrace{\nabla_x \log Z_\theta}_{=0} = -\nabla_x f_\theta(x)$$

- Score matching (Fisher divergence)

$$\mathbb{E}_{p_{\mathrm{data}}(x)}[||s_\theta(x) - \nabla_x \log p_{\mathrm{data}}(x)||_2^2]$$

Score matching





Вспомним что хотели шуметь… Denosing SM

- Зашумим данные

$$p_\sigma(\tilde{x}) = \int_x q_\sigma(\tilde{x}|x) p_{\mathrm{data}}(x) \mathrm{d}x, \quad q_\sigma(\tilde{x}|x) = \mathcal{N}(\tilde{x}; x, \sigma^2 I)$$

- Score matching для зашумленных данных

$$\mathbb{E}_{p_\sigma(x)}[||s_\theta(x) - \nabla_x \log p_\sigma(x)||_2^2]$$

- Можно переписать как

$$\mathbb{E}_{q_\sigma(\tilde{x}|x) p_{\mathrm{data}}(x)}[||s_\theta(x) - \nabla_{\tilde{x}} \log q_\sigma(\tilde{x}|x)||_2^2]$$

Еще раз о том зачем шумим…

- Динамика Ланжевена с трудом переходит между модами

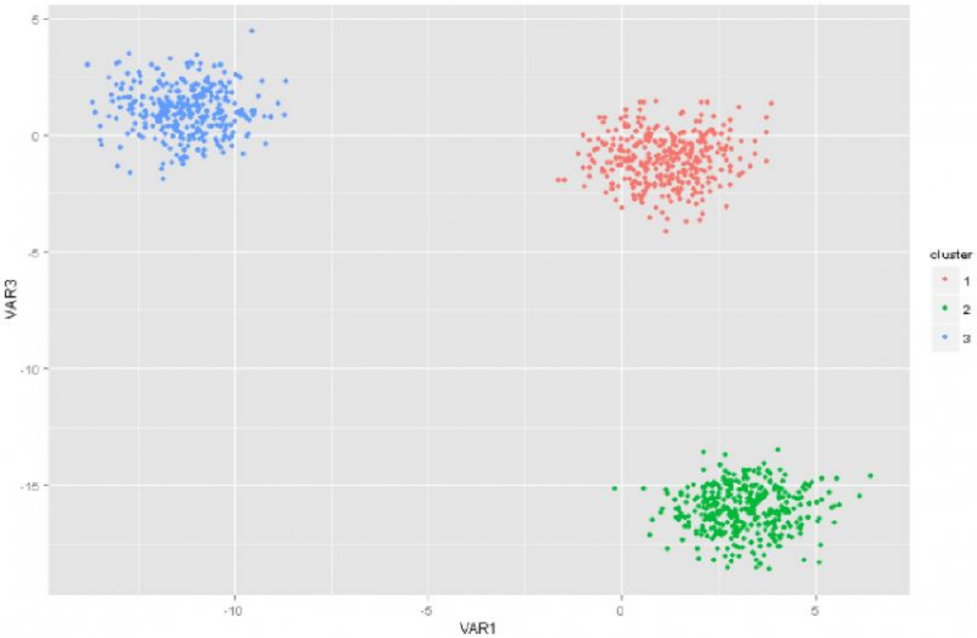
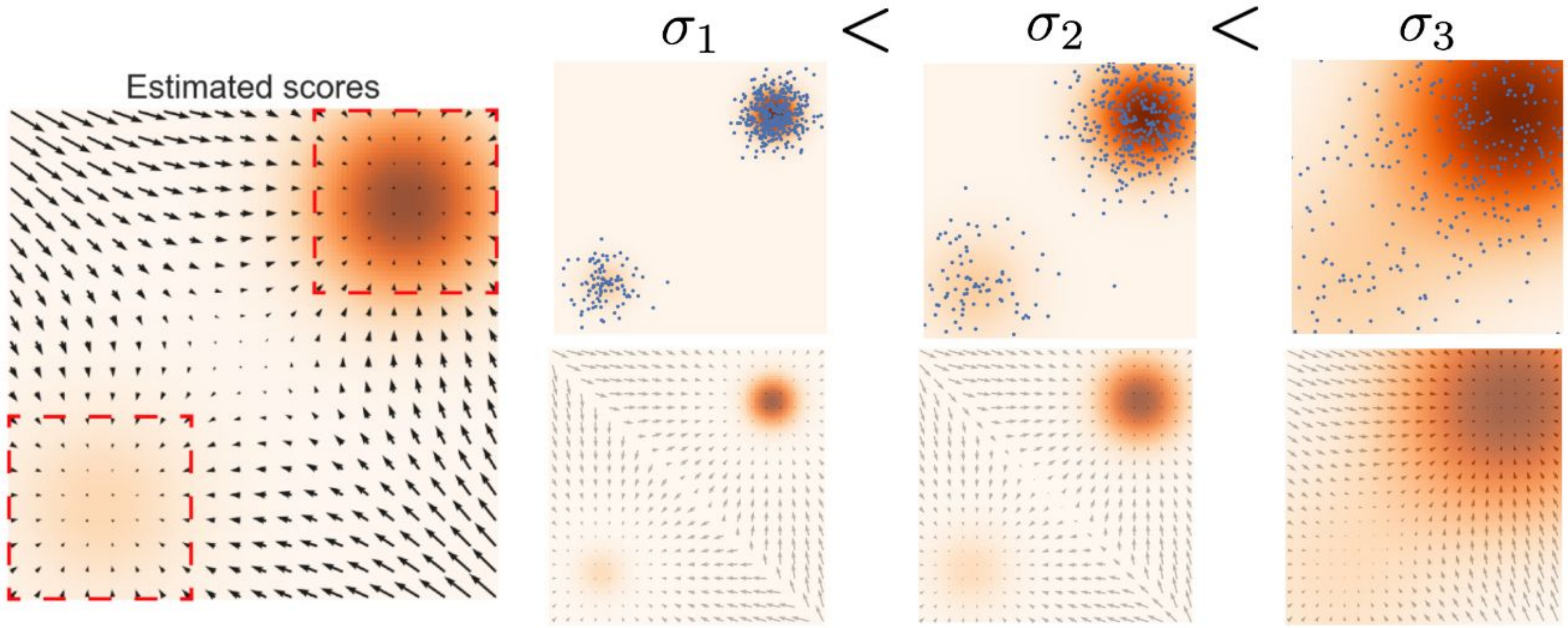


Иллюстрация как шумим



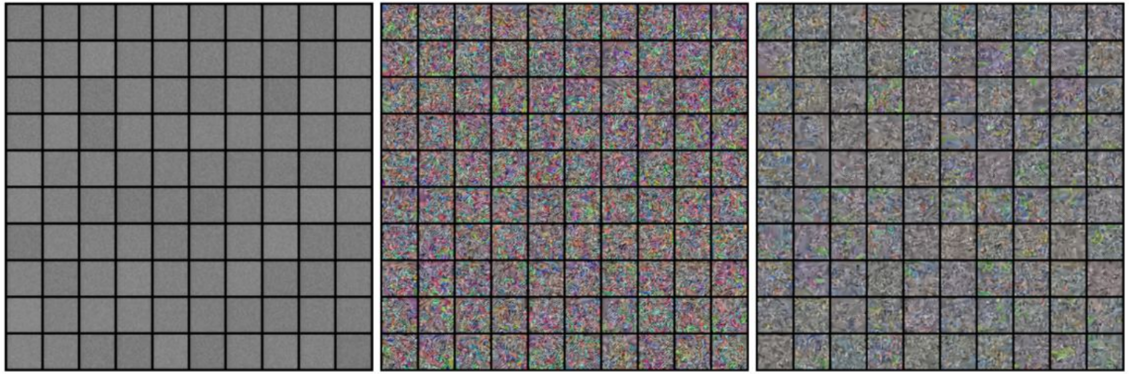
Учили без шума

$$\mathbb{E}_{p_{\mathrm{data}}(x)}[||s_\theta(x) - \nabla_x \log p_{\mathrm{data}}(x)||_2^2]$$

А теперь с шумом! (Noise Conditional Score Network)

$$\sum_{i=1}^{L} \lambda(i)\mathbb{E}_{p_{\sigma_i}(x)}[||\nabla_x \log p_{\sigma_i}(x) - s_\theta(x, i)||_2^2], \quad \lambda(i) = \sigma_i^2$$

А как это выглядит в итоге?



(a) MNIST (b) CelebA (c) CIFAR-10

(a) MNIST (b) CelebA (c) CIFAR-10

Хьюстон, у нас проблемы…

Чтобы всем вышеперечисленным пользоваться, нужно:

1) перебирать параметры зашумлений ручками
$\mathcal{N}(0, \sigma_1^2 I), ..., \mathcal{N}(0, \sigma_L^2 I), : \sigma_1 < ... < \sigma_L$
2) учить под каждый параметр нейронную сеть

Но ведь зашумление это же некоторый стохастический процесс!!!

$$dx = f(x, t)dt + g(t)dw$$

винеровский процесс

функция дрейфа

коэф. диффузии

А решения этого уравнения и есть наши $\mathcal{N}(0, \sigma_1^2 I), ..., \mathcal{N}(0, \sigma_L^2 I), : \sigma_1 < ... < \sigma_L$

И как же выбрать …

$$dx = f(x, t)dt + g(t)dw$$

выбери подходящую функцию

этот выбор определяет судьбу модели …

$$\mathcal{N}(0, \sigma_1^2 I), ..., \mathcal{N}(0, \sigma_L^2 I), : \sigma_1 < ... < \sigma_L$$

А как из шума получить данные в непрерывном мире?

это знаем

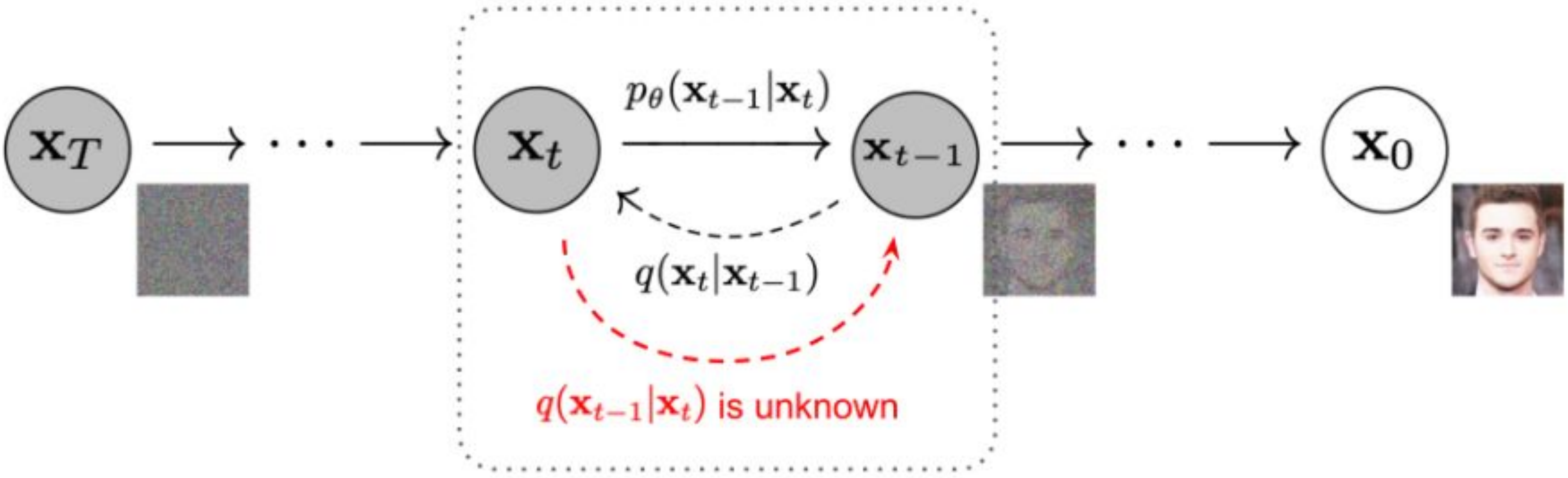
Reverse SDE

[Anderson, 1982] The Reverse SDE of forward SDE looks like as follow:

$$dx = [f(x,t) - g^2(t)\nabla_x \log p(x(t))]dt + g(t)dw$$

а как посчитать градиенты?
да мы уже умеем (Score Matching)

В общем виде диффузионки



Forward

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$$

$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) = \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1})$$

x_0 x x_T

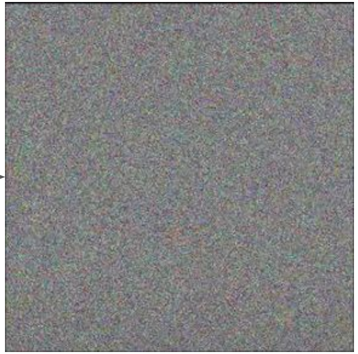
Faster forward

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$$

$$\{\beta_t \in (0,1)\}_{t=1}^T \qquad \alpha_t = 1 - \beta_t \qquad \bar{\alpha}_t = \prod_{i=1}^t \alpha_i:$$

$$q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I})$$

x_0



x_T

Faster forward

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

$$\boxed{x_{t-1} = \sqrt{\alpha_{t-1}}\, x_{t-2} + \sqrt{1 - \alpha_{t-1}}\, \epsilon_{t-2}}$$

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

$$\boxed{\bar{\epsilon}_{t-2} \in N\left(0, \left(\sqrt{1 - \alpha_t}\right)^2 + \left(\sqrt{1 - \alpha_{t-1}}\right)^2\right)}$$

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

$$= \sqrt{\alpha_t \alpha_{t-1} \alpha_{t-2}}\, x_{t-3} + \sqrt{1 - \alpha_t \alpha_{t-1} \alpha_{t-2}}\, \bar{\epsilon}_{t-3} =$$

$$= \boxed{\sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1 - \bar{\alpha}_t}\, \epsilon}$$

Faster forward

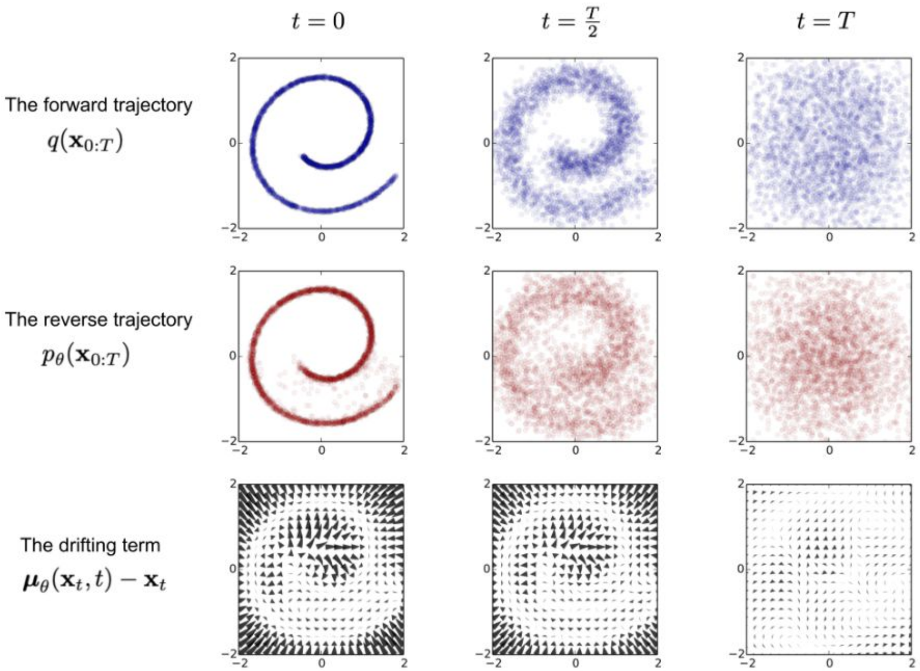
$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I}) \longrightarrow q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I})$$

Reverse

будем предсказывать
нейронкой

$$p_\theta(\mathbf{x}_{0:T}) = p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) \quad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boxed{\boldsymbol{\mu}_\theta(\mathbf{x}_t, t)}, \boxed{\boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t)})$$

это β_t из forward, его
мы задаем сами



Reverse

$$q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_{t-1}; \boxed{\tilde{\boldsymbol{\mu}}(\mathbf{x}_t, \mathbf{x}_0)}, \boxed{\tilde{\beta}_t}\mathbf{I})$$

$$\begin{aligned}
q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) &= q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0)\frac{q(\mathbf{x}_{t-1}|\mathbf{x}_0)}{q(\mathbf{x}_t|\mathbf{x}_0)} \\
&\propto \exp\Big(-\frac{1}{2}\big(\frac{(\mathbf{x}_t - \sqrt{\alpha_t}\mathbf{x}_{t-1})^2}{\beta_t} + \frac{(\mathbf{x}_{t-1} - \sqrt{\bar{\alpha}_{t-1}}\mathbf{x}_0)^2}{1-\bar{\alpha}_{t-1}} - \frac{(\mathbf{x}_t - \sqrt{\bar{\alpha}_t}\mathbf{x}_0)^2}{1-\bar{\alpha}_t}\big)\Big) \\
&= \exp\Big(-\frac{1}{2}\big(\frac{\mathbf{x}_t^2 - 2\sqrt{\alpha_t}\mathbf{x}_t\mathbf{x}_{t-1} + \alpha_t\mathbf{x}_{t-1}^2}{\beta_t} + \frac{\mathbf{x}_{t-1}^2 - 2\sqrt{\bar{\alpha}_{t-1}}\mathbf{x}_0\mathbf{x}_{t-1} + \bar{\alpha}_{t-1}\mathbf{x}_0^2}{1-\bar{\alpha}_{t-1}} - \frac{(\mathbf{x}_t - \sqrt{\bar{\alpha}_t}\mathbf{x}_0)^2}{1-\bar{\alpha}_t}\big)\Big) \\
&= \exp\Big(\boxed{-\frac{1}{2}\big((\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar{\alpha}_{t-1}})\mathbf{x}_{t-1}^2 - (\frac{2\sqrt{\alpha_t}}{\beta_t}\mathbf{x}_t + \frac{2\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}\mathbf{x}_0)\mathbf{x}_{t-1}} + C(\mathbf{x}_t, \mathbf{x}_0)\big)\Big)
\end{aligned}$$

$$\boxed{\begin{aligned}&f(x) = e^{ax^2+bx+c}, \\ &\text{где } a = -\frac{1}{2\sigma^2},\ b = \frac{\mu}{\sigma^2},\end{aligned}}\ c = -\left(\ln\sigma + \frac{1}{2}\ln 2\pi + \frac{1}{2}\frac{\mu^2}{\sigma^2}\right)$$

Reverse

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

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

$$\boxed{\tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t, \mathbf{x}_0) = (\frac{\sqrt{\alpha_t}}{\beta_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}\mathbf{x}_0)/(\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar{\alpha}_{t-1}})}$$

$$= (\frac{\sqrt{\alpha_t}}{\beta_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}\mathbf{x}_0)\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t} \cdot \beta_t$$

$$= \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\mathbf{x}_0$$

Reverse

$$\frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\boxed{\mathbf{x}_0}$$

$$\mathbf{x}_0 = \frac{1}{\sqrt{\bar{\alpha}_t}}(\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t)$$

$$\tilde{\boldsymbol{\mu}}_t = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\frac{1}{\sqrt{\bar{\alpha}_t}}(\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t)$$

$$= \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_t\right)$$

можно считать имея x_t и предсказывая ϵ_t

Reverse: variational lower bound

$$p_\theta(\mathbf{x}_{0:T}) = p(\mathbf{x}_T) \prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) \longrightarrow -\log p_\theta(\mathbf{x}_0)$$

хотим вычисл $p_\theta(\mathbf{x}_0)$
и минизирова
(т.е. максимизаровать макс. правдоподобие)

$$-\log p_\theta(\mathbf{x}_0) \le -\log p_\theta(\mathbf{x}_0) + \underbrace{\boxed{D_{\mathrm{KL}}(q(\mathbf{x}_{1:T}|\mathbf{x}_0)\|p_\theta(\mathbf{x}_{1:T}|\mathbf{x}_0))}}_{>0}$$

variational lower bound: минимизируя что-то большее, меньшая часть также будет уменьшаться

$$= -\log p_\theta(\mathbf{x}_0) + \mathbb{E}_{\mathbf{x}_{1:T}\sim q(\mathbf{x}_{1:T}|\mathbf{x}_0)}\Big[\log \frac{q(\mathbf{x}_{1:T}|\mathbf{x}_0)}{p_\theta(\mathbf{x}_{0:T})/p_\theta(\mathbf{x}_0)}\Big]$$

$$= -\log p_\theta(\mathbf{x}_0) + \mathbb{E}_q\Big[\log \frac{q(\mathbf{x}_{1:T}|\mathbf{x}_0)}{p_\theta(\mathbf{x}_{0:T})} + \log p_\theta(\mathbf{x}_0)\Big]$$

$$= \boxed{\mathbb{E}_q\Big[\log \frac{q(\mathbf{x}_{1:T}|\mathbf{x}_0)}{p_\theta(\mathbf{x}_{0:T})}\Big] = L_{\mathrm{VLB}}}$$

Reverse

константа-здесь нет обучаемых параметров

отдельно получаемый [декодер]

а вот это надо учить, каждое слагаемое-новая сеть (но так не хочется)

$$
\begin{aligned}
L_{\mathrm{VLB}} &= \mathbb{E}_{q(\mathbf{x}_{0:T})}\Big[\log \frac{q(\mathbf{x}_{1:T}|\mathbf{x}_0)}{p_\theta(\mathbf{x}_{0:T})}\Big] \\
&= \mathbb{E}_q\Big[\log \frac{\prod_{t=1}^T q(\mathbf{x}_t|\mathbf{x}_{t-1})}{p_\theta(\mathbf{x}_T)\prod_{t=1}^T p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)}\Big] \\
&= \mathbb{E}_q\Big[-\log p_\theta(\mathbf{x}_T) + \sum_{t=1}^T \log \frac{q(\mathbf{x}_t|\mathbf{x}_{t-1})}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)}\Big] \\
&= \mathbb{E}_q\Big[-\log p_\theta(\mathbf{x}_T) + \sum_{t=2}^T \log \frac{q(\mathbf{x}_t|\mathbf{x}_{t-1})}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} + \log \frac{q(\mathbf{x}_1|\mathbf{x}_0)}{p_\theta(\mathbf{x}_0|\mathbf{x}_1)}\Big] \\
&= \mathbb{E}_q\Big[-\log p_\theta(\mathbf{x}_T) + \sum_{t=2}^T \log \Big(\frac{q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} \cdot \frac{q(\mathbf{x}_t|\mathbf{x}_0)}{q(\mathbf{x}_{t-1}|\mathbf{x}_0)}\Big) + \log \frac{q(\mathbf{x}_1|\mathbf{x}_0)}{p_\theta(\mathbf{x}_0|\mathbf{x}_1)}\Big] \\
&= \mathbb{E}_q\Big[-\log p_\theta(\mathbf{x}_T) + \sum_{t=2}^T \log \frac{q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} + \sum_{t=2}^T \log \frac{q(\mathbf{x}_t|\mathbf{x}_0)}{q(\mathbf{x}_{t-1}|\mathbf{x}_0)} + \log \frac{q(\mathbf{x}_1|\mathbf{x}_0)}{p_\theta(\mathbf{x}_0|\mathbf{x}_1)}\Big] \\
&= \mathbb{E}_q\Big[-\log p_\theta(\mathbf{x}_T) + \sum_{t=2}^T \log \frac{q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} + \log \frac{q(\mathbf{x}_T|\mathbf{x}_0)}{q(\mathbf{x}_1|\mathbf{x}_0)} + \log \frac{q(\mathbf{x}_1|\mathbf{x}_0)}{p_\theta(\mathbf{x}_0|\mathbf{x}_1)}\Big] \\
&= \mathbb{E}_q\Big[\log \frac{q(\mathbf{x}_T|\mathbf{x}_0)}{p_\theta(\mathbf{x}_T)} + \sum_{t=2}^T \log \frac{q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} - \log p_\theta(\mathbf{x}_0|\mathbf{x}_1)\Big] \\
&= \boxed{\mathbb{E}_q[\underbrace{D_{\mathrm{KL}}(q(\mathbf{x}_T|\mathbf{x}_0)\parallel p_\theta(\mathbf{x}_T))}_{L_T}} + \boxed{\sum_{t=2}^T \underbrace{D_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)\parallel p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))}_{L_{t-1}}} \boxed{\underbrace{-\log p_\theta(\mathbf{x}_0|\mathbf{x}_1)}_{L_0}]}
\end{aligned}
$$

Reverse

$$L_t = D_{\mathrm{KL}}(q(\mathbf{x}_t|\mathbf{x}_{t+1}, \mathbf{x}_0) \parallel p_\theta(\mathbf{x}_t|\mathbf{x}_{t+1})) \quad 1 \le t \le T-1$$

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

$$\boldsymbol{\mu}_\theta(\mathbf{x}_t, t) = \frac{1}{\sqrt{\alpha_t}}\Big(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\Big)$$

$$\begin{aligned}
L_t &= \mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}\Big[\frac{1}{2\|\boldsymbol{\Sigma}_\theta(\mathbf{x}_t,t)\|_2^2}\|\tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t,\mathbf{x}_0) - \boldsymbol{\mu}_\theta(\mathbf{x}_t,t)\|^2\Big] \\
&= \mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}\Big[\frac{1}{2\|\boldsymbol{\Sigma}_\theta\|_2^2}\|\frac{1}{\sqrt{\alpha_t}}\Big(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_t\Big) - \frac{1}{\sqrt{\alpha_t}}\Big(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)\Big)\|^2\Big] \\
&= \mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}\Big[\frac{(1-\alpha_t)^2}{2\alpha_t(1-\bar{\alpha}_t)\|\boldsymbol{\Sigma}_\theta\|_2^2}\|\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)\|^2\Big] \\
&= \mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}\Big[\frac{(1-\alpha_t)^2}{2\alpha_t(1-\bar{\alpha}_t)\|\boldsymbol{\Sigma}_\theta\|_2^2}\|\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t, t)\|^2\Big]
\end{aligned}$$

Формула Байеса

Формула Байеса:

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)},$$

где

$P(A)$ — априорная вероятность гипотезы *A* (смысл такой терминологии см. ниже);
$P(A \mid B)$ — вероятность гипотезы *A* при наступлении события *B* (апостериорная вероятность);
$P(B \mid A)$ — вероятность наступления события *B* при истинности гипотезы *A*;
$P(B)$ — полная вероятность наступления события *B*.

Новая математика

- марковские цепи
- стохастические уравнения
- винеровский процесс
- energy base models
- score matching

А что почитать?

1) https://arxiv.org/pdf/2006.11239.pdf - 2020 год. Первая статья по диффузионкам, привлекшая большое внимание.

2) https://arxiv.org/pdf/2105.05233.pdf - 2021 год. Та самая "Diffusion models beat GANs on image synthesis" от авторов 3 статьи. Там всякие эвристики + classifier guidance.

3) https://arxiv.org/pdf/2011.13456.pdf - 2021 год. Про связь диффузии и score matching.

4) https://arxiv.org/pdf/2010.02502.pdf - изначально 2021 год. DDIM sampling.

5) https://arxiv.org/pdf/2106.15282.pdf - 2021 год. Каскадная диффузия с upsampling-ом.

6) https://arxiv.org/pdf/2112.10752.pdf - 2022 год. Статья по Stable Diffusion.

7) https://arxiv.org/pdf/2201.09865.pdf - 2022 год. Inpainting с помощью предобученной диффузионки.

8) https://arxiv.org/pdf/2207.12598.pdf - 2022 год. Classifier-free guidance.

https://lilianweng.github.io/posts/2021-07-11-diffusion-models/

https://www.youtube.com/watch?v=pEIwO-2vyus