Семинар 3

2024 курс: Генеративные модели

Семинарист: Кондратенко Владимир

$$1 + \underset{\substack{= \\ \text{tanh}}}{h'}(w^T z + b) w^T u \neq 0 \quad \forall z \in \mathbb{R}^d$$

$$\tanh = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

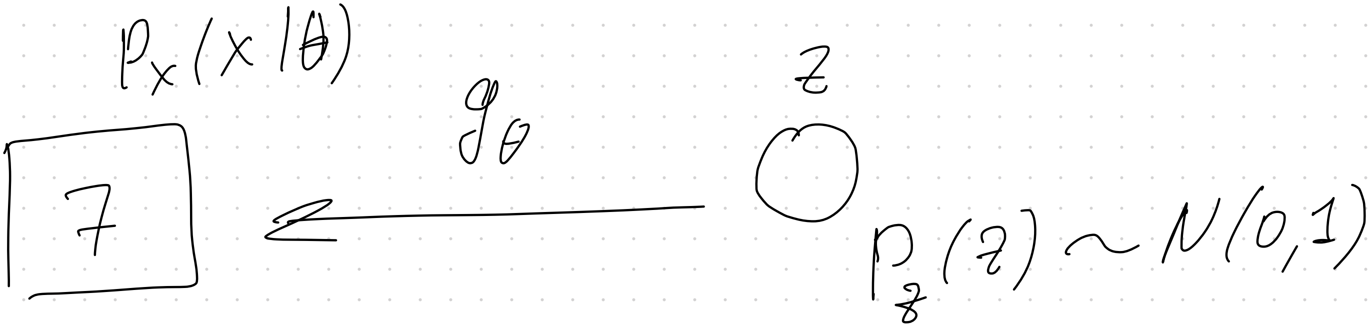
$$\boxed{0 < h' < 1 \quad \forall z}$$

$$1 + h'(\dots) w^T u > 0 \Rightarrow h'(\dots) w^T u > -1$$

$$\Downarrow$$

$$w^T u \geq -1 \qquad w^T u > \frac{-1}{h'}$$

$$h' \in (-\infty, -1)$$



$\pi(x)$ - страшное, непонятное, sampling

$p_z(x) \sim N(0,1)$

$\Pi(x)$

$$P(x|\theta) = P_z(f(x)) \cdot |\det J_f|$$

$$P(z|\theta) = P_x(g(z)) \cdot |\det J_g|$$

$$P(z) \sim N(0, 1)$$

$$\boxed{\begin{aligned} P_x\, dx &= P_z\, dz \\ P_x &= P_z \frac{dz}{dx} \end{aligned}}$$

Задача $\pi(x)$ - ?

$$p_x(x|\theta) \approx \pi(x)$$

1. $X \to z$, в z умеем оценивать плотн.

2. $z \to X$, p_z мы знаем

$$KL[\pi; p(x|\theta)] = \int \pi(x) \log \frac{\pi(x)}{p(x|\theta)} dx =$$

$$= -\int \pi(x) \log p(x|\theta) dx + \underbrace{\int \pi(x) \log \pi(x)}_{const}$$

$$-\mathbb{E}_{\pi(x)} \log p(x|\theta) dx = -\mathbb{E}_{\pi(x)}[\log p_z(f(x)) + \log|J_f|]$$

$$-\mathbb{E}_{\pi(x)}\left[\log p_z(f(x)) + \log|J_f|\right]$$

FKL

$\mathbb{E}_{\pi(x)}$ – семплирование из $\pi(x)$

$z = f(x)$ – поток

↳ $x \to z$

$p_z(z = f(x))$ – правдоподобие $z \sim N(0,1)$

Что нужно?

1. $z = f(x)$
2. sampling $x \sim \pi(x)$
3. $p_z(z)$ ↳ плотность
4. считать $\det J_f$

$g(z) = x = f^{-1}(z)$ в явном виде не нужно

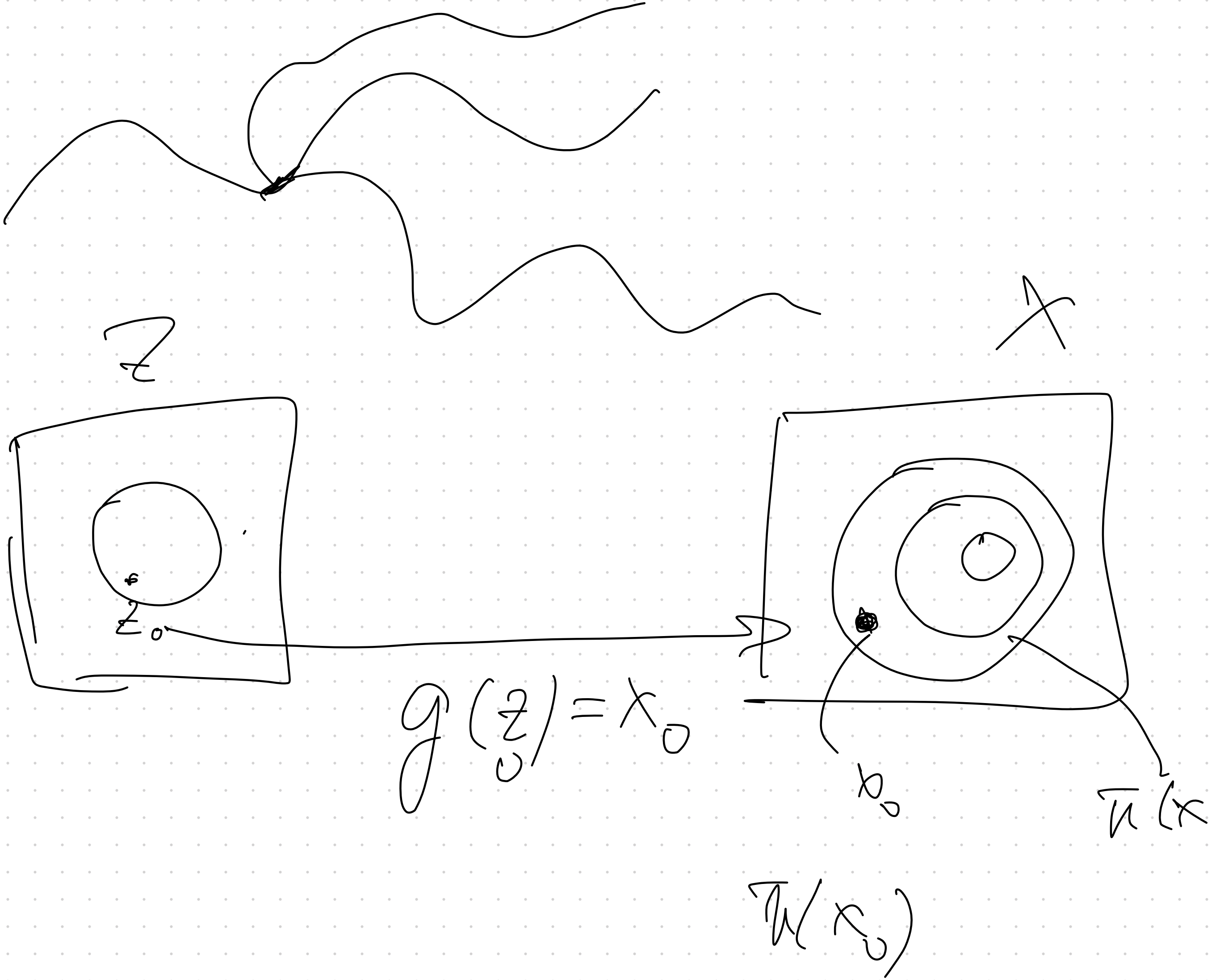
Reverse KL

$$KL[p(x|\theta), \pi(x)] = KL[p(z); p(z|\theta)]$$

$$\mathbb{E}_{p(z)}\left[\log p_z(z) - \log|\det J_g| - \log \pi(g(z))\right]$$

1. $g(z) = x$
2. $\det J_g$
3. $\pi(g(z))$ — умеем оценивать плотность
4. sampling z
5. $p_z(z)$

$\boxed{\text{не MLE}}$



Z
X
z_0
$g(z_0) = x_0$
x_0
$\pi(x)$
$\pi(x_0)$