Title: Diffusion Generative Models and potential applications in physics

Speakers: Kirill Neklyudov

Series: Machine Learning Initiative

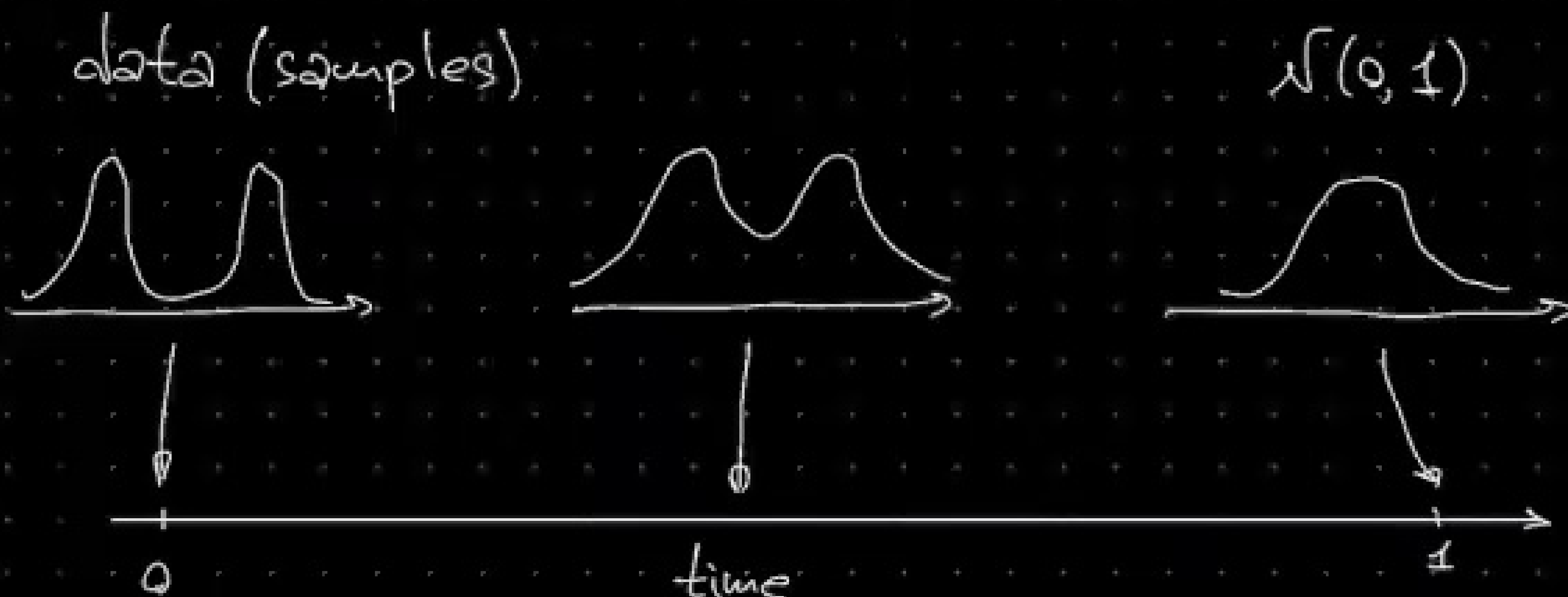
Date: March 24, 2023 - 11:00 AM

URL: https://pirsa.org/23030113

Abstract: Generative modeling via diffusion processes is already a vast field of literature. In this introduction, I will give an entry point to this field by going over the main concepts and deriving the essential results of the area. Thus, by the end of the talk, we would have a minimal pipeline for implementing the generative model. Furthermore, I will outline several alternative ways for learning such models that the community has developed in recent years. These directions bring novel perspectives and new capabilities of generative modeling.

Zoom link: https://pitp.zoom.us/j/98786491081?pwd=U1cvZzBQT2VUZDl5Ykd0c1lqY29aZz09

Setup



forward process $\rightarrow$

- fixed (not trainable)

Forward process (diffusion)

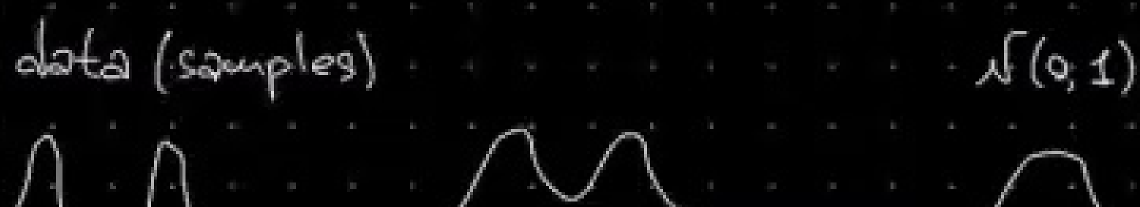
$$\dot{q}_t(x) = -\langle \nabla, q_t(x)\, u_t(x) \rangle + \frac{\beta_t}{2} \Delta q_t(x)$$

(Fokker-Planck equation)

density

drift

$$q_0(x) \rightsquigarrow q_{dt}(x)$$



If we don't have any noise then all the samples $q_t \to \delta(x)$. However, since we add noise, $q_t \to \mathcal{N}(0,1)$.

Conditional distribution

Given $x_0 \sim q_0$. What's $q_t(x_t | x_0)$?

In other words, $q_0 = \delta(x - x_0)$, find $q_t(x)$.

We know

<u>Exercise</u> $\ldots \int_0^t d\tau \, \beta_\tau \Big)$.

Now let's guess

$$q_t(x) = \mathcal{N}\left(x \,\Big|\, x_0 \exp\Big(-\tfrac{1}{2}\int_0^t d\tau \, \beta_\tau\Big), 1 - \exp\Big(-\int_0^t d\tau \, \beta_\tau\Big)\right)$$

<u>Exercise</u> check that this is indeed a solution of Fokker-Planck

Now we have, for $q_0(x) = \delta(x - x_0)$,

$$q_t(x) = \mathcal{N}\left(x \,\Big|\, x_0 \exp\Big(-\tfrac{1}{2}\int_0^t d\tau \, \beta_\tau\Big), 1 - \exp\Big(-\int_0^t d\tau \, \beta_\tau\Big)\right)$$

In other w[...] $q_t(x)$.

We know

1D

$$\dot{q}_t = -\partial_x\left(q_t u_t\right) + \frac{\beta_t}{2}\partial^2_{xx} q_t =$$

$$= -\partial_x\left(q_t\left(-\frac{1}{2}\beta_t x\right)\right) + \frac{\beta_t}{2}\partial^2_{xx} q_t$$

Let's see how the mean changes

$$\mu_t = \mathbb{E}\, q_t = \int_{\mathbb{R}^d} dx\, q_t(x)\cdot x$$

$$\frac{d}{}(\mu_t) = \int dx\, \frac{d}{}\left(q_t(x)\, x\right)$$

For $q_0(x) = p(x)$ (data distribution), we have

$$q_t(x) = \int dx_0 \; p(x_0) \; \underbrace{k_t(x|x_0)}_{\mathcal{N}\left(x \,|\, x_0 \exp\left(-\frac{1}{2}\int_0^t d\tau\, \beta_\tau\right),\, 1-\exp\left(-\int_0^t d\tau\, \beta_\tau\right)\right)}$$

$$\mathcal{N}(x \,|\, x_0 \cdot \xi_t,\, 1-\eta_t)$$

Note that we can easily sample from $q_t(x)$!

1. sample $x_0 \sim p(x)$
2. sample $x_t \sim k(x|x_0)$

09:41 Tue 9 Jan

100 %

Note that we can easily sample from $q_t(x)$!

1. sample $x_0 \sim p(x)$
2. sample $x_t \sim k(x \mid x_0)$

Thus, we have defined the forward process!

Reversing the diffusion

$$\dot{q}_t = -\langle \nabla, q_t u_t \rangle + \beta_t \Delta q_t = \qquad (\text{FP eq.})$$

$$= -\langle \nabla q_t u_t \rangle + \beta_t \langle \nabla, \nabla q_t \rangle =$$

For $q_0(x) = p(x)$ (data distribution), we have

$$q_t(x) = \int dx_0 \, p(x_0) \, k_t(x \mid x_0)$$

$$\mathcal{N}\left(x \mid x_0 \exp\left(-\frac{1}{2}\int_0^t d\tau \, \beta_\tau\right), 1 - \exp\left(-\int_0^t d\tau \, \beta_\tau\right)\right)$$

$$\mathcal{N}(x \mid x_0 \cdot \xi_t, 1 - \eta_t)$$

Note that we can easily sample from $q_t(x)$!

1. sample $x_0 \sim p(x)$
2. sample $x_t \sim k(x \mid x_0)$

Note that we can easily sample from $q_t(x)$!

1. sample $x_0 \sim p(x)$
2. sample $x_t \sim k(x \mid x_0)$

Thus, we have defined the forward process!

Reversing the diffusion

$$\dot{q}_t = -\langle \nabla, q_t u_t \rangle + \beta_t \Delta q_t = \qquad (\text{FP eq.})$$

$$= -\langle \nabla q_t, u_t \rangle + \beta_t \langle \nabla, \nabla q_t \rangle =$$

v_t

$$= -\langle \nabla, q_t v_t \rangle \qquad \text{(the continuity eq.)}$$

It means that the samples evolve as

$$\frac{dx}{dt} = v_t(x)$$

$$x(t=1) = x(t=0) + \int_0^1 dt\, v_t(x(t)) \qquad \text{forward integration}$$

$x(t=0) = x(t$ … rd integration

Hence, our goal ∇

$$v_t = \overbrace{-\frac{1}{2}\beta_t x}^{u_t} + \beta_t \boxed{\nabla \log q_t}$$

Score matching

Given samples from $q(x)$, estimate $\nabla \log q(x)$

$$\text{Loss} = \frac{1}{2}\mathbb{E}_q \|\vec{s}(x,\theta) - \nabla \log q(x)\|^2 =$$

09:41 Tue 9 Jan 100 %

$$\ominus \ \frac{1}{2}\,\mathbb{E}_q \|s\|^2 + \mathbb{E}_q \langle \nabla, s \rangle =$$

$$= \mathbb{E}_q \left[\frac{1}{2} \|s(x,\theta)\|^2 + \langle \nabla, s(x,\theta) \rangle \right]$$

Depends only on samples!

<u>Exercise</u> prove denoising score matching: when

$q(x) = \int dx_0 \, p(x_0)\, k(x|x_0)$, then

$$\text{Loss} = \mathbb{E} \quad \| s(x,\theta) - \nabla \log k(x|x_0) \|^2 \sim$$

09:41 Tue 9 Jan ••• 100 %

Summary

- Using the conditional $k_t(x|x_0)$ and the samples from data $p(x_0)$, we can learn $\nabla \log q_t$ $\forall t$.

$\text{Loss} : \| \mathbb{E} \quad s(x, \theta, t) - \nabla \log k_t(x|x_0) \|^2$

09:41 Tue 9 Jan 100 %

- Using the conditional $k_t(x|x_0)$ and the samples from data $p(x_0)$, we can learn $\nabla \log q_t$ $\forall t$.

$$\text{Loss} = \int dt \, \mathbb{E}_{p(x_0)\, k_t(x|x_0)} \| s(x, \theta, t) - \nabla_x \log k_t(x|x_0) \|^2$$

- Then we know the vector field

$$v_t(x) = -\frac{1}{2} \beta_t x + \nabla \log q_t$$

Action Matching

- What if we have only samples $x \sim q_t(x)\ \forall t$?

W_2^2 - space