

Deep Generative Models

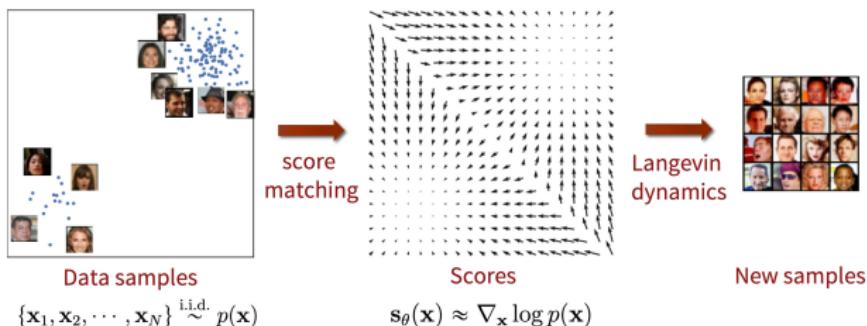
Lecture 12

Roman Isachenko

Moscow Institute of Physics and Technology

2023, Autumn

Recap of previous lecture



Theorem (implicit score matching)

$$\frac{1}{2} \mathbb{E}_\pi \|s_\theta(\mathbf{x}) - \nabla_{\mathbf{x}} \log \pi(\mathbf{x})\|_2^2 = \mathbb{E}_\pi \left[\frac{1}{2} \|s_\theta(\mathbf{x})\|_2^2 + \text{tr}(\nabla_{\mathbf{x}} s_\theta(\mathbf{x})) \right] + \text{const}$$

1. The left hand side is intractable due to unknown $\pi(\mathbf{x})$ – **denoising score matching**.
2. The right hand side is complex due to Hessian matrix – **sliced score matching (Hutchinson's trace estimation)**.

Recap of previous lecture

Let perturb original data by normal noise $p(\mathbf{x}'|\mathbf{x}, \sigma) = \mathcal{N}(\mathbf{x}'|\mathbf{x}, \sigma^2 \mathbf{I})$

$$\pi(\mathbf{x}'|\sigma) = \int \pi(\mathbf{x}) p(\mathbf{x}'|\mathbf{x}, \sigma) d\mathbf{x}.$$

Then the solution of

$$\frac{1}{2} \mathbb{E}_{\pi(\mathbf{x}'|\sigma)} \| \mathbf{s}_\theta(\mathbf{x}', \sigma) - \nabla_{\mathbf{x}'} \log \pi(\mathbf{x}'|\sigma) \|_2^2 \rightarrow \min_{\theta}$$

satisfies $\mathbf{s}_\theta(\mathbf{x}', \sigma) \approx \mathbf{s}(\mathbf{x}', \theta, 0) = \mathbf{s}(\mathbf{x}', \theta)$ if σ is small enough.

Theorem (denoising score matching)

$$\begin{aligned} \mathbb{E}_{\pi(\mathbf{x}'|\sigma)} \| \mathbf{s}_\theta(\mathbf{x}', \sigma) - \nabla_{\mathbf{x}'} \log \pi(\mathbf{x}'|\sigma) \|_2^2 &= \\ &= \mathbb{E}_{\pi(\mathbf{x})} \mathbb{E}_{p(\mathbf{x}'|\mathbf{x}, \sigma)} \| \mathbf{s}_\theta(\mathbf{x}', \sigma) - \nabla_{\mathbf{x}'} \log p(\mathbf{x}'|\mathbf{x}, \sigma) \|_2^2 + \text{const}(\theta) \end{aligned}$$

Here $\nabla_{\mathbf{x}'} \log p(\mathbf{x}'|\mathbf{x}, \sigma) = -\frac{\mathbf{x}' - \mathbf{x}}{\sigma^2}$.

- ▶ The RHS does not need to compute $\nabla_{\mathbf{x}'} \log \pi(\mathbf{x}'|\sigma)$ and even more $\nabla_{\mathbf{x}'} \log \pi(\mathbf{x}')$.
- ▶ $\mathbf{s}_\theta(\mathbf{x}', \sigma)$ tries to **denoise** a corrupted sample.
- ▶ Score function $\mathbf{s}_\theta(\mathbf{x}', \sigma)$ parametrized by σ .

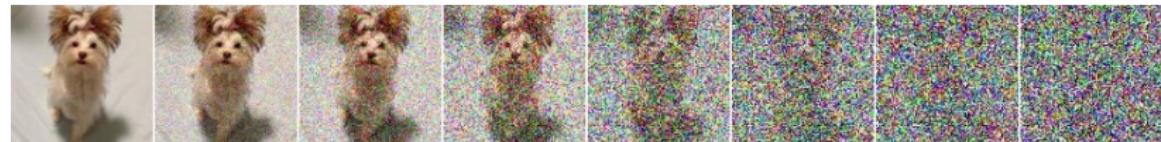
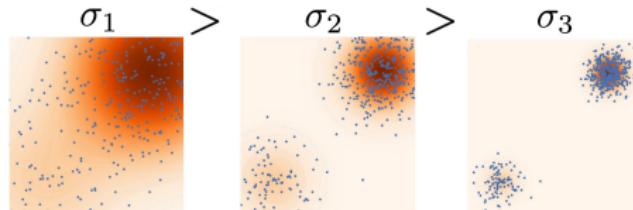
Recap of previous lecture

Noise conditioned score network

- ▶ Define the sequence of noise levels: $\sigma_1 > \sigma_2 > \dots > \sigma_L$.
- ▶ Train denoised score function $s_\theta(\mathbf{x}', \sigma)$ for each noise level:

$$\sum_{l=1}^L \sigma_l^2 \mathbb{E}_{\pi(\mathbf{x})} \mathbb{E}_{p(\mathbf{x}'|\mathbf{x}, \sigma_l)} \| s_\theta(\mathbf{x}', \sigma_l) - \nabla'_{\mathbf{x}} \log p(\mathbf{x}'|\mathbf{x}, \sigma_l) \|_2^2 \rightarrow \min_{\theta}$$

- ▶ Sample from **annealed** Langevin dynamics (for $l = 1, \dots, L$).



Outline

1. Gaussian diffusion process
2. Denoising diffusion probabilistic model (DDPM)
 - Objective of DDPM
 - Reparametrization of DDPM
3. Overview of DDPM and link to score matching
4. The worst course overview

Outline

1. Gaussian diffusion process
2. Denoising diffusion probabilistic model (DDPM)
 - Objective of DDPM
 - Reparametrization of DDPM
3. Overview of DDPM and link to score matching
4. The worst course overview

Forward gaussian diffusion process

Let $\mathbf{x}_0 = \mathbf{x} \sim \pi(\mathbf{x})$, $\beta \in (0, 1)$. Define the Markov chain

$$\mathbf{x}_t = \sqrt{1 - \beta} \cdot \mathbf{x}_{t-1} + \sqrt{\beta} \cdot \boldsymbol{\epsilon}, \quad \text{where } \boldsymbol{\epsilon} \sim \mathcal{N}(0, 1);$$

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

Statement 1

Applying the Markov chain to samples from any $\pi(\mathbf{x})$ we will get $\mathbf{x}_\infty \sim p_\infty(\mathbf{x}) = \mathcal{N}(0, 1)$. Here $p_\infty(\mathbf{x})$ is a **stationary** distribution:

$$p_\infty(\mathbf{x}) = \int q(\mathbf{x} | \mathbf{x}') p_\infty(\mathbf{x}') d\mathbf{x}'.$$

Statement 2

Denote $\bar{\alpha}_t = \prod_{s=1}^t (1 - \beta_s)$. Then

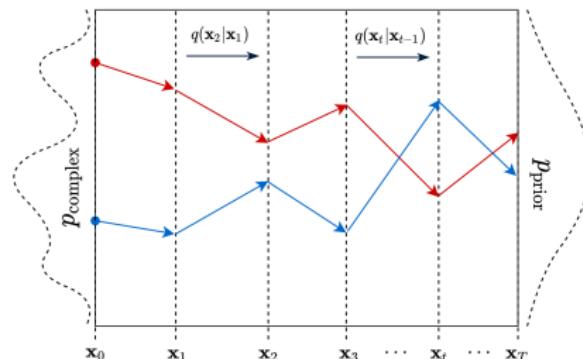
$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t} \cdot \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \cdot \boldsymbol{\epsilon}, \quad \text{where } \boldsymbol{\epsilon} \sim \mathcal{N}(0, 1)$$

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

We could sample from any timestamp using only \mathbf{x}_0 !

Forward gaussian diffusion process

Diffusion refers to the flow of particles from high-density regions towards low-density regions.

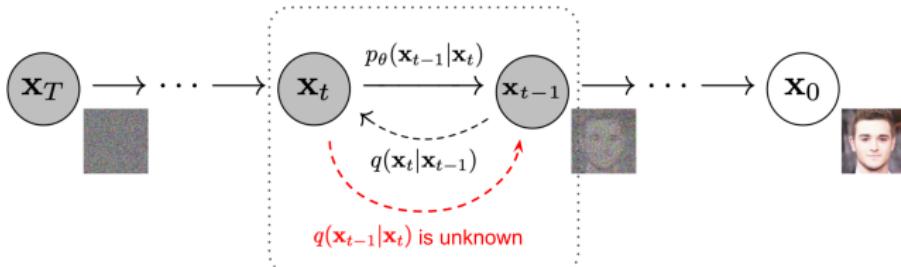


1. $\mathbf{x}_0 = \mathbf{x} \sim \pi(\mathbf{x});$
2. $\mathbf{x}_t = \sqrt{1 - \beta} \cdot \mathbf{x}_{t-1} + \sqrt{\beta} \cdot \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim \mathcal{N}(0, 1)$, $t \geq 1$;
3. $\mathbf{x}_T \sim p_\infty(\mathbf{x}) = \mathcal{N}(0, 1)$, where $T \gg 1$.

If we are able to invert this process, we will get the way to sample $\mathbf{x} \sim \pi(\mathbf{x})$ using noise samples $p_\infty(\mathbf{x}) = \mathcal{N}(0, 1)$.

Now our goal is to revert this process.

Reverse gaussian diffusion process



Let define the reverse process

$$p(\mathbf{x}_{t-1}|\mathbf{x}_t, \theta) = \mathcal{N}(\mathbf{x}_{t-1}|\mu_\theta(\mathbf{x}_t, t), \sigma_\theta^2(\mathbf{x}_t, t))$$

Forward process

1. $\mathbf{x}_0 = \mathbf{x} \sim \pi(\mathbf{x});$

2. $\mathbf{x}_t = \sqrt{1 - \beta} \cdot \mathbf{x}_{t-1} + \sqrt{\beta} \cdot \boldsymbol{\epsilon},$ where $\boldsymbol{\epsilon} \sim \mathcal{N}(0, 1), t \geq 1;$

3. $\mathbf{x}_T \sim p_\infty(\mathbf{x}) = \mathcal{N}(0, 1).$

Reverse process

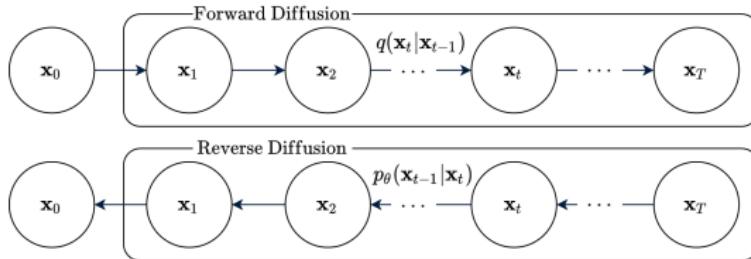
1. $\mathbf{x}_T \sim p_\infty(\mathbf{x}) = \mathcal{N}(0, 1);$

2. $\mathbf{x}_{t-1} = \sigma_\theta(\mathbf{x}_t, t) \cdot \boldsymbol{\epsilon} + \mu_\theta(\mathbf{x}_t, t);$

3. $\mathbf{x}_0 = \mathbf{x} \sim \pi(\mathbf{x});$

Note: The forward process does not have any learnable parameters!

Gaussian diffusion model as VAE



- ▶ Let treat $\mathbf{z} = (x_1, \dots, x_T)$ as a latent variable (**note**: each x_t has the same size).
- ▶ Variational posterior distribution (**note**: there is no learnable parameters)

$$q(\mathbf{z}|\mathbf{x}) = q(x_1, \dots, x_T | x_0) = \prod_{t=1}^T q(x_t | x_{t-1}).$$

- ▶ Probabilistic model

$$p(\mathbf{x}, \mathbf{z} | \boldsymbol{\theta}) = p(\mathbf{x} | \mathbf{z}, \boldsymbol{\theta}) p(\mathbf{z} | \boldsymbol{\theta})$$

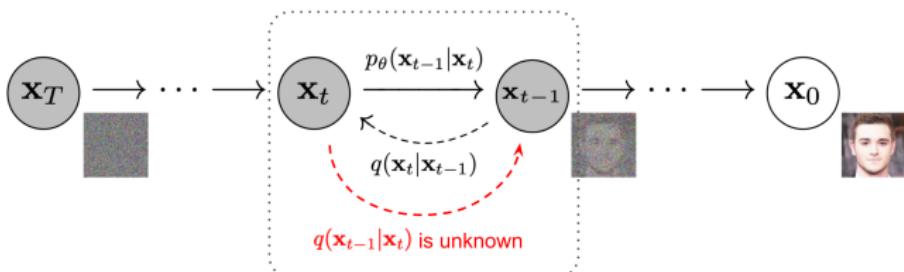
- ▶ Generative distribution and prior

$$p(\mathbf{x} | \mathbf{z}, \boldsymbol{\theta}) = p(x_0 | x_1, \boldsymbol{\theta}); \quad p(\mathbf{z} | \boldsymbol{\theta}) = \prod_{t=2}^T p(x_{t-1} | x_t, \boldsymbol{\theta}) \cdot p(x_T)$$

Outline

1. Gaussian diffusion process
2. Denoising diffusion probabilistic model (DDPM)
 - Objective of DDPM
 - Reparametrization of DDPM
3. Overview of DDPM and link to score matching
4. The worst course overview

Reverse gaussian diffusion process



Forward process

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

Reverse process

$$q(\mathbf{x}_{t-1} | \mathbf{x}_t) = \frac{q(\mathbf{x}_t | \mathbf{x}_{t-1}) q(\mathbf{x}_{t-1})}{q(\mathbf{x}_t)} \approx p(\mathbf{x}_{t-1} | \mathbf{x}_t, \theta)$$

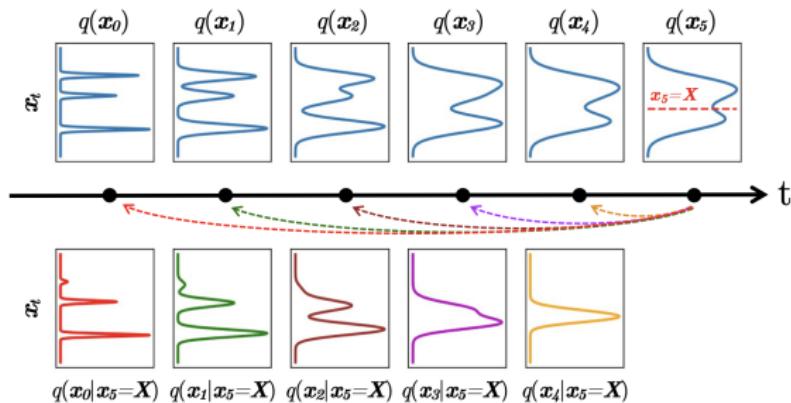
- ▶ $q(\mathbf{x}_{t-1})$, $q(\mathbf{x}_t)$ are intractable.
- ▶ If β_t is small enough, $q(\mathbf{x}_{t-1} | \mathbf{x}_t)$ will be Gaussian (Feller, 1949).

Feller W. On the theory of stochastic processes, with particular reference to applications, 1949

Reverse gaussian diffusion process

Reverse process

$$q(\mathbf{x}_{t-1}|\mathbf{x}_t) \approx p(\mathbf{x}_{t-1}|\mathbf{x}_t, \theta) = \mathcal{N}(\mathbf{x}_{t-1} | \mu_\theta(\mathbf{x}_t, t), \sigma_\theta^2(\mathbf{x}_t, t))$$



Important distribution

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

Outline

1. Gaussian diffusion process
2. Denoising diffusion probabilistic model (DDPM)
 - Objective of DDPM
 - Reparametrization of DDPM
3. Overview of DDPM and link to score matching
4. The worst course overview

Objective of DDPM

ELBO

$$\log p(\mathbf{x}|\theta) \geq \mathbb{E}_{q(\mathbf{z}|\mathbf{x})} \log \frac{p(\mathbf{x}, \mathbf{z}|\theta)}{q(\mathbf{z}|\mathbf{x})} = \mathcal{L}(q, \theta) \rightarrow \max_{q, \theta}$$

Derivation

$$\begin{aligned}\mathcal{L}(q, \theta) &= \mathbb{E}_{q(\mathbf{x}_{1:T}|\mathbf{x}_0)} \log \frac{p(\mathbf{x}_0, \mathbf{x}_{1:T}|\theta)}{q(\mathbf{x}_{1:T}|\mathbf{x}_0)} = \mathbb{E}_q \log \frac{\prod_{t=1}^T p(\mathbf{x}_{t-1}|\mathbf{x}_t, \theta) p(\mathbf{x}_T)}{\prod_{t=1}^T q(\mathbf{x}_t|\mathbf{x}_{t-1})} \\ &= \mathbb{E}_q \left[\log p(\mathbf{x}_T) + \log \frac{p(\mathbf{x}_0|\mathbf{x}_1, \theta)}{q(\mathbf{x}_1|\mathbf{x}_0)} + \sum_{t=2}^T \log \left(\frac{p(\mathbf{x}_{t-1}|\mathbf{x}_t, \theta)}{q(\mathbf{x}_t|\mathbf{x}_{t-1})} \right) \right] \\ q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) &= \frac{q(\mathbf{x}_t|\mathbf{x}_{t-1})q(\mathbf{x}_{t-1}|\mathbf{x}_0)}{q(\mathbf{x}_t|\mathbf{x}_0)} \\ \log \left(\frac{p(\mathbf{x}_{t-1}|\mathbf{x}_t, \theta)}{q(\mathbf{x}_t|\mathbf{x}_{t-1})} \right) &= \log \left(\frac{p(\mathbf{x}_{t-1}|\mathbf{x}_t, \theta)q(\mathbf{x}_{t-1}|\mathbf{x}_0)}{q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)q(\mathbf{x}_t|\mathbf{x}_0)} \right)\end{aligned}$$

Objective of DDPM

Derivation

$$\begin{aligned}\mathcal{L}(q, \theta) &= \mathbb{E}_q \left[\log p(\mathbf{x}_T) + \log \frac{p(\mathbf{x}_0|\mathbf{x}_1, \theta)}{q(\mathbf{x}_1|\mathbf{x}_0)} + \sum_{t=2}^T \log \left(\frac{p(\mathbf{x}_{t-1}|\mathbf{x}_t, \theta)}{q(\mathbf{x}_t|\mathbf{x}_{t-1})} \right) \right] \\ &= \mathbb{E}_q \left[\log p(\mathbf{x}_T) + \log \frac{p(\mathbf{x}_0|\mathbf{x}_1, \theta)}{q(\mathbf{x}_1|\mathbf{x}_0)} + \sum_{t=2}^T \log \left(\frac{p(\mathbf{x}_{t-1}|\mathbf{x}_t, \theta)}{q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)} \right) + \right. \\ &\quad \left. + \sum_{t=2}^T \log \left(\frac{q(\mathbf{x}_{t-1}|\mathbf{x}_0)}{q(\mathbf{x}_t|\mathbf{x}_0)} \right) \right] = \mathbb{E}_q \left[\log \frac{p(\mathbf{x}_T)}{q(\mathbf{x}_T|\mathbf{x}_0)} + \log \frac{p(\mathbf{x}_0|\mathbf{x}_1, \theta)}{q(\mathbf{x}_1|\mathbf{x}_0)} + \right. \\ &\quad \left. + \sum_{t=2}^T \log \left(\frac{p(\mathbf{x}_{t-1}|\mathbf{x}_t, \theta)}{q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)} \right) \right] = \mathbb{E}_q \left[-KL(q(\mathbf{x}_T|\mathbf{x}_0)||p(\mathbf{x}_T)) + \right. \\ &\quad \left. + \log p(\mathbf{x}_0|\mathbf{x}_1, \theta) - \sum_{t=2}^T KL(q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)||p(\mathbf{x}_{t-1}|\mathbf{x}_t, \theta)) \right]\end{aligned}$$

Objective of DDPM

$$\mathcal{L}(q, \theta) = \mathbb{E}_q \left[\log p(\mathbf{x}_0 | \mathbf{x}_1, \theta) - KL(q(\mathbf{x}_T | \mathbf{x}_0) || p(\mathbf{x}_T)) - \sum_{t=2}^T \underbrace{KL(q(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0) || p(\mathbf{x}_{t-1} | \mathbf{x}_t, \theta))}_{\mathcal{L}_t} \right]$$

- ▶ First term is a decoder distribution

$$\log p(\mathbf{x}_0 | \mathbf{x}_1, \theta) = \log \mathcal{N}(\mathbf{x}_0 | \mu_\theta(\mathbf{x}_1, t), \sigma_\theta^2(\mathbf{x}_1, t))$$

- ▶ Second term is constant ($p(\mathbf{x}_T)$ is a standard Normal, $q(\mathbf{x}_T | \mathbf{x}_0)$ is a non-parametrical Normal).
- ▶ \mathcal{L}_t is a KL between two normal distributions:

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

$\tilde{\mu}_t(\mathbf{x}_t, \mathbf{x}_0)$ and $\tilde{\beta}_t$ have analytical formulas (we omit it) and they are both dependent on β_t .

Outline

1. Gaussian diffusion process
2. Denoising diffusion probabilistic model (DDPM)
 - Objective of DDPM
 - Reparametrization of DDPM
3. Overview of DDPM and link to score matching
4. The worst course overview

Gaussian diffusion model as VAE

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

$$p(\mathbf{x}_{t-1} | \mathbf{x}_t, \theta) = \mathcal{N}(\mathbf{x}_{t-1} | \mu_\theta(\mathbf{x}_t, t), \sigma_\theta^2(\mathbf{x}_t, t))$$

- ▶ Assume $\sigma_\theta^2(\mathbf{x}_t, t) = \tilde{\beta}_t \mathbf{I}$.
- ▶ Use KL formula between two normal distributions:

$$\mathcal{L}_t = KL\left(\mathcal{N}(\tilde{\mu}_t(\mathbf{x}_t, \mathbf{x}_0), \tilde{\beta}_t \mathbf{I}) || \mathcal{N}(\mu_\theta(\mathbf{x}_t, t), \tilde{\beta}_t \mathbf{I})\right)$$

$$= \mathbb{E}_\epsilon \left[\frac{1}{2\tilde{\beta}_t} \|\tilde{\mu}_t(\mathbf{x}_t, \mathbf{x}_0) - \mu_\theta(\mathbf{x}_t, t)\|^2 \right]$$

$$= \mathbb{E}_\epsilon \left[\frac{1}{2\tilde{\beta}_t} \left\| \frac{1}{\sqrt{1-\beta_t}} \left(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \epsilon \right) - \mu_\theta(\mathbf{x}_t, t) \right\|^2 \right]$$

Here we used the analytic expression for $\tilde{\mu}_t(\mathbf{x}_t, \mathbf{x}_0)$.

Reparametrization

$$\mu_\theta(\mathbf{x}_t, t) = \frac{1}{\sqrt{1-\beta_t}} \left(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \epsilon_\theta(\mathbf{x}_t, t) \right)$$

Reparametrization of DDPM

KL term

$$\begin{aligned}\mathcal{L}_t = \mathbb{E}_{\epsilon} \left[\frac{1}{2\tilde{\beta}_t} \left\| \frac{1}{\sqrt{1-\beta_t}} \left(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \epsilon \right) - \right. \right. \\ \left. \left. - \frac{1}{\sqrt{1-\beta_t}} \left(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \epsilon_{\theta}(\mathbf{x}_t, t) \right) \right\|^2 \right] = \\ \mathbb{E}_{\epsilon} \left[\frac{\beta_t^2}{2\tilde{\beta}_t(1-\beta_t)} \left\| \frac{\epsilon}{\sqrt{1-\bar{\alpha}_t}} - \frac{\epsilon_{\theta}(\textcolor{teal}{x}_t, t)}{\sqrt{1-\bar{\alpha}_t}} \right\|^2 \right]\end{aligned}$$

$$\textcolor{teal}{x}_t = \sqrt{\bar{\alpha}_t} \cdot \mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t} \cdot \epsilon, \quad \text{where } \epsilon \sim \mathcal{N}(0, 1)$$

At each step of reverse diffusion process we try to predict the noise ϵ that we used in forward process!

Outline

1. Gaussian diffusion process
2. Denoising diffusion probabilistic model (DDPM)
 - Objective of DDPM
 - Reparametrization of DDPM
3. Overview of DDPM and link to score matching
4. The worst course overview

Denoising diffusion probabilistic model (DDPM)

DDPM is a VAE model

- ▶ Encoder is a fixed Gaussian Markov chain.
- ▶ Latent variable is a hierarchical (in each step the dim. of the latent equals to the dim of the input).
- ▶ Decoder is a simple Gaussian model.
- ▶ Prior distribution is given by parametric Gaussian Makov chain.

Forward process

1. $\mathbf{x}_0 = \mathbf{x} \sim \pi(\mathbf{x});$
2. $\mathbf{x}_t = \sqrt{1 - \beta} \cdot \mathbf{x}_{t-1} + \sqrt{\beta} \cdot \boldsymbol{\epsilon},$
where $\boldsymbol{\epsilon} \sim \mathcal{N}(0, 1), t \geq 1;$
3. $\mathbf{x}_T \sim p_\infty(\mathbf{x}) = \mathcal{N}(0, 1).$

Reverse process

1. $\mathbf{x}_T \sim p_\infty(\mathbf{x}) = \mathcal{N}(0, 1);$
2. $\mathbf{x}_{t-1} = \sigma_\theta(\mathbf{x}_t, t) \cdot \boldsymbol{\epsilon} + \mu_\theta(\mathbf{x}_t, t);$
3. $\mathbf{x}_0 = \mathbf{x} \sim \pi(\mathbf{x});$

Gaussian diffusion model vs Score matching

$$\mathcal{L}_t = \mathbb{E}_{\epsilon} \left[\frac{\beta_t^2}{2\tilde{\beta}_t(1-\beta_t)} \left\| \frac{\epsilon}{\sqrt{1-\bar{\alpha}_t}} - \frac{\epsilon_{\theta}(\mathbf{x}_t, t)}{\sqrt{1-\bar{\alpha}_t}} \right\|^2 \right]$$

- ▶ Result from Statement 2

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

- ▶ Score of noised distribution

$$\nabla_{\mathbf{x}_t} \log q(\mathbf{x}_t | \mathbf{x}_0) = -\frac{\epsilon}{\sqrt{1-\bar{\alpha}_t}}, \quad \text{where } \epsilon \sim \mathcal{N}(0, 1).$$

- ▶ Let reparametrize our model:

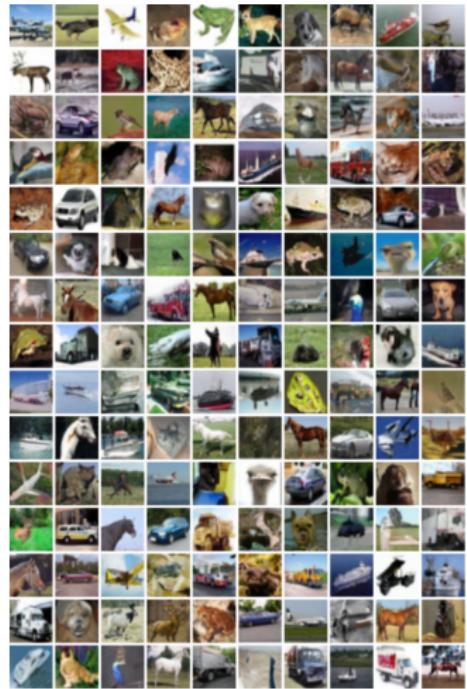
$$\mathbf{s}_{\theta}(\mathbf{x}_t, t) = -\frac{\epsilon_{\theta}(\mathbf{x}_t, t)}{\sqrt{1-\bar{\alpha}_t}}.$$

Noise conditioned score network

$$\mathbb{E}_{p(\mathbf{x}'|\mathbf{x}, \sigma_I)} \left\| \mathbf{s}(\mathbf{x}', \theta, \sigma_I) - \nabla_{\mathbf{x}'} \log p(\mathbf{x}'|\mathbf{x}, \sigma_I) \right\|_2^2 \rightarrow \min_{\theta}$$

Denoising diffusion probabilistic model (DDPM)

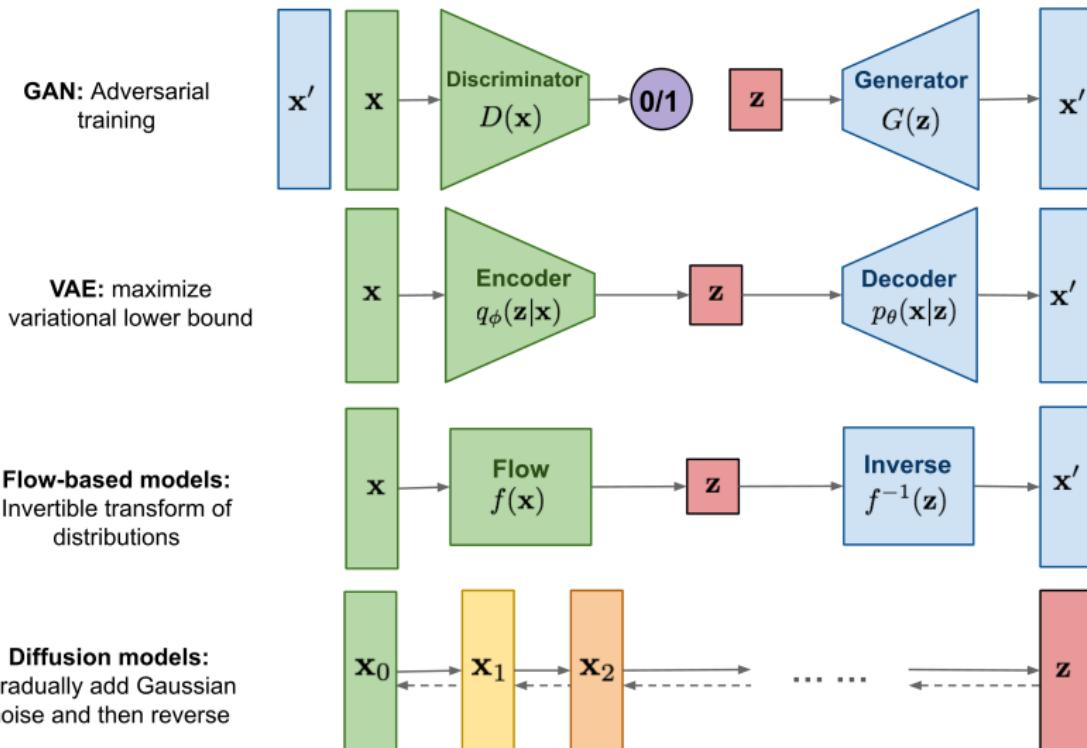
Samples



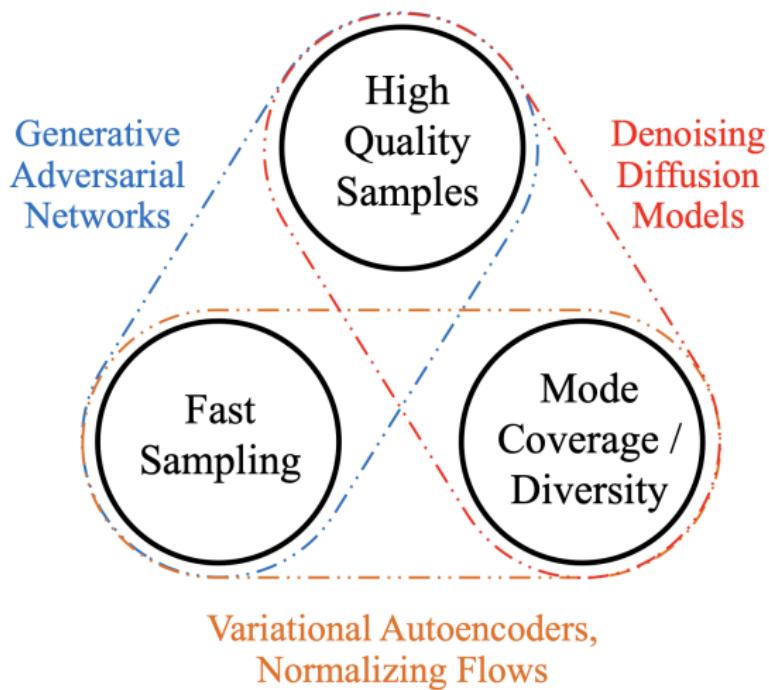
Outline

1. Gaussian diffusion process
2. Denoising diffusion probabilistic model (DDPM)
 - Objective of DDPM
 - Reparametrization of DDPM
3. Overview of DDPM and link to score matching
4. The worst course overview

The worst course overview :)



The worst course overview :)



Summary

- ▶ Gaussian diffusion process is a Markov chain that injects special form of Gaussian noise to the samples.
- ▶ Reverse process allows to sample from the real distribution $\pi(\mathbf{x})$ using samples from noise.
- ▶ Diffusion model is a VAE model which reverts gaussian diffusion process using variational inference.
- ▶ ELBO of DDPM is a sum of KL terms.
- ▶ At each step DDPM predicts the noise used in forward process.
- ▶ Objective of DDPM is closely related to the noise conditioned score network and score matching.