Diffusion Models

Fatim Majumder

Emory University

November 18, 2025

Background: Variational Autoencoders

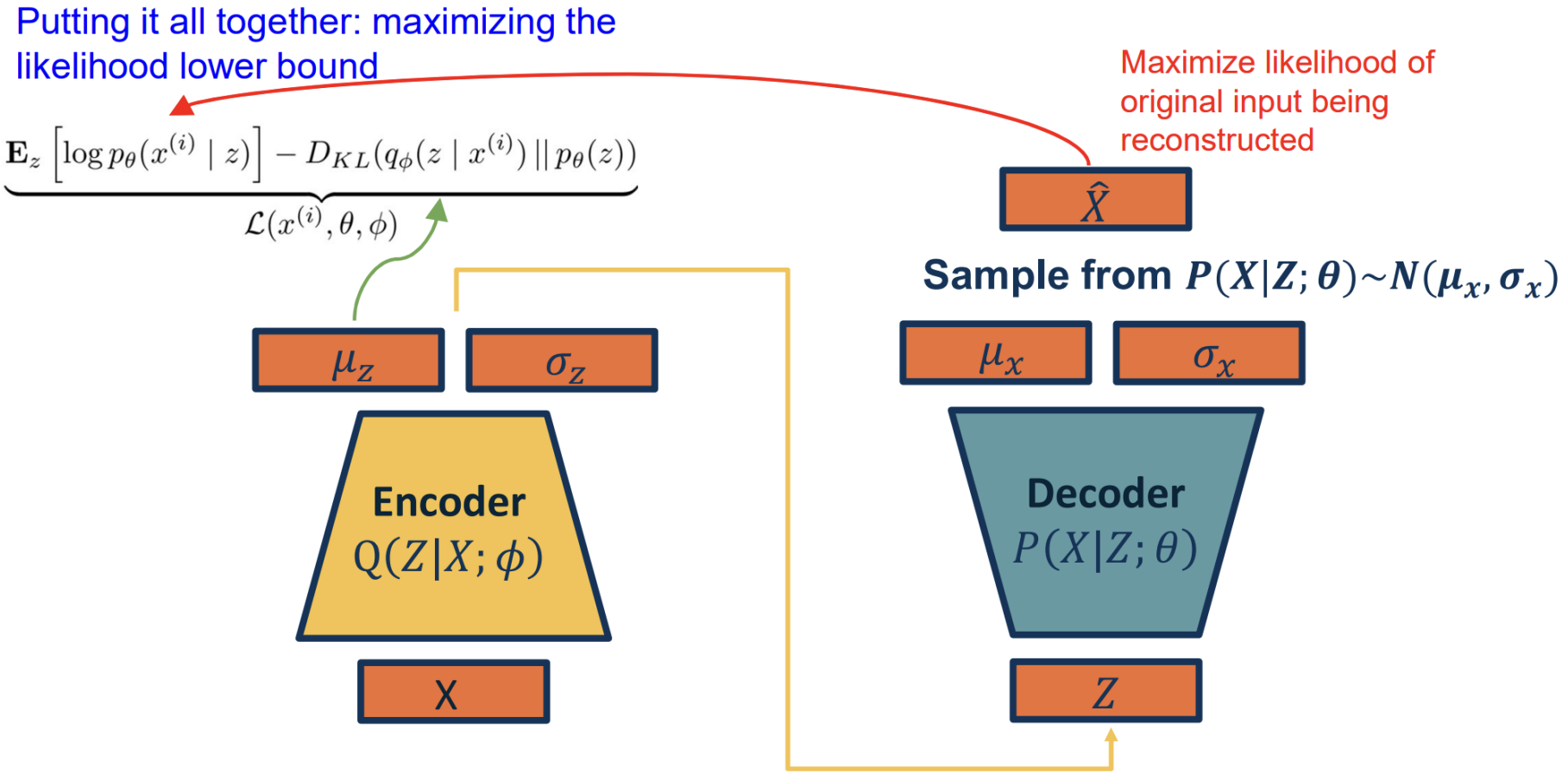
Key Idea: Learn latent representations by maximizing a lower bound on the data likelihood.

Evidence Lower Bound (ELBO)

$$\mathcal{L}(x, \theta, \phi) = \mathbb{E}_{q_\phi(z|x)} \left[\log p_\theta(x|z)\right] - D_{KL}\big(q_\phi(z|x) \| p_\theta(z)\big)$$

- $q_\phi(z|x)$: Encoder approximating the posterior.
- $p_\theta(x|z)$: Decoder reconstructing x from z.
- The KL term regularizes z-space to the prior $p_\theta(z)$.

Variational Autoencoders



Diffusion Process

image from dataset

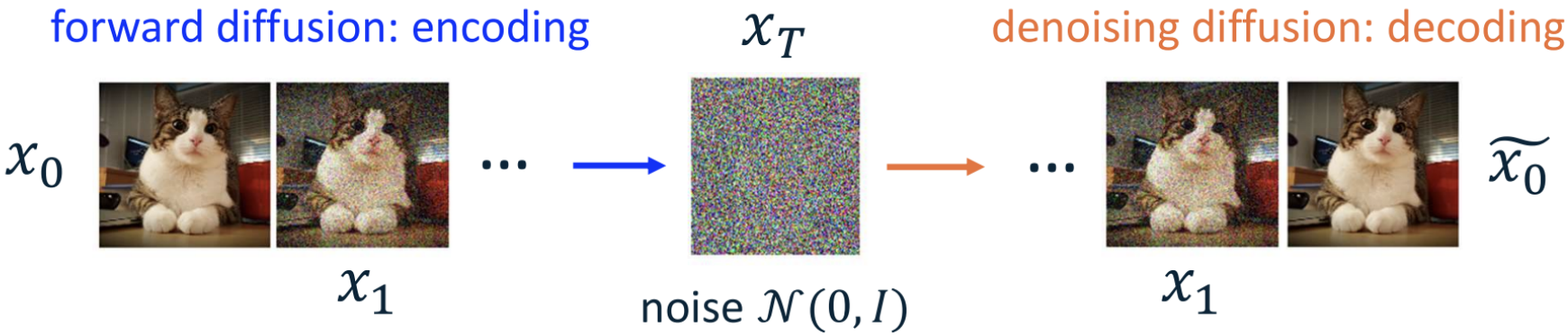
The "forward diffusion" process: add Gaussian noise each step

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

The "denoising diffusion" process: generate an image from noise by *denoising* the gaussian noises

Ties/inspiration form Annealed Imporantce Sampling in physics

Variational Auto Encoding



Known / predefined:
$q(x_{1:T}|x_0)$

Unknown / learned:

$$p_\theta(x_{0:T}) = p(x_T) \prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)$$

Forward Process: Noise Injection

Idea: Starting from a clean sample x_0, we construct a noisy sequence $(x_1, \ldots, x_T)$.

Forward (noising) step:

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

By applying this recursively:

$$q(x_{1:T}|x_0) = \prod_{t=1}^{T} q(x_t|x_{t-1}).$$

Key property: We can directly compute the distribution at an arbitrary step t:

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}\, x_0, (1-\bar{\alpha}_t) I),$$

where $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$.

This closed-form solution helps us understand how noise accumulates at each step.

Reverse Process: Denoising

Goal: Learn a model $p_\theta(x_{t-1}|x_t)$ that reverses the noising process.

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \sigma_t^2 I).$$

By chaining these steps backwards:

$$p_\theta(x_{0:T}) = p(x_T) \prod_{t=1}^{T} p_\theta(x_{t-1}|x_t).$$

We typically fix or simplify σ_t^2 and focus on learning $\mu_\theta(x_t, t)$, guiding the model from noisy x_T back to the clean data x_0. The model effectively learns to peel off the accumulated noise one step at a time.

Variational Inference and the ELBO

Big Picture: Variational inference provides a framework for handling intractable likelihoods by optimizing a lower bound.
From Jensen's inequality:

$$\log p_\theta(x) = \log \int q_\phi(z|x) \frac{p_\theta(x,z)}{q_\phi(z|x)} dz \geq \mathbb{E}_{q_\phi(z|x)} \left[\log \frac{p_\theta(x,z)}{q_\phi(z|x)} \right] = \mathcal{L}_{\theta,\phi}(x).$$

This $\mathcal{L}_{\theta,\phi}(x)$ is the Evidence Lower Bound (ELBO):

$$\mathcal{L}_{\theta,\phi}(x) \leq \log p_\theta(x).$$

By minimizing the KL divergence $D_{KL}(q(z|x) \| p_\theta(z|x))$, we maximize $\mathcal{L}$, making it a tractable surrogate for the full likelihood.

Learning Objective for Diffusion Models

High-Level Goal: Align the learned denoising model $p_\theta(x_{t-1}|x_t)$ with the *true* reverse distribution $q(x_{t-1}|x_t, x_0)$.

$$\arg\min_\theta D_{KL}(q(x_{t-1}|x_t, x_0) \parallel p_\theta(x_{t-1}|x_t)).$$

Under certain assumptions, this KL minimization simplifies to minimizing a mean-squared error:

$$\arg\min_\theta \mathbb{E}_w \|\mu_q(t) - \mu_\theta(x_t, t)\|^2,$$

which is equivalent to:

$$\arg\min_\theta \|\epsilon - \epsilon_\theta(x_t, t)\|^2,$$

where $\epsilon \sim \mathcal{N}(0, I)$.

Interpretation: Predict and remove the noise added at step t.

Training Algorithm

Algorithm 1 Training Algorithm (Noise Prediction)

Require: Training data $\{x_0\}$, parameters θ, schedule $\{\beta_t\}$ (hence $\{\alpha_t, \bar{\alpha}_t\}$)

1: **for all** data samples x_0 **do**

2: Sample a timestep $t \sim \text{Uniform}(\{1, \ldots, T\})$.

3: Sample noise $\epsilon \sim \mathcal{N}(0, I)$.

4: Form the noisy sample:

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

5: Use the model to predict $\epsilon_\theta(x_t, t)$.

6: Update θ by minimizing the mean-squared error:

$$\|\epsilon - \epsilon_\theta(x_t, t)\|^2.$$

7: **end for**

Sampling Algorithm

Algorithm 2 Sampling Algorithm (Denoising Process)

Require: Learned parameters θ, schedule $\{\beta_t\}$ (and thus $\{\alpha_t, \bar{\alpha}_t\}$).

1: Sample $x_T \sim \mathcal{N}(0, I)$ (pure noise).
2: **for** $t = T, \ldots, 1$ **do**
3: Predict $\epsilon_\theta(x_t, t)$ using the model.
4: Compute $\mu_\theta(x_t, t)$ from the predicted noise:

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

5: Sample $x_{t-1} \sim \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \sigma_t^2 I)$.
6: **end for**
7: Return x_0 as the generated sample.

References

Jonathan Ho, Ajay Jain, and Pieter Abbeel.

Denoising Diffusion Probabilistic Models.

In *NeurIPS*, 2020.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli.

Deep Unsupervised Learning using Nonequilibrium Thermodynamics.

In *ICML*, 2015.

Diederik P. Kingma and Max Welling.

Auto-Encoding Variational Bayes.

In *ICLR*, 2014.