Diffusion Models

📅Date: 12-30-2022
🕐Time: 20:39

Bipin Koirala

Table of Contents

Preliminary 🔍

Discriminant Models Vs Generative Models

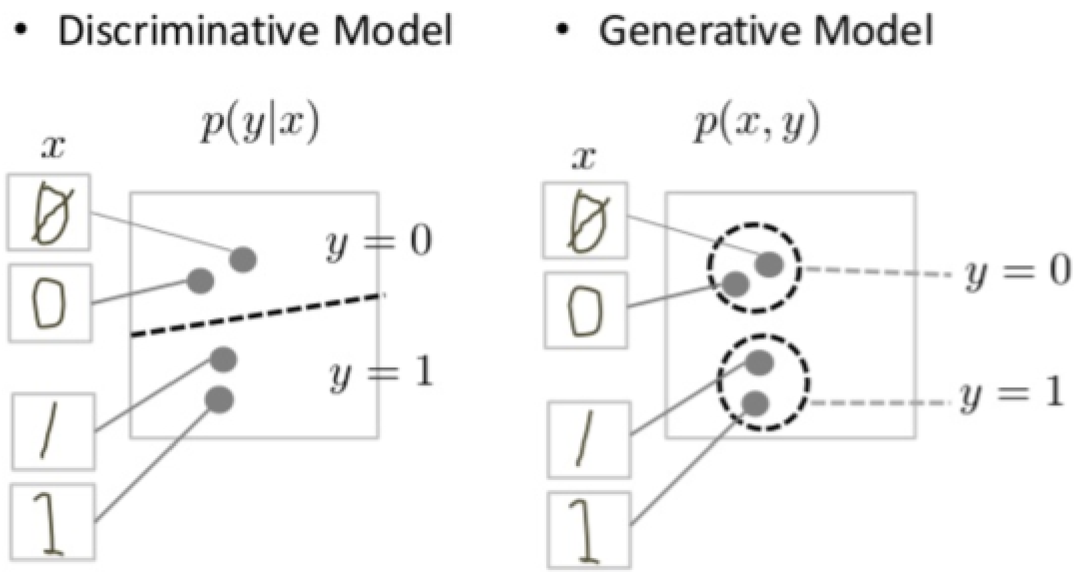
Most of the time we are concerned with predicting a label given an instance of a dataset. Statistical models that operate on this notion are #discriminant models. Unlike these models there exists a different paradigm where we want to learn the joint probability distribution between the data and its label (holds true even if the data has no label). These models are called #generative models and can generate new data instances.

For a set of data instances X and set of labels Y, we have the following:

$$\text{Discriminant Model} : \mathbb{P}(label|data) = \mathbb{P}(Y|X)$$

$$\text{Generative Model} : \mathbb{P}(data, label) = \mathbb{P}(X, Y) \;\text{ or }\; \mathbb{P}(X) \;\text{ if no labels}$$

For example, a discriminant model could tell a picture of a bird from a horse but generative model could generate a new pictures of animals that look like real animals. For data $x \in \mathcal{D}$, discriminant model aims to draw a boundary in $\mathcal{D}$ whereas generative model aims to model how a data is placed throughout the space $\mathcal{D}$.



Some types of generative models are:

- Gaussian Mixture Model (GMM)
- Bayesian Network (Naive Bayes, Auto-regressive models)
- Boltzmann Machine
- Generative Adversarial Network (GAN)
- Variational Auto-encoder (VAE)
- Diffusion Models
- Energy Based Models (EBM) etc.

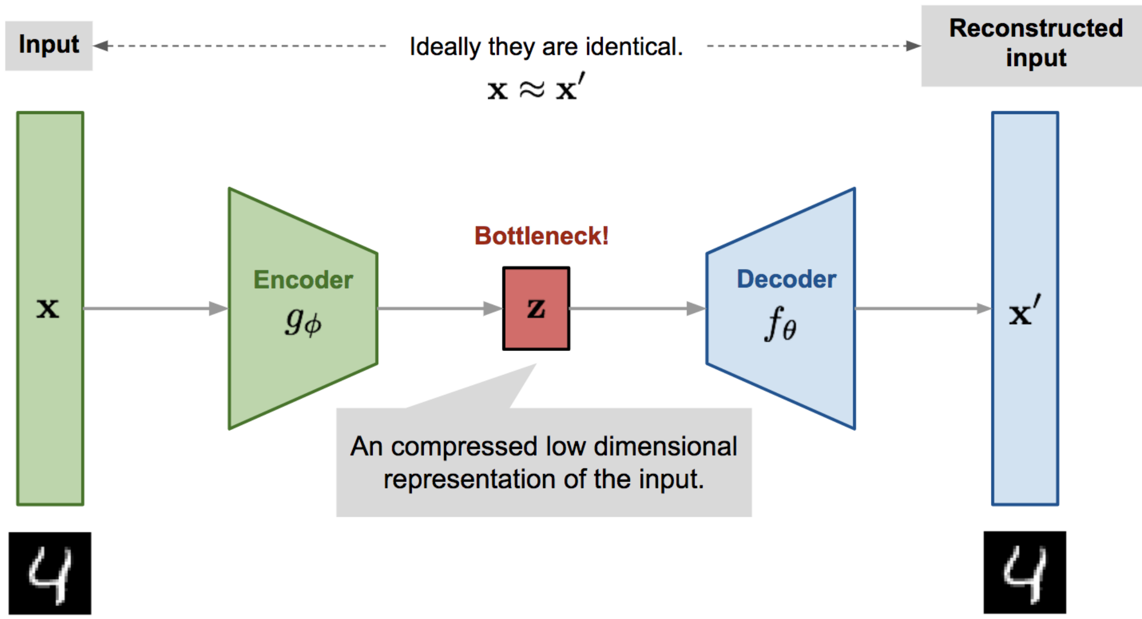
Below are summary of few generative models.

GANs

These are primarily used to replicate real-world contents such as images, languages, and musics. Two agents, #generator and #discriminator play a min-max game to attain equilibrium. It is difficult to train GAN because of training instability and failure to converge. #Wasserstein GAN (WGAN) provides improved results over traditional GAN.

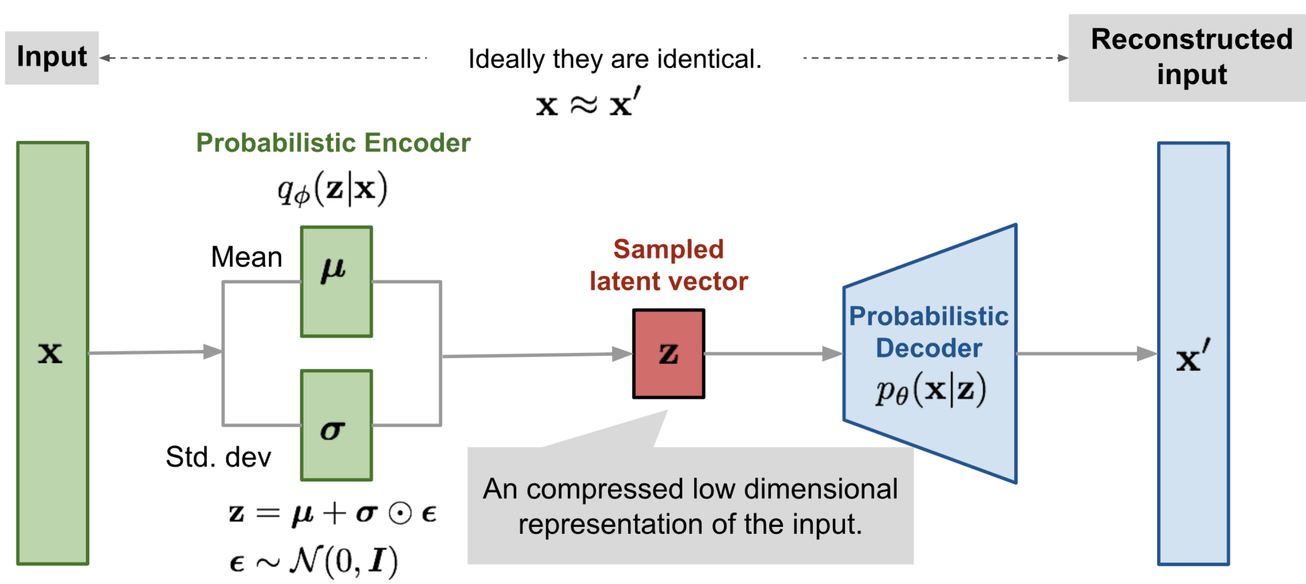
VAE

#Auto-encoder is a neural network which attempts to reconstruct a given data via compressing the input in the process so as to discover a latent/sparse representation. This latent representation of data can later be used in various downstream tasks.

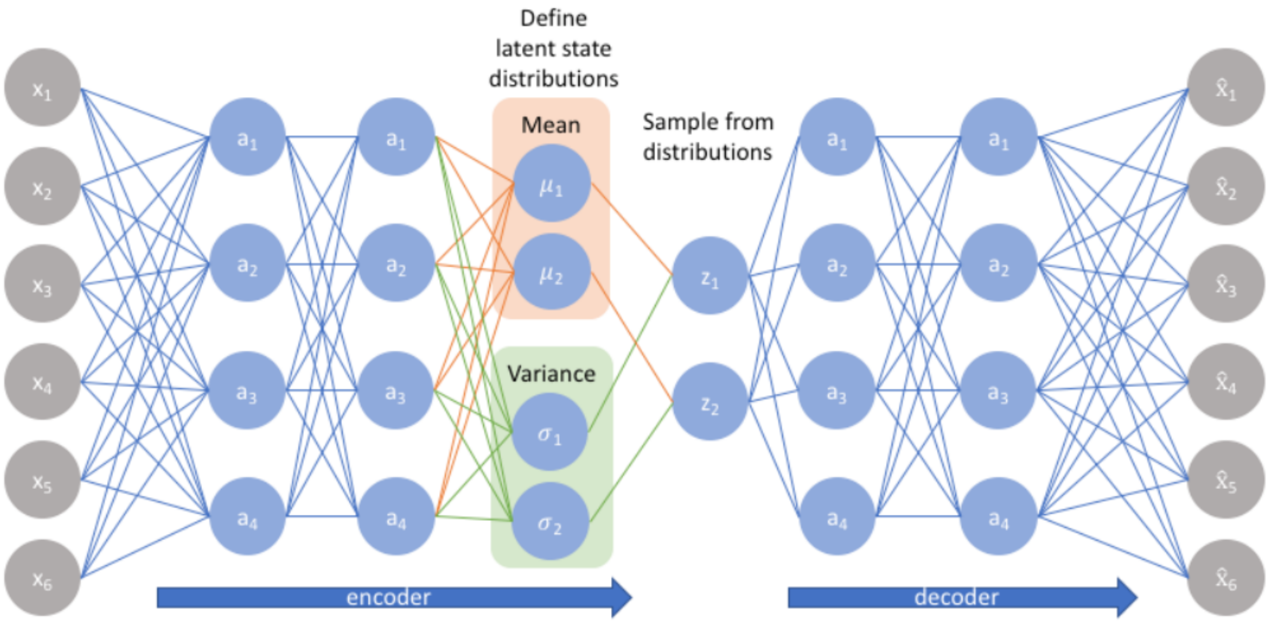


The latent space in auto-encoders are primarily discrete and does not allow for an easy interpolation. The generative part or the decoder of the auto-encoder works by randomly sampling points from the latent space and it can be challenging if the latent space is itself discontinuous or has gaps.

#Variational-Auto-Encoder (VAE) solves this issue because its latent space is continuous in nature which makes VAE powerful in generating new data instances. Instead of generating a latent vector $z \in \mathbb{R}^N$, VAE generates two vectors i.e. mean (μ) vector and standard deviation (σ) vector followed by decoder sampling from this distribution.



Neural Network architecture of VAE showing how decoder samples from latent vectors.



Flow-Models

At its core, #Flow-Models make use of **Normalizing Flow (NF)**, a technique used to build a complex probability distributions by transforming simple distributions.
Let, $z \sim \mathbb{P}_\theta(z)$ and $z \in Z$ be a probability distribution, generally taken something simple like $\mathcal{N}(z; \mu, \sigma)$. The key idea here is to transform this simple distribution to a complex distribution $x = f(z)$, where f is a bijective map. We formulate f as a composition of sequence of invertible transformations.

$$x = f_K \circ f_{K-1} \circ \ \ldots \ f_2 \circ f_1(z)$$

Now,

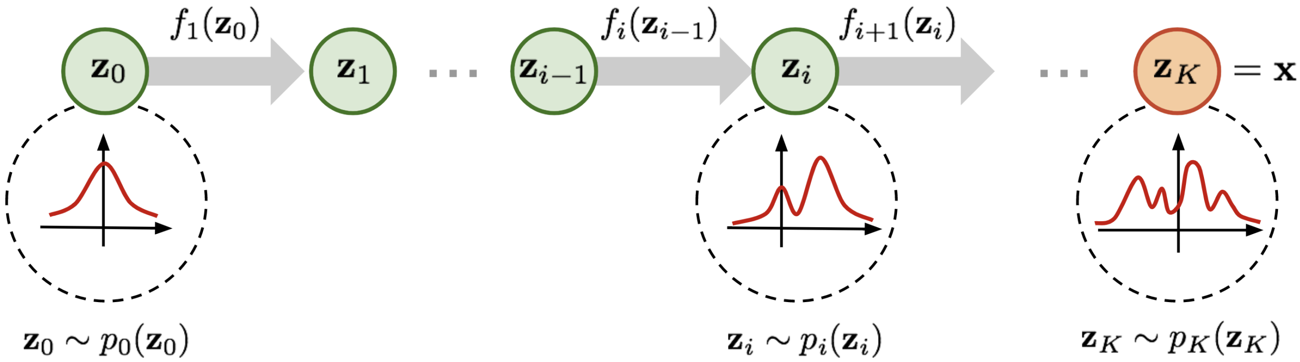
$$\int p_\theta(x)dx = \int p_\theta(z)dz = 1$$

$$p_\theta(x)dx = p_\theta(f^{-1}(x))dz$$

$$p_\theta(x) = p_\theta(f^{-1}(x))\left|\frac{dz}{dx}\right| = p_\theta(f^{-1}(x))\left|\frac{df^{-1}}{dx}\right|$$

Multivariable formulation of the above expression gives us;

$$p_\theta(\mathbf{x}) = p_\theta(f^{-1}(\mathbf{x}))\left|det\left(\frac{df^{-1}}{d\mathbf{x}}\right)\right|$$



Diffusion Models

Notes 📝

Sources 📚

[1]. [Generative Models]()