L13 Variational Autoencoders

Autoencoders - limitations

- Autoencoders may suffer from overfitting and may not be able to generate new data points.
- There is also a fundamental problem with autoencoders, for generation
- The latent space may not be continuous, or allow easy interpolation.

Variational Autoencoders

- VAE is a generative system and serves purpose similar to that of a generative adversarial network (GAN).
- Variational Autoencoders (VAEs) are generative models explicitly designed to capture the underlying probability distribution of a given dataset and generate novel samples.
- Similar to a standard autoencoder, a variational autoencoder is essentially an architecture that consists of an encoder as well as a decoder.
- It is trained to minimize the reconstruction error between the encoded-decoded data and the initial data.
- But, for the purpose of introducing a regularization of the latent space, there is an adjustment made to the encoding-decoding process.
- Rather than encoding an input as a single point, it is encoded as a distribution over the latent space
- This probabilistic approach to encoding the input allows VAEs to learn a structured and continuous latent space representation, which is useful for generative modeling and data synthesis.
- AE is a deterministic model, while VAE is probabilistic model

Variational Autoencoders

- In AE, encoder network is outputting a single value for each encoding dimension.
- Rather than building an encoder which outputs a single value to describe each latent state attribute, VAE encoder describes a probability distribution for each latent attribute
- Ex: suppose we have trained an autoencoder model on a large dataset of faces with a encoding dimension of 4.

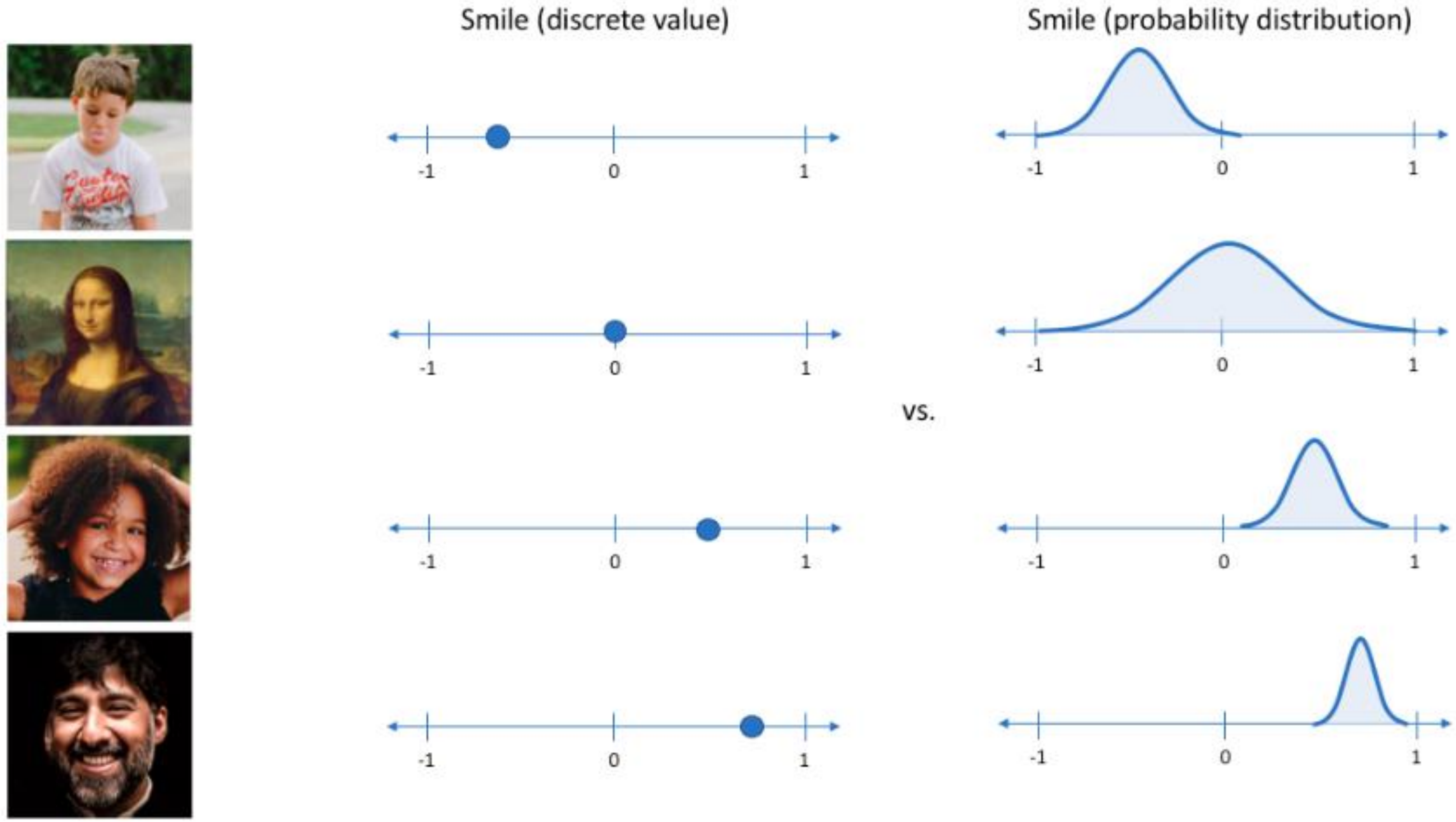
Autoencoders

Latent attributes

Variational Autoencoders

- An autoencoder will learn descriptive attributes of faces such as skin color, whether or not the person is wearing glasses, etc. in an attempt to describe an observation in some compressed representation.
- A variational autoencoder represents each latent attribute for a given input as a probability distribution.
- When decoding from the latent state, it randomly samples from each latent state distribution to generate a vector as input for decoder model.

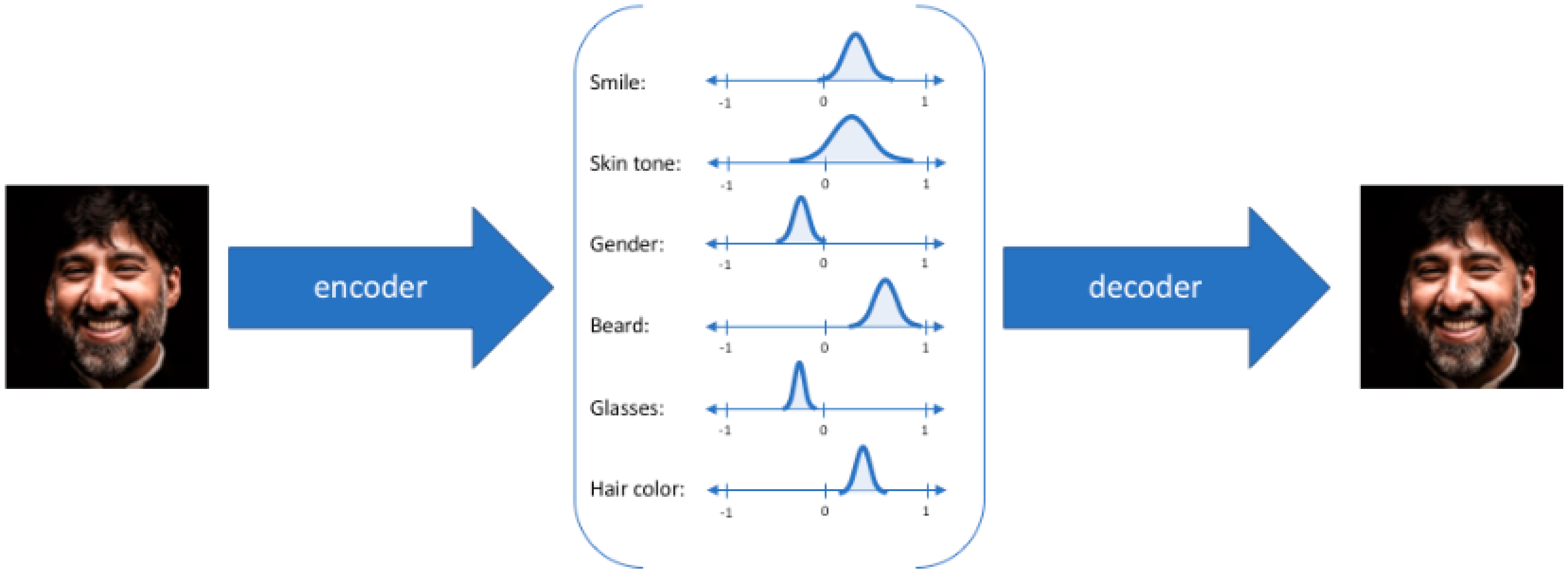
AE vs VAE



Representing the input image using a single value to describe each attribute in terms of its latent attributes

Represent each latent attribute as a range of possible values.

Variational Autoencoders

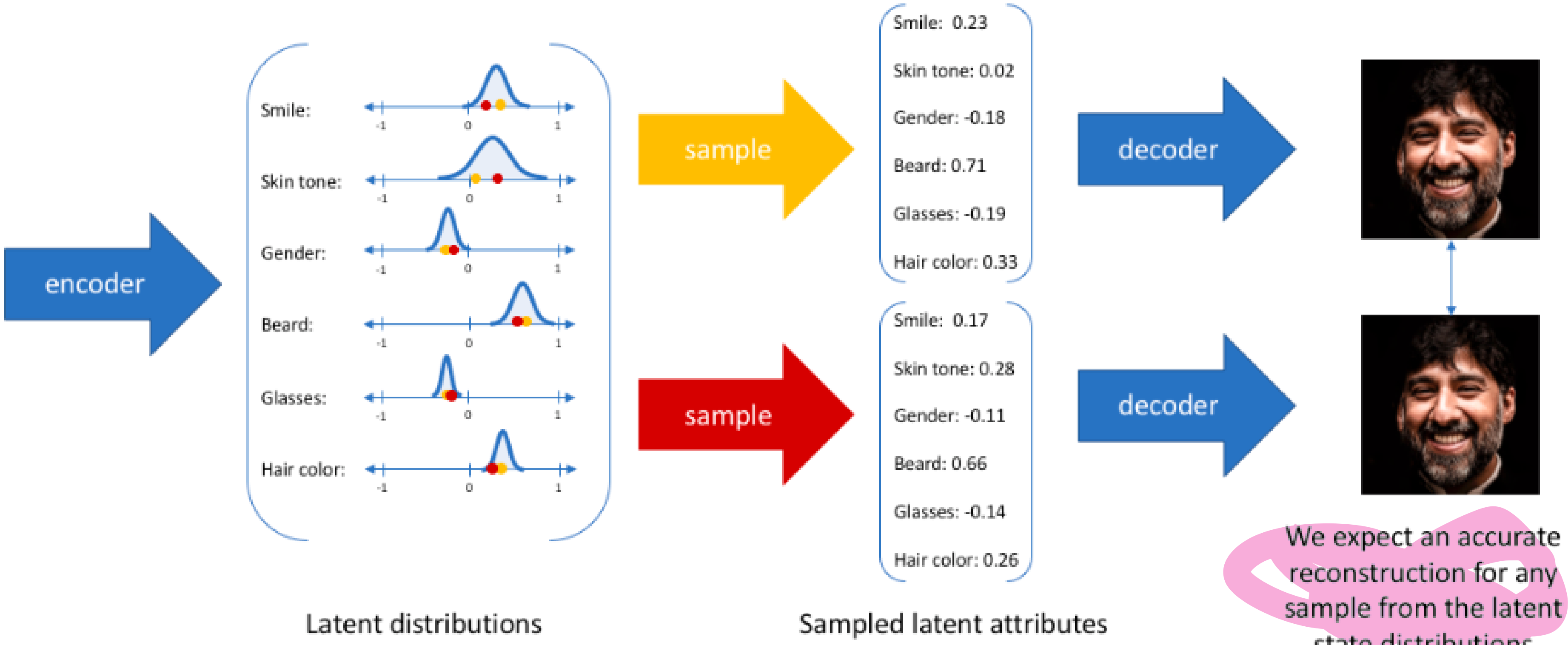


Latent attributes

Variational Autoencoders

- By constructing encoder model to output a range of possible values (a statistical distribution) which will be randomly sampled to feed into decoder model, we are essentially enforcing a continuous, smooth latent space representation.
- For any sampling of the latent distributions, we are expecting our decoder model to be able to accurately reconstruct the input.
- Thus, values which are nearby to one another in latent space should correspond with very similar reconstructions.

Variational Autoencoders

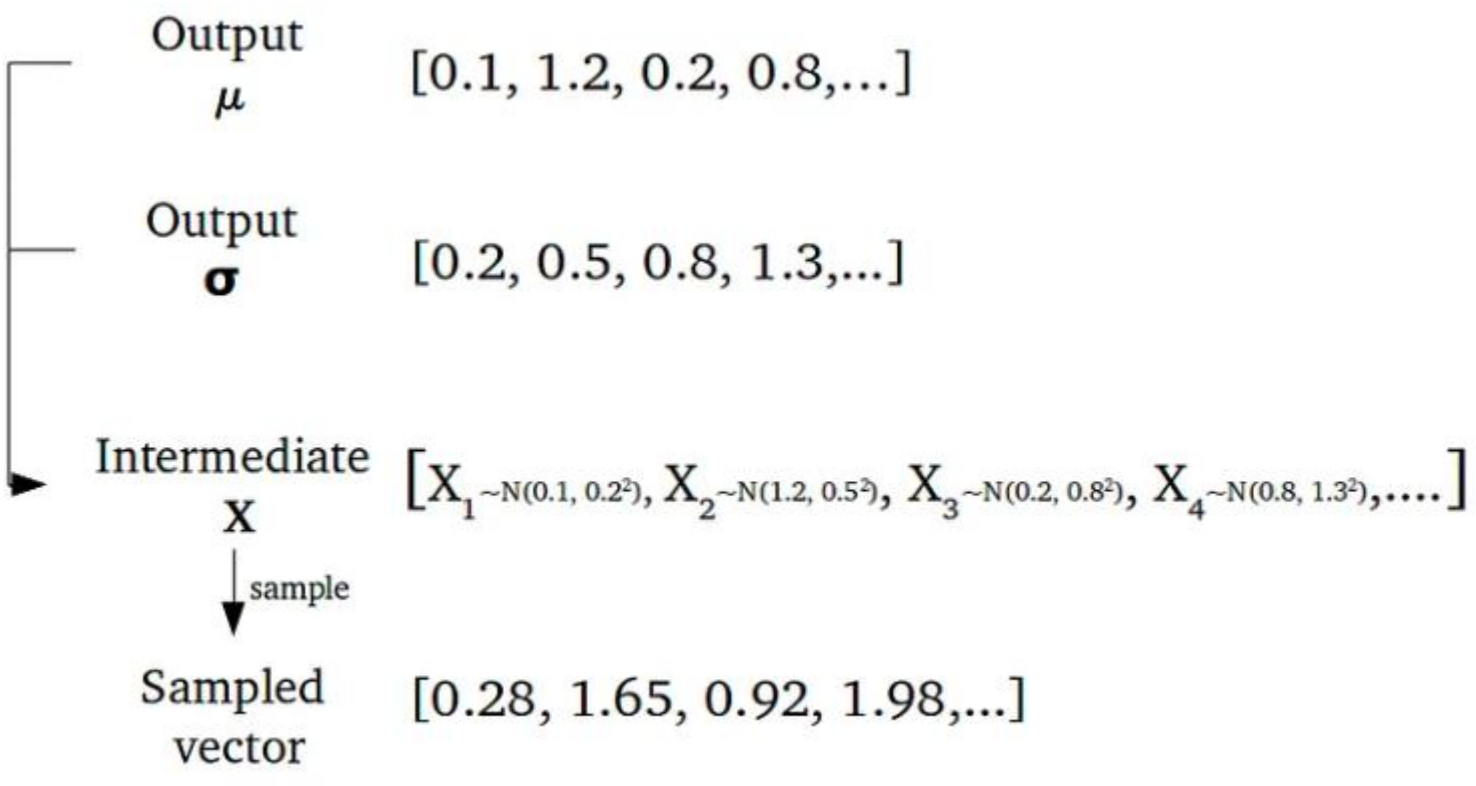


We expect an accurate reconstruction for any sample from the latent state distributions

Variational Autoencoders

- VAE encoder does not output an encoding vector of size n, but
- outputs two vectors of size n:
  - a vector of means, μ, and
  - another vector of standard deviations, σ

Variational Autoencoders



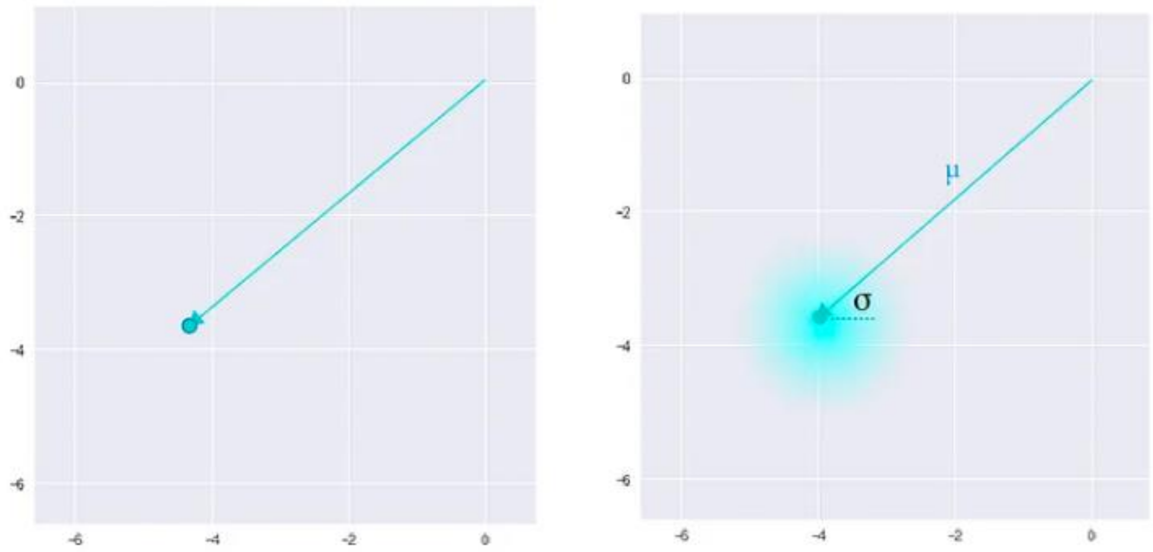
Stochastically generating encoding vectors

Form the parameters of a vector of random variables of length n,

With the i th element of μ and σ being the mean and standard deviation of the i th random variable, X i, from which we sample, to obtain the sampled encoding which we pass onward to the decoder

Variational Autoencoders

- In stochastic generation, even for the same input while the mean and standard deviations remain the same, the actual encoding will vary on every single pass due to sampling.



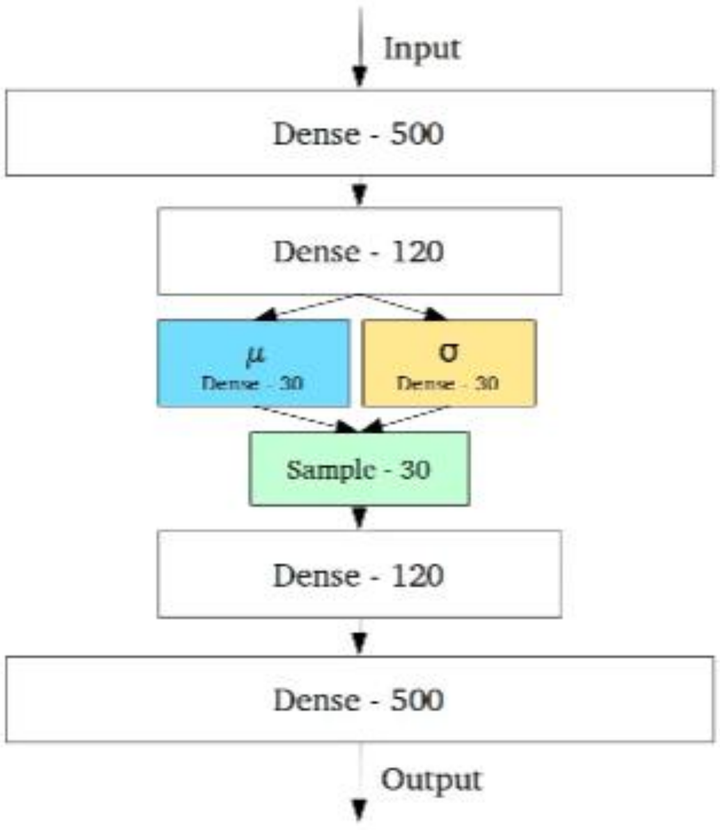
Standard Autoencoder
(direct encoding coordinates)

Variational Autoencoder
(μ and σ initialize a probability distribution)

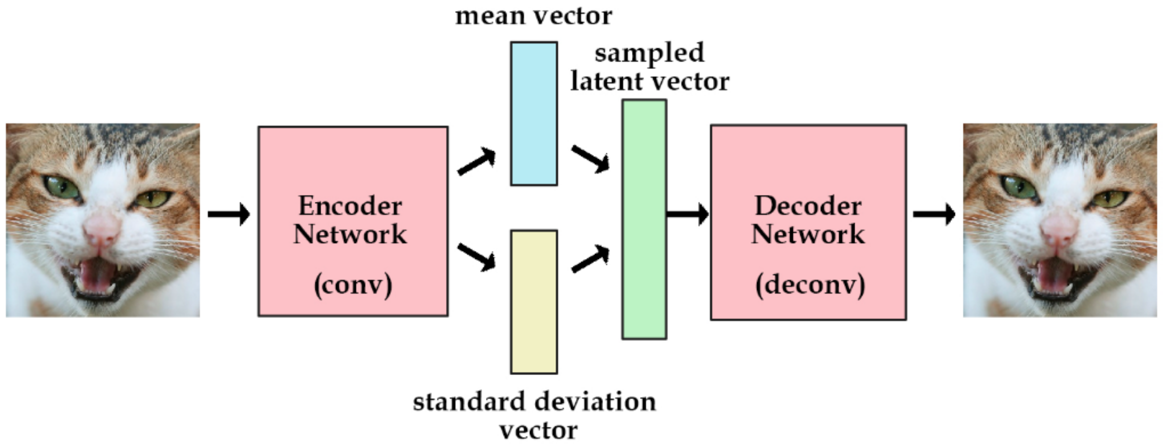
Variational Autoencoders

- The mean vector controls where the encoding of an input should be centered around
- while the standard deviation controls the “area”, how much from the mean the encoding can vary.
- As encodings are generated at random from anywhere inside the “circle” (the distribution), the decoder learns that not only is a single point in latent space referring to a sample of that class, but all nearby points refer to the same as well.
- This allows the decoder to not just decode single, specific encodings in the latent space but ones that slightly vary too, as the decoder is exposed to a range of variations of the encoding of the same input during training.

Variational Autoencoders



Variational Autoencoder



Variational Autoencoders

- How the model is trained?

1. It starts by encoding the input as distribution over the latent space
2. After that, a point from the latent space is sampled from that distribution
3. Next, the point that was sampled gets decoded and the reconstruction error is computed.
4. Then, the reconstruction error is backpropagated through the network.

- The input is encoded as a distribution with a bit of variance rather than a single point so that it is possible to naturally express the latent space regularization.
- The encoder returns the distributions which are enforced to be close to a standard normal distribution.

Variational Autoencoders - Disadvantages

- VAEs can be more difficult to train and require more computational resources.
- Additionally, the learned latent space representation can be difficult to interpret, and the quality of generated data can be limited by the model architecture and training data.
- VAEs may suffer from mode collapse, where the model generates similar outputs for different inputs.

Variational Autoencoders – Advantages over AEs

VAEs have several advantages over traditional autoencoders.

1. They allow for generative modeling, meaning they can generate new data points from the learned latent space distribution.
2. They also allow for continuous latent space representations, which means that we can interpolate between different points in the latent space to generate novel data points.
3. VAEs are less susceptible to overfitting than traditional autoencoders since the probabilistic nature of the encoding forces the model to learn a more robust representation of the data.

Applications AE vs VAE

- AEs are better suited for tasks like dimensionality reduction and feature extraction.
- VAEs are better suited for generative tasks like image and text generation, where we want to generate new data points
- VAEs find applications in various domains like density estimation and text generation