RV College of Engineering®

Variational AutoEncoders

Variational AutoEncoders

Variational Autoencoders (VAEs) are generative models in Machine Learning (ML) that create new data similar to the input they are trained on. Along with data generation they also perform common autoencoder tasks like **denoising.** Like all autoencoders

VAEs consist of:

•**Encoder:** Learns important patterns (latent variables) from input data.

•**Decoder:** It uses those latent variables to reconstruct the input.

Variational AutoEncoders

- Traditional Autoencoders that encode a fixed representation .
- VAEs learn a **continuous probabilistic** representation of latent space.
- VAE is a special kind of autoencoder that can generate new data instead of just compressing and reconstructing it.

Variational AutoEncoders

VAE Main Parts

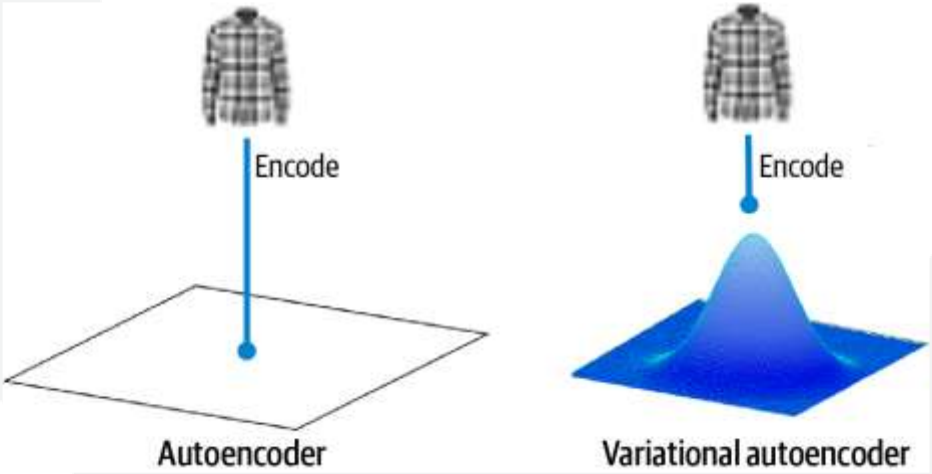
1. Encoder (Understanding the Input)

•The encoder takes the input data like an image or text and tries to understand its most important features.

•Instead of creating a fixed compressed version like a normal autoencoder it creates two things:

- **Mean (μ):** A central value representing the data.
- **Standard Deviation (σ):** It is a measure of how much the values can vary.

•These two values define a range of possibilities instead of a single number.



Variational AutoEncoders

VAE Main Parts

1. Encoder (Understanding the Input)

The Multivariate Normal Distribution

A *normal distribution* (or *Gaussian distribution*) μ, σ is a probability distribution characterized by a distinctive *bell curve* shape, defined by two variables: the *mean* (μ) and the *variance* ($\sigma 2$). The *standard deviation* (sigma) is the square root of the variance.

The probability density function of the normal distribution in one dimension is:

$$f\left(x \mid \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

We can sample a point z from a normal distribution with mean μ and standard deviation σ using the following equation: $z = \mu + \sigma\epsilon$

VAE Main Parts

1. Encoder (Understanding the Input)

To summarize, the encoder will take each input image and encode it to two vectors that together define a multivariate normal distribution in the latent space:

`z_mean`

The mean point of the distribution

`z_log_var`

The logarithm of the variance of each dimension

We can sample a point z from the distribution defined by these values using the following equation:

```
z = z_mean + z_sigma * epsilon
```

where:

```
z_sigma = exp(z_log_var * 0.5)
epsilon ~ N(0,I)
```

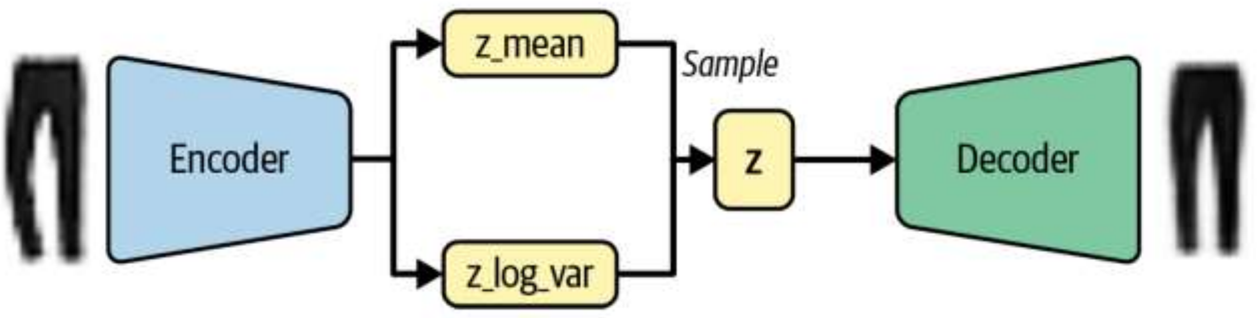
The derivation of the relationship between `z_sigma` (σ) and `z_log_var` ($\log\left(\sigma^2\right)$) is as follows:

$$\sigma = \exp\left(\log(\sigma)\right) = \exp\left(2\log(\sigma)/2\right) = \exp\left(\log\left(\sigma^2\right)/2\right)$$

Variational AutoEncoders

VAE Main Parts

1. **Encoder (Understanding the Input)**



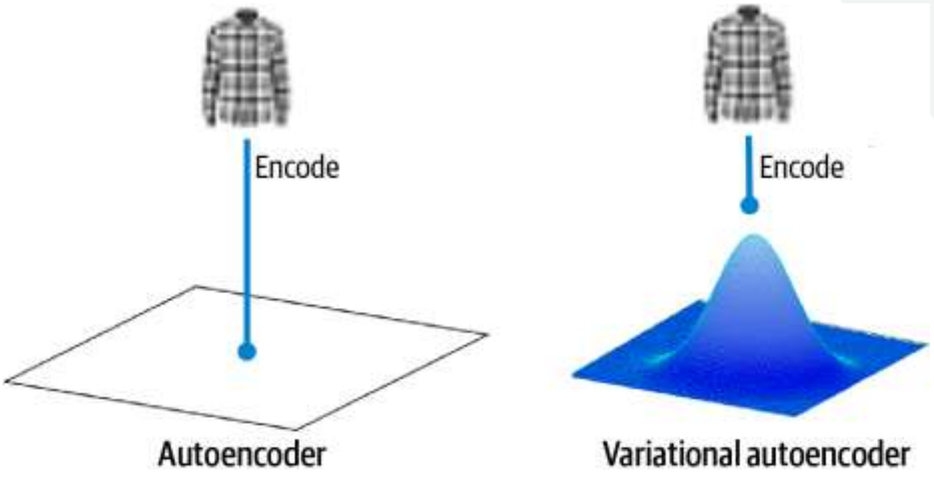
VAE architecture diagram

Variational AutoEncoders

VAE Main Parts

2. Latent Space (Adding Some Randomness)

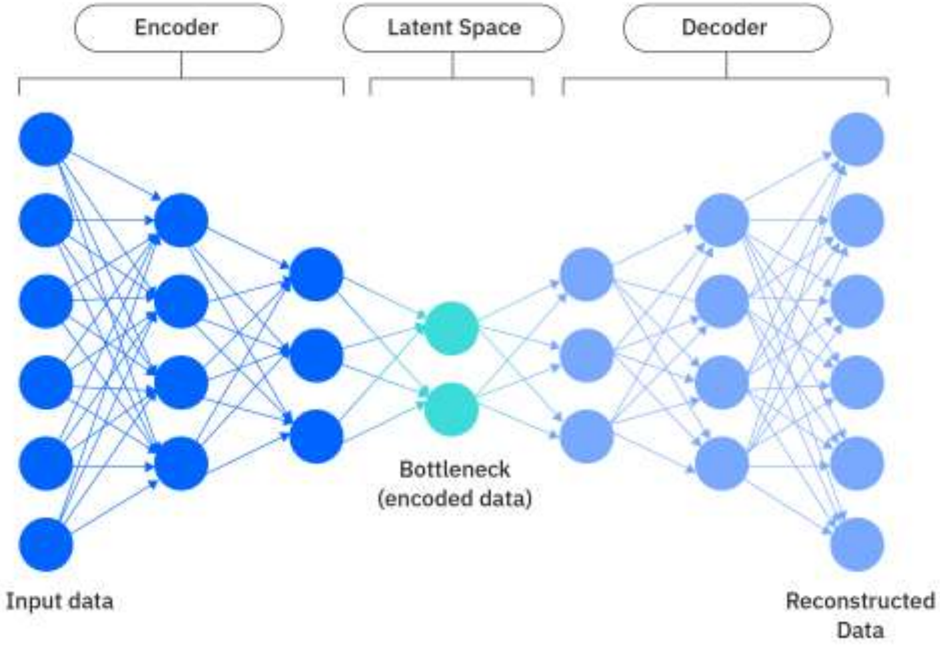
- Instead of encoding input into a fixed number VAEs introduce randomness to create variations.
- The model picks a point from the range to create different variations of the data.
- This is what makes VAEs great for generating new slightly different but realistic data.



Variational AutoEncoders

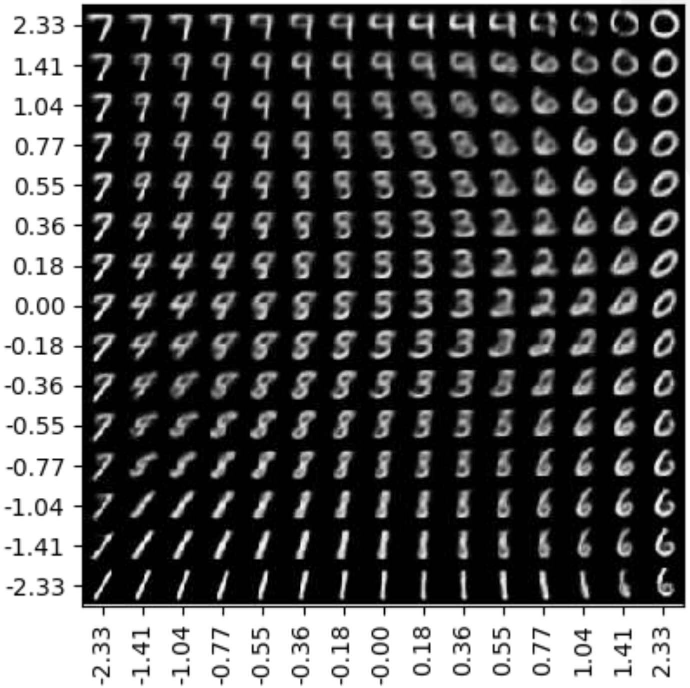
VAE Main Parts

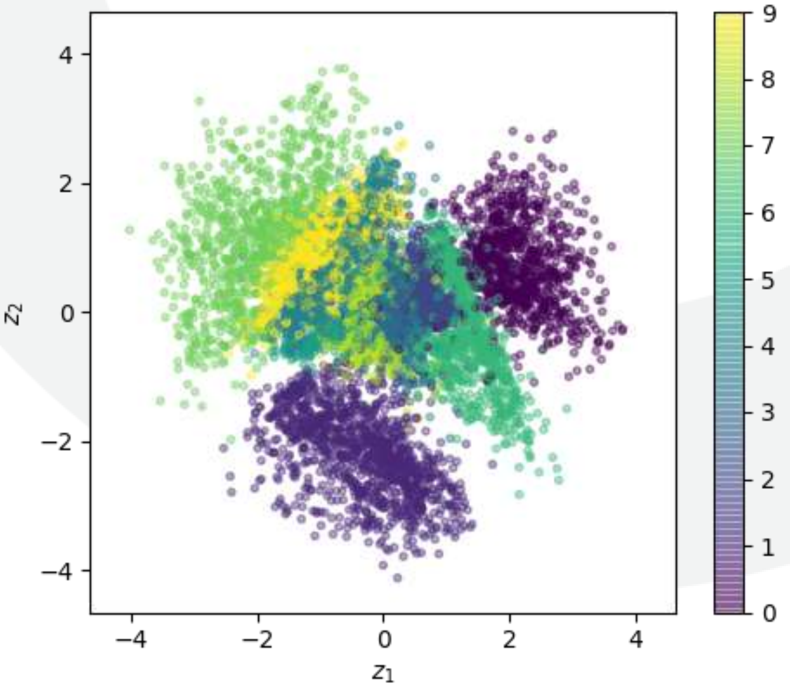
2. Latent Space (Adding Some Randomness)



Variational AutoEncoders

2. Latent Space (Adding Some Randomness)





Variational AutoEncoders

3. Decoder (Reconstructing or Creating New Data)

The **decoder** takes this sampled value and tries to reconstruct the original input.

Since the encoder creates a range of possibilities instead of a fixed number the decoder can generate new similar data instead of just memorizing the input.