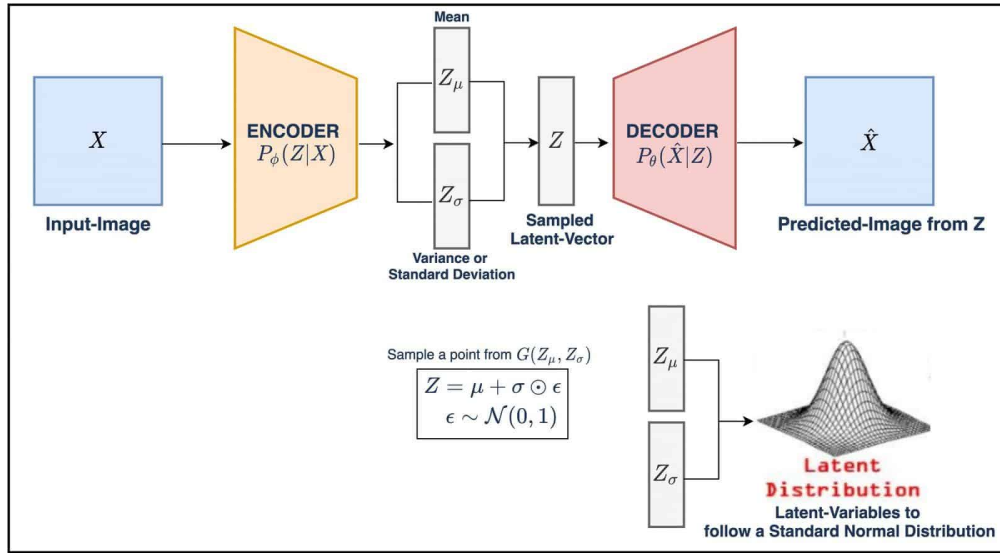# Mathematical Formulation and code explanation of Variational Autoencoders

May 20, 2025

1 Architecture

A Variational Autoencoder consists of:

- An encoder network $q_\phi(z|x)$ that maps input x to a latent distribution.

- A decoder network $p_\theta(x|z)$ that reconstructs data from latent variable z.

Latent variables z are sampled from a prior $p(z)$, typically a standard normal distribution:

$$p(z) = \mathcal{N}(0, I)$$

The encoder produces parameters $\mu(x)$ and $\sigma^2(x)$ to approximate a Gaussian posterior distribution over latent variable z given input x:

$$q_\phi(z|x) = \mathcal{N}(z; \mu_\phi(x), \text{diag}(\sigma_\phi^2(x)))$$

where $\mu(x)$ is the mean and $\sigma^2(x)$ is the variance of the latent variable, ϕ are the parameters of the encoder network, and the covariance matrix is diagonal.