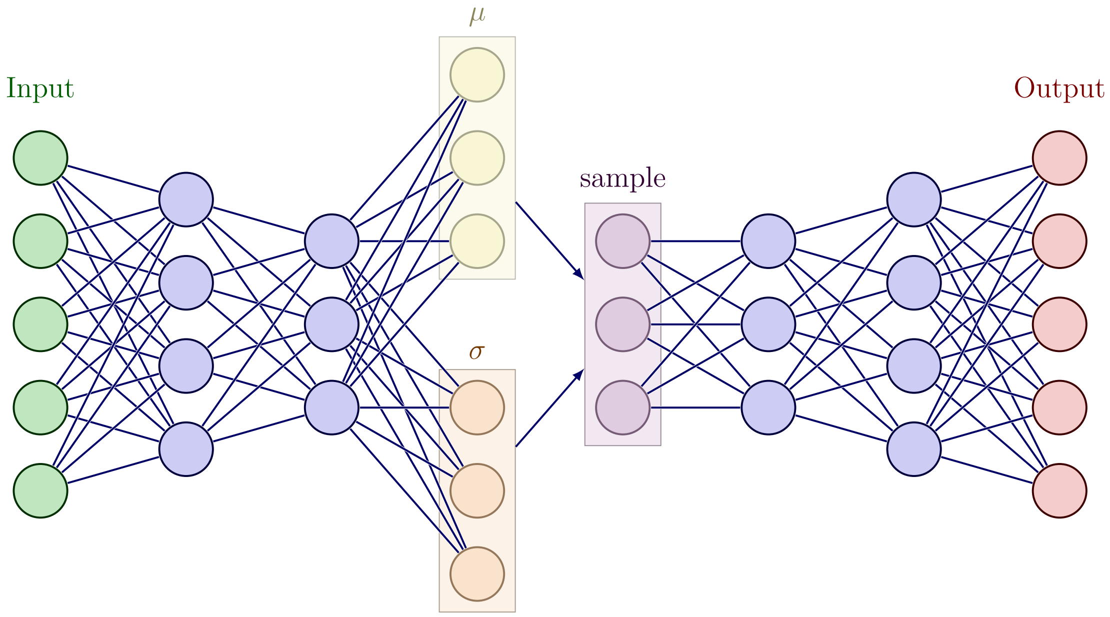

Variational Autoencoders Architecture

latent representation learned from the training samples and not from the data itself. Consequently, we do not have direct access to the data on which we are integrating.

To explain this problem, let us consider the formula in detail. It involves two probability distributions: $p_\theta(z)$ (**prior distribution**) represents the latent space, while $p_\theta(x|z)$ represents the Decoder, which samples z from the prior distribution $p_\theta(z)$ and transforms it into a sample x that must be similar to the training data.

Since the Decoder samples from the prior distribution, the latent space must be a simple probability distribution to facilitate this process. Therefore, the latent space is generally a **Gaussian Multivariate**. However, we cannot directly produce $p_\theta(z)$ since it is derived from the training data. **This is the integration problem**.

Here is a possible solution! We could obtain $p_\theta(z)$ using Bayes' theorem. However, another problem would arise when we try to calculate the a posteriori probability distribution $p_\theta(z|x)$, since the denominator is intractable (🥲):

$$p_\theta(z|x) = \frac{p_\theta(x|z)p_\theta(z)}{p_\theta(x)}$$

In this context, $p_\theta(z|x)$ represents the **latent space distribution extracted from the given input**.

As usual in machine learning, the solution is VERY simple: we throw in a neural network. In this case, the network is the **Encoder**, which is trained to map the input image to the latent space, $q_\phi(z|x)$. In this way, the encoder approximates the a posteriori distribution $p_\theta(z|x)$. **Beware that the encoder does not provide a global solution to the optimization problem, only an approximation: it provides a lower bound on the likelihood of the data!**

Since we are modeling probabilistic data generation, both the encoder and the decoder are probabilistic models and are described by their own parameters, ϕ and θ, respectively. The Encoder learns the mean and variance associated with the Gaussian distribution of the