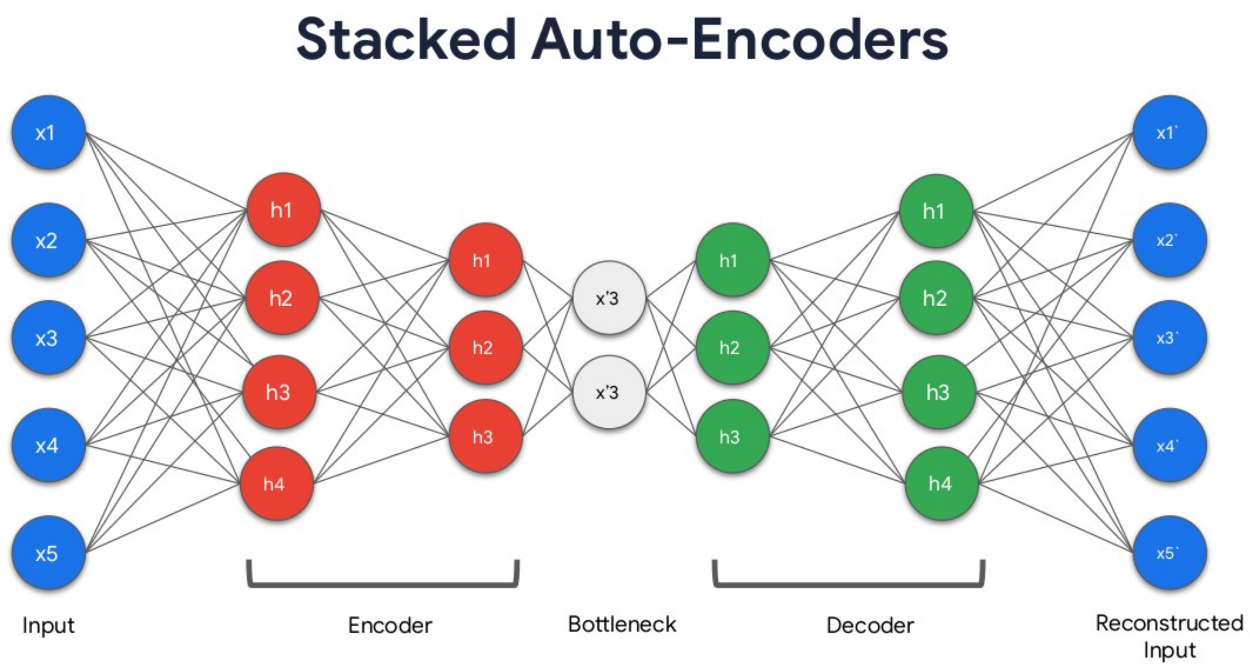
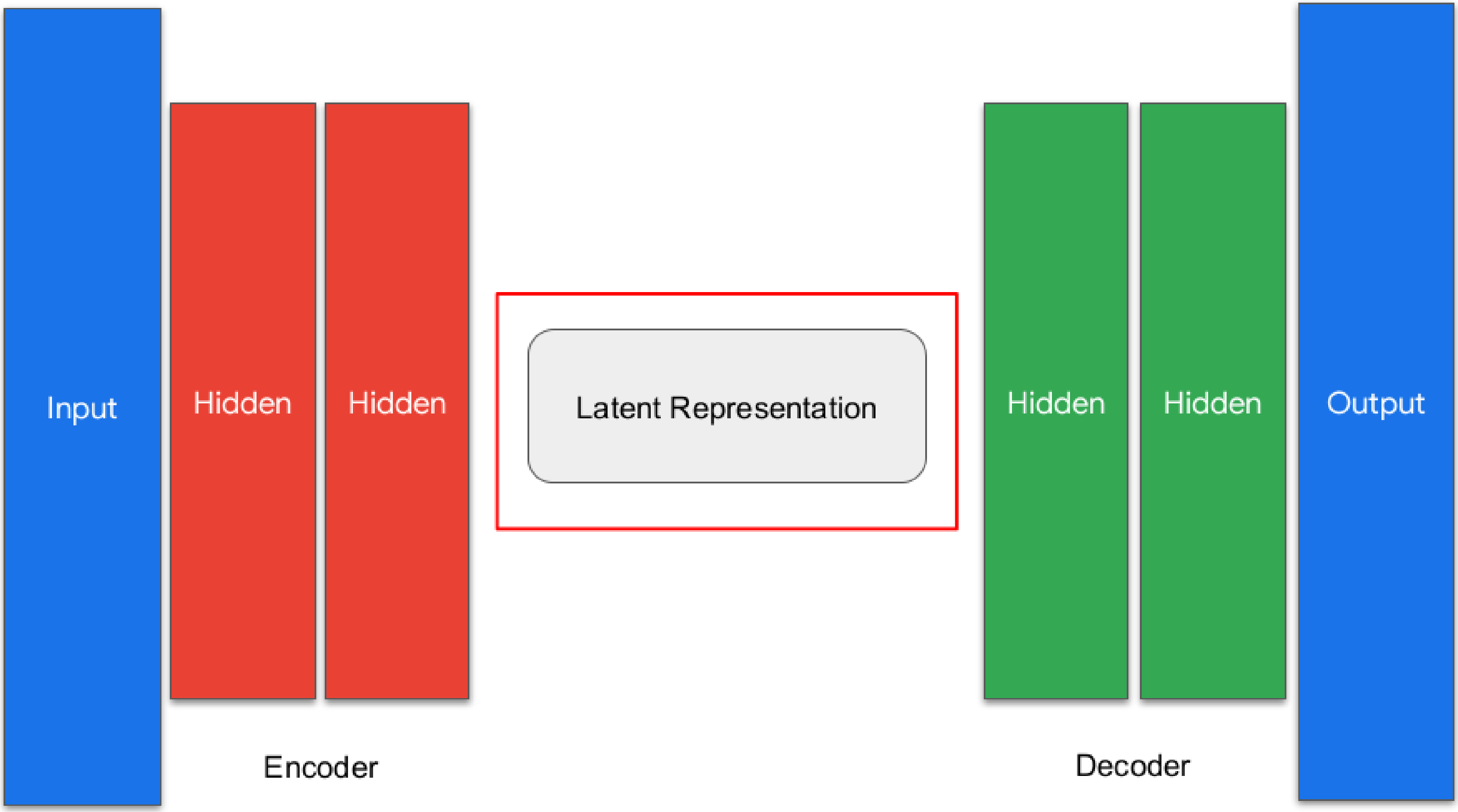
Variational autoencoder

By eng.Ahmed Hisham

Recap autoencoders



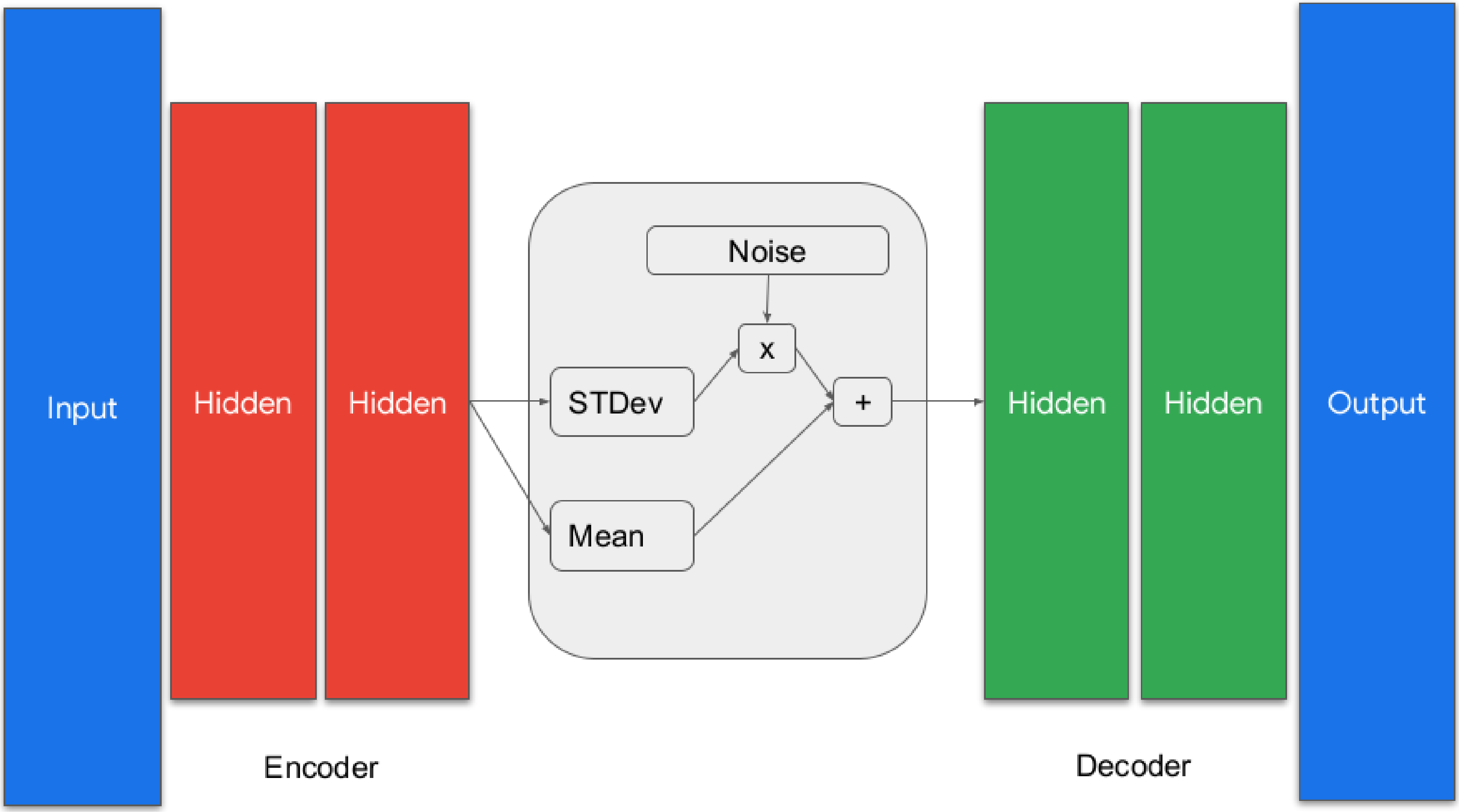
Variational autoencoder



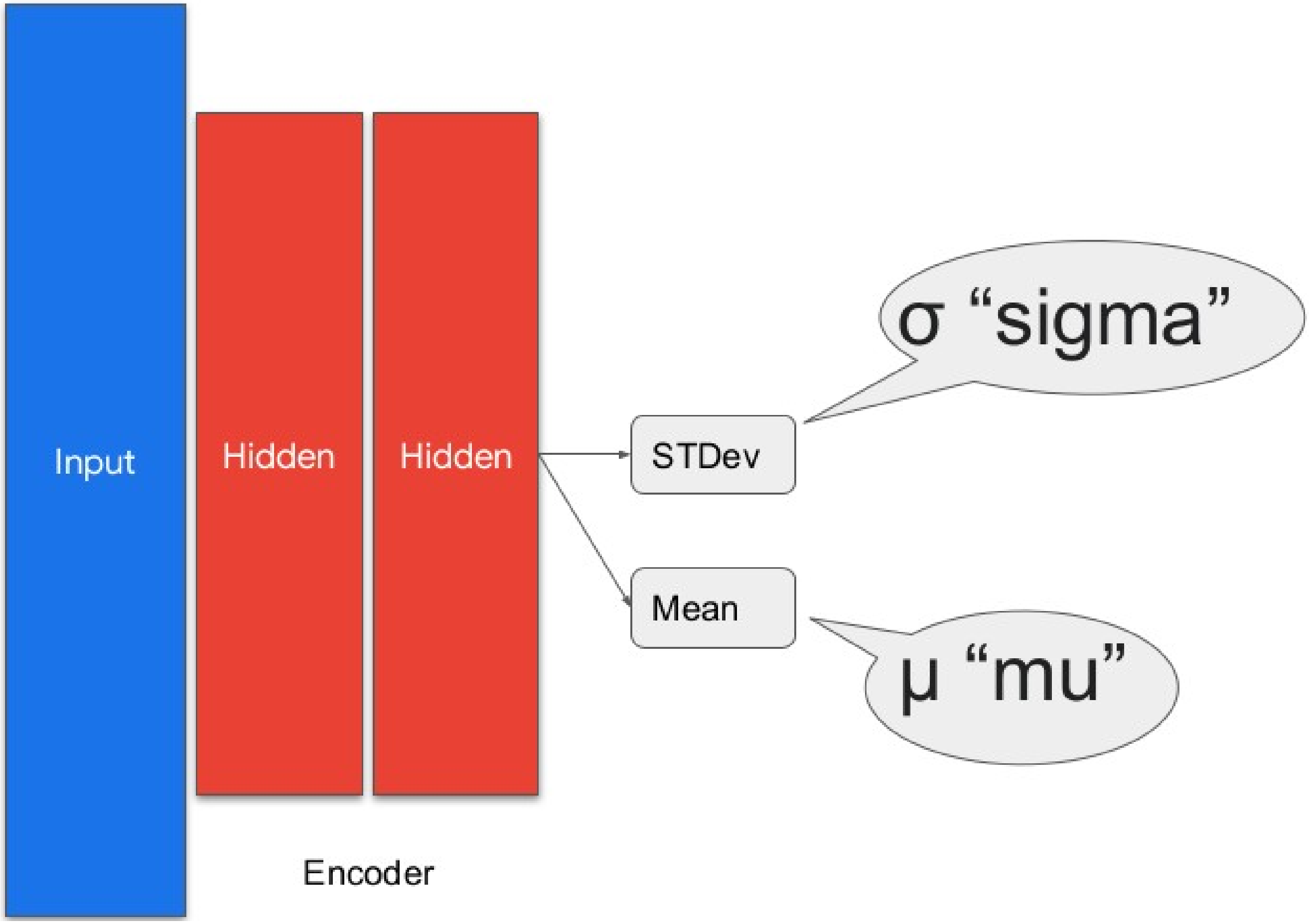
Key difference

This model tends to deal with statistical computations rather than bottle neck layer (downsampling)

VAE latent layer



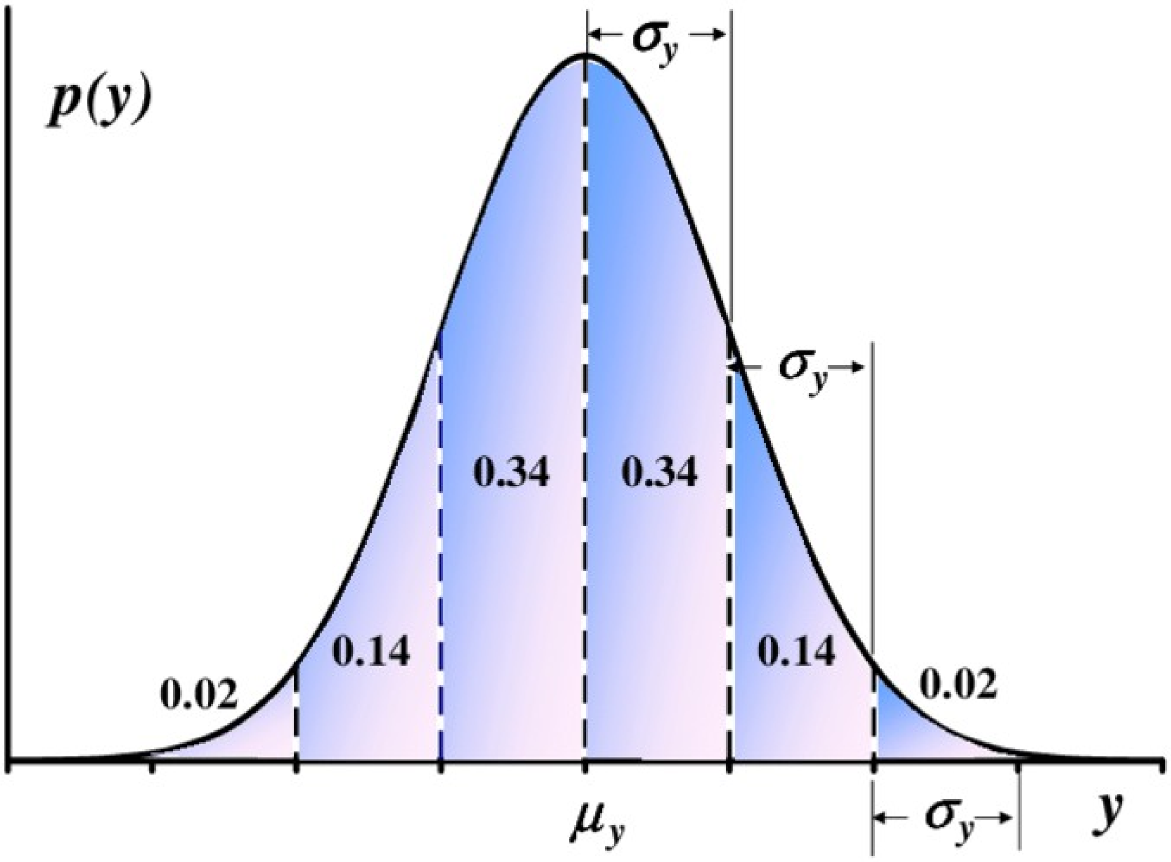
Encoder + latent layer



Sigma and Mu

Sigma and Mu usually are linked to a probability density function → Normal distribution or what is known as Gaussian density probability function

Gaussian distribution



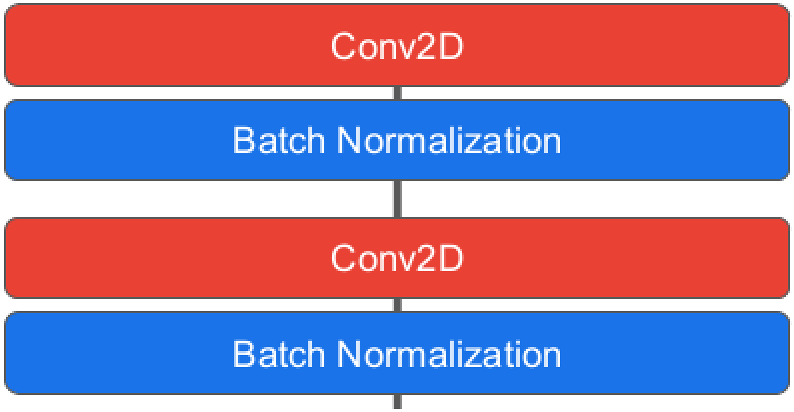
Terminology

Mu→ if we looked at the distribution , we notice that at the middle of the curve there is a function named Mu(y) , since its at the middle exactly , → Mu→ mean value of the curve

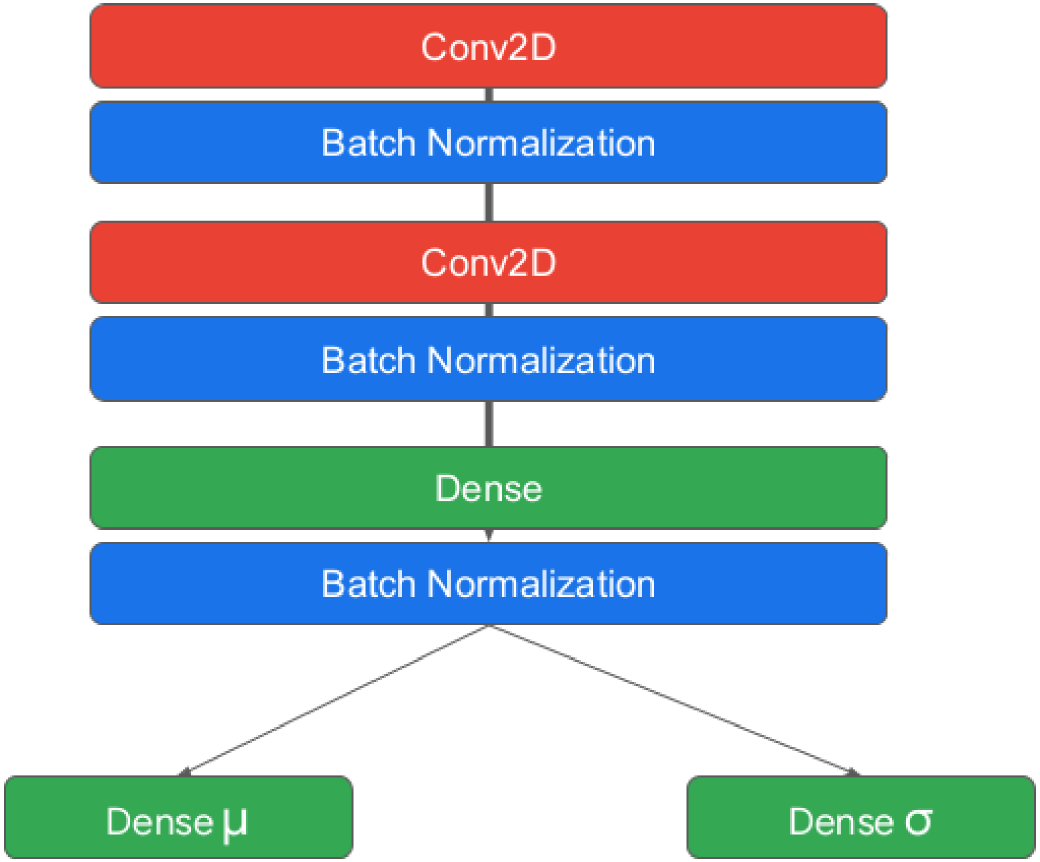
Sigma(y)→ if we looked to the curve slopping up and down , taking a slope of a deviating curve , therefore its known as the standard deviation pf the curve

Encoder

No maxpooling here why ?



Full encoder architecture



Important notes

The encoder function here

Must take 2 parameters (input shape and latent dimension)

Recall latent_dimension is the sigma and Mu

must return the following

Mu,Sigma and the flattened batch shape of the last conv2d layer (no maxpooling return)

Latent layer

Sampling layer (Mu and Sigma)

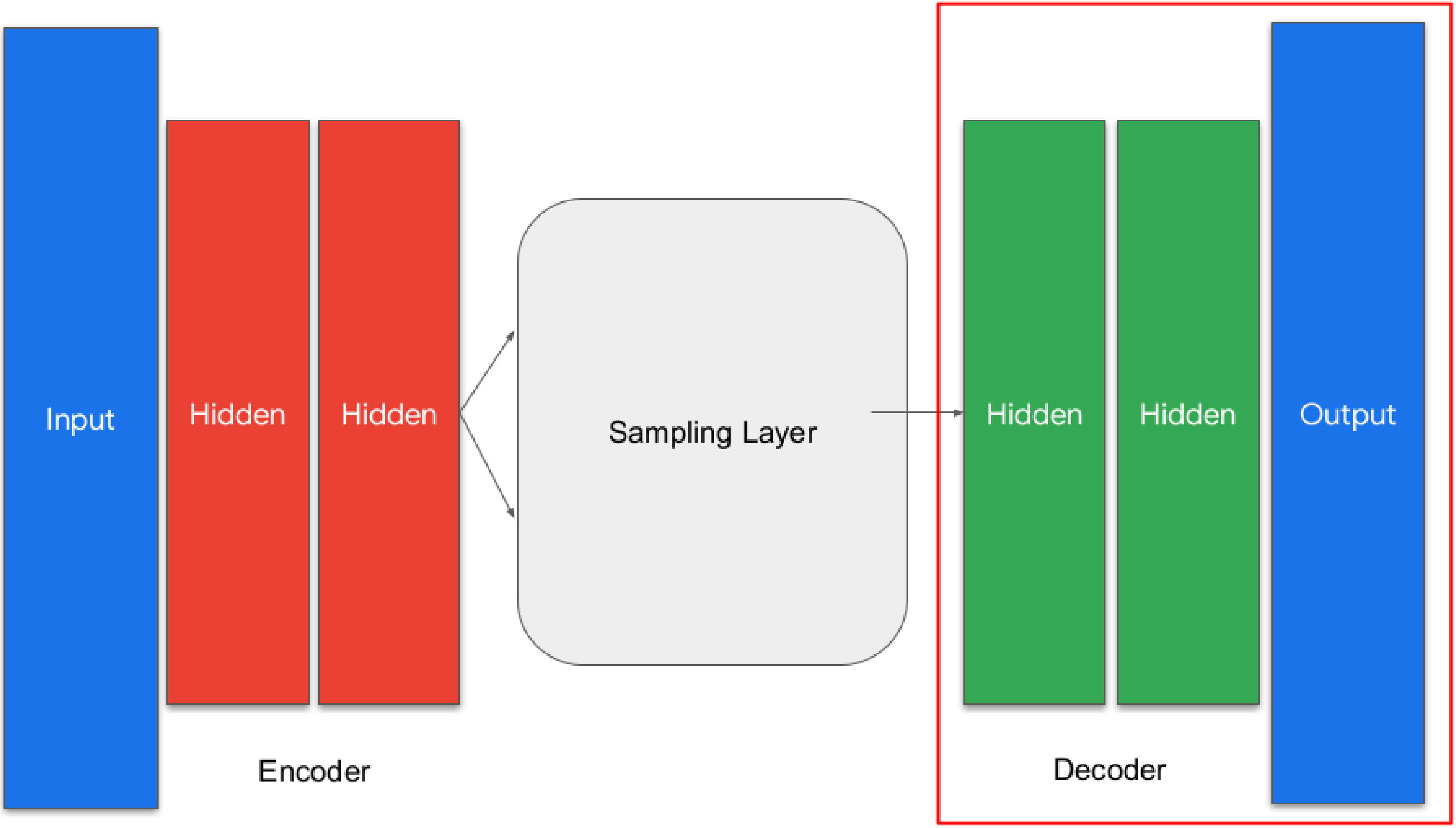
1st output of the encoder → Mu and sigma

2nd get the size and dimensions of the batch

3rd generate a random tensor (generated from Mu and sigma , using gaussian distribution)

4th combine the inputs and noise

Decoder


Input
Hidden
Hidden
Sampling Layer
Hidden
Hidden
Output
Encoder
Decoder

Decoder

Traditional decoder except it uses conv2D transpose in its layers

Loss Kullback leiber loss

Kullback Leibler divergence

- P = true distribution;
- Q = alternative distribution that is used to encode data
- KL divergence is the expected extra message length per datum that must be transmitted using Q

$$
\begin{aligned}
D_{KL}(P \,||\, Q) &= \Sigma_i \, P(x_i) \log \,(P(x_i)/Q(x_i)) \\
&= \Sigma_i \, P(x_i) \log P(x_i) - \Sigma_i \, P(x_i) \log Q(x_i) \\
&= H(P,Q) - H(P) \\
&= \text{Cross-entropy - entropy}
\end{aligned}
$$

- Measures how different the two distributions are

To simplify KLD

-1/2*mean of (1+sigma – (Mu)^2 - exp(sigma))