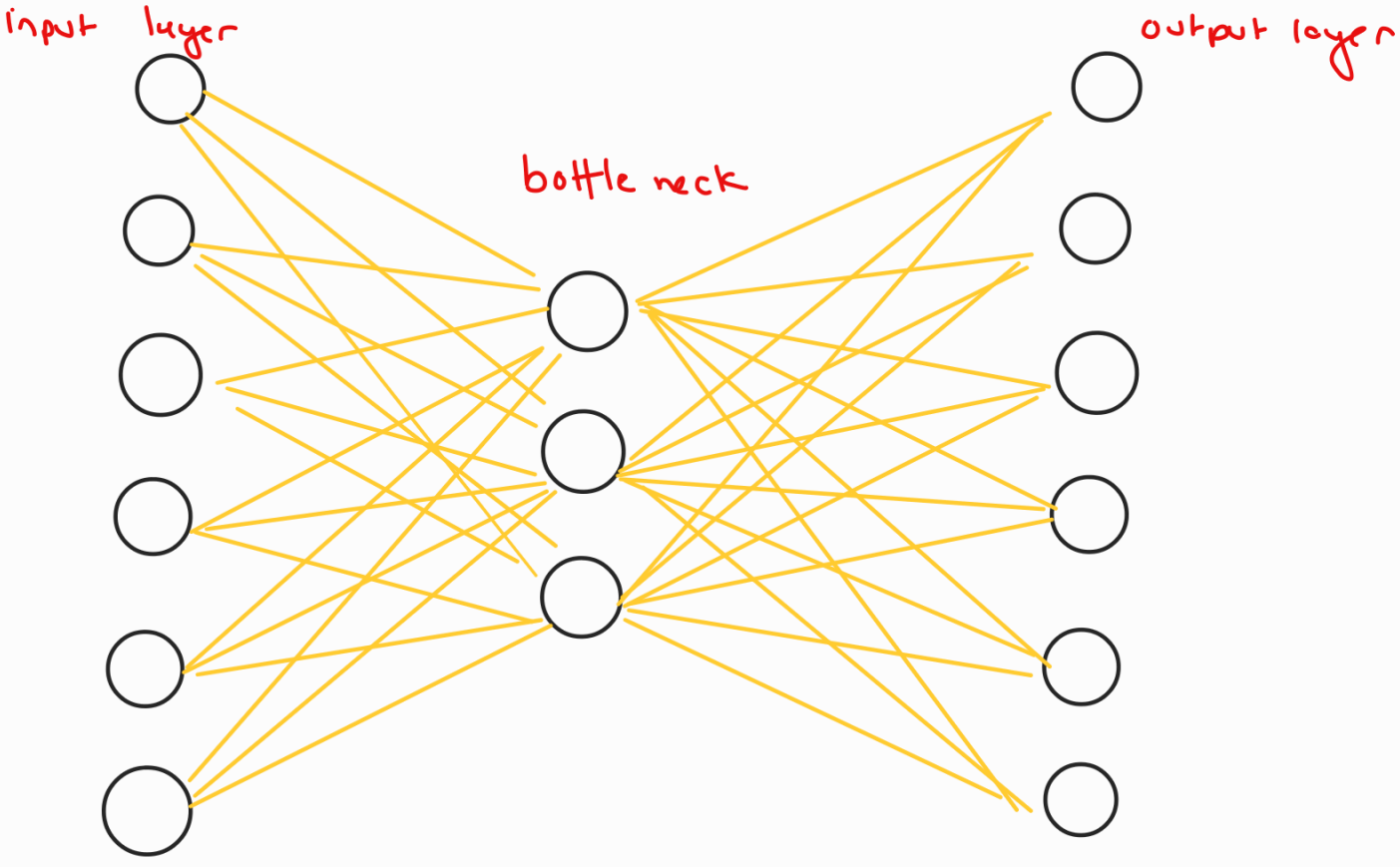
Auto encoders

An autoencoder is a type of artificial neural network used to learn efficient coding of unlabeled (unsupervised learning)

An autoencoder learns two functions : an encoding fn. that transforms the input data, and a decoding fn that recreates the input data from the encoded representation.



Encoder :

The encoder takes the input and compresses it to the lower - dimensional representation, known as latent space or encoded representation

Decoder

The decoder takes the compressed representation and tries to reconstruct the original input. The goal is to minimize the difference between the original and the reconstructed data.

Hidden Layer (Latent Space)

The hidden layer is the "bottleneck" of the network, where the compressed representation (also called the latent vector or encoding) resides.

This representation contains the most important features of the original data.

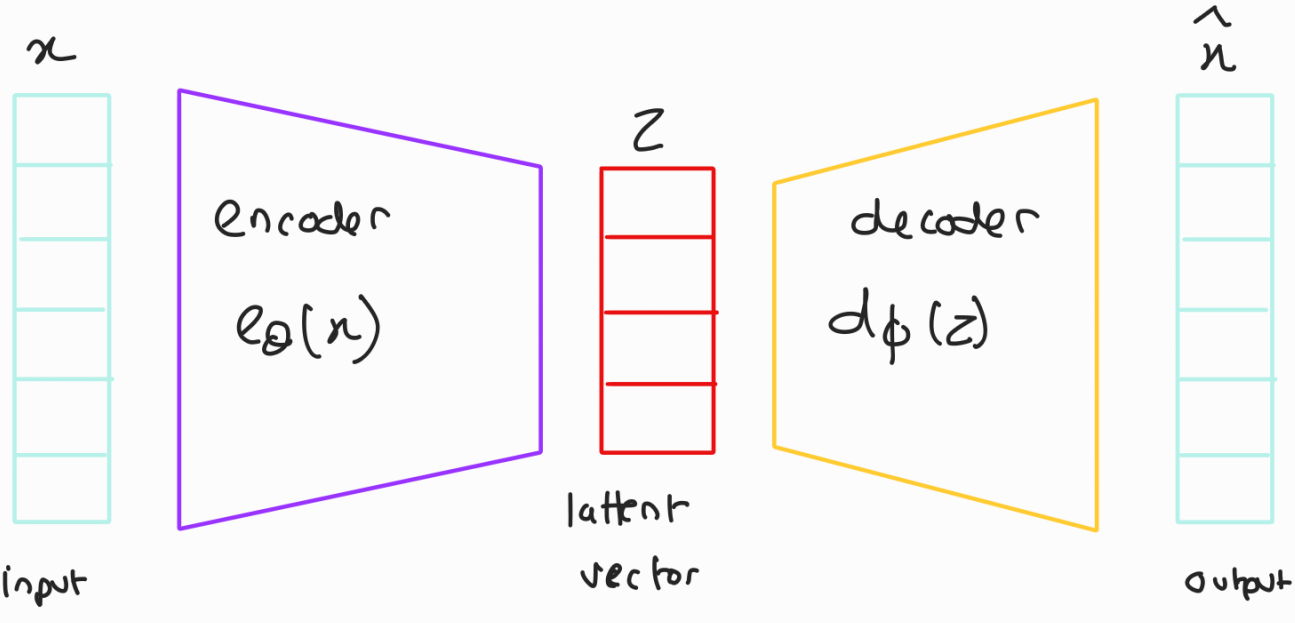
The size of this hidden layer is critical parameter in autoencoder design.

Undercomplete Autoencoder.

The size of the hidden layer is smaller than the input leading to a more compact encoding.

Overcomplete Autoencoder.

The size of the hidden layer is larger than the input allowing the network to potentially capture more complex features.



Variational AutoEncoder

Variational autoencoder addresses the issue of non-regularized latent space in autoencoder and provides the generative capability to the entire space. The encoder in the AE outputs latent vectors Instead of outputting the vectors in the latent space, the encoder of VAE outputs the parameter of a pre-defined distribution in the latent space for every input The VAE the imposes a constraint on this latent distribution forcing it to be normal distribution. This constraint make sure that the latent space is regularized.