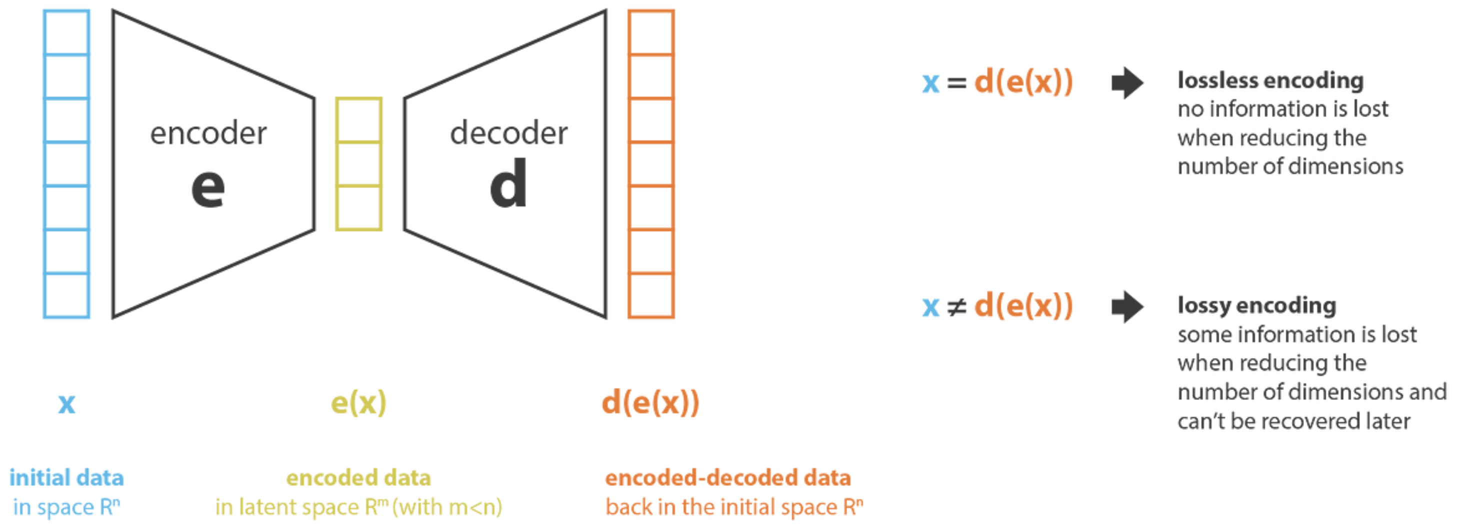
# Latent variables models, dimensionality reduction

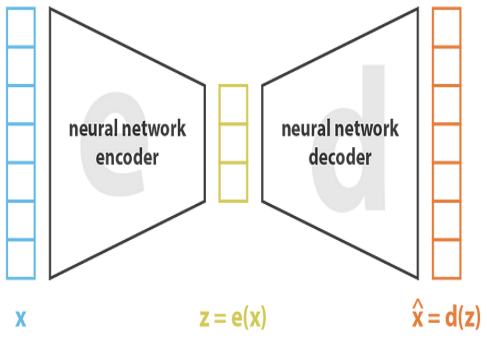
Dimensionality reduction is the process of reducing the number of features that describe some data.

Let's call **encoder** the process that produce the "new features" representation from the "old features" representation and **decoder** the reverse process. Dimensionality reduction can then be interpreted as data compression where the encoder compress the data (from the initial space to the **encoded space**, also called **latent space**) whereas the decoder decompress them.

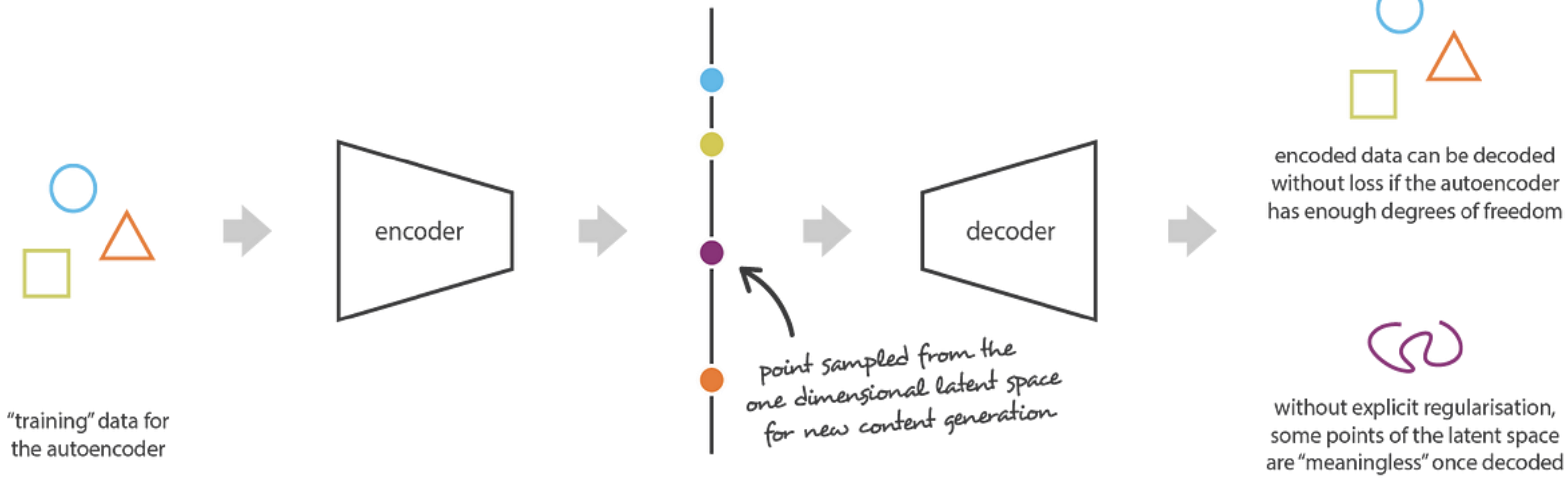


# Neural network for dimensionality reduction, Autoencoders

The general idea of autoencoders is pretty simple and consists in **setting an encoder and a decoder as neural networks** and to **learn the best encoding-decoding scheme using an iterative optimization process**.

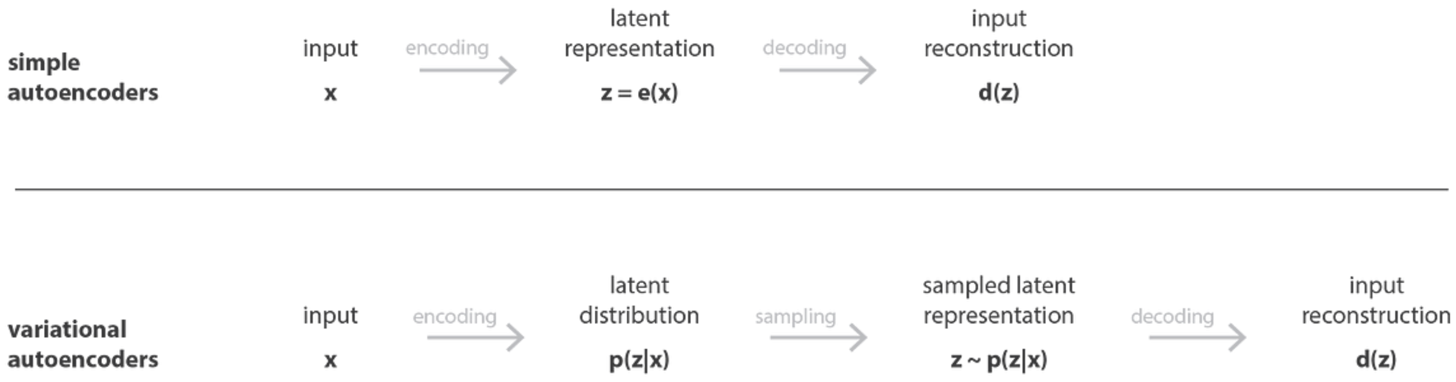



$$\text{loss} = \| x - \hat{x} \|^2 = \| x - d(z) \|^2 = \| x - d(e(x)) \|^2$$



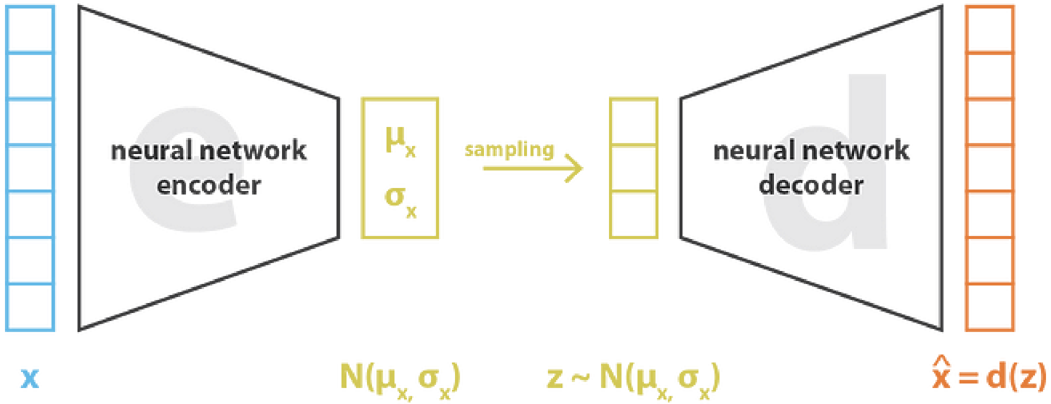
# Regularizing the latent space, Variational Autoencoders

A **variational autoencoder can be defined as being an autoencoder whose training is regularized to avoid overfitting and ensure that the latent space has good properties that enable generative process.**



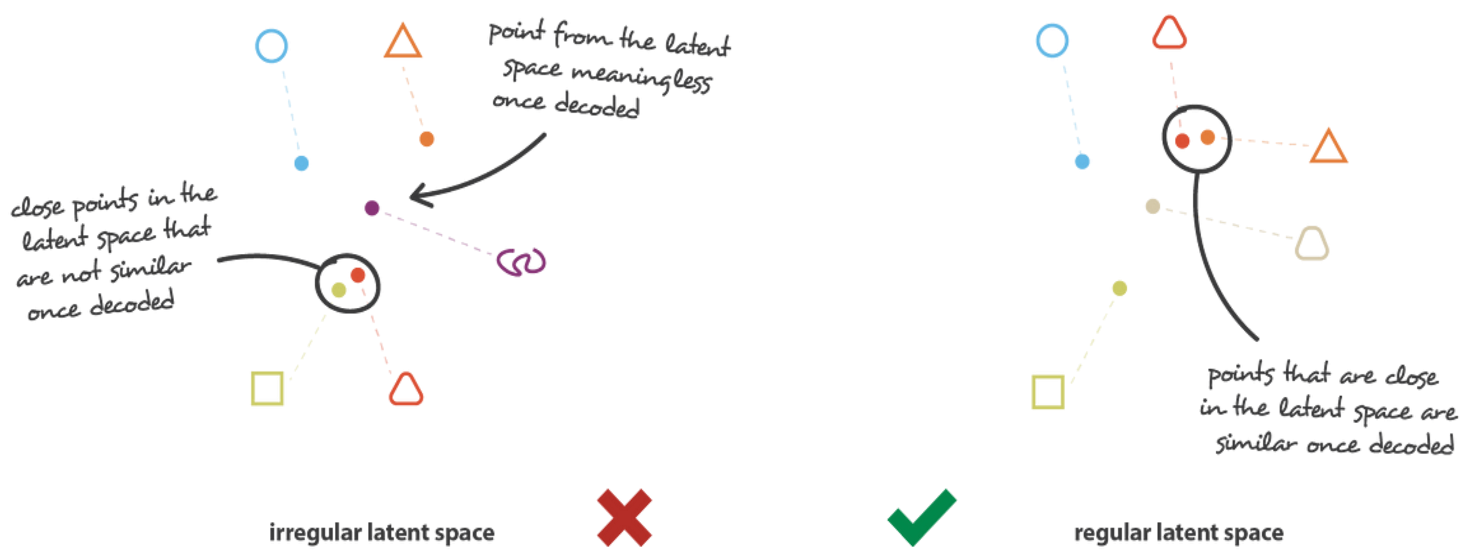

In practice, the encoded distributions are chosen to be normal so that the encoder can be trained to return the mean and the covariance matrix that describe these Gaussians. The reason why an input is encoded as a distribution with some variance instead of a single point is that it makes possible to express very naturally the latent space regularization: the distributions returned by the encoder are enforced to be close to a standard normal distribution

**The loss function that is minimized when training a VAE is composed of a "reconstruction term" (on the final layer), that tends to make the encoding-decoding scheme as performant as possible, and a "regularization term" (on the latent layer), that tends to regularize the organization of the latent space by making the distributions returned by the encoder close to a standard normal distribution. That regularization term is expressed as the Kulback-Leibler divergence between the returned distribution and a standard Gaussian**




$$\text{loss} = \| x - \hat{x} \|^2 + \mathrm{KL}[\, N(\mu_x, \sigma_x), N(0, I) \,] = \| x - d(z) \|^2 + \mathrm{KL}[\, N(\mu_x, \sigma_x), N(0, I) \,]$$

**Regularization tends to create a “gradient” over the information encoded in the latent space.**




Points in the latent space follows the desidered distribution