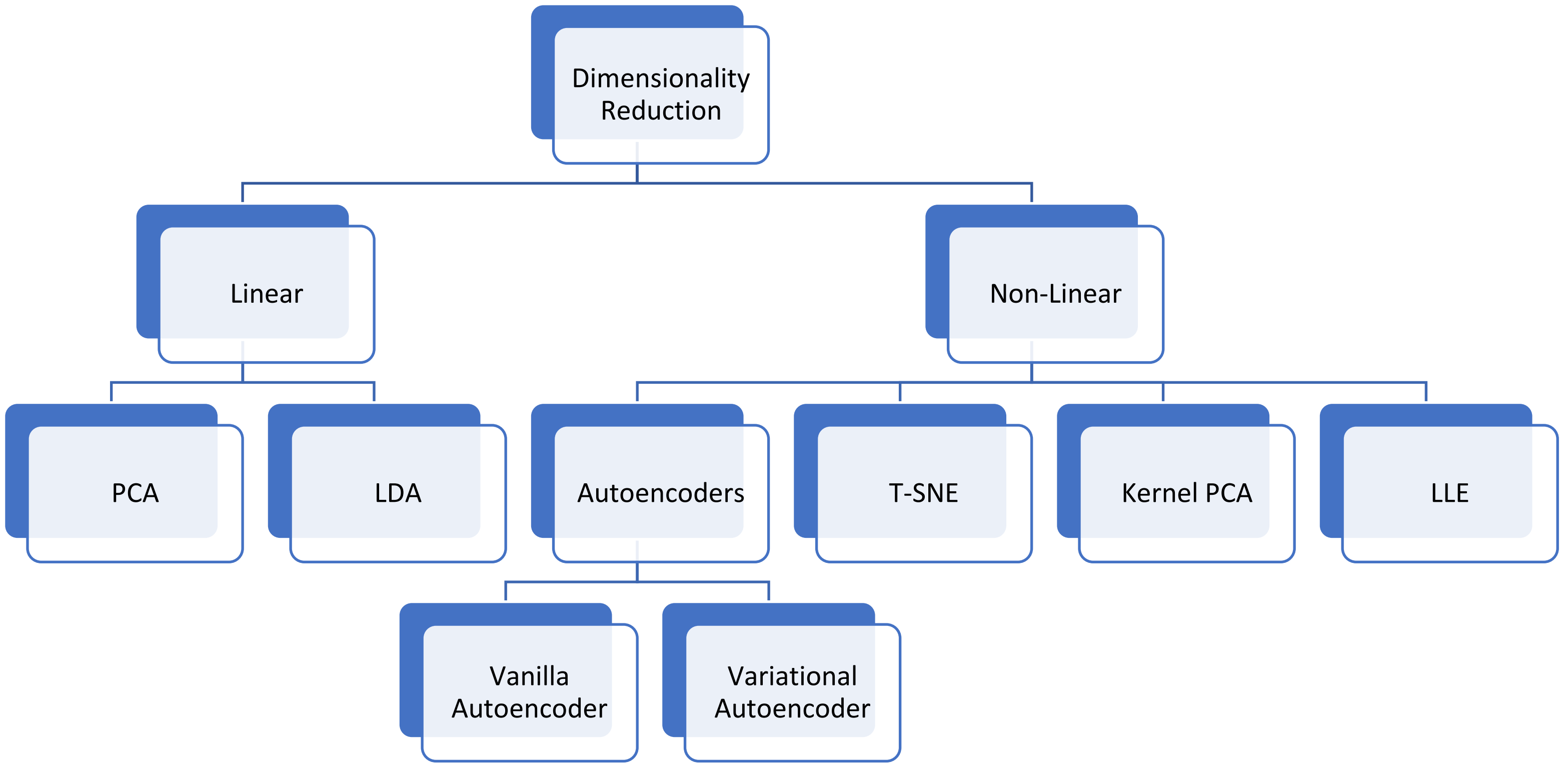
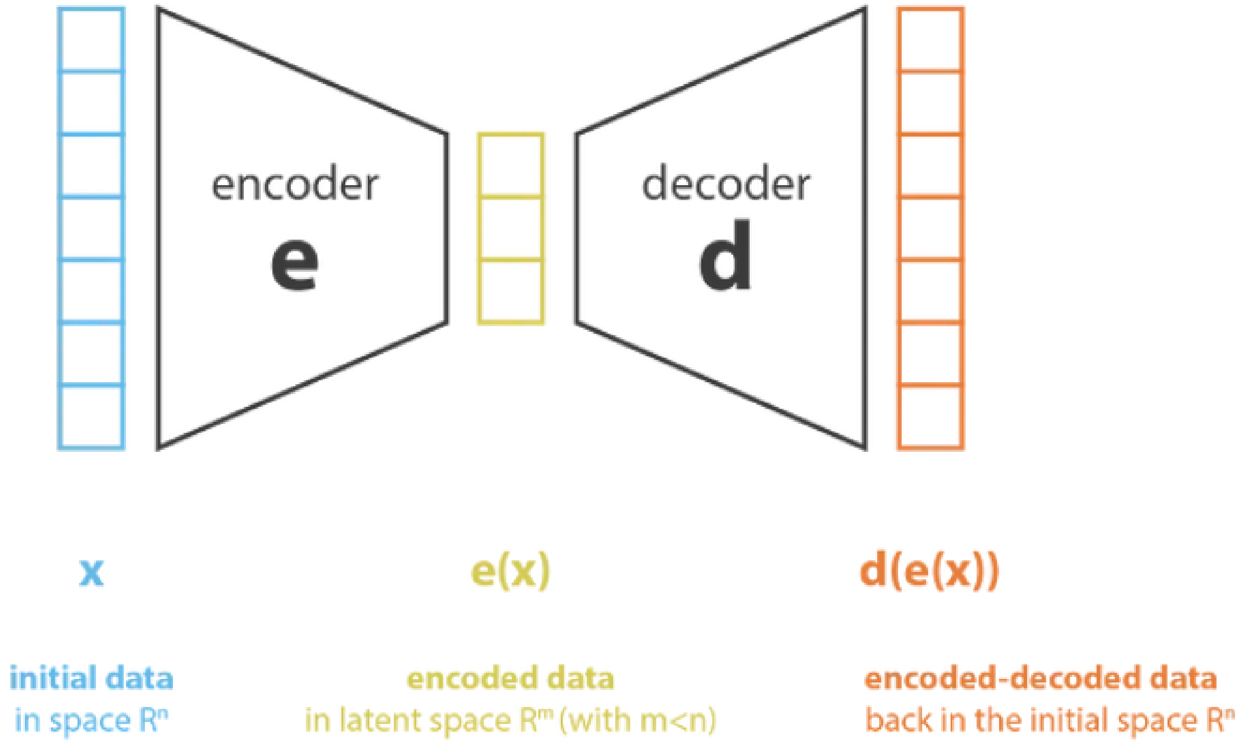
Non-Linear Dimensionality Reduction: Variational Autoencoders

Mostafa Tavassolipour

Dimensionality (Feature) Reduction



Dimensionality Reduction

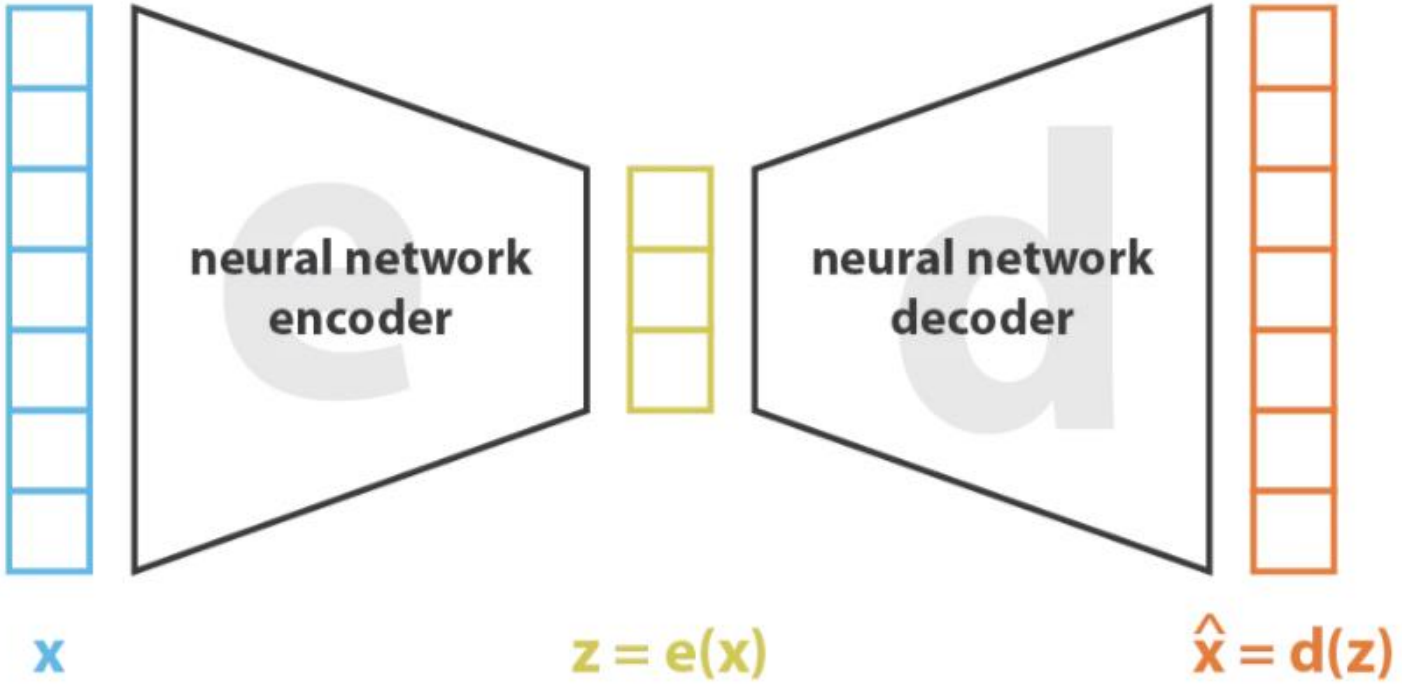


$x = d(e(x))$ ➔ **lossless encoding**
no information is lost when reducing the number of dimensions

$x \neq d(e(x))$ ➔ **lossy encoding**
some information is lost when reducing the number of dimensions and can't be recovered later

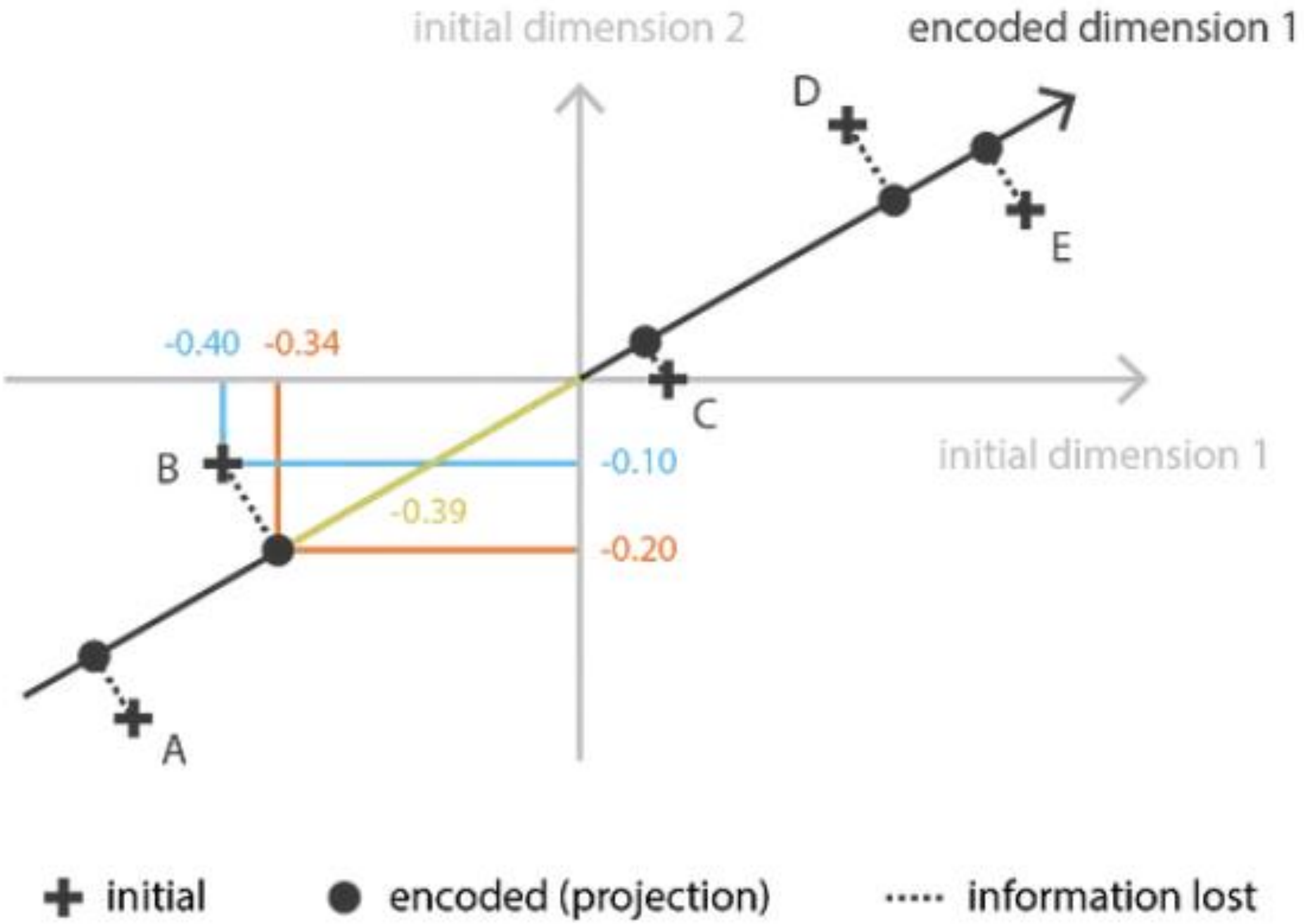
$$(e^*, d^*) = \arg\min \ \epsilon(x, d(e(x)))$$

Autoencoder



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

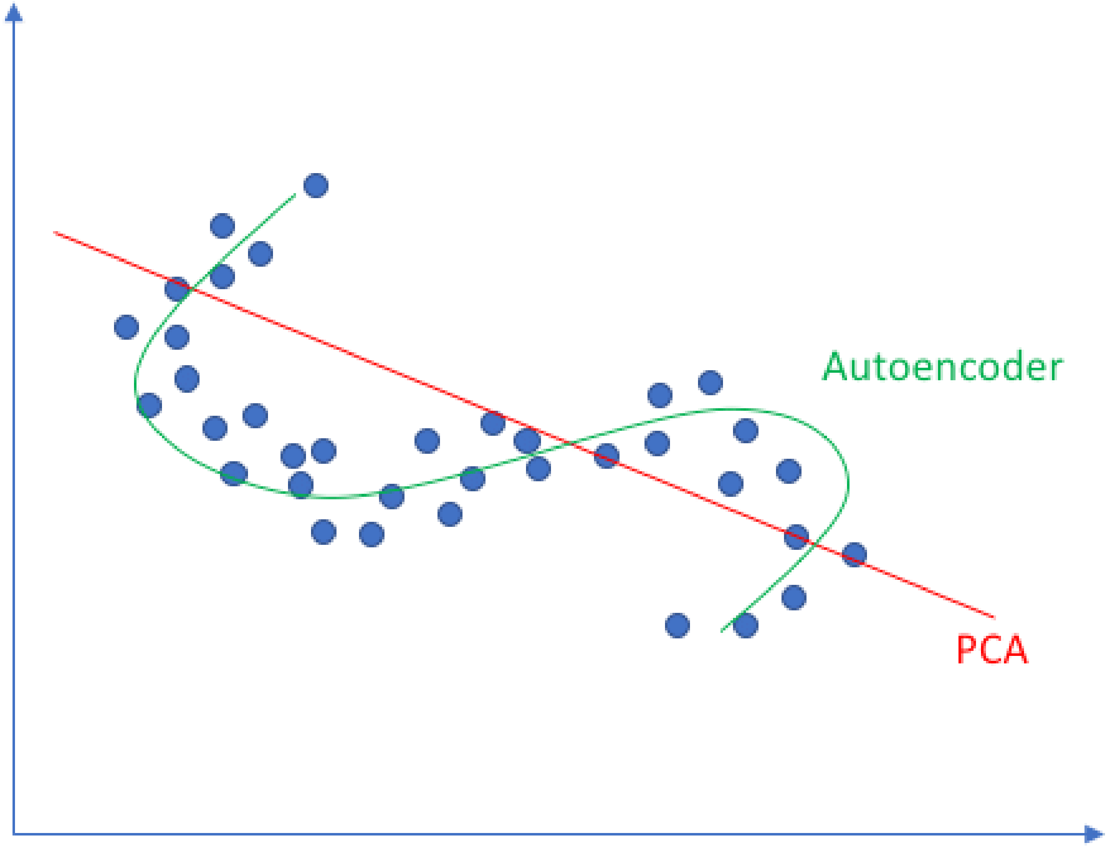
PCA



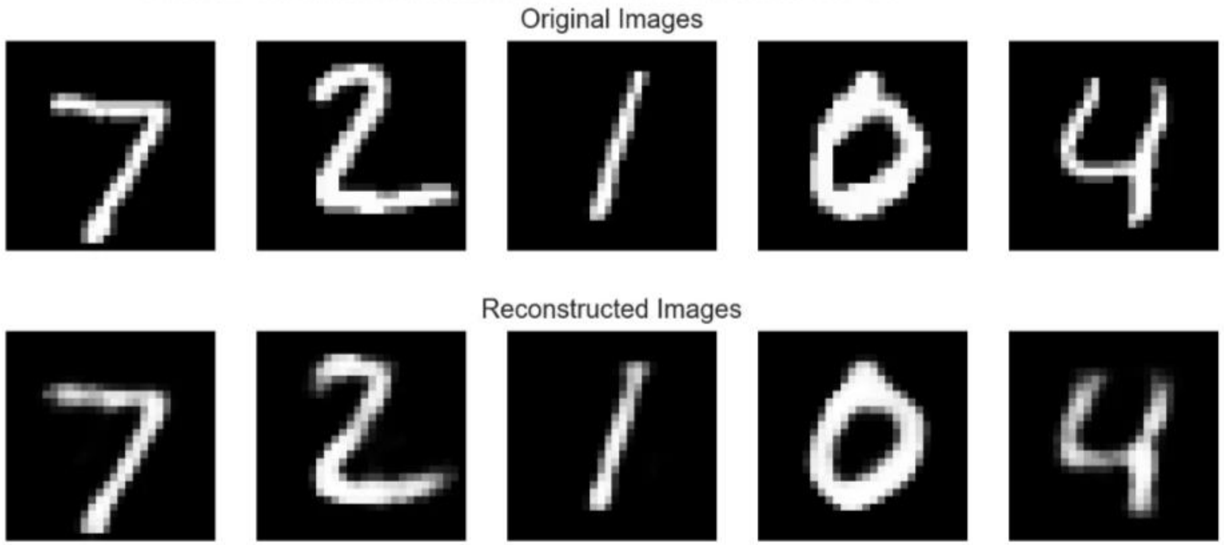
initial dimension 2
encoded dimension 1
D
E
-0.40
-0.34
C
initial dimension 1
B
-0.10
-0.39
-0.20
A
initial
encoded (projection)
information lost

Autoencoder vs PCA

Linear vs nonlinear dimensionality reduction

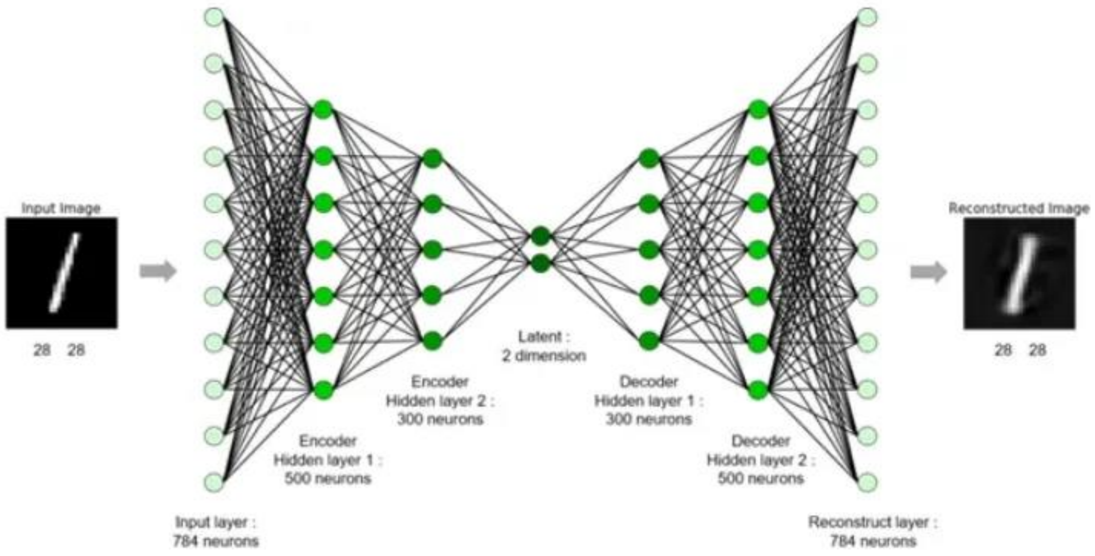


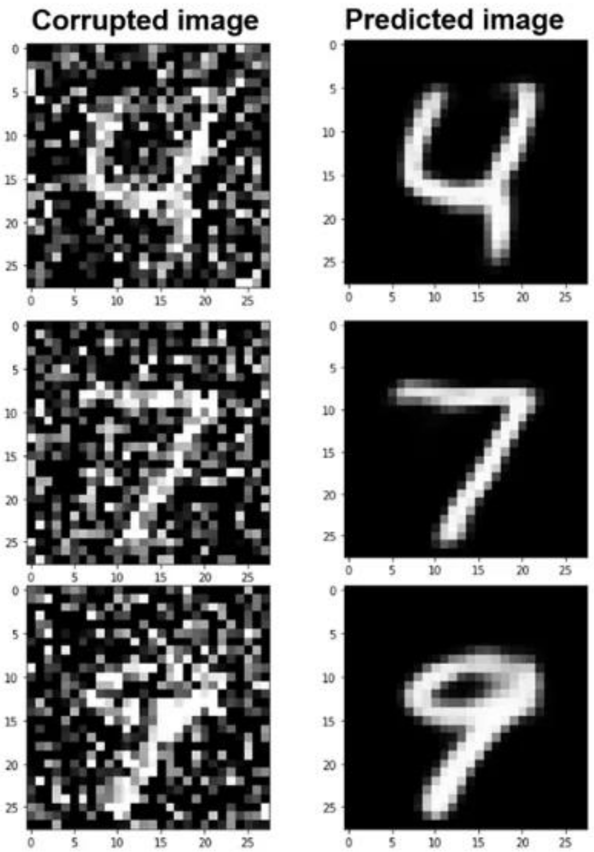
Autoencode on MNIST Dataset



Denoising Autoencoder

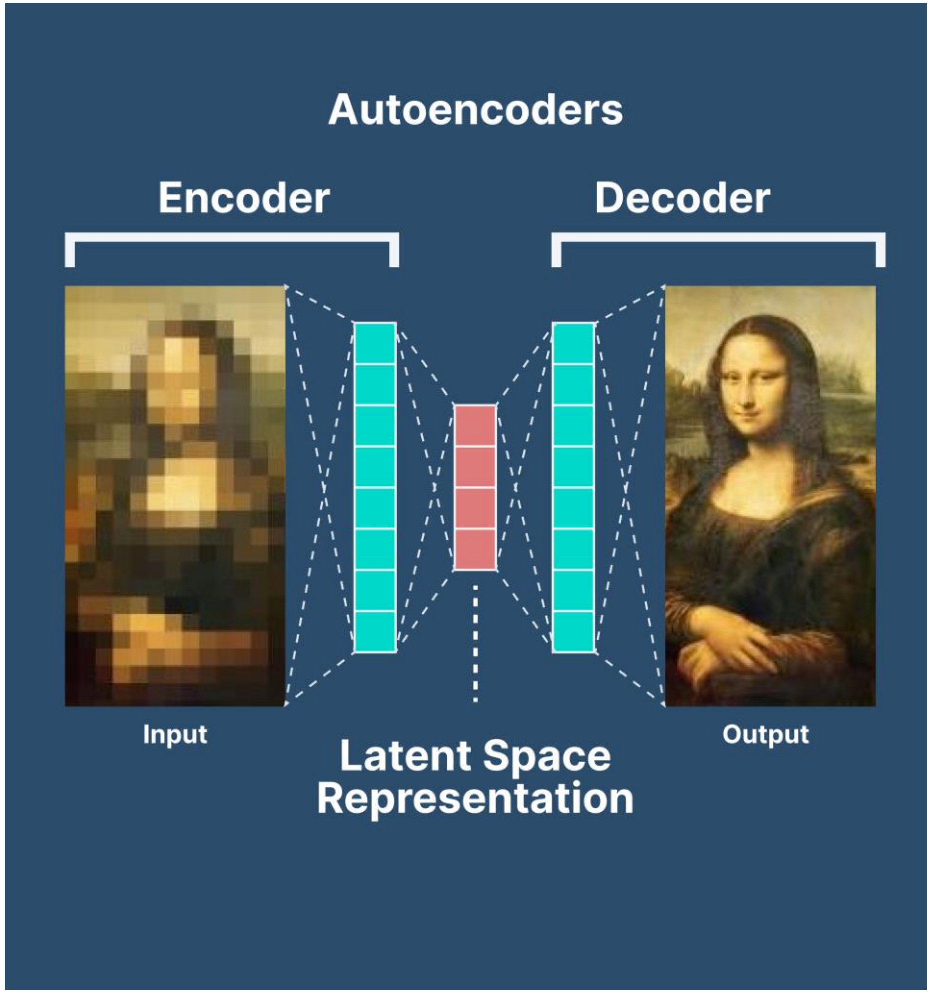
Self-Supervised



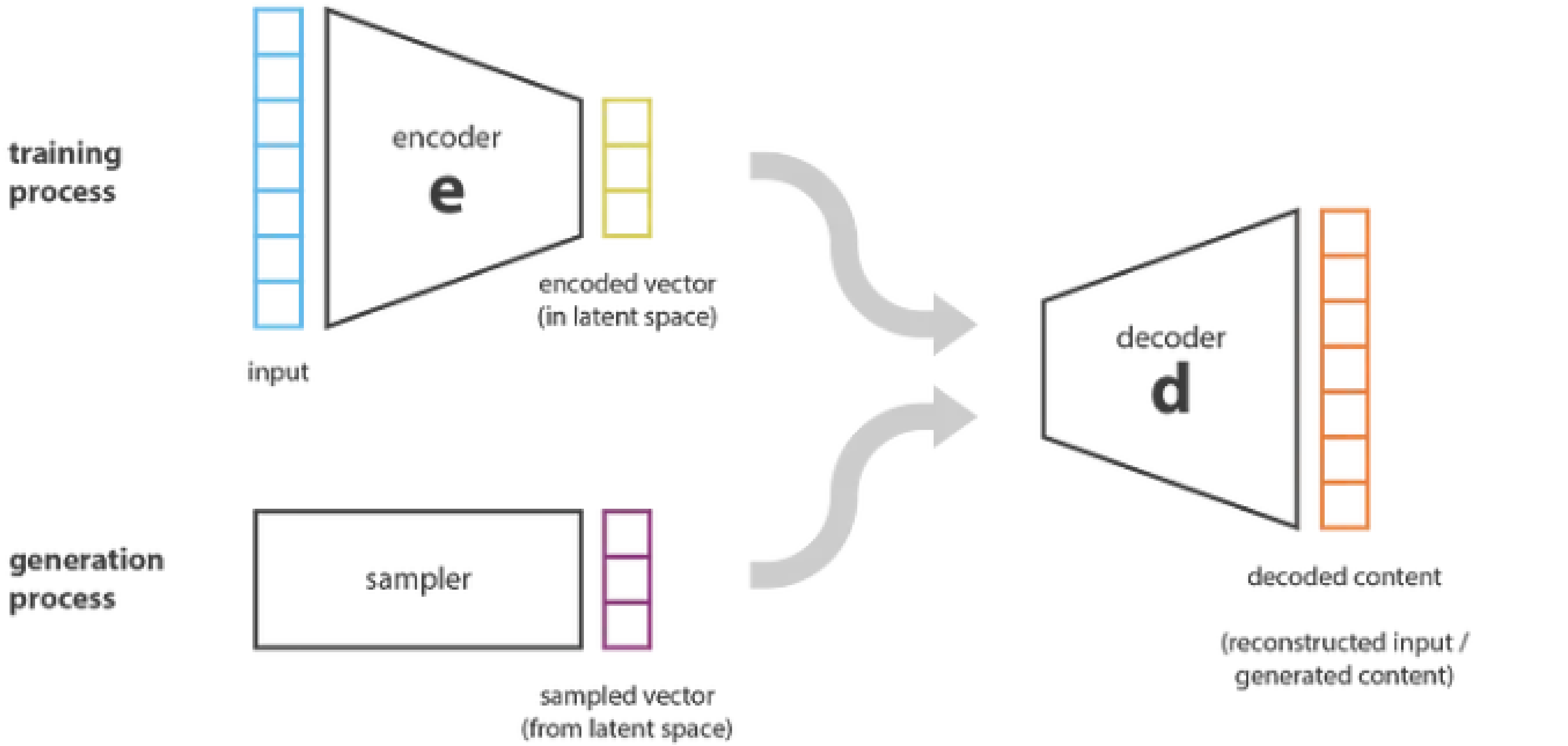


Super-Resolution using Autoencoder

- Self-Supervised



Limitations of Autoencoders for Generation



We can generate new data by decoding points that are randomly sampled from the latent space. The quality and relevance of generated data depend on the regularity of the latent space.

Autoencoders

- **IDEA**: Setting an encoder and a decoder as neural networks

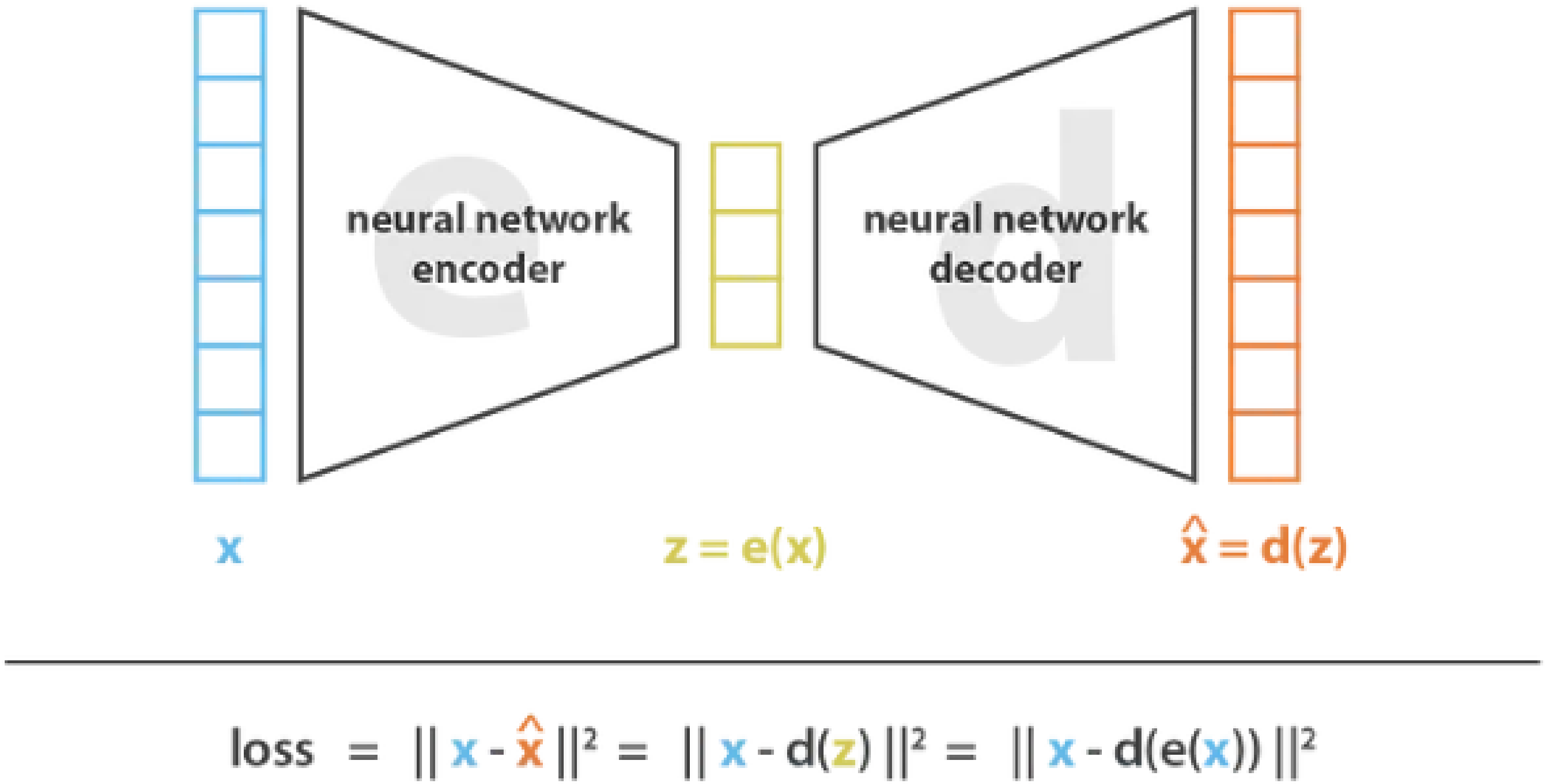
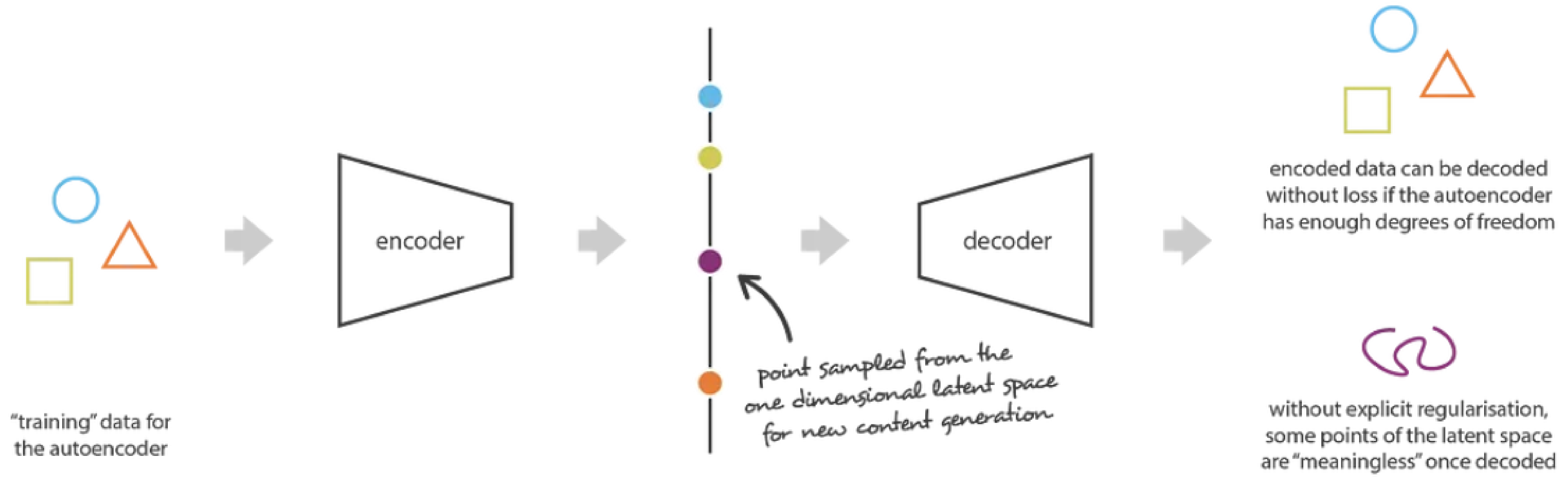


Illustration of an autoencoder with its loss function.

Sampling from Autoencoders

- The autoencoder is solely trained to encode and decode with as few loss as possible, no matter how the latent space is organized.

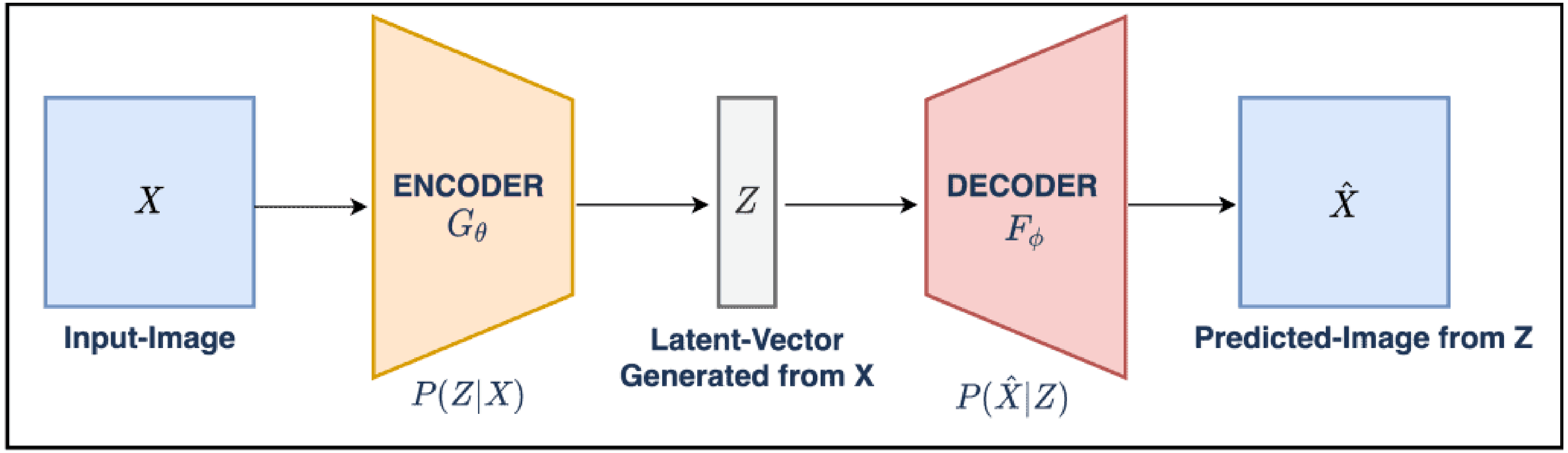


Irregular latent space prevent us from using autoencoder for new content generation.

Variational Autoencoder

- 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.

- **Encoding-decoding process**: instead of encoding an input as a single point, we encode it as a distribution over the latent space.



Variational Autoencoder

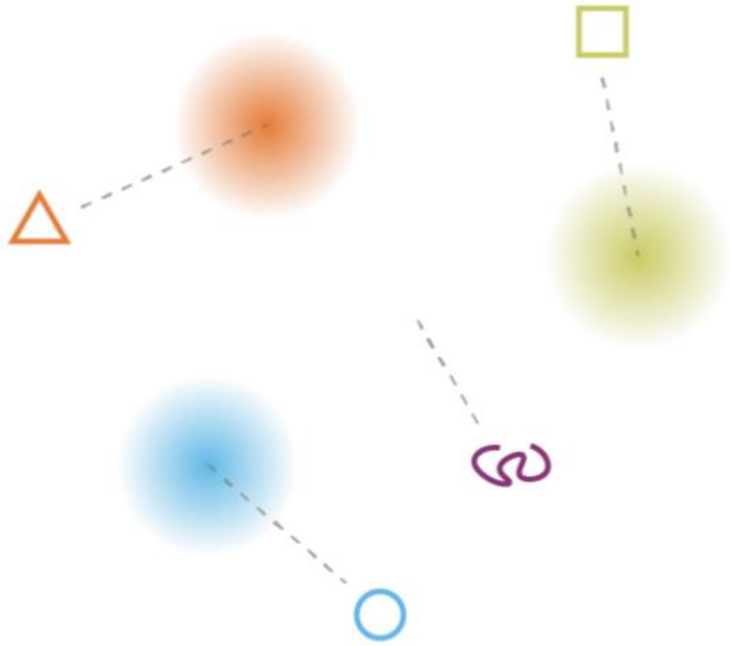
Face images generated with a Variational Autoencoder (source: Wojciech Mormul on Github).

Training VAE

- The input is encoded as a distribution over the latent space.
- A point from the latent space is sampled from that distribution.
- The sampled point is decoded and the reconstruction error can be computed.
- Finally, the reconstruction error is backpropagated through the network.

Regular Latent Space

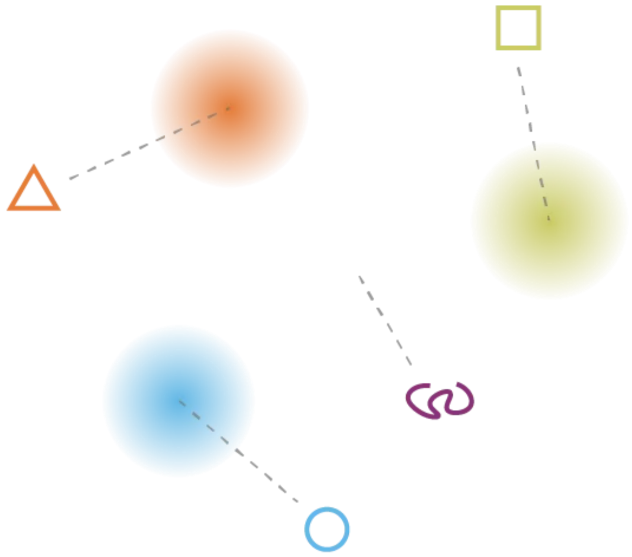
Variation Autoencoder: Latent Space



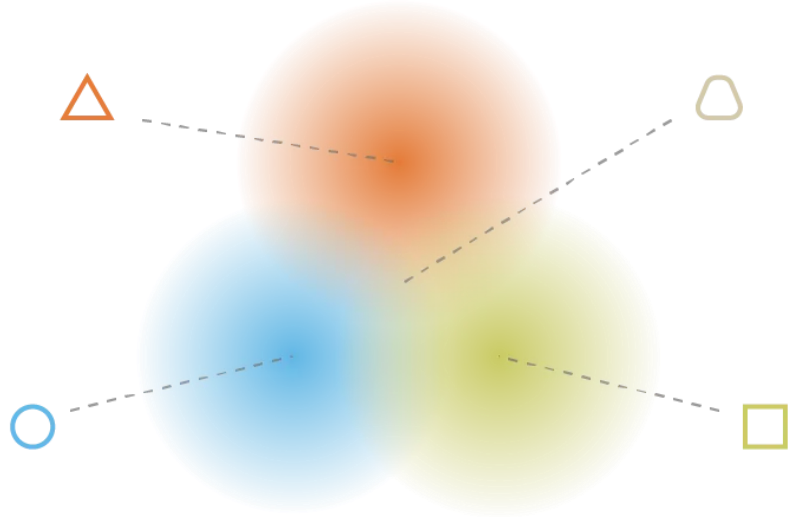
Autoencoder vs Variational autoencoder

simple autoencoders	input **x**	encoding →	latent representation $\mathbf{z = e(x)}$	decoding →	input reconstruction $\mathbf{d(z)}$			
variational autoencoders	input **x**	encoding →	latent distribution $\mathbf{p(z \mid x)}$	sampling →	sampled latent representation $\mathbf{z \sim p(z \mid x)}$	decoding →	input reconstruction $\mathbf{d(z)}$	

Regularized Latent Space

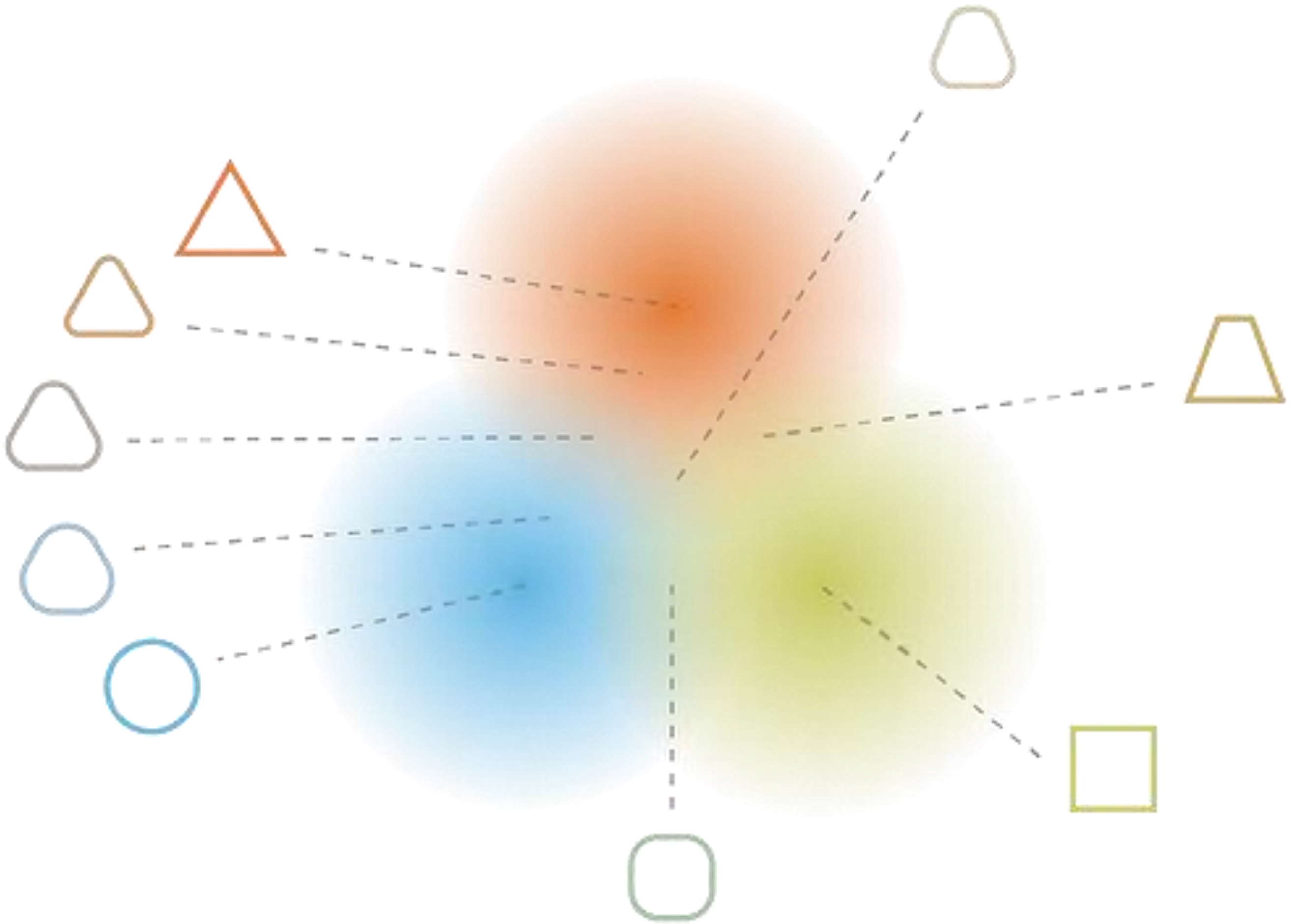


what can happen without regularisation

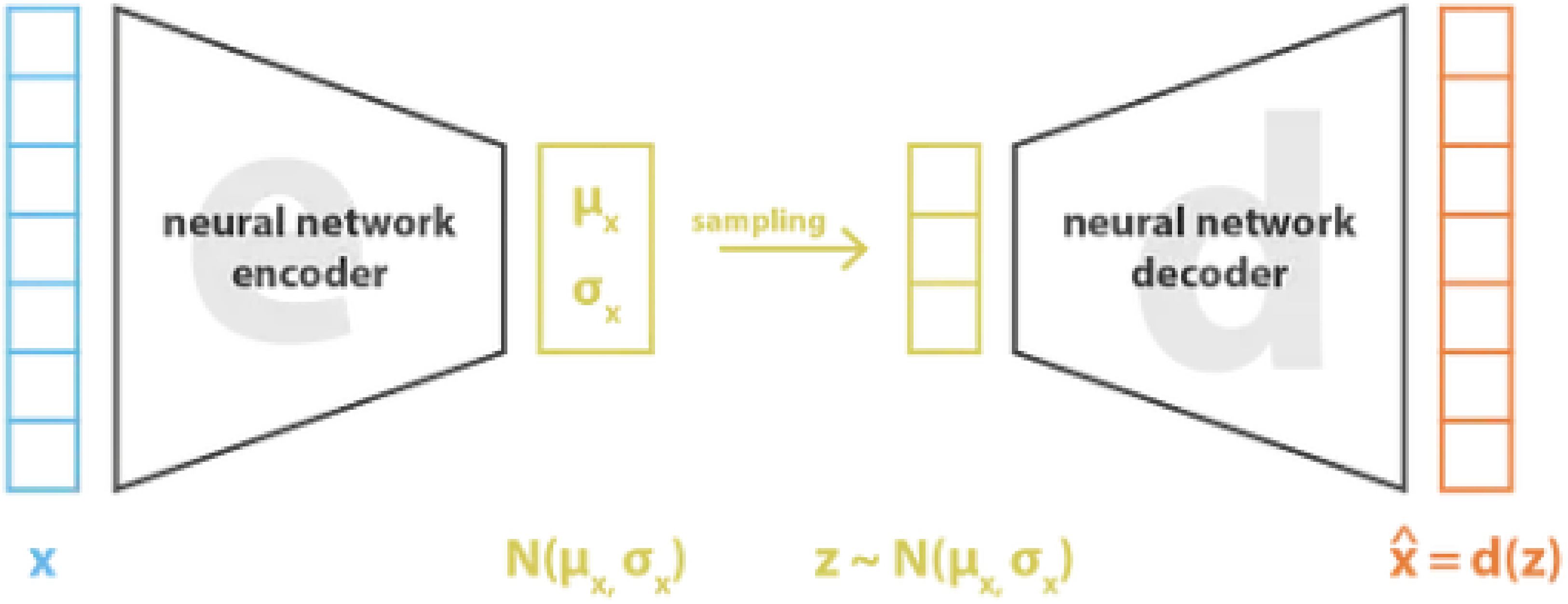


what we want to obtain with regularisation

Gradient over Latent Space



VAE: Loss Function



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

In variational autoencoders, the loss function is composed of a reconstruction term (that makes the encoding-decoding scheme efficient) and a regularisation term (that makes the latent space regular).

VAE on MNIST Dataset