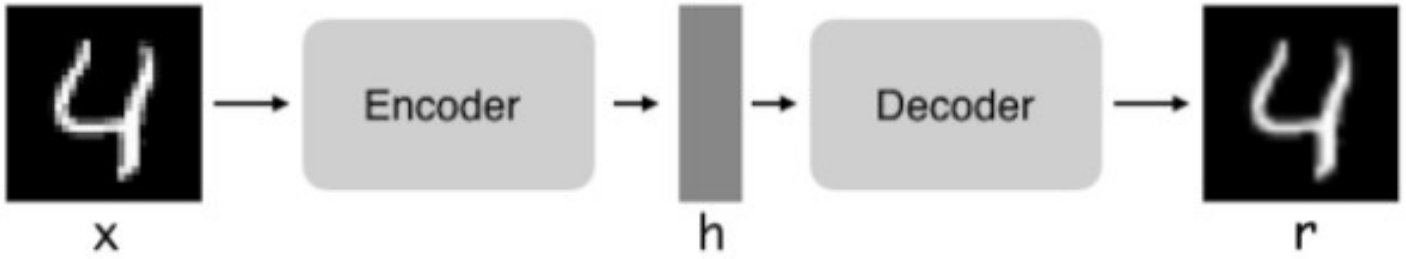
Module 4

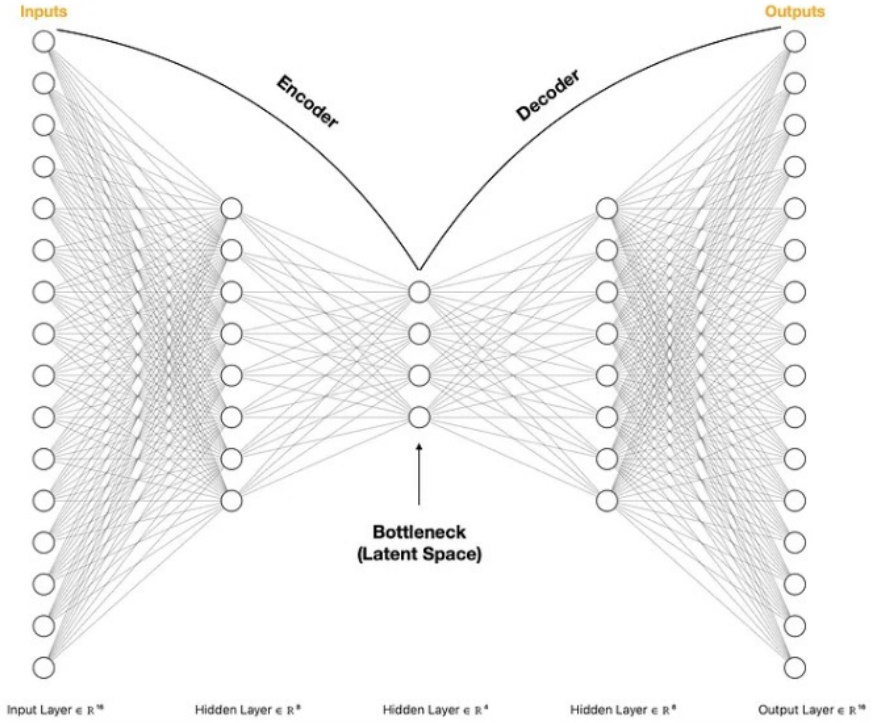
Deep Learning: AutoEncoders

DL: AutoEncoders



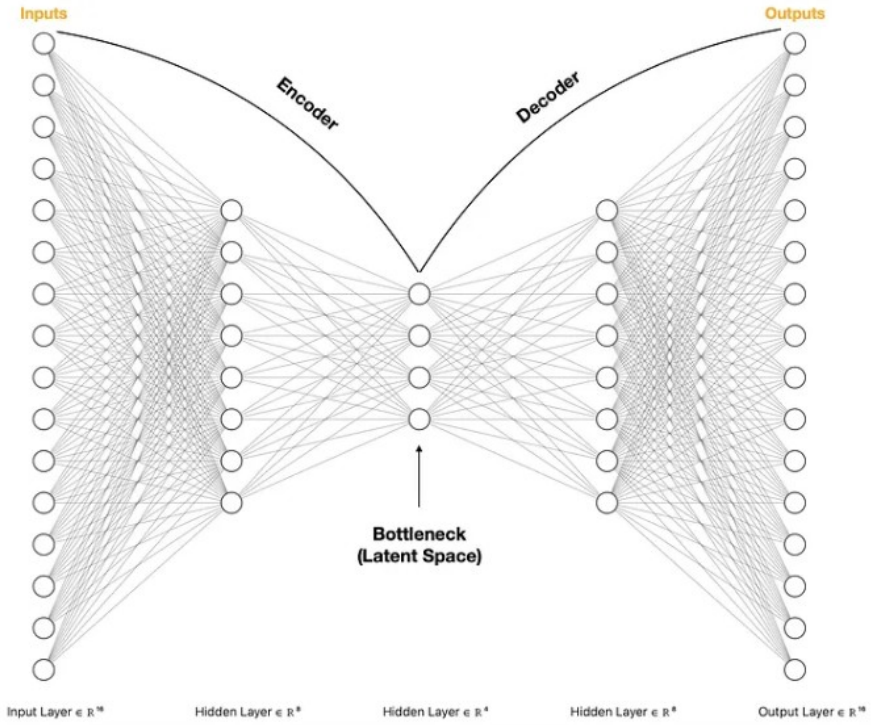
AutoEncoders

- Unsupervised learning NN architecture used to learn data codings
- Trained to reconstruct input data after compression
- Learns two functions: Encoder and Decoder



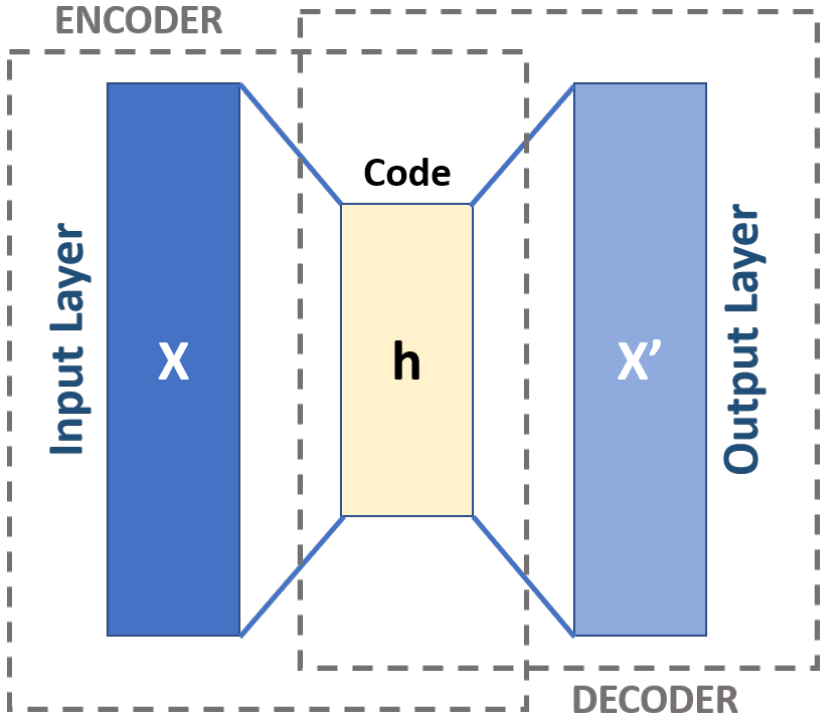
AutoEncoders

- Encoder: transforms the input data to a lower dimension
- Decoder: recreates the input data.



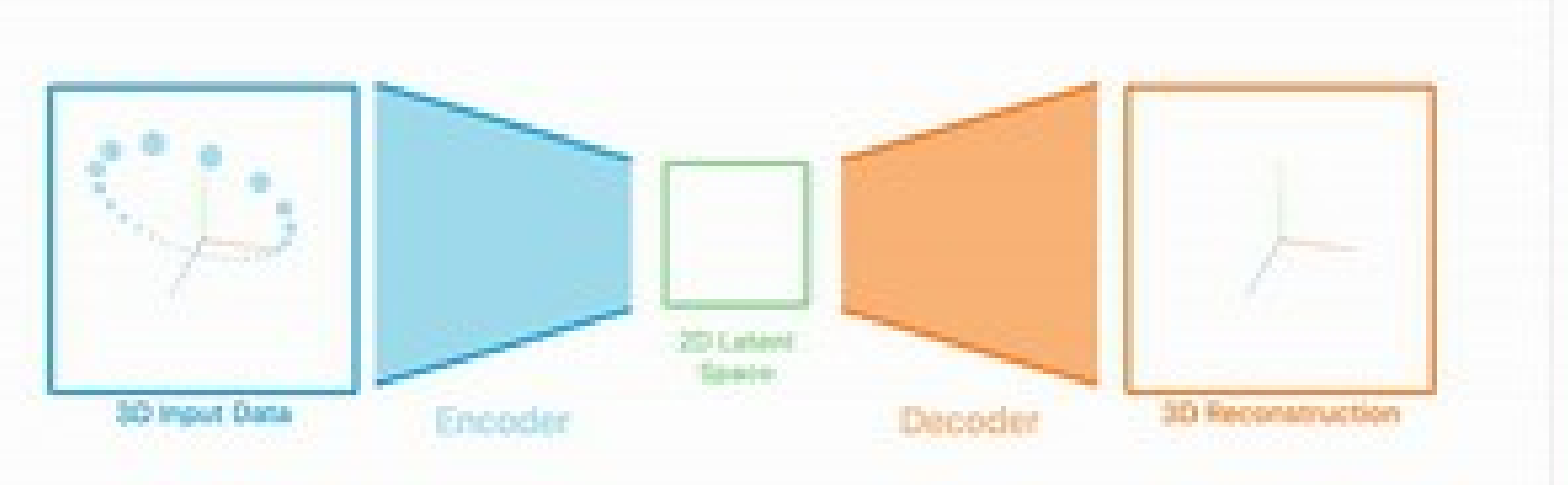
Schema of an AutoEncoder

- Input and Output layers have the same neurons (input vars)
- Encoder maps the message to a code (latent space or bottleneck)
- Decoder reconstructs the code



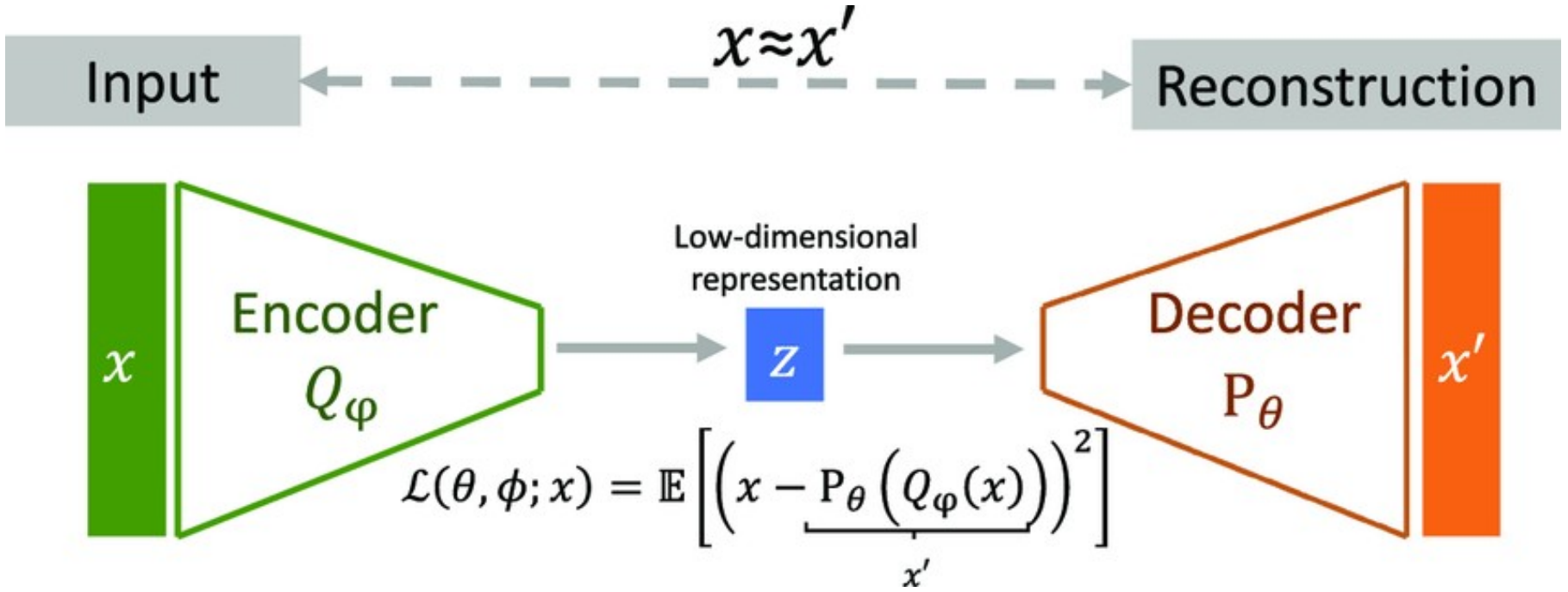
Layers

- Between encoder and decoder, the layer(s) act as a bottleneck
- Compressed knowledge representation of the original input
- The compressed data is also called the latent space



AutoEncoders

- Goal of autoencoder is to recreate the input values
- The reconstruction loss measures how good the autoencoder is

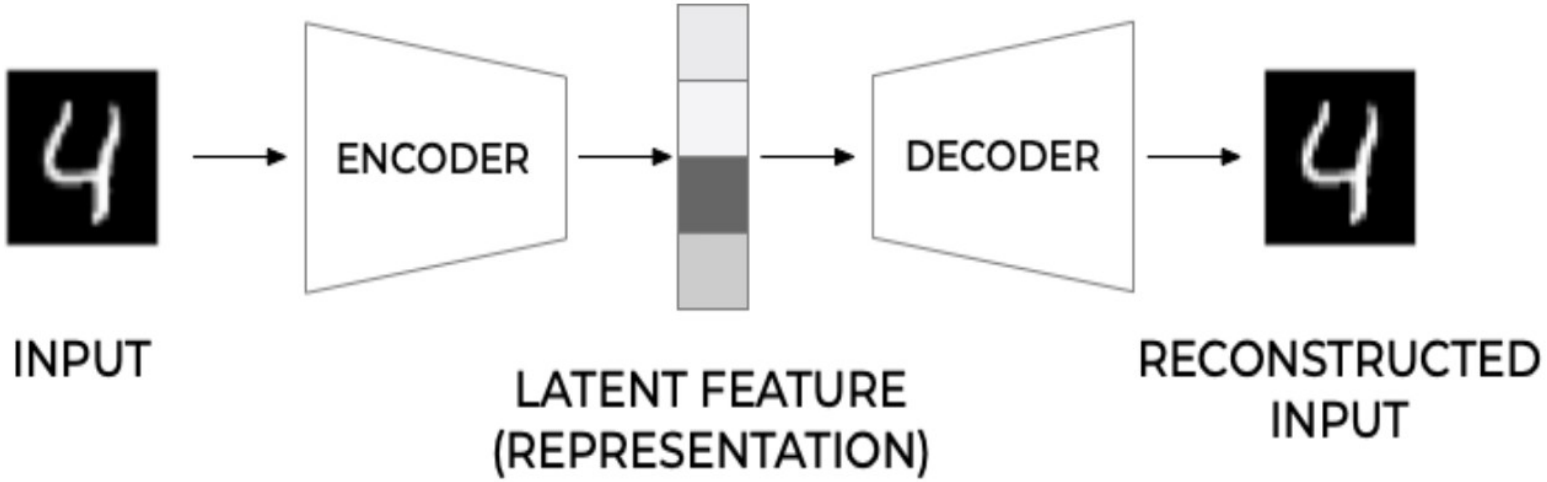


Why use AutoEncoders?

- Dimension reduction
- Noise reduction (denoising)
- Anomaly detection
- Feature extraction

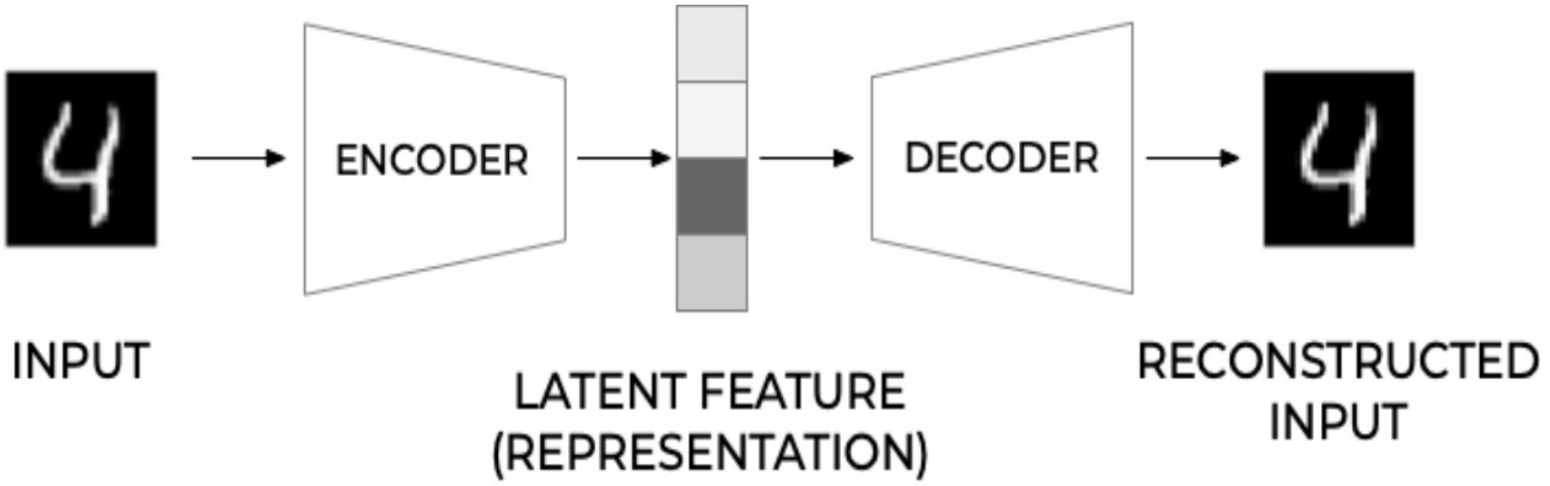
Mathematical Formulation

- Encoder: $h = g(x)$
- Decoder: f
- Reconstruction: $\tilde{x} = f(h)$
- Loss: Minimize Cost Function (MSE or Binary Cross-Entropy)



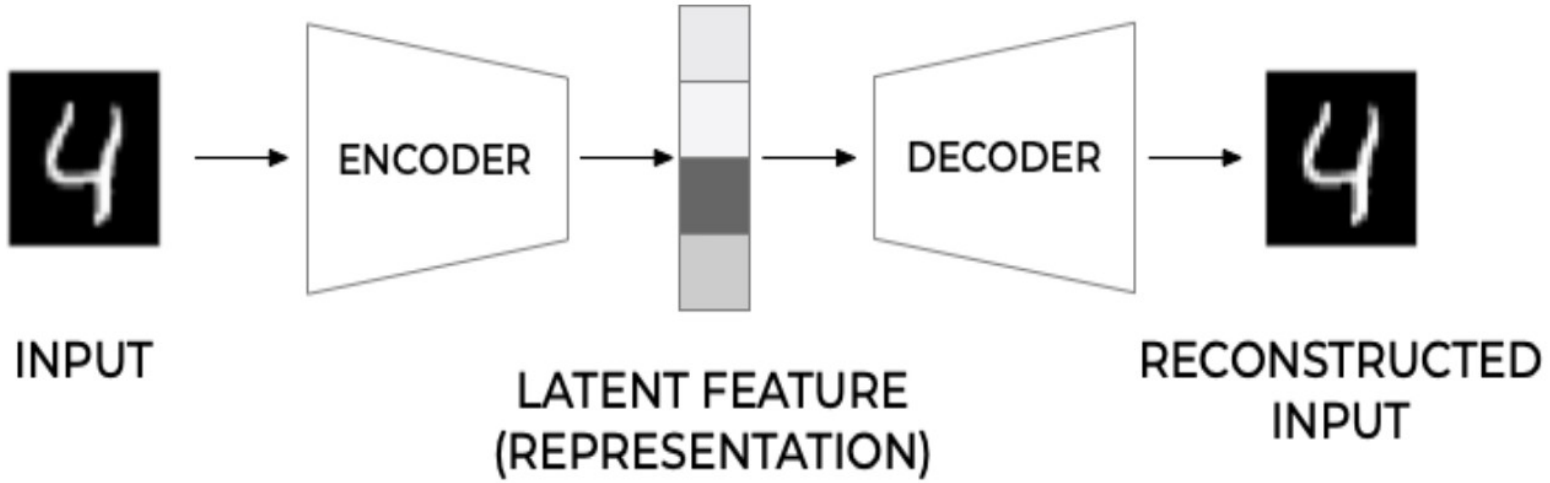
Dimensionality Reduction

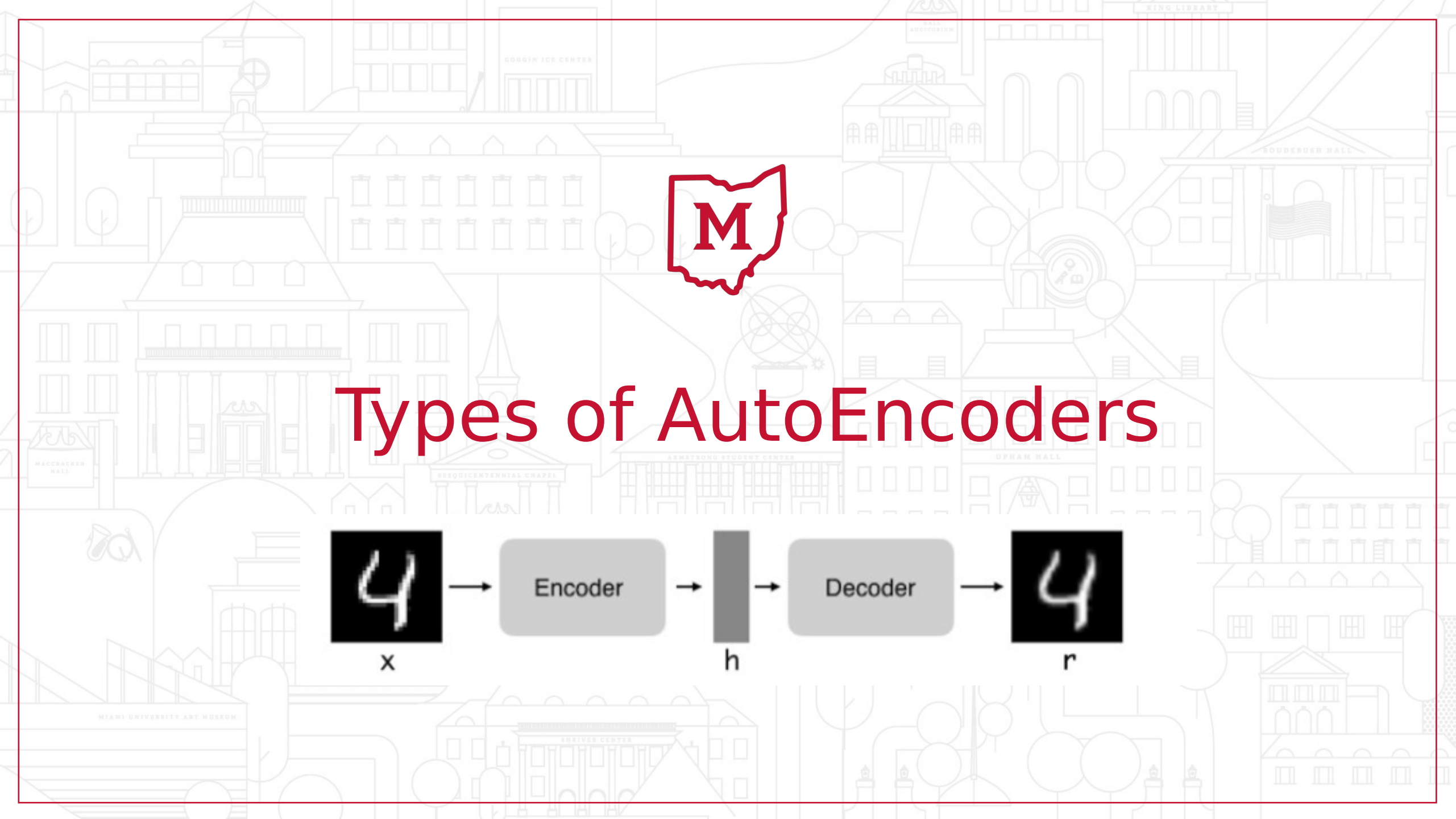
- Bottleneck: Latent space has lower dimension than inpu
- Latent space also captures key features
- Autoencoders outperform PCA on large or nonlinear data



Reconstruction Error

- Difference between the original input $\boldsymbol{x}$ and reconstructed $\tilde{\boldsymbol{x}}$.
- The observation MSE reconstruction error is:
- $\mathbf{MSE_i} = \frac{1}{k}\|x - \tilde{x}\|_2^2$

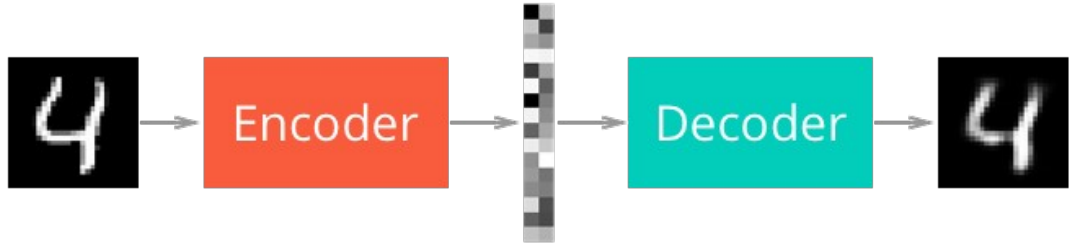


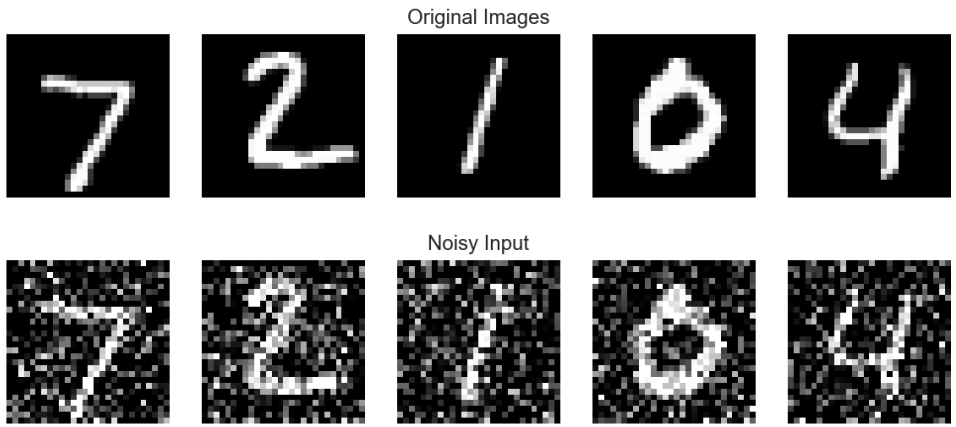


Types of AutoEncoders

Types of AutoEncoders

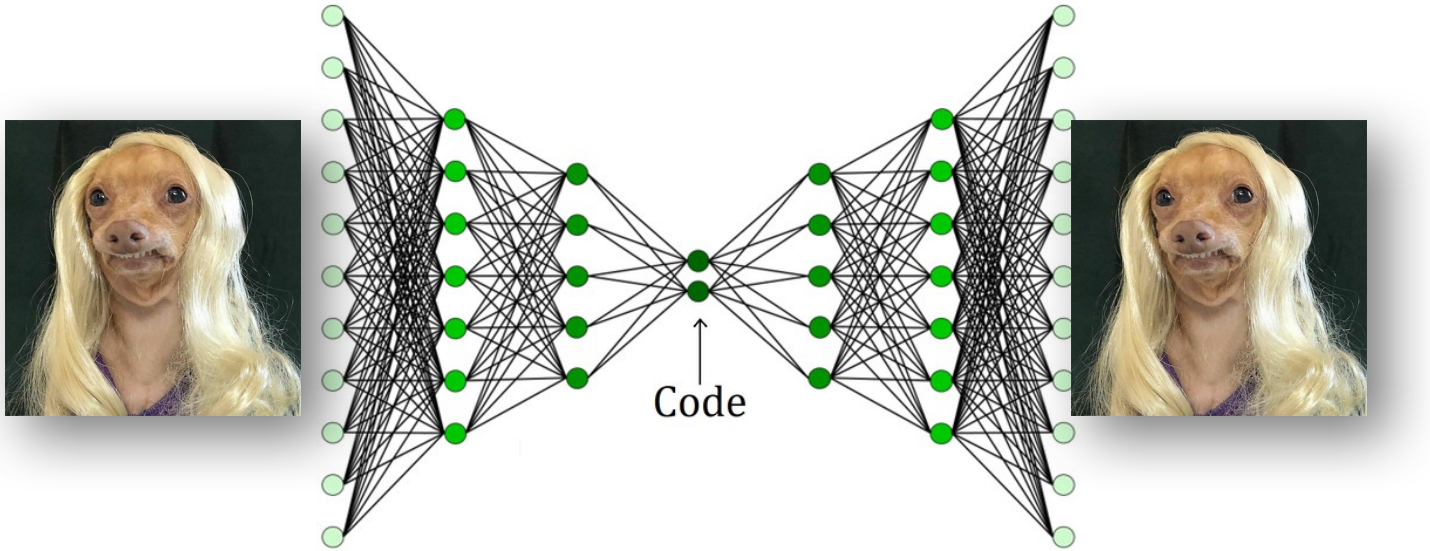
- There are many types of autoencoders, we will focus on:
  - Undercomplete autoencoders
  - Sparse autoencoders
  - Denoising autoencoders





Deep AutoEncoders (Undercomplete)

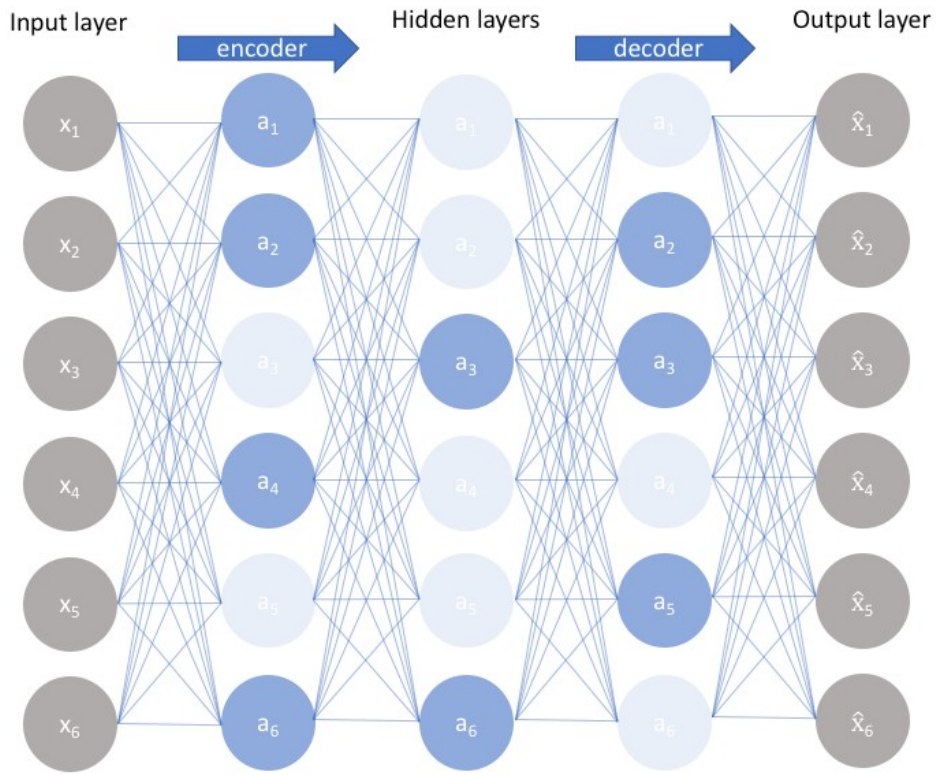
- Take an input and predicts it as output.
- Using multiple hidden layers we call these deep autoencoders.



Python

Sparse AutoEncoders

- Autoencoder with a sparsity penalty
- Common sparsity penalties include L1 and L2 norms



Python

Denoising AutoEncoders

- The goal is to remove noise from a signal
- We add noise to the images but predict images without the noise.

Python