Carnegie Mellon University

AutoEncoders and VAE

IDL Recitation 12

Puru Samal, Gabriel Zencha, Romerik Lokossou
Slide credits :Dareen Alharthi, Harshith Kumar, Yuzhou Wang

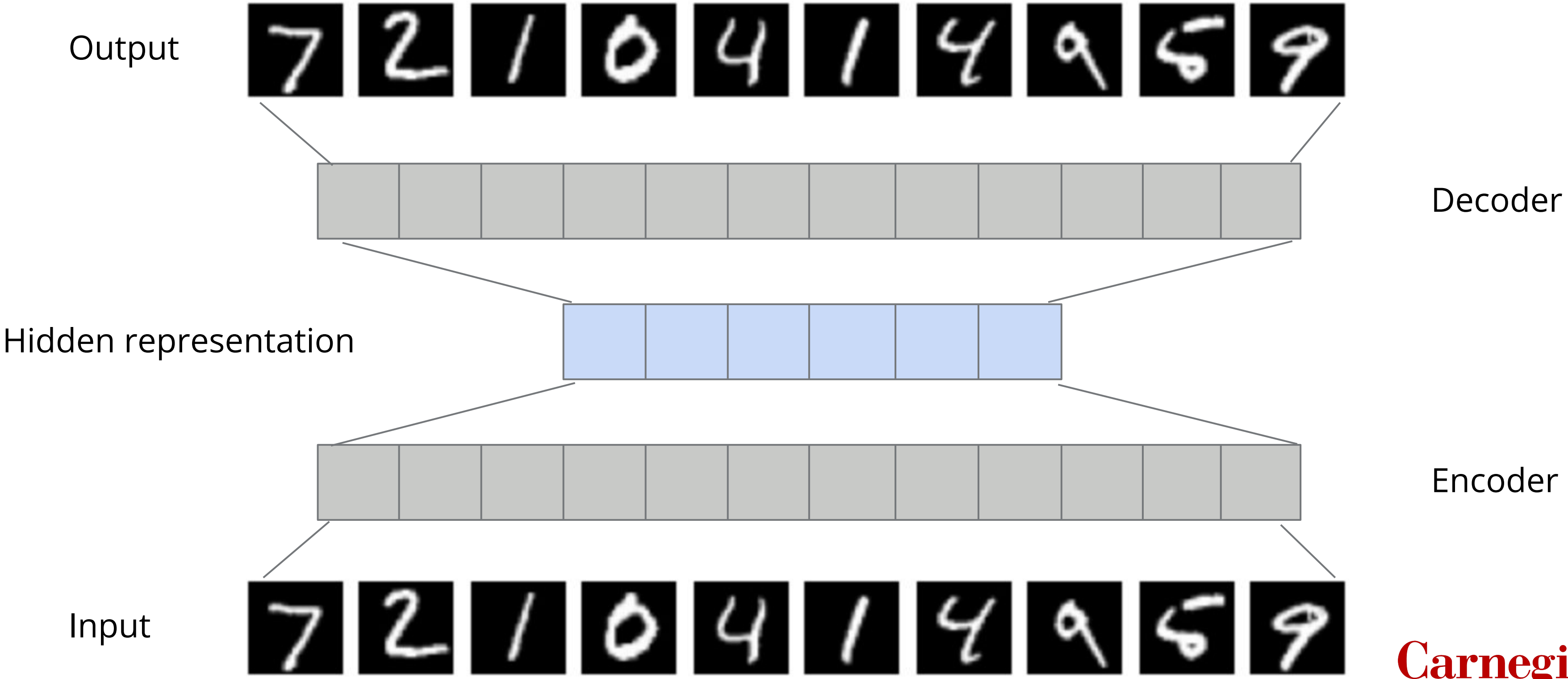
Autoencoder

- A type of neural network that is trained to learn a compressed representation of input data and generate output matching the input
- Unsupervised learning

Encoder: Takes in the input data and maps it to a lower-dimensional latent space representation

Decoder: Takes the encoded representation and generates a reconstructed output that closely matches the original input

Autoencoder



Implementation and training

- A sequence of linear or convolutional layers
- Decoder is often a mirror of Encoder

Training objective: Minimize the difference between inputs and outputs
Loss function: MSE Loss, L1 Loss, Binary CrossEntropy Loss (for [0,1] data)

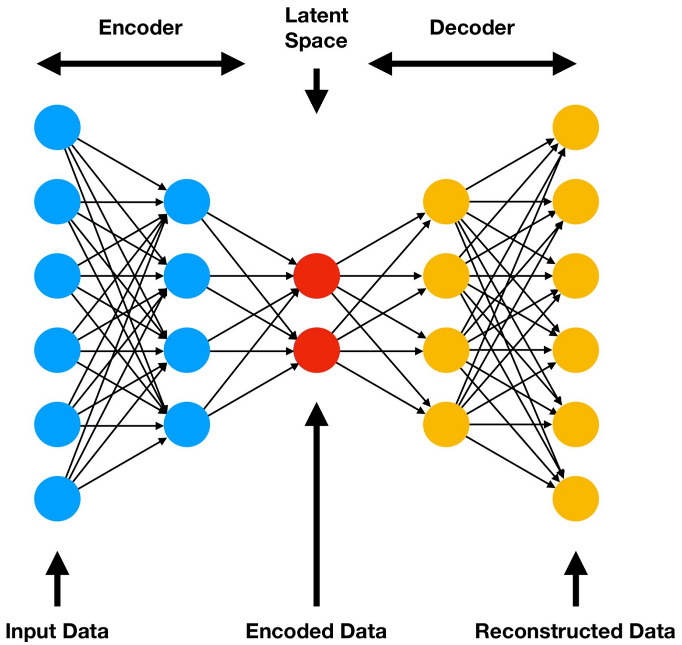
Applications

- Feature Extraction/Dimension Reduction
- Data Denoising
    - Input data with noise, calculate loss on clean data and output
- Anomaly Detection
    - Detect inputs with high loss values

Drawbacks of the AE

Limitations of the Auto Encoder :

1. Limited control over the latent space representation
2. Overfitting
3. Generation not Possible



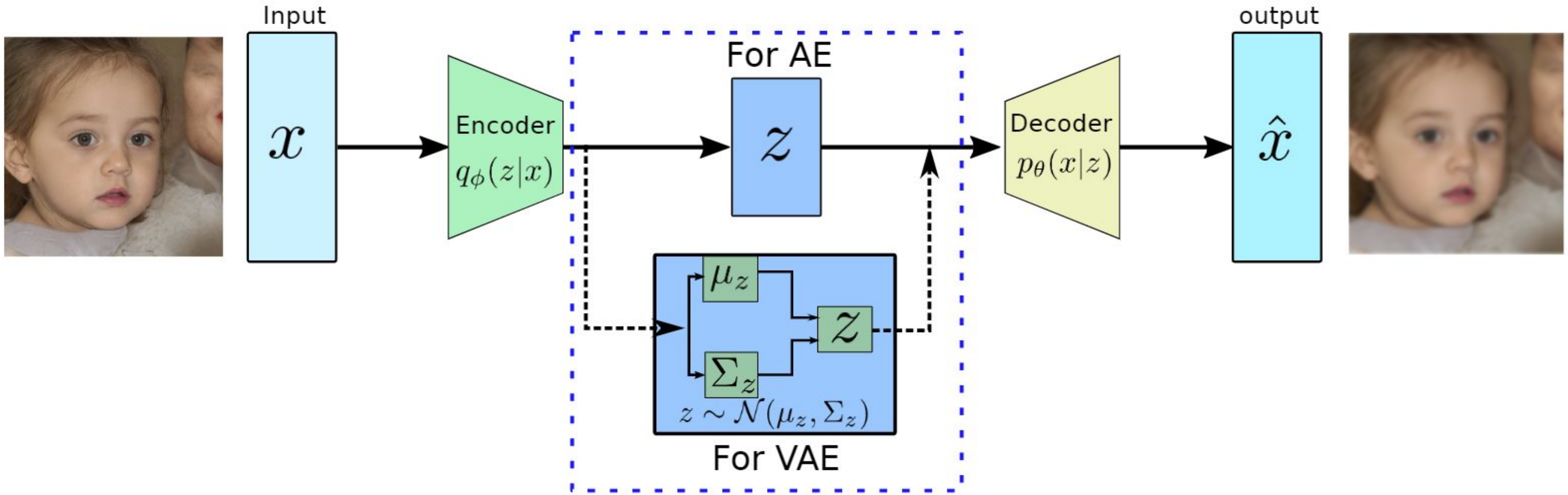
Variational Autoencoders

An Auto Encoder but better

Aim : Learn the data distribution of the input data

Why?
1. Data and Image Generation
2. Data Compression
3. Anomaly Detection

Variational Autoencoders



Variational Autoencoders

Loss Function

$$\mathcal{L}(\boldsymbol{\theta}, \phi; \mathbf{x}^{(i)}) = -D_{KL}(q_{\phi}(\mathbf{z}|\mathbf{x}^{(i)})||p_{\boldsymbol{\theta}}(\mathbf{z})) + \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x}^{(i)})}\left[\log p_{\boldsymbol{\theta}}(\mathbf{x}^{(i)}|\mathbf{z})\right]$$

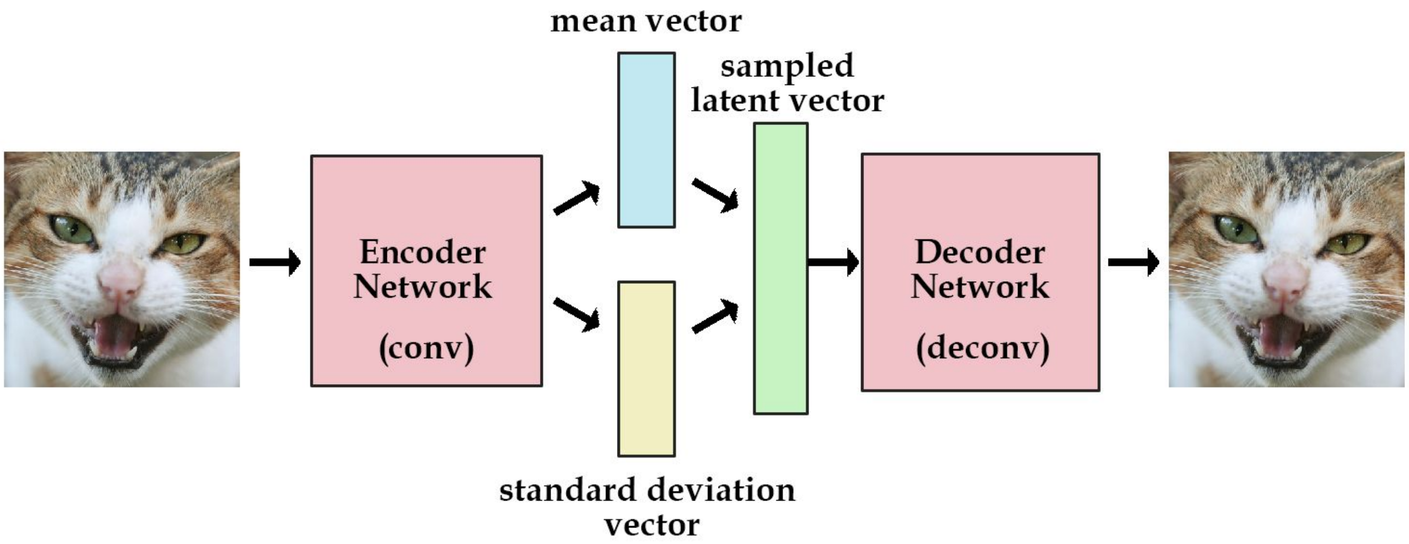
First Term : Forcing latent vector distribution to be Normal(0, 1)

Second Term : Reconstruction Loss

Variational Autoencoders

- VAEs are probabilistic models to learn data distribution
- Map inputs to a probability distribution
- Objective is to maximize the evidence lower bound (ELBO) : NLL of Data
- Allows learning of a structured latent space representation

Variational Autoencoders

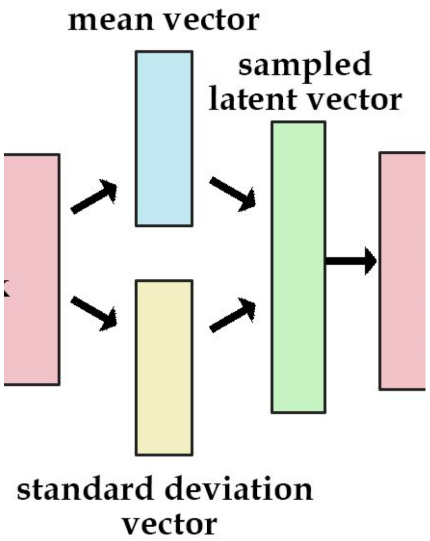


Variational Autoencoders

Re-parametrization Trick

Sampled Vector : Mean Vector + Std deviation vector x Noise

Sampled Vector : $\boldsymbol{\mu}$ + $\boldsymbol{\sigma}$ x $\boldsymbol{\epsilon}$

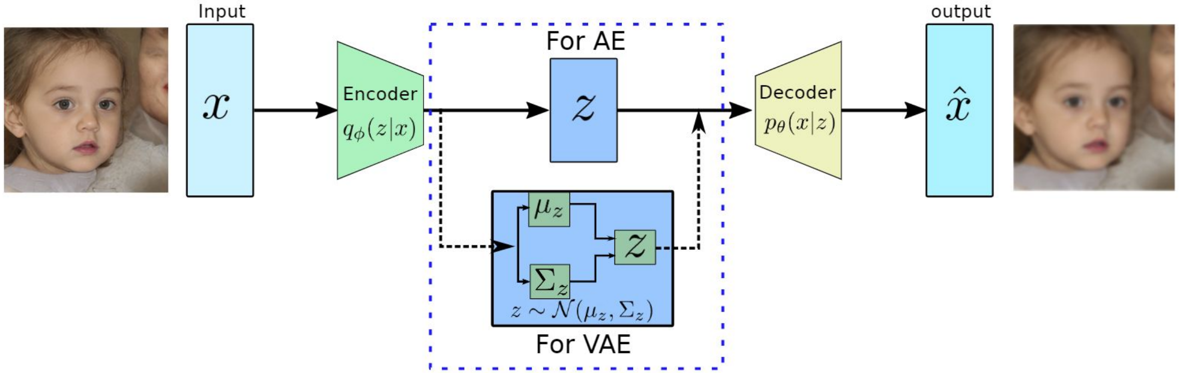


Variational Autoencoders

Issues with AE :

1. Latent Space is unstructured and difficult to interpret

SOLUTION : Map it to a Normal distribution



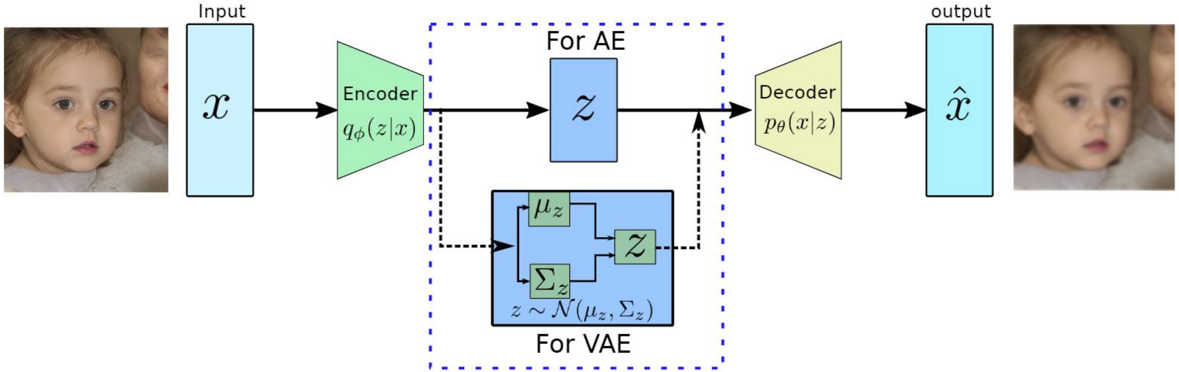
Carnegie Mellon University

Variational Autoencoders

Issues with AE :

2. Overfitting

SOLUTION : Normal Distribution mapping restricts freedom



Variational Autoencoders

Issues with AE :

3. Data Generation?

SOLUTION : Give a normal sampled vector as an input to the decoder