Johns Hopkins
Whiting School of Engineering

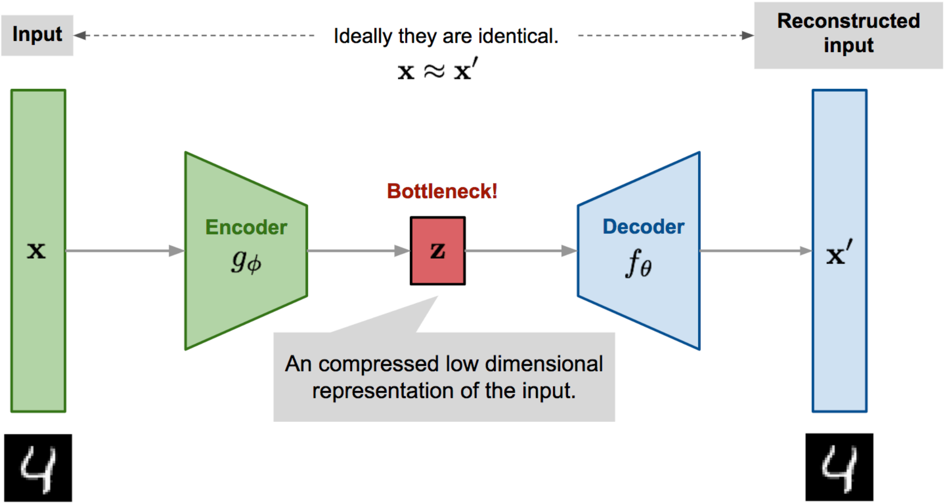
685.621 Algorithms for Data Science

Variational Autoencoders (VAEs)

Introduction to VAEs

VAEs are generative models that learn probabilistic representation of data rather than deterministic mapping to generate inputs.

- **Basic Structure**
  - Encoder: Compresses input into a latent distribution q(z|x).
  - Latent Space: probabilistic space where data variations exist.
  - Decoder: Reconstructs data from latent variables z.



Medium, 2021

VAE Objective

VAEs use reconstruction loss and regularization in order to construct their loss function, known as maximizing ELBO (Evidence Lower Bound).

$$\log p(x) \geq \mathbb{E}_{q(z|x)}[\log p(x|z)] - KL(q(z|x)||p(z))$$

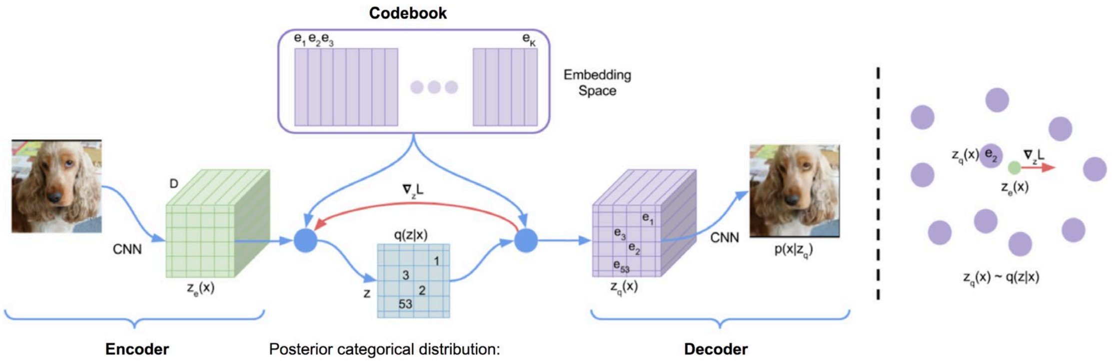
- The first term is Reconstruction loss which measures how well the Decoder reconstructs an ouput as close to the input. The second term is KL Divergence Loss which forces latent space to be smooth and continuous. This balance of accurate reconstruction and well structured latent space optimizes the VAE to reconstruct accurately and regularize the latent space.

Step 1: Define the Encoder and Decoder

The Encoder compresses input data into lower-dimensional latent space. The Decoder reconstructs sampled latent data into an image.

- **Example of Encoder**
  - Begin with a 784D vector and reduce it into 2D latent space. Pass it through a fully connected layer to reduce the image data. Add a final layer to output the mean of the latent space and log-variance. Apply reparameterization trick to obtain the latent compressed representation of the input image (z vector).
- **Example of Decoder**
  - Begin with the sampled z vector and map it from latent space to a fully connected layer. Use ReLU activation to introduce non-linearity and map to final fully connected layer back to 784D vector.

Step 2: Training/Evaluating the VAE



Lil'Log, 2018

1. **Encoding**: Pass an input image through the encoder and compress into latent representation. Instead of a single value for each latent variable, the encoder learns mu and logvar of distribution. This is the compression stage.
2. **Sample Latent Space**: Use reparameterization trick to sample from the latent representation to ensure smooth learning. The output of this is the latent vector, z.
3. **Decoding**: Use latent vector, z, to reconstruct image through the use of a fully connected neural network. This is the reconstruction stage.

Johns Hopkins Whiting School of Engineering

Step 3: Advantages

1. **Easy training**: SGD is used to optimize ELBO objective function making it easier to train compared to GANs.
2. **Robust Modeling Framework**: Enables generation of diverse and coherent text with a solid probabilistic model which can facilitate beneficial tasks like unsupervised learning from interpretable latent representations.

Step 4: Challenges

1. **Blurry Images**: Compared to GANs, VAE’s tend to produce less sharp images
2. **Less expressive**: Tend to struggle with complex distributions given overly smoothed latent space.
3. **Training Instability**: ELBO objective function requires careful tuning to balance

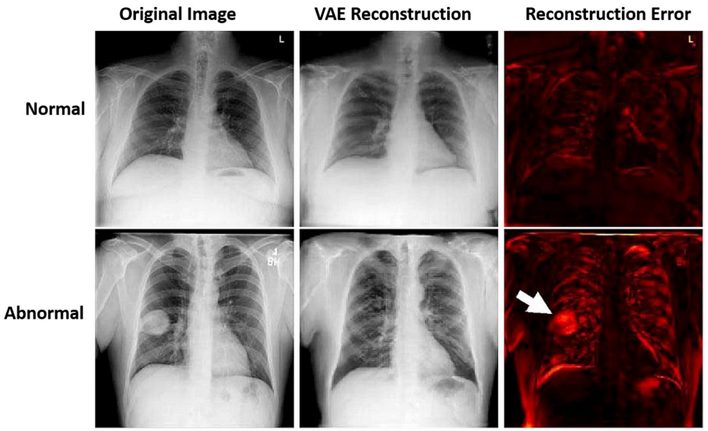
Original image VAE Re-construct

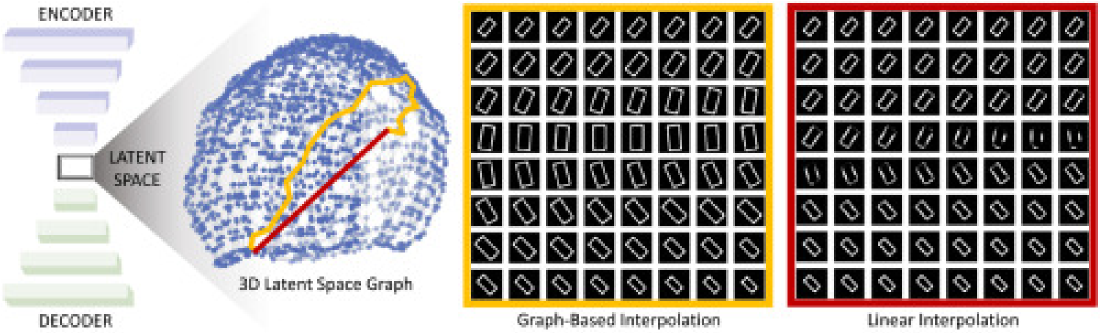
Step 5: Real World Applications

1. **Medical Imaging**: Realistic MRI scans with super-resolution and de-noising

2. **Latent Space Interpolation**: Smooth transitions between images

3. **Anomaly Detection**: bank fraud detection or medical abnormal structure detection





JOHNS HOPKINS
WHITING SCHOOL of ENGINEERING