CS-583: Deep Learning
AutoEncoder

Abdul Rafae Khan
Department of Computer Science
Stevens Institute of Technology
akhan4@stevens.edu

November 21, 2023

Autoencoder

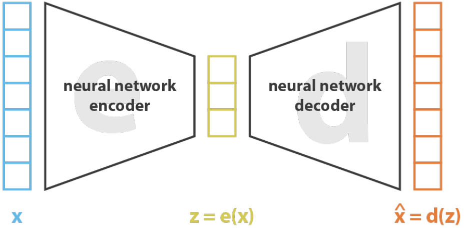
Motivation:
- Deep neural networks require lot of data
- Sometimes not very much labeled data, but plenty of unlabeled data (text, images, videos)
- Humans rarely get direct supervision; can learn from raw sensory information?

Autoencoder

Analogy:

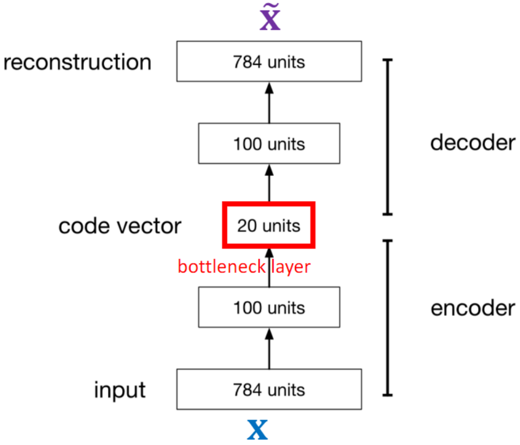
A A A A B B B B B $\Rightarrow$ 4 A's, 5 B's $\Rightarrow$ A A A A B B B B B

Key idea: If we can compress a data point and still reconstruct it, then we have learned something generally useful



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

Autoencoder



- An autoencoder is a neural net taking an input x and reconstruct x
- For dim reduction, we need a bottleneck layer whose output shape is much smaller than the input
- Loss function: $\sum_{j=1}^{x} |x_j - \tilde{x}_j|_2^2$

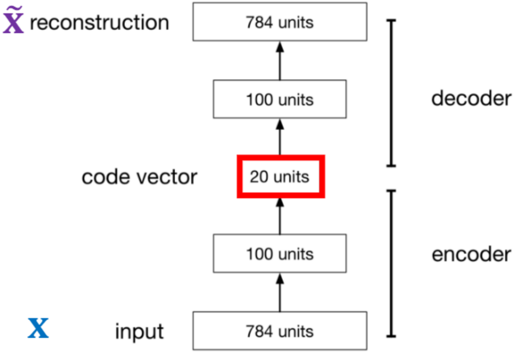
Autoencoder

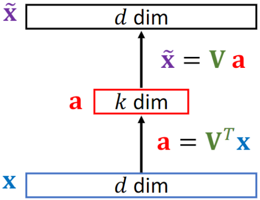
Autoencoder is nonlinear

$\tilde{x} = decoder(encoder(x))$

PCA is linear

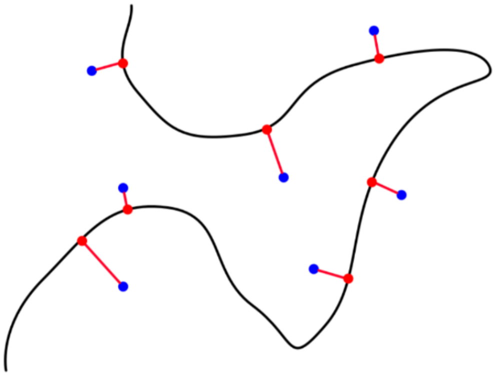
$\tilde{x} = VV^T x$



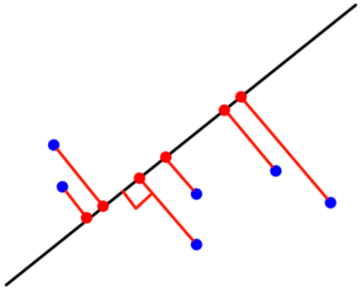


Autoencoder

Autoencoder projects data onto nonlinear manifold

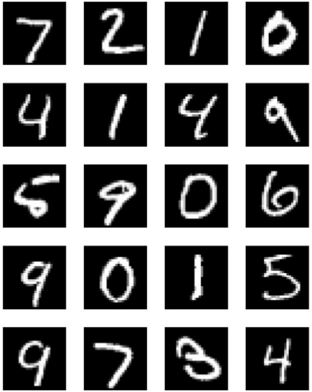


PCA projects data onto a subspace

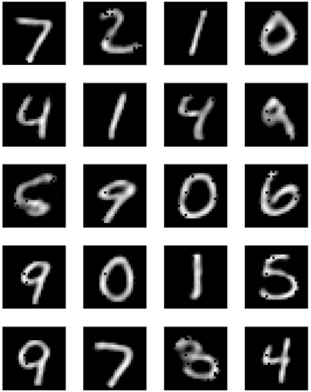


Autoencoder

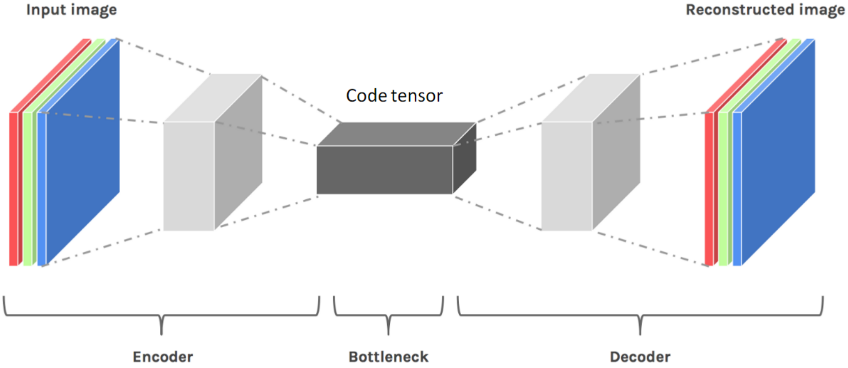
Input



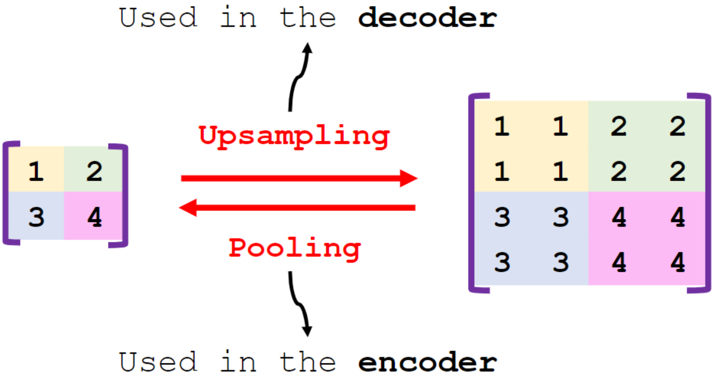
Reconstructed



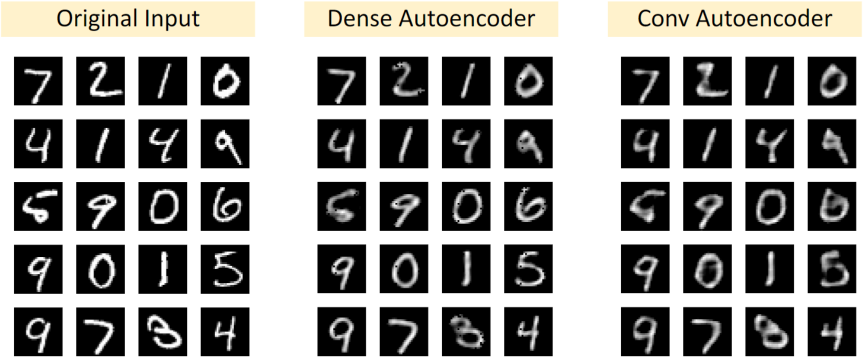
Convolutional Autoencoder



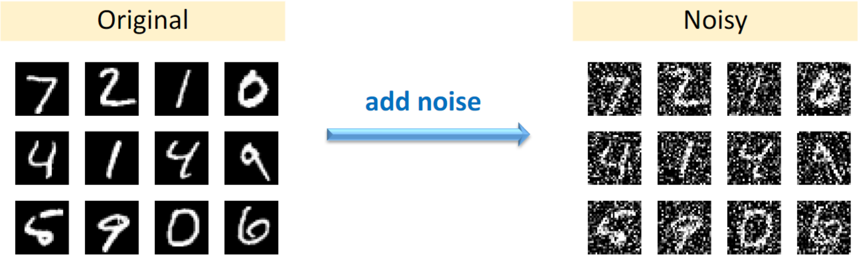
Convolutional Autoencoder



Convolutional Autoencoder



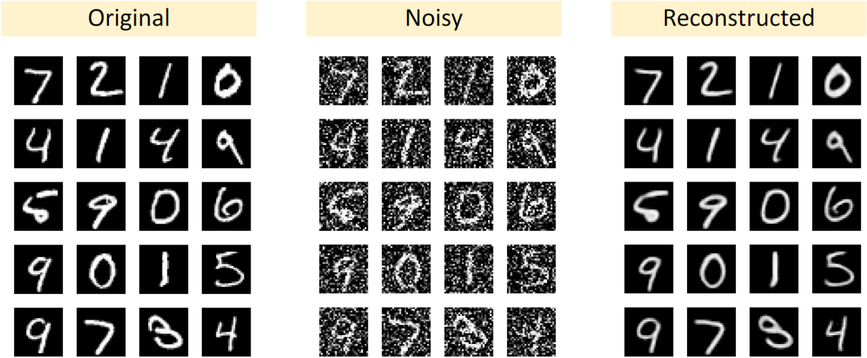
Denoising Autoencoder



Used as targets

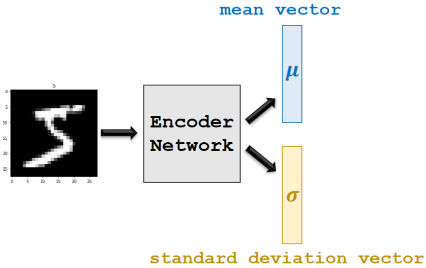
Used as inputs

Denoising Autoencoder



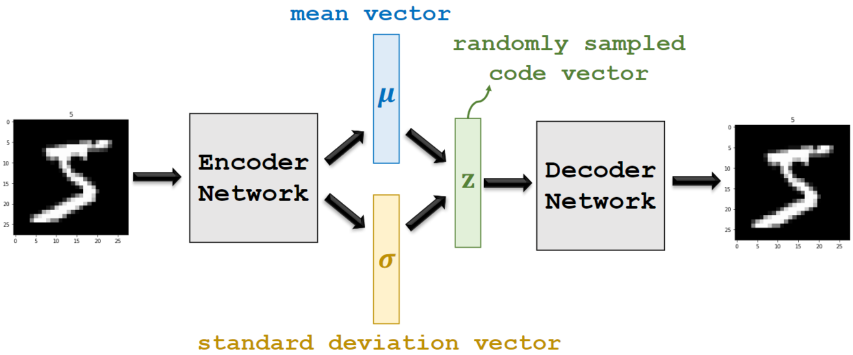
Variational Autoencoder

Instead of a single vector output by the encoder, VAE outputs a distribution

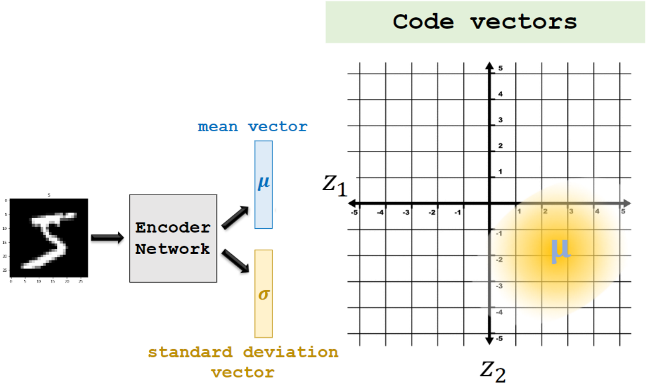


Variational Autoencoder

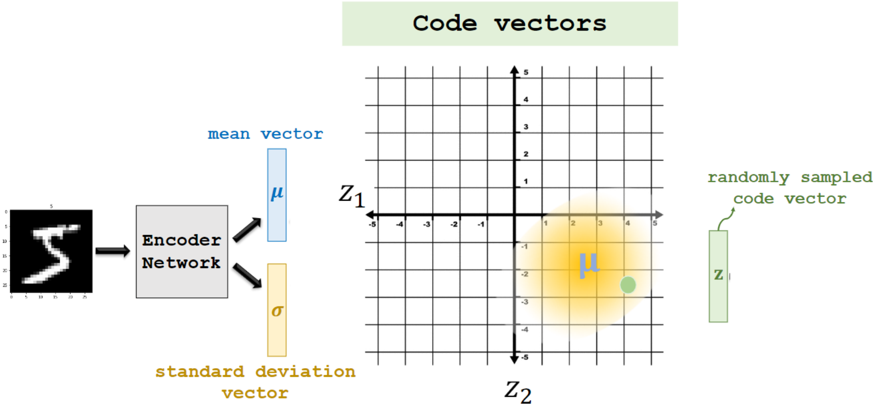
Sample a code vector from the distribution and reconstruct the original



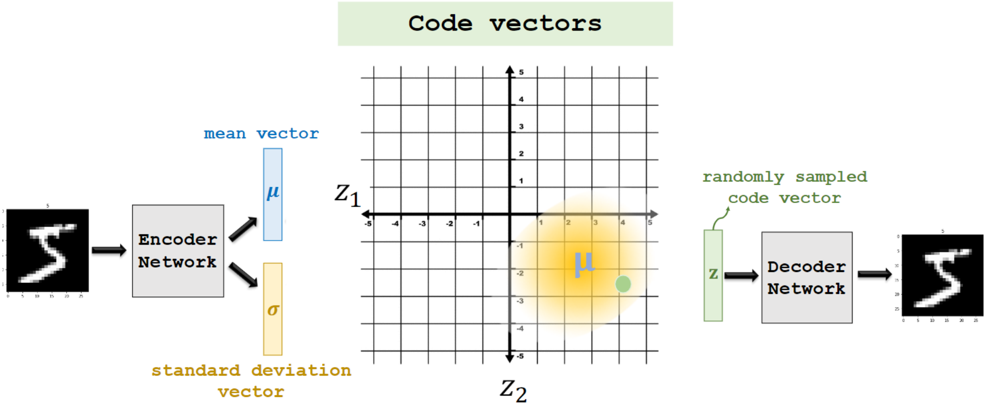
Variational Autoencoder



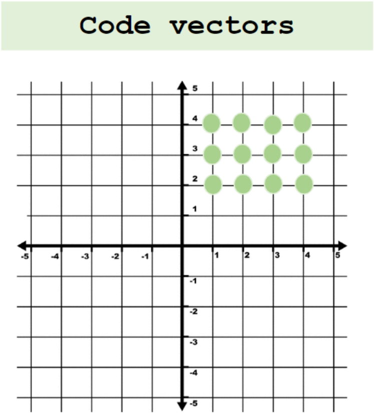
Variational Autoencoder



Variational Autoencoder



Variational Autoencoder



- Get a set of code vectors from the grid:

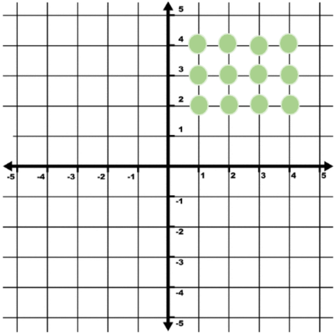
$$z_1, z_2, z_3, z_4, z_5, \cdots$$

For every code vector z_i, map it to an image using the decoder:

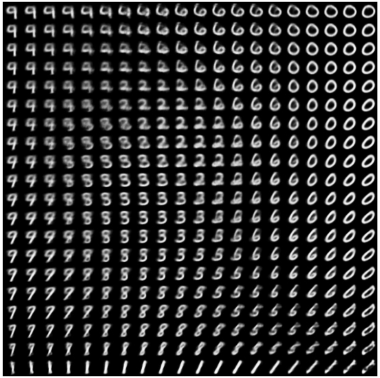
$$image_i = decoder(z_i)$$

Variational Autoencoder

Code vector



Images



Variational Autoencoder

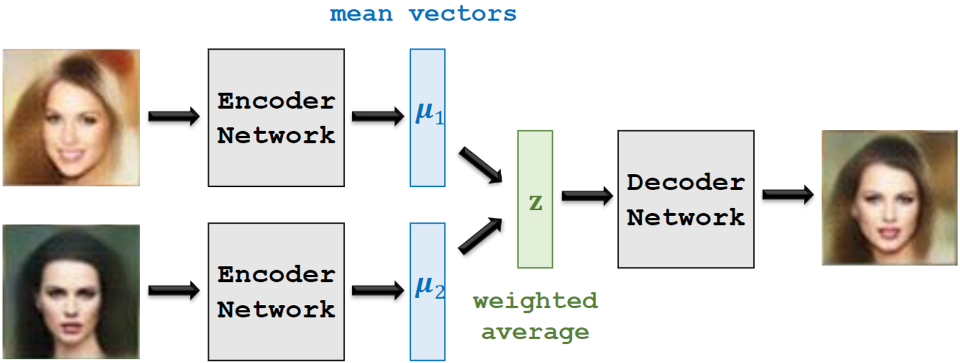
- Function f is continuous if a small change in z (input) results in a small change in $f(z)$ function value)
- The decoder network is trained to be (almost) continuous
- If the code vectors z and z' are similar, then the images

$$Decoder(z) \quad \text{and} \quad Decoder(z')$$

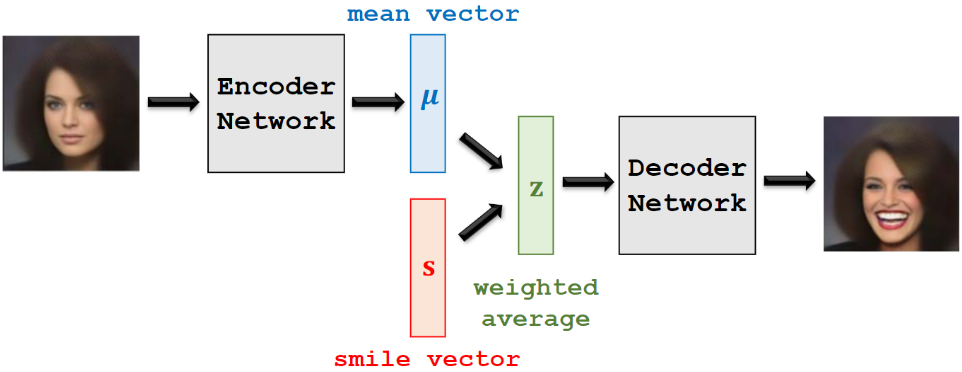
are similar as well

Variational Autoencoder

Variational Autoencoder

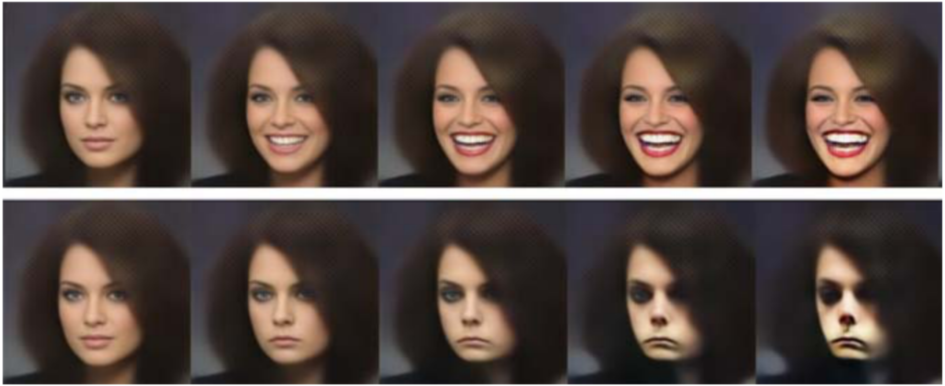


Variational Autoencoder



Variational Autoencoder

$\boldsymbol{\mu}$ $\quad$ $\boldsymbol{\mu} + \mathbf{s}$ $\quad$ $\boldsymbol{\mu} + 2\mathbf{s}$ $\quad$ $\boldsymbol{\mu} + 3\mathbf{s}$ $\quad$ $\boldsymbol{\mu} + 4\mathbf{s}$



$\boldsymbol{\mu}$ $\quad$ $\boldsymbol{\mu} - \mathbf{s}$ $\quad$ $\boldsymbol{\mu} - 2\mathbf{s}$ $\quad$ $\boldsymbol{\mu} - 3\mathbf{s}$ $\quad$ $\boldsymbol{\mu} - 4\mathbf{s}$

Summary

- Autoencoder: generalizations of (linear) PCA
- Autoencoder $=$ Encoder $+$ Decoder
- Original/noisy images $\rightarrow$ original images
- Application: Dimensionality reduction or Denoising

- VAE $=$ AE $+$ probability tricks
- The decoder network behaves like a continuous function
- Application: Edit images (average faces/add smile etc.)

References

Stuart Russell and Xiaodong Song (2021)
CS 188 — Introduction to Artificial Intelligence
University of California, Berkeley

Chelsea Finn and Nima Anari (2021)
CS221 — Artificial Intelligence: Principles and Techniques
Stanford University

The End