In the name of God

Auto Encoders

Mehrnaz Faraz

Faculty of Electrical Engineering

K. N. Toosi University of Technology

Milad Abbasi

Faculty of Electrical Engineering

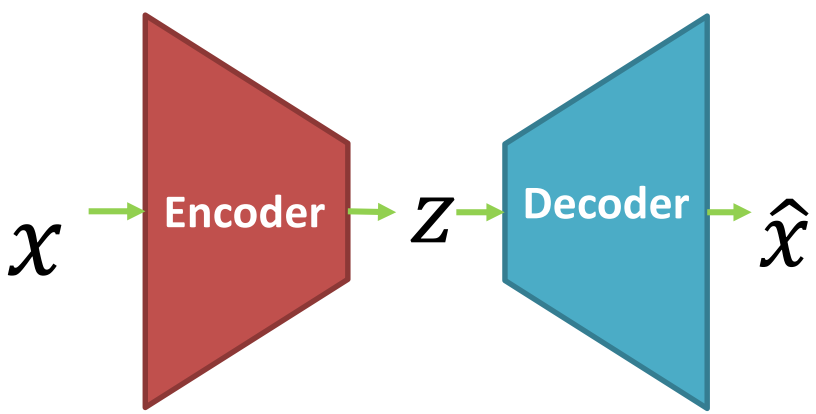
Sharif University of Technology

Auto Encoders

- An unsupervised deep learning algorithm ⟶ Unlabeled data
- Are artificial neural networks
- Useful for dimensionality reduction and clustering

$\hat{x}$ is x's reconstruction

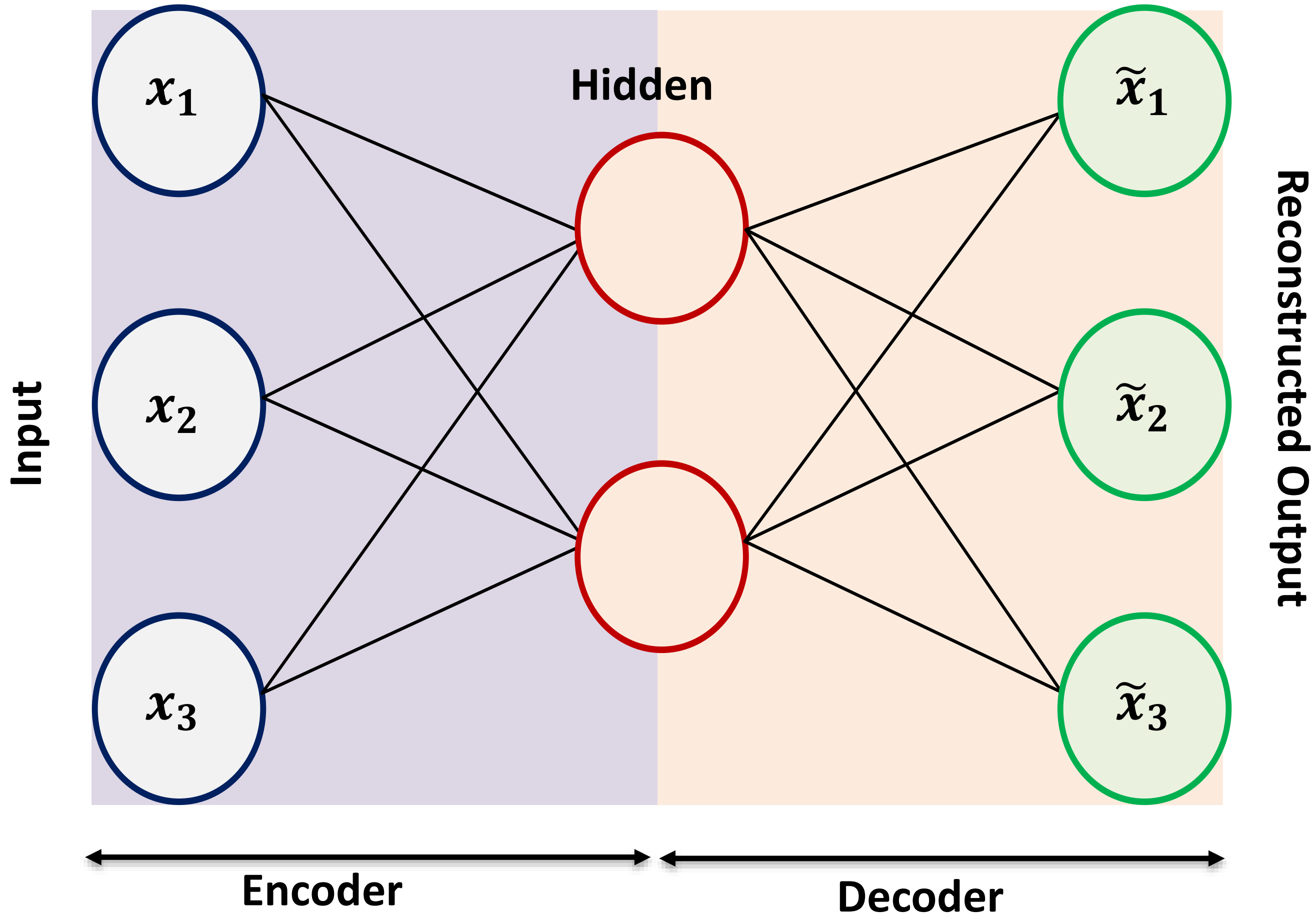
z is some latent representation or code and s is a non-linearity such as the sigmoid



$$z = s(wx + b)$$
$$\hat{x} = s(w'z + b')$$

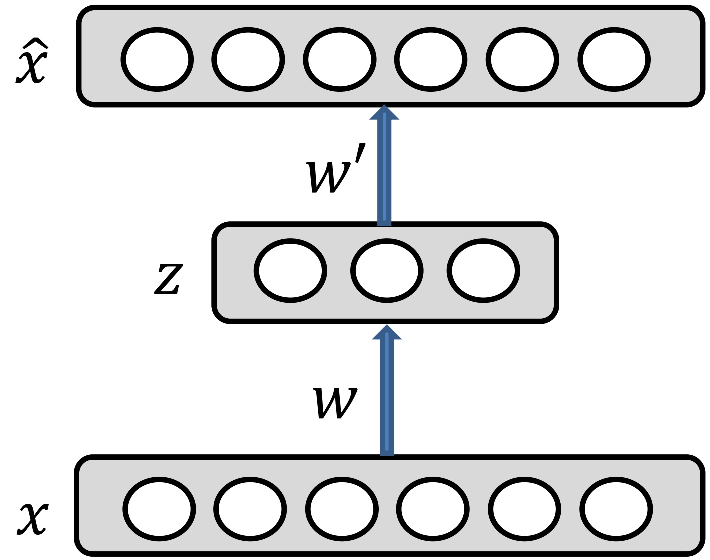
Auto Encoders

- Simple structure:



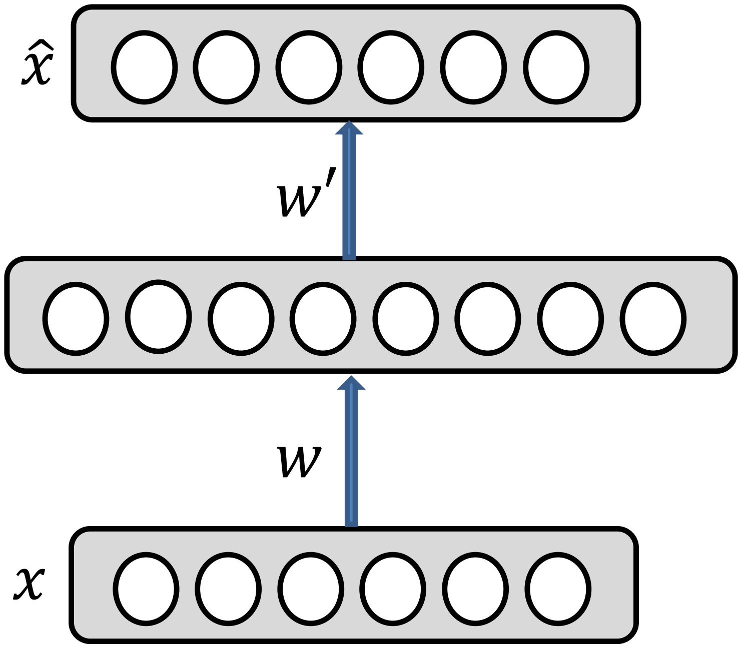
Undercomplete AE

- Hidden layer is **Undercomplete** if smaller than the input layer
  - Compresses the input
  - Hidden nodes will be Good features for the training



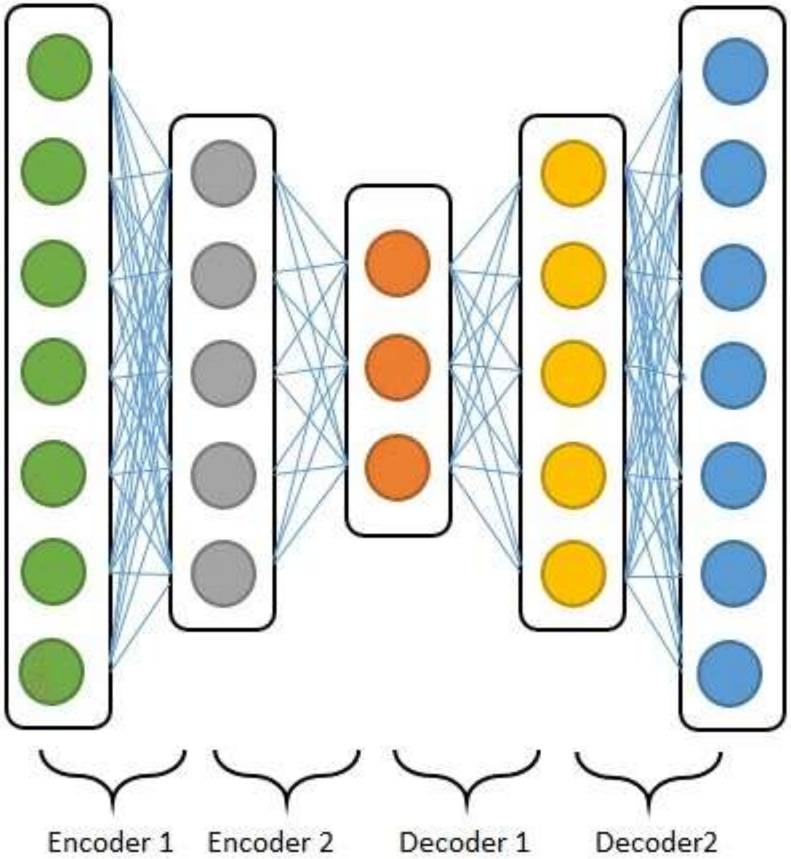
Overcomplete AE

- Hidden layer is **Overcomplete** if greater than the input layer
  - No compression in hidden layer.
  - Each hidden unit could copy a different input component.



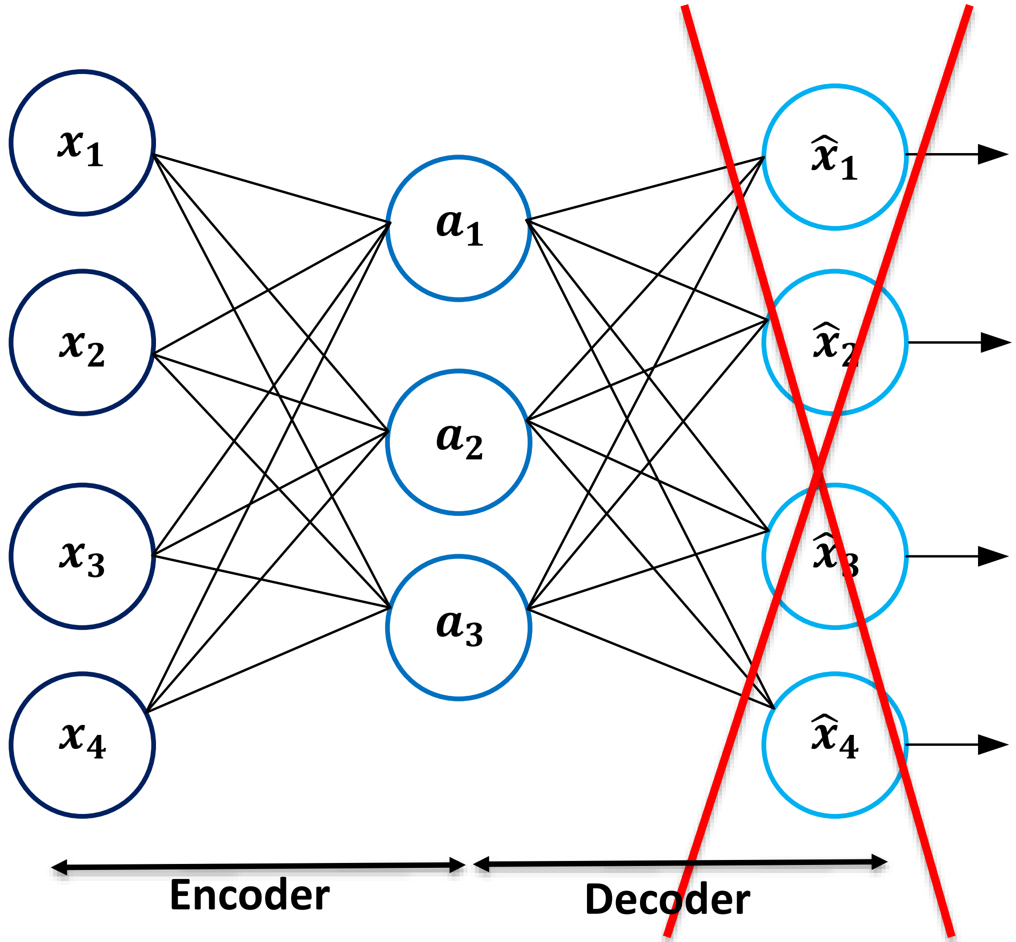
Deep Auto Encoders

- Deep Auto Encoders (DAE)
- Stacked Auto Encoders (SAE)



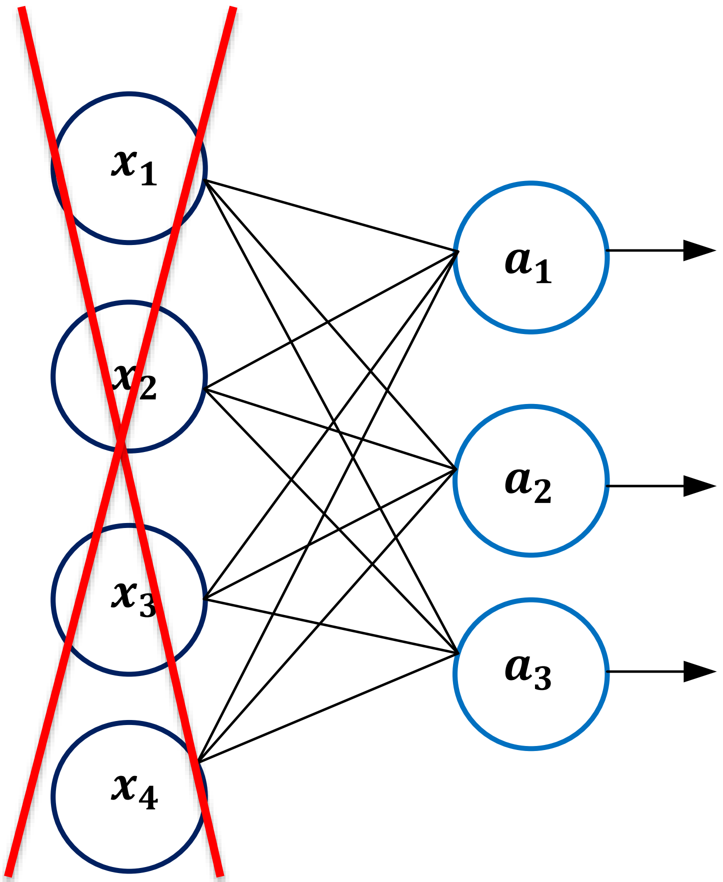
Training Deep Auto Encoder

- First layer:



Training Deep Auto Encoder

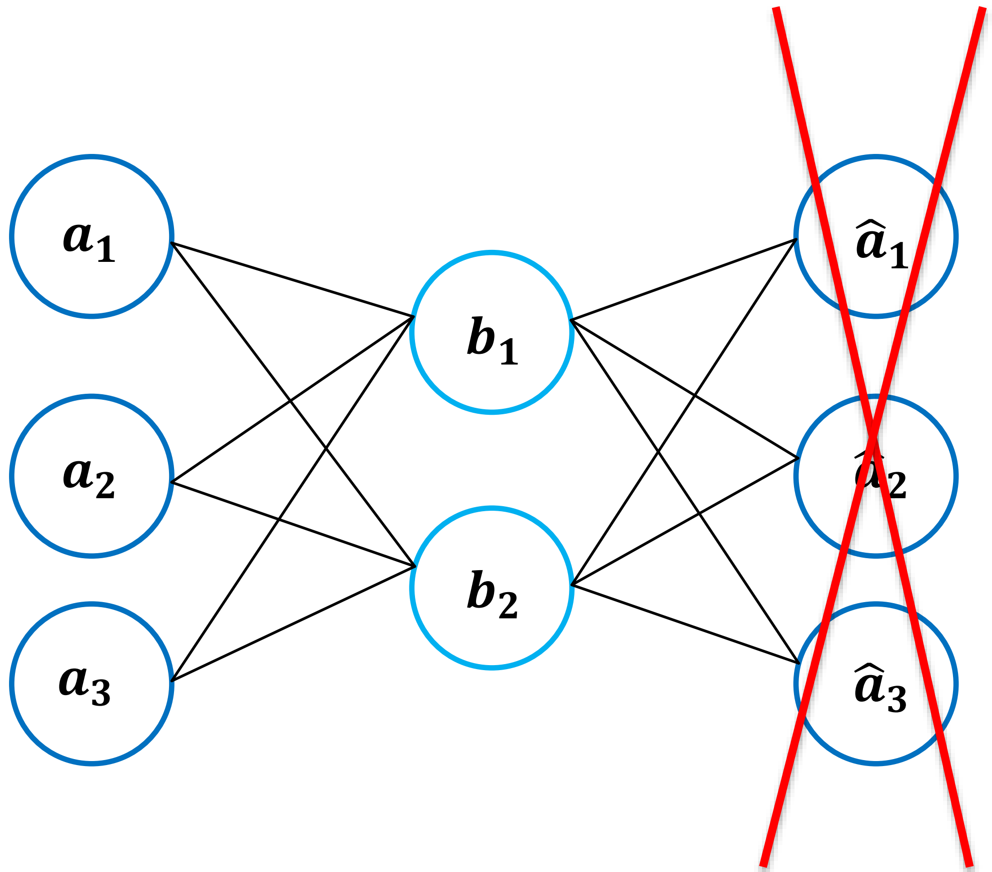
- Features of first layer:



$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$$

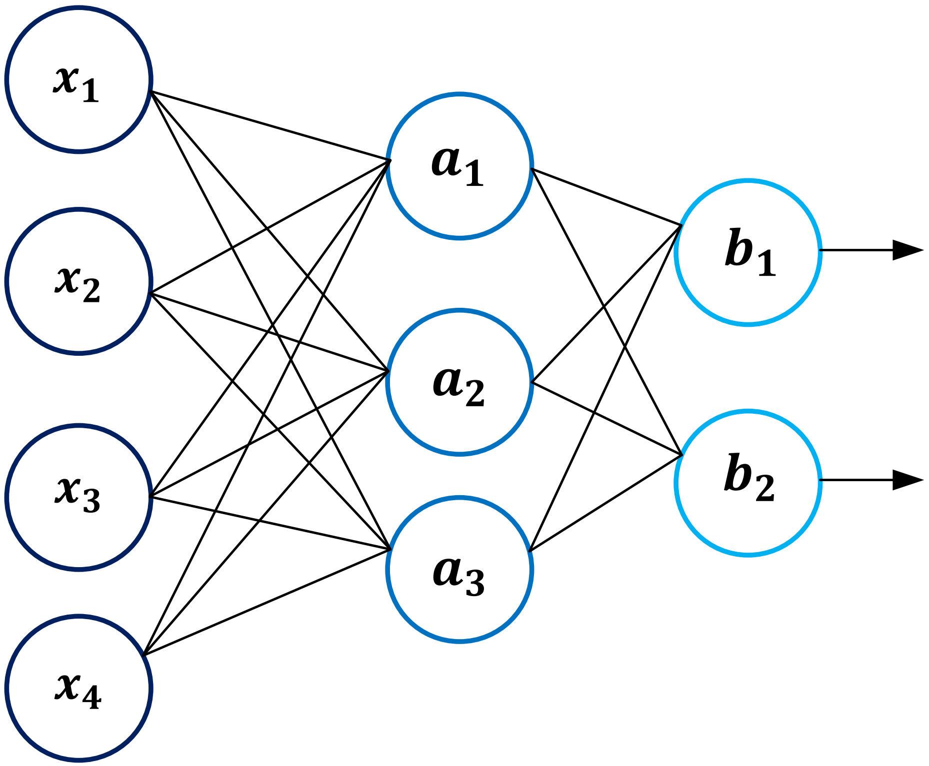
Training Deep Auto Encoder

- Second layer:



Training Deep Auto Encoder

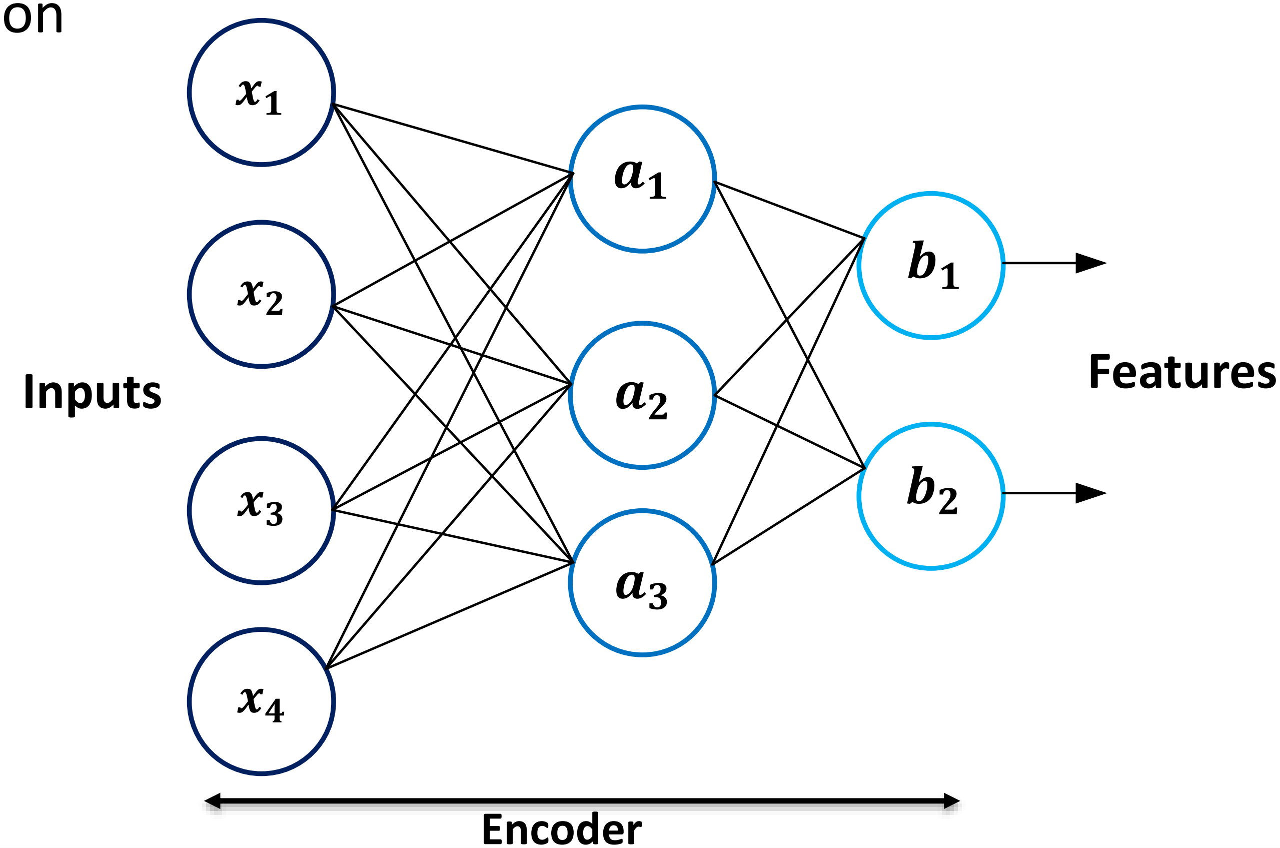
- Features of second layer:



$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

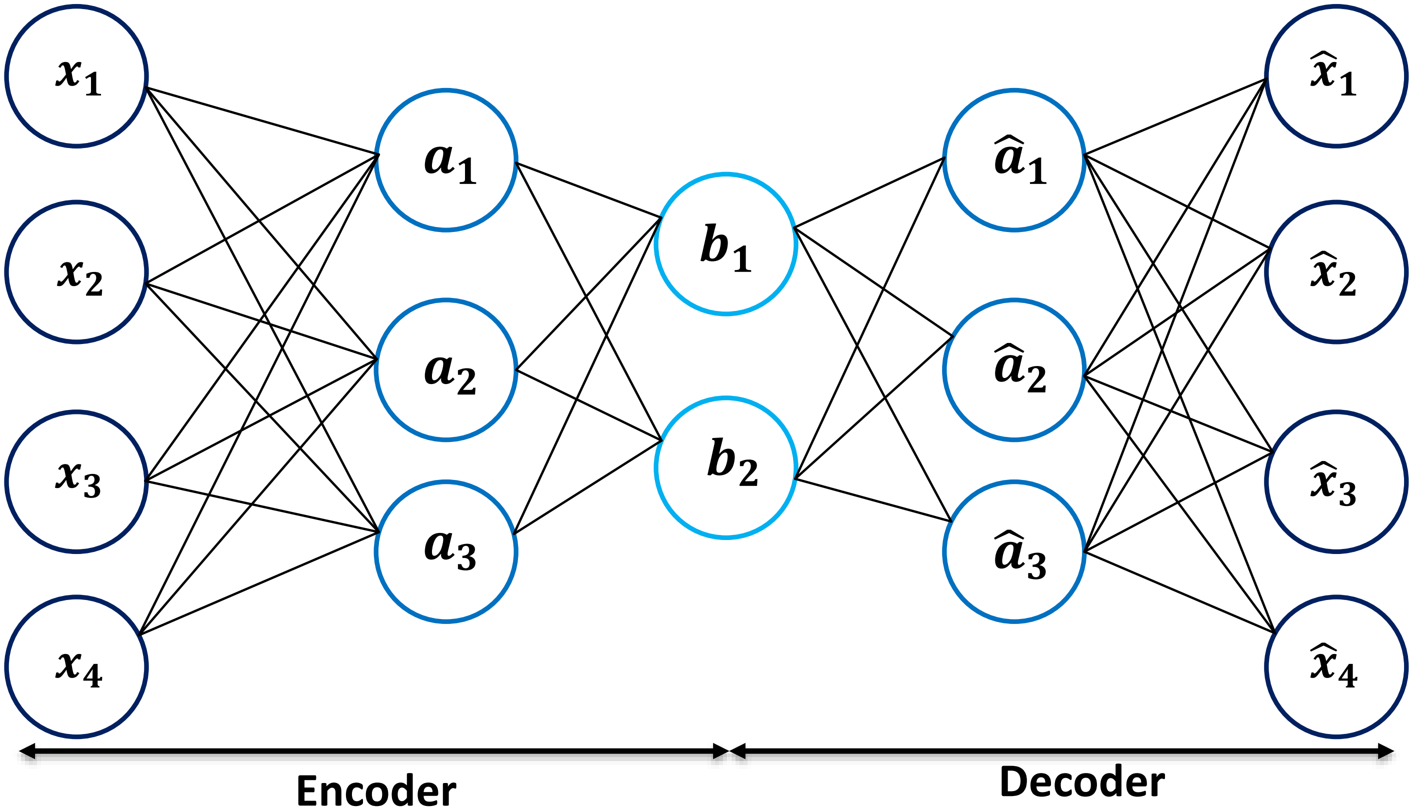
Using Deep Auto Encoder

- Feature extraction
- Dimensionality reduction
- Classification



Using Deep Auto Encoder

- Reconstruction



Using AE

- Denoising
- Data compression
- Unsupervised learning
- Manifold learning
- Generative model

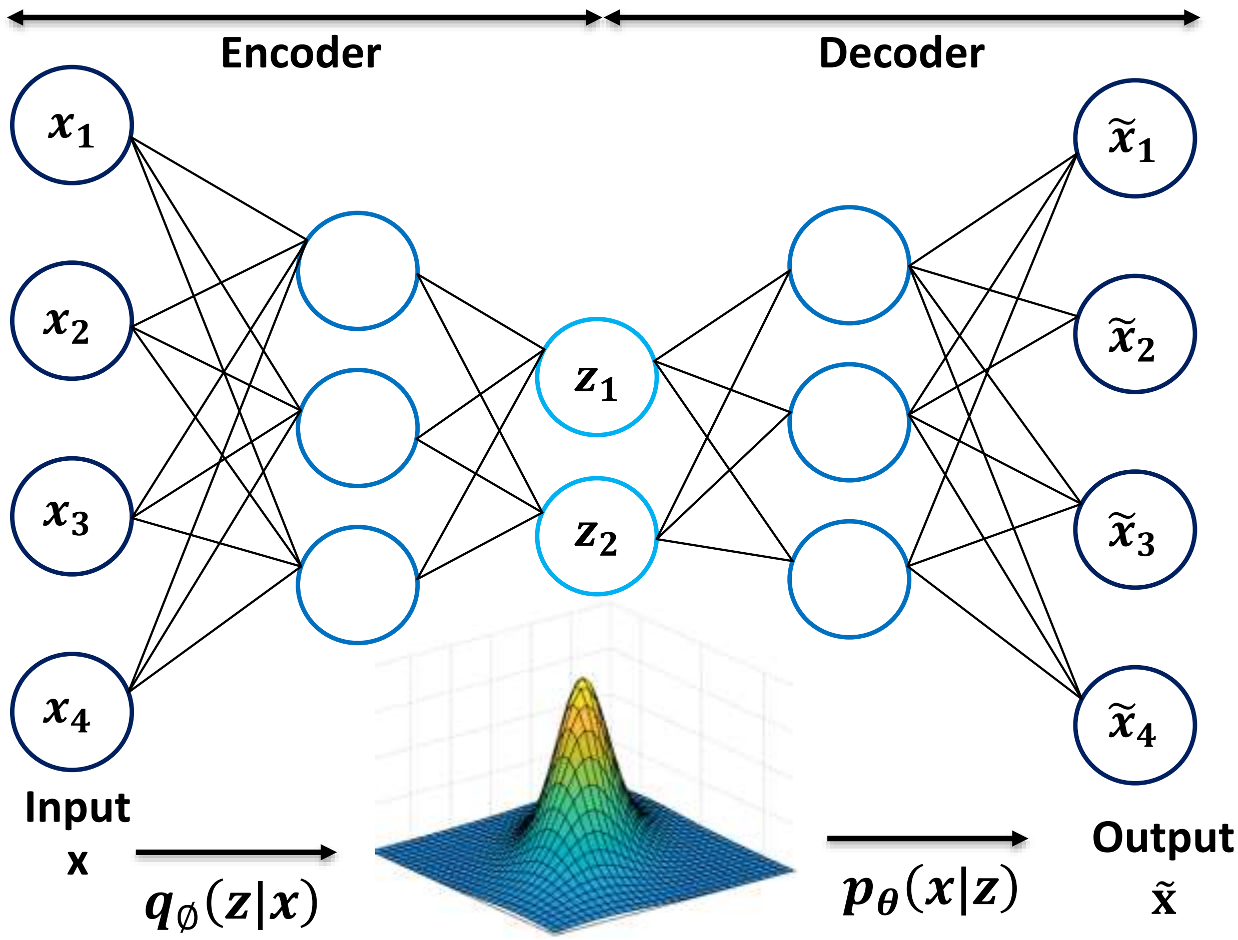
Types of Auto Encoder

- Stacked auto encoder (SAE)
- Denoising auto encoder (DAE)
- Sparse Auto Encoder (SAE)
- Contractive Auto Encoder (CAE)
- Convolutional Auto Encoder (CAE)
- Variational Auto Encoder (VAE)

Generative Models

- Given training data, generate new samples from same distribution
  - Variational Auto Encoder (VAE)
  - Generative Adversarial Network (GAN)

Variational Auto Encoder



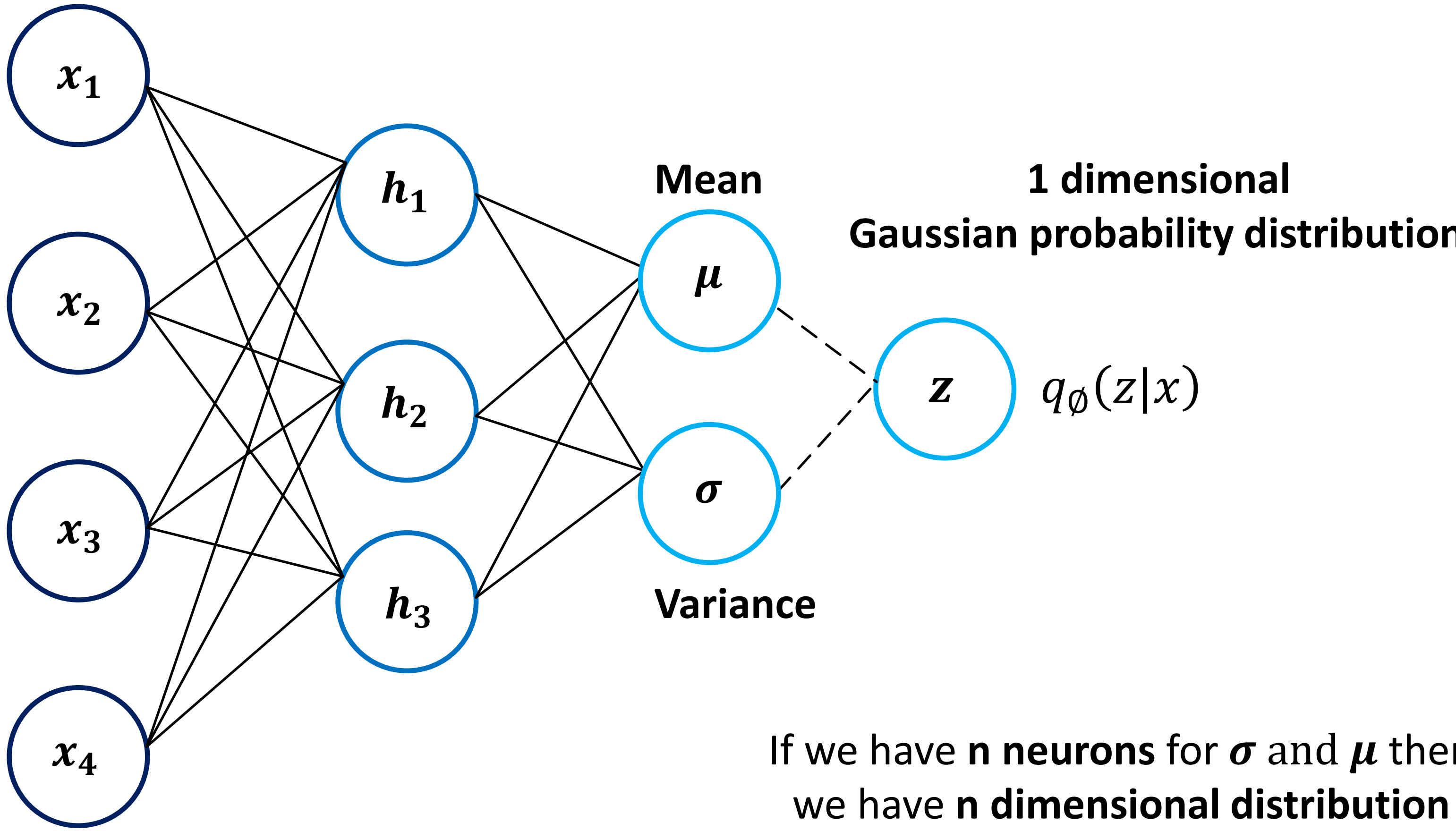
Variational Auto Encoder

- Use probabilistic encoding and decoding
  - Encoder: $q_{\emptyset}(z|x)$
  - Decoder: $p_{\theta}(x|z)$

- x: Unknown probability distribution
- z: Gaussian probability distribution

Training Variational Auto Encoder

- **Latent space:**



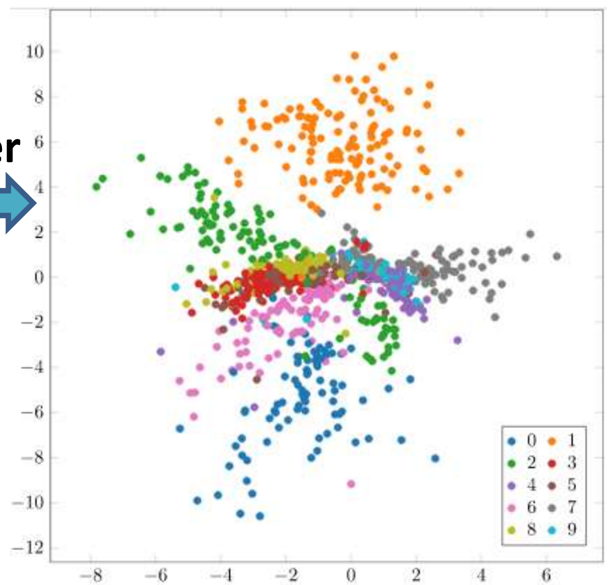
If we have **n neurons** for $\boldsymbol{\sigma}$ and $\boldsymbol{\mu}$ then we have **n dimensional distribution**

Training Variational Auto Encoder

- **Generating new data:**
  - Example: MNIST Database

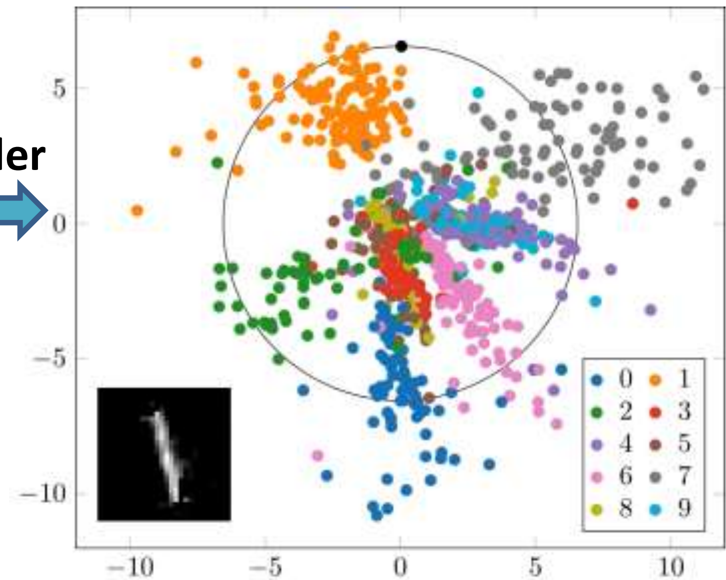
Encoder

x



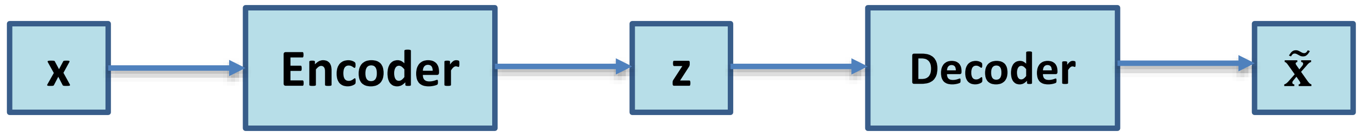
Latent space

Decoder

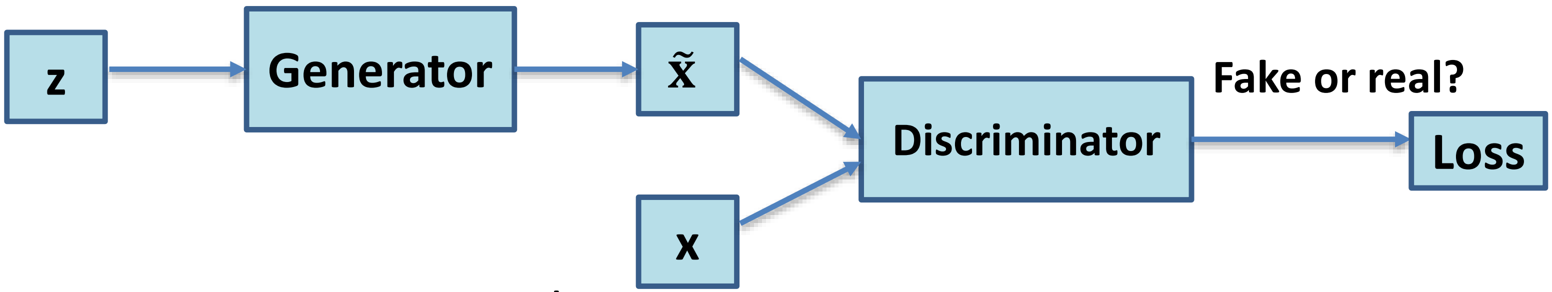


Generative Adversarial Network

- **VAE:**



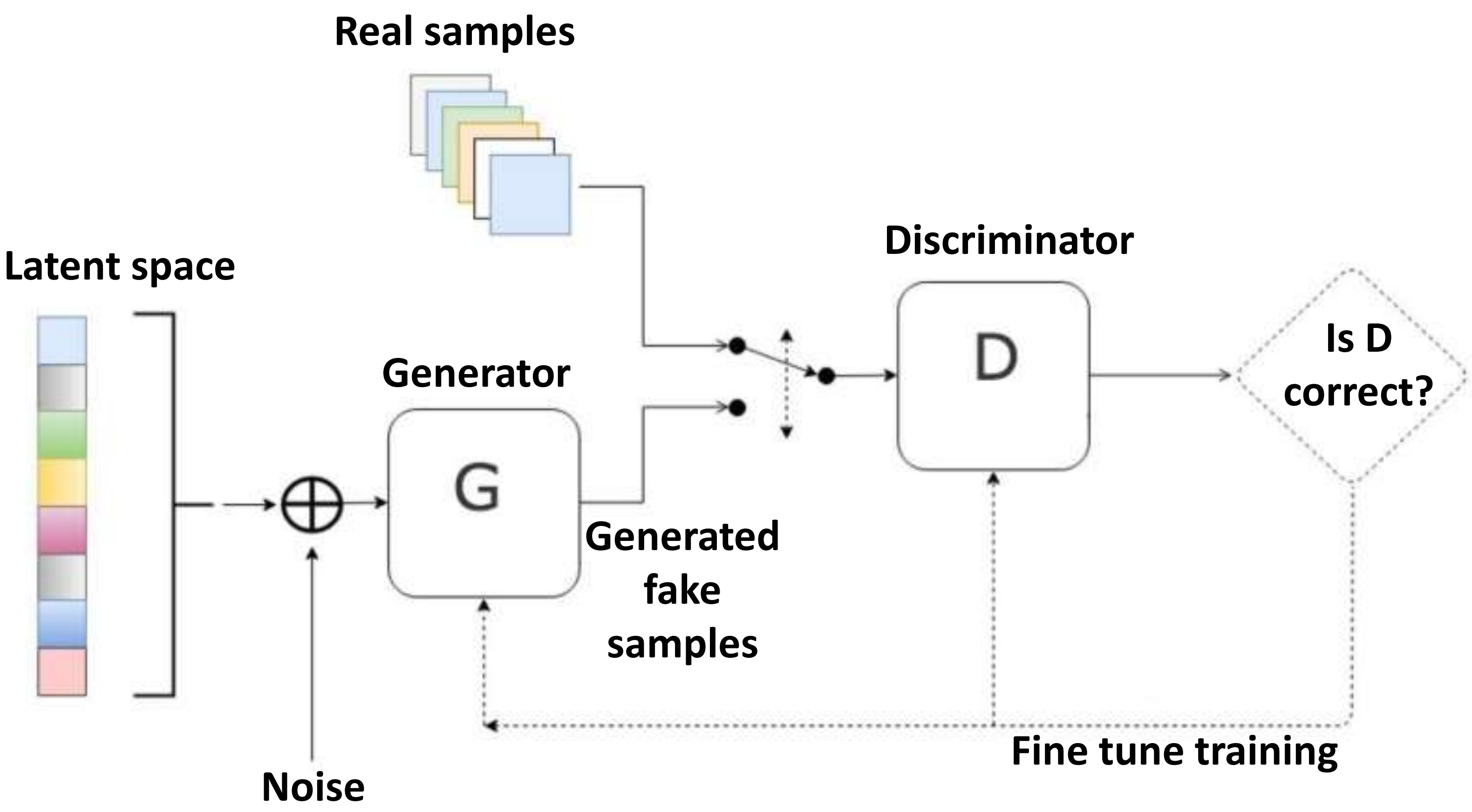
- **GAN:**



  - Can generate samples
  - Trained by competing each other
  - Use neural network
  - **Z** is some **random noise** (Gaussian/Uniform).
  - **Z** can be thought as the **latent representation** of the image.

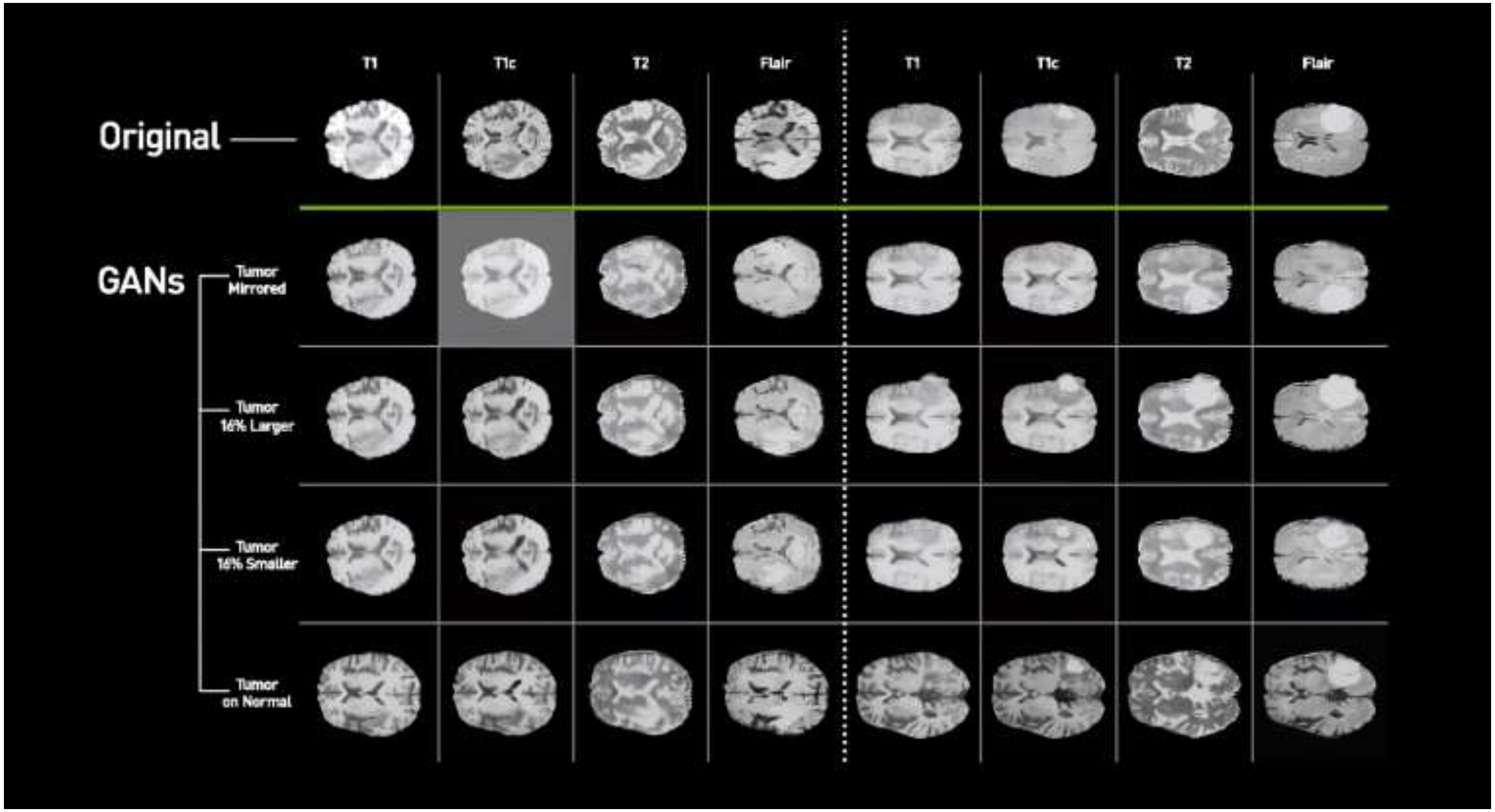
GAN’s Architecture

- **Overview:**



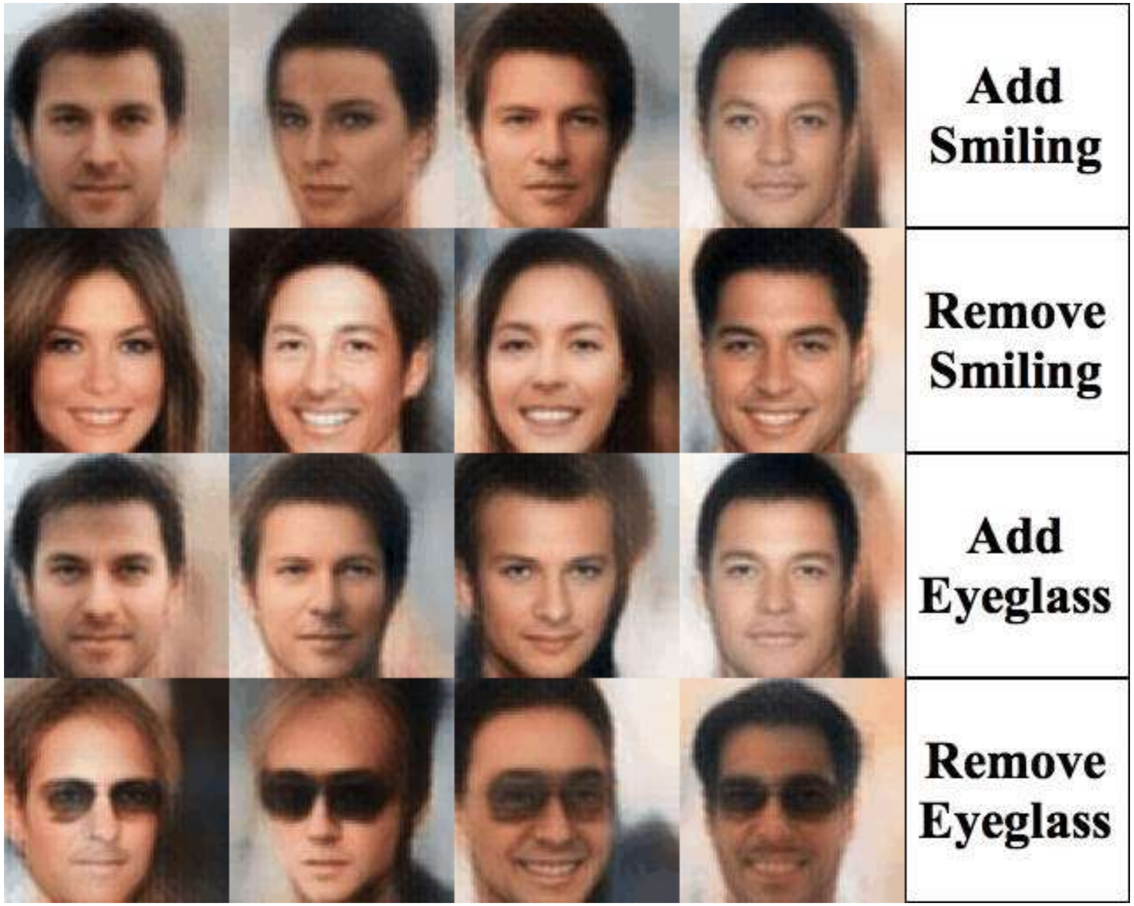
Using GAN

- Image generation:



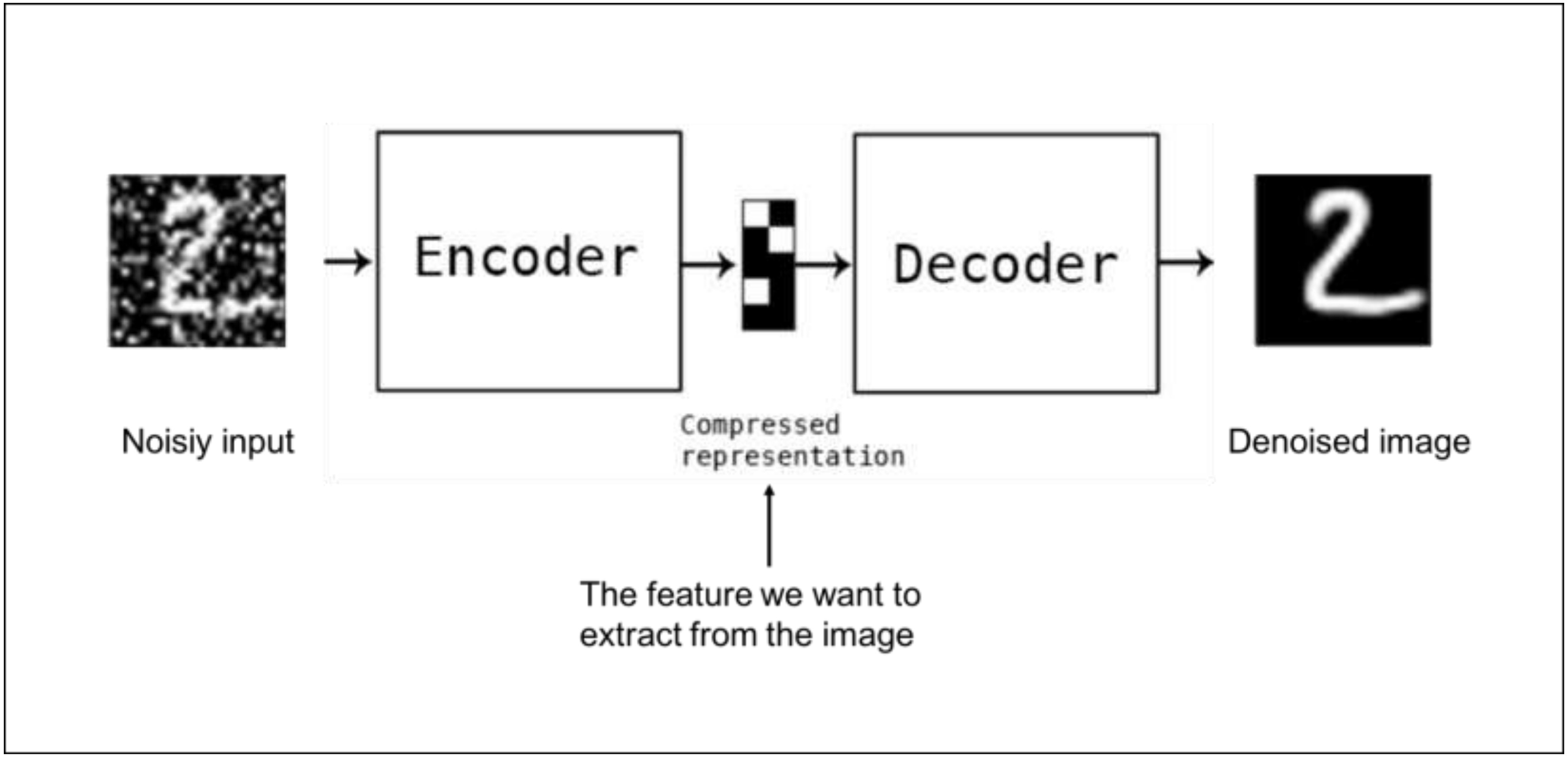
Using GAN

- Data manipulation:

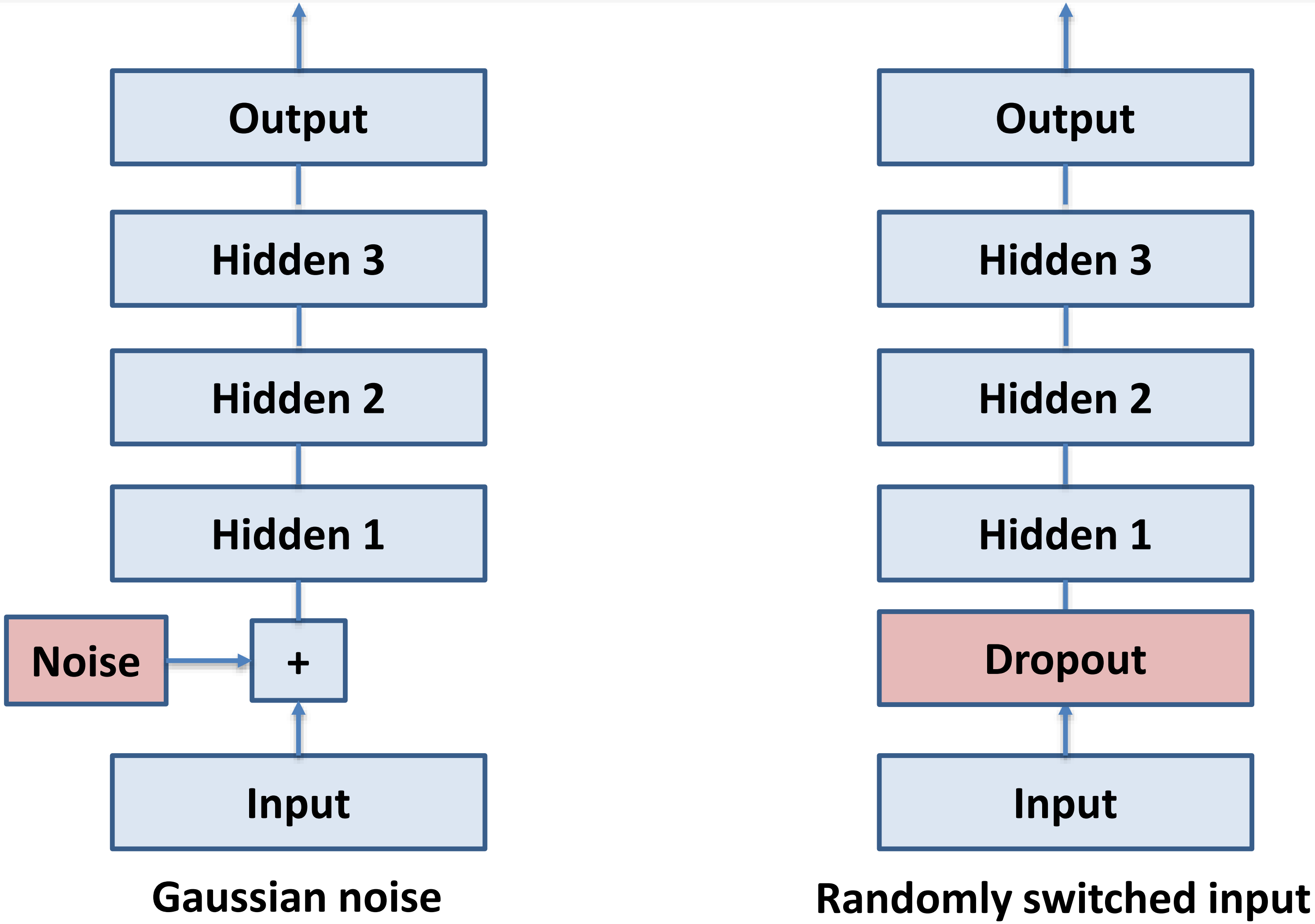


Denoising Auto Encoder

- Add noise to its input, and train it to recover this original.



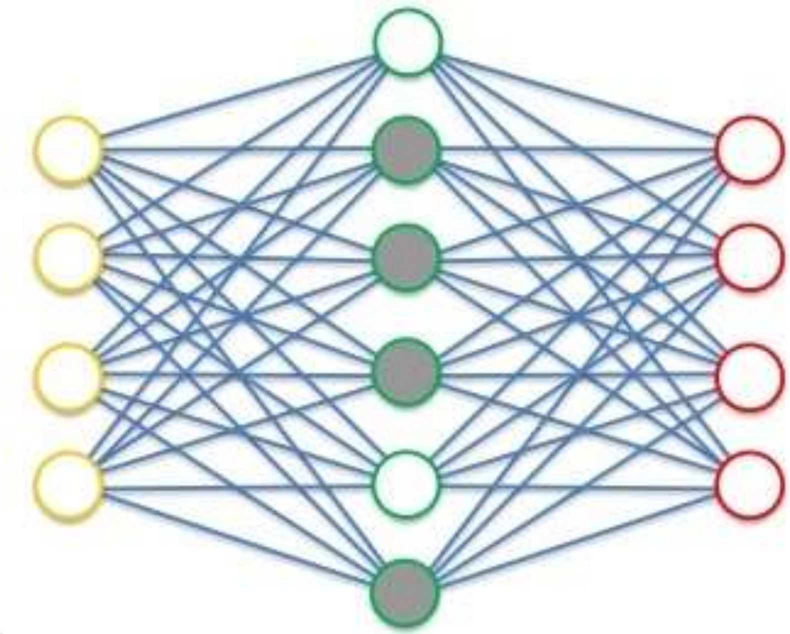
Denoising Auto Encoder



Sparse Auto Encoder

- Reduce the number of active neurons in the coding layer.
  - Add sparsity loss into the cost function.

- Sparsity loss:
  - Kullback-Leibler(KL) divergence is commonly used.



Sparse Auto Encoder

$$J_{sparse}(w,b) = J(w,b) + \beta \sum_{j=1} KL(\rho \,\square\, \hat{\rho}_j)$$

$$KL(\rho \,\square\, \hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1-\rho) \log \frac{1-\rho}{1-\hat{\rho}_j}$$