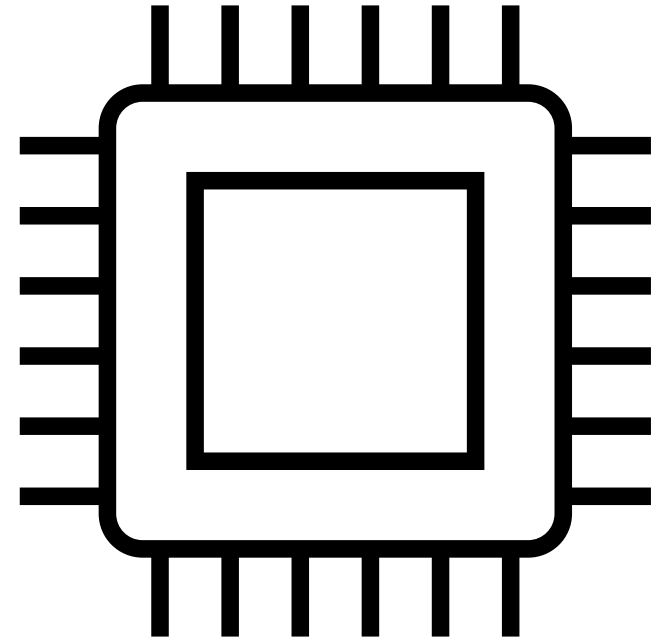


# Manifold Learning (Latent Space)

2025.07.09.

Copyright©2025 by 고재균



# Manifold Learning [1]

## • Manifold Learning

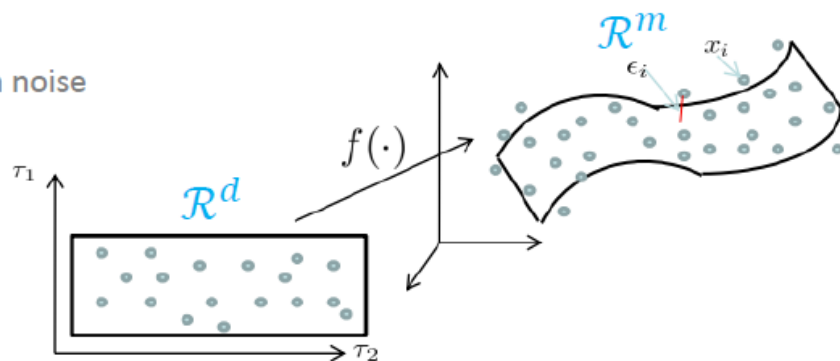
- 고차원 데이터가 있을 때 고차원 데이터를 데이터 공간에 뿌리면 sample들을 잘 아우르는 subspace가 있을 것이라는 가정에서 학습을 진행
- 이렇게 찾은 manifold는 데이터의 차원을 축소

### ▶ $d \rightarrow \text{Latent}$

- A  $d$  dimensional manifold  $\mathcal{M}$  is embedded in an  $m$  dimensional space, and there is an explicit mapping  $f: \mathcal{R}^d \rightarrow \mathcal{R}^m$  where  $d \leq m$

- We are given samples  $x_i \in \mathcal{R}^m$  with noise

$$x_i = f(\tau_i) + \epsilon_i$$



- $f(\cdot)$  is called embedding function,  $m$  is the extrinsic dimension,  $d$  is the intrinsic dimension or the dimension of the latent space
- Finding  $f(\cdot)$  or  $\tau_i$  from the given  $x_i$  is called manifold learning
- We assume  $p(\tau)$  is smooth, is distributed uniformly, and noise is small → **Manifold Hypothesis**

# Manifold Learning [2]

- **Manifold Learning**

- Application

- ▶ Data Compression

- ▶ Data Visualization

- ▶ Curse of Dimensionality

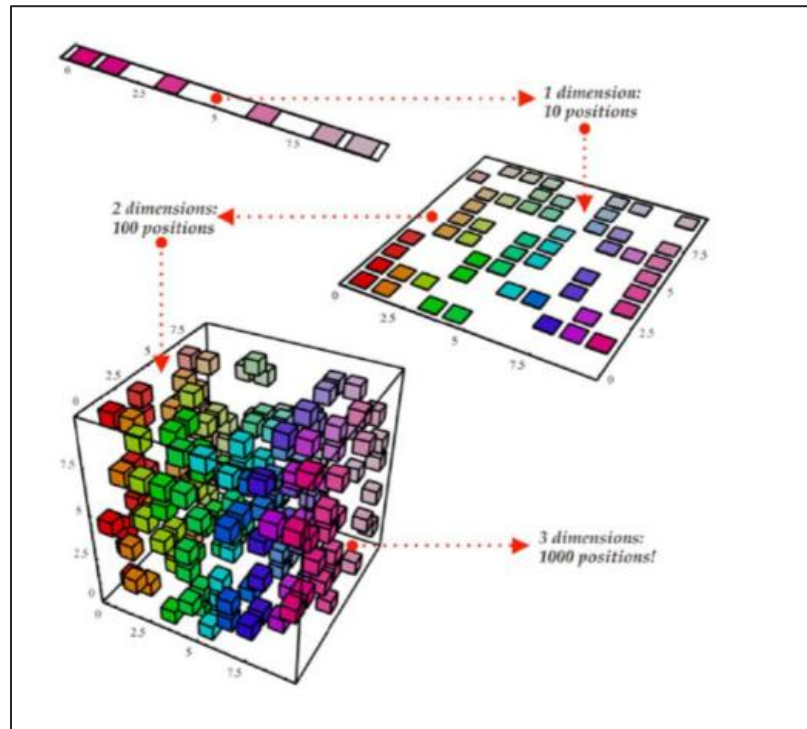
- ▶ Discovering Most Important Features

# Manifold Learning [3]

- Manifold Learning

- Curse of Dimensionality

- ▶ 데이터의 차원이 증가할수록 해당 공간의 크기(부피)는 기하급수적으로 증가
    - ▶ 동일한 개수의 데이터의 밀도는 차원이 증가할수록 급속도로 희박해지게 됨
    - ▶ 따라서 차원이 증가할수록 데이터 분포 분석 또는 모델 추정에 필요한 샘플 데이터의 개수가 기하급수적으로 증가

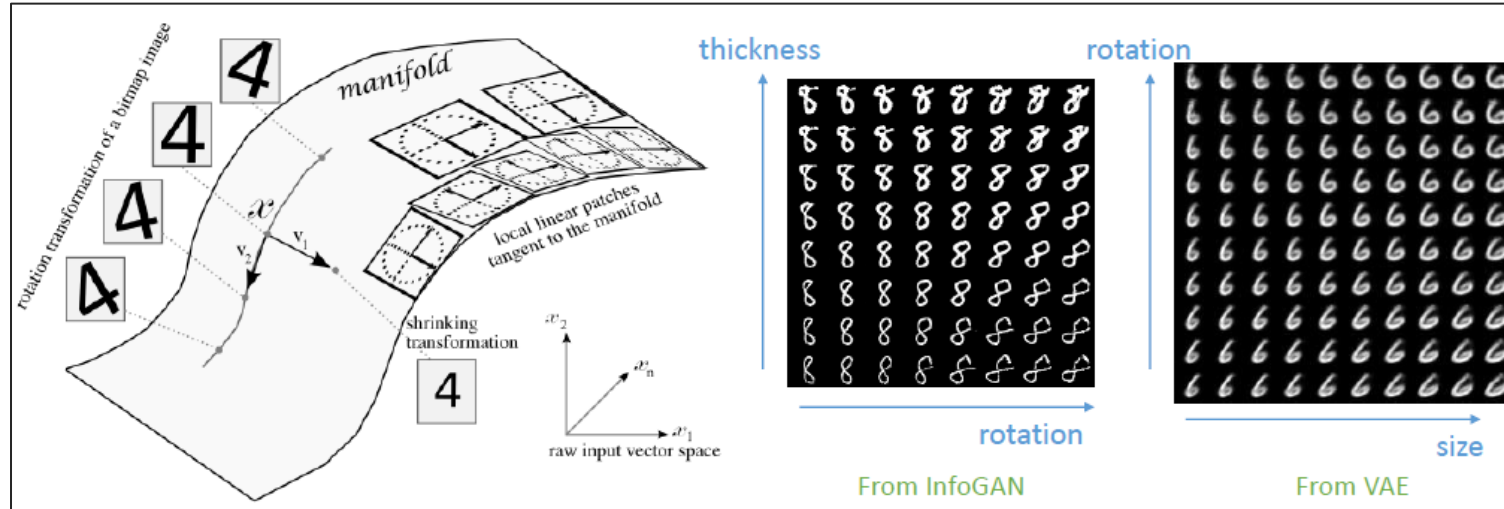


# Manifold Learning [4]

## • Manifold Learning

### ▪ Discovering Most Important Features

- ▶ 데이터의 차원이 증가할수록 해당 공간의 크기(부피)는 기하급수적으로 증가
- ▶ 이 저차원의 Manifold를 벗어나는 순간 밀도는 급격히 낮아지게 됨
- ▶ 즉 고차원의 데이터를 잘 표현하는 Manifold를 통해 우리는 샘플 데이터의 특징을 파악 가능
- ▶ 고차원의 데이터를 잘 표현한다 함은 바로 데이터의 중요한 특징을 발견하는 것
- ▶ 고차원 데이터의 Manifold 좌표들을 조정해보면 Manifold의 변화에 따라 학습 데이터도 유의미하게 조금씩 변하는 것을 확인 가능

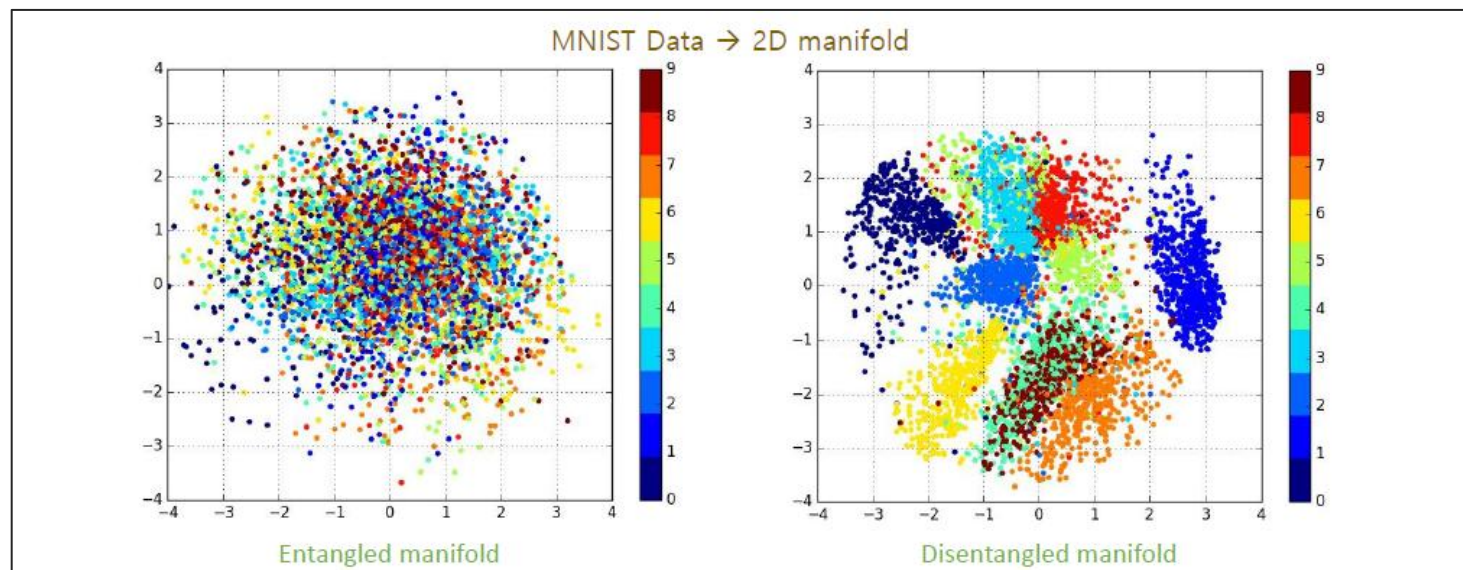


# Manifold Learning [5]

## • Manifold Learning

### ▪ Key Features of Manifold Learning

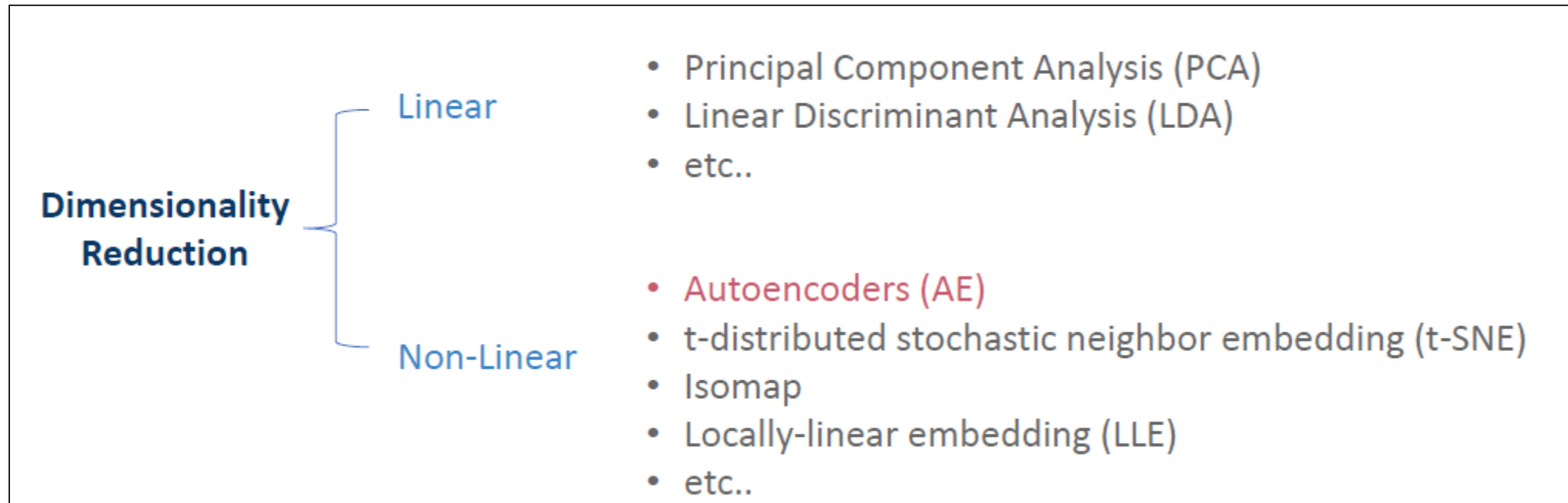
- ▶ Data (샘플)를 잘 아우르는 Manifold를 찾게 되면 얻는 효과
  - ▷ Feature를 잘 찾았기 때문에 Manifold의 좌표를 조금씩 변경해가면서 데이터를 유의미하게 조금씩 변화 가능
  - ▷ 역으로 Manifold를 잘 찾았다면 Dominant Feature가 유사한 Sample들을 찾아보는 것이 가능
- ▶ 일반적으로 학습된 Manifold는 얽혀 읽는 관계 (예 : 복잡한 방식으로 데이터 공간에 인코딩 됨)
- ▶ Manifold 가 풀리면 해석이 쉽고 작업에 쉽게 적용 가능 (Manifold Disentangling)



# Manifold Learning [6]

- **Manifold Learning**

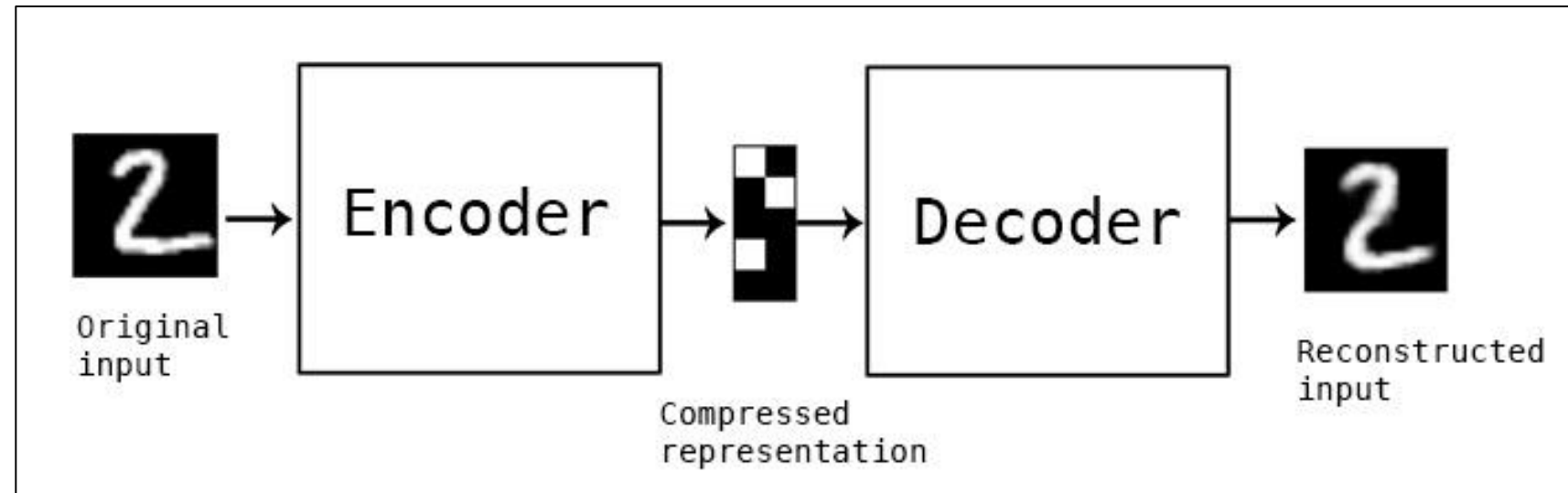
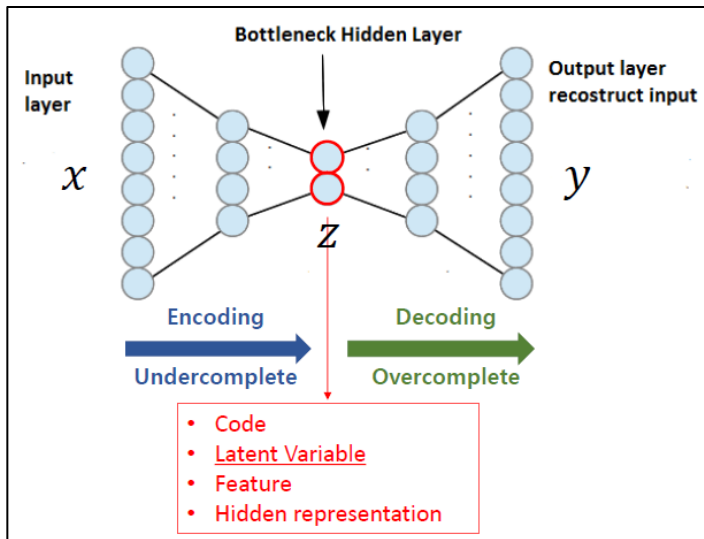
- Types of Dimension Reduction



# Manifold Learning [7]

## • Autoencoder

- 입력을 출력으로 복사하는 신경망
- 인코더 (Encoder)와 디코더 (Decoder)로 구성
- 인코더는 잠재적 공간 (Latent Space)로 입력 데이터의 차원을
- 디코더는 축소된 데이터를 원래 차원으로 복원





# Manifold Learning [8]

## • Autoencoder

- 인코더를 통해 알지 못하는 Latent Variable를 추출
- Latent Variable을 찾는 과정은 훈련을 통해 자동으로 진행
- 이러한 과정으로 인해 Autoencoder라는 명칭을 받음
- 이러한 훈련 과정을 Self-Supervised Learning이라고 칭함
  - ▶ Label 되어있지 않은 데이터로부터 좋은 Representation을 얻고자 하는 학습 방식
  - ▶ Autoencoder는 입력 데이터와 복원 데이터의 Distance를 최소화 하는 방식으로 훈련

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \|f_{\theta}(x) - x\|_n^n$$

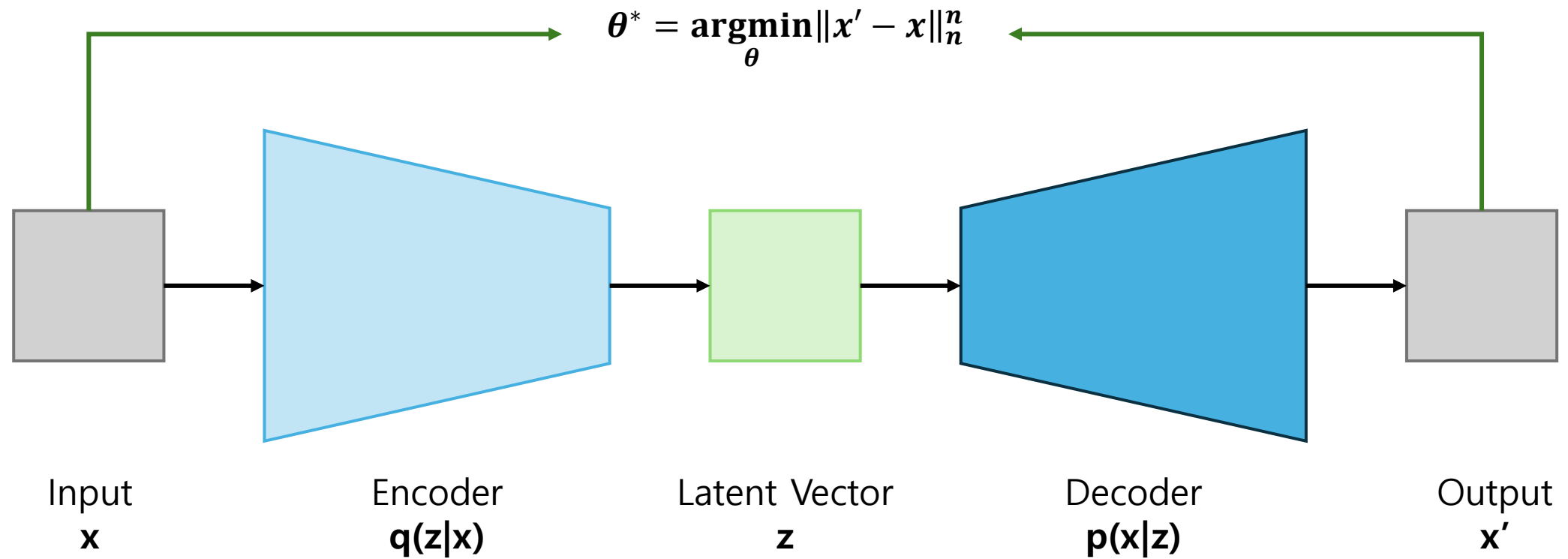
Symbol	Meaning
$x$	Input
$f_{\theta}$	Autoencoder
$\theta$	Model Parameter
$n$	Distance Dimension ( $n = 1, 2$ )

# Manifold Learning [9]



- Autoencoder

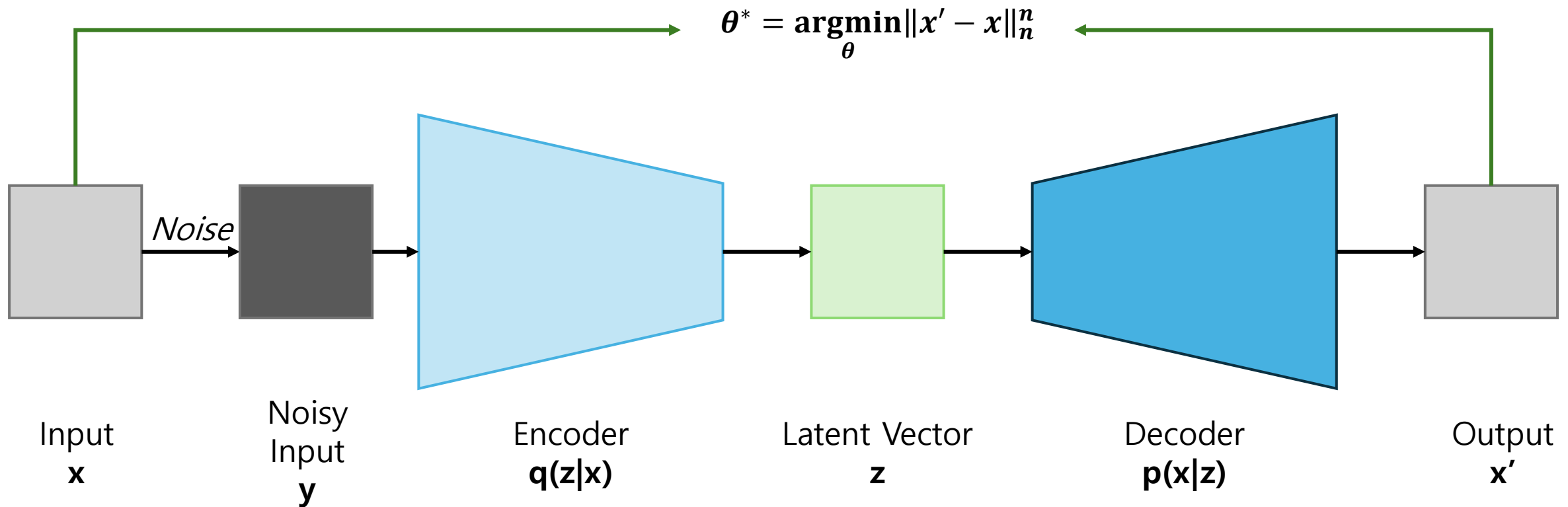
- Overall Framework



# Manifold Learning [10]

- Denoising Autoencoder

- More Robust Network Compared to Vanilla Autoencoder
- Overall Framework



# Manifold Learning [11]

- Autoencoder

- 개인 실습 (1시간 소요)

- ▶ CIFAR10 데이터셋을 사용하여 3-D Latent Space를 학습하세요.

- ▷ [Exercise\_1]Latent\_Space\_Learning\_with\_CIFAR10.ipynb를 활용하세요.

- ▷ Autoencoder 모델 설계부터 Hyperparameter 세팅까지 직접 진행하여 학습을 진행하세요.

- ▷ 학습을 통해 얻은 CIFAR10 Test 데이터의 3-D Latent Space를 시각화하세요.

