인공지능 기반 설계 이론 및 사례 연구

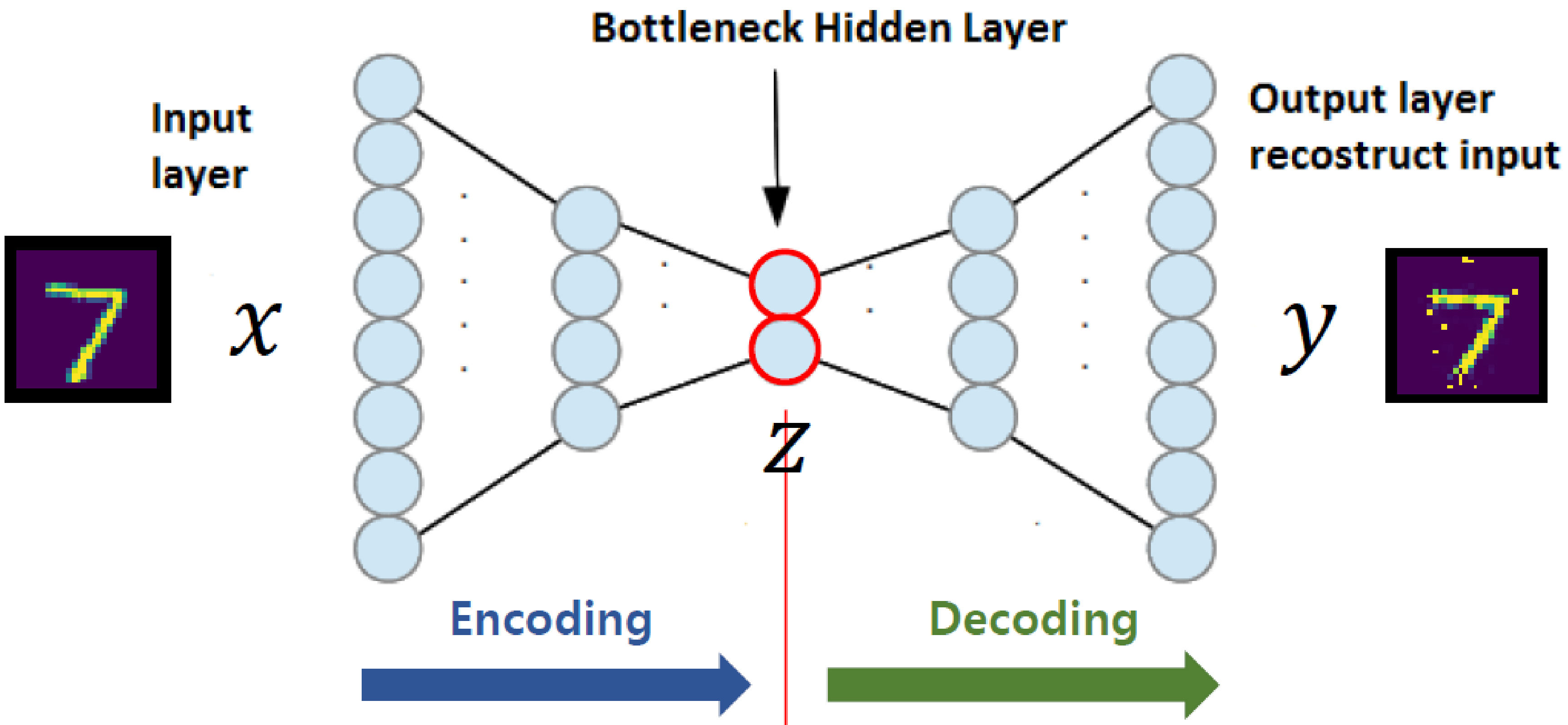
8차) Autoencoder와 Anomaly Detection

2020년 10월

강남우

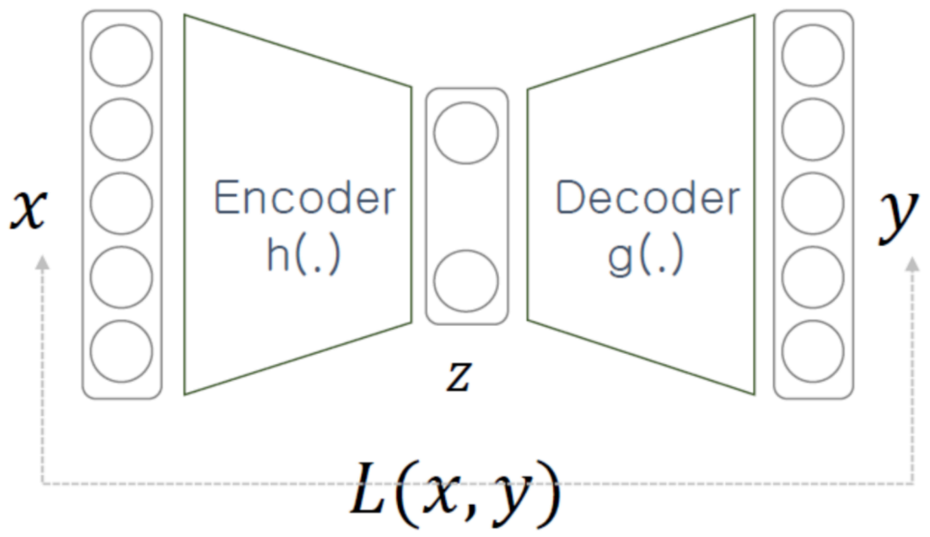
기계시스템학부
숙명여자대학교

Autoencoder – How to work



- Code
- Latent Variable (잠재변수)
- Feature
- Hidden representation

Autoencoder – How to work



$$z = h(x) \in \mathbb{R}^{d_z}$$

$$y = g(z) = g(h(x))$$

$$L_{AE} = \sum_{x \in D} L(x,y)$$

MSE or cross-entropy

$$L_{AE} = \| x - y \|^2$$

Make output layer same size as input layer

$$x, y \in \mathbb{R}^d$$

Loss encourages output to be close input

$$L(x,y)$$

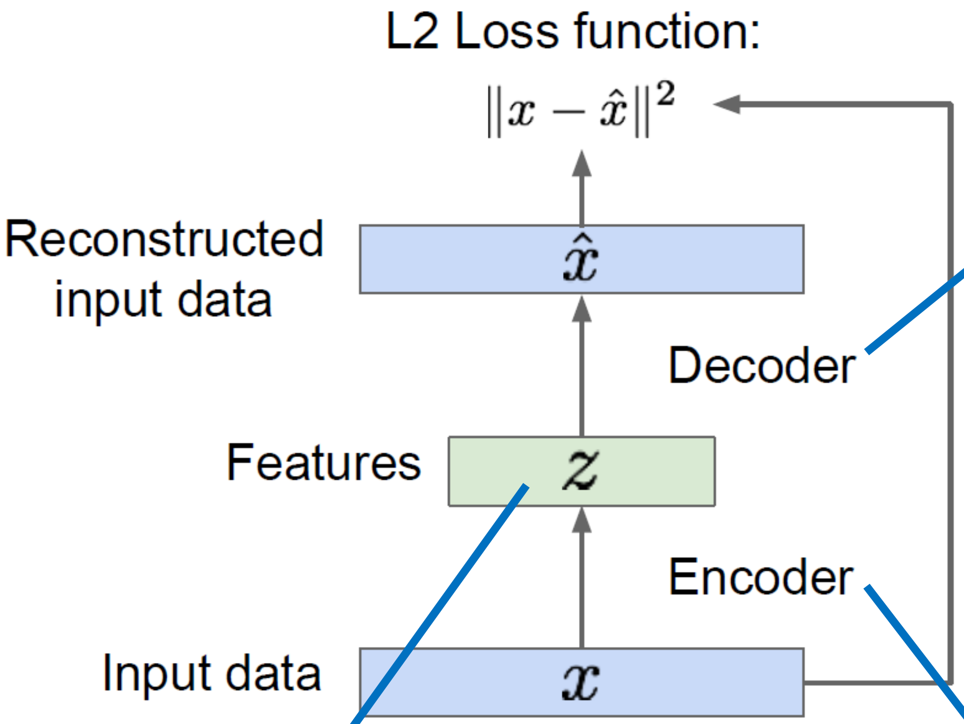
입출력이 동일한 네트워크

Unsupervised Learning ➔ Supervised Learning

비지도학습문제를 지도학습문제로 바꾸어서 해결

- Decoder가 최소한 학습 데이터는 생성해 낼 수 있게 된다.
  ➔ 생성된 데이터가 학습 데이터를 좀 닮아있다.
- Encoder가 최소한 학습 데이터는 잘 latent vector로 표현 할 수 있게 된다.
  ➔ 데이터의 추상화를 위해 많이 사용된다.

Autoencoder – How to work

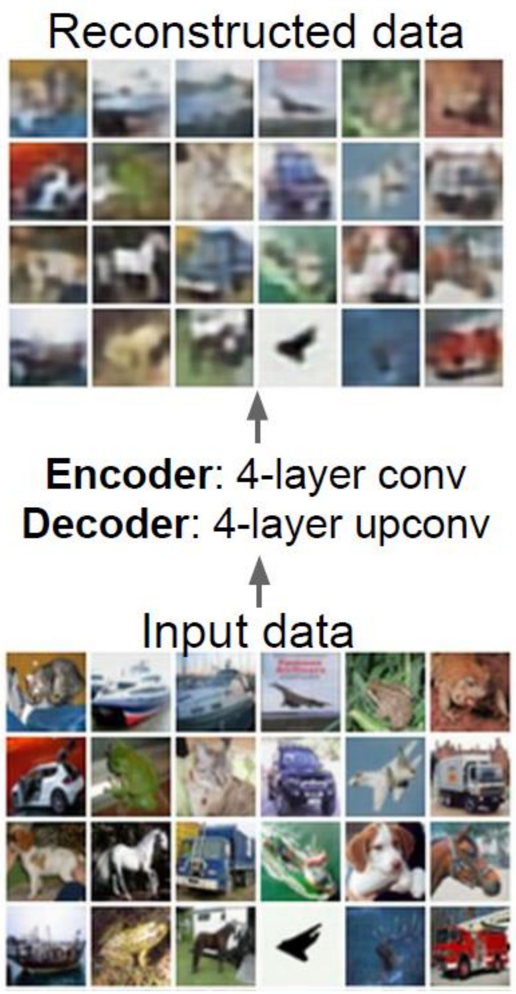


Originally: Linear + nonlinearity (sigmoid)
Later: Deep, fully-connected
Later: ReLU CNN (upconv)

Originally: Linear + nonlinearity (sigmoid)
Later: Deep, fully-connected
Later: ReLU CNN

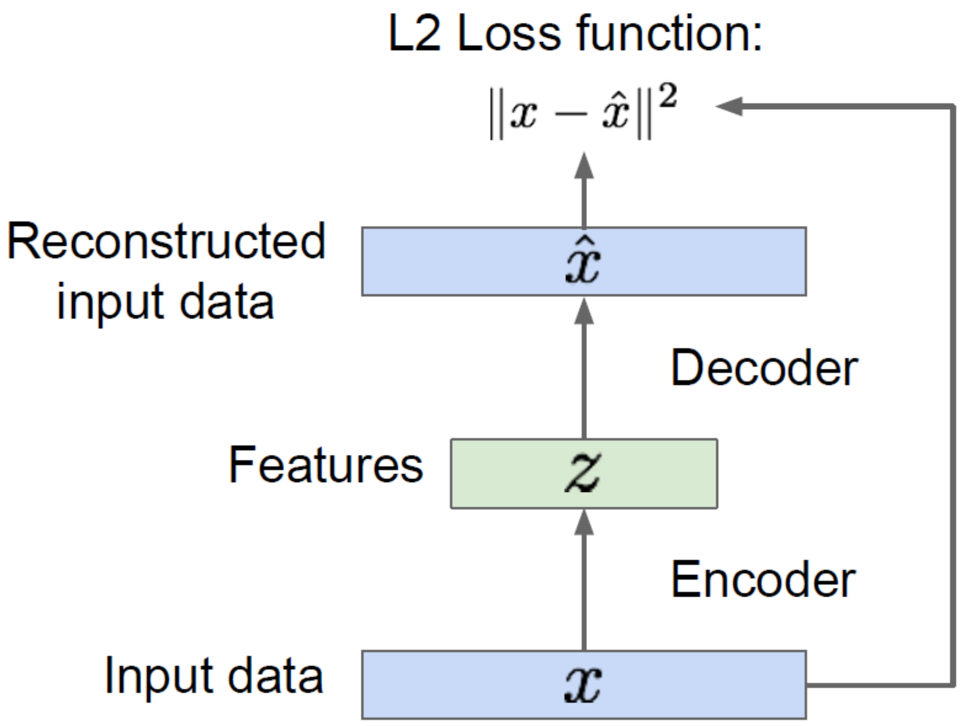
z usually smaller than **x** (dimensionality reduction)

Doesn't use labels!

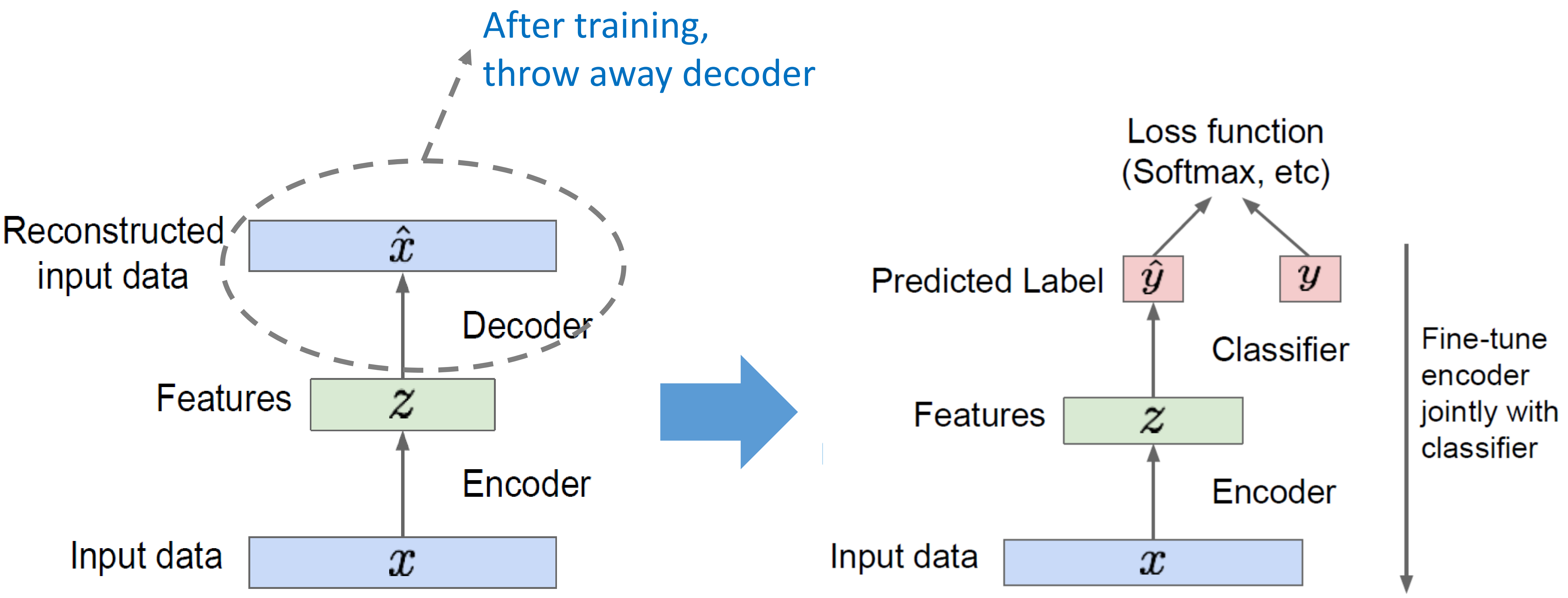


Autoencoder Coding

```
# 모델 학습
hist = model.fit(x_train, x_train, nb_epoch=num_epochs, batch_size=batch_size, shuffle=True, verbose=1)
```



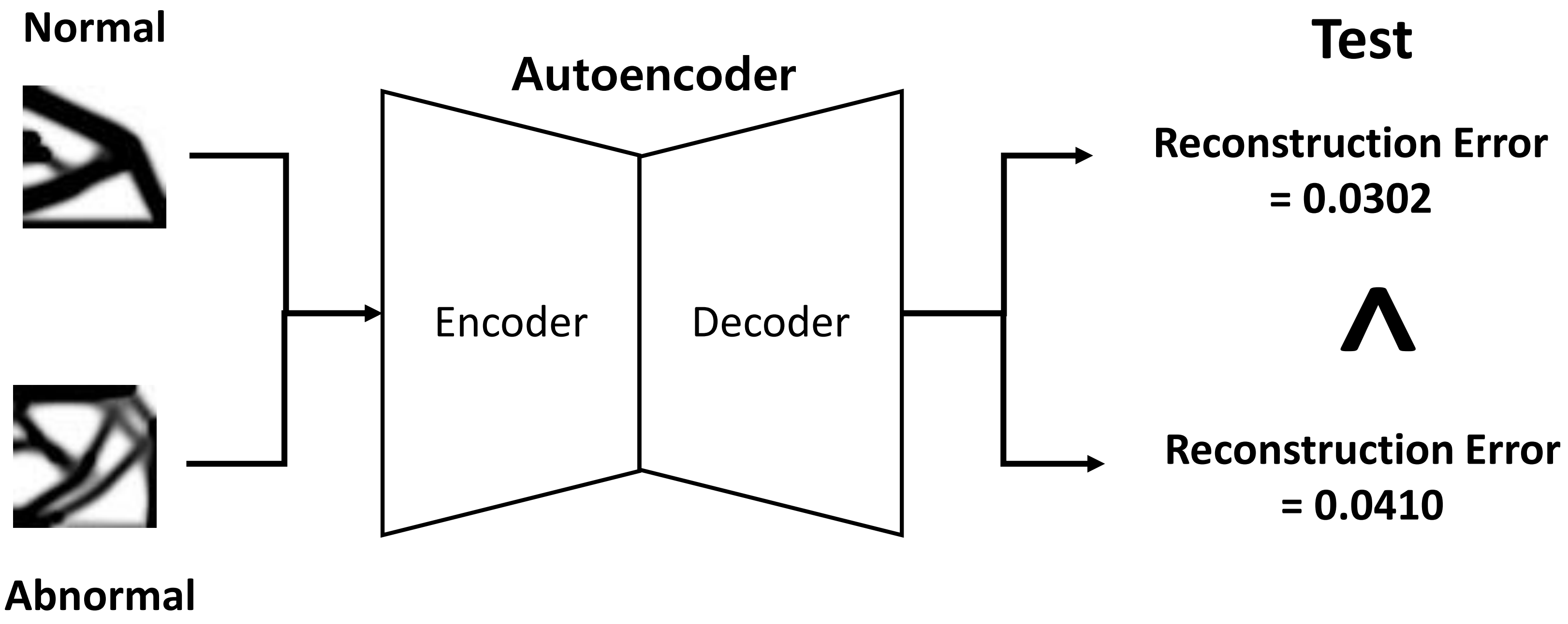
Application 1: Supervised Learning



Encoder can be used to initialize a **supervised** model

Define:

- Normal: Brackets with *small* compliance ➔ Normal data만 사용해서 AE 학습시키기
- Abnormal: Brackets with *large* compliance



➔ Reconstruction Error가 임계치보다 크면 Abnormal로 분류

SMART DESIGN LAB

Application 2: Anomaly Detection

Confusion Matrix

n=165	Predicted: Negative	Predicted: Positive	
Actual: Negative	TN = 50	FP = 10	60
Actual: Positive	FN = 5	TP = 100	105
	55	110	

- **true positives (TP):** These are cases in which we predicted yes (they have the disease), and they do have the disease.
- **true negatives (TN):** We predicted no, and they don't have the disease.
- **false positives (FP):** We predicted yes, but they don't actually have the disease. (Also known as a "Type I error.")
- **false negatives (FN):** We predicted no, but they actually do have the disease. (Also known as a "Type II error.")

SMART DESIGN LAB

Application 2: Anomaly Detection

Go to http://www.smartdesignlab.org/dl.html

Confusion Matrix

n=165	Predicted: Negative	Predicted: Positive	
Actual: Negative	TN = 50	FP = 10	60
Actual: Positive	FN = 5	TP = 100	105
	55	110	

성능지표

- **Accuracy** (실제 이상/정상인지 맞게 예측한 비율)

  = (TP+TN)/(TP+FN+FP+TN) = 90.9%
- **Precision** (이상으로 예측한 것중에 실제 이상인 샘플의 비율)

  = TP/(TP+FP) = 90.9%
- **Recall** (실제 이상 샘플중에 이상으로 예측한 비율)

  = TP/(TP+FN) = 95.20%

What Questions Do You Have?

nwkang@sm.ac.kr

www.smartdesignlab.org