Autoencoder & Anomaly Detection

Namwoo Kang

Smart Design Lab
CCS Graduate School of Green Transportation
KAIST

Reference

❑ 강의 슬라이드 및 실습코드는 아래의 링크에서 받으실 수 있습니다

- http://www.smartdesignlab.org/dl_aischool_2021.html
- Contributors: 김성신, 유소영, 이성희, 김은지

❑ 강의 소스

- Andrew Ng의 ML Class (www.holehouse.org/mlclass/)
- Fei-Fei Li & Justin Johnson & Serena Yeung, CS231n: Convolutional Neural Networks for Visual Recognition, Stanford (http://cs231n.stanford.edu/)
- Stefano Ermon & Aditya Grover, CS 236: Deep Generative Models , Stanford (https://deepgenerativemodels.github.io/)
- 모두를 위한 딥러닝 (https://hunkim.github.io/ml/)
- 모두를 위한 딥러닝 시즌 2 (https://deeplearningzerotoall.github.io/season2/lec_tensorflow.html)
- 이활석, Autoencoders (https://www.slideshare.net/NaverEngineering/ss-96581209)
- 최윤제, 1시간만에 GAN(Generative Adversarial Network) 완전 정복하기 (https://www.slideshare.net/NaverEngineering/1-gangenerative-adversarial-network?qid=c53ce33f-6643-4437-8e93-88776c9cebb1&v=&b=&from_search=5)
- 김성범, [핵심 머신러닝] Principal Component Analysis (PCA, 주성분 분석) (https://youtu.be/FhQm2Tc8Kic)

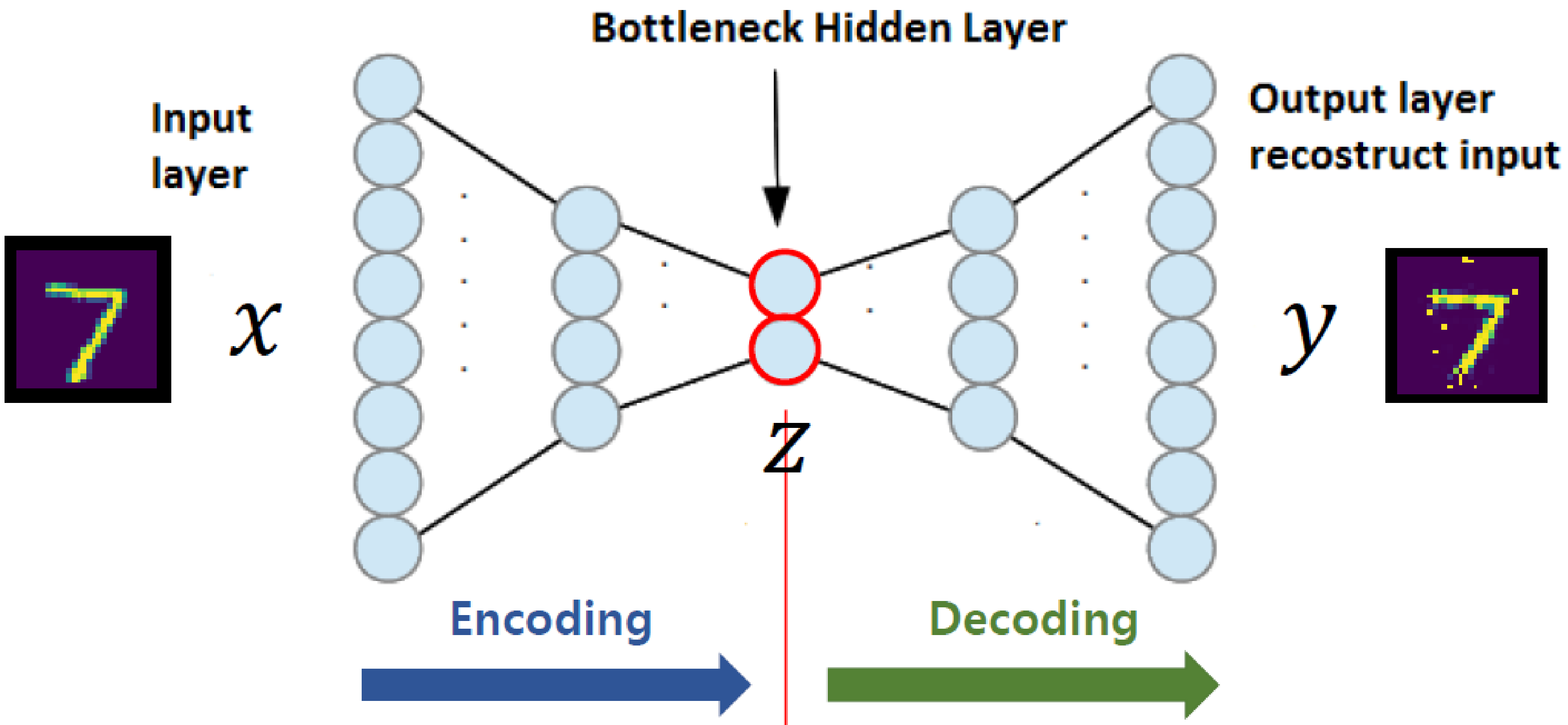
Contents

- **Ch1: Introduction to Unsupervised Learning Part I** → Probability & Maximum Likelihood
- **Ch2: Introduction to Unsupervised Learning Part II** → Generative Model & Dimensionality Reduction
- **Ch3: Principal Component Analysis (PCA)** → Machine Learning Model
- **Ch4: Autoencoder & Anomaly Detection**
  + 실습
- **Ch5: Variational AutoEncoder (VAE)**
  + 실습
- **Ch6: Generative Adversarial Network (GAN)**
  + 실습

→ Deep Learning Models

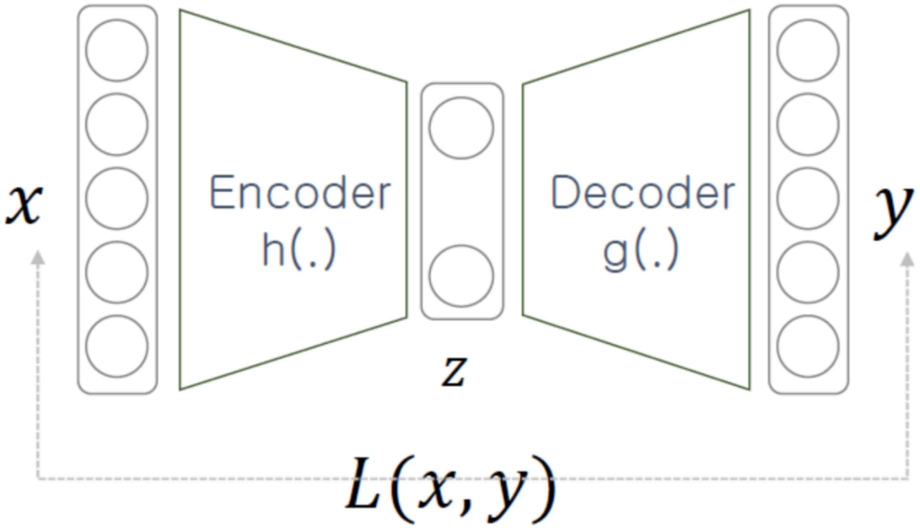
- **Ch7: Application: Mechanical Design + AI** → CAD/CAM/CAE/Design Optimization + AI

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

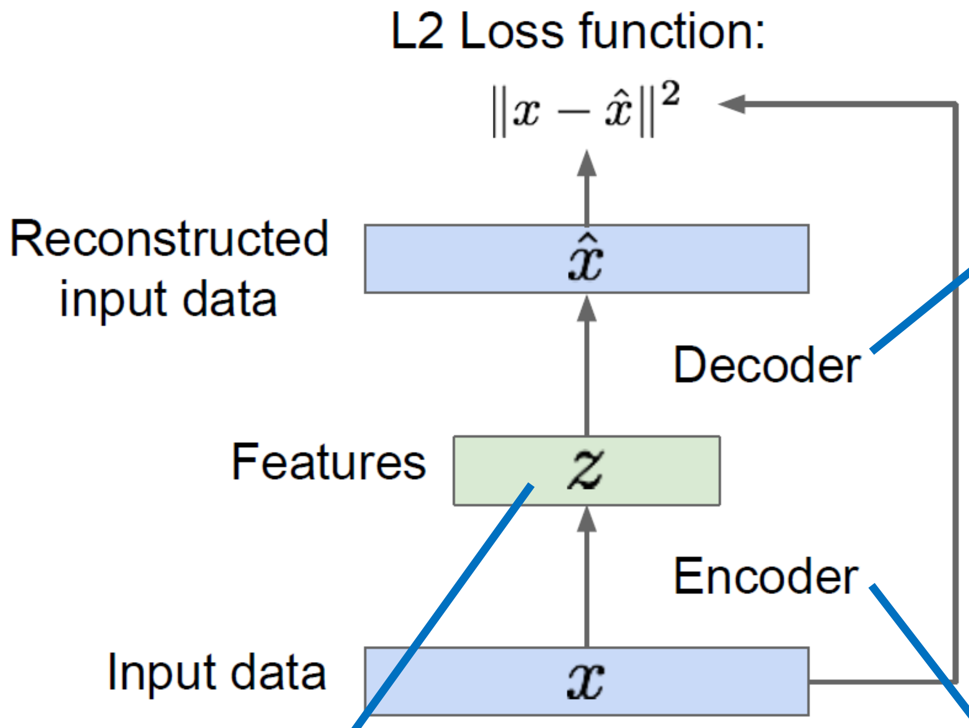
Unsupervised Learning ➔ Supervised Learning

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.

SMART DESIGN LAB KAIST

Autoencoder – How to work

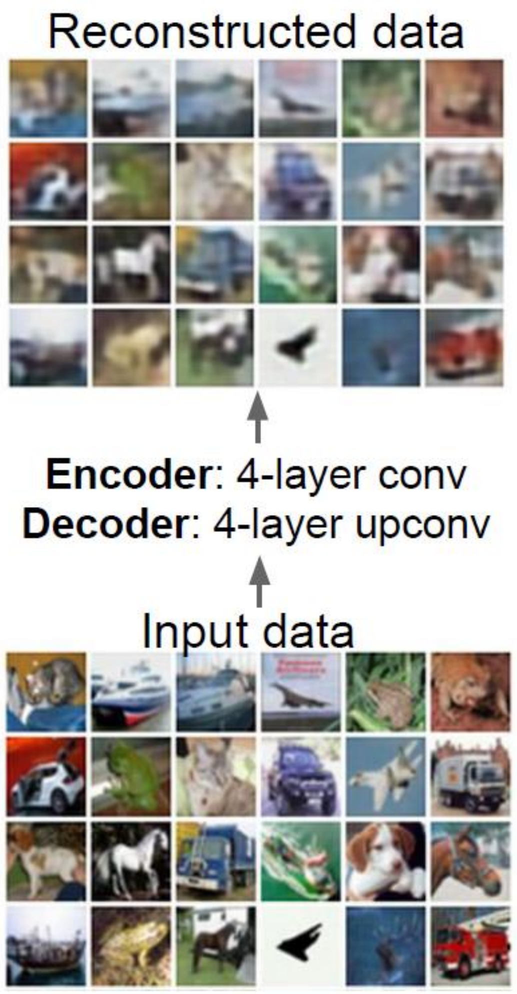


Originally: Linear + nonlinearity (sigmoid)
Later: Deep, fully-connected
Later: ReLU CNN (upconv)

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

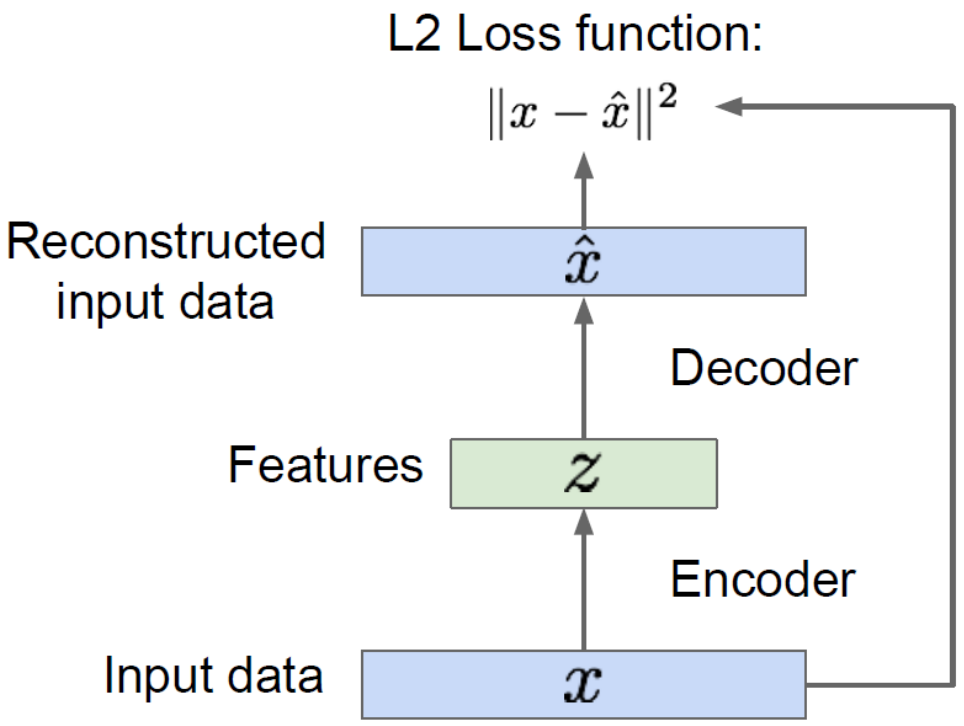
z usually smaller than **x** (dimensionality reduction)

Doesn't use labels!



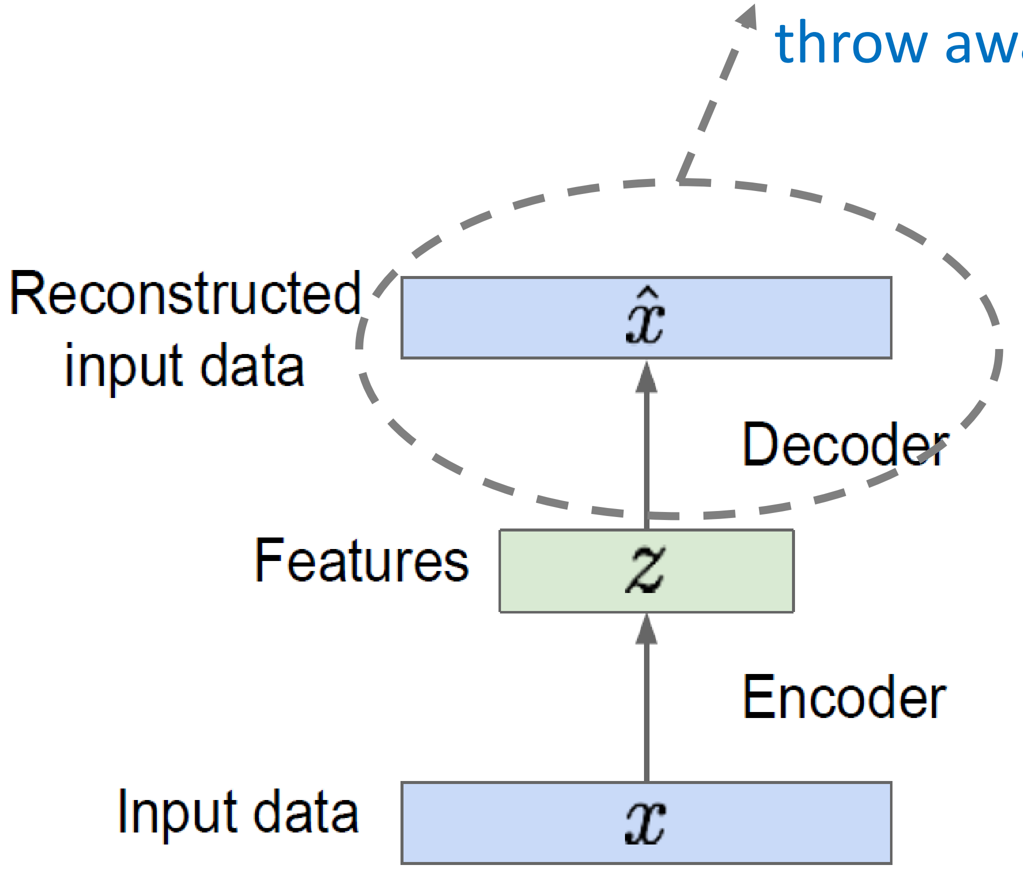
Autoencoder Coding

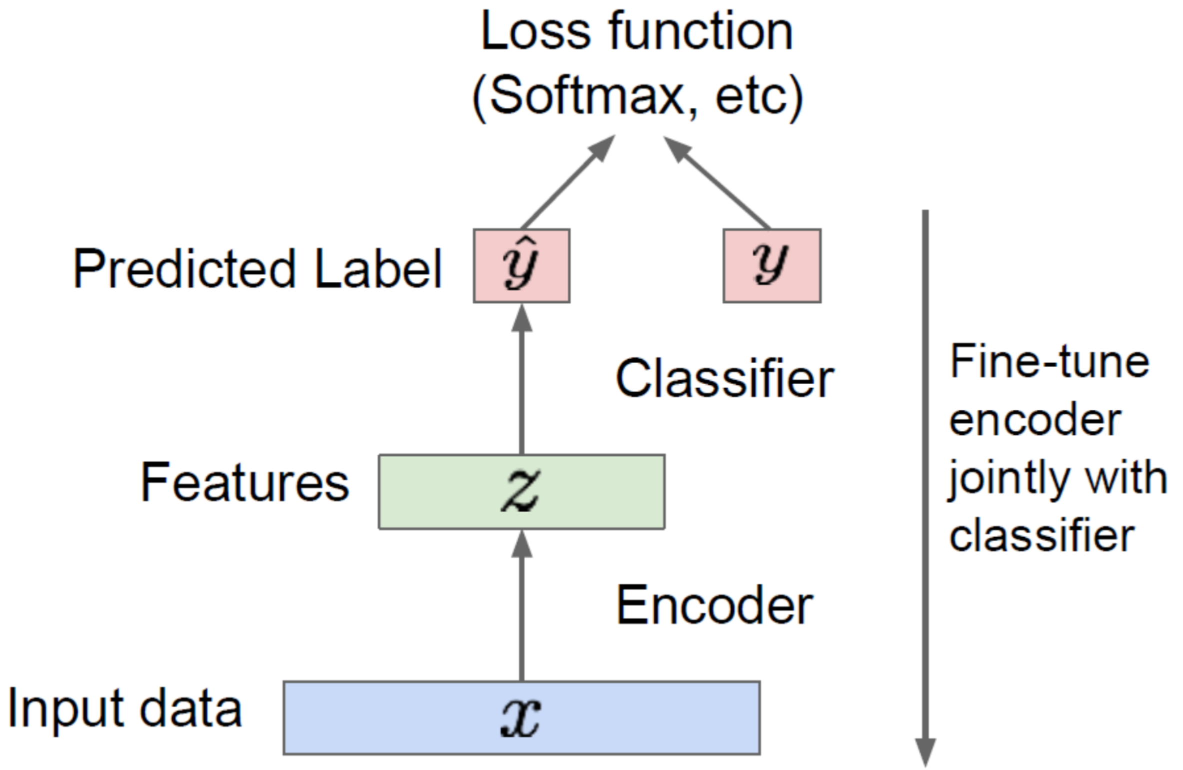
```
# 모델 학습
hist = model.fit(x_train, x_train, nb_epoch=num_epochs, batch_size=batch_size, shuffle=True, verbose=1)
```



Application 1: Supervised Learning

After training,
throw away decoder



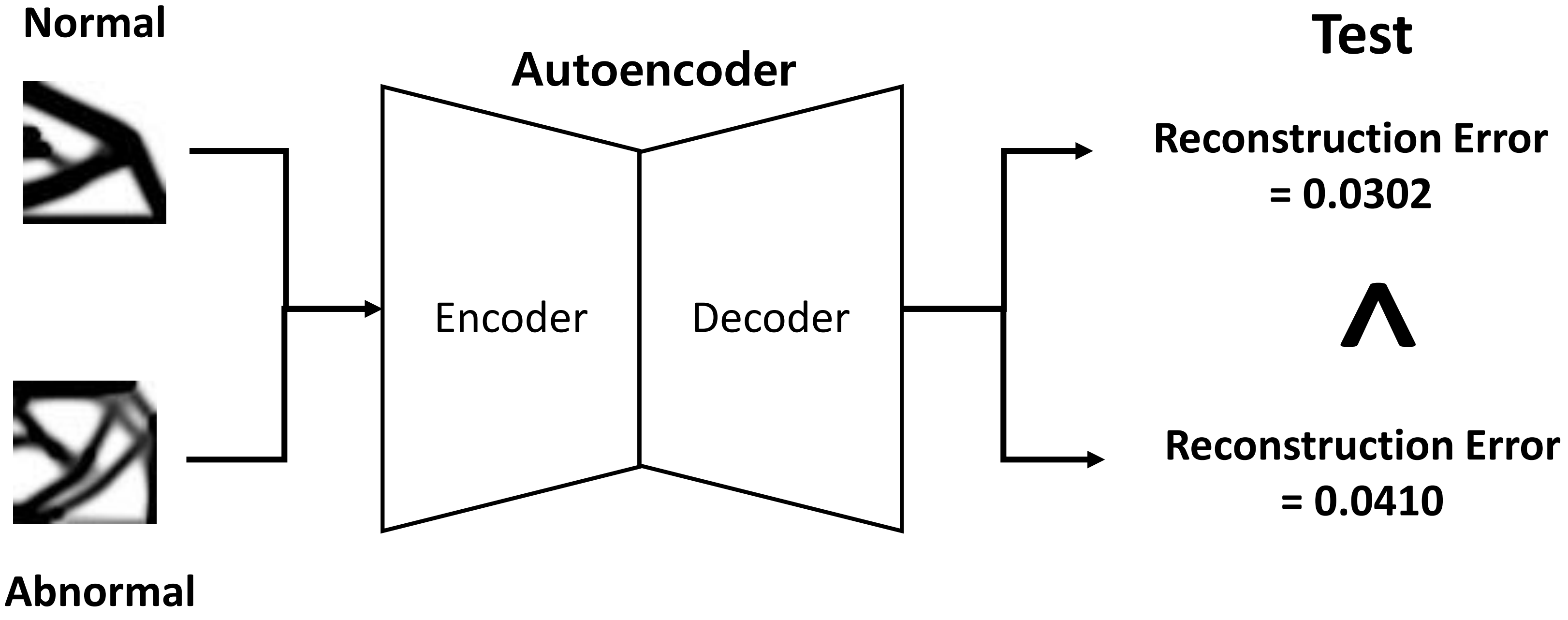


Encoder can be used to initialize
a **supervised** model

Application 2: Anomaly Detection

Define:

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



➔ Reconstruction Error가 임계치보다 크면 Abnormal로 분류

What Questions Do You Have?

nwkang@kaist.ac.kr

www.smartdesignlab.org

SMART DESIGN LAB KAIST