고용노동부 SK planet

Reconstruction Error Based Anomaly Detection
재구축 오차 기반 이상탐지 방법론

9주차 스터디 세션

2024.01.05(금) 09:30

김설진/안정현/이승용/이지흔/이혜준

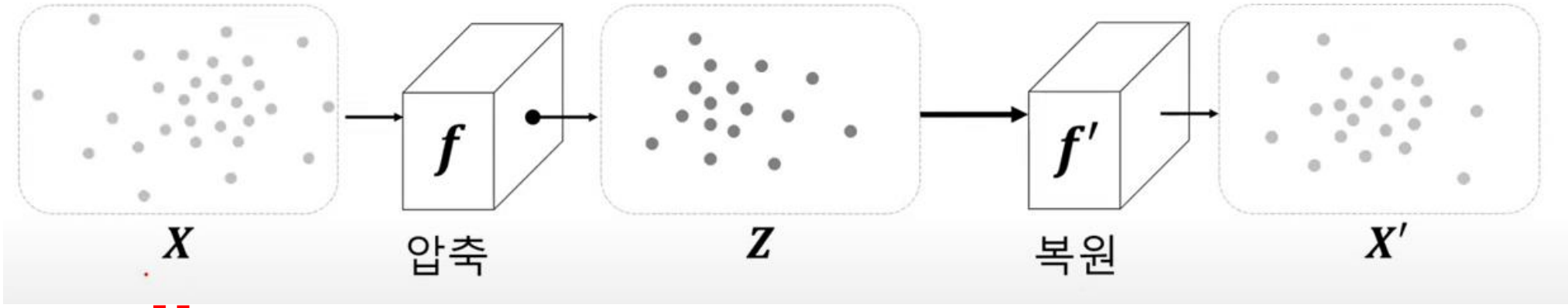
Index

학습 목차

8.1.1 재구축 오차 원리

재구축 오차 기반 알고리즘의 핵심 아이디어

원본 데이터를 압축하고 복원하는 과정에서 발생하는 오차를 이용한 이상탐지 방법론



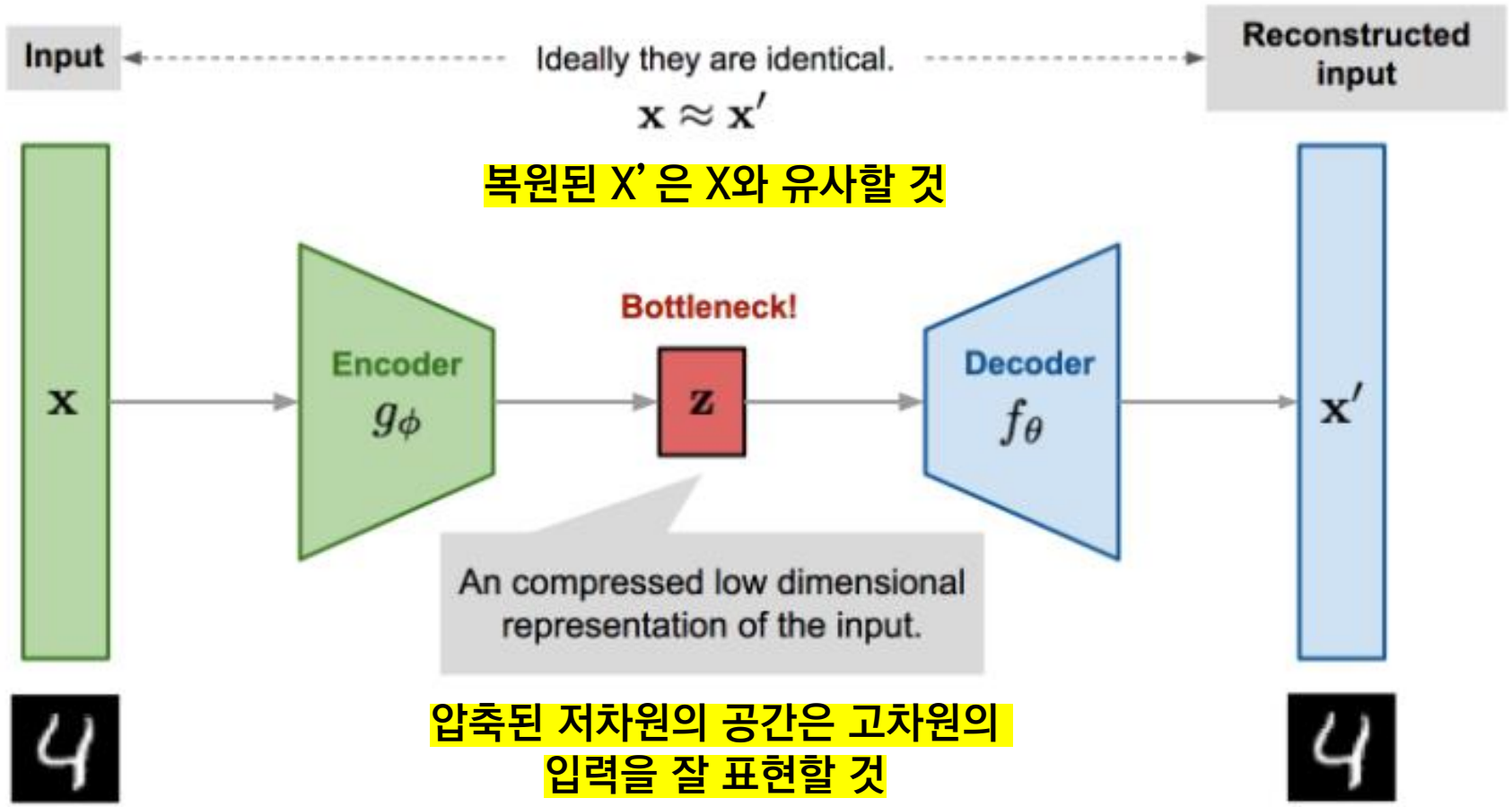
X의 차원 〉 Z의 차원

X’ 의 차원 = X의 차원

8.1.1 재구축 오차 원리

재구축 오차 접근법의 2가지 가정

복원된 데이터는 원본 데이터와 유사할 것이며, 압축된 차원은 고차원 공간을 잘 표현할 것



8.1.1 재구축 오차 원리

재구축 오차를 통한 이상치 탐지 방법

정상 데이터에 대한 **표현**을 잘 학습한 알고리즘은 **정상 데이터를 잘 복원**하지만,
이상 데이터는 잘 복원하지 못함

8.1.2 용어 정리

알고리즘 이해를 위한 중요 개념

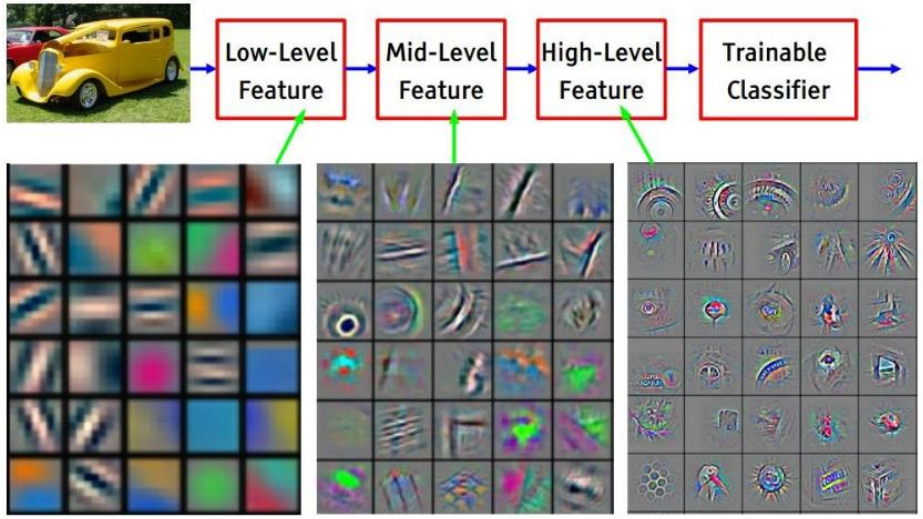
Representation Learning (표현 학습)

딥러닝 = representation learning?

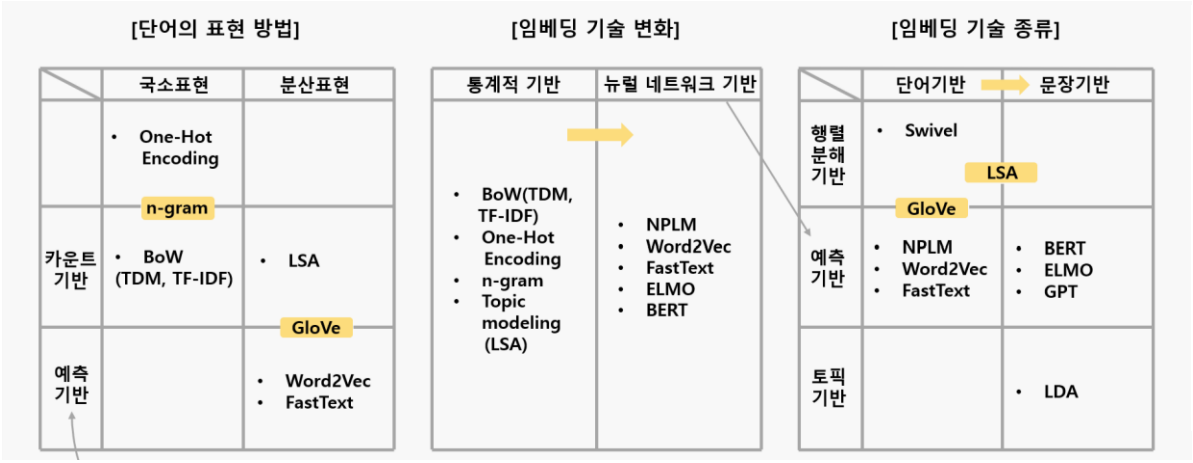
사람이 직접(feature engineering) vs 모델이 스스로

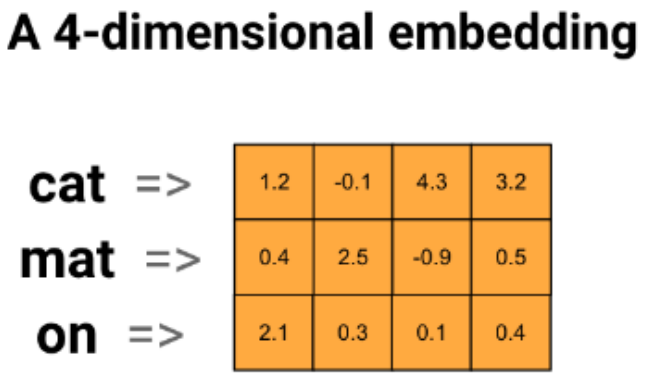
해석의 어려움...

▼CNN의 convolution filter



단어 임베딩 ▶





8.1.2 용어 정리

알고리즘 이해를 위한 중요 개념

Manifold Learning (매니폴드 학습)

데이터 압축

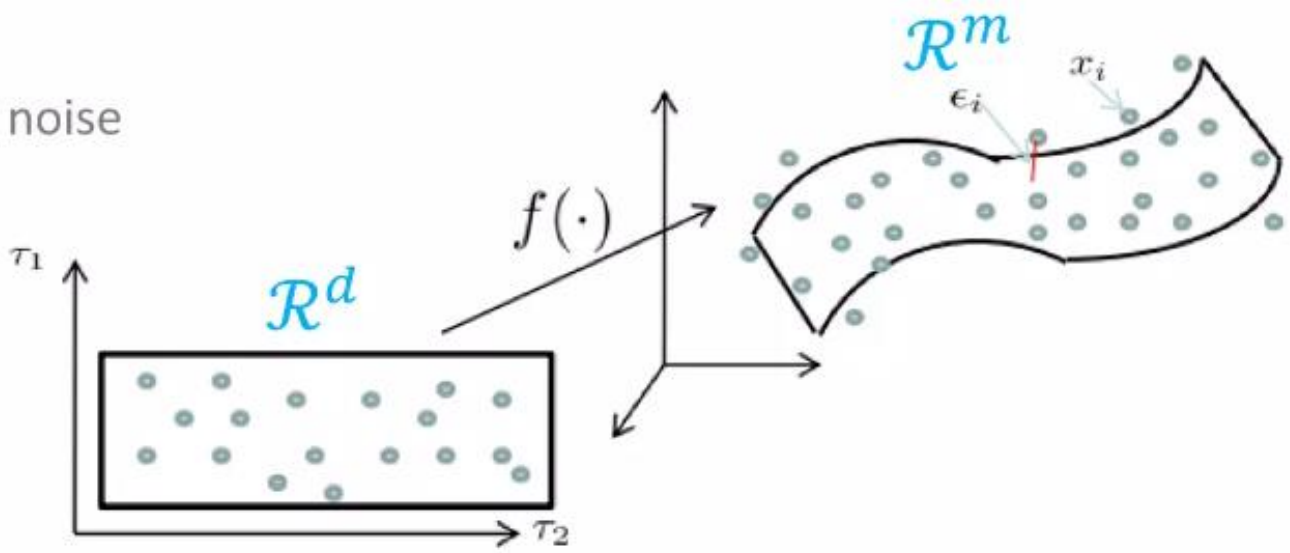
데이터 시각화

차원의 저주

Representation, 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

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



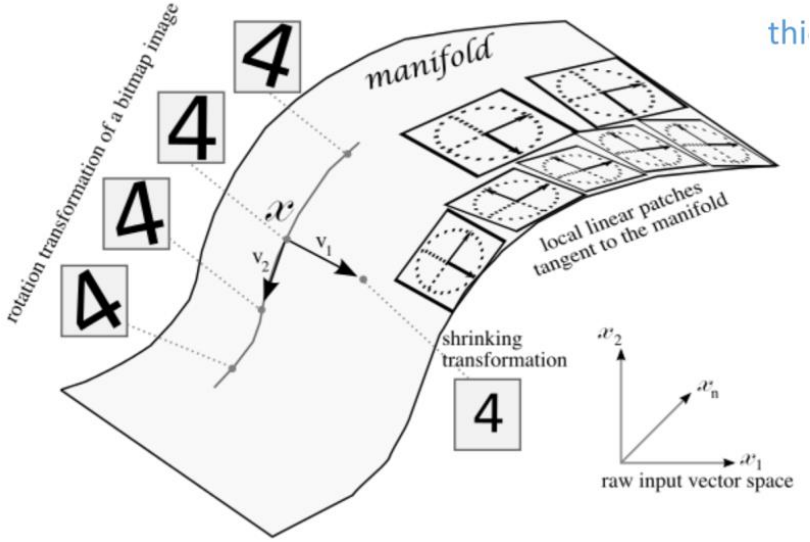
8.1.2 용어 정리

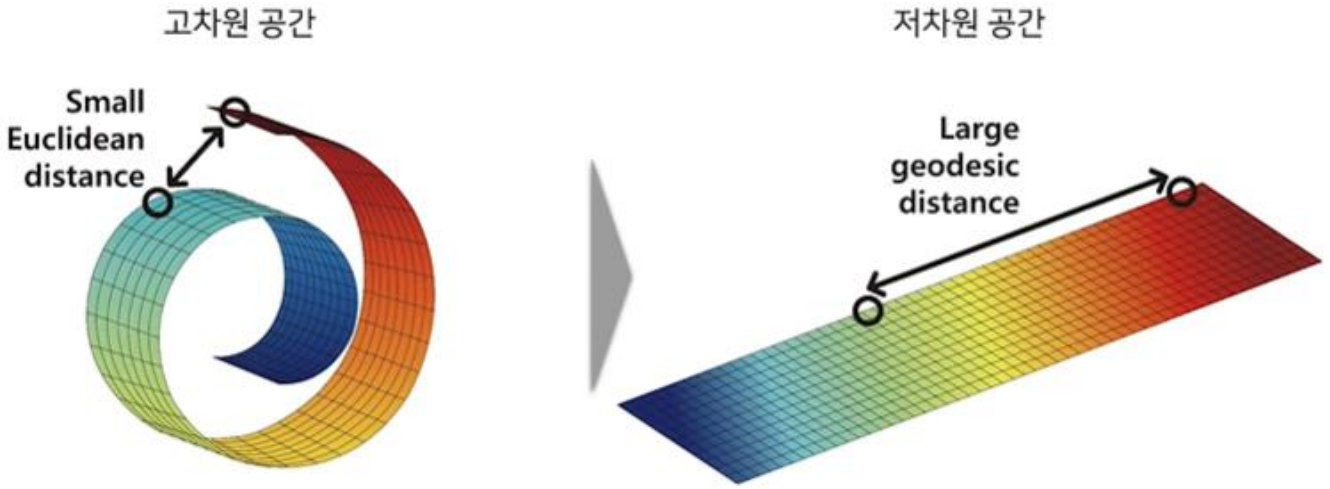
알고리즘 이해를 위한 중요 개념

Manifold Learning (매니폴드 학습)

Manifold Hypothesis

- 고차원의 데이터의 밀도는 낮지만, 이들의 집합을 포함하는 저차원의 매니폴드가 존재
- 저차원의 매니폴드를 벗어나는 순간 급격히 밀도는 낮아짐





8.1.2 용어 정리

알고리즘 이해를 위한 중요 개념

Dimensionality Reduction

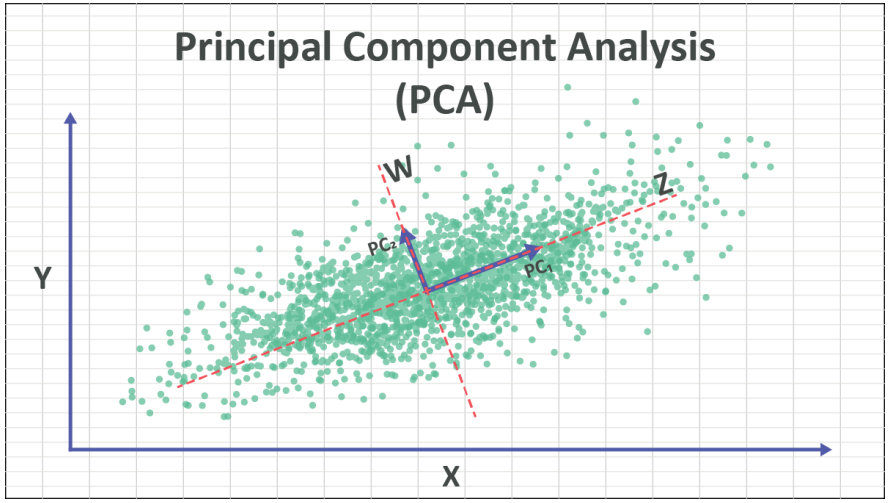
- Linear
  - Principal Component Analysis (PCA)
  - Linear Discriminant Analysis (LDA)
  - etc..
- Non-Linear
  - Autoencoders (AE)
  - t-distributed stochastic neighbor embedding (t-SNE)
  - Isomap
  - Locally-linear embedding (LLE)
  - etc..

8.2 재구축 오차 기반 이상탐지 알고리즘

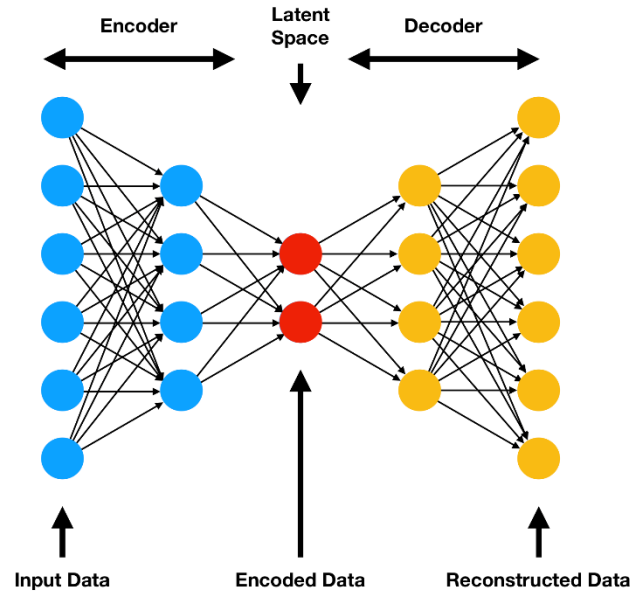
재구축 오차 기반 이상탐지에 활용되는 알고리즘

재구축 오차 기반 이상탐지에 활용되는 대표적인 알고리즘으로는 PCA와 Autoencoder가 있음

Principal Component Analysis



Autoencoder



두 방법론 모두 특징을 잘 요약하는 공통점을 가짐

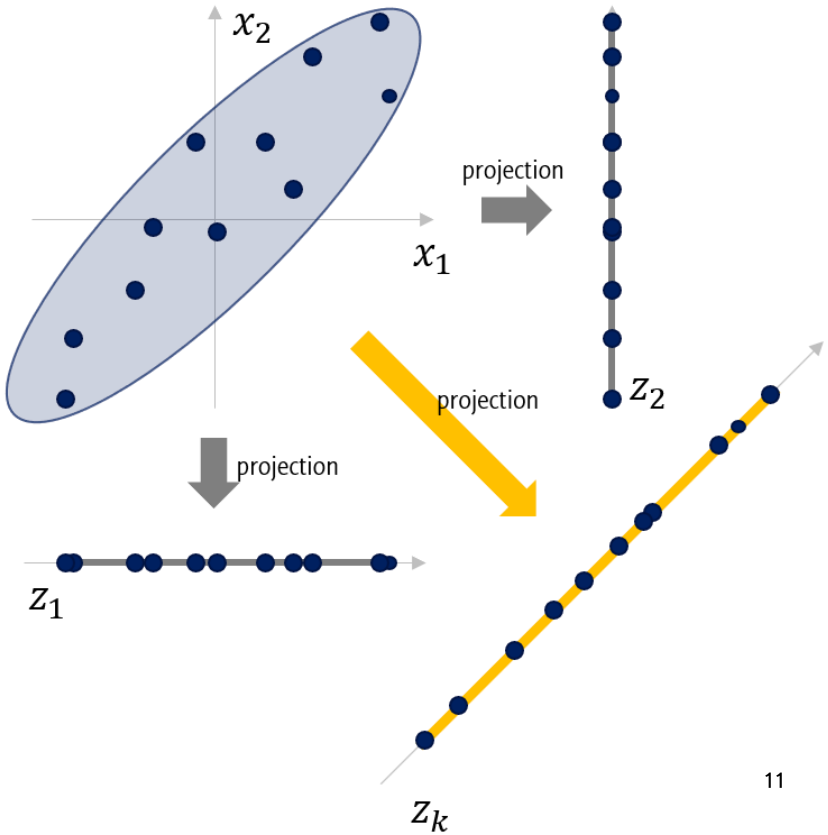
8.2.1 Principal Component Analysis

PCA를 통한 재구축 기반 이상탐지

PCA 원리

- 데이터의 분산(정보)을 최대한 보존하는 새로운 축을 찾고, 그 축에 데이터를 사영(projection, 사상, 투사) 시키는 기법
  - 정보 손실을 최소화하는 방향으로 차원을 축소하는 것이 중요
  - 선형적으로 차원을 축소 (cf. 비선형 차원 축소)
  - 고유값 분해(eigen decomposition)를 통해 데이터를 잘 사영시킬 수 있는 벡터들의 집합을 찾는 것이 목표
  - PCA의 축들은 서로 직교(독립)함으로써, 서로 다른 정보를 보존

Find space z_k which maximize variance of projected data



8.2.1 Principal Component Analysis

PCA를 통한 재구축 기반 이상탐지

주성분 분석을 통한 변환 행렬 구성 단계

1. N개의 D차원 자료들 x_n 으로부터 공분산 행렬 $\sum$(DxD)를 계산한다.

$$\mu = \frac{1}{N}\sum_{n=1}^{N} x_n \qquad \Sigma = \frac{1}{N-1}\sum_{n=1}^{N}(x_n-\mu)(x_n-\mu)^T \quad \text{where} N\text{: No. of sample}$$

2. 고유값 분석을 행한다.

$$\mathbf{\Sigma} = U\Lambda U^T = [\mathbf{u}_1\mathbf{u}_2...\mathbf{u}_D]\begin{bmatrix}\lambda_1 & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & \lambda_D\end{bmatrix}[\mathbf{u}_1\mathbf{u}_2...\mathbf{u}_D]^T$$

3. D개의 고유값들 중에서 가장 큰 고유값 C(<D)개를 $(\lambda_1,....,\lambda_C)$ 을 선택한다.

4. 선택된 고유값과 관련된 고유벡터를 구하고 연결하여 변환 행렬 W을 만든다.

$$W = [u_1...u_C]$$

5. 특징 벡터를 다음의 변환식으로 변환한다.

$$y = W^T x$$

결과적으로 선형 변환

source : 머신러닝 세미나 3회차 - 차원축소를 이용한 데이터 압축

8.2.1 Principal Component Analysis

PCA를 통한 재구축 기반 이상탐지

PCA를 통해 축소된 차원 복원

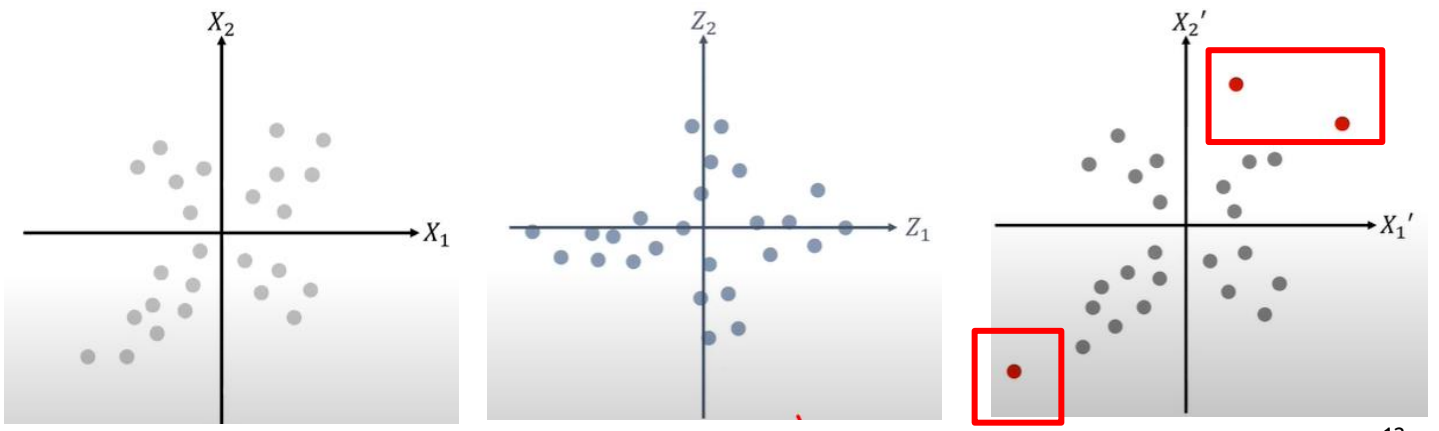
- PCA는 선형결합을 통해 기존의 X 데이터의 차원을 저차원의 Z 공간으로 축소
  - 1) 분산을 최대로 보존하는 새로운 축에 데이터를 사영
  - 2) 새로운 축으로 사영된 데이터를 다시 기존 데이터의 형태로 재구축
  - 3) 재구축된 데이터와 기존 데이터 사이의 차이를 재구축 오차로 정의
  - 4) 재구축 오차가 큰 관측치를 이상치로 정의

Let $Z=\alpha^T X$ be the projection of X onto the direction α

$$Z = \alpha^T X$$

$$X \rightarrow Z = \alpha^T X \rightarrow X' = \alpha\,\alpha^T X$$

original projection reconstruction

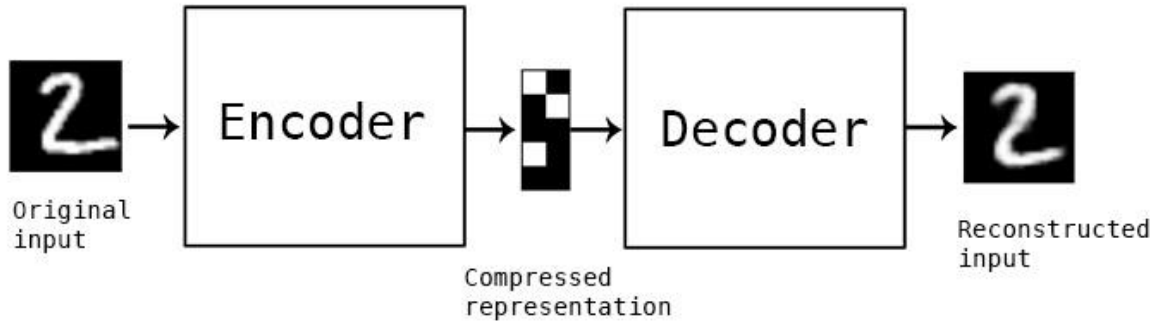


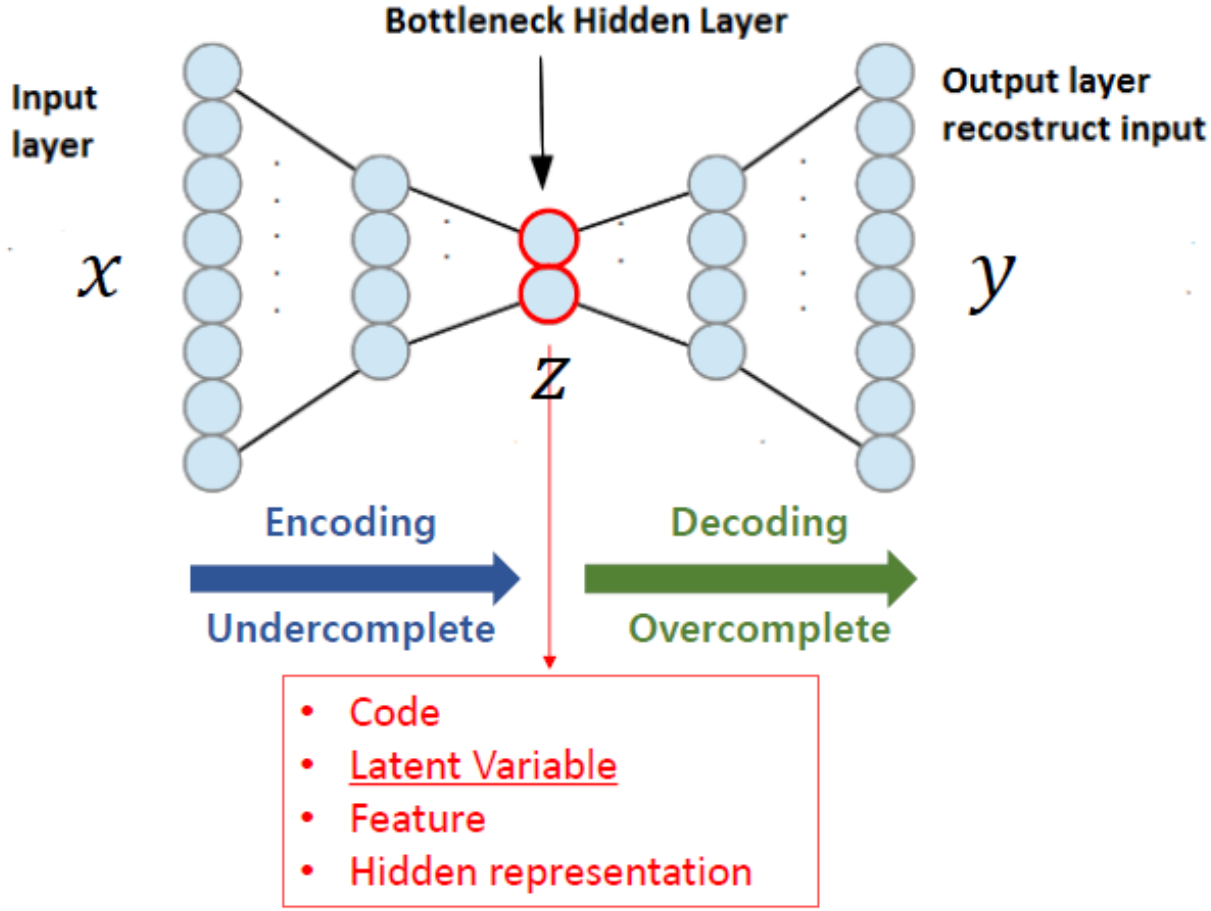
8.2.2 Autoencoder

Autoencoder를 통한 재구축 기반 이상탐지

Autoencoder 원리

- 입력된 데이터의 특성을 요약하는 인코더와 요약된 정보를 복원하는 디코더로 네트워크 구성
- Bottleneck에 해당하는 레이어는 입력 데이터에 대한 정보(representation)를 잘 축약하고 있음
- 정답 레이블 없이 학습이 가능한 비지도학습(unsupervised learning)이자 자가지도학습(self supervised learning)

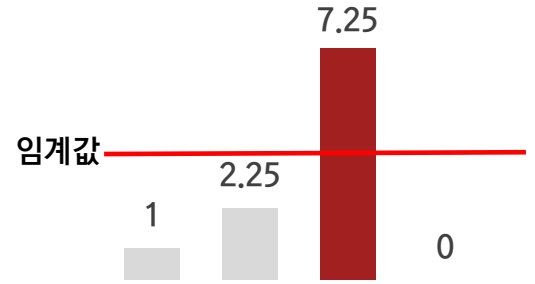




8.2.2 Autoencoder

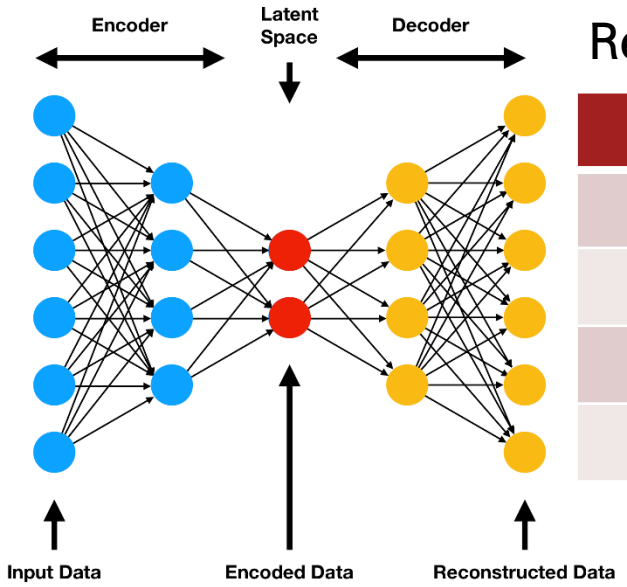
Autoencoder를 통한 재구축 기반 이상탐지

Autoencoder을 통한 이상 탐지 과정



Original Data(X)

X1	X2	X3
1	4	2
2.5	5	3
1.5	5	4
2	3	5



Reconstruction Data(X')

X1'	X2'	X3'
1	3	2
2	4	2
0	4	2
2	3	0.5

|X-X' |^2
1
2.25
7.25
0

이상탐지 결과

이상탐지
정상
정상
이상
정상

Reconstruction Error

MSE 오차 계산

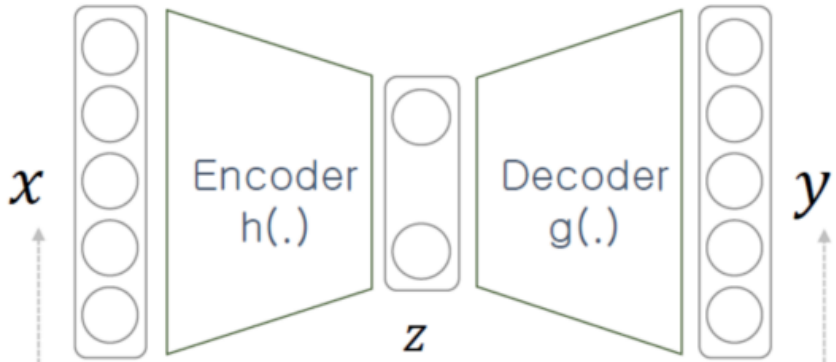
$$\frac{1}{N}\sum_{n=1}^{N}\|\mathbf{x}_n - \tilde{\mathbf{x}}_n\|$$

8.2.2 Autoencoder

Autoencoder를 통한 재구축 기반 이상탐지

Autoencoder의 네트워크 구조

- Output layer의 크기는 Input layer의 크기와 동일하게 유지
  - 일반적으로 Regression에서는 1개, Classification에서는 class 개수만큼 output layer 구성
- 입출력이 동일한 네트워크 구조 (X-> X)
  - 비지도 학습을 지도학습으로 전환 (자가지도 학습)
- Loss function은 MSE 또는 Cross-Entropy 사용
  - VAE에서는 ELBO/KL-Divergence를 사용
- 학습 데이터는 정상 데이터로만 구성
  - Latent Vector(Z)는 정상 데이터에 대한 표현 학습



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

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

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

MSE or cross-entropy

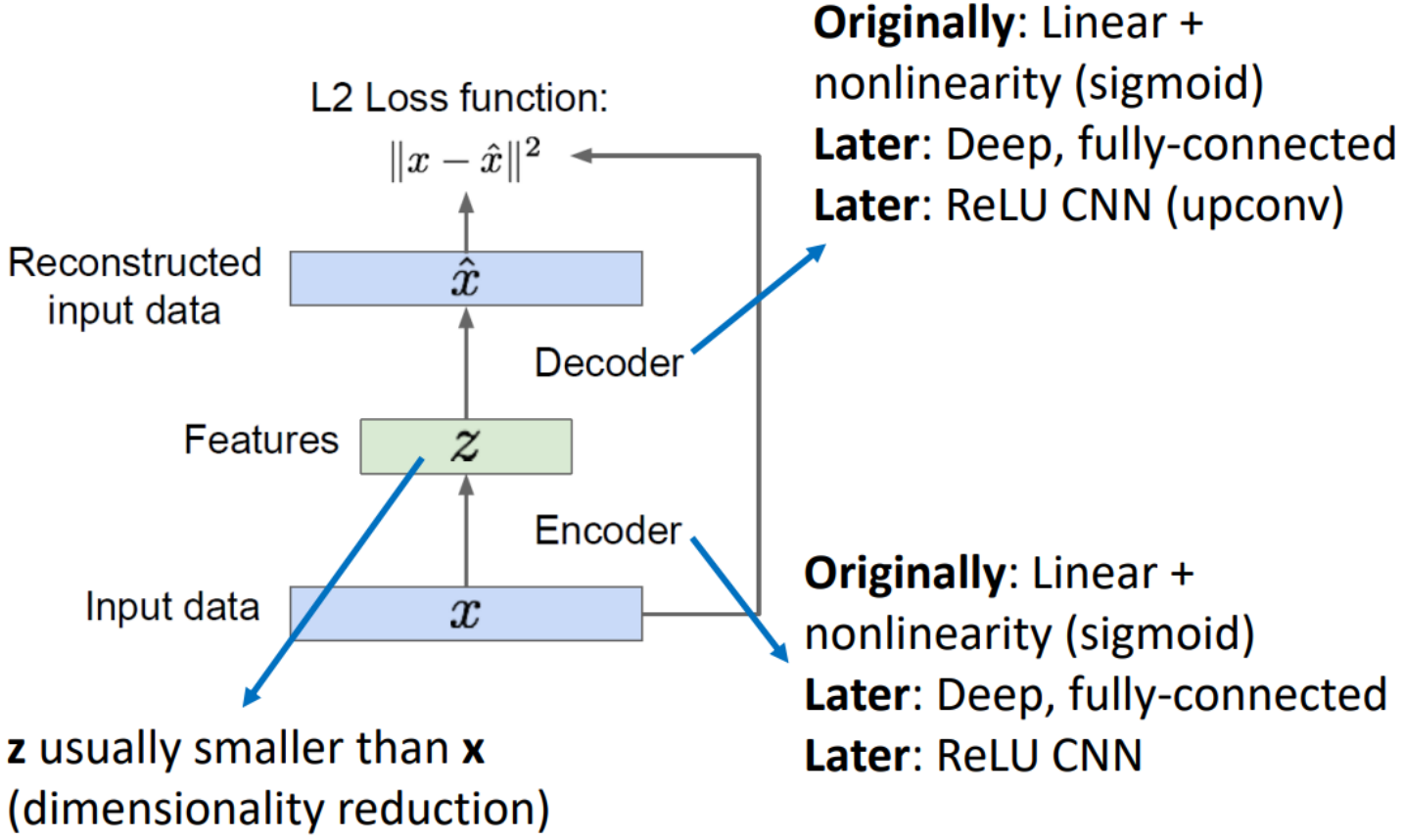
$$L_{AE} = \| x - y \|^2$$

8.2.2 Autoencoder

Autoencoder를 통한 재구축 기반 이상탐지

Autoencoder의 네트워크 구조

- 기본적으로 인코더, 디코더에 심층 신경망 네트워크(DNN) 구성
- 이상 탐지에 적용되는 데이터 특성에 따라 다양한 네트워크 사용 가능
  - 시계열 데이터 : LSTM
  - 이미지 데이터 : CNN



Special thanks to
김성범 교수님, 강남우 교수님, 이활석 CTO

EOD