

DMQA open seminar

Introduction to Autoencoder

2021. 08. 13

Data Mining & Quality Analytics Lab.

발표자: 임새린

momo_om@korea.ac.kr

Contents

1. Introduction to Autoencoder
2. Types of Autoencoders
3. Applications of Autoencoders
4. Summary

Introduction

❖ 발표자 소개



- 임새린
- 고려대학교 산업경영공학과
- Data Mining & Quality Analytics Lab. (김성범 교수님)
- M.S student (2021.03 ~)

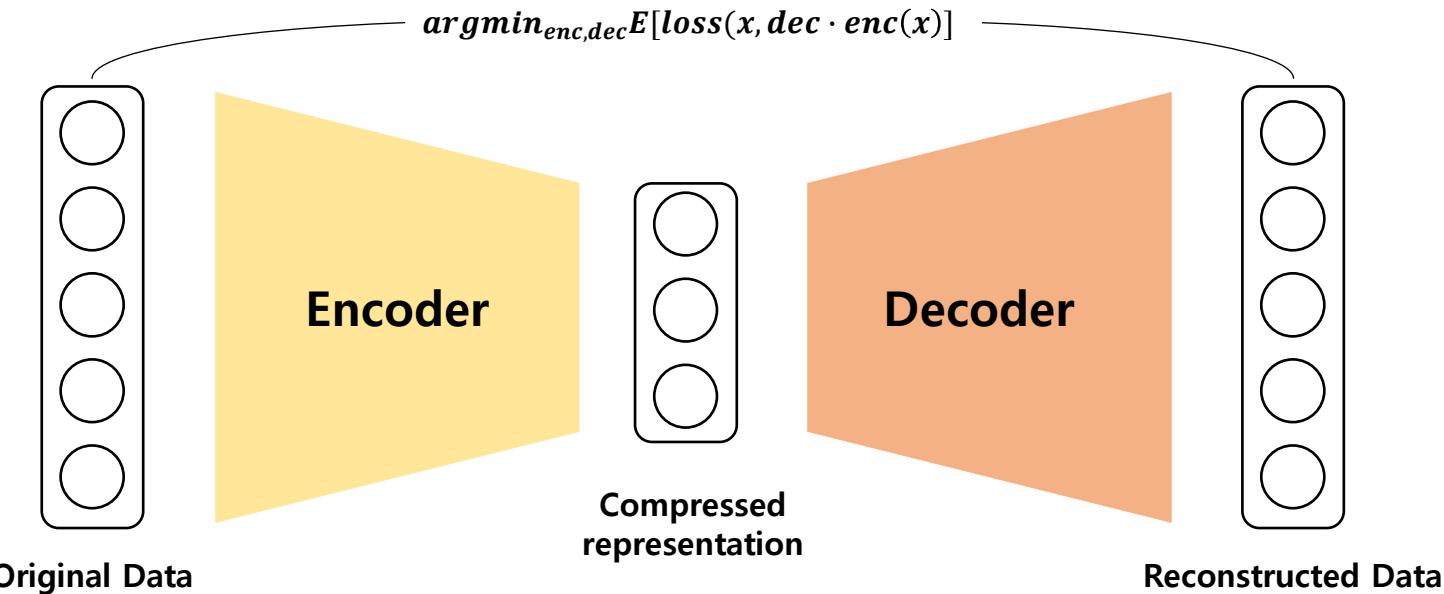
✓ 관심 연구 분야

- Machine Learning and Deep Learning
- Self-supervised Learning

Introduction

❖ Introduction to Autoencoder

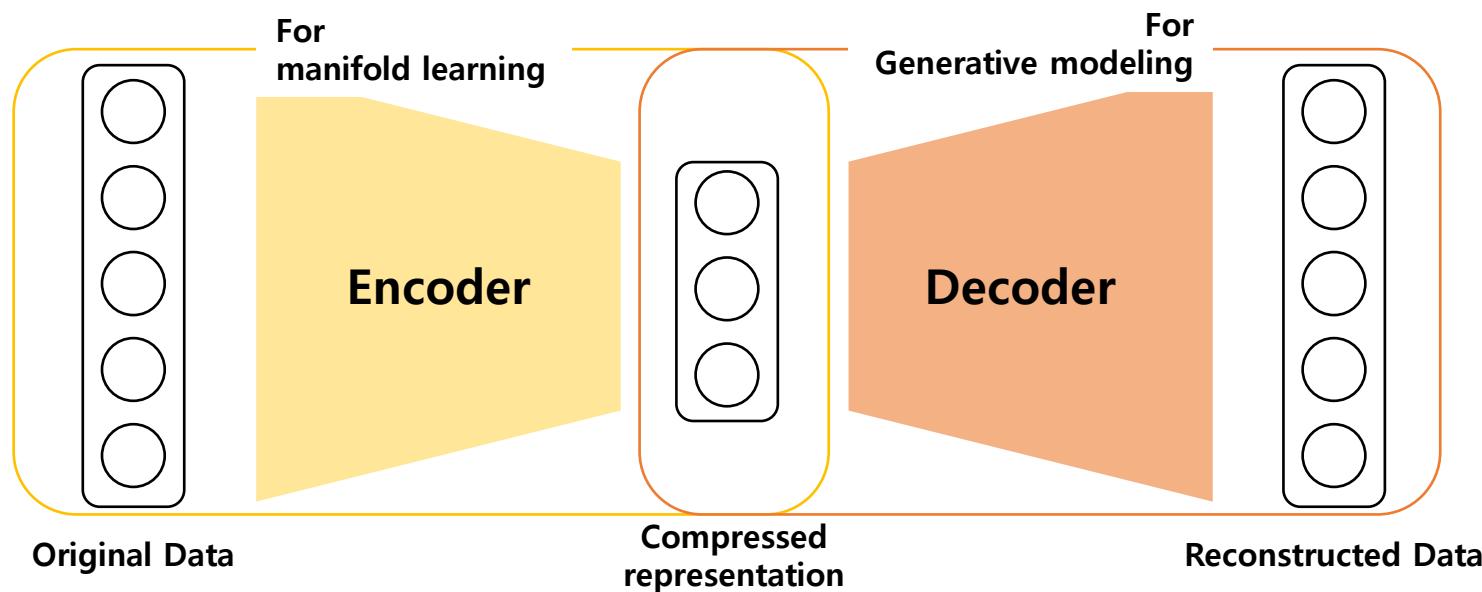
- 정의 : Unlabeled data의 효과적인 코딩을 배우는데 사용되는 인공신경망의 한 형태
- 구조 : **Encoder**와 **Decoder**가 bottleneck 구조를 이룸
- **Encoder** : 입력데이터를 meaningful하고 compressed된 representation으로 encoding
- **Decoder** : Encoding된 representation을 다시 입력데이터로 복원
- 입력 데이터만을 활용하는 Unsupervised learning으로 입력데이터와 복원된 데이터의 차 이를 최소화 하는 방향으로 학습



Introduction

❖ Autoencoder for what?

- **Autoencoder for manifold learning**
 - 입력데이터를 meaningful and compressed representation으로 mapping 해주는 **encoder**를 학습시키는 것이 목적
- **Autoencoder for generative model**
 - 어떤 latent variable을 실제 data distribution으로 mapping 해주는 **decoder**를 학습시키는 것이 목적



Introduction

❖ Manifold Learning

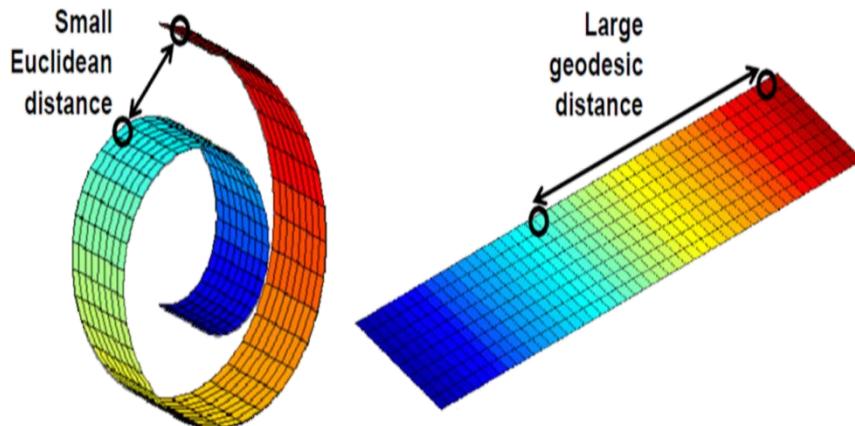
- Autoencoder for manifold learning
- Autoencoder for generative model

- **manifold**

- 두 점사이의 거리 혹은 유사도가 근거리에서는 유클리디안 거리를 따르지만 원거리에서는 그렇지 않은 공간, 고차원 공간에 내재한 저차원 공간

- **Manifold Hypothesis**

- 고차원 공간에 주어진 실제 세계의 데이터는 고차원 입력 공간 R^d 에 내재한 훨씬 저차원인 d_M 차원 매니폴드의 인근에 집중되어 있다. – Yoshua Bengio 몬트리올 대학교
 - “데이터를 자동차, 산을 고차원 공간, 도로를 매니폴드 공간이라고 생각하면 차의 대부분은 도로 위의 공간에 존재할 것이다.”



https://indico.cern.ch/event/967970/contributions/4118959/attachments/2151681/3628080/Burnaev_Manifold_Knowledge_Transfer_v2.pdf



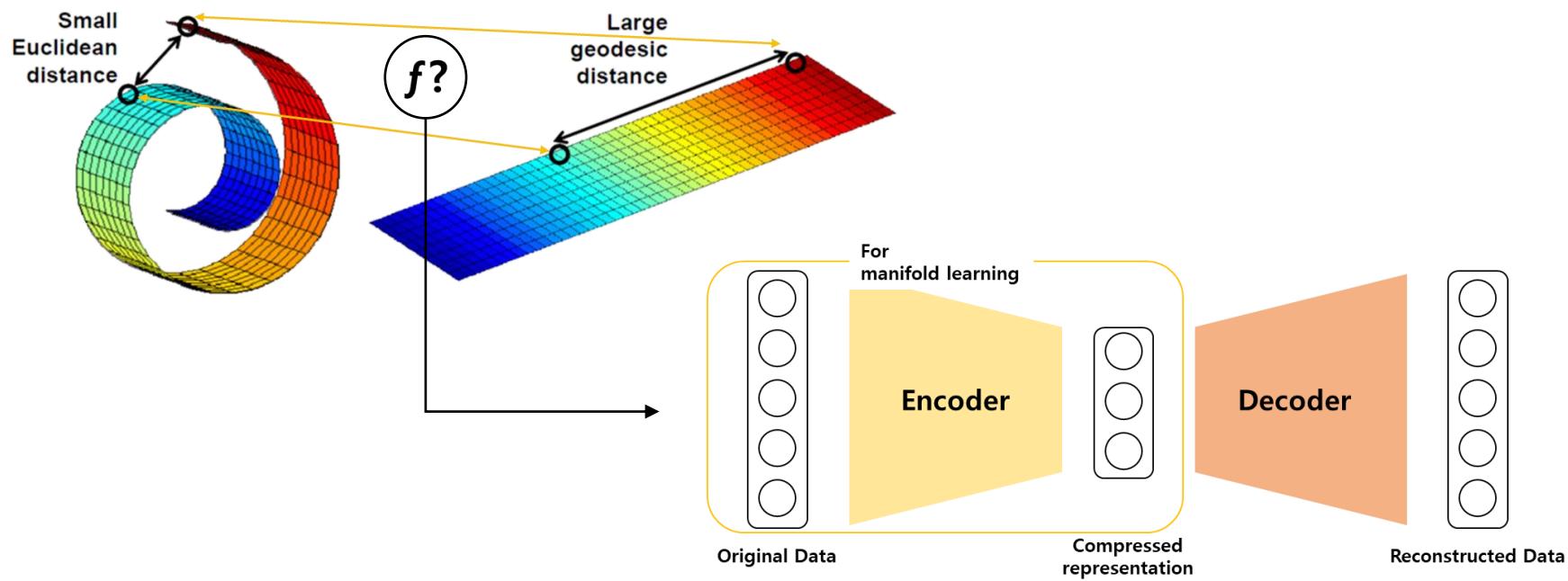
<https://www.wallpaperflare.com/sport-cars-on-wet-mountain-road-in-rainy-weather-clouds-fog-wallpaper-wkmc>

Introduction

❖ Manifold Learning

- Autoencoder for manifold learning
- Autoencoder for generative model

- 고차원 공간의 데이터를 저차원 manifold 공간으로 mapping시키는 함수를 찾는 과정
- 학습이 끝난 Autoencoder의 **encoder**를 **mapping 함수**로 사용
- 데이터를 Manifold 공간으로 mapping함으로써 **dimensionality reduction, feature extraction, representation learning** 등 역할을 수행

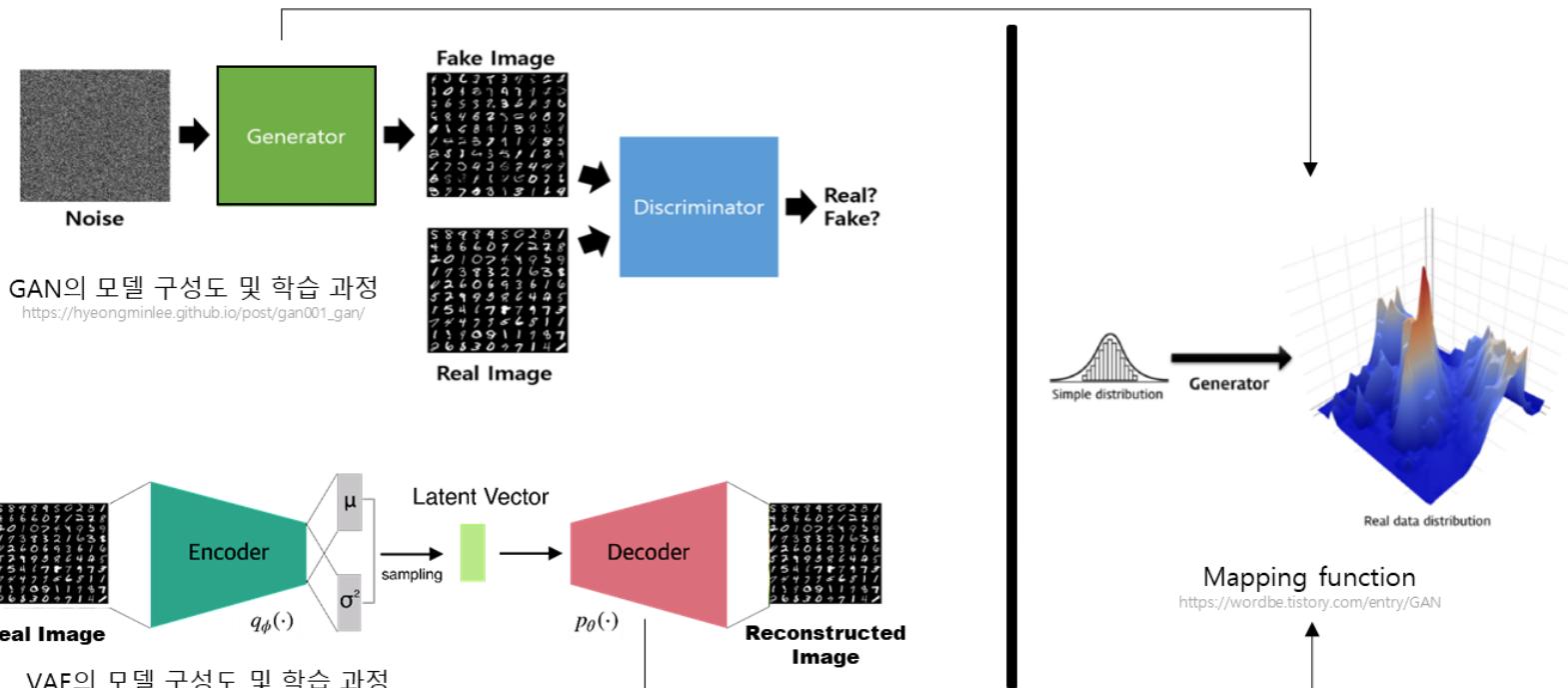


Introduction

❖ Generative Model

- Autoencoder for manifold learning
- Autoencoder for generative model

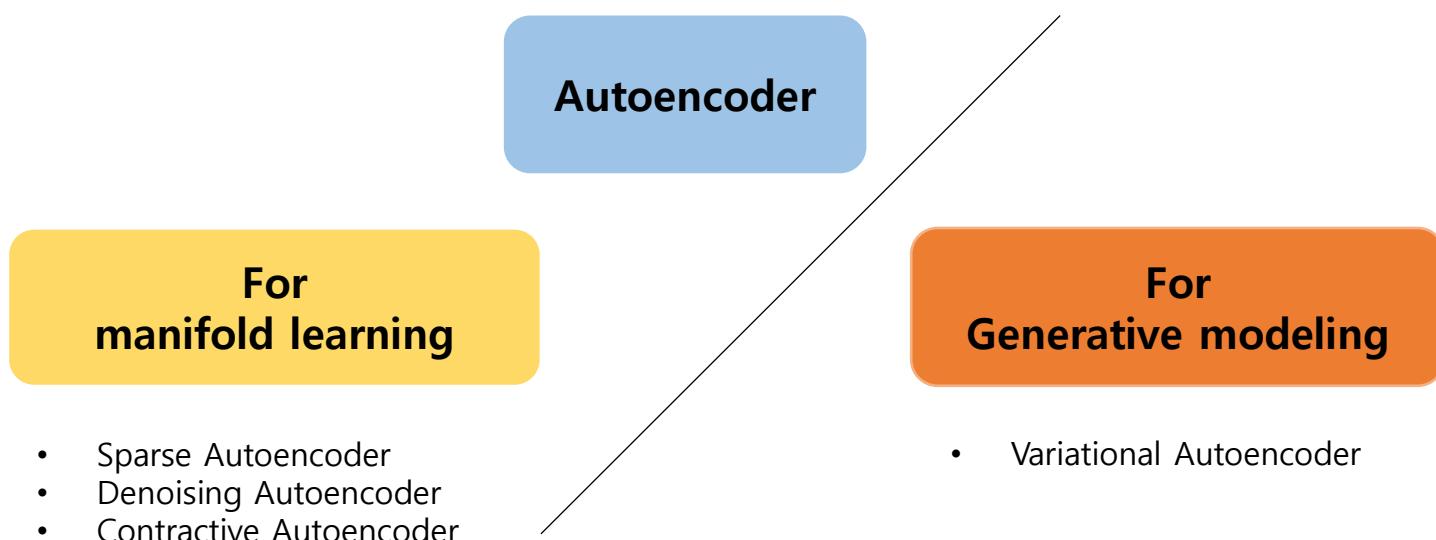
- 실제 데이터의 분포를 학습하여 데이터를 생성하는 생성모델로 대표적으로 Generative Adversarial Networks(GAN)와 Variational Autoencoder(VAE)가 있음
- Latent variable을 실제 데이터 분포로 mapping시키는 함수를 찾는 것이 목적
- 학습이 끝난 Autoencoder의 **decoder**를 **mapping 함수**로 사용



Types of Autoencoders

❖ 다양한 Autoencoder

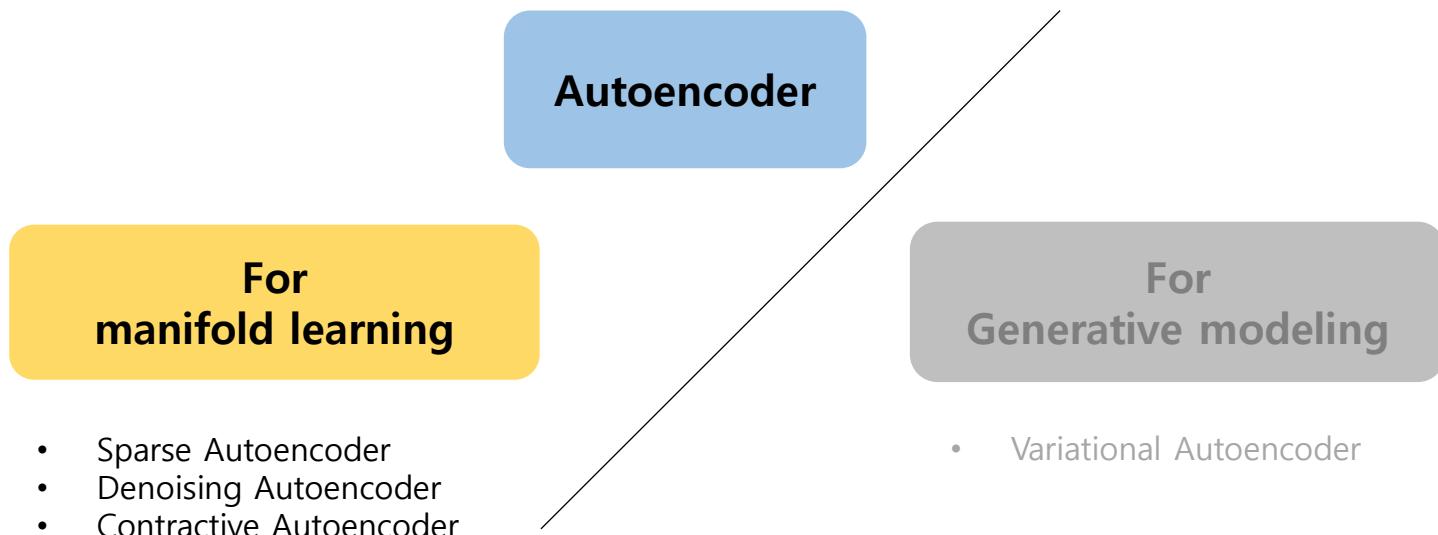
- For manifold learning : Autoencoder의 **과적합 문제를 해결하기** 위해 다양한 규제 기법을 적용하는 과정에서 여러 변형이 발생
- For generative modeling : Variational Autoencoder로써 Autoencoder의 정의와 다르게 생성 모델을 학습하는 과정에서 모델의 구조가 autoencoder와 유사해짐
- 실제로 generative model인 VAE는 Autoencoder와 기원이 다르며 이름과 구조가 비슷하여 혼동하기 쉬움



Types of Autoencoders

❖ 다양한 Autoencoder

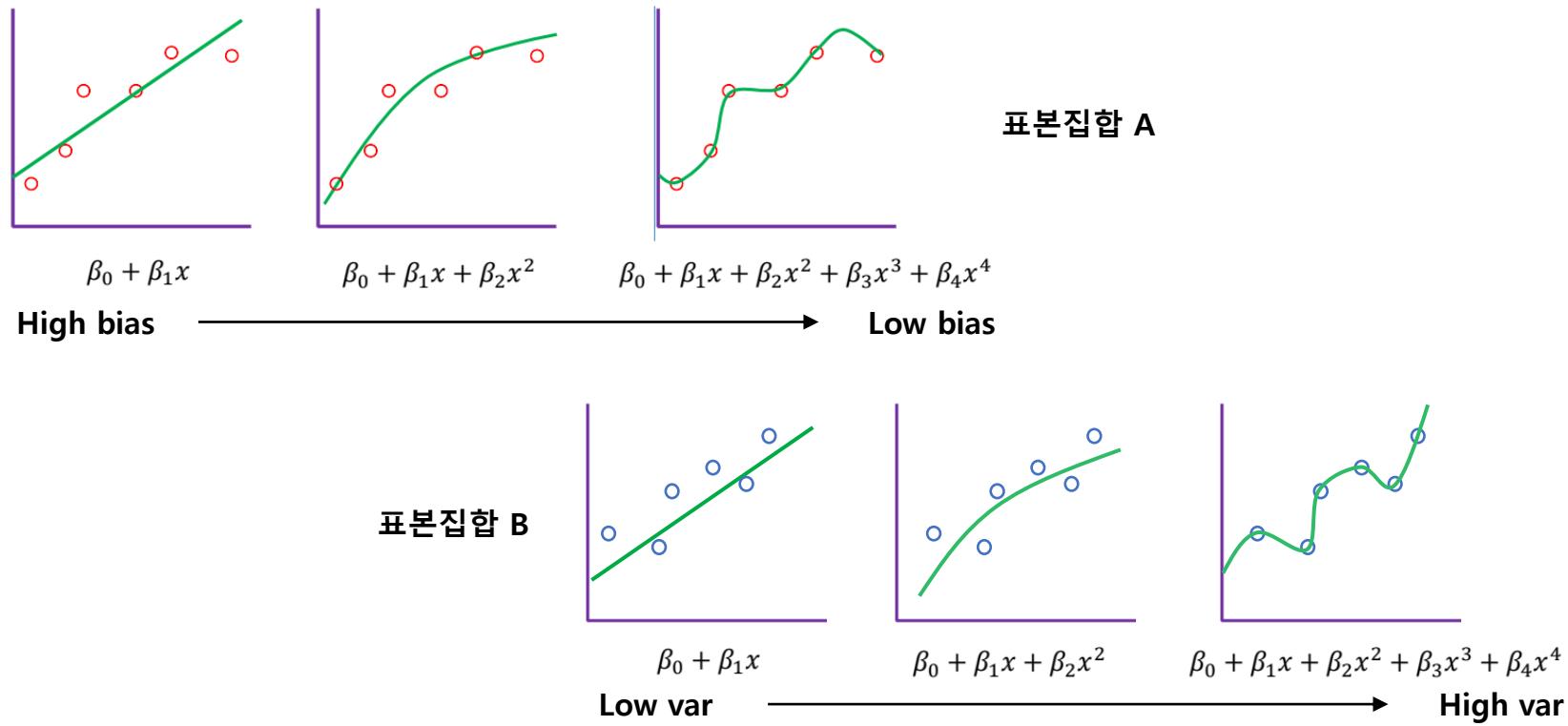
- For manifold learning : Autoencoder의 **과적합 문제를 해결하기** 위해 다양한 규제 기법을 적용하는 과정에서 여러 변형이 파생
- For generative modeling(VAE) : Autoencoder의 정의와 다르게 생성모델을 학습하는 과정에서 모델의 구조가 autoencoder와 유사해 짐
- 실제로 generative model인 VAE는 Autoencoder와 유래가 다르며 이름과 구조가 비슷하여 혼동하기 쉬움



Types of Autoencoders : for manifold learning

❖ Bias-variance tradeoff in autoencoder

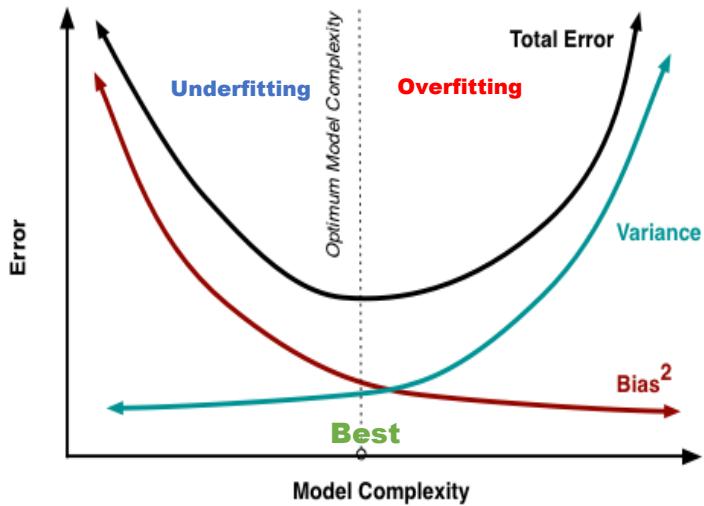
- $Total\ Error = Bias^2 + Variance + Irreducible\ Error$
- Bias : 실제 데이터를 표현하는 모델과 가정한 모델의 차이에서 발생하는 오류, 실제값과 평균 예측값의 차이
- Variance : 모델링에 사용되는 여러 표본 데이터 집합에 대한 추정을 할 때 발생하는 오류, 예측값에 대한 분산
- Irreducible Error : 줄일 수 없는 자연적인 오류



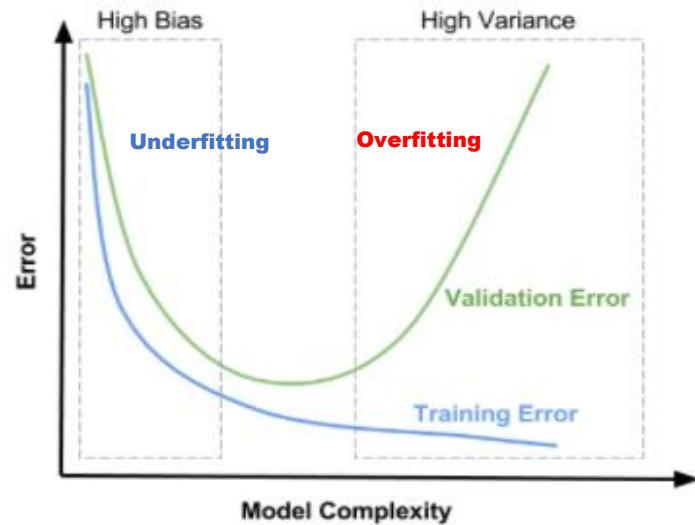
Types of Autoencoders : for manifold learning

❖ Bias-variance tradeoff in autoencoder

- $Total\ Error = Bias^2 + Variance + Irreducible\ Error$
- Bias-variance tradeoff : Model complexity(=capability) 관점에서 반비례 관계를 가짐
- 훈련집합과 검증집합 관점에서 훈련집합에 대한 오류율이 너무 작아지면 검증집합에 대한 variance가 커져 다른 표본집합에 대한 일반화 성능이 떨어지는 **overfitting** 발생



<http://scott.fortmann-roe.com/docs/BiasVariance.html>

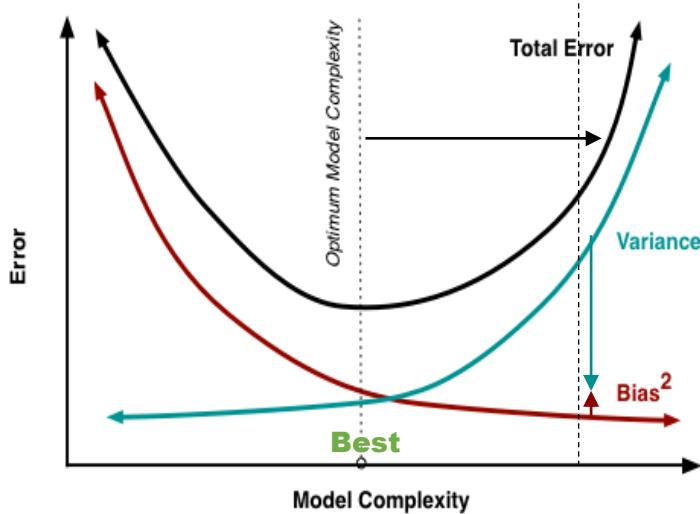


<https://opentutorials.org/module/3653/22071>

Types of Autoencoders : for manifold learning

❖ Bias-variance tradeoff in autoencoder

- 데이터에 알맞은 모델의 복잡도를 찾는 것은 매우 어려움. 딥러닝은 모델의 복잡도를 높인 뒤, 다양한 규제 기법을 통해서 bias를 높이고 variance를 낮추는 방향으로 발전
- Autoencoder가 입력 데이터를 잘 복원하길 바라는 동시에(low bias) 처음보는 데이터(테스트)도 아무를 수 있는 representation을 학습하길(low variance) 바람



<http://scott.fortmann-roe.com/docs/BiasVariance.html>

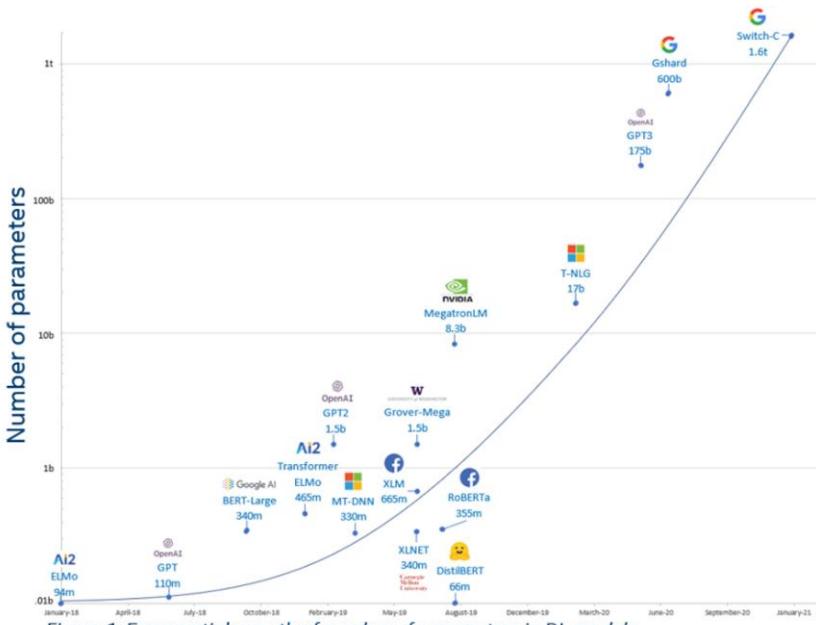


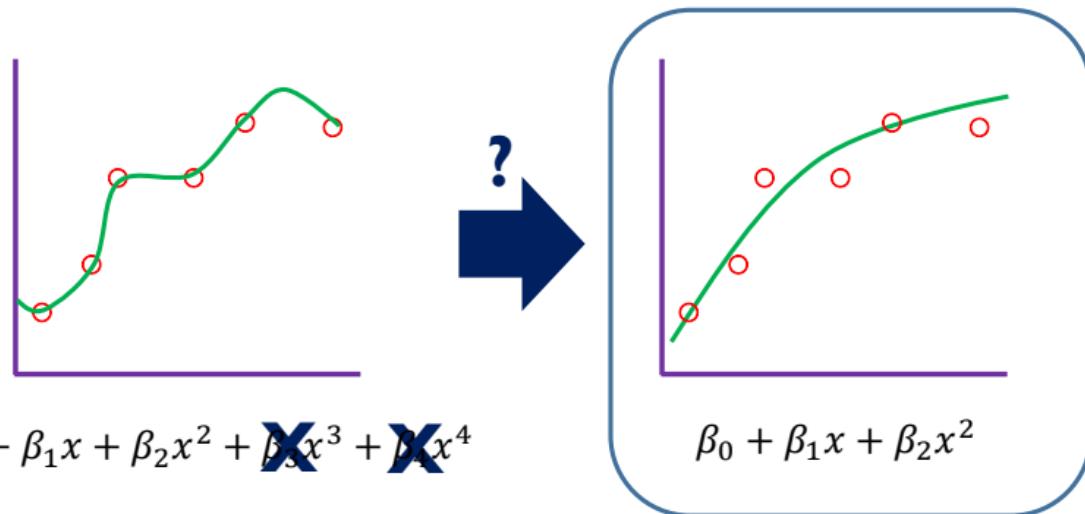
Figure 1: Exponential growth of number of parameters in DL models

<https://www.microsoft.com/en-us/research/blog/turing-nlg-a-17-billion-parameter-language-model-by-microsoft/>

Types of Autoencoders : for manifold learning

❖ Bias-variance tradeoff in autoencoder

- 데이터에 알맞은 모델의 복잡도를 찾는 것은 매우 어려움. 딥러닝은 모델의 복잡도를 높인 뒤, 다양한 규제 기법을 통해서 bias를 높이고 variance를 낮추는 방향으로 발전
- Autoencoder가 입력 데이터를 잘 복원하길 바라는 동시에(low bias) 처음보는 데이터(테스트)도 아우를 수 있는 representation을 학습하길(low variance) 바람



Types of Autoencoders : for manifold learning

❖ Bias-variance tradeoff in autoencoder

- 현재의 딥러닝은 모델의 복잡도를 높인 뒤, 다양한 규제 기법을 통해서 bias를 높이고 variance를 낮추는 방향으로 발전
- Autoencoder가 입력 데이터를 잘 복원하길 바라는 동시에(low bias) 처음보는 데이터(테스트)도 아우를 수 있는 representation을 학습하길(low variance) 바람
- 다른 인공신경망과 마찬가지로 모델의 복잡도를 높이고 규제 기법을 적용하는 형태의 다양한 파생 autoencoder 모델이 개발

For
manifold learning

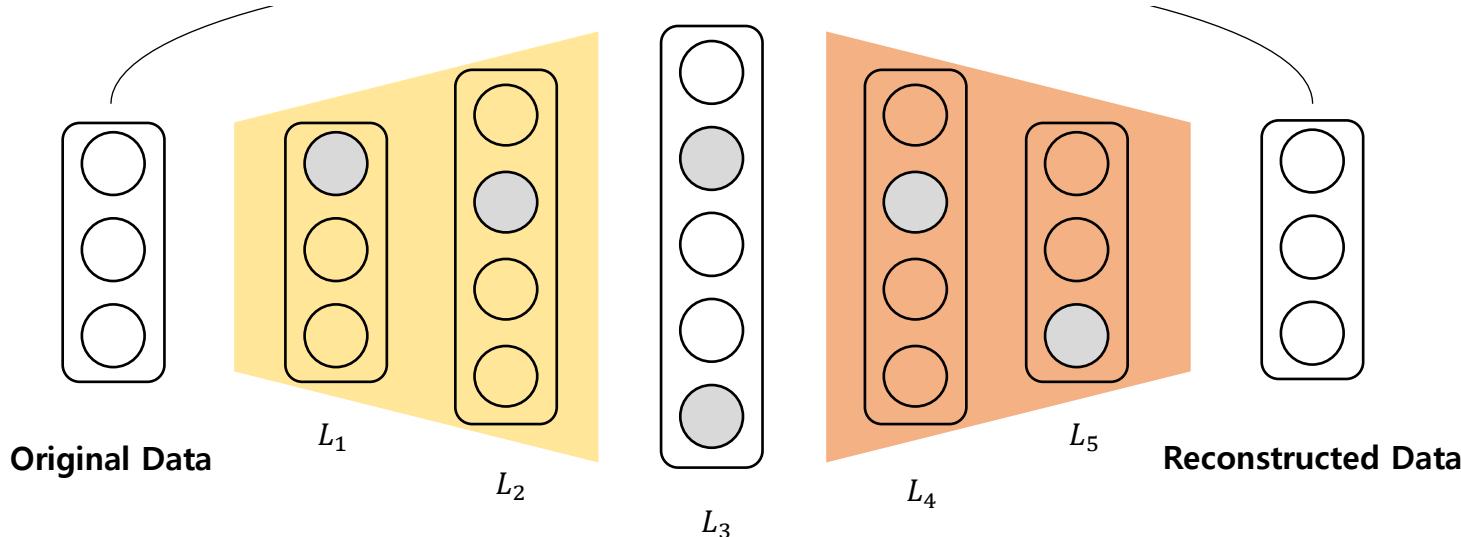
- **Sparse Autoencoder : AE + Sparsity regularization**
- **Denoising Autoencoder : AE + Random noise**
- **Contractive Autoencoder : AE + Jacobian Matrix**

Types of Autoencoders : for manifold learning

❖ Sparse Autoencoder

- Hidden layer의 node 수가 input layer의 node 수보다 많은 overcomplete autoencoder 구조
- Sparsity parameter를 제어하여 **은닉층의 활성화에 규제를 가하는 방법**
- ρ 와 특정 node의 활성화값 $a_j^{(h)}$ 를 관측치에 대해 평균을 낸 $\hat{\rho}_j$ 가 같아지도록 제약
- $\hat{\rho}_j = \frac{1}{m} \sum_{i=1}^m [a_j^{(2)}(x^{(i)})]$ (m : 관측치의 수)
- KL-divergence나 L₁-regularization을 Loss term에 추가하여 사용

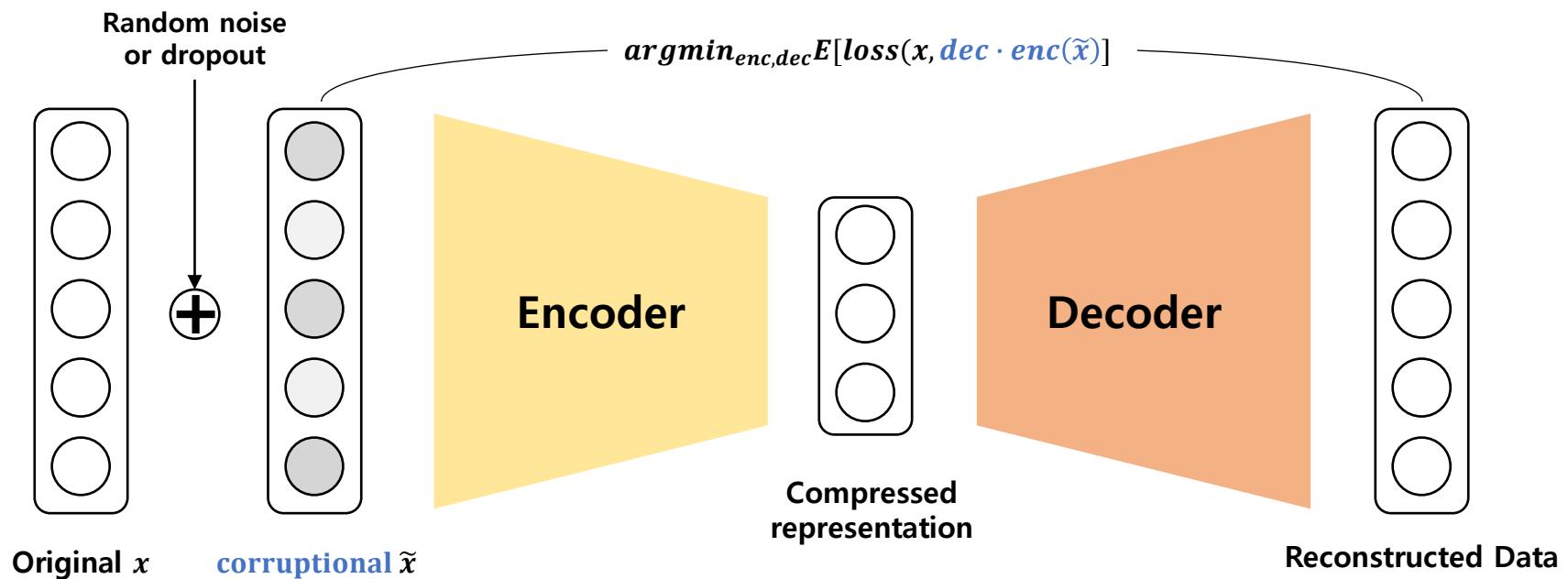
$$\operatorname{argmin}_{enc, dec} E[\text{loss}(x, dec \cdot enc(x))] + \sum_{h=1}^5 \sum_{j=1}^{\# \text{ of } L_h} KL(\rho \mid \mid \hat{\rho}_j)$$



Types of Autoencoders : for manifold learning

❖ Denoising Autoencoder

- 입력 데이터에 **random noise**나 **dropout**을 추가하는 규제기법 적용
- 입력 데이터에 어떠한 noise를 부여하더라도 **manifold** 상에서 같은 곳에 위치해야 한다는 가정
- 입력 데이터에 작은 변화를 주어 손상된 데이터를 만들고 모델을 통해 손상되지 않은 데이터를 출력하는 방법으로 이를 통해 **작은 변화에 대해 덜 민감한, 강건한 모델**을 만들 수 있음

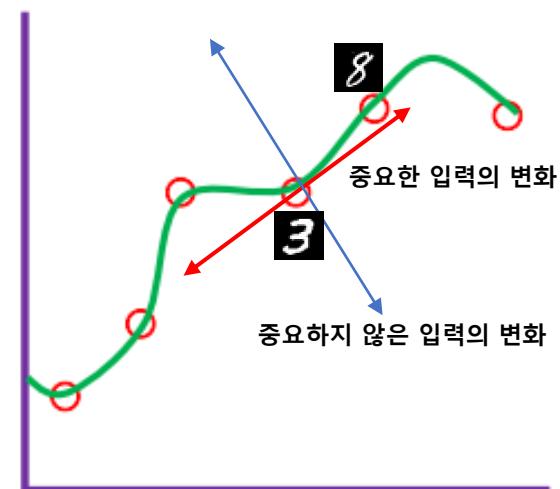
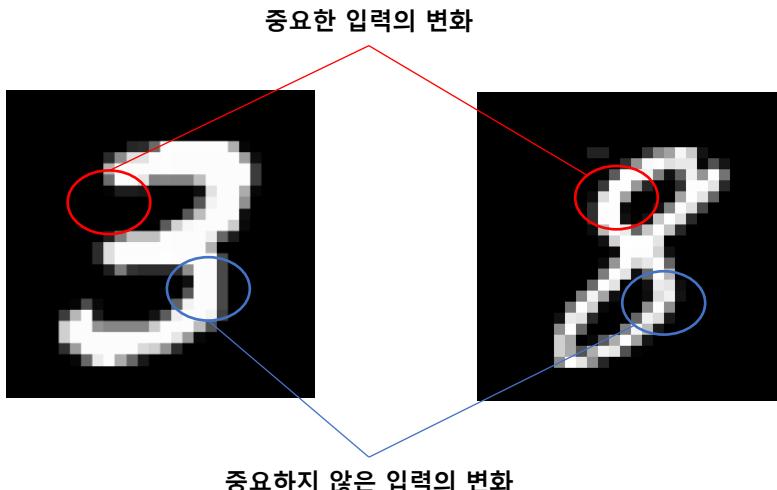


Types of Autoencoders : for manifold learning

❖ Contractive Autoencoder

- Denoising Autoencoder와 같이 작은 변화에 강건한 모델을 학습하는 것이 목적
- DAE에서는 Encoder가 **입력 데이터의 작은 변화**에 저항하도록 하는 데 중점
- Contractive Autoencoder에서는 Encoder가 Decoder에서 재구성할 때 **중요하지 않은 입력의 변화**를 무시하도록 하여 특징을 추출할 때 작은 변화에 덜 민감하도록 중점을 둠
- Encoder의 Jacobian matrix를 손실함수에 추가하여 구현
- New loss function : $\text{argmin}_{enc, dec} E[\text{loss}(x, dec \cdot enc(x))] + \lambda \|\nabla_x enc(x)\|_2^2$

$$\text{where, } \|\nabla_x enc(x)\|_2^2 = \sum_j \sum_k \left(\frac{\partial enc(x)_j}{\partial x_k} \right)^2$$



Applications of autoencoders

❖ Use of autoencoders

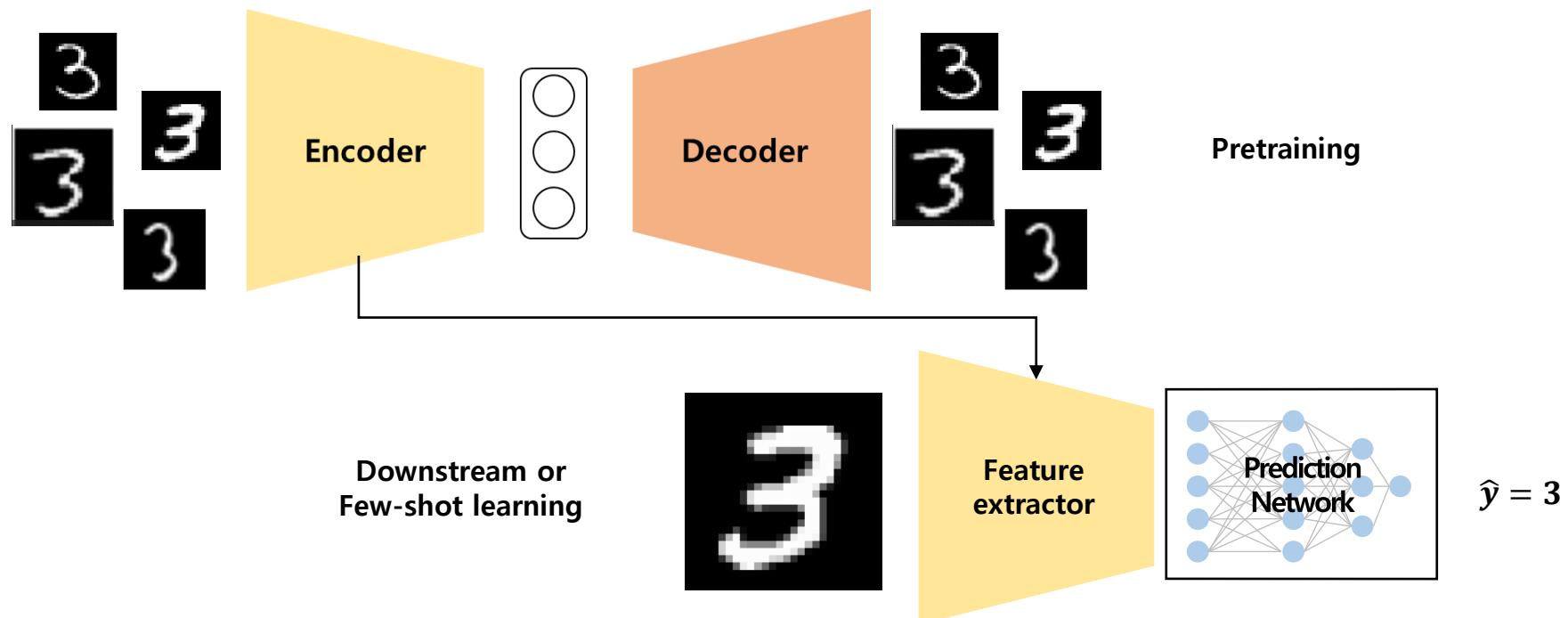
- **Autoencoder for manifold learning**
 - 입력데이터를 **meaningful and compressed representation**으로 바꿔주는 **encoder**를 학습시키는 것이 목적
- **Classification**
- **Clustering**
- **Anomaly detection**
- **Dimensionality reduction**



Applications of autoencoders

❖ Use of autoencoders for classification

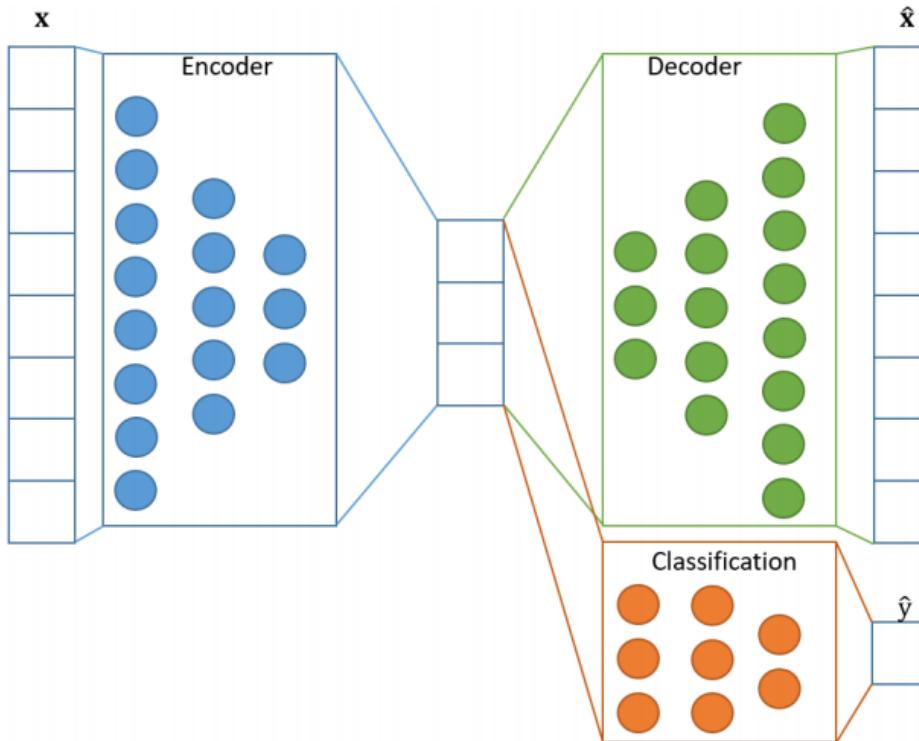
- Autoencoder를 이용해 학습시킨 Encoder를 **Feature extractor**로 사용하여 self-supervised learning 혹은 semi-supervised learning에 활용
- 같은 Class를 가지는 관측치들은 드러나지 않았지만 autoencoder를 통해 이를 근사할 수 있는 **latent structure**가 있을 것(**meaningful representation**)이라 가정



Applications of autoencoders

❖ Use of autoencoders for classification

- Autoencoder를 통한 reconstruction을 규제 기법으로 활용
- 일반적인 classification loss term에 reconstruction loss를 추가하여 구현
- $Loss = Classification\ Loss(y, \hat{y}) + \lambda \cdot Reconstruction\ Loss(x, \hat{x})$



Applications of autoencoders

❖ Use of autoencoders for classification

- **SiT : Self-supervised Vision Transformer**
- Vision Transformer를 활용한 self-supervised learning 방법론 개발
- Transformer의 구조적 이점을 활용하여 마치 Autoencoder처럼 활용한 논문

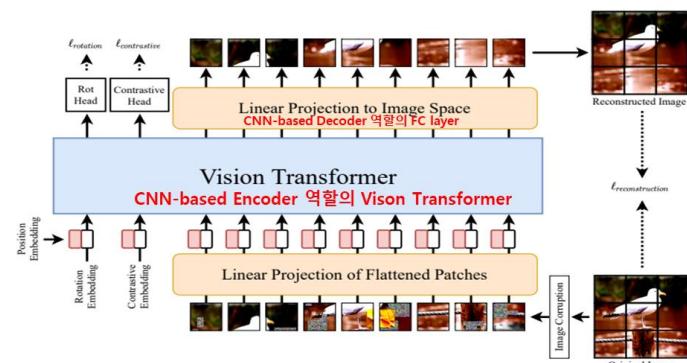
SiT: Self-supervised vlsion Transformer

Sara Atito, Member IEEE, Muhammad Awais, and Josef Kittler, Life Member, IEEE

Abstract—

Self-supervised learning methods are gaining increasing traction in computer vision due to their recent success in reducing the gap with supervised learning. In natural language processing (NLP) self-supervised learning and transformers are already the methods of choice. The recent literature suggests that the transformers are becoming increasingly popular also in computer vision. So far, the vision transformers have been shown to work well when pretrained either using a large scale supervised data [1] or with some kind of co-supervision, e.g. in terms of teacher network. These supervised pretrained vision transformers achieve very good results in downstream tasks with minimal changes [1], [2], [3]. In this work we investigate the merits of self-supervised learning for pretraining image/vision transformers and then using them for downstream classification tasks. We propose Self-supervised vision Transformers (SiT) and discuss several self-supervised training mechanisms to obtain a pretext model. The architectural flexibility of SiT allows us to use it as an autoencoder and work with multiple self-supervised tasks seamlessly. We show that a pretrained SiT can be finetuned for a downstream classification task on small scale datasets, consisting of a few thousand images rather than several millions. The proposed approach is evaluated on standard datasets using common protocols. The results demonstrate the strength of the transformers and their suitability for self-supervised learning. We outperformed existing self-supervised learning methods by large margin. We also observed that SiT is good for few shot learning and also showed that it is learning useful representation by simply training a linear classifier on top of the learned features from SiT. Pretraining, finetuning, and evaluation codes will be available under: <https://github.com/Sara-Ahmed/SiT>.

Index Terms—Vision Transformer, Self-supervised Learning, Discriminative Learning, Image Classification, transformer based autoencoders.



Applications of autoencoders

main 1 branch 0 tags Go to file Code

dudwojae Update README.md 40cac62 yesterday 55 commits

VisionTransformer Add files via upload 6 days ago

README.md Update README.md yesterday

README.md

NeverMind_DMQA

This repository contains key summary and description files for various papers.

Vision Transformer

Key Summary and Description of Paper on Vision Transformer

- 2021-07-02
 - Young Jae Lee / Visual Transformers: Token-based Image Representation and Processing for Computer Vision [paper][presentation]
 - Saerin Lim / TransGAN: Two Transformers Can Make One Strong GAN [paper][presentation]
- 2021-07-09
 - Jaehoon Kim / Rethinking Spatial Dimensions of Vision Transformers [paper][presentation]
 - Jongkook Heo / Emerging Properties in Self-Supervised Vision Transformers [paper][presentation]
- 2021-07-16
 - Yongwon Jo / Inpainting Transformer for Anomaly Detection [paper][presentation]
 - Eunji Koh / Training Data-Efficient Image Transformers & Distillation through Attention [paper][preser
- 2021-07-23
 - Jinhyeok Park / End-to-End Object Detection with Transformers [paper][presentation]
 - Leekyung Yoo / Mlp-Mixer: An All-Mlp Architecture for Vision [paper][presentation]
- 2021-08-06
 - Young Jae Lee / Swin Transformer: Hierarchical Vision Transformer using Shifted Windows [paper][presentation]
 - Sae Rin Lim / SiT: Self-supervised vision Transformer [paper][presentation]

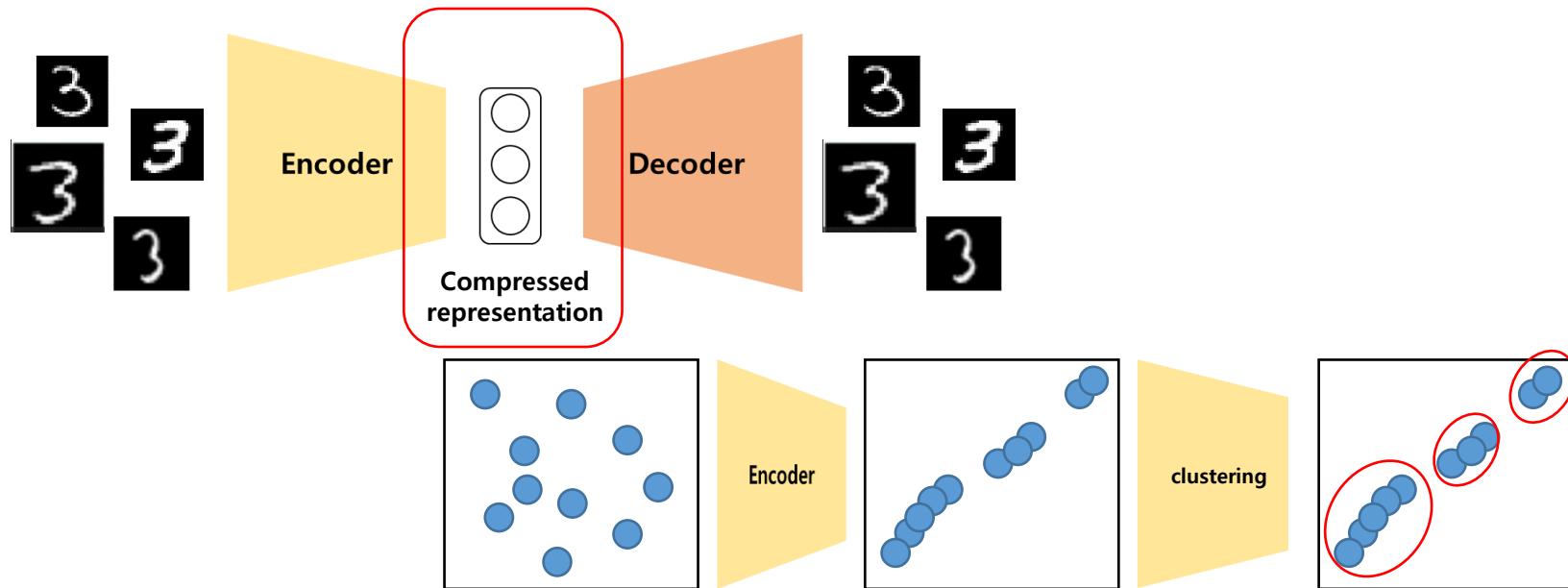
https://github.com/dudwojae/NeverMind_DMQA



Applications of autoencoders

❖ Use of autoencoders for clustering

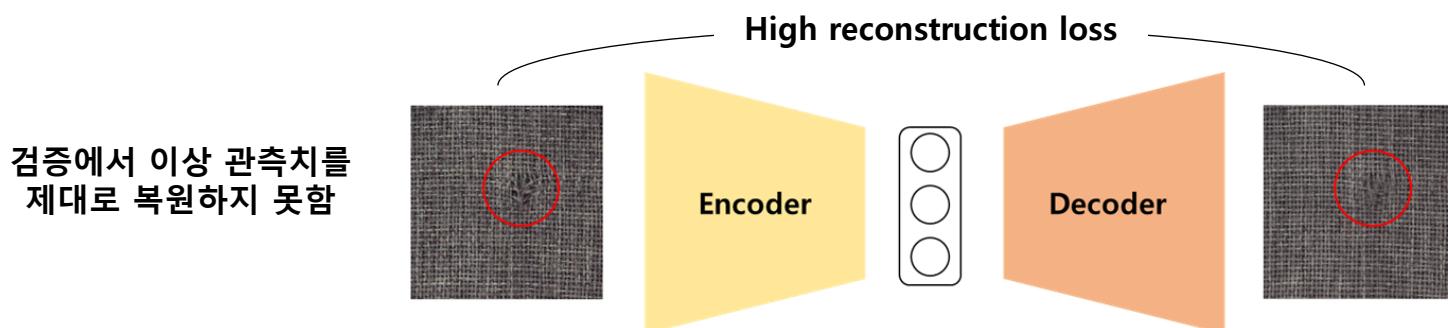
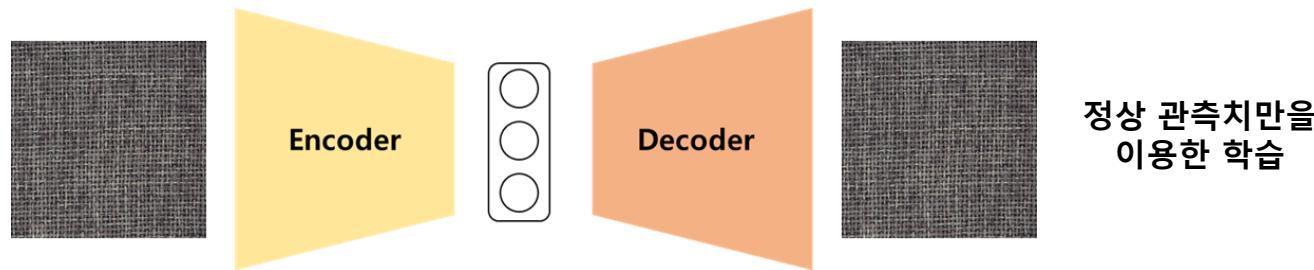
- 비슷한 Unlabeled data간에는 숨겨진 **row-dimensional latent representation**(manifold)가 존재 할 것이라는 가정
- Autoencoder를 이용해 학습시킨 Encoder를 통해 Unlabeled data를 Embedding
- **Embedding된 관측치**를 이용해 기존의 군집화 알고리즘 수행
- Autoencoder의 학습이 군집화가 아닌 reconstruction이 주 목적이기에 **군집화를 위한 추가적인 기법들이 적용**(ex. 군집 중심과 embedding된 관측치의 거리를 loss term에 추가)



Applications of autoencoders

❖ Use of autoencoders for anomaly detection

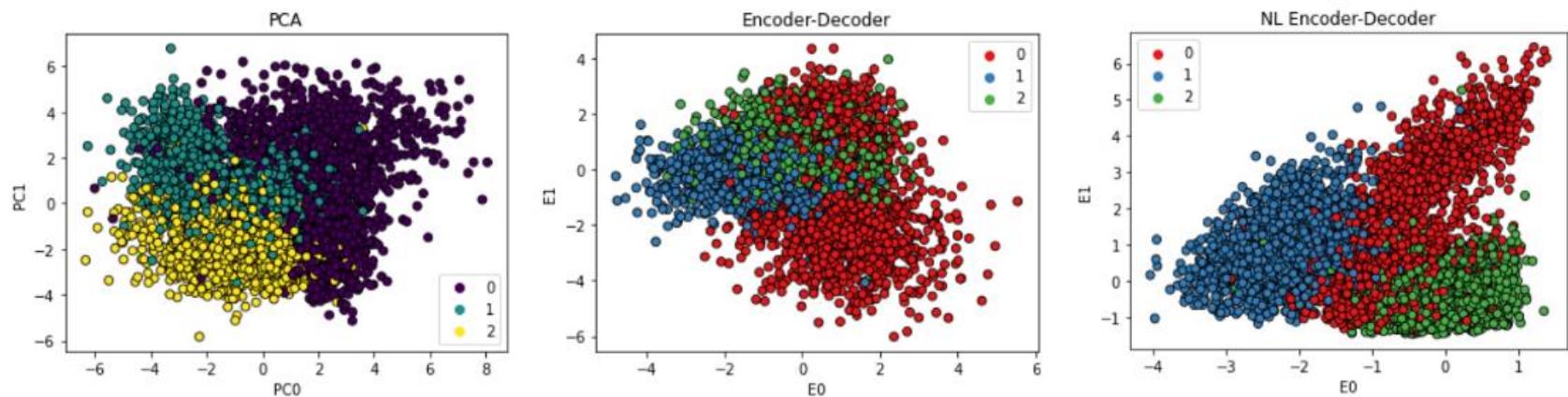
- 정상 관측치와 이상 관측치 간에 latent structure가 다를 것이라 가정
- 정상 관측치만을 학습데이터로 사용함으로써 **Autoencoder**가 이상 관측치는 잘 복원하지 못할 것이라는 기대
- 검증과정에서 비정상 관측치가 입력된다면 reconstruction loss가 클 것이라고 예상



Applications of autoencoders

❖ Use of autoencoders for dimensionality reduction

- 실제 데이터들(이미지, 텍스트)들은 매우 높은 차원의 sparse한 벡터로 표현되는 경우가 많음
- 이러한 고차원 벡터들을 통해 학습을 하면 *curse of dimensionality*에 빠지게 될 수 있음
- 이러한 데이터들을 훨씬 낮은 차원의 "intrinsic dimensionality(본질적 차원)"으로 mapping시키는 것이 목적
- 고전적인 방법으로 관측치를 낮은 차원으로 linear projection하는 PCA가 있으며 여기에 비선형 변환을 추가한 것이 Autoencoder로 해석할 수 있음
- Autoencoder의 encoder를 mapping function으로 사용함으로써 구현



<https://ichi.pro/ko/autoencoderwa-pcaleul-tonghan-chawon-chugso-268789370849091>

Summary

❖ Summary of Autoencoder

- Autoencoder는 **manifold learning**을 위한 부류와 **generative learning**을 위한 부류로 나뉨
- Manifold learning을 통해 representation을 학습할 때 여러 규제 기법을 사용함으로써 다양한 Autoencoder가 등장
- Autoencoder의 종류
 - ✓ **Sparse AE** : Sparse parameter를 이용하여 노드가 활성화 될 확률을 제어함으로써 규제
 - ✓ **Denoising AE** : Noise를 추가한 손상된 입력데이터를 복원함으로써 noise에 강건한 모델을 학습
 - ✓ **Contractive AE** : Jacobian matrix 활용해 중요하지 않은 입력의 변화를 버림으로써 강건한 모델을 학습
- Autoencoder의 활용
 - ✓ **Classification** : Semi or self-supervised learning에서 data의 latent structure를 찾는데 이용
 - ✓ **Clustering** : Data를 row-dimensional latent representation로 mapping 시키는데 이용
 - ✓ **Anomaly detection** : 정상만을 학습시킨 AE가 이상치를 잘 복원하지 못할 것이라는 가정으로 이용
 - ✓ **Dimensionality reduction** : 고차원의 Sparse data를 intrinsic dimension으로 mapping 시키는데 이용

Reference

1. Bank, D., Koenigstein, N., & Giryes, R. (2020). Autoencoders. *arXiv preprint arXiv:2003.05991*.
2. Atito, S., Awais, M., & Kittler, J. (2021). Sit: Self-supervised vision transformer. *arXiv preprint arXiv:2104.03602*.
3. <https://d2.naver.com/news/0956269>
4. https://github.com/dudwojae/NeverMind_DMQA
5. <https://www.youtube.com/watch?v=pJCcGK5omhE&list=PLpIPLT0Pf7IoTxTCi2MEQ94MZnHaxrP0j&index=21>

감사합니다