

딥러닝 세미나 Season #11

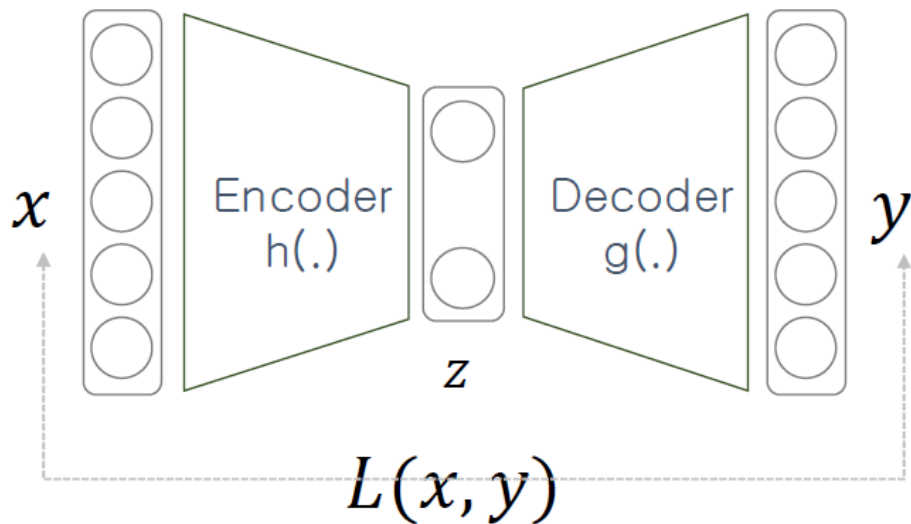
Auto Encoder

조건희

Introduction

- Auto encoder

- 입력과 출력이 동일한 네트워크 구조
- Unsupervised learning 을 Supervised learning 으로 바꾸어 해결
- z - code, latent variable, feature, hidden representation



- 키워드

- Unsupervised learning - 비지도 학습
- Manifold learning (차원 축소)
 - 학습된 오토인코더의 인코더는 차원 축소 역할을 수행
- Generative model learning
 - 학습된 오토인코더의 디코더는 생성 모델 역할을 수행

- 매니폴드 러닝

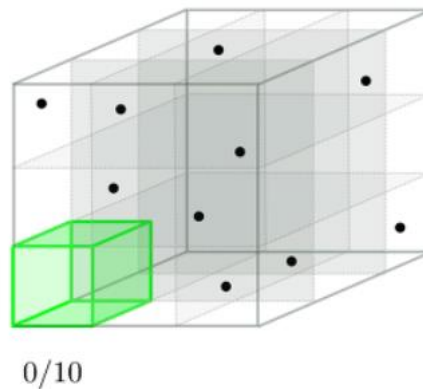
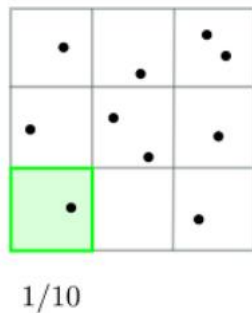
- 데이터 차원(dimension)을 축소(reduction)하는 방법 중 하나
- 데이터 차원을 왜 축소하나?
 - 데이터 압축 - 데이터를 줄이자
 - 데이터 시각화 - 우리는 꼭해야 4차원까지 밖에 구분 못하니까..
 - 차원의 저주
 - 주요 feature의 발견

Manifold learning

▪ 매니폴드 러닝

▪ 차원의 저주

- 데이터 차원이 증가할수록 데이터 공간의 부피가 기하급수적으로 증가하기 때문에, 동일한 개수의 데이터의 밀도는 차원 증가에 따라 급속히 희소(sparse)해짐.
- 따라서, 차원이 증가할수록 데이터의 분포 분석 또는 모델 추정에 필요한 샘플데이터의 개수가 기하급수적으로 증가

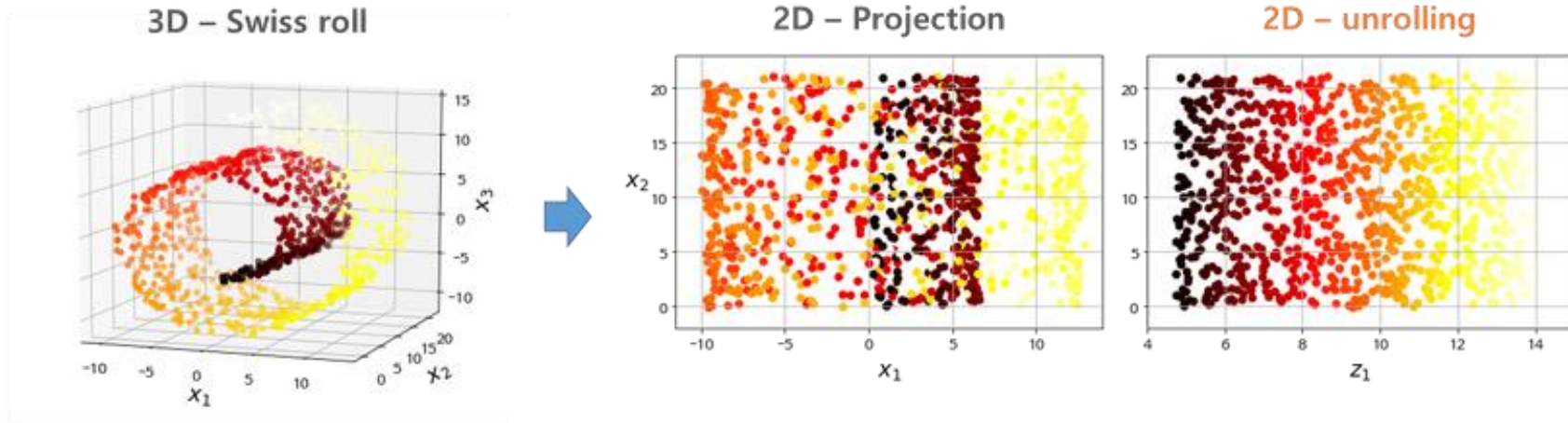


Manifold learning

- 매니폴드 러닝

- 매니폴드

- 가정: 고차원 데이터 공간에 데이터가 존재할 때, 이들을 잘 포함할 수 있는 저차원의 non-linear한 서브스페이스가 존재할 것이다. 그 서브스페이스가 매니폴드.
 - 매니폴드 러닝 : 고차원의 데이터를 잘 표현할 수 있는 매니폴드를 모델링(학습)



Manifold learning

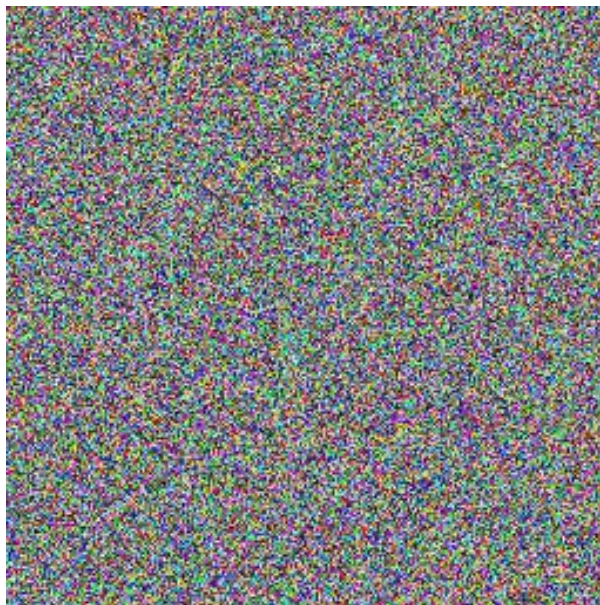
▪ 매니폴드 러닝

▪ 매니폴드 가정이 맞는 가정인가?

▪ 200x200 RGB 이미지 : 10^{96329} possible state(고차원 공간)

▪ 여기서 random sampling 아무리 해봐야 noisy한 이미지밖에 안 나옴.

▪ 얼굴 이미지 샘플들(=서로 비슷한 특성을 가진 애들)은 고차원 공간 중 어딘가 몰려 있을 것 같다.(골고루 분포가 아님)

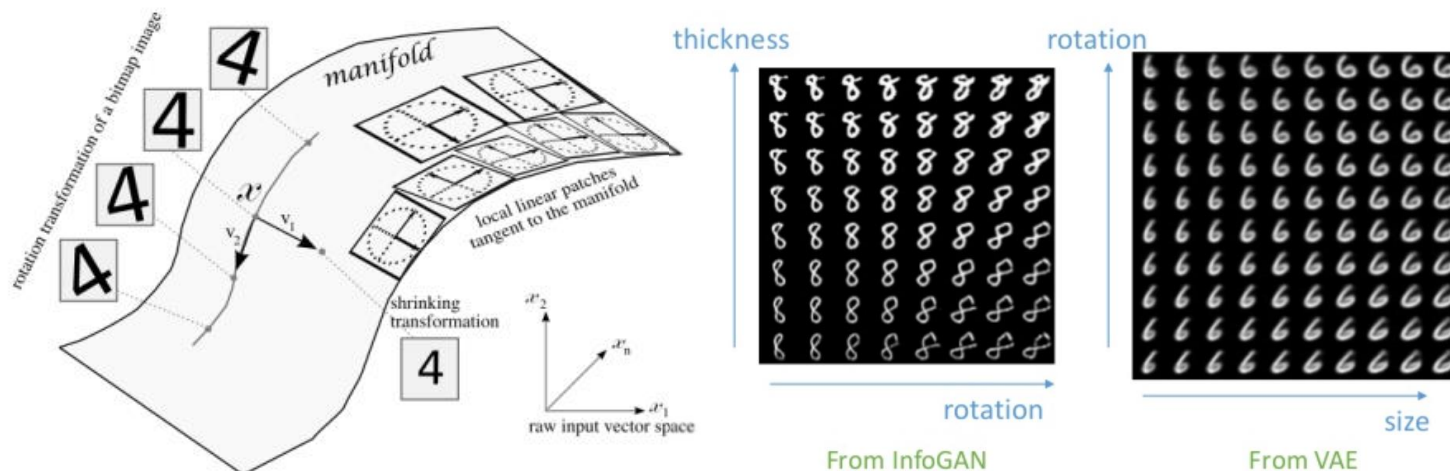


Manifold learning

▪ 매니폴드 러닝

▪ 주요 feature의 발견

- 매니폴드 학습 결과: 매니폴드 상의 좌표가 조금씩 변할 때 원데이터도 유의미하게 조금씩 변함을 보인다.



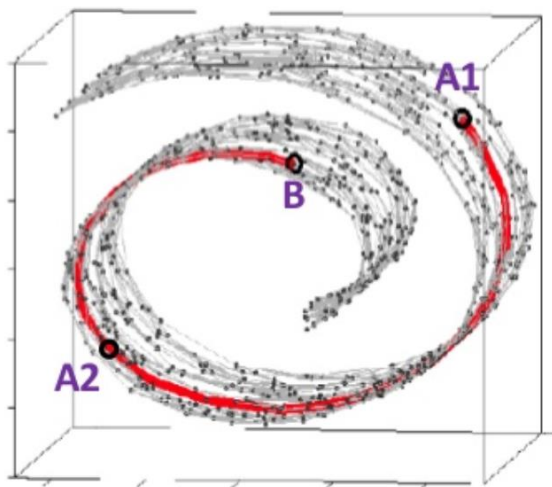
- 오토인코더가 학습을 잘 해서 인코딩 후 디코딩하여 원본 데이터를 잘 복원한다면, 저차원 공간으로 고차원 데이터를 표현한 것이니 그 feature를 잘 찾았다고 봐도 무방

Manifold learning

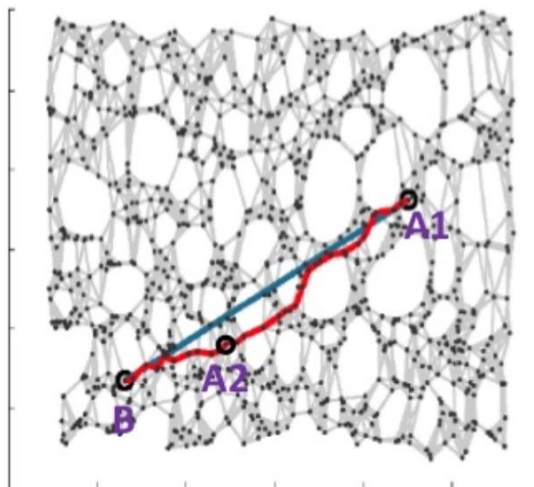
▪ 매니폴드 러닝

▪ 주요 feature의 발견

- 고차원 공간에서 의미적으로 가깝다고 생각되는 두 샘플 간의 원본 데이터 공간에서의 유클리드 거리는 실제로는 먼 경우가 있다.
- 매니폴드에서 유클리드 거리가 가깝다는 것은 dominant feature 들이 가깝다는 뜻으로 해석 가능



Distance in high dimension

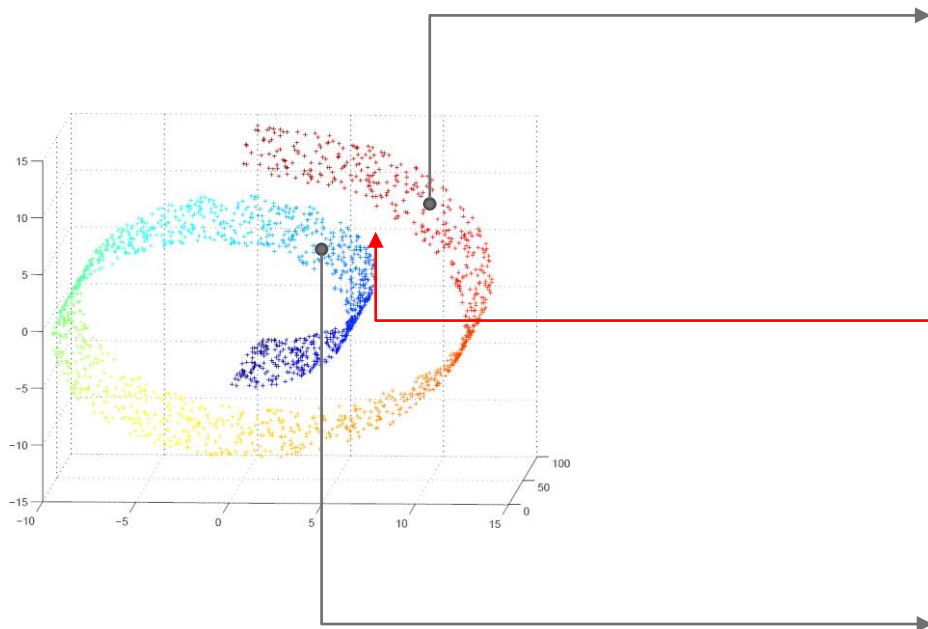


Distance in manifold

Manifold learning

- 매니폴드 러닝

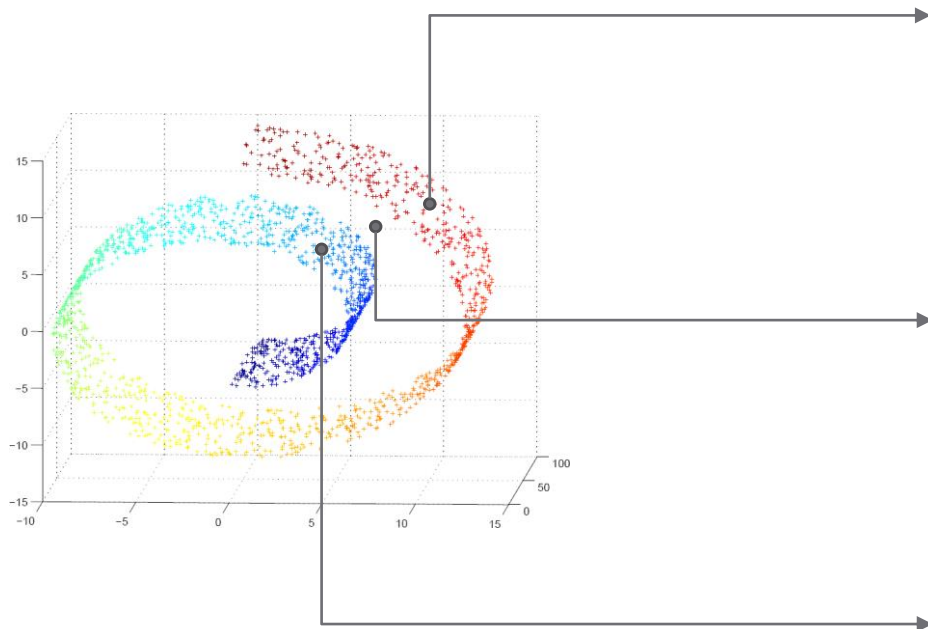
- 주요 feature의 발견



Manifold learning

- 매니폴드 러닝

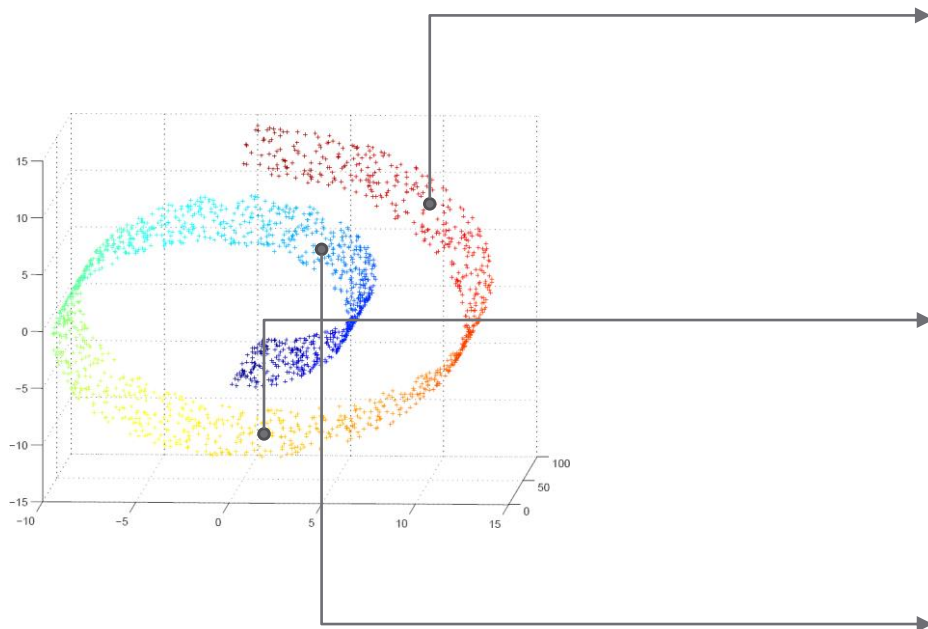
- 주요 feature의 발견



Manifold learning

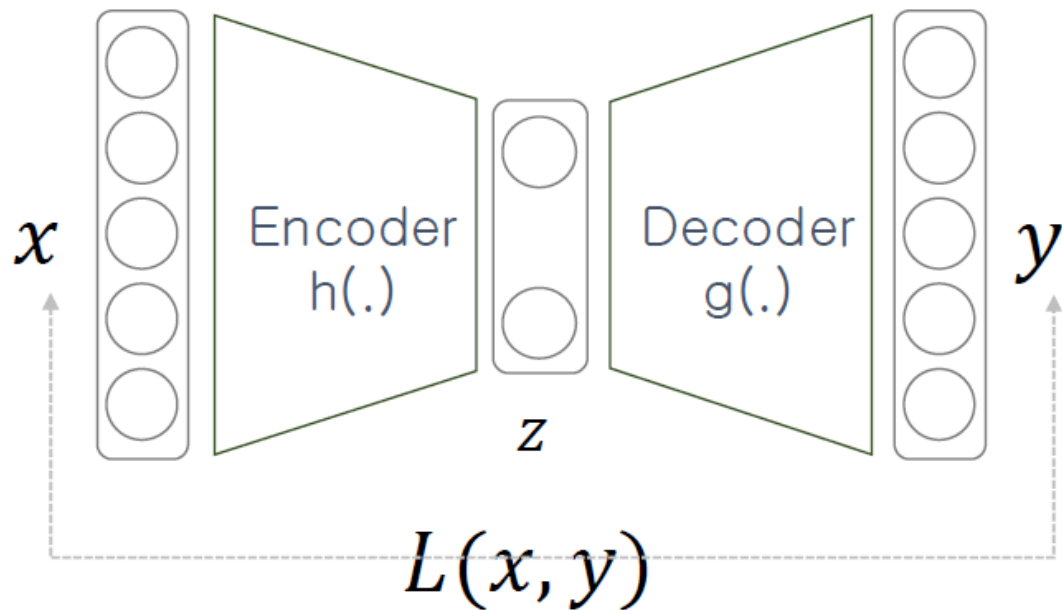
- 매니폴드 러닝

- 주요 feature의 발견



Auto encoder

- 입력과 출력이 동일한 네트워크 구조
- Unsupervised learning 을 Supervised learning 으로 바꾸어 해결
- z
 - code
 - latent(잠재) variable
 - feature
 - hidden representation

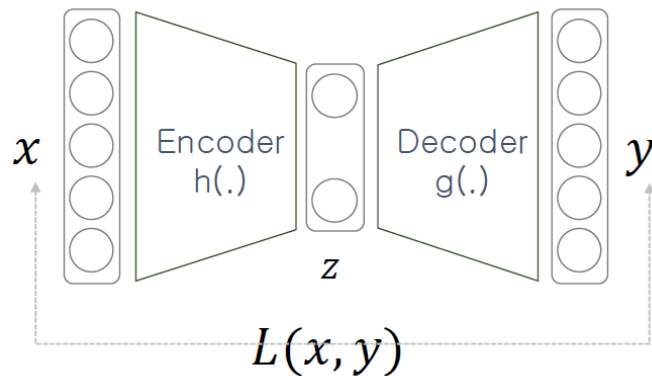


Auto encoder

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

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

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



- 오토인코더는 학습 후 인코더 디코더를 따로 떼어내서 사용.

- 인코더가 최소한 학습 데이터는 latent vector 로 잘 표현할 수 있게 된다.

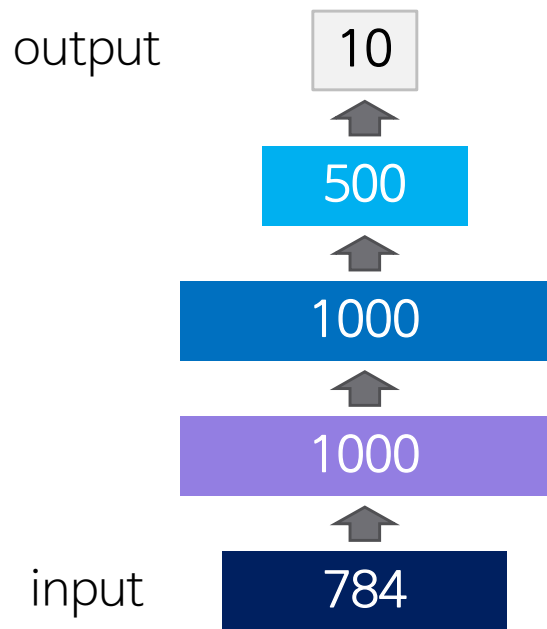
- 데이터의 추상화를 위해 사용할 수 있다.

- 디코더가 최소한 학습데이터는 생성해 낼 수 있게 된다.(minimum 성능 보장)

- 생성된 데이터가 학습데이터를 닮아 있다.(새로운 것보다는 기존 학습데이터와 비슷한 것만 생성)

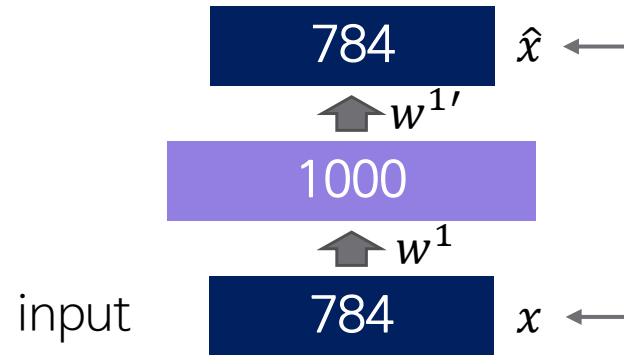
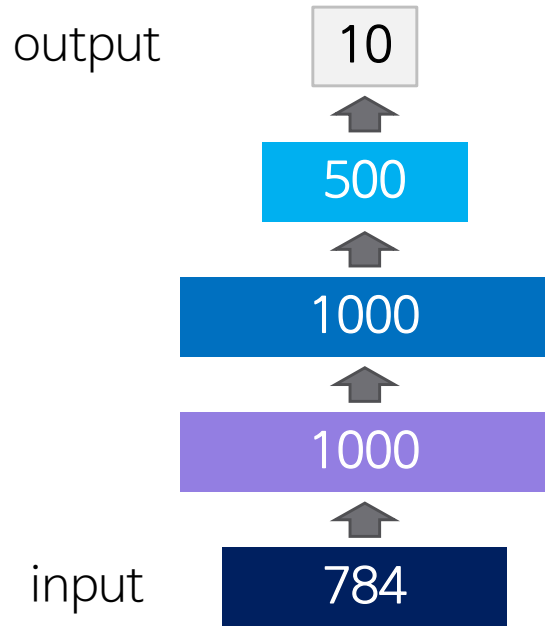
Auto encoder

- Stacking AutoEncoder (SAE)
 - 초기 오토인코더의 활용
 - 네트워크 파라미터 초기화(=pre-training)



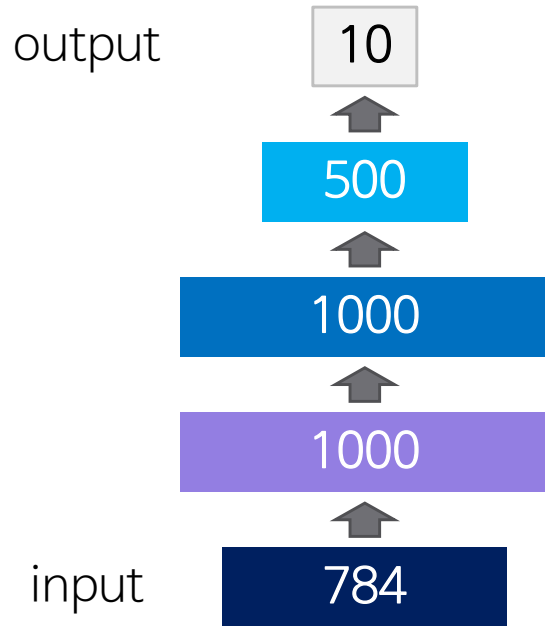
Auto encoder

- Stacking AutoEncoder (SAE)
 - 초기 오토인코더의 활용
 - 네트워크 파라미터 초기화(=pre-training)



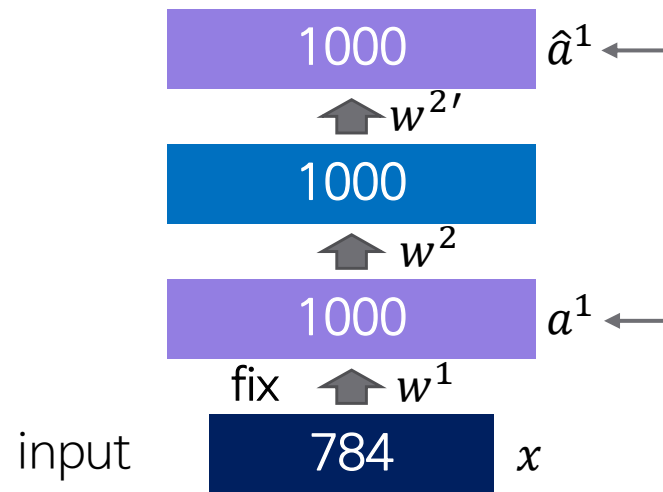
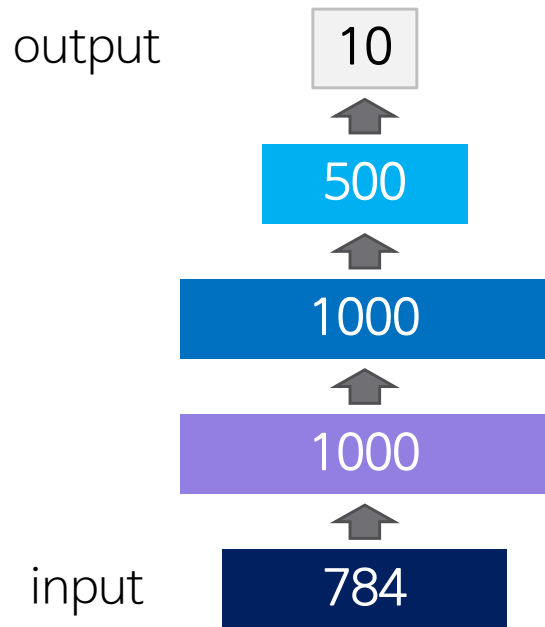
Auto encoder

- Stacking AutoEncoder (SAE)
 - 초기 오토인코더의 활용
 - 네트워크 파라미터 초기화(=pre-training)



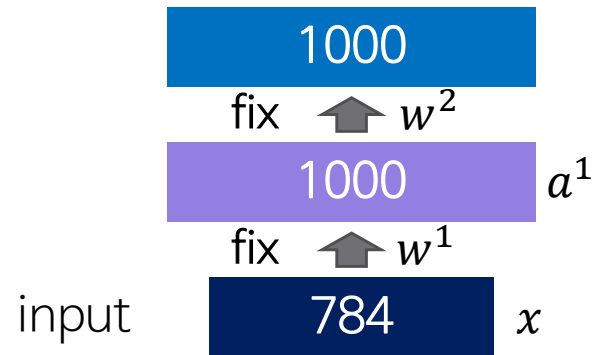
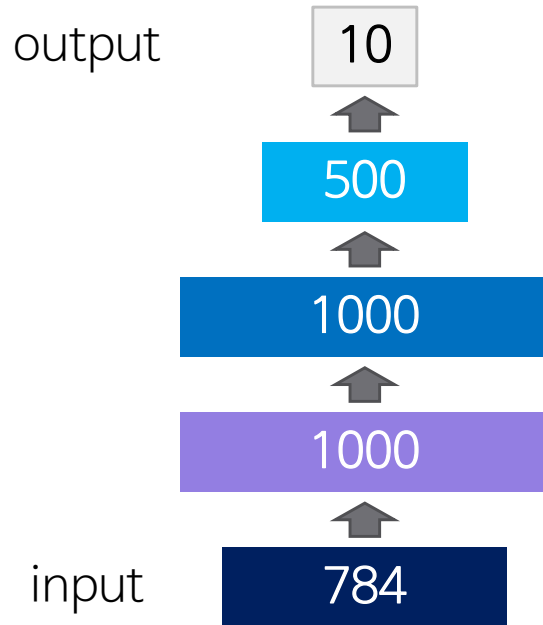
Auto encoder

- Stacking AutoEncoder (SAE)
 - 초기 오토인코더의 활용
 - 네트워크 파라미터 초기화(=pre-training)



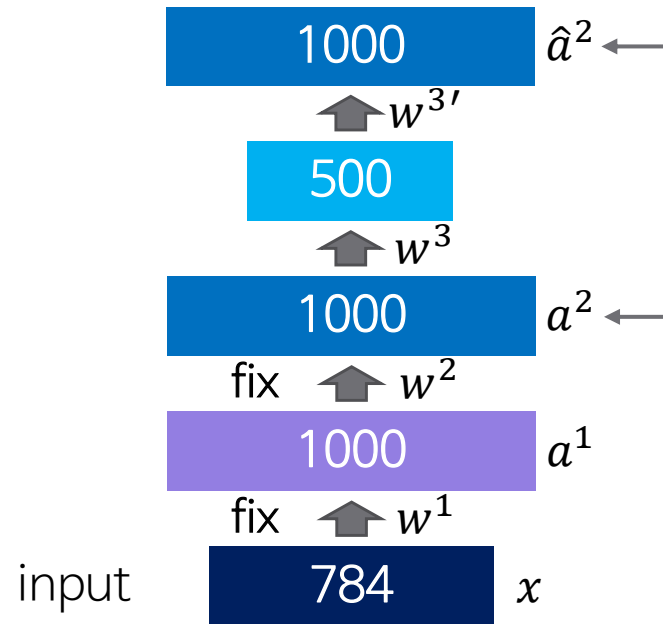
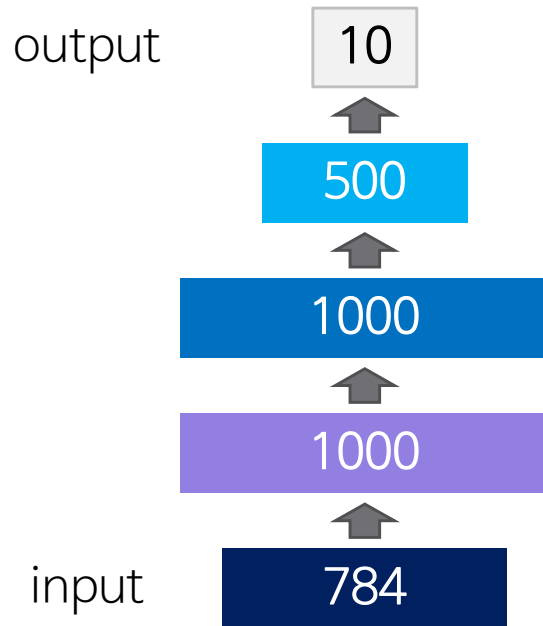
Auto encoder

- Stacking AutoEncoder (SAE)
 - 초기 오토인코더의 활용
 - 네트워크 파라미터 초기화(=pre-training)



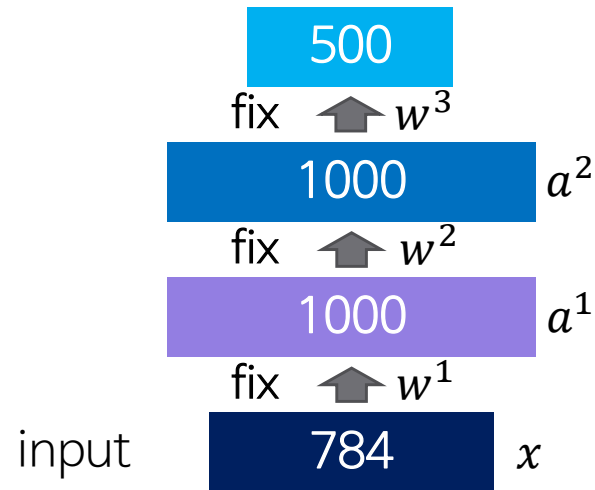
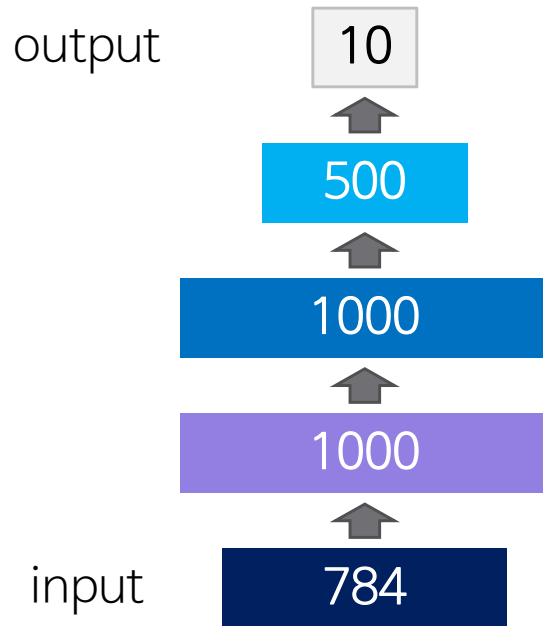
Auto encoder

- Stacking AutoEncoder (SAE)
 - 초기 오토인코더의 활용
 - 네트워크 파라미터 초기화(=pre-training)



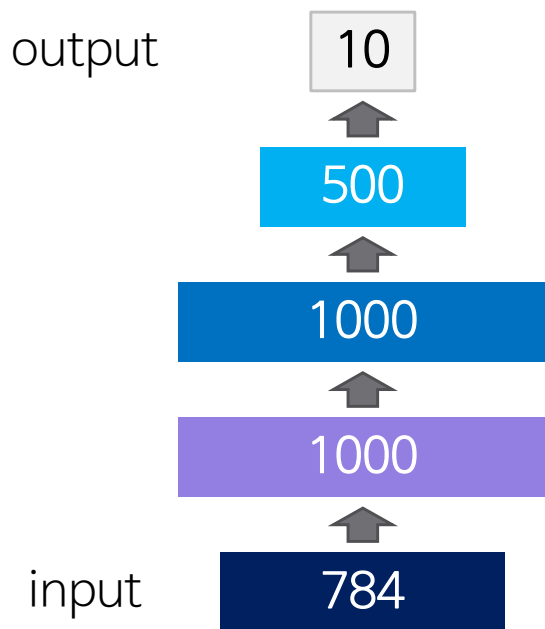
Auto encoder

- Stacking AutoEncoder (SAE)
 - 초기 오토인코더의 활용
 - 네트워크 파라미터 초기화(=pre-training)

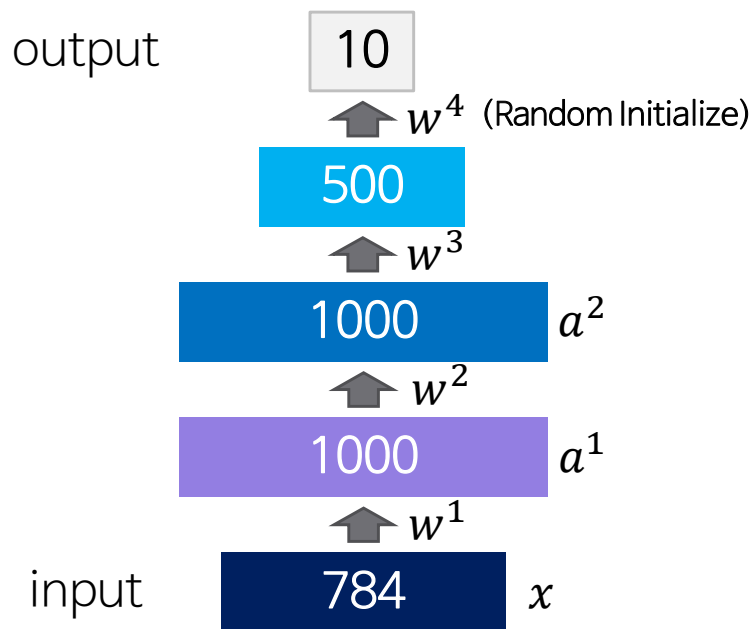


Auto encoder

- Stacking AutoEncoder (SAE)
 - 초기 오토인코더의 활용
 - 네트워크 파라미터 초기화(=pre-training)

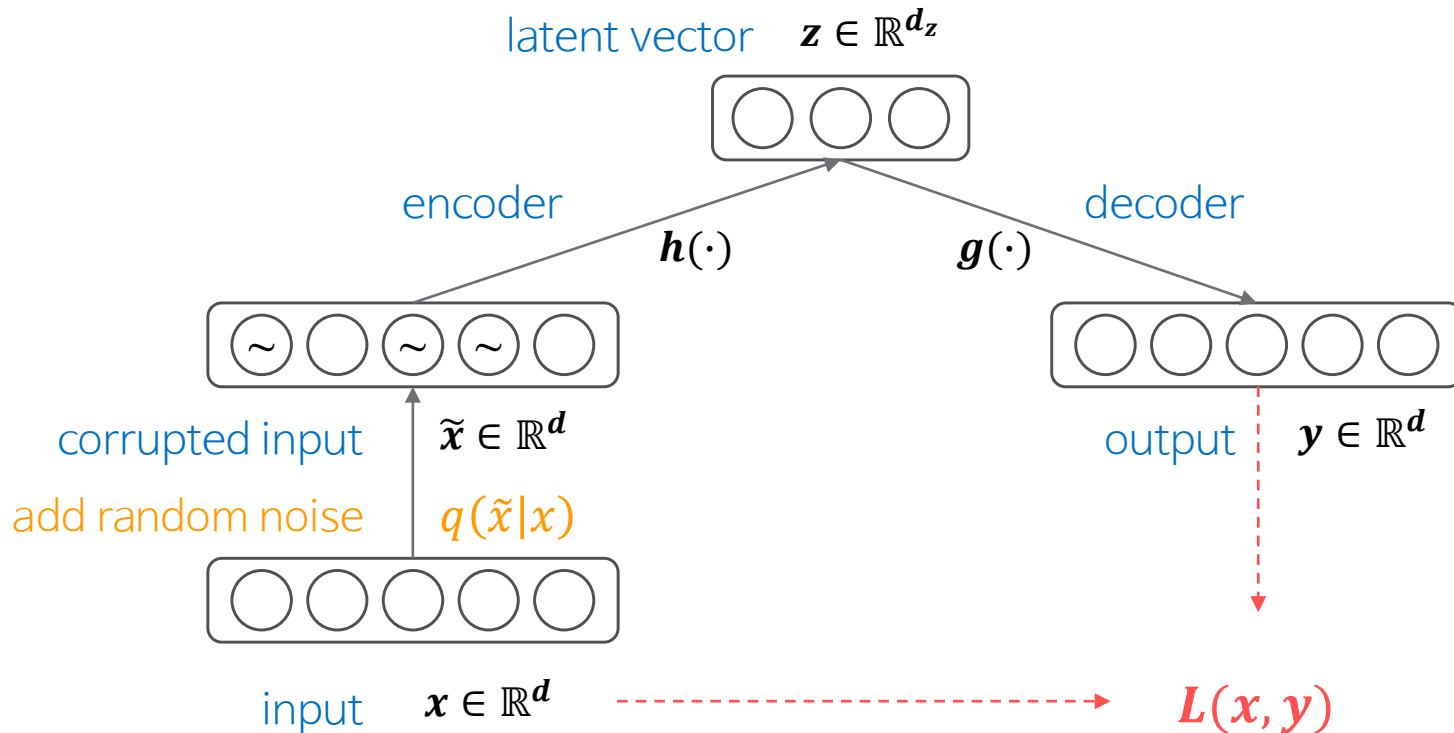


Fine-tuning by backpropagation



Auto encoder

- Denoising AutoEncoder (DAE)

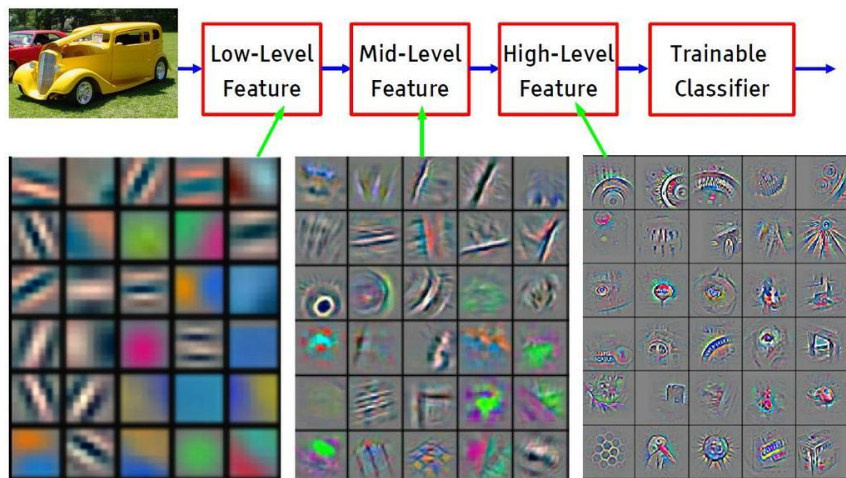


$$L_{DAE} = \sum_{x \in D} E_{q(\tilde{x}|x)} [L(x, g(h(\tilde{x})))]$$

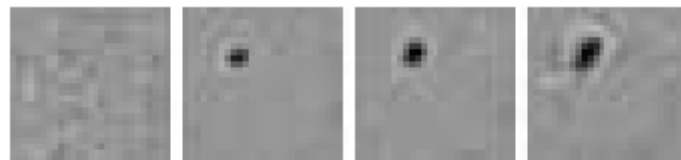
Auto encoder

▪ Denoising AutoEncoder (DAE)

- 원래 DAE의 목적은 매니폴드 러닝
- 매니폴드 러닝이 잘 되었다는 것 = 데이터의 feature 를 잘 찾는다는 것
- 노이즈를 추가해도 의미적으로는 원본 이미지와 같은 것이므로 같은 매니폴드에 매핑되도록 학습



Performance on mnist dataset



(d) Neuron A (0%, 10%, 20%, 50% corruption)



(e) Neuron B (0%, 10%, 20%, 50% corruption)

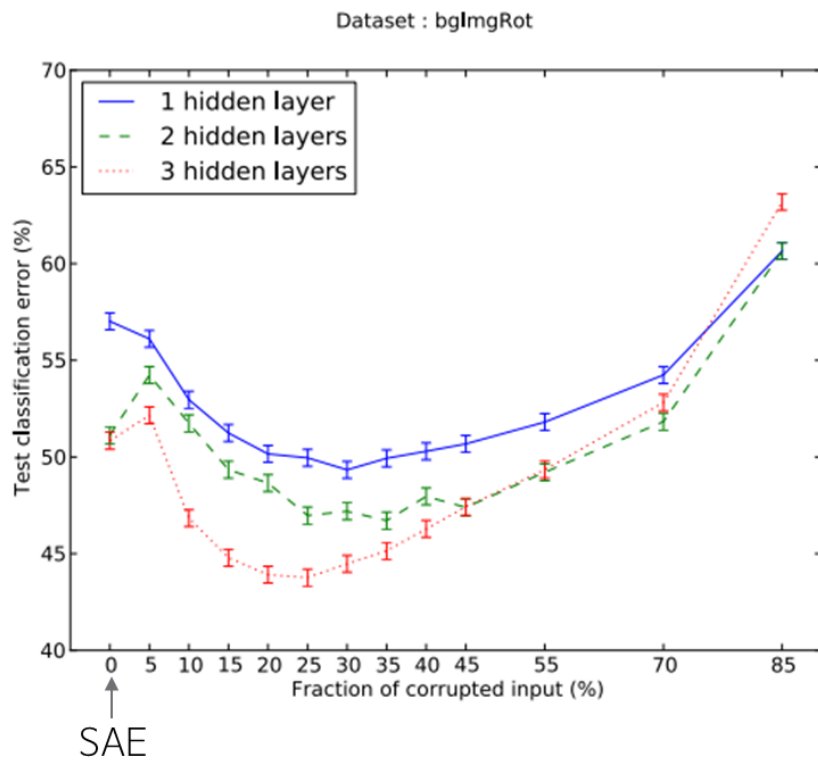
Auto encoder

- Stacked Denoising AutoEncoder (SDAE)

- Performance

- 노이즈를 어느정도 주었을 때 에러가 떨어지는 것으로 보아 DAE 가 pre-training 에 더 좋다.

(= 데이터 feature를 더 잘 파악한다)



Auto encoder

- Stochastic Contractive AutoEncoder (SCAE)

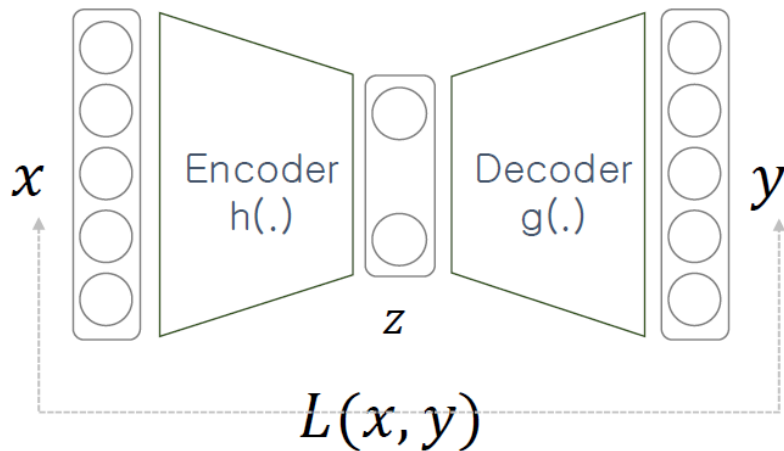
- DAE 의 loss 수정

- x 가 약간 바뀌더라도 매니폴드 위에서 같은 샘플로 매칭되도록 인코더 h 를 학습.

- $$L_{SCAE} = \underbrace{\sum_{x \in D} L(x, g(h(x)))}_{\text{Reconstruction Error (기존 오토인코더 loss)}} + \underbrace{\lambda \mathbf{E}_{q(\tilde{x}|x)} [\|h(x) - h(\tilde{x})\|^2]}_{\text{Stochastic Regularization}}$$

Reconstruction Error
(기존 오토인코더 loss)

Stochastic Regularization



Auto encoder

- Contractive AutoEncoder (CAE)

- SCAE regularization term

- $E_{q(\tilde{x}|x)}[\|h(x) - h(\tilde{x})\|^2]$

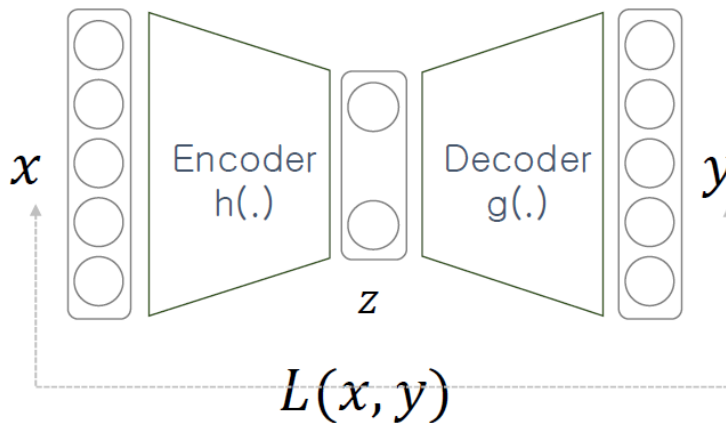
- CAE regularization term

- $h(\tilde{x}) = h(x + \epsilon) \approx h(x) + \frac{\partial h}{\partial x} \epsilon$

- $E_{q(\tilde{x}|x)}[\|h(x) - h(\tilde{x})\|^2]$ $\approx \left\| \frac{\partial h}{\partial x}(x) \right\|_F^2$

Stochastic Regularization

Analytic Regularization



- AI Lab. seminar - github
 - Season #3 - 04. Variational_Auto_Encoder
 - Season #6 - 03. Auto-Encoder
 - Season #6 - 12. Variational Auto-Encoder
- 오토인코더의 모든 것 - youtube
 - https://youtu.be/o_peo6U7IRM
- 차원 축소 - 블로그
 - <https://excelsior-cjh.tistory.com/167>

딤러닝 세미나 Season #11

Thank you