# Graph Auto-Encoder

고우람
ALDE @ PNU
2021.08.10

# Overview

- Auto-Encoder
- **Graph** Auto-Encoder
- CiteSeer Node Embedding by Link Prediction

# Neural Network가 하는 일?

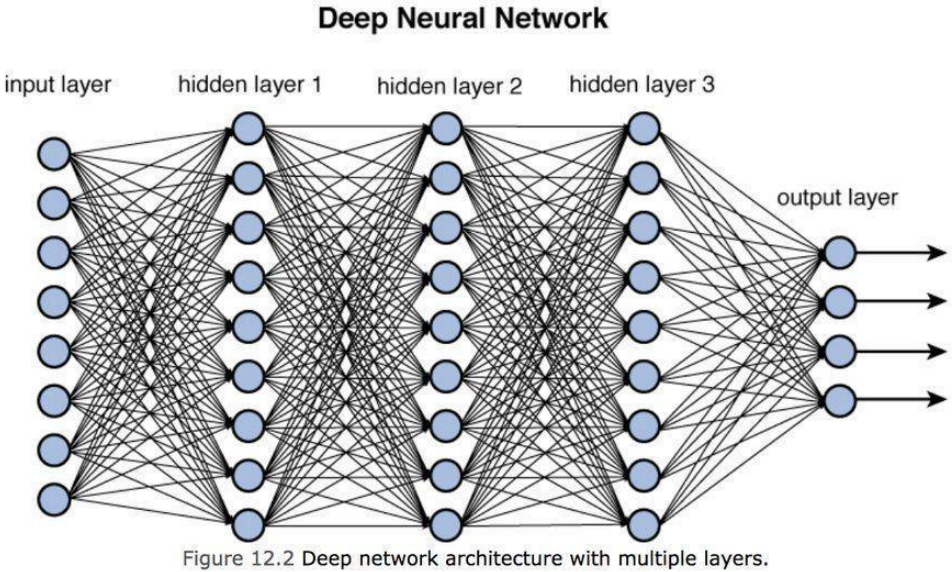


Figure 12.2 Deep network architecture with multiple layers.

[1
2
3
4]

???

# Neural Network가 하는 일

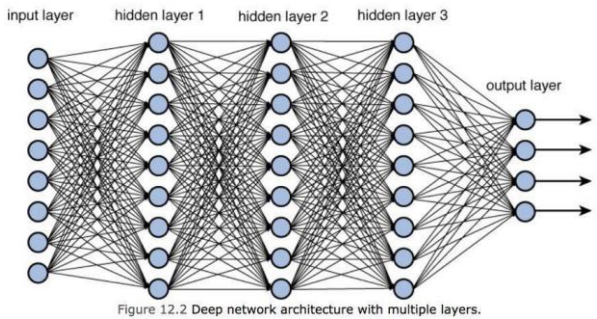


Figure 12.2 Deep network architecture with multiple layers.

```
[1
 2
 3
 4]
```

```
[9
 0
 8
 0]
```

입력에서 “어떤 중요한" feature 학습

- 예시: 분류를 위한 feature 학습

# Neural Network가 하는 일

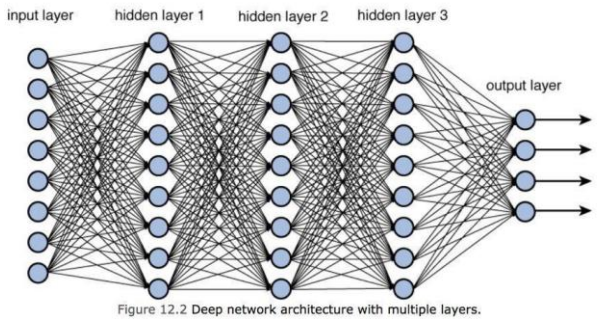


Figure 12.2 Deep network architecture with multiple layers.

```
[1
 2
 3
 4]
```

```
[9
 0
 8
 0]
```

embedded latent variable
in latent space

입력에서 “어떤 중요한" feature 학습

- 예시: 분류를 위한 feature 학습

# 지도 학습

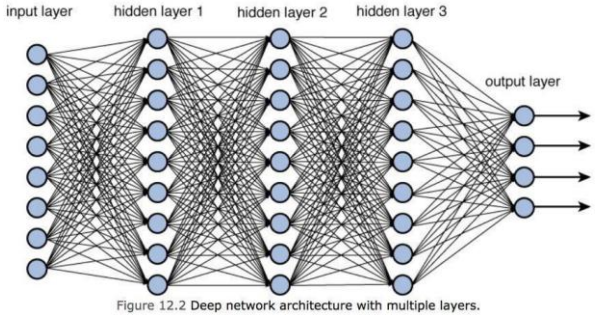


Figure 12.2 Deep network architecture with multiple layers.

```
[1        [1
 2         1
 3         1
 4]        1]

[9        [0
 0         0
 8         0
 0]        0]
```

출력 정답

feature 학습을 위해서는 label이 필요

# 비지도 학습?

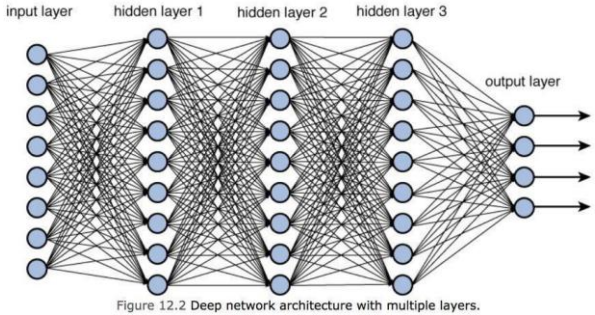
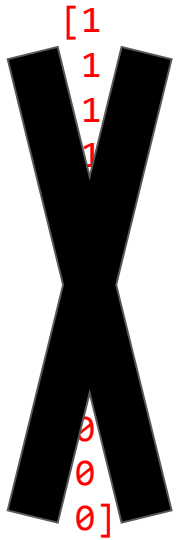


Figure 12.2 Deep network architecture with multiple layers.

[1
2
3
4]

[9
0
8
0]



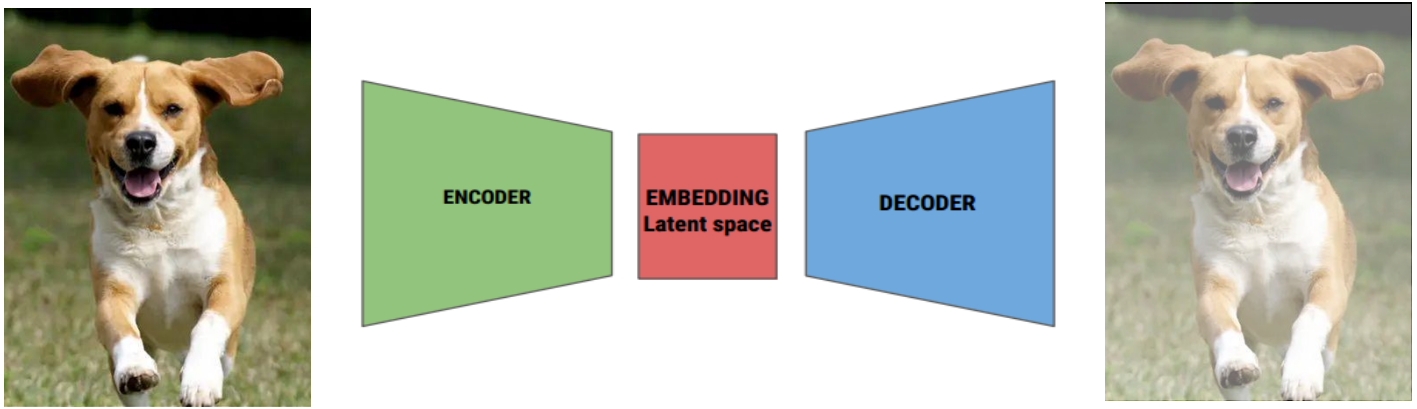

출력

정답

feature 학습을 위해서는 label이 필요

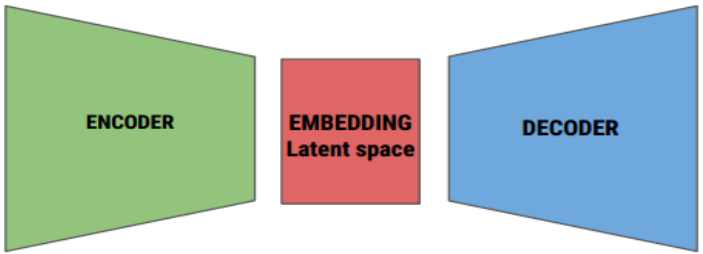
**label 없이 feature 학습**을 하는 방법은?

# Auto-Encoder




Encoder - Decoder 구조로 입력을 생성

# Auto-Encoder

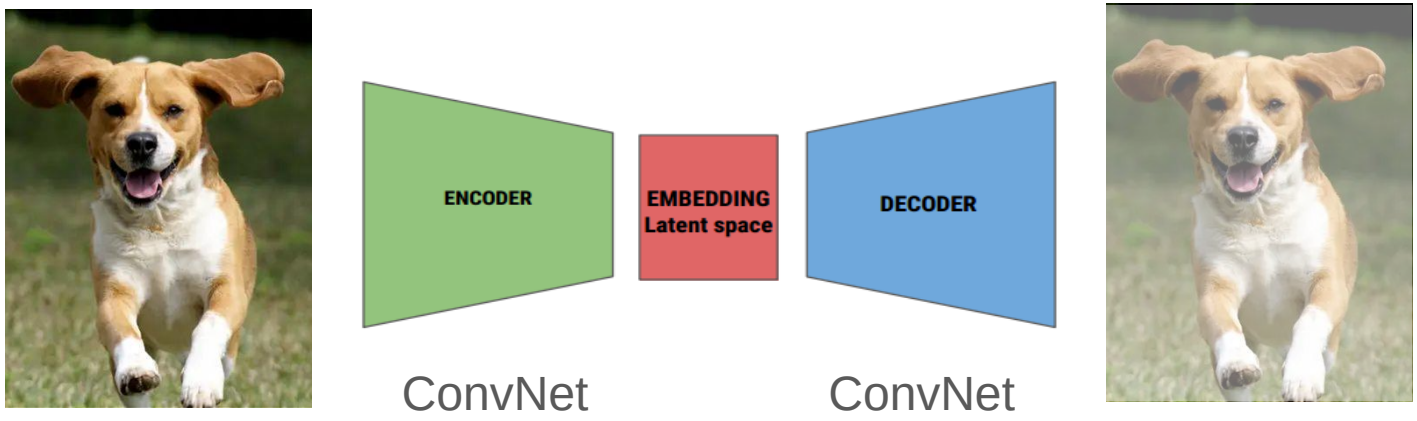



**Loss = Similarity( output , input )**

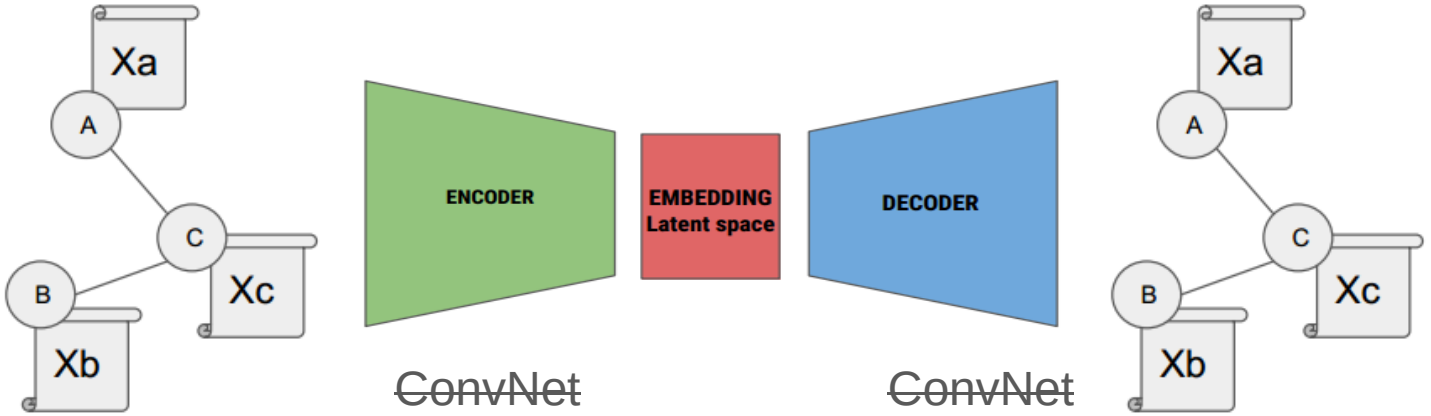
Encoder - Decoder 구조로 입력을 생성

입력과 출력을 비교.

# Auto-Encoder



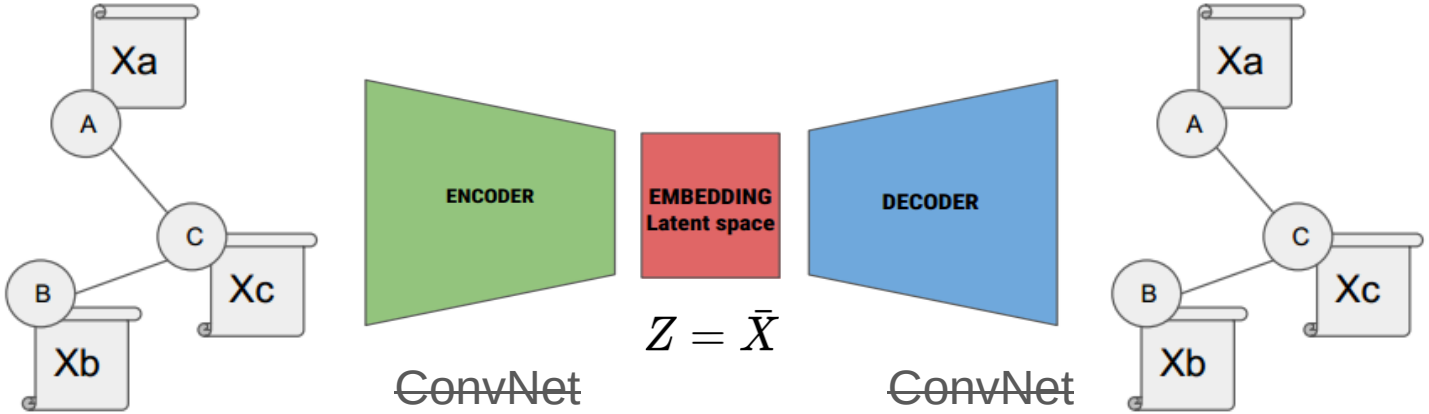
# **Graph** Auto-Encoder




~~ConvNet~~ ~~ConvNet~~

**GCNConv** **InnerProductDec**

# **Graph** Auto-Encoder




$Z = \bar{X}$

~~ConvNet~~ ~~ConvNet~~

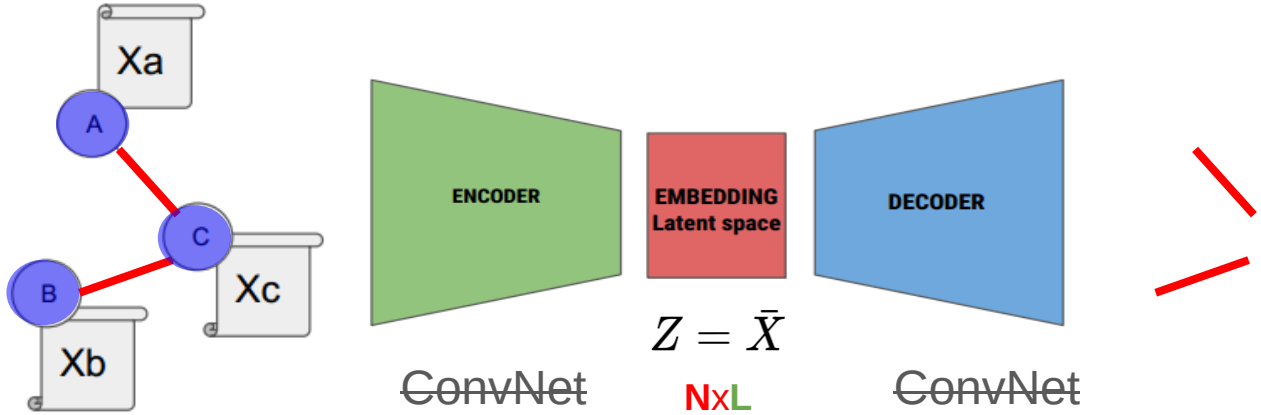
**GCNConv** **InnerProductDec**

$$\bar{X} = GCN(A, X) = ReLU(\tilde{A}XW_0)$$
$$\text{with } \tilde{A} = D^{-1/2}AD^{-1/2}$$

$$\hat{A} = \text{logistic sigmoid}\,(zz^T)$$

(실제 논문과 예시에서는 2-layer GCN을 씀)

# **Graph** Auto-Encoder




NxN NxF

**GCNConv**

$$\bar{X} = GCN(\boldsymbol{A}, \boldsymbol{X}) = ReLU(\tilde{A}XW_0)$$
$$\text{with } \tilde{A} = D^{-1/2}AD^{-1/2}$$

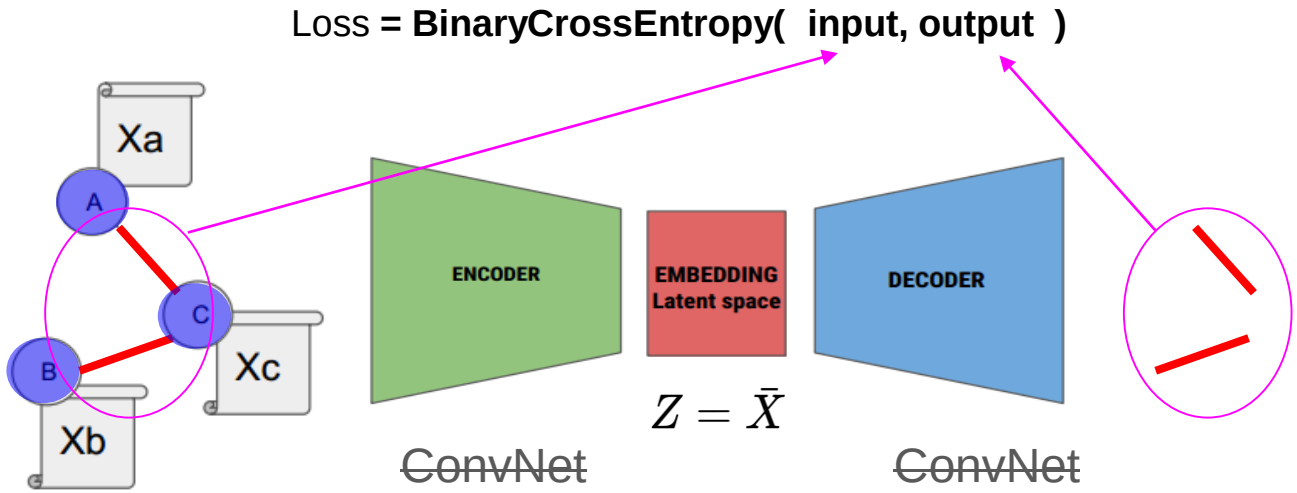
**InnerProductDec**

$$\hat{A} = \text{logistic sigmoid}\,(zz^T)$$

NxN NxL LxN

(실제 논문과 예시에서는 2-layer GCN을 씀)

# **Graph** Auto-Encoder




**GCNConv**

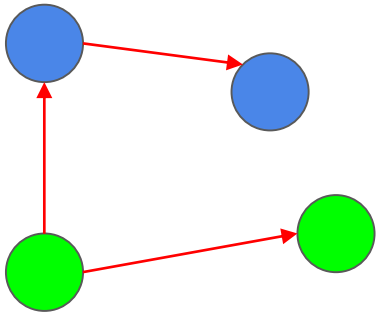
$$\bar{X} = GCN(\boldsymbol{A}, \boldsymbol{X}) = ReLU(\tilde{A}XW_0)$$
$$\text{with } \tilde{A} = D^{-1/2}AD^{-1/2}$$

(실제 논문과 예시에서는 2-layer GCN을 씀)

**InnerProductDec**

$$\hat{A} = \text{logistic sigmoid}\,(zz^T)$$

# CiteSeer Node Feature Embedding



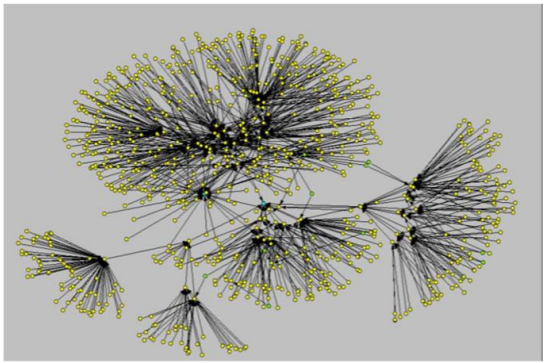
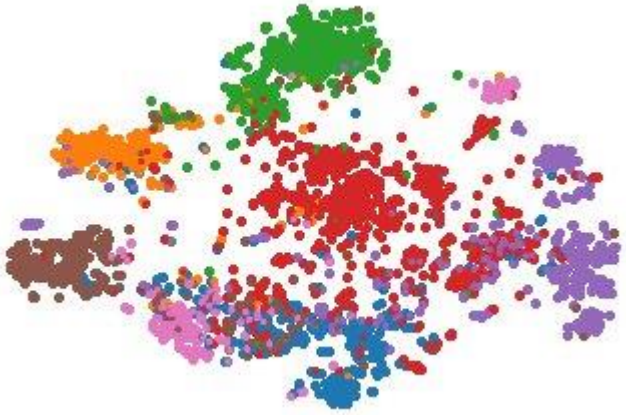


Fig. 1 Pajek output with Kamada Kawi free algorithm

Khalid et al.



Zheng et al.

*CiteSeer Dataset*

node: 논문에 있는 유일 단어(unique word) 존재성 벡터
node label: 분야(AI, ML, DB, ..)에 따라 분류

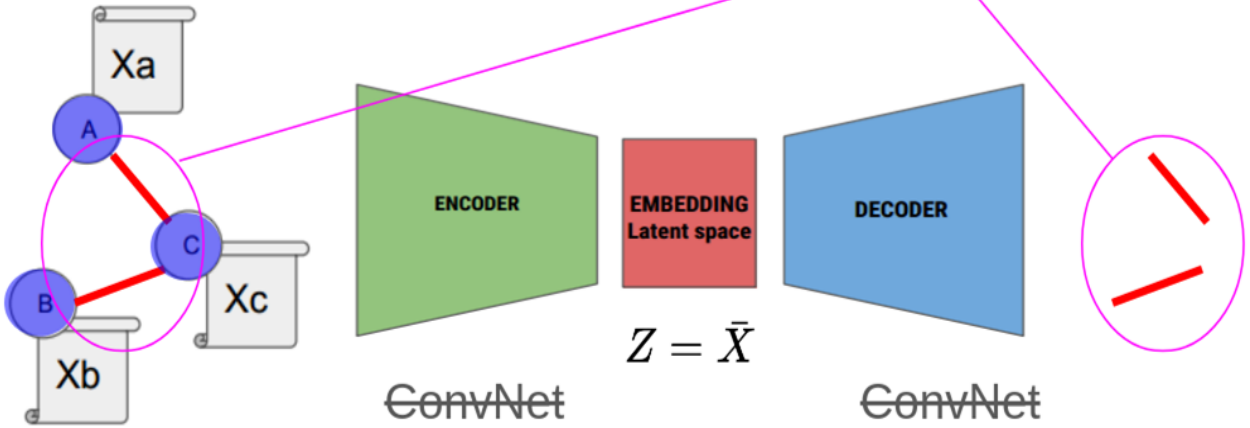
edge: 논문간 Citation을 표시

목표: **node label 없이** node를 **분야에 따라 분류**(클러스터링) 해보자

# CiteSeer Node Feature Embedding by Link Prediction

$$H_p(q) = -\frac{1}{(N_{pos} + N_{neg})}\left[\sum_{i=1}^{N_{pos}} log(p(y_i)) + \sum_{i=1}^{N_{neg}} log(1 - p(y_i))\right]$$

Loss = **BinaryCrossEntropy( input, output )**




$$\bar{X} = GCN(\boldsymbol{A}, \boldsymbol{X}) = ReLU(\tilde{A}XW_0)$$
$$\text{with } \tilde{A} = D^{-1/2}AD^{-1/2}$$

$$\hat{A} = \text{logistic sigmoid}\,(zz^T)$$

[GAE my Colab]

목적: 노드 피처 임베딩

피처: 노드(논문)별
유일 단어 존재성

학습 방법:
GAE로 Link Prediction

# 결론

GAE는 그래프의 피처를 임베딩할 수 있다.

GAE는 Encoder-Decoder 구조를 이용하여 비지도 학습 방식으로 학습할 수 있다.

GAE는 기존 Auto-Encoder처럼 클러스터링 등에 응용할 수 있다.