

# 입력 데이터 무작위도 특성이 기본 오토인코더와 적층 오토인코더의 학습 성능에 미치는 영향

이주홍\*, 정다운\*, 박형곤<sup>o</sup>

## Impact of Input Data Randomness on Training Performance of Basic Autoencoder and Stacked Autoencoder

Joohong Rhee<sup>\*</sup>, Daeun Jung<sup>\*</sup>, Hyunggon Park<sup>o</sup>

### 요약

본 논문에서는 입력 데이터 특성 중 하나인 무작위도가 오토인코더의 학습 성능에 미치는 영향을 살펴본다. 가우시안 분포를 따르는 입력 데이터를 가정하여 입력 데이터의 무작위도와 오토인코더의 학습 성능이 선형 관계에 있음을 이론적으로 증명하였다. 이론적 증명을 실험적으로 확인하기 위하여, 입력 데이터의 표준편차와 평균을 변화시켜가며 실험을 진행하였다. 그 결과, 입력 데이터의 평균에 관계없이 입력 데이터의 표준편차와 오토인코더의 학습 손실이 선형 관계임을 보임으로써 이론적 증명과 실험 결과가 동일함을 확인하였다. 기본 구조의 오토인코더에 대한 실험과 적층 오토인코더에 대한 실험을 각각 진행하여 은닉층의 수에 관계없이 입력 데이터의 무작위도가 오토인코더의 학습 성능과 선형 관계에 있음을 확인하였다. 이때 은닉층의 노드 개수가 증가할수록 오토인코더의 학습 손실이 감소하며, 입력 데이터의 무작위도에 따른 학습 손실의 변화가 적음을 알아내었다.

**Key Words** : Autoencoder, Stacked autoencoder, Training loss, Gaussian distribution, Randomness

### ABSTRACT

In this paper, we examine the impact of the input data randomness on the training performance of the autoencoder. Assuming the input data follow a Gaussian distribution, we mathematically analyze that the randomness of the input data and the training performance of the autoencoder have a linear relationship, and demonstrate this finding through experiments. Experiments on the basic autoencoder and the stacked autoencoder confirm that the input data randomness and the training performance of the autoencoder have a linear relationship regardless of the number of hidden layers. In addition, we examine the training performance of autoencoder according to the mean and standard deviation of the input data. The results support that the mean of the input data has a negligible effect on the training loss of the autoencoder. Therefore, we ensure consistency between mathematical analysis and the experimental results that the number of nodes in the hidden layer increases, the training loss of the autoencoder and the impact of input data randomness decrease.

※ 본 연구는 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(2021-0-00739)과 한국연구재단의 지원(NRF-2020R1A2B5B01002528)을 받아 수행되었습니다.

• First Author : Ewha Womans University Department of Electronic and Electrical Engineering, Graduate Program in Smart Factory, joohong.rhee@ewhain.net, 학생회원

◦ Corresponding Author : Ewha Womans University Department of Electronic and Electrical Engineering, Graduate Program in Smart Factory, hyunggon.park@ewha.ac.kr, 종신회원

\* Ewha Womans University Department of Electronic and Electrical Engineering, daeun.jung@ewhain.net, 학생회원

논문번호 : 202110-279-A-RE, Received September 30, 2021; Revised October 18, 2021; Accepted October 18, 2021

## I. 서 론

사물인터넷(IoT, Internet of Things)의 발달, 컴퓨팅 성능 향상, 데이터 저장 가능 용량 증가로 빅데이터에 기반한 초연결 사회가 도래하고 있다. 5G/B5G 통신·네트워크를 이용하는 사용자가 폭발적으로 증가하고, 다양한 기능을 제공하는 새로운 서비스가 등장하면서 비정형 데이터가 대량으로 발생하고 있다<sup>1)</sup>. 이러한 변화로 인해 더 높은 데이터 전송 속도, 향상된 커버리지(coverage), 더 나은 비용 효율성(cost efficiency) 등이 요구되고 이를 해결하기 위하여 머신러닝을 적용하는 노력이 계속되고 있다<sup>2)</sup>. 구체적으로 LSTM(Long Short-Term Memory) 기반 순환신경망(RNN, Recurrent Neural Network)을 활용한 서비스별 네트워크 자원 예측<sup>3)</sup>, 강화학습 기반 기지국 전원 제어<sup>4)</sup> 등이 제안된 바 있다. 이 외에도 머신러닝을 적용하여 트래픽 분류(traffic classification), 트래픽 라우팅(traffic routing), 혼잡 제어(congestion control), 결함 예측(fault prediction), QoS(Quality of Service) 관리, QoE(Quality of Experience) 관리, 이상 탐지(anomaly detection), 침입 탐지(intrusion detection) 등의 문제를 해결하려는 시도가 있다<sup>5)</sup>.

다양한 머신러닝 기법 중, 오토인코더는 입력 데이터를 압축하는 인코딩(encoding) 과정을 거친 후 디코딩(decoding) 과정을 통해 다시 원래의 데이터로 복원시킬 수 있도록 표현 학습(representation learning)을 하는 인공 신경망이다. 오토인코더는 라벨이 없는 입력 데이터 및 적은 입력 데이터의 표현을 학습할 수 있는 장점이 있어, 비지도 학습 방식 기반의 딥러닝 알고리즘 사용에 효과적이다. 따라서 오토인코더 기반의 소형 셀 주파수 및 전력 자원 할당 시스템<sup>6)</sup> 및 데이터 특성 간의 비선형 상관 관계(correlation)를 분석하여 정확도를 높인 네트워크 이상 탐지 시스템<sup>7)</sup>이 제안된 바 있으며 잡음제거 오토인코더(denoising autoencoder), 적층 오토인코더(stacked autoencoder), 변분 오토인코더(variational autoencoder)와 같은 다양한 종류의 오토인코더를 활용한 네트워크 침입 탐지 알고리즘<sup>8)</sup> 등의 연구가 진행된 바 있다.

오토인코더를 최적으로 사용하기 위해서는 오토인코더의 학습 성능을 결정하는 활성화 함수(activation function), 손실함수(loss function), 은닉층(hidden layer)의 구조, 에포크(epoch), 학습률(learning rate) 등을 적절히 설정하는 것이 매우 중요하다. 오토인코더의 구성 요소 중 활성화 함수와 손실함수는 입력 데이터의 특성과 사용 목적에 따라 결정할 수 있다. 또

한, 은닉층의 구조는 오토인코더에서 일반적으로 은닉층의 노드 개수가 증가할수록 학습 손실(training loss)이 작아지는 경향이 있다는 사실이 알려져 있으니<sup>9)</sup>, 입력 데이터의 특성에 따른 최적의 은닉층 노드 개수 결정에 관한 연구는 충분하지 않다. 오토인코더의 은닉층의 노드 개수가 증가함에 따라, 학습 손실은 감소하지만 학습량은 증가하는 트레이드 오프(trade-off)가 발생하므로, 효율적인 학습을 위하여 입력 데이터에 따라 적합한 은닉층의 노드 개수를 결정할 필요가 있다.

본 논문에서는 오토인코더를 활용할 때 입력 데이터의 특성에 맞는 적절한 은닉층의 노드 개수를 효율적으로 설정하기 위하여, 입력 데이터 특성 중 하나인 무작위도가 오토인코더의 학습 성능에 미치는 영향을 해석적으로 분석하였다. 가우시안 분포를 따르는 입력 데이터에 대하여, 오토인코더의 학습 손실이 입력 데이터의 무작위도를 의미하는 지표인 표준편차와 선형 관계가 있음을 증명하였다. 이론적 학습 손실을 실험적으로 확인하기 위하여, 가우시안 분포를 따르는 입력 데이터의 평균과 표준편차를 변화시켜가며 오토인코더에 적용하였다. 이때 오토인코더의 은닉층의 수에 관계없이 입력 데이터의 무작위도와 오토인코더의 학습 손실이 선형 관계에 있음을 확인하기 위해, 기본 구조의 오토인코더에 대한 실험과 적층 오토인코더에 대한 실험을 각각 진행하였다. 또한, 동일한 조건 하에 은닉층의 노드 개수를 변화시켰을 때 오토인코더의 학습 손실에 미치는 영향을 알아보았다. 이를 통하여 반복적인 은닉층 노드 개수 변화를 통한 실험 없이, 입력 데이터의 무작위도를 계산하여 적합한 은닉층의 노드 개수를 설정할 수 있다.

본 논문은 다음과 같이 구성되어 있다. II장에서는 본 논문의 실험에 활용되는 오토인코더의 구조와 오토인코더의 학습에 대한 이론적 배경을 제시한다. III장에서는 오토인코더의 학습 성능 척도를 정의하고, 가우시안 분포를 따르는 입력 데이터와 오토인코더 학습 성능 간의 관계를 분석한다. IV장에서는 III장에서 분석한 입력 데이터의 무작위도와 오토인코더의 학습 성능 간의 관계를 검증하기 위하여 실험을 수행하고, 실험 결과에 대해 논의한다. 마지막으로 V장에서는 입력 데이터의 무작위도가 오토인코더에 미치는 영향에 대한 결론을 제시한다.

## II. 오토인코더의 이론적 배경

### 2.1 오토인코더 구조

오토인코더는 입력 데이터의 차원을 압축하는 인코

더 부분과 인코더를 통하여 압축된 데이터를 원본의 입력 데이터로 다시 재구성하는 디코더 부분으로 이루어져 있다<sup>10)</sup>. 그림 1은 오토인코더의 기본 구조를 보여준다. 기본 구조의 오토인코더는 단일 입력층(input layer), 출력층(output layer), 은닉층으로 이루어져 있다. 인코더를 통과한 입력 데이터의 정보가 압축된 데이터 층을 은닉층이라 한다<sup>11)</sup>. 입력  $X$ 가 신경망을 통과하여 압축된 잠재 변수(latent variable)  $Z$ 가 되고,  $Z$ 가 다시 신경망을 통과하여 출력  $\hat{X}$ 이 된다. 오토인코더는 학습을 통해 출력  $\hat{X}$ 이 입력  $X$ 와 유사하게 출력되도록 하는 가중치를 찾기 위해 손실 함수  $L(\cdot)$ 를 이용하여  $\hat{X}$ 과  $X$ 의 차이를 줄이는 방향으로 학습한다. 손실함수를 통해 오토인코더의 학습 손실을 계산하고, 이를 최소화하는 잠재 변수  $Z$ 를 학습한다.

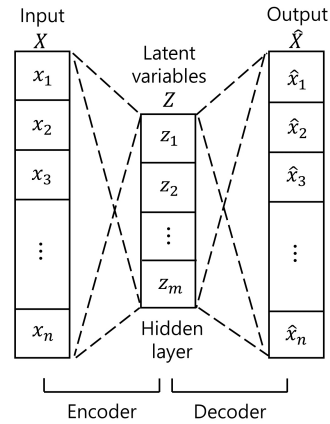


그림 1. 기본 오토인코더의 구조  
Fig. 1. The structure of basic autoencoder

은닉층이 하나인 기본적인 오토인코더 외에도, 다양한 종류의 오토인코더가 존재한다. 그 중, 적층 오토인코더는 그림 2와 같이 은닉층을 2개 이상 쌓아 올린 형태이다<sup>12)</sup>. 적층 오토인코더는 은닉층이 여러 개이기 때문에 마치 기본 구조의 오토인코더를 여러 개 쌓아놓은 것처럼 여러 단계의 인코딩, 디코딩 과정을 거치며, 가운데의 은닉층을 기준으로 대칭 형태로 이루어져 있다.

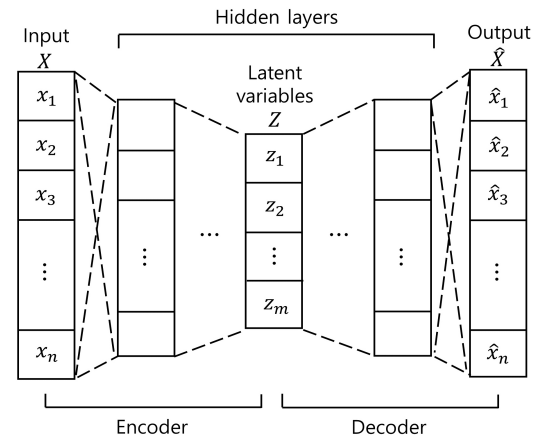


그림 2. 적층 오토인코더의 구조  
Fig. 2. The structure of stacked autoencoder

## 2.2 오토인코더의 학습에 영향을 미치는 요소

오토인코더의 구성 요소는 활성화 함수, 손실함수, 은닉층의 구조, 에포크, 학습률 등이 있다. 오토인코더의 학습 성능이 구성 요소 설정 방식에 따라 영향을 받기 때문에 오토인코더 사용 목적과 입력 데이터에 맞는 구성 요소를 설정하는 것이 매우 중요하다. 이를 위해 적절한 구성 요소를 결정하는 방식에 관한 다양한 연구가 진행되었다.

대표적인 오토인코더의 활성화 함수로는 ELU (Exponential Linear Unit), ReLU(Rectified Linear Unit), LeakyReLU, 시그모이드(sigmoid) 함수, 쌍곡 탄젠트(hyperbolic tangent) 함수가 있다. 활성화 함수에 내재된 특성으로 인해 출력값과 그래디언트(gradient) 값의 범위가 상이하기에, 입력 데이터와

목적 함수(objective function)에 따라 우수한 성능을 보이는 활성화 함수를 선택해야 한다<sup>13)</sup>. 이를 위해 기존의 활성화 함수를 결합하여 성능을 높이는 방안도 제안되었다<sup>14)</sup>.

오토인코더와 같은 인공 신경망을 이용한 딥러닝은 에포크를 적절하게 설정하는 것이 중요하다. 에포크를

작게 설정하면 데이터에 대한 학습이 충분히 이루어지지 않아 과소적합(underfitting)이 발생하고, 에포크를 크게 설정하면 학습 과정에서 과대적합(overfitting)이 발생하기 때문이다. 반복적인 실험을 통해 에포크를 결정하는 것을 피하고자 자기조직화 지도(self-organized map)를 활용하여 최적의 에포크를 설정하는 방안이 제안되었다<sup>15)</sup>.

은닉층의 수가 오토인코더의 학습 성능에 미치는 영향을 알아보기 위하여, 기본 오토인코더와 적층 오토인코더의 학습 성능을 비교하는 연구가 진행되었다<sup>16)</sup>. 은닉층의 수가 증가할수록 신경망의 학습 능력(learning capability)이 향상되고, 은닉층의 수가 감소하면 수렴 속도가 증가하며 학습 과정이 빠르게 진행된다는 사실이 밝혀졌다.

이와 같이 오토인코더 적용 모델의 정확성 및 효율

성을 향상시키기 위해 활성화 함수, 에포크, 은닉층의 개수 설정과 관련된 다양한 선행 연구가 진행되었지만, 오토인코더의 학습 성능은 오토인코더의 구성 요소 뿐만 아니라 입력 데이터에 따라서도 달라진다. 따라서 본 논문에서는 입력 데이터의 특성이 오토인코더의 학습 성능에 미치는 영향에 관하여 분석하고자 한다.

### III. 입력 데이터의 무작위도와 오토인코더의 학습 성능에 대한 분석

본 논문에서는 입력 데이터의 특성 중에서도 데이터 무작위도가 오토인코더의 학습 성능에 미치는 영향을 분석한다. 이를 위해 입력 데이터와 출력 데이터의 차이인 학습 손실과 입력 데이터의 무작위도 간 관련성을 도출하였다. 본 논문에서는 입력 데이터가 가우시안 분포를 따른다고 가정한다. 그러므로 데이터의 무작위도는 표준편차에 의해 결정되며, 입력 데이터의 무작위도를 나타내는 표준편차와 오토인코더의 학습 손실 간의 관계를 분석한다.

#### 3.1 오토인코더의 학습 성능 척도 정의

본 논문에서는 오토인코더의 학습 성능을 분석하기 위하여  $l_1$ -norm을 이용한 손실함수를 정의한다. 오토인코더의 입력을 벡터  $\mathbf{X}$ , 출력을 벡터  $\hat{\mathbf{X}}$ 이라 하고, 오토인코더의 입력층과 출력층의 노드 개수를  $n$ , 입력층의  $i$ 번째 노드를  $X_i$ , 출력층의  $i$ 번째 노드를  $\hat{X}_i$ 라 하고  $l_1$ -norm을 이용한 손실함수  $L(\mathbf{X}, \hat{\mathbf{X}})$ 은 식(1)과 같이 정의한다.

$$L(\mathbf{X}, \hat{\mathbf{X}}) = \sum_{i=1}^n |X_i - \hat{X}_i| \quad (1)$$

$\Omega$ 를 확률 변수(random variable),  $\omega$ 를 실현값(realization)이라 할 때, 임의의 분포  $p(\omega)$ 를 따르도록 입력 데이터  $\mathbf{X}$ 를 생성한다. 오토인코더를 통과한 출력 데이터  $\hat{\mathbf{X}}$ 은 임의의 분포  $q(\omega)$ 를 따른다. 즉,  $\mathbf{X} \sim p(\omega)$ ,  $\hat{\mathbf{X}} \sim q(\omega)$ 이다. 이때  $\mathbf{X}, \hat{\mathbf{X}}$ 의 데이터 집합의 크기가 충분히 크다면 식(2)와 같이 근사적으로 표현될 수 있다.

$$\sum_{i=1}^n |X_i - \hat{X}_i| \approx \sum_{\omega \in \Omega} |p(\omega) - q(\omega)| \quad (2)$$

식(2)와 같은 근사식이 성립한다면 손실함수  $L(\mathbf{X}, \hat{\mathbf{X}})$ 은 두 분포  $p(\omega), q(\omega)$  간의  $l_1$  거리를 계산한 것과 같다.  $p(\omega), q(\omega)$ 의 확률 측도(probability measure)를 각각  $P, Q$ 라 정의하면 식(3)과 같이 표현 가능하다.

$$\sum_{\omega \in \Omega} |p(\omega) - q(\omega)| = \|P - Q\|_1 \quad (3)$$

#### 3.2 가우시안 분포를 따르는 입력 데이터에서의 오토인코더 학습 성능 분석

본 논문에서는 입력 데이터  $\mathbf{X}$ 가 가우시안 분포  $\mathbf{X} \sim N(\mu_X, \sigma_X^2)$ 를 따른다고 가정한다. 오토인코더에서 은닉층의 압축된 데이터를 통해 출력 데이터가 입력 데이터와 유사하게 복원되도록 학습이 되었다면, 출력 데이터 또한 입력 데이터와 동일한 형태의 분포를 따를 것이므로 출력 데이터  $\hat{\mathbf{X}}$ 은 평균  $\mu_{\hat{X}}$ , 표준편차  $\sigma_{\hat{X}}$ 의 가우시안 분포를 갖는다. 즉,  $\hat{\mathbf{X}} \sim N(\mu_{\hat{X}}, \sigma_{\hat{X}}^2)$ 이다. 입력 데이터  $\mathbf{X}$ 와 출력 데이터  $\hat{\mathbf{X}}$ 의 차이를 의미하는 오토인코더의 학습 손실을 계산하기 위하여 식(4)와 같이  $\bar{\mathbf{X}}, \mathbf{Y}$ 를 정의한다.

$$\begin{aligned} \bar{\mathbf{X}} &= -\hat{\mathbf{X}}, \\ \mathbf{Y} &= \mathbf{X} + (-\hat{\mathbf{X}}) = \mathbf{X} + \bar{\mathbf{X}} \end{aligned} \quad (4)$$

이때,  $\mathbf{Y}$ 는 입력 데이터와 출력 데이터의 차를 나타내는 변수이며,  $\bar{\mathbf{X}}$ 는 출력 데이터  $\hat{\mathbf{X}}$ 에 음수를 취한 값으로, 식(7)의 합성곱(convolution)의 계산을 위하여 정의한 변수이다. 입력 데이터 분포와 출력 데이터의 분포 간의 거리를 계산하기 위해 입력 데이터  $\mathbf{X}$ 의 확률밀도함수를 나타낸 식(5)와 출력 데이터  $\hat{\mathbf{X}}$ 의 음수 값  $\bar{\mathbf{X}}$ 의 확률밀도함수를 나타낸 식(6)을 정의한다.

$$f_X(x) = \frac{1}{\sigma_X \sqrt{2\pi}} e^{-\frac{(x - \mu_X)^2}{2\sigma_X^2}} \quad (5)$$

$$f_{\bar{X}}(\bar{x}) = f_{\hat{X}}(-\hat{x}) = \frac{1}{\sigma_{\hat{X}} \sqrt{2\pi}} e^{-\frac{(\hat{x} + \mu_{\hat{X}})^2}{2\sigma_{\hat{X}}^2}} \quad (6)$$

$\mathbf{Y}$ 의 확률밀도함수  $f_Y$ 는 입력 데이터 확률밀도함수  $f_X$ 와 출력 데이터 확률밀도함수  $f_{\hat{X}}$  간의 거리를

의미한다. 즉,  $f_Y$ 는 식(7)과 같이  $\mathbf{X}$ 의 확률밀도함수  $f_X$ 와  $\bar{\mathbf{X}}$ 의 확률밀도함수  $f_{\bar{X}}$ 의 합성곱 형태이다.

$$f_Y(y) = f_X(x) * f_{\bar{X}}(\bar{x}) \quad (7)$$

입력 데이터  $\mathbf{X}$ 와 출력 데이터  $\hat{\mathbf{X}}$ 이 모두 가우시안 분포를 가지므로, 합성곱인  $\mathbf{Y}$  또한 가우시안 분포를 가진다. 식(7)에 입력 데이터  $\mathbf{X}$ 의 확률밀도함수를 나타낸 식(5)와 출력 데이터  $\hat{\mathbf{X}}$ 의 음수값  $\bar{\mathbf{X}}$ 의 확률밀도함수를 나타낸 식(6)을 대입하여 계산하면, 입력 데이터와 출력 데이터 간 거리  $\mathbf{Y}$ 의 확률밀도함수  $f_Y$ 는 식(8)과 같다.

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{\bar{X}}(y-x)f_X(x)dx \\ &= \frac{1}{\sqrt{2\pi(\sigma_X^2 + \sigma_{\bar{X}}^2)}} e^{-\frac{(y - (\mu_X + \mu_{\bar{X}}))^2}{2(\sigma_X^2 + \sigma_{\bar{X}}^2)}} \\ &= \frac{1}{\sqrt{2\pi(\sigma_X^2 + \sigma_{\bar{X}}^2)}} e^{-\frac{(y - (\mu_X - \mu_{\bar{X}}))^2}{2(\sigma_X^2 + \sigma_{\bar{X}}^2)}} \quad (8) \\ &= \frac{1}{\sqrt{2\pi\sigma_Y^2}} e^{-\frac{(y - \mu_Y)^2}{2\sigma_Y^2}} \end{aligned}$$

식(8)에서 오토인코더의 학습 과정에서 발생한 입력 데이터와 출력 데이터 간 오차의 분포  $f_Y$ 는 평균  $\mu_Y = \mu_X + \mu_{\bar{X}} = \mu_X - \mu_{\hat{X}}$ 이고, 분산  $\sigma_Y^2 = \sigma_X^2 + \sigma_{\bar{X}}^2 = \sigma_X^2 + \sigma_{\hat{X}}^2$ 인 가우시안 분포이다. 즉,  $\mathbf{Y} \sim N(\mu_X - \mu_{\hat{X}}, \sigma_X^2 + \sigma_{\hat{X}}^2)$ 이다.

오토인코더의 학습 손실은 식(3)과 같이 입력 데이터와 출력 데이터의 차이의 절댓값으로 정의된다. 그러므로 오토인코더의 학습 손실 값은  $\mathbf{Y}$ 에 절댓값을 취한 후 기댓값을 계산한 식(9)와 같다.

$$E(|\mathbf{Y}|) = \sum_{i=1}^n |y_i| p_i \quad (9)$$

오토인코더의 학습 손실 분포를 나타내기 위해 식(10)과 같이  $\hat{\mathbf{Y}}$ 을 설정한다.

$$\hat{\mathbf{Y}} = |\mathbf{Y}| \quad (10)$$

$\hat{\mathbf{Y}}$ 의 확률밀도함수  $f_{\hat{Y}}$ 은  $y \geq 0$ 인 부분은  $f_Y$ 와 같고,  $y < 0$ 인 부분은  $y$ 축 기준으로 대칭이동 시킨 형태이므로 식(11)과 같이 나타낼 수 있다.

$$\begin{aligned} f_{\hat{Y}}(y) &= \frac{1}{\sigma_Y\sqrt{2\pi}} e^{-\frac{(y - \mu_Y)^2}{2\sigma_Y^2}} \\ &+ \frac{1}{\sigma_Y\sqrt{2\pi}} e^{-\frac{(y + \mu_Y)^2}{2\sigma_Y^2}} \quad (11) \end{aligned}$$

오토인코더의 학습 손실 값을 도출하기 위하여, 본문에서는 식(12)와 같이 정의된 오차함수(error function)  $erf(x)$ 를 사용한다.

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (12)$$

이를 이용하여  $f_{\hat{Y}}$ 의 기댓값  $E(\hat{\mathbf{Y}})$ 은

$$\begin{aligned} E(\hat{\mathbf{Y}}) &= \sigma_Y \sqrt{\frac{2}{\pi}} e^{-\frac{\mu_Y^2}{2\sigma_Y^2}} \\ &+ \mu_Y erf\left(\frac{\mu_Y}{\sqrt{2}\sigma_Y}\right) \quad (13) \end{aligned}$$

와 같이 나타낼 수 있다. 평균  $\mu_Y = \mu_X - \mu_{\hat{X}}$ 이고, 분산  $\sigma_Y^2 = \sigma_X^2 + \sigma_{\hat{X}}^2$ 이므로,

$$\begin{aligned} E(\hat{\mathbf{Y}}) &= \sqrt{\sigma_X^2 + \sigma_{\hat{X}}^2} \sqrt{\frac{2}{\pi}} e^{-\frac{(\mu_X - \mu_{\hat{X}})^2}{2(\sigma_X^2 + \sigma_{\hat{X}}^2)}} \\ &+ (\mu_X - \mu_{\hat{X}}) erf\left(\frac{\mu_X - \mu_{\hat{X}}}{\sqrt{2(\sigma_X^2 + \sigma_{\hat{X}}^2)}}\right) \quad (14) \end{aligned}$$

와 같이 나타낼 수 있다.

오토인코더의 구성에 의하여 결정되는 출력 데이터의 특성을 파라미터  $\sigma_{\hat{X}}, \mu_{\hat{X}}$ 로 나타내고,  $a = \sigma_X^2 - \sigma_{\hat{X}}^2$ ,  $b = \mu_X - \mu_{\hat{X}}$ 로 정의하여 식(14)에 대입하면 아래 식(15)와 같이 나타낼 수 있다. 식(15)는 입력 데이터의 표준편차  $\sigma_X$ 에 관한 함수이다.

$$\begin{aligned} E(\hat{\mathbf{Y}}) &= \sqrt{2\sigma_X^2 + a} \sqrt{\frac{2}{\pi}} e^{-\frac{b^2}{4\sigma_X^2 + 2a}} \\ &+ b erf\left(\frac{b}{\sqrt{4\sigma_X^2 + 2a}}\right) \quad (15) \end{aligned}$$

오토인코더는 입력 데이터와 출력 데이터의 차이를 최소화하는 방향으로 학습하기 때문에, 오토인코더의 학습이 목적에 맞게 이루어졌다면 입력 데이터와 출력 데이터의 무작위도가 유사할 것이다. 따라서  $a \approx 0$ ,  $b \approx 0$ 로 근사하여, 식(16)과 같이 나타낼 수 있다.

$$E(\hat{Y}) \approx \sqrt{2\sigma_X^2 + a} \sqrt{\frac{2}{\pi}} \quad (16)$$

식(16)에서  $E(\hat{Y})$ 의 값은  $\sqrt{2\sigma_X^2 + a}$ 의 값에 따라 달라지는 것을 알 수 있다. 이때  $a$ 는 고정된 값이므로 식(17)과 같이  $\sigma_X \geq 0$ 일 때  $E(\hat{Y})$ 은  $\sigma_X$ 와 비례한다.

$$E(\hat{Y}) \propto \sigma_X \quad (17)$$

$E(\hat{Y})$ 와  $\sigma_X$ 가 비례하므로 둘 사이의 선형 관계가 성립한다. 선형 관계의 기울기를 나타내는  $\alpha$ 를 정의하여, 식(17)을 식(18)과 같이 나타낼 수 있다.

$$E(\hat{Y}) = \alpha \sigma_X \quad (18)$$

오토인코더의 학습 손실  $L(\mathbf{X}, \hat{\mathbf{X}})$ 을 의미하는  $E(\hat{Y})$ 은  $\sigma_X$ 와 선형 관계이며 직선의 경사도  $\alpha$ 는 오토인코더의 구성 요소들인 활성화 함수, 손실함수, 은닉층의 구성, 에포크 등에 의해 결정된다.

#### IV. 입력 데이터의 무작위도와 오토인코더 학습 성능 관계 실험

##### 4.1 오토인코더 입출력 데이터 분포에 대한 실험

III장 2절에서 논의한 것과 같이, 은닉층의 압축된 데이터를 통해 출력 데이터가 입력 데이터와 유사하게 복원되도록 오토인코더의 학습이 이루어졌다면, 출력 데이터 또한 입력 데이터와 동일한 형태의 분포를 따른다는 것을 실험적으로 확인하고자 한다.

본 실험에서는 평균이 10, 표준편차가 1인 가우시안 분포를 따르는 500,000개의 데이터  $\mathbf{X} \sim N(10, 1)$ 를 생성한 후, 이를 오토인코더의 입력 데이터로 사용한다. 이때 활성화 함수, 학습률, 에포크를 각각 LeakyReLU, 0.01, 100으로 설정하였으며, 은닉층의 노드 개수는 50개로 설정하였다.

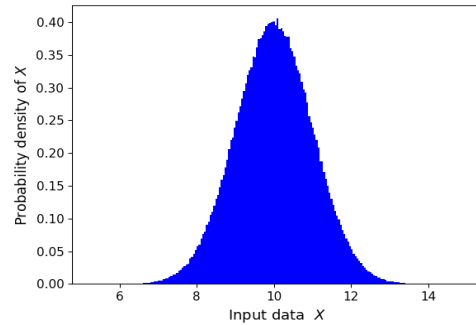


그림 3. 입력 데이터  $\mathbf{X} \sim N(10, 1)$ 의 확률밀도함수  
Fig. 3. Probability density function of input data  $\mathbf{X} \sim N(10, 1)$

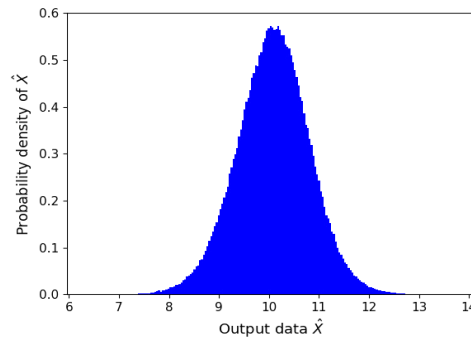


그림 4.  $\mathbf{X} \sim N(10, 1)$ 를 오토인코더의 입력으로 사용하였을 때의 출력 데이터  $\hat{\mathbf{X}}$ 의 확률밀도함수  
Fig. 4. Probability density function of output data  $\hat{\mathbf{X}}$  as an outcome of  $\mathbf{X} \sim N(10, 1)$

그림 3은 오토인코더의 입력으로 사용된 데이터  $\mathbf{X}$ 에 대한 확률밀도함수  $f_X$ 를 나타낸 것이고 그림 4는  $\mathbf{X}$ 를 입력 데이터로 사용하였을 때의 오토인코더의 출력  $\hat{\mathbf{X}}$ 에 대한 확률밀도함수  $f_{\hat{X}}$ 를 나타낸 것이다. 그림 4가 그림 3과 같은 가우시안 분포 형태인 것을 통해, 오토인코더의 출력 데이터의 분포는 입력 데이터의 분포와 동일한 형태임을 확인할 수 있다.

##### 4.2 입력 데이터의 무작위도에 따른 오토인코더 학습 성능 실험

III장에서 이론적으로 증명한 입력 데이터의 무작위도와 오토인코더의 학습 손실이 선형 관계에 있음을 실험적으로 확인하고자 한다. 무작위도가 서로 다른 가우시안 분포를 따르는 입력 데이터를 생성한 후, 동일한 조건의 실험을 반복하여 입력 데이터의 무작위도가 오토인코더의 학습 성능에 미치는 영향을 확인

한다. 또한 입력 데이터의 무작위도와 오토인코더의 학습 손실이 오토인코더의 은닉층의 구조에 영향을 받지 않음을 확인하기 위하여 기본 오토인코더와 적층 오토인코더를 이용한 실험을 각각 진행한다.

평균  $\mu_X$ , 표준편차  $\sigma_X$ 의 가우시안 분포를 가지는 500,000개의 데이터를 생성하였다. 생성한 데이터를 오토인코더의 입력으로 사용하기 위해, 수를 100개씩 나누어 길이가 100인 5,000개의 데이터  $\mathbf{X}$ 를 만들었다.  $\mathbf{X}$ 를 오토인코더의 입력 데이터로 사용하였으며, 오토인코더의 활성화 함수는 LeakyReLU, 에포크는 100으로 설정하였다. 이는 오토인코더 학습 과정에서 과적합이 일어나지 않고, 에포크가 증가할수록 학습 손실이 줄어드는 안정적인 학습 조건을 실험적으로 확인한 것이다. 본 실험에서는 입력 데이터  $\mathbf{X}$ 의 평균은  $\mu_X \in \{10, 20, 30\}$ 으로 설정하고 순차적으로 번갈아 가며 변화시킨다. 이때 입력 데이터  $\mathbf{X}$ 의 표준편차는  $\sigma_X \in \{1, 2, 3, \dots, 30\}$ 로,  $\sigma_X$ 를 1부터 30까지 1씩 증가시킨다. 활성화 함수 LeakyReLU는 입력이 음수일 때는 출력을 작은 값으로 축소시키는 특성이 있으므로,  $\mu_X > 0$ 인 데이터만 사용한다.

4.2.1 입력 데이터의 무작위도에 따른 기본 오토인코더 학습 성능의 변화 실험

은닉층이 1개인 기본 구조의 오토인코더에서 입력 데이터의 무작위도가 오토인코더의 학습 성능에 미치는 영향을 알아보기 위한 실험을 진행한다. 학습률은 0.01로 설정하였고, 은닉층의 노드 개수는  $h \in \{10, 20, 30, 40, 50, 60, 70\}$ 로 변화시켜가며 실험하였다.

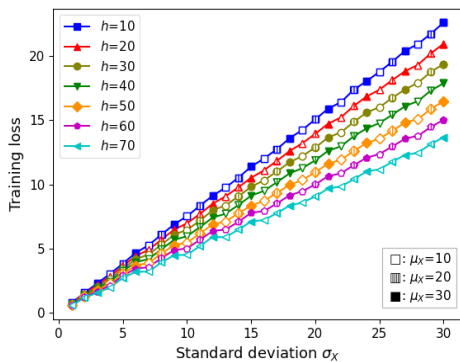


그림 5. 기본 오토인코더에서 은닉층의 노드 개수를  $h$ 개로 변화시켰을 때 표준편차에 따른 학습 손실 변화  
Fig. 5. Training loss according to standard deviation from different number of nodes in single hidden layer  $h$  of basic autoencoder

그림 5는 은닉층의 노드 개수( $h$ )를 변화시켰을 때 표준편차에 따른 오토인코더의 학습 손실을 나타낸다.  $\sigma_X$ 가 커질수록 오토인코더의 학습 손실이 증가하는데, 이는  $\sigma_X$ 가 클수록 입력 데이터의 무작위도가 커져 학습 성능이 저하되기 때문이다. 또한, 그림 5에서  $\sigma_X$ 가 클수록 은닉층의 노드 개수 증가에 따른 학습 성능 개선율이 높아지는 것을 확인할 수 있다.  $\mu_X \in \{10, 20, 30\}$ 의 값에 관계없이 학습 성능은  $\sigma_X$ 와 선형인 것을 관찰할 수 있다. 따라서 입력데이터  $\mathbf{X} \sim N(\mu_X, \sigma_X^2)$ 의 무작위도에 영향을 미치지 않는 평균값은 오토인코더의 학습 성능에도 영향을 미치지 않는다는 것을 확인할 수 있다.

4.2.2 입력 데이터의 무작위도에 따른 적층 오토인코더 학습 성능의 변화 실험

은닉층이 3개인 적층 오토인코더에서 입력 데이터의 무작위도가 오토인코더의 학습 성능에 미치는 영향을 알아보기 위한 실험을 진행한다. 적층 오토인코더는 대칭 구조이므로 첫 번째와 세 번째 은닉층의 노드 개수는 동일하게 설정하고, 두 번째 은닉층의 노드 개수는 첫 번째와 세 번째 은닉층의 노드 개수보다 작거나 같게 설정한다. 적층 오토인코더의 첫 번째와 세 번째 은닉층의 노드 개수는 70개로 고정하였고, 두 번째 은닉층의 노드 개수를  $h \in \{10, 20, 30, 40, 50, 60, 70\}$ 로 변화시켜 실험하였으며, 학습률은 0.001로 설정하였다. 오토인코더 학습 과정에서 과적합이 일어나지 않고, 에포크가 증가할수록 학습 손실이 줄어드

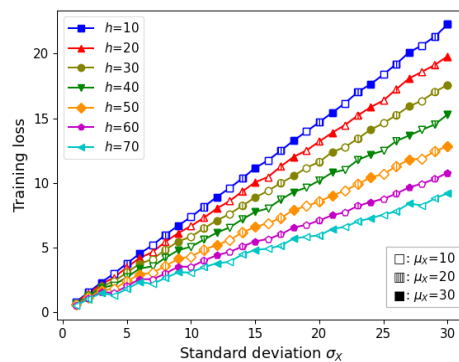


그림 6. 적층 오토인코더에서 은닉층의 노드 개수를  $h$ 개로 변화시켰을 때 표준편차에 따른 학습 손실 변화  
Fig. 6. Training loss according to standard deviation from different number of nodes in hidden layer  $h$  of stacked autoencoder

는 안정적인 학습 조건을 실험적으로 확인한 것이다.

그림 6은 은닉층의 노드 개수( $h$ )를 변화시켰을 때 표준편차에 따른 적층 오토인코더의 학습 손실을 나타낸다. 그림 6에서  $\sigma_X$ 가 커질수록 적층 오토인코더의 학습 손실이 증가하는 결과를 관찰할 수 있다. 또한, 그림 6에서  $\sigma_X$ 가 클수록 은닉층의 노드 개수 증가에 따른 학습 성능 개선율이 높아지는 것을 확인할 수 있다.  $\mu_X \in \{10, 20, 30\}$ 와 관계없이 학습 성능이  $\sigma_X$ 와 선형인 결과를 바탕으로 은닉층의 개수가 여러 개인 적층 오토인코더의 경우에도  $\mu_X$ 는 학습 성능에 영향을 미치지 않는다는 것을 알 수 있다.

2절 1항의 그림 5와 2항의 그림 6을 통해 기본 구조의 오토인코더와 적층 오토인코더 모두  $\sigma_X$ 에 대하여 학습 성능이 식(15)에 나타난 결과와 같이 선형으로 나타남을 알 수 있다. III장에서 이론적으로 증명한 것과 같이 은닉층의 수와 관계없이 입력 데이터의 무작위도가 증가할수록 오토인코더의 학습 성능이 저하되는 결과를 관찰할 수 있으며, 은닉층의 노드 개수( $h$ )가 증가할수록 학습 손실은 줄어든다는 사실을 실험적으로 확인하였다. 표 1은 은닉층의 노드 개수에 따른 선형 계수  $\alpha$ 를 정리한 것이다. 본 논문에서는 은닉층의 노드 개수 이외에 나머지 구성 요소는 모두 고정시켰으므로,  $\alpha$ 는 은닉층의 노드 개수에 따라 결정된다. 표 1에서 은닉층의 수와 관계없이  $h$ 가 커질수록  $\alpha$ 가 감소하는 것을 통해, 은닉층의 노드 개수가 증가할수록 입력 데이터의 무작위도의 영향을 적게 받는다는 것을 알 수 있다. 또한,  $h$ 가 동일한 경우 적층 오토인코더의 학습 성능이 더 뛰어나다는 것을 확인하였다.

표 1. 은닉층의 노드 개수  $h$ 에 따른 선형 계수  $\alpha$ 값  
Table 1. Value of  $\alpha$  for different number of nodes in the hidden layer  $h$

		$\alpha$	
		Basic autoencoder	Stacked autoencoder
$h$	10	0.7491	0.7359
	20	0.6919	0.6566
	30	0.6383	0.5807
	40	0.5869	0.4984
	50	0.5373	0.4209
	60	0.4882	0.3447
	70	0.4408	0.2900

## V. 결 론

본 논문에서는 입력 데이터의 무작위도가 오토인코더의 학습 성능에 미치는 영향에 대해 분석하고, 입력 데이터의 무작위도를 이용하여 은닉층의 노드 개수를 설정하는 방안에 관한 연구를 진행하였다. 본 논문에서는 가우시안 분포를 가지는 입력 데이터를 활용하여 입력 데이터의 무작위도를 의미하는 지표로 표준편차를 사용해 오토인코더의 학습 손실과 선형 관계임을 증명하였다. 실험을 통해 입력 데이터의 표준편차와 학습 손실이 양의 선형 관계를 가지는 것을 확인하였다. 또한 은닉층의 노드 개수가 증가할수록, 무작위도에 따른 오토인코더의 학습 성능의 변화가 적다는 사실을 알아내었다. 이 결과는 적층 오토인코더에도 적용되는 것을 실험적으로 확인하였다. 또한 은닉층의 노드 개수가 증가할수록, 무작위도에 따른 오토인코더의 학습 성능의 변화가 적다는 사실을 알아내었다. 이 결과는 적층 오토인코더에도 적용되는 것을 실험적으로 확인하였다. 따라서 오토인코더의 은닉층의 노드 개수를 결정할 때 입력 데이터의 무작위도를 계산하여 학습 성능을 예측할 수 있다.

## References

- [1] C. Wang, F. Haider, X. Gao, X. You, Y. Yang, D. Yuan, H. M. Aggoune, H. Haas, S. Fletcher, and E. Hepsaydir, "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 122-130, Feb. 2014.
- [2] C. Wang, M. D. Renzo, S. Stanczak, S. Wang, and E. G. Larsson, "Artificial intelligence enabled wireless networking for 5G and beyond: Recent advances and future challenges," *IEEE Wireless Commun.*, vol. 27, no. 1, pp. 16-23, Feb. 2020.
- [3] D. Choi, S. Chun, M. Kang, E. Cho, and T. T. Kwon, "A proposal of LSTM-based recurrent neural network model to predict per-service network resources in data center," in *Proc. Symp. KICS*, pp. 1376-1377, Yongpyung, Korea, Jan. 2019.
- [4] H. Park and Y. Lim, "Cell activation algorithm with reinforcement learning in



- mobile ultra dense network,” *J. KICS*, vol. 45, no. 2, pp. 293-302, 2020.
- [5] R. Boutaba, M. A. Salahuddin, N. Limam, S. Ayoubi, N. Shahriar, F. E. Solano, and O. M. Caicedo, “A comprehensive survey on machine learning for networking: Evolution, applications and research opportunities,” *J. Internet Serv. and Appl.*, vol. 9, no. 16, pp. 1-99, 2018.
- [6] K. I. Ahmed, H. Tabassum, and E. Hossain, “Deep learning for radio resource allocation in multi-cell networks,” *IEEE Netw.*, vol. 33, no. 6, pp. 188-195, Nov.-Dec. 2019.
- [7] Z. Chen, C. K. Yeo, B. S. Lee, and C. T. Lau, “Autoencoder-based network anomaly detection,” *2018 WTS*, pp. 1-5, Phoenix, AZ, USA, Apr. 2018.
- [8] H. Choi, M. Kim, G. Lee, and W. Kim, “Unsupervised learning approach for network intrusion detection system using autoencoder,” *J. Supercomputing*, vol. 75, no. 9, pp. 5597-5621, Sep. 2019.
- [9] N. Bahadur and R. Paffenroth, “Dimension estimation using autoencoders with applications to financial market analysis,” *2020 19th IEEE ICMLA*, pp. 527-534, Miami, FL, USA, Dec. 2020.
- [10] D. Kwon, S. Jin, H. Kim, and S. Hong, “Improving non-profiled side-channel analysis using auto-encoder based noise reduction preprocessing,” *J. KIISC*, vol. 29, no. 3, pp. 491-501, Jun. 2019.
- [11] H. Lee and Y. Jung, “Comparison of deep learning-based autoencoders for recommender systems,” *The Korean J. Appl. Statist.*, vol. 34, no. 3, pp. 329-345, Jun. 2021.
- [12] G. Liu, H. Bao, and B. Han, “A stacked autoencoder-based deep neural network for achieving gearbox fault diagnosis,” *Math. Problems in Eng.*, vol. 2018, pp. 1-10, Jul. 2018.
- [13] L. Vu and Q. U. Nguyen, “An ensemble of activation functions in autoencoder applied to IoT anomaly detection,” *2019 6th NAFOSTED Conf. Inf. and Comput. Sci. (NICS)*, pp. 534-539, Hanoi, Vietnam, Dec. 2019.
- [14] W. W. Ng, G. Zeng, J. Zhang, D. S. Yeung, and W. Pedrycz, “Dual autoencoders features for imbalance classification problem,” *Pattern Recognition*, vol. 60, pp. 875-889, Dec. 2016.
- [15] S. Sinha, T. N. Singh, V. K. Singh, and A. K. Verma, “Epoch determination for neural network by self-organized map (SOM),” *Computational Geosci.*, vol. 14, no. 1, pp. 199-206, Jan. 2010.
- [16] Q. Xu, C. Zhang, L. Zhang, and Y. Song, “The learning effect of different hidden layers stacked autoencoder,” *2016 8th Int. Conf. IHMSC*, pp. 148-151, Hangzhou, China, Aug. 2016.

이 주 흥 (Joohong Rhee)



2021년 8월 : 이화여자대학교 전  
자전기공학과 졸업  
2021년 8월~현재 : 이화여자대  
학교 전자전기공학과 석사  
<관심분야> 딥러닝, 머신러닝,  
머신러닝을 활용한 데이터 분  
석

[ORCID:0000-0003-1798-0247]

정 다 은 (Daeun Jung)



2019년 2월 : 이화여자대학교 전  
자전기공학과 졸업  
2021년 8월 : 이화여자대학교 전  
자전기공학과 석사  
2021년 8월~현재 : 이화여자대  
학교 전자전기공학과 석사 후  
연구원

<관심분야> 시계열 데이터, 분산 학습, 표현 학습  
[ORCID:0000-0002-6701-0380]

박형곤 (Hyunggon Park)



2004년 2월 : 포항공과대학교 전자전기공학과 졸업

2006년 3월 : University of California, Los Angeles (UCLA) M.S.

2008년 12월 : University of California, Los Angeles (UCLA) Ph.D.

2010년~현재 : 이화여자대학교 전자전기공학과 교수  
<관심분야> 멀티에이전트 시스템 최적화, 머신러닝, 인공지능, 게임이론

[ORCID:0000-0002-5079-1504]