

# Batch Normalization

practice 인공지능,머신러닝 · 2024. 10. 12. 10:47

## Batch Normalization

- Batch Normalization은 학습 과정에서 각 층의 입력 분포를 정규화 (Normalisation) 하여 학습 속도를 향상시키고 안정성을 높이는 기법입니다.
- 입력의 평균과 분산을 조정하여, 신경망의 층마다 동일한 분포를 유지하도록 합니다.

-----

- 따라서, 모든 값들을 Normalisation 할 필요가 있음

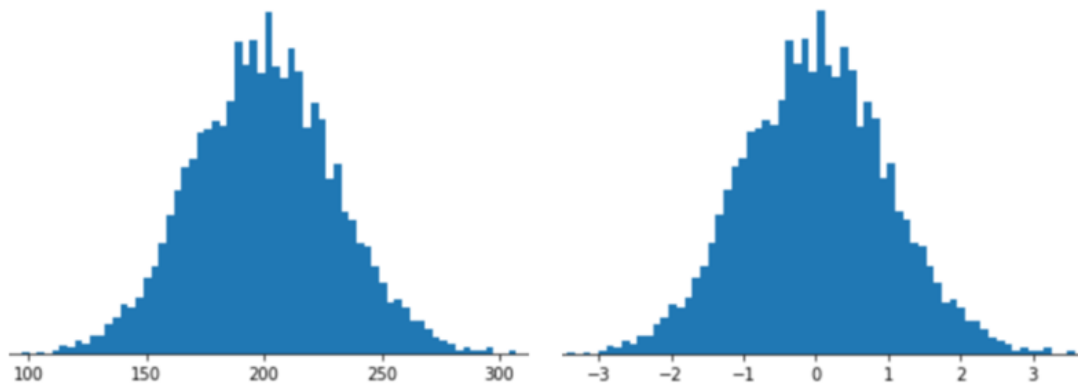
- Normalisation을 하면, 데이터의 평균은 0, 표준편차는 1이 됨

- 다만 이때 데이터의 “모양”은 그대로 유지됨

- 예를 들어, 어떤 집이 가장 넓은 크기를 가지고 있었으면, 이 정보는 Normalisation 후에도 그대로 유지됨

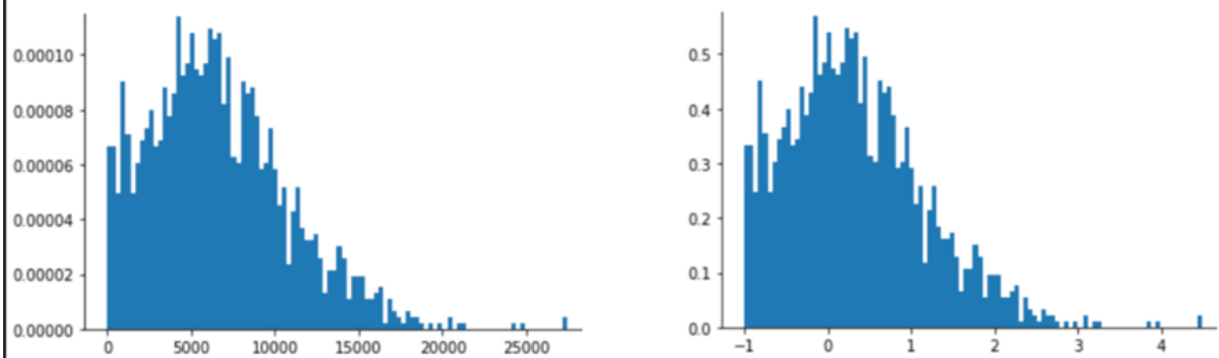
- 인공지능 학습에서, 일반적으로 데이터를 여러 개의 Batch로 나누어서 학습한다
- 이때 데이터의 값들이 다양한 범위를 가지고 있다면, 모델의 학습 속도가 느려질 수 있다
- 또한, 값들이 매우 크면 역전파 과정에서 기울기도 매우 커질 것이고
- 반대로 작은 값들은 기울기 또한 매우 작아질 것이다 (기울기 소실)

- On the left, a classic normal distribution.
- On the right, the same distribution normalized by Z-score scaling.

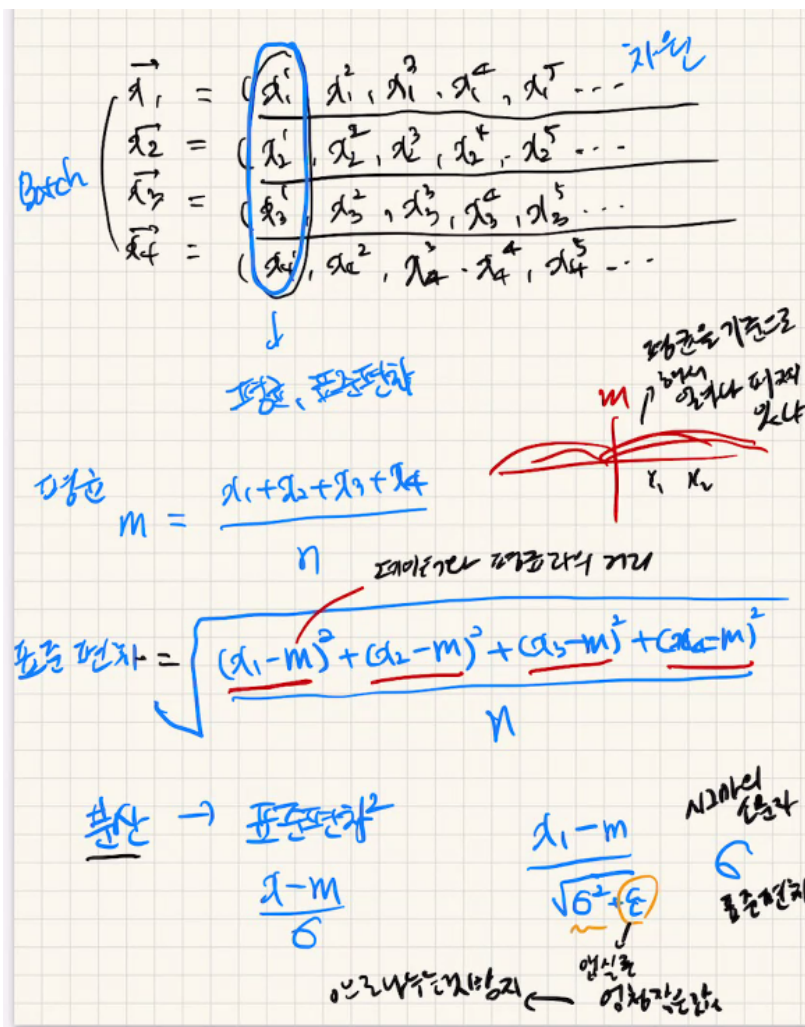


**Figure 4.** Raw data (left) versus Z-score (right) for a normal distribution.

Z-score scaling is also a good choice for data like that shown in the following figure, which has only a vaguely normal distribution.



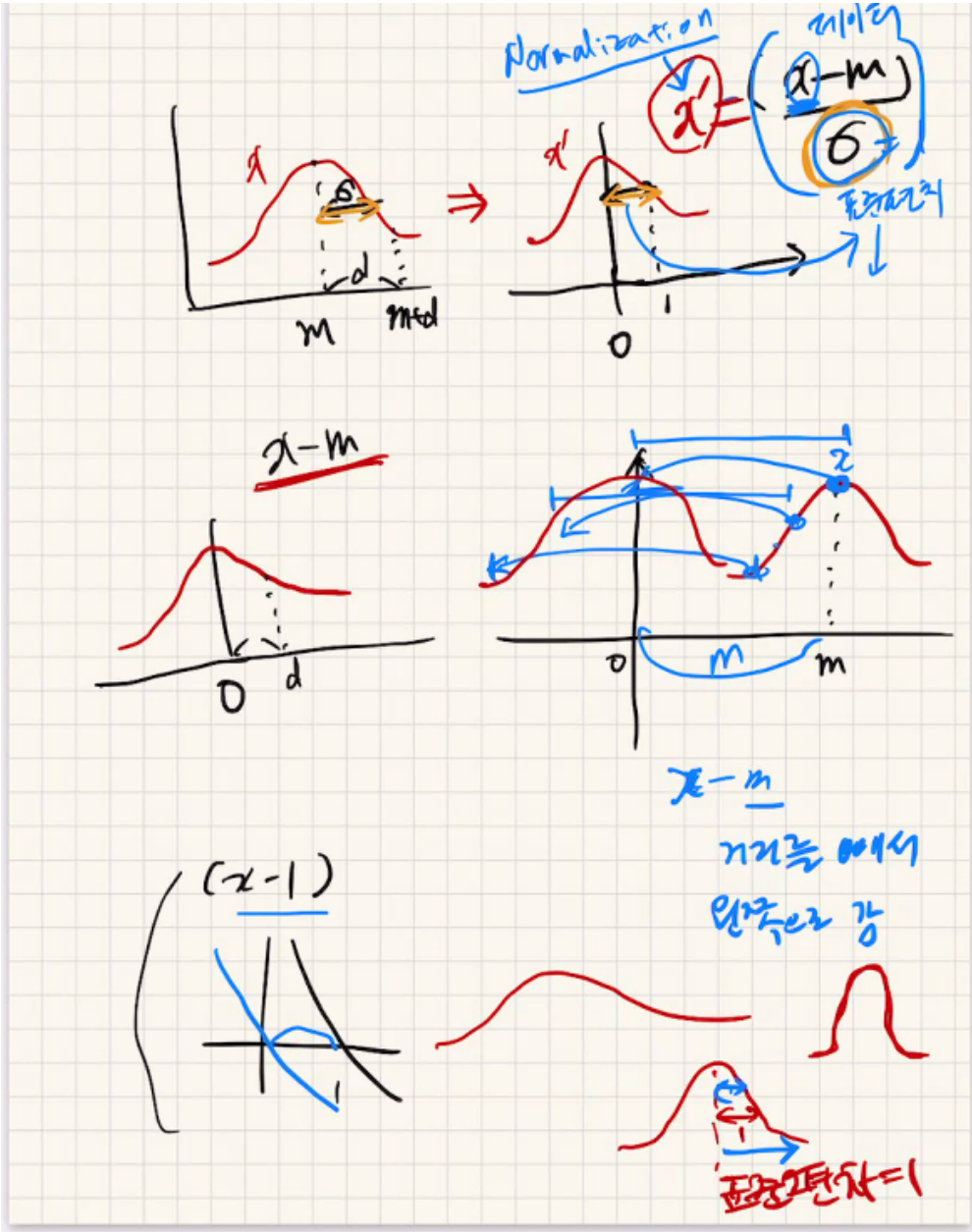
**Figure 5.** Raw data (left) versus Z-score scaling (right) for a non-classic normal distribution.



normalization

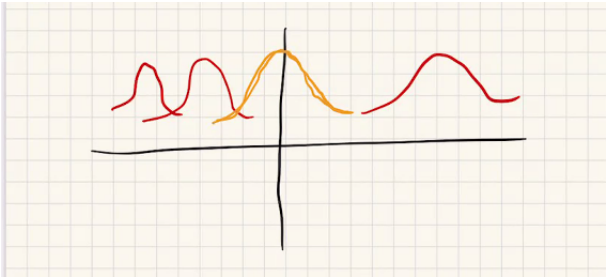
$m(\text{평균}) = 0$ , **표준편차=1**로 만들어준다. (원래 표준편차는 1이 같거나 아니었는데 1로만들)

batch normalization



batch normalization

- 데이터의 batch를 normalization한다.
- 정리를 하니 overfitting 이 된다.



Batch

$$\begin{aligned}\vec{x}_1 &= (x_1^1, x_1^2, x_1^3, x_1^4, x_1^5 \dots) \\ \vec{x}_2 &= (x_2^1, x_2^2, x_2^3, x_2^4, x_2^5 \dots) \\ \vec{x}_3 &= (x_3^1, x_3^2, x_3^3, x_3^4, x_3^5 \dots) \\ \vec{x}_4 &= (x_4^1, x_4^2, x_4^3, x_4^4, x_4^5 \dots)\end{aligned}$$

↓  
평균, 표준편차

평균  $m = \frac{x_1 + x_2 + x_3 + x_4}{n}$

표준편차 =  $\sqrt{\frac{(x_1 - m)^2 + (x_2 - m)^2 + (x_3 - m)^2 + (x_4 - m)^2}{n}}$

분산  $\rightarrow$  표준편차<sup>2</sup>  
 $\frac{x - m}{\sigma}$

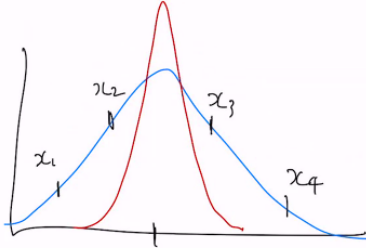
$\hat{x} = \frac{x_1 - m}{\sqrt{\sigma^2 + 1}}$

평균을 기준으로  
m  
x<sub>1</sub> x<sub>2</sub>

이제부터 평균과 분산을  
알아야 한다

이제부터 평균과 분산을  
알아야 한다

$x_1, x_2, x_3, x_4$



$$\frac{(x_1 - m) + (x_2 - m) + (x_3 - m) + (x_4 - m)}{n}$$

이제부터 데이터를 평균으로 바꿔 줄거나 데이터인 줄이 아니냐.

$\rightarrow$  평균으로 바꿔 편차의 평균.

근리 위의 식에서,  $(x_1 - m), (x_2 - m)$ 은 음수,  $(x_3 - m), (x_4 - m)$ 은 양수라서, 서로 제곱하면 0이 된다.

$\rightarrow$  제곱해서 편차들을 제곱해서 (음수  $\rightarrow$  양수 변환) 제곱한 것.

분산: 
$$\frac{(x_1 - m)^2 + (x_2 - m)^2 + (x_3 - m)^2 + (x_4 - m)^2}{n}$$

$=$  평균으로 바꿔 편차의 제곱의 평균

이때 표준편차 (편차의 평균)은, 단순히 제곱한 것  
원래대로 풀리는 것 (주로 쓰이기) 표준편차  $= \sqrt{\text{분산}}$

♡ 1



...

구독하기

'practice 인공지능,머신러닝' 카테고리의 다른 글

[Transformer, RNN 차이](#) (0)

2024.10.12

[Adam Optimizer](#) (1)

2024.10.12

[강화학습](#) (0)

2024.10.12

[앙상블 학습](#) (0)

2024.10.11

[SVM\(Support Vector Machine\)](#) (0)

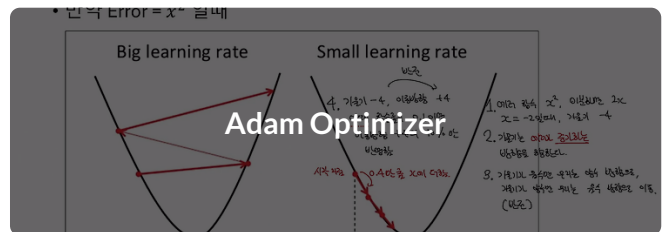
2024.10.11

관련글

[관련글 더보기](#)

Transformer, RNN 차이

강화학습



자연어(NLP)

네이쳐2024 님의 블로그입니다.

구독하기 +



익명



비밀댓글입니다.

2024. 10. 12. 11:13



익명



비밀댓글입니다.

2024. 10. 13. 14:40



익명



비밀댓글입니다.

2024. 10. 13. 14:40



이름

비밀번호

내용을 입력하세요.



답글