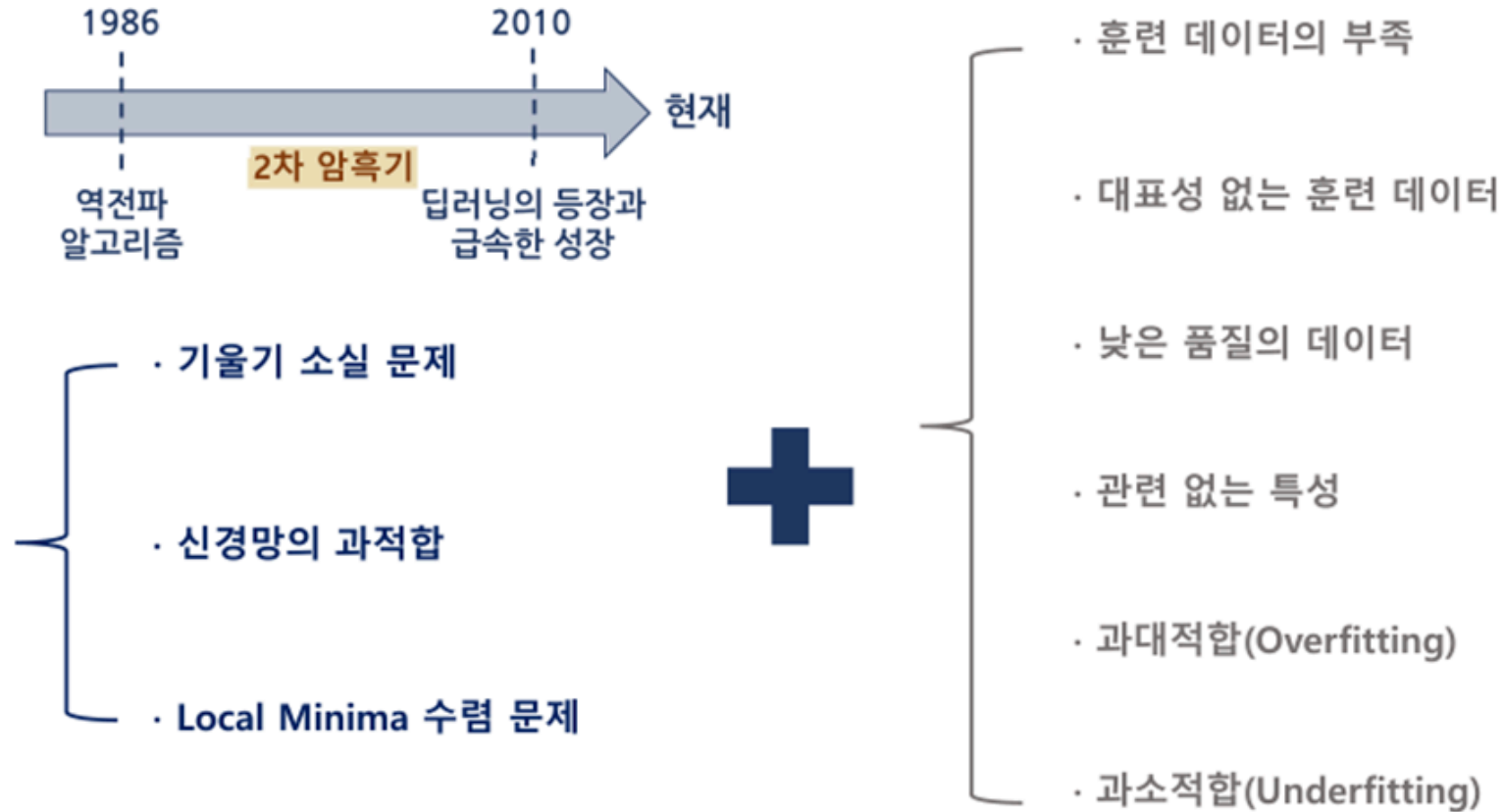


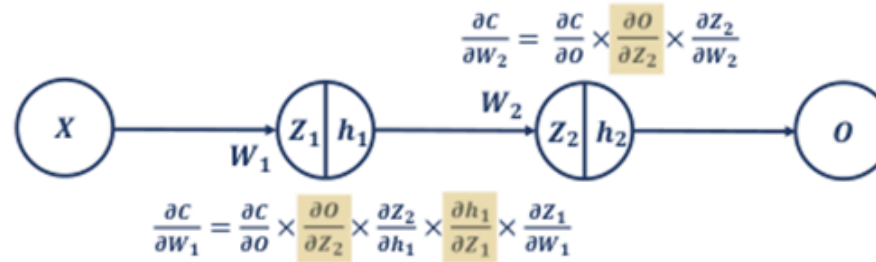
17강. 딥러닝 학습에서 마주하는 문제들

- 기울기 소실 문제
- RELU / 배치정규화
- 드롭아웃 / 데이터 증강 / 규제 / 조기종료
- 경사하강법 옵티마이저(Optimizer)

■ 딥러닝 학습에서 마주하는 문제들



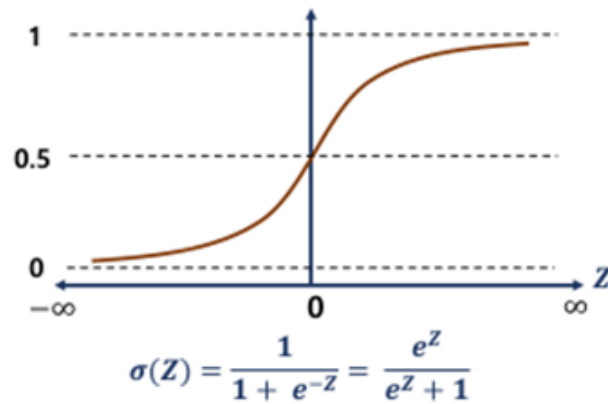
기울기 소실 문제



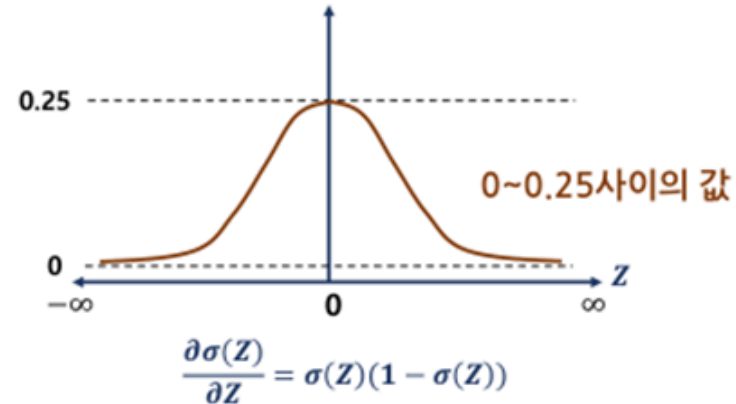
활성함수의 미분

층이 깊어질수록 활성화함수의 미분 값을 여러 번 곱해야 함

시그모이드 함수



시그모이드의 미분함수



■ 기울기 소실 문제

활성함수로 시그모이드를 사용하면 층이 깊어질수록 기울기가 전파가 되지 않는 현상



$$(1\text{층}) \quad \frac{\partial h}{\partial Z} \quad 0 \sim 0.25$$

$$(2\text{층}) \quad \frac{\partial h}{\partial Z} \times \frac{\partial h}{\partial Z} \quad 0 \sim 0.0625$$

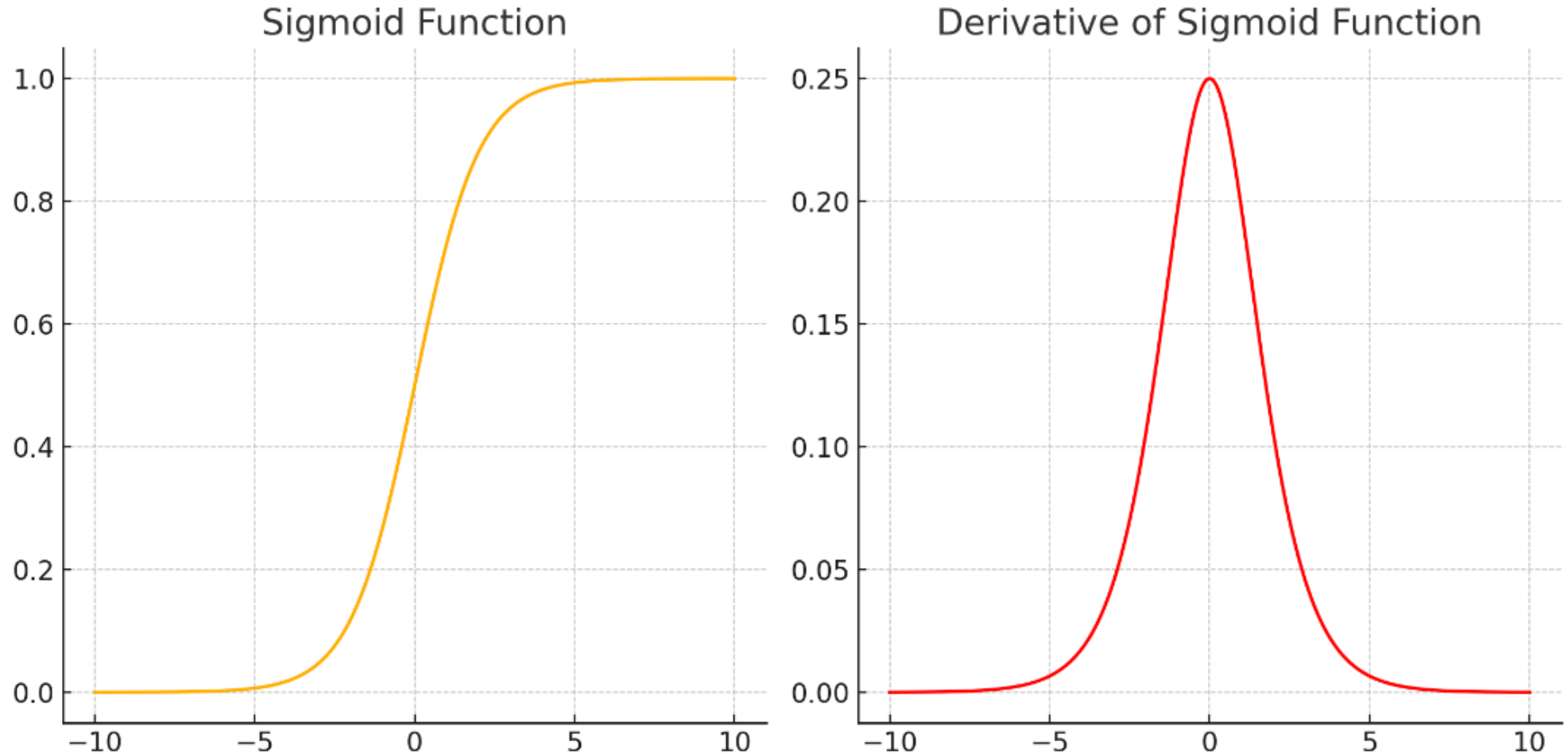
$$(3\text{층}) \quad \frac{\partial h}{\partial Z} \times \frac{\partial h}{\partial Z} \times \frac{\partial h}{\partial Z} \quad 0 \sim 0.015625$$

⋮

$$(\infty\text{층}) \quad \lim_{n \rightarrow \infty} \left(\frac{\partial h}{\partial Z} \right)^n = 0$$

✓ 활성화 함수

- Sigmoid나 tanh와 같은 활성화 함수는 입력 값이 매우 크거나 작을 때 출력의 변화가 매우 작다. 예를 들어 Sigmoid의 도함수는 아래와 같다.
- 여기서 $\sigma(x)$ 가 0 또는 1에 가까워지면 도함수가 0에 가까워진다. 이러한 활성화 함수들은 입력 값이 크거나 작을 때 기울기를 거의 0으로 만들기 때문에, 역전파 동안 기울기가 점점 더 작아진다.



$$W_i = W_{i-1} - \eta \frac{\delta Error}{\delta W}$$

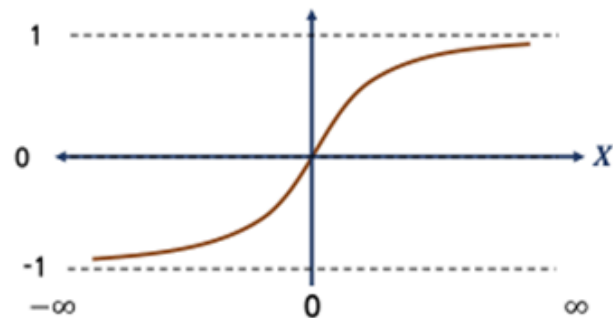
■ 기울기 소실 문제

극복 방안

극복 가능한 활성화함수

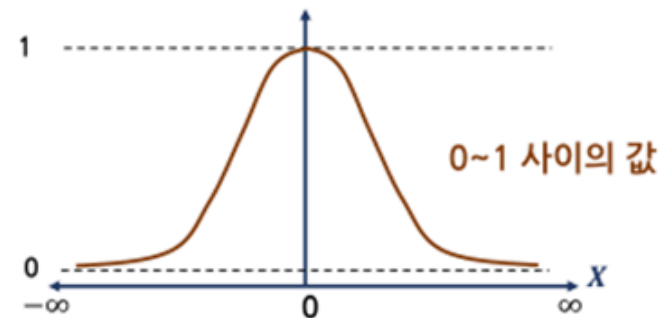
배치 정규화

하이퍼볼릭 탄젠트



$$\tanh(X) = \frac{e^X - e^{-X}}{e^X + e^{-X}}$$

하이퍼볼릭 탄젠트의 미분함수



0~1 사이의 값

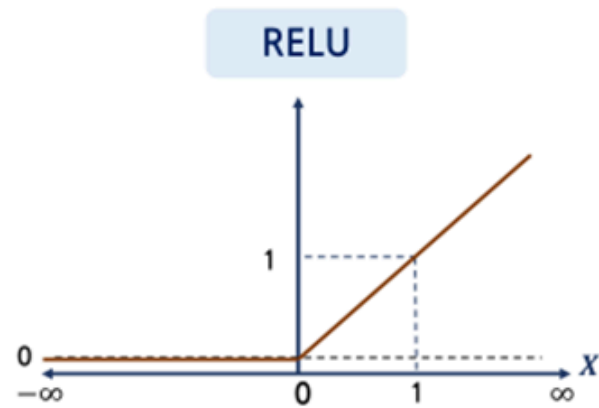
시그모이드 함수보다 기울기 소실을 지연

기울기 소실 문제

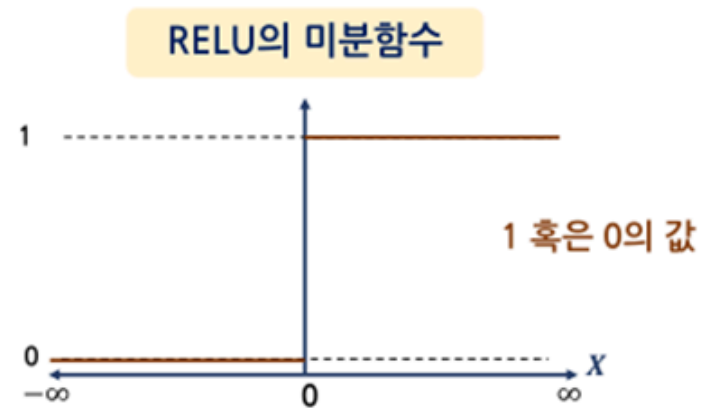
극복 방안

극복 가능한 활성화함수

배치 정규화

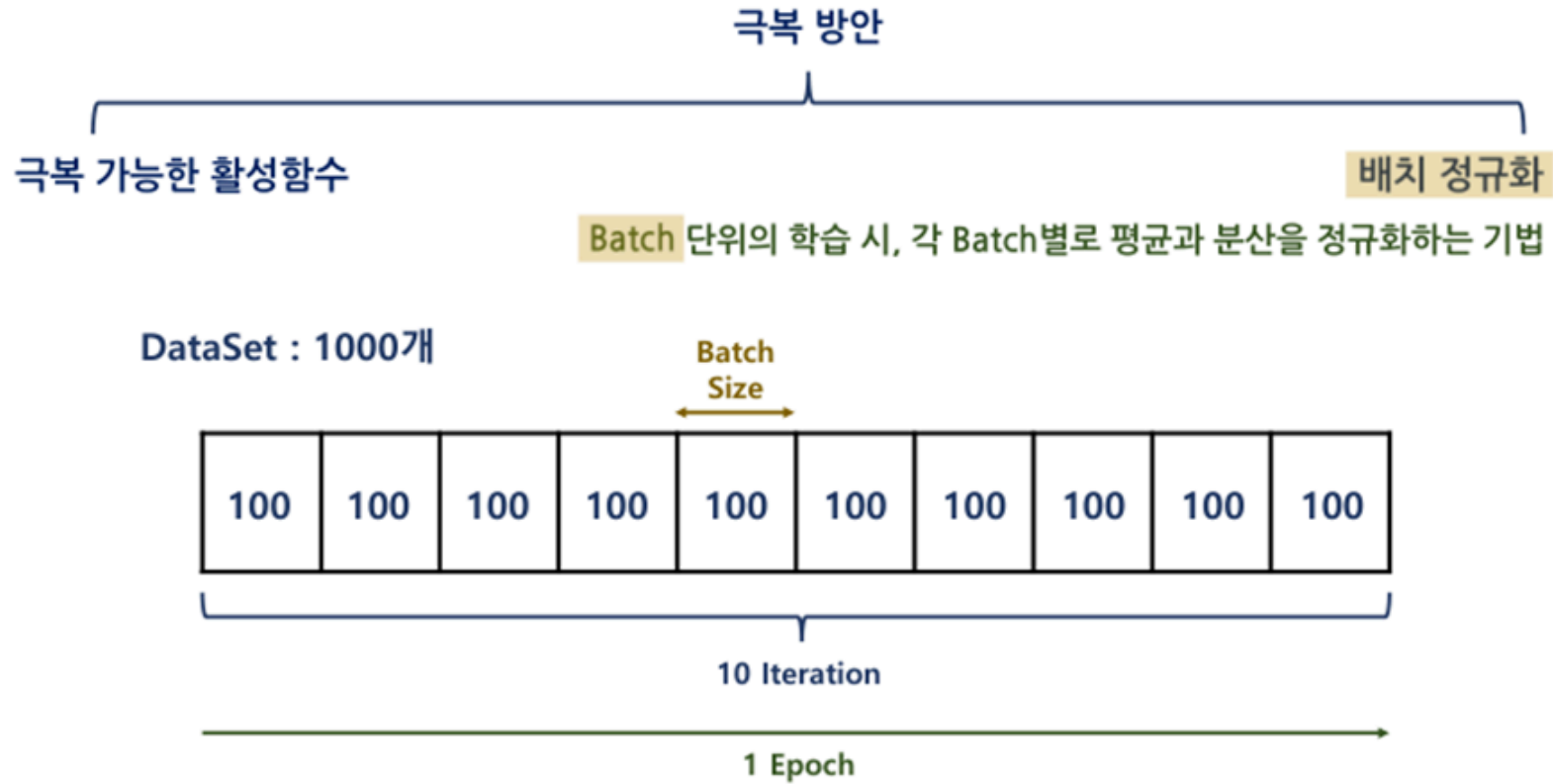


$$ReLU(X) = \text{Max}(0, X)$$



비선형성을 유지하며 기울기 소실문제를 극복

■ 기울기 소실 문제



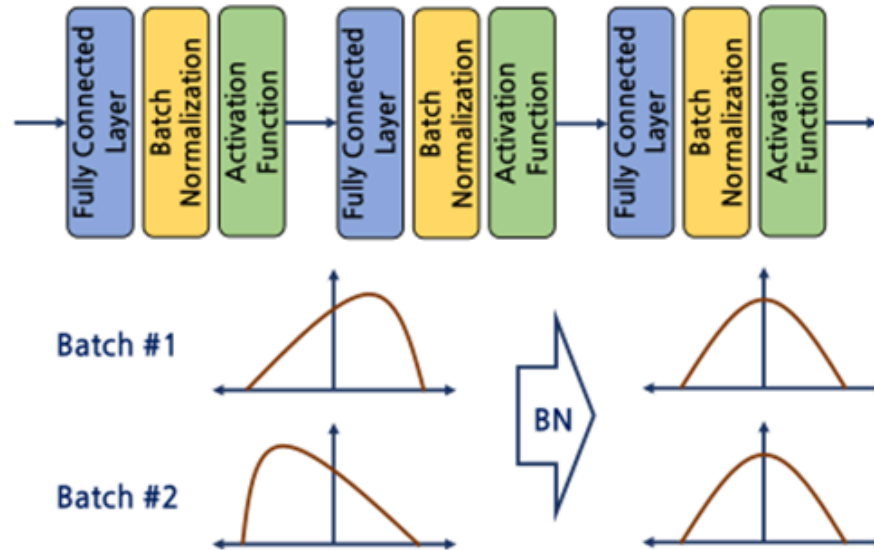
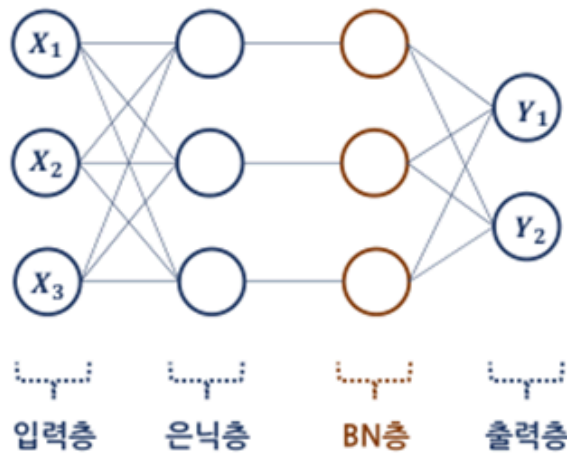
기울기 소실 문제

극복 방안

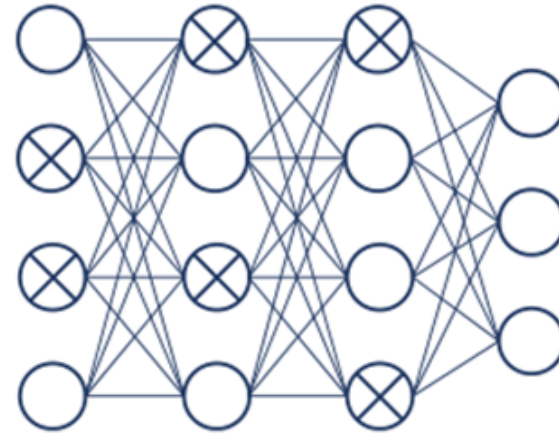
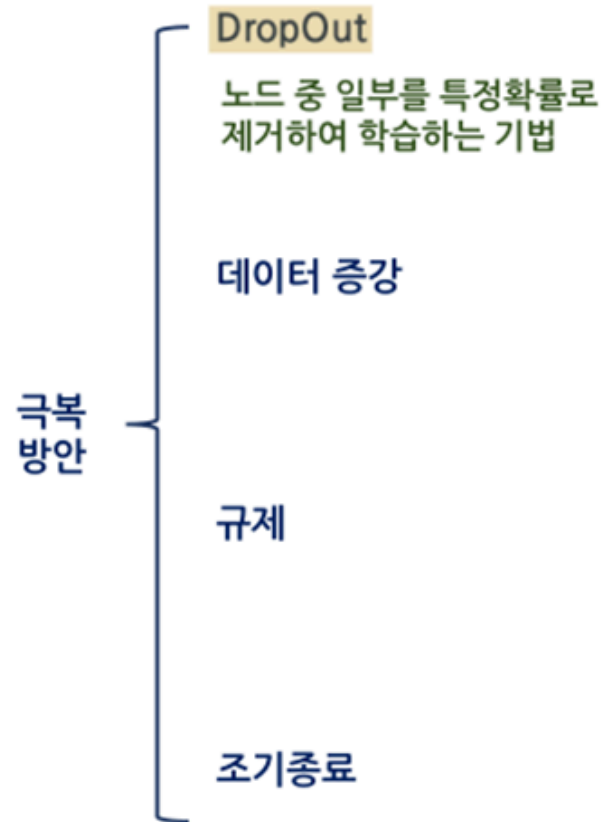
극복 가능한 활성화함수

배치 정규화

Batch 단위의 학습 시, 각 Batch별로 평균과 분산을 정규화하는 기법



■ 신경망의 과적합



- 특정 노드 사이의 관계가 불필요하게 강해지는 것 방지
- 특정 노드의 영향력 감소
- 강건한 신경망 구성 가능
- 학습 속도 저하

■ 신경망의 과적합

극복
방안

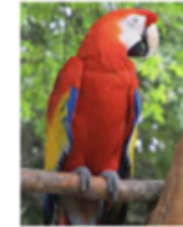
DropOut

데이터 증강

기존 데이터에서 새로운 데이터를 인위적으로
생성하여 데이터 양을 증가시키는 기법

규제

조기종료



Flip



Crop



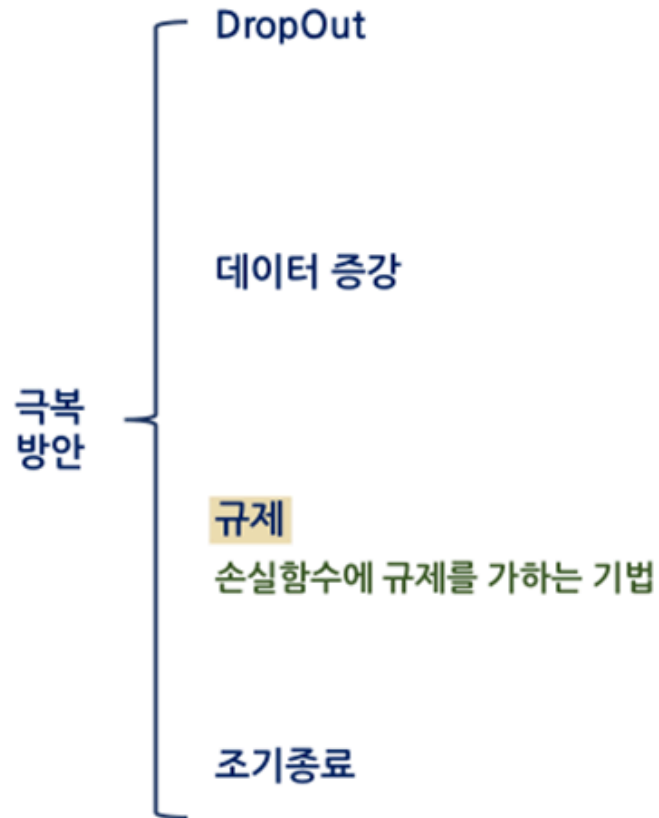
Blur



Contrast

※ 출처 : What is image augmentation and how it can improve the performance of deep neural networks
Link - [What is image augmentation - Albumentations Documentation](https://docs.albumentations.ai/en/latest/getting_started/what_is_image_augmentation.html)

■ 신경망의 과적합



L1 규제

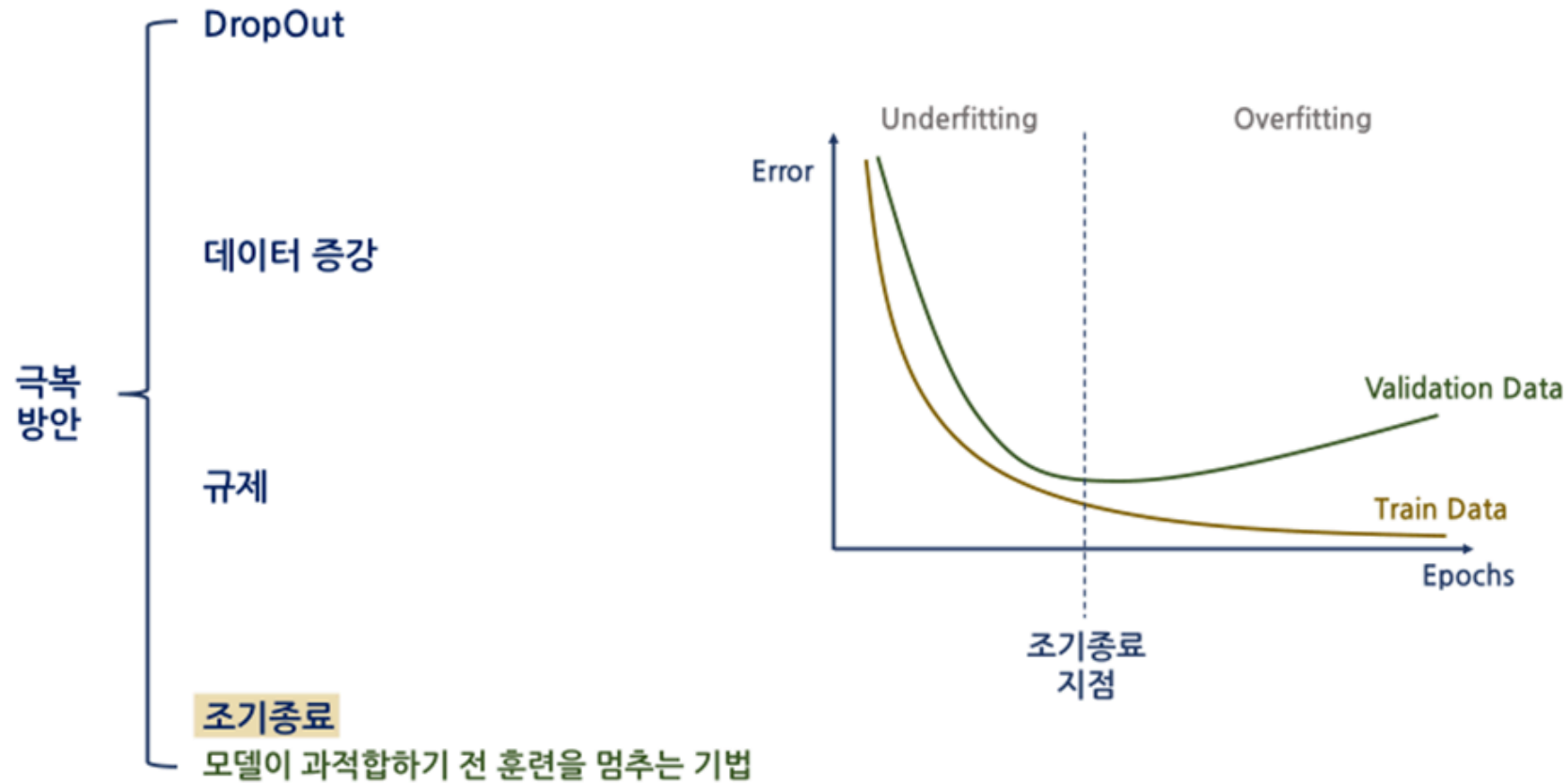
- 라쏘(Lasso) 회귀
- 비용함수 + $\alpha \sum |W|$
- 맨하탄 거리 기반

L2 규제

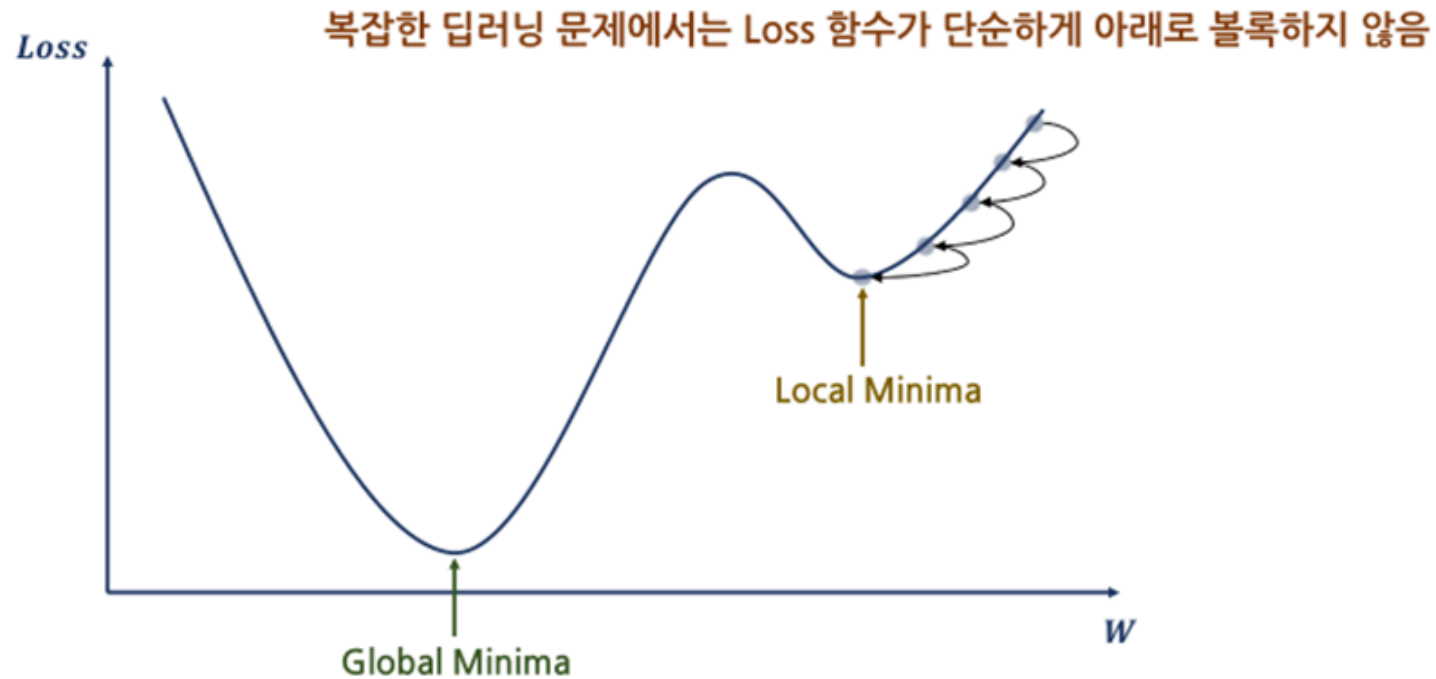
- 릿지(Ridge) 회귀
- 비용함수 + $\alpha \frac{1}{2} \sum W^2$
- 유클리드 거리 기반

L2 규제가 일반적으로 학습 시 더 좋은 성능

■ 신경망의 과적합



■ 극소점(Local Minima) 수렴 문제



경사하강법으로 학습 시 원하는 건 Global Minima에 수렴
Local Minima에서 기울기가 0으로 학습이 종료되는 문제

옵티마이저(Optimizer)

