

Bitamin 4주차 복습 세션

2주차, 3주차 복습

# 심층 신경망의 학습 review

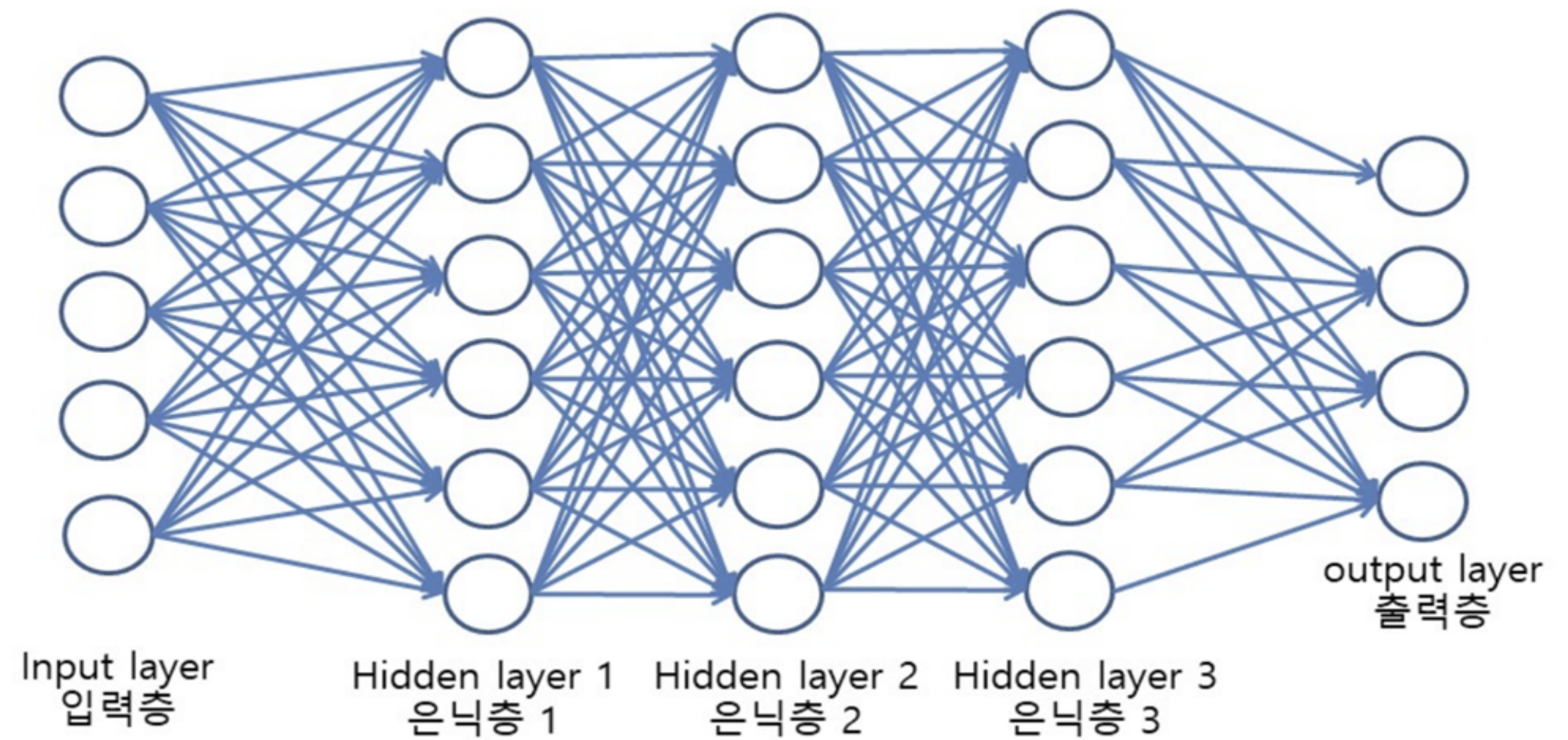
활성화 함수, 손실 함수, 옵티마이저와 좋은 학습을 위한 방법들

10기 4조

노지에 임청수 한세림

## 심층 신경망 (DNN)

입력층과 출력층 사이에  
여러 개의 은닉층으로 이루어진  
인공신경망



## 활성화 함수 (Activation Function)

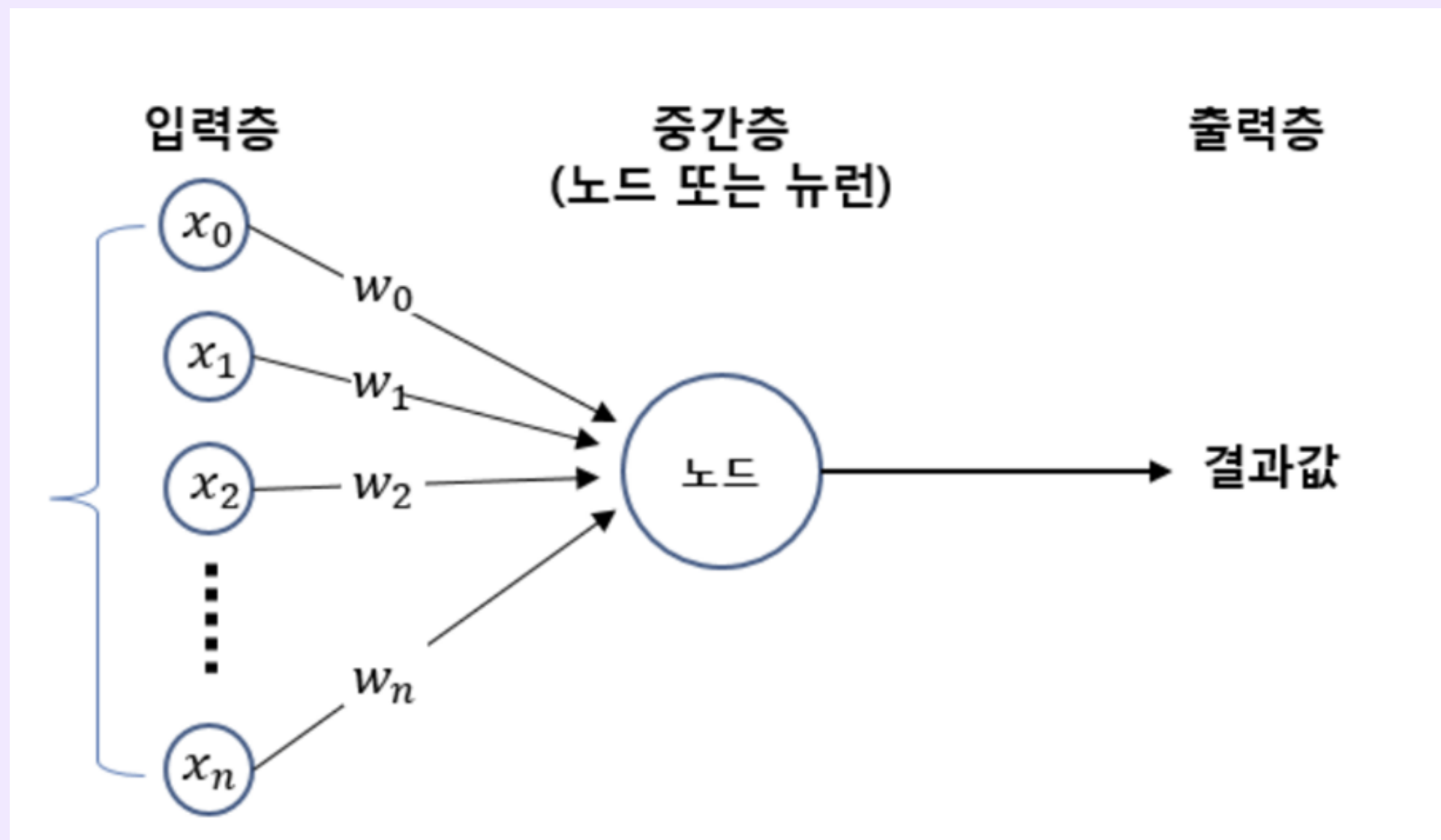
- 입력 신호의 총합을 출력 신호의 총합으로 변환해주는 비선형 함수  
(선형함수일 경우 신경망의 층을 쌓는 의미가 없어짐.)

ex. Sigmoid, tanh, ReLU 함수

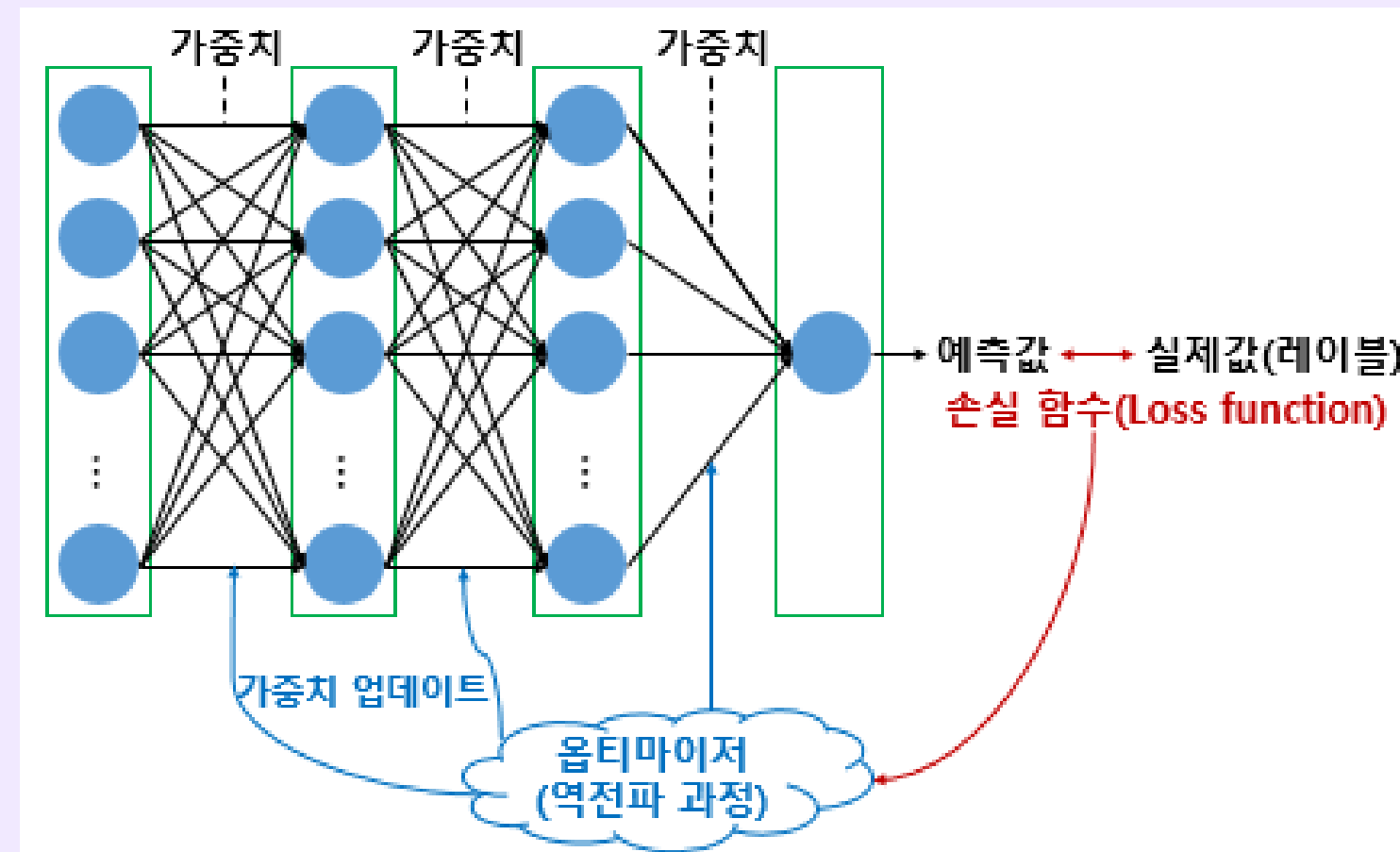
(1) Sigmoid – Gradient Vanishing, not zero-centered

(2) tanh – Gradient Vanishing 발생

(3) ReLU – 가장 많이 사용함. not zero-centered



## 가중치 업데이트



### 손실 함수 (Loss Function)

학습 알고리즘이 얼마나 정답과 다른 지 알려주는 지표  
ex. MSE, RMSE, MAE, Cross Entropy 등.

### 옵티마이저 (Optimizer)

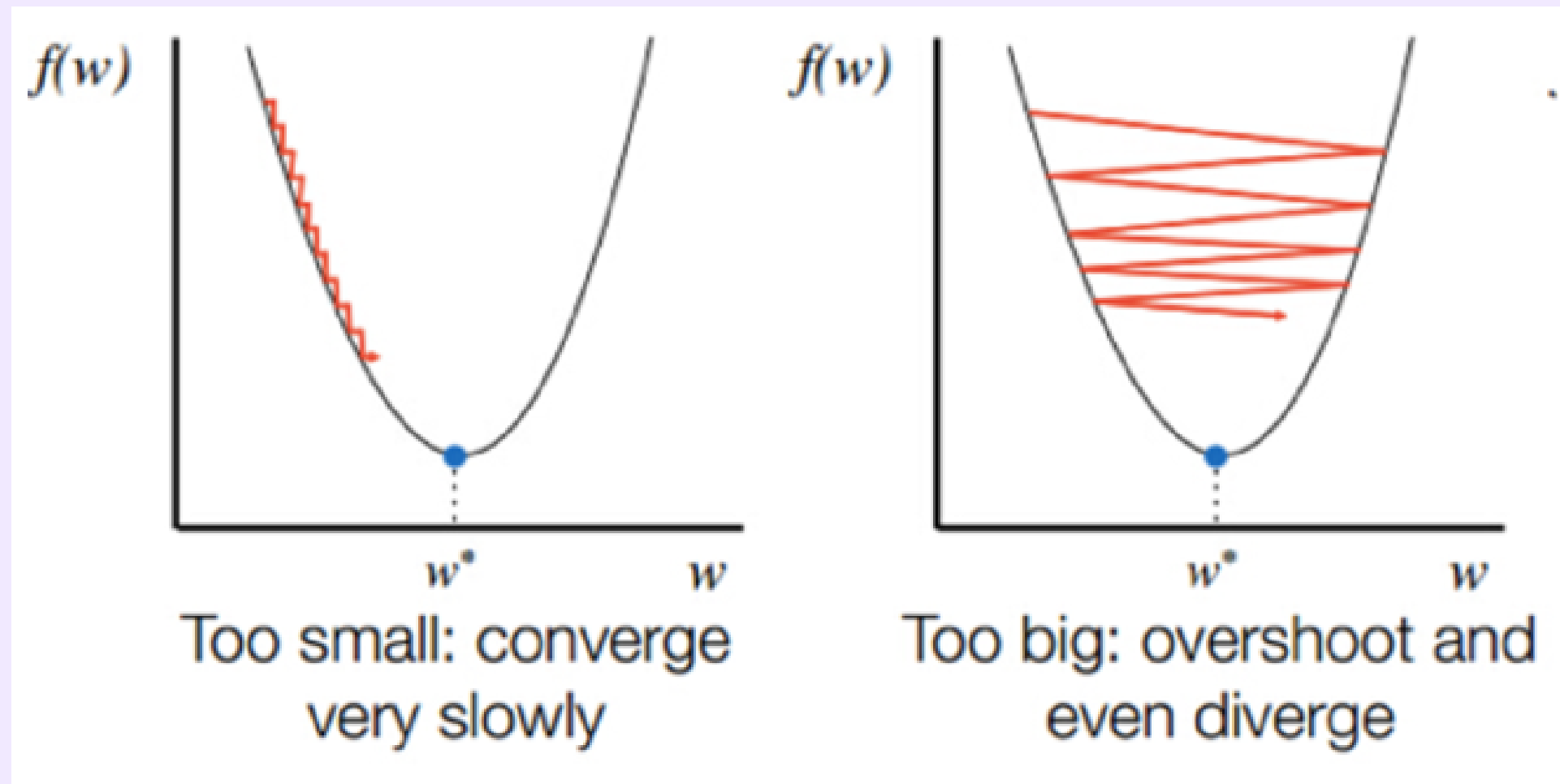
Loss function의 최솟값에 가까워지도록 모델의  
가중치와 편향을 변화시키는 최적화 함수.  
ex. SGD, Momentum, AdaGrad, Adam 등

## 경사하강법 (Gradient Descent)

손실함수의 기울기를 구하여 기울기가 낮은 쪽으로 이동시키는 방법

학습률: 손실함수의 기울기 값을 경사하강법에 적용하는 정도.

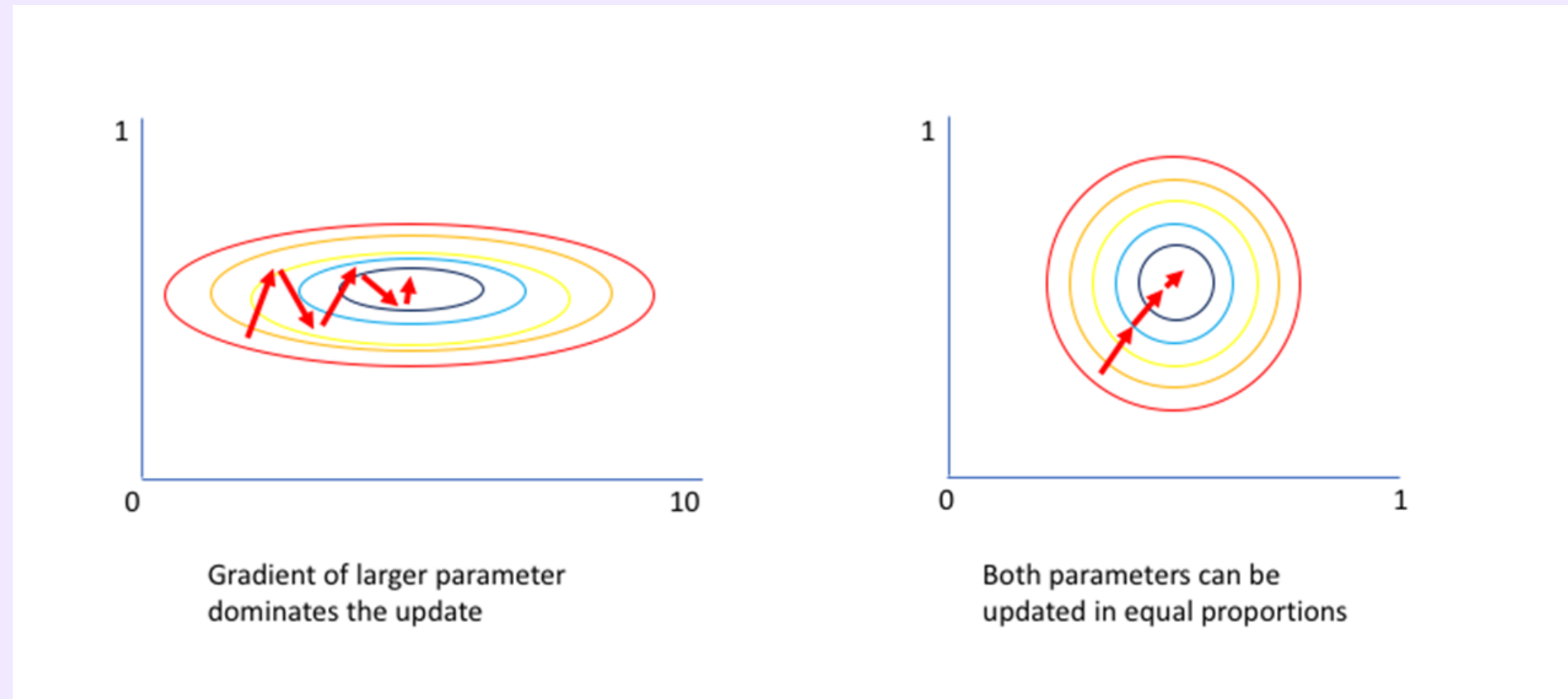
- 학습률이 너무 작을 경우 학습 속도가 느리다
- 학습률이 너무 큰 경우 발산하여 최소값에 도달하지 못할 수 있다.



## Why Normalize?

데이터 값 간의 범위 차이가 클 경우,  
비효율적인 최적화 과정을 거치게 됨  
(지그재그 모양으로 그래프가 수렴함)

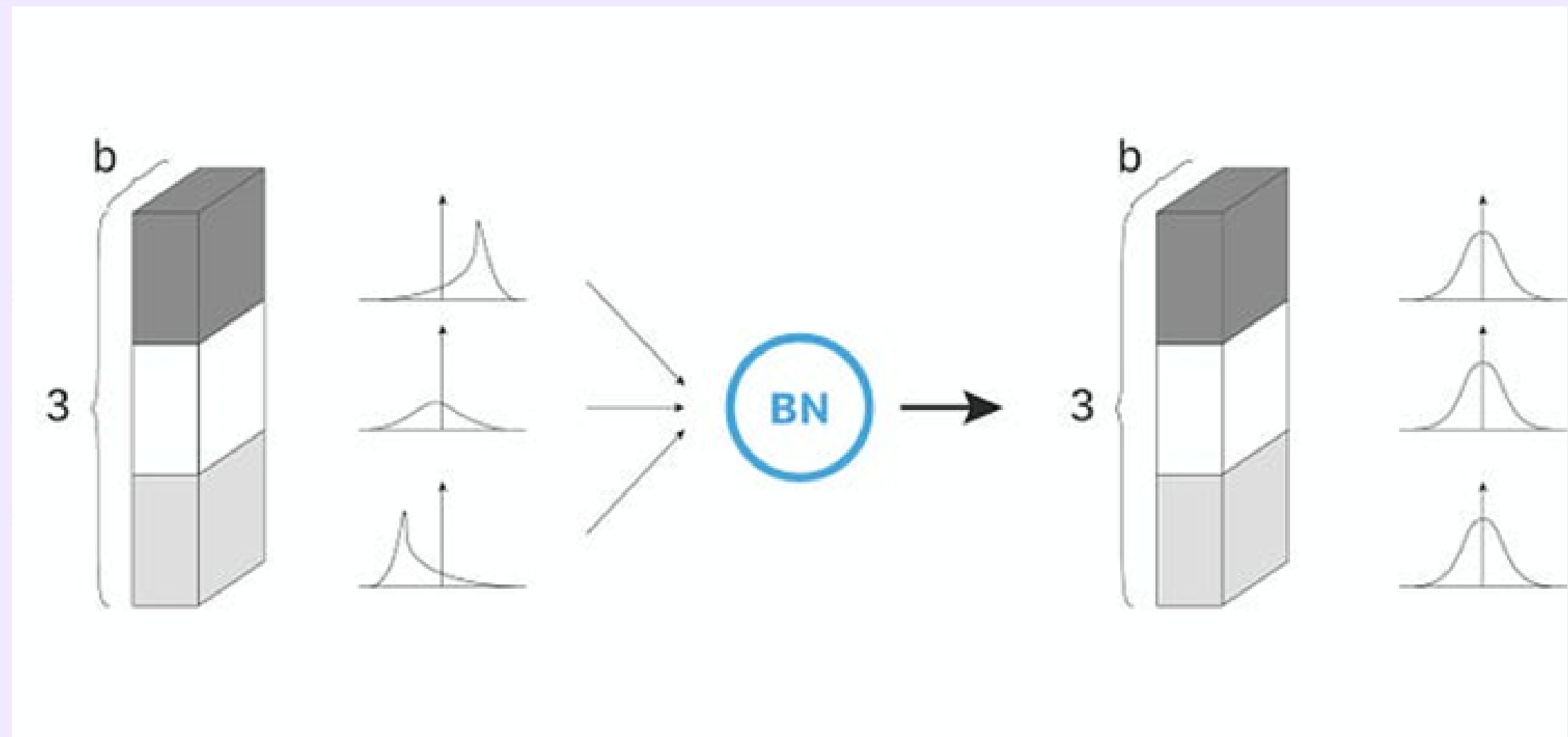
-> 데이터를 'Batch' 단위로 정규화



## 배치 정규화 (Batch Normalization)

배치 정규화의 장점:

- (1) 빠른 학습 진행
- (2) 초깃값 의존도 감소
- (3) 오버피팅 억제



## 배치 정규화와 경사하강법

### (1) Batch Gradient Descent

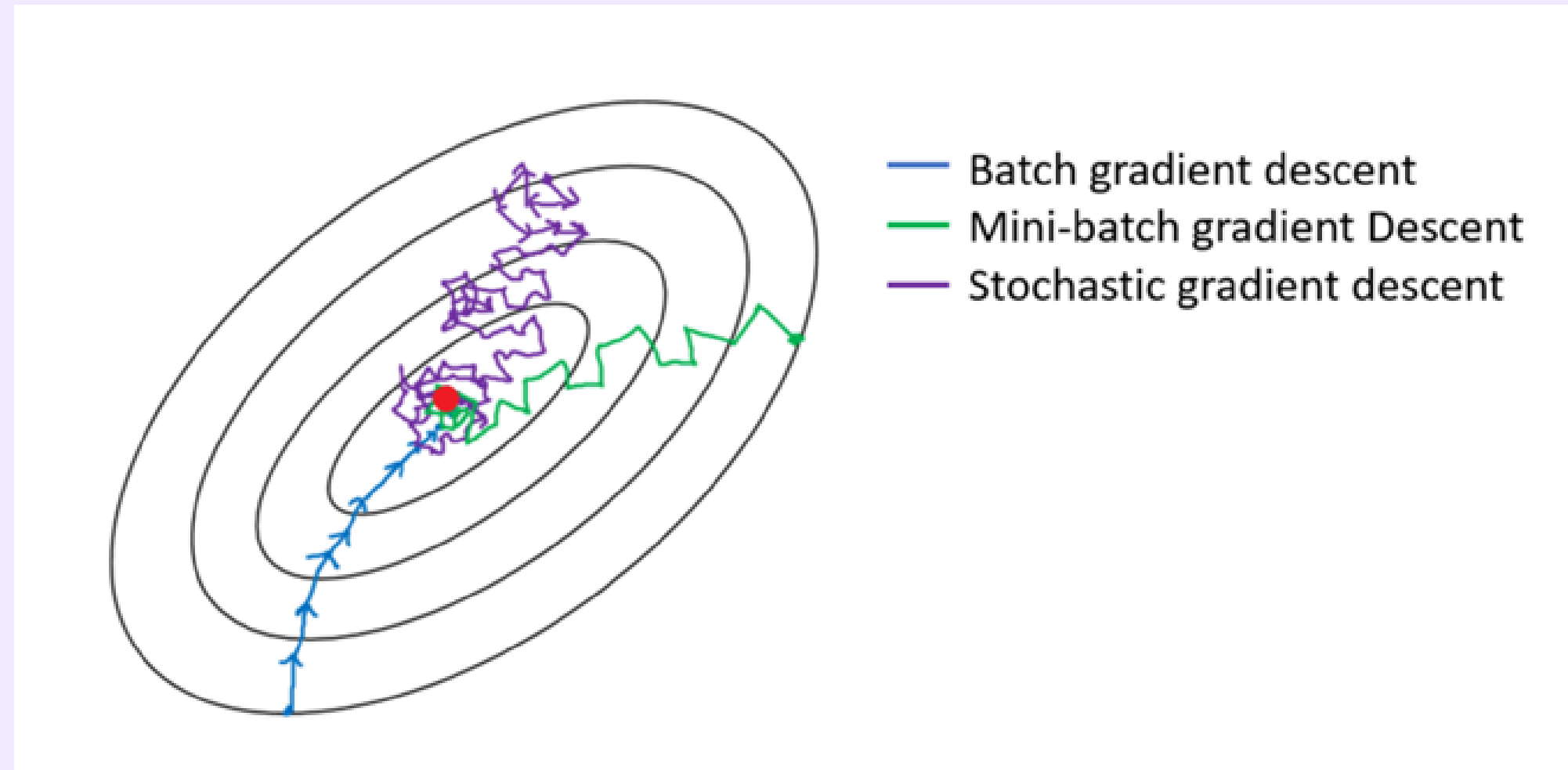
전체 데이터의 gradient의 평균을 구해 수렴함. 부드럽게 수렴하지만 속도가 느림.

### (2) Stochastic Gradient Descent

데이터를 한 개씩 추출해서 가중치를 갱신함. 속도는 빠르지만 오차율이 커서 제대로 수렴하지 않을 수 있음.

### (3) Mini-Batch Gradient Descent

전체 데이터를 배치 크기로 나누어 경사하강법을 진행함. 둘의 절충안.





# 과소적합과 과대적합

## 과소적합 (Underfitting)

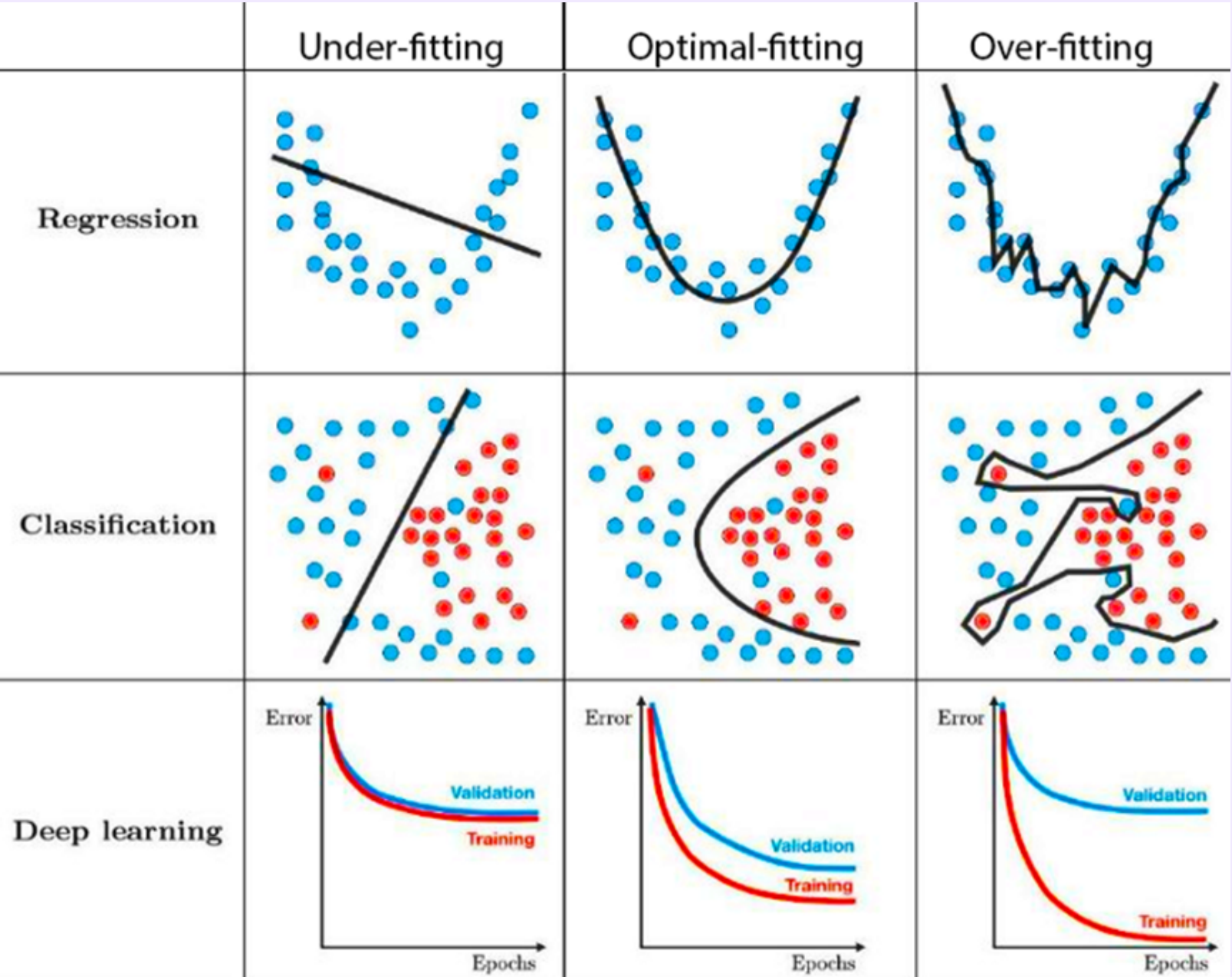
Train set을 제대로 학습하지 못한 경우.

- 학습 횟수가 충분하지 못할 때
- 데이터의 특성에 비해 모델이 너무 단순할 때
- 데이터의 양이 충분하지 않을 때

## 과대적합 (Overfitting)

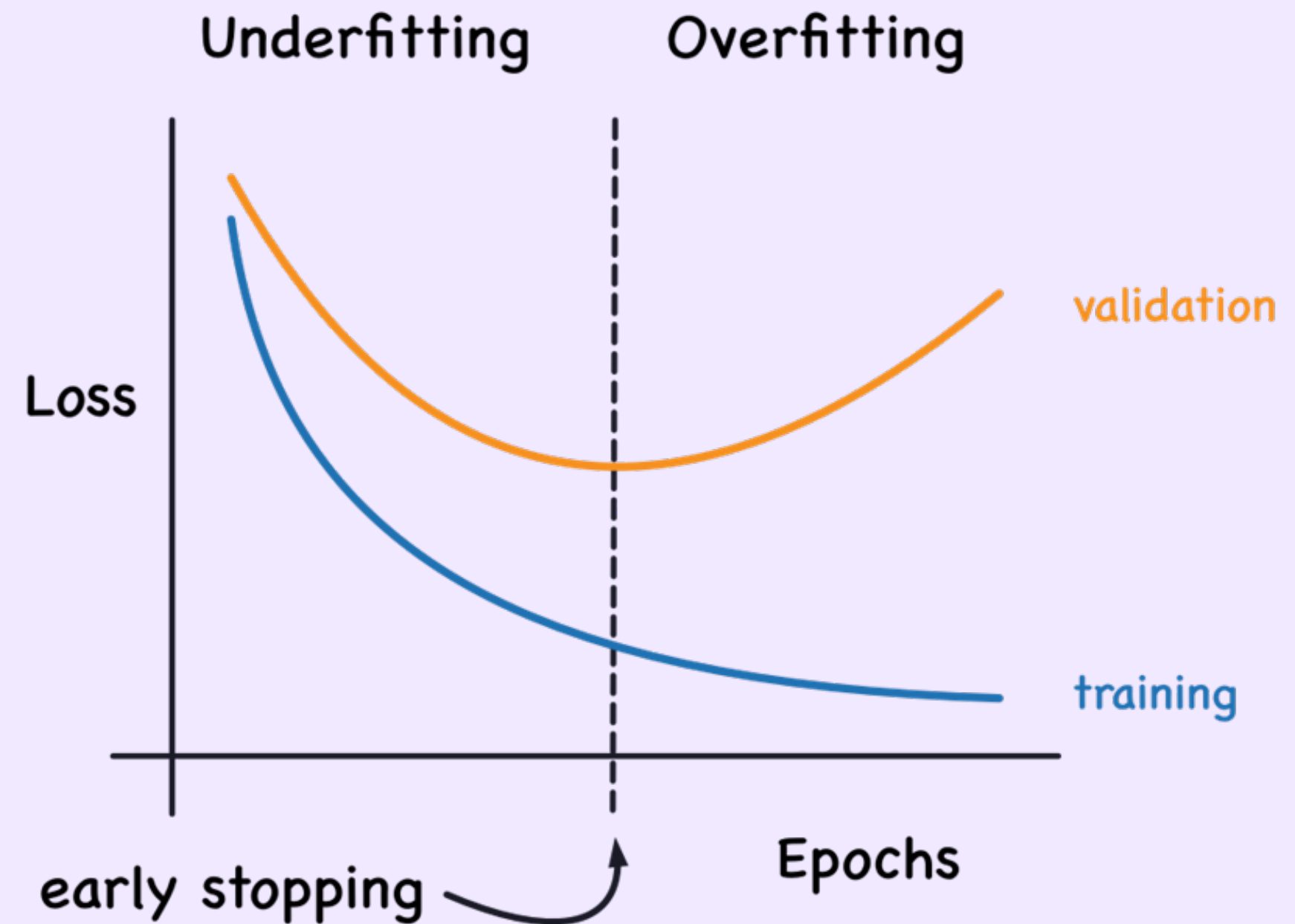
Train set을 과도하게 학습한 경우.

- 데이터의 특성에 비해 모델이 너무 복잡할 때
- 데이터의 양이 충분하지 않을 때

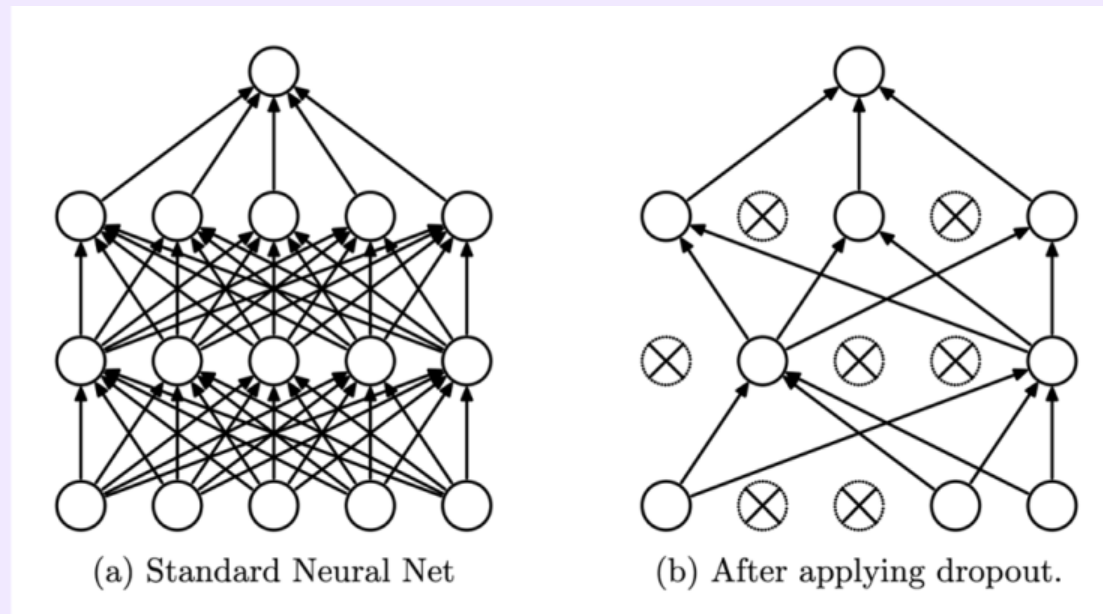


## Validation Set

- 검증 데이터셋 (validation set)을 이용하여 underfitting, overfitting 정도를 확인하기
- Early Stopping: Overfitting이 되기 전에 학습을 멈춤.

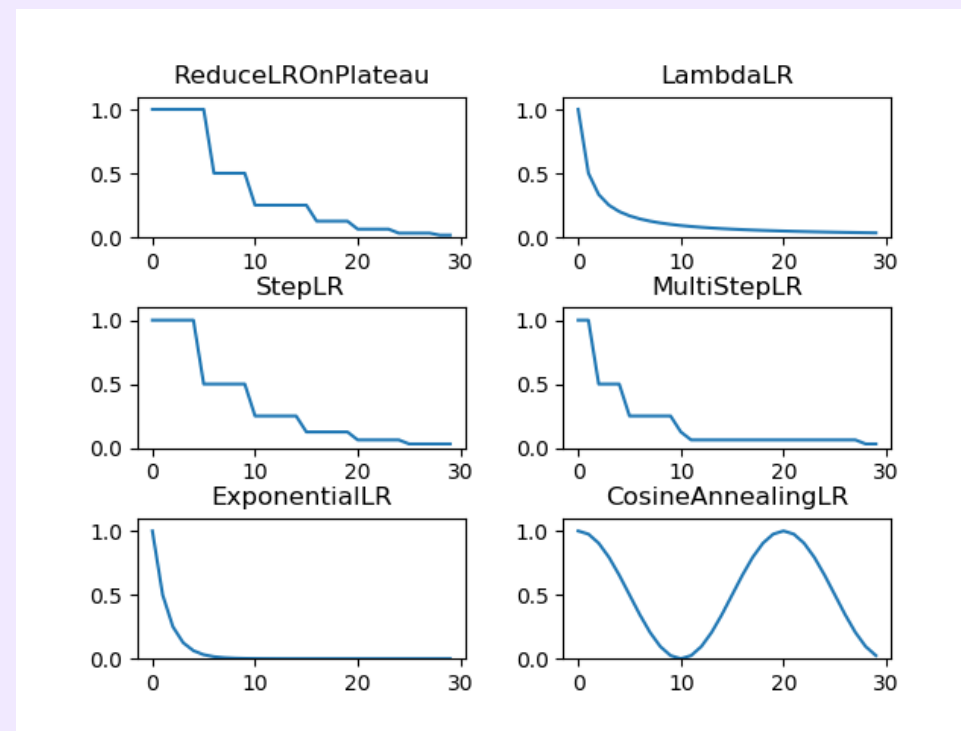


## 모델 성능 향상을 위한 방법들



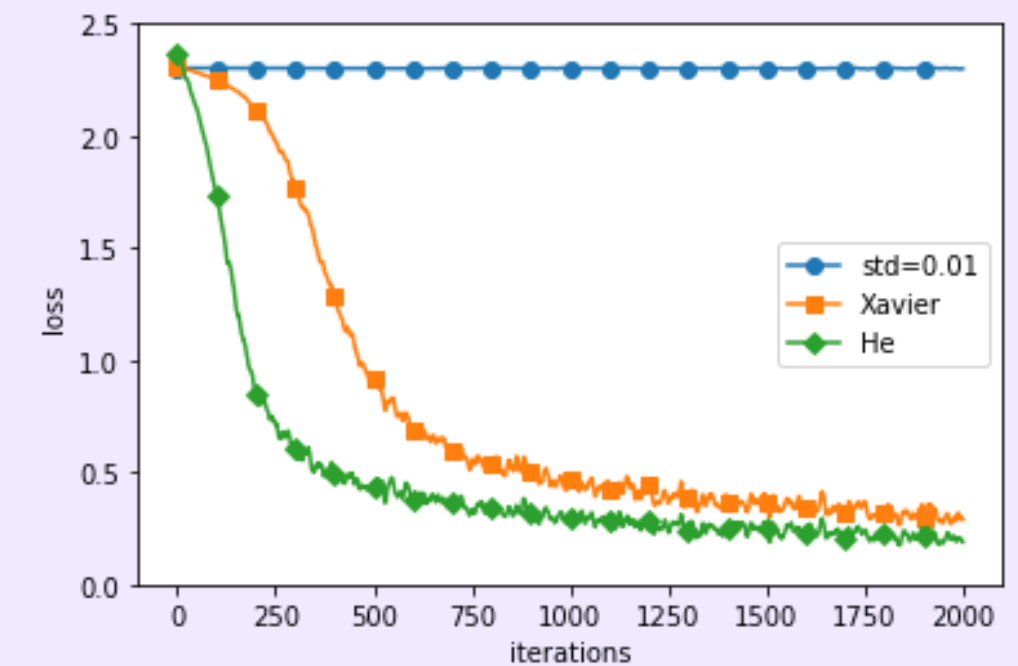
### 드롭아웃 (Dropout)

- 0부터 1 사이의 확률로 랜덤하게 신경망의 뉴런 부분을 제거하는 방법.
- Overfitting을 방지하기 위해 사용함.



### 스케줄러 (Scheduler)

- 학습 과정에서 학습률을 조정하는 장치
- ex. LambdaLR, MultiplicativeLR, StepLR, CyclicLR 등



### Weight Initializer

- 가중치의 초기값을 지정해주는 함수
- ex. 모두 0 또는 1로 설정하기, Xavier initialization, He initialization 등

Bitamin 4주차 복습 세션

2주차, 3주차 복습

---

**Thank You**  
이어서 CNN으로