

# 밑바닥부터 시작하는 딥러닝1

4장. 신경망 학습

2024. 04. 09

# 신경망 학습

- 특징: 데이터 주도 학습
  - ML: 사람이 특징 추출에 개입
  - DL: 학습과정의 처음부터 끝까지 사람 개입 X
- 오버피팅 (overfitting): 한 데이터셋에 지나치게 최적화된 상태

# 손실함수 (loss function)

- 신경망 성능의 '나뭇'을 나타내는 지표
- 신경망 훈련 목표: 손실을 최소화 하는 것
- 대표적 손실함수
  - 평균 제곱 오차 (Mean Squared Error, MSE)

$$E = \frac{1}{N} \sum_i^N (y_i - \hat{y}_i)^2$$

- 교차 엔트로피 오차 (Cross Entropy Error, CEE)

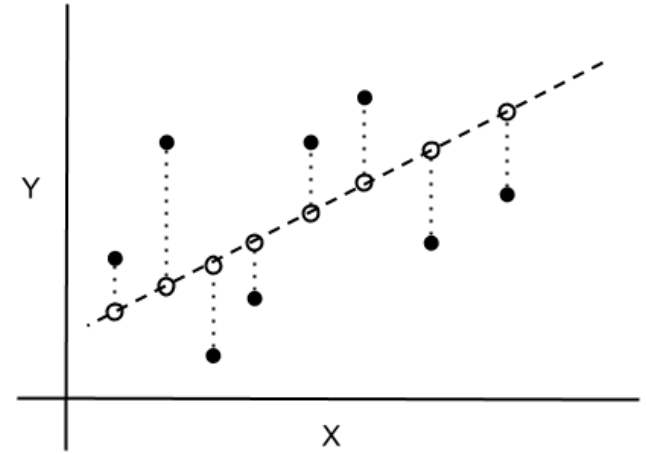
$$E = - \sum_i^{class} y_i \cdot \log \hat{y}_i$$

# 손실함수 (loss function)

회귀

- 평균 제곱 오차 (Mean Squared Error, MSE)

$$E = \frac{1}{N} \sum_i^N (y_i - \hat{y}_i)^2$$



분류

- 교차 엔트로피 오차 (Cross Entropy Error, CEE)

$$E = - \sum_i^{class} y_i \cdot \log \hat{y}_i$$

예)

	개	고양이	닭	곰
y	1	0	0	0
y_hat	0.8	0.1	0.0	0.1

손실:  $E = -(1 \cdot \log(0.8))$

# 미니 배치 (mini-batch)

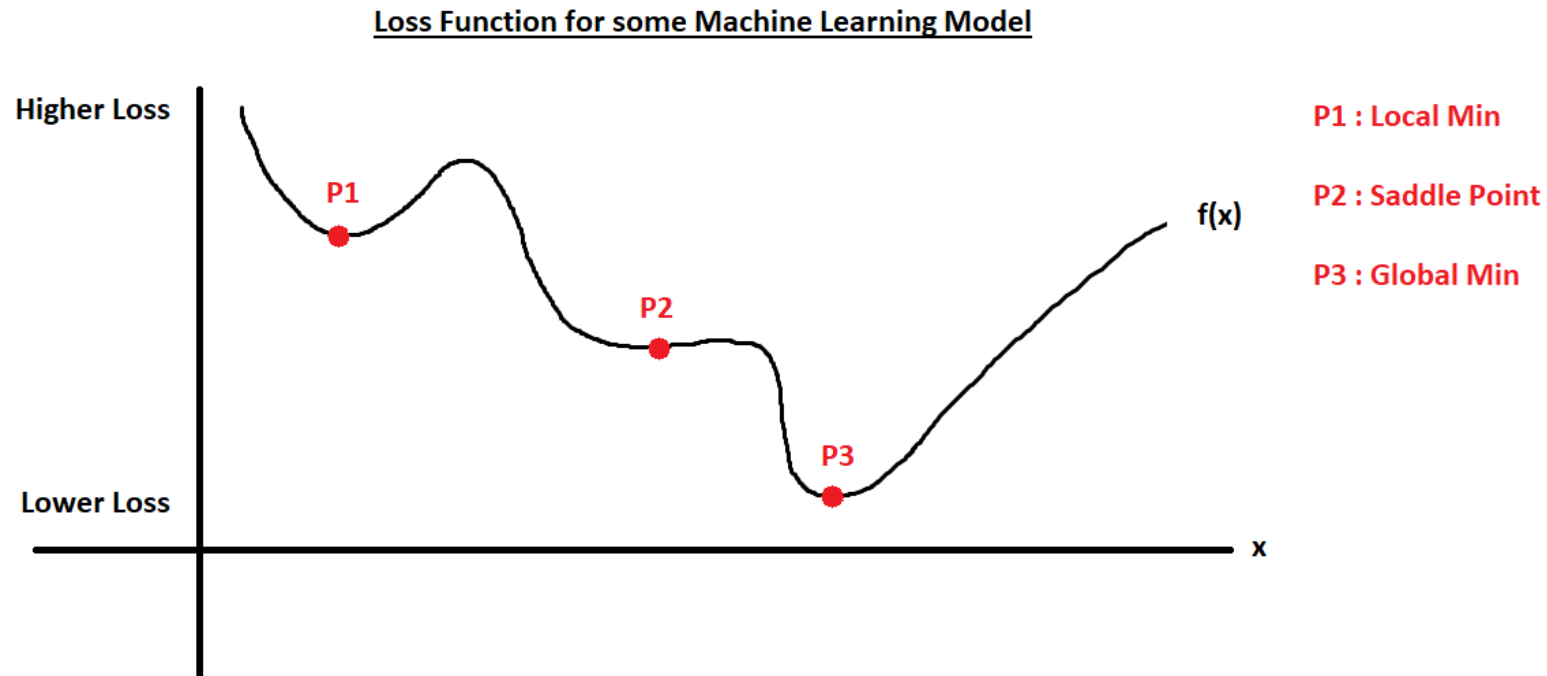
- 데이터의 일부를 추려 전체의 '근사치'로 학습에 사용
  - 예) 데이터 60,000개 중 무작위 100개만 사용
- 데이터가 N개일 때, 전체 훈련 데이터에 대한 Cross Entropy Error :

$$E = -\frac{1}{N} \sum_n \underbrace{\sum_i^{class} y_{n,i} \cdot \log \hat{y}_{n,i}}_{\text{기존 수식}}$$

- n: 전체 데이터 N개 중 n번째
- k (=클래스 개수): n번째 데이터의 k번째 클래스의 확률값
- 1/N: 평균 손실 구하기 위함

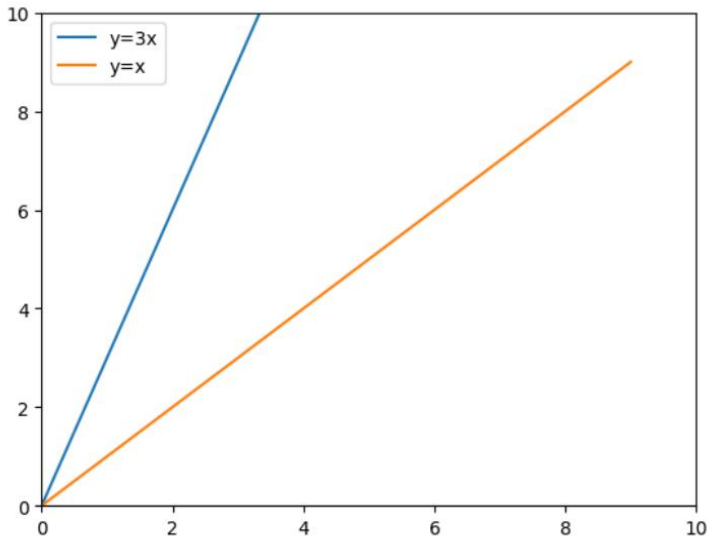
# 경사 하강법 (Gradient Descent method)

- 미분 (기울기)를 이용해 손실함수의 최솟값을 찾는 방법
  - 기울기가 0일때가 최적의 파라미터( $W, b$ )
    - 단, 항상 그런 것은 아님



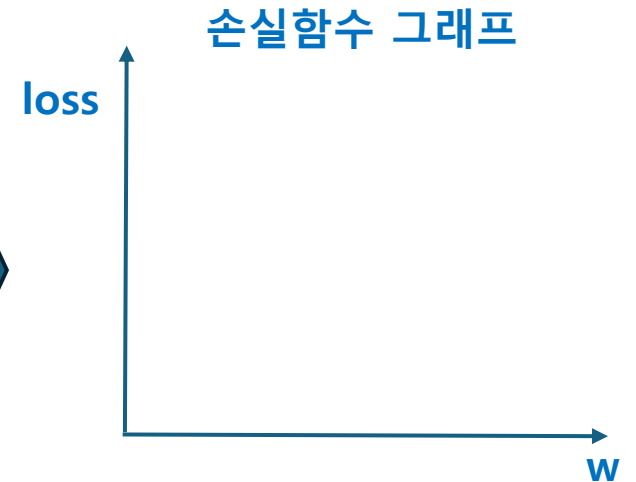
# 경사 하강법 (Gradient Descent method)

- 예시) 딥러닝의 학습 과정을 하나씩 따라가보자
  - 풀어야 할 회귀 문제:  $y=3x$
  - 식의 형태가  $y=w \cdot x$  라고 했을 때 딥러닝이 찾아야될 가중치  $w=?$
  - $w=1$  부터 시작해보자



x	1	2	3	...	N
y	3	6	9	...	3N
y_hat					

- 사용할 손실함수:  $E = \frac{1}{N} \sum_i (y_i - \hat{y}_i)^2$
- 계산해보자:



$$E = \frac{1}{N} \sum_i^N (y_i - \hat{y}_i)^2$$

# 경사 하강법 (Gradient Descent method)

- 미분 방법
  - 해석적 미분:
  - 수치적 미분:

$\epsilon = 0.1$

i)  $w + \epsilon = 1.1$

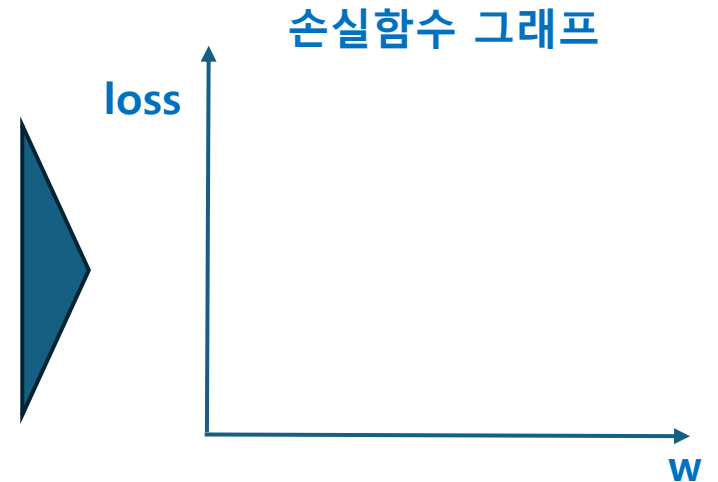
<b>x</b>	1	2	3	...	N
<b>y</b>	3	6	9	...	3N
<b>y_hat</b>	1.1	2.2	3.3	...	1.1N

➤ loss=

ii)  $w - \epsilon = 0.9$

<b>x</b>	1	2	3	...	N
<b>y</b>	3	6	9	...	3N
<b>y_hat</b>	0.9	1.8	2.7	...	0.9

➤ loss=





# 확률적 경사하강법 Stochastic Gradient Descent

- 데이터를 미니 배치로 묶은 후 무작위 선정하는 방식
- 예)

① 데이터셋 shuffle



② 미니배치로 묶기



③ 미니배치 단위로 무작위 선정하여 경사하강법에 사용 (= 확률적 경사하강법)

