

밑바닥 딥러닝1

5장. 오차역전법

2024. 04.12

저번 시간에 설명 부족했던 부분

경사하강법 Gradient Descent

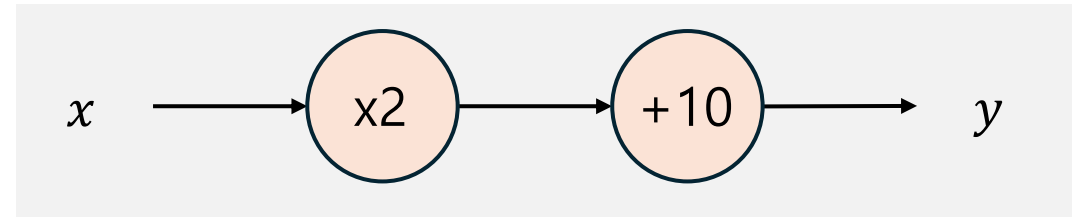
- **GD:** 1 epoch시, 전체 데이터의 손실을 경사 하강에 사용
- **SGD:** 무작위 1개 데이터의 손실을 경사 하강에 사용. (단, 1 epoch시, iteration=전체 데이터 N개)
- **mini-batch GD**
 - 무작위 1개 미니배치를 경사하강에 사용
 - 예) 전체 데이터셋 60,000개, batch_size=100일 때, mini-batch 개수=?
1 epoch에서 600 iter 만큼 경사하강 수행. (= 600 iter 번 가중치 갱신)
이 때, 한 미니 배치가 여러 번 선택될 수 있음.
 - 참고: 보통 batch_size=2의 제곱수 사용
- SGD는 batch_size=1인 미니배치 경사하강법이라고도 할 수 있다.
- 일반적으로 SGD라고 하면 '미니배치 경사하강법'을 의미함.

5장. 오차 역전법 `backpropagation`

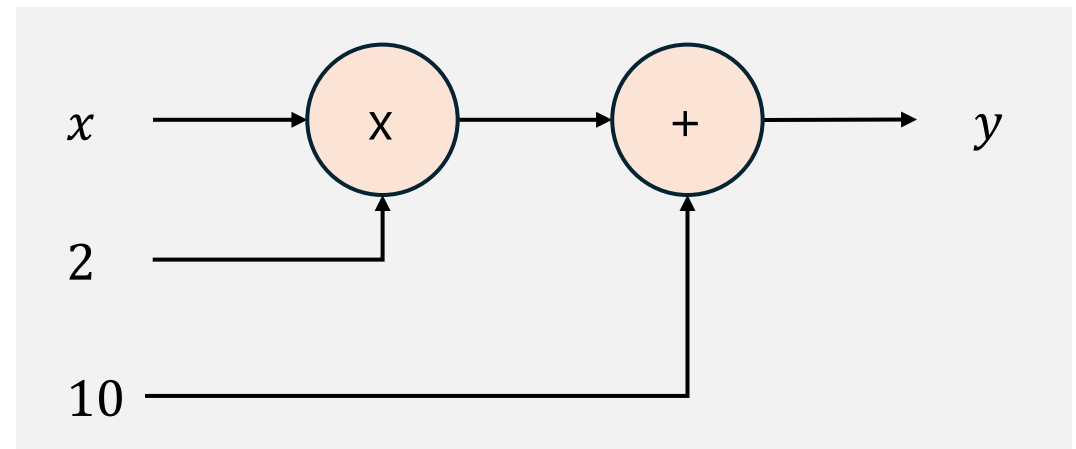
- ❖ 수치 미분은 단순하고 구현이 쉽지만, 계산 시간이 오래 걸림
- ❖ 오차 역전법을 사용하면 효율적으로 계산 가능!

1. 계산 그래프 computational graph

- 오차 역전법을 시각적으로 이해하기 위한 그래프
- 기본 형태
 - 수식: $y = 2 \cdot x + 10$
 - 계산 그래프:



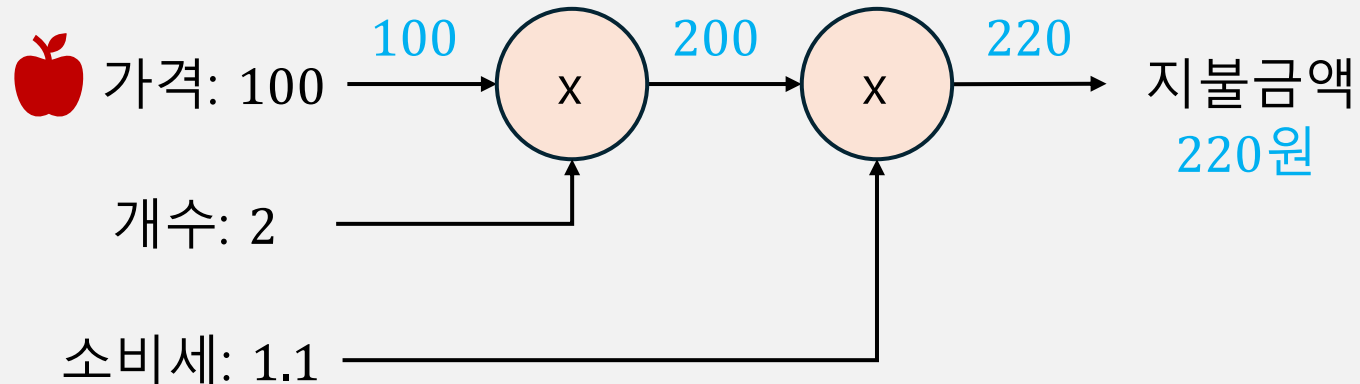
연산자, 피연산자를 분리



1. 계산 그래프

- 문제1: 사과 2개를 구매할 때 지불금액은 얼마인가?

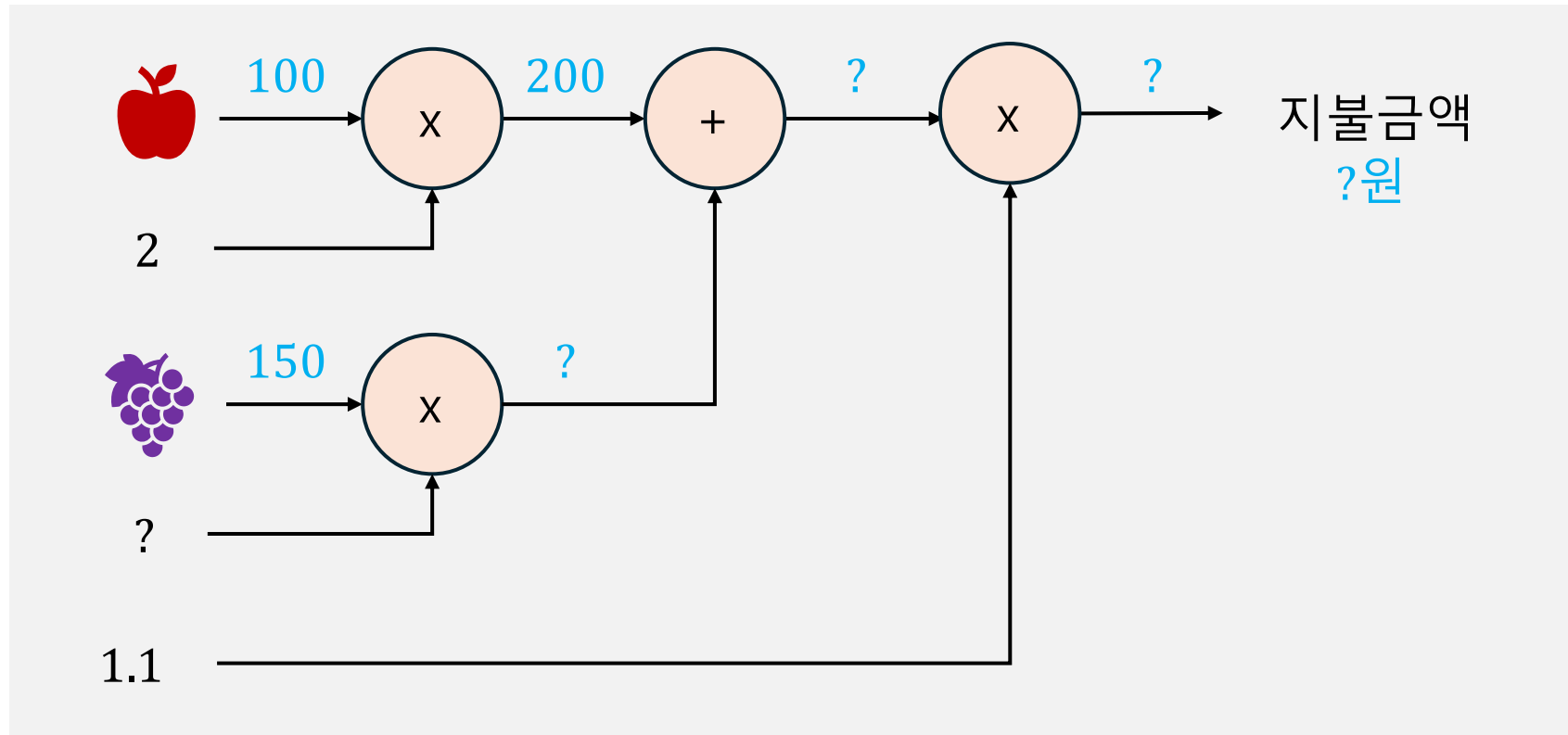
사과 (개)	100 원
소비세	10%



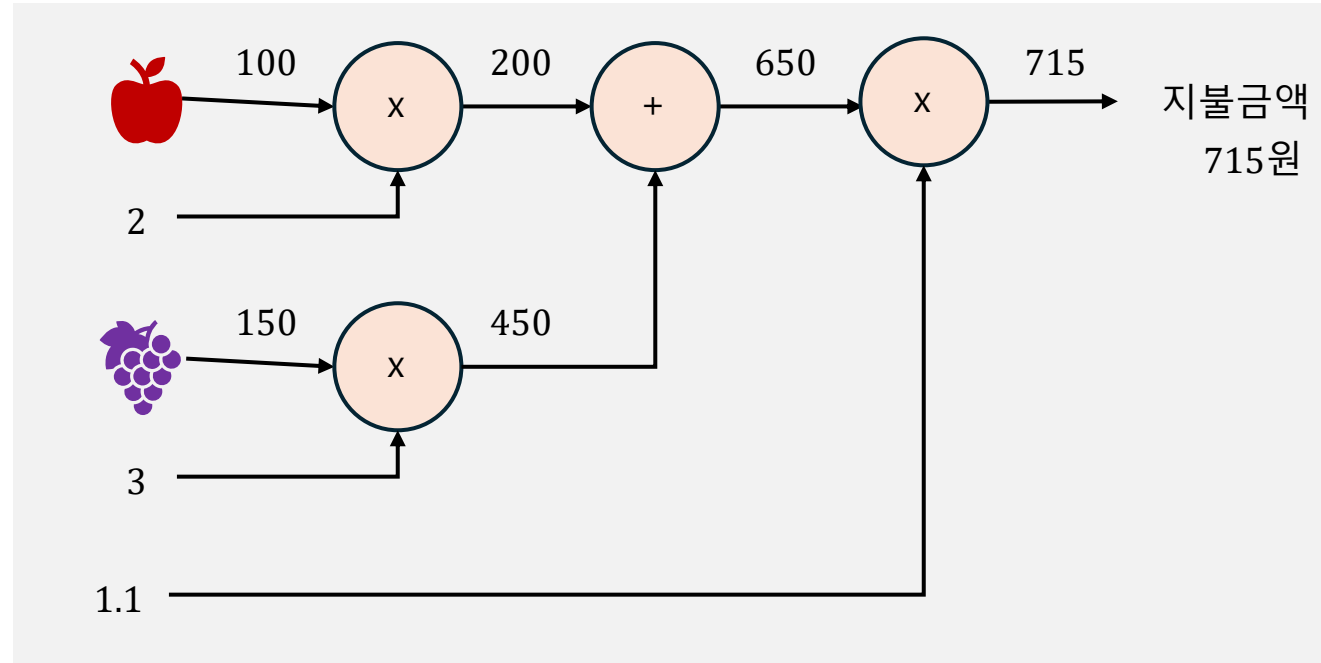
1. 계산 그래프

- 문제2: 사과 2개, 포도 3개를 구매할 때 지불금액은 얼마인가?

사과 (개)	100원
포도 (개)	150원
소비세	10%



1. 계산 그래프



순전파 (forward propagation)



역전파 (backward propagation)



1. 계산 그래프

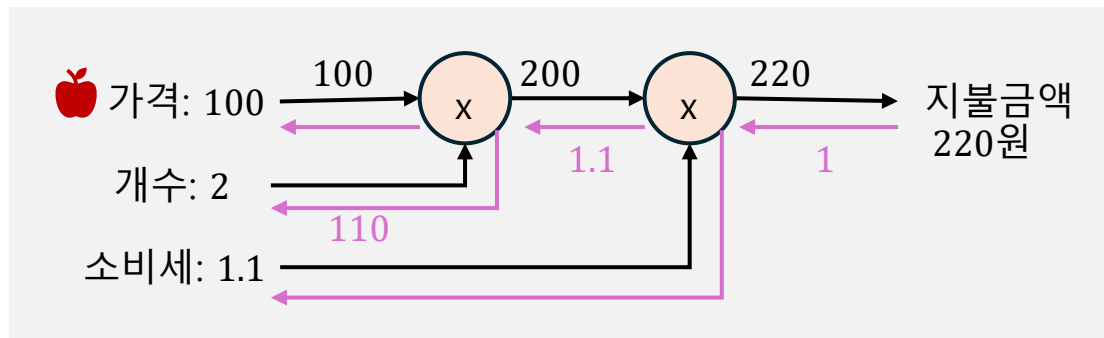
- 계산 그래프의 이점

- ① 전체 계산 과정이 아무리 복잡해도, 각 노드에서의 계산에 충실하면 됨.
- ② 중간 계산 결과를 전부 보관함 = 국소적 계산

- 역전파, 미분의 의미

- 역전파: 국소적 미분을 전달하는 전파
- 미분: 변수의 변화가 출력값에 영향을 끼치는 정도.

예) 사과 개수 (변수)가 지불금액 (출력값)에 영향을 끼치는 정도.
사과 개수가 1개 늘어날 때마다 110원 만큼 지불금액 증가.



➔ 이때, 효율적 미분을 위해 '연쇄법칙' 사용

2. 연쇄 법칙 chain rule

- 합성 함수: 여러 함수로 구성된 함수

$$z = (x + y)^2$$



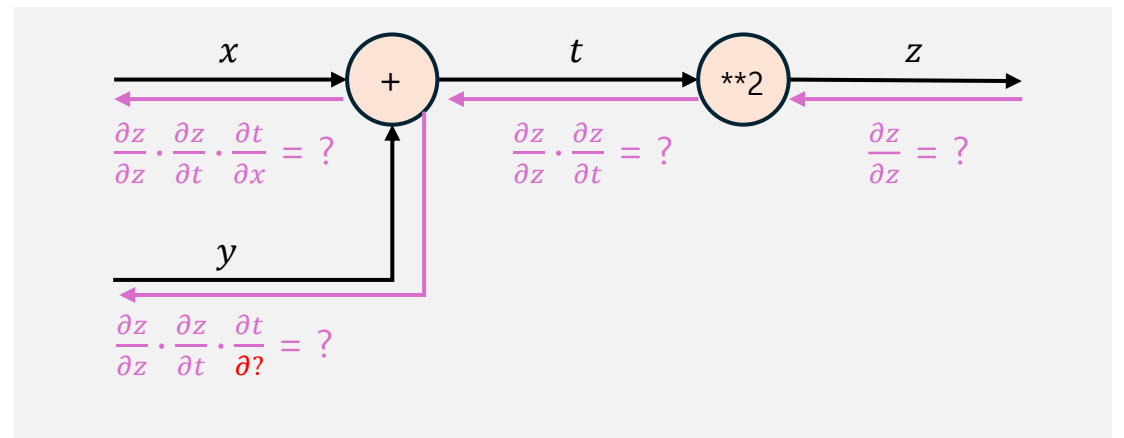
$$\begin{cases} z = t^2 \\ t = x + y \end{cases}$$

- 합성 함수의 미분: 합성 함수를 구성하는 각 함수의 미분의 곱으로 나타낼 수 있다.
= 연쇄 법칙

- 수식:

$$\begin{aligned} \frac{\partial z}{\partial x} &= \frac{\partial z}{\partial t} \cdot \frac{\partial t}{\partial x} \\ &= 2t \cdot 1 \\ &= 2(x + y) \end{aligned}$$

- 계산 그래프:



3. 역전파

- 덧셈 노드의 역전파

- 예) $z = x + y$ 일 때, $\frac{\partial z}{\partial x} = \text{---}$, $\frac{\partial z}{\partial y} = \text{---}$

즉, 입력값을 그대로 흘려보냄.

덧셈 노드에서 미분 결과는 항상 1.

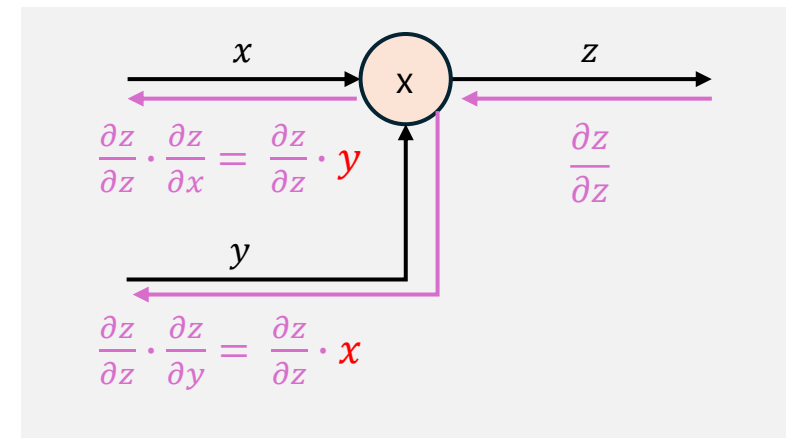
미분 결과가 1이 아닌 때도 있지만,
편의상 1로 계산

- 곱셈 노드의 역전파

- 예) $z = xy$ 일 때, $\frac{\partial z}{\partial x} = \text{---}$, $\frac{\partial z}{\partial y} = \text{---}$

즉, 다른 입력신호 값이 미분값.

따라서 순방향 입력신호값 저장 필요.

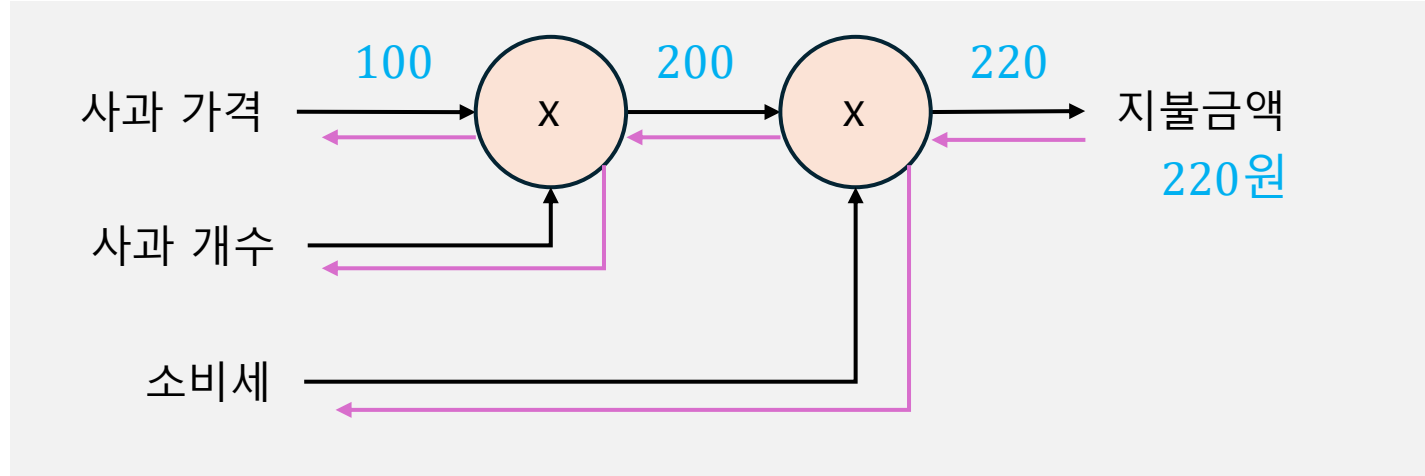


덧셈, 곱셈 노드 계산법을 활용해 문제를 풀어보자

문제1. 사과 쇼핑의 예

- 문제: 지불금액에 변수 $\begin{pmatrix} \text{사과 가격} \\ \text{사과 개수} \\ \text{소비세} \end{pmatrix}$ 각각이 얼마나 영향을 미치는지 구하시오.

- 풀이 (계산그래프):



- 정답: 사과 가격은 ____ 만큼,
사과 개수는 ____ 만큼,
소비세는 ____ 만큼 지불액에 영향 미침.

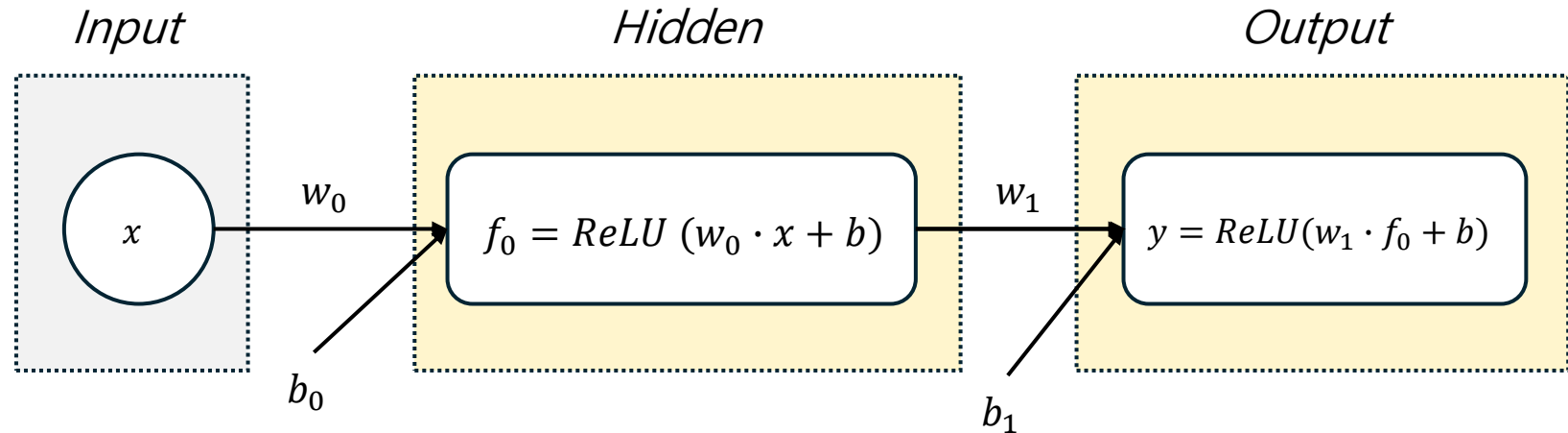


사과 가격이 1(원) 만큼 변화하면, 지불 금액은 ____ (원)만큼,
사과 개수가 1(개) 만큼 변화하면, " ____ (원)만큼
소비세가 1(100%) 만큼 변화하면, " ____ (원)만큼 변화

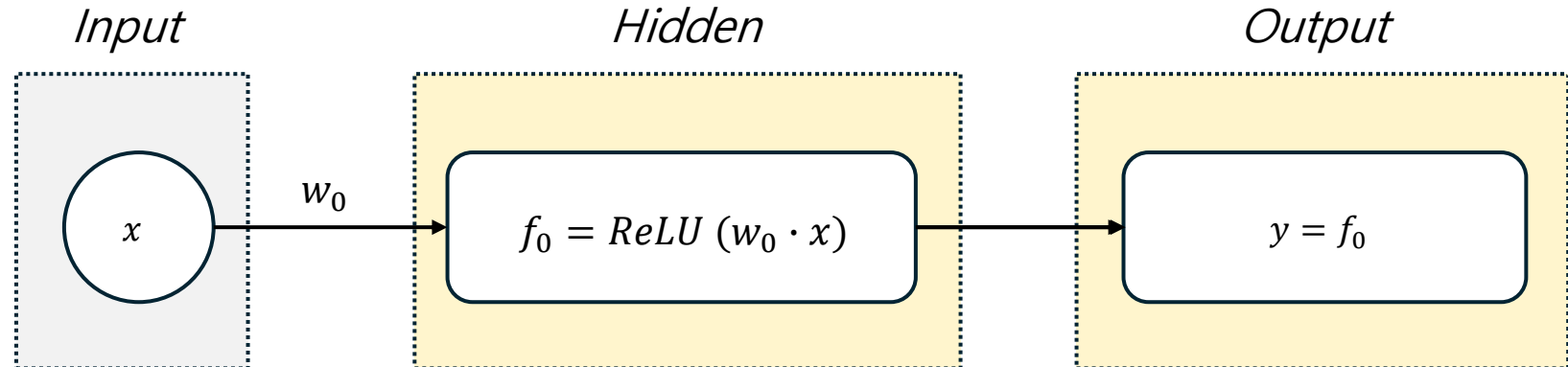
문제2. 딥러닝에서의 역전파

- 문제: $y = 2x$ 를 해결하는 신경망

원래는 이런데



이렇게 간략하게
풀어보자



문제2. 딥러닝에서의 역전파

- 문제: $y = 2x$ 를 해결하는 신경망
- 풀이:

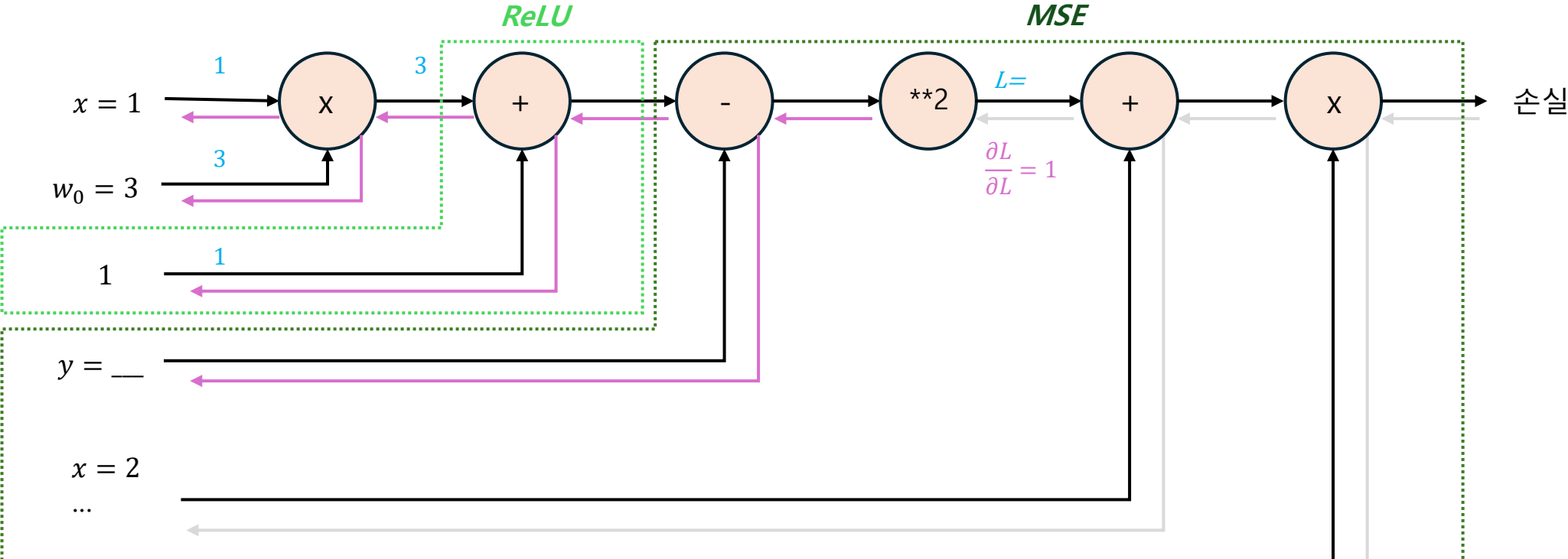
✓ 초기 가중치 $w_0 = 3$ 일때, 초기 식 $y = w_0 \cdot x = \underline{\hspace{2cm}}$

✓ 초기 식에 대한 계산그래프:

- ✓ 참고:
- ReLU: $y = \begin{cases} 0: & z \leq 0 \\ z: & z > 0 \end{cases}$
 - MSE: $\frac{1}{N} \sum_1^N (y - \hat{y})^2$

✓ Dataset:

<i>train_X</i>	1	2	3
<i>train_y</i>	—	—	—
<i>pred_y</i>	—	—	—



[부록] 질문...

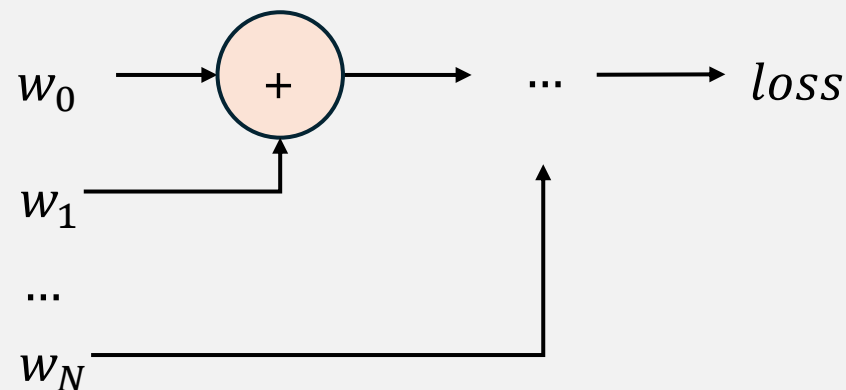
Q. '역전파'와 '오차 역전법' 은 같은건가요?

➤ 답변: 역전파 = 오차 역전법 = backpropagation

Q. '오차 역전법'에서 '오차'는 무슨 뜻인지?

➤ 답변: '오차'는 손실 함수의 값을 의미함.

(오차 = 손실 = Error = Loss = Objective = Cost ... 이름이 참 많음..)



감사합니다