Machine Learning

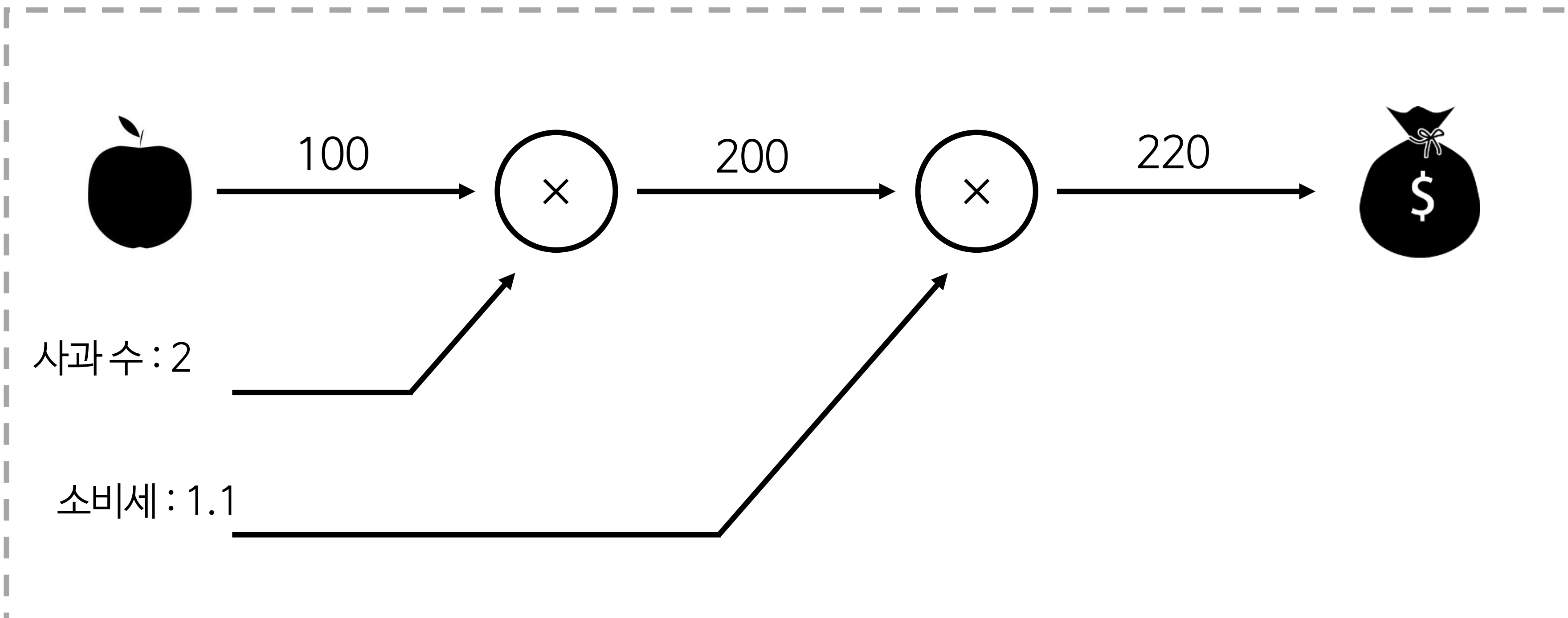
Chapter 5

오차역전파법

5. 오차역전파법

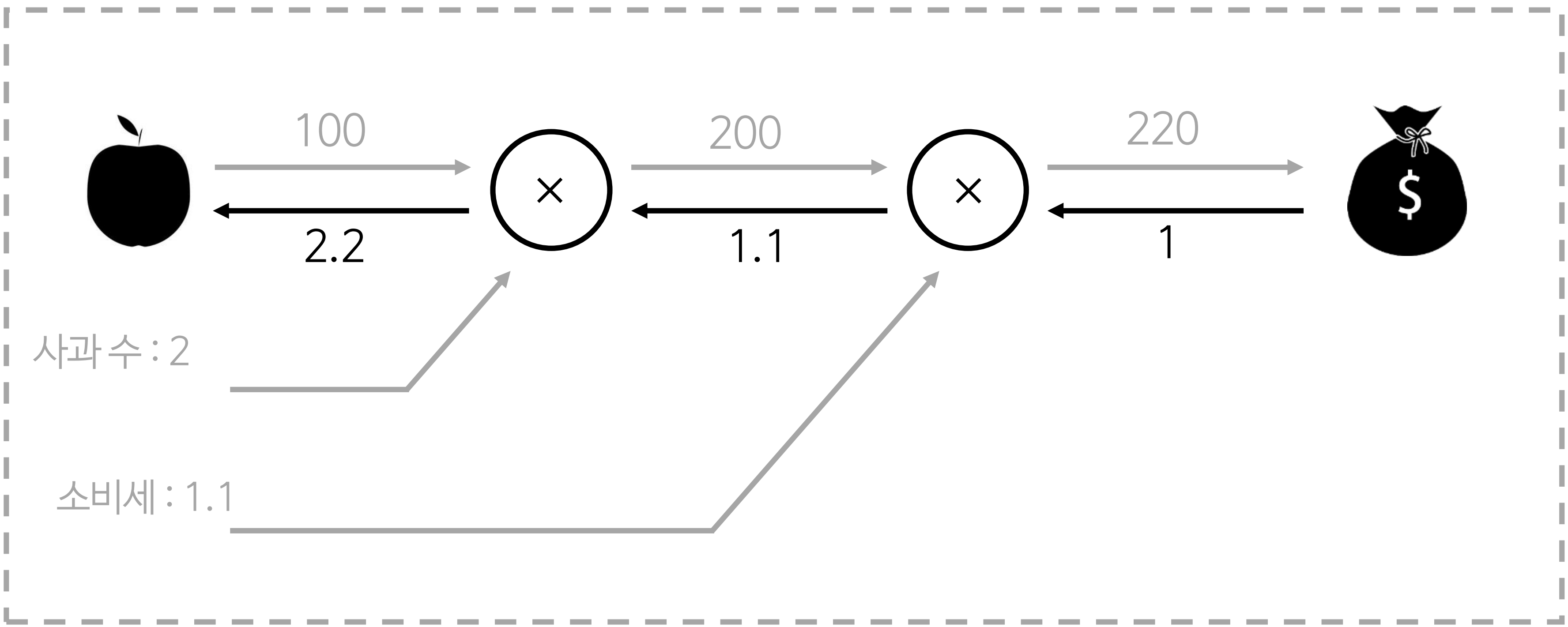
5.1 계산 그래프

- 노드(node), 엣지(edge) 구성
- 순전파(forward propagation)
  : 왼쪽에서 오른쪽으로 진행



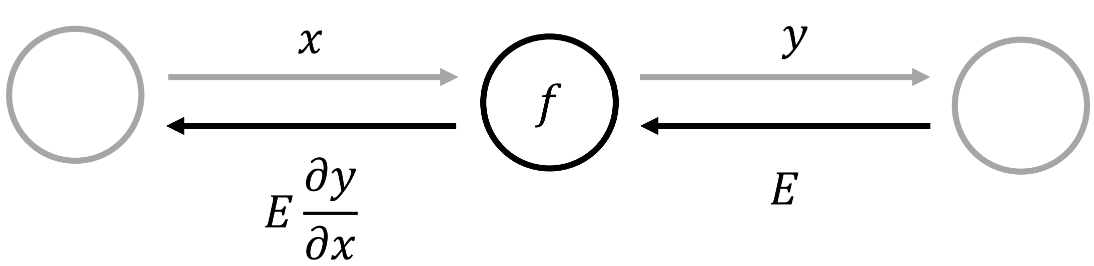
5.1 계산 그래프

- 사과 가격이 오르면 최종 금액에 어떤 영향을 끼칠까?
  → 사과 가격에 대한 지불 금액의 미분
- 역전파(back propagation)에 의한 미분 값의 전달



5.2 연쇄법칙

- 계산그래프 상 역전파의 미분전달 원리
  → 연쇄법칙을 따름
- 연쇄법칙 : 최종 변화율은 중간 변화율들의 곱으로 나타내어짐

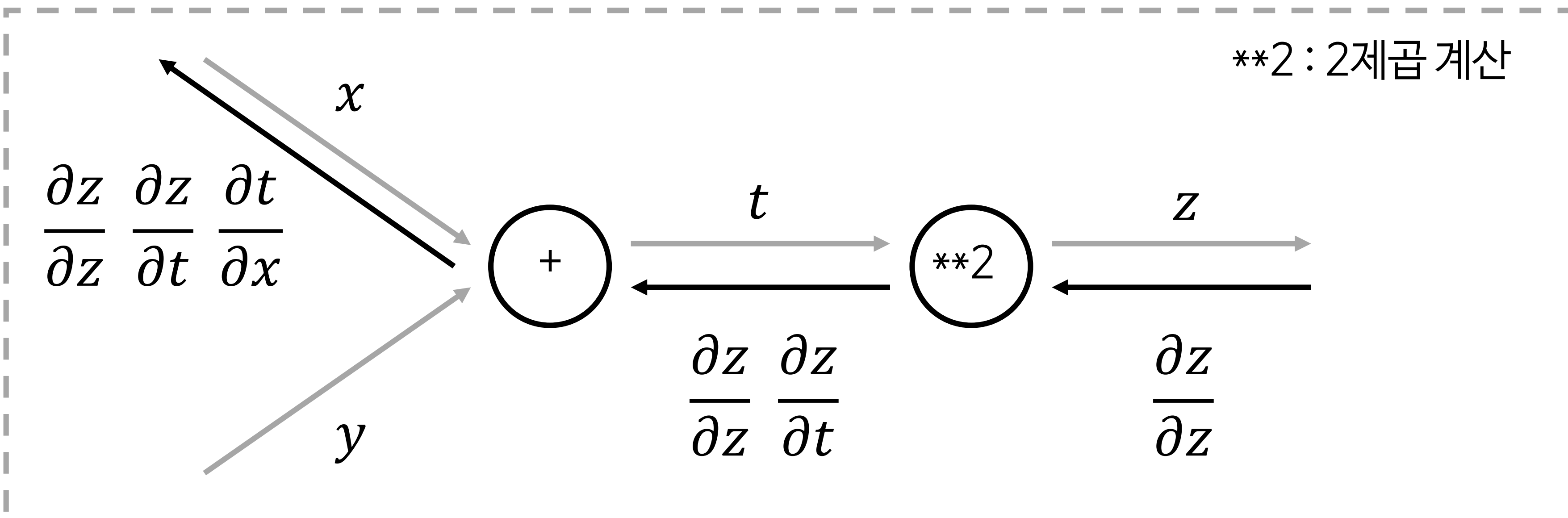


- 순방향과는 반대 방향으로 국소적 미분을 곱한다

5.2 연쇄법칙

- 계산그래프의 역전파를 통한 합성함수의 미분

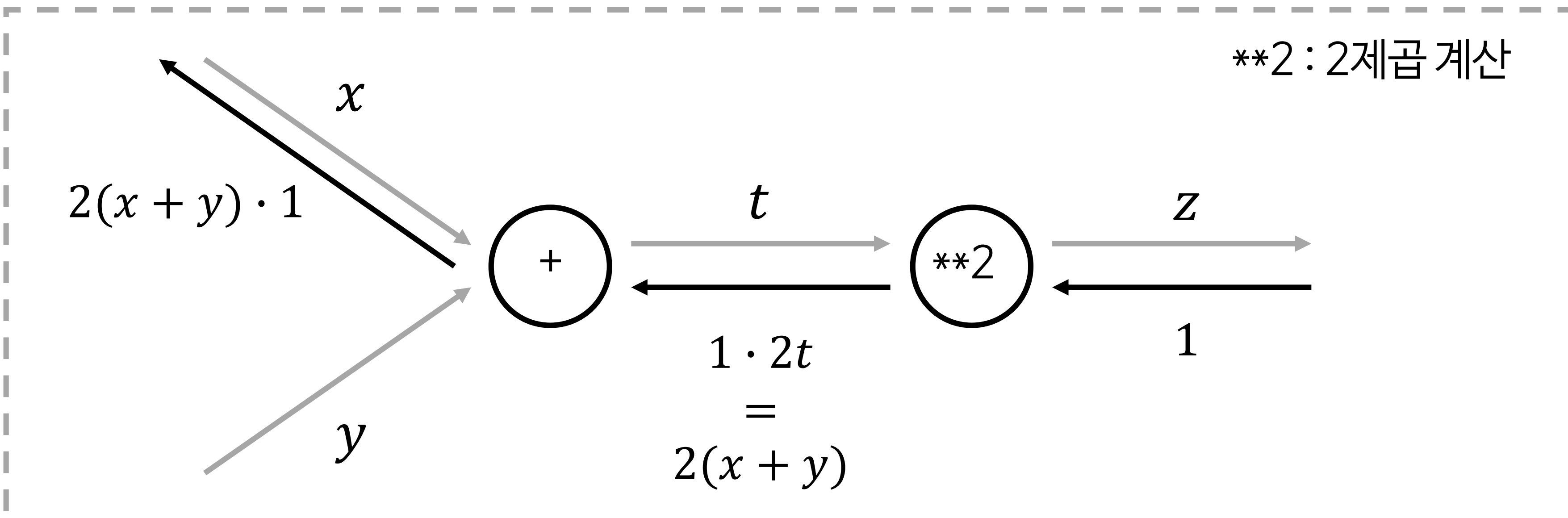
  - 예) $z = (x+y)^2$ $\quad \begin{cases} t = x + y \\ z = t^2 \end{cases}$ 일 때, $\dfrac{\partial z}{\partial x}$?



5.2 연쇄법칙

$$\begin{cases} t = x + y \\ z = t^2 \end{cases} \longrightarrow \begin{cases} \dfrac{\partial t}{\partial x} = 1 \\ \dfrac{\partial z}{\partial t} = 2t \end{cases}$$

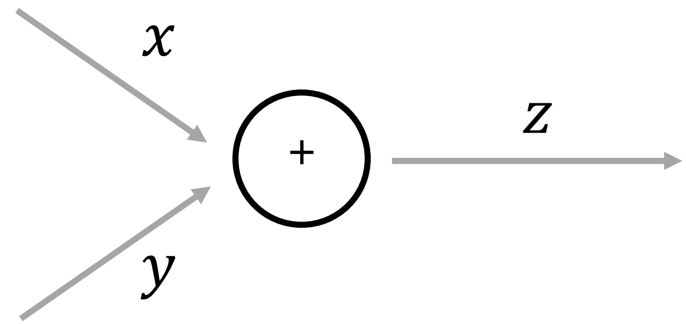
$$\therefore \frac{\partial z}{\partial x} = \frac{\partial z}{\partial t}\frac{\partial t}{\partial x} = 2t \cdot 1 = 2(x+y)$$



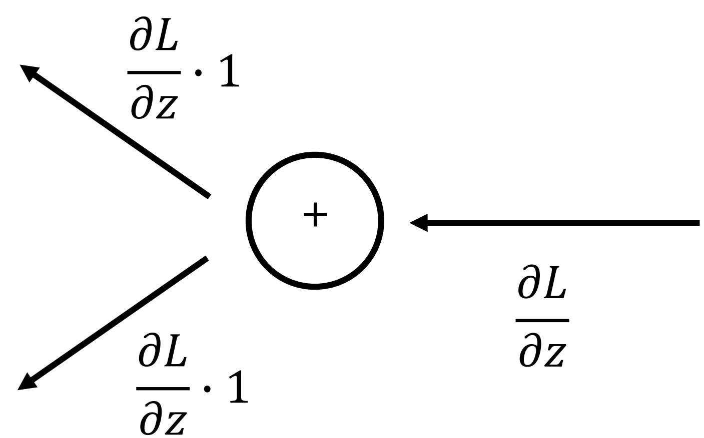
5.3 역전파

- 덧셈노드의 역전파

  - $z = x + y$



  - $\begin{cases} \dfrac{\partial z}{\partial x} = 1 \\ \dfrac{\partial z}{\partial y} = 1 \end{cases}$

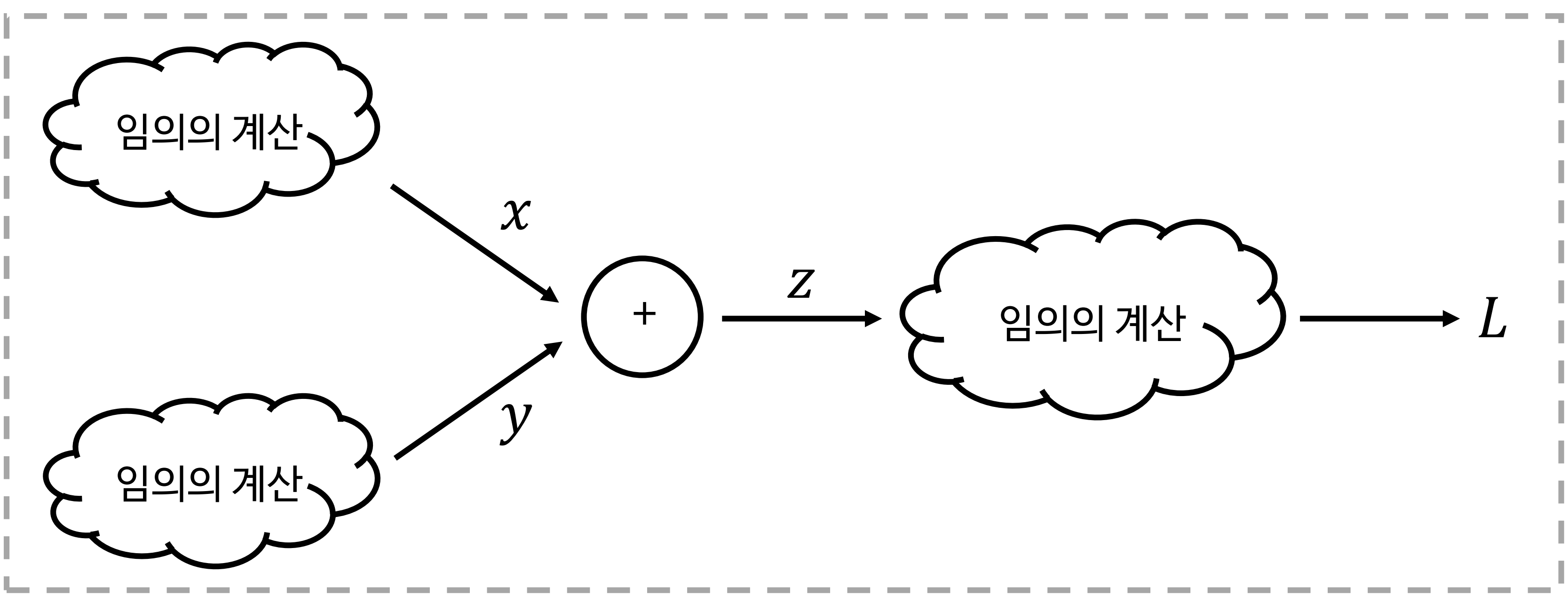


– 덧셈노드의 역전파는 입력값을 그대로 흘려보낸다.

5.3 역전파

▪ 계산그래프의 특징 – 국소적 미분

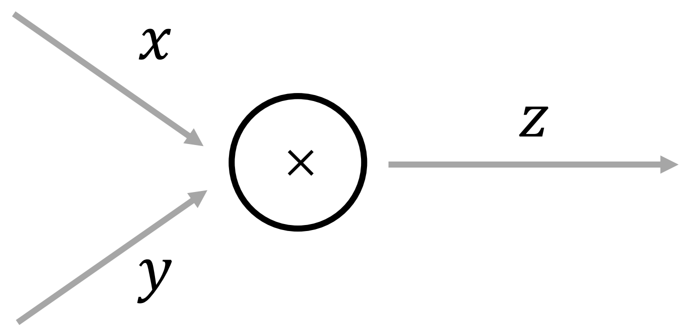
- 최종적으로 L이라는 값을 출력하는 큰 계산그래프를 가정
- $z = x + y$ 은 큰 계산 그래프의 어딘가에 존재
- 상류로부터 $\frac{\partial L}{\partial z}$ 값이 전달되어 $z = x + y$에 대한 국소적 미분을 수행



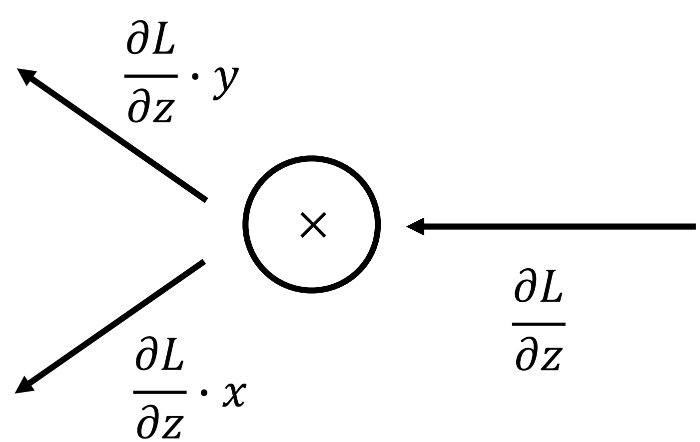
5.3 역전파

- **곱셈노드의 역전파**
  - 순전파 때의 입력신호들을 서로 바꾼 값을 곱한다.

- $z = xy$



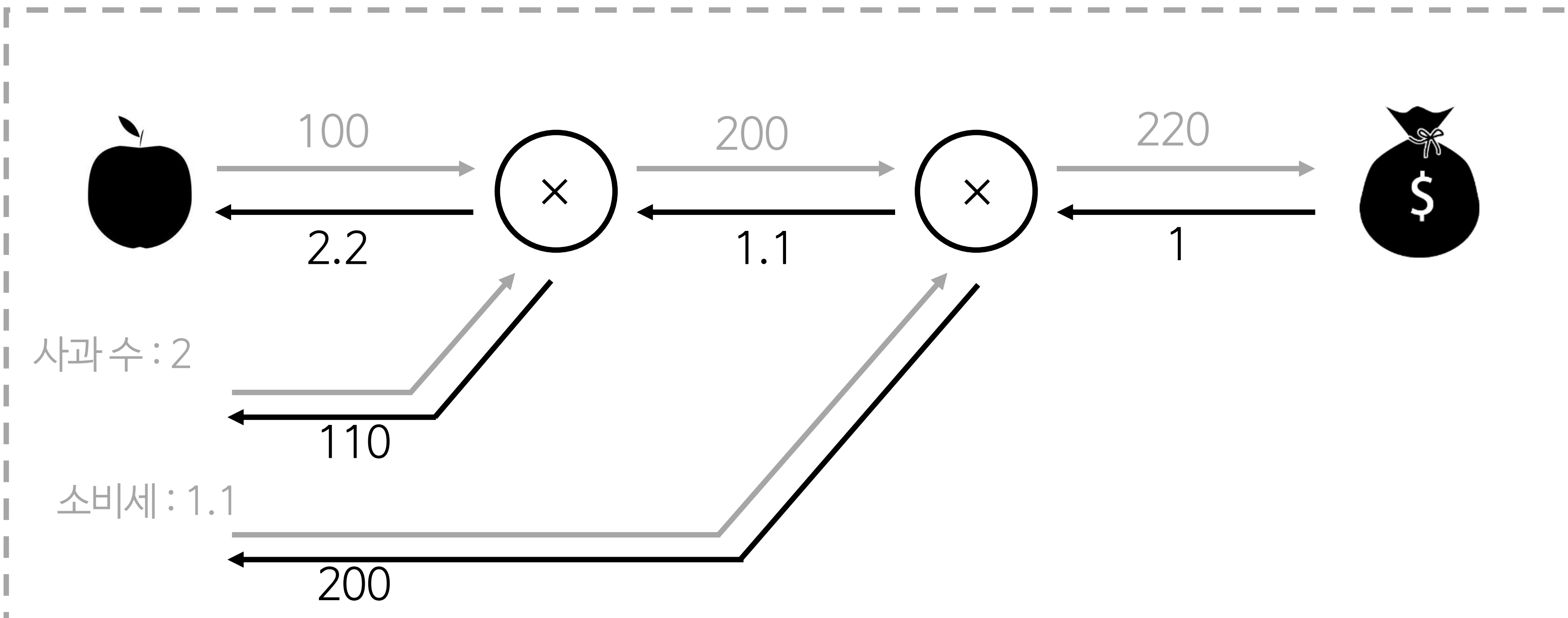
- $$\begin{cases} \dfrac{\partial z}{\partial x} = y \\ \dfrac{\partial z}{\partial y} = x \end{cases}$$



5.3 역전파

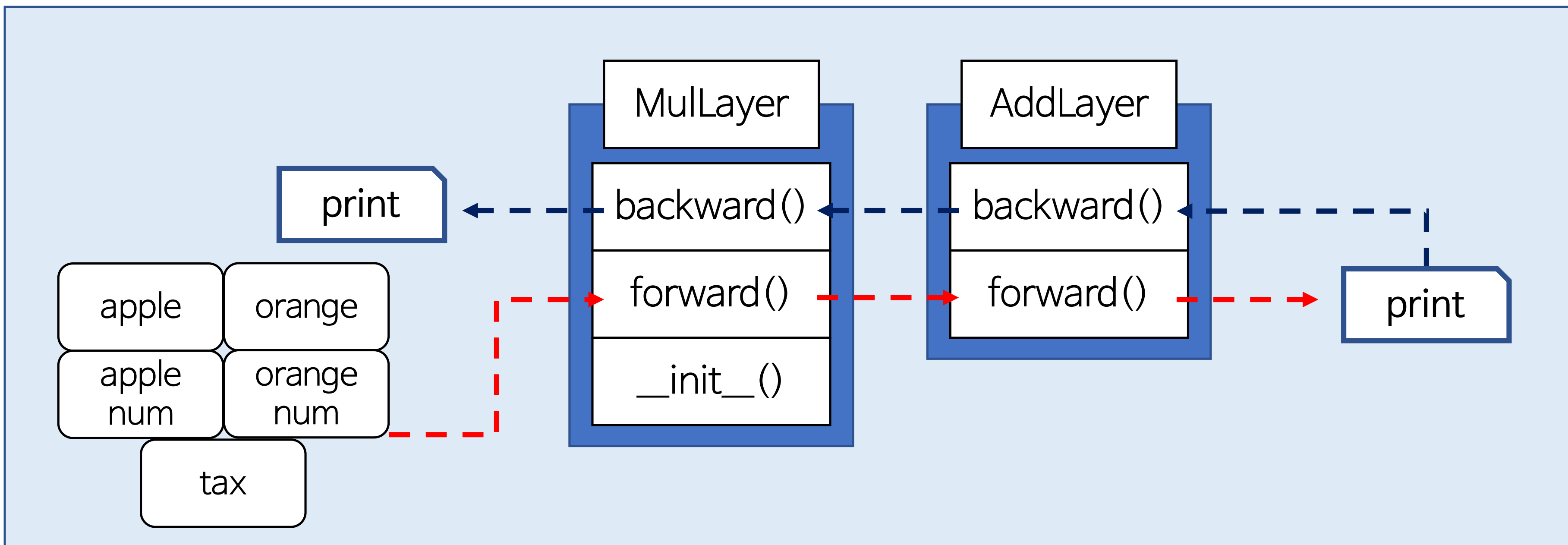
예)

- (사과 가격의 미분) = 2.2
- (사과 개수의 미분) = 110
- (소비세의 미분) = 200



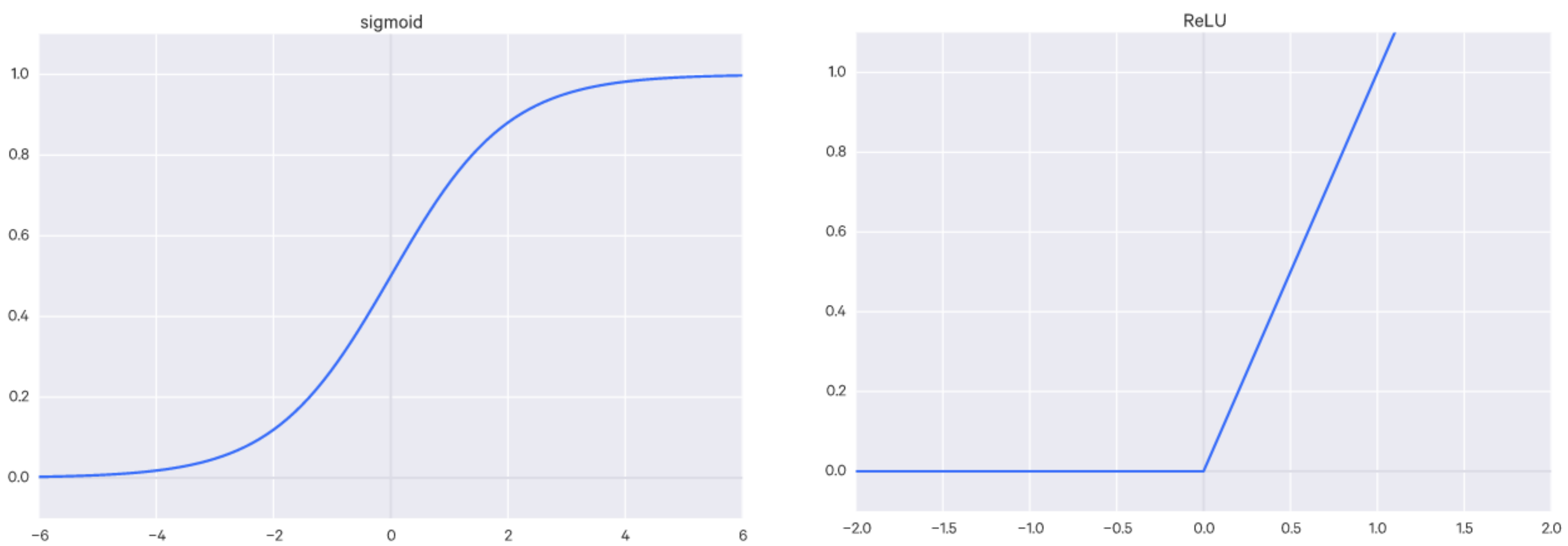
5.4 단순한 계층 구현하기

- 계층별로 순방향 전파된 계산값과 역방향 전파된 미분값 확인
- 계산 그래프를 이용해 복잡한 미분도 계산할 수 있다. (신경망)



5.5 활성화 함수 계층 구현하기

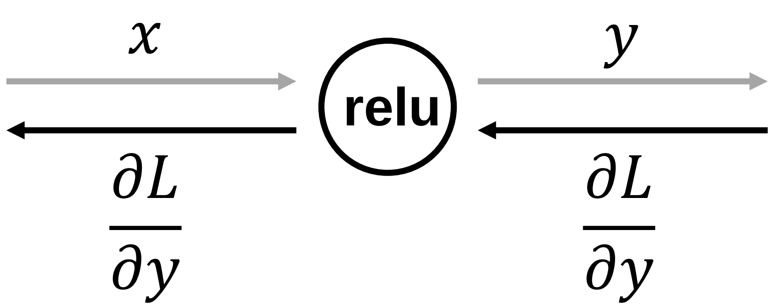
- Sigmoid계층, ReLU계층



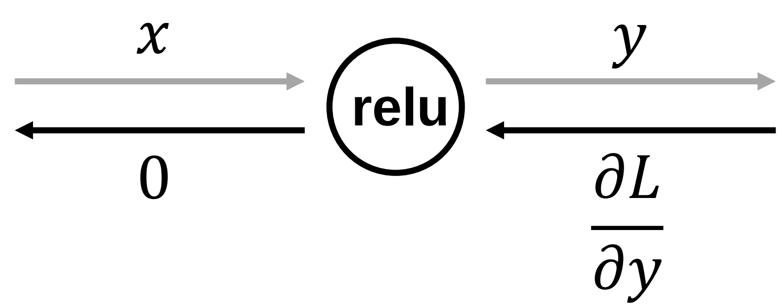
5.5 활성화 함수 계층 구현하기

- ReLU계층의 계산
  - $y = \begin{cases} x \ (x > 0) \\ 0 \ (x \le 0) \end{cases}$
  - $\frac{\partial y}{\partial x} = \begin{cases} 1 \ (x > 0) \\ 0 \ (x \le 0) \end{cases}$

- ReLU계층 미분 계산그래프



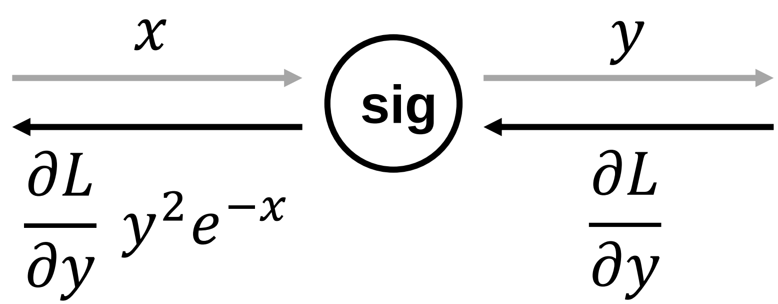
$x > 0$



$x \le 0$

5.5 활성화 함수 계층 구현하기

- Sigmoid계층의 계산
  - $y = \dfrac{1}{1+e^{-x}}$
- Sigmoid계층 미분 계산그래프



- 다음과 같이 정리할 수 있다.

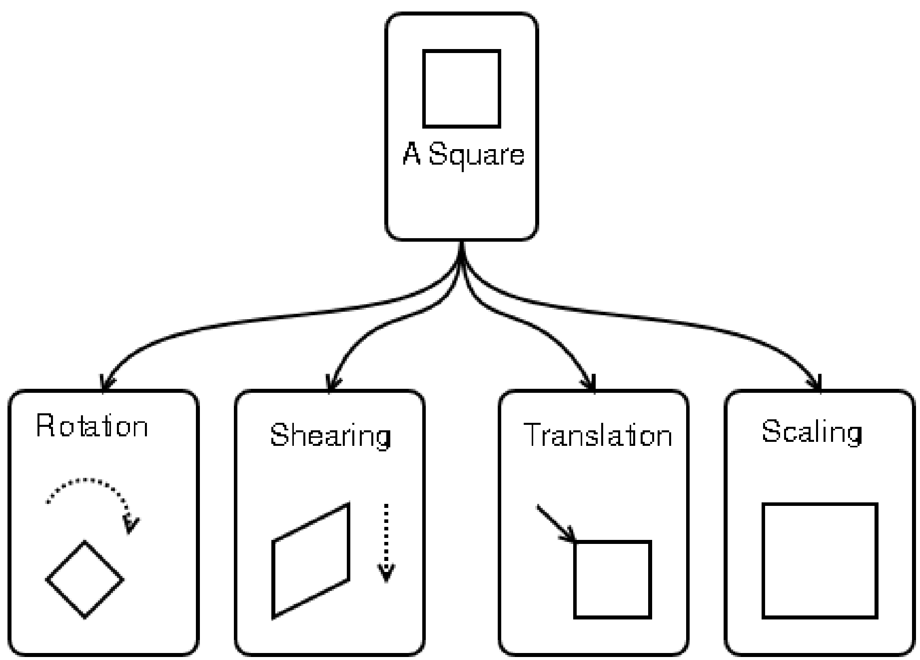
$$\frac{\partial L}{\partial y}\frac{1}{(1+e^{-x})^2}e^{-x} = \frac{\partial L}{\partial y}\frac{1}{1+e^{-x}}\frac{e^{-x}}{1+e^{-x}}$$

$$= \frac{\partial L}{\partial y}\frac{1}{1+e^{-x}}\left(1-\frac{1}{1+e^{-x}}\right)$$

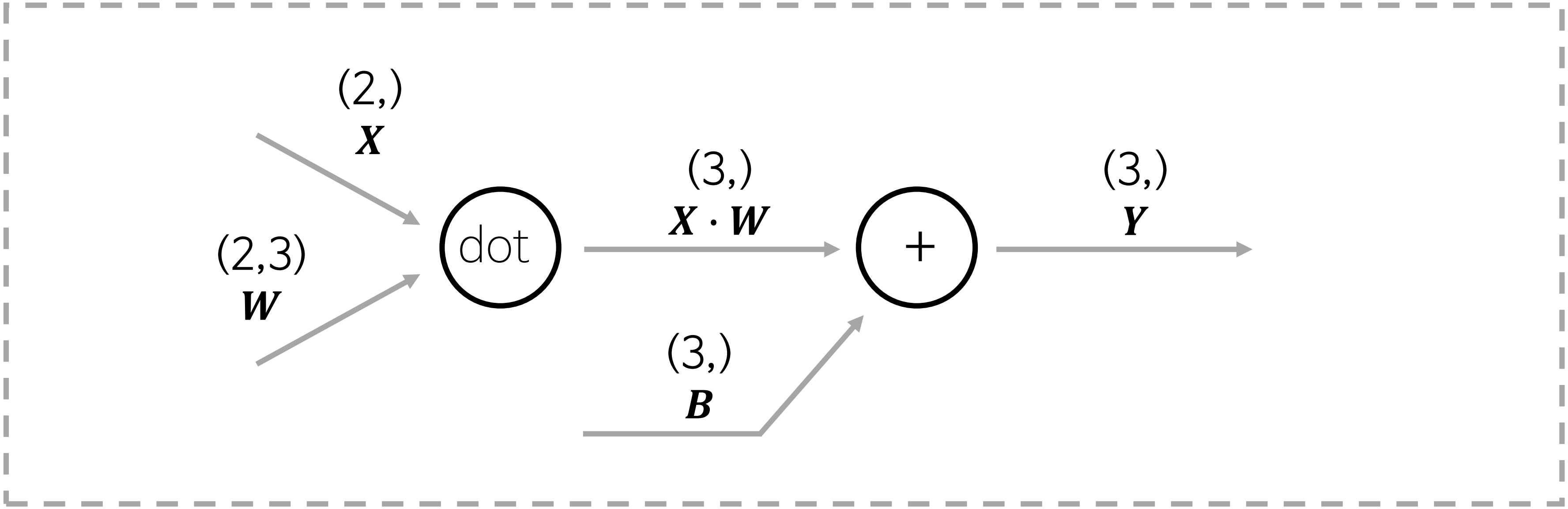
$$= \frac{\partial L}{\partial y}y(1-y)$$

5.6 Affine / Softmax 계층 구현하기

- Affine 변환
  - 벡터공간 상에서 병진(translation), 축척(scaling), 회전(rotation), 층 밀리기(shearing)등과 같은 변환을 수행
  - 직선(선형성)이 보존된다.



- Affine계층 : 행렬의 내적을 이용해 신경망의 순전파를 수행하는 계층

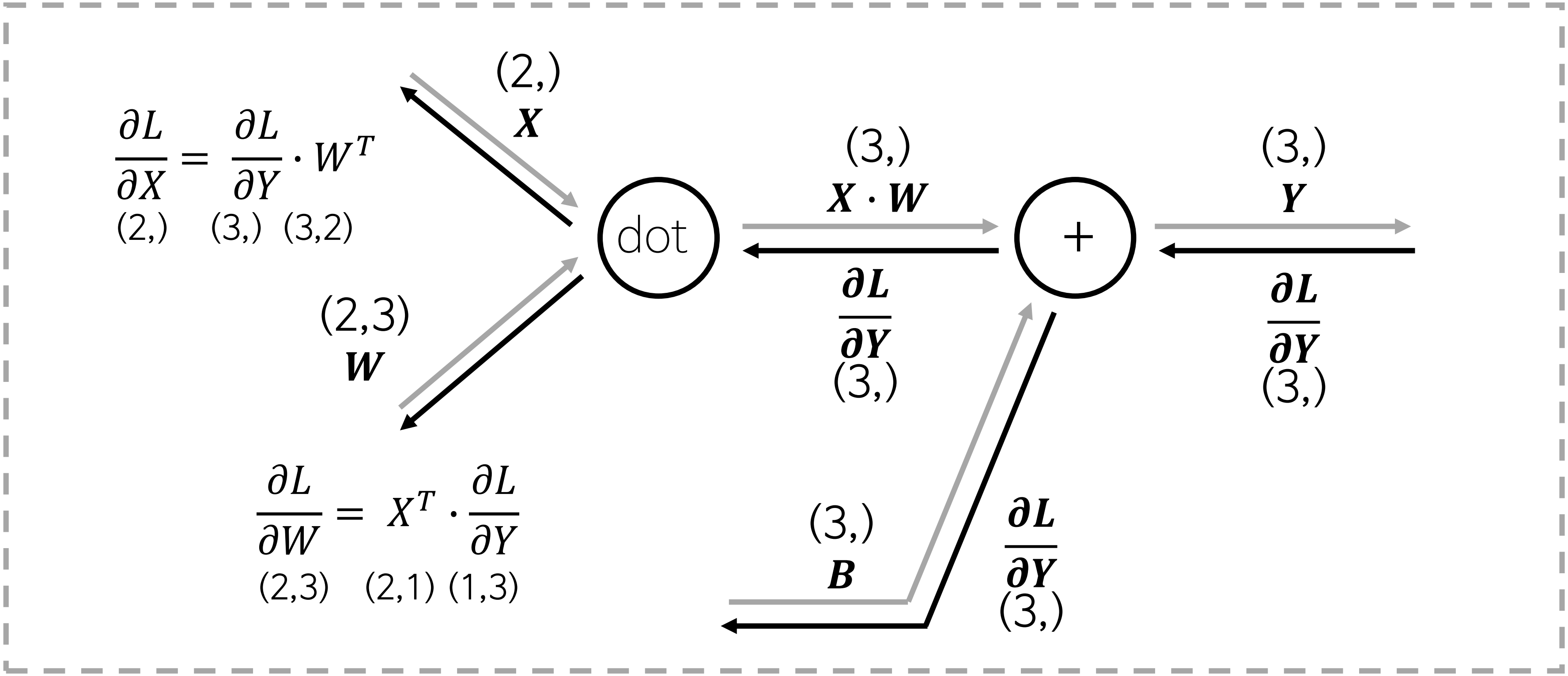


- 입력(X)과 가중치(W)의 역전파 미분값은 다음과 같다.

  - $\frac{\partial L}{\partial X} = \frac{\partial L}{\partial Y} \cdot W^T$
  - $\frac{\partial L}{\partial W} = X^T \cdot \frac{\partial L}{\partial Y}$

- Affine계층의 역전파

  – 행렬 내적의 역전파는 행렬의 대응하는 차원의 원소 수가 일치하도록 내적을 조립



- Softmax-with-Loss계층

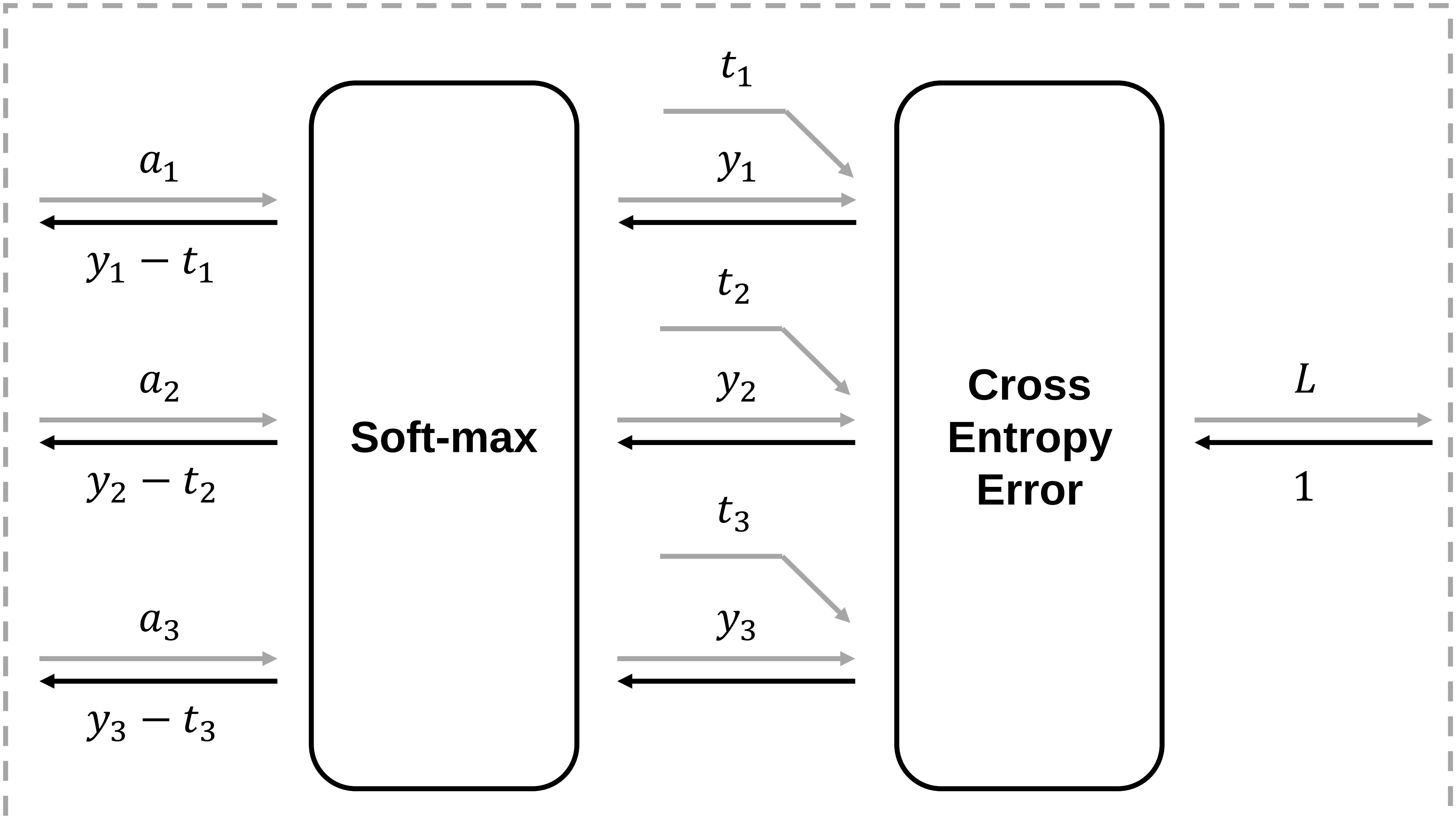
  - 확률로 정규화된 출력과 그에 대한 손실 교차엔트로피를 출력하는 계층.

1. 입력 (a_1, a_2, a_3)를 정규화하여 (y_1, y_2, y_3)를 출력
2. Cross Entropy Error 계층은 Softmax의 출력 (y_1, y_2, y_3)과 정답레이블 (t_1, t_2, t_3)을 받고, 이들 데이터로부터 손실 L을 출력
3. Softmax 계층의 출력과 정답 레이블의 차분인 $(y_1 - t_1, y_2 - t_2, y_3 - t_3)$ 값이 역전파를 통해 앞 계층으로 전달.

→ 신경망의 출력이 정답 레이블과 가까워지도록 가중치 매개변수의 값을 조정하기 위함

- Softmax-with-Loss계층

  - 확률로 정규화된 출력과 그에 대한 손실 교차엔트로피를 출력하는 계층.



5.7 오차역전파법 구현하기

전제

신경망에는 적응 가능한 가중치와 편향이 있고, 이 가중치와 편향을 훈련 데이터에 적응하도록 조정하는 과정을 학습이라 한다. 신경망 학습은 다음과 같이 4단계로 수행한다.

1단계 – 미니배치

훈련 데이터 중 일부를 무작위로 가져온다. 이렇게 선별한 데이터를 미니배치라 하며, 그 미니배치의 손실 함수 값을 줄이는 것이 목표이다.

2단계 – 기울기 산출

미니배치의 손실 함수 값을 줄이기 위해 각 가중치 매개변수의 기울기를 구한다. 기울기는 손실 함수의 값을 가장 작게 하는 방향을 제시한다.

3단계 – 매개변수 갱신

가중치 매개변수를 기울기 방향으로 아주 조금 갱신한다.

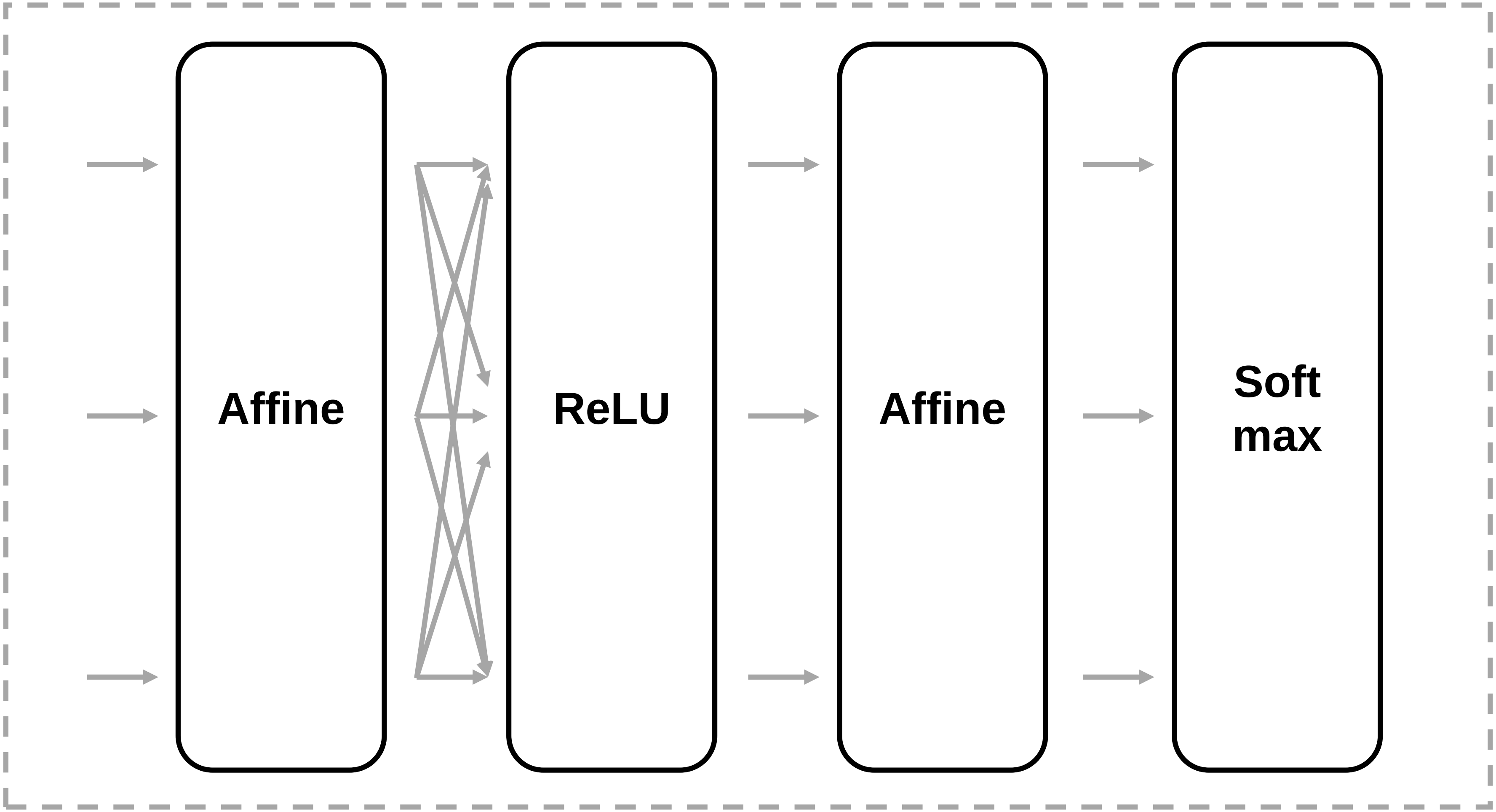
4단계 – 반복

1~3단계를 반복한다.

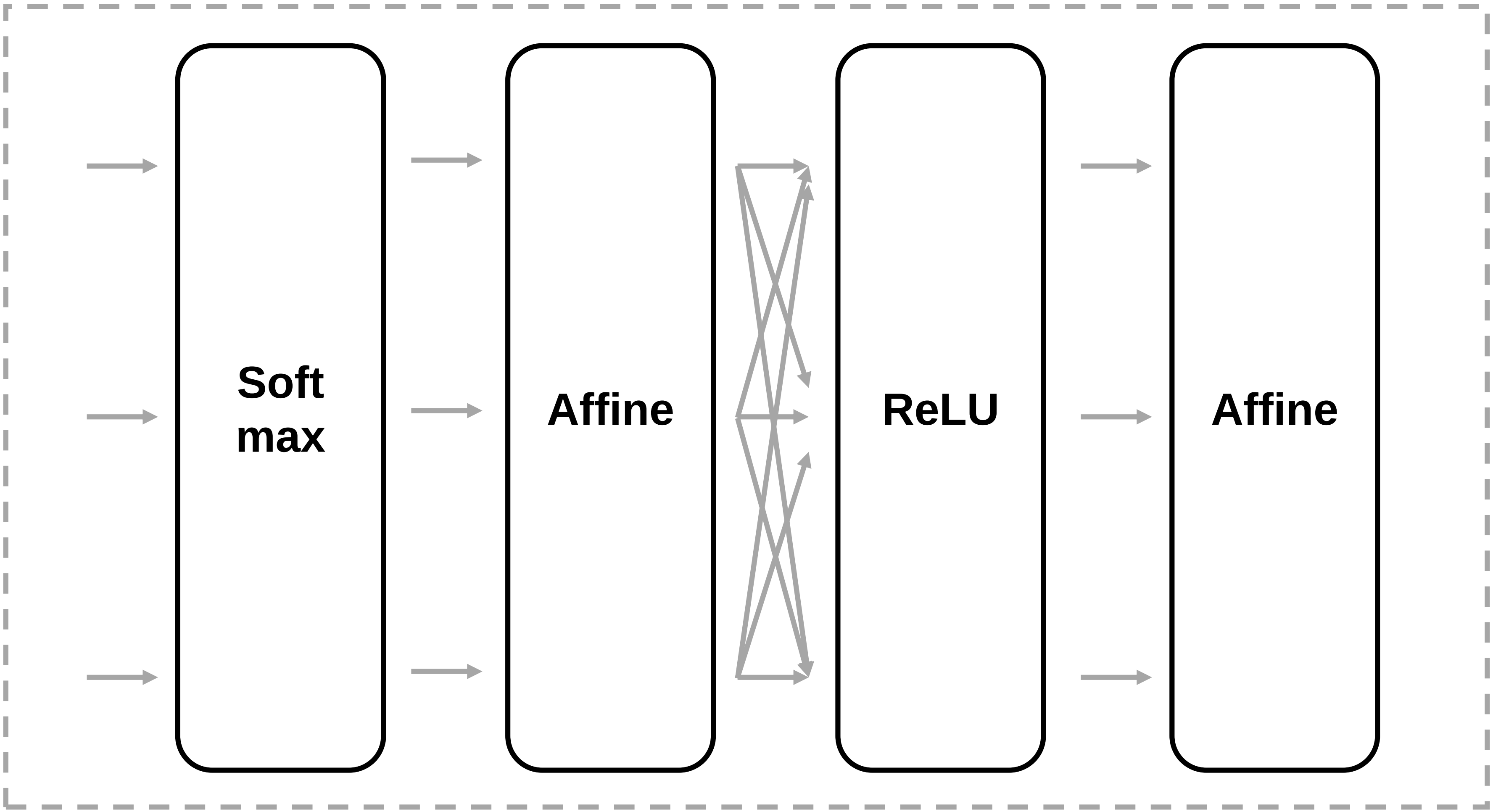
- 오차 역전파법 적용

  - 2단계(기울기 산출)에 있어 수치 미분보다 효율적으로 기울기를 산출.

- layer.forward()



- 역전파(backward propagation)
  - 계층을 반대 순서로 호출한다.
- layer.backward()



- 기울기 확인(gradient check)
  - 오차역전파법은 구현하기 복잡해 버그가 숨어있을 수 있다.
  - 결과의 검증을 위해 '수치 미분' 결과와 비교하여 검증 필요.

- (오차역전파법으로 구한 기울기) - (수치 미분으로 구한 기울기)

= avg (abs ($(w_{역전파}, b_{역전파}) - (w_{수치}, b_{수치})$))

5.8 정리

이번 장에서 배운 내용

- 계산 그래프를 이용하면 계산 과정을 시각적으로 파악할 수 있다.
- 계산 그래프의 노드는 국소적 계산으로 구성된다. 국소적 계산을 조합해 전체 계산을 구성한다.
- 계산 그래프의 순전파는 통상의 계산을 수행한다. 한편, 계산 그래프의 역전파로는 각 노드의 미분을 구할 수 있다.
- 신경망의 구성 요소를 계층으로 구현하여 기울기를 효율적으로 계산할 수 있다(오차역전파법).
- 수치 미분과 오차역전파법의 결과를 비교하면 오차역전파법의 구현에 잘못 없는지 확인할 수 있다(기울기 확인).