

# 제1 고지 : 미분 자동 계산

## STEP 5 : 역전파 이론

### 5.1 연쇄 법칙(Chain Rule)

그림 5-1 합성 함수의 예



$$\frac{dy}{dx} = \frac{dy}{db} \frac{db}{da} \frac{da}{dx} = \frac{dy}{dy} \frac{dy}{db} \frac{db}{da} \frac{da}{dx}$$

- 위에서 알 수 있듯이 합성함수의 미분은 구성함수 각각의 미분값을 연쇄적으로 모두 곱한 값

### 5.2 역전파와 원리 도출

그림 5-2 출력 쪽의 미분부터 순서대로 계산

$$\begin{aligned} \frac{dy}{dx} &= \left( \left( \left( \frac{dy}{dy} \frac{dy}{db} \right) \frac{db}{da} \right) \frac{da}{dx} \right) \\ \textcircled{1} \quad \frac{dy}{dy} \frac{dy}{db} &= \frac{dy}{db} \\ \textcircled{2} \quad \frac{dy}{db} \frac{db}{da} &= \frac{dy}{da} \\ \textcircled{3} \quad \frac{dy}{da} \frac{da}{dx} &= \frac{dy}{dx} \end{aligned}$$

그림 5-3  $\frac{dy}{dx}$ 를 구하는 계산 그래프

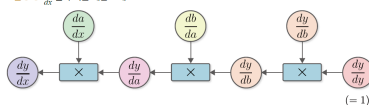
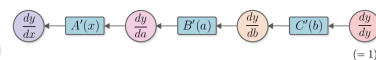


그림 5-4 단순화된 역전파 계산 그래프(A'(x)의 곱셈을 A'(x)라는 노드로 간략하게 표현)



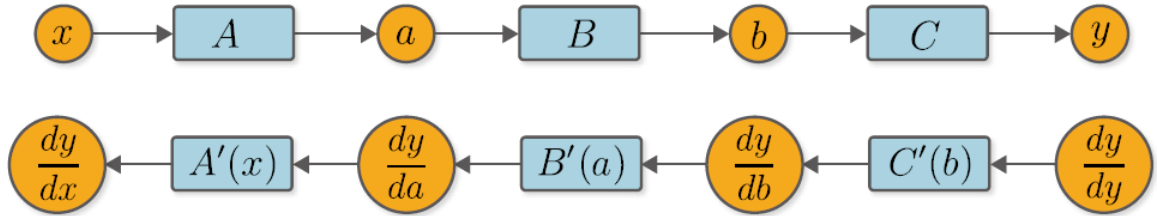
출력 쪽의 미분부터 순서대로 계산

- 위의 그림에서 알 수 있듯이,  $y$ 의 각 변수에 대한 미분값이 오른쪽에서 왼쪽으로 전파되는 것을 알 수 있다.

- 머신러닝은 주로 대량의 매개변수를 입력받아서 마지막 **손실함수(loss function)** 을 거쳐 출력을 내는 형태로 진행된다. 이때, 손실함수의 출력은 많은 경우 단일한 스칼라값이며, 이 값이 **중요요소** 이다. 즉, 손실함수의 각 매개변수에 대한 미분을 계산해야 한다. 이런 경우 미분값을 출력에서 입력방향으로 전파하면 한 번의 전파만으로 모든 매개변수에 대한 미분을 효율적으로 계산할 수 있다.

### 5.3 계산 그래프로 살펴보기

그림 5-5 순전파(위)와 역전파(아래)



- 순전파 시의 변수  $a$  는 역전파 시의 미분  $\frac{dy}{da}$  ,  $b$  는  $\frac{dy}{db}$  ,  $x$  는  $\frac{dy}{dx}$  가 대응한다.
- 또한 함수  $A$  는  $A'(x)$  /  $B$  는  $B'(x)$  /  $C$  는  $C'(x)$  가 대응한다.
- 주목해야 할것은 **역전파를 계산하기 위해서는 순전파 시 이용한 데이터가 필요하다**는 것이다.
  - 함수  $C$  를 살펴보면,  $C'(b)$  를 구하기 위해선 순전파 시  $C$  에 입력된  $b$  값이 필요하다.

💡 역전파를 계산하기 위해선 **\*\*순전파 시의 입력값을 기억\*\***해둬야 한다. (= 순전파가 진행되어야 역전파가 진행될 수 있다)