ch17_7_partial_derivative_chain_rule

Tags

ch17_7 parital derivative, chain rule
편미분 기초
Chain Rule
정리

ch17_7 parital derivative, chain rule

역전파는 딥 러닝 모델을 학습시키는 가장 핵심적인 알고리즘입니다. 오류 역전파를 이해하기 위해서는 편미분과 chain rule을 알아야합니다. 직접 구현할 일은 없지만, 핵심적인 개념을 이해하면 딥 러닝 모델이 어떻게 학습되는지 이해하는데 큰 도움이 됩니다.

편미분 기초

역전파는 기본적으로 편미분 개념을 사용합니다. 간단하게 미분 개념을 짚고 넘어가겠습니다. 미분 먼저 살펴보겠습니다. 미분을 직관적으로 이해하면 x가 아주 작은값으로 변할 때의 순간 변화량입니다.

$$\frac{d}{dx}f(x) = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

$$f(x) = 3,\ \frac{\partial f}{\partial x} = 0$$
$$f(x) = x,\ \frac{\partial f}{\partial x} = 1$$
$$f(x) = 2x,\ \frac{\partial f}{\partial x} = 2$$
$$f(x) = x + 3,\ \frac{\partial f}{\partial x} = 1$$

위에 x에 대한 함수를 각각 미분하면 0, 1, 2가 나옵니다. (자세한 풀이는 생략합니다. 미분에 대한 개념을 보충하고 싶으신 분들은 아래 강의를 참고해주세요)

https://www.youtube.com/watch?v=rG0msnwLF68

이번엔 편미분을 살펴보겠습니다. 편미분은 영어로 partial derivative로 미분하고자 하는 변수를 제외한 나머지는 상수로 취급한 뒤, 미분하는 기법입니다.

$$f(x) = 2x,\ \frac{\partial f}{\partial x} = 2$$
$$f(x, y) = xy,\ \frac{\partial f}{\partial x} = y$$
$$f(x, y) = xy,\ \frac{\partial f}{\partial y} = x$$

f(x)=2x를 x에 대해서 편미분하면 그냥 미분했을 때와 동일하게 2가 나옵니다. f(x, y)=xy를 x에 대해서 편미분을 해주면 x를 제외한 나머지 변수들을 상수 취급을 해주므로 y가 나옵니다. 반대로 y에 대해서 편미분을 해주면, x를 상수 취급해주니까 x가 나옵니다.

$$f(x, y) = x + y + 3, \frac{\partial f}{\partial x} = 1$$
$$f(x, y) = x + 2y, \frac{\partial f}{\partial x} = 1$$
$$f(x, y) = x + 2y, \frac{\partial f}{\partial y} = 2$$

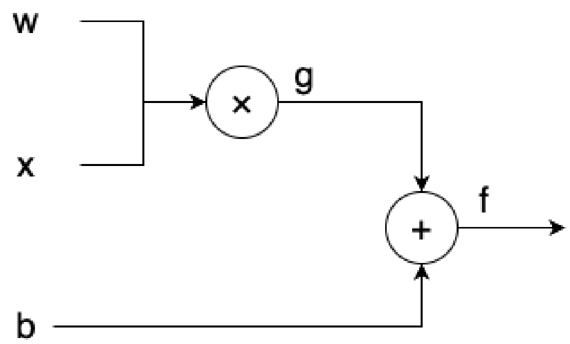
이제 x와 y가 더하기로 분리되어 있을 때입니다. 마치 f(x)=x+3를 x에 대해 미분했을 때 3을 날려준 것 처럼, f(x)=x+y+3을 x에 대해 편미분 할 때는 y를 상수 취급하여 날려주면 됩니다. y에 대해서 편미분을 한다면 반대로 x를 날려주면 됩니다.

Chain Rule

$$f(g(x))$$
$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial g} \cdot \frac{\partial g}{\partial x}$$

chain rule은 합성 함수의 미분을 구하고 싶을 때, 적용할 수 있는 법칙입니다. 각각의 미분을 구한 다음 곱해주면 됩니다. 간단하게 f(x) = wx + b 함수에 chain rule을 적용해보겠습니다.

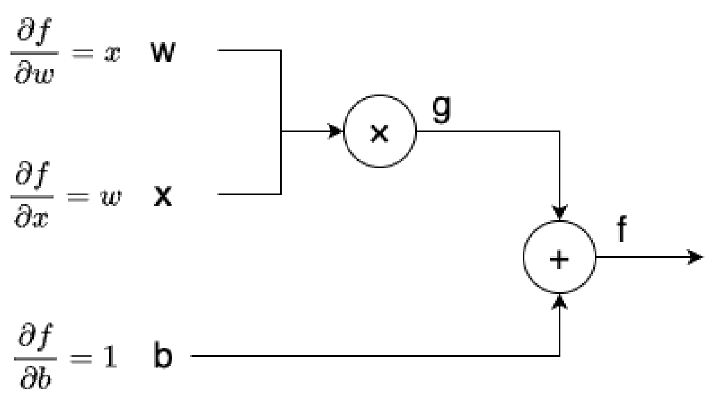
$$g(x) = wx$$
$$f(g) = g + b$$
$$f(g(x)) = wx + b$$



간단한 linear 함수를 합성함수로 표현해 보았습니다. 여기서 각 변수에 대한 편미분 값은 최종 결과값에 미치는 영향이라고 해석할 수 있습니다. 한번 각 변수별로 편미분 값을 구해보겠습니다.

간단한 linear 함수를 합성함수로 표현해 보았습니다. 여기서 각 변수에 대한 편미분 값은 최종 결과값에 미치는 영향이라고 해석할 수 있습니다. 한번 각 변수별로 편미분 값을 구해보겠습니다.

$$\frac{\partial f}{\partial g} = 1$$
$$\frac{\partial g}{\partial w} = x$$
$$\frac{\partial g}{\partial x} = w$$
$$\frac{\partial f}{\partial w} = \frac{\partial f}{\partial g} \cdot \frac{\partial g}{\partial w} = 1 \cdot x = x$$
$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial g} \cdot \frac{\partial g}{\partial x} = 1 \cdot w = w$$
$$\frac{\partial f}{\partial b} = 1$$

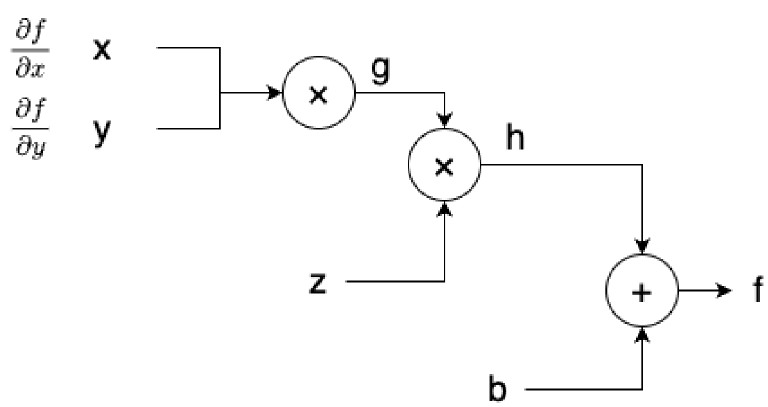


각 변수의 편미분 값이 최종 결과값에 미치는 영향력이라는 말을 예시를 통해서 알아보겠습니다.

w	x	b	f(g(x))
-2	5	3	-7
-1	5	3	-2
-2	6	3	-9
-2	5	4	-6

자 처음에 w, x, b 값이 각각 -2, 5, 3으로 주어졌을 때 최종 결과 값이 -7이 나왔습니다. 그 다음, w 값을 1 증가 시켜서 -1, 5, 3이 주어졌을 때 최종 결과값은 -2가 되었습니다. 즉, w가 한 단위 증가했을 때 최종 결과값은 5만큼 증가했으며, 이는 편미분을 통해서 구한 값과 동일합니다. 나머지 x와 b도 각각 한 단위 씩 증가시켜보면 각각의 편미분 값만큼 최종 결과값이 변화하는 것을 확인할 수 있습니다.

더 복잡한 함수는 어떻게 될까요?



$$g(x, y) = xy$$
$$h(g) = gz$$
$$f(h) = h + b$$
$$f(h(g(x, y), z))$$

합성 함수 f의 최종 결과 값에 대해서 x가 미치는 영향력을 편미분으로 계산해보겠습니다. chain rule을 이용해서 차근차근 계산해보면 아래처럼 구할 수 있습니다.

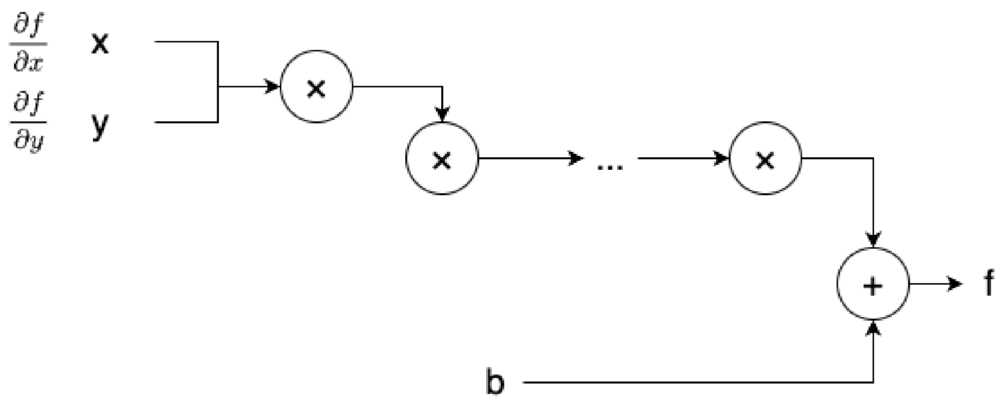
$$\frac{\partial f}{\partial x} = \frac{\partial g}{\partial x} \cdot \frac{\partial h}{\partial g} \cdot \frac{\partial f}{\partial h} = y \cdot z \cdot 1 = yz$$

실제로 x를 한 단위 변경했을 때, 최종 결과값이 yz만큼 변화하는지 확인해보겠습니다.

x	y	z	b	f(h(g(x, y), z))
1	2	3	4	10

2	2	3	4	16

실제로 x가 1만큼 증가하면 x에 대한 편미분 값인 yz=6 만큼 최종 결과값이 변경되었습니다. 이보다 훨씬 더 복잡한 함수라면 어떻게 될까요?



이런 상황에서도 chain rule을 이용하면 각 변수들의 편미분 값을 쉽게 구할 수 있습니다. 그리고 이는 각 변수들이 최종 결과 값에 미치는 영향력이라고 해석할 수 있습니다.

정리

이번 챕터에서는 미분과 편미분의 개념에 대해서 배워봤습니다. 그리고 chain rule을 이용하여 복잡한 형태의 합성함수에서 각 변수의 편미분 값을 계산할 수 있다는 것을 알아보았습니다. 너무 어렵게 생각할 필요는 없고, 직관적으로 편미분과 chain rule이 이렇게 동작하는 구나 정도만 이해하고 넘어가도 무방합니다.