인공지능 수학

미분, 편미분

미분의 정의

미분의 정의를 쉽게 이해하려면 먼저 도함수의 개념을 알아야 합니다. 도함수는 어떤 함수 안에 포함된 값 각각이 0에 한없이 가까워지는 극한값(미분계수)을 구하는 함수를 말합니다.[22]

함수 $y = f(x)$의 도함수 $f'(x)$는 다음처럼 정의합니다.

식 2-15[23]

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

참고로 '$\lim_{\Delta x \to 0}$ (Δx의 식)'은 "Δx라는 수가 한없이 0과 가깝다"라는 의미입니다.

그럼 예를 살펴보겠습니다. 첫 번째는 $f(x) = 3x$일 때의 도함수 계산 과정입니다.

$$f'(x) = \lim_{\Delta x \to 0} \frac{3(x+\Delta x) - 3x}{\Delta x} = \lim_{\Delta x \to 0} \frac{3\Delta x}{\Delta x} = \lim_{\Delta x \to 0} 3 = 3$$

두 번째 예는 $f(x) = x^2$일 때의 도함수 계산 과정입니다.

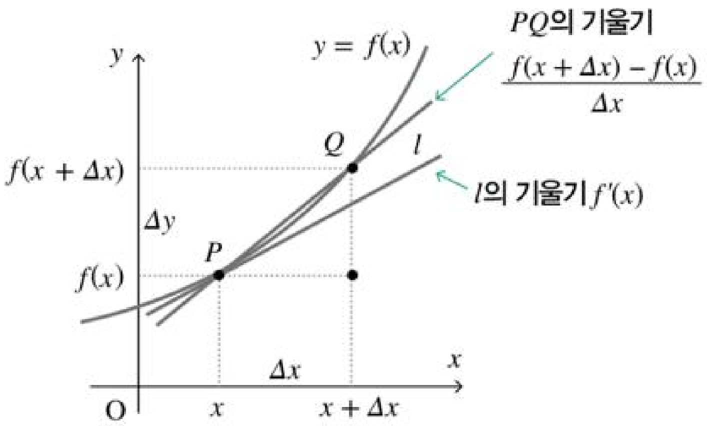
$$f'(x) = \lim_{\Delta x \to 0} \frac{(x+\Delta x)^2 - x^2}{\Delta x} = \lim_{\Delta x \to 0} \frac{2x\Delta x + (\Delta x)^2}{\Delta x} = \lim_{\Delta x \to 0} (2x + \Delta x) = 2x$$

미분의 정의

함수 $f(x)$의 도함수 $f'(x)$를 구하는 것을 "함수 $f(x)$를 미분한다"라고 합니다. 또한 [식 2-15]의 값을 계산할 수 있다면 미분 가능이라고 합니다.

그럼 도함수의 의미를 [그림 2-31]로 나타내보겠습니다. 함수 $f(x)$를 그래프로 그릴 때 $f'(x)$는 해당 그래프 **접선의 기울기**를 표현합니다. 따라서 연속되는 형태의 그래프를 갖는 함수는 미분 가능합니다.

그림 2-31



도함수의 의미. $f'(x)$는 그래프 접선의 기울기를 표현합니다. 실제로 Q를 한없이 P에 가깝게 하면 직선 PQ는 접선 l에 한없이 가까워집니다.

미분 기호

[식 2-15]에서 함수 $y = f(x)$의 도함수 $f'(x)$를 극한 개념으로 표현했는데 분수로 표현하는 방법도 있습니다. $f'(x) = dy/dx$입니다.[24]

이 표기법은 매우 편리합니다. 복잡한 함수의 미분을 분수 형식으로 계산할 수 있기 때문입니다.

예를 들어 [식 2-16]의 $(c)' = 0$은 $dc/dx = 0$으로 표현할 수 있습니다(c는 상수). 또한 $(x)' = 1$은 $dx/dx = 1$로 표현할 수 있습니다.

미분의 성질

[식 2-17]을 이용하면 미분 가능한 함수의 세계가 비약적으로 넓어집니다. 이 공식을 미분의 선형성이라고 합니다(선형성의 기본 개념은 2장 03 참고).

식 2-17[25]

$$\{f(x)+g(x)\}' = f'(x)+g'(x),\ \{cf(x)\}' = cf'(x) \quad (c\text{는 상수})$$

기억하기 어렵다면 다음처럼 기억하기 바랍니다.

- 함수 합의 미분은 각 함수를 미분한 합과 같습니다.
- 상수를 곱한 함수의 미분은 미분한 함수에 상수를 곱한 것과 같습니다.

'미분의 선형성'은 나중에 알아볼 오차역전파법의 기반이 되는 개념이기도 합니다.

그럼 예를 살펴보겠습니다. 첫 번째는 함수 $C = (2-y)^2$ (y는 변수)일 때 C'을 계산해보겠습니다. 다음과 같습니다.

$$C' = (4-4y+y^2)' = (4)'-4(y)'+(y^2)' = 0-4+2y = -4+2y$$

미분의 성질

두 번째로는 함수 $f(x) = 2x^2+3x+1$을 미분해보겠습니다. 계산 결과는 다음과 같습니다.

$$f'(x) = (2x^2)'+(3x)'+(1)' = 2(x^2)'+3(x)'+(1)' = 4x+3$$

세 번째로는 $f(x) = 1+e^{-x}$를 미분해보겠습니다. 계산 결과는 다음과 같습니다.

$$f'(x) = (1+e^{-x})' = (1)'+(e^{-x})' = -e^{-x}$$

분수 함수와 시그모이드 함수의 미분

분수 형태의 함수를 미분할 때 도움 되는 것이 분수 함수의 미분 공식입니다. [식 2−18]과 같습니다.

식 2-18[27]

$$\left\{\frac{1}{f(x)}\right\}' = -\frac{f'(x)}{\{f(x)\}^2}$$

신경망에서 가장 유명한 활성화 함수 중 하나는 시그모이드 함수입니다. 시그모이드 함수 $\sigma(x)$는 2장 01에서 $1/(1+e^{-x})$로 정의했습니다. 그런데 나중에 알아볼 경사하강법에서는 이 함수를 미분할 필요가 있습니다. 그럴 때 편리한 것이 [식 2−19]입니다.

식 2-19

$$\sigma'(x) = \sigma(x)(1-\sigma(x))$$

[식 2−19]를 이용하면 미분하지 않아도 시그모이드 함수의 도함숫값을 $\sigma(x)$의 값에서 얻을 수 있습니다. [식 2−18]의 $f(x)$에 $1+e^{-x}$을 대입한 후 [식 2−16]의 지수함수 미분 공식 $(e^{-x})' = -e^{-x}$를 이용하면 다음과 같은 식을 계산할 수 있습니다.

시그모이드 함수의 미분

$$\sigma'(x) = -\frac{(1+e^{-x})'}{(1+e^{-x})^2} = \frac{e^{-x}}{(1+e^{-x})^2}$$

이는 다음처럼 변형할 수 있습니다.

$$\sigma'(x) = \frac{1+e^{-x}-1}{(1+e^{-x})^2} = \frac{1}{1+e^{-x}} - \frac{1}{(1+e^{-x})^2} = \sigma(x) - \sigma(x)^2$$

이렇게 $\sigma(x)$의 값으로 [식 2-19]를 얻을 수 있습니다.

최소값의 필요조건

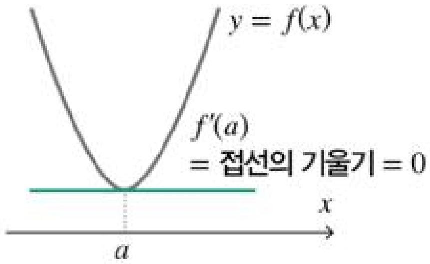
앞에서 도함수 $f'(x)$는 접선의 기울기를 나타낸다고 했습니다. 여기에서 나중에 알아볼 '최적화' (2장 12 참고)에서 이용하는 원리([식 2-20])를 얻을 수 있습니다.

식 2-20

함수 $f(x)$가 $x = a$일 때 최솟값이라면 $f'(a) = 0$

그럼 왜 0이 될까요? [그림 2-32]를 보면 분명하게 알 수 있습니다.

그림 2-32



$x=a$이고 $f(x)$가 최솟값일 때, 그 점에서 접선의 기울기(즉 도함수의 값)는 0이 된다.

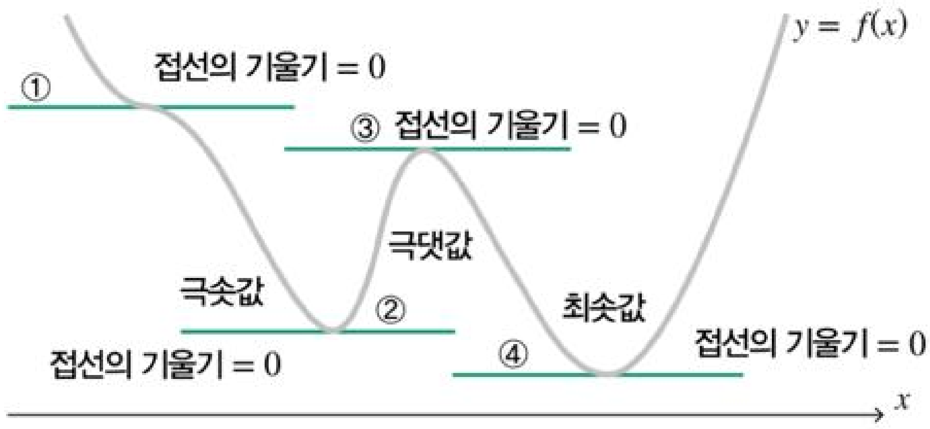
이 원리를 응용할 때는 다음 사항도 기억해둡시다.

$f'(a) = 0$는 함수 $f(x)$가 $x = a$에서 최솟값이 되기 위한 '필요' 조건[28]입니다.

최소값의 필요조건

이는 접선의 기울기가 0이더라도 꼭 최솟값이라는 보장이 없다는 의미입니다. [그림 2−33] 함수 $y = f(x)$의 그래프를 보면 분명하게 알 수 있습니다.

그림 2-33



$f'(a)=0$(접선의 기울기가 0, 즉 접선이 x축에 평행)이라도, ①, ②, ③일 때는 함수의 최솟값이 되지 않습니다.

참고로 경사하강법은 접선의 기울기가 낮은 쪽으로 계속 이동시켜서 최솟값을 구합니다. 그런데 자칫 함수 전체의 최솟값과 값이 커지거나 작아질 때 발생하는 극솟값/극댓값을 혼동할 수 있습니다. 경사하강법으로 최솟값을 구할 때 주의해야 하는 부분입니다.

최소값의 필요조건

그럼 이러한 혼동을 막기 위해 3차 함수 $f(x) = 3x^4-4x^3-12x^2+32$의 최솟값을 구하는 과정을 살펴보겠습니다. 도함수는 다음과 같습니다.

$$f'(x) = 12x^3-12x^2-24x = 12x(x+1)(x-2)$$

그리고 함수 $f(x)$와 도함수 $f'(x)$ 값의 증감표([표 2-1])[29]를 만들어 극솟값, 극댓값, 최솟값을 정확하게 확인해보겠습니다.

표 2-1[30]

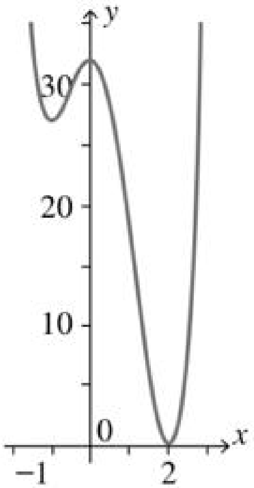
x	⋯	−1	⋯	0	⋯	2	⋯
$f'(x)$	−	0	+	0	−	0	+
$f(x)$	↘	27	↗	32	↘	0	↗
		극솟값		극댓값		최솟값	

최소값의 필요조건

앞 표를 확인하면 $x = 2$일 때 최솟값이 0이라는 사실을 정확하게 확인할 수 있습니다.

또한 증감표는 그래프를 그리는 데 도움을 받을 수 있습니다. [표 2-1]을 이용해 $f(x) = 3x^4 + 4x^3 - 12x^2 + 32$의 그래프를 그려보면 다음과 같습니다.

그림 2-34



최소값의 필요조건

이번에는 2차 함수 $f(x) = 2x^2-4x+3$의 최솟값을 구해보겠습니다. 도함수는 다음과 같습니다.

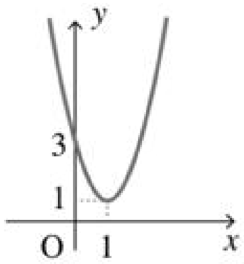
$$f'(x) = 4x-4$$

이를 이용해 [표 2-2]와 같은 증감표를 만들 수 있습니다. 즉 $x = 1$일 때 최솟값이 1입니다. 그래프는 [그림 2-35]와 같습니다.

표 2-2

x	…	1	…
$f'(x)$	−	0	+
$f(x)$	↘	1	↗
		최솟값	

그림 2-35

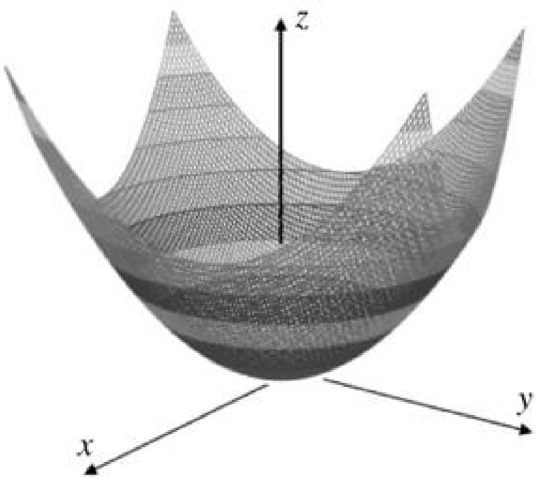


다변수 함수의 편미분

2장 06에서 미분을 설명할 때는 독립변수[31]가 하나인 함수를 다뤘습니다. 여기서는 독립변수가 2개 이상인 다변수 함수를 다루겠습니다.[32]

다변수 함수의 특징 중 하나는 시각화하기는 어렵다는 점입니다. 예를 들면 $z = x^2+y^2$과 같은 단순한 함수도 그래프는 다음처럼 3차원의 복잡한 모습으로 표현됩니다.

그림 2-36



$z = x^2+y^2$의 그래프

다변수 함수의 편미분

다변수 함수도 미분할 수 있습니다. 단, 변수가 여러 개 있으므로 어떤 변수를 미분할지 명시해야 합니다. 이렇게 특정 변수를 명시해 미분하는 것을 편미분partial derivative이라고 합니다.

예를 들어 두 변수 x, y가 있는 함수 $z=f(x, y)$를 생각해봅시다. 변수 x를 미분하고 y를 상수로 취급하는 것을 'x에 관한 편미분'이라고 하며 다음처럼 표현합니다.

$$\frac{\partial z}{\partial x}=\frac{\partial f(x, y)}{\partial x}=\lim_{\Delta x \to 0}\frac{f(x+\Delta x, y)-f(x, y)}{\Delta x}$$

'y에 관한 편미분'도 마찬가지입니다.

$$\frac{\partial z}{\partial y}=\frac{\partial f(x, y)}{\partial y}=\lim_{\Delta y \to 0}\frac{f(x, y+\Delta y)-f(x, y)}{\Delta y}$$

그럼 신경망에서 이용하는 편미분의 대표적인 예를 몇 가지 살펴보겠습니다. 첫 번째는 $z=wx+b$일 때 각 독립변수에 관한 편미분 결과입니다.

$$\frac{\partial z}{\partial x}=w,\ \frac{\partial z}{\partial w}=x,\ \frac{\partial z}{\partial b}=1$$

다변수 함수의 편미분

다음은 $f(x, y) = 3x^2+4y^2$일 때 각 독립변수에 관한 편미분 결과인 $\partial f(x, y)/\partial x$, $\partial f(x, y)/\partial y$를 구하겠습니다.

$$\frac{\partial f(x, y)}{\partial x} = 6x, \quad \frac{\partial f(x, y)}{\partial y} = 8y$$

이번에는 $z = w_1x_1+w_2x_2+b_1$일 때, x_1, w_2, b_1에 관해 각각 편미분하겠습니다.

$$\frac{\partial z}{\partial x_1} = w_1, \quad \frac{\partial z}{\partial w_2} = x_2, \quad \frac{\partial z}{\partial b_1} = 1$$

다변수 함수 최소값의 필요조건

미분 가능한 일변수 함수 $y = f(x)$의 어떤 x가 최솟값인 필요조건은 도함수가 0이 되는 것이었습니다(2장 06 참고). 다변수 함수도 마찬가지입니다. 예를 들어 이변수 함수는 [식 2-21] 같이 표현할 수 있습니다.

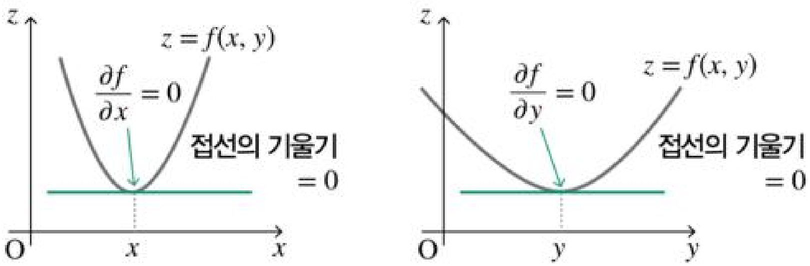
식 2-21

$$\frac{\partial f}{\partial x} = 0, \quad \frac{\partial f}{\partial y} = 0 \qquad \text{함수 } z = f(x, y)\text{가 최솟값이 되는 필요조건}$$

다변수 함수라면 [식 2-21]을 일반화해 n 변수의 경우로 확장하면 됩니다.

[식 2-21]의 개념은 [그림 2-37]을 보면 분명해집니다. 함수 $z = f(x, y)$가 최솟값이 되는 점 x 및 y의 그래프 방향을 보면 접선의 기울기가 0이기 때문입니다.

그림 2-37



[식 2-21]의 의미

2장 06에서 확인한 것처럼 [식 2-21]은 필요조건입니다. 즉, 함수 $f(x, y)$를 미분한 값이 0이라도 최솟값이 된다고 보장할 수 없습니다.

다변수 함수 최소값의 필요조건

그럼 함수 $z = x^2+y^2$가 최솟값일 때의 x, y의 값을 구해보겠습니다. 먼저 x, y에 관해 편미분하면 다음과 같습니다.

$$\frac{\partial z}{\partial x} = 2x, \quad \frac{\partial z}{\partial y} = 2y$$

그럼 [식 2-21]에서 함수가 최솟값이 될 필요조건은 $x = 0$, $y = 0$입니다. 이때 $z = x^2+ y^2 \geq 0$이므로, $z = 0$이 최솟값인 것을 알 수 있습니다([그림 2-36] 그래프에서 이를 확인합시다).