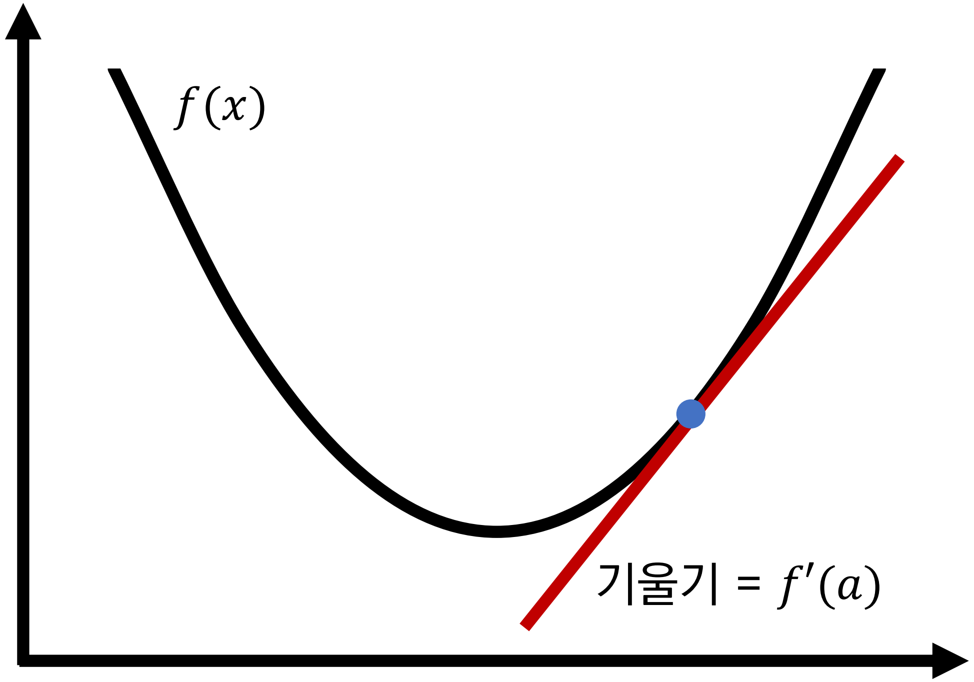
# 미분 (Ordinary derivative)

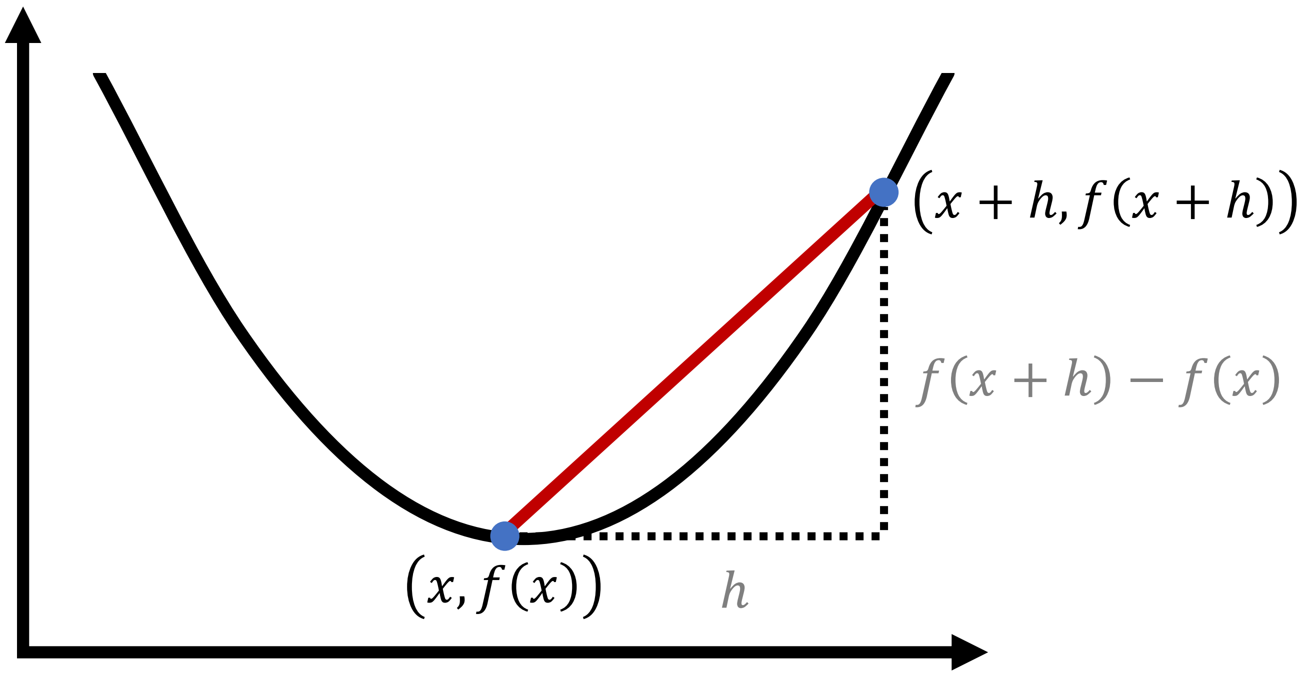
- 도함수(Derivative) $f'(x)$란?
  - 입력 $x$에 대하여, 함수 $f$의 기울기(gradient)를 알려주는 함수
  - 입력 $x$에 대하여, 함수 $f$가 얼마나 민감하게 변화하는지(순간 변화율)를 알려주는 함수
- 미분(Differentiation)이란? 도함수 $f'(x)$를 계산하는 일




따라서 어떠한 함수 $f(x)$가 있을 때, 특정한 점 $a$의 위치에서의 기울기(gradient) 혹은 순간 변화율 값을 구하고 싶다면 $f'(a)$를 계산하면 됩니다.

# 미분 (Ordinary derivative)

- 기울기의 정의는? '$f(x)$의 변화량/$x$의 변화량'
- 특정 함수에서 $x$가 $h$만큼 변할 때의 기울기를 계산해 봅시다.




- $x$가 $h$만큼 변할 때의 기울기는?

$$\frac{f(x+h)-f(x)}{h}$$

도함수란 $x$에서의 순간 변화율을 알려주는 함수이므로, $h=0$일 때의 값을 계산해야 합니다.

나동빈

# 미분 (Ordinary derivative)

- 따라서 도함수 $f'(x)$는 다음과 같이 계산할 수 있습니다.

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

- 미분은 인공지능 분야에서, 뉴럴 네트워크를 학습(training)을 시키기 위한 과정에서 사용됩니다.
- 뉴럴 네트워크의 파라미터를 기울기 값을 기준으로 학습을 시키는 방법을 주로 이용합니다.
- 실제로 $h$를 $0.0001$ 정도로 설정하여 근사할 수도 있으나 머신러닝 라이브러리에서는 실제로 도함수를 계산하여 학습을 진행하며, 레이어가 많으므로 연쇄법칙(chain-rule)을 이용하게 됩니다.

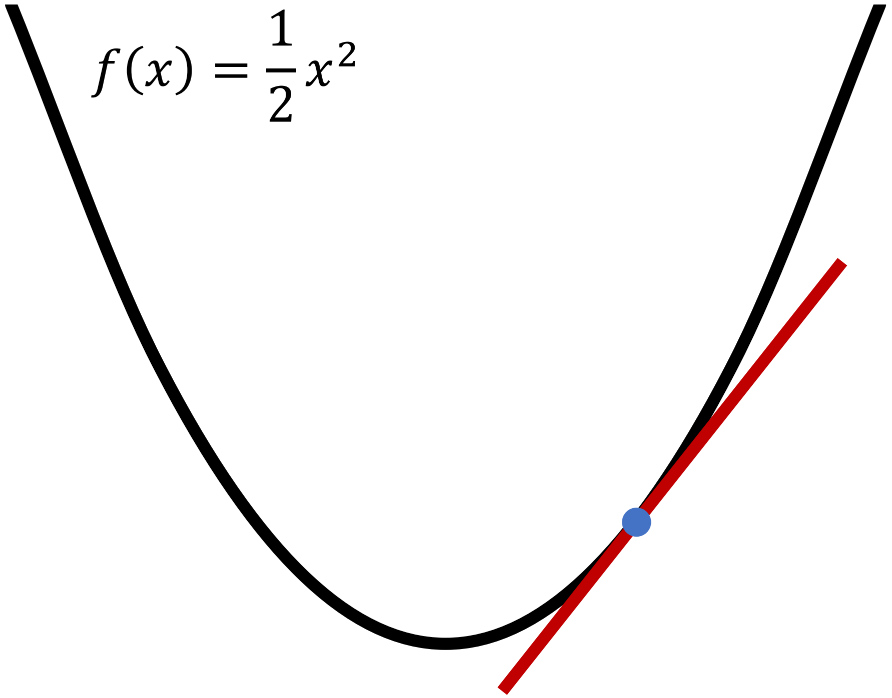
- 기본 미분 공식

$$(constant)' = 0 \qquad \left(ax^k\right)' = kax^{k-1}$$

Q. 다음을 계산하세요.

$(a)\ f(x) = \frac{1}{2}x^2$일 때 $f'(x)$

$(b)\ f(x) = \frac{1}{2}x^2$일 때 $x = 1$에서의 기울기(gradient)



# 주요 함수에 대한 미분

- 기본 미분 공식

$(constant)' = 0$ $\quad$ $\left(ax^k\right)' = kax^{k-1}$

- 지수 함수 미분 공식

$(e^x)' = e^x$

$(a^x)' = a^x \ln a$

$\left(\mathrm{e}^{f(x)}\right)' = \mathrm{e}^{f(x)} * f'(x)$

$\left(a^{f(x)}\right)' = a^{f(x)} * \ln a * f'(x)$

- 로그 함수 미분 공식

$(\ln x)' = \frac{1}{x}$

$(\log_a x)' = \frac{1}{x \ln a}$

$(\ln f(x))' = \frac{f'(x)}{f(x)}$

- 삼각 함수 미분 공식

$(\sin x)' = \cos x$

$(\cos x)' = -sin\, x$

$(\tan x)' = sec^2\, x$

# 편미분 (Partial derivative)

- 편미분이란? 다변수 함수(multivariate  function)에서, 하나의 변수를 기준으로 미분하는 것입니다.
- 미분을 할 때 다른 변수는 모두 상수(constant) 취급합니다.

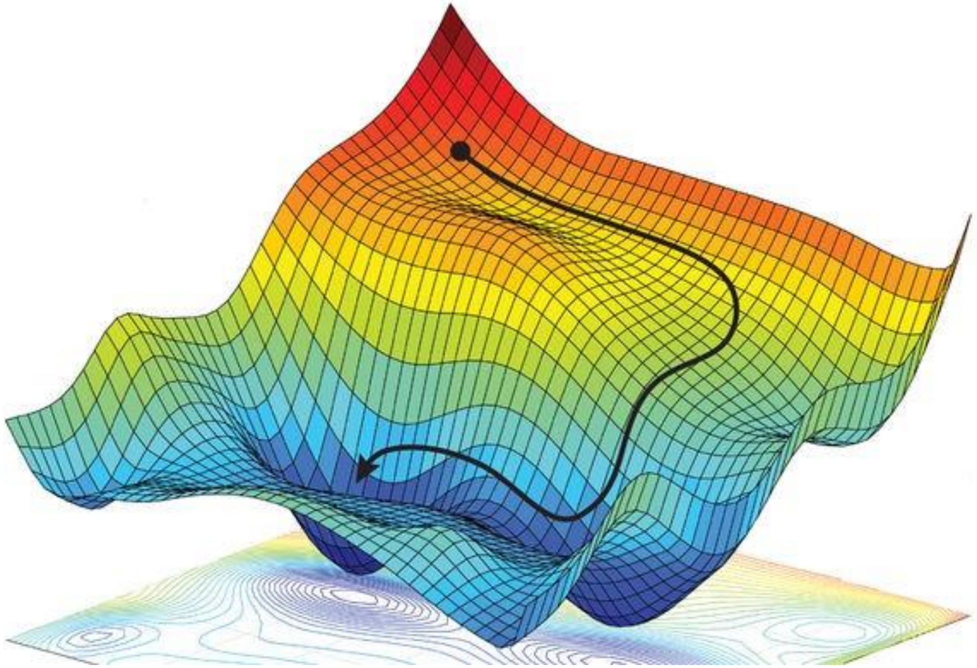
$$f(x, y) = 2x^2 + xy + 5y$$

$$\frac{\partial f(x, y)}{\partial x} = \frac{\partial(2x^2 + xy + 5y)}{\partial x} = 4x + y$$

$$\frac{\partial f(x, y)}{\partial y} = \frac{\partial(2x^2 + xy + 5y)}{\partial y} = 5$$

# 편미분 (Partial derivative)

- 실제로 머신러닝 모델은 입력(input)이나 가중치(weight) 값들이 다변수 벡터 형태입니다.
    - 따라서 머신러닝 모델에서 학습(training) 과정은 편미분을 통해 이루어집니다.



- 또한 머신러닝 모델은 다수의 레이어로 구성되어 있기 때문에 연쇄법칙(chain-rule)을 이용합니다.