Machine Learning & Deep Learning

Lecture 14

딥러닝을 위한 기초 수학

김희숙 (H.S.Kim)

- 딥러닝의 수학 원리를 이해하기 위해서는 당연히 기본적인 수학 지식이 필요함
- 어떤 원리로 입력 값의 패턴을 분석하고 학습하는지를 이해하려면 그 배경이 되는 수학 연산을 살펴봐야 하고, 여기에 사용되는 함수들을 알아야 하기 때문임

- 함수 : 두 집합 사이의 관계를 설명하는 수학 개념
  - ✓ 변수 x 와 y 가 있을 때, x 가 변하면 이에 따라 y 는 어떤 규칙으로 변하는지를 나타냄
  - ✓ 보통 함수를 나타낼 때는 function의 f 와 변수 x를 사용해 **y = f(x)** 라고 표시함

- 일차 함수 : y 가 x 에 관한 일차식으로 표현된 경우를 말함

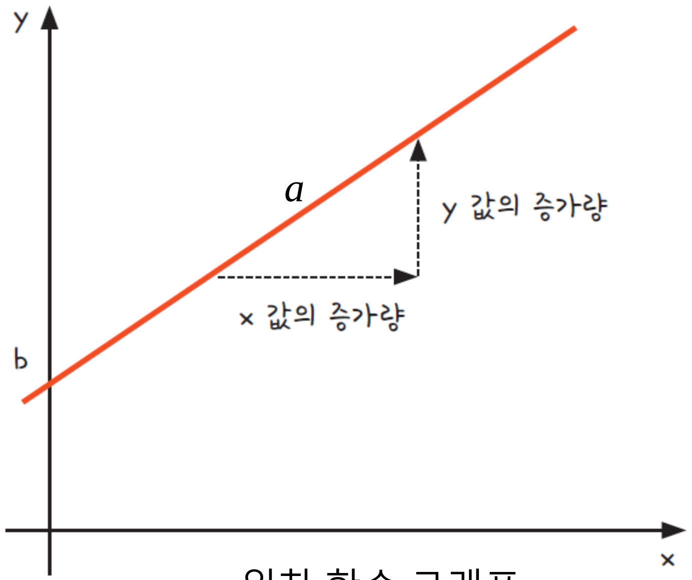
- X 가 일차인 형태이며, x 가 일차로 남으려면 a 는 0 이 아니어야 함

$$y = ax + b \ (a \neq 0)$$

- 일차 함수식 **$y = ax + b\ (a \neq 0)$** 에서 a 는 기울기, b 는 절편이라고 함

x 값이 증가할 때
y 값이 어느 정도 증가하는지에 따라
그래프의 기울기 a 가 정해진다.

절편은
그래프가
축과 만나는
지점



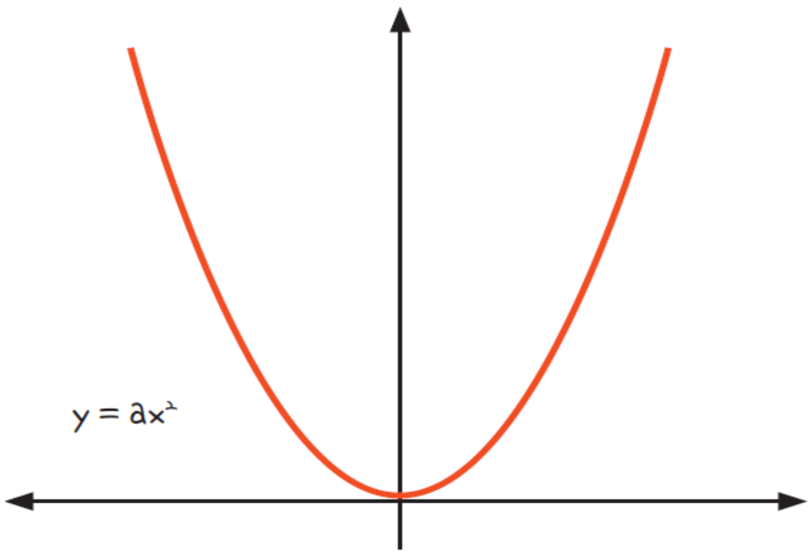
일차 함수 그래프

- 딥러닝의 수학 원리를 배울 때 초반부터 이 식이 등장함
- x 가 주어지고 원하는 y 값이 있을 때 적절한 a 와 b 를 찾는 것,
  이것이 바로 딥러닝을 설명하는 가장 간단한 표현임

- 이차 함수란 y 가 x 에 관한 이차식으로 표현되는 경우를 말함

$$y = ax^2 \quad (a \neq 0)$$

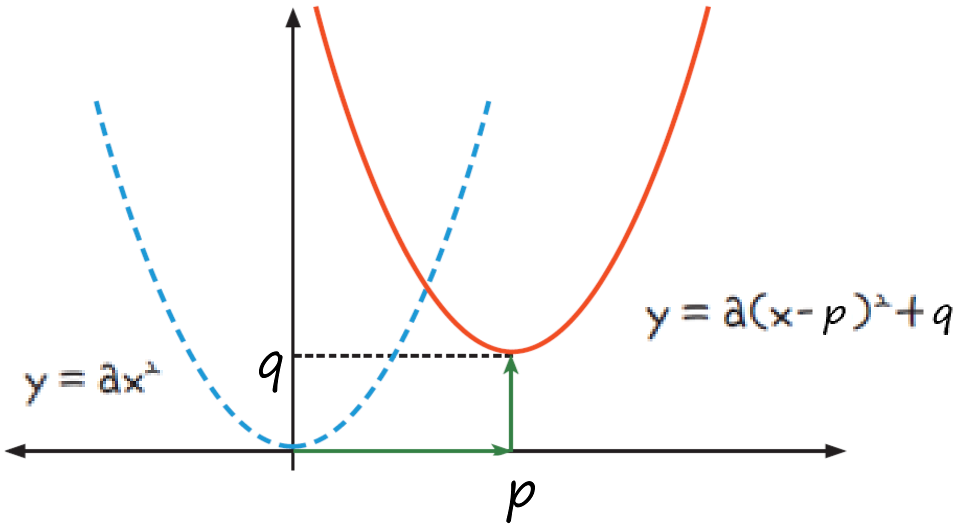
- a > 0 이면 아래로 볼록한 그래프가 됨



이차 함수 그래프

- 포물선의 맨 아래에 위치한 지점이 **최소값**이 되는데,

  딥러닝을 실행할 때는 이 **최소값**을 찾아내는 과정이 매우 중요함



이차 함수 그래의 평행이동과 **최소값**

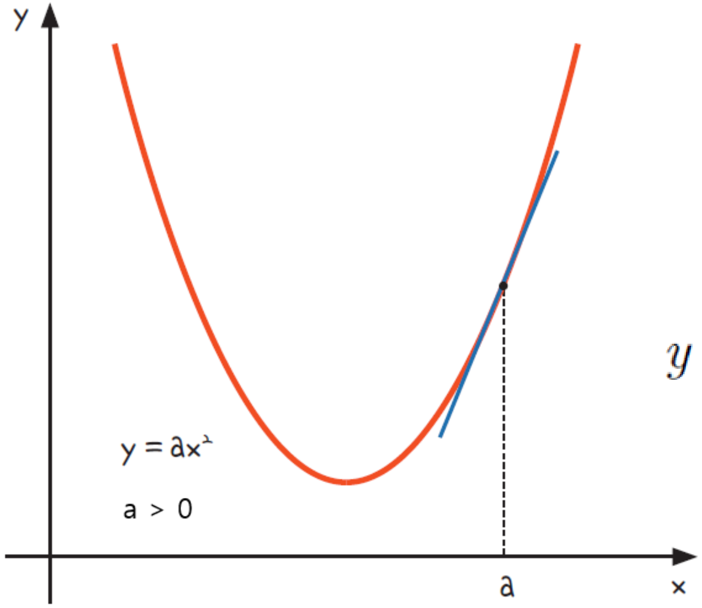
- 딥러닝에서 최소값은 **최소 제곱법(Method of Least Squares)** 공식으로 알아낼 수 있음

  하지만 **최소 제곱법**을 계산하기 위해 **꼭 필요한 조건들을 모두 알 수 없기 때문**에 **미분**과 **기울기**를 이용해야 함

3. 미분 & 순간 변화율과 기울기

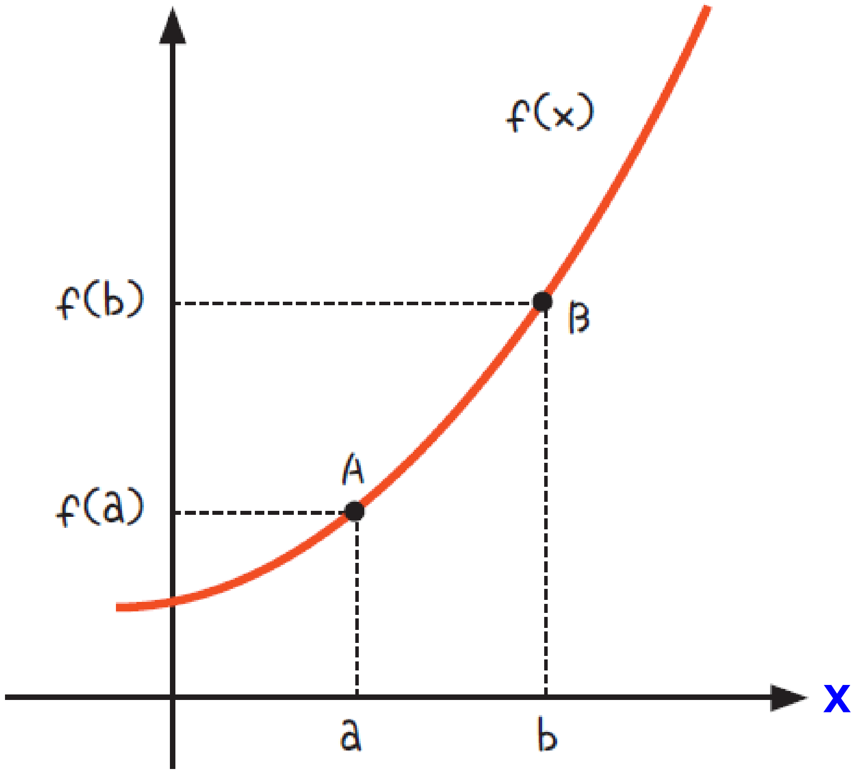
- 딥러닝을 이해하는 데 가장 중요한 수학 원리는 **미분**이라고 할 수 있음
- 너무 **미세**해서 실제로 움직이는 게 아니라 **방향만 드러내는 정도의 순간적인 변화**만 있을 것임
- 이 순간의 변화를 놓고 **순간 변화율**이라는 이름을 붙임
- 순간 변화율은 어느 쪽을 향하는 **방향성**을 지니고 있으므로,
  이 **방향을 따라** 직선을 길게 그려주면 그래프와 맞닿는 **접선**이 그려짐
- 이 선이 바로 이 점에서의 **기울기**가 됨

- **미분**을 한다는 것은 쉽게 말해 이 **'순간 변화율'을 구한다**는 것임
- **미분 계수** : 어느 순간에 어떤 변화가 일어나고 있는지를 **숫자**로 나타낸 것
- 이 **미분 계수**는 곧 그래프에서의 **기울기**를 의미함

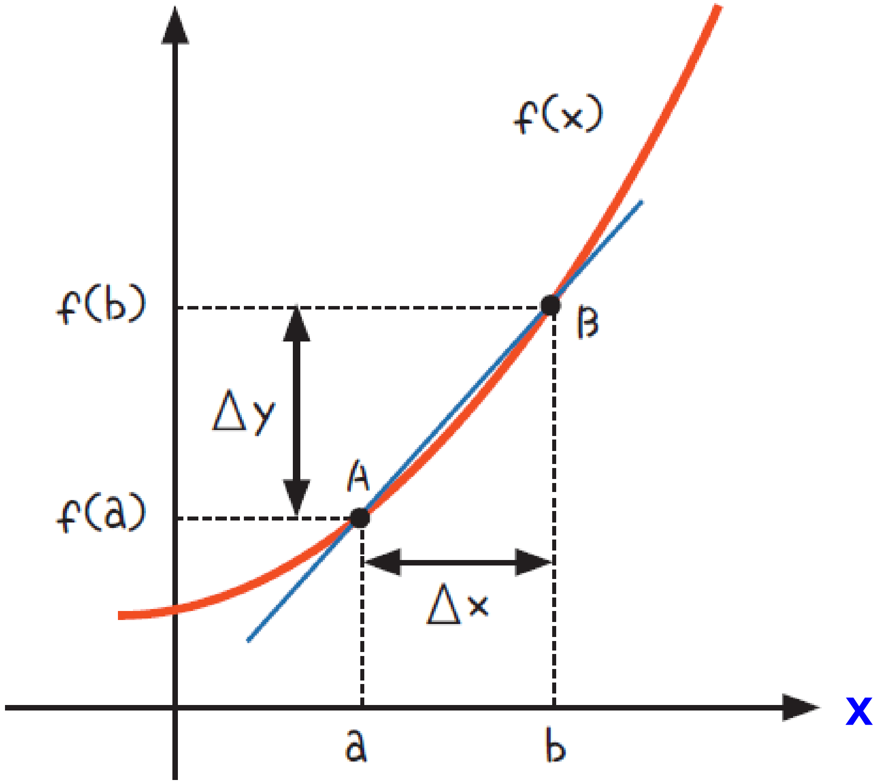


$$y = ax^2 \, (a \neq 0)$$

a 에서의 순간 변화율은 곧 기울기다!



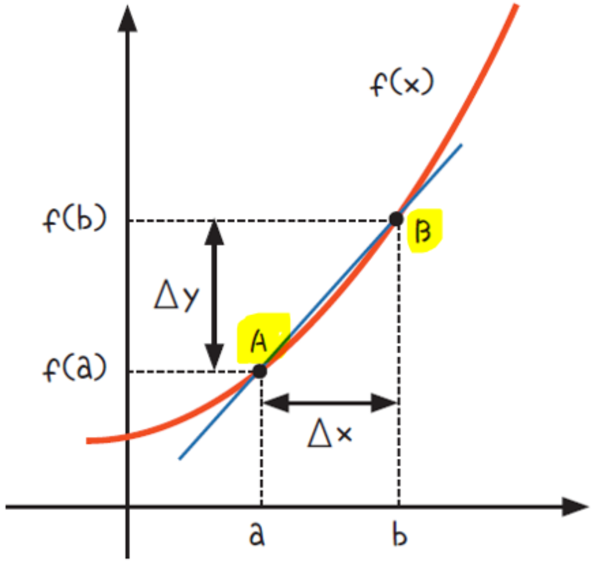
함수 f(x)의 **x축** 위에 두 **실수 a와 b**를 대입



A, B를 지나는 직선은 이 두 점 간의 기울기,
곧 **평균 변화율**을 의미

여기서 Δ(델타)는 변화량을 나타내는 기호임

- 이 그래프에서 x 값의 증가량은 b-a 이고,
 y 값의 증가량은 f(b) – f(a) 이다.



$$\text{직선 AB의 기울기} = \frac{y\text{값의 증가량}}{x\text{값의 증가량}} = \frac{f(b)-f(a)}{b-a} = \frac{f(a+\Delta x)-f(a)}{\Delta x}$$

- 이때 직선 AB의 기울기를 A와 B 사이의
 '평균 변화율'이라고도 부름
- 미분을 배우고 있는 우리에게 필요한 것은 순간 변화율임
- 순간 변화율 : x 의 증가량 (Δx 0에 가까울 만큼 아주 작을 때의 순간적인 기울기를 말함

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

- $\lim_{\Delta x \to 0}$ **x 의 증가량이 0 에 가까울 만큼 작을 때**라는 뜻
- 기울기는 $\dfrac{y\text{값의 증가량}}{x\text{값의 증가량}}$ 므로 순간 기울기는 $\lim_{\Delta x \to 0} \dfrac{y\text{값의 증가량}}{x\text{값의 증가량}}$ 이 되며,

  이것은 $\lim_{\Delta x \to 0} \dfrac{\Delta y}{\Delta x} = \dfrac{dy}{dx}$ 고도 쓸 수 있음
- **함수** $f(x)$**를 미분하라**는 것을 $f'(x)$ 또는 $\dfrac{d}{dx}f(x)$ 라함

1 함수 f(x)를 x로 미분하라는 것은

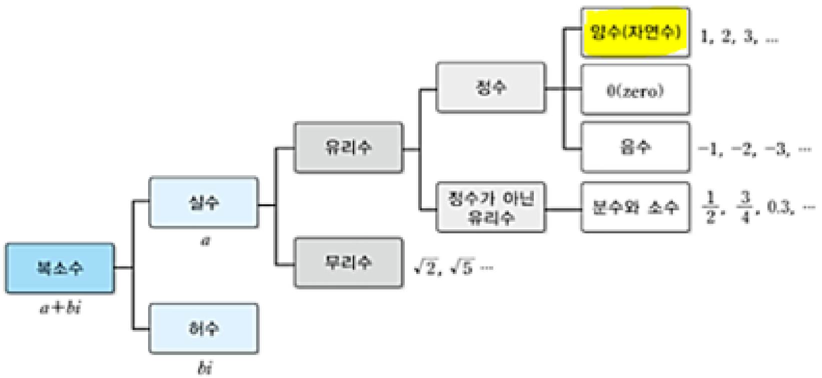
2 x의 변화량이 0에 가까울 만큼 작을 때

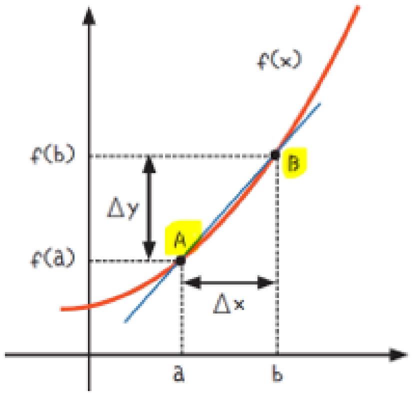
3 y의 변화량을

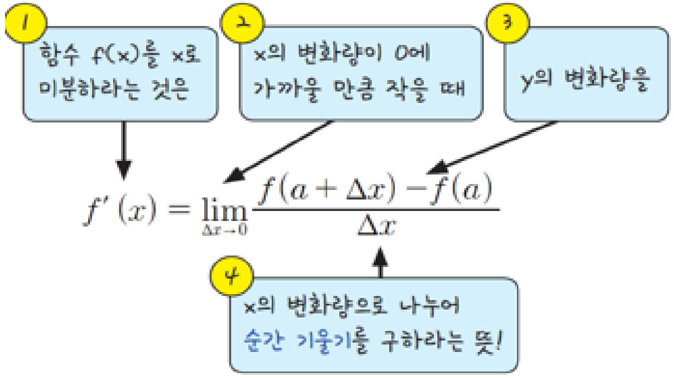
$$f'(x) = \lim_{\Delta x \to 0} \frac{f(a+\Delta x) - f(a)}{\Delta x}$$

4 x의 변화량으로 나누어 순간 기울기를 구하라는 뜻!

- 다음은 딥러닝을 공부하는 과정 중에 자주 만나게 되는 중요한 4가지







미분의 성질

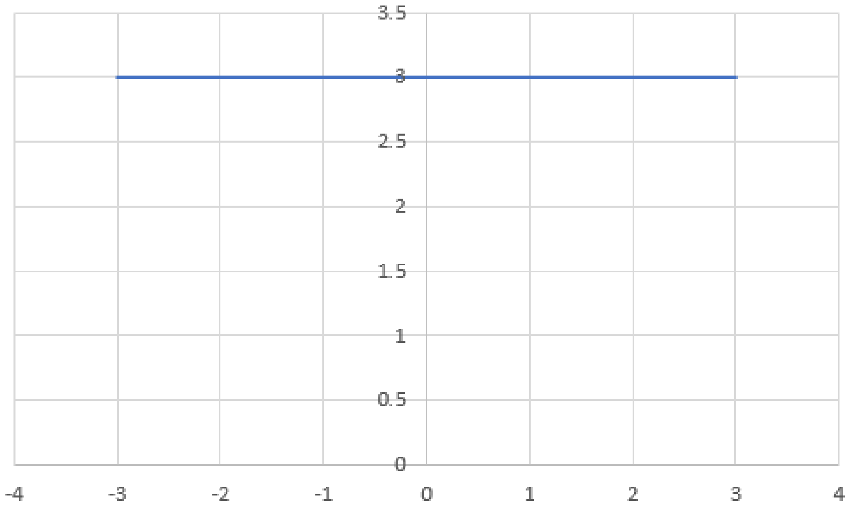
성질 1. $f(x) = a$ 에서 a 가 상수일 때 미분 값은 0임

성질 2. $f(x) = x$ 일 때의 미분 값은 1임

성질 3. $f(x) = ax$ 에서 a가 상수이면 미분 값은 a 임

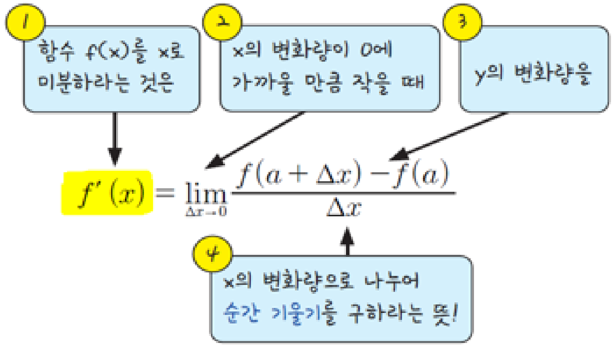
성질 4. $f(x) = x^a$에서 a가 자연수이면 미분 값은 ax^{a-1}이 됨

성질 1. $f(x) = a$ 에서 a 가 상수일 때 미분 값은 0임



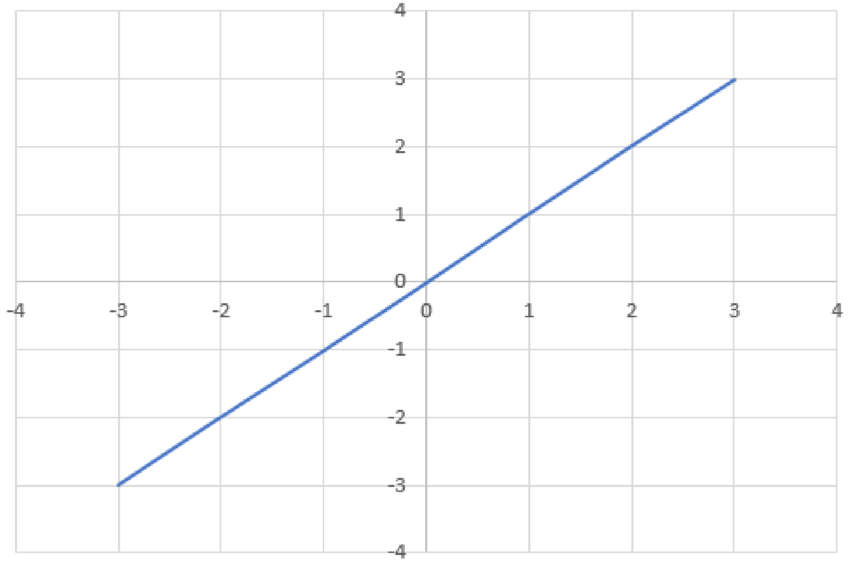
a 가 3 이면

x	f(x)	미분 값
-3	3	0
-2	3	0
-1	3	0
0	3	0
1	3	0
2	3	0
3	3	0



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

성질 2. $f(x) = x$ 일 때의 미분 값은 1임



x	f(x)	미분 값
-3	-3	1
-2	-2	1
-1	-1	1
0	0	1
1	1	1
2	2	1
3	3	1

1 함수 f(x)를 x로 미분하라는 것은

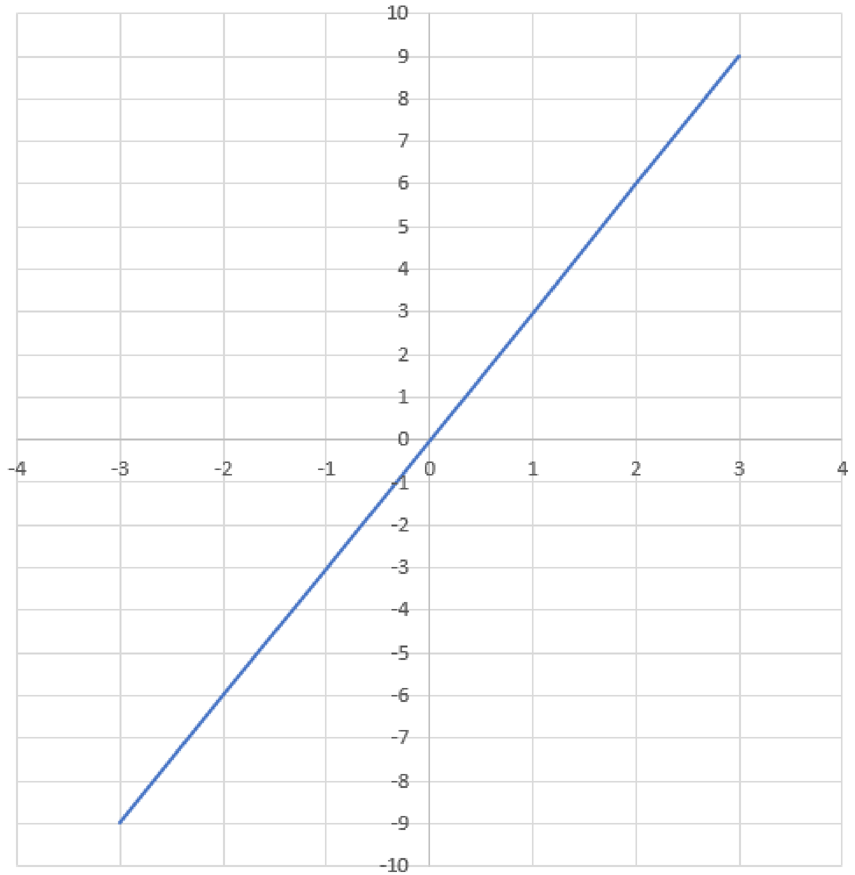
2 x의 변화량이 0에 가까울 만큼 작을 때

3 y의 변화량을

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

4 x의 변화량으로 나누어
순간 기울기를 구하라는 뜻!

성질 3. $f(x) = ax$ 에서 a가 상수이면 미분 값은 a임



a가 3이면

x	f(x)	미분 값
-3	-9	3
-2	-6	3
-1	-3	3
0	0	3
1	3	3
2	6	3
3	9	3

① 함수 f(x)를 x로 미분하라는 것은

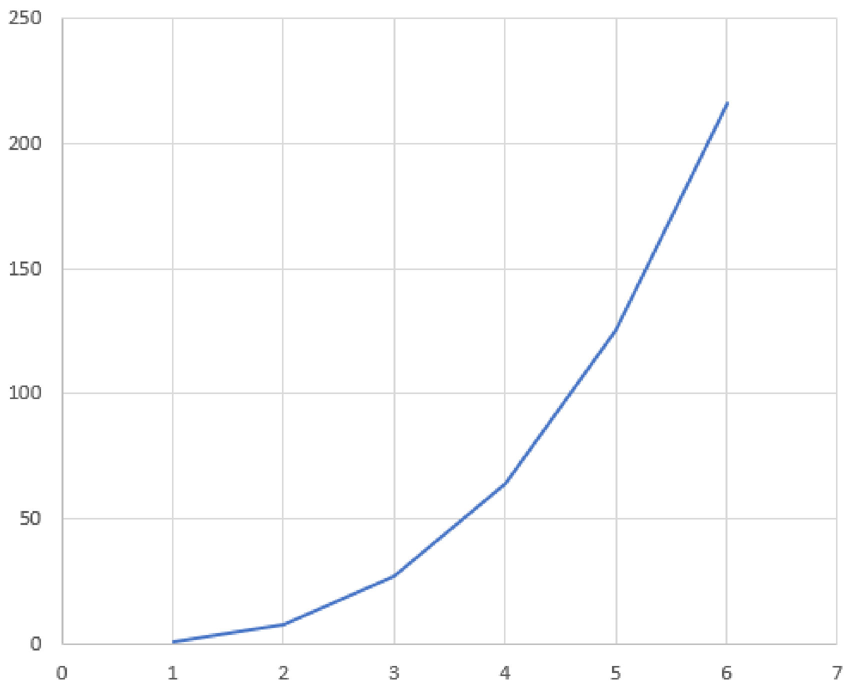
② x의 변화량이 0에 가까울 만큼 작을 때

③ y의 변화량을

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

④ x의 변화량으로 나누어 순간 기울기를 구하라는 뜻!

성질 4. $f(x) = x^a$에서 a가 자연수이면 미분 값은 ax^{a-1}이 됨



a 가 3 이면

미분값은 $3x^2$

x	f(x)	미분 값
1	1	9
2	8	36
3	27	81
4	64	144
5	125	225
6	216	324

- 미분과 더불어 딥러닝을 공부할 때 가장 자주 접하게 되는 또 다른 수학 개념은 바로 **편미분**임
- 미분과 편미분 모두 '미분하라'는 의미에서는 다를 바가 없음

$$\boxed{f(x, y) = x^2 + yx + a \text{ (a는 상수)}}$$

- **여러 가지 변수가 식 안에 있을 때,** 모든 변수를 미분하는 것이 아니라
  우리가 **원하는 한 가지 변수만 미분**함**, 그 외에는 모두 상수**로 취급하는 것이 바로 편미분임

- 변수가 x 와 y 중 어떤 변수로 미분해야 하는지를 정해야 하므로 **편미분**을 사용하는 것
- 만일 이 식처럼 **여러 변수 중에서 x 에 관해서만 미분**하고 싶다면,
  함수 f 를 x 에 관해 편미분하라고 함

$$\boxed{\frac{\partial f}{\partial x}}$$

$f(x, y) = x^2 + yx + a$ (a는 상수)

$\frac{\partial f}{\partial x}$

성질 4. $f(x) = x^a$에서 a가 자연수이면 미분 값은 ax^{a-1}이 됨

- 함수 $f(x, y) = x^2 + yx + a$를 x 에 관해 **편미분** 하는 과정은 미분의 성질 4번에 따라 x^2 항은 $2x$ 가 됨

성질 3. $f(x) = ax$ 에서 a가 상수이면 미분 값은 a임

- x 에 관해 미분하면 **다른 모든 항은 상수**로 취급하므로 y 는 **상수** 가 됨
  미분의 성질 3번에 따라 yx 는 y 가 됨

성질 1. $f(x) = a$ 에서 a 가 상수일 때 미분 값은 0임

- 마지막 항 a 는 미분의 성질 1번에 따라 0 이 됨

함수 f를 x에 관해 편미분

$f(x, y) = x^2 + yx + a$ (a는 상수) ➡ $\frac{\partial f}{\partial x} = 2x + y$

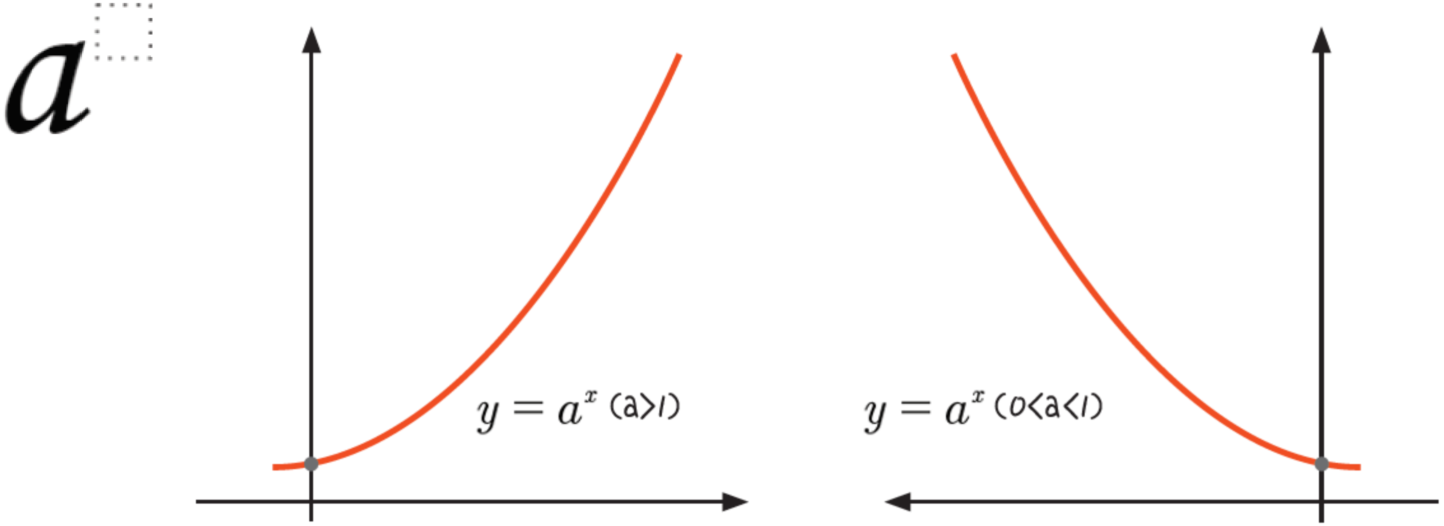
- 지수란 다음과 같은 형태를 말함

$$a^{\square}$$

- 여기서 a 를 **'밑'**이라 하고 $\square$ '지수'라고 부름
- a 를 $\square$만큼 반복해서 곱한다는 뜻
- **지수 함수(exponential function)** : 변수 x가 지수 자리에 있는 경우를 말함

$$\boxed{y = a^x (a \neq 1,\ a > 0)}$$

- 지수 함수에서는 밑(a)의 값이 무엇인지가 중요함
- 이 값(밑)이 **1이면 함수가 아님**
- 이 값(밑)이 **0보다 작으면** 허수를 포함하게 되므로 **안 됨**
- 이 값(밑)은 **a>1** 이거나 **0<a<1**, 둘 중 하나가 되어야 함



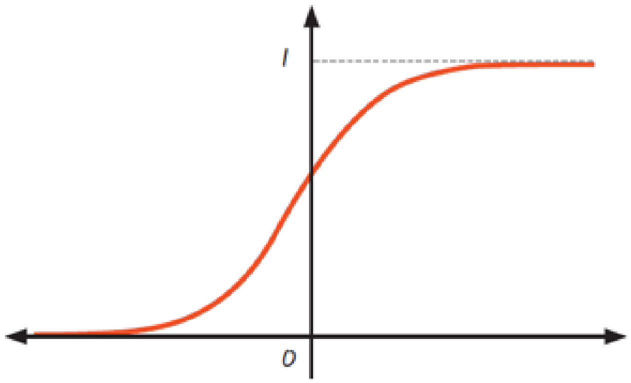
a>1 인 경우와 **0<a<1** 인 경우의 **지수 함수**

$a^{\square}$

- **시그모이드 함수** : **지수 함수에서 밑**의 값이 **자연 상수** e 인 함수를 말함
- **자연 상수** e : **'자연 로그의 밑', '오일러의 수'** 등 여러 이름으로 불림
  - ✓ 파이(π)처럼 수학에서 중요하게 사용되는 무리수임
  - ✓ 그 값은 대략 **2.718281828...** 근사값으로 표현할 수 있음
- **자연 상수** e 가 **지수 함수**의 **밑**에 **포함되어 분모**에 들어가면 **시그모이드 함수**가 됨

$$\boxed{f(x) = \frac{1}{1+e^{-x}}}$$

- 시그모이드 함수를 그래프로 그려보면 **S** 자 형태로 나타남



시그모이드 함수의 그래프

- 로그를 이해하려면 먼저 지수부터 이해해야 함

$$a^x = b$$

- x 를 구하기 위해 사용하는 방법이 log 이다.
- 영어로 **logarithm**이라고 하는데 앞 세 글자 **log**를 씀
- 지수 식에서 a와 b의 위치를 다음과 같이 바꾸어 써주면 됨

$$a^x = b$$

$$\log_a b = x$$

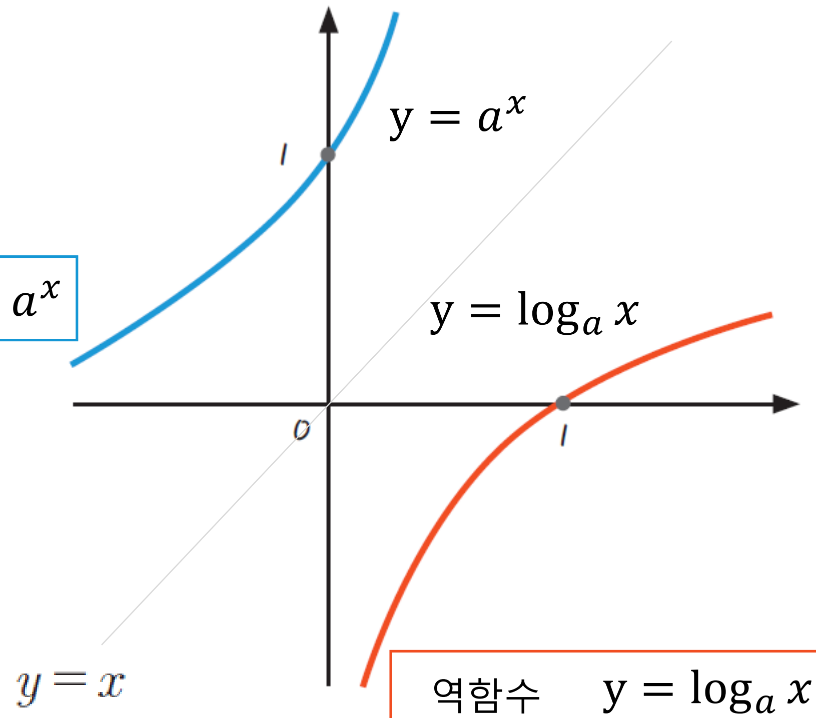
- 로그가 지수와 이렇게 밀접한 관계가 있듯, **로그 함수** 역시 **지수 함수**와 밀접한 관계에 있는데 바로 **역함수**의 관계임
- 역함수는 x 와 y 를 서로 바꾸어 가지는 함수임

$$a^x = b$$
$$\log_a b = x$$

- 지수 함수 $\boxed{y = a^x (a \neq 1,\ a > 0)}$ 그의 정의를 따라 $\boxed{x = \log_a y}$ 있음

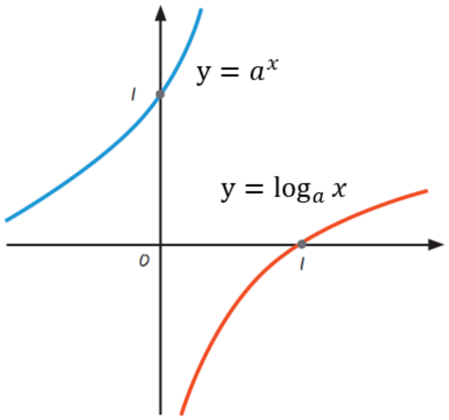
- **역함수를 만들기 위해 x와 y를 서로 바꾸어 주면 됨** $\boxed{y = \log_a x}$

- 지수 함수 $\boxed{y = a^x}$ 그래프를 $y = x$|칭으로 이동시킨 로그 함수 $\boxed{y = \log_a x}$
- **역함수의 그래프**는 $y = x$| 대하여 대칭인 선으로 나타남

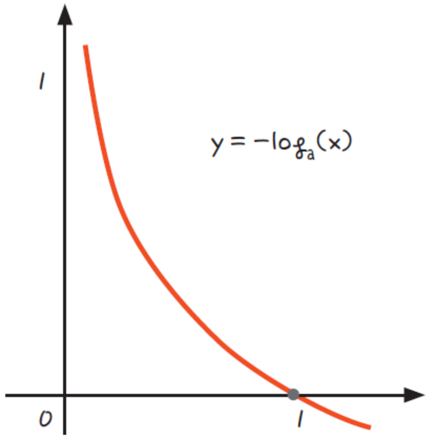


지수 함수 $y = a^x$

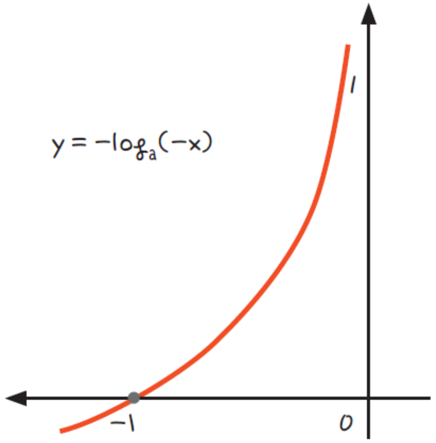
역함수 $y = \log_a x$



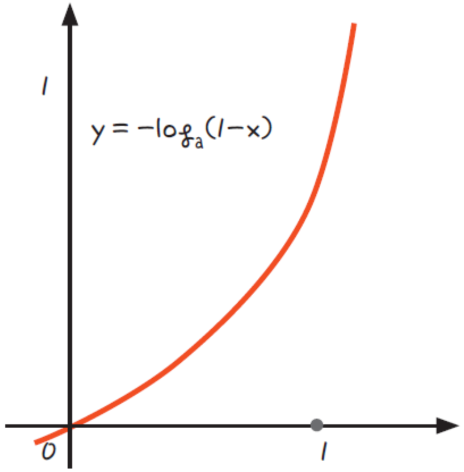
1 x축에 대하여 대칭 이동



2 x축과 y축에 대하여 대칭 이동



3 2번 그래프를 x축 오른쪽 방향으로 1만큼 평행 이동



4 1번과 3번을 함께 나타낸 그래프