첫째마당
딥러닝 시작을 위한 준비 운동

3장 딥러닝을 위한 기초 수학

1 일차 함수, 기울기와 y 절편

2 이차 함수와 최솟값

3 미분, 순간 변화율과 기울기

4 편미분

5 지수와 지수 함수

6 시그모이드 함수

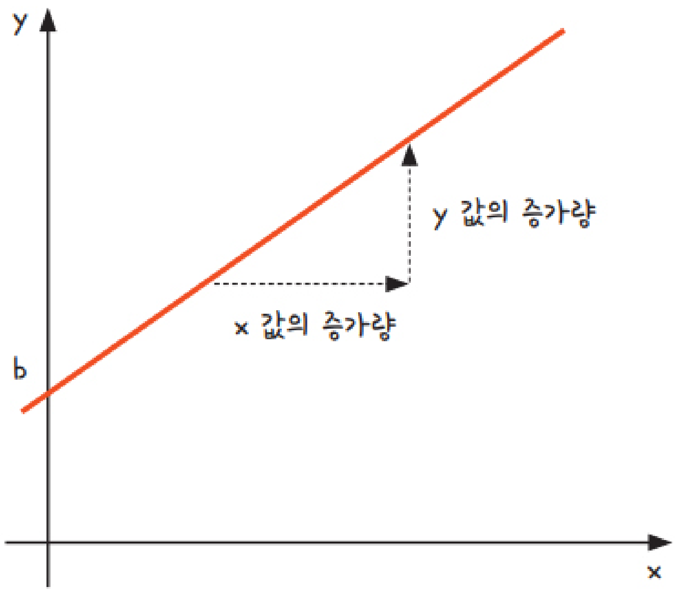
7 로그와 로그 함수

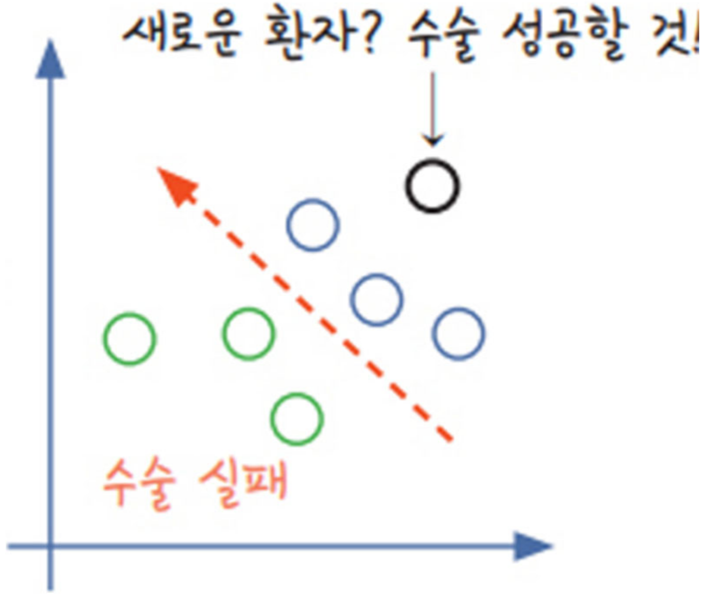
1 일차 함수, 기울기와 y 절편

1 일차 함수, 기울기와 y 절편

- **일차 함수, 기울기와 y 절편**
  - 딥러닝을 이야기 할 때는 1차 함수를 이야기하게 됨.

    $y = ax + b \ (a \neq 0)$
  - X가 일차인 형태이며 x가 일차로 남으려면 a는 0이 아니어야 함
  - 아래와 같은 그래프가 1차함수 그래프가 됨.

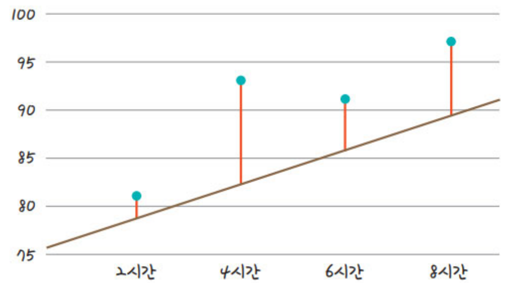




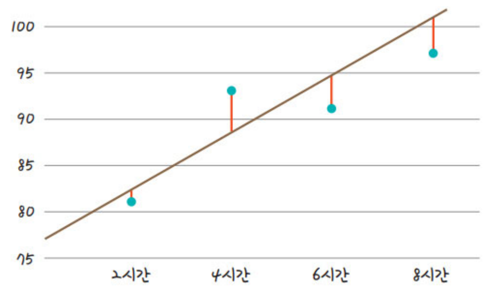
  - 일차 함수로 만들어진 선형 직선이 기준선이 되어 어디에 위치 하냐에 따라서 새로운 데이터에 대해 예측을 할 수 있다. 즉 선을 그려주는 것이 중요 함.

1 일차 함수, 기울기와 y 절편

- **일차 함수, 기울기와 y 절편**
  - 일차 함수식 y = ax + b에서 a는 **기울기**, b는 **절편**이라고 함
  - 기울기는 기울어진 정도를 의미하는데 x 값이 증가할 때 y 값이 어느 정도 증가하는지에 따라 그래프의 기울기 a가 정해짐
  - 절편은 그래프가 축과 만나는 지점을 의미
  - x가 주어지고 원하는 y 값이 있을 때 적절한 a와 b를 찾는 것, 이것이 바로 딥러닝을 설명하는 가장 간단한 표현이다.



  - 선과 데이터 사이를 계산해서 계산된 값을 오차라고 한다.
  - 그러므로 선과 데이터 사이의 값이 작은 최적의 선을 찾는 것이 딥러닝이다.



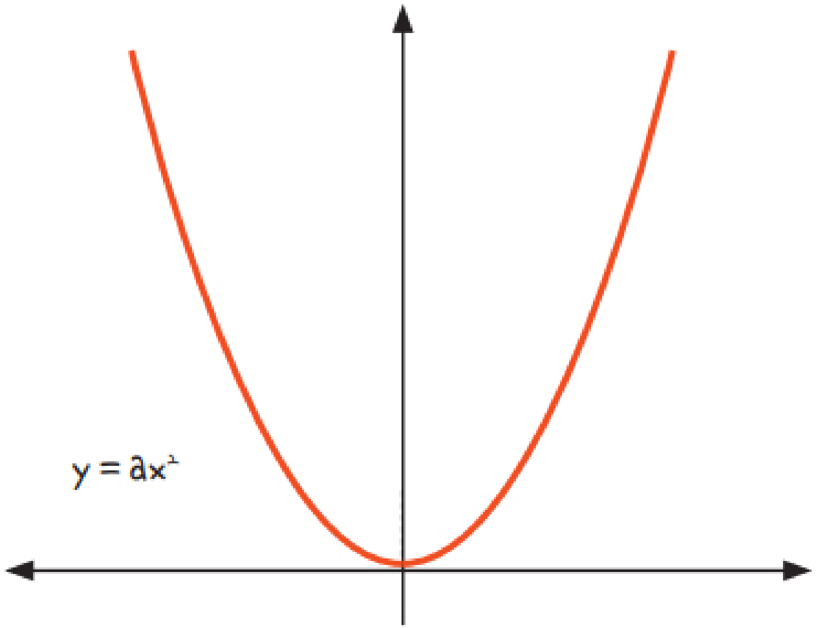
  - 이렇게 최적의 선을 구하기 위해서 필요한 것이 2차 함수임.

2 이차 함수와 최솟값

2 이차 함수와 최솟값

- **이차 함수와 최솟값**
  - **이차 함수**란 y가 x에 관한 이차식으로 표현되는 경우를 의미
  - 다음과 같은 함수식으로 표현할 수 있음

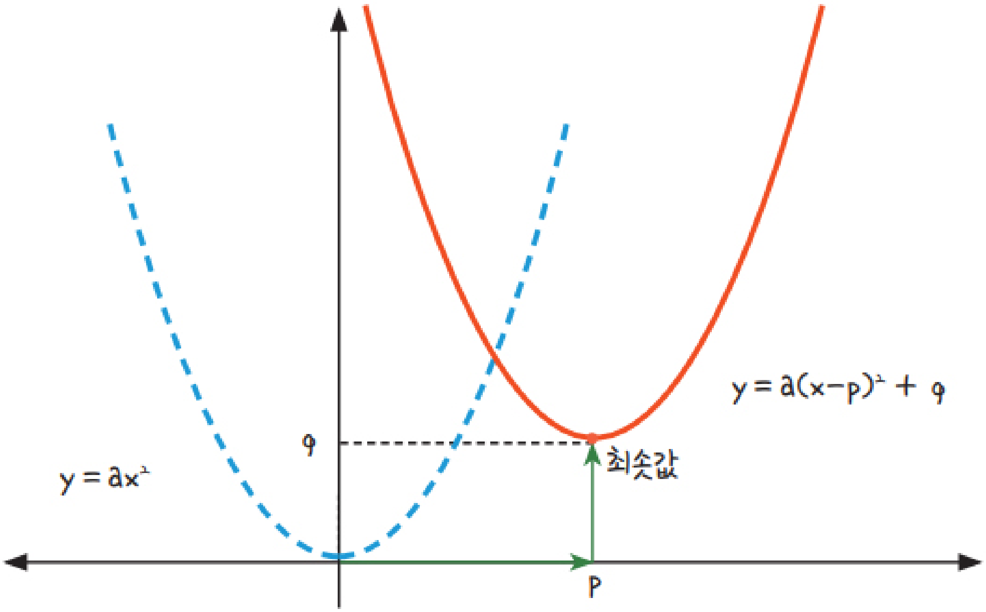
    $y = ax^2 \, (a \neq 0)$
  - a > 0이면 아래로 볼록한 그래프가 됨



  - 1차함수의 최적의 그래프는 2차 함수의 꼭지점이 된다.

2 이차 함수와 최솟값

- **이차 함수와 최솟값**
  - $y = ax^2$의 그래프를 x축 방향으로 p만큼, y축 방향으로 q만큼 평행 이동시키면 점 p와 q를 꼭짓점으로 하는 포물선이 됨
  - 이때 포물선의 맨 아래에 위치한 지점이 **최솟값**이 되는데, 딥러닝을 실행할 때는 이 최솟값을 찾아내는 과정이 매우 중요함

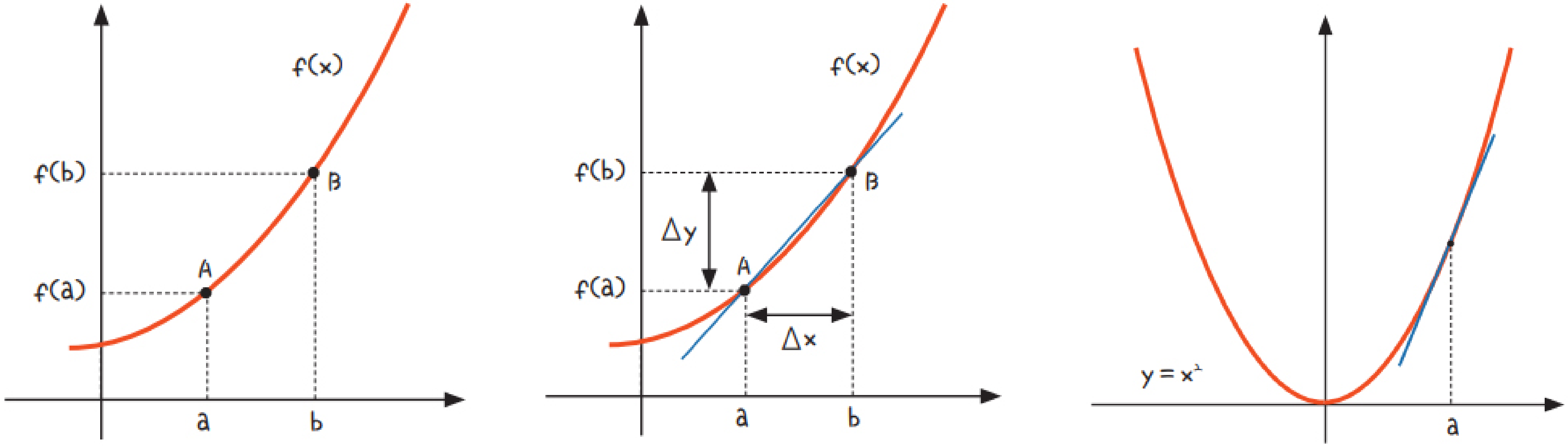


  - 만약 p값과 q값을 알지 못할 때 최소값을 구하기 위해서는 미분이 필요한 것이다.
  - 일차함수에 오차가 없는 선을 구하는 것이 중요한데 이때 최소값이 일차함수의 최적의 기울기가 된다

3 미분, 순간 변화율과 기울기

3 미분, 순간 변화율과 기울기

- **미분, 순간 변화율과 기울기**
  - 딥러닝은 결국 일차 함수의 a와 b 값을 구하는 것인데, a와 b 값은 이차 함수 포물선의 최솟값을 구하는 것
  - 이 최솟값을 미분으로 구하기 때문에 미분이 딥러닝에서 중요한 것

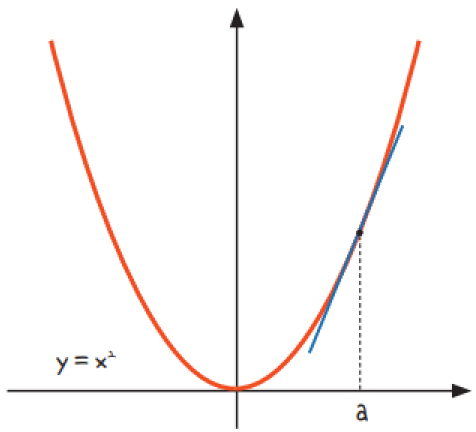


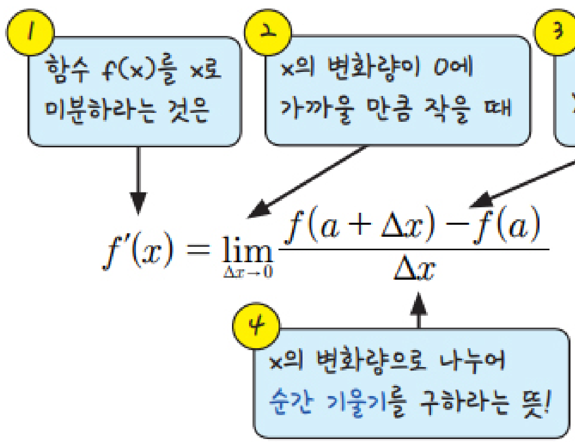
  - 기울기란 A점과 B점을 연결하는 것이 기울기 이다.
  - 기울기를 정하는 것은 y축의 증가값에 x의 증가값을 나눠주는 것이다.
  - Δ를 써서 표현하면 x 값의 증가량은 Δx로, y 값의 증가량은 f(a + Δx) - f(a)로 나타낼 수 있음

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

3 미분, 순간 변화율과 기울기

- **미분, 순간 변화율과 기울기**
  - a가 미세하게 '0에 가까울 만큼' 움직였이면 y 값 역시 매우 미세하게 변화를 할 텐데, 너무 미세해서 순간적인 변화만 있음
  - 이 순간의 변화를 놓고 **순간 변화율**이라는 이름을 붙였음
  - 순간 변화율은 어느 쪽을 향하는 방향성을 지니고 있으므로, 이 방향을 따라 직선을 길게 그려 주면 그래프와 맞닿는 접선이 그려짐
  - 이 선이 바로 이 점에서의 **기울기**가 됨
  - 미분을 한다는 것은 쉽게 말해 이 '순간 변화율'을 구한다는 것
  - 순간 변화율은 x의 증가량(Δx)이 0에 가까울 만큼 아주 작을 때의 순간적인 기울기를 의미하므로, 극한(limit) 기호를 사용해 다음과 같이 나타냄
  - 최소값을 구한다는 것은 미분값의 기울기가 0인 곳을 찾는 것이다.





3 미분, 순간 변화율과 기울기

미분, 순간 변화율과 기울기

- 다음은 딥러닝을 공부하는 과정 중에 자주 만나게 되는 중요한 다섯 가지 미분의 기본 공식

미분의 기본 공식

1 | $f(x) = x$일 때 $f'(x) = 1$

2 | $f(x) = a$에서 a가 상수일 때 $f'(x) = 0$

3 | $f(x) = ax$에서 a가 상수일 때 $f'(x) = a$

4 | $f(x) = x^a$에서 a가 자연수일 때 $f'(x) = ax^{a-1}$

5 | $f(g(x))$에서 $f(x)$와 $g(x)$가 미분 가능할 때 $\{f(g(x))\}' = f'(g(x)) \times g'(x)$

4 편미분

4 편미분

- **편미분**
  - 여러 가지 변수가 식 안에 있을 때, 모든 변수를 미분하는 것이 아니라 우리가 원하는 한 가지 변수만 미분하고 그 외에는 모두 상수로 취급하는 것이 바로 편미분
  - 예를 들어 f(x) = x와 같은 식을 미분할 때는 변수가 x 하나뿐이어서 미분하라는 의미에 혼란이 없음

$f(x,\ y) = x^2 + yx + a$ (a는 상수)

  - 여기에는 변수가 x와 y, 이렇게 두 개 있음
  - 두개의 변수 중 하나의 변수만 미분하는 것을 편미분이라 함
  - 만일 이 식처럼 여러 변수 중에서 x에 관해서만 미분하고 싶다면, 함수 f를 ‘x에 관해 편미분하라’고 하며 다음과 같이 식을 씀

$f(x,\ y) = x^2 + yx + a$일 때

$$\frac{\partial f}{\partial x} = 2x + y$$

5 지수와 지수 함수

5 지수와 지수 함수

지수와 지수 함수

$$a^{\square}$$

- 여기서 a를 '밑'이라 하고 $\square$ 를 '지수'라고 함
- a를 $\square$ 만큼 반복해서 곱한다는 뜻
- **지수 함수**란 변수 x가 지수 자리에 있는 경우를 의미
- 식으로 나타내면 다음과 같은 형태

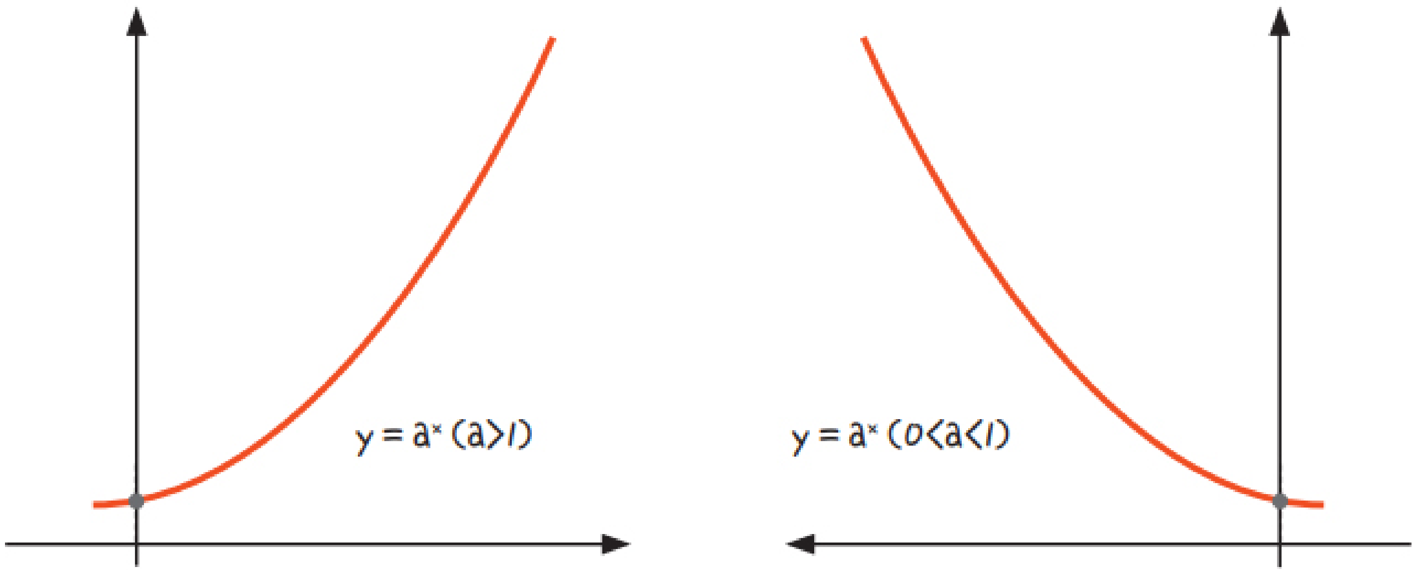
$$y = a^x \ (a \neq 1,\ a > 0)$$

- 지수 함수에서는 밑(a) 값이 무엇인지가 중요하다.
- 이 값이 1이면 함수가 아니다. 1에 x승은 무조건 1이므로 함수라 할 수 없다.
- 또 0보다 작으면 허수를 포함하게 되므로 안 된다.
- 밑의 값은 a > 1이거나 0 < a < 1, 둘 중 하나가 되어야 한다.

5 지수와 지수 함수

지수와 지수 함수

- 밑의 값은 a > 1이면 왼쪽 그래프, 0 < a < 1이면 오른쪽 그래프

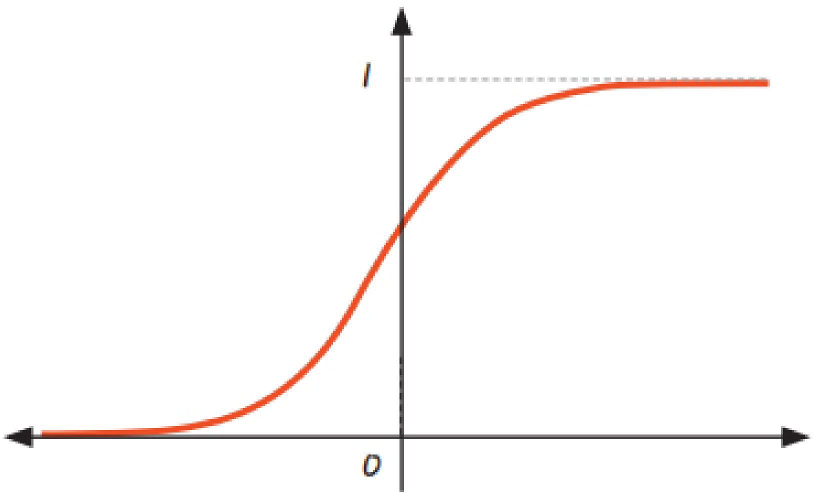


6 시그모이드 함수

6 시그모이드 함수

- **시그모이드 함수**
  - 자연 상수 e는 '자연 로그의 밑', '오일러의 수' 등 여러 이름으로 불리는데, 파이(π)처럼 수학에서 중요하게 사용되는 무리수이며 그 값은 대략 2.718281828...
  - x가 큰 값을 가지면 f(x)는 1에 가까워지고, x가 작은 값을 가지면 f(x)는 0에 가까워지는 그래프를 말한다.
  - S자 형태로 그려지는 이 함수의 속성은 0 또는 1, 두 개의 값 중 하나를 고를 때 유용하게 쓰임
  - 자연 상수 e가 지수 함수에 포함되어 분모에 들어가면 **시그모이드 함수**가 되는데, 이를 식으로 나타내면 다음과 같고 그래프는 그림과 같다.

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



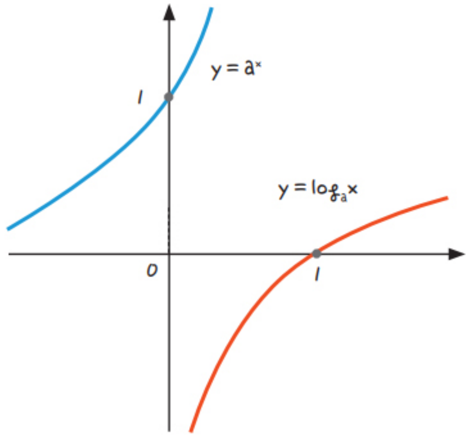
7 로그와 로그 함수

7 로그와 로그 함수

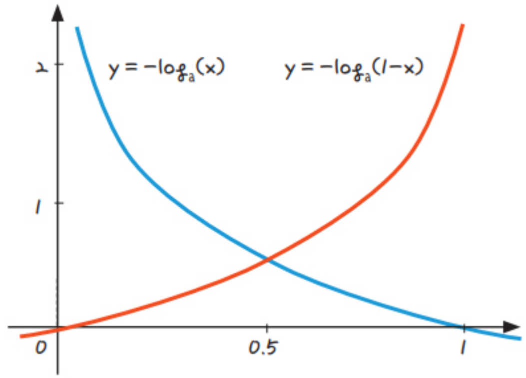
- **로그와 로그 함수**
  - A를 x만큼 거듭제곱한 값이 b라고 할 때 x를 구하기 위해 사용하는 방법이 로그이다.

$$a^x = b$$

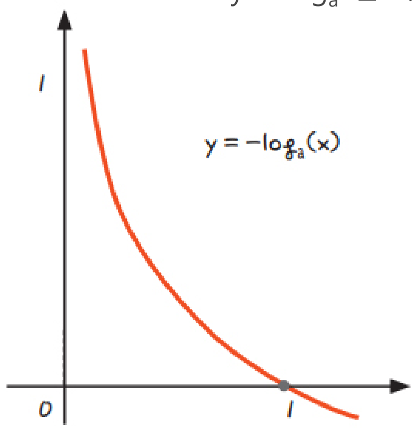
$$\log_a b = x$$



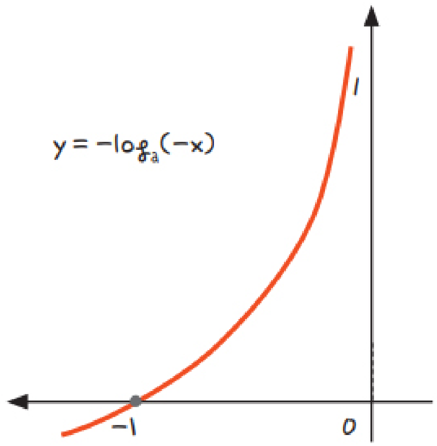
지수 함수 y = a^x의 그래프와 로그 함수
y = $\log_a x$는 대칭



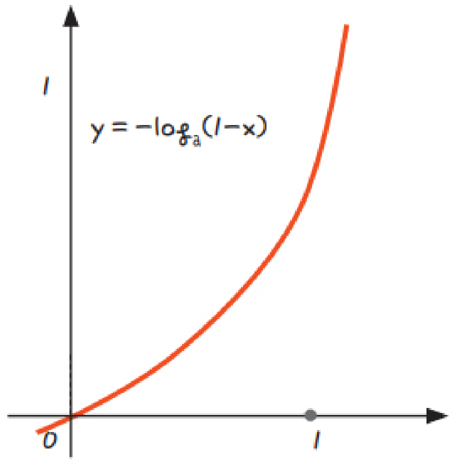
파란색은 x값이 0에 가까울 수록 오차가 커짐
빨간색은 x값이 1에 가까울 수록 오차가 커짐



위 왼쪽 그래프의 로그함수를 x축으로 대칭 이동한 그래프



왼쪽 그래프의 로그함수를 y축으로 대칭 이동한 그래프



위 왼쪽 그래프의 로그함수를 y축으로 대칭 이동한 그래프르 x축으로 1만큼 이동