회귀2

2장 딥러닝을 위한 기초 수학

1 | 일차 함수, 기울기와 y절편

2 | 이차 함수와 최솟값

3 | 미분, 순간 변화율과 기울기

4 | 편미분

5 | 지수와 지수 함수

6 | 시그모이드 함수

7 | 로그와 로그 함수

2 딥러닝을 위한 기초 수학

- 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

1 | 일차 함수, 기울기와 y 절편

- 함수 :

  두 집합 사이의 관계를 설명하는 수학 개념
- 변수 x와 y가 있을 때, x가 변하면 이에 따라 y는 어떤 규칙으로 변하는지를 나타냄
- 보통 함수를 나타낼 때는 function의 f와 변수 x를 사용해 $y=f(x)$ 라고 표시함
- $y = ax + b\,(a \neq 0)$

  x가 y에 관한 일차식으로 표현된 경우를 말함

- x가 일차인 형태이며 x가 일차로 남으려면 a는 0이 아니어야 함

1 | 일차 함수, 기울기와 y 절편

- 일차 함수식 $y = ax + b$ 에서 a는 기울기, b는 절편이라고 함

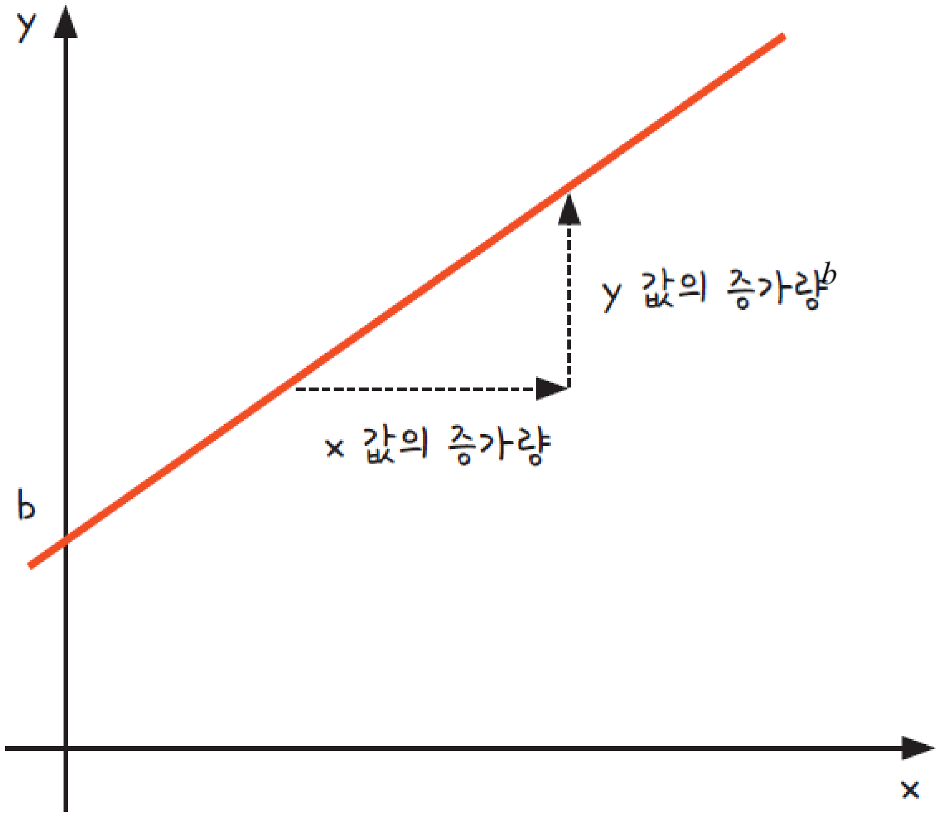


그림 2-1 일차 함수 그래프

1 | 일차 함수, 기울기와 y 절편

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

2 | 이차 함수와 최솟값

- 이차 함수란 y가 x에 관한 이차식으로 표현되는 경우를 말함

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

- a > 0이면 아래로 볼록한 그래프가 됨

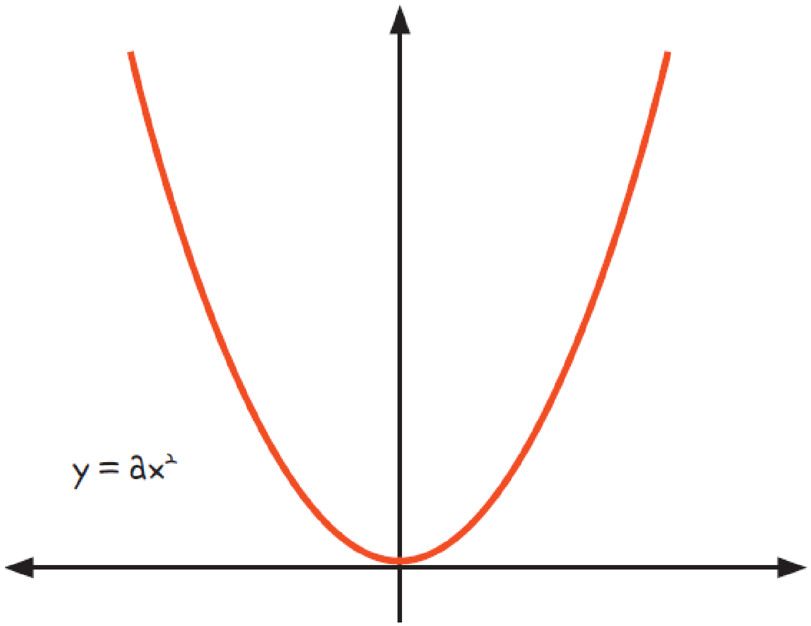


그림 2-2 이차 함수 그래프

2 | 이차 함수와 최솟값

- 포물선의 맨 아래에 위치한 지점이 최솟값이 되는데, 딥러닝을 실행할 때는 이 최솟값을 찾아내는 과정이 매우 중요함

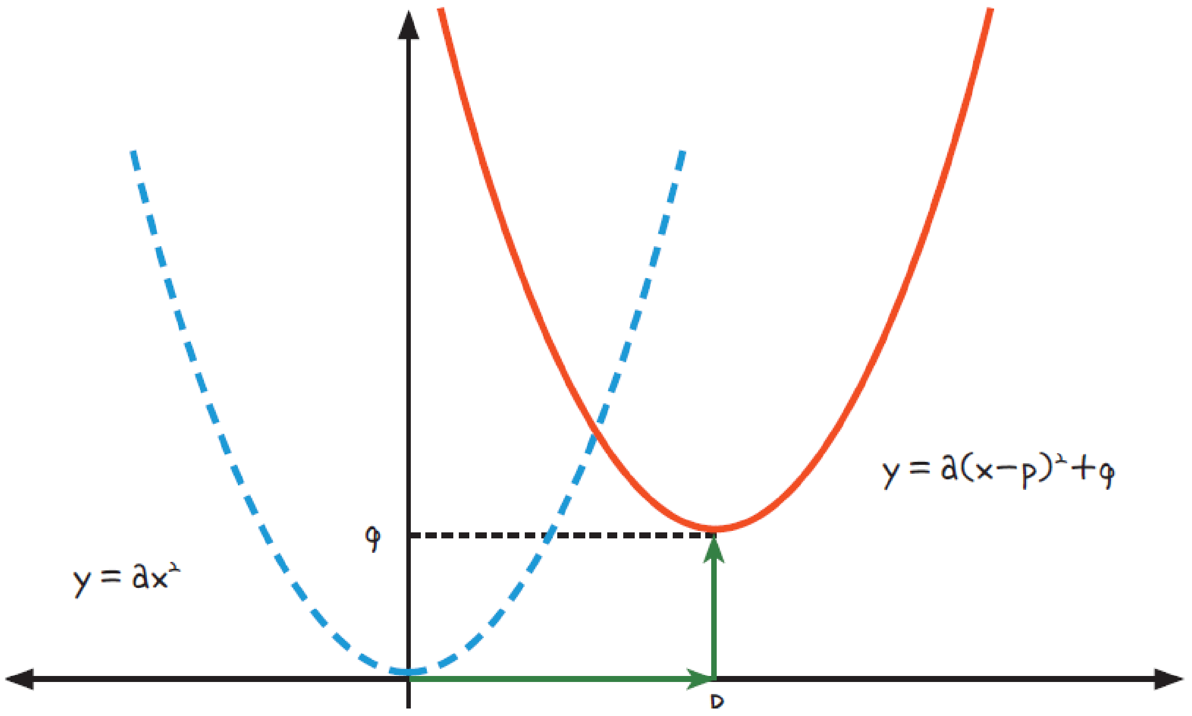


그림 2-3 이차 함수 그래프의 평행이동과 최솟값

3 | 미분, 순간 변화율과 기울기

- 딥러닝을 이해하는 데 가장 중요한 수학 원리는 미분이라고 할 수 있음
- 너무 미세해서 실제로 움직이는 게 아니라 방향만 드러내는 정도의 순간적인 변화만 있을 것임
- 이 순간의 변화를 놓고 순간 변화율이라는 이름을 붙임
- 순간 변화율은 어느 쪽을 향하는 방향성을 지니고 있으므로, 이 방향을 따라 직선을 길게 그려주면 그래프와 맞닿는 접선이 그려짐
- 이 선이 바로 이 점에서의 기울기가 됨

3 | 미분, 순간 변화율과 기울기

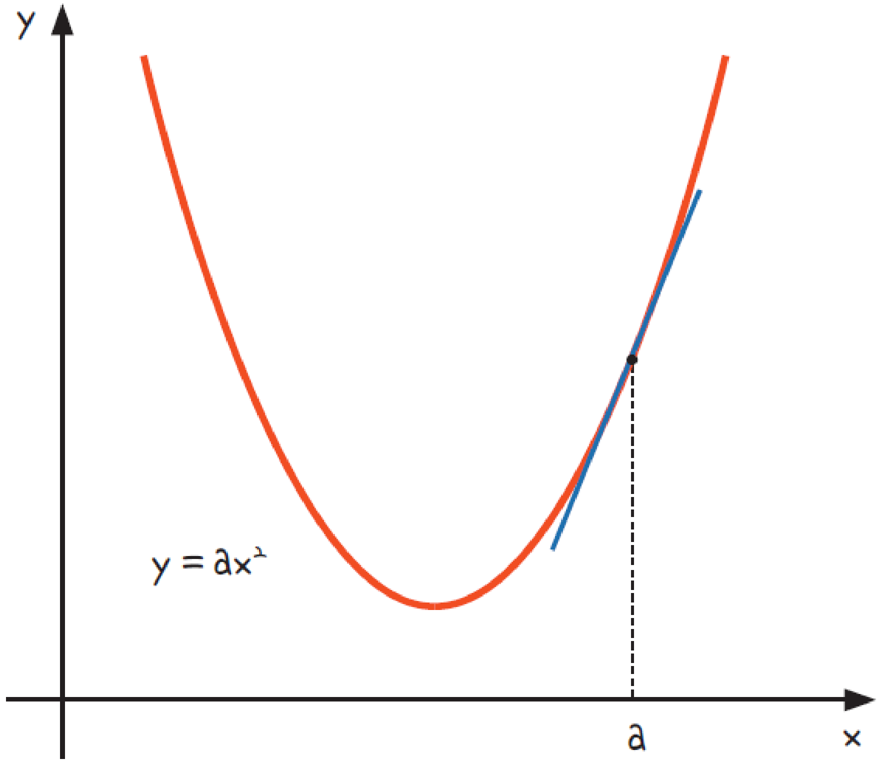


그림 2-4 a에서의 순간 변화율은 곧 기울기다!

3 | 미분, 순간 변화율과 기울기

- 미분을 한다는 것은 쉽게 말해 이 '순간 변화율'을 구한다는 것임
- 미분 계수 :

  어느 순간에 어떤 변화가 일어나고 있는지를 숫자로 나타낸 것
- 이 미분 계수는 곧 그래프에서의 기울기를 의미함

3 | 미분, 순간 변화율과 기울기

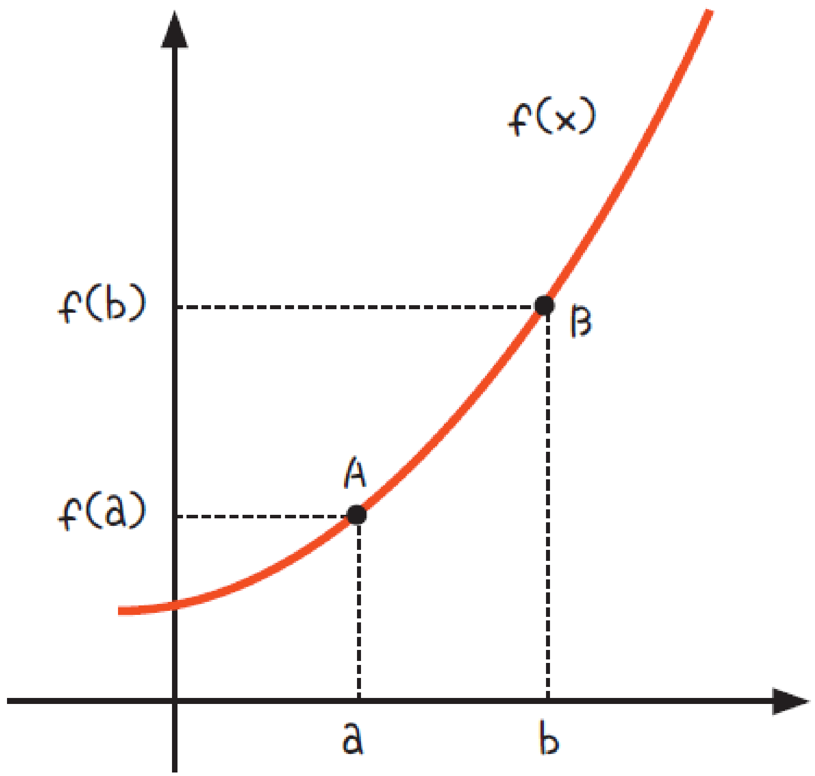


그림 2-5 함수 f(x)의 x축 위에 두 실수 a와 b를 대입

3 | 미분, 순간 변화율과 기울기

- 여기서 Δ(델타)는 변화량을 나타내는 기호임

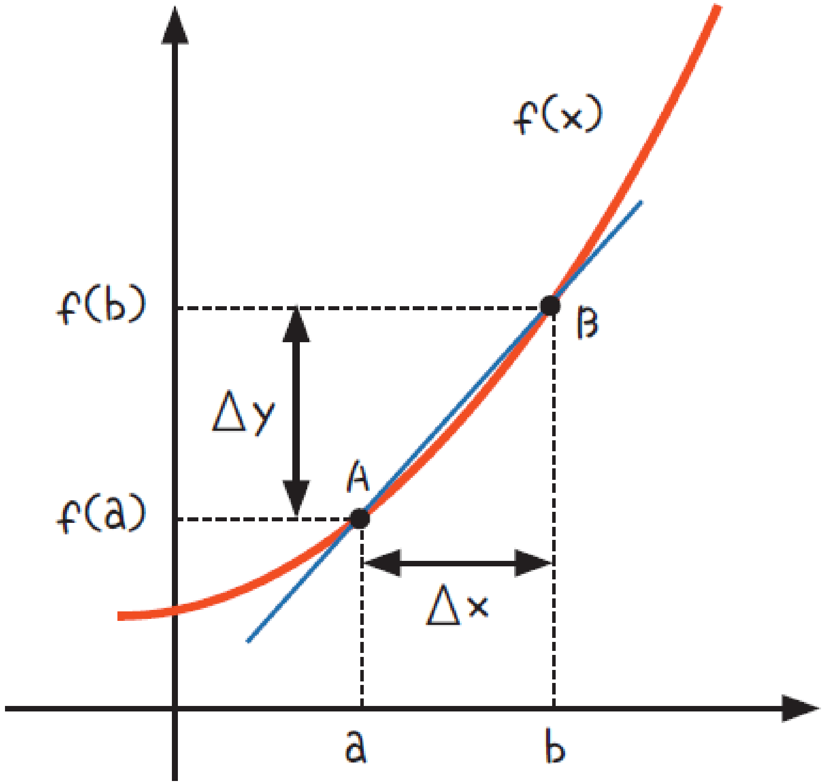


그림 2-6 A, B를 지나는 직선은 이 두 점 간의 기울기, 곧 평균 변화율을 의미

3 | 미분, 순간 변화율과 기울기

- 이 그래프에서 x값의 증가량은 $b-a$이고, y값의 증가량은 $f(b)-f(a)$ 임

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

- 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)이 0에 가까울 만큼 아주 작을 때의 순간적인 기울기를 말함

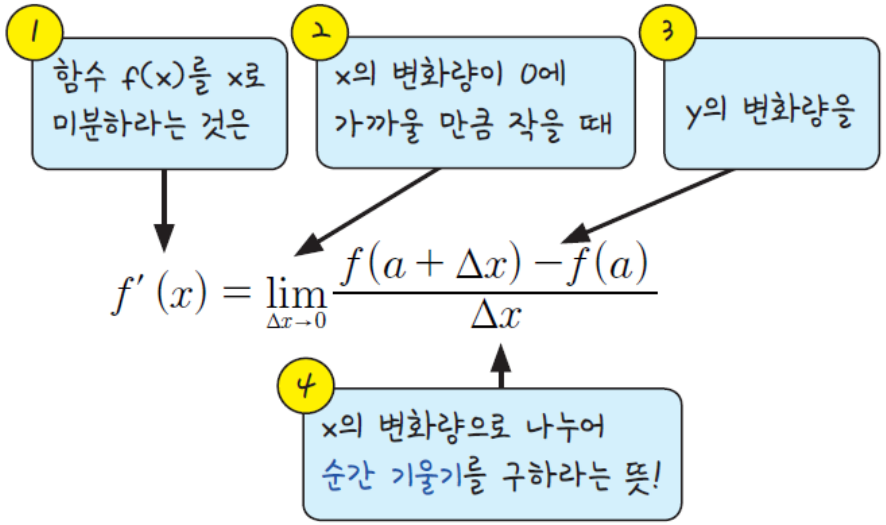
$$\lim_{\Delta x \to 0} \frac{f(a+\Delta x)-f(a)}{\Delta x}$$

3 | 미분, 순간 변화율과 기울기

- $\lim\limits_{\Delta x \to 0}$:

  'x의 증가량이 0에 가까울 만큼 작을 때'라는 뜻
- 기울기는 $\dfrac{y\text{값의 증가량}}{x\text{값의 증가량}}$이므로 순간 기울기는 $\lim\limits_{\Delta x \to 0}\dfrac{y\text{값의 증가량}}{x\text{값의 증가량}}$으로 표현되며,

  이것은 $\lim\limits_{\Delta x \to 0}\dfrac{\Delta y}{\Delta x} = \dfrac{dy}{dx}$라고도 쓸 수 있음
- "함수 $f(x)$ 를 미분하라"는 것을 $f'(x)$ 또는 $\dfrac{d}{dx}f(x)$로 표기함



3 | 미분, 순간 변화율과 기울기

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

1 | $f(x)=a$ 에서 가 상수일 때 미분 값은 0임

2 | $f(x)=x$ 일 때의 미분 값은 1임

3 | $f(x)=ax$ 에서 a가 상수이면 미분 값은 a 임

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

4 | 편미분

- 미분과 더불어 딥러닝을 공부할 때 가장 자주 접하게 되는 또 다른 수학 개념은 바로 편미분임
- 미분과 편미분 모두 '미분하라'는 의미에서는 다를 바가 없음
- 여러 가지 변수가 식 안에 있을 때, 모든 변수를 미분하는 것이 아니라 우리가 원하는 한 가지 변수만 미분함
- 그 외에는 모두 상수로 취급하는 것이 바로 편미분임

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

4 | 편미분

- 변수가 f와 y중 어떤 변수로 미분해야 하는지를 정해야 하므로 편미분을 사용하는 것
- 만일 이 식처럼 여러 변수 중에서 x에 관해서만 미분하고 싶다면, 함수 f를 ‘ x에 관해 편미분하라’고 함

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

4 | 편미분

- 함수 $f(x, y) = x^2 + yx + a$ 에 관해 편미분 하는 과정은 미분의 성질 4번에 따라 x^2 항은 $2x$가 됨
- x 에 관해 미분하면 다른 모든 항은 상수로 취급하므로 y는 상수가 됨
- 미분의 성질 3번에 따라 xy 는 y가 됨
- 마지막 항 a 는 미분의 성질 1번에 따라 0이 됨

$$f(x,y) = x^2 + xy + a \text{ 일 때,}$$

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

5 | 지수와 지수 함수

- 지수란 다음과 같은 형태를 말함

$$a^{\square}$$

- 여기서 a 를 '밑'이라 하고 $\square$를 '지수'라고 부름
- a 를 $\square$만큼 반복해서 곱한다는 뜻
- 지수 함수 :

  변수 x 가 지수 자리에 있는 경우를 말함

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

5 | 지수와 지수 함수

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

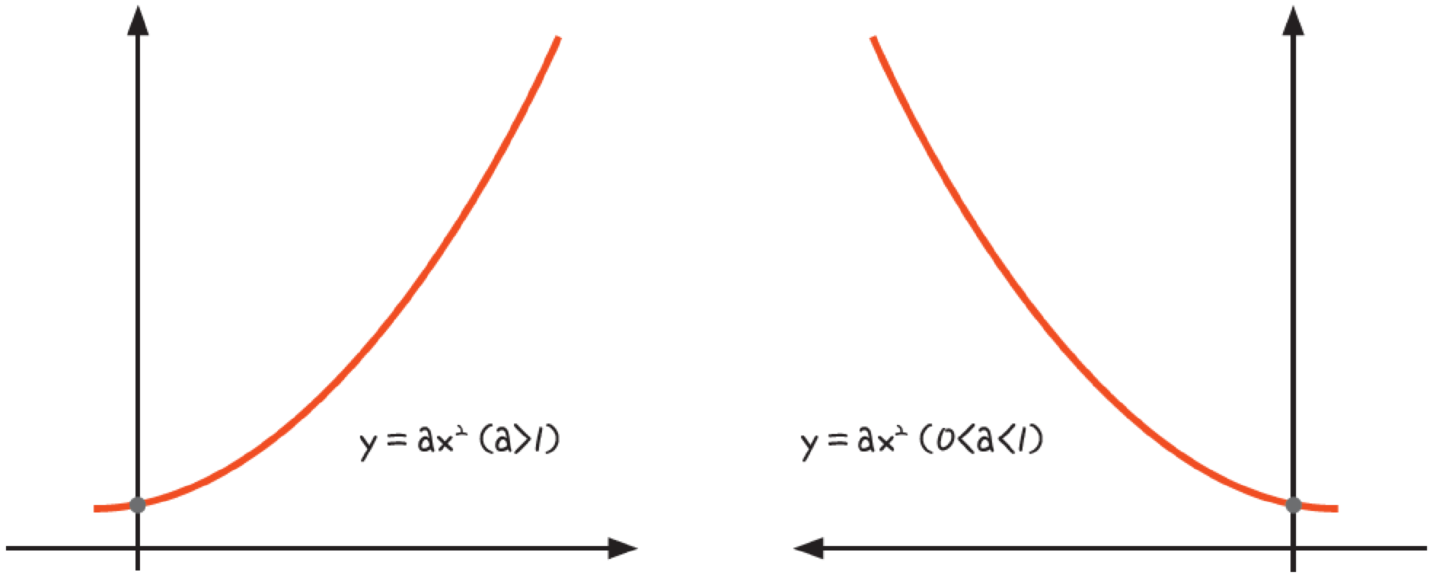


그림 2-7 a>1일 때와 0<a<1일때의 지수 함수

6 | 시그모이드 함수

- 시그모이드 함수 :

  지수 함수에서 밑의 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…임
- 자연 상수 e가 지수 함수에 포함되어 분모에 들어가면 시그모이드 함수가 됨

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

6 | 시그모이드 함수

- 시그모이드 함수를 그래프로 그려보면 그림 2-8과 같이 S자 형태로 나타남

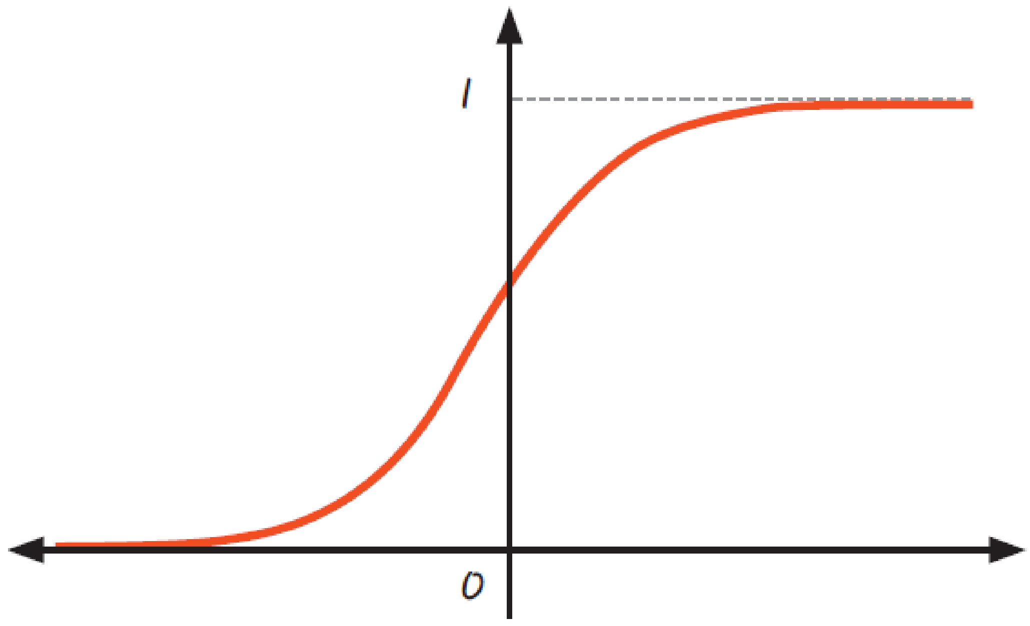


그림 2-8 시그모이드 함수의 그래프

7 | 로그와 로그 함수

- 로그를 이해하려면 먼저 지수부터 이해해야 함

$$a^x = b$$

- 로그 :

x를 구하기 위해 사용하는 방법

- 영어로 logarithm이라고 하는데 앞 세 글자 log를 씀
- 지수 식에서 a와 b의 위치를 다음과 같이 바꾸어 써주면 됨

$$a^x = b$$

$$\log_a b = x$$

7 | 로그와 로그 함수

- 로그가 지수와 이렇게 밀접한 관계가 있듯, 로그 함수 역시 지수 함수와 밀접한 관계에 있는데 바로 역함수의 관계임
- 역함수는 x 와 y를 서로 바꾸어 가지는 함수임
- 지수 함수 $y = a^x (a \neq 1,\ a > 0)$로그의 정의를 따라 $x = \log_a y$로 바꿀 수 있음
- 역함수를 만들기 위해 x와 y를 서로 바꾸어 주면 됨 $y = \log_a x$

7 | 로그와 로그 함수

- 역함수의 그래프는 $y = x$ 에 대하여 대칭인 선으로 나타남

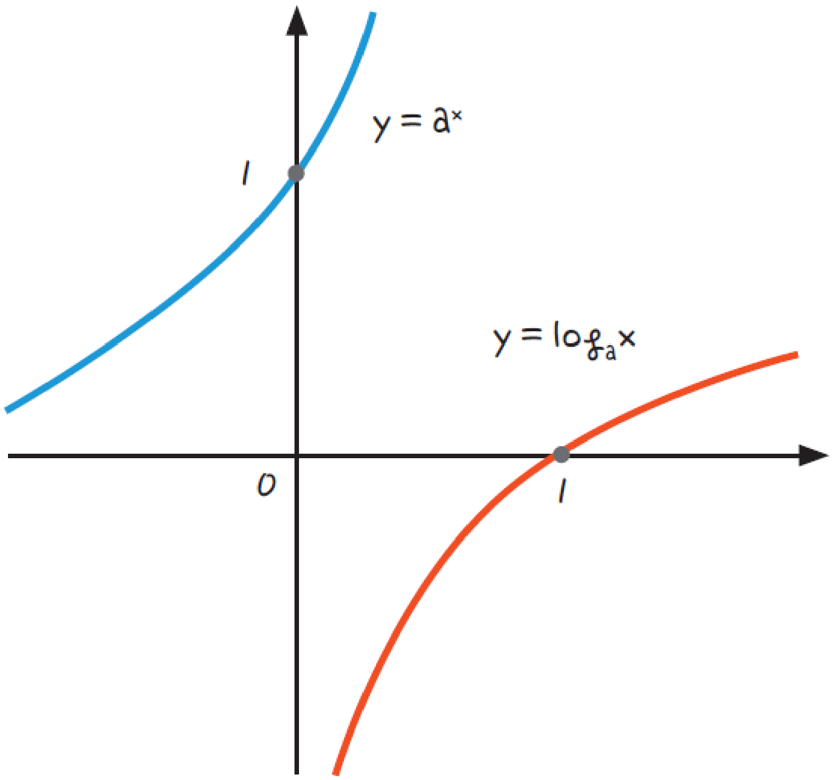


그림 2-9 지수 함수 y = a x와 로그 함수 y = log a x

7 | 로그와 로그 함수

- 우리는 x 가 1에 가까워지거나, 0에 가까워질수록 오차가 커지는 그래프가 필요함

1 | x축에 대하여 대칭 이동

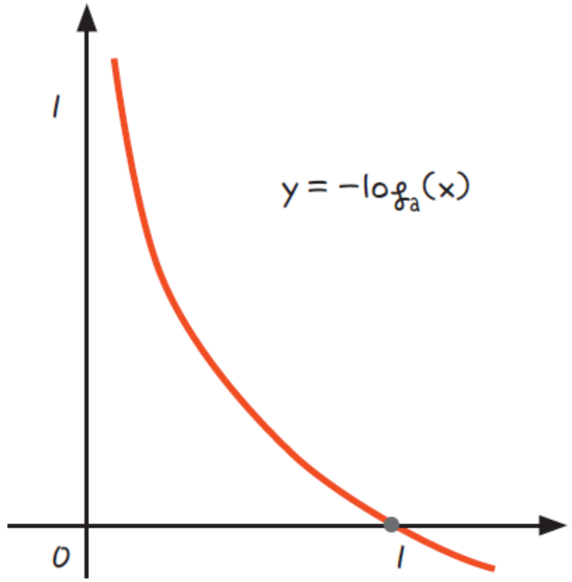


그림 2-10 y = -log a x 그래프

7 | 로그와 로그 함수

2 | x축과 y축에 대하여 대칭 이동

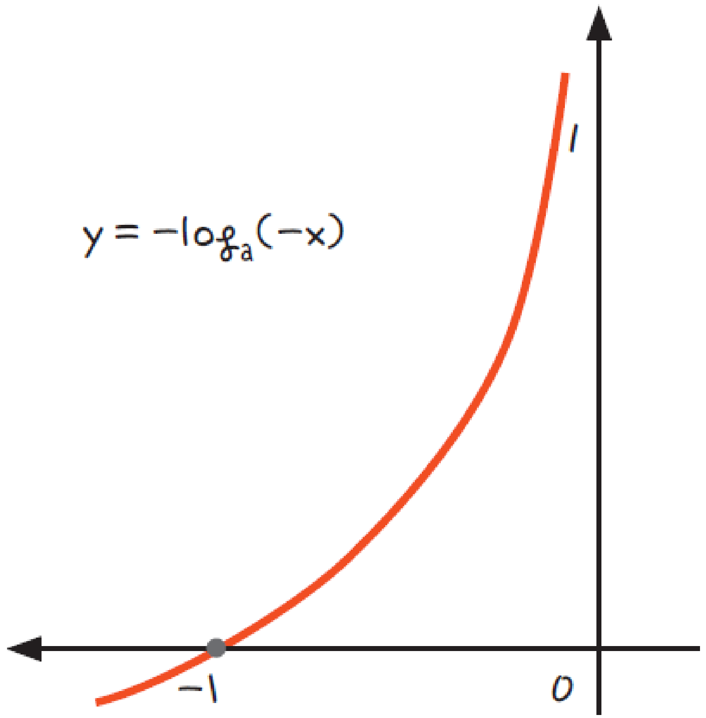


그림 2-11 y = -log a (-x) 그래프

7 | 로그와 로그 함수

3 | 2번 그래프를 x축 오른쪽 방향으로 1만큼 평행 이동

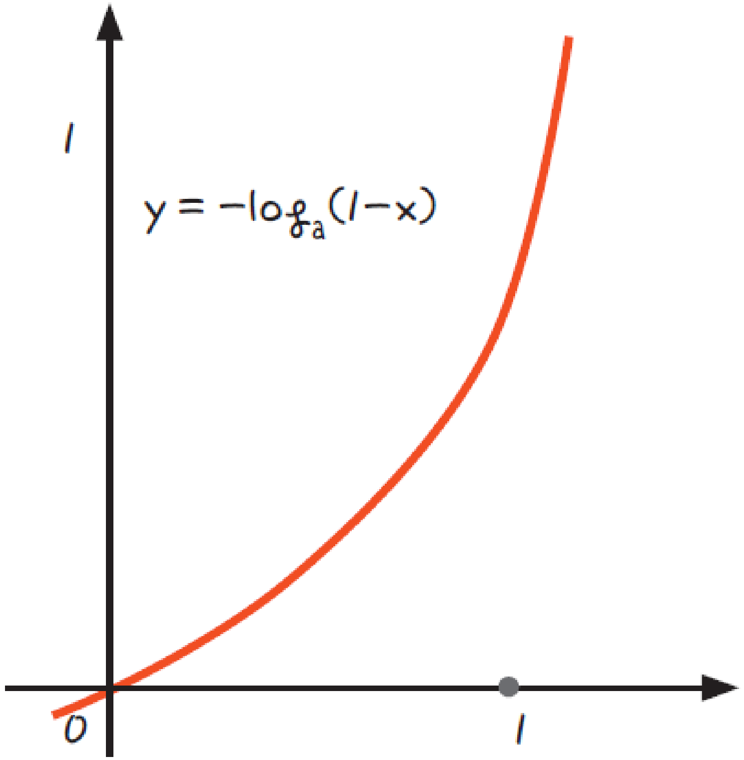


그림 2-12 y = -log a (1-x) 그래프

7 | 로그와 로그 함수

4 | 1번과 3번을 함께 나타낸 그래프

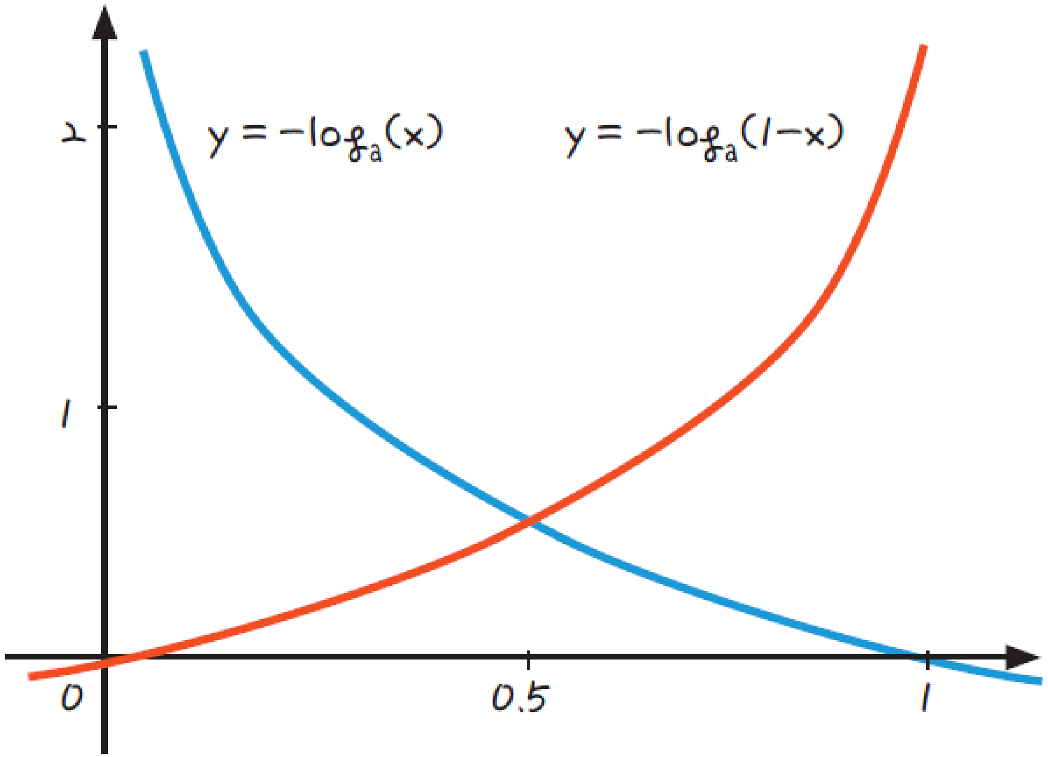


그림 2-13 y = -log a x 와 y = -log a (1-x) 그래프