4.2 심파이를 사용한 함수미분

- 데이터분석에 미적분이 필요한 이유? → 최적화에 사용되는 중간단계!

→ 초보적 수준의 미적분만 필요함

"개념 용어 주요공식 에 집중할 것"

$\boxed{\text{최소자승문제}}$

$x \rightarrow y$

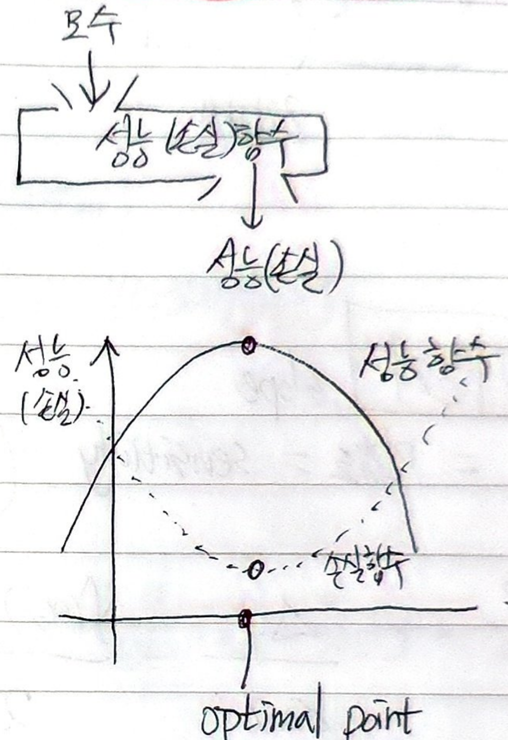
$\hat{y} = w^T x$

$= w_1 x_1 + \cdots + w_N x_N$

목적 $\hat{y} \approx y$

$e^2 = (y - \hat{y})^2$

$w = (X^T X)^{-1} X y$



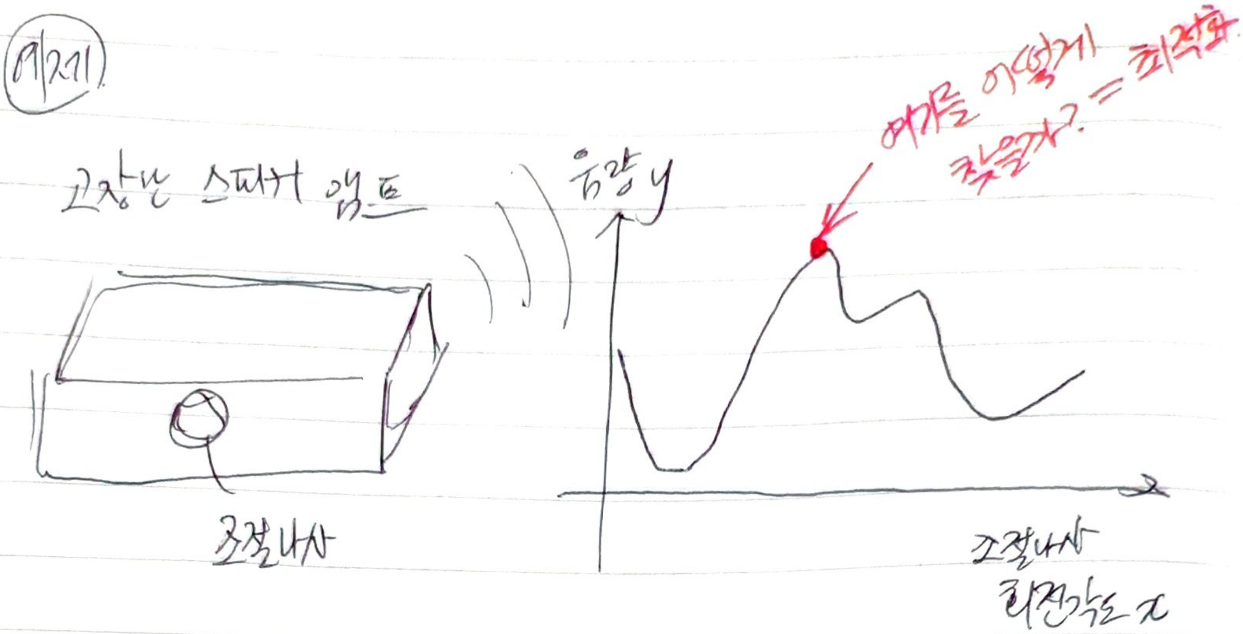
미분의 사용목적

→ 성능(손실)함수에 투입하는 모수를 조금 더 투입하면 성능(손실) 출력값이 증가하는가, 혹은 감소하는가를 확인

→ 투입 모수를 늘려야 할지 줄여야 할지에 대한 정보 확인

→ 최적화 iteration.

예제



기울기 slope

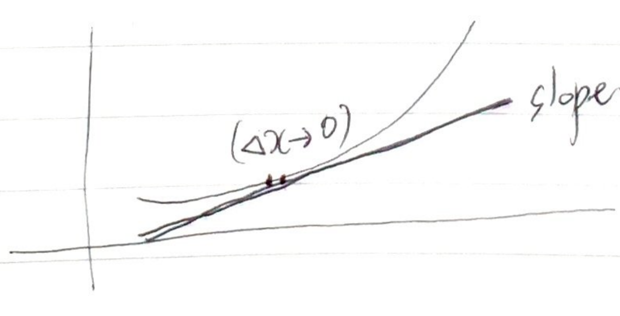
= 민감도 = sensitivity "반응"

$$\frac{\Delta y}{\Delta x} = \frac{f(x_2) - f(x)}{x_2 - x} = \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

- x를 요만큼 바꾸면 y는 얼만큼 바뀌는가?

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

. 이렇게 계산하면 Δx 크기에 따라 기울기가 변화되는 문제를 해결할 수 있다. (일관적인 slope)



수치미분

scipy.misc 패키지의 derivative() 명령.

$$\boxed{\text{Slope} \approx \frac{f(x+\frac{1}{2}dx) - f(x-\frac{1}{2}dx)}{dx}}$$

(numerical differentiation)

— 대략적인 기울기를 구한다 (정확하지 않음)

미분 (differentiation)

$$f' = \frac{d}{dx}(f) = \frac{d}{dx}f = \frac{df}{dx} = \frac{d}{dx}(y) = \frac{d}{dx}y = \frac{dy}{dx}$$

f' = "f 프라임 (prime)"

$\frac{df}{dx}$ = "df over dx"

미분하다 = differentiate (동사)

도함수 = derivative (명사)

$f = x^2$

differentiate

$f' = \frac{df}{dx} = 2x$

(derivative)

미분가능

"ReLU"

"NOT DIFFERENTIABLE AT $x=0$"

"Soft plus function"

☆ 미분공식

핵심 4가지!

- 기본 미분공식
- 선형조합법칙
- 곱셈법칙
- 연쇄법칙

"기본 미분공식"

: 상수, 거듭제곱, 로그함수, 지수함수 등 간단한 함수에 대한 미분공식.

- 상수

$$\frac{d}{dx}(c)=0$$

- 거듭제곱

$$\frac{d}{dx}(x^n)=nx^{n-1}$$

예제

거듭제곱 공식의 응용.

$$\frac{d}{dx}(x^3) = 3x^2$$

$$\frac{d}{dx}(x^2) = 2x$$

$$\frac{d}{dx}(x) = 1$$

$$\frac{d}{dx}\left(\frac{1}{x}\right) = \frac{d}{dx}x^{-1} = -1 \cdot x^{-2} = \frac{-1}{x^2}$$

$$\frac{d}{dx}\left(\frac{1}{x^2}\right) = \frac{d}{dx}x^{-2} = -2 \cdot x^{-3} = -\frac{2}{x^3}$$

$$\frac{d}{dx}\left(\frac{1}{x^3}\right) = \frac{d}{dx}x^{-3} = -3 \cdot x^{-4} = \frac{-3}{x^4}$$

"선형조합법칙"

$$\frac{d}{dx}(cf) = c \cdot \frac{df}{dx}$$

$$\frac{d}{dx}(f_1 + f_2) = \frac{df_1}{dx} + \frac{df_2}{dx}$$

$$\frac{d}{dx}(c_1 f_1 + c_2 f_2) = c_1 \frac{df_1}{dx} + c_2 \frac{df_2}{dx}$$

예제 $y = 1 + 2x + 3x^2 + 4\exp(x) + 5\log(x)$

$$\frac{d}{dx}y = 0 + 2 + 6x + 4\exp(x) + \frac{5}{x}$$

곱셈법칙

어떤 함수의 형태가 두 함수를 곱한 것과 같을 때는 다음과 같이 개별 함수의 도함수를 사용하여 원래 함수의 도함수를 구한다. 이를 곱셈법칙이라고 한다.

$$\frac{d}{dx}(f \cdot g) = f \cdot \frac{dg}{dx} + g \cdot \frac{df}{dx}$$

$$f = xe^x$$

$$\frac{df}{dx} = xe^x + e^x$$

* 연쇄법칙 chain rule.

미분하고자 하는 함수의 입력변수가 다른 함수의 출력변수인 경우 적용할 수 있다.

$$f(x) = h(g(x))$$

인 도함수는 다음과 같이 구한다.

g 때문에 h가 바뀐 양

x 때문에 g가 바뀐 양

$$\boxed{\frac{df}{dx} = \frac{dh}{dg} \cdot \frac{dg}{dx}}$$

||

x 때문에 f (= h·g)가 바뀐 양

예제 나중에 다룰 정규분포 (Gaussian normal distribution)의 확률밀도함수 (probability density function)는 기본적으로 다음과 같은 형태라고 볼 수 있다.

$$f = \exp \frac{(x-\mu)^2}{\sigma^2}$$

$f = \exp\left(\frac{(x-\mu)^2}{\sigma^2}\right)$ 도함수 구하기.

$f = \exp(z)$, $z = \frac{y^2}{\sigma^2}$, $y = x - \mu$

연쇄법칙을 적용하면,

$$\frac{df}{dx} = \frac{df}{dz}\cdot\frac{dz}{dy}\cdot\frac{dy}{dx}$$

연쇄법칙에 사용된 각각의 도함수는 다음과 같다. 이때 임시로 사용된 중간 변수는 없앤다.

$$\frac{df}{dz} = \exp(z) = \exp\left(\frac{(x-\mu)^2}{\sigma^2}\right)$$

$$\frac{dz}{dy} = \frac{2y}{\sigma^2} = \frac{2(x-\mu)}{\sigma^2}$$

$$\frac{dy}{dx} = 1$$

$$\boxed{\therefore \frac{df}{dx} = \frac{2(x-\mu)}{\sigma^2}\exp\left(\frac{(x-\mu)^2}{\sigma^2}\right)}$$

☆

예제 로그 함수에 연쇄법칙을 적용하면 다음과 같다. (규칙)

$$\boxed{\frac{d}{dx}\log f(x) = \frac{f'(x)}{f(x)}}$$

$$\because \frac{d}{dx}\log f(x) = \frac{d}{df}(\log f)\cdot\frac{d}{dx}(f)$$

연습문제 4.2.3

다음 함수를 미분하라. 이 식에서 k, a, b는 변수가 아니라 상수다.

(1) $f(x) = x^3 - 1$

$$\frac{d}{dx} f(x) = 3x^2$$

(2) $f(x) = \log(x^2 - 3k)$

$$\frac{d}{dx}\log f \cdot \frac{d}{dx} f = \frac{1}{x^2-3k} \cdot 2x$$

(3) $f(x) = \exp(ax^b)$

$g(x) = ax^b$ 라 할 때,

$f(g) = \exp g$.

$$f'(x) = \frac{df}{dg} \cdot \frac{dg}{dx} = \exp g \cdot abx^{b-1}$$

$$= abx^{b-1} \cdot \exp(ax^b)$$

2차 도함수

도함수를 한번 더 미분 → 2차 도함수

f'' or $\frac{d^2}{dx^2}$

ex. $y=f(x)$ 라는 함수의 2차 도함수 표기.

$$f'' = \frac{d^2}{dx^2}(f) = \frac{d^2}{dx^2} f = \frac{d^2 f}{dx^2} = \frac{d^2}{dx^2}(y)$$

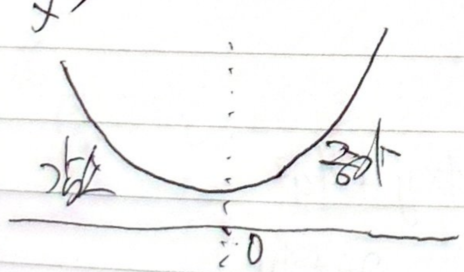
$$= \frac{d^2}{dx^2} y = \frac{d^2 y}{dx^2}$$

(2차 도함수 = 도함수의 기울기)

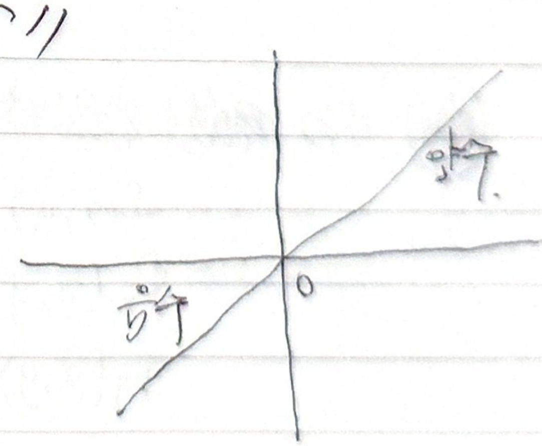
- 도함수 값 증가 $\Longleftrightarrow$ 2차 도함수 > 0 "CONVEX" 볼록
- 도함수 값 감소 $\Longleftrightarrow$ 2차 도함수 < 0 "CONCAVE" 오목

2차 도함수 값을 볼록도 'convexity' 라고 부르기도 함!

f'



f''



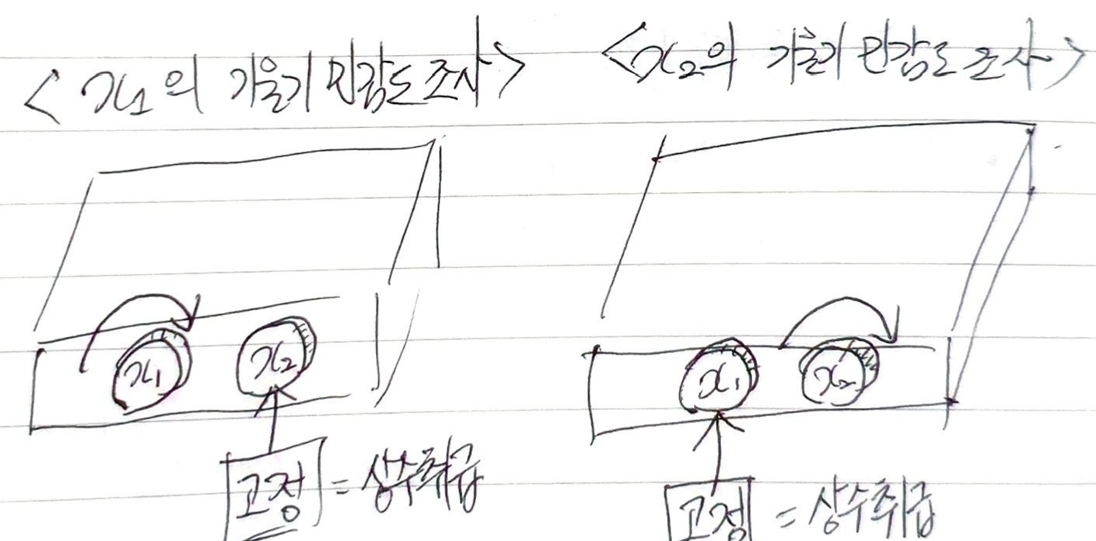
편미분

둘 이상의 독립변수를 갖는 다변수 함수인 경우에도,
미분 (=기울기)는 하나의 변수에 대해서만 구할 수도 있다.
이를 편미분 (partial difference)이라고 한다.
따라서, 편미분의 결과로, 하나의 함수에 대해 여러개의
도함수가 나올 수 있다.

round symbol

$$f_x(x, y) = \frac{\partial f}{\partial x}$$

$$f_y(x, y) = \frac{\partial f}{\partial y}$$



예제) 다음은 편미분의 간단한 예다.

$$f(x, y) = x^2 + 4xy + 4y^2$$

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

$$f_y(x, y) = \frac{\partial f}{\partial y} = 4x + 8y$$

$\boxed{\text{다변수 함수의 연쇄법칙}}$

다변수 함수의 미분을 구할 때도 함수가 연결되어 있으면 연쇄법칙이 적용된다. 예를 들어, 변수 x를 입력으로 가지는 함수가 $f_1, f_2, \dots, f_N$과 같이 N개가 있고 각각의 출력을 $y_1, y_2, \cdots y_N$이라고 하자.

$$
\begin{aligned}
y_1 &= f_1(x) \\
y_2 &= f_2(x) \\
&\vdots \\
y_N &= f_N(x)
\end{aligned}
$$

$$z = g(y_1, y_2, \cdots y_N) \text{ 일 때,}$$

변수 x값의 변화에 따른 z값의 변화는 다음처럼 계산한다.

$$\frac{dz}{dx} = \frac{\partial z}{\partial y_1}\frac{dy_1}{dx} + \frac{\partial z}{\partial y_2}\frac{dy_2}{dx} + \cdots + \frac{\partial z}{\partial y_N}\frac{dy_N}{dx}$$

$$
\begin{tikzcd}
x \arrow[r, "f_1"] \arrow[dr, "f_2"] \arrow[ddr, "f_3"] & y_1 \arrow[r, "g"] & z \\
& y_2 \arrow[ur, "g"] & \\
& y_3 \arrow[uur, "g"] &
\end{tikzcd}
$$

"연쇄적인 변화량의 합"

$$\frac{dz}{dy_1}\frac{dy_1}{dx} + \frac{dz}{dy_2}\frac{dy_2}{dx} + \frac{dz}{dy_3}\frac{dy_3}{dx} \cdots$$

이번에는 함수 $f_1, f_2, \cdots f_N$이 $x_1, x_2, \cdots x_M$을 입력으로 갖는 다변수 함수라고 하자.

$$\begin{pmatrix} y_1 = f_1(x_1, x_2, \cdots, x_M) \\ y_2 = f_2(x_1, x_2, \cdots, x_M) \\ y_3 = f_3(x_1, x_2, \cdots, x_M) \\ \vdots \\ y_N = f_N(x_1, x_2, \cdots, x_M) \end{pmatrix}$$

이때의 변수 x_1 값의 변화에 따른 z값의 변화도 마찬가지로 계산할 수 있다.

round 기호로 바뀜 (다른 애들이 있다는 뜻) ($y_2, y_3, \cdots$ = 상수)

$$\frac{\partial z}{\partial x_1} = \frac{\partial z}{\partial y_1}\frac{\partial y_1}{\partial x_1} + \frac{\partial z}{\partial y_2}\frac{\partial y_2}{\partial x_1} + \cdots + \frac{\partial z}{\partial y_N}\frac{\partial y_N}{\partial x_1}$$

2차 편미분

편미분에 대해서도 2차 도함수를 정의할 수 있다.

- 편미분의 2차 도함수를 구할 때는 각각의 미분에 쓰이는 독립변수를 자유롭게 선택할 수 있다.

• 첫번째 미분과 두번째 미분에서 모두 y에 대해 미분하면 다음과 같이 표기한다.

$$f_{xx}(x, y) = \frac{\partial^2 f}{\partial x^2}$$

첫번째, 두번째 미분 — y에 대해서

$$f_{yy}(x,y) = \frac{\partial^2 f}{\partial y^2}$$

첫번째 x, 두번째 y에 대해 미분

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

$$\begin{tikzcd}
 & & f(x,y) \arrow[dl, "\frac{\partial}{\partial x}"'] \arrow[dr, "\frac{\partial}{\partial y}"] & & \\
 & f_x(x,y) \arrow[dl, "\frac{\partial}{\partial x}"'] \arrow[d, "\frac{\partial}{\partial y}"] & & f_y(x,y) \arrow[d, "\frac{\partial}{\partial x}"'] \arrow[dr, "\frac{\partial}{\partial y}"] & \\
f_{xx}(x,y) & f_{xy}(x,y) & & f_{yx}(x,y) & f_{yy}(x,y)
\end{tikzcd}$$

함수가 연속하고 미분 가능하면,

$$f_{xy}(x,y) = f_{yx}(x,y)$$

; 슈바르츠 정리 (Schwartz's theorem)

예제

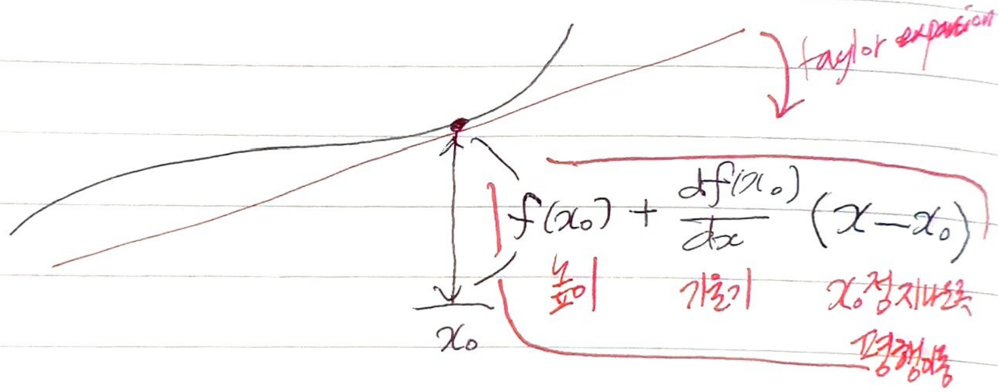
위 다변수 함수에 대해 2차 및 2차 편미분 도함수를 구하면 다음과 같다.

$$f_{xx}(x, y) = 2$$
$$f_{yy}(x, y) = 8$$
$$f_{xy}(x, y) = 4$$
$$f_{yx}(x, y) = 4$$

슈바르츠 정리 성립!

테일러전개

Taylor expansion.



- 함수의 기울기 (1차 미분값)를 알고 있다면, 함수의 모양을 다음처럼 근사화할 수 있다. x_0는 함수값과 기울기를 구하는 x 위치이며 사용자가 마음대로 설정할 수 있다.

$$f(x) \approx f(x_0) + \frac{df(x_0)}{dx}(x - x_0)$$

다변수 함수의 경우, 다음과 같이 테일러 전개한다

$$f(x, y) \approx f(x_0, y_0) + \frac{\partial f(x_0, y_0)}{\partial x}(x - x_0) + \frac{\partial f(x_0, y_0)}{\partial y}(y - y_0)$$