Calculus

Theory: Differential & Integral

Section #1

통계학과 미분

- 모델링

- 통계학에서 미분은 숨 쉬듯이 사용하지만 주로 모델링, 변수 효과 추적, 최적화 세가지 목적을 가지고 사용한다.
- 수학적 모델링은 어떤 수학적 모델을 사용하느냐가 중요한데, 우리가 아는 다항함수를 시작으로 지수함수, 삼각함수 등 여러 함수를 사용할 수 있다.

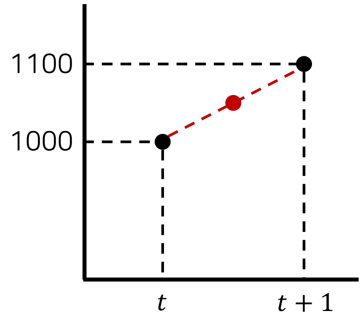
$$F = ma$$

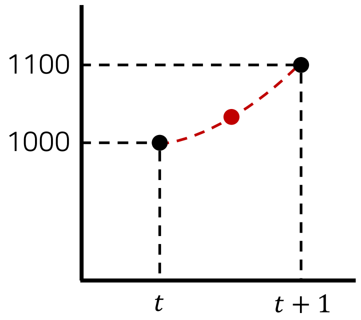
- 위의 방정식은 질량이 고정된 상태를 가정한 뉴턴 역학에서 힘과 가속도의 관계를 일차 함수를 사용하여 모델링한 결과물이다. 위 모델과 같이 독립 변수(가속도)가 같을 때, 종속변수(가속도)가 항상 같은 모델을 결정론적 모델(deterministic model)이라고 한다.
- 여기서 공기저항, 측정기의 오류 등을 가정해 결과가 항상 같지 않음을 고려한다면 확률변수를 추가하여 아래와 같이 확률적 모델링(stochastic model)을 할 수도 있다.

$$F = ma + \epsilon$$

- 즉, 모델은 가정을 통해 현실을 단순화 시킨 것이며, 얼마만큼 단순화시킬 것(Level of abstraction)인지에 따라 모델의 복잡도가 결정된다.

- 어떤 사람이 연이자율 1%인 예금 1000만원을 넣었다고 하자. 넣은 지 6개월만에 예금을 찾는다고 하면 상환받는 원리금은 얼마인가라는 질문에 답은 크게 두 가지이다.





- 이자율을 일차함수로 모델링하는 단리를 통해 계산한 사람은 1050만원으로, 이자율을 지수함수로 모델링하는 복리를 통해 계산한 사람은 1048.8만원 정도로 계산했을 것이다.

$$a(t) = a(0)\times(1 + it)$$
$$a(t) = a(0)\times(1 + i)^t$$

i: 이자율, t: 시간

선형 근사 (Linear Approximation)

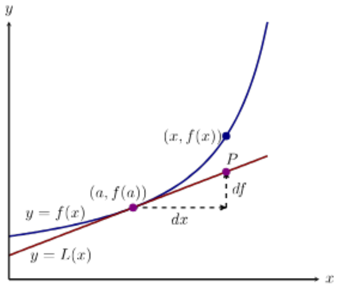
- 일변수 함수의 선형 근사와 오차

- 위 이자율 예시가 모든 t에 대해 함수를 가정한 것이라면, 어떤 t 근방에서만 국소적으로 함수를 가정할 수 있다. 이때 가정되는 함수의 형태가 일차함수이면 **선형 근사**라고 정의한다.

- 이제 원래 함수를 근사하는 가장 좋은 일차함수를 찾아보자. 가장 좋은 일차함수는 a 근방에서 $f(a)$와 같아야 하므로 다음과 같은 조건을 만족해야 한다.

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

$$\Rightarrow f(a) = pa + q,\ f'(a) = p$$



- 이를 통해 $px + q$를 구하면 선형 근사(Best Linear approximation)는 아래와 같이 정의된다.

$$f(x) \fallingdotseq \mathrm{L}(x) = f(a) + f'(a)(x-a)$$

- 위 식이 최적의 선형 근사임에도 필연적으로 원래의 값과의 오차를 지닐 수 밖에 없다. 다음 예제를 통해 선형 근사의 오차 추정을 알아보자.

$$y = x^2 + 3x$$

- 위의 함수를 $x = 2$ 근방에서 ± 0.2의 오차까지 허용하여 선형 근사를 하면,

$$L(x) = 10 + 7(x-2)$$

$$\Delta y \fallingdotseq \Delta x \left[\frac{dy}{dx}\right]_{x=2} = (0.2)(7) = 1.4$$

- 즉, x에 0.2의 오차가 있을 때 y에는 거의 1.4의 오차가 발생하게 된다. 위의 x와 y의 계산 순서를 반대로 하게 되면 종속변수에서 오차 수준을 만족하기 위해 독립 변수의 오차를 얼마까지 허용할 수 있는지 계산할 수 있다.

- 아래의 표를 통해 1989년의 네팔 인구를 추정해보자.

t	1985	1990	1995	2000
N(t)	17.04	19.33	21.91	24.7

(1) 선형 근사

$$\begin{aligned} L(x) &= N(1990) + N'(1990)(t-1990) \\ &= 19.33 + N'(1990)(t-1990) \end{aligned}$$

(2) 기울기를 평균 내어 $N'(1990)$ 추정

$$N'(1990) \approx \frac{1}{2}\left(\frac{21.91-19.33}{1995-1990} + \frac{19.33-17.04}{1990-1985}\right)$$

(3) $N(1989)$ 추정

$$L(x) = 19.33 + 0.4870(x-1990)$$
$$L(1989) = 19.33 + 0.4870(1989-1990)$$

테일러 전개 (Taylor Expansion)

- 일변수 함수의 테일러 전개(Taylor Expansion)와 확률변수 근사

- 하지만, 일차함수만으로는 함수를 충분히 근사하기 어려울 수 있다. 이때는 다항함수를 이용하여 함수를 근사할 수 있다. 이를 **테일러 전개**라고 한다.

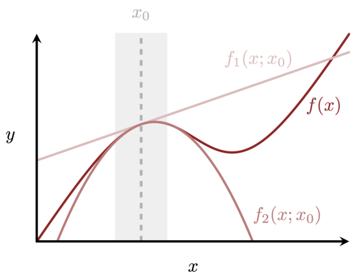
$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}}{n!}(x-a)^n = f(a) + f'(a)(x-a) + \frac{1}{2!}f''(a)(x-a)^2 + \cdots$$

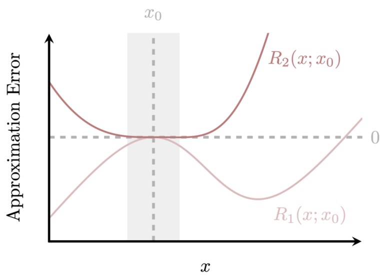
$f^{(n)}(x)$은 함수 $f(x)$의 n계 도함수이다.

- 하지만 실제로 함수의 모든 n계 도함수를 구할 수 없으므로 m차까지 절삭하여 사용하게 된다. 이때 m+1차부터는 모두 오차($R(x;a)$)에 해당이 된다. 아래의 그림을 통해 2차 테일러 전개 형태가 오차가 더 작은 것을 알 수 있다.

$$f(x) = \sum_{n=0}^{m} \frac{f^{(n)}}{n!}(x-a)^n + R(x;a)$$

$$R(x;a) = \frac{1}{(m+1)!}f^{(m+1)}(a)(x-a)^{m+1} + \cdots$$





- 이러한 테일러 전개를 이용해서 통계학에서 많이 쓰는 감마함수를 근사(Stirling's formula)하기도 하고 적률(평균, 분산 등)을 생성(MGF)할 수도 있다.

- 이를 응용하면 아래와 같이 $E[f(X)]$를 테일러 전개를 통해 근사할 수 있다.

$$E[f(X)]$$
$$\approx E[f(\mu_X) + f'(\mu_X)(X-\mu_X) + \frac{1}{2}f''(\mu_X)(X-\mu_X)^2]$$
$$= f(\mu_X) + f'(\mu_X)E[X-\mu_X] + \frac{1}{2}f''(\mu_X)E[(X-\mu_X)^2]$$

- $E[X-\mu_X] = 0$이고 $E[(X-\mu_X)^2] = \sigma_X^2$이므로 정리하면

$$E[f(X)] \approx f(\mu_X) + \frac{f''(\mu_X)}{2}\sigma_X^2$$

- 원래는 $E[f(X)]$를 구하기 위해 Montecarlo Integration 같은 수치적분을 활용해야 하는데, 함수와 도함수에 평균, 분산을 대입하는 방식을 통해 계산을 효율적으로 할 수 있다.

편미분(Partial Differential)과 변수 추적

- 다른 변수들이 고정된 상황에서 변수 효과 확인

Discrete Case – 밀의 차이법, 이중차분법

	a	b	c	Y
철수	O	O	O	O
영희	O	O	X	X

	이전	이후
처치	50	85
통제	35	55

- 왼쪽 표에서 철수와 영희의 차이는 c의 처치 유무이고, Y에서 결과 차이가 발생하였다. 따라서 c가 Y의 원인이라고 말할 수 있다. 이렇게 차이를 통해 인과를 논증하는 방식을 **차이법**이라고 한다. 이때 시간의 효과까지 고려한다면 **이중차분법**을 사용할 수 있다.
- 오른쪽 표의 값은 어떤 쇼핑몰의 집단 별 평균 구매 액수이다. 이때 처치의 효과에 대해 알아보자.

(1) 처치 집단의 처치 효과

$$85 - 50 = 35$$

- 여기서는 시간으로 인한 효과가 포함되므로

(2) 통제 집단을 통해 추세 추정

$$55 - 35 = 20$$

(3) 이중 차분을 통해 처치 효과 추정

$$(85 - 50) - (55 - 35) = 15$$

- 즉, 최종적으로 추정한 효과는 15이다.
- 단, 이중차분법을 사용할 때는 두 집단이 동일한 추세를 따라야한다는 가정이 들어간다.

Continuous Case – 편미분

- 독립변수가 연속적인 경우에는 독립변수가 한 단위 변화할 때 종속변수가 얼만큼 변하는 지에 대해 관심을 갖는다. 이때 인과를 논증하기 위해서는 다른 변수들이 고정되어 있어야 한다.
- 수학에서는 다른 변수들을 상수 취급하고 관심 있는 변수로 미분하는 **편미분**을 통해 이를 수행할 수 있다.

$$\lim_{h \to 0} \frac{f(x+h, y) - f(x, y)}{h} = \frac{\partial f}{\partial x}$$

- 아래의 두 재화 X, Y에 따른 효용 함수를 편미분을 통해 분석하면,

$$\pi(X, Y) = 50X - 4Y^2 + 75Y$$

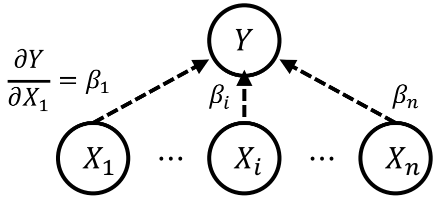
$$\frac{\partial \pi}{\partial X} = 50, \frac{\partial \pi}{\partial Y} = -8Y + 75$$

- X재의 경우 한 단위 증가할 때마다 50만큼의 효용이 증가하고, Y재의 경우는 구매할수록 점차 얻는 효용(한계효용)이 줄어든다. (두 재화의 가격이 똑같다면 이 사람은 세 단위까지는 Y를 구매하겠지만 이후부터는 X만을 구입할 것이다.)

편미분(Partial Differential)과 변수 추적

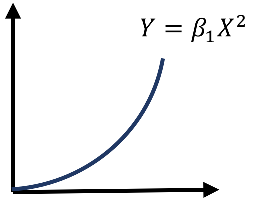
- 다른 변수들이 고정된 상황에서 변수 효과 확인

- 많은 모델링에서 선형 모델(일차 함수)을 가정하는 경우가 많다. 그 이유는 어떤 지점에서나 변수들의 효과가 똑같기 때문이다.



$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$$

- 이차 함수를 예시로 비선형 함수에서 미분을 통해 변수들의 효과를 구해보면,

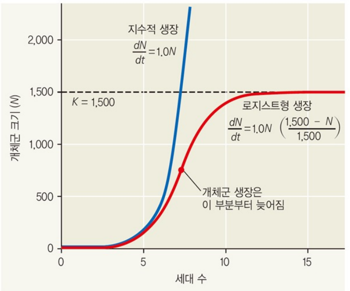


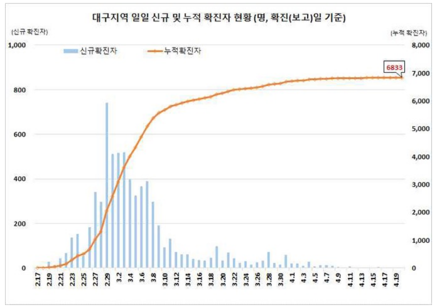
$$Y = \beta_1 X^2$$
$$\frac{\partial Y}{\partial X_1} = \boxed{2\beta_1}\,\boxed{X}$$

효과의 크기 위치

- 위 그림과 같이 비선형 함수에서는 X의 값, 즉 위치에 따라 변수의 효과가 다르게 나타난다. 예를 들어, 자유 낙하하는 물체가 이동한 거리는 0초에서 10초까지 이동한 거리와 10초에서 20초까지 이동한 거리가 다르므로 이런 경우 시점이 이동한 거리에 영향을 끼치게 된다.

- 아래 그림과 같이 인구, 확진자 수와 같은 집단의 경우 지수함수와 로지스틱 함수를 이용해 모델링할 수 있다. 이론적으로는 시간에 따라 지수적으로 개체 수가 늘어나지만 (출산율이 2가 넘는 경우, 바이러스의 확산), 현실에서는 자원(토지, 미감염자 수)이 한정되어 있으므로 로지스틱 함수와 비슷하게 증가한다.





- 로지스틱 함수를 미분해보면,

$$f(x) = \frac{1}{1 + e^{-\beta_1 x}}$$

$$f'(x) = \beta_1 e^{-\beta_1 x} \cdot \left(1 + e^{-\beta_1 x}\right)^{-2}$$
$$= \beta_1 f(x) \cdot (1 - f(x))$$

- 즉, 집단의 성장 ($f'(x)$)은 집단이 크면 클수록 더 커짐 ($f(x)$) 과 동시에 집단의 크기가 너무 커질 경우 자원 요구에 압박을 받아 성장($1 - f(x)$) 에 제한을 받는 형태이다. 이 경우, β_1은 성장 속도를 의미하게 된다.

전미분(Total Differential)과 변수 추적

- 다른 변수들이 고정되지 않은 상황에서의 변화량

- 그러나 모든 상황에서 변수가 고정되는 것은 아니기 때문에 한 독립변수가 변화할 때 다른 독립 변수도 변화하는 상황을 고려해보자.

- (허니버터칩을 사기 위해서 다른 미끼 상품을 사야 했던 것처럼) X, Y재는 더 이상 따로 구입할 수 없고, 두 재화를 함께 구매할 수 밖에 없는 상황에서 두 재화의 세트의 개수를 T로 두고 효용의 변화량을 구해보자.

- T가 한 단위 증가한다는 것은 X와 Y 모두 한 단위 증가하는 것이므로,

$$\frac{d\pi}{dT} = \frac{d\pi}{dX} + \frac{d\pi}{dY} = 50 + (-8Y + 75) = 125 - 8T$$

- 이 경우 16개부터는 T의 한 단위가 증가하더라도 효용이 감소하므로 15개까지 구매할 것이다. 만약 세트의 구성이 X재 한 개와 Y재 두 개의 경우라면, 변화량은 어떻게 변할까?

$$\frac{d\pi}{dT} = \frac{d\pi}{dX} + 2\frac{d\pi}{dY} = 50 + 2(-8Y + 75) = 200 - 16T$$

- 일반화하면, X와 Y에 대한 T의 변화율이 효용함수에 영향을 주므로,

$$\frac{d\pi}{dT} = \frac{\partial\pi}{\partial X}\frac{dX}{dT} + \frac{\partial\pi}{\partial Y}\frac{dY}{dT}$$

- 즉, **전미분**이란 다변수 함수의 모든 변수의 변화에 따라 변화하는 다변수 함수의 형태를 근사하는 양이며 기호로는 다음과 같이 나타난다.

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i$$

- 이때까지는 한 사람에게 두 재화 X,Y가 영향을 미치는 것을 고려했는데 $(\mathbb{R}^2 \to \mathbb{R})$ 이제 두 사람에게 동시에 영향을 미치는 것을 고려해보자. $(\mathbb{R}^2 \to \mathbb{R}^2)$

$$\pi_1(X, Y) = 50X - 4Y^2 + 75Y$$
$$\pi_2(X, Y) = 20X - 4XY + 35Y$$

$$\frac{\partial\pi_1}{\partial X} = 50, \frac{\partial\pi_1}{\partial Y} = -8Y + 75$$
$$\frac{\partial\pi_2}{\partial X} = 20 - 4Y, \frac{\partial\pi_2}{\partial Y} = -4X + 35$$

- X, Y재의 변화량이 두 사람의 효용함수에 영향을 미치는 것을 표현해야 하는데, 2x2의 행렬을 이용해서 쉽게 표현할 수 있다.

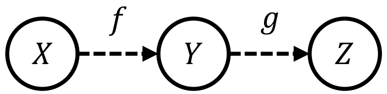
$$d\pi = \begin{bmatrix} d\pi_1 \\ d\pi_2 \end{bmatrix} = \begin{bmatrix} \frac{\partial\pi_1}{\partial X} & \frac{\partial\pi_1}{\partial Y} \\ \frac{\partial\pi_2}{\partial X} & \frac{\partial\pi_2}{\partial Y} \end{bmatrix} \begin{bmatrix} dX \\ dY \end{bmatrix} = \begin{bmatrix} 50 & -8Y + 75 \\ 20 - 4Y & -4X + 35 \end{bmatrix} \begin{bmatrix} dX \\ dY \end{bmatrix}$$

연쇄법칙(Chain Rule)

- 여러 선후관계 속에서 변수 효과 확인

One Variable Case (Scalar Function)

- 이때까지는 독립변수가 종속변수에 직접적으로 영향을 미치는 경우를 살펴보았다. 그러나 X가 Y에 영향을 미치고, 다시 Y가 Z에 영향을 미치는 경우를 보자.



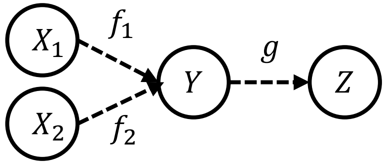
- 예를 들어, 코카콜라의 원자재 값이 올랐다고 하자. 이로 인해 코카콜라의 가격이 올랐고, 경쟁사인 펩시를 찾는 경우가 많아졌다. 이때 코카콜라 원자재 1원 만큼 올랐을 때, 펩시를 찾는 사람들은 얼마나 많아졌는지 구해보자.

- 원자재 비용 인상이 가격에 미치는 영향은 $\frac{df}{dx}$로 구할 수 있다. 마찬가지로 코카콜라의 가격이 펩시에 수요에 미치는 영향은 $\frac{dg}{dy}$를 통해 구할 수 있다. 즉, 코카콜라 비용 1원 이상은 $\frac{df}{dx}$원의 가격 인상을 가져오고, 1원의 가격 인상은 $\frac{dg}{dx}$명의 수요를 증가시킨다.

$$\frac{dZ}{dX} = \frac{df \circ g}{dX} = \frac{df}{dX} \times \frac{dg}{dY} = \frac{df}{dX} \times \frac{dg}{df}$$

Multi Variable Case (Vector Function)

- 이제 여러 원인이 매개 변수를 거쳐 결과에 영향을 주는 경우를 살펴보자.



- 이동에 따른 기체의 부피를 구하는 문제에서 기체의 좌표를 (위도, 고도)라고 하자. 이때 위도가 높아질수록(북극, 남극과 가까워질수록) 기체의 온도가 감소하고, 마찬가지로 고도의 변화도 기체의 온도에 영향을 미친다. 또한 기체의 온도와 부피는 반비례한다. 이제 편미분과 전미분을 이용해 이동과 부피 사이의 관계를 구해보자.

① 위도와 부피 사이의 관계: $\frac{\partial Z}{\partial X_1} = \frac{\partial Z}{\partial Y} \times \frac{\partial Y}{\partial X_1} = g'(f_1(x)) \times f_1'(x)$

② 고도와 부피 사이의 관계: $\frac{\partial Z}{\partial X_2} = \frac{\partial Z}{\partial Y} \times \frac{\partial Y}{\partial X_2} = g'(f_2(x)) \times f_2'(x)$

③ 위도,고도와 부피 사이의 관계: $dZ = \frac{\partial Z}{\partial X_1} dX_1 + \frac{\partial Z}{\partial X_2} dX_2$

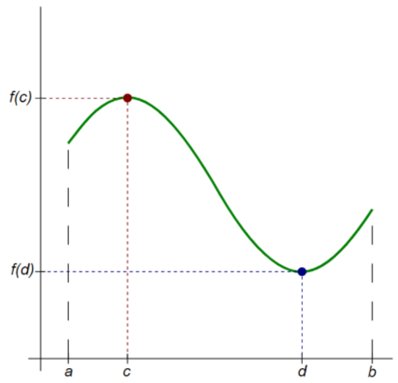
일변수함수의 최대·최소

- 최적화에 들어가기에 앞서..

미분을 활용하여 해결할 수 있는 분야 중의 하나는 최댓값, 혹은 최솟값을 찾는 일이다. 그런데 어떻게 미분이 최댓값과 최솟값을 찾아줄 수 있는 것일까?

정해진 영역에서 함수 f(x)의 최댓값과 최솟값과 최대-최소 정리를 알아보자

- 함수 f(x)가 집합 I 에서 정의되어 있고, I 내의 모든 x에 대하여 $f(x) \leq f(d)$를 만족하는 d가 I 내에 존재하면 $f(d)$를 I 위에서 $f(x)$의 최댓값이라 한다. 같은 방법으로, I 내의 모든 x에 대하여 $f(x) \geq f(c)$를 만족하는 c가 I 내에 존재하면 $f(c)$를 I 위에서 $f(x)$의 최솟값이라 한다.
- 함수 $f(x)$가 폐구간 $[a, b]$ 에서 연속이라 하자. 그러면 f는 폐구간 $[a, b]$ 위에서 최댓값과 최솟값을 갖는다.



이때 일변수함수의 최댓값과 최솟값은 폐구간의 양 끝 점(a ~or ~ b) 혹은 임계점 ($f'(x_0) = 0$ 또는 존재하지 않음)의 함숫값 중에서 찾을 수 있다. 임계점의 함숫값에 해당할 수 있는, 극값의 정의를 간단하게 살펴보자.

- $\delta > 0$ 이 존재하여 구간 $[c-\delta, c+\delta]$에서 f의 최댓값이 $f(c)$이면 함수 f는 c에서 극댓값(relative maximum value) $f(c)$를 갖는다고 한다. 같은 방법으로 $\delta > 0$ 이 존재하여 구간 $[c-\delta, c+\delta]$에서 f의 최솟값이 $f(c)$이면 함수 f는 c(relative minimum value) $f(c)$를 갖는다. 극댓값과 극솟값을 통틀어 극값이라 한다.

- 이제 일변수함수의 최대와 최소를 요약하면 다음과 같다

1. 최적화에 사용되는 중요한 문제이다.
2. 최대 최소 정리에 의해서 함수 f 는 폐구간에서 최댓값과 최솟값을 가진다.
3. 최댓값과 최솟값의 후보는 구간의 양 끝점이나 임계점의 함숫값이다.

즉 미분값이 0이 되는 지점이 최적 값이 될 가능성이 있기 때문에 미분을 이용하게 되는 것이다.

다변수함수의 최대·최소

- 최적화에 들어가기에 앞서..

앞에서 다룬 일변수함수의 최대·최소를 확장해 다변수함수의 최대·최소를 이해할 수 있다.

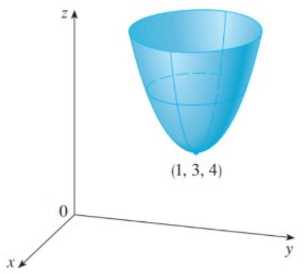


FIGURE 2
$z = x^2 + y^2 - 2x - 6y + 14$

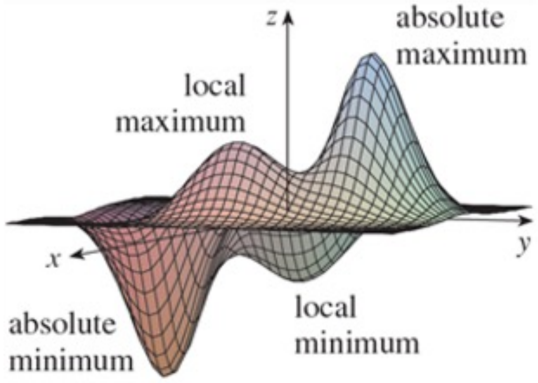
- $z = f(x, y)$를 영역 R 상에서 정의된 이변수함수라 하자. $f(x, y)$ 가 점 (a,b)의 적당한 근방 D 내의 모든 점 (x, y) 에 대하여 $f(a, b) \geq f(x, y)$일 때, 점 (a,b)에서 극댓값을 갖는다고 한다. 지금 $f(x, y)$ 가 점 (a,b)에서 극값을 갖는다면,

- $g(x) = f(x, b), h(y) = f(a, y)$라 할 때 $g(x)$는 a에서 극값을 갖고, $h(y)$는 b에서 극값을 갖는다. 따라서 만약 $f_x(a, b)$, $f_y(a, b)$가 존재하면,

$$f_x(a, b) = g'(a) = 0, \quad f_y(a, b) = h'(b) = 0$$

그러므로 $f(x, y)$의 극값은 $f(x, y)$의 1계 편도함수 모두가 0이 되는 점 또는 1계 편도함수가 존재하지 않는 점에서만 갖는다.

일변수함수의 예시에서 유추 가능하듯이, 최댓값과 최솟값의 후보는 경계점이나 임계점의 함숫값이다. 최적화 문제에서는 경계점보다 임계점을 통해 최대와 최소를 구하는 경우가 흔하다.



다변수함수의 최대·최소

- 최적화에 들어가기에 앞서..

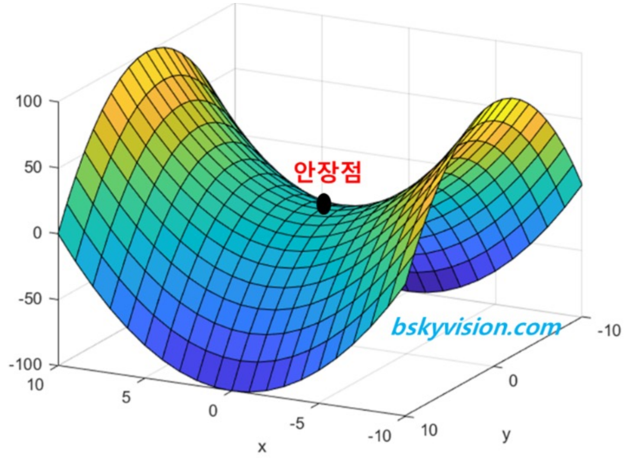
- 다변수함수의 최대와 최소를 요약하면 다음과 같다

1. 일변수함수와 유사하게, 다변수함수 역시 함수 f의 임계점 혹은 R의 경계점에서 최대와 최소를 찾을 수 있다.
2. 결국 다변수함수의 최대 최소 문제(최적화 문제) 역시, 기본적으로는 미분을 통해 임계점을 구하고 비교하여 해결한다.

하지만 모든 임계점의 값이 극대 혹은 극소를 갖는 것은 아니다. 일변수함수의 경우, 임계점은 극댓값 또는 극솟값에 해당하거나 둘 다 아닐 수도 있다. 이는 다변수함수 역시 마찬가지인데, 어느 한 축에서는 극대에 이르면서 다른 축에서는 극소에 이르기 때문에 방향에 따라 극대로, 혹은 극소로 보이는 지점이 존재하게 된다. 일변수함수에서 해당 지점은 변곡점이 되지만, 다변수함수에서는 변곡점 혹은 안장점이 된다.

- 안장점

이변수함수에서 함수의 양쪽 방향에서 동시에 기울기가 0이 되어 극대와 극소 모두에 해당하지 않는 임계점(critical point)



안장점은 헤시안 행렬(Hessian Matrix)를 통해 계산될 수 있는데, 이는 다음 시간에 알아보도록 하자.