04.01 함수

04.01 함수

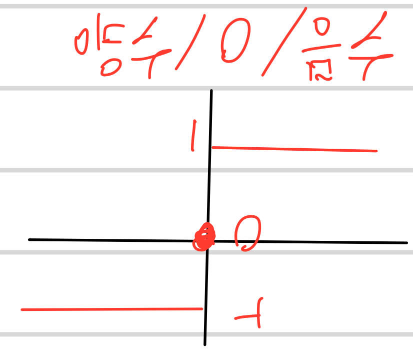
0) 연속과 불연속

불연속 : 함수의 값이 중간에 갑자기 변하는 것(discontinuous)

- 부호함수, 단위계단함수, 지시함수

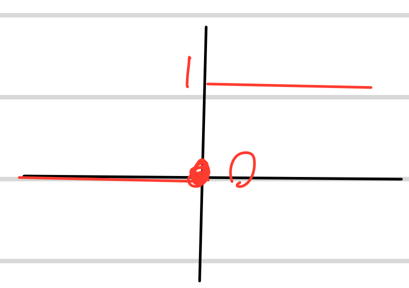
연속 : 그렇지 않은 것

1. 입력이 1개인 경우



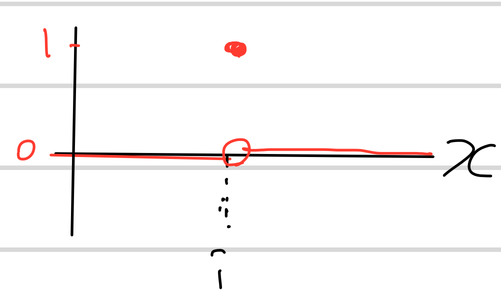
1) 부호함수

- 입력이 양수이면 1, 음수이면 -1, 0이면 0 출력
- np.sign()
- 표시 : sgn(x)



2) 단위계단함수

- 양수이면 1, 아니면 0 출력
- 직접 구현(넘파이 구현x)
- H(x)



3) 지시함수(indicator function)

- 1 (if x==i)
- 0 (if x!=i)

04.01 함수

1. 데이터분석에서 많이 사용되는 함수들

1) 다항함수 (polynomial)

2) 최대함수와 최소함수
- ReLU (Rectified Linear Unit)

3) 지수함수
- 단조증가 X1>X2 이면, exp(X1)>exp(X2)

4) 로지스틱함수
- 시그모이드 함수 중 하나 (로지스틱함수가 제일 많이 쓰임)
- 그래프 그려보기

5) 로그함수
- 단조증가
- 확률론에서 가장 많이 사용되는 함수
  1) 로그함수는 곱하기를 더하기로 변환함 logAB = logA + logB
  2) 어떤 함수에 로그를 적용해도 함수의 최고점, 최저점의 위치는 변하지 않음 (최적화 시, 로그 취한 함수에 대해 최적화 가능)
  3) 로그함수는 0부터 1사이의 작은 값을 확대시켜 보여준다.

6) 소프트플러스 함수
- 지수함수와 로그함수의 결합(ReLU와 유사)

2. 입력이 여러개인 경우

04.01 함수

7) 다변수 함수

- z = f(x,y) (높이까지, z)

- 2차원 함수는 평면상의 지면과 같기 때문에, 3차원의 서피스 플롯(surface plot), 컨투어 플롯(contour, 등고선 plot) 으로 표현

8) 분리가능 다변수 함수

- 단변수 함수의 곱으로 표현 가능한 다변수 함수

= 분리가능 다변수 함수(단변수로 분리가능하다는 의미)

- f(x,y) = f1(x)f2(y)

3. 다변수 다출력 함수

9) 입력과 출력모두 여러개 <== 이 경우, 출력을 벡터나 행렬로 나타낼 수 있음

ex) 소프트맥스 함수

- 소프트맥스 함수 : 다차원 벡터를 입력받아, 다차원 벡터를 출력

- 소프트맥스 함수 특징

  1) 지수함수로 다 감싸고 있기 때문에, 출력값은 항상 양수

  2) 각 출력값은 0과 1사이이고, 모든 출력(출력값이 여러개, 벡터를 이룸)의 합은 1이다. => 확률이 아닌 것을 확률처럼 보이게 함

4. 함수의 평행이동, 스케일링

10) 단변수함수 (f(x) = (x+2)**2 + 2x) 를 오른쪽으로 a만큼, 위로 b만큼 평행이동 시키기

11) x축 방향으로 a배 만큼 늘리기 / y축 방향으로 b배 만큼 늘리기

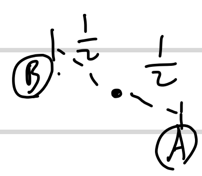
(ex_로지스틱함수)

04.02 심파이를 사용한 함수 미분

1) 미분 목적 1 : 예측 모형의 성능
- 오류의 크기 출력하는 함수
  손실함수(loss function), 비용함수(cost function), 오차함수(error function)
  ==> 최적화의 대상이 되는 함수(목적함수)
    *최적화 : 목적함수의 max값, min값을 만들어내는 모수(입력값)을 구하는 것 (미분을 통해서)

2) 기울기 (민감도, sensitivity, slope)

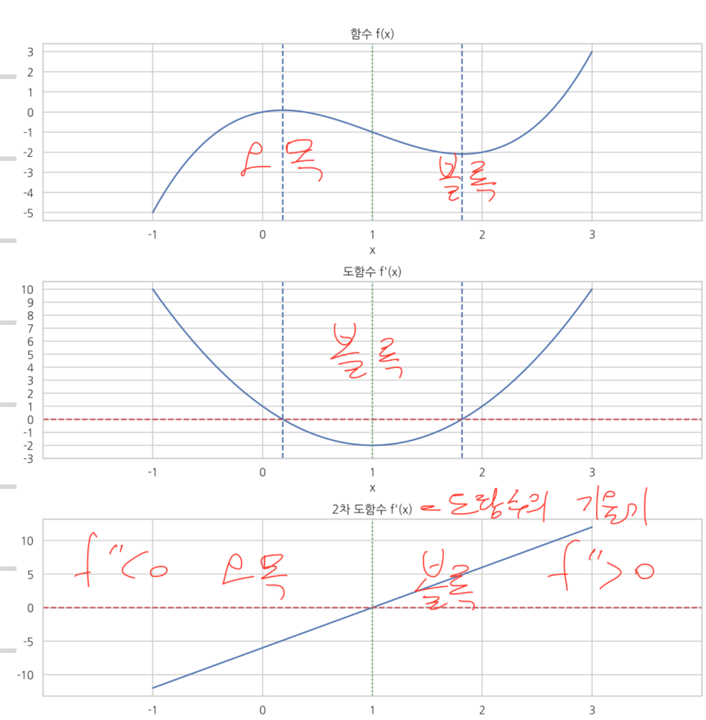
3) 수치미분



Ⓐ계산, Ⓑ계산 → 기울기 구하기

4) 미분
- 도함수(f')와 미분(f'(x))
- 미분 가능, 미분 불가능
  ex) ReLU함수 -> 미분 불가능 (대안 : softplus 함수)
  - 연속의 기준 : 함수값 = 우극한값 = 좌극한값
  - 미분가능 기준 : 좌미분계수(h -> +0) -0 = 우미분계수(h -> -0) +0

- 상수, 거듭제곱, 로그함수, 지수함수 미분



5) 미분 곱셈법칙, 체인룰

6) 2차도함수와 Convexity
- 2차도함수 > 0 <==> 볼록
- 2차도함수 = 도함수의 기울기, 도함수값이 증가하면 2차도함수(기울기) 값은 양수, 감소하면 2차도함수(기울기) 값은 음수
- 04.02 13page, 15page

04.02 심파이를 사용한 함수 미분

7) 편미분

- 편미분 시에는, 델타 대신 라운드를 사용해 표기

8) 이차 편미분

- (슈와르츠 정리) f 함수가 연속 + 미분 가능하면, fxy = fyx 이다. 미분의 순서는 상관 없다.
- ex) ReLU함수는 미분불가능 하기 때문에 해당 안됨.

9) 테일러 정리 (=테일러 전개, 테일러 급수)

- 테일러 정리 = 한 점에서의 직선의 방정식
- 함수를 근사할 때 사용됨 (테일러 정리 = Taylor approximation)

- 우리가 잘 모르거나 복잡한 함수를 테일러 전개를 통해 비교적 다루기 쉬운 다항함수로 대체하기 위해 쓰인다. 이렇게 어떤 함수를 테일러 급수로 바꿔 표현하면, 그 함수의 특성을 분석/이해하기 더 쉬운 경우가 있기 때문에 종종 쓰이는 것

- 모든 함수를 다 테일러 급수로 표현할 수 있는건 아니고 f(x)가 x=x0인 곳에서 미분 가능하면 저렇게 표현하는 것이 가능하다.

즉, 함수가 모든 구역에서 미분될 필요 없고, anchor point(x0) 에서만 미분되어도 테일러 전개로 근사할 수 있다는 것이다.

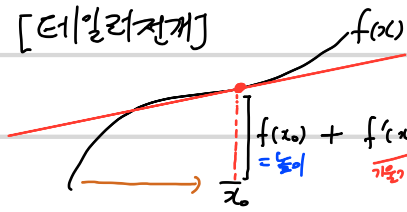
"테일러 급수를 활용하면 한 점에서만 미분 가능해도 함수값을 구할 수 있다"는 것은 아주아주 매력적이다.

04.02 심파이를 사용한 함수 미분

- 테일러 급수 식

$$f(x) \approx f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)(x-x_0)^2}{2!} + \cdots$$

- 테일러 급수 식 이해



$$\underbrace{f(x_0)}_{=\text{높이}} + \underbrace{f'(x_0)}_{\text{기울기}} \times \underbrace{(x-x_0)}_{\text{평행이동}} + \frac{f''(x_0)}{2!}(x-x_0)^2 + \cdots \doteqdot b + a(x-x_0) \quad \text{at } x_0$$

- 테일러 급수 식 (다변수 함수 경우)

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

- 맥클로린 급수 식

$x_0 = 0$ 일 때, 가정

$$f(x) \approx f(0) + f'(0)x + \frac{f''(0)x^2}{2!} + \cdots$$

10) sympy - symbolic 연산

- x라는 문자를 문자 그대로 인정하는 것(symbolize)
- x = 2 처럼, 문자에 숫자 대입하면 망가짐

04.03 적분

04.03 적분

1) 부정정분

- 부정적분 : anti-derivative
- 편미분의 부정적분

  4.3.6

- 다차도함수와 다중적분

  4.3.10

2) sympy 이용한 부정적분

3) 정적분

- 면적을 구하는 것 (부정적분은 함수를 구하는 과정. 정적분은 면적(수치)를 구하는 과정)
- 4.3.14

4) 다변수 정적분

- 4.3.18

5) 다차원 함수의 단일 정적분

- 4.3.22

04.04 행렬의 미분

- 스칼라, 벡터, 행렬 모두 입력과 출력에 사용가능

- 행렬 미분 : 행렬을 입력이나 출력으로 갖는 함수를 미분하는 것
  (정확하게는 편미분을 하는 것)
  - 분모중심 표현법(Denominator-layout notation)을 활용할 것

1) *스칼라를 벡터로 미분하는 경우*
  - 다변수 함수 : 벡터 입력 --> 스칼라 출력 (f = 스칼라, 입력변수인 벡터로 결과값을 미분(스칼라를 미분), 편미분)
  - 다변수 함수의 미분 : 기울기가 변수의 갯수만큼 생성(편미분)
  - 스칼라의 벡터미분 : 그레디언트 벡터(기울기로 이루어진 벡터

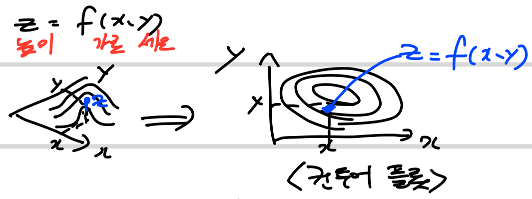
$$\nabla f = \frac{\partial f}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_N} \end{bmatrix}$$

2) 컨투어플롯, 퀴버플롯, 그레디언트 벡터
  - 컨투어플롯 : 등고선, 2차원함수의 평면상 표기 플롯
  - 퀴버플롯 : 벡터로 현 위치에서 경사 높은 방향 보여줌
    (컨투어플롯 위에 그레디언트 벡터를 화살표로 나타낸 것)
  - 그레디언트 벡터 : 다변수 함수의 출력값(스칼라)를 벡터로 미분해서 얻은 벡터
    (각 원소는 1차 편미분값. 기울기. 현재 위치의 기울기. 증감 여부)

변수 갯수만큼
기울기 생성
=> 그레디언트 벡터

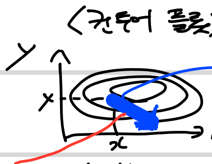
<컨투어 플롯> 등고선 표기
: 2차원 함수의 평면 상 표기 플롯

ex) 2차원 함수 $z = f(x, y)$
높이 가로 세로



<컨투어 플롯>

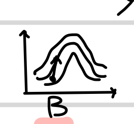
<퀴버플롯>
: 컨투어플롯 위에
그레디언트 벡터를
화살표로 나타낸 플롯 $\nabla f = \begin{bmatrix} \frac{\partial f(x,y)}{\partial x} \\ \frac{\partial f(x,y)}{\partial y} \end{bmatrix}$

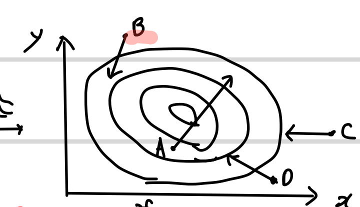


x, y 위치에서의
그레디언트 벡터
∇f
가로성분(벡터1원소)
$= \frac{\partial f}{\partial x}$
세로성분(벡터2원소)
$= \frac{\partial f}{\partial y}$

신치로
반대방향가리킬것.

<퀴버플롯>





$\nabla f(B) = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$ = 단위 당 f값 한계 증가가 가장 큰 방향

$\nabla f(A) = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$

=> 등고선의 방향과 직교!

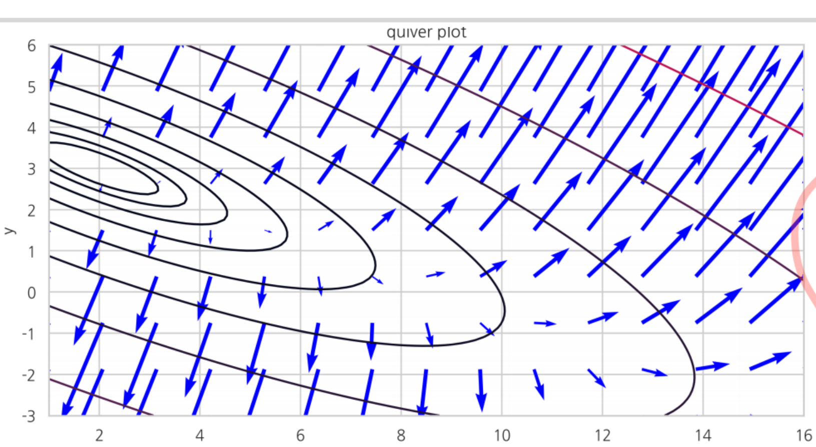
$\nabla f(D) = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix} = \begin{pmatrix} -2 \\ 3 \end{pmatrix}$

$\nabla f(C) = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$

(그레디언트 벡터 방향 = 산에서 발을 디딜 때 경사가 가장 높은 방향)
(함수값=높이)

  - 퀴버플롯의 역할 : 퀴버플롯을 보고 실제 곡면의 모양을 유추할 수 있어야함

*4page필기



<곡면의 모양>
=> '분지' 일 것

04.04 행렬의 미분

[주요 공식 정리] (04.04 11page)

1. $\nabla f = \dfrac{\partial f}{\partial x_i}$

2. $J f(x) = \begin{bmatrix} -\nabla f_1^T- \\ -\nabla f_2^T- \\ \vdots \\ -\nabla f_m^T- \end{bmatrix}$

3. $H = \dfrac{\partial^2 f}{\partial x_i \partial x_j}$

<정리>

$w^T x$	벡터 x → 스칼라	$(w^T)^T$
$x^T A x$	벡터 x → 스칼라	$(A+A^T)x$
Ax	벡터 x → 벡터	$(A)^T$
$tr(AX)$	행렬 X → 스칼라	$(A)^T$
$\log\lvert X\rvert$	행렬 X → 스칼라	$(X^{-1})^T$

[스칼라를 벡터로 미분 : 다변수함수 (선형모형, 이차형식)]

1) 선형모형 (W.T*x)

- f : 스칼라 / input : 벡터x / f를 벡터로 미분 (스칼라를 벡터로 미분)
- 04.04 5page

행렬미분법칙 1: 선형 모형

선형 모형을 미분하면 그레디언트 벡터는 가중치 벡터이다.

· 벡터 x → 스칼라 f

· $\frac{\partial f}{\partial x} = \nabla f$

$$f(x) = w^T x \tag{4.4.25}$$

$$\nabla f = \frac{\partial w^T x}{\partial x} = \frac{\partial x^T w}{\partial x} = w \tag{4.4.26}$$

(증명)

상수

$$\frac{\partial (w^T x)}{\partial x} = \begin{bmatrix} \boxed{\frac{\partial (w^T x)}{\partial x_1}} \\ \boxed{\frac{\partial (w^T x)}{\partial x_2}} \\ \vdots \\ \boxed{\frac{\partial (w^T x)}{\partial x_N}} \end{bmatrix} = \begin{bmatrix} \boxed{\frac{\partial (w_1 x_1 + \cancel{w_2 x_2} + \cdots + \cancel{w_N x_N})}{\partial x_1}} \\ \boxed{\frac{\partial (\cancel{w_1 x_1} + w_2 x_2 + \cdots + \cancel{w_N x_N})}{\partial x_2}} \\ \vdots \\ \boxed{\frac{\partial (\cancel{w_1 x_1} + \cancel{w_2 x_2} + \cdots + w_N x_N)}{\partial x_N}} \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix} =$$

2) 이차형식 (x.T A x)

- f : 스칼라 / input : 벡터x / f를 벡터로 미분 (스칼라를 벡터로 미분)
- 04.04 6page, 8page

행렬미분법칙 2: 이차 형식

이차 형식을 미분하면 행렬과 벡터의 곱으로 나타난다.

· 벡터 x → 스칼라 f

$$f(x) = x^T A x \tag{4.4.28}$$

$$\nabla f(x) = \frac{\partial x^T A x}{\partial x} = (A + A^T)x \tag{4.4.29}$$

[벡터를 스칼라로 미분 : 여러 함수에 입력]

- 각 벡터의 원소를 스칼라로 미분 (표시는 행벡터로 표시. 그레디언트벡터와 헷갈리지 않도록)
- 04.04 8page

벡터를 스칼라로 미분하는 경우

벡터

스칼라 벡터

x_1 → f_1, f_2, f_3

$$f(x) = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_M \end{bmatrix} \tag{4.4.35}$$

를 스칼라 x로 미분하는 경우에는 결과를 행 벡터로 표시한다.

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_2}{\partial x} & \cdots & \frac{\partial f_M}{\partial x} \end{bmatrix} \tag{4.4.36}$$

[벡터를 벡터로 미분]

- 04.04 8page

벡터를 벡터로 미분하면 미분을 당하는 벡터의 원소가 여러개($i = 1, \ldots, N$)이고 미분을 하는 벡터의 원소도 여러개($j = 1, \ldots, M$)이므로 미분의 결과로 나온 도함수는 2차원 배열 즉, 행렬이 된다.

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_2}{\partial x} & \cdots & \frac{\partial f_N}{\partial x} \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_M} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} & \cdots & \frac{\partial f_N}{\partial x_1} \\ \frac{\partial f_1}{\partial x_2} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_N}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial x_M} & \frac{\partial f_2}{\partial x_M} & \cdots & \frac{\partial f_N}{\partial x_M} \end{bmatrix} \tag{4.4.37}$$

04.04 행렬의 미분

3) 행렬과 벡터의 곱(Ax)의 미분 : 벡터를 벡터로 미분

- 04.04 9page

행렬미분법칙 3: 행렬과 벡터의 곱의 미분

행렬 A와 벡터 x의 곱 Ax를 벡터 x로 미분하면 행렬 A^T가 된다.

$$f(x) = Ax \tag{4.4.38}$$

$$\nabla f(x) = \frac{\partial (Ax)}{\partial x} = A^T \tag{4.4.39}$$

(증명)

$$Ax = c_1 x_1 + c_2 x_2 + \cdots + c_M x_M \tag{4.4.40}$$

$$\frac{\partial (Ax)}{\partial x} = \begin{bmatrix} \frac{\partial (Ax)}{\partial x_1} \\ \frac{\partial (Ax)}{\partial x_2} \\ \cdots \\ \frac{\partial (Ax)}{\partial x_M} \end{bmatrix} = \begin{bmatrix} \frac{\partial (c_1 x_1 + c_2 x_2 + \cdots + c_M x_M)^T}{\partial x_1} \\ \frac{\partial (c_1 x_1 + c_2 x_2 + \cdots + c_M x_M)^T}{\partial x_2} \\ \cdots \\ \frac{\partial (c_1 x_1 + c_2 x_2 + \cdots + c_M x_M)^T}{\partial x_M} \end{bmatrix} = \begin{bmatrix} {c_1}^T \\ {c_2}^T \\ \cdots \\ {c_M}^T \end{bmatrix} = A^T \tag{4.4.41}$$

다차원함수의 1차미분 (기울기, ∇)

- 자코비안 행렬 : 벡터함수를 벡터로 미분해서 생기는 행렬의 전치행렬

    -벡터를 벡터로 미분 = 행렬

(04.04 9page)

함수의 출력변수와 입력변수가 모두 벡터(다차원) 데이터인 경우에는 입력변수 각각과 출력변수 각각의 조합에 대해 모두 미분이 존재한다. 따라서 도함수는 행렬 형태가 된다. 이렇게 만들어진 도함수의 행렬을 **자코비안 행렬(Jacobian matrix)** 이라고 한다. 자코비안 행렬은 벡터함수를 벡터변수로 미분해서 생기는 행렬의 **전치행렬**이다. 따라서 행/열의 방향이 다르다는 점에 유의한다.

$$Jf(x) = J = \left(\frac{\partial f}{\partial x}\right)^T = \begin{bmatrix} \left(\frac{\partial f_1}{\partial x}\right)^T \\ \vdots \\ \left(\frac{\partial f_M}{\partial x}\right)^T \end{bmatrix} = \begin{bmatrix} \nabla f_1^T \\ \vdots \\ \nabla f_M^T \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_M}{\partial x_1} & \cdots & \frac{\partial f_M}{\partial x_N} \end{bmatrix} \tag{4.4.42}$$

$Jf(x)$

=> 그레디언트 벡터들 만들어라!

$$\left(\frac{\partial f_M}{\partial x}\right)^T = \nabla f_M^T = \begin{bmatrix} \frac{\partial f_M}{\partial x_1} \\ \frac{\partial f_M}{\partial x_2} \\ \vdots \\ \frac{\partial f_M}{\partial x_N} \end{bmatrix}^T$$

$$J = \left(\frac{\partial f}{\partial x}\right)^T = \begin{bmatrix} -\nabla f_1^T- \\ \vdots \\ -\nabla f_M^T- \end{bmatrix}$$

자코비안 = (∇ 모음)^T

다차원함수의 2차미분 (곡률)

- 헤시안 행렬 : 다변수함수의 2차 도함수 (대칭행렬 <==> 실수 고유값)

    -벡터를 벡터로 미분 = 행렬

(04.04 10page)

다변수함수

벡터 x $\xrightarrow{\text{입력}}$ 스칼라 f

↓ 1차미분 (벡터 x로)

$\nabla f \longrightarrow J$

↓ 2차미분

$H = \nabla^2 f$ or $\nabla\nabla f$

: 헤시안행렬

- 대칭행렬 (고유값 실수) $H_{ij} = H_{ji}$

- 다변수함수의 2차도함수 수학으로

- 스칼라를 행렬로 미분

    - 행렬 X의 각 원소들로 스칼라를 미분해주면 그만

    (04.04 10page)

출력변수 f가 스칼라값이고 입력변수 X가 행렬인 경우에는 도함수 행렬의 모양이 입력변수 행렬 X와 같다.

$$\frac{\partial f}{\partial X} = \begin{bmatrix} \frac{\partial f}{\partial x_{1,1}} & \frac{\partial f}{\partial x_{1,2}} & \cdots & \frac{\partial f}{\partial x_{1,N}} \\ \frac{\partial f}{\partial x_{2,1}} & \frac{\partial f}{\partial x_{2,2}} & \cdots & \frac{\partial f}{\partial x_{2,N}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial x_{M,1}} & \frac{\partial f}{\partial x_{M,2}} & \cdots & \frac{\partial f}{\partial x_{M,N}} \end{bmatrix}$$

4) 행렬 곱의 대각성분

(04.04 11page)

행렬미분법칙 4: 행렬 곱의 대각성분

두 정방행렬을 곱해서 만들어진 행렬의 대각성분(trace)는 스칼라이다. 이 스칼라를 뒤의 행렬로 미분하면 앞의 행렬의 전치행렬이 나온다.

$$f(X) = \text{tr}(WX) \tag{4.4.49}$$

$$W \in R^{N \times N}, X \in R^{N \times N} \tag{4.4.50}$$

$$\frac{\partial f}{\partial X} = \frac{\partial\, \text{tr}(WX)}{\partial X} = W^T \tag{4.4.51}$$

(증명)

$$\text{tr}(WX) = \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ji} x_{ij} \tag{4.4.52}$$

$$\frac{\partial \text{tr}(WX)}{\partial x_{ij}} = w_{ji} \tag{4.4.53}$$

5) 행렬식의 로그

(04.04 11page)

행렬식(determinant)은 스칼라값이고 이 값의 로그 값도 스칼라이다. 이 값을 원래의 행렬로 미분하면 원래 행렬의 역행렬의 전치 행렬이 된다.

$$f(X) = \log |X| \tag{4.4.54}$$

$$\frac{\partial f}{\partial X} = \frac{\partial \log |X|}{\partial X} = (X^{-1})^T \tag{4.4.55}$$

[gradient, J, H]

1. gradient 다변수 스칼라 함수의 미분 (1차미분을 다변수 함수로 확장)
$(x_1 \cdots x_n)$ f

$\frac{\partial f}{\partial x_1}$ ⋮ $\frac{\partial f}{\partial x_n}$

2. J 다변수 벡터 함수의 미분 $(\nabla f_1, \nabla f_2, \cdots \nabla f_m)$
기울기의 모음 $(x_1, \cdots x_n)$ $f_1 \cdots f_m$

$$\begin{bmatrix} -\nabla f_1^T- \\ \vdots \\ -\nabla f_m^T- \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & & \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

3. H 다변수 함수의 2차 미분
곡률 확인가능 $(x_1, \cdots x_n)$ f

$$\begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & & \\ & & \ddots & \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & & & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

$$\nabla f_i = \frac{\partial f}{\partial x}$$

$$J_{ij} = \left(\frac{\partial f}{\partial x}\right)^T$$

$$H_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j} = H_{ji}$$ ($\because$ 슈바르츠)

H = 대칭행렬
(실수 고윳값)

Why $H_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$?

- $J f(x) = \left(\frac{\partial f}{\partial x_i}\right)^T$
- $H_{ij} = J(\nabla f)^T = \left(\frac{\partial}{\partial x} \nabla f\right)^T = \left(\frac{\partial}{\partial x} \frac{\partial f}{\partial x}\right)^T = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)^T = \frac{\partial^2 f}{\partial x_i \partial x_j}$