< 파이토치 인강 >

< 3강 - 딥러닝을 위한 기초 수학 > → 필요한 부분만 골라 들음.

1. 함수

$$\begin{array}{l} x \searrow \\ \quad\ f \to yx^2 = f(x, y) \\ y \nearrow \end{array}$$

2. 로그함수

3. 미분

4. 연쇄법칙 "Chain Rule" → 분수 형태로 표현한다.

$$\frac{\partial (x^2+1)^2}{\partial x} = \frac{\partial (x^2+1)^2}{\partial (x^2+1)} \cdot \frac{\partial (x^2+1)}{\partial x^2} \cdot \frac{\partial x^2}{\partial x}$$

$$= 2(x^2+1) \cdot 2x \cdot 1 \cdot 2x$$

5. Gradient

6. 확률변수

$$확률함수 \begin{cases} \text{probability mass function} \\ \text{probability density function} \end{cases}$$

7. 평균과 분산

8. 균등분포와 정규분포.

< 4강 - 알고리즘 소개 >

1. gradient descent.

→ optimization 고려 요소 { 방향 "gradient" ⇔ 미분. "가장 가파른 방향으로 간다"
얼마나 "Step Size" (= learning rate) α → α를 너무 크게, 혹은 너무 작게 잡아서는 안된다.

2. 알고리즘 순서

i) initialize x_0

ii) $x_{k+1} = x_k - \alpha \cdot \frac{\partial f}{\partial x}$

iii) repeat 2

3. 예시

$z = Ax + n$

i) $f = (z - Ax)^T (z - Ax)$ ⇝ 행렬 곱셈 맞춰주기 위해 transpose 필요.

ii) $\partial f = \frac{\partial f}{\partial x^T} \cdot \partial x$

$= (-A\partial x)^T (z - Ax)^T - (z - Ax)^T A \partial x$

$= -2(z - Ax)^T A \partial x$

4. Directional derivative

✲ 왜 gradient descent (미분)가 가장 가파른 방향을 의미할까.

→ contour의 접선에 수직한 방향.

why? 같은 값을 이은 선이므로, 접선은 그 순간 변화가 없는 방향

⇔ 방향 미분이 0이 되도록 하는 방향

⇔ 내적이 0이 되도록 하는 방향

⇔ gradient에 수직인 방향