

# 정보이론과 딥러닝 Loss 함수 이해하기

2022.04.16

최종욱

# CONTENTS

---

- 정보이론
  1. 정보량
  2. 엔트로피(entropy)
  3. Cross-entropy / Kullback-Leibler Divergence
  4. Jensen-Shannon Divergence
- Cross-entropy vs MSE
  - 관점 1: Backpropagation Algorithm
  - 관점 2: MLE

# 1. 정보이론

# 정보량 $I(X)$

- 통계학 관점의 사건의 정보량
  - 놀라움의 정도
- 정보량의 정의에 의한 성질
  1. 정보량은 확률에 반비례 한다.
  2. 연달아 발생하는 독립사건의 정보량은 각각의 사건의 합으로 표현 가능해야 한다.
    - 동전과 주사위 예시
- If base: 2 -> bit, e -> nit, 10 -> dit

# 엔트로피 entropy

- 통계학 관점의 엔트로피
  - 평균 정보량
    - 이산랜덤변수  $X$  의 샘플 공간  $\{x_1, x_2, \dots, x_n\}$  의 엔트로피?
    - $H(X) ? : E() \rightarrow \Sigma$
  - 발생확률? 이벤트 값(정보량)?
    - 주사위, 100원

# 크로스엔트로피 / Kullback-Leibler Divergence

- 크로스엔트로피 (Cross-entropy)
  - 정답과 예측이 달라서 놀라는 정도
  - 즉, 정답이 이미 존재하고 예측과 달라서 얻는 평균 정보량
  - P:정답, Q:예측,  $H(P, Q)$
- Kullback-Leibler Divergence
  - 두 분포의 **차이**를 계산하는데 사용
  - $$D_{KL}(P||Q) = E_{x \sim p}[\log \frac{P(x)}{Q(x)}]$$
  - Cross-entropy, KLD 관계 파악
  - 왜 차이?

# Jensen-Shannon Divergence

- 젠슨-샤넌 발산
  - KLD 는 차이 -> 거리 x
  - 그럼 분포의 **거리**는 어떻게 표현?

$$JSD(P \parallel Q) = \frac{1}{2} D_{KL}(P \parallel M) + \frac{1}{2} D_{KL}(Q \parallel M), \text{ where } M = \frac{1}{2}(P + Q)$$

## 2. Cross-entropy vs MSE



# Cross-entropy vs MSE

---

- TBD