Information Theory

Additional mini-project in deeplearning math

03-1. 정보 엔트로피(entropy) 개념, 표기, 연산 - 이론

소프트웨어 꼰대 강의

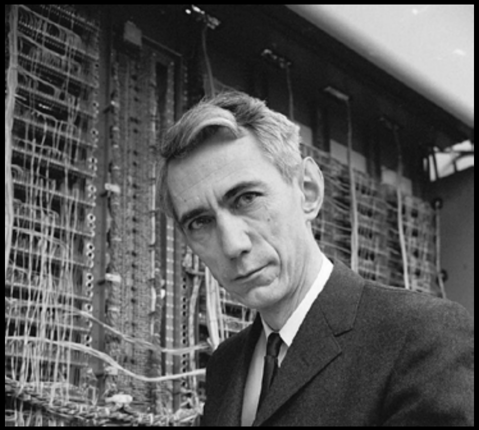
노기섭 교수

(kafa46@cju.ac.kr)

Course Overview

Topic	Contents
01. Orientation 오리엔테이션	Motivations & Course introduction 동기부여, 과정 소개
02. Information 정보	What is the information? Concept & definition 정보란 무엇인가? 개념과 정의
03. Information Entropy 정보 엔트로피	Concepts, notation, and operations on information entropy 정보 엔트로피의 개념, 표기, 연산
04. Entropy in Deeplearning 딥러닝에서의 엔트로피	How to apply the information entropy into Deeplearning? 어떻게 정보 엔트로피를 딥러닝에 적용하는가?
05. Entropy Loss 엔트로피 손실	Loss function using entropy, BCE, and cross entropy 엔트로피를 이용한 손실 함수, BCE, 크로스 엔트로피
06. KL Divergence KL 발산	Concept & definition of KL divergence KL 발산의 개념과 정의
07. Summary & Closing 요약 및 마무리	Summary & closing on this project, 'Information Theory' 정보 이론 요약 및 마무리

엔트로피의 등장



Claude Shannon (1916~2001)
새넌에 의해 제안되어
'새넌 엔트로피'라고 불리기도 함

여러 개 정보들이 있을 때,
평균 정보량은 얼마일까? ⇨ Information Entropy

평균 공식과 기댓값을 생각해 보면 쉽다~

이전 강의를 참고하세요 ^^.

[Probability]_01. Random Variables (확률변수)... 이게 뭔가요?
https://youtu.be/iTxTGBOhzCA

[Probability]_02. Expected Value (기댓값)... 이게 뭔가요?
https://youtu.be/nvHyIScyQxs

Recap: Random variable & expectation

기댓값 공식: 확률변수 값에 확률을 곱하여 모두 더한다.

$$E[X] = x_1\ p_1 + x_2\ p_2 + \cdots + x_n\ p_n = \sum_{i=1}^{n} x_i p_i$$

- X: $a\ random\ variable$
- x_i: $possible\ outcome\ of\ X$
- p_i: $probability\ of\ x_i$

가능한 점수

Ω
100
90
80
70
60

$\xrightarrow{X}$

E	확률
100	1/6
90	1/6
80	1/6
70	1/6
60	1/6

$$E[X] = \sum_{i=1}^{n} x_i \times p(x_i)$$

$$= 100 \cdot \frac{1}{6} + 90 \cdot \frac{1}{6} + 80 \cdot \frac{1}{6} + 70 \cdot \frac{1}{6} + 60 \cdot \frac{1}{6} + 50 \cdot \frac{1}{6}$$

$$= \frac{100 + 90 + 80 + 70 + 60 + 50}{6} = \frac{450}{6} = 75$$

Entropy formulation

$$H(P) = H(x) = E_{X\sim P}[I(x)] = E_{X\sim P}[-\log P(X = x)]$$

좀 더 간략하게

$$= E_x[-\log P(x)]$$

$$H(x) = \sum_{i=1}^{n} -p\log P(x_i) = -\sum_{i=1}^{n} p\log P(x_i)$$

참고: 일반적으로 정보의 표현은 비트(bit, 0과 1)로 하기 때문에

대부분 밑(base)이 2인 로그를 사용

→ 하지만 밑(base)는 어떤 것을 사용해도 문제되지 않습니다.

로그의 밑 변환 공식을 이용하면 쉽게 변환 가능

(증명은 생략)

Information Entropy 값의 범위

Probability axioms

1. $P(x) \in \mathbb{R} \ \wedge P(x) \geq 0$

2. $P(\Omega) = 1$

3. $P\left(\bigcup_{i=1}^{\infty} x_i\right) = \sum_{i=1}^{\infty} P(x_i)$

따라서,

$$0 \leq p(x) \leq 1$$

$$0 \leq I(x) \leq \infty$$

$$0 \leq H(x) \leq \infty$$

동전 던지기에서 Information Entropy

이미지 저작자: FLATICON
이미지 출처: https://www.flaticon.com/kr/free-icon/coin-toss_2471534

Binary Case…

Random variable

동전 앞면 (head) $\xrightarrow{X}$ p

동전 뒷면 (tail) $\xrightarrow{X}$ $q = 1 - p$

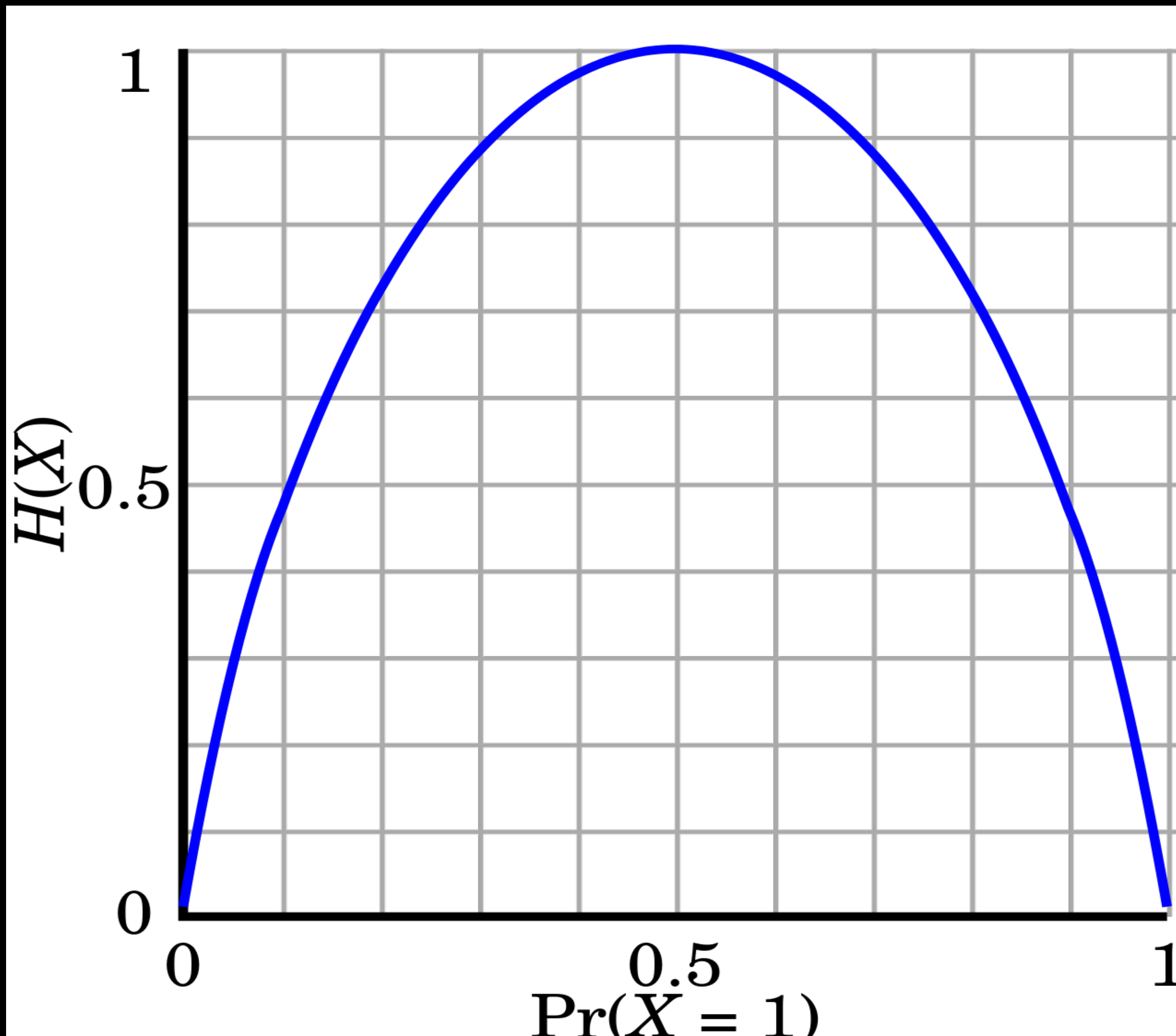
$p = q = \frac{1}{2}$ $\quad H(X) = \left(-\frac{1}{2}\log\frac{1}{2}\right) + \left(-\frac{1}{2}\log\frac{1}{2}\right) = \left(-\frac{1}{2}\log 2^{-1}\right) + \left(-\frac{1}{2}\log 2^{-1}\right) = \frac{1}{2} + \frac{1}{2} = 1$

$p = \frac{1}{4}, q = \frac{3}{4}$ $\quad H(X) = \left(-\frac{1}{4}\log\frac{1}{4}\right) + \left(-\frac{3}{4}\log\frac{3}{4}\right) = -\frac{1}{4} \times (-2) + \left(-\frac{3}{4}\right) \times (-0.415) = 0.811$

$p = 1, q = 0$ $\quad H(X) = (-1\log 1) + (-0\log 0) = 0$

교수님 ~~

모든 p에 대한 엔트로피는 어떻게 되나요?



Random variable

Head $\xrightarrow{X}$ 1

예제로 이해하는 Information Entropy

$$X = \begin{cases} a \;\; with \;\; probabiility \;\; p_1 = \dfrac{1}{2} \\ b \;\; with \;\; probabiility \;\; p_2 = \dfrac{1}{4} \\ c \;\; with \;\; probabiility \;\; p_3 = \dfrac{1}{8} \\ d \;\; with \;\; probabiility \;\; p_4 = \dfrac{1}{8} \end{cases}$$

$$H(X) = -\sum_i p_i \log p_i = \left(-\frac{1}{2}\log\frac{1}{2}\right) + \left(-\frac{1}{4}\log\frac{1}{4}\right) + \left(-\frac{1}{8}\log\frac{1}{8}\right) + \left(-\frac{1}{8}\log\frac{1}{8}\right)$$

$$= \left(-\frac{1}{2}\log 2^{-1}\right) + \left(-\frac{1}{4}\log 2^{-2}\right) + \left(-\frac{1}{8}\log 2^{-3}\right) + \left(-\frac{1}{8}\log 2^{-3}\right)$$

$$= \left(-\frac{1}{2}\times(-1)\right) + \left(-\frac{1}{4}\times(-2)\right) + \left(-\frac{1}{8}\times(-3)\right) + \left(-\frac{1}{8}\times(-3)\right)$$

$$= \frac{1}{2} + \frac{1}{2} + \frac{3}{8} + \frac{3}{8} = \frac{7}{4} \;\; bits$$

주사위를 동시에 던지는 경우?

Random variable: n 개 주사위를 동시에 던졌을 때 나온 값의 합

Entropy???

경우의 수가 많아서.... ㅠㅠ

Python으로 실습해 보겠습니다~~
(실습 동영상은 별도로 구성)

수고하셨습니다 ..^^..