엔트로피 성질

1) "엔트로피"의 최솟값: 0 ⇒ when? 특정한 하나의 값이 1이며, 나머지가 0일 때.

2) "엔트로피"의 최댓값: K ⇒ why? 랜덤변수가 취할 수 있는 클래스가 2^K개이고, 최대값은 각 class가 모두 같은 확률을 가지는 때이다.

$$H = -2^K\left(\frac{1}{2^K}\log_2\frac{1}{2^K}\right) = K$$

따라서, $H = K$.

엔트로피와 정보량 → 엔트로피는 확률변수가 갖을 수 있는 정보량을 표시한다.

"정보량"이란 – 확률변수의 표본값을 관측하여 얻을 수 있는 정보의 총량을 말한다.

ex) 알파벳 "A", "B", "C", "D"로 된 이루어진 문서

4이 문서를 이진수 0과 1로 변환할 때, 다음과 같이 encoding

- A = 00
- B = 01
- C = 10
- D = 11

⇒ 1,000글자로 이루어진 문서를 이진수로 표현하면, 2000자리.

P(A) P(B) P(C) P(D)

그런, 알파벳이 나올 확률이 동일하지 않고, $\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\}$ 라고.

⇒ 이진수의 수를 줄여보기 위해, 새로운 개념인 "가변길이 인코딩"을 활용.
(Variable Length Encoding)

- A = 0
- B = 10
- C = 110
- D = ~~110~~ 111

⇒ $(1000\times\frac{1}{2})\cdot 1 + (1000\times\frac{1}{4})2 + (1000\times\frac{1}{8})3 + (1000\times\frac{1}{8})3$

= 1,750자

⇒ $H = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{4}\log_2\frac{1}{4} - \frac{1}{8}\log_2\frac{1}{8} - \frac{1}{8}\log_2\frac{1}{8} = 1.75$

⇒ 1.75는 Alphabet 한 글자를 인코딩 하는데 필요한 평균 bit 수이며, 확률변수의 엔트로피의 값과 같다.

- 가변길이 Encoding (Variable Length Encoding): 많이 쓰인 글자는 비트 최소화하여 짧게 늘리는 것.

ex)

A	$\frac{1}{2}$	0	$\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}\}$
B	$\frac{1}{4}$	10	
C	$\frac{1}{8}$	110	$-1\times\frac{1}{2} = \frac{1}{2}$
D	$\frac{1}{16}$	1110	
E	$\frac{1}{64}$	111100	$-2\times\frac{1}{4} = \frac{1}{2}$
F	$\frac{1}{64}$	111101	
G	$\frac{1}{64}$	111110	$-4\times\frac{1}{16} = \frac{1}{4}$
H	$\frac{1}{64}$	111111	$-6\times\frac{1}{64} = \frac{1}{4}$

⇒ 2 (H)