Shannon Entropy (정보량과 불확실성의 상관관계)

요점 : 불확실성이 높을수록 많은 정보량을 얻게 된다

Shannon Entropy : 정보량을 측정하는 대표적인 방법

$$H(X) = -\sum_{i=1}^{n} p_i log_2 p_i = \sum_{i=1}^{n} (log_2 \frac{1}{p_i}) p_i$$

기대값 공식 : $E[X] = \sum_{i=1}^{k} x_i p_i$

엔트로피 공식 : $H(X) = \sum_{i=1}^{n} log_2 \frac{1}{p_i} p_i \rightarrow$ "**정보량의 기댓값**"을 의미함

$log_2 \frac{1}{p_i}$ 가 정보량을 의미하는 이유

$H(X) = \sum_{i=1}^{n} I(x_i) \times p_i$, $I(x_i)$는 x_i의 정보량을 계산하는 함수라고 가정하고 다음 두 조건을 생각해보자

1. The more uncertain an event is, the more information if carries : 불확실성이 높을수록 많은 정보량을 가지게 된다

   $$p(x_i) > p(x_2) \rightarrow I(x_1) < I(x_2)$$

   즉, x_i보다 x_2의 불확실성이 더 크기 때문에 x_2의 정보량이 더 크다.

   이 때 $I(x)$를 $\frac{1}{p(x)}$라고 정의하면 확률이 높을수록 정보량이 작아지기에 $p(x_i) > p(x_2) \rightarrow I(x_1) < I(x_2)$를 만족함

2. If x_1 and x_2 are independent events, $I(x_1, x_2) = I(x_1) + I(x_2)$

   log를 이용하게 되면 $logA \times B$를 $logA + logB$로 계산 가능함. ($logA \times B = logA + logB$)

   so, $I(x) = log(\frac{1}{p(x)})$

log의 base가 왜 2인가

이용하고자 하는 정보량을 bit의 수로 나타낼 수 있기 때문

ex) what is the max entropy for random variable has 4 differnt result?

$$0 \leq \sum_{i=1}^{n} log_2 \frac{1}{p_i} * p_i \leq 2$$

위 수식 중 2는 4개($4 = 2^2$)의 랜덤변수의 정보량을 저장하는데 필요한 bit의 갯수를 의미한다.

- 최대 엔트로피가 $log_2 n$이 되는 이유
  엔트로피가 최대값이 되기 위해선 n개의 random variable의 모든 확률이 $\frac{1}{n}$이어야 한다.
  so, 최대 엔트로피는 $\sum_{i=1}^{n} log_2 \frac{1}{\frac{1}{n}} \times \frac{1}{n} = log_2 \frac{1}{\frac{1}{n}} \times \frac{1}{n} \times n = log_2 \frac{1}{\frac{1}{n}} = log_2 n$로 유도된다.

출처 : https://www.youtube.com/watch?v=CdH7U3IjRI8&list=PLVNY1HnUlO241gILgQloWAs0xrrkqQfKe&index=51