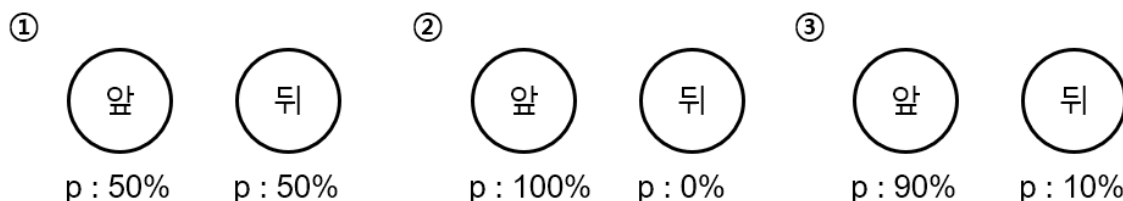


엔트로피의 이해 및 머신러닝 사용 예제 (액티브 러닝, 의사결정트리)

예제 : 동전던지기



불확실성의 측도를 나타내는 단위 → 엔트로피

(① > ③ > ②)

엔트로피 계산

- shannon entropy 식 :

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

앞	뒤	계산	entropy
50%	50%	$-(0.5 \log 0.5 + 0.5 \log 0.5)$	1
100%	0%	$-(1.0 \log 1.0 + 0.0 \log 0.0)$	0
0%	100%	$-(0.9 \log 0.9 + 0.1 \log 0.1)$	0.47

머신러닝에서의 entropy 사용예제

1. Building **Decision Tree** → 엔트로피 값이 작은 Decision Tree를 만든다. (효율성 ↑)

2. Active Learning

- 정의 (in wikipedia) : Active Learning as a special case of machine learning in which a learning algorithm is able to interactively query the user (or some other information source) to obtain the desired outputs at new data points. In statistics literature, it is sometimes also called optimal experimental design.
- 엔트로피를 이용해 불확실성이 큰 데이터만을 걸러 labeling 작업을 한 뒤 학습시킴
 - 특징 : train data 관리가 효율적이고 아주 적은 data만을 labeling하기에 자원 절약
 - ex : 해당 문장이 긍정문인지 부정문인지 분류

문장	happy label p	unhappy label p	entropy
----	---------------	-----------------	---------

문장	happy label p	unhappy label p	entropy
I love you	0.99	0.01	0.08
I am angry	0.05	0.95	0.29
I am sad	0.3	0.7	0.88
I am feeling blue	0.6	0.4	0.97

entropy가 높은 'I am sad','I am feeling blue'만을 labeling작업을 한 뒤 학습시킨다.

출처 : <https://www.youtube.com/watch?v=r3iRRQ2ViQM&list=PLVNY1HnUIO241gILgQloWAs0xrrkqQfKe&index=52>