

| *Entropy*

정보, 정보량 그리고 불확실성

- ✓ 어떤 일이 일어날 확률이 높다면, 그 일이 일어나는 것은 확실
- ✓ 어떤 일이 일어날 확률이 낮다면, 그 일이 일어나는 것은 불확실
- ✓ 확실한 정보 vs 확실하다고 말하는 정보
 - 내일 아침 태양이 동쪽에서 뜰 겁니다. ==> 당연!
 - 이건 확실한 정보인데요, 내일 아침 태양이 서쪽에서 뜰 겁니다. ==> 뭘소리야?
 - 근데 만약 확실하다고 말하는 정보가 진짜라면, 완전 대박 고급정보!
- ✓ 그래서, 정보는 확률에 반비례 합니다. $\frac{1}{p(x)}$

정보, 정보량 그리고 불확실성

✓ 그리고 정보를 량으로 수치화 하면,

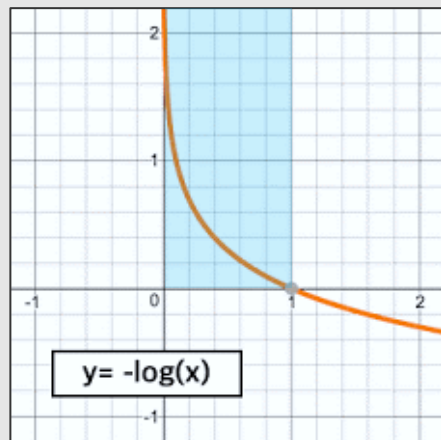
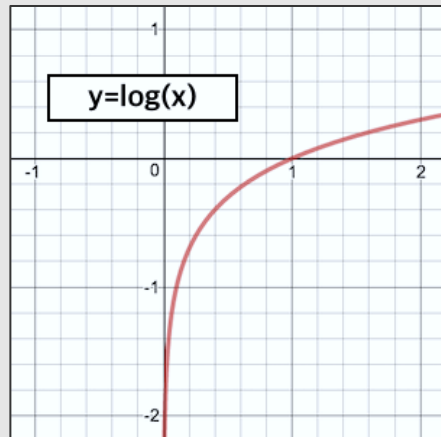
$$\text{정보량} = \log \frac{1}{p(x)} = -\log p(x)$$

✓ 확률에 Log 계산을 한다는 것은..

- 확률 $0 \sim 1 \Rightarrow \infty \sim 0$ 으로 변환 0에서 1사이의 값을
무한대~0사이의 값으로 변환
- 정보량이 크다(확률이 적다) $\rightarrow \infty$
- 정보량이 적다(확률이 크다) $\rightarrow 0$

✓ 불확실성이 큰 상황에서 주어지는 정보가 고급정보!

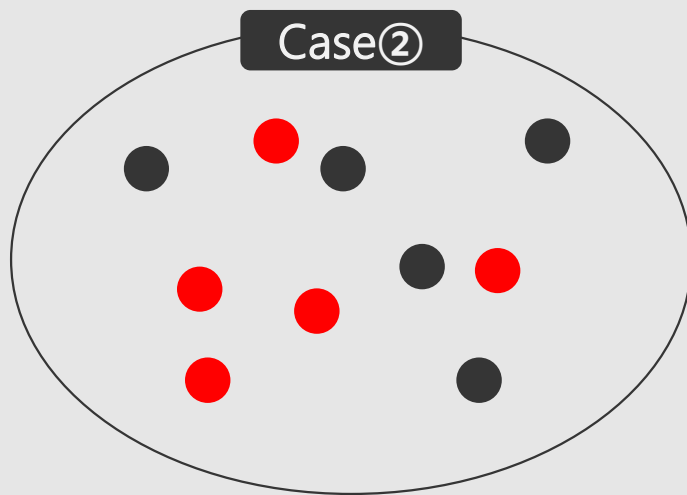
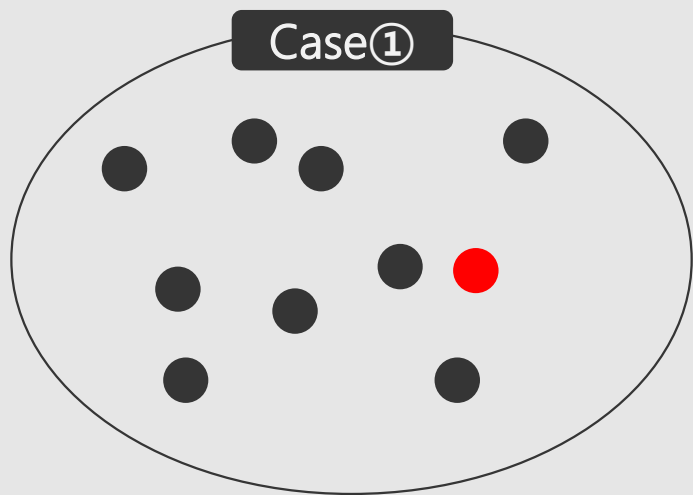
- 정보량이 크다 = 일어날 확률이 적다 = 불확실성이 크다.



평균 정보량 = 엔트로피

✓ 만약 주머니에 공이 10개고, 하나를 뽑으려 할 때,

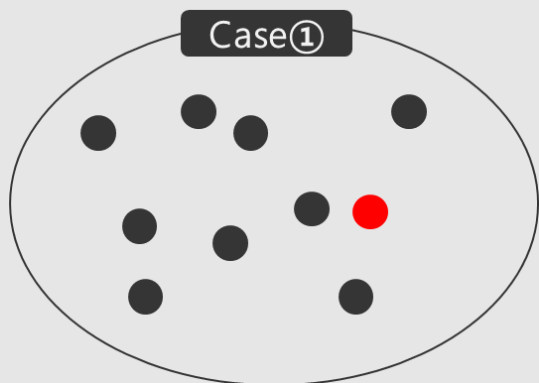
- Case① : 빨간공 1개, 검은공 9개 → 웬만하면 검은공이 뽑힘. (불확실성 낮음)
- Case② : 빨간공 5개, 검은공 5개 → 검은공 빨간공 어떤 게 뽑힐지 불확실. (불확실성이 높음)



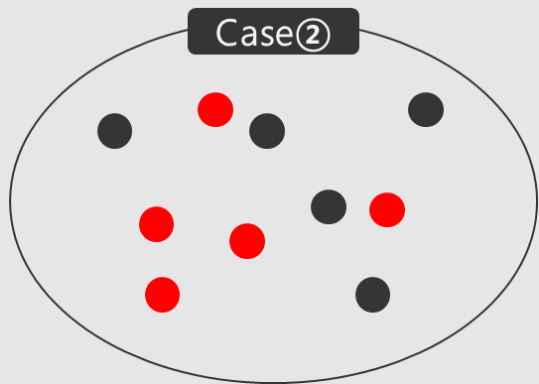
평균 정보량 = 엔트로피

✓ 얼마나 불확실 한거야?

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



- 검은 공의 비율 : $9/10 = 0.9$
- 빨간 공의 비율 : $1/10 = 0.1$
- 엔트로피 = $-(0.9 \cdot \log 0.9 + 0.1 \cdot \log 0.1) = 0.325$
가중치·검은 공 정보량 + 가중치·빨간 공 정보량



- 검은 공의 비율 : $5/10 = 0.5$
- 빨간 공의 비율 : $5/10 = 0.5$
- 엔트로피 = $-(0.5 \cdot \log 0.5 + 0.5 \cdot \log 0.5) = 0.693$

평균 정보량 = 엔트로피

✓ 얼마나 불확실 한거야?

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

✓ 이 값을 다른 말로 엔트로피 불순도(Impurity) 라고 부릅니다.

- 의사결정나무에서는
- 전체적으로 이 값을 떨어뜨려 가는 것이 모델링의 목표 입니다.
- 그래서 split 시 불순도를 가장 많이 떨어뜨려주는 변수와 값으로 기준을 결정합니다.
- 부모의 불순도에서 자식의 불순도를 뺀 것이 Information Gain(정보 이득) 입니다.

Cross Entropy

분류 모델에 대한 loss function

예측결과와 실제값의 차이가 클수록 오차(정보량)이 크다.
이를 낮추는 방향으로 학습이 되어야함.

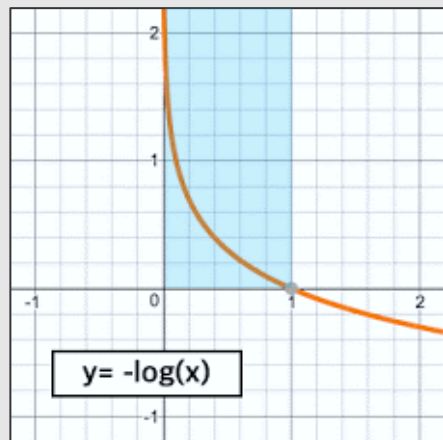
✓ 예측결과와 실제 값에 대한 오차를 어떻게 계산하면 좋을까요?

- y 가 1이라면, = 정보량
 - \hat{y} 이 1에 가까울 수록 오차가 0에 가깝고
 - \hat{y} 이 0에 가까울 수록 오차가 ∞ 에 가깝도록 만드는 방법이 있습니다.
- 그렇다면 y 가 0일 때는 어떻게 계산하는 것이 좋을까요?
 - \hat{y} 이 1에 가까울 수록 오차가 ∞ 에 가깝고
 - \hat{y} 이 0에 가까울 수록 오차가 0에 가깝도록 만들려면

$$err_1 = -\log \hat{y}_1$$

$$err_0 = -\log(1 - \hat{y}_0)$$

y	\hat{y}	$-\log \hat{y}_1$	$-\log(1 - \hat{y}_0)$
1	0.9	0.11	
0	0.3		0.36
0	0.4		0.51
1	0.7	0.36	
0	0.5		0.69
0	0.7		1.2
1	0.5	0.69	



분류 모델에 대한 loss function

- ✓ 이 오차의 식을 일반화 시키고(err_1 과 err_0 을 하나의 식으로 합치고)
- ✓ 평균을 계산한 오차식 : Log Loss, 혹은 Cross Entropy 입니다.

$$-\frac{1}{n} \sum (y \cdot \log \hat{y} + (1 - y) \cdot \log(1 - \hat{y}))$$

$$err_1 = -\log \hat{y}_1$$

$$err_0 = -\log(1 - \hat{y}_0)$$

y	\hat{y}	$-\log \hat{y}_1$	$-\log(1 - \hat{y}_0)$
1	0.9	0.11	
0	0.3		0.36
0	0.4		0.51
1	0.7	0.36	
0	0.5		0.69
0	0.7		1.2
1	0.5	0.69	

