

## • 엔트로피의 정의

$Y=0$  또는  $Y=1$  인 두 값은 각각 확률분포가 다음과 같이 세 종류가 있다고 가정한다

• 확률분포  $Y_1 = P(Y=0)=0.5, P(Y=1)=0.5$

• 확률분포  $Y_2 = P(Y=0)=0.8, P(Y=1)=0.2$

• 확률분포  $Y_3 = P(Y=0)=1.0, P(Y=1)=0.0$

이 확률분포가 베이지안 확률이라면 확률분포  $Y_1$ 은 불확실성이 모두 같은 상태,  $Y_3$ 은

불확실성이 100% 확정된 상태,  $Y_2$ 은 불확실성이 0이 아니라고 생각한 확률분포를 갖는 상태를 나타내고 있을 것이다.

이처럼 확률분포들이 가지는 확률의 정도를 수치로 표현하는 엔트로피(entropy)이다.

확률변수의 여러가지 값이 나온 확률이 대략을 비슷한 경우에는 엔트로피가 높아진다.

반대로, 특정값이 나올 확률이 높아지고 나머지 값의 확률이 낮아지면 엔트로피가 낮아진다.

- 엔트로피의 측정

## • 엔트로피의 측정량

엔트로피는 확률변수가 가질 수 있는 정보의 양을 의미한다고 볼 수도 있다.

엔트로피가 ↓ → 얻을 수 있는 정보가 줄어든다.

엔트로피가 ↑ → 얻을 수 있는 정보가 늘어난다.

## • 지니불균등 (gini impurity)

엔트로피와 유사한 개념. 엔트로피처럼 확률분포가 어떻게 구성되어 있는지를 측정하는 것으로 불균등이 지니는 엔트로피 대용으로 쓰인다.