# <Decision Tree - Choosing a split : Information Gain>

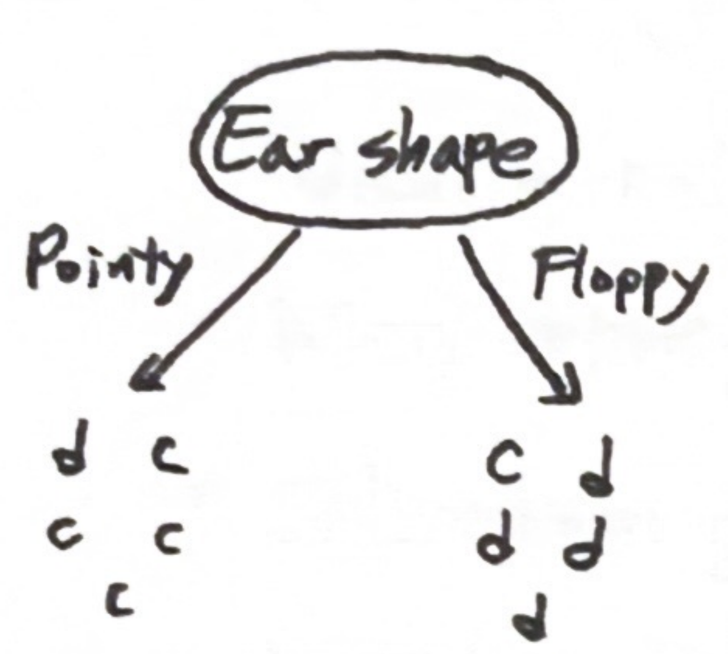
- the way we'll decide what feature to split on at a node will be based on what choice of feature reduces entropy the most

  in decision tree learning
  the reduction of entropy = "information gain"

* Choosing Split ex) 5 cats, 5 dogs (c=cat, d=dog) * root node ⇒ $P_1 = 5/10 = 0.5$, $H(0.5) = 1$

10개 example을 아직 split 하기전

①




$P_1 = 4/5 = 0.8$ $P_1 = 1/5 = 0.2$

$H(0.8) = 0.72$ $H(0.2) = 0.72$

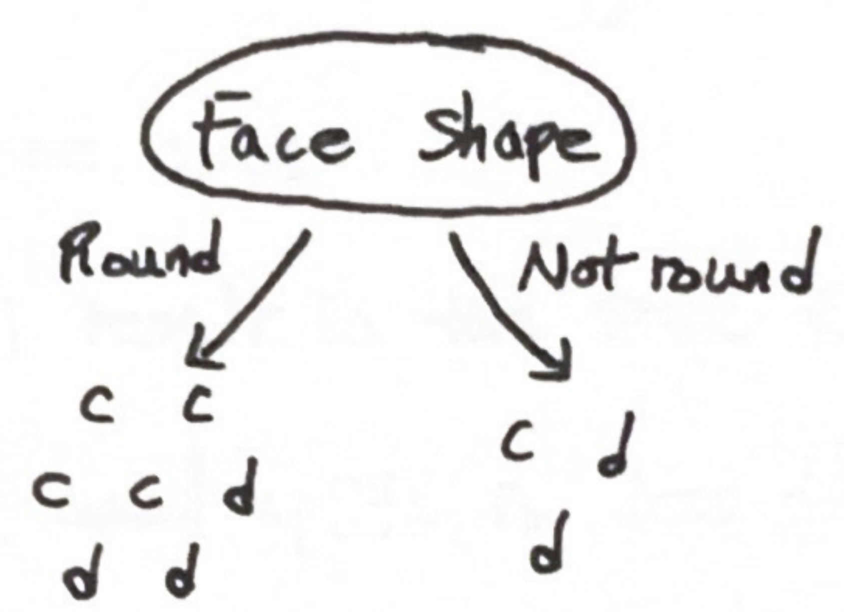
⇓ weighted average

$H(0.5) - \left(\frac{5}{10} H(0.8) + \frac{5}{10} H(0.2)\right)$

root node entropy / left branch weight / right branch weight

$= 0.28$

②




$P_1 = 4/7 = 0.57$ $P_1 = 1/3 = 0.33$

$H(0.57) = 0.99$ $H(0.33) = 0.92$

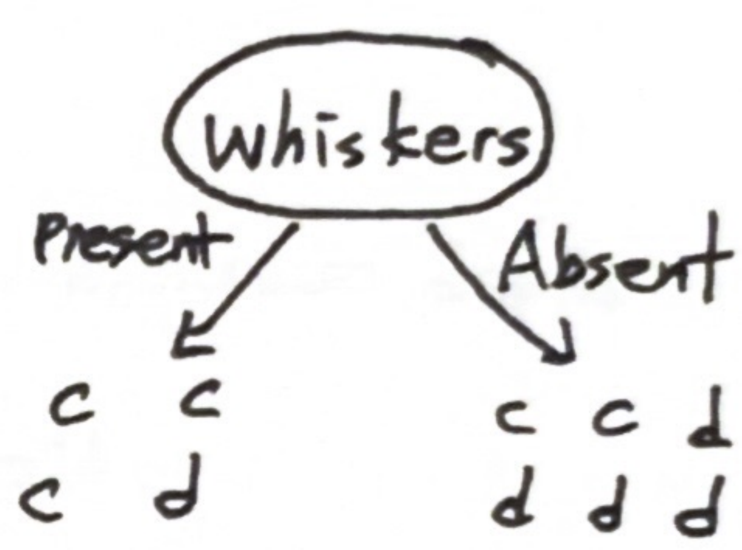
⇓ weighted average

$H(0.5) - \left(\frac{7}{10} H(0.57) + \frac{3}{10} H(0.33)\right)$

root node entropy / left branch weight / right branch weight

$= 0.03$

③




$P_1 = 3/4 = 0.75$ $P_1 = 2/6 = 0.33$

$H(0.75) = 0.81$ $H(0.33) = 0.92$

⇓ weighted average

$H(0.5) - \left(\frac{4}{10} H(0.75) + \frac{6}{10} H(0.33)\right)$

root node entropy / left branch weight / right branch weight

$= 0.12$

"information gain" $= H(p_1^{root}) - \left(w^{left} H(p_1^{left}) + w^{right} H(p_1^{right})\right)$

↳ information gain : measures the reduction in entropy that you get in your tree resulting from making a split

(original root entropy - node entropy = reduction in entropy)

node entropy: value of impurity ∴ 작은 값일수록 pure ⇒ reduction in entropy: information gain ∴ 큰 값일수록 좋음

* 그냥 entropy가 아닌 reduction in entropy 사용이유

⇒ entropy의 감소 (reduction in entropy)가 너무 작은 경우 더 이상 split 하지 않는 기준으로 삼기 위해서

⇒ reduction in entropy (information gain)을 사용함으로써 불필요하게 tree를 split하여 크기를 키워 overfitting 해버리는 risk를 줄일 수 있음 (지정한 threshold 보다 information gain이 낮으면 무시해버리기)