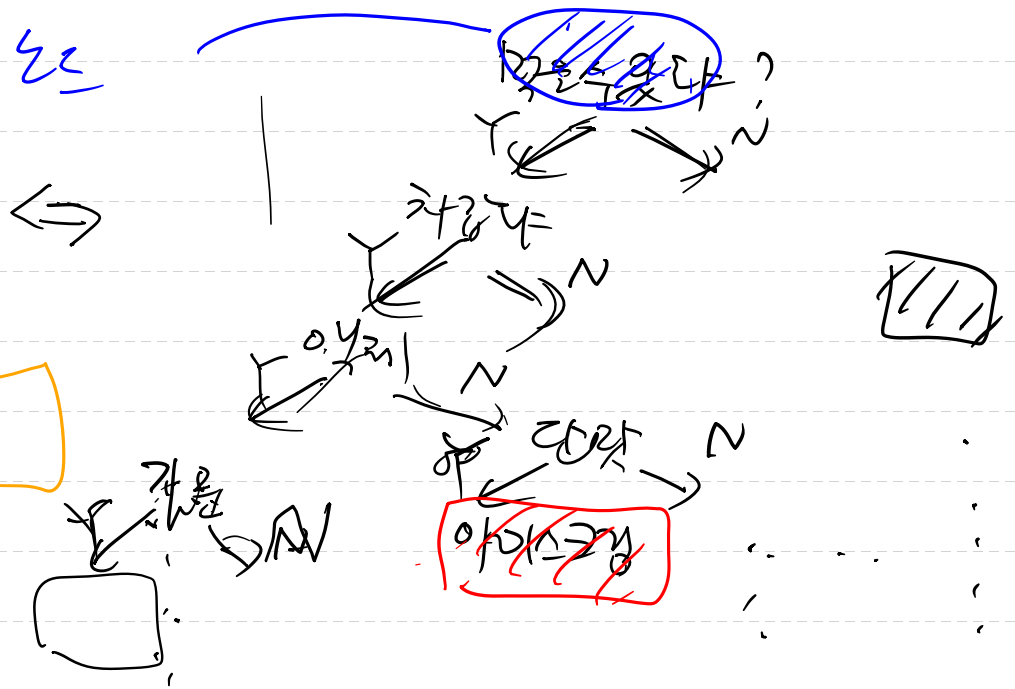
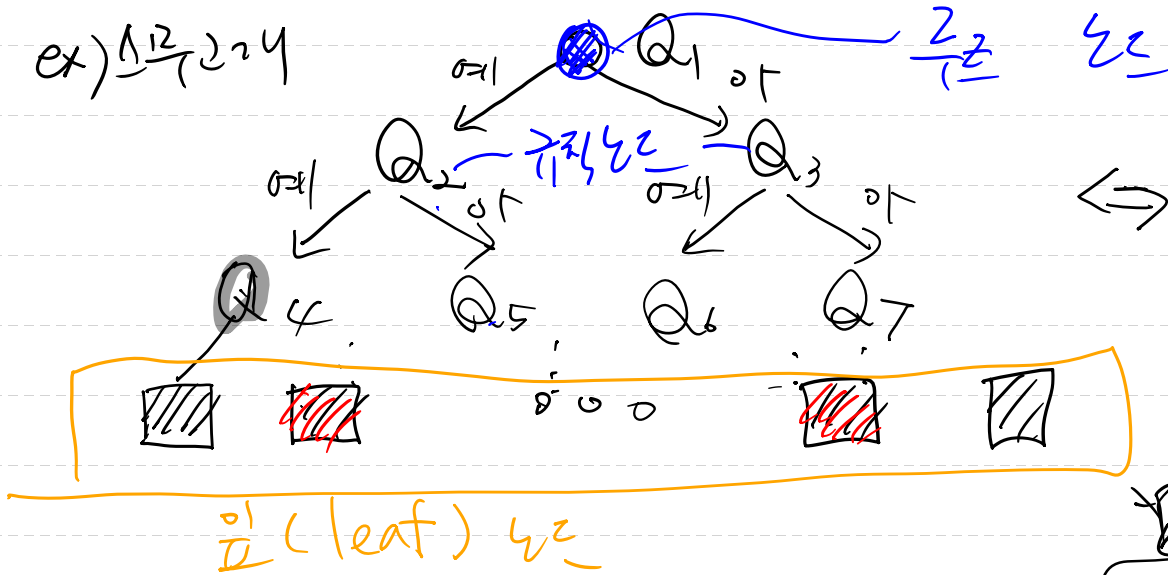


- 의사결정나무 (decision tree) : 최치, 분지.

- 기이러 패턴을 찾아낸 다음, 그 기반의 처리를 생각.
- 직각삼각, 정육각(?) , 사다리꼴 주의! (drumming)

[illegible]

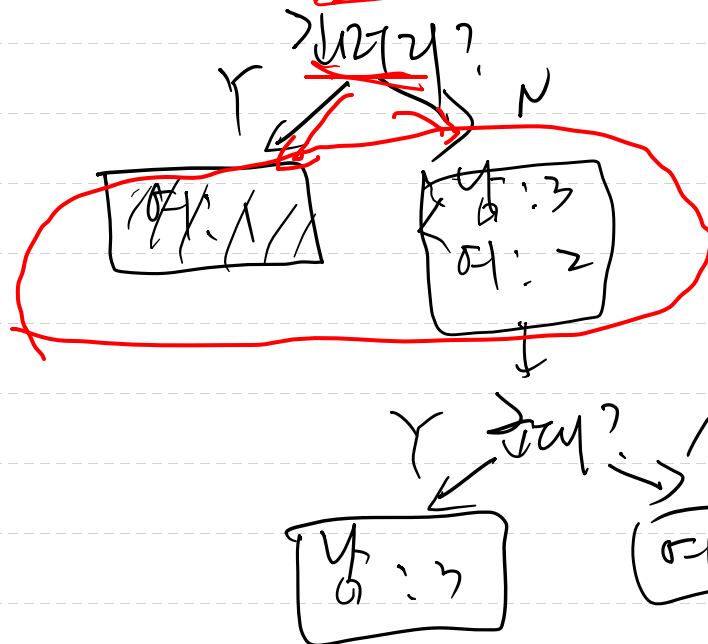
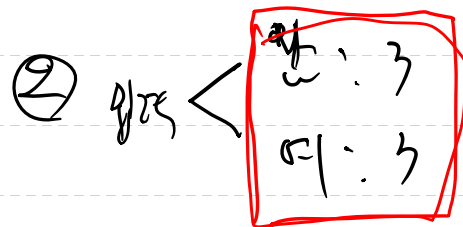
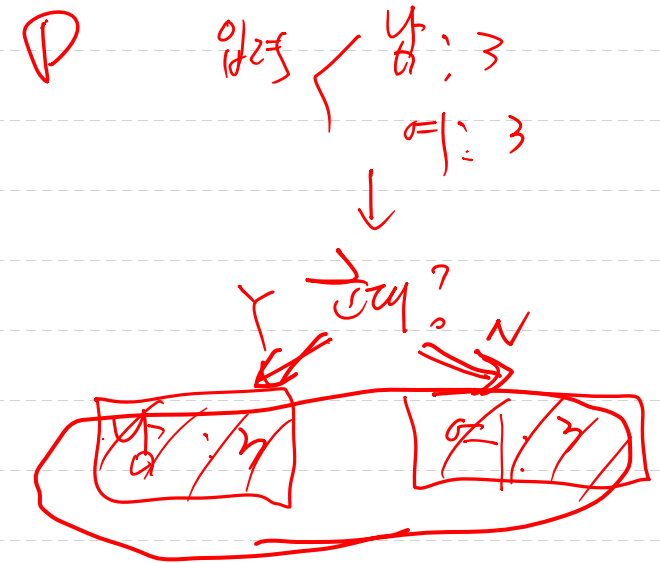
ex)  $\frac{12}{7} \approx 1,71$



중요 의사결정 트리 ?   
 1. 질문이 질문이냐 : 양의 양으로 양하는 양을  
 2. 양의 양을 양하는 양을

\* ID 3 알려줌 : 죽은 자의 자녀 가 아 는 아 니 다 .

이름	특징	특징 (feature)	성격
xx	기	아니오	나
xx	아니오	아니오	아
xx	기	아니오	나
xx	아니오	기	아
xx	기	아니오	나
xx	아니오	아니오	아



2y tree  $\frac{2}{6}$

① free 分

질문? 고이치의 특징 (수업방식, 기러기)

★ 영향력,  $\frac{1}{n}$  (관 분포) 특징은 항상  
상위  $\frac{1}{n}$  (정제) 작은 특징은 하위  $\frac{1}{n}$ 로

57 7/17

ID3 알고리즘 : 정렬된 feature를 선택하는 알고리즘.

- 엔트로피 (Entropy) : 불확실성.

Information	<div style="border: 1px solid blue; padding: 5px; display: inline-block;">             정렬 전의 불확실성  <math>\left( \begin{smallmatrix} \text{남:3, 여:3} \end{smallmatrix} \right)</math> </div>	—	<div style="border: 1px solid blue; padding: 5px; display: inline-block;">             정렬 후의 불확실성  <math>\left( \begin{smallmatrix} \text{여:1} \\ \text{남:3, 여:2} \end{smallmatrix} \right)</math> </div>
Gain (IG)			
정렬을 하는 이유			Purity

IG < 고라? 많이 불확실성을 줄 수 있는 (순수)

$$\text{Entropy} = \sum_{i=1}^m -P_i * \log_2 P_i$$

\* 엔트로피 (Entropy) =  $\sum_{i=1}^n -P_i \log_2 P_i$

(복잡도)

$\sum_{i=1}^n$  → 클래스 개수

\* 정보 획득량 (I.G) =  $\text{정보 전 복잡도} - \text{정보 후 복잡도}$

이름	특성	나이 (feature)	성별
xx	남	아름	남
xx	아름	아름	여
xx	남	아름	남
xx	아름	남	여
xx	남	아름	남
xx	아름	아름	남

\* 정보 전 복잡도

$$= -(P_{\text{남}} \log_2 P_{\text{남}}) + (-P_{\text{여}} \log_2 P_{\text{여}})$$

$$= -\left(\frac{3}{6} \times \log_2 \left(\frac{3}{6}\right)\right) + \left(-\frac{3}{6} \times \log_2 \left(\frac{3}{6}\right)\right)$$

$$= -\left(\frac{1}{2} \times \log_2 \frac{2}{1}\right) + \left(-\frac{1}{2} \times \log_2 \frac{2}{1}\right)$$

$$= \frac{1}{2} + \frac{1}{2} = 1$$

- 특징에 대한 엔트로피 계산.  
(자바스크립트)

$$\Rightarrow \text{Entropy} = \sum_{c \in X} P(c) E(c)$$

$X$ : 특징 (길이, 주파수...),  $c$ : 선택된 특징에 의해 생성된 하위 노드

$P(c)$ : 선택된 특징에 의해 생성된 하위 노드의 기/노드 수를 확률

$E(c)$ : " " " " "의 엔트로피

이름	특징 (feature)	길이	생성
x1	비	아니오	노
x1	아니오	아니오	아니
x2	비	아니오	노
x3	아니오	비	아니
x4	비	아니오	아니
x5	아니오	아니오	아

예) 순서 특징으로 분리했을 때 엔트로피

$$\frac{3}{6} * E[3, 0] + \frac{3}{6} * E[0, 3]$$

$$= \frac{3}{6} * \left( -\left(\frac{3}{3}\right) * \log\left(\frac{3}{3}\right) - \left(\frac{0}{3}\right) * \log\left(\frac{0}{3}\right) \right) + \frac{3}{6} * \left( -\frac{0}{3} * \log\left(\frac{0}{3}\right) - \frac{3}{3} * \log\left(\frac{3}{3}\right) \right)$$

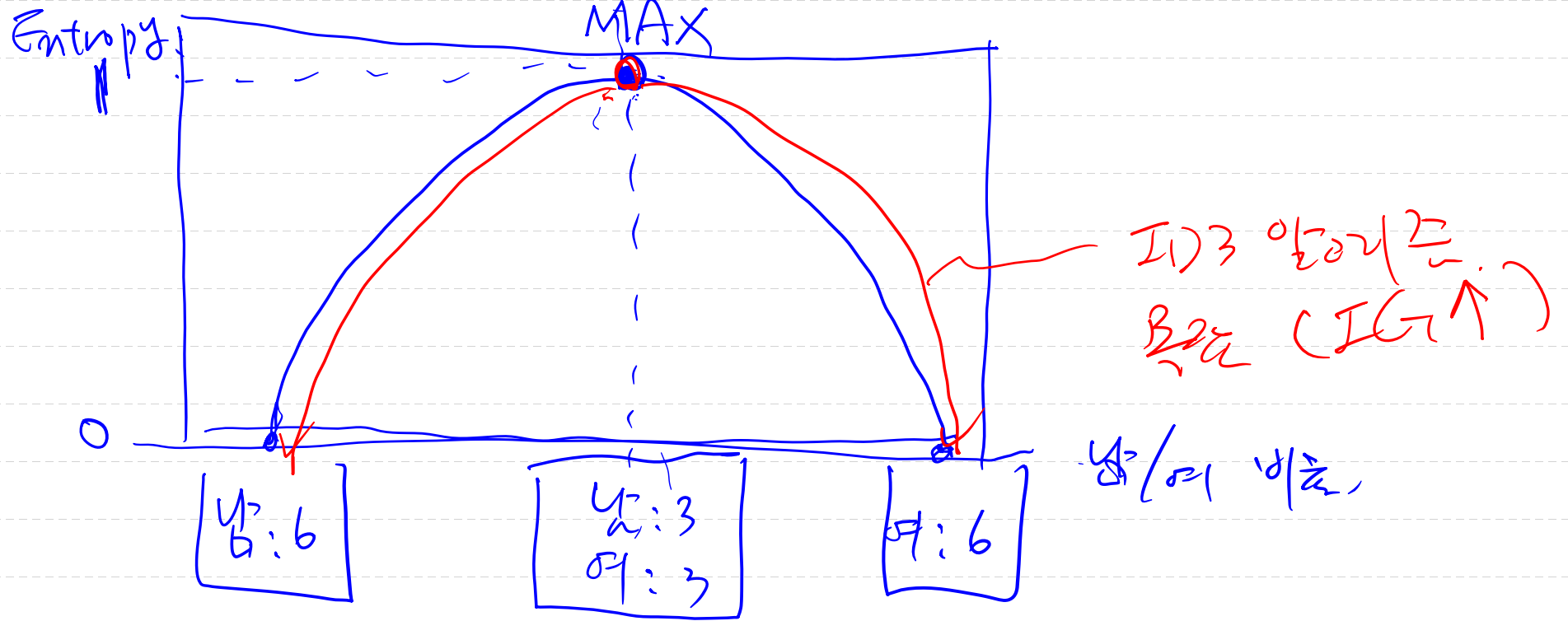
예) 길이로

$$\frac{1}{6} * E[0, 1] + \frac{5}{6} * E[3, 2]$$

$\log 1 = 0$

$$= \frac{1}{6} * \left( -\left(\frac{0}{1}\right) * \log\left(\frac{0}{1}\right) - \left(\frac{1}{1}\right) * \log\left(\frac{1}{1}\right) \right) + \frac{5}{6} * \left( -\left(\frac{3}{5}\right) * \log\left(\frac{3}{5}\right) - \left(\frac{2}{5}\right) * \log\left(\frac{2}{5}\right) \right)$$

결과



IG 3 - 알파는 부족

IG - 알파는 부족 + 알파는 부족

# 점의 계수 (gini)

α) 점의 계수 < 대역들  
대역들이 양호함.

ex) 점의 계수 / 양호  
아양.

$$① \text{ 점의 } \underline{\text{대역들}} = \left( \frac{0}{3} \right)^2 + \left( \frac{3}{3} \right)^2 = \underline{1}$$

← 양호  
← 양호  
← 점의 대역들 양호함

$$② \text{ 점의 } \text{대역들이 양호함} = \left( \frac{3}{3} \right)^2 + \left( \frac{0}{3} \right)^2 = \underline{1}$$

← 대역들이 양호함  
← 대역들이 양호함

$$★ ③ \text{ 점의 계수의 리나계수} = \frac{3}{6} * 1 + \frac{3}{6} * 1 = \underline{1}$$

← 점의 계수  
← 대역들이 양호함  
← 대역들이 양호함  
← 대역들이 양호함

$$① \text{ 리나계수} = \left( \frac{0}{1} \right)^2 + \left( \frac{1}{1} \right)^2 = 1$$

$$② \text{ 리나계수} = \left( \frac{3}{5} \right)^2 + \left( \frac{2}{5} \right)^2 = \frac{13}{25}$$

③ 리나계수

$$= \frac{1}{6} * 1 + \frac{5}{6} * \frac{13}{25} = \frac{1}{6} + \frac{13}{30} = \frac{18}{30} = \frac{3}{5} = \underline{0.6}$$

날짜	날씨	온도	습도	바람	참가여부
D1	맑음	더움	높음	약함	X
D2	맑음	더움	높음	강함	X
D3	흐림	더움	높음	약함	O
D4	비	포근	높음	약함	O
D5	비	서늘	정상	약함	O
D6	비	서늘	정상	강함	X
D7	흐림	서늘	정상	강함	O
D8	맑음	포근	높음	약함	X
D9	맑음	서늘	정상	약함	O
D10	비	포근	정상	약함	O
D11	맑음	포근	정상	강함	O
D12	흐림	포근	높음	강함	O
D13	흐림	더움	정상	약함	O
D14	비	포근	높음	강함	X



날짜	날씨	온도	습도	바람	참가여부
D1	맑음	더움	높음	약함	X
D2	맑음	더움	높음	강함	X
D3	흐림	더움	높음	약함	O
D4	비	포근	높음	약함	O
D5	비	서늘	정상	약함	O
D6	비	서늘	정상	강함	X
D7	흐림	서늘	정상	강함	O
D8	맑음	포근	높음	약함	X
D9	맑음	서늘	정상	약함	O
D10	비	포근	정상	약함	O
D11	맑음	포근	정상	강함	O
D12	흐림	포근	높음	강함	O
D13	흐림	더움	정상	약함	O
D14	비	포근	높음	강함	X











